



**HAL**  
open science

## Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation

Y. Tang, P. Reed, Thibaut Wagener, K. van Werkhoven

► **To cite this version:**

Y. Tang, P. Reed, Thibaut Wagener, K. van Werkhoven. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences Discussions*, 2007, 11 (2), pp.793-817. hal-00305053

**HAL Id: hal-00305053**

**<https://hal.science/hal-00305053>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation

Y. Tang, P. Reed, T. Wagener, and K. van Werkhoven

Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA

Received: 29 September 2006 – Published in Hydrol. Earth Syst. Sci. Discuss.: 1 November 2006

Revised: 11 January 2007 – Accepted: 30 January 2007 – Published: 5 February 2007

**Abstract.** This study seeks to identify sensitivity tools that will advance our understanding of lumped hydrologic models for the purposes of model improvement, calibration efficiency and improved measurement schemes. Four sensitivity analysis methods were tested: (1) local analysis using parameter estimation software (PEST), (2) regional sensitivity analysis (RSA), (3) analysis of variance (ANOVA), and (4) Sobol's method. The methods' relative efficiencies and effectiveness have been analyzed and compared. These four sensitivity methods were applied to the lumped Sacramento soil moisture accounting model (SAC-SMA) coupled with SNOW-17. Results from this study characterize model sensitivities for two medium sized watersheds within the Juniata River Basin in Pennsylvania, USA. Comparative results for the 4 sensitivity methods are presented for a 3-year time series with 1 h, 6 h, and 24 h time intervals. The results of this study show that model parameter sensitivities are heavily impacted by the choice of analysis method as well as the model time interval. Differences between the two adjacent watersheds also suggest strong influences of local physical characteristics on the sensitivity methods' results. This study also contributes a comprehensive assessment of the repeatability, robustness, efficiency, and ease-of-implementation of the four sensitivity methods. Overall ANOVA and Sobol's method were shown to be superior to RSA and PEST. Relative to one another, ANOVA has reduced computational requirements and Sobol's method yielded more robust sensitivity rankings.

## 1 Introduction

In this paper we apply and evaluate the differences between four popular sensitivity analysis methods, selected to represent the variety of methods currently used. The four sensitivity analysis methods include: (1) local analysis using the parameter estimation software (PEST), (2) regional sensitivity analysis (RSA), (3) analysis of variance (ANOVA), and (4) Sobol's method. The methods are applied to the Sacramento soil moisture accounting model, a medium complexity spatially lumped rainfall-runoff model used for river forecasting throughout the USA. The model is implemented in two watersheds in the Susquehanna River Basin in Pennsylvania and run at hourly, six hourly, and daily time steps.

Broadly, models of watershed hydrology are irreplaceable components of water management studies including flood and drought prediction, water resource assessment, climate and land use change impacts, or non-point source pollution analysis (e.g., Singh and Woolhiser, 2002). Hydrologic models are evolving from single purpose tools to complex decision support systems that can perform all (or at least many) of the tasks mentioned above in a single software package. Hydrological models vary in complexity from lumped conceptual models to distributed models that include close coupling of surface and groundwater flow processes, feedbacks with the atmosphere, transport of water and solutes, and spatially explicit representations of system characteristics and states (e.g., Duffy, 1996, 2004; Koren et al., 2004; Panday and Huyakorn, 2004). In integrated assessment applications models may even include socioeconomic components to integrate human behavior (Wagener et al., 2005). In general, hydrologic models are highly non-linear, contain thresholds, and often have significant parameter interactions. These properties make it difficult to evaluate how models of hydrologic systems behave and which parameters control this behavior during different response modes (e.g., Demaria et al.,

---

Correspondence to: P. Reed  
(preed@engr.psu.edu)

2007). The increasing trend towards more complex models and its potential consequences in terms of computational constraints and obfuscating model impacts on decision making motivates the need for enhanced model identification and evaluation tools (Beven and Freer, 2001; Vrugt et al., 2003; Saltelli et al., 2004; Wagener and Kollat, 2007).

Hydrologic models play an important role in elucidating the dominant controls on watershed behavior and in this context it is important for hydrologists to identify the dominant parameters controlling model behavior. One approach to gain this understanding is through the use of sensitivity analysis, which evaluates the parameter's impacts on the model response (Hornberger and Spear, 1981; Freer et al., 1996; Wagener et al., 2001; Liang and Guo, 2003; Hall et al., 2005; Pappenberger et al., 2005; Sieber and Uhlenbrook, 2005). Sensitivity analysis results can be used to decide which parameters should be the focus of model calibration efforts, or even as an analysis tool to test if the model behaves according to underlying assumptions (e.g., Wagener et al., 2003). Ultimately, sensitivity methods should serve as diagnostic tools that help to improve mathematical models and potentially help us to identify where gaps in our knowledge are most severe and are most strongly affecting prediction uncertainty. Data gaps are particularly important in the context of guiding field measurement campaigns (Langbein, 1979; Moss, 1979; Wagener and Kollat, 2007; Reed et al., 2006). Section 2 provides a more detailed review of existing sensitivity analysis methods and a detailed discussion of the four methods compared in this study.

## 2 Sensitivity analysis tools and sampling schemes

### 2.1 Overview

Model sensitivity analysis characterizes the impact that changes in model inputs have on the model outputs in a strict sense. Model inputs include model parameters, forcing, initial conditions, boundary conditions, etc. In this study, we focus on analyzing the sensitivities of model parameters. Sensitivity measures are determined mathematically, statistically, or even graphically. There are several prior studies that have broadly reviewed and classified the sensitivity analysis methods that exist (Saltelli et al., 2000, 2004; Helton and Davis, 2003; Oakley and O'Hagan, 2004; Frey and Patil, 2002; Christiaens and Feyen, 2002). Any sensitivity analysis approach can be broken up into two components (Wagener and Kollat, 2007): (1) a strategy for sampling the model parameter space (and/or state variable space), and (2) a numerical or visual measure to quantify the impacts of sampled parameters on the model output of interest. The implementation of these two components varies immensely (e.g., Freer et al., 1996; Frey and Patil, 2002; Hamby, 1994; Patil and Frey, 2004; Pappenberger et al., 2006; Vandenberghe et al., 2007), and guidance is currently lacking to help modelers

decide which approach is best suited to the needs of a particular study. Generally, the approaches can be categorized into two main groups – local methods and global methods (Saltelli et al., 1999; Muleta and Nicklow, 2005).

The nominal range and differential analysis methods are two well known local parameter sensitivity analysis methods (Frey and Patil, 2002; Helton and Davis, 2003). Nominal range sensitivity analysis calculates the percentage change of outputs due to the change of model inputs relative to their baseline (nominal) values. The percentage change is seen as the sensitivity of the corresponding input. Differential analysis utilizes partial derivatives of the model outputs with respect to the perturbations of the model input. The derivative values are themselves the metrics of sensitivity. Further analysis can be conducted by approximating the simulation model using Taylor's series (Helton and Davis, 2003).

The nominal range and differential analysis methods have the advantages of being straightforward to implement while maintaining modest computational demands. The major drawback of these methods is their inability to account for parameter interactions, making them prone to underestimating true model sensitivities. Alternatively, global parameter sensitivity analysis methods vary all of a model's parameters in predefined regions to quantify their importance and potentially the importance of parameter interactions.

There are a variety of global sensitivity analysis methods such as regional sensitivity analysis (RSA) (Young, 1978; Hornberger and Spear, 1981), variance based methods (Saltelli et al., 2000), regression based approaches (Spear et al., 1994; Helton and Davis, 2002), and Bayesian sensitivity analysis (Oakley and O'Hagan, 2004). Global methods attempt to explore the full parameter space within predefined feasible parameter ranges. In this paper, our goal is to test a suite of sensitivity methods and discuss their relative benefits and limitations for advancing lumped watershed model identification and evaluation.

The four sensitivity analysis approaches include one local method termed PEST and three global methods consisting of RSA, analysis of variance (ANOVA), and Sobol's method. These sensitivity analysis methods were selected for comparison due to their popularity and their common application in a variety of scientific domains (Doherty, 2003; Doherty and Johnston, 2003; Moore and Doherty, 2005; Wagener et al., 2003; Lence and Takyi, 1992; Freer et al., 1996; Pappenberger et al., 2005; Mokhtari and Frey, 2005; Sobol', 1993, 2001; Fieberg and Jenkins, 2005; Hall et al., 2005). The sensitivity analysis methods tested in this study range from local to global and capture a broad range of analysis methodologies (differential analysis, RSA, and variance-based analysis). The main characteristics of these four methods are summarized in Table 1. In Sect. 2.2, each of these approaches and the associated statistical sampling schemes used in this study are discussed in more detail. In the context of this paper we assume that the selection of an appropriate numerical measure, is satisfied through two chosen objec-

**Table 1.** Summary of sensitivity analysis tools in the study.

Tools	Local/Global	Interactions	Sampling	Quantify/Visualize
PEST	Local	No	Local Perturbation	Composite Sensitivity
RSA	Global	No	Latin Hypercube Sampling	CDF Plots
ANOVA	Global	Yes	Iterated Fractional Factorial Design	F-Values
Sobol's	Global	Yes	Quasirandom	Sensitivity Indices

tive functions based on the root mean square error (RMSE) (see Sect. 5.2). Readers interested in how parameter sensitivity changes with different objective functions can reference the following studies (Wagener et al., 2001; Demaria et al., 2007).

## 2.2 Sensitivity analysis tools

### 2.2.1 PEST

PEST, which stands for parameter estimation, is a model independent nonlinear parameter estimation tool (Doherty, 2003; Doherty and Johnston, 2003; Doherty, 2004; Moore and Doherty, 2005). PEST was developed to facilitate data interpretation, model calibration and predictive analysis. Like many other parameter estimation or model calibration tools, PEST aims to match the model simulation with an observed set of data by minimizing the weighted sum of squared differences between the two. The optimization problem is iteratively solved by linearizing the relationship between a model's output and its parameters. The linearization is conducted using a Taylor series expansion where the partial derivatives of each model output with respect to every parameter must be calculated at every iteration. For each iteration, the solution of the linearized problem is the current optimal set of parameters. The current optimal set is then compared to that of the previous time step to determine when to terminate the optimization process. During the linearization step, the forward difference or central difference operators can be used for calculating the derivatives. Parameter ranges, initial parameter values, and parameter increments must be provided by the user. The parameter vector is updated at each step using the Gauss-Marquardt-Levenberg algorithm (Marquardt, 1963; Levenberg, 1944). The derivatives of the model outputs with respect to its parameters are calculated and provide a measure of the parameter sensitivities at each iteration. The "composite sensitivity" is provided by PEST as a byproduct of the parameter estimation results. The composite sensitivity of parameter  $i$  is defined as:

$$s_i = (\mathbf{J}^t \mathbf{Q} \mathbf{J})_{ii}^{1/2} / m \quad (1)$$

where  $\mathbf{J}$  is the Jacobean matrix and  $\mathbf{Q}$  is the cofactor matrix which in most cases is a diagonal matrix whose elements are composed of squared weights for model outputs. If the

model outputs are equally weighted,  $\mathbf{Q}$  is equal to the identity matrix. The number of outputs,  $m$ , is the number of data records in the time series in this study. Thus  $s_i$  is the normalized magnitude of the Jacobean matrix column with respect to parameter  $i$ . As expected for a local sensitivity analysis method, Eq. (1) is a univariate analysis of parameter impacts on model outputs (i.e., no parameter interactions are considered).

### 2.2.2 Regional sensitivity analysis using Latin hypercube sampling

RSA (Young, 1978; Hornberger and Spear, 1981) is also called generalized sensitivity analysis (GSA) (Freer et al., 1996) and has been widely used in hydrology (e.g., Lence and Takyi, 1992; Spear et al., 1994; Freer et al., 1996; Pappenberger et al., 2005; Sieber and Uhlenbrook, 2005; Ratto et al., 2006). Monte Carlo sampling and "behavioral/nonbehavioral" partitioning are the two major components of this method. Monte Carlo sampling is used to generate  $n$  parameter sets in the feasible parameter space defined using a multi-variate uniform distribution. After model evaluations using these parameters, the sets of parameters are decomposed into two separate groups (behavioral/good and nonbehavioral/bad) according to the model's performance or behavior. RSA identifies the difference between the underlying distributions of the behavioral and nonbehavioral groups. Either graphical methods (e.g., marginal cumulative distribution function plots) or statistical methods such as Kolmogorov-Smirnov (KS) testing (Kottegoda and Rosso, 1997) are then used to characterize if a parameter significantly impacts behavioral results.

Freer et al. (1996) extended the original RSA by breaking the behavioral parameter sets into ten equally sized groups. (Wagener et al., 2001) modified this approach further by including all parameter sets and avoiding the need to specify behavioral and non-behavioral sets. Instead, the population is divided into ten bins of equal size based on a sorted model performance measure (Wagener and Kollat, 2007). Conclusions about parameter sensitivities are made qualitatively by examining differences in the marginal cumulative distributions of a parameter within each of the ten groups. Ten lines in the RSA plot represent the cumulative distributions of a parameter with respect to ten sampled sub-ranges. If the lines are clustered, the parameter is not sensitive to a spe-

cific model performance measure. Conversely, the degree of dispersion of the lines is a visual measure of a model's sensitivity to an input parameter. Wagener and Kollat (2007) implemented the original idea of Freer et al. (1996) visually using the Monte Carlo analysis toolbox (MCAT) (Wagener et al., 2001, 2003, 2004) where the marginal cumulative distributions of the ten groups are plotted as the likelihood value versus the parameter values.

In this study, Latin hypercube sampling (LHS) was used to sample the feasible parameter space for testing RSA based on the recommendations and findings of prior studies (e.g., Osidele and Beck, 2001; Sieber and Uhlenbrook, 2005). LHS integrates random sampling and stratified sampling (Mckay et al., 1979; Helton and Davis, 2003) to make sure that all portions of the parameter space are considered. The method divides the parameters' ranges into  $n$  disjoint intervals with equal probability  $1/n$  from which one value is sampled randomly in each interval. LHS is generally recommended for sparse sampling of the parameter space and the parameter interactions are neglected as noted by William et al. (1999). More details about LHS are available in the following papers (Mckay et al., 1979; Helton and Davis, 2003; William et al., 1999).

### 2.2.3 Analysis of variance using iterated fractional factorial design sampling

Assuming model response (e.g., RMSE of streamflow in this study) is normally distributed, the role of ANOVA is to quantify the differences of the mean model responses that result from samples of each parameter. In ANOVA, parameters are "grouped" into particular ranges of parameter values representing intervals with equal parameter value width, contrasting to RSA in which parameter sets are "grouped" based on model response measures such as the RMSE of streamflow predictions used in this study. According to ANOVA terminology, a parameter is called a "factor" and a parameter group is termed a "level" of the factor. ANOVA essentially partitions the model output or response into the overall mean, main factor effects, factor interactions, and an error term (Neter et al., 1996; Mokhtari and Frey, 2005). Theoretically, ANOVA can capture a range from the first order (main effects from single parameters) to the total order of effects (i.e., all parameter impacts including all interactions). However, it is not feasible to calculate all of the effects for a complex model in practice due to computational limitations. Fortunately, prior studies have shown that second order interactions are usually sufficient for capturing a model's output variance (Box et al., 1978; Henderson-Sellers et al., 1993; Liang and Guo, 2003). Therefore, our analysis focuses on first order and second order effects within the ANOVA model. The model response variable  $Y$  is decomposed into main and second order effects of two factors according to

2-way ANOVA model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + \varepsilon_{ijk} \quad (2)$$

where  $i$  and  $j$  indicate the levels of factors  $A$  and  $B$  respectively,  $\alpha_i$  is the main effect of  $i$ th level of  $A$ ,  $\beta_j$  is the main effect of  $j$ th level of  $B$ ,  $(\alpha \times \beta)_{ij}$  represents the interaction of  $A$  and  $B$ . The error term,  $\varepsilon_{ijk}$ , reflects the effects that are not explained by the main effects and interactions of the two factors. Variable  $k$  represents the  $k$ th value of  $Y$ .

