



## Predictive streamflow uncertainty in relation to calibration-constraint information, model complexity, and model bias

MICHAEL J. FRIEDEL, *U.S. Geological Survey, P.O. Box 25046, MS 964, Denver Federal Center, Lakewood, Colorado 80225.*  
*Tel.: 011-303-236-7790; fax: 011-303-236-1425; e-mail: mfriedel@usgs.gov*

### ABSTRACT

Reliance on basin models by hydrologists to study the effects of changing landscapes on sustainability of water resources is increasing. The success of basin models to provide reliable predictions depends on adequacy of the calibrated model and residual model uncertainty. In this paper, reductions in predictive streamflow uncertainty are demonstrated using streamflow quantities integrated over different sample frequencies as calibration constraints and iterative updating of prior information. Respective model-quality results in terms of coefficient of model-fit efficiencies for cases using published parameter values and no calibration, traditional calibration constrained by daily streamflow measurements, and calibration constrained by various derived streamflow quantities were  $-0.44$  (very poor),  $0.66$  (fair), and  $0.66$  (fair) to  $0.91$  (excellent). Incrementally reducing the number of active excellent-quality-model parameters (from 130 to 6 parameters) reduced the range of uncertainty (from  $287.7 \text{ m}^3 \text{ s}^{-1}$  to  $54.5 \text{ m}^3 \text{ s}^{-1}$ ) when predicting the 500-year peakflow discharge within the Blackberry Creek basin near Chicago, Illinois. The reduction in range of predictive uncertainty, however, coincided with increased model bias for which the 6-parameter model could not predict the 500-yr peakflow discharge. Thus, the hydrologist needs to strike a balance between an acceptable level of model complexity that gives rise to predictive uncertainty and model bias.

*Keywords:* Model calibration; calibration constraints; model-fit efficiencies; peak-flow discharge; predictive uncertainty.

### 1 Introduction

Reliance on basin models by hydrologists to study the effects of changing landscapes and climatic conditions on the sustainability of water resources is increasing (Steuer and Hunt, 2001; Guay, 2002; Johnston *et al.*, 2003). The success of these basin models to provide reliable prediction largely depends on the adequacy of the model structure, parameterization, and content of calibration-constraint information. Assuming that model structure is adequate and calibration-constraint information available, parameterization proceeds by supplying parameter values that control the rate exchange of fluxes within the basin. Parameter values that are supplied to the basin model can be obtained from direct and (or) indirect information sources. Whereas a few parameter values may be directly related to observable characteristics of the basin, most basin model parameter values are abstract conceptual representations of nonmeasurable basin characteristics that must be assigned published values or estimated indirectly through a calibration process.

In general, basin models can be calibrated using trial and error (manual), a mathematically-based search algorithm, or a combination of these inverse-modeling approaches. Despite the time intensive and unsatisfying task of trial and error basin model calibration, the prevailing model calibration paradigm outside academic circles involves inverse modeling using the trial and error approach (Duncker *et al.*, 1995; Anderson, 1997; Duncker

and Melching, 1998; Steuer and Hunt, 2001, Bossong *et al.*, 2003). During the trial and error calibration of a basin model, individual groups of model parameter values are adjusted based on hydrologic reasoning until a reasonable match is achieved between the simulated and observed streamflow hydrographs. Whereas the trial and error calibration approach implicitly honors the information content of the streamflow hydrograph, the trial and error calibration approach is subjective and determining an optimal parameter set of parameter values that best represents the basin hydrology is nearly impossible.

To ensure that an optimal set of model parameter values are determined, hydrologists use various objective mathematically-based parameter-estimation approaches that include genetic algorithm, nonlinear regression, shuffled complex evolution, and simulated annealing (Wang, 1991; Duan *et al.*, 1992; Sumner *et al.*, 1997; Yapo *et al.*, 1997; Thiemann *et al.*, 2001; Doherty and Johnston, 2003). Whereas the genetic algorithm is robust and rapid, it is generally not able to estimate the actual parameter values. In contrast, the shuffled complex evolution and simulated annealing algorithms are comparatively slow, but can eventually arrive at the correct global minimum. Whereas simulated annealing is not well-suited to evaluating parameter nonuniqueness, the premise of the shuffled complex evolution algorithm is that estimated parameter values are nonunique and for this reason an inverse Monte Carlo approach is used to compute the most likely parameter values. In this study, a faster alternative two-step

approach is used in which nonlinear regression is first used to estimation optimal parameter values from which the range of predictive uncertainty can be calculated by finding the associated critical points in parameter space (Doherty, 2000).

In general, attempts to calibrate basin models using an objective mathematically-based parameter-estimation approach experience challenges that are not found to the same extent as with most ground-water and other environmental models. For example, basin models have (1) a high degree of nonlinearity with respect to adjustable parameters (extrinsic), (2) a large number of parameters that require calibration (tens to hundreds), (3) a large number of measurements (thousands to tens of thousands), and (4) a large range in measurement magnitude (several orders of magnitude). Despite the potentially large number of measurements available to constrain the model-calibration process, the objective calibration of most basin models results in estimated parameters values with a comparatively high degree of uncertainty (Kuczera, 1983a; Kuczera and Mroczkowski, 1988; Gupta *et al.*, 1998). Some reasons for the high degree of parameter uncertainty include inappropriate model structure and limited information content of input variables (for example, precipitation measurements) and calibration constraints (for example, streamflow measurements).

The limited information content of measurement variables used in the model-calibration process result in the estimation of a filtered version of the true basin model transfer function during inverse modeling (deconvolution). Because the simulated streamflow hydrograph represents a convolution of input variables with the basin model transfer function (described by the conceptual model and parameter values), the simulated hydrograph will be limited by the frequency content (amplitude and phase) of the basin model transfer function. For these reasons, the hydrologist must supply information to the model-calibration process that is rich in frequency content. Practically speaking, the information supplied to the inverse modeling process is limited in frequency content due to time and spatial sampling constraints and noise that exists in the model structure and observed measurements. In most cases, the information associated with extreme events may be missing entirely from the streamflow record. In addition to the natural and physical filtering of calibration-constraint information, the inverse-modeling process results in the explicit filtering of calibration-constraint information content through weighting of streamflow measurements and use of *a priori* parameterization schemes.

One consequence of inverse modeling with filtered and frequency-limited information is the introduction and (or) enhancement of correlation among parameters (Kuczera and Mroczkowski, 1988). The correlation among parameters results in parameter nonuniqueness reflecting the fact that it is possible to calibrate a model against a set of measurements using optimal yet alternate sets of parameter values (McKenna *et al.*, 2003; Friedel, 2005). The fact that calibrated model-parameter values are likely to be nonunique despite their optimality suggests that predictions made while using a calibrated model also are likely to be nonunique and therefore uncertain. For these reasons, the goal of this paper is to understand how the uncertainty

in basin model predictions can be reduced. One hypothesis is that the explicit filtering of calibration-constraint information can be counterbalanced by reintroducing streamflow information integrated over selected sample frequencies. Another hypothesis is that additional information content can be extracted from the calibration constraints if simplifications in model parameterization are allowed to occur only when it is necessary for the achievement of numerical stability through the parameter estimation process (Friedel, 2005). The objectives therefore are to: (1) calibrate a basin model using various streamflow quantities integrated over different sample frequencies as constraint information with iterative parameter updating of prior temporal and spatial parameter relations; (2) quantify and compare the calibrated model quality using coefficient of model-fit efficiency (sometimes called the Nash-Sutcliffe criteria; Nash and Sutcliffe, 1970); and (3) quantify the relation between model complexity, bias, and uncertainty associated with a rare peak-flow discharge event.

## 2 Hydrologic simulation

The basin processes were simulated using the Hydrologic Simulation Program-Fortran (HSPF) numerical basin model (Bicknell *et al.*, 1997). In simulating the basin processes, the principle of continuity was used to balance precipitation, evapotranspiration, storage, and runoff given by

$$\text{Precipitation} - \text{Actual evapotranspiration} \pm \text{Storage} = \text{Runoff}. \quad (1)$$

The storage zones represented in this basin model included interception, surface, interflow, upper, lower, and active ground-water (Fig. 1). The respective maximum to minimum evapotranspiration demand exerted on these storage zones were active ground-water outflow, interception, upper, active ground water, and lower. The three runoff components simulated in the basin model were surface runoff, interflow, and active ground-water flow. Surface runoff was direct runoff across the surface of the soil, whereas interflow was direct runoff through macropores and

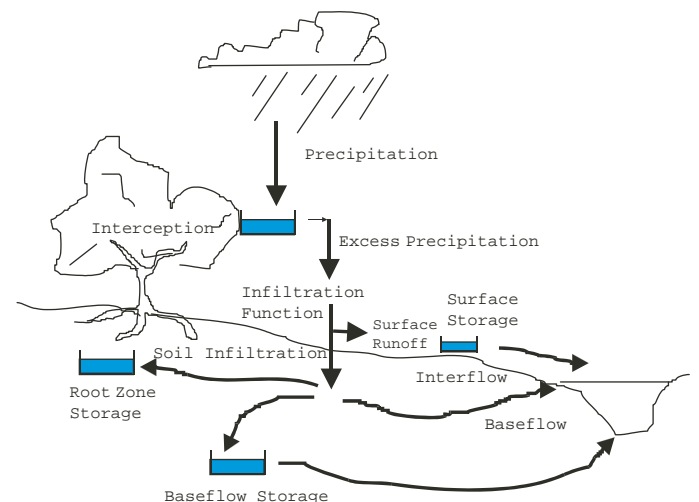


Figure 1 Schematic of conceptual storage and fluxes simulated in the hydrologic software.

active ground-water flow is baseflow that maintained streamflow between storm events.

