WATER RESOURCES RESEARCH, VOL. 46, W01505, doi:10.1029/2009WR008066, 2010

On the use of spatial regularization strategies to improve calibration of distributed watershed models

Prafulla Pokhrel[1] and Hoshin V. Gupta[1]

Received 1 April 2009; revised 19 August 2009; accepted 3 September 2009; published 8 January 2010.

[1] Hydrologic models require the specification of unknown model parameters via calibration to historical input-output data. For spatially distributed models, the large number of unknowns makes the calibration problem poorly conditioned. Spatial regularization can help to stabilize the problem by facilitating inclusion of additional information. While a common regularization approach is to apply a scalar multiplier to the prior estimate of each parameter field, this can cause problems by simultaneously changing both the mean and the variance of the distribution. This paper explores a multiple-criteria regularization approach that facilitates adjustment of the mean, variance, and shape of the parameter distribution, using prior information to constrain the problem while providing sufficient degrees of freedom to enable model performance improvements. We also test simple squashing functions to help in maintaining conceptually reasonable parameter values throughout the spatial domain. We apply the method to three basins in the context of the Distributed Model Intercomparison Project (DMIP2), obtaining considerable performance improvements at the basin outlet. However, the prior parameter estimates are found to give much better performance at the interior points (treated as ungauged), suggesting that the spatial information has not been properly exploited. The results also suggest that basin outlet hydrographs may not be particularly sensitive to spatial parameter variability and that an overall basin mean value may be sufficient for flow forecasting at the outlet, although not at the interior points. We discuss weaknesses in our study approach and suggest diagnostically more powerful strategies to be pursued.

Citation: Pokhrel, P., and H. V. Gupta (2010), On the use of spatial regularization strategies to improve calibration of distributed watershed models, *Water Resour. Res.*, *46*, W01505, doi:10.1029/2009WR008066.

1. Introduction

1.1. Background

[2] Hydrologic models typically assume some universality in model structure [*Gupta et al.*, 2003], meaning that there is an expectation that the model can provide acceptable input-state-output (ISO) simulations for the watersheds of interest. Therefore, for a particular watershed, the modeling problem becomes "simply" that of specifying values for the model parameters. Because many parameters are not directly observable (or easily inferred from measurement data), they must be estimated indirectly by an inverse (calibration) process that conditions the parameter estimates (and hence the model response) on historically observed input-output data [e.g., *Beven*, 1996; *Gupta et al.*, 1998, 2003, *Madsen*, 2003; *Vieux and Moreda*, 2003; *Refsgaard and Storm*, 1996]. Some degree of calibration is therefore generally unavoidable in hydrologic modeling [e.g., *Beven*, 1989; *Gupta et al.*, 1998; *Refsgaard and Storm*, 1996].

[3] When implementing a spatially distributed hydrologic model, the large number of unknowns results in a high-dimensional parameter search space, which significantly complicates the optimization problem [e.g., *Beven*, 1996; *Leavesly et al.*, 2003; *Pokhrel et al.*, 2008; *Refsgaard*, 1997; *Vieux*, 2001]. When this dimension becomes so large that the unknowns cannot be uniquely constrained by the data, the problem is said to be ill posed or poorly conditioned. "Regularization" is a mathematical strategy that helps to stabilize ill-posed problems by the enabling the inclusion of additional information [e.g., *Tikhonov and Arsenin*, 1977; *Lawson and Hanson*, 1995; *Weiss and Smith*, 1998; *Doherty and Skahill*, 2006; *Linden et al.*, 2005; *Tonkin and Doherty*, 2005; *Isaaks and Srivastava*, 1989]. By using prior information (either direct or indirect) related to the parameters, regularization is able to "better condition" the objective function response surface, either via some kind of penalty function [*Carrera and Neuman*, 1986; *Doherty and Skahill*, 2006] or by imposing constraints that reduce the dimensionality of the parameter search space [*see Pokhrel et al.*, 2008, 2009].

1.2. Strategies for Spatial Regularization

[4] In spatially distributed watershed modeling, a common and simple approach to regularization is to apply a scalar multiplier m to each prior parameter field (specified from data describing watershed characteristics: soils, vegetation, topography, land use, etc.) and to estimate a "best" value for this multiplier via calibration [e.g., *Bandaragoda*

[1]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA.

0043-1397/10/2009WR008066$09.00

et al., 2004; *Cornfield and Lopes*, 2004; *Eckhardt and Arnold*, 2001; *Francés et al.*, 2007; *Pokhrel et al.*, 2008; *Vieux et al.*, 2004; *Yatheendradas et al.*, 2008]. In a model with N_g grid cells and N_p parameters per cell, the dimension of the calibration problem is thereby reduced from N_g*N_p to $1*N_p$. This so-called ''multiplier'' approach makes the (hopefully reasonable) assumption that the prior parameter field properly describes the spatial pattern of parameter variation (the pattern of relative magnitudes from cell to cell), but that the magnitudes of *all* the parameter values must be adjusted to achieve a better simulation of the model response. For example, the calibration approach described in the manual for the Hydrology Lab Distributed Hydrology Model [*NOAA*, 2007] uses prior spatial estimates of the model parameters derived via the [*Koren et al.*, 2000] method from watershed soils information (STATSGO/SSURGO), land use, vegetation and other data. *Koren et al.* [2000] and *Anderson et al.* [2006] report that these estimates provide good initial approximations, which can be refined via manual and/or automated calibration using the multiplier approach. *Bandaragoda et al.* [2004] used the multiplier approach to refine spatial parameter estimates for the TOPNET model derived from remote sensing and GIS information. *Vieux et al.* [2004] used an ''ordered physics-based parameter adjustment'' methodology in which scalar multipliers, applied to the prior estimates of a parameter field, were adjusted in stepwise fashion to calibrate a physically based hydrologic model.

[5] In essence, the multiplier approach reduces the dimension of the ''distributed'' model optimization problem to that of an equivalent ''lumped'' model, so that existing optimization algorithms (developed for lumped watershed modeling) can be applied. But under what circumstances should the multiplier approach be used, and what are the implications of its use? Simple examination shows that the multiplier simultaneously changes both the *mean* and the *variance* of the parameter field. Let $\theta = [\theta_1 \ldots \theta_g]$ represent the prior values of parameters in the N_g grid cells and $\phi = [\phi_1 \ldots \phi_g]$ represent the corresponding posterior values after adjustment by the multiplier m (i.e., $\phi_j = m \times \theta_j$). The mean and variance of the posterior distribution are given by

$$E\{\phi\} = m \cdot E\{\theta\} \tag{1}$$

$$Var\{\phi\} = m^2 \cdot Var\{\theta\} \tag{2}$$

showing clearly that the mean and variance of the distribution are adjusted in a linked fashion, whereby ''small'' prior values are adjusted by only small amounts while ''large'' prior values are adjusted by much larger amounts.

[6] One way to avoid this problem, less widely reported, is to instead apply a scalar additive constant a to each spatial parameter field so that only the mean of the parameter distribution is adjusted (i.e., $\phi_j = \theta_j + a$) while leaving the variance (and the absolute differences between the parameter values in the grid) unchanged [*Vieux*, 2001]. By extension, the two approaches can be combined:

$$\phi_j = m \cdot \left(\theta_j - E\{\theta\}\right) + a \tag{3}$$

to allow the mean and variance of each parameter field to be adjusted independently [*Vieux*, 2001]; in this case there are two values to be adjusted per spatial parameter field, and the dimension of the regularized problem becomes $2*N_p$.

[7] By further extension, nonlinear transformations of the following form can also be used:

$$\phi_j = m \cdot \left(\theta_j\right)^b + a \tag{4}$$

[8] This approach has three values to be adjusted per spatial parameter field (the multiplier m, additive constant a and power term b), increasing the dimension of the regularized problem to $3*N_p$. By increasing the degrees of freedom, this strategy allows the optimization method to deform the probability distribution of each parameter field in more complex ways. *Pokhrel et al.* [2008] developed this nonlinear regularization approach for application to watershed models in a different manner, where the θ_j terms refer instead to the soils properties and other watershed characteristics rather than the inferred prior parameter estimates.

[9] Regularization approaches of the kind mentioned above are reasonable ways for using prior information to constrain the calibration problem, making the optimization problem better conditioned while allowing the model sufficient degrees of freedom to match the observed behavior of the watershed. To ensure clarity of communication, we adopt the terminology introduced by *Tonkin and Doherty* [2005] and refer to the coefficients of the regularization constraint equations as ''superparameters.''

1.3. Dealing With Parameter Constraints

[10] When applying regularization strategies an interesting question is how to deal with ''feasible parameter bounds.'' For reasons of conceptual-physical consistency and realism, model parameters are usually expected to remain within (or close to) some range described by an upper and lower bound. In spatially distributed models, these constraints must be obeyed by the values in *each and every cell*. However, when adjusting an entire parameter field, some of the grid cell values can move outside the feasible range, even though the spatial mean (and much of the distribution) remains feasible. If the boundaries can be considered ''soft'' (not well specified by prior information), this might not matter much. But when the boundaries are ''hard'' (e.g., to prevent storage capacities from becoming negative or to restrict drainage rate parameters to a [0–1] range), violations of the boundaries can have severe computational or conceptual implications.

[11] If the feasible parameter range constraints are treated as hard boundaries, the optimization procedure can become severely constrained, particularly if the variance of the priori parameter distribution is initially large (with values close to the feasible parameter bounds). This will prevent it from adjusting the shape and location of the parameter field. One solution is to ''soften'' the constraint boundaries and allow the parameter probability distribution to partially violate the constraints, but this cannot be done for capacity and rate parameters that must obey constraints imposed by computational and conceptual consistency. An alternative approach is to deform or ''squash'' the shape of the parameter probability distribution to keep it within the boundaries, thereby maintaining the degrees of freedom facilitated by

Table 1. Distributed SACSMA Parameters Used for Calibration

Name	Description	Parameter Boundary Settings: Tight Boundary Setting	Parameter Boundary Settings: Loose Boundary Settings
UZTWM	upper zone tension water capacity (mm)	10–300	0.01–1000
UZFWM	upper zone free water capacity (mm)	5–150	0.01–1000
UZK	upper zone free water withdrawal rate (mm/h)	0.1–0.8	0.001–0.99
REXP	percolation equation exponent	1–5	1–5
LZTWM	lower zone tension water capacity (mm)	5–500	0.01–2000
LZFSM	lower zone supplemental free water capacity (mm)	5–400	0.01–2000
LZFPM	lower zone primary free water capacity (mm)	10–1000	0.01–2000
LZSK	lower zone supplemental free water withdrawal rate (mm/h)	0.01–0.35	0.0001–0.4
LZPK	lower zone primary free water withdrawal rate (mm/h)	0.001–0.05	0.00001–0.1
PFREE	fraction of percolated water going directly to lower zone free water storage	0.0–0.8	0.0–1.0
ZPERC	maximum percolation rate coefficient	5–350	5–1000

parameter regularization. To our knowledge the use of squashing functions for this purpose has not been examined in the literature.

1.4. Model Performance at Interior Forecast Points

[12] A significant advantage of spatially distributed models is their ability to simulate hydrologic responses of the watershed at interior locations. *Brath et al.* [2004] reported obtaining good performance in simulating streamflows at interior locations of a midsized catchment in north-central Italy, based on calibrations at the outlet. Similarly, findings were reported by *Bandaragoda et al.* [2004], *Behnaz et al.* [2009] and *Reed et al.* [2004] although they generally reported interior point performance to be worse than at the outlet. Results from phase 1 of DMIP [*Reed et al.*, 2004] also suggest that reasonable interior point performance can be achieved based on calibration to watershed outlet hydrographs, but with a few notable exceptions. In general, not enough investigations into this matter have been reported to confidently say whether spatially distributed watershed models are currently able to provide reliable estimates of hydrological fluxes at ungauged interior locations.

