



Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis

Y. Tang,¹ P. Reed,¹ K. van Werkhoven,¹ and T. Wagener¹

Received 8 December 2006; revised 30 March 2007; accepted 9 April 2007; published 20 June 2007.

[1] This study provides a step-wise analysis of a conceptual grid-based distributed rainfall-runoff model, the United States National Weather Service (US NWS) Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM). It evaluates model parameter sensitivities for annual, monthly, and event time periods with the intent of elucidating the key parameters impacting the distributed model's forecasts. This study demonstrates a methodology that balances the computational constraints posed by global sensitivity analysis with the need to fully characterize the HL-RDHM's sensitivities. The HL-RDHM's sensitivities were assessed for annual and monthly periods using distributed forcing and identical model parameters for all grid cells at 24-hour and 1-hour model time steps respectively for two case study watersheds within the Juniata River Basin in central Pennsylvania. This study also provides detailed spatial analysis of the HL-RDHM's sensitivities for two flood events based on 1-hour model time steps selected to demonstrate how strongly the spatial heterogeneity of forcing influences the model's spatial sensitivities. Our verification analysis of the sensitivity analysis method demonstrates that the method provides robust sensitivity rankings and that these rankings could be used to significantly reduce the number of parameters that should be considered when calibrating the HL-RDHM. Overall, the sensitivity analysis results reveal that storage variation, spatial trends in forcing, and cell proximity to the gauged watershed outlet are the three primary factors that control the HL-RDHM's behavior.

Citation: Tang, Y., P. Reed, K. van Werkhoven, and T. Wagener (2007), Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resour. Res.*, 43, W06415, doi:10.1029/2006WR005813.

1. Introduction

[2] Over the past decade the increasing availability of spatially distributed hydrometeorological data (e.g., precipitation, air temperature and soil properties) coupled with advances in computational resources has resulted in increasing interest in the development of spatially distributed hydrological models [e.g., *Beven and Kirkby*, 1979; *Abbott et al.*, 1986; *Boyle et al.*, 2001; *Panday and Huyakorn*, 2004; *Duffy*, 2004]. Developers of distributed models seek to better simulate watershed behavior by taking advantage of spatially distributed forcing as well as distributed watershed parameters for a broader array of processes such as surface flow, groundwater flow, sediment transport, and solute transport. The increasing complexity of distributed models poses several challenges in terms of (1) their severe computational demands relative to lumped watershed models [*Apostolopoulos and Georgakakos*, 1997; *van Griensven et al.*, 2006], (2) their potential for overparameterization [*Beven*, 1989], and (3) their high-dimensional, nonlinear parametric spaces and structural uncertainties [*Carpenter et al.*, 2001].

[3] For many distributed hydrologic models, the number of model parameters can range from hundreds to several thousand per watershed model depending on the grid resolution and number of state predictions. These high-dimensional parametric spaces make it extremely difficult to assess the impacts of parameters or combinations of the parameters on watershed model behavior. In an operational context, it becomes more challenging to characterize spatially distributed parameters directly using field measurements, which then yields very high-dimensional model calibration problems [*Doherty*, 2003; *Doherty and Johnston*, 2003; *Madsen*, 2003; *Ajami et al.*, 2004; *Muleta and Nicklow*, 2005; *Tonkin and Doherty*, 2005; *Tang et al.*, 2006a]. Early studies [*Duan et al.*, 1992; *Gupta et al.*, 1998] have highlighted that in the context of optimization, the hydrologic model calibration problem is ill-posed, often highly nonlinear, and multimodal (i.e., numerous local optima exist), especially for high-dimensional parameter spaces.

[4] Model parameter sensitivity analysis has long been recognized as a helpful parameter screening tool that can be used to identify the key parameters controlling model performance [*Young*, 1978; *Hornberger and Spear*, 1981; *Freer et al.*, 1996; *Archer et al.*, 1997; *Saltelli et al.*, 1999; *Carpenter et al.*, 2001; *Sieber and Uhlenbrook*, 2005; *Muleta and Nicklow*, 2005; *van Griensven et al.*, 2006; *Pappenberger et al.*, 2006; *Demaria et al.*, 2007; *Wagener*

¹Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.

and Kollat, 2007]. This approach is particularly important with the current shift toward distributed hydrologic models. Sieber and Uhlenbrook [2005] have highlighted that sensitivity analysis cannot only clarify the most important parameters but also help understand and improve the model structure potentially. However, to date, the computational demands and spatial complexity of distributed hydrologic models have limited our ability to understand their parametric interactions and sensitivities. The limited body of recent literature applying sensitivity analysis to spatially distributed hydrologic models highlights the importance and significant challenges posed by this problem [Carpenter *et al.*, 2001; Doherty, 2003; Tonkin and Doherty, 2005; Sieber and Uhlenbrook, 2005; Muleta and Nicklow, 2005; van Griensven *et al.*, 2006].

[5] Tang *et al.* [2006b] comprehensively compared state-of-the-art sensitivity analysis tools including Sobol's [1993] method, a Jacobian-based local method, regional sensitivity analysis, and Analysis of Variance (ANOVA). Sobol's method was found to be the most effective approach in globally characterizing single parameter and multiparameter interactive sensitivities. Building on this prior study, the present work extends our use of Sobol's global sensitivity analysis method to characterize the spatial and temporal variations of single parameter and multiparameter interactions for the United States National Weather Service's (US NWS) distributed hydrological watershed model termed the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) [Koren *et al.*, 2004; Smith *et al.*, 2004; Reed *et al.*, 2004; Moreda *et al.*, 2006]. This study seeks to carefully characterize the HL-RDHM's sensitivities while balancing the computational demands associated with Sobol's method. Both issues are important to both long-term and short-term forecasts. In the context of long-term forecasts, this study characterizes HL-RDHM's sensitivities over annual time periods using 24-hour model time steps and monthly time periods using 1-hour model time steps. In these cases, computational demands required our use of distributed forcing and lumped model parameterizations. For event-level time periods, a detailed spatial analysis of the HL-RDHM's sensitivities is presented for two events using hourly model time steps. The events were selected to explore how the spatial heterogeneity of forcing impacts the HL-RDHM's spatial sensitivities.

[6] In the remainder of this paper, section 2 provides an overview of the HL-RDHM distributed hydrologic model. Section 3 provides a detailed description of Sobol's method as well as the statistical sampling scheme used in this study. Sections 4 and 5 present the case studies and the computational experiment used to characterize the HL-RDHM, respectively. Section 6 presents the sensitivity analysis results for annual, monthly, and event timescales. Section 7 discusses the implications of the HL-RDHM's temporal and spatial sensitivity trends in terms of their potential value for simplifying model calibration and enhancing operational forecasting.

2. Overview of the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM)

[7] The HL-RDHM was developed by the US NWS [Koren *et al.*, 2004; Smith *et al.*, 2004; Reed *et al.*, 2004;

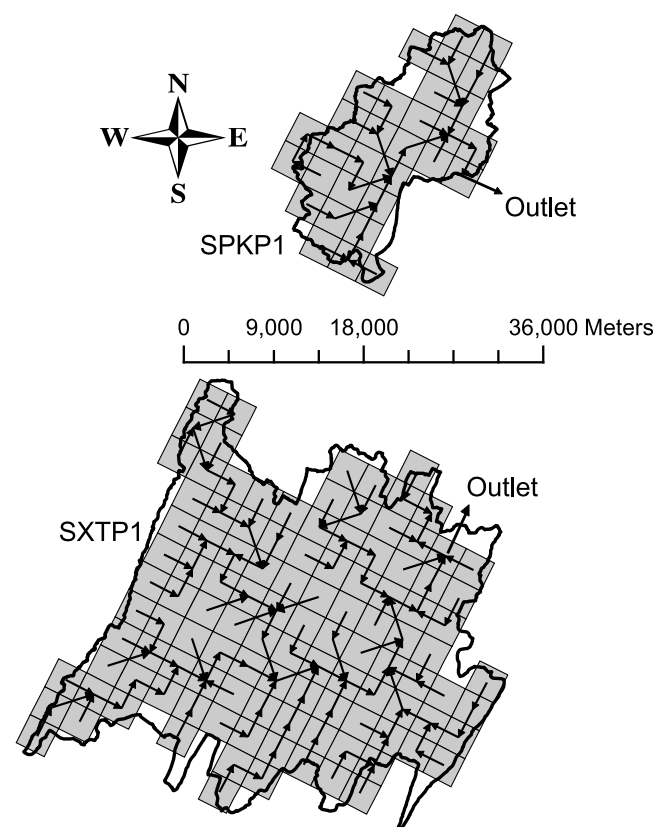


Figure 1. HL-RDHM model grid for the Saxton (SXTP1) and Spruce Creek (SPKP1) watersheds.

Moreda *et al.*, 2006]. It is a flexible modeling framework for building lumped, semidistributed, and fully distributed hydrological models. The structure of the modeling system is based on the Hydrologic Rainfall Analysis Project (HRAP) rectangular grid. The HRAP grid is defined at the 4 km \times 4 km resolution that corresponds directly to the US NWS' Next Generation Weather Radar (NEXRAD) precipitation products. Figure 1 presents the HRAP grid cells for the Saxton (SXTP1) and Spruce Creek (SPKP1) headwaters for the Juniata river in Pennsylvania. SXTP1 and SPKP1 are the case study watersheds used to evaluate the HL-RDHM in this study. The arrows in the grid indicate the direction of surface flows. More information about HRAP gridded NEXRAD data are given by Reed and Maidment [1999].