In pervious areas, only the interception, upper, and lower storage zones were user specified. Interception represented storage associated with vegetal cover above the surface of the soil, whereas the upper zone storage reflected storage in the soil layers that were above the layer controlling infiltration. The upper zone determined the proportion of potential direct runoff and water flux entering the upper zone. Drainage of the upper zone was dependent on evaporation and percolation to the lower and active ground-water zones. The lower zone storage represented lower soil layers that controlled the rate of infiltration. Water available for infiltration was excess water that occurred when the interception storage capacity was exceeded. Water that did not infiltrate the subsurface was divided between the upper zone, surface runoff, and interflow, whereas infiltrating water was drained by the lower and active ground-water zones. The division of infiltrated and percolated water between lower and active ground-water zones was a function of fullness in the lower zone. The lower zone was drained by evapotranspiration, whereas a portion of active ground water also could be diverted out of the basin as recharge to the deep aquifer.

Impervious areas were represented independently of pervious areas with losses by retention storage. Retention storage was associated with cracks and wetting of the surface (analogous to interception storage for pervious areas). The retention storage was drained by evaporation at the potential rate and all rainfall in excess of the user-specified storage capacity became runoff. Surface runoff both for impervious and pervious surfaces was routed across an overland flow plane of user specified length, slope and roughness. Because the flow rate across the plane was a function of the depth of flow, the runoff response was nonlinear and partly dependent on the runoff intensity.

### 3 Model calibration and validation

Solution of the basin inverse problem was achieved by applying a nonlinear regression algorithm (PEST) (Doherty, 2004) to the hydrologic simulation model. The purpose for using nonlinear regression was to estimate values of model parameters that minimized a predefined objective function. When the objective function was minimized, the resulting vector of parameter values was considered to be estimated and the model calibrated.

#### 3.1 Objective function

In this study, the objective function to be minimized was defined by

$$\begin{aligned} \Phi(Q, Va, Vm, Vs, D, d) \\ = \lambda \left\{ \sum_{i=1}^{NP} w_{Q_i} [Q_m(t, x, y)_i - Q_s(p)_i]^2 \right. \\ \left. + \sum_{j=1}^{NVa} w_{Va_j} [Va_m(t, x, y)_j - Va_s(p)_j]^2 \right. \end{aligned}$$

$$\begin{aligned} & + \sum_{j=1}^{NVm} w_{Vm_j} [Vm_m(t, x, y)_j - Vm_s(p)_j]^2 \\ & \times \sum_{j=1}^{NVs} w_{Vs_j} [Vs_m(t, x, y)_j - Vs_s(p)_j]^2 \\ & + \sum_{j=1}^{ND} w_{D_j} [D_m(t, x, y)_j - D_s(p)_j]^2 \left. \right\} \\ & + \sum_{j=1}^{NR} w_{rl} [d_{ml} - d_{sl}(p)]^2 \end{aligned} \quad (2)$$

where  $\Phi$  is the objective function composed of the sum of weighted squared differences between simulated and measured time-dependent variables (first five terms on right-hand-side known as the measurement objective function) and set of prior information defined in this report as preferred relationships (fourth term on right-hand-side known as the regularization objective function);  $\lambda$  is the Lagrange multiplier that was applied to perform nonlinear parameter estimation as an unconstrained problem (Doherty, 2004);  $Q, Va, Vm, Vs, D$  are daily streamflow, annual streamflow volumes, monthly streamflow volumes, storm streamflow volumes, and streamflow duration;  $d$  was the vector of preferred conditions;  $p$  is the vector of parameter values being estimated for each pervious land segment (lzs, infiltr, uzsn, intfw, irc, lzsetp, agwrc, lsur, deepfr, nsur, kvary, basetp, agwetp, cepsc, infexp);  $m$  and  $s$  are subscripts indicating measured and simulated values;  $NQ, NVa, NVm, NVs$ , and  $ND$  are number of daily streamflow, annual streamflow volume, monthly streamflow volume, storm streamflow volumes, and streamflow duration fractions;  $NR$  is the number of soft prior conditions;  $w_{Q_i}, w_{Va_j}, w_{Vm_j}, w_{Vs_j}, w_{D_j}$  are weights for daily streamflow, annual streamflow volumes, monthly streamflow volumes, storm streamflow volumes, and streamflow duration fractions that were determined iteratively to form a single population with a uniform population variance (Gailey and Gorelick, 1991);  $w_{rl}$  are soft prior information weights that were estimated as part of the nonlinear regression process (Doherty, 2004). The actual number of measurement objective function terms used during the model calibration process was dependent on the number of measurement types used to constrain the inverse modeling procedure.

#### 3.2 Prior information and iterative updating

In this study, the basin model was regularized by introducing preferred relations as soft prior information. The soft prior information is defined by equations where the difference between weighted logarithms of model parameters is initially set equal to a preferred value for temporal and spatial parameter relations. A general form of the equations used to describe preferred conditions can be written as

$$pr_i = w_{ri} \log(p_j) - w_{ri} \log(p_k) \quad (3)$$

where  $pr_i$  is the prior information (set initially equal to zero),  $w_{ri}$  are regularization weights (set initially to 1),  $p_j$  is a parameter in pervious land segment  $j$ , and  $p_k$  is a parameter in pervious

land segment  $k$ . The prior information is considered soft because these preferred values have the possibility of changing following a nonlinear iteration. That is, if there is sufficient information content in the calibration-constraint measurements to improve the estimation of parameter values and weights (continued reduction in the objective function), the parameter values would change and prior information accordingly. In this way, the model parameterization structure is iteratively updated in a similar way to Bayesian updating as described by Thiemann *et al.* (2001). Using the prior information equations (3) to describe preferred temporal and spatial relations is analogous to using spatial measures of dissimilarity described by a variogram during the calibration of groundwater (Hoeksema and Kitanidis, 1984; Carrea and Neuman, 1986; McKenna *et al.*, 2003) and vadose-zone models (Friedel, 2005).

### 3.3 Streamflow quantities and calibration constraints

One possible reason for the poor outcome of past objective basin model calibration attempts was that traditional nonlinear regression weighting schemes, in an effort to apply weights evenly to daily (or hourly) discharge measurements, rendered the information content used as calibration constraints to be frequency-limited; that is, on-the-average the daily (or hourly) discharge measurements each contributed equivalently to the objective function thereby implicitly filtering the low- and peak-flow discharge contributions. In an attempt to avoid the problems associated with frequency-limited constraint-information, it was important to use meteorological and streamflow quantities that spanned the widest variety of hydrologic conditions (time periods and magnitudes). In this study, the vector of observations supplied to the measurement objective function differed so that the effectiveness of additional frequency-related information to constrain the model-calibration process could be evaluated. For example, the vector of observations included daily streamflow measurements (very high sample frequency) with one or more integrated quantities derived from the streamflow hydrograph at different sample frequencies; such as, annual streamflow volumes (low sample frequency), monthly streamflow volumes (intermediate sample frequency), storm volumes (high sample frequency), and flow-duration values (broad sample frequency). Combining these streamflow quantities in different ways resulted in the following

calibration-constraint scenarios: daily streamflow (Q), daily streamflow and annual streamflow volumes (QVa); daily streamflow and monthly streamflow volumes (QVm); daily streamflow and flow-duration values (QD); daily streamflow, annual streamflow volumes, and monthly streamflow volumes (QVaVm); daily streamflow, annual streamflow volumes, monthly streamflow volumes, and storm volumes (QVaVmVs); daily streamflow, annual streamflow volumes, monthly streamflow volumes, storm volumes, and flow-duration values (QVaVmVsD) (Table 1).

### 3.4 Diagnostic and inferential statistics

Progress of the model calibration process was evaluated by monitoring changes in the objective function and composite-scaled sensitivities (Mehl and Hill, 2001). The overall model quality of estimated parameter values was evaluated quantitatively by considering mean of weighted residuals, standard error of the regression and correlation coefficient (Hill, 1998). However, the correlation coefficient (CC) did not consider the magnitude of differences between observed and simulated values (only values that were greater than and less than their respective mean values), the simulated time series could increase and decrease in the same pattern as the observed time series yielding a high correlation coefficient even though the two time series were in relatively poor agreement.

In contrast to the correlation coefficient, the coefficient of model-fit efficiency (CMFE) is a direct measure of the fraction of the variance of the original data series explained by the model (Nash and Sutcliffe, 1970). Because the CMFE considers the magnitude of differences between observed and simulated values, the CFME provided a more rigorous evaluation of the calibrated basin model quality. The CMFE for streamflow on a monthly basis was calculated (after Steuer and Hunt, 2001) by

$$CMFE = \frac{\sum_1^N (Q_{o_i} - Q_o)^2 - \sum_1^N (Q_{o_i} - Q_{s_i})^2}{\sum_1^N (Q_{o_i} - Q_o)^2} \quad (4)$$

where  $Q_{o_i}$  is the observed streamflow volume for month  $i$ ,  $Q_{s_i}$  is the simulated streamflow volume for month  $i$ ,  $Q_o$  is the average observed monthly streamflow volume,  $Q_s$  is the average simulated monthly streamflow volume, and  $N$  is the number of months in the calibration period. If the calibrated model residuals were

Table 1 Calibration-constraint scenarios. (Q – daily streamflow discharge, D – streamflow duration, Va – annual volumes, Vm – monthly volumes, Vs – storm volumes).

Number	Calibration constraint scenario	Daily streamflow	Annual volume	Monthly volume	Storm volume	Duration fraction
1	Q	Yes	No	No	No	No
2	QD	Yes	No	No	No	Yes
3	QVa	Yes	Yes	No	No	No
4	QVm	Yes	No	Yes	No	No
5	QVmVs	Yes	No	Yes	Yes	No
6	QVaVmVs	Yes	Yes	Yes	Yes	No
7	QVaVmVsD	Yes	Yes	Yes	Yes	Yes

normally distributed with a mean error close to zero (minimum bias), the CMFE would nearly equal the square of the correlation coefficient. A CMFE value that exceeded 0.9 was considered excellent, whereas the CMFE values in the range of 0.8–0.9, 0.7–0.8, 0.6–0.7, 0.5–0.6, and less than 0.5 were considered very good, good, fair, poor, and very poor, respectively (Duncker *et al.*, 1995). In addition to these statistical measures, the quality of the calibrated model also was checked graphically by comparing the observed and simulated time series and runoff-duration curves.