The  $F$ -test is used to evaluate the statistical significance of differences in the mean responses among the levels of each parameter or parameter interaction. The  $F$  values are calculated for all parameters and parameter interactions. The higher the  $F$  values are, the more significant the differences are and therefore the more sensitive the parameter or parameter interaction is. Detailed presentation of the ANOVA calculation table for main effects and second order effects can be found in other studies (Neter et al., 1996; Mokhtari and Frey, 2005). In addition to the  $F$ -test, the coefficient of determination ( $r^2$ ) quantifies how the ANOVA model shown in Eq. (2) captures the total variation of model responses with the inclusion of the second order parameter interactions. In cases where parameter interactions are important the coefficient of determination should improve (or increase) with the addition of the interaction term  $(\alpha \times \beta)_{ij}$  from Eq. (2).

When applying the ANOVA method the statistical sampling scheme used to quantify the model response is a key determinant of the method's computational feasibility and accuracy. If one parameter or parameter interaction is analyzed at a time in succession, the total number of model runs will be excessively large and most hydrologic applications would be computationally intractable. In this study, the iterated fractional factorial design (IFFD) sampling scheme (Andres and Wayne, 1993; Andres, 1997; Saltelli et al., 1995) was used to limit the computational burden posed by ANOVA while seeking high quality results.

IFFD works well when first and second order parameter effects dominate (Andres, 1997). Using IFFD in ANOVA allows users to neglect higher order interactions not included in the model (Liang and Guo, 2003; Andres, 1997) while generating highly repeatable results (Saltelli et al., 1995). Consequently, the number of model runs required can be reduced substantially. IFFD as implemented in this study samples the parameters at three different levels: low, middle, and high. The parameter levels are defined as equally spaced intervals within the predefined parameter ranges. Using a small number of factor levels enables the sampling scheme to attain statistically significant results efficiently and accurately (Mokhtari and Frey, 2005; Andres, 1997). IFFD extends the basic orthogonal fractional factorial design (FFD) by conducting multiple iterations. The basic operations in IFFD include orthogonalization, folding, replication and random sampling (Andres and Wayne, 1993; Andres, 1997; Saltelli et al., 1995). The orthogonalized design guarantees equal

frequency for two parameter combinations but also differentiates the main effects from two-way interactions. A detailed presentation of IFFD is beyond the scope of this paper. Readers interested in detailed descriptions of IFFD are referred to the following papers (Andres and Wayne, 1993; Andres, 1997; Saltelli et al., 1995).

#### 2.2.4 Sobol’s method using quasi-random sequence sampling

In Sobol’s method (Sobol’, 1993), the variance of the model output is decomposed into components that result from individual parameters as well as parameter interactions. Conventionally, the direct model output is replaced by a model performance measure such as RMSE as used in this study. The sensitivity of each parameter or parameter interaction is then assessed based on its contribution (measured as a percentage) to the total variance computed using a distribution of model responses. Assuming the parameters are independent, the Sobol’s variance decomposition is:

$$D(\mathbf{y}) = \sum_i D_i + \sum_{i < j} D_{ij} + \sum_{i < j < k} D_{ijk} + D_{12\dots m} \quad (3)$$

where  $D_i$  is the measure of the sensitivity to model output  $\mathbf{y}$  due to the  $i$ th component of the input parameter vector denoted as  $\Theta$ ,  $D_{ij}$  is the portion of output variance that results due to the interaction of parameter  $\theta_i$  and  $\theta_j$ . The variable  $m$  defines the total number of parameters. The variance decomposition shown in Eq. (3) can be used to define the sensitivity indices of different orders as:

$$\text{first order } S_i = \frac{D_i}{D} \quad (4)$$

$$\text{second order } S_{ij} = \frac{D_{ij}}{D} \quad (5)$$

$$\text{total } S_{Ti} = 1 - \frac{D_{\sim i}}{D} \quad (6)$$

where  $S_i$  denotes the sensitivity that results from the main effect of parameter  $\theta_i$ . The second order sensitivity index,  $S_{ij}$ , defines the sensitivity that results from the interaction of parameters  $\theta_i$  and  $\theta_j$ . The average variance,  $D_{\sim i}$ , results from all of the parameters except for  $\theta_i$ . The total order sensitivity,  $S_{Ti}$ , represents the main effect of  $\theta_i$  as well as its interactions up to  $m^{\text{th}}$  order of analysis. A parameter which has a small first order index but large total sensitivity index primarily impacts the model output through parameter interactions.

The variances in Eq. (3) can be evaluated using approximate Monte Carlo numerical integrations. The Monte Carlo approximations for  $D$ ,  $D_i$ ,  $D_{ij}$ , and  $D_{\sim i}$  are defined as presented in the following prior studies (Sobol’, 1993, 2001; Hall et al., 2005):

$$\widehat{f}_0 = \frac{1}{n} \sum_{s=1}^n f(\Theta_s) \quad (7)$$

$$\widehat{D} = \frac{1}{n} \sum_{s=1}^n f^2(\Theta_s) - \widehat{f}_0^2 \quad (8)$$

$$\widehat{D}_i = \frac{1}{n} \sum_{s=1}^n f(\Theta_s^{(a)}) f(\Theta_{(\sim i)s}^{(b)}, \Theta_{is}^{(a)}) - \widehat{f}_0^2 \quad (9)$$

$$\widehat{D}_{ij}^c = \frac{1}{n} \sum_{s=1}^n f(\Theta_s^{(a)}) f(\Theta_{(\sim i, \sim j)s}^{(b)}, \Theta_{(i,j)s}^{(a)}) - \widehat{f}_0^2 \quad (10)$$

$$\widehat{D}_{ij} = \widehat{D}_{ij}^c - \widehat{D}_i - \widehat{D}_j \quad (11)$$

$$\widehat{D}_{\sim i} = \frac{1}{n} \sum_{s=1}^n f(\Theta_s^{(a)}) f(\Theta_{(\sim i)s}^{(a)}, \Theta_{is}^{(b)}) - \widehat{f}_0^2 \quad (12)$$

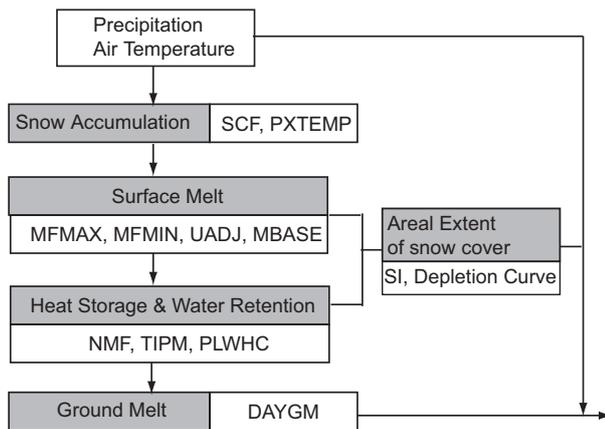
where  $f$  is the model response, term  $n$  is the sample size,  $\Theta_s$  denotes the sampled individual in the scaled unit hypercube, and  $(a)$  and  $(b)$  are two different samples. All of the parameters take their values from sample  $(a)$  are represented by  $\Theta_s^{(a)}$ . The variables  $\Theta_{is}^{(a)}$  and  $\Theta_{is}^{(b)}$  denote that parameter  $\theta_i$  uses the sampled values in sample  $(a)$  and  $(b)$ , respectively. The symbols  $\Theta_{(\sim i)s}^{(b)}$  and  $\Theta_{(\sim i)s}^{(b)}$  represent cases when all of the parameters except for  $\theta_i$  use the sampled values in sample  $(a)$  and  $(b)$ , respectively. The symbol  $\Theta_{(i,j)s}^{(a)}$  represents parameters  $\theta_i$  and  $\theta_j$  with sampled values in sample  $(a)$ . Finally,  $\Theta_{(\sim i, \sim j)s}^{(a)}$  represents the case when all of the parameters except for  $\theta_i$  and  $\theta_j$  utilize sampled values from sample  $(b)$ .

The original Sobol’s method required  $n \times (2m + 1)$  model runs to calculate all the first order and the total order sensitivity indices. An enhancement of the method made by Saltelli (2002) provides the first, second and total order sensitivity indices using  $n \times (2m + 2)$  model runs. In this study, we implemented this modified version of Sobol’s methodology to compute the first order, second order and total order indices.

The convergence of the Monte Carlo integrations used in Sobol’s method is heavily affected by the sampling scheme selected. The error term in the Monte Carlo integration decreases as a function of  $1/\sqrt{n}$  given uniform, random samples at  $n$  points in the  $m$ -dimensional space. However, in this study we elected to use Sobol’s quasi-random sequence (Sobol’, 1967; Sobol, 1994) to increase the convergence rate to nearly  $1/n$ . The quasi-random sequence samples points more uniformly along the Cartesian grids than uncorrelated random sampling. Details about Sobol’s quasi-random sequence can be found in the following studies (Sobol’, 1967; Sobol, 1994; Bratley and Fox, 1988; William et al., 1999).

### 3 Overview of the lumped hydrologic models

The SNOW-17 (Anderson, 1973) and the Sacramento soil moisture accounting (SAC-SMA) models (Burnash, 1995) are popular and the United States National Weather Service (US NWS) uses them for river forecasting (Moreda et al., 2006; Koren et al., 2004; Smith et al., 2004; Reed et al., 2004). In this study, lumped versions of these models have



**Fig. 1.** Major SNOW-17 processes and their corresponding parameters. MBASE-Base temperature for snowmelt computations during nonrain periods (degc). NMF-Maximum negative melt factor (mm/degc/6hr). TIPM-Antecedent temperature index parameter. PLWHC-Percent (decimal) liquid-water holding capacity. DAYGM – Constant amount of melt which occurs at the snow-soil interface whenever snow is present (mm). The full description of other parameter names can be found in Sect. 3.1 and in Table 2. Shaded boxes represent the states or processes.

been coupled where SAC-SMA uses SNOW-17's outputs as forcing. Sections 3.1 and 3.2 provide brief overviews of both models.

### 3.1 SNOW-17

SNOW-17 (Anderson, 1973) is a conceptual model that simulates the energy balance of a snowpack using a temperature index method. Air temperature and precipitation are the model inputs. The states and processes include snow melt, snow cover accumulation, surface energy exchange during non-melt periods, snow cover heat storage, areal extent of snow cover, retention and transmission of liquid water, and heat exchange at the snow-soil interface. Snow melt, snow cover accumulation, and areal extent are the three most influential components in the model.

Snow melt is calculated separately for rain-on-snow periods and non-rain periods. The snow melt during rain-on-snow periods is computed based on energy and mass balance equations with average wind function (UADJ) as the only parameter. In contrast, snow melt during non-rain periods is calculated empirically. The maximum melt factor (MFMAX) and the minimum melt factor (MFMIN) control this calculation. When calculating the accumulation of snow cover, the form of precipitation is simply determined by a threshold temperature (PXTEMP). The snowfall correction factor (SCF) adjusts gage precipitation estimates for biases during snowfall. To determine the areal extent of snow cover, a pre-defined depletion curve relates the areal extent to areal water equivalent based on the historical maximum wa-

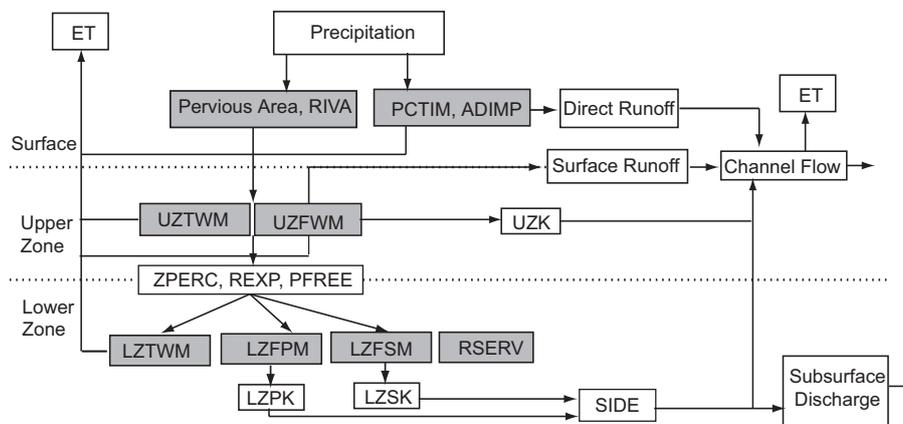
ter equivalent and the water equivalent above which 100% of snow cover exists. Process calculations are described in more detail in Anderson (1973). The main processes and corresponding twelve model parameters in SNOW-17 are shown in Fig. 1. Based on the prior work of Anderson (2002), we have focused our sensitivity analysis on five of SNOW-17's parameters (excluding the areal depletion curve). These five parameters and their allowable ranges (Anderson, 2002) are summarized in Table 2.

### 3.2 Sacramento soil moisture accounting model

The SAC-SMA model (Burnash, 1995) is a sixteen parameter lumped conceptual watershed model used for operational river forecasting by the US NWS. It represents the soil column by an upper and lower zone of multiple storages. The upper zone is divided into free water and tension water storages. The tension water spills into the free water storage only when the tension water storage (UZTWM) is filled. The free water in the upper zone can then move laterally as interflow or move vertically down to the lower zone as percolation. Capacities of the two storages are model parameters (UZFWM and UZTWM), while the volume of water in each at any time step are model states. Similar to the upper zone, the lower zone also has tension water and free water storages. The free water in the lower zone is further partitioned into two types: primary and supplemental free water storages, both of which can contribute to base-flow but drain independently at different speeds following Darcy's law. The maximum storages for these different types of lower zone free water are the lower zone maximum tension water (LZTWM), the primary free water (LZFPM), and the supplemental free water (LZFSM). SAC-SMA's processes and parameters are illustrated in more detail in Fig. 2. It is indicated in the figure that there are four principal forms of runoff generated by SAC-SMA: 1) direct runoff on the impervious area, 2) surface runoff when the upper zone free water storage is filled and the precipitation intensity is greater than percolation and interflow rate, 3) the lateral interflow from upper zone free water storage, and 4) primary baseflow. The direct runoff is composed of the impervious runoff over the permanent impervious area and the direct runoff on the temporal impervious area. The permanent impervious area, represented by parameter PC-TIM (percent of impervious area), represents constant impervious areas such as pavements. The temporal impervious area, represented by parameter ADIMP (additional impervious area), includes the filling of small reservoirs, marshes, and temporal seepage outflow areas which become impervious when the upper zone tension water is filled. Prior work (Peck, 1976) has shown that thirteen out of sixteen parameters control model performance and must be calibrated. Feasible ranges of these thirteen parameters are presented by Boyle et al. (2000) and also used in the calibration studies of Tang et al. (2006, 2007) and Vrugt et al. (2003) (see Table 2). As shown in Table 2, the maximum allowable value of

**Table 2.** Summary of SNOW-17 and SAC-SMA parameters.

Model	Parameters	Unit	Description	Allowable Range
SNOW-17	SCF		Gage catch deficiency adjustment factor	1.0–1.3
	MFMAX	mm/°C/6 h	Maximum melt factor during non-rain periods	0.5–1.2
	MFMIN	mm/°C/6 h	Minimum melt factor during non-rain periods	0.1–0.6
	UADJ	mm/mb/6 h	Average wind function during rain-on-snow periods	0.02–0.2
	SI	mm	Mean water-equivalent above which 100% cover exists	10–120
SAC-SMA	UZTWM	mm	Upper zone tension water maximum storage	1.0–150.0
	UZFWM	mm	Upper zone free water maximum storage	1.0–150.0
	UZK	day <sup>-1</sup>	Upper zone free water lateral depletion rate	0.1–0.5
	PCTIM		Impervious fraction of the watershed area	0.0–0.1
	ADIMP		Additional impervious area	0.0–0.4
	ZPERC		Maximum percolation rate	1.0–250.0
	REXP		Exponent of the percolation equation	0.0–5.0
	LZTWM	mm	Lower zone tension water maximum storage	1.0–500.0
	LZFSM	mm	Lower zone free water supplemental maximum storage	1.0–1000.0
	LZFPM	mm	Lower zone free water primary maximum storage	1.0–1000.0
	LZSK	day <sup>-1</sup>	Lower zone supplemental free water depletion rate	0.01–0.25
	LZPK	day <sup>-1</sup>	Lower zone primary free water depletion rate	0.0001–0.025
	PFREE		Fraction of water percolating from upper zone directly to lower zone free water storage	0.0–0.6



**Fig. 2.** Major SAC-SMA processes and their corresponding parameters. RIVA-Riparian vegetation area. SIDE-Ratio of deep recharge to channel base flow. RSERV-Fraction lower zone free water not transferable to tension water. The full description of other parameter names can be found in the Table 2. The parameters in the shaded boxes pertain to storages or states.