[8] For each HL-RDHM grid cell, a snow model, a rainfall-runoff model, and hillslope and channel routing models are used to simulate the rainfall-runoff processes. Fast model responses such as overland flow and direct runoff on impervious area are routed from the hillslope and drained into a conceptual channel which has the length of the cell diagonal distance. The intercell channel routing is conducted by using a connectivity file which reflects the surface flow directions (Figure 1). A modified version of the algorithm developed by Wang *et al.* [2000] is used to generate the connectivity file [Koren *et al.*, 2004]. Essentially, the fine-resolution Digital Elevation Model (DEM) cells were aggregated into the coarser HRAP grid cells. In the algorithm, the DEM defined flowpath determines a grid cell's flow direction. The algorithm constrains the grid flow

Table 1. 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/degc/6hr	maximum melt factor during nonrain periods	0.5–1.2
	MFMIN	mm/degc/6hr	maximum melt factor during nonrain periods	0.1–0.6
	UADJ	mm/mb	average wind function during rain-on-snow periods	0.02–0.2
SAC-SMA	SI	mm	mean water-equivalent above which 100% cover exists	10–120
	UZTWM	mm	upper zone tension water maximum storage	25.0–125.0
	UZFWM	mm	upper zone free water maximum storage	10.0–75.0
	UZK	day ⁻¹	upper zone free water lateral depletion rate	0.2–0.5
	PCTIM		impervious fraction of the watershed area	0.0–0.01
	ADIMP		additional impervious area	0.0–0.2
	ZPERC		maximum percolation rate	20.0–300.0
	REXP		exponent of the percolation equation	1.4–3.5
	LZTWM	mm	lower zone tension water maximum storage	75.0–300.0
	LZFSM	mm	lower zone free water supplemental maximum storage	15.0–300.0
	LZFPM	mm	lower zone free water primary maximum storage	40.0–600.0
	LZSK	day ⁻¹	lower zone supplemental free water depletion rate	0.03–0.2
	LZPK	day ⁻¹	lower zone primary free water depletion rate	0.001–0.015
	PFREE		fraction of water percolating from upper zone directly to lower zone free water storage	0.0–0.5

directions to closely match the flow pattern that would be predicted using the high-resolution DEM. The slow model responses such as interflow and baseflow go straight into the conceptual channel without going through hillslope routing.

[9] In the version of the HL-RDHM used in this study, SNOW-17 [Anderson, 1973] is used to simulate the energy balance of a snowpack using a temperature index method. Two models, the Sacramento Soil Moisture Accounting (SAC-SMA) model developed by Burnash [1995] and the Continuous Antecedent Precipitation Index (CONT-API) developed by Anderson [1994], are available for rainfall-runoff modeling. In this study, the SAC-SMA model is adopted because it is widely used by the river forecast centers of the US NWS [Smith et al., 2004; Reed et al., 2004; Moreda et al., 2006]. Hillslope and channel routing processes are modeled within HL-RDHM using a kinematic wave approximation to the St. Venant equations. The feasible ranges of the parameters listed in Table 1 are based on the recommendations of Anderson [2002].

3. Sobol's Sensitivity Analysis

3.1. Sobol's Method

[10] Sobol's method is a variance-based sensitivity analysis approach that represents a model in the following functional form:

$$\mathbf{y} = f(\mathbf{x}, \Theta), \quad (1)$$

where \mathbf{y} is the model outputs, \mathbf{x} is the input state variables, and Θ is the parameter set. Sobol's method evaluates parametric sensitivity by evaluating the variance of \mathbf{y} due to changes of parameter vector Θ . As described by Sobol' [1993], the total variance of the model output is decomposed into component variances that result from individual parameters and parameter interactions. Typically, the direct model output \mathbf{y} is replaced by a model performance measure such as the root mean square error (RMSE) of the streamflow model predictions. Single parameters or para-

meter interactions are then evaluated according to their percentage contribution to the total variance of model responses. The Sobol's variance decomposition shown in equation (2) assumes that the parameters are independent:

$$D(\mathbf{y}) = \sum_i D_i + \sum_{i < j} D_{ij} + \sum_{i < j < k} D_{ijk} + D_{12\dots m}, \quad (2)$$

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 Θ , D_{ij} is the portion of output variance due to the interaction of parameters θ_i and θ_j . The variable m stands for the total number of parameters. Sobol's sensitivity indices are computed using the following equations derived from equation (2).

$$\text{first-order } S_i = \frac{D_i}{D} \quad (3)$$

$$\text{total-order } S_{T_i} = 1 - \frac{D_{\sim i}}{D}, \quad (4)$$

where S_i indicates the sensitivity that results from the main effect of parameter θ_i . The average variance, $D_{\sim i}$, results from all of the parameters except for θ_i . The total-order sensitivity, S_{T_i} , defines the independent and interactive effects up to the m th order of parameter θ_i . A parameter mainly influences the model output by parameter interactions if it has a small first-order index and a large total-order sensitivity index.

[11] When implementing Sobol's method, the parameter ranges are scaled to be between 0 and 1. Equation (1) is partitioned as follows:

$$f(\theta_1, \dots, \theta_m) = f_0 + \sum_i^m f_i(\theta_i) + \sum_{1 \leq i < j \leq m} f_{ij}(\theta_i, \theta_j) + \dots + f_{1,2,\dots,m}(\theta_1, \dots, \theta_m) \quad (5)$$

and the normalized total variance D can be evaluated by equation (6),

$$D = \int_0^1 \cdots \int_0^1 f(\Theta)d\Theta - f_0^2 \quad (6)$$

and the variances D_{i_1, \dots, i_s} are

$$D_{i_1, \dots, i_s} = \int_0^1 \cdots \int_0^1 f_{i_1, \dots, i_s}^2(\theta_{i_1}, \dots, \theta_{i_s})d\theta_{i_1}d\theta_{i_s} \quad (7)$$

$$1 \leq i_1 < \cdots < i_s \leq m, \quad s = 1, \dots, m. \quad (8)$$

When the model is highly nonlinear and complex, Monte Carlo numerical integration is the most suitable method for evaluating the integrals represented in equations (6)–(8) above. The Monte Carlo approximations for D , D_i , and $D_{\sim i}$ are given in the following equations as presented in prior studies by Sobol' [1993, 2001] and Hall et al. [2005]:

$$\hat{f}_0 = \frac{1}{n} \sum_{s=1}^n f(\Theta_s) \quad (9)$$

$$\hat{D} = \frac{1}{n} \sum_{s=1}^n f^2(\Theta_s) - \hat{f}_0^2 \quad (10)$$

$$\hat{D}_i = \frac{1}{n} \sum_{s=1}^n f(\Theta_s^{(a)})f(\Theta_{(\sim i)s}^{(b)}, \Theta_{is}^{(a)}) - \hat{f}_0^2 \quad (11)$$

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

In equations (9)–(12), the variable n defines the Monte Carlo sample size, Θ_s represents the sampled individual in the scaled unit hypercube, and (a) and (b) are two different samples. The parameters with values drawn from sample (a) are denoted by $\Theta_s^{(a)}$. The variables $\Theta_{is}^{(a)}$ and $\Theta_{is}^{(b)}$ designate that parameter θ_i draws values from sample (a) and (b), respectively. The symbols $\Theta_{(\sim i)s}^{(a)}$ and $\Theta_{(\sim i)s}^{(b)}$ define cases when parameter θ_i is not sampled and the remaining parameters do get their sample values from samples (a) and (b), respectively.

[12] Although Sobol's method can identify important parameter interactions, the method becomes computationally expensive when high-order interactions (or indices) must be calculated for models with large parameter sets (e.g., distributed hydrologic models). The original Sobol's method [Sobol', 1993] required $n \times (2m + 1)$ model runs to calculate all of the first-order and the total-order sensitivity indices. Recall that n is the number of Monte Carlo samples and m is the number of model parameters. An enhancement of the method made by Saltelli [2002] provides the first, second, $(m - 2)$ th, and total-order sensitivity indices using $n \times (2m + 2)$ model runs. Saltelli [2002] provides an alternate method to calculate the first, $(m - 2)$ th, and total order sensitivity at a reduced cost of $n \times (m + 2)$ model

simulations. For this study, the latter method is desirable because it not only reduces the run time for analysis but it also provides sufficient information regarding main effects and parameter interactions. Readers interested in more details in the numerical implementation of Sobol's method should reference the following studies [Saltelli, 2002; Sobol', 2001].

3.2. Latin Hypercube Sampling (LHS)

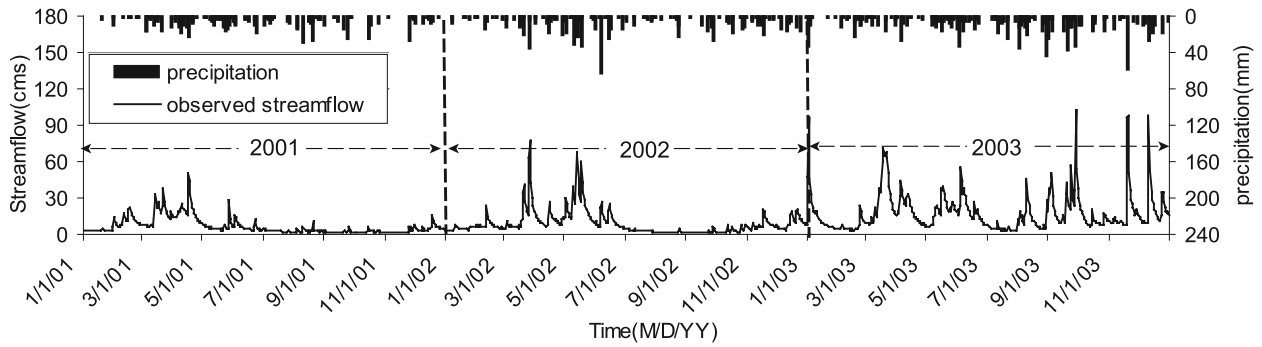
[13] In this study, Latin hypercube sampling (LHS) was used to sample the feasible parameter space because it was found to be effective in prior studies [e.g., Osidele and Beck, 2001; Sieber and Uhlenbrook, 2005; Tang et al., 2006b]. LHS integrates the strength of random sampling and stratified sampling [McKay et al., 1979; Helton and Davis, 2003] to make sure that all portions of the parameter space are considered. The method divides the m parameters' ranges into n disjoint intervals with equal probability $1/n$ from which one value is sampled randomly in each interval. To create one sample, a value of one parameter from a specific interval is picked and combined randomly with a value from another parameter from an interval, then this pair of values are combined with a value of the third parameter and the process repeated till each parameter has a value in the combined parameter set. The values which have been picked to form an individual parameter draw do not participate in generating other individuals (i.e., sampling without replacement). After all values of the variables have been chosen to create individuals, a sample of size n is created. The process can be repeated r times by so that a sample of total size $r \times n$ is created. A major benefit of the LHS method is its ability to divide parameter spaces into hypercubes to ensure a well spread parameter sample. More details about LHS are available in the following papers [McKay et al., 1979; Helton and Davis, 2003; Press et al., 1999]. Readers should note that the traditional implementation of LHS used in this study assumes that parameters are independent. This assumption is very common in hydrologic sensitivity analysis because in many cases it is not possible to a priori specify a covariance matrix for hydrologic model parameters. In cases where there are large numbers of spatially correlated observations for a model parameter (e.g., groundwater hydraulic conductivities) LHS can be modified to account parameter covariance [see Zhang and Pinder, 2003; Hall et al., 2005].