### 3.5 Split sample validation

Following the successful calibration of a basin model, the conventional wisdom is to perform model validation. Although model validation has various definitions (Anderson and Bates, 2001), the essential meaning of validation is the process of ensuring that a calibrated model accurately represents the processes and response of the study site. Validation differs from calibration in that the parameters are not adjusted and performance of the calibrated model is evaluated using a different data set than that used to calibrate the model. Perhaps the most common approach is the so-called split-sample validation. In the split-sample validation approach, a portion of the measured field data are withheld from the calibration stage in favor of comparing these field data to simulated time series responses that arise when using the calibrated model (Anderson and Bates, 2001).

In general, there are several reasons why the split-sample validation approach is considered a poor test of the calibrated model's predictive ability. First, even if there is no noise in the conceptual model or measurements, the less than perfect measurement information content such as number, type, frequency content, and hydrologic stresses, that is used to calibrate the basin model results in introducing correlation among parameters. One result of parameter correlation between subsets of fitted parameters is that there are many best-fit sets of parameter values that can produce identically simulated streamflow (nonuniqueness), although the relative contributions of fluxes constituting the streamflow may vary greatly (Kuczera and Mroczkowski, 1988). For this reason, the nonuniqueness in estimated parameter values introduced by parameter correlation is stochastic and therefore associated with a range of predictive uncertainty. Second, there is no agreement as to what critical statistical values define acceptable levels of model accuracy determined as part of the validation process. Third, post-audit analysis following validation and application of numerical models has mixed results: some predictions are relatively poor while other predictions are reasonably good (Donigian and Rao, 1990; Anderson and Woesner, 1992). For these reasons, the test of predictability for the calibrated basin model in this study is based on computing and analyzing the model and predictive uncertainty and not the split-sample validation.

## 4 Uncertainty

### 4.1 Model uncertainty

The interrelation between model parameter correlation and model parameter uncertainty is more easily understood using the

conceptualized linear confidence ellipse (Carrera and Neuman, 1986). The linear confidence ellipse represents a prescribed level of certainty associated with the two-parameter response surface (sum of squared residuals) for which the idealized objective function minimum is a point. At this prescribed level of uncertainty, the linear confidence ellipse is defined by semi-axes (eigenvectors) whose magnitudes are proportional to the square roots of eigenvalues derived from the covariance matrix. Because the nonlinear regression algorithm used in this study is only an approximate solution to the parameter estimation problem, in practical nonlinear basin studies involving more than 2 parameter values, the normalized eigenvector matrix (derived from the covariance matrix) provides a means of evaluating approximate uncertainty and correlation among multiple parameters (Friedel, 2005).

In practical basin studies involving more than 2 parameter values, the normalized eigenvector matrix provides a means of evaluating approximate nonlinear uncertainty and correlation among multiple parameters (Friedel, 2005). For example, those parameters with a shared predominance among a particular eigenvector represent correlation among a group of parameters that contribute to the model uncertainty. Whereas the angle between the major and minor eigenvalues and horizontal axis describes the degree of correlation between parameters, the ratio of major to minor eigenvalues (called the condition number) of the normalized eigenvector matrix describes the relative magnitude of model uncertainty. In general, larger condition numbers represent greater model uncertainty and smaller condition numbers represent lesser model uncertainty. Whereas the condition number describes the relative magnitude of model uncertainty, the nonlinear uncertainty associated with specific streamflow predictions must be determined in another way.

### 4.2 Predictive uncertainty

The predictive estimation technique, first described by Vecchia and Cooley (1987), is used to estimate the limits of nonlinear uncertainty associated with peak-flow predictions. A model action used to compute the prediction limit in this approach can be represented (Doherty, 2004) by

$$d = Mp, \quad (5)$$

where  $d$  is a scalar prediction value,  $M$  is a  $1 \times n$  matrix describing the model ( $n$  is the number of parameters requiring estimation), and  $p$  is the parameter vector (previously defined). The goal of the predictive estimation process is to maximize (or minimize)  $Mp$ , while maintaining the basin model in an optimally calibrated state, as defined by a value  $\Phi_p$  that is about 5 percent larger than the objective function minimum  $\Phi_r$ . Using this approach, the set of optimal parameter values estimated determined during the model calibration procedure are used as initial (starting) parameter values in the predictive estimation problem together with their associated measurement and regularization weights.

The parameter vector that maximizes (or minimizes) the predictive estimation problem is given by

$$p = (X^T QX)^{-1} \left[ X^T Qc - \frac{K}{2\lambda} \right], \quad (6)$$

where the first term is the matrix representation of the normal equations,  $X$  is the matrix operator that contains equations for the solution of the basin model as a function of the parameters describing the model domain as well as the discretization scheme and boundary conditions,  $X^T$  is the transpose of matrix  $X$ ,  $Q$  is the diagonal weighting matrix, and  $c$  is the vector of model-simulated observations. The Lagrange parameter is computed using an analytical equation presented by Doherty (2004).

## 5 Basin model

The model used in this study represents the Blackberry Creek basin that is located about 10 miles west of metropolitan Chicago. The Blackberry Creek basin has a temperate, humid, continental climate with a respective long-term (1927–99) average annual precipitation and temperature of 889 mm and 9.4 Celsius, respectively. Climate of the area results in storms with seasonal rainfall characteristics typical of the Midwestern United States. Land-use in the Blackberry Creek basin (drainage area of about 189 km<sup>2</sup>) is largely agricultural (84.2%); however, the basin is developing rapidly with population and proportion of urbanized land area in the basin expected to double by the year 2020. The Blackberry Creek basin boundaries, land-use categories, and land-use percentages are important in developing the conceptual model used in the basin model calibration process (Fig. 2). Other important information that is needed to construct the basin model includes field measurements and model parameter values. Collectively, the model conceptualization and parameterization, and measurement (input and calibration constraint) information account for the total uncertainty in the calibrated basin model.

## 5.1 Conceptualization

Although HSPF is classified as a lumped parameter model, this basin model reproduces spatial variability by dividing the basin into hydrologically homogeneous pervious land, impervious land, and river reach segments. For this reason, the Blackberry Creek basin boundary delineation and percentages of land-use types were determined using a geographical information system and multiresolution land cover information with a 30-meter pixel resolution (Illinois Department of Natural Resources, 1996). Six homogeneous land-segment types were identified based on percent impervious area and five pervious land segments: flat grassland (less than 4.7 m/km), moderately steep grassland (4.7 to 37.9 m/km), steep grassland (greater than 37.9 m/km), forest, and cropland. In determining the pervious land segments, the amount of imperviousness was assumed to be 40 percent, 50 percent, and 85 percent for the respective low-intensity residential, high intensity residential, and commercial/industrial land-use categories. In these urban land-use categories, the portion of residential land that was not impervious was assumed to represent grass and forest. In addition to conceptualization of land-segment types, the basin model required conceptualization of an appropriate number of stream reaches that drain the pervious land segments.

Two conceptual models of the Blackberry Creek basin were constructed for evaluation in this study. These two conceptual Blackberry Creek models differed in the total number of stream reaches used to describe the drainage system: conceptual model 1 used 3 stream reaches (1 mainstem and 2 tributary reaches), whereas conceptual model 2 used 21 stream reaches (11 mainstem and 10 tributary reaches). In both conceptual models, the stream reaches were represented as one-dimensional straight-line segments requiring information such as discharge, depth (stage), surface area, and volume. In conceptual model 1, the estimates of stream reach surface area and volume for various discharge

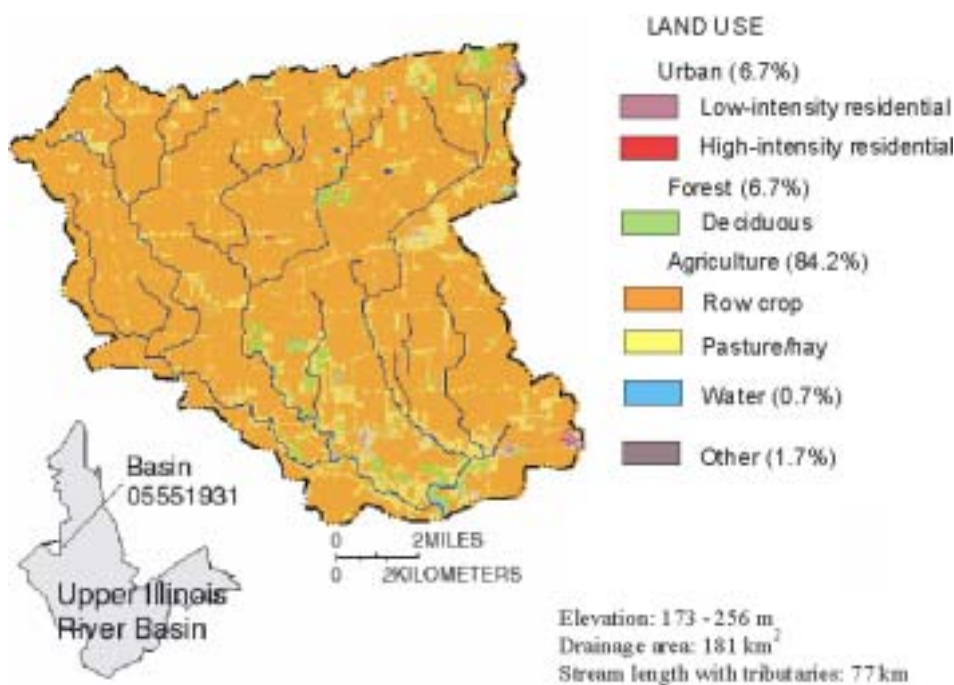


Figure 2 Blackberry Creek basin boundaries, tributaries, and land use.

values were computed using the stage-discharge relations for the streamflow gage station at Yorkville. This same information was used for all three reaches. In conceptual model 2, the measured cross sections and Manning's roughness coefficients were used together with the HEC-RAS program (Brunner, 2001) to derive rating tables for input to the model.