1.5. Objectives and Scope of This Paper

[13] This paper is part of a broad investigation into how prior information about the spatial distribution of model parameters can be used to enhance the performance of a distributed hydrological model. In particular, we seek a spatial parameter estimation strategy that will (1) preserve and exploit information regarding spatial ordering and variability provided by prior parameter distributions, (2) constrain the dimensionality of the optimization problem to enable conventional search algorithms to be used, (3) allow sufficient degrees of freedom to enable significant improvements in model performance at the gauging points where observational data are available, and (4) provide accurate estimates of streamflow at "ungauged" locations along the river network.

[14] Specifically, this paper examines the degree to which the spatial parameter information provided by the *Koren et al.* [2000] approach can serve as a suitable prior for the Distributed Hydrology Model University of Arizona (DHMUA) when applied to three test basins in the Oklahoma-Arkansas area. The *fundamental assumption* is that the Koren approach provides valuable information regarding the relative spatial ordering and magnitude of the parameter fields, but that adjustment of the shape and position of their frequency distributions can help to improve the behavior and performance of the model. This adjustment is achieved by multiple-criteria calibration of the superparameters of a nonlinear spatial regularization strategy, constrained using squashing functions, to obtain improved model performance at the basin outlet. Model performance is evaluated at both the basin outlet from which data were used for model calibration, *and* at interior gauge points for which data are available but withheld from use during calibration. This work was performed in the context of Distributed Model Intercomparison Project (DMIP2), which encouraged the development of spatially distributed watershed-scale models, with the specific goal of improving operational flood forecasting by the U.S. National Weather Service [*Smith et al.*, 2006] (see also http://www.nws.noaa.gov/oh/hrl/dmip/2/index.html).

2. Study Approach and Methods

2.1. Spatially Distributed Watershed Model

[15] The DHMUA model is a research version of the Hydrology Lab Distributed Hydrologic Model (HL-DHM), programmed in MATLAB^{TM} (version 7.0.1, www.mathworks.com) and designed to run at an hourly time step on a personal computer. The water balance component consists of the Sacramento Soil Moisture Accounting Model (SACSMA) [*Burnash et al.*, 1973] having 16 parameters (see Tables 1, 2, and 3) and 6 state variables, applied to each grid cell. Eleven of the parameters are spatially distributed and five are spatially lumped. Prior estimates for the spatially distributed parameter fields were computed using the *Koren et al.* [2000] approach (based mainly on soils information in the top 2 m) and values for the lumped parameters were fixed at values specified by the NWS [*Reed et al.*, 2004]. The routing component has been simplified by removing the within-grid hillslope routing and by using the Muskingum approach (with 2 spatially lumped parameters) instead of kinematic wave for channel routing; *Pokhrel* [2007] showed the impact of this modification to be

Table 2. Uniform SACSMA Parameters Used for Calibration

Name	Description	Illinois River Basin	Baron Fork River Basin	Blue River Basin
RSERV	fraction of lower zone free water not transferable to lower zone tension water storage	0.3	0.3	0.3
SIDE	ratio of deep recharge to channel base flow	0	0	0
PCTIM	minimum impervious area[a]	0.005	0	0.005
ADIMP	additional impervious area[a]	0.1	0.0	0.0
RIVA	riparian vegetation area[a]	0.02	0.035	0.03

[a]Decimal fraction.

insignificant. Applying equation (4) with 3 regularization superparameters to 11 distributed parameter fields and including 2 routing parameters results in 35 calibration terms (3*11+2) that can be adjusted to improve the performance of the model.

2.2. Study Watersheds and Data

[16] The study was conducted on three DMIP2 study basins, the Illinois and Baron Fork River basins straddling the Oklahoma-Arkansas border and the Blue River basin in southern Oklahoma (Figure 1). The gently sloping 1645 km^2 Illinois River basin is mostly composed of silty clay and silty clay loam, and ranges in elevation from 202 to 486 m. The gently sloping 795 km^2 Baron Fork river basin is composed mostly of silty clay and silt loam, and ranges in elevation from 214 to 443 m. The 1233 km^2 Blue River basin is a narrow elongated river valley, ranging in elevation from 154 to 427 m, composed mostly of clay and loam (see *Smith et al.* [2004] for details).

[17] Spatially distributed precipitation estimates are derived from a combination of Next Generation Weather Radar (NEXRAD) and rain gauge data, and quality controlled by the NWS. The data have a temporal resolution of 1 h, and spatial resolution of approximately 4×4 km^2 over a rectilinear Hydrologic Rainfall Analysis Project (HRAP) grid with 100 units over the Illinois, 50 units over the Baron Fork, and 78 units over the Blue. Potential evaporation estimates are based on annual free water surface (FWS) evaporation maps and mean monthly station data (V. Koren et al., PET upgrades to NWSRFS, unpublished report, National Weather Service, NOAA Washington, D. C., 13 August 1998) available from the DMIP Web site (www.weather.gov/oh/hrl/dmip/2/evap.html), and corrected to account for the effects of vegetation.

[18] For each basin, the U.S. Geological Survey records hourly instantaneous river discharge estimates at the basin outlet. The average annual rainfall at Illinois, Baron Fork and Blue are 1175 mm, 1160 mm and 1036 mm, respectively, and the average annual flow at their outlets are 302 mm, 340 mm and 176 mm, respectively. The long-term runoff ratios are approximately 0.26, 0.29 and 0.17 [*Smith et al.*, 2004]. Additional streamflow data are available at interior gauging points for the Illinois and Baron Fork. For the Illinois, the interior point gauge at Savoy drains an area of 433 km^2 (about a quarter of the entire basin), and for Baron Fork, the interior point gauge at Dutch drains an area of 105 km^2 (less than a quarter of the basin). The long-term runoff ratios at Savoy and Dutch are approximately 0.29 and 0.23, respectively.

[19] Six years of data (1 October 1999 to 30 September 2005) were used for this study: the first water year (WY 99–00) for model spin-up to reduce sensitivity to state initialization errors, the next 2 years (WY 00–02) for model calibration, and the final 3 years (WY 02–05) to evaluate model performance. Model performance for the Illinois and Baron Fork was also evaluated at interior points, to examine to what degree the distributed modeling approach (with regularized parameter fields) was able to provide satisfactory streamflow estimates at ungauged locations.

2.3. Regularization Methods Tested

[20] It should be pointed out that no *hydrologically relevant* reason has yet been presented in the literature for why any *specific* kind of spatial regularization (multiplicative, additive, power transformation, or any other form) should be applied in the case of watershed modeling. In general, however, we can assume that (1) the regularization strategy should preserve the relative spatial ordering of the *magnitudes* of the parameter values from grid cell to grid cell (i.e., the parameter values should retain their sort order in the probability distribution function, regardless of spatial location), and (2) the adjusted parameter distribution must not violate boundaries imposed by conceptual-computational consistency. Based on this, we adopt a general nonlinear regularization approach (equation (4)) with three superparameters per parameter field, similar to that proposed by *Pokhrel et al.* [2008]. Specifically, we explore the impact (and relative value) of using each of the three superparameter types (multiplicative, additive and power) independently and in combinations, to determine (1) if the dimension of the superparameter optimization problem can be further reduced and (2) whether certain types of superparameters (or combinations) are better suited to

Table 3. State Variables of the SACSMA Model Used for Calibration

	Description
ADIMC	water contents of the ADIMP area (mm)
UZTWC	upper zone tension water contents (mm)
UZFWC	upper zone free water contents (mm)
LZTWC	lower zone tension water contents (mm)
LZFSC	lower zone free supplemental contents (mm)
LZFPC	lower zone free primary contents (mm)

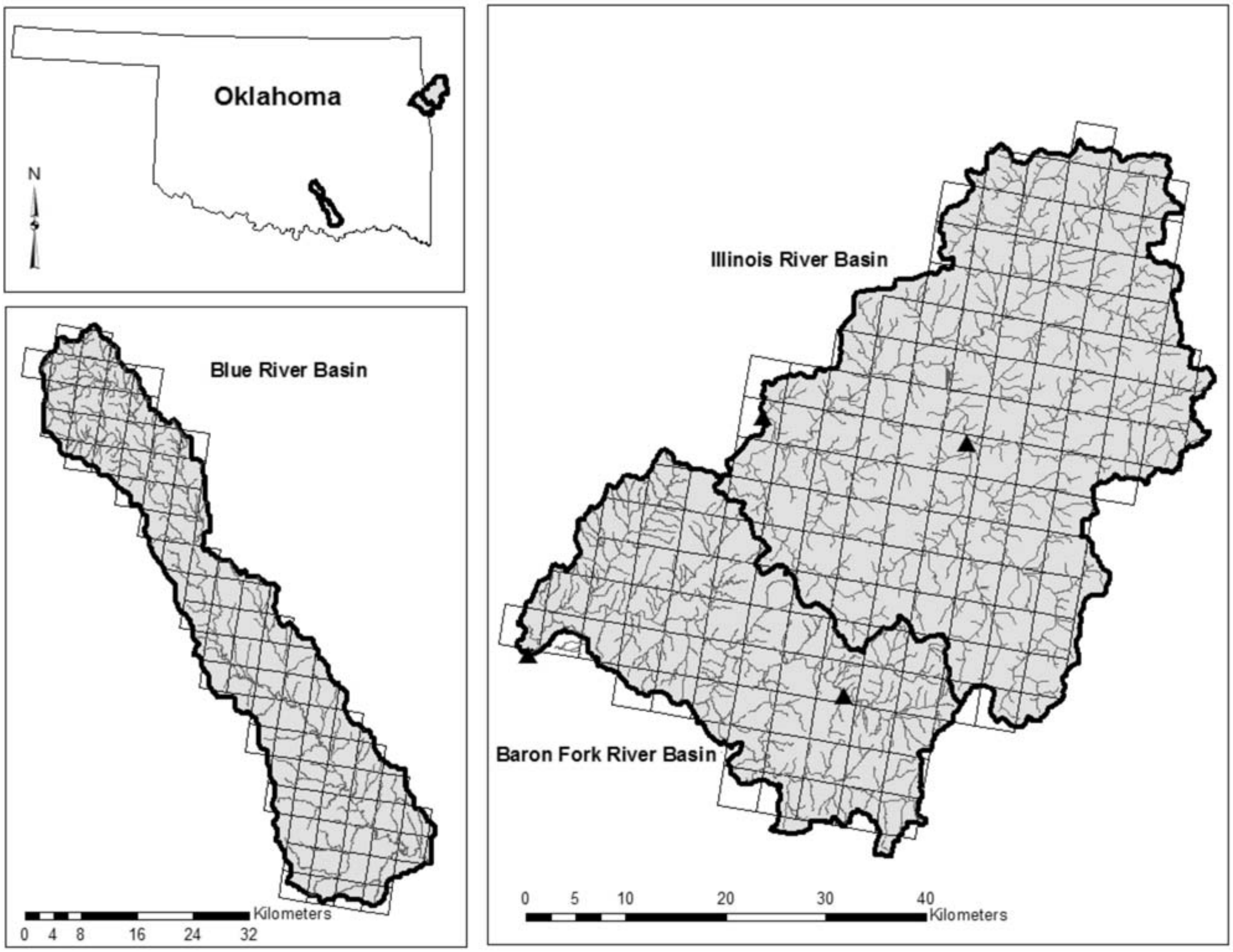

Figure 1. (top left) Oklahoma State showing the positions of the Blue, Illinois, and Baron Fork River basins, (bottom left) Blue River basin with the model grid overlay (HRAP grids, extending from −33°54′N, 96°48′W bottom left corner to 34°42′N, 96°12′W top right corner), and (right) Illinois and Baron Fork River basins (extending from −35°44′N, 94°52W bottom left to 36°24′N, 94°40′W top right) with grid overlays and the location of the gauging stations.

certain kinds of model parameter fields (capacities, rate constants, etc).