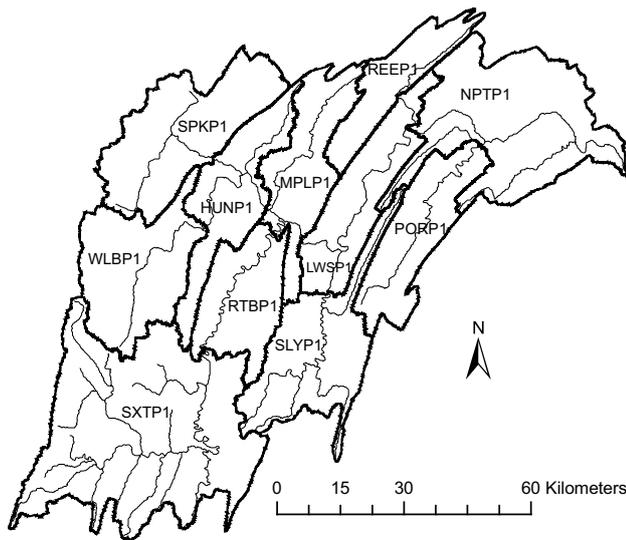
ADIMP specified by the author is 0.4 indicating that 40% of the watershed area is the additional impervious area, which can lead to large direct runoff under wet conditions.

## 4 Case study

### 4.1 Juniata watershed description

The Juniata Watershed, part of the Susquehanna River Basin, covers an area of 8800 km<sup>2</sup> in the ridge and valley region of the Appalachian Mountains of south central Pennsylvania. The watershed is within the US NWS mid-Atlantic

river forecast center (MARFC) area of forecast responsibility. The primary aquifer formations are composed of sedimentary and carbonate rocks that are presented in alternating layers of sandstone, shale, and limestone. Approximately, 67 percent of the watershed is forested, 23 percent is agricultural, 7 percent is developed area, and the rest is mine lands, water, or miscellaneous. There are 11 major sub-watersheds (see Fig. 3), among which, RTBP1, LWSP1, MPLP1, and NPTP1 have heavily controlled flows from reservoirs. Our preliminary analysis of the watershed focused on 7 headwater sub-watersheds where flows are not managed. Figure 4 illustrates our preliminary analysis of the hydrologic condi-



**Fig. 3.** Sub-watersheds in the Juniata river basin.

tions within the seven sub-watersheds by plotting their flow duration curves as well as monthly averages for streamflow, potential evaporation, and precipitation. Figure 4 shows that the SPKP1 (Spruce Creek) and SXT1 (Saxton) watersheds have distinctly different flow regimes. In the remainder of our study, we have evaluated the model sensitivities within these two watersheds using the four sensitivity analysis tools introduced in Sect. 2. As will be discussed in more detail in Sect. 5, our analysis evaluates model sensitivities for different temporal and spatial scales (i.e., SPKP1 and SXT1 have drainage areas of 570 and 1960 km<sup>2</sup>, respectively).

#### 4.2 Data set

The SAC-SMA/SNOW-17 lumped model used required input forcing data consisting of precipitation, potential evapotranspiration (PE), and air temperature. The precipitation data are next generation radar (NEXRAD) multisensor precipitation estimator data from the US NWS. Hourly data for precipitation and air temperature were available from 1 January 2001 to 31 December 2003. The observed streamflow in the same period was obtained from United States Geological Survey (USGS) gauge stations located at the outlets of the SPKP1 and SXT1 watersheds.

### 5 Computational experiment

#### 5.1 Model setup and parameterizations

In this study, we used a Linux computing cluster with 133 computer nodes composed of dual or quad AMD Opteron processors and 64 GB of RAM. Two month warmup periods (1 January to 28 February 2001) were used to reduce the influence of initial conditions. Model performance was evalu-

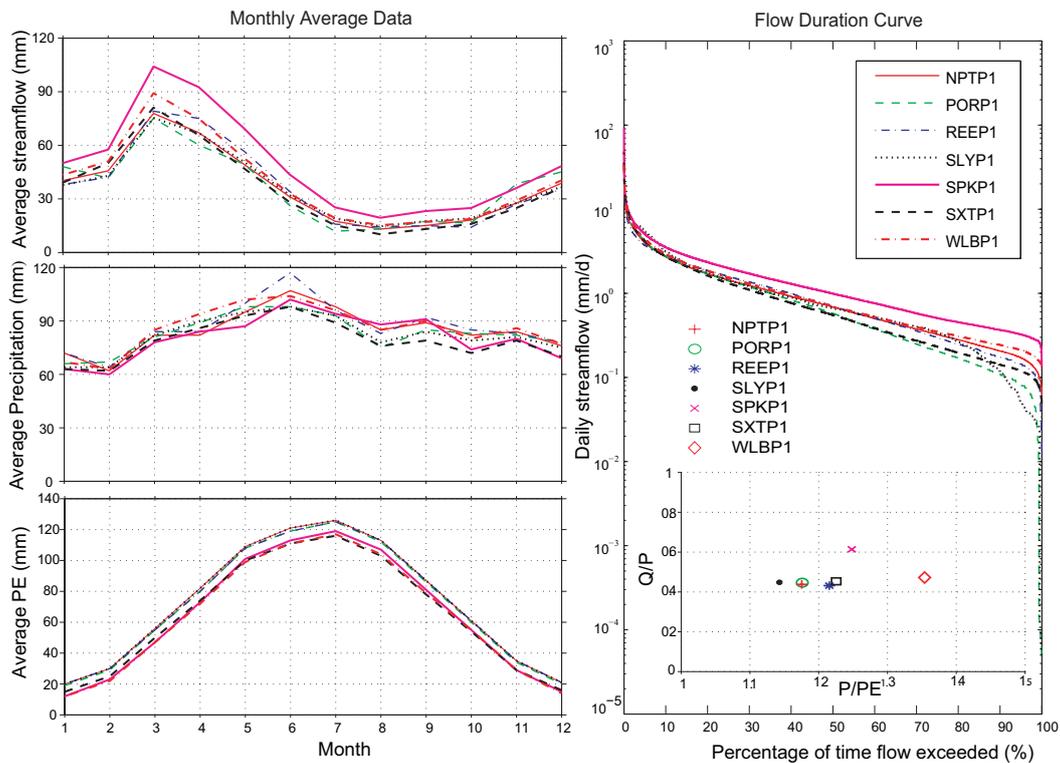
ated using three different time intervals (1 h, 6 h, 24 h) to test how parameter sensitivities change due to different prediction time scales. The a priori parameter settings used for the SNOW-17 and SAC-SMA models were based on the recommendations of the Mid-Atlantic River Forecasting Center of the US NWS.

The primary algorithmic parameters for PEST were set based on the recommendations of Doherty (2004). The initial Marquardt lambda and its adjust factor were set to be 5 and 2 respectively. When calculating the derivatives, the parameters were incremented by a fraction of the current parameters' values subject to the absolute increment lower bounds. The fraction is 0.01 and the lower bounds vary from parameter to parameter based on their magnitudes. The parameter estimation process terminates if one of the following conditions is satisfied: 1) the number of iterations exceeds 30; 2) the relative difference between the objective value of the current iteration and the minimum objective value achieved to date is less than 0.01 for 3 successive iterations; 3) the algorithm fails to lower the objective value over 3 successive iterations; 4) the magnitude of the maximum relative parameter change between optimization iterations is less than 0.01 over 3 successive iterations.

Statistical sample sizes are key parameters for RSA, ANOVA, and Sobol's method. In this study, the sample sizes were configured based on both literature recommendations and experiments by observing the convergence and reproducibility of the sensitivity analysis results. Sieber and Uhlenbrook (2005) used a sample size of 10 times the number of perturbed parameters while doing sensitivity analysis on a distributed catchment model using LHS. However, the experimental analysis showed this is far from enough for our study. Examining statistical convergence as a function of increasing sample size, we determined a size of 10 000 was sufficient for LHS in RSA. For the ANOVA method, typically the *F* values increase for the sensitive parameters with increases in sample size (Mokhtari and Frey, 2005). Our analysis of convergence for the ANOVA method's *F*-values and parameter sensitivity rankings showed that a sample size of 1,000 was sufficient when using IFFD sampling. For Sobol's quasi-random sequence Sobol' (1967) states that additional uniformity can be obtained if the sample size is increased according to the function  $n=2^k$ , where *k* is an integer. Building on this recommendation, our analysis showed that Sobol's sensitivity indices converged and were reproducible using a sample size 8,192 (2<sup>13</sup>).

#### 5.2 Objective functions

Two different model performance objective functions were used to screen the sensitivity of SAC-SMA and SNOW17 for high streamflow and low streamflow. The first objective was the non-transformed root mean square error (RMSE) objective, which is largely dominated by peak flow prediction errors due to the use of squared residuals. The second ob-



**Fig. 4.** Hydrologic conditions of headwater sub-watersheds in the Juniata River basin.

jective was formulated using a Box-Cox transformation of the hydrograph ( $z=[(y+1)^\lambda-1]/\lambda$  where  $\lambda=0.3$ ) as recommended by Misirli et al. (2003) to reduce the impacts of heteroscedasticity in the RMSE calculations (also increasing the influence of low flow periods).

### 5.3 Bootstrap confidence intervals

For ANOVA and Sobol’s method, the  $F$  values and sensitivity indices can have a high degree of uncertainty due to random number generation effects (Archer et al., 1997; Fieberg and Jenkins, 2005). In this study, we used the bootstrap method (Efron and Tibshirani, 1993) to provide confidence intervals for the parameter sensitivity rankings for both ANOVA and Sobol’s method. Essentially, the samples generated by IFFD or Sobol’s sequence were resampled  $N$  times when calculating the  $F$  values or sensitivity indices for each parameter, resulting in a distribution of the  $F$  values or indices. The moment method (Archer et al., 1997) was adopted for acquiring the bootstrap confidence intervals (BCIs) for this paper. The moment method is based on large sample theory and requires a sufficiently large resampling dimension to yield symmetric 95% confidence intervals. In this study, the resample dimension  $N$  was set to 2000 based on prior literature discussions as well as computational experiments that confirmed a symmetric distribution for standard errors. Readers interested in detailed descriptions of the bootstrap-

ping method used in this paper can reference the following sources (Archer et al., 1997; Efron and Tibshirani, 1993).

### 5.4 Evaluation of sensitivity analysis results

As argued by Andres (1997), good sensitivity analysis tools should generate repeatable results using a different sample set to evaluate model sensitivities. The effectiveness of a sensitivity analysis method refers to its ability to correctly identify the influential parameters controlling a model’s performance. Building on Andres (1997), we have tested the effectiveness of each of the sensitivity methods using an independent LHS-based random draw of 1000 parameter groups for the 18 parameters analyzed in this study.

The independent sample and the sensitivity classifications from each of the sensitivity analysis methods were combined to develop three parameter sets. Set 1 consists of the full randomly generated independent sample set of 1000 parameter groups. In Set 2, the parameters classified as highly sensitive or sensitive are set to a priori values while the remaining insensitive parameters are allowed to vary randomly. Lastly, in Set 3 the parameters classified as being highly sensitive or sensitive vary randomly and the insensitive parameters are set to a priori values.

Varying parameters that are correctly classified as insensitive in Set 2 should theoretically yield a zero correlation with the full random sample of Set 1 (i.e., plot as a horizontal

**Table 3.** PEST sensitivities based on the RMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold value of 1.0. Light gray designates sensitive parameters defined using a threshold value of 0.01. White cells in the table designate insensitive parameters.

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	0.04	0.18	0.31	0.11	0.35	0.55
	MFMAX	0	0.07	0.01	0	0.05	0.11
	MFMIN	0.03	0.16	0.42	0.03	0.18	0.64
	UADJ	0.01	0.10	0.50	0.02	0.13	0.35
	SI	0	0	0	0	0	0
	UZTWM	0	0	0.01	0	0	0.01
SAC-SMA	UZFWM	0	0	0	0	0	0
	UZK	0.03	0.07	0.16	0.55	1.15	2.26
	PCTIM	0.30	0.57	1.03	0.69	1.65	3.05
	ADIMP	0.17	0.34	0.60	0.36	0.79	1.71
	ZPERC	0	0	0	0.01	0.01	0.02
	REXP	0.01	0.01	0.07	0	0.01	0.04
	LZTWM	0	0	0	0	0	0
	LZFSM	0	0	0	0	0	0
	LZFPM	0	0	0	0	0	0
	LZSK	0.14	0.20	0.58	1.30	5.04	10.06
	LZPK	1.59	1.11	5.54	0.21	0.14	32.61
	PFREE	0.01	0.01	0.08	0.03	0.10	0.03

line). If some parameters are incorrectly classified as insensitive then the scatter plots show deviations from a horizontal line and increased correlation coefficients. Conversely, if the correct subset of sensitive parameters is sampled randomly (i.e., Set 3) then they should be sufficient to capture model output from the random samples of the full parameter set in Set 1 yielding a linear trend with an ideal correlation coefficient of 1. We extended the evaluation methodology of Andres (1997) by calculating the corresponding correlation coefficients instead of using scatter plots.

## 6 Results

Sections 6.1–6.4 present the results attained for each of the four sensitivity analysis methods tested in this study. Results are presented for the SPKP1 and SXTPI watersheds at 1 h, 6 h, and 24 h timescales. Section 6.5 then provides a detailed analysis of how the results from each sensitivity method compare in terms of their selection of highly sensitive, sensitive, and insensitive parameters for the SAC-SMA/SNOW-17 lumped model. Additionally, Sect. 6.5 builds on the work of Andres (1997) to evaluate the relative effectiveness of the methods in identifying the key input parameters controlling model performance. Detailed conclusions on how individual watershed properties impact model performance are beyond the scope of this paper.