4. Case Study

4.1. Juniata Watershed Description

[14] Two headwater watersheds SPKP1 (drainage area 570 km², mean elevation 485 m) and SXTPI (drainage area 1960 km², mean elevation 457 m) contributing to the Juniata River where used as test cases in this study. The watershed boundaries and the associated HL-RDHM model grids are shown in Figure 1. The Juniata River Basin has a drainage area of 8800 km² in south central Pennsylvania, and is a major tributary to the Susquehanna River. As highlighted by Tang et al. [2006b], the SPKP1 and SXTPI watersheds have different hydrologic conditions and basin characteristics which define their response behavior [also see Seo et al., 2006; Moreda et al., 2006; Tang et al., 2006b].

(a) SPKP1



(b) SXTP1

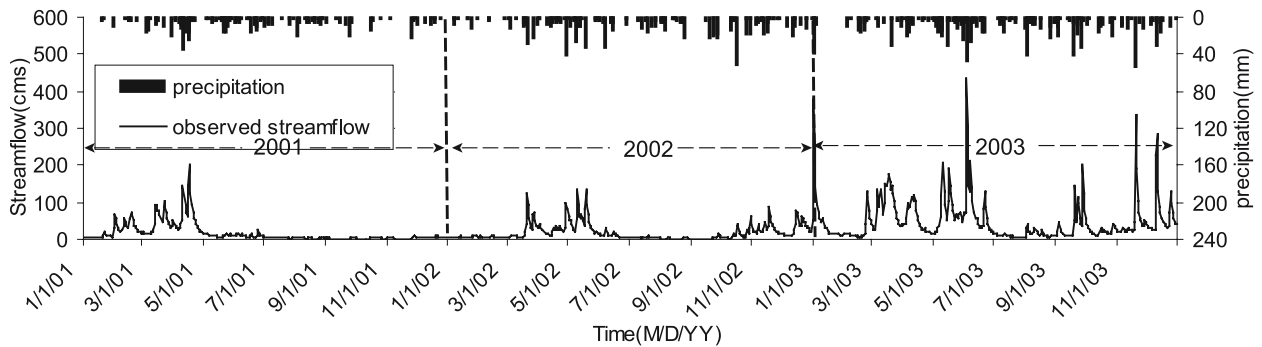


Figure 2. Hydrographs for the two studied watersheds from year 2001 to year 2003. (a) Hydrograph for SPKP1. (b) Hydrograph for SXTP1.

4.2. Data Set

[15] The input forcing data of the HL-RDHM consist of precipitation, monthly potential evapotranspiration (PE), and air temperature. The precipitation data are based on the NEXRAD multisensor precipitation estimator data developed on a HRAP grid with a 4 km × 4 km resolution as shown in Figure 1. The selected data set for precipitation and air temperature were attained from the US NWS mid-Atlantic river forecast center (MARFC) and are defined for an hourly time interval over the three year period spanned 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 SXTP1 watersheds. The time series of the 3 years’ precipitation and observed streamflow are presented in Figure 2. The statistics of the precipitation and streamflow presented in Table 2 show that 2001 was the driest year and 2003 was the wettest year among the three years. Both years 2001 and 2002 had a similar rainfall-runoff ratio. Table 3 illustrates the percentages of time when

air temperature is below zero and precipitation is larger than zero based on the mean area precipitation and temperature. It reveals that frozen conditions generally occur from October to December and from January to April.

5. Computational Experiment

5.1. Test Cases and Parameterization

[16] Computation time poses a severe constraint for analyzing the HL-RDHM’s sensitivity. Consequently, we have carefully designed our analysis as a step-wise progression from long-term to short-term time periods to maintain the computational tractability of Sobol’s method. Three test cases were configured on the basis of these considerations: (1) annual sensitivity analysis using 3 years of observations and precipitation data modeled using daily time steps for the SPKP1 and SXTP1 watersheds, (2) monthly sensitivity analysis using 3 years observation and precipitation data modeled using hourly time steps for the SPKP1 and SXTP1 watersheds, and (3) event sensitivity analysis using two

Table 2. Statistics of the Precipitation and Streamflow Data From Year 2001 to Year 2003 and Long-Term Averages

	SPKP1				SXTP1			
	2001	2002	2003	Average	2001	2002	2003	Average
Precipitation, mm	794.46	947.69	1251.14	971	711.17	943.12	1243.37	946
Runoff, mm	377.45	496.93	988.86	660	258.30	286.69	814.71	438
Runoff/precipitation	0.48	0.52	0.79	0.68	0.36	0.30	0.66	0.46

Table 3. Percentage of Time When Air Temperature is Below Zero and Precipitation is Larger Than Zero^a

Month	Time Percentage as $T_{air} < 0$ (%)						Time Percentage as $T_{air} < 0$ and $P > 0$ (%)					
	SPKP1			SXTP1			SPKP1			SXTP1		
	2001	2002	2003	2001	2002	2003	2001	2002	2003	2001	2002	2003
1	79.8	49.1	88.8	81.5	50.7	90.3	5.6	5	10.2	9.1	5.2	12.2
2	58	56.4	87.5	58.9	51	92	3.4	6.5	20.2	4.2	2.8	21
3	48.3	28.9	37.9	50.7	30.5	38.7	9.1	4.4	7	8.6	4.6	5.6
4	10	10.8	10.7	10	10	9.9	2.1	0.6	2.5	2.1	0.8	2.2
5	0	4	0	0	4.8	0	0	0	0	0	0.5	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	4.8	3.4	2.4	7.3	5	9.7	0	1.3	0.3	0	1.2	0.3
11	11.7	23.3	17.5	15	27.5	22.6	0.6	4.6	1.8	0.6	6.1	3.3
12	44.5	74.1	69.5	46.1	73.4	75	2.6	12.2	13.3	2	12.6	13.7

^a T_{air} denotes air temperature and P defines precipitation. The statistics are based on hourly mean area precipitation and air temperature.

events with hourly observed data and model time steps for the SPKP1 watershed. For the first 2 test cases, the parameters of the HL-RDHM were spatially lumped (i.e., every grid cell takes the same value for a specific parameter) in order to make Sobol's method computationally tractable. For test case (3), the analysis was conducted using spatially distributed parameters (i.e., the parameter value varied from one model cell to another for a specific parameter). In test case (3), only 13 SAC-SMA parameters were analyzed because no snow occurred during these two events. For all of the test cases, the precipitation and the air temperature were spatially distributed. The hydrographs of the two selected events for test case (3) are presented in Figure 3.

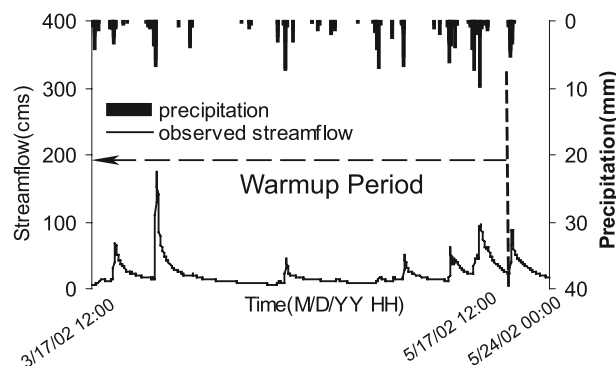
[17] The period between 1 January and 28 February 2001 were used as the warmup periods for the first 2 test cases to limit the impact of initial conditions. Similarly, a 2 month warmup period was used for both the May 2002 event and the September 2003 events as shown in Figure 3. The a priori parameter settings used for the SNOW-17, SAC-SMA, and routing models were set on the basis of the recommendations of the MARFC of the US NWS.