[21] Our test of regularization methods was conducted as follows. First, we evaluated the impact and relative value of using each superparameter independently, to see if there are clear advantages to using only multipliers ($\phi_j = m \cdot (\theta_j)$), adders ($\phi_j = (\theta_j) + a$), or power terms ($\phi_j = (\theta j)^b$) to regularize the fields. These three regularized multiple-criteria optimization cases are called RMO-m, RMO-a, and RMO-b, respectively. Next, we examine the marginal benefit of increasing degrees of freedom, by using the linear two-superparameter strategy ($\phi_j = m \cdot (\theta_j) + a$; RMO-ma) and the nonlinear three-superparameter strategy ($\phi_j = m \cdot (\theta_j)^b + a$; RMO-mba). In each calibration case, we also optimize the two routing parameters along with the superparameters.

2.4. Methods to Handle Parameter Boundaries

[22] For reasons of conceptual-physical consistency and realism, the model parameters must remain within (or close to) their feasible parameter ranges specified in terms of upper and lower bounds. When the values in one or more grid cells try to pass outside of the feasible parameter bounds, we can either make the feasible ranges wider (while still maintaining conceptual realism such as nonnegativity, etc.) or permit the shapes of the parameter distributions to deform. For this study we explore the following four boundary settings:

2.4.1. Setting 1: Tight Bounds (TB)

[23] The parameter constraints are set as "hard" and fixed at the feasible bounds used by the NWS (Tables 1–3) [*Koren et al.*, 2003], with one exception: because the largest prior value for parameter UZK in the Blue was already at the upper bound, the bound was increased from 0.75 to 0.8 to allow more space for the distribution to move.

2.4.2. Setting 2: Loose Bounds (LB)

[24] Where possible, the parameter bounds were widened to create a *relaxed* set of hard constraints (Tables 1–3). Capacity parameter ranges were expanded more than double maximum values and to bring minimum values close to zero. Rate parameter ranges were increased to the maximum extents allowable, while ensuring computational stability and preserving their physical-conceptual properties.

2.4.3. Setting 3: Loose Bounds With Tight Initialization (LBTI)

[25] The loose bounds specified in setting 2 were used, but the optimization was initialized at points wholly inside the tight bounds specified in setting 1. This biases the search to the region of the NWS feasible values (tight bounds) but allows it wander outside to improve calibration performance.

2.4.4. Setting 4: Tight Bound Squash (TBS)

[26] The tight bounds specified in setting 1 were used, in conjunction with a simple ''squashing function'' that deforms the parameter distribution as it approaches the boundaries [*Leavesly et al.*, 2003]. If a parameter tries to pass outside the NWS feasible range it is constrained to remain at the boundary. Hence, the distribution can vary freely within the boundaries, and in the extreme case all the grid cells can be assigned the same upper or lower boundary value.

2.5. Method for Multiple-Criteria Optimization

[27] Model calibration was performed via multiple-criteria optimization using the Multiple Objective Shuffled Complex Evolution Metropolis (MOSCEM) global search algorithm [*Vrugt et al.*, 2003]. To obtain balanced model performance on both high and low flows at the basin outlet, two criteria were used: the mean squared error (MSE) to emphasize high-flow performance (equation (5)) and the MSE applied to log-transformed flows (MSEL) to emphasize low-flow performance (equation (6)):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - S_i)^2 \tag{5}$$

$$MSEL = \frac{1}{n}\sum_{i=1}^{n}(\log(O_i) - \log(S_i))^2 \tag{6}$$

where n is the number of time steps used for the evaluation, S is the simulated flow and O is the observed flow. Optimizing both criteria at once, results in a set of mutually nondominated parameter combinations [*Gupta et al.*, 1998, 2003], also called Pareto-optimal solutions, which have the property that when moving from one to another point at least one criterion is improved while another gets worse (a trade-off occurs) and it is not possible to find any other solution within the feasible parameter space for which all the criteria can be simultaneously improved. The method has been widely applied in calibration of hydrological models [e.g., *Vrugt et al.*, 2003; *Leplastrier et al.*, 2002; *Xia et al.*, 2002; *Demarty et al.*, 2005].

[28] Four additional criteria were used in manual selection of a single compromise solution from the set of nondominated solutions: the percentage volume bias (PVB) (equation (7)) [*Boyle et al.*, 2000], the modified correlation coefficient (MCC) (equation (9)) [*Smith et al.*, 2004; *McCuen and Snyder*, 1975], the cumulative sum of absolute monthly volume bias (CMVB) (equation (11)), and the difference between the observed and simulated variance ($\Delta\sigma^2$):

$$PVB = \frac{\sum_{i-1}^{N}(S_i - O_i)}{\sum_{i=1}^{N} O_i} \tag{7}$$

where N is the total number of time steps, S is the simulated flow and O is the observed flow:

$$r = \frac{N\sum_{i=1}^{N} S_iO_i - \sum_{i=1}^{N} S_i \sum_{i=1}^{N} O_i}{\sqrt{\left[N\sum_{i=1}^{N} S_i^2 - \left(\sum_{i=1}^{N} S_i\right)^2\right]\left[N\sum_{i=1}^{N} O_i^2 - \left(\sum_{i=1}^{N} O_i\right)^2\right]}} \tag{8}$$

$$MCC = r\frac{\min(\sigma_{sim}, \sigma_{obs})}{\max(\sigma_{sim}, \sigma_{obs})} \tag{9}$$

where S is the simulated flow, O is the observed flow, N is the number of time steps, σ_{sim} is the standard deviation of the simulated flow and σ_{obs} is the standard deviation of the observed flow:

$$MB_k = \frac{\sum_{j=1}^{M_k}(S_j - O_j)}{\sum_{j=1}^{M_k} O_j} \tag{10}$$

$$\mathrm{CVMB} = \sum_{\mathrm{k}=1}^{12} |MB_k| \tag{11}$$

where MB is the monthly volume bias and M is the number of time steps in the kth month. The compromise solution was subjectively chosen so that PVB, CVMB and $\Delta\sigma^2$ were close to zero and MCC was as close to 1.0 as possible.

2.6. Method for Model Performance Evaluation

[29] The basin outlet (calibration point) performance of the regularized-calibrated model was evaluated over an independent 3 year period not used for model calibration. For the Illinois and Baron Fork basins performance was also evaluated at interior gauging points for which data were available but withheld from use in model calibration.

[30] In each case, model performance was compared against three benchmark runs. The first benchmark (labeled NWS-U), used the parameter values obtained by NWS via *manual* calibration of the ''lumped'' SACSMA model [*Reed et al.*, 2004] applied in uniform fashion to all of the grids of our distributed structure model. This corresponds to transferring parameters from a lumped to distributed version of the model. The second benchmark (labeled NWS-K), applies the *Koren et al.* [2000] spatially distributed parameter estimates to the basin; this corresponds to use of prior parameter estimates without calibration. The third benchmark (labeled NWS-KU), applies the *means* of the Koren spatially distributed parameter values *in uniform fashion over all grids* of the distributed model; when compared to the previous case this illustrates the marginal loss of using uniform instead of distributed parameter fields. For all three benchmark cases, the two routing parameters were adjusted to optimize model performance using the same multiple-criteria calibration approach described above.

[31] Finally, one additional comparison case was run (labeled MO-U), in which the model parameters were treated as uniform over the basin, but calibrated to optimize

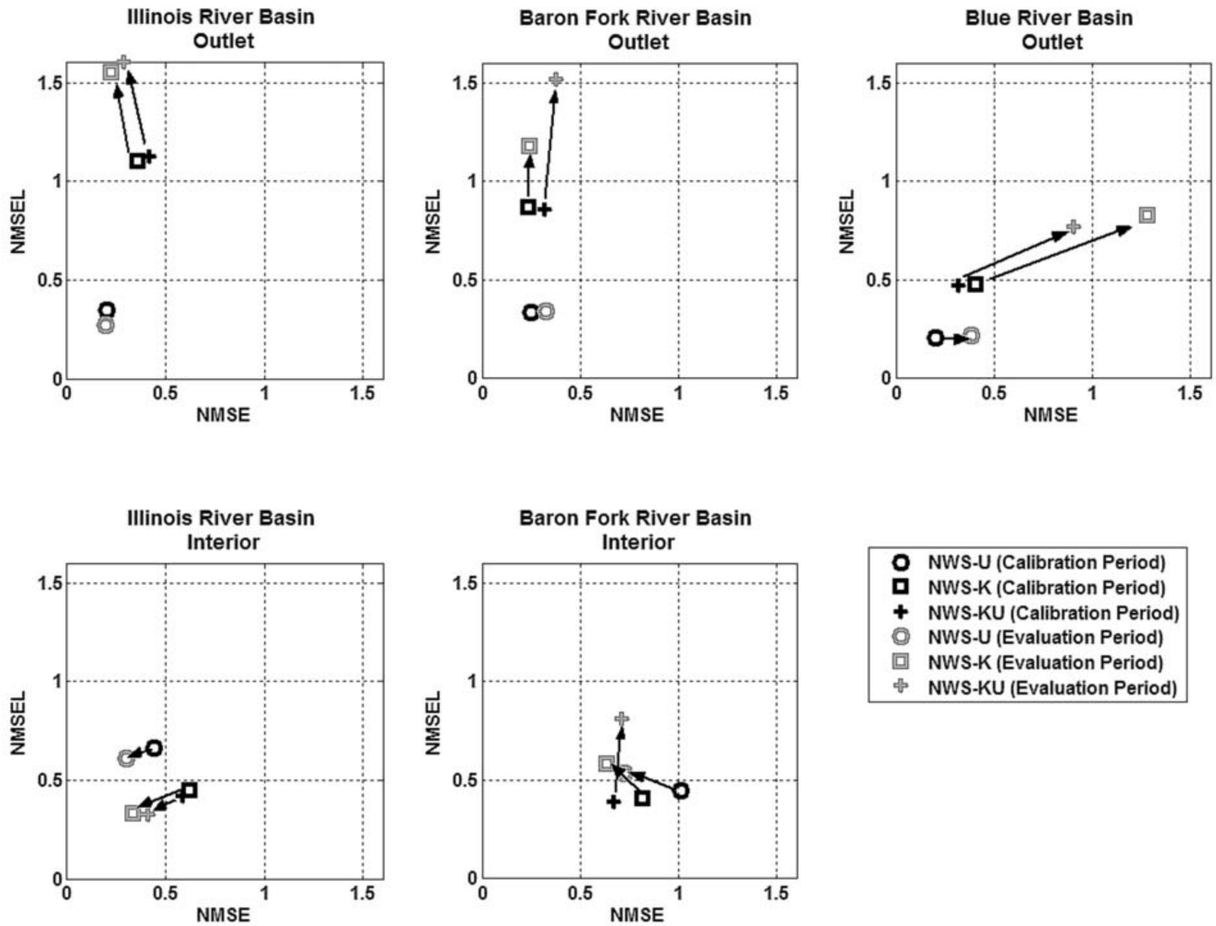

Figure 2. Benchmark runs in all three basins. Arrows indicate the change in function values when going from calibration to evaluation period.

performance at the basin outlet; this corresponds a lack of prior parameter information to constrain the calibration.