Before discussing the sensitivity results in detail, it is worth noting that there are several ways that sensitivity anal-

**Table 4.** PEST sensitivities based on the TRMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold value of 0.1. Light gray designates sensitive parameters defined using a threshold value of 0.001. White cells in the table designate insensitive parameters.

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	0.004	0.014	0.022	0.007	0.020	0.033
	MFMAX	0.001	0.006	0.001	0	0.002	0.001
	MFMIN	0.002	0.008	0.019	0.008	0.016	0.025
	UADJ	0.001	0.005	0.009	0.001	0.007	0.010
	SI	0	0	0	0	0	0
	UZTWM	0	0	0.001	0	0	0.001
SAC-SMA	UZFWM	0	0	0	0	0	0
	UZK	0.007	0.008	0.047	0.061	0.095	0.134
	PCTIM	0.041	0.080	0.159	0.100	0.227	0.419
	ADIMP	0.020	0.038	0.078	0.020	0.045	0.121
	ZPERC	0	0	0.003	0.001	0.002	0
	REXP	0.001	0.001	0.001	0	0.001	0.005
	LZTWM	0	0	0.001	0	0	0.001
	LZFSM	0	0	0	0	0	0
	LZFPM	0	0	0	0	0	0
	LZSK	0.014	0.024	0.075	0.059	0.353	0.006
	LZPK	0.258	0.449	0.713	0.332	5.024	2.513
	PFREE	0.004	0.007	0.016	0.005	0.006	0.009

ysis methods can be evaluated and used in the context of watershed model identification and evaluation. The current study builds on the optimization research of Tang et al. (2006) by focusing on how well PEST, RSA, ANOVA, and Sobol’s method can identify the set of model input parameters that control model performance. Successful screening of the relative importance of input parameters and their interactions can help to limit the dimensionality of calibration search problems and serve to enhance the efficiency of uncertainty analysis. Recall from Sect. 5 that the model performance objectives used in this study evaluate the influence of high streamflow conditions via the RMSE measure and low streamflow conditions via the Box-Cox transformed RMSE (TRMSE). Small values of these measures implies that the SAC-SMA/SNOW-17 streamflow projections closely match observations in the simulated period.

### 6.1 Sensitivity results for PEST

In the case of PEST, sensitivities are computed using the Jacobean derivative-based composite measures defined in Eq. (1). The method is termed local because the composite derivatives are evaluated at a single point in the parameter space deemed locally optimal by the Gauss-Marquardt-Levenberg algorithm. Tables 3 and 4 provide the sensitivities computed by PEST for the RMSE and TRMSE objectives, respectively. In the tables, highly sensitive parameters are designated with dark grey shading, sensitive parameters have light grey shading, and insensitive parameters are not shaded. The SNOW-17 and SAC-SMA parameters are listed

**Table 5.** RSA sensitivities based on the RMSE measure. Dark gray shading designates highly sensitive (HS) parameters. Light gray designates sensitive (S) parameters. White cells in the table designate parameters that are not sensitive (NS).

Model	Parameter	SPKP1			SXTP1		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	S	S	NS	S	S	NS
	MFMAX	NS	NS	NS	NS	NS	NS
	MFMIN	NS	NS	S	NS	S	NS
	UADJ	NS	NS	NS	NS	NS	NS
	SI	NS	NS	S	NS	NS	S
SAC-SMA	UZTWM	S	S	S	NS	S	S
	UZFWM	NS	NS	NS	NS	NS	S
	UZK	NS	NS	NS	NS	NS	NS
	PCTIM	S	S	S	S	S	S
	ADIMP	S	S	S	S	S	S
	ZPERC	NS	NS	NS	S	NS	NS
	REXP	NS	NS	NS	NS	NS	NS
	LZTWM	HS	HS	HS	HS	HS	HS
	LZFSM	S	NS	S	NS	S	S
	LZFPM	NS	NS	NS	S	S	NS
	LZSK	S	S	S	HS	S	S
	LZPK	NS	NS	NS	NS	NS	S
	PFREE	NS	NS	NS	S	NS	NS

separately as are the 1 h, 6 h, and 24 h results for each watershed.

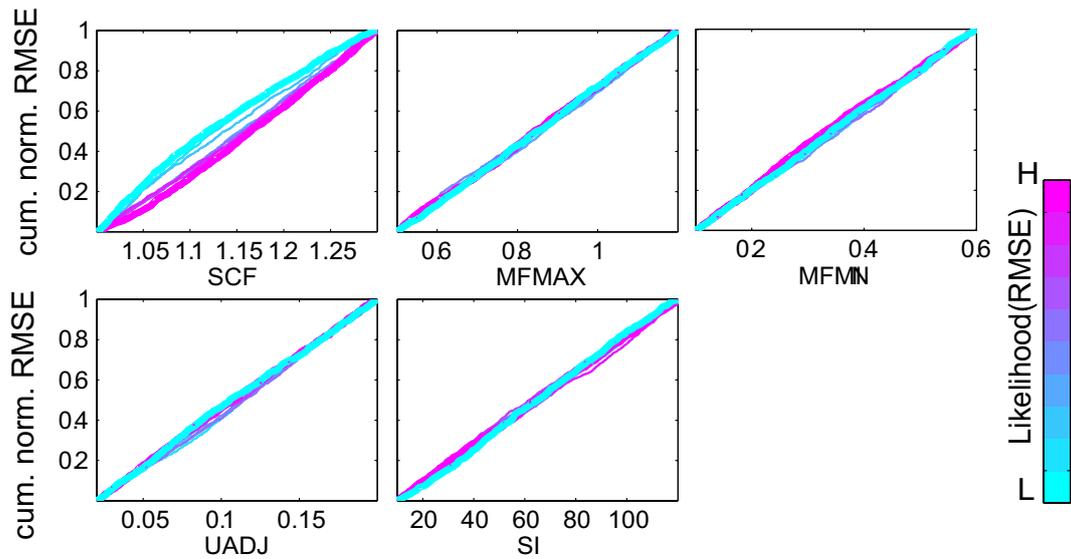
As a caveat, the thresholds used to differentiate highly sensitive, sensitive, and insensitive parameters are based only on the relative magnitudes of the derivatives given in each column, making them subjective and somewhat arbitrary. The thresholds were determined by ranking each column in ascending order and then plotting the relative magnitudes of the derivatives. Results were classified as either highly sensitive or sensitive where the derivative values changed the most significantly. Insensitive parameters had small derivative values that could not be distinguished. Note different thresholds were used for Tables 3 and 4 since the Box-Cox transformation reduced the original range of RMSE by approximately an order of magnitude. The results in Tables 3 and 4 show that PEST did not detect significant changes in parameter sensitivities for high flow (RMSE) versus low flow (TRMSE) conditions. Also differences in the time-scales of predictions as well as watershed locations did not significantly change the PEST sensitivity designations in both tables. Overall PEST found the parameters for impervious cover (PCTIM, ADIMP) and those for storage depletion rates (UZK, LZPK, LZSK) significantly impacted model performance, especially for daily time-scale predictions. The mean water-equivalent threshold for snow cover (SI), upper zone storage parameters (UZTWM, UZFWM), and lower zone storage parameters (LZTWM, LZFSM, and LZFPM) were classified by PEST as being the least sensitive.

**Table 6.** RSA sensitivities based on the TRMSE measure. Dark gray shading designates highly sensitive (HS) parameters. Light gray designates sensitive (S) parameters. White cells in the table designate parameters that are not sensitive (NS).

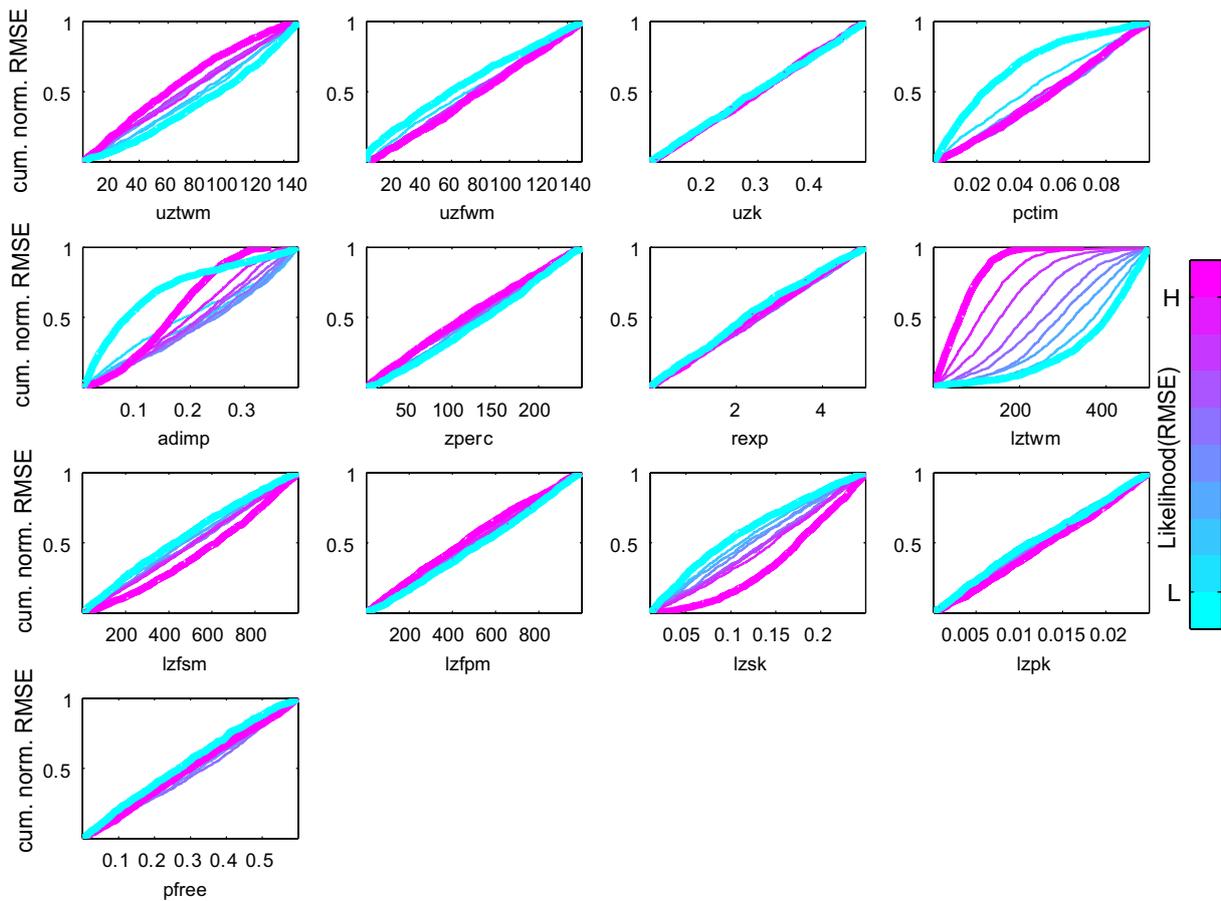
Model	Parameter	SPKP1			SXTP1		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	S	S	NS	S	S	NS
	MFMAX	NS	NS	NS	NS	NS	NS
	MFMIN	NS	NS	S	NS	NS	S
	UADJ	NS	NS	NS	NS	NS	NS
	SI	NS	NS	NS	NS	NS	NS
SAC-SMA	UZTWM	S	S	S	S	S	S
	UZFWM	NS	NS	NS	NS	NS	NS
	UZK	NS	NS	NS	NS	NS	NS
	PCTIM	NS	S	S	S	S	S
	ADIMP	HS	S	S	S	NS	S
	ZPERC	NS	NS	NS	NS	NS	NS
	REXP	NS	NS	NS	NS	NS	NS
	LZTWM	HS	HS	HS	HS	HS	HS
	LZFSM	NS	NS	NS	NS	S	NS
	LZFPM	S	S	S	S	S	S
	LZSK	NS	NS	NS	S	S	NS
	LZPK	NS	NS	NS	S	S	S
	PFREE	S	S	S	S	S	S

## 6.2 RSA Results

As described in Sect. 2.2.2, a visual extension of RSA (Young, 1978; Hornberger and Spear, 1981; Freer et al., 1996; Wagener and Kollat, 2007) was used to evaluate parameter sensitivities for the SAC-SMA/SNOW-17 lumped model. Results were computed for the same timescales and watersheds as were presented for PEST. Given the large number of results analyzed, Figs. 5 and 6 provide sample plots for our RSA analysis, whereas the full sensitivity classifications are summarized in Tables 5 and 6. In Figs. 5 and 6 each model parameter has its own plot with its range on the horizontal axis and its cumulative normalized RMSE distribution value on the vertical axis. In the plots, color shading is used to differentiate the likelihoods of each one of the ten bins used to divide the input parameter samples. High likelihood bins plotted in purple represent portions of the parameters' ranges where low RMSE values are expected. In the context of sensitivity analysis, RSA measures the distribution of model responses that result from the 10 000 Latin hypercube input parameter groups sampled. When parameters are insensitive (see the SNOW-17 results shown in Fig. 5) each of the 10 sample bins plot over each other in linear trend lines that are representative of uniformly distributed RMSE values. Sensitive parameters produced highly dispersed bin lines such as those shown for LZTWM shown in Fig. 6.



**Fig. 5.** RSA (Regional Sensitivity Analysis) plot for Snow17 parameters in the SPKP1 watershed. The objective function is RMSE based on a 1-hour time interval.



**Fig. 6.** RSA (Regional Sensitivity Analysis) plot for SAC-SMA parameters in the SPKP1 watershed. The objective function is RMSE based on a 1-hour time interval.

**Table 7.** ANOVA single parameter sensitivities based on the RMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold F value of 460. Light gray designates sensitive parameters defined using a threshold F value of 4.6. White cells in the table designate insensitive parameters. The values in the brackets provide the 95% confidence interval for the F-values (i.e., the unbracketed value  $\pm$  the bracketed value yields the confidence interval).