## 5.2 Parameterization

Based on an initial evaluation of composite-scaled sensitivities, there was determined to be sufficient information content in the calibration-constraint measurements to estimate 16 parameter types (of the thirty-nine total) during model calibration. Model parameters that would be estimated for each model land segment included: ground-water evapotranspiration (AGWETP), daily ground-water recession constant (AGWRC), baseflow evapotranspiration (BASETP), nominal interception storage (CEPSC), fraction groundwater inflow that becomes recharge (DEEPR), exponent affecting sensitivity of infiltration function (INFEXP), index to infiltration capacity of soil (INFILT), interflow parameter (INTFW), interflow recession constant (IRC), ground-water recession shape factor (KVARY), length of overland flow plane (LSUR), lower zone storage (LZSN), lower zone evapotranspiration (LZETP), roughness of overland flow plane (NSUR), and upper zone storage (UZSN). Many of these parameters were interactive and therefore associated with more than one component of the streamflow hydrograph. Parameters not incorporated into the estimation process were assigned reasonable published values as were the initial parameter values. The assignment of parameters values was based on those reported for a neighboring basin by the Northeastern Illinois Planning Commission (Price, 1997). The upper and lower bounds supplied for adjustable parameters replicated those described by Bicknell (1997). To enhance the rate of numerical convergence and prevent parameters from becoming negative, all parameters were log transformed.

The application and use of these model parameter values was the same for both conceptual models with the exception of upper zone storage. Because the application of upper zone storage differed for these two conceptual models, the total number of parameter values that were estimated during model calibration also differed. Specifically, conceptual model 1 assumed that there was no monthly difference in upper zone storage. For this case, a single UZSN parameter value was assigned to each pervious land segment giving a total of 75 parameter values to be estimated during model calibration. In contrast, the conceptual model 2 assumed that the upper zone storage differed on a monthly basis. For this case, twelve UZSN parameter values were assigned to each pervious land segment resulting in a total of 130 parameter values to be estimated during model calibration. The estimation of parameter values using these conceptual models was performed with the constraints derived from two different calibration periods that differed only in the inclusion (or exclusion) of streamflow quantities related to the maximum streamflow recorded in Blackberry Creek on July 19, 1996.

## 5.3 Measurements

Adequacy of the basin model calibration process is highly dependent on the temporal and spatial quality of input measurements driving the basin simulations and recorded observations used as calibration constraints. Input measurements driving the Blackberry Creek basin model were related to meteorology, whereas those measurements used to constrain the Blackberry Creek basin model calibration were related to streamflow. The meteorological time-series information used in this study spanned the period from 1989 to 2000. Information such as sky cover, solar radiation, air temperature, and wind movement were input to the basin model so that evapotranspiration and snowmelt could be calculated. These time-series meteorological measurements (assumed to be homogeneous over the basin) were recorded at the Chicago International O'Hare airport and obtained from the Midwest Climate Information System (Kunkel *et al.*, 1990).

Because of the large spatial variability in rainfall that is often associated with Midwest storms, the model calibration process can be difficult when based on a sparse number of rainfall measurements. Generally speaking, the less detail available to describe spatial rainfall requires a longer calibration period to average-out spatial variability between storms (Duncker *et al.*, 1995). Even long-period rainfall records, however, could introduce model bias if errors in rainfall measurements were present (Troutman, 1982). In this study, no rain-gage stations were located within the Blackberry Creek basin. Stations with long-period rainfall records were available for three stations that surrounded the Blackberry Creek basin: Aurora, Illinois (southeast), St. Charles, Illinois (northeast), and DeKalb, Illinois (northwest). The spatial variability in rainfall associated with these three sites was evaluated by comparing correlation coefficients between rainfall events.

Over the period of 1989–96, only 14 storms appeared highly correlated ( $> 0.70$ ) between pairs of rain gage stations (Aurora–St. Charles, St. Charles–DeKalb, Aurora–DeKalb), and only three dates (2/17/1994–2/24/1994, 6/23/1994–6/28/1994, and 1/13/1995–1/22/1995) had high concurrent correlation coefficients for all three pairs of rain gage stations indicating the presence of a relatively uniform storm over the Blackberry Creek basin. The fact that most storms were poorly correlated between rain-gage stations underscored the heterogeneous nature of rainfall over the Blackberry Creek basin. Because of the heterogeneous rainfall over the Blackberry Creek basin, a 7-year time series (1989–96) from the closest rain gage station at Aurora (about 8 km from the basin centroid) was used as input to the basin-model (Fig. 3a).

Streamflow observations for Blackberry Creek were recorded at the U.S. Geological Survey station gage (00050700) near Yorkville, Illinois (Fig. 3b). Hourly streamflow for the period 1989–2000 was interpolated to daily streamflow for determination of summary statistics and provide information to construct calibration-constraint scenarios for two periods: 10/01/1989–6/30/1996 and 10/01/1989–7/30/1996. The summary statistics for Blackberry Creek indicated that streamflow was perennial and daily flows ranged from  $0.04 \text{ m}^3 \text{ s}^{-1}$  to  $98 \text{ m}^3 \text{ s}^{-1}$  (Table 2).

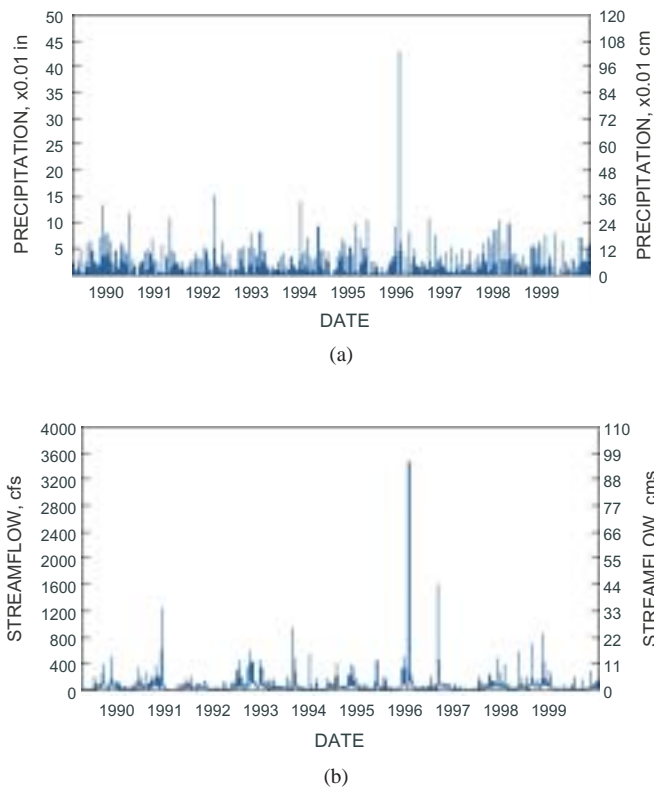


Figure 3 Daily (a) Daily precipitation over the period 1989–2000, Aurora, Illinois. (b) Mean streamflow measured at gage near Yorkville, Illinois.

Whereas the historical and calibration periods had similar minimum, mean, and median streamflow, the maximum peak flow in the first calibration period was less than half that of the second calibration period. Both calibration data sets had at least one discharge value in the 1st percentile range (streamflow exceeding  $11 \text{ m}^3 \text{ s}^{-1}$ ), and the second calibration data set had the largest historic peak flow of  $98 \text{ m}^3 \text{ s}^{-1}$  (July 19, 1996) resulting in a greater standard deviation and coefficient of variation. In general, there

was considerable variation in daily streamflow between all data sets as indicated by the respective historical, calibration 1, and calibration 2, coefficient of variations of 1.59, 1.35, and 2.46.

In addition to streamflow statistics, the range of hydrologic conditions was determined from an annual streamflow exceedence probability curve, where the 0–25, 25–75 and 75–100 percent ranges indicated wet, average, and dry hydrologic conditions, respectively. Based on these exceedence criteria, the Blackberry Creek daily streamflow data sets were considered unbiased because they included streamflow information from one wet, one dry, and four average water years. Using these daily streamflow data sets, the constraint scenarios constructed for the two respective calibration periods had a total of 2,502 and 2,629 streamflow measurements (Q). The derivation of integrated streamflow quantities with different sample frequencies from daily records for each calibration period provided the same number of additional measurements for each period: 10 annual streamflow volumes ( $V_a$ ), 82 monthly streamflow volumes ( $V_m$ ), 26 storm volumes ( $V_s$ ), and 15 streamflow duration fractions (D).

## 6 Results

### 6.1 Calibration of model 1

For comparative purposes, streamflow was simulated (base case) using the Blackberry Creek basin model and published parameter values (uncalibrated) for similar land-use covers in an adjacent basin (Table 3). This uncalibrated base case resulted in a very poor relation (CMFE of  $-0.43$  and correlation coefficient of  $0.66$ ) between the simulated and observed streamflow. Following the application of various calibration-constraint scenarios, the quality of the calibrated Blackberry Creek basin model improved. In general, the quality of the calibrated model increased with the number of types of streamflow quantities included in the calibration process. For example, application of the Q, QVa, QVaVm, QVaVmVs constraint scenarios resulted in enhancing

Table 2 Summary statistics for daily streamflow at Blackberry Creek gage near Yorkville, Illinois. (cm – centimeter, hr – hour).