3. Results

3.1. Baseline Performance Provided by the Benchmark Runs

[32] The benchmark runs provide a basis against which to compare performance of the various regularized calibration strategies. The top plots in Figure 2 show basin outlet performance at each study basin and the bottom plots show interior gauging point performance (no interior point available for Blue); calibration period is indicated by black symbols and evaluation period by gray symbols. Performance is evaluated using two normalized criteria: the NMSE statistic that has a high-flow emphasis (MSE normalized by the variance of the observed flows) and the NMSEL statistic that has a low-flow emphasis (MSEL normalized by the variance of the log transformed flows). The normalization facilitates comparison across locations and time periods, while partially accounting for temporal nonstationarity of the data variance between data periods (see *Pokhrel et al.* [2009] for a discussion of data nonstationarity in the Blue River basin). Note that 1.0-NMSE corresponds to the popular Nash Sutcliffe efficiency (NSE) measure [*Nash and Sutcliffe*, 1970]; see limitations discussed by *Schaefli and Gupta* [2007] and 1.0-NMSEL corresponds to its counterpart in the log-transformed space (NSEL). Therefore NMSE > 1 and NMSEL > 1 each indicate very poor performance in their corresponding flow transformation space.

[33] Surprisingly, at the outlets of all three basins, the NWS-U benchmark (black circles) consistently provides best model performance ($\sim$0.2–0.4) on both calibration and evaluation periods, with similar performance during both calibration and evaluation. In contrast, the two cases based on Koren prior parameter estimates (black squares and pluses) give relatively poor low-flow performance (NMSEL criterion), particularly on Illinois and Baron Fork, and show marked deterioration in performance from calibration-to-evaluation periods. In the Blue, we see somewhat different behavior with significant deterioration of high-flow performance. Meanwhile, the uniform (NWS-KU, gray pluses) and distributed (NWS-K, gray squares) versions of the Koren prior parameter estimates give very similar performance to each other.

[34] At the interior gauging points the results are not so clear-cut. For the Baron Fork calibration period (black symbols), the NWS-KU (pluses) is better than NWS-K (squares), which is better than NWS-U (circles); in other

words the spatially distributed and uniform applications of the Koren prior parameter estimates both perform better than the NWS manually calibrated estimates. For the evaluation period, NWS-U (gray circles) and NWS-K (gray squares) are similar and better than NWS-KU (gray pluses). For Illinois, the evaluation performance is better than calibration for all cases and the Koren prior parameter estimates perform better than NWS-U.

[35] Overall, the tendency is for the NWS-U uniform parameter set to give better performance at the outlet and for the Koren-based NWS-K spatially distributed prior parameter estimates to give better performance at the interior point. Also, the best model performance is generally poorer at the interior point than at the basin outlet.

3.2. Performance Using Regularized Multiple-Criteria Optimization

[36] The multiple-criteria optimization cases illustrate the impact and relative value of using different types of parameter regularization, and also whether a preferred boundary setting exists for parameter constraining. Each calibration case was run with the 4 boundary settings. For the RMO-m, RMO-a, RMO-b, and MO-U cases, the problem dimension was 13 (1 superparameter for each parameter field, plus 2 routing parameters), and MOSCEM was run with 20 complexes (population size 540 points) and generally terminated after 65 loops, requiring $\sim$24 h on a Macintosh Xserve Quad Xeon 2.66GHz, 4GB RAM machine. For the RMO-ma case, the problem dimension was 24 (2 superparameters per field plus 2 routing parameters), and MOSCEM was run with 35 complexes requiring $\sim$72 h of computer time. For the RMO-mba case, the problem dimension was 35 (3 superparameters per field plus 2 routing parameters), and MOSCEM was run with 45 complexes requiring $\sim$144 h of computer time.

[37] The results are first presented for the Illinois and Baron Fork River basins which are quite similar (and for which interior point data are available), and then for the Blue River basin. For each case we examine the multiple-criteria function tradeoff curve (Pareto frontier), flow duration curve, and posterior parameter distribution.

3.2.1. Illinois and Baron Fork River Basins

3.2.1.1. Multiple-Criteria NMSE Versus NMSEL Performance at the Basin Outlets

[38] Figure 3 shows calibration cases for each basin. In each plot, we show both calibration period (closed dots) and evaluation period (squares) performance. The open black circle corresponds to the best performing benchmark (NWS-U) with calibration performance indicated by black and evaluation performance by gray. For the Illinois and Baron Fork river basins, these plots show the following.

[39] 1. The four boundary settings performed comparably well, due to several related reasons: (1) the prior parameter distributions for these basins are narrow compared to their feasible parameter ranges (relative spatial variation is small), (2) the optimization did not try to push the parameter distributions outside of the bounds, and (3) parameter sensitivity was low for the few parameter fields that did hit the bounds. Therefore, to keep the plots simple we show only the TB (black symbols) and TBS cases (gray symbols).

[40] 2. All six calibration cases (the five regularized cases RMO-m, RMO-a, RMO-b, RMO-ma, and RMO-mba, and the unregularized MO-U case) generally perform better than the best benchmark run (NWS-U) at the outlet, during both calibration and evaluation, indicating that the calibration strategy is successful at finding improved (posterior) parameter fields. Further, the performance change from calibration to evaluation periods is similar for all five regularization cases.

[41] 3. There is no significant performance difference between the multiplier (RMO-m), power (RMO-b) and adder (RMO-a) types of regularized calibration (the positions of the NMSE versus NMSEL Pareto frontiers are almost identical).

[42] 4. There is only slight improvement when increasing the degrees of freedom to RMO-ma, and no added benefit in using RMO-mba. For this reason, the latter results are not shown.

[43] 5. The MO-U calibration case, which uses a *uniform* (as opposed to spatially varying) parameter distribution, gives comparable performance during calibration but does not perform as well during evaluation for Baron Fork.

[44] Admittedly these are only two study locations, but the results suggest that calibration using regularized spatially distributed parameter fields gives some performance benefit at the basin outlet (compared to spatially uniform fields), albeit marginal. The surprising result is that the different kinds of individual superparameter regularization (RMO-m, RMO-b, and RMO-a) give very similar NMSE versus NMSEL performance.

3.2.1.2. Parameter Values

[45] Figure 4 shows the posterior parameter distributions (we show only the TB case for the Illinois). Each set of plots shows a different model parameter, arranged from top to bottom as UZTWM, UZFWM, UZK, REXP, PFREE, LZTWM, LZFPM, LZFSM, LZPK, LZSK and ZPERC (parameter functions explained in Table 1). The x axis shows the parameter range, the gray scale indicates the shape of the parameter frequency distribution (darker values indicate higher frequency), and the y axis shows how the parameter distribution varies from one end of the Pareto frontier to the other (indicated in terms of the NMSE value; high NMSE corresponds to low NMSEL and vice versa). These plots show the following.

[46] 1. The RMO-m, RMO-b, and RMO-a strategies have all arrived at similar posterior parameter distributions (compare the three left sets of plots).

[47] 2) For some parameters (UZK, REXP, LZFPM, LZFSM), the distribution does not vary significantly across the Pareto frontier (top to bottom along the y axis), indicating little sensitivity to high-flow/low-flow performance trade-off. Other parameters (UZFWM, PFREE, and LZPK) show strong high-flow/low-flow performance sensitivity.

[48] 3.) The three sets of plots on the right show the corresponding optimized values of the superparameters, leading to a similar conclusion. Here, m = 1, b = 1 or a = 0 indicates no change of the posterior parameter distribution from its prior value. A sloping line indicates high-flow/low-flow tradeoff sensitivity across the Pareto frontier while a vertical line indicates none. The degree of identifiability of each parameter field is indicated by the thickness of the top to bottom line; a thin line with little scatter indicates greater identifiability (UZFWM, LZPK) than a line having large scatter (ZPERC, REXP, LZPK).

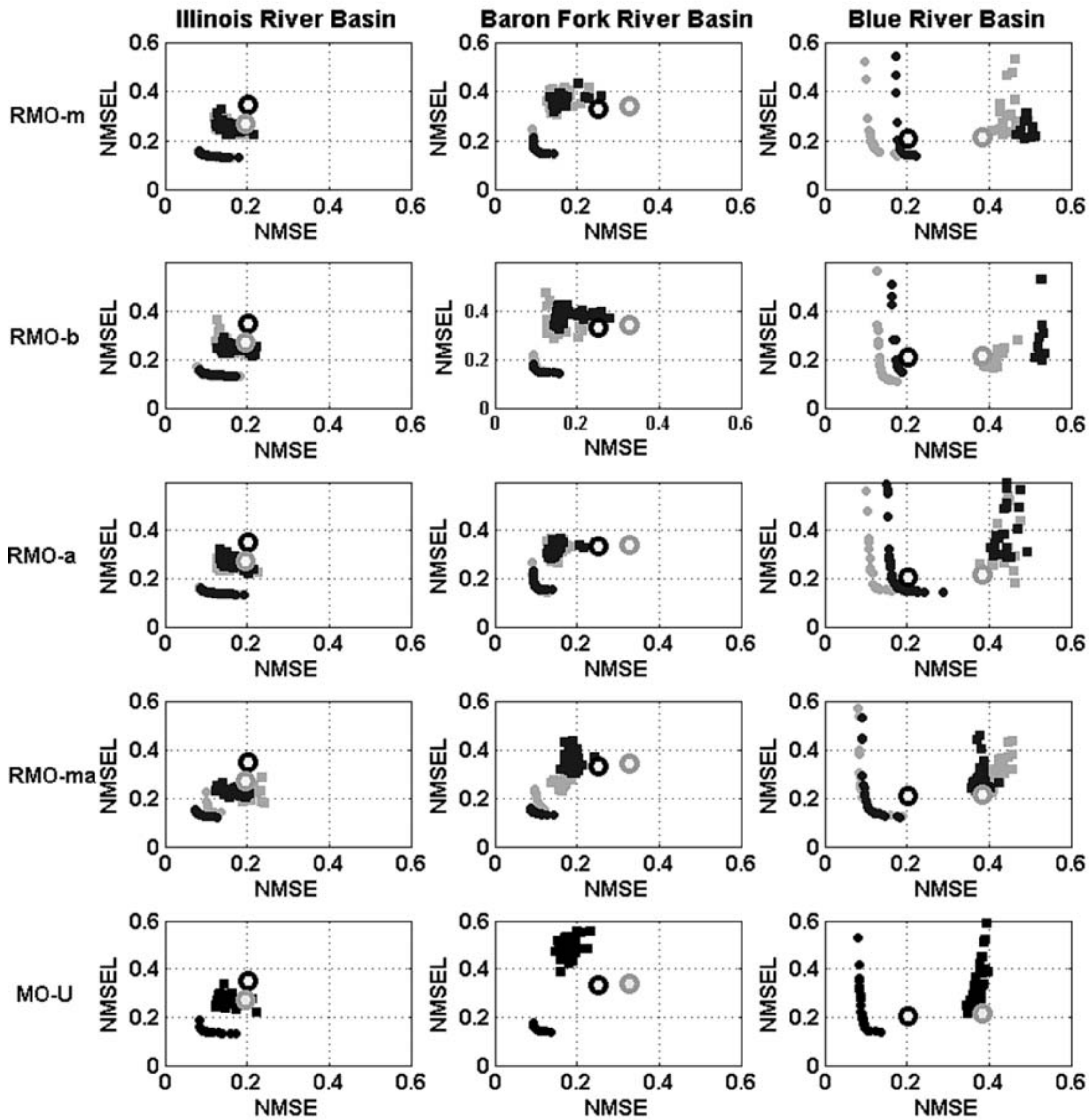

Figure 3. Pareto frontier obtained (during calibration and evaluation periods) from model simulations using the data from the outlet of Illinois, Baron Fork, and Blue River basins: Pareto frontier obtained during (top to bottom) RMO-m, RMO-b, RMO-a, RMO-ma, and MO-U. Black, TB; gray, TBS. Solid dots represent Pareto frontier at the calibration period, and the solid squares represent the evaluation period. Benchmark symbols (NWS-U) are given by open circles (black, calibration period; gray, evaluation period).