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	57.02 [21.78]	94.63 [27.57]	0.24 [2.40]	80.38 [25.71]	71.50 [23.80]	0.65 [2.79]
	MFMAX	0.64 [2.90]	1.79 [4.22]	1.27 [3.67]	1.40 [3.99]	1.15 [3.60]	1.03 [3.40]
	MFMIN	1.51 [3.69]	13.52 [10.26]	178.03 [36.16]	0.25 [2.23]	19.53 [12.04]	9.96 [9.14]
	UADJ	1.25 [3.63]	1.76 [4.22]	0.86 [3.27]	1.18 [3.52]	1.01 [3.50]	1.89 [4.46]
	SI	1.25 [3.60]	0.03 [2.00]	9.94 [9.18]	5.27 [6.83]	7.58 [7.62]	1.83 [4.53]
	SAC-SMA	UZTWM	61.70 [23.35]	97.69 [27.41]	85.25 [26.54]	41.01 [17.68]	27.95 [14.84]
UZFWM		57.14 [21.34]	22.91 [13.43]	44.90 [19.07]	31.87 [15.70]	64.83 [23.10]	119.42 [29.68]
UZK		2.51 [4.83]	2.96 [5.06]	1.56 [3.98]	23.66 [13.23]	3.40 [5.33]	1.03 [3.33]
PCTIM		123.52 [34.54]	463.48 [63.95]	232.04 [45.87]	107.64 [29.38]	121.19 [32.73]	9.81 [8.78]
ADIMP		173.27 [42.48]	757.63 [82.30]	127.19 [33.93]	426.75 [60.13]	273.77 [50.98]	210.94 [46.35]
ZPERC		11.91 [9.57]	18.38 [11.71]	9.59 [8.49]	64.16 [22.80]	30.37 [15.09]	9.79 [8.44]
REXP		7.81 [8.31]	7.77 [8.20]	8.23 [8.13]	19.62 [12.48]	13.86 [10.86]	7.36 [7.84]
LZTWM		3047.02 [326.83]	3870.49 [255.63]	4025.30 [356.21]	1748.99 [143.26]	2722.24 [240.93]	1265.10 [199.25]
LZFSM		54.29 [20.53]	22.75 [13.32]	51.53 [20.32]	12.46 [10.16]	33.93 [16.37]	36.64 [17.31]
LZFPM		15.24 [11.28]	39.95 [17.75]	21.01 [13.16]	363.73 [56.37]	153.42 [35.83]	14.38 [10.67]
LZSK		196.73 [39.09]	178.52 [36.40]	200.72 [39.07]	489.21 [66.64]	456.82 [62.66]	242.11 [48.62]
LZPK		19.18 [11.97]	13.57 [10.32]	13.28 [10.26]	8.94 [8.55]	39.64 [17.31]	37.17 [17.45]
PFREE		13.02 [10.35]	3.46 [5.53]	26.09 [14.49]	140.19 [31.35]	0.25 [2.44]	29.44 [16.21]

The SAC-SMA/SNOW-17 sensitivity classifications resulting from RSA are presented in Tables 5 and 6. The classifications represent our qualitative interpretation of visual plots similar to those in Figs. 5 and 6 for each timescale and each watershed. As is standard in hydrologic applications of RSA (e.g., Freer et al., 1996; Wagener and Kollat, 2007), only individual parameter impacts on model performance are considered and parameter interactions have been neglected. Analysis of Tables 5 and 6 show changes in sensitivity when comparing across timescales, watersheds, and model performance objectives. Examples of these changes include the increased importance of the SNOW-17 parameters such as the mean water-equivalent above which 100-percent cover exists (SI) for the RMSE measure (i.e., high flow) and the minimum melt factor for non-rain periods (MFMIN) for the TRMSE measure (i.e., low flow) at the daily timescale. Both the RMSE measure and the TRMSE measure identified the vadose zone storage (LZTWM) as the most sensitive parameter in all of the tested cases. Shifting the focus from high flow to low flow using the TRMSE measure resulted in the percolation factor (PFREE) and lower zone free water primary maximum storage (LZFPM) being classified as being sensitive.

### 6.3 Sensitivity results for ANOVA

Recall that ANOVA is a parametric analysis of variance that uses the assumption of normally distributed model responses

(RMSE and TRMSE for streamflow in this study) to partition variance contributions between single parameters and parameter interactions. In this study, a second order ANOVA model (i.e., a model that considers pair wise parameter interactions) was fitted to the model outputs and the F-test is used to evaluate the statistical significance of each parameter's or parameter interaction's impact on the model output. Higher F-values indicate higher significance or sensitivity. Additionally, the coefficient of determination  $r^2$  can be used to measure if incorporating parameter interactions into the ANOVA model improves its ability to represent model output variability (Mokhtari and Frey, 2005). Because random sampling can introduce significant uncertainty into the calculation of F-values, we have followed the recommendations of Archer et al. (1997) and used statistical bootstrapping to provide 95% confidence intervals for our ANOVA sensitivity rankings. Tables 7 and 8 provide F-values for each parameter as well as its bootstrapped confidence interval. Tabular presentation of the ANOVA results improved their clarity since the F-values ranged over 4 orders of magnitude [0.25–4000] making plots difficult to interpret.

Tables 7 and 8 are formatted similarly to the prior sensitivity tables where highly sensitive parameters have dark grey shading, sensitive parameters have light grey shading, and insensitive parameters have no shading. These classifications were based on the F-distribution where a threshold of 4.6 represents less than a 1-percent chance of misclassi-

**Table 8.** ANOVA single parameter sensitivities based on the TRMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold F value of 460. Light gray designates sensitive parameters defined using a threshold F value of 4.6. White cells in the table designate insensitive parameters. The values in the brackets provide the 95% confidence interval for the F-values (i.e., the unbracketed value  $\pm$  the bracketed value yields the confidence interval).

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	72.17 [24.01]	81.30 [25.20]	12.21 [9.91]	73.38 [23.90]	53.47 [20.35]	2.07 [4.49]
	MFMAX	0.80 [3.17]	1.02 [3.45]	0.88 [3.26]	1.62 [4.04]	0.32 [2.51]	0.36 [2.51]
	MFMIN	2.75 [4.81]	10.04 [8.87]	67.12 [22.84]	2.12 [4.44]	16.74 [11.47]	39.02 [17.84]
	UADJ	0.58 [2.84]	1.19 [3.58]	0.99 [3.40]	2.05 [4.37]	1.53 [3.94]	1.90 [4.32]
	SI	1.07 [3.45]	0.82 [3.20]	1.39 [3.85]	0.60 [2.95]	5.41 [6.78]	3.71 [5.69]
	SAC-SMA	UZTWM	521.40 [66.33]	510.03 [65.56]	363.17 [56.17]	133.85 [32.53]	99.84 [28.13]
UZFWM		3.27 [5.48]	9.91 [9.21]	13.50 [10.73]	3.00 [5.20]	20.38 [12.90]	22.24 [13.72]
UZK		8.33 [8.36]	2.45 [4.84]	1.73 [4.21]	5.11 [6.43]	1.43 [3.77]	2.85 [4.95]
PCTIM		11.81 [9.95]	192.84 [39.35]	121.95 [31.21]	12.18 [9.07]	10.91 [8.74]	5.90 [6.49]
ADIMP		609.07 [71.54]	101.10 [27.57]	142.23 [32.79]	92.64 [26.66]	25.52 [13.52]	178.11 [37.74]
ZPERC		4.71 [6.33]	1.45 [3.89]	1.64 [4.08]	0.85 [3.33]	0.97 [3.28]	0.58 [2.91]
REXP		1.33 [3.73]	2.90 [5.07]	3.33 [5.32]	1.59 [3.83]	3.24 [5.21]	2.25 [4.49]
LZTWM		3024.57 [188.15]	3989.52 [227.00]	3898.47 [228.48]	2736.39 [177.17]	2801.18 [184.95]	2208.46 [164.19]
LZFSM		16.18 [11.77]	10.77 [9.60]	9.63 [9.10]	25.33 [14.60]	38.08 [17.60]	2.70 [4.89]
LZFPM		353.09 [54.52]	265.59 [47.14]	256.97 [46.11]	93.66 [28.50]	35.10 [16.60]	283.82 [50.95]
LZSK		3.38 [5.63]	1.58 [4.09]	1.38 [3.85]	11.40 [9.49]	88.67 [25.12]	1.95 [4.33]
LZPK		11.93 [10.13]	9.04 [8.74]	4.30 [6.31]	201.48 [40.04]	143.88 [33.50]	74.30 [23.05]
PFREE		384.08 [58.19]	368.17 [56.55]	483.47 [66.90]	394.68 [58.04]	429.42 [62.26]	563.84 [73.40]

fying a parameter as sensitive. As can be seen in the tables, some parameters' F-values were up to three orders of magnitude larger than 4.6. A threshold of 460 was used to classify parameters as being highly sensitive. Although the threshold used to classify highly sensitive parameters is subjective, it accurately captures those parameters with very large F-values.

Analysis of Table 7 shows that for the high-flow RMSE objective, the most significant differences in sensitivities across timescales and across watersheds involved SNOW-17 parameters. The results show increasing sensitivities for the minimum melt factor for non-rain periods (MFMIN) at the 6-hour and daily timescales. Overall, Table 7 shows that most of the SAC-SMA parameters are sensitive for high flow conditions regardless of timescale or watershed. The high flow RMSE analysis identified the lower zone storage (LZTWM) as having the highest influence on model variance while the upper zone free water lateral depletion rate (UZK) is rated to have the least impact.

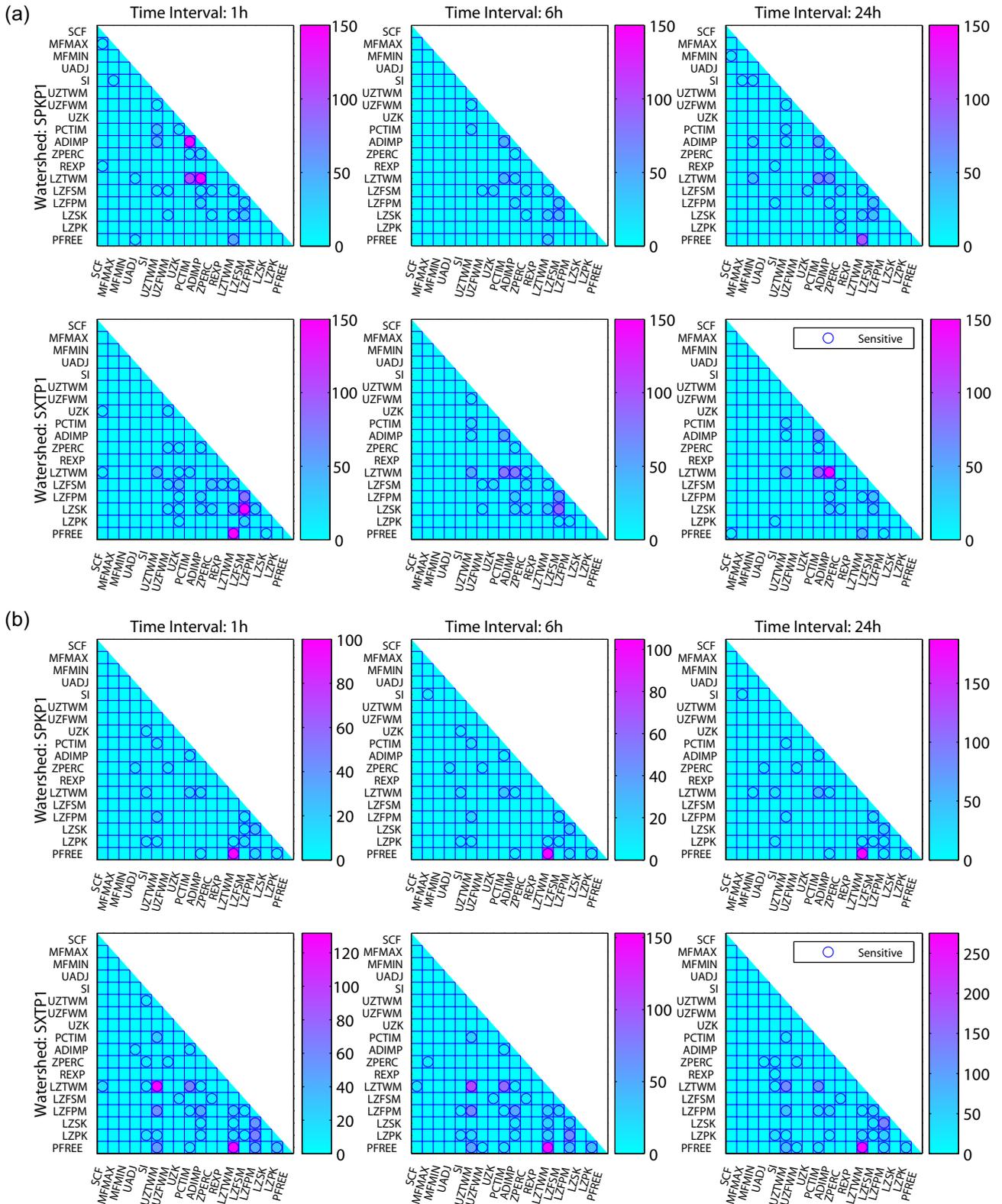
In Table 8 the ANOVA results using the low flow TRMSE objective are substantially different from those for high flow in Table 7. For low flow conditions, fewer parameters are classified as being sensitive. Table 8 shows a general reduction relative to Table 7 in the influence of the upper zone free water storage (UZFWM) and an increase in the importance of the upper zone tension water storage (UZTWM) as well as the percolation factor (PFREE).

Beyond single parameter sensitivities, the coefficients of determination in Table 9 show that 2nd order interactions (or pairwise parameter interactions) improve the accuracy of the ANOVA model, which means the model better represents the total variance of the SAC-SMA/SNOW-17 model output. The coefficients of determination show that 2nd order parameter interactions improve the ANOVA models' performances by up to 40%. Figure 7 illustrates the 2nd order parameter interactions impacting the SAC-SMA/SNOW-17 model. Second order analysis changes the degrees of freedom used when analyzing the F-distribution making it necessary to define a new threshold in Fig. 7. An F-value threshold of 3.32 designates at least a 99% likelihood of being sensitive. Again higher F-values imply higher sensitivity.

Figure 7 provides a more detailed portrayal of how parameter sensitivities change across timescales for each of the watershed models. The RMSE results in Fig. 7a show that interactively sensitive parameters varied across watersheds as well as timescales. The results in Fig. 7b and Table 9 show that the SXTPI watershed possesses more TRMSE-based ANOVA interactions than the SPKP1 watershed. The results imply each watershed model has a "unique" set of parameter interactions impacting its performance (Beven, 2000).

#### 6.4 Sensitivity results for Sobol's method

Recall from Sect. 2.2.4, that Sobol's method decomposes the overall variance of the sampled SAC-SMA/SNOW-17 model



**Fig. 7.** (a) ANOVA second order parameter interactions based on the RMSE measure. (b) ANOVA second order parameter interactions based on the TRMSE measure. Circles represent statistically significant F-values defined using the threshold value of 3.32. The color legends and shading represent the F-value magnitudes and ranges.

**Table 9.** Coefficients of determination for the ANOVA model. R1 designates a 1st order ANOVA model that neglects parameter interactions. R2 designates a 2nd order ANOVA model that accounts for pairwise parameter interactions.

Order	RMSE						TRMSE					
	SPKP1			SXTPI			SPKP1			SXTPI		
	1h	6h	24h									
R1	0.535	0.755	0.643	0.601	0.604	0.543	0.745	0.739	0.718	0.556	0.546	0.541
R2	0.730	0.861	0.778	0.790	0.753	0.655	0.864	0.856	0.850	0.777	0.760	0.750

output to compute 1st order (single parameter), 2nd order (two parameter), and total order sensitivity indices. These indices are presented as percentages and have straightforward interpretations as representing the percent of total model output variance contributed by a given parameter or parameter interaction. The total order indices are the most comprehensive measures of a single parameter's sensitivity since they represent the summation of all variance contributions involving that parameter (i.e., its 1st order contribution plus all of its pairwise interactions).

Table 10 shows the relative importance of 1<sup>st</sup> and 2nd order effects for all of the cases analyzed. Readers should note that the truncation and Monte Carlo approximations of the integrals required in Sobol's method can lead to small numerical errors (e.g., see Archer et al., 1997; Sobol', 2001; Fieberg and Jenkins, 2005) such as slightly negative indices or for example in Table 10 the few cases where 1st and 2nd order effects sum to be slightly larger than 1. In this study these effects were very small and did not impact parameter rankings. Table 10 supports our analysis assumption that 1st and 2nd order parameter sensitivities explain nearly all of the variance in the SAC-SMA/SNOW-17 model's output distributions. The table also shows that the importance of 2-parameter interactions ranged from 3% to 40% of the total variance depending on the model performance objective, the prediction timescale, and the watershed. Except for SXTPI 6-hour test case, the results indicate that there were more parameter interactions for the RMSE measure compared to the TRMSE measure.