5.2. Sensitivity Analysis Implementation

[18] Statistical sample size is a key parameter for Sobol's method. In this study, the sample sizes were configured on

the basis of both literature recommendations and experiments that tested the convergence and reproducibility of the sensitivity analysis results. In this study, LHS replaced the Sobol's quasirandom sequence sampling [Sobol', 1967] used in our prior study [Tang et al., 2006b] because Sobol's sequence sampling can sample a maximum of 100 parameters, which is insufficient given that test case (3) has 403 parameters (31 cells \times 13 parameters). In our prior study, an extremely conservative sample size of 8192 was used. However, in this study, we cannot afford as many model runs as the prior study because the model execution time is significantly higher when switching from a lumped to a distributed model. A LHS sample size of 2000 was used in this study for all three test cases resulting in $2000 \times (18 + 2) = 40,000$ model runs per watershed for the first 2 test cases and $2000 \times (403 + 2) = 810,000$ model runs per event model runs for the third test case. The results of section 6.4 confirm that this sample size is sufficient to maintain the accuracy and repeatability of Sobol's method. The most computationally intensive cases were the monthly analysis in the SXTP1 watershed (113 cells) and the single event (May 2002 or September 2003) analysis in SPKP1 (31 cells). Their estimated evaluation times are about 28 days and 16 days respectively on a single processor, which demon-

(a) 2002.5 Event in SPKP1



(b) 2003.9 Event in SPKP1

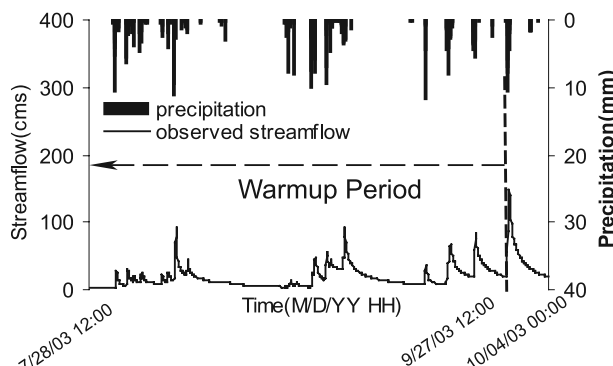


Figure 3. Hydrographs for the two analyzed events in SPKP1. (a) Hydrograph for May 2002 event. (b) Hydrograph for September 2003 event.

Table 4. Annual First-Order Sensitivity Indices From Sobol's Method Computed Using the RMSE Measure and 24-Hour Model Time Steps^a

Model	Parameter	SPKP1			SXTP1		
		2001	2002	2003	2001	2002	2003
SNOW-17	SCF	0.00 [0.01]	0.00 [0.00]	<i>0.01 [0.01]</i>	0.00 [0.01]	0.00 [0.00]	0.00 [0.01]
	MFMAX	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>	0.00 ^b [0.00]	0.00 ^b [0.00]	<i>0.01 [0.01]</i>
	MFMIN	0.00 [0.00]	0.00 [0.00]	<i>0.02 [0.01]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.03 [0.01]</i>
	UADJ	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 ^b [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]
	UZTWM	0.19 [0.03]	<i>0.06 [0.02]</i>	<i>0.01 [0.01]</i>	0.15 [0.02]	<i>0.05 [0.01]</i>	<i>0.01 [0.01]</i>
	UZFWM	0.00 [0.00]	0.00 [0.01]	<i>0.01 [0.01]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>
	UZK	0.00 [0.00]	0.00 [0.00]	0.00 [0.01]	0.00 [0.00]	0.00 [0.00]	0.00 [0.01]
	PCTIM	0.16 [0.02]	<i>0.04 [0.01]</i>	<i>0.08 [0.02]</i>	0.13 [0.02]	<i>0.06 [0.01]</i>	<i>0.08 [0.02]</i>
	ADIMP	0.00 [0.01]	<i>0.07 [0.02]</i>	0.20 [0.03]	<i>0.01 [0.01]</i>	<i>0.06 [0.02]</i>	0.25 [0.03]
ZPERC	0.00 [0.00]	0.00 [0.01]	<i>0.02 [0.01]</i>	0.00 [0.01]	0.00 [0.01]	<i>0.03 [0.02]</i>	
SAC-SMA	REXP	0.00 ^b [0.00]	0.00 ^b [0.01]	0.00 ^b [0.01]	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 ^b [0.01]
	LZTWM	<i>0.06 [0.02]</i>	0.49 [0.04]	0.17 [0.03]	<i>0.07 [0.03]</i>	0.48 [0.04]	0.14 [0.02]
	LZFSM	<i>0.03 [0.01]</i>	<i>0.02 [0.01]</i>	0.02 [0.03]	<i>0.04 [0.02]</i>	<i>0.02 [0.01]</i>	<i>0.05 [0.03]</i>
	LZFPM	<i>0.04 [0.02]</i>	0.13 [0.03]	0.15 [0.03]	<i>0.08 [0.02]</i>	<i>0.08 [0.02]</i>	<i>0.06 [0.03]</i>
	LZSK	<i>0.01 [0.01]</i>	<i>0.01 [0.01]</i>	0.02 [0.02]	<i>0.02 [0.01]</i>	<i>0.01 [0.01]</i>	0.00 ^b [0.02]
	LZPK	<i>0.02 [0.01]</i>	<i>0.02 [0.01]</i>	0.00 [0.01]	<i>0.01 [0.01]</i>	<i>0.03 [0.01]</i>	<i>0.01 [0.02]</i>
	PFREE	0.34 [0.04]	<i>0.02 [0.01]</i>	0.00 [0.01]	0.30 [0.04]	<i>0.07 [0.02]</i>	0.00 [0.01]

^aBoldface designates highly sensitive parameters defined using a threshold value of 0.1. Italics designate 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).

^bNegative mean value was set to zero.

strates the importance of using parallel computing to support this analysis.

[19] The bootstrap method [Efron and Tibshirani, 1993] was used to provide confidence intervals for the parameter sensitivity rankings for the Sobol's method. Essentially, the samples generated by LHS were resampled N times when calculating the sensitivity indices for each parameter, resulting in a distribution of the indices. The moment method [Archer et al., 1997] was used for attaining the bootstrap confidence intervals (BCIs). 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 on the basis of prior literature discussions and our prior study [Tang et al., 2006b]. Readers interested in detailed descriptions of the bootstrapping method used in this paper can check the referenced sources [Archer et al., 1997; Efron and Tibshirani, 1993].

6. Results

[20] Sections 6.1–6.4 present a step-wise analysis of the HL-RDHM's sensitivities from annual to event level time-scales with the intent of elucidating the key parameters impacting the model's forecasts. In the context of long-term forecasts, sections 6.1 and 6.2 focus on HL-RDHM's sensitivities for annual and monthly periods where computational demands required our use of distributed forcing and lumped model parameterizations for the SPKP1 and SXTP1 watersheds. Alternatively, section 6.3 provides a detailed spatial analysis of the HL-RDHM's sensitivities for the SPKP1 watershed for two events. These events were selected to explore how the spatial heterogeneity of forcing impacts the model's spatial sensitivities. Section 6.4 continues our

spatial analysis of event sensitivities and evaluates how well Sobol's sensitivity method performs in identifying the principle input parameters controlling the HL-RDHM's response. Our evaluation of Sobol's method extends the SA repeatability test recommended by Andres [1997] to a spatially distributed modeling context.

6.1. Annual Sensitivities Based on Distributed Forcing and Lumped Parameters

[21] In this section, the computational demands associated with Sobol's method were made tractable by lumping the HL-RDHM's parameters (i.e., all cells had the same parameter values) while maintaining spatially distributed forcing and model structure. The first-order and total-order Sobol's indices are reported in Tables 4 and 5, respectively, for the 18 parameters analyzed. Recall that the first-order indices measure single parameter contributions to the HL-RDHM's output variance, whereas the total-order indices also include the influence of parameter interactions. 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. The years 2001–2003 analyzed in Tables 4 and 5 capture a gradient from the end of a regional drought in 2001 through a transition year in 2002 to a wet year in 2003 with increased rainfall and snowfall. This gradient is pronounced for the SNOW-17 parameters where 2003 is the only year when they influence the HL-RDHM results by contributing

Table 5. Annual Total-Order Sensitivity Indices From Sobol'’s Method Computed Using the RMSE Measure and 24-Hour Model Time Steps^a

Model	Parameter	SPKP1			SXTPI		
		2001	2002	2003	2001	2002	2003
SNOW-17	SCF	<i>0.01 [0.01]</i>	0.00 [0.00]	<i>0.03 [0.01]</i>	0.00 [0.01]	0.00 [0.00]	<i>0.02 [0.01]</i>
	MFMAX	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>
	MFMIN	0.00 [0.00]	0.00 [0.00]	<i>0.03 [0.01]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.03 [0.01]</i>
	UADJ	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.00]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.00]</i>
	SI	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 [0.00]	0.00 ^b [0.00]	0.00 ^b [0.00]	0.00 [0.00]
SAC-SMA	UZTWM	0.22 [0.03]	<i>0.07 [0.02]</i>	<i>0.01 [0.01]</i>	0.18 [0.03]	<i>0.06 [0.01]</i>	<i>0.01 [0.01]</i>
	UZFWM	0.00 [0.00]	<i>0.850.01 [0.01]</i>	<i>0.02 [0.01]</i>	0.00 [0.01]	0.00 [0.00]	<i>0.02 [0.01]</i>
	UZK	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>	0.00 [0.00]	0.00 [0.00]	<i>0.01 [0.01]</i>
	PCTIM	0.16 [0.02]	<i>0.04 [0.01]</i>	<i>0.08 [0.02]</i>	0.14 [0.02]	<i>0.06 [0.01]</i>	<i>0.08 [0.02]</i>
	ADIMP	<i>0.02 [0.01]</i>	<i>0.07 [0.02]</i>	0.21 [0.03]	<i>0.03 [0.01]</i>	<i>0.06 [0.02]</i>	0.26 [0.03]
	ZPERC	<i>0.01 [0.01]</i>	<i>0.02 [0.01]</i>	<i>0.05 [0.02]</i>	<i>0.03 [0.02]</i>	<i>0.02 [0.01]</i>	<i>0.06 [0.02]</i>
	REXP	0.00 [0.00]	<i>0.01 [0.01]</i>	<i>0.05 [0.01]</i>	0.00 [0.00]	<i>0.01 [0.00]</i>	<i>0.05 [0.01]</i>
	LZTWM	0.11 [0.02]	0.57 [0.04]	0.21 [0.02]	0.13 [0.03]	0.60 [0.04]	0.18 [0.02]
	LZFSM	<i>0.04 [0.02]</i>	<i>0.04 [0.01]</i>	0.15 [0.03]	<i>0.07 [0.02]</i>	<i>0.03 [0.01]</i>	0.20 [0.03]
	LZFPM	<i>0.04 [0.02]</i>	0.15 [0.03]	0.22 [0.03]	<i>0.08 [0.02]</i>	0.10 [0.02]	0.19 [0.03]
	LZSK	<i>0.02 [0.01]</i>	<i>0.03 [0.01]</i>	<i>0.08 [0.02]</i>	<i>0.05 [0.02]</i>	<i>0.02 [0.01]</i>	<i>0.08 [0.02]</i>
	LZPK	<i>0.03 [0.01]</i>	<i>0.03 [0.01]</i>	<i>0.04 [0.02]</i>	<i>0.02 [0.01]</i>	<i>0.03 [0.01]</i>	<i>0.05 [0.02]</i>
	PFREE	0.40 [0.04]	<i>0.03 [0.01]</i>	<i>0.01 [0.01]</i>	0.36 [0.04]	0.12 [0.02]	<i>0.01 [0.01]</i>

^aBoldface designates highly sensitive parameters defined using a threshold value of 0.1. Italics designate 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).