3.2.1.3. Multiple-Criteria NMSE Versus NMSEL Performance Results at Interior Points

[49] Figure 5 shows the multiple-criteria performance results, with calibration and evaluation period results separated into different plots. The following points are immediately clear.

[50] 1. Simulation performance is considerably worse at the interior basin points than at the basin outlets, in terms of both high-flow (NMSE) and low-flow (NMSEL) performance (compare Figure 5 with Figure 3).

[51] 2. The model performance obtained by the multiplier, power and adder superparameter regularization strategies (RMO-m, RMO-b, and RMO-a) and the unregularized case (MO-U) are similar.

[52] 3. The multiplier-plus-adder regularization case (RMO-ma) gives slightly better high-flow and worse low-flow performance.

[53] 4. When compared with the *prior* parameter estimates, the calibration runs generally provide better reproduction of high flows (lower NMSE) while the Koren-based NWS-K (gray squares) and NWS-KU (gray pluses) parameter estimates provide better reproduction of low flows (lower NMSEL). However, there is some inconsistency: on Baron Fork, the NWS-KU (gray pluses) prior parameter

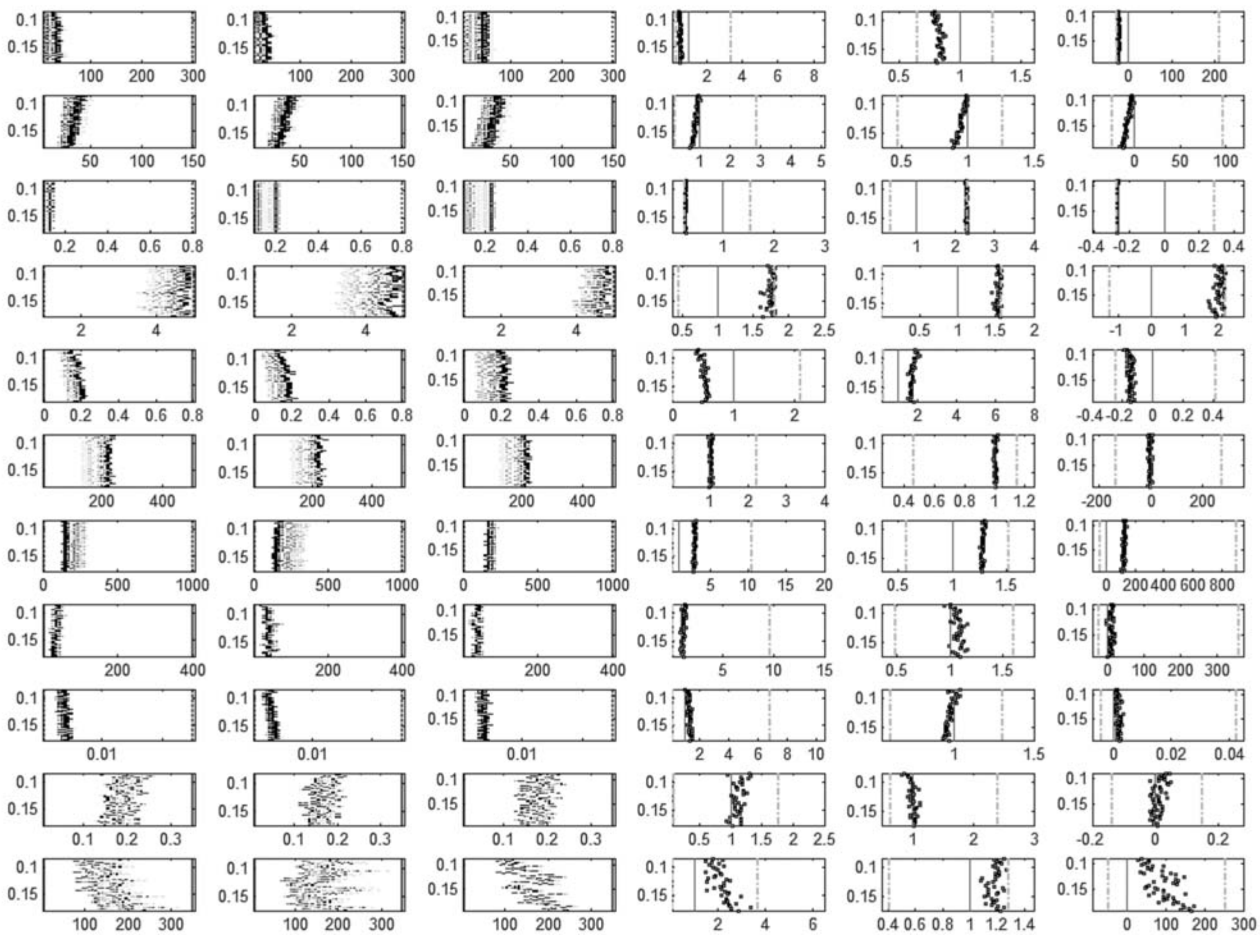

Figure 4. Two-dimensional parameter distribution plots for Illinois River basin using the TB boundary settings for SACSMA parameters (top to bottom) UZTWM, UZFWM, UZK, REXP, PFREE, LZTWM, LZFPM, LZFSM, LZPK, LZSK, and ZPERC. From left to right, the first three sets of plots represent SACSMA parameter distribution for RMO-m, RMO-b, and RMO-a cases, and the next three sets of plots represent superparameter distributions in the same order. The gray vertical dashed lines in the last three sets of plots correspond to the range within which the superparameters were allowed to vary for the TB boundary setting (the total range of the plot corresponds to TBS setting). The solid gray lines correspond to prior values of the superparameters (either 0 or 1).

set gives better overall performance during calibration but worse overall performance during evaluation.

[54] 5. The NWS-K (gray squares) provides the most consistent overall performance (although poor on high flows).

3.2.1.4. Flow Duration Curves

[55] Figure 6 shows the flow duration curves (FDCs) (more correctly called flow exceedance plots or cumulative distribution curves) for the compromise solution parameter set selected for the Illinois Basin. The top plots show the benchmark cases and the bottom plots show the calibration cases. The first and second sets of plots correspond to the basin outlet while the third and fourth sets of plots correspond to the interior point. For ease of comparison, a thick black line is used to indicate the observations. The results show the following.

[56] 1. The three regularized calibration cases (different kinds of stars) all show remarkably good (and very similar) fits to the observed FDC at the basin outlet during calibration, but with a slight tendency toward underestimation during evaluation. At the interior point, they show a strong tendency toward overestimation of low flows.

[57] 2. The NWS-K (squares) and NWS-KU (pluses) benchmarks behave almost alike and tend to underestimate the FDC at the basin outlet (on both calibration and evaluation periods). At the interior point, they give the best performance of all the cases (calibrated and benchmark).

[58] 3. The NWS-U (circles) benchmark parameter set and the MO-U (triangles) calibrated parameter set track the observed FDC well at the basin outlet (on both calibration and evaluation periods), but tend to overestimate low flows at the interior point.

[59] In summary, the regularized calibration runs give best performance at the basin outlet while the NWS-K and NWS-KU benchmark parameter sets (based on the Koren spatially distributed prior parameter estimates) give the best performance at the interior point. The three types of superparameter regularization give similar calibration and evalu-

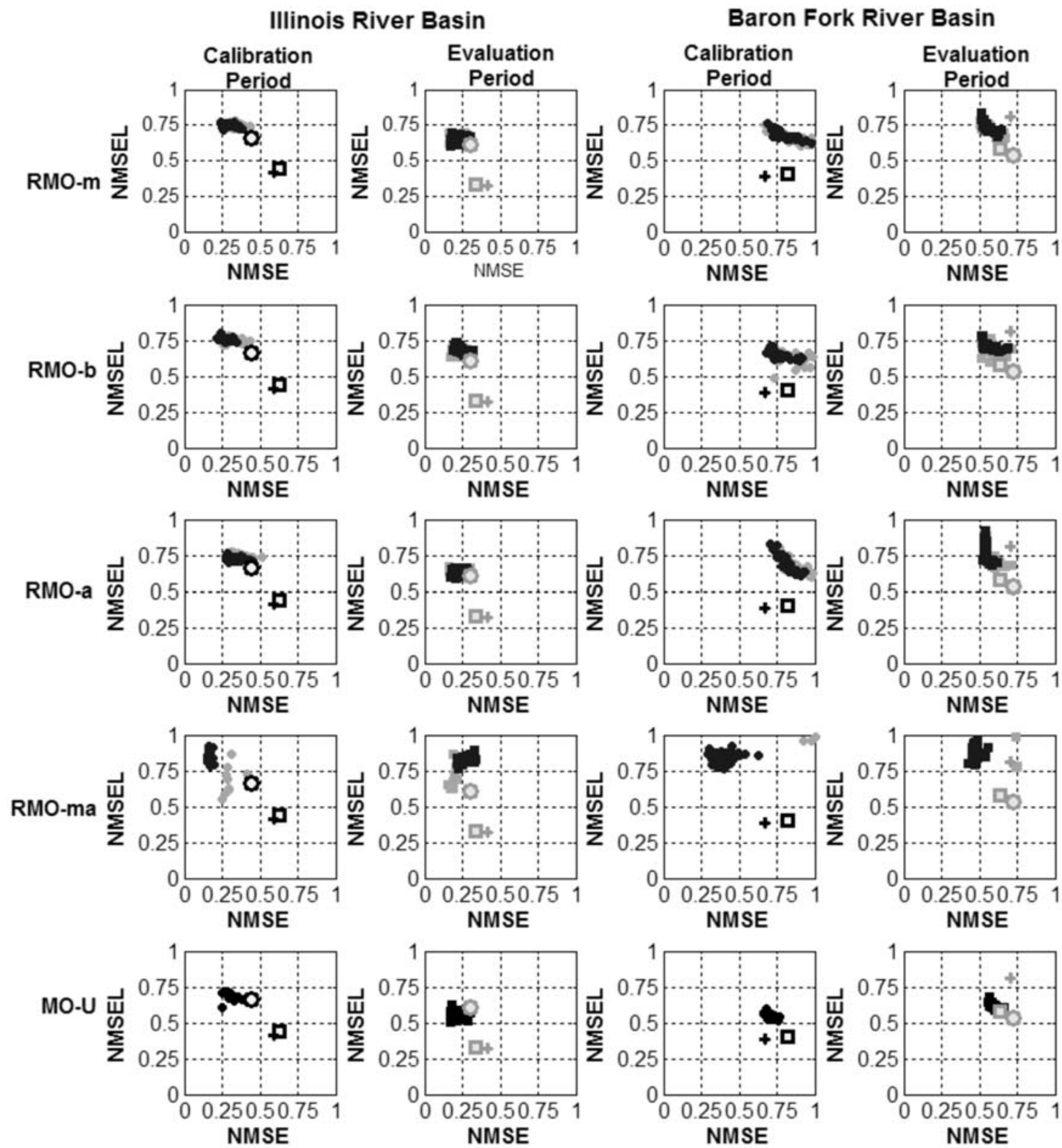

Figure 5. Pareto frontier obtained (during the calibration and evaluation periods) from model simulations using the data from the interior points of Illinois and Baron Fork River basins for (top to bottom) RMO-m, RMO-b, RMO-a, RMO-ma, and MO-U. Black, TB settings; gray, TBS settings. Benchmark symbols: pluses, NWS-KU; squares, NWS-K; open circles, NWS-U.

ation period performance. Increasing the degrees of freedom (by increasing complexity of the regularization equation) did not provide significant performance improvements for these two basins. The use of different boundary settings did not provide added benefit over the use of hard constraints.