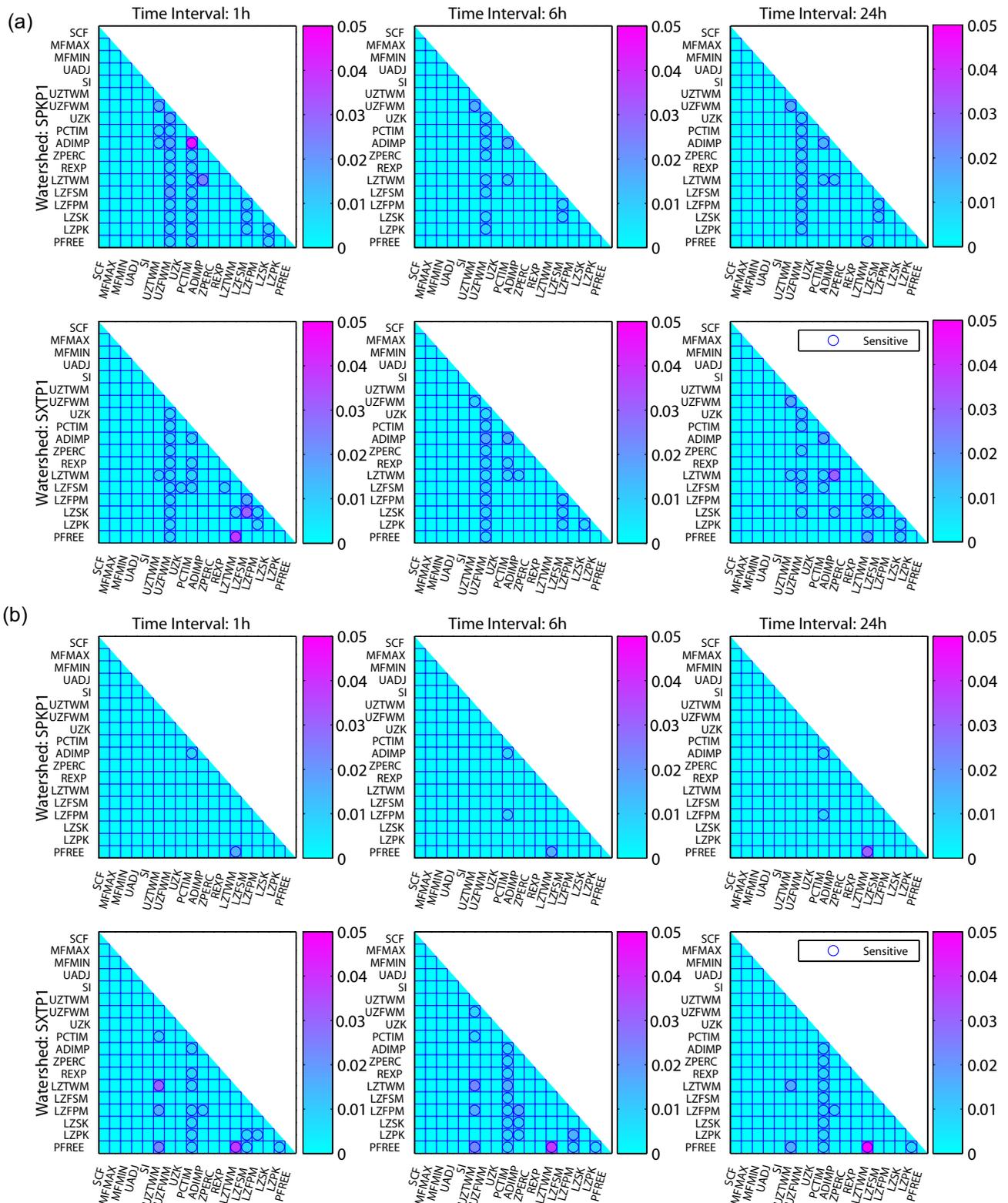
Tables 11 and 12 summarize the total order indices (i.e., total variance contributions) for the SAC-SMA/SNOW-17 parameters analyzed. Again highly sensitive parameters are designated with dark grey shading, sensitive parameters have light grey shading, and insensitive parameters are not shaded. In all of the results presented for Sobol's method, parameters classified as highly sensitive had to contribute on average at least 10-percent of the overall model variance and sensitive parameters had to contribute at least 1-percent. These thresholds are subjective and their ease-of-satisfaction decreases with increasing numbers of parameters or parameter interactions. In Tables 11 and 12 the total order indices again show that the model performance objective, the prediction timescale, and the watershed all heavily impact the SAC-SMA/SNOW-17 sensitivities.

In both tables, the SNOW-17 parameters contributed minimally to the overall variance of the simulation model's output. Only the minimum melt factor for non-rain periods (MFMIN) parameter has a statistically significant sensitivity when the bootstrapped confidence intervals are considered. Tables 11 and 12 also insinuate that most of the SAC-SMA model parameters are sensitive. For the high-flow RMSE results, the lower zone tension water storage (LZTWM) and the additional impervious area (ADIMP) were the most sensitive SAC-SMA parameters. The upper zone storage parameters (UZTWM, UZFWM) and all of the lower zone parameters dominate model response for the low-flow TRMSE measure. In particular, the lower zone tension water storage (LZTWM) appears to be the dominant overall parameter as it explains about 50% of the output's variance for each test case. Similar to ANOVA's results, there are fewer parameters classified as being sensitive for the TRMSE measure versus RMSE.

Figure 8 provides a more detailed understanding of the total order indices presented in Tables 11 and 12. Similar to the ANOVA interaction plots in Section 6.3, these figures show the matrix of parameter interactions where circles designate pairings that contribute at least 1% of the overall model output variance. The actual 2nd order indices' values are shown with the color shading defined in the plots' legends. These plots show how the dominant parameters for both the RMSE and TRMSE measures tend to have the greatest number of interactions (e.g., LZTWM and PFREE in Fig. 8). Interestingly, there are very distinct differences for the parameter interactions for the two watersheds. When comparing the RMSE results in Fig. 8a with TRMSE results in Fig. 8b the shift from high-flow to low flow analysis tends to substantially decrease the importance of parameter interactions for the SPKP1 watershed, whereas no signification reduction was found for the SXTPI watershed. Readers should note that our 1% threshold for Sobol's method is particularly conservative when analyzing Figs. 8a and b since the number of variables analyzed increases from 18 for 1st order analysis to 162 parameter interactions in 2nd order analysis.

## 6.5 Comparative summary of sensitivity methods

Sections 6.1-6.4 present classifications of SAC-SMA/SNOW-17 model parameters into three categories: (1) highly sensitive, (2) sensitive, and (3) insensitive. Given



**Fig. 8.** (a) Second order parameter interactions based on the RMSE measure computed using Sobol’s method. (b) Second order parameter interactions based on the TRMSE measure computed using Sobol’s method. Circles represent interactions that contribute at least 1% of the overall model output variance. The color legends and shading represent the Sobol indices’ magnitudes and ranges.

**Table 10.** Summations of Sobol's sensitivity indices for 1st order and 2nd order contributions to model output variance.

Order	RMSE						TRMSE					
	SPKP1			SXTPI			SPKP1			SXTPI		
	1h	6h	24h									
Summation of 1st order effects	0.61	0.84	0.79	0.71	0.79	0.69	0.91	0.85	0.86	0.68	0.69	0.73
Summation of 2nd order effects	0.40	0.15	0.19	0.32	0.22	0.25	0.03	0.04	0.06	0.27	0.33	0.24

**Table 11.** Total order sensitivity indices from Sobol's method computed using the RMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold value of 0.1. Light gray designates sensitive parameters defined using a threshold value of 0.01. White cells in the table designate insensitive parameters. The values in the brackets provide the 95% confidence interval for the indices' values (i.e., the unbracketed value  $\pm$  the bracketed value yields the confidence interval).

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	0.00 [0.00]	0.01 [0.00]	0.00* [0.01]	0.01 [0.00]	0.01 [0.00]	0.02 [0.03]
	MFMAX	0.00 [0.00]	0.00 [0.00]	0.00* [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	MFMIN	0.00* [0.00]	0.01 [0.00]	0.07 [0.01]	0.00 [0.00]	0.01 [0.00]	0.01 [0.01]
	UADJ	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00* [0.01]
	SI	0.00 [0.00]	0.00 [0.00]	0.01 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
SAC-SMA	UZWWM	0.06 [0.03]	0.05 [0.01]	0.06 [0.02]	0.03 [0.01]	0.05 [0.02]	0.05 [0.04]
	UZFWWM	0.20 [0.04]	0.07 [0.02]	0.16 [0.03]	0.04 [0.01]	0.16 [0.03]	0.43 [0.06]
	UZK	0.01 [0.00]	0.01 [0.00]	0.00 [0.00]	0.03 [0.01]	0.01 [0.00]	0.00 [0.00]
	PCTIM	0.11 [0.01]	0.12 [0.01]	0.08 [0.01]	0.03 [0.01]	0.06 [0.01]	0.05 [0.01]
	ADIMP	0.30 [0.02]	0.33 [0.02]	0.13 [0.01]	0.19 [0.01]	0.20 [0.01]	0.23 [0.04]
	ZPERC	0.03 [0.01]	0.01 [0.00]	0.01 [0.00]	0.05 [0.01]	0.02 [0.00]	0.01 [0.00]
	REXP	0.01 [0.01]	0.00 [0.00]	0.00 [0.00]	0.03 [0.01]	0.01 [0.00]	0.01 [0.00]
	LZTWM	0.48 [0.03]	0.48 [0.02]	0.52 [0.03]	0.40 [0.02]	0.45 [0.02]	0.34 [0.04]
	LZFSM	0.08 [0.02]	0.04 [0.01]	0.05 [0.01]	0.13 [0.02]	0.08 [0.01]	0.05 [0.01]
	LZFPM	0.03 [0.02]	0.02 [0.00]	0.02 [0.00]	0.12 [0.01]	0.06 [0.01]	0.03 [0.01]
	LZSK	0.08 [0.01]	0.07 [0.01]	0.08 [0.01]	0.19 [0.01]	0.15 [0.01]	0.09 [0.01]
LZPK	0.02 [0.01]	0.01 [0.00]	0.01 [0.00]	0.03 [0.01]	0.02 [0.00]	0.02 [0.00]	
PFREE	0.02 [0.00]	0.01 [0.00]	0.02 [0.00]	0.12 [0.01]	0.01 [0.00]	0.03 [0.00]	

\* Negative mean value was set to zero

the large number of cases analyzed in this study, Fig. 9 was developed to provide a comparative summary of the results attained from the four sensitivity analysis methods. These figures show that there are distinct similarities and differences between the sensitivity classifications attained using each method. For example, despite the subjective decisions required to differentiate highly sensitive and sensitive parameters, generally RSA, ANOVA, and Sobol's method agree on their classifications of the most sensitive parameters for each scenario.

All three global sensitivity methods (RSA, ANOVA, and Sobol's method) show that the SAC-SMA/SNOW-17 model's responses are "uniquely" determined by the performance objective specified, prediction timescale, and specific watershed being modeled (Beven, 2000). Differences be-

tween the four sensitivity methods' classifications as illustrated in Fig. 9 are particularly pronounced for parameters at the threshold between sensitive and insensitive. Significant differences exist between the classifications of PEST and those of the other three algorithms. One of the biggest discrepancies shown in the plots is that PEST generally found the SNOW-17 parameters to be sensitive. Another distinct difference is that PEST found the storage parameters (UZWWM, UZFWWM, LZTWM, LZFSM, LZFPM) to be insensitive while other methods identified these parameters to be sensitive or highly sensitive. It is likely that the PEST results are reflective of local optima in the model's response surface, which would be expected to be highly multimodal (Duan et al., 1992; Tang et al., 2006).

**Table 12.** Total order sensitivity indices from Sobol’s method computed using the TRMSE measure. Dark gray shading designates highly sensitive parameters defined using a threshold value of 0.1. Light gray designates sensitive parameters defined using a threshold value of 0.01. White cells in the table designate insensitive parameters. The values in the brackets provide the 95% confidence interval for the indices’ values (i.e., the unbracketed value  $\pm$  the bracketed value yields the confidence interval).

Model	Parameter	SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h
SNOW-17	SCF	0.01 [0.00]	0.01 [0.00]	0.00 [0.00]	0.01 [0.00]	0.01 [0.00]	0.00 [0.00]
	MFMAX	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	MFMIN	0.00 [0.00]	0.00 [0.00]	0.01 [0.00]	0.00 [0.00]	0.00 [0.00]	0.01 [0.00]
	UADJ	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	SI	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
SAC-SMA	UZTWM	0.14 [0.01]	0.14 [0.01]	0.10 [0.01]	0.14 [0.01]	0.13 [0.01]	0.08 [0.01]
	UZFWM	0.02 [0.01]	0.03 [0.01]	0.05 [0.01]	0.02 [0.01]	0.07 [0.02]	0.10 [0.02]
	UZK	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	PCTIM	0.02 [0.00]	0.06 [0.01]	0.04 [0.01]	0.05 [0.01]	0.05 [0.01]	0.04 [0.01]
	ADIMP	0.22 [0.01]	0.05 [0.01]	0.06 [0.01]	0.08 [0.01]	0.05 [0.01]	0.09 [0.01]
	ZPERC	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.01 [0.00]	0.00 [0.00]	0.00 [0.00]
	REXP	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	LZTWM	0.43 [0.02]	0.49 [0.02]	0.51 [0.02]	0.51 [0.02]	0.51 [0.02]	0.46 [0.02]
	LZFSM	0.02 [0.01]	0.02 [0.00]	0.02 [0.00]	0.03 [0.01]	0.04 [0.01]	0.02 [0.01]
	LZFPM	0.10 [0.01]	0.08 [0.01]	0.08 [0.01]	0.12 [0.01]	0.11 [0.01]	0.13 [0.01]
	LZSK	0.02 [0.00]	0.02 [0.00]	0.02 [0.00]	0.05 [0.01]	0.06 [0.01]	0.04 [0.01]
	LZPK	0.04 [0.01]	0.03 [0.01]	0.03 [0.01]	0.09 [0.01]	0.08 [0.01]	0.05 [0.01]
	PFREE	0.13 [0.01]	0.13 [0.01]	0.16 [0.01]	0.23 [0.02]	0.23 [0.01]	0.27 [0.02]