Model usage period	Historical 1960–2000		Calibration period 1 10/01/1989–6/30/1996		Calibration period 2 10/01/1989–7/30/1996	
	Statistic		Statistic		Statistic	
	Daily streamflow		Daily streamflow		Daily streamflow	
	$\text{m}^3 \text{ s}^{-1}$	cm/hr	$\text{m}^3 \text{ s}^{-1}$	cm/hr	$\text{m}^3 \text{ s}^{-1}$	cm/hr
Minimum	0.0	0.0	0.1	0.0	0.0	0.0
Maximum	98.0	4.7	34.8	1.7	98.0	4.7
Mean	1.5	0.1	1.5	0.1	1.7	0.1
Median	0.9	0.0	1.0	0.0	0.9	0.0
Standard deviation	2.4	0.1	2.1	0.1	4.1	0.2
99th Percentile	11.0	0.5	10.9	0.5	12.6	0.6
90th Percentile	3.1	0.1	3.4	0.2	3.1	0.1
10th Percentile	0.3	0.0	0.3	0.0	0.3	0.0
Coefficient of variation	0.0	4.0	0.0	3.4	0.1	6.2
Number of measurements	413.8	37109.4	70.9	6355.1	74.5	6677.7

Table 3 Summary of performance for Blackberry Creek conceptual model 1 constrained by period 1 (10/01/1989–6/30/1996) information. (Q – daily streamflow discharge, D – streamflow duration, Va – annual volumes, Vm – monthly volumes, Vs – storm volumes).

Calibration constraint scenario	Number of measurements	Coefficient of model-fit efficiency	Correlation coefficient	Model quality
Base case	None	-0.431	0.66	Very poor
Q	2,502	0.592	0.873	Poor
QVa	2,509	0.574	0.859	Poor
QVaVm	2,590	0.631	0.887	Fair
QVaVmVs	2,616	0.777	0.905	Good

Base case – published parameter values used in forward simulation only.

the respective calibrated model quality from poor (CMFE of 0.59 and correlation coefficient of 0.87), poor (CMFE of 0.57 and correlation coefficient of 0.86), fair (CMFE of 0.63 and correlation coefficient of 0.89), and good (CMFE of 0.78 and correlation coefficient of 0.91). Because the total number of daily streamflow measurements was more than an order of magnitude greater than the total of all other derived streamflow quantities, the improvement in model quality was attributed to enhance spectral information through the addition of streamflow quantities sampled at different frequencies.

The fact that the CMFE values were about equal to the square of the correlation coefficient for these calibrated models indicates

that the residuals (difference between simulated and observed streamflow quantities) had a zero bias. In general, the distribution of residuals for these and other models was essentially random with an expected value of the weighted residuals close to zero (unbiased). For example, the plot of weighted simulated and observed streamflow information for the basin model constrained by the QVaVmVsD scenario revealed that values of daily streamflow, annual volume, monthly volume, storm volume, and flow duration fell along a straight line. The fact that the simulated and observed streamflow quantities fell on a line of slope equal to 1 with an intercept of zero visually corroborates the low mean value of weighted residuals (-0.04) and unbiasedness of the model (Fig. 4).

Inspection of the simulated and observed streamflow hydrographs and flow duration curves (exceedence probability for flows) for the uncalibrated (Fig. 5a-d) and calibrated model cases (Fig. 6a-d) indicated significant differences. For this conceptual model, the use of published parameter values and no calibration resulted in significantly overestimating high flows and underestimating low flows with a crossover at about 0.001 in/hr and 30-percent chance for flow exceedence. In contrast, application of the QVaVmVsD calibration-constraint scenario resulted in better matching the peak and low flows, as indicated visually by the hydrograph and flow-duration curve. Given the improved match between simulated and observed hydrographs following model calibration with selected constraint scenarios and iterative updating of parameter relations, it is of interest to evaluate the effect of the conceptual model on model calibration quality.

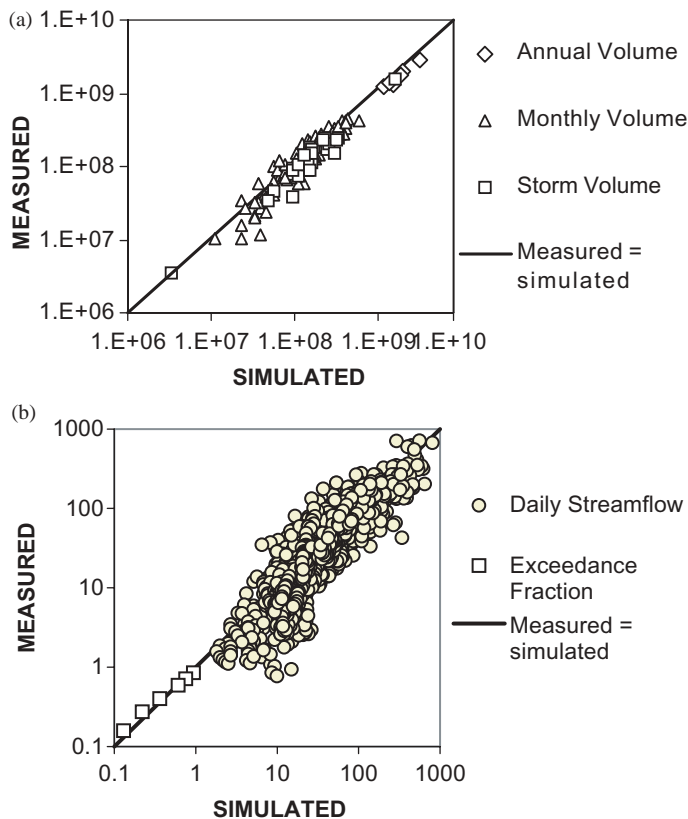


Figure 4 Simulated versus observed streamflow values (a) daily and exceedence fraction, (b) storm volume, monthly volume, and annual volume.

### 6.2 Calibration of model 2

In this test series, the application of calibration-constraint scenarios to the second conceptual Blackberry Creek basin model resulted in additional improvements to the estimated model parameter values over calibrations when using conceptual model 1 (Fig. 7a-d). As with the first conceptual model, the calibration quality of the second conceptual model was dependent on the combination of streamflow quantities used to constrain the calibration process (Table 4). For example, calibration of the second conceptual model subject to the respective Q, QVa, QVaVmVs, and QVaVmVsD constraint scenarios increased the model quality over the first conceptual model from poor, poor, and good (CMFE values of 0.592, 0.574, and 0.777) to Fair, Fair, and Very Good (CMFE values of 0.664, 0.857, and 0.926). Whereas the use of monthly upper zone storage and additional stream reaches enhanced the respective model quality when applying the Q, QVa, and QVaVmVs constraint-scenarios by about 11 percent, 42 percent, and 10 percent, the CMFE value for the calibrated basin model constrained by the QVaVmVs with no monthly storage was similar to the calibration test using conceptual model 1 suggesting that the addition of stream reaches may not be as important as the introduction of time varying storage for this basin.

The additional constraint scenarios used to calibrate the second conceptual Blackberry Creek model were included to

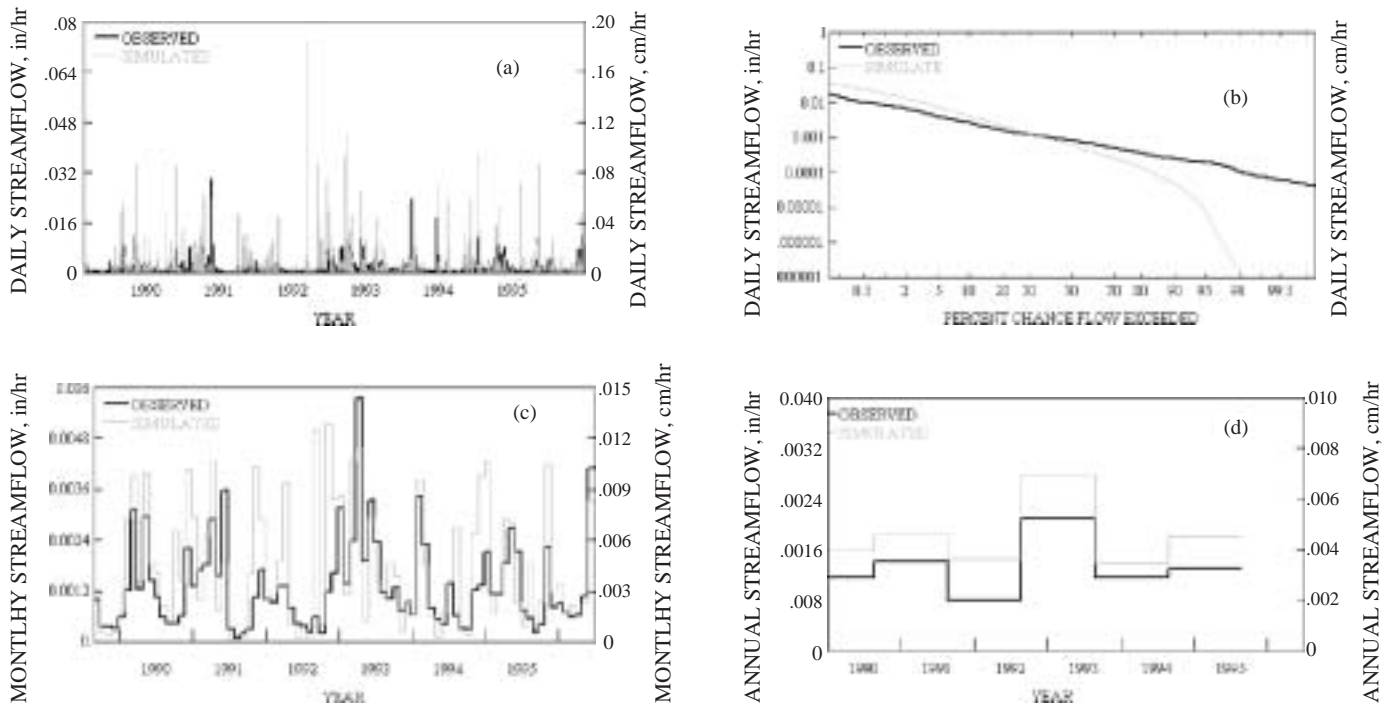


Figure 5 Simulated versus observed: (a) streamflow hydrograph with no calibration of conceptual model 1 (three stream reaches, no monthly storage) and published values; (b) flow-duration curves with no calibration of conceptual model 1 (three stream reaches, no monthly storage) and published values; (c) monthly mean streamflow hydrographs with no calibration of conceptual model 1 (three stream reaches, no monthly storage) and published values; (d) annual mean streamflow hydrographs with no calibration of conceptual model 1 (three stream reaches, no monthly storage) and published values.