^bNegative mean value was set to zero.

approximately 4 percent of the model output variance for both watersheds. This result makes sense given the significant regional snowfall in 2003 (see Table 3).

[22] It is interesting to note the similarity of the sensitivity results for the two watersheds for every year analyzed. This actually differs with the fully lumped SAC-SMA/SNOW-17 sensitivity analysis of *Tang et al.* [2006b] for the same two watersheds and the same set of analyzed parameters. *Tang et al.* [2006b] concluded that each watershed had a unique set of sensitivities for the fully lumped SAC-SMA/SNOW-17 model. In contrast, the distributed forcing/lumped parameter results in Tables 4 and 5 show nearly identical sensitivity classifications for the two watersheds for each given year when spatially distributed forcing is considered.

[23] The most significant changes in sensitivity for the watersheds occurred in the transition from dry conditions to wet conditions. Under the dry conditions of 2001, the HL-RDHM's upper zone storages had to be filled before significant interflow and/or subsurface flow was possible. The first-order and total-order indices in Tables 4 and 5 show that the upper zone tension water storage (UZTWM), impervious cover (PCTIM), and the fraction of water percolating from the upper to the lower zone (PFREE) control 60 to 70 percent of the HL-RDHM's response. As would be expected the fraction of percolating water (PFREE) is highly interactive during the dry conditions of 2001 with approximately 6 percent of its influence on model output coming from interactions with other parameters.

[24] In the 2002 transitional year, Tables 4 and 5 show a strong shift in parametric sensitivity from the upper zone storage to the lower zone storage. In the transitional year, the maximum storage in the lower zone tension water (LZTWM) and free water (LZFPM) influence about 60 percent of the HL-RDHM's response. Comparing Tables 4

and 5 for 2002, shows that LZTWM is highly interactive and influences an additional 10 percent of the model's variance with its interactions with other parameters. Conceptually, this makes sense given that 2002 is in general a wetting period for the subsurface in both watersheds.

[25] The wet conditions in 2003 shift HL-RDHM's sensitivities to being nearly equally distributed between all of the lower zone storage parameters (LZTWM, LZFSM, LZFPM) in both watersheds. Additionally, the wet conditions of 2003 also increased the incidence of surface saturation, which is reflected in the model's sensitivity to additional impervious area (ADIMP). Overall in 2003, LZTWM, LZFSM, LZFPM, and ADIMP explain approximately 80 percent of the HL-RDHM's output variance. The next section further elucidates HL-RDHM's changes in sensitivity across the transition from dry to wet conditions using a per-month analysis.

6.2. Monthly Sensitivities Based on Distributed Forcing and Lumped Parameters

[26] The results shown in Figure 4 provide a more detailed description of the temporal trends in the HL-RDHM's sensitivities. These plots classify highly sensitive and sensitive parameters on a monthly basis using lumped parameters and distributed forcing at an hourly time step. The magnitude and ranges of Sobol's indices are shown by the color legends. In the prior section, the transition from dry conditions in 2001 to wet conditions in 2003 resulted in a shift in the dominant parameters controlling the HL-RDHM response from the upper zone storage (UZTWM), impervious cover (PCTIM), and percolating water (PFREE) to lower zone storages (LZTWM, LZFSM, LZFPM). Figure 4 shows that these parameters' are dominant in a year when they are classified as highly sensitive over a majority of its component months. Although this result is expected, it is interesting to note the temporal transitions in

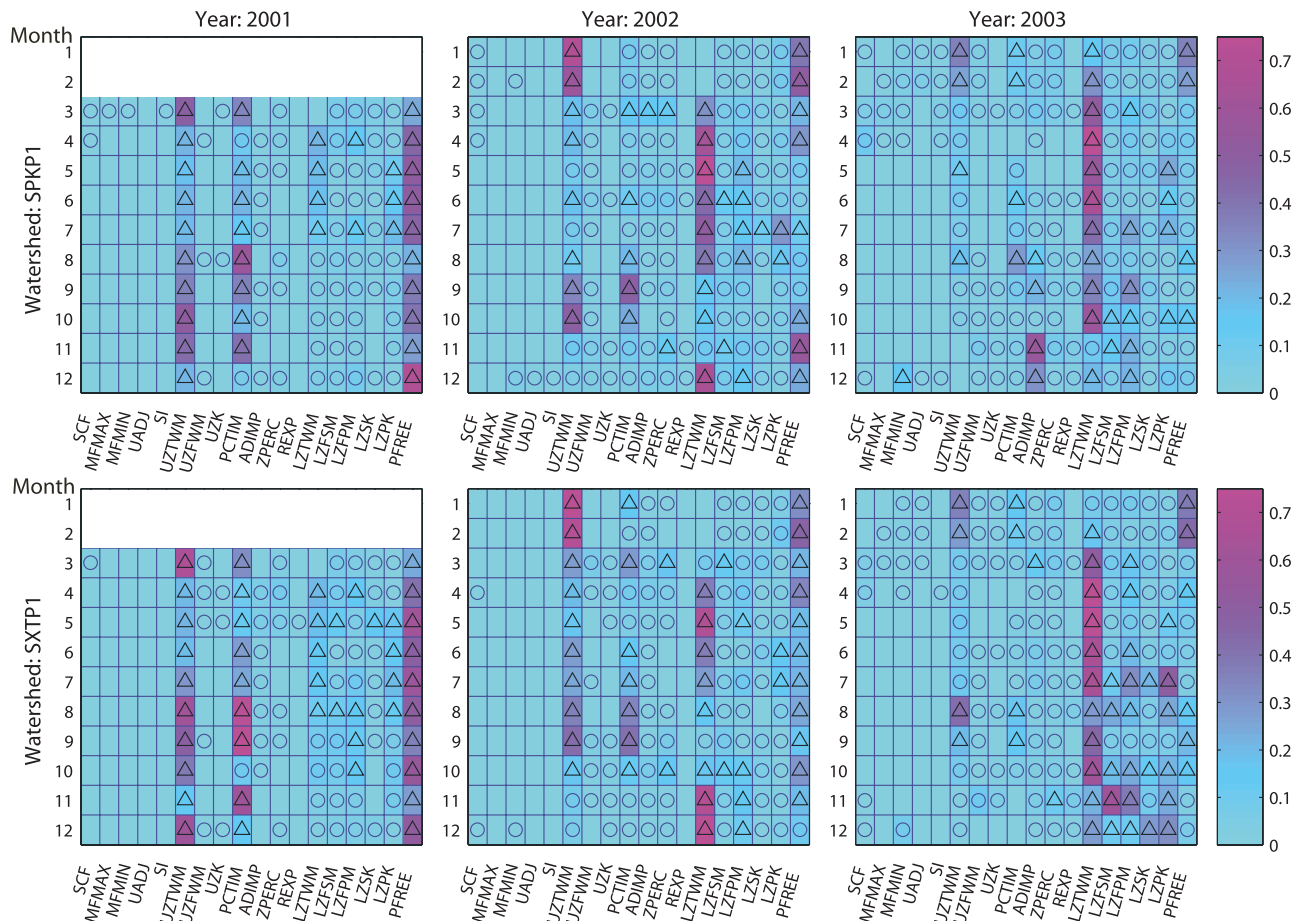


Figure 4. Monthly HL-RDHM sensitivities using Sobol’s total-order indices. Sobol’s indices were computed using the RMSE measure and an hourly model time step. Triangles represent highly sensitive parameters that contribute at least 10% of the overall model output variance. Circles represent sensitive parameters that contribute at least 1% of the overall model output variance. The color legends and shading represent the Sobol’ indices’ magnitudes and ranges. Each row represents one month and each column represents one parameter. January and February of 2001 are missing because they represent the model warm-up period.

sensitivity from the dry conditions in 2001 to the wet conditions in 2003.

[27] For example, note that the upper zone storage (UZTWM) strongly influences the HL-RDHM response from March 2001 to February 2002. From January 2002 to March 2002, UZTWM’s total-order indices (see Figure 4) decrease from approximately 0.6 in both watersheds to approximately 0.2 (a three-fold decrease). A similar trend exists for water percolating from the model’s upper zone to its lower zone (i.e., PFREE). Note in March 2002 and continuing through 2003 that as UZTWM’s influence decreases, the lower zone tension water storage (LZTWM) indices increase from smaller than 0.01 to greater than 0.6 (i.e., it controls 60 percent of the model’s variance).

[28] Figure 4 also shows how seasonal trends impact predictions in the summer and winter months. As an example UZTWM’s influence in the summer months shows a seasonal increasing trend until the transition to fall. This model sensitivity trend reflects the expected impact of reduced precipitation and the resultant reduction in shallow storage within both watersheds. In the winter months, the

SNOW-17 parameters are only sensitive from December through April. The sensitivity of the SNOW-17 parameters is heavily influenced by the transition from dry conditions in 2001 to the wet conditions of 2003. Interestingly, the biggest difference between the SXTPI and SPKP1 sensitivities is associated with the SNOW-17 parameters, which are maximally sensitive in the wet winter months of 2003. In general, the model’s snow response in SPKP1 is more sensitive than SXTPI, especially in dry years.