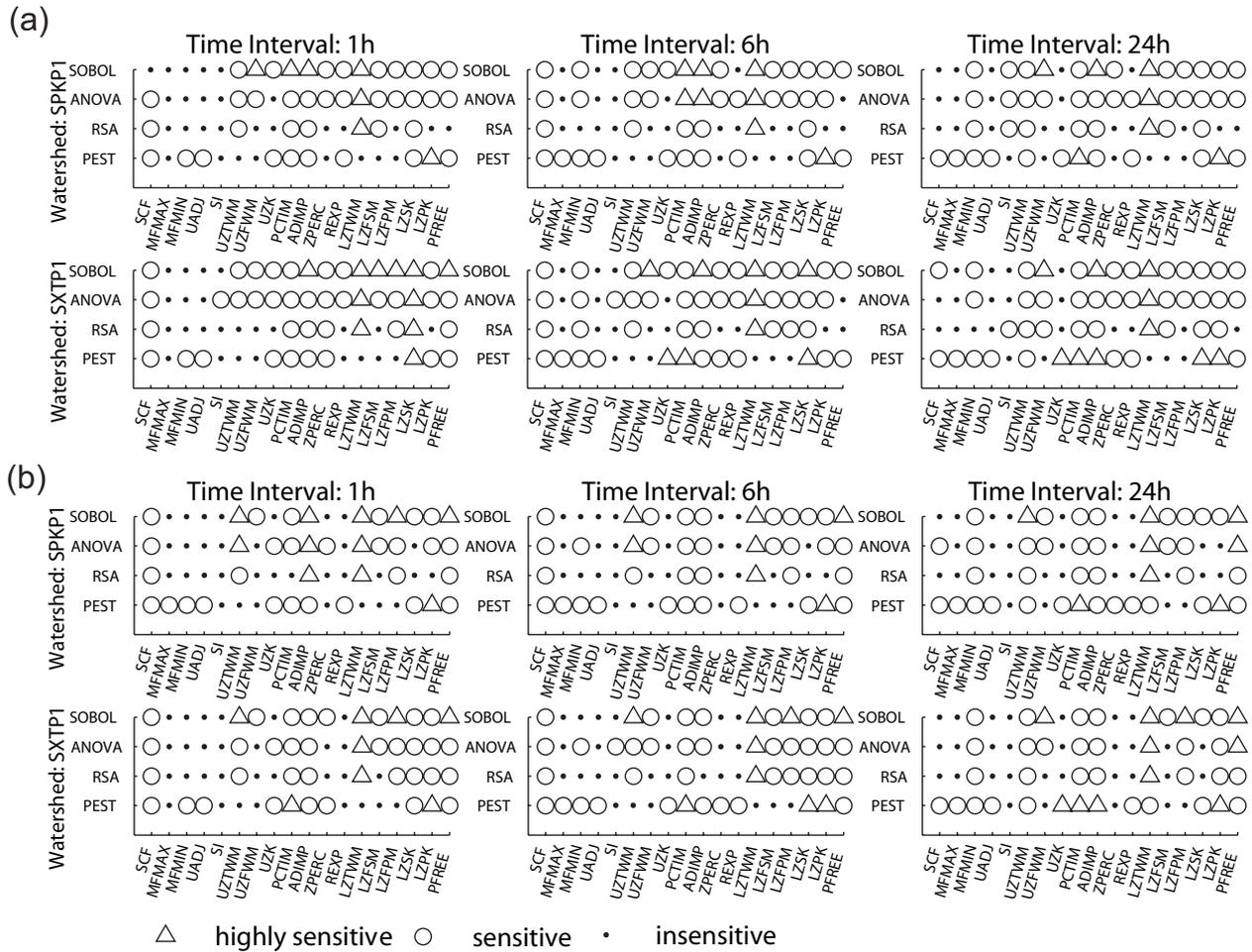
\* Negative mean value was set to zero

Figure 9 shows that RSA generally defined the smallest subset of SAC-SMA/SNOW-17 parameters as being sensitive or highly sensitive. The RSA version used in this study is unique among the four tested sensitivity methods in the sense that our classifications required qualitative assessments of a visual representation of results. As noted above RSA yields very similar rankings for highly sensitive results, but the qualitative interpretation of sensitivity becomes more challenging for parameters that show modest sensitivity.

Although Fig. 9 provides a comparative synopsis of the different results attained by the four sensitivity methods, it does not allow for any quantitative analysis of their relative effectiveness as screening tools. Building on Andres (1997), we have tested the effectiveness of each of the sensitivity methods used in this study. We have used the sensitivity classifications given in Fig. 9 in combination with an independent LHS-based random draw of 1000 parameter groups for the 18 parameters analyzed in this study. Recall from Sect. 5.4, that the independent sample and the sensitivity classifications in Fig. 9 were used to develop three parameter sets. Set 1 consists of the full randomly generated independent sample set. In Set 2, the parameters classified as highly sensitive or sensitive are set to a priori fixed values while the remaining insensitive parameters are allowed to vary randomly. Lastly, in Set 3 the parameters classified as being highly sensitive or sensitive vary randomly and the insensitive parameters are set to a priori values.

Figure 10 illustrates that by plotting Set 2 versus Set 1 as well as Set 3 versus Set 1 we can test the effectiveness of the sensitivity analysis methods. As shown for the Sobol’s method and ANOVA results in Fig. 10a varying parameters that are correctly classified as “insensitive” in Set 2 should theoretically yield a zero correlation with the full random sample of Set 1 (i.e., plot as a horizontal line). If some parameters are incorrectly classified as insensitive then the scatter plots show deviations from a horizontal line and increased correlation coefficients as is the case for the PEST and RSA results in Fig. 10a. Conversely, if the correct subset of sensitive parameters is sampled randomly (i.e., Set 3) then they should be sufficient to capture model output from the random samples of the full parameter set in Set 1 yielding a linear trend with an ideal correlation coefficient of 1. Figure 10b shows that the Sobol’s method yields the highest correlation between Set 3 and Set 1 followed closely by ANOVA. PEST yields the worst correlations for the hourly SPKP1 watershed’s results shown in Fig. 10. More generally, the plots in Fig. 10 show that this analysis can be quantified using correlation coefficients.

Table 13 provides a summary of correlation coefficients for all of the test cases analyzed in this study. PEST shows the worst performance overall and in a few cases PEST performed comparably to RSA for the RMSE-based evaluations. The correlation coefficients in Table 13 show that Sobol’s method and ANOVA perform very similarly in terms of their



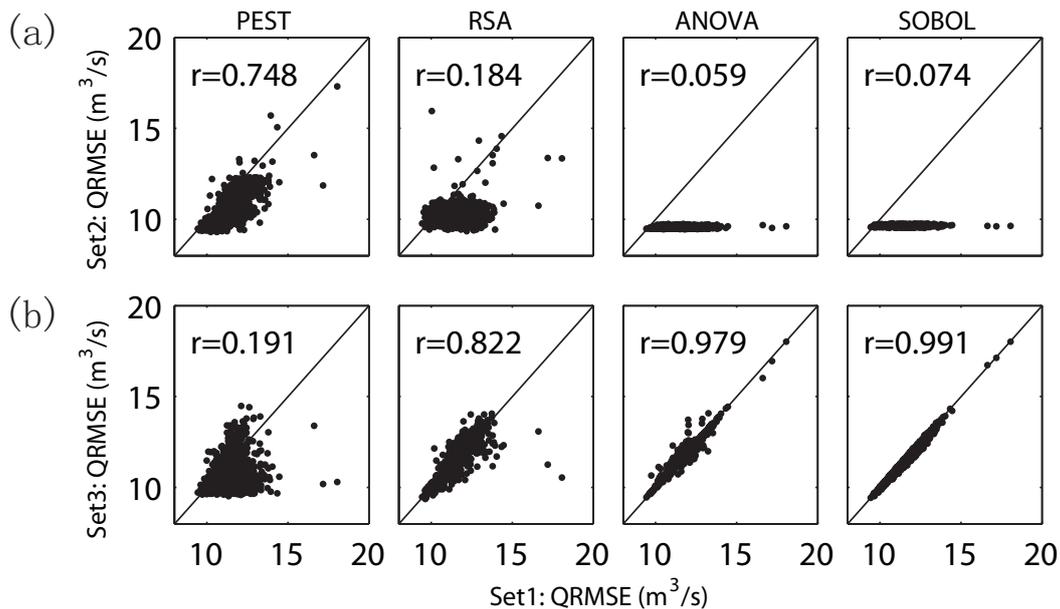
**Fig. 9.** (a) Comparative summary of sensitivity classifications based on the high-flow RMSE model performance objective. (b) Comparative summary of sensitivity classifications based on the high-flow TRMSE model performance objective.

**Table 13.** Summary of correlation coefficients from the independent testing of each sensitivity method’s effectiveness.

Sets	Methods	RMSE						TRMSE					
		SPKP1			SXTPI			SPKP1			SXTPI		
		1h	6h	24h	1h	6h	24h	1h	6h	24h	1h	6h	24h
Set 2 vs. Set 1 ideally R=0	PEST	0.748	0.749	0.782	0.282	0.327	0.395	0.798	0.844	0.306	0.499	0.504	0.183
	RSA	0.184	0.242	0.264	0.029	0.165	0.009	0.147	0.159	0.195	0.062	0.130	0.115
	ANOVA	0.059	0.046	0.050	-0.059	-0.066	-0.049	0.000	-0.001	0.029	0.004	-0.003	0.041
	SOBOL	0.074	0.023	0.023	-0.013	0.016	-0.058	0.067	0.016	0.022	-0.024	0.004	0.011
Set 3 vs. Set 1 ideally R=1	PEST	0.191	0.316	0.272	0.426	0.240	0.283	0.331	0.208	0.882	0.303	0.261	0.719
	RSA	0.822	0.738	0.831	0.699	0.565	0.711	0.922	0.935	0.926	0.905	0.852	0.790
	ANOVA	0.979	0.981	0.984	1.000	0.984	0.988	0.961	0.972	0.961	0.850	0.911	0.817
	SOBOL	0.991	0.985	0.950	1.000	0.999	0.991	0.982	0.991	0.990	0.991	0.909	0.963

effectiveness. Sobol’s method has a slight advantage in that its sensitivity classifications are more certain or robust relative to ANOVA’s sensitivity rankings. Sobol’s method and ANOVA have consistently superior correlations compared to

both PEST and RSA, which should be expected given the importance of parameter interactions in the problems analyzed.



**Fig. 10.** Example illustration of an independent test of the sensitivity classifications found for the SPKPI watershed's model at the 1-hour timescale. The scatter plots show the RMSE of streamflow predictions. Set 1 consists of 1000 randomly drawn Latin hypercube samples. Set 2 is composed of constant reference values for sensitive parameters and random samples of the remaining insensitive parameters. In Set 3 sensitive parameters are allowed to vary randomly and insensitive parameters are set to constant reference values. Term  $r$  represents correlation coefficient.

## 7 Discussion

The results of this study show that model parameter sensitivities are heavily impacted by the choice of analysis method as well as the selected model time interval. Differences between the two adjacent watersheds also suggest strong influences of local physical characteristics on the sensitivity methods' results. As the only local sensitivity approach analyzed, it is not surprising that PEST yielded results that were often significantly different from the global sensitivity methods. As noted in Sect. 6.5, the most contradictory PEST result in this study was its classification of the SNOW-17 parameters and the SAC-SMA storage parameters as being sensitive. This apparent misclassification of the snow parameters reflects the biggest and to a degree most expected limitation for PEST. Readers should be aware that the linearization of the relationship between a model's output and its parameters will adversely impact PEST applications for hydrologic models with thresholds because of their impacts on the derivatives in the Taylor's series expansion. As a local sensitivity approach, PEST is more prone to misclassify sensitivities for highly multimodal response surfaces since the method's derivatives are computed at a single point determined to be locally optimal by the Gauss-Marquardt-Levenberg algorithm (Abbaspour et al., 2001; Doherty and Johnston, 2003). The model response surface is defined as the mapping of all sampled parameter inputs for a model into the set of RMSE (or TRMSE) values that define the accuracy

of its performance. Hydrologic models have been shown to yield complex surfaces with a large number of local optima (Duan et al., 1992; Vrugt et al., 2003; Tang et al., 2006). As the search space dimension and complexity increases, PEST's composite Jacobean-based measure has an increasing likelihood of identifying sensitivities that represent local anomalies in the model's response. PEST's strengths lie in its computational efficiency, ease-of-implementation, and ease-of-use [for an application of inversion of a groundwater flow-and-transport model see Tonkin and Doherty (2005)].

Overall the results in Table 13 indicate that RSA is more effective than PEST, but less effective than ANOVA and Sobol's method. While RSA often identifies the same highly sensitive parameters as ANOVA and Sobol's method, less sensitive parameters were often neglected. This reflects a methodological limitation of the RSA version we used where the qualitative interpretation of sensitivity becomes more challenging for parameters that show modest impacts on model performance. The qualitative nature of the RSA version we used prevents quantitative rankings of parameter sensitivities. Moreover, the qualitative interpretations ignore parameter interactions which further biases RSA to yield smaller "sensitive" parameter sets. The dimensionality of the parameter space being sampled in this study also seems to have a large impact on the RSA sensitivity classifications. As has been shown in numerous prior studies (Duan et al., 1992; Vrugt et al., 2003; Tang et al., 2006, 2007), it is extremely challenging to identify regions of the SAC-SMA

model's response surface that are near optimal or optimal. Although the Latin hypercube sampling used in RSA showed statistical convergence at 10 000 samples, the plots used to classify parameters as being sensitive (e.g., see Figs. 5 and 6) are guaranteed to under represent the true dispersion within their likelihood distributions due to the high-dimension of the sample space (i.e., 18 parameters). The strengths of RSA that have motivated its popularity in the hydrologic literature are its ease-of-implementation, its highly visual results, and its complementary support of uncertainty analysis.

ANOVA and Sobol's method have a clear methodological distinction relative to PEST and RSA in that they readily account for 2nd order parameter interactions. Clearly, parameter interactions contribute to ANOVA and Sobol's method classifying more parameters as being sensitive relative to RSA.

The F-values attained for ANOVA varied significantly in their computed confidence intervals, making the method's ranking of moderately sensitive parameters highly uncertain. We found that Sobol's parameter sensitivity rankings still showed variability, but that the overall rank order did not change as significantly as the ANOVA rankings. As noted by Mokhtari and Frey (2005) and verified in this study, ANOVA is fairly robust for highly nonlinear models with thresholds, despite its normality assumptions. ANOVA's use of the F-distribution in assigning sensitivities has positive and negative impacts on analysis. The F-values attained in ANOVA are easily ranked and sensitivity thresholds can be easily defined by users based on the likelihood of misclassification. Unfortunately, the F-values vary significantly with bootstrapping and sample size, impacting their rank ordering and meaningfulness as sensitivity metrics.

Alternatively, Sobol's indices have very direct and meaningful interpretations in terms of the overall contribution to model output variance. Relative to the other methods tested, Sobol's indices provided the most detailed description of how individual parameters and their interactions impact model performance. Although Sobol's method is the most complex and computationally expensive [requiring  $8192 \times (2 \times 18 + 2) = 311\,296$  model runs] sensitivity analysis technique tested, its robustness, ease-of-interpretation, and detailed results distinguish it among the methods tested in this study. Readers should note that the relatively small computational burden posed by the SAC-SMA/SNOW 17 model allowed us to be very conservative in our choice of sample size for Sobol's method. Generally, the relative ranking of parameter sensitivities was stable for a much smaller sample size. Our use of 8192 samples may be overly rigorous and reflects our minimization of minor numerical errors that did not appreciably change sensitivity classifications.

## 8 Conclusions

This study tested four sensitivity analysis methods: (1) local analysis using parameter estimation software (PEST), (2) regional sensitivity analysis (RSA), (3) analysis of variance (ANOVA), and (4) Sobol's method. These four sensitivity methods were applied to the lumped Sacramento soil moisture accounting model (SAC-SMA) coupled with SNOW-17. Results from this study characterize model sensitivities for the two medium sized watersheds within the Juniata River Basin in Pennsylvania, USA. Comparative results for the 4 sensitivity methods are presented for a 3-year time series with 1 hour, 6 h, and 24 h time intervals.

In this study, we classified the SAC-SMA/SNOW-17 model parameters into three categories: (1) highly sensitive, (2) sensitive, and (3) insensitive. The sensitivity results presented in Sect. 6 show that the fully lumped SAC-SMA/SNOW-17 model's responses are "uniquely" determined by the performance objective specified, prediction timescale, and specific watershed being modeled. Generally, the global methods (RSA, ANOVA, and Sobol's method) agreed on their classifications of the most sensitive parameters for each case. The local method, PEST, generated results that were often significantly different from the global sensitivity analysis methods.