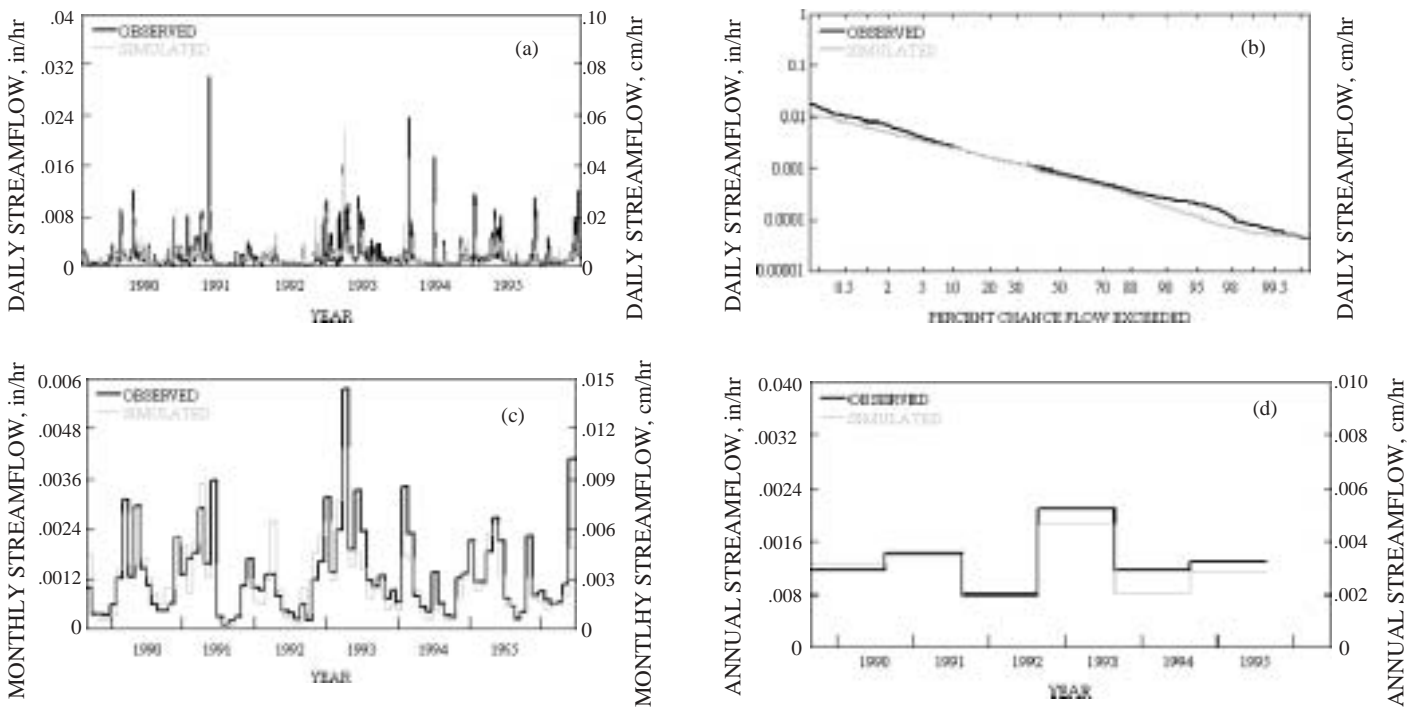


Figure 6 Simulated versus observed: (a) streamflow hydrographs following calibration of conceptual model 1 (three stream reaches, no monthly storage) calibrated using new approach and QVaVmVsD objective function contributions; (b) flow duration curves following calibration of conceptual model 1 (three stream reaches, no monthly storage) using the new approach and QVaVmVsD objective function contributions; (c) monthly mean streamflow hydrographs following calibration of conceptual model 1 (three stream reaches, no monthly storage) using the new approach and QVaVmVsD objective function contributions; (d) annual mean streamflow hydrographs following calibration of conceptual model 1 (three stream reaches, no monthly storage) using the new approach and QVaVmVsD.

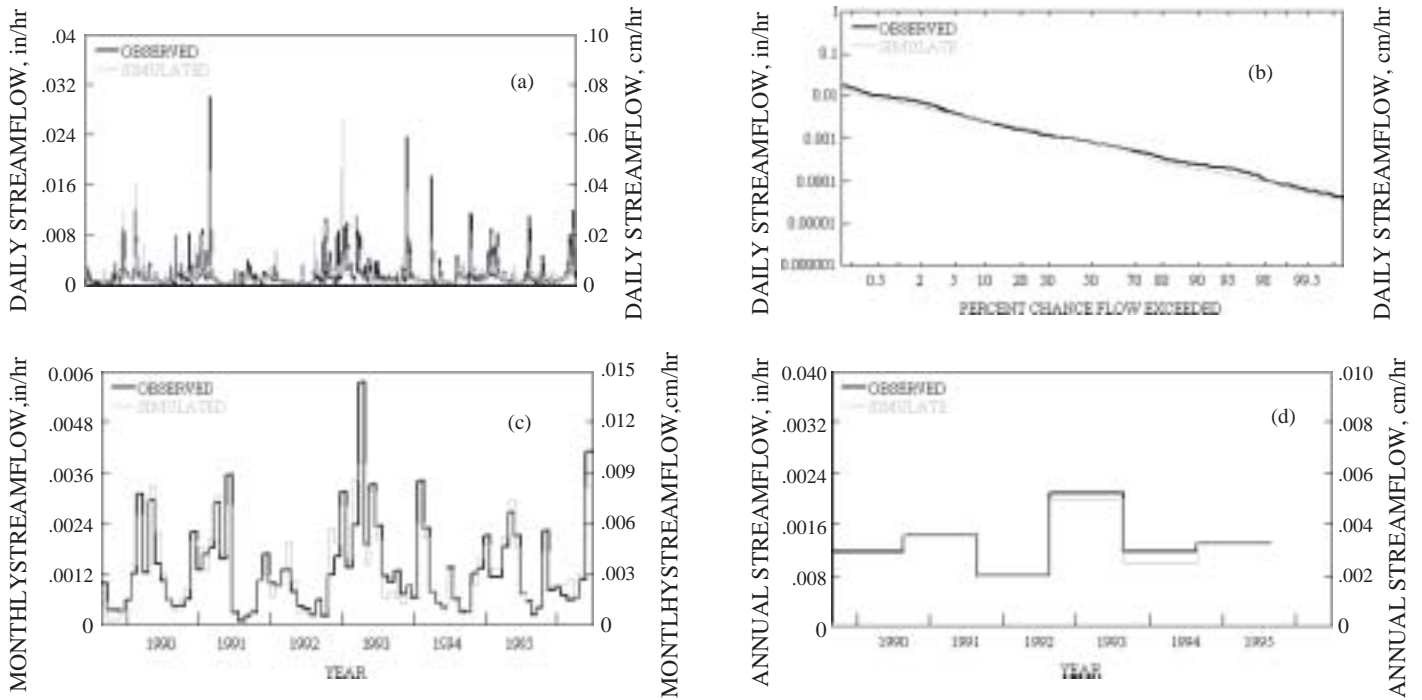


Figure 7 Simulated versus observed: (a) streamflow hydrographs following calibration of conceptual model 2 (21 stream reaches and monthly upper-zone storage) calibrated using new approach and QVaVmVsD objective function contributions; (b) flow duration curves following calibration of conceptual model 2 (21 stream reaches and monthly upper-zone storage) using the new approach and QVaVmVsD objective function contributions; (c) monthly mean streamflow hydrographs following calibration of conceptual model 2 (21 stream reaches and monthly upper-zone storage) using the new approach and QVaVmVsD objective function contributions; (d) annual mean streamflow hydrographs following calibration of conceptual model 2 (21 stream reaches and monthly upper-zone storage) using the new approach and QVaVmVsD.

Table 4 Summary of performance for Blackberry Creek conceptual model 2 constrained by period 1 (10/01/1989–6/30/1996) information. (Q – daily streamflow discharge, D – streamflow duration, Va – annual volumes, Vm – monthly volumes, Vs – storm volumes).

Calibration constraint scenario	Number of observations	Coefficient of model-fit efficiency	Correlation coefficient	Model quality
Q	2,502	0.664	0.834	Fair
QD	2,527	0.685	0.854	Fair
QVa	2,509	0.816	0.905	Good
QVm	2,583	0.759	0.913	Good
QVmVs	2,608	0.840	0.931	Very Good
QVaVmVsD	2,629	0.857	0.926	Very Good
QVaVmVs <sup>a</sup>	2,616	0.751	0.877	Good
QVaVmVs	2,616	0.862	0.929	Very Good

<sup>a</sup>No monthly upper zone storage used.

evaluate the potential model improvements over the widest possible range of streamflow observation quantity combinations. Inspection of these results indicated that incorporating annual volumes (QVa) may be more useful in model calibration than incorporating monthly volumes (QVm), however, adding storm volumes to the monthly volumes further improved the model over the inclusion of annual volumes. Whereas using the three types of volume measurements to constrain the calibration resulted in the best overall model calibration improvement (QVaVmVs), adding the flow duration values to the monthly volume observations did

not significantly change the quality of the calibrated basin model, possibly because much of the duration information was already accounted for through the derived streamflow quantities sampled at varying frequencies. Despite the improvements when calibrating the Blackberry Creek basin model various derived streamflow quantities as calibration constraints, the CMFE values were still less than 1 indicating that limitations existed in the conceptual model and (or) information content used to constrain the model calibration process.