6.3. Event Sensitivities Based on Distributed Forcing and Distributed Parameters

[29] Two flood events have been selected and analyzed to better understand the spatial distribution of the HL-RDHM sensitivities. The May 2002 event is fairly uniform and the September 2003 event has a heterogeneous spatial distribution. Our focus on the SPKP1 watershed was largely motivated by the extremely large computing demands posed by using Sobol’s method for spatial analysis of the HL-RDHM’s sensitivities. To further improve the tractability of this analysis the SNOW-17 component was excluded

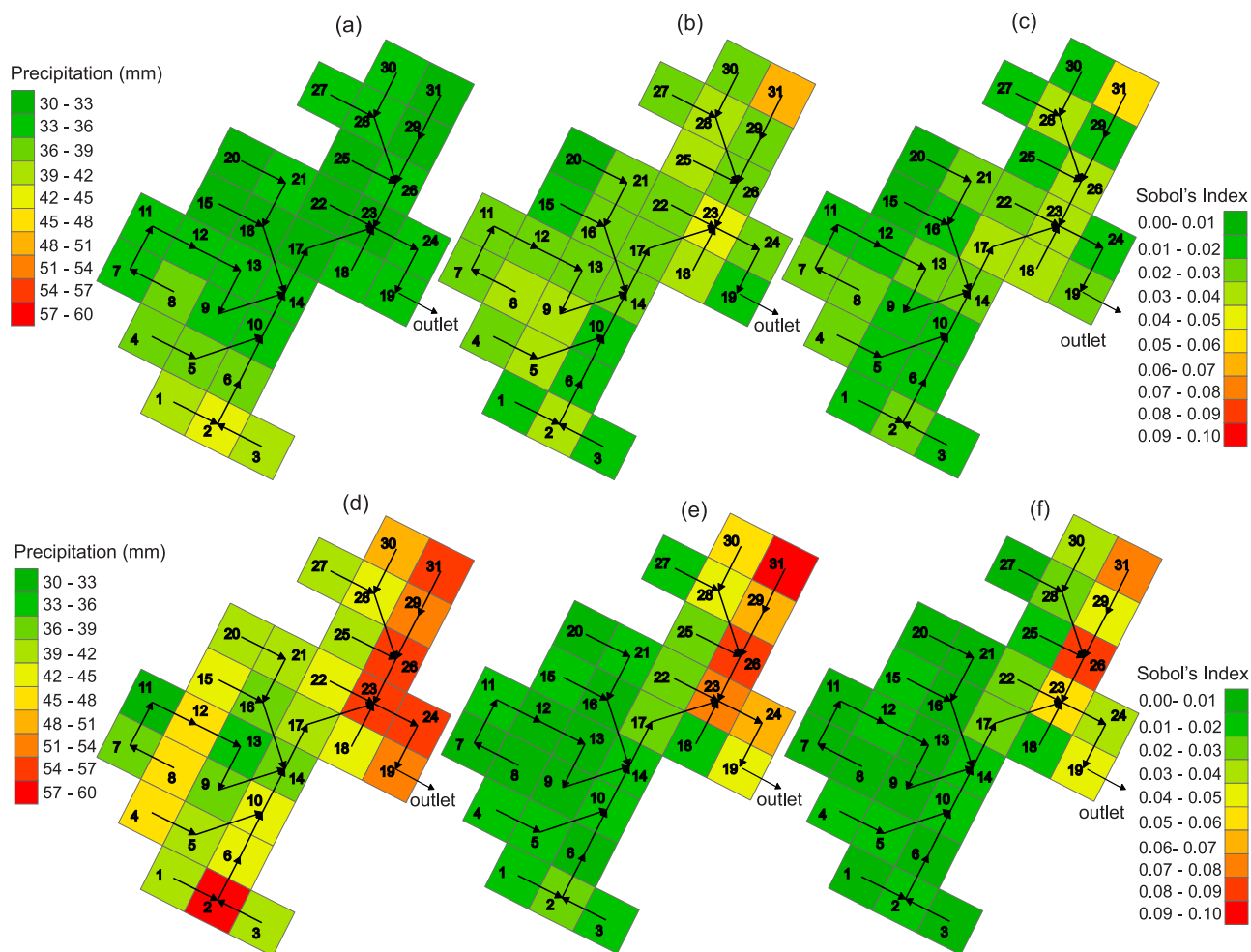


Figure 5. Spatial distribution of the total event precipitation and cell-level sensitivities for the SPKP1 watershed. The May 2002 event is represented by (a) its spatial precipitation distribution, (b) the first-order Sobol's indices for each model cell, and (c) the cell level interactions. The September 2003 event is represented by (d) its spatial precipitation distribution, (e) the first-order Sobol's indices for each model cell, and (f) the cell level interactions. Note cell level interactions were computed as the difference between each cell's total-order and the first-order Sobol's indices. The cell-level Sobol's indices were computed by summing over all of individual parameter indices analyzed in each cell. The arrows in the cells designate surface flow directions.

neglected since neither event was impacted by snowfall. Figure 5 provides a detailed spatial mapping of precipitation, first-order sensitivities, and cell-level interactive sensitivities for the two events.

[30] Figures 5a and 5d illustrate that the May 2002 and September 2003 events do represent a uniform and a heterogeneous distribution of precipitation, respectively. Overall the mean area of precipitation (MAP) of the May 2002 event is smaller than the September 2003 event. It is quite evident that the September 2003 event shown in Figure 5d has its most significant forcing in the northeastern boundary of the SPKP1 watershed near the outlet. The uniformly forced event in May 2002 yielded fairly uniform sensitivities as shown in Figures 5b and 5c. After the 2 month warm up period used to model both events, the HL-RDHM's initial conditions for the lower zone free primary water storage showed a slight increasing trend from the southwestern portion of SPKP1 (i.e., cells 1, 2, and 3) to the northeastern

portion of the watershed (i.e., cells 29, 30, and 31). For the May 2002 event cells 23 and 31 had the maximum initial storage, which implies these cells were initially the wettest cells and that they should have an increased influence on the HL-RDHM's response. Figure 5b confirms this expectation and shows that cells 23 and 31 by themselves account for more than 10 percent of the HL-RDHM's response. Beyond the initial wetness of these cells, their close proximity to the modeled outlet for the watershed also increased their influence on predictions.

[31] Overall the first-order Sobol's indices and interactivity results in Figures 5b and 5c for the May 2002 event highlight that two factors largely control the HL-RDHM's sensitivity to a cell: (1) the cell's initial wetness and (2) the cell's proximity to the gauged outlet. Beyond these two factors, the spatially heterogeneous event in September 2003 also shows that significant spatial differences in forcing is a third factor that strongly impacts cell-level

Table 6. Sobol's Indices for Each of the 13 Parameters Analyzed in the May 2002 and September 2003 Events^a

Event	Order	Parameters												
		UZTWM	UZFWM	UZK	PCTIM	ADIMP	ZPERC	REXP	LZTWM	LZFSM	LZFPM	LZSK	LZPK	PFREE
May 2002	1st	0.118	0.011	0.006	0.084	0.046	0.029	0.003	0.231	0.015	0.084	0.008	0.023	0.183
May 2002	total	0.212	0.017	0.012	0.063	0.066	0.048	0.006	0.346	0.066	0.126	0.041	0.026	0.27
Sept. 2003	1st	0.099	0.025	0.006	0.068	0.081	0.047	0.013	0.311	0.039	0.081	0.013	0.006	0.064
Sept. 2003	total	0.163	0.045	0.018	0.074	0.094	0.097	0.013	0.411	0.091	0.078	0.054	0.022	0.133

^aThe first- and total-order indices for each parameter were computed by summing their individual cell-level indices over the SPKP1 watershed's model domain.

sensitivities. In Figure 5d, cells 24, 23, 29, 30, and 31 are very close to the SPKP1 watershed's gauged outlet and receive significantly more rainfall than nearby cells, which strongly increases their sensitivities as shown in Figures 5e and 5f. These five cells cover only 16 percent of the modeled area but account for nearly half of the HL-RDHM's output variance. It is interesting to note that cell 2 in the southwestern portion of the grid domain received the most rainfall overall (see Figure 5d), but had a relatively small impact on model predictions due to its initial conditions (i.e., it initially had a slightly increased storage capacity to capture much of the rainfall) and its significant distance from the gauged outlet.

[32] Table 6 provides a summary of the first and total order Sobol's indices for the 13 parameters sampled for the May 2002 and September 2003 events. The indices in Table 6 reflect the total watershed-level influence of each parameter on the event predictions. The table shows many similar trends to the prior lumped parameter results from sections 6.1 and 6.2. For both events the model's upper zone tension water storage (UZTWM), lower zone tension water storage (LZTWM), and the fraction of percolating water (PFREE) are the dominant parameters impacting between 16 to more than 40 percent of the HL-RDHM's variance.

[33] As was also seen in the annual and monthly results, the model's upper zone tension water storage (UZTWM) and fraction of percolating water (PFREE) are more important for the relatively drier system conditions in the May 2002 event versus the wetter conditions of the September 2003 event. PFREE is not activated when LZTWM is full which is more likely to happen during wet conditions. In both events, the lower zone tension water storage (LZTWM) is the dominant parameter explaining 35 to 41 percent of the HL-RDHM's output variance. The spatial maps of Figure 5 as well as the Sobol's indices for each parameter in Table 6 show that the spatial distributions of forcing and model cell wetness significantly control the HL-RDHM's sensitivity.