3.2.2. Blue River Basin

3.2.2.1. Multiple-Criteria NMSE Versus NMSEL Performance Results at Basin Outlets

[60] The results show the following (third sets of plots in Figure 3).

[61] 1. There is little apparent difference in performance among the three individual superparameter regularized calibration cases.

[62] 2. The best overall performance is provided by the RMO-ma calibration (during both calibration and evaluation periods), indicating that the increased degrees of freedom results in improved performance, mainly in reproduction of high flows.

[63] 3. The Blue behaves differently from the Illinois or Baron Fork. In contrast with those basins, the main deterioration in calibration to evaluation period performance is in the reproduction of high flows (NMSE).

[64] 4. Because the prior parameter distributions for the Blue are much wider relative to the feasible parameter ranges, the squashing function enables significant improvements in high-flow model performance by permitting distortions of the parameter distributions (compare TBS Pareto frontier (gray dots) to TB Pareto frontier (black dots)). This performance improvement persists from calibration to evaluation periods.

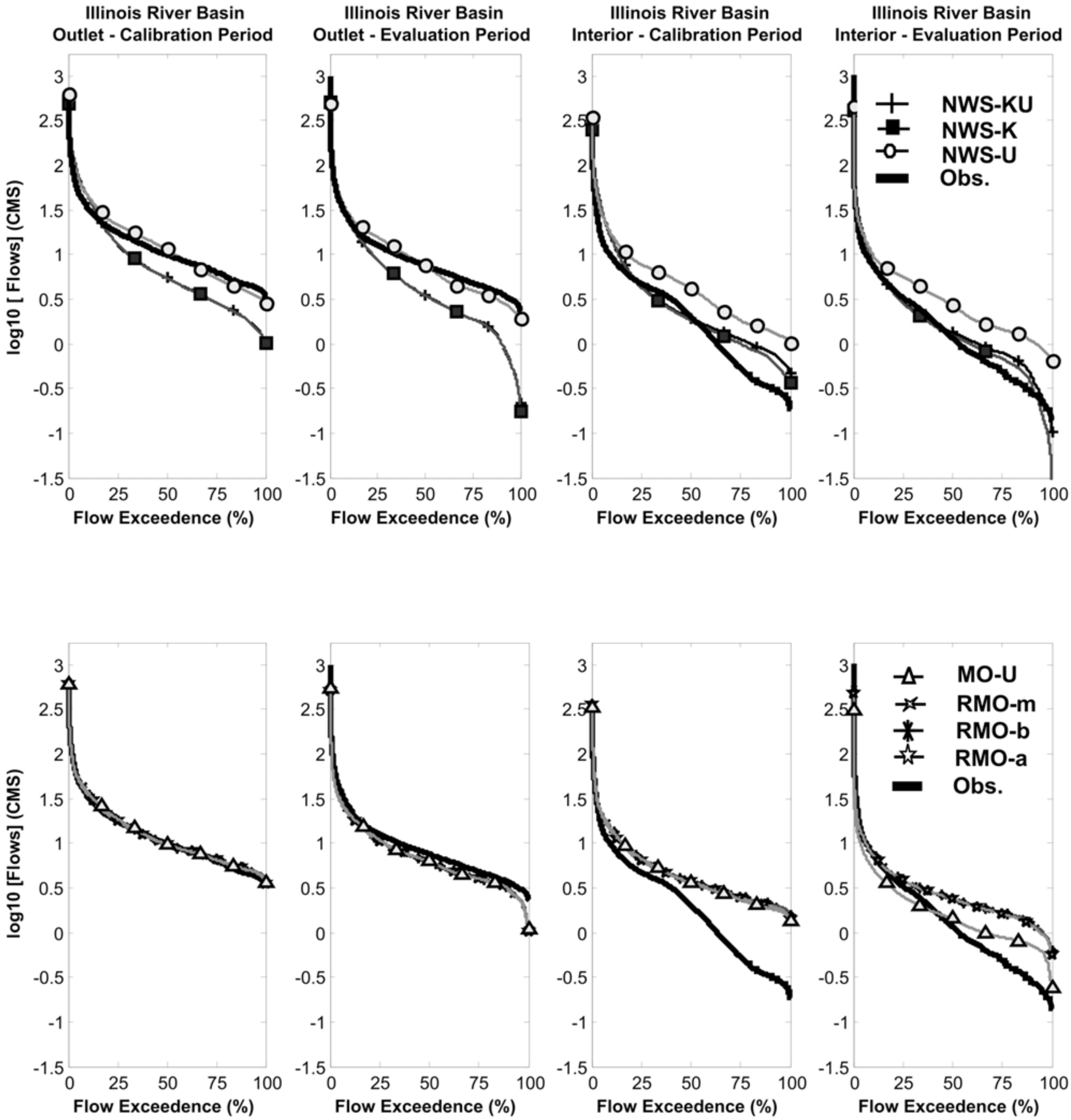

Figure 6. Flow duration curves obtained using the simulations from the outlet and the interior of the Illinois River Basin.

[65] 5. The MO-U calibration (using uniform parameter sets) achieves similar performance to the RMO-ma calibration (using spatially distributed parameter sets), at the basin outlet.

[66] In addition, FDCs at an interior point (selected roughly at the middle of the basin) showed that the RMO-ma results in lower-flow volumes compared to MO-U (plots not shown). Unfortunately, no data exist to test the model performance at this (interior) point.

3.2.2.2. Parameter Values

[67] Figure 7 shows the posterior parameter distributions obtained by regularized calibration using the tight boundary setting (TB). This case is shown because it clearly illustrates the problems of imposing feasible parameter constraints when performing distributed parameter calibration.

[68] 1. Contrary to Illinois and Baron Fork, the parameter distributions obtained by the different regularization schemes are quite different.

[69] 2. Variations in parameters UZTWM (top set of plots in Figure 7) and UZK (third set of plots in Figure 7) account for most of the calibration performance improvements obtained on this basin, particularly with respect to high flows (NMSE).

[70] These results are explained by the following facts. First, the Blue has a lower runoff ratio than the other two basins (0.17 compared to 0.26–0.29), and its fractional

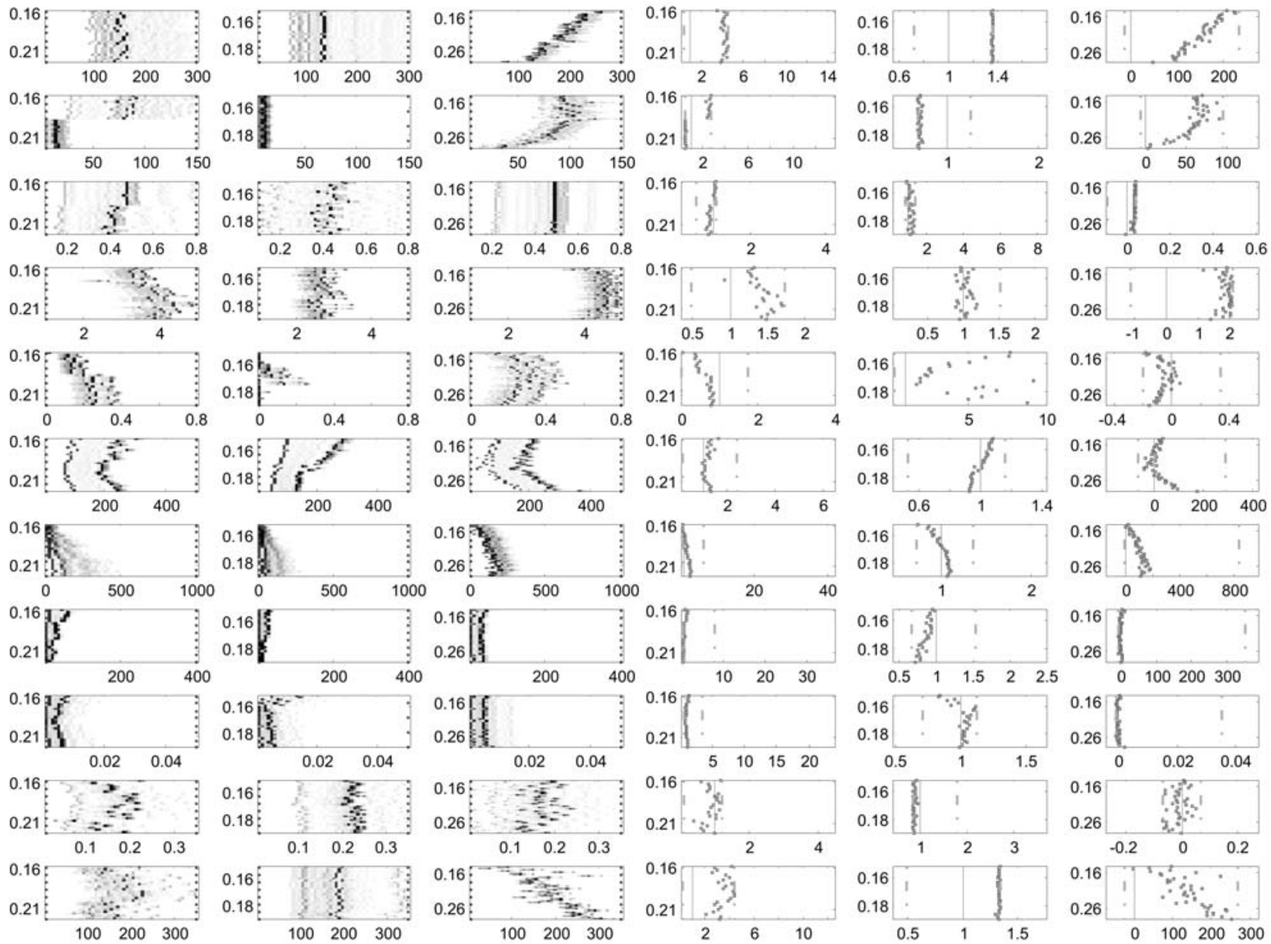

Figure 7. Two-dimensional parameter distribution plots for Blue River basin using the TB boundary settings for SACSMA parameters (top to bottom) UZTWM, UZFWM, UZK, REXP, PFREE, LZTWM, LZFPM, LZFSM, LZPK, LZSK, and ZPERC. From left to right, the first three sets of plots represent SACSMA parameter distributions for RMO-m, RMO-b, and RMO-a cases, and the next three sets of plots represent superparameter distributions in the same order. The gray vertical dashed lines in the last three sets of plots correspond to the range within which the superparameters were allowed to vary for the TB boundary setting (the total range of the plot corresponds to TBS setting). The solid gray lines correspond to prior values of the superparameters (either 0 or 1).

contribution of interflow is around 65–70% (compared to 15–20%). UZTWM, which controls about 90% of the total evapotranspiration (ET), therefore needs to be large to simulate the larger ET losses. This in turn reduces the fraction of time that the upper zone is saturated, and results in the need for larger values of the parameter UZK to generate sufficient interflow to match the shape and timing of the flood peaks. These findings are consistent with *Yilmaz et al.* [2008] and *Wagener et al.* [2009].