Application of the constraint-scenarios with the largest storm event (calibration period 2) to the conceptual model 2 resulted in additional improvements in the Blackberry Creek basin model quality. For example, the respective quality of calibrated models when using the new QVs, QVmVs, QVaVmVs, and QVaVmVsD constraint-scenarios (period 2) improved quality from good, very good, very good, and very good to very good, excellent, excellent, and excellent (Table 5). The respective CMFE when the model was constrained by the QVmVs, QVaVmVs, and QVaVmVsD scenarios increased by 9.88 percent, 6.03 percent, and 5.20 percent. These increases in model-calibration quality were attributed to a shift in the hydrologic regime due to the largest storm forcing the model to operate over a wider range of hydrologic states. This finding underscores the importance of recalibrating a model when the basin experiences a disturbance not previously associated with the former calibration constraints. Taking advantage of the additional information due to shifts in hydrologic fluxes may result in reduced model and predictive uncertainty.

### 6.3 Model and predictive uncertainty

Despite the demonstrated improvements in model quality that occurred when enhancing the calibration-constraint information, the calibrated models were not perfect. Some possible reasons for the less than perfect match between simulated and observed streamflow included poor spatial representation of rainfall, noise in the spatial and temporal measurements, and residual noise in the conceptual model. Even if a basin model could be calibrated with a perfect match between simulated and observed hydrographs (CMFE and CC of 1), there would likely be some level of parameter nonuniqueness due to correlation among parameters that contributed to model and predictive uncertainty.

The effect of parameter nonuniqueness on model and predictive uncertainty was evaluated by considering the Blackberry Creek basin model calibrated using the QVaVmVsD constraint scenario. In this case, 130 parameter values were estimated with the uncertainty a function of the constraint scenario derived from the first period of streamflow record without the largest historic peak. Following model calibration subject to this constraint, the maximum eigenvalue was on the order of  $10^6$  indicating significant parameter uncertainty associated with that eigenvector (Table 6). A number of elements in this eigenvector shared predominance implying that the correlation among these parameters was the primary contributor to parameter nonuniqueness and therefore model uncertainty. To reduce the present level of model uncertainty (and therefore parameter nonuniqueness), one or more of these correlated parameters needed to be removed (by assigning physically realistic values) from the nonlinear

regression and (or) additional information must be made available to better constrain the model calibration process.

Because there was no additional information available to constrain the model-calibration process, the 109 correlated parameters associated with eigenvectors having the largest eigenvalues ( $>1$ ) were removed from the basin model-calibration process. For this case, there was a significant reduction in the maximum uncertainty of the calibrated model (indicated by a reduction in eigenvalue of the maximum eigenvector from  $1.3 \times 10^6$  to about 0.88) by reducing the total number of estimated parameter values from 130 to 21. An additional reduction in the maximum uncertainty was achieved (indicated by a further reduction in eigenvalue of the maximum eigenvector from 0.88 to about 0.18) by further reducing the number of estimated parameter values from 21 to 6. The removal of these additional 15 parameters from the model calibration resulted in a single predominant element value in each of the remaining 6 eigenvectors. The predominance of a single element in each eigenvector indicated that an additional reduction in the maximum uncertainty was not possible without changes to the conceptual model and that only 6 of the original 130 parameters controlled the simulated basin hydrology response. These reductions in the maximum uncertainty were coincident with reductions in the overall model uncertainty indicated by reductions in the model condition numbers from 120,367,108,557 to 5,880,270 (reduced the number of estimated parameter values from 130 to 21) and 5,880,270 to 50,986 (reduced the number of estimated parameter values from 21 to 6).

Table 5 Summary of performance for Blackberry Creek conceptual model 2 constrained by period 1 and period 2 information. (Q – daily streamflow discharge, D – streamflow duration, Va – annual volumes, Vm – monthly volumes, Vs – storm volumes).

Calibration constraint scenario	Time step	Calibration without extreme event (Period 1: 10/01/1989–6/30/1996)			Calibration with extreme event (Period 2: 10/01/1989–7/30/1996)		
		Coefficient of model-fit efficiency	Correlation coefficient	Model quality	Coefficient of model-fit efficiency	Correlation coefficient	Model quality
QVs	Monthly	—	—	—	0.896	0.951	Very Good
QVmVs	Monthly	0.840	0.931	Very Good	0.923	0.961	Excellent
QVaVmVs	Monthly	0.857	0.926	Very Good	0.904	0.956	Excellent
QVaVmVsD	Monthly	0.862	0.929	Very Good	0.914	0.956	Excellent

Table 6 Summary of uncertainty for Blackberry Creek conceptual model 2 constrained by period 1 information (10/01/1989–6/30/1996). (Q – daily streamflow discharge, D – streamflow duration, Va – annual volumes, Vm – monthly volumes, Vs – storm volumes).

Calibration constraint scenario	Number estimated parameters	Coefficient of model-fit efficiency	Correlation coefficient	Calibration model quality	Eigenvalues of Eigenvector matrix		Condition number	Prediction quality
					Minimum	Maximum		
QVaVmVsD	130	0.862	0.929	Very Good	$1.08 \times 10^{-5}$	$1.30 \times 10^6$	120,367,108,557	Poor
QVaVmVsD	21	0.847	0.928	Very Good	$1.50 \times 10^{-7}$	$8.82 \times 10^{-1}$	5,880,270	Fair
QVaVmVsD	6	0.865	0.931	Very Good	$3.60 \times 10^{-6}$	$1.84 \times 10^{-1}$	50,986	Good

Whereas the previous analysis demonstrated that a reduction in the Blackberry Creek model parameterization complexity coincided with a reduction in the model uncertainty, there was considerable residual model uncertainty even for the 6-parameter model. Because of the varying amounts of residual model uncertainty in these three calibrated Blackberry Creek models, an important test was to quantify and compare the uncertainty limits associated with predictions of a streamflow event not included in the calibration data set. To evaluate the usefulness of reducing parameter complexity in the Blackberry Creek model, the predictive estimation technique was used to compute the uncertainty limits for the largest magnitude streamflow recorded (500-year peakflow discharge) in the Blackberry Creek basin. For these three calibrated models, the estimated minimum prediction limits were  $10.6 \text{ m}^3 \text{ s}^{-1}$ ,  $45.9 \text{ m}^3 \text{ s}^{-1}$ , and  $156 \text{ m}^3 \text{ s}^{-1}$ ; whereas the estimated maximum prediction limits were  $298.8 \text{ m}^3 \text{ s}^{-1}$ ,  $253.6 \text{ m}^3 \text{ s}^{-1}$ , and  $210.4 \text{ m}^3 \text{ s}^{-1}$ .

Despite the consistency in model calibration quality (CMFE of about 0.86 and CC of about 0.93) among these three models, the reduction in model parameterization complexity from 130 to 21 and 21 to 6 parameters coincided with a reduction in the respective range of predictive uncertainty from about  $288.2 \text{ m}^3 \text{ s}^{-1}$  to  $207.7 \text{ m}^3 \text{ s}^{-1}$ , and  $207.7$  to  $54.4 \text{ m}^3 \text{ s}^{-1}$  (Fig. 8). Whereas the reduction in model parameter complexity reduced the range of predictive uncertainty, the reduction in range of predictive uncertainty appeared to occur at the expense of model bias; that is, as the range of predictive uncertainty decreased, the model bias increased. The increase in model bias indicated that there were probably one or more model parameters that control basin hydrology but were not accounted for because of a lack of relevant calibration-constraint information content. This finding implied the following basin modeling conundrum: model complexity guarantees prediction within the limits of uncertainty, but reduced prediction uncertainty at the expense of model complexity guarantees model bias outside the limits of uncertainty. These findings

underscore the need to strike a balance between an acceptable level of model complexity that gives rise to predictive uncertainty and model bias.

### 7 Conclusions

Based on results from this study, several summary statements and conclusions were formulated to assist hydrologists engaged in using calibrated basin models to predict the effect of changing landscapes and climatic conditions on the sustainability of water resources. *First*, the substitution of published parameter values for the same or similar land-use cover types into the basin model gave very poor model results that significantly overestimated high flows and underestimated low flows. For this reason, the application of these uncalibrated basin models should be used only to understand average system behavior. Published parameter values, however, could be valuable when used as initial parameter values in the nonlinear model calibration process. *Second*, improvements in the basin model quality could be achieved by augmenting traditional streamflow measurements used as calibration-constraint information with integrated streamflow quantities derived at different sample frequencies. The continuing improvements in basin-model quality with an increase in the number of integrated streamflow quantities used to constrain the calibration process underscores the importance of broadening the spectrum of the transfer response function. The fact that addition of flow duration values to the various volume observations did not significantly change the quality of the calibrated basin model is attributed to the duration information being already accounted for through the derived streamflow quantities sampled at varying frequencies. *Third*, the similarity in calibrated quality of basin models when constrained by the same measurement types, no monthly storage, but different conceptual models of drainage networks suggests that the addition of stream reaches may not be as important as the introduction of time varying storage for this basin. Also, evaluating conceptual models using the CMFE is more appropriate than the objective function because of the differences in number of estimated parameter values. *Fourth*, despite the improvements when calibrating the basin model using various derived streamflow quantities as calibration constraints, the model correlation coefficients and coefficient of model fit efficiencies were less than 1 indicating that limitations exist in the conceptual model and (or) information content used to constrain the model calibration process. Because changes in the conceptual model did not change the coefficient of model fit efficiencies, the calibration limitation was attributed to the information content associated with system stresses not otherwise captured in the streamflow record or filtered as a result of the inverse modeling procedure. *Fifth*, because there was no additional information available to constrain the best-calibrated basin model, the correlated parameters associated with eigenvectors having the largest eigenvalues ( $>1$ ) were removed from the basin model-calibration process. Whereas the reduction in model parameter complexity reduced the range of predictive uncertainty, the reduction in range of predictive uncertainty appeared to occur

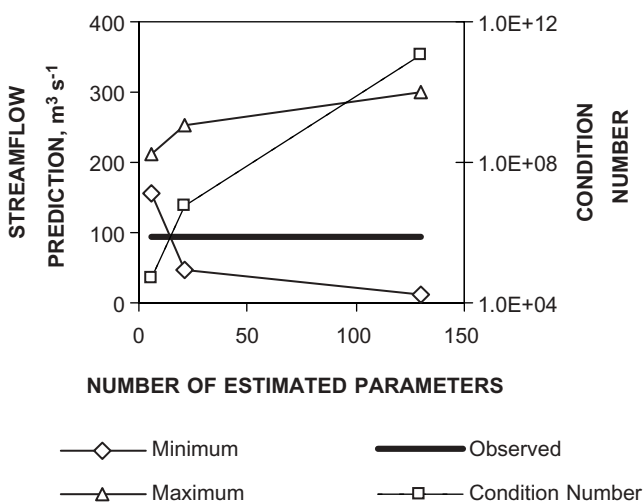


Figure 8 Prediction of the largest peak flow ( $98 \text{ m}^3 \text{ s}^{-1}$ , July 19, 1996) using the best-quality calibrated Blackberry Creek models (coefficient of model-fit efficiency of 0.86, and correlation coefficients of 0.93) based on the conceptual model 2 with 21 stream reaches and QVaVmVsD constraint-scenario derived from October 1, 1989 to June 30, 1996).

at the expense of model bias. The increase in model bias with a decrease in model complexity indicated that there were probably one or more model parameters that control basin hydrology that were not accounted for because of a lack of relevant calibration-constraint information content. This finding suggests that model complexity guarantees prediction within the limits of uncertainty, but reduced prediction uncertainty at the expense of model complexity guarantees model bias outside the limits of uncertainty. These findings underscore the need to strike a balance between an acceptable level of model complexity that gives rise to predictive uncertainty and model bias.