6.4. Verification of Event Analysis Sensitivity Rankings

[34] Building on work by *Andres* [1997] and *Tang et al.* [2006b], we have tested the repeatability and screening effectiveness of the Sobol's sensitivity method. We have used the sensitivity classifications given in section 6.3 in combination with independent LHS-based random draws to develop the verification plots for event analysis sensitivity rankings given in Figure 6. Our overall intent for this analysis is to use independent random samples to test if the parameters and model cells classified as being sensitive in section 6.3 do in fact control the HL-RDHM's response.

Repeating the analysis for the May 2002 and September 2003 events also provides some insights on how the spatial heterogeneity of forcing impacts model parameter screening using Sobol's method.

[35] Our analysis uses four randomly drawn parameter sets: (1) Set 1 consists of 1000 randomly drawn Latin hypercube samples for the 13 SAC-SMA parameters analyzed for the SPKP1 watershed model, (2) Set 2 consists of random samples for the subset of model parameters consisting of the top 6 most sensitive SAC-SMA parameters which are perturbed across all model cells, (3) Set 3 consists of random samples for all 13 SAC-SMA parameters analyzed using only the top 15 most sensitive model cells, and (4) Set 4 consists of random samples for the subset of model parameters representing the top 6 most sensitive parameters within the top 15 most sensitive model cells. The overall rationale for using these four sets is to simply reduce the number of model parameters (i.e., 13 parameters per cell \times 31 model cells) by approximately 50 percent for Sets 2 and 3. Set 4 then combines the parameter-based screening strategy of Set 2 with the cell-based screening strategy of set 3 to reduce the overall number of model parameters being considered by more than 75 percent.

[36] In Figure 6, Set 1 provides the performance baseline representing the full randomly generated independent sample set. The premise of this analysis is that if the Sobol's method results from section 6.3 are correct then the method's rankings should provide sufficient information to identify the correct subset of sensitive parameters. When the correct subset of sensitive parameters is sampled randomly (i.e., Set 2, Set 3, or Set 4) 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. Several interesting observations are evident in Figure 6 that depend on the type of strategy used for reducing the number of parameters considered as well as the spatial heterogeneity of the events' forcing.

[37] The parameter-based screening strategy represented by Set 2 worked well yielding correlation coefficients of 0.873 and 0.856 for the May 2002 and September 2003 events, respectively. The Set 2 parameter-based screening strategy selected the top 6 most sensitive parameters whose total-order effects explained more than 90 percent of the HL-RDHM's output variance. For the cell-based screening strategy represented by Set 3, performance depended on the event analyzed (see Figure 6). For the uniform spatial distribution of precipitation in the May 2002 event, the cell-based ranking strategy actually degraded performance for capturing the HL-RDHM's output. This result is intuitive since uniformly distributed precipitation did not lead to

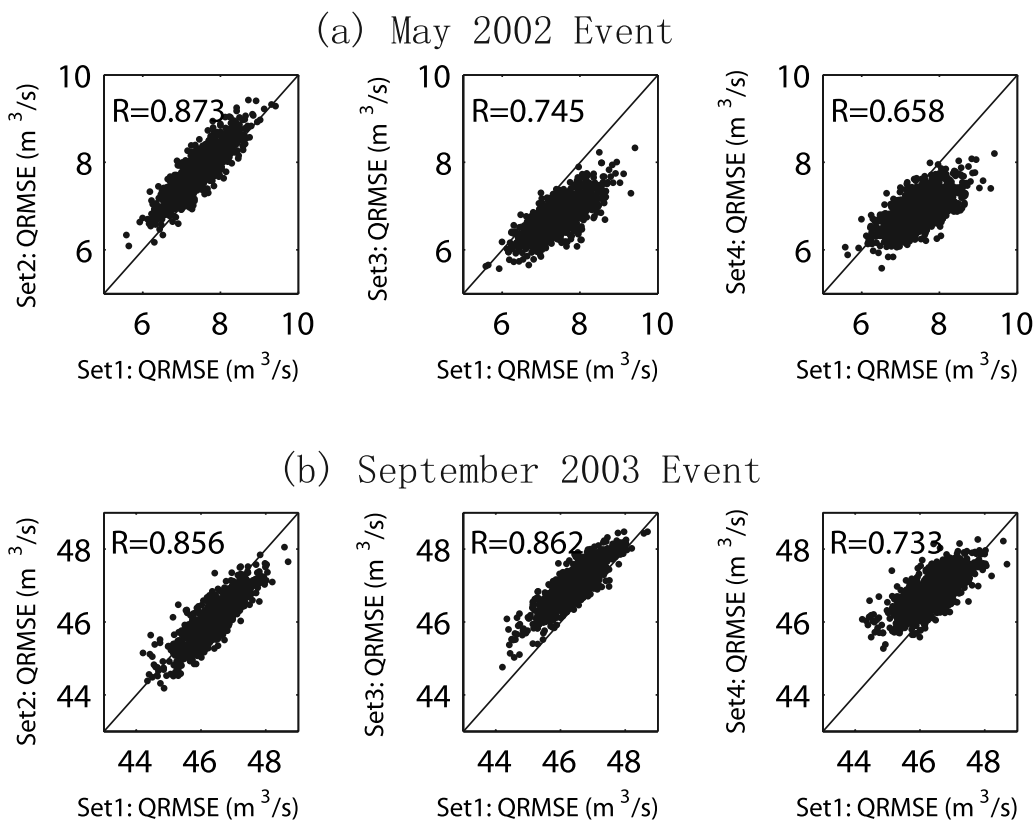


Figure 6. Verification plots for event analysis sensitivity rankings based on hourly model time steps for the HL-RDHM. The scatterplots show the RMSE of streamflow predictions. Set 1 consists of 1000 randomly drawn Latin hypercube samples for the 13 SAC-SMA parameters analyzed for the SPKP1 watershed model. Set 2 consists of random samples for the subset of model parameters composed by top 6 most sensitive SAC-SMA parameters perturbed across all model cells. Set 3 consists of random samples for all 13 SAC-SMA parameters analyzed using only the top 15 most sensitive model cells. Set 4 consists of random samples for the subset of model parameters representing the top 6 most sensitive parameters on the top 15 most sensitive model cells.

any signature spatial trends in sensitivity. Alternatively, the Set 3 cell-based strategy showed an improved screening performance for the September 2003 event since this event had signature spatial trends for precipitation and model sensitivities.

[38] Combining the parameter screening strategies to yield Set 4 served to reduce the overall set of parameters being analyzed to less than 100 parameters while still maintaining correlation coefficients of 0.658 and 0.733 for the May 2002 and September 2003 events, respectively. It is interesting to note that strong spatial trends in precipitation appear to improve the identifiability of the HL-RDHM-based watershed models. Evidence to support this claim can be drawn from Figure 6b, which shows that the Set 4 parameter screening strategy yielded a much higher correlation coefficient for the September 2003 event. This result implies that a much smaller set of parameters can be used to approximate HL-RDHM's response for the heterogeneously forced event. Overall, the subset of parameters used in the Set 4 screening strategy for both events explained approximately 70 percent of HL-RDHM's output variance. The results of Figure 6 confirm that Sobol's method provides robust sensitivity rankings. In future extensions of this

work, we will explore how Sobol's method can be used to enhance calibration methodologies for HL-RDHM.

7. Discussion and Conclusions

[39] This study provides a step-wise analysis of the US NWS's distributed modeling framework's (HL-RDHM) sensitivities from annual to event level time periods with the intent of elucidating the key parameters impacting the model's forecasts. This study demonstrates a methodology that balances the computational constraints posed by Sobol's sensitivity analysis with the need to fully characterize the HL-RDHM's forecasting sensitivities. In the context of long-term forecasts, HL-RDHM's sensitivities were assessed for annual periods using 24-hour model time steps and monthly periods using 1-hour model time steps. For the annual and monthly analysis, the HL-RDHM's computational demands required our use of distributed forcing and model structure, but lumped model parameterizations for two case study watersheds within the Juniata River basin in central Pennsylvania, USA. In the context of event forecasts, this study provides detailed spatial analysis of the HL-RDHM's sensitivities for two flood events simulated using a 1-hour model time step. The events were

selected to demonstrate how the spatial heterogeneity of forcing has a significant influence on the model's spatial sensitivities. Our spatial analysis of event sensitivities also included an evaluation of how well Sobol's sensitivity method performs in identifying the principle input variables controlling the HL-RDHM's response. The robustness of Sobol's method was tested using an extension of the SA repeatability test recommended by Andres [1997]. The results of this analysis demonstrate that the method provides robust sensitivity rankings and that the rankings could be used to significantly reduce the number of parameters that must be considered when calibrating the HL-RDHM.

[40] The annual to event-level time periods analyzed in the prior section show that the HL-RDHM's sensitivities are controlled largely by three factors: (1) variations in the model's storage for both its upper and lower zones that occur for transitions from dry to wet conditions (or vice versa), (2) strong spatial trends in the model's NEXRAD forcing data, and (3) the model cells' proximities to the gauged outlet where the model performance objective is computed (i.e., RMSE in this study). It provides some insights on the types of information that should guide future forecasting methodologies. Sensitivities that arise from a cell's proximity to the "outlet gauge" and "spatial trends in precipitation" indicate that our ability to evaluate and identify appropriate forecasting models is a function of the observation network. It is very common to adapt modeling frameworks but it is far less common to advance forecasting through adaptive design and improvement of the terrestrial and riverine observation network. The results of this study demonstrate that a spatially distributed simulation can be controlled by as few as 3 or 4 model cells very near the drainage basin outlet. It is likely that increased gauging would more fully activate the spatial distribution of model sensitivities. It has been widely recognized in prior studies that the resolution, magnitude, and quality of NEXRAD observations of forcing also heavily impact distributed models' responses [Carpenter *et al.*, 2001; Young *et al.*, 2000]. Improved gauging of both the terrestrial states predicted by the HL-RDHM as well as the atmospheric forcing driving the model appear to be important needs for advancing our ability to develop distributed forecasts.