PEST and RSA both neglect parameter interactions and as a consequence yield a far less nuanced description of the models they evaluate. In a broader context, sensitivity analysis shapes the manner in which hydrologists view the processes and watershed properties impacting their model results. The basic assumptions, used in PEST and RSA, such as neglecting parameter interactions, may manifest themselves in the subsequent myriad of potential uses of the hydrologic model (e.g., flood forecasting, observation network design, reservoir management, etc.) by providing an overly simplified view of the controls on a hydrologic system.

Overall ANOVA and Sobol's method were shown to be superior to RSA and PEST. Relative to one another, ANOVA has reduced computational requirements and Sobol's method yielded more robust sensitivity rankings. The results from Sobol's method clearly show that second order parameter interactions explained between 3 to 40% of the SAC-SMA/SNOW-17 model's variance. The implication of this result is that a larger number of parameters and processes within the model are impacting its performance. This study shows that as prediction problems in hydrology grow in complexity, our analysis techniques need to evolve to better represent and understand how models behave.

*Acknowledgements.* This work was partially supported by the National Science Foundation under grant EAR-0310122. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the writers and do not necessarily reflect the views of the National Science Foundation. Partial support for the third author was provided by SAHRA under NSF-STC grant EAR-664 9876800, and the National Weather Service Office of

Hydrology under grant numbers 665 NOAA/NA04NWS4620012 and NOAA/DG 133W-03-666 SE-0916. Support for the fourth author was provided by the Henry Luce Foundation in the form of a Clare Booth Luce Fellowship, a Penn State College of Engineering Fellowship, and a GE Faculty for the Future Fellowship.

Edited by: R. T. Clarke

## References

- Abbaspour, K. C., Schulin, R., and van Genuchten M. Th.: Estimating unsaturated soil hydraulic parameters using ant colony optimization, *Adv. Water Resour.*, 24, 827–841, 2001.
- Anderson, E. A.: National Weather Service River Forecast System - Snow Accumulation and Ablation Model, Tech. rep., NOAA Technical Memorandum NWS Hydro-17, Dept. of Commerce, Silver Spring, Maryland, USA, 1973.
- Anderson, E. A.: Calibration of Conceptual Hydrologic Models for Use in River Forecasting, Tech. rep., NOAA Technical Report, NWS 45, Hydrology Laboratory, Silver Spring, Maryland, USA, 2002.
- Andres, T. H.: Sampling method and sensitivity analysis for large parameter sets, *J. Statist. Comput. Simul.*, 57, 77–110, 1997.
- Andres, T. H. and Wayne, C. H.: Using Iterated Fractional Factorial Design to Screen Parameters in Sensitivity analysis of a Probabilistic Risk Assessment Model, in: Proceedings of the Joint International Conference on Mathematical Methods and Supercomputing in Nuclear Applications, edited by: Kusters, H., Stein, E., and Werner, W., vol. 2, pp. 328–337, 1993.
- Archer, G. E. B., Saltelli, A., and Sobol', I. M.: Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Statist. Comput. Simul.*, 58, 99–120, 1997.
- Beven, K.: Uniqueness of place and process representations in hydrological modeling, *Hydrol. Earth Syst. Sci.*, 4, 203–214, 2000, <http://www.hydrol-earth-syst-sci.net/4/203/2000/>.
- Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249, 11–29, 2001.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S.: *Statistics for Experiments: An Introduction to Design, Data Analysis, and Modeling Building*, Wiley, New York, 1978.
- Boyle, D., Gupta, H., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, 36, 3663–3674, 2000.
- Bratley, P. and Fox, B. L.: Implement Sobol's Quasirandom Sequence Generator, *ACM Transactions on Mathematical Software*, 14, 88–100, 1988.
- Burnash, R. J. C.: The NWS river forecast system-Catchment model, in: *Computer Models of Watershed Hydrology*, edited by Singh, V. P., Water Resour. Publ., Highlands Ranch, CO, 1995.
- Christiaens, K. and Feyen, J.: Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model, *Water Resour. Res.*, 38, 1169, doi:10.1029/2001WR000478, 2002.
- Demaria, E., Nijssen, B., and Wagener, T.: Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model, *J. Geophys. Res. – Atmos.*, in press, 2007.
- Doherty, J.: Groundwater model calibration using pilot points and regularization, *Ground Water*, 41, 170–177, 2003.
- Doherty, J.: *PEST-Model Independent Parameter Estimation User Manual: 5th Edition*, Watermark Numerical Computing, Brisbane, Australia, 2004.
- Doherty, J. and Johnston, J. M.: Methodologies for calibration and predictive analysis of a watershed model, *J. Amer. Water Resour. Assoc.*, 39, 251–265, 2003.
- Duan, Q., Gupta, V. K., and Sorooshian, S.: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031, 1992.
- Duffy, C. J.: A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain, *Water Resour. Res.*, 32, 2421–2434, 1996.
- Duffy, C. J.: Semi-Discrete Dynamical Model for Mountain-Front Recharge and Water Balance Estimation (Rio Grande of Southern Colorado and New Mexico), in: *Groundwater Recharge in a Desert Environment: The Southwestern United States*, edited by: Hogan, J., Phillips, F., and Scanlon, B., Water Science and Application Monograph, 9, American Geophysical Union, pp. 236–255, 2004.
- Efron, B. and Tibshirani, R.: *An introduction to the bootstrap*, Chapman Hall, New York, USA, 1993.
- Fieberg, J. and Jenkins, K. J.: Assessing uncertainty in ecological systems using global sensitivity analyses: a case example of simulated wolf reintroduction effects on elk, *Ecological Modelling*, 187, 259–280, 2005.
- Freer, J., Beven, K. J., and Ambrose, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, 32, 2161–2173, 1996.
- Frey, H. and Patil, S.: Identification and Review of Sensitivity Analysis Methods, *Risk Analysis*, 22, 553–578, 2002.
- Hall, J., Tarantola, S., P.D.Bates, and M.S.Horritt: Distributed sensitivity analysis of flood inundation model calibration, *J. Hydraulic Eng.*, doi:10.1061/(ASCE)0733-9429(2005)131:2(117), 2005.
- Hamby, D.: A review of techniques for sensitivity analysis of environmental models, *Environmental Modelling and Assessment*, 32, 135–154, 1994.
- Helton, J. and Davis, F.: Illustration of sampling-based methods for uncertainty and sensitivity analysis, *Risk Analysis*, 22, 622–691, 2002.
- Helton, J. and Davis, F.: Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering and System Safty*, 81, 23–69, 2003.
- Henderson-Sellers, A., Yang, Z.-L., and Dickinson, R.: The project of intercomparison of land-surface parameterization schemes, *Bull. Am. Meteorol. Soc.*, 74, 1335–1349, 1993.
- Hornberger, G. and Spear, R.: An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, 12, 7–18, 1981.
- Koren, V., S., R., Smith, M., Zhang, Z., and Seo, D. J.: Hydrology laboratory research modeling system (HL-RMS) of the US national weather service, *J. Hydrol.*, 291, 297–318, 2004.
- Kottogoda, N. and Rosso, R.: *Statistics, Probability and Reliability for Civil and Environmental Engineering*, McGraw-Hill, New York, 1997.
- Langbein, W. B.: Overview of Conference on Hydrologic Data Net-

- works, *Water Resour. Res.*, 15, 1867–1871, 1979.
- Lence, B. and Takyi, A.: Data requirements for seasonal discharge programs: an application of a regionalized sensitivity analysis, *Water Resour. Res.*, 28, 1781–1789, 1992.
- Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.*, 2, 164–168, 1944.
- Liang, X. and Guo, J.: Intercomparison of land-surface parameterization schemes: sensitivity of surface energy and water fluxes to model parameters, *J. Hydrol.*, 279, 182–209, 2003.
- Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11, 431–441, 1963.
- Mckay, M., Beckman, R., and Conover, W.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21, 239–245, 1979.
- Misirli, F., Gupta, H. V., Sorooshian, S., and Thiemann, M.: Bayesian recursive estimation of parameter and output uncertainty for watershed models, in: *Calibration of Watershed Models*, edited by: Duan, Q. and Gupta, H. V., Vol. 6, pp. 113–124, American Geophysical Union, Washington, D.C., 2003.
- Mokhtari, A. and Frey, H. C.: Sensitivity Analysis of a Two-Dimensional Probabilistic Risk Assessment Model Using Analysis of Variance, *Risk Analysis*, 25, 1511–1529, 2005.
- Moore, C. and Doherty, J.: Role of the calibration process in reducing model predictive error, *Water Resour. Res.*, 41, W05020, doi:10.1029/2004WR003501, 2005.
- Moreda, F., Koren, V., Zhang, Z., Reed, S., and Smith, M.: Parameterization of distributed hydrological models: learning from the experiences of lumped modeling, *J. Hydrol.*, 320, 218–237, 2006.
- Moss, M. E.: Space, Time, and the Third Dimension (Model Error), *Water Resour. Res.*, 15, 1797–1800, 1979.
- Muleta, M. and Nicklow, J. W.: Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, 306, 1–4, 127–145, 2005.
- Neter, J., Kutner, M., Nachtsheim, C., and Wasserman, W.: *Applied Linear Statistical Models*, 4th ed., McGraw-Hill, Chicago, IL, 1996.
- Oakley, J. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach, *Journal of Royal Statistical Society Series B - statistical Methodology*, 66, 751–769, 2004.
- Osiede, O. and Beck, M. B.: Identification of model structure for aquatic ecosystems using regionalized sensitivity analysis, *Water Sci. Technol.*, 43, 271–278, 2001.
- Panday, S. and Huyakorn, P.: A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow, *Adv. Water Resour.*, 27, 361–382, 2004.
- Pappenberger, F., Iorgulescu, I., and Beven, K. J.: Sensitivity analysis based on regional splits and regression trees (SARS-RT), *Environmental Modelling and Software*, 21, 976–990, 2005.
- Pappenberger, F., Iorgulescu, I., and Beven, K. J.: Sensitivity analysis based on regional splits and regression trees (SARS-RT), *Environmental Modelling & Software*, 21, 976–990, 2006.
- Patil, S. and Frey, H.: Comparison of sensitivity analysis methods based upon applications to a food safety risk model, *Risk Analysis*, 23, 135–154, 2004.
- Peck, E. L.: Catchment modeling and initial parameter estimation for the National Weather Service river forecast system, Tech. rep., Tech. Memo, NWS Hydro-31, Natl. Oceanic and Atmos. Admin., Silver Spring, Maryland, USA, 1976.
- Ratto, M., Young, P. C., Romanowicz, R., Pappenberger, F., Saltelli, A., and Pagano, A.: Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci.*, 3, 3099–3146, 2006, <http://www.hydrol-earth-syst-sci.net/3/3099/2006/>.
- Reed, P., Brooks, R., Davis, K., DeWalle, D. R., Dressler, K. A., Duffy, C. J., Lin, H. S., Miller, D., Najjar, R., Salvage, K. M., Wagener, T., and Yarnal, B.: Bridging River Basin Scales and Processes to Assess Human-Climate Impacts and the Terrestrial Hydrologic System, *Water Resour. Res.*, 42, W07418, doi:10.1029/2005WR004153, 2006.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D., and Participants, D.: Overall distributed model intercomparison project results, *J. Hydrol.*, 298, 27–60, 2004.
- Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications*, 145, 280–297, 2002.
- Saltelli, A., Andres, T. H., and Homma, T.: Sensitivity analysis of model output. Performance of the iterated fractional factorial design method, *Computational Statistics & Data Analysis*, 20, 387–407, 1995.
- Saltelli, A., Tarantola, S., and Chan, K. P.-S.: A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, 41, 39–56, 1999.
- Saltelli, A., Tarantola, S., and Campolongo, F.: Sensitivity Analysis as an Ingredient of Modeling, *Statistical Science*, 15, 377–395, 2000.
- Saltelli, A., Tarantola, A., Campolongo, F., and Ratto, M.: *Sensitivity Analysis in Practice - A Guide to Assessing Scientific Models*, John Wiley and Sons, Chichester, 2004.
- Sieber, A. and Uhlenbrook, S.: Sensitivity analyses of a distributed catchment model to verify the model structure, *J. Hydrol.*, 310, 216–235, 2005.
- Singh, V. and Woolhiser, D.: Mathematical modeling of watershed hydrology, *J. Hydrol. Eng.*, 7, 270–292, 2002.
- Smith, M., Seo, D., Koren, V., Reed, S., Zhang, Z., Duan, Q., Moreda, F., and Cong, S.: The distributed model intercomparison project (DMIP): motivation and experiment design, *J. Hydrol.*, 298, 4–26, 2004.
- Sobol', I.: Sensitivity estimates for nonlinear mathematical models, *Math Model Comput. Exp.*, 1, 407–417, 1993.
- Sobol', I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation*, 55, 271–280, 2001.
- Sobol', I. M.: On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.*, 7, 86–112, 1967.
- Sobol', I. M.: *A primer for the Monte Carlo method*, CRC Press, Boca Raton, Fla., 1994.
- Spear, R., Grieb, T. M., and Shang, N.: Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, 30, 3159–3169, 1994.
- Tang, Y., Reed, P., and Kollat, J.: Parallelization Strategies for Rapid and Robust Evolutionary Multiobjective Optimization in Water Resources Applications, *Adv. Water Resour.*, 30(3), 335–353, 2007.

- Tang, Y., Reed, P., and Wagener, T.: How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration, *Hydrol. Earth Syst. Sci.*, 10, 289–307, 2006, <http://www.hydrol-earth-syst-sci.net/10/289/2006/>.
- Tonkin, M. J. and Doherty, J.: A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, 41, W10412, doi:10.1029/2005WR003995, 2005.
- Vandenberghe, V., Bauwens, W., and Vanrolleghem, P.: Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns, *Environmental Modelling & Software*, 22(5), 725–732, 2007.
- Vrugt, J., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, 39, 1214, doi:10.1029/2002WR001746, 2003.
- Wagener, T. and Kollat, J.: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environmental Modeling and Software*, in press, 2007.
- Wagener, T., Boyle, D., Lees, M. J., Wheater, Gupta, H. S., and H. V. Sorooshian, S.: A framework for development and application of hydrological models, *Hydrol. Earth Syst. Sci.*, 5, 13–26, 2001, <http://www.hydrol-earth-syst-sci.net/5/13/2001/>.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, 17, 455–476, 2003.
- Wagener, T., Wheater, H. S., and Gupta, H. V.: Rainfall-runoff modeling in gauged and ungauged catchments, Imperial College Press, London, UK, 2004.
- Wagener, T., Liu, Y., Gupta, H., Springer, E., and Brookshire, D.: Multi-resolution integrated assessment modeling for water resources management in arid and semi-arid regions, in: Regional hydrologic impacts of climate change - Impact assessment and decision making, edited by: Wagener, T., Franks, S., Bøgh, E., Gupta, H., Bastidas, L., Nobre, C., and Oliveira Galvão, C., IAHS Redbook Publ., 295, pp. 265–272, 2005.
- William, H., Teukolsky, S. A., William, T. V., and Brian, P. F.: Numerical recipes in C (2nd edition), Cambridge University Press, New York, USA, 1999.
- Young, P. C.: A general theory of modelling for badly defined dynamic systems, in: Modeling, Identification and Control in Environmental Systems, edited by: Vansteenkiste, G. C., North Holland, Amsterdam, pp. 103–135, 1978.