[71] Second, a problem arises in the way that the optimization wants to modify the prior spatial distributions (Figures 7 and 8, top). For UZTWM (size of the upper zone tension storage capacity), the multiplier (RMO-m) and power (RMO-b) regularization strategies cause the variance (width) of the UZTWM parameter distribution to increase, so that the distribution runs up against the feasible bounds and the mean value is prevented from changing. When the adder (RMO-a) strategy is used, the parameter variance remains constant allowing the parameter mean to change without being constrained by the bounds; hence the mean of the UZTWM distribution can vary significantly across the Pareto frontier (Figure 7, top), preferring higher values at lower NMSE (high-flow performance), consistent with a tradeoff between high- and low-flow fitting.

[72] Third, the prior distribution of UZK (upper zone lateral drainage rate) covers most of the feasible parameter range (Figure 8, third set of plots from the top), so that none of the individual superparameter regularization strategies can distort the distribution sufficiently to allow the parameter mean to increase. We conducted an additional regularized calibration run (not shown here), in which all parameters were treated as spatially distributed while UZK was made spatially uniform; the value of UZK moved very close to the upper bound.

[73] Together, this explains why the squashing function (TBS setting) and looser bounds (LB and LBTI settings) improved the NMSE (high-flow) performance on the Blue River basin. It also explain why RMO-ma and MO-U calibrations did so well; RMO-ma can vary the mean and variance of the distribution independently thereby remain-

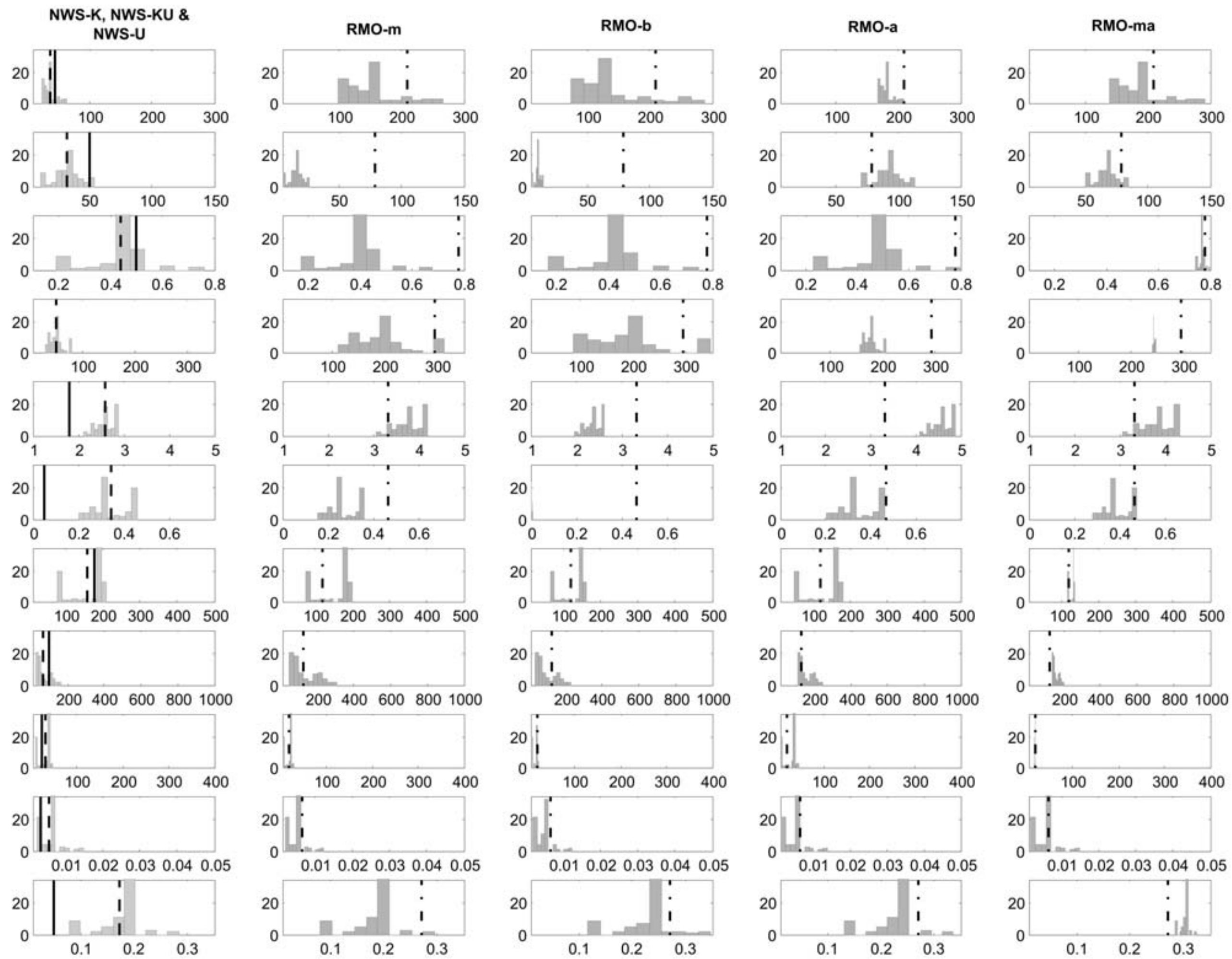

Figure 8. Blue River basin parameter distribution plot for RMO calibration (TB settings) for (top to bottom) UZTWM, UZFWM, UZK, ZPERC, REXP, PFREE, LZTWM, LZFPM, LZFSM, LZPK, and LZSK. From left to right, the first set of plots represents NWS-K (gray histogram), NWS-U (solid black line), and NWS-KU (dashed black line). The gray histograms in the next three sets of plots represent RMO-m, RMO-b, and RMO-a, respectively, and the dash-dotted line represents MO-U. In the far right set of plots, the gray histogram represents RMO-ma, and the dash-dotted line represents MO-U.

ing relatively unconstrained by the boundaries (the final distribution for UZK has a mean close to the upper boundary and a very small variability; see Figure 8, third plot from the top on the far right).

[74] Overall, for the Blue River at the outlet, the mean values of the parameter fields may be more important than their spatial distributions. Of course, this result is very likely *not* true for flow simulations at interior points, which cannot be checked without the existence of interior point data.

4. Summary, Conclusions, and Discussion

4.1. Summary

[75] This study has examined whether a regularized multiple-criteria calibration strategy, based on the use of prior spatial parameter distribution information, can improve the performance of a distributed hydrological model. Our nonlinear regularization strategy has three superparameters per parameter field and is designed to preserve the relative spatial ordering of the *magnitudes* of the parameter values across grid cells. We explored the impact (and relative value) of using each of the three superparameter types (multiplicative, additive and power terms) independently and in combinations. We also examined the use of simple squashing functions to ensure that the adjusted parameter distribution does not violate any boundaries imposed by conceptual-computational consistency, thereby maintaining reasonable values throughout the spatial domain.

[76] The method was tested on three study basins, using spatial parameter distribution information provided by the *Koren et al.* [2000] approach. Our *fundamental assumption* is that these prior parameter estimates provide valuable information about relative spatial ordering and magnitude of the parameters, but that adjusting the shape and position of their frequency distributions can help to improve model performance. The calibration was performed using streamflow data at the basin outlets, and performance was evaluated on an independent evaluation period *and* at the basin interior points.

[77] The overall results can be summarized as follows.

[78] 1. The regularized multiple-criteria calibration considerably improved performance *at the basin outlet*, indicating success in finding improved (posterior) parameter fields.

[79] 2. *At interior points*, the Koren prior parameter estimates give better performance than the calibrated posterior fields, although model performance is generally much poorer than at the basin outlet.

[80] 3. None of the three superparameter types showed clear advantages when used individually. They tended to arrive at similar posterior parameter distributions.

[81] 4. Increasing degrees of freedom to allow more complex distortions of the parameter distributions did not help for basins where prior spatial variability of the parameter fields was *small* compared to the feasible ranges, but did help where the prior spatial variability of the parameter fields was *large*.

[82] 5. Use of a squashing function in conjunction with the commonly used "tight bounds" parameter constraining strategy helped to improve model performance.

[83] 6. Unregularized multiple-criteria calibration, using uniform parameter fields (ignoring prior spatial information), provided basin outlet performance comparable to that achieved by regularized calibration (except for Baron Fork). However, interior point performance was poor, especially for high flows.

4.2. Conclusions and Discussion

[84] In conclusion, multiple-criteria regularized calibration significantly altered the prior distributions of parameter values, giving consistent improvements in model performance at the basin outlet that persisted during evaluation. However, calibration using uniform parameters gave comparable performance, indicating that spatial parameter variability was not particularly helpful in improving basin outlet performance. These results suggest that the *outlet response* for *these basins* is not strongly sensitive to spatial parameter variability so that an overall basin mean parameter value performs just as well. However, we cannot conclude from these results that the spatial parameter information is not valuable, because our approach has the characteristic of lumping together all the temporal hydrograph information into a single time-aggregated measure, making it difficult to extract information about spatial parameter variability from the basin outlet hydrographs. Further, there can be several causes of the lack of outlet sensitivity to spatial parameter variation, including structural considerations such as the smoothing effects of channel routing, and in general the reasons will require considerable additional investigation to uncover. The temporal sensitivity of hydrographs to spatial information [e.g., *van Werkhoven et al.*, 2008a] is one factor that needs to be pursued, so that methods for improving information extraction can be designed.

[85] In this regard, an important result is that the Koren prior parameter estimates (NWS-K and NWS-KU) gave better performance at the interior points, than obtained by any of the calibrations. Clearly these prior parameter distributions provide considerable value, and calibration made matters worse at the interior points. Two major questions that arise from this are (1) Do outlet hydrographs provide sufficient information to properly infer the catchment behavior in the interior? (2) What calibration methods can be used to obtain better representation of watershed behavior at interior points?

[86] Clearly, there is a clear need to design improved calibration strategies that achieve performance improvements at the calibration points, incorporate regularization based on spatial parameter patterns, and prevent prior parameter estimates from being changed in locations showing little sensitivity to calibration point data. By extension, this sensitivity should be evaluated for both spatial and temporal variations (rather than treated as fixed, as in this study), to improve overall information extraction (suggested also by *Gupta et al.* [2006] and *Wagener and Gupta* [2005]).

[87] Overall, our results support the use of a regularization strategy having the form $\phi_j = m \cdot (\theta_j - E\{\theta\}) + E\{\theta\} + a$, where the adder a shifts the distribution mean and the multiplier m adjusts the distribution variance. In this way, the superparameters will have a clearer interpretation, such that

$$E\{\phi\} = E\{\theta\} + a \tag{12}$$

$$Var\{\phi\} = m^2 \cdot Var\{\theta\} \tag{13}$$

[88] Finally, this work has explored using two kinds of information to enhance the calibration of distributed models: (1) information in hydrographs and (2) information in prior estimates of spatially distributed parameter fields. However, more sophisticated ways of extracting information from these two sources need to be explored, including use of diagnostically more powerful criteria that have better hydrological relevance [*Gupta et al.*, 2008; *Yilmaz et al.*, 2008; *Pokhrel et al.*, 2009; *Herbst et al.*, 2009; *Gupta et al.*, 2009] and are better able to recognize and exploit spatial and temporal variability in the input activations [*van Werkhoven et al.*, 2008a, 2008b]. Other strategies for constraining the parameter space could include the use of interior point data [*Khu et al.*, 2008], soft information [*Seibert and McDonnell*, 2002; *Dunn*, 1999], and manual calibration within a Tikhonov regularization framework [*Tikhonov and Arsenin*, 1977; *Tonkin and Doherty*, 2005]. As always, we invite dialog on these and related issues related to hydrological model identification.