### Acknowledgments

This study was supported in part by the National Water Quality Assessment program of the Water Resources Discipline, U.S. Geological Survey. Several persons provided valuable assistance in this effort including J. Dunker and T. Straub. This manuscript benefited from comments by J. Doherty, B. Troutman, R. Webb, J. Tindall, and two anonymous reviewers.

### References

- ANDERSON, E.A. (1997). Hydrologic Model Calibration Using the Interactive Calibration Program (ICP), *Report, Hydrol. Res. Lab.*, U.S. National Weather Service, Silver Spring, MD.
- ANDERSON, M.G. and BATES, P.D. (2001). *Model Validation: Perspectives in Hydrological Science*. Wiley: Chichester.
- ANDERSON, M.P. and WOESSNER, W.W. (1992). "The Role of the Post Audit in Model Validation," *Advances in Water Resources*, 15, 167–173.
- BICKNELL, B.R., IMHOFF, J.C., KITTLE, J.L., DONIGIAN, A.S. and JOHANSON, R.C. (1997). Hydrologic Simulation Program-Fortran, User's Manual for Version 11, U.S. Environmental Protection Agency, EPA/600/R-97/-90, 755 pp.
- BOSSONG, C.R., CAINE, J.S., STANNARD, D.I., FLYNN, J.I., STEVENS, M.R. and HEINY-DASH, J.S. (2003). Hydrologic Conditions and Assessment of Water Resources in the Turkey Creek Basin, Jefferson County, Colorado, 1998–2001, U.S. Geological Survey Water-Resources Investigations Report 03-4034, 148 pp.
- BRUNNER, G.W. (2001). HEC-RAS, River Analysis System User's Manual, U.S. Army Corps of Engineers, CPD-68, 320 pp.
- CARRERA, J. and NEUMAN, S.P. (1986). "Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 2. Uniqueness, Stability, and Solution Algorithm," *Water Resources Research*, 38(28), 1–14.
- DOHERRY, J. (2004). PEST: Model-Independent Parameter Estimation. Version 5 of User Manual. Watermark Numerical computing, Brisbane, Australia, 213 pp.
- DOHERTY, J. and JOHNSTON, J.M. (2003). "Methodologies for Calibration and Predictive Analysis of a Basin Model," *J. American Water Resources Association*, 39(2), 251–265.
- DONIGIAN, A.S. and RAO, P.S.C. (1990). Selection, application, and validation of environmental models, in *Proceedings of the International Symposium on Water Quality Modeling of Agricultural Non-Point Sources, Part 2*, D.G. De Coursey, (ed.), pp. 577–600, USDA-ARS ARS-81, U.S. Gov. Print. Office, Washington, D.C.
- DUAN, Q., SOROOSHIAN, S. and GUPTA, V. (1992). "Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models," *Water Resources Research*, 28(4), 1015–1031.
- DUNCKER, J.J., VAIL, T.J. and MELCHING, C.S. (1995). Regional rainfall-runoff relations for simulation of streamflow for watersheds in Lake County, Illinois, U.S. Geological Survey Water Resources Investigation Report, 95-4023, 71 pp.
- DUNCKER, J.J. and MELCHING, C.S. (1998). Regional rainfall-runoff relations for simulations of streamflow for watersheds in Du Page County, Illinois: U.S. Geological Survey Water Resources Investigation Report, 98-4035, 80 pp.
- FRIEDEL, M.J. (2005). "Coupled Inverse Modeling of Vadose Zone Water, Heat, and Solute Transport: Calibration Constraints, Parameter Nonuniqueness, and Predictive Uncertainty," *Journal of Hydrology*, 312(1–4), 148–175.
- GAILEY, R.M. and GORELICK, S.M. (1991). "Coupled Process Parameter Estimation and Prediction Uncertainty Using Hydraulic Head and Concentration Data," *Advances in Water Resources*, 14(5), 301–314.
- GUAY, J.R. (2002). Rainfall-runoff characteristics and effects of increased urban density on streamflow and infiltration in the eastern part of San Jacinto River Basin, Riverside County, CA. U.S. Geological Survey, CA. Water Resources Division. USGS/WRI-02-4090. 60 p.
- GUPTA, H.V., SOROOSHIAN, S. and YAPO, P.O. (1998). "Toward Improved Calibration of Hydrologic Models: Multiple and on Commensurable Measures of Information," *Water Resources Research*, 34(4), 751–763.
- HILL, M. (1998). Methods and guidelines for effective model calibration: U.S. Geological Survey Water Resources Investigation Report, 98-4005, 90 pp.
- HOEKSEMA, R.J. and KITANIDIS, P.K. (1984). "An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling," *Water Resources Research*, 20(7), 1003–1020.
- Illinois Department of Natural Resources (1996). *Illinois Land Cover – An Atlas*. Illinois Department of Natural Resources, Springfield, IL, IDNR/EEA-96/05, 152 pp.
- JAMES, L.D. and BURGESS, S.K. (1992). Selection, calibration and testing of hydrologic models in hydrologic modeling of small watersheds: American Society of Agricultural Engineers, 466 pp.
- JOHNSTON, R.K., WANG, P.F., HALKOLA, H., RICHTER, K.E., WHITNEY, V.S., SKAHILL, B.E., CHOI, W.H., ROBERTS, M., AMBROSE, R. and KAWASE, M. (2003). An integrated

- watershed-receiving water model for Sinclair and Dyes inlets, Puget Sound, Washington, USA. Presentation at Estuarine research federation 2003 conference estuaries on the edge: convergence of ocean, land, and culture, September 14–18, 2003, Seattle, WA.
23. KUCZERA, G. (1983). "Improved Parameter Inference in Catchment Models, 1, Evaluating Parameter Uncertainty," *Water Resources Research*, 19(5), 1151–1162.
  24. KUCZERA, G. and MROCKOWSKI, B. (1988). "Assessment of Hydrologic Parameter Uncertainty and the Worth of Multiresponse Data," *Water Resources Research*, 34, 1481–1489.
  25. KUNKEL, K.E., CHANGNON, S.A., LONQUIST, C.G. and ANGEL, J.A. (1990). A Real-Time Climate Information System for Midwestern United States: Bulletin of the American Meteorological Society, Vol. 71, No. 11, pp. 1601–1609.
  26. MCKENNA, S.A., DOHERTY, J. and HART, D.B. (2003). "Non-Uniqueness of Inverse Transmissivity Field Calibration and Predictive Transport Modeling," *Journal of Hydrology*, 281, 263–280.
  27. MEHL, S. and HILL, M. (2001). "A Comparison of Solute-Transport Solution Techniques and Their Effect on Sensitivity Analysis and Inverse Modeling Results," *Ground Water*, 39(2), 300–307.
  28. NASH, J.E., and SUTCLIFFE, J.V. (1970). "River Flow Forecasting Through Conceptual Models; Part 1 – A Discussion of Principles," *Journal of Hydrology*, 10(3), 282–290.
  29. PRICE, T.H. (1997). Unit area detention volumes based on continuous simulation recommended for use with the DuPage countywide stormwater management ordinance. Northeastern Illinois Planning Commission, 29 pp.
  30. STEURER, J.J. and HUNT, R.J. (2001). Use of a watershed modeling approach to assess hydrologic effects of urbanization, North Fork Pheasant Branch Basin near Middleton, Wisconsin, U.S. Geological Survey Water Resources Investigation Report, 01-4113, 49 pp.
  31. SUMNER, N.R., FLEMING, P.M. and BATES, B.C. (1997). "Calibration of a Modified SFB Model for 25 Australian Catchments Using Simulated Annealing," *Journal of Hydrology*, 197(1), 166–188.
  32. THIEMANN, M., TROSSET, M., GUPTA, H. and SOROOSHIAN, S. (2001). "Bayesian Recursive Parameter Estimation for Hydrologic Models," *Water Resources Research*, 37, 2521–2535.
  33. TROUTMAN, B. (1982). "An Analysis of Input Errors in Precipitation-Runoff Models Using Regression with Errors in the Independent Variables," *Water Resources Research*, 18, 947–963.
  34. VECCHIA, A.V. and COOLEY, R.L. (1987). "Simultaneous Confidence Intervals and Prediction Intervals for Nonlinear Regression Models with Application to a Groundwater Flow Model," *Water Resources Research*, 23(7), 1237–1250.
  35. WANG, Q.J. (1991). "The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-Runoff Models," *Water Resources Research*, 27(9), 2467–2471.
  36. YAPO, P.O, GUPTA, H.V. and SOROOSHIAN, S. (1997). "Multi-Objective Global Optimization for Hydrologic Models," *Journal of Hydrology*, 204, 83–97.