[41] Conceptual challenges in calibrating the HL-RDHM arise because the model exhibits a highly nonlinear response behavior that is heavily dependent on the individual as well as the interactive sensitivities of its parameters. As would be expected, the number and magnitude of individual and interactive parameter sensitivities increases for the HL-RDHM when modeling wet conditions. Across all of the time periods analyzed, the HL-RDHM's sensitivities showed that while upper zone storage and percolation parameters are very important for dry conditions, a transition to wet conditions decreases the importance of these parameters. For wet conditions, the HL-RDHM's lower zone storages become the dominant parameters controlling the model's response. The annual, monthly, and event-level sensitivity trends presented in this study highlight that static calibration strategies that do not incorporate changing model sensitivities would likely be suboptimal in extracting information from available data [see Wagener *et al.*, 2003]. Operational forecasts based on the HL-RDHM would benefit from the joint use of a robust sensitivity analysis

framework directly integrated into new calibration methodologies.

[42] The emerging trend in surface hydrology toward spatially distributed simulation will require the field to embrace and advance high-performance computing to ensure these tools can be used in both scientific and operational applications [e.g., see Apostolopoulos and Georgakakos, 1997; Tang *et al.*, 2006c]. Algorithmic design and implementation for high-performance computing architectures will need to be carefully considered. In this study, our Sobol's method code was implemented using a highly portable MPI parallel framework [Gropp, 1999] so that a nonexpert could utilize our computational cluster in sensitivity analysis applications. Operational use of the HL-RDHM will require computational support tools that are parallel, highly portable, and implemented in a manner that minimizes the computational expertise needed by operational personnel.

[43] **Acknowledgments.** The first and second authors of this work were partially supported by the National Science Foundation under grants EAR-0310122, EAR-0609791, and EAR-0609741. 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. Support for the third 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. Partial support for the fourth author was provided by SAHRA under NSF-STC grant EAR-664 9876800 and the National Weather Service Office of Hydrology under grants 665 NOAA/NA04NWS4620012 and NOAA/DG 133W-03-666 SE-0916.

References

- Abbott, M. B., J. A. Bathurst, and P. E. Cunge (1986), An introduction to the European hydrological system-systeme hydrologique europeen "SHE" 1: History and philosophy of a physically based distributed modeling system, *J. Hydrol.*, *87*, 45–59.
- Ajami, N. K., H. V. Gupta, T. Wagener, and S. Sorooshian (2004), Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *J. Hydrol.*, *298*, 112–135.
- Anderson, E. A. (1973), National Weather Service river forecast system-snow accumulation and ablation model, technical report, U.S. Dep. of Commer., Washington, D. C.
- Anderson, E. A. (1994), Continuous API model: National Weather Service river forecasting system users manual, technical report, U.S. Natl. Weather Serv., Washington, D. C.
- Anderson, E. A. (2002), Calibration of conceptual hydrologic models for use in river forecasting, technical report, *NOAA Tech. Rep. NWS 45*, Hydrol. Lab., Silver Spring, Md.
- Andres, T. H. (1997), Sampling method and sensitivity analysis for large parameter sets, *J. Stat. Comput. Simul.*, *57*, 77–110.
- Apostolopoulos, T. K., and K. P. Georgakakos (1997), Parallel computation for streamflow prediction with distributed hydrologic models, *J. Hydrol.*, *197*, 1–24.
- Archer, G. E. B., A. Saltelli, and I. M. Sobol' (1997), Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Stat. Comput. Simul.*, *58*, 99–120.
- Beven, K. J. (1989), Changing ideas in hydrology—The case of physically-based models, *J. Hydrol.*, *105*, 157–172.
- Beven, K. J., and M. J. Kirkby (1979), A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*(1), 43–69.
- Boyle, D. P., H. V. Gupta, S. Sorooshian, V. Koren, Z. Zhang, and M. Smith (2001), Toward improved streamflow forecasts: Value of semidistributed modeling, *Water Resour. Res.*, *37*(11), 2749–2759.
- Burnash, R. J. C. (1995), The NWS river forecast system-catchment model, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 311–366, Water Resour. Publ., Highlands Ranch, Colo.
- Carpenter, T. M., K. P. Georgakakos, and J. A. Sperflage (2001), On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use, *J. Hydrol.*, *253*, 169–193.

- Demaria, E., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model, *J. Geophys. Res.*, doi:10.1029/2006JD007534, in press.
- Doherty, J. (2003), Groundwater model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.
- Doherty, J., and J. M. Johnston (2003), Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, *39*(2), 251–265.
- Duan, Q., V. K. Gupta, and S. Sorooshian (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.
- Duffy, C. J. (2004), 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*, *Water Sci. Appl. Ser.*, vol. 9, edited by J. F. Hogan, F. Phillips, and B. Scanlon, pp. 255–271, AGU, Washington, D. C.
- Efron, B., and R. Tibshirani (1993), *An Introduction to the bootstrap*, Chapman Hall, New York.
- Freer, J., K. J. Beven, and B. Ambrose (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach, *Water Resour. Res.*, *32*, 2161–2173.
- Gropp, W. (1999), *Using MPI: Portable Parallel Programming with Message-Passing Interface*, 2nd ed., MIT Press, Cambridge, Mass.
- Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*, 751–763.
- Hall, J., S. Tarantola, P. D. Bates, and M. S. Horritt (2005), Distributed sensitivity analysis of flood inundation model calibration, *J. Hydraul. Eng.*, *131*(2), 117–126.
- Helton, J., and F. Davis (2003), Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliab. Eng. Syst. Safety*, *81*(1), 23–69.
- Hornberger, G. M., and R. C. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*, 7–18.
- Koren, V., S. Reed, M. Smith, Z. Zhang, and D. J. Seo (2004), Hydrology Laboratory Research Modeling System (HL-RMS) of the U.S. National Weather Service, *J. Hydrol.*, *291*, 297–318.
- Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modeling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*, 205–216.
- McKay, M., R. Beckman, and W. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245.
- Moreda, F., V. Koren, Z. Zhang, S. Reed, and M. Smith (2006), Parameterization of distributed hydrological models: Learning from the experiences of lumped modeling, *J. Hydrol.*, *320*, 218–237.
- Muleta, M., and J. W. Nicklow (2005), Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, *306*, 1–19.
- Osidele, O. O., and M. B. Beck (2001), Identification of model structure for aquatic ecosystems using regionalized sensitivity analysis, *Water Sci. Technol.*, *43*(7), 271–278.
- Panday, S., and P. Huyakorn (2004), A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow, *Adv. Water Resour.*, *27*, 361–382.
- Pappenberger, F., I. Iorgulescu, and K. Beven (2006), Sensitivity analysis based on regional splits and regression trees (SARS-RT), *Environ. Modell. Software*, *21*, 976–990.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1999), *Numerical Recipes in C*, 2nd ed., Cambridge Univ. Press, New York.
- Reed, S. M., and D. R. Maidment (1999), Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling, *J. Hydrol. Eng.*, *4*(2), 174–182.
- Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D. Seo, and DMIP Participants (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, *298*, 27–60.
- Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*, 280–297.
- Saltelli, A., S. Tarantola, and K. P.-S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, *41*(1), 39–56.
- Seo, D. J., H. D. Herr, and J. C. Schaake (2006), A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, *Hydrol. Earth Syst. Sci. Discuss.*, *3*, 1987–2035.
- Sieber, A., and S. Uhlenbrook (2005), Sensitivity analyses of a distributed catchment model to verify the model structure, *J. Hydrol.*, *310*, 216–235.
- Smith, M., D. Seo, V. Koren, S. Reed, Z. Zhang, Q. Duan, F. Moreda, and S. Cong (2004), The distributed model intercomparison project (DMIP): Motivation and experiment design, *J. Hydrol.*, *298*, 4–26.
- Sobol', I. M. (1967), On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys., Engl. Transl.*, *7*, 86–112.
- Sobol', I. M. (1993), Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.*, *1*(4), 407–417.
- Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*, 271–280.
- Tang, Y., P. Reed, and T. Wagener (2006a), How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration, *Hydrol. Earth Syst. Sci.*, *10*, 289–307.
- Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2006b), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci. Discuss.*, *3*, 3333–3395.
- Tang, Y., P. Reed, and J. Kollat (2006c), Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications, *Adv. Water Resour.*, *30*, 335–353.
- Tonkin, M. J., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.
- van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*, 10–23.
- Wagener, T., and J. Kollat (2007), Numerical and visual evaluation of hydrological and environmental models using the monte carlo analysis toolbox, *Environ. Modell. Software*, *22*, 1021–1033.
- Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*, 455–476.
- Wang, M., A. Hjelmfelt, and J. Garbrecht (2000), DEM aggregation for watershed modeling, *J. Am. Water Resour. Assoc.*, *36*(3), 579–584.
- Young, B., A. A. Bradley, W. F. Krajewski, A. Kruger, and M. Morrissey (2000), An evaluation study of NEXRAD multisensor precipitation estimates for operational hydrologic forecasting, *J. Hydrometeorol.*, *1*(3), 241–254.
- Young, P. C. (1978), A general theory of modelling for badly defined dynamic systems, in *Modeling, Identification and Control in Environmental Systems*, edited by G. C. Vansteenkiste, pp. 103–135, North Holland, Amsterdam.
- Zhang, Y., and G. F. Pinder (2003), Latin hypercube lattice sample selection strategy for correlated random hydraulic conductivity fields, *Water Resour. Res.*, *39*(8), 1226, doi:10.1029/2002WR001822.

P. Reed, Y. Tang, K. van Werkhoven, and T. Wagener, Department of Civil and Environmental Engineering, 212 Sackett Building, Pennsylvania State University, University Park, PA 16802, USA. (yxt132@psu.edu)