[89] **Acknowledgments.** Partial support for this work was provided by the Hydrology Laboratory of the National Weather Service (grant NA04NWS4620012) and by SAHRA (Center for Sustainability of semi-Arid Hydrology and Riparian Areas) under the STC program of the National Science Foundation (agreement EAR 9876800). The first author was also supported by a fellowship provided by the Salt River Project, Phoenix, Arizona, and by the World Laboratory International Center for Scientific Culture, Geneva, Switzerland. These sources of support are gratefully acknowledged.

References

Anderson, R. M., V. I. Koren, and S. M. Reed (2006), Using SSURGO data to improve Sacramento Model a priori parameter estimates, *J. Hydrol.*, *320*, 103–116, doi:10.1016/j.jhydrol.2005.07.020.

Bandaragoda, C., D. G. Tarboton, and R. Woods (2004), Application of TOPNET in the distributed model comparison project, *J. Hydrol.*, *298*, 178–201, doi:10.1016/j.jhydrol.2004.03.038.

Behnaz, K., B. Imam, K. Hsu, and S. Sorooshian (2009), From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models, *J. Hydrol.*, in press.

Beven, K. J. (1989), Changing ideas in hydrology—The case of physically based models, *J. Hydrol.*, *105*, 157–172, doi:10.1016/0022-1694(89)90101-7.

Beven, K. J. (1996), A discussion of distributed hydrological modeling, in *Distributed Hydrological Modelling*, edited by M. B. Abbott and J. C. Refsgaard, pp. 255–278, Kluwer Acad., Dordrecht, Netherlands.

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

Brath, A., A. Montanari, and E. Toth (2004), Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially distributed hydrological model, *J. Hydrol.*, *291*, 232–253, doi:10.1016/j.jhydrol.2003.12.044.

Burnash, R., R. Ferral, and R. McGuire (1973), A generalized streamflow simulation system conceptual modeling for digital computers, State of California, report, Dep. of Water Resour., U.S. Dep. of Commer., Natl. Weather Serv., Sacramento, Calif.

Carrera, J., and S. P. Neuman (1986), Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199–210, doi:10.1029/WR022i002p00199.

Cornfield, H. E., and V. L. Lopes (2004), Parameter identification in a two-multiplier sediment yield model, *J. Am. Water Resour. Assoc.*, *40*(2), 321–332, doi:10.1111/j.1752-1688.2004.tb01032.x.

Demarty, J., C. Ottl, I. Braud, A. Olioso, J. P. Frangi, H. V. Gupta, and L. A. Bastidas (2005), Constraining a physically based SVAT model with surface water content and thermal infrared brightness temperature measurements using a multiobjective approach, *Water Resour. Res.*, *41*, W01011, doi:10.1029/2004WR003695.

Doherty, J., and B. E. Skahill (2006), An advanced regularization methodology for use in watershed model calibration, *J. Hydrol.*, *327*, 564–577, doi:10.1016/j.jhydrol.2005.11.058.

Dunn, S. M. (1999), Imposing constraints on parameter values of a conceptual hydrological models using baseflow response, *Hydrol. Earth Syst. Sci.*, *3*(2), 271–284.

Eckhardt, K., and J. G. Arnold (2001), Automatic calibration of a distributed catchment model, *J. Hydrol.*, *251*, 103–109, doi:10.1016/S0022-1694(01)00429-2.

Francés, F., J. I. Vélez, and J. J. Vélez (2007), Split-parameter structure for the automatic calibration of distributed hydrological models, *J. Hydrol.*, *332*, 226–240, doi:10.1016/j.jhydrol.2006.06.032.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Towards improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763, doi:10.1029/97WR03495.

Gupta, H. V., S. Sorooshian, T. S. Hogue, and D. P. Boyle (2003), Advances in automatic calibration of watershed models, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 9–28, AGU, Washington, D. C.

Gupta, H. V., K. J. Beven, and T. Wagener (2006), Model calibration and uncertainty estimation, in *Encyclopedia of Hydrologic Sciences*, pp. 2015–2031, John Wiley, New York.

Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*(18), 3802–3813, doi:10.1002/hyp.6989.

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

Herbst, M., H. V. Gupta, and M. C. Casper (2009), Mapping model behaviour using self-organizing maps, *Hydrol. Earth Syst. Sci.*, *13*, 395–409.

Isaaks, E. H., and R. M. Srivastava (1989), *Applied Geostatistics*, Oxford Univ. Press, New York.

Khu, S. T., H. Madsen, and F. di Pierro (2008), Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, *Adv. Water Resour.*, *31*, 1387–1398, doi:10.1016/j.advwatres.2008.07.011.

Koren, V. I., M. Smith, D. Wang, and Z. Zhang (2000), Use of soil property data in the derivation of conceptual rainfall runoff model parameters, paper presented at 15th Conference on Hydrology, Am. Meteorol. Soc., Long Beach, Calif.

Koren, V., M. Smith, and Q. Duan (2003), Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 239–254, AGU, Washington, D. C.

Lawson, C. L., and J. R. Hanson (1995), *Solving Least Squares Problems*, *Classics Appl. Math.*, vol. 15, 337 pp., Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Leavesly, G. H., L. E. Hay, R. J. Viger, and S. L. Markstrom (2003), Use of a priori parameter-estimation methods to constrain calibration of distributed-parameter models, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 9–28, AGU, Washington, D. C.

Leplastrier, M., A. J. Pitman, H. V. Gupta, and Y. Xia (2002), Exploring the relationship between complexity and performance in a land surface model using the multi-criteria method, *J. Geophys. Res.*, *107*(D20), 4443, doi:10.1029/2001JD000931.

Linden, J., B. Vinsonneau, and K. J. Burnham (2005), Review and enhancement of cautious parameter estimation for model based control: A specific realisation of regularization, paper presented at International Conference on Systems Engineering, Inst. of Electr. and Electron. Eng., Las Vegas, Nev., 16–18 Aug.

Madsen, H. (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives, *Adv. Water Resour.*, *26*, 205–216, doi:10.1016/S0309-1708(02)00092-1.

McCuen, R. H., and W. M. Snyder (1975), A proposed index for comparing hydrographs, *Water Resour. Res.*, *11*(6), 1021–1024, doi:10.1029/WR011i006p01021.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models. Part I: A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.

NOAA (2007), DHM grid scalar editor program—User manual, Off. of Hydrol. Dev., Natl. Weather Serv., Silver Spring, Md. (Available at http://www.nws.noaa.gov/oh/hrl/hseb/docs/DHM_Grid_Scalar_Editor_User_Manual.pdf)

Pokhrel, P. (2007), Estimation of spatially distributed model parameters using a regularization approach, M.S. thesis, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.

Pokhrel, P., H. V. Gupta, and T. Wagener (2008), A spatial regularization approach to parameter estimation for a distributed watershed model, *Water Resour. Res.*, *44*, W12419, doi:10.1029/2007WR006615.

Pokhrel, P., K. K. Yilmaz, and H. V. Gupta (2009), Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures, *J. Hydrol.*, doi:10.1016/2008.12.004, in press.

Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, D.-J. Seo, and DMIP participants (2004), Overall distributed model intercomparison project results, *J. Hydrol.*, *298*(1–4), 27–60, doi:10.1016/j.jhydrol.2004.03.031.

Refsgaard, J. C. (1997), Parameterization, calibration and validation of distributed hydrological models, *J. Hydrol.*, *198*(1–4), 69–97, doi:10.1016/S0022-1694(96)03329-X.

Refsgaard, J. C., and B. Storm (1996), Construction, calibration and validation of hydrological models, in *Distributed Hydrological Modeling*, edited by M. B. Abbott and J. C. Refsgaard, pp. 41–54, Kluwer Acad., Dordrecht, Netherlands.

Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processes*, *21*(15), 2075–2080, doi:10.1002/hyp.6825.

Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.

Smith, M. B., D.-J. Seo, V. I. Koren, S. Reed, Z. Zhang, Q.-Y. Duan, S. Cong, F. Moreda, and R. Anderson (2004), The Distributed Model Intercomparison Project (DMIP): Motivation and experiment design, *J. Hydrol.*, *298*(1–4), 4–26, doi:10.1016/j.jhydrol.2004.03.040.

Smith, M., V. Koren, S. Reed, Z. Zhang, D. J. Seo, F. Moreda, and Z. Cui (2006), The Distributed Model Intercomparison Project: Phase 2 science plan, Hydrol. Lab., Off. of Hydrol. Dev., Natl. Weather Serv., NOAA, Silver Spring, Md. (Available at http://www.weather.gov/oh/hrl/dmip/2/docs/dmip_2_plan_march10_06.pdf)

Tikhonov, A., and V. Arsenin (1977), *Solutions of Ill-Posed Problems*, Halsted, New York.

Tonkin, M., and J. Doherty (2005), A hybrid regularized inversion methodology for highly parameterized environmental models, *Water Resour. Res.*, *41*, W10412, doi:10.1029/2005WR003995.

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

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008b), Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, *Geophys. Res. Lett.*, *35*, L11403, doi:10.1029/2008GL034162.

Vieux, B. E. (2001), *Distributed Hydrologic Modeling Using GIS*, *Water Sci. Technol. Ser.*, vol. 38, 1st ed., Kluwer Acad., Norwell, Mass.

Vieux, B. E., and F. G. Moreda (2003), Ordered physics-based parameter adjustment of a distributed model, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 9–28, AGU, Washington, D. C.

Vieux, B. E., Z. Cui, and A. Gaur (2004), Evaluation of a physics-based distributed hydrologic model for flood forecasting, *J. Hydrol.*, *298*(1–4), 155–177, doi:10.1016/j.jhydrol.2004.03.035.

Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, *39*(8), 1214, doi:10.1029/2002WR001746.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *J. Stochastic Environ. Res. Risk Assess.*, *19*, 378–387, doi:10.1007/s00477-005-0006-5.

Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang (2009), Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling, *Water Resour. Res.*, *45*, W02501, doi:10.1029/2008WR007347.

Weiss, R., and L. Smith (1998), Parameter space methods in joint parameter estimation for groundwater flow models, *Water Resour. Res.*, *34*(4), 647–661, doi:10.1029/97WR03467.

Xia, Y., A. J. Pitman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas (2002), Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw dataset, *J. Hydrometeorol.*, *3*, 181–194, doi:10.1175/1525-7541(2002)003 <0181:CALSMO>2.0.CO;2.

Yatheendradas, S., T. Wagener, V. Gupta, K. Unkrich, D. Goodrich, M. Schaeffer, and A. Stewart (2008), Understanding uncertainty in distributed flash-flood forecasting for semi-arid regions, *Water Resour. Res.*, *44*, W05S19, doi:10.1029/2007WR005940.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, *44*, W09417, doi:10.1029/2007WR006716.

H. V. Gupta and P. Pokhrel, Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA. (pokhrel@email.arizona.edu)