Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model

Misgana K. Muleta   John W. Nicklow

Abstract

Distributed watershed models should pass through a careful calibration procedure before they are utilized as a decision making aid in the planning and management of water resources. Although manual approaches are still frequently used for calibration, they are tedious, time consuming, and require experienced personnel. This paper describes an automatic approach for calibrating daily streamflow and daily sediment concentration values estimated by the US Department of Agriculture's distributed watershed simulation model, Soil and Water Assessment Tool (SWAT). The automatic calibration methodology applies a hierarchy of three techniques, namely screening, parameterization, and parameter sensitivity analysis, at the parameter identification stage of model calibration. The global parameter sensitivity analysis is conducted using a stepwise regression analysis on rank-transformed input–output data pairs. Latin hypercube sampling is used to generate input data from the assigned distributions and ranges, and parameter estimation is performed using genetic algorithm. The Generalized Likelihood Uncertainty Estimation methodology is subsequently implemented to investigate uncertainty of model estimates, accounting for errors due to model structure, input data and model parameters. To demonstrate their effectiveness, the parameter identification, parameter estimation, model verification, and uncertainty analysis techniques are applied to a watershed located in southern Illinois.

1. Introduction

Hydrologic models are particularly useful tools in that they enable us to investigate many practical and pressing issues that arise during planning, design, operation, and management of water resources systems. Models are, however, simplifications of reality, and no matter how sophisticated they may be, models undergo

some aspect of conceptualization or empiricism, and their results are only as realistic as model assumptions and algorithms, detail and quality of inputs, and parameter estimates. For most models, it is imperative that a mechanism that improves accuracy of model estimates, based on observed information available to the modeler, be implemented before using models for their intended purposes. The common approach to accomplishing this useful task is to identify values of model parameters so that model simulations closely match observed behavior of the study area, a process commonly referred to as calibration.

Parameter specification and parameter estimation are the two most important stages of calibration (Sorooshian and Gupta, 1995). The former refers to the stage at which parameters that need to be adjusted are selected, while the latter is the subsequent process of identifying the 'optimal' or 'near-optimal' values of the specified parameters. Approaches and difficulties of conducting these two calibration stages depend on, among other factors, the type and complexity of the watershed simulation model being calibrated. Most previous calibration studies have dealt with lumped, empirical (i.e. black box) models and lumped, conceptual models (Klemes, 1986; Sorooshian and Gupta, 1995; Gupta et al., 1998). However, watershed variables and inputs that may affect hydrologic responses may vary spatially, as well as temporally. Therefore, accounting for heterogeneity of environmental variables such as soil types, land uses, topographic features, and weather parameters is essential in order to properly simulate the effect of spatially varying properties. Distributed, long-term, continuous simulation models, such as the US Department of Agriculture's (USDA) Soil and Water Assessment Tool (SWAT), which are capable of describing this spatial and temporal variability, should generally be used for complex watershed simulation problems. The number of parameters and variables in a distributed model is, however, much higher than that of a lumped model for the same watershed, making calibration of such models, particularly the parameter specification stage, far more complex. Yet, limited work has been completed with regard to calibration of distributed models; Refsgaard and Knudsen (1996), Refsgaard (1997), Senarath et al. (2000), and Eckhardt and Arnold (2001) are some of the most recent contributions that focus on this area of study.

Perhaps the most important task that needs to be accomplished during parameter specification is reducing the number of parameters that should be carried over to the parameter estimation stage. This paper reports on the use of three hierarchical methods, namely parameter screening, spatial parameterization, and global parameter sensitivity analysis to reduce calibrable parameters of SWAT. The parameter sensitivity analysis was performed using stepwise regression analysis, which was carried out on ranks of input–output data pairs that were generated based on a Monte Carlo technique with Latin hypercube sampling.

Parameter estimation follows the decision of which parameters of the simulation model to calibrate. Manual calibration and automatic calibration are two types of parameter estimation approaches. Manual calibration is by far the most widely used approach for complex models, including those of the distributed type (Refsgaard and Knudsen, 1996; Refsgaard, 1997; Senarath et al., 2000). Manual calibration, however, is time consuming and very subjective, and its success highly depends on the experience of the modeler and their knowledge of the study watershed, along with model assumptions and its algorithms. Automatic calibration involves the use of a search algorithm to determine best-fit parameters, and it offers a number of advantages over the manual approach. Automatic calibration is fast, it is less subjective, and since it makes an extensive search of the existing parameter possibilities, it is highly likely that results would be better than that which could be manually obtained. Senarath et al. (2000) and Eckhardt and Arnold (2001) have implemented automatic calibration for distributed models. Both studies used a Shuffled Complex Evolution (Duan et al., 1992) search algorithm. In this study, an automatic calibration module is developed using genetic algorithms (GAs) (Holland, 1975). The resulting module is applied to calibrate streamflow and sediment concentration estimates of SWAT using data from a southern Illinois watershed.

Unfortunately, model calibration does not guarantee reliability of model predictions. The parameter values obtained during calibration and the subsequent predictions made using the calibrated model are only as realistic as the validity of the model assumptions for the study watershed

and the quality and quantity of actual watershed data used for calibration and simulation. Therefore, even after calibration, there is potentially a great deal of uncertainty in results that arises simply because it is too unlikely to find error-free observational data (e.g. precipitation, streamflow, topography) and because no simulation model is an entirely true reflection of the physical process being modeled. This study used the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992; Beven and Freer, 2001) in order to investigate uncertainties involved with predicting streamflow and sediment concentration for the study watershed.

2. The watershed simulation model

SWAT is a continuous-time, spatially distributed simulator developed to assist water resource managers in predicting impacts of land management practices on water, sediment and agricultural chemical yields (Neitsch et al., 2001; Arnold et al., 1998; ASCE, 1999). SWAT makes use of watershed information such as weather, soil, topography, vegetation, and land management practices to simulate watershed processes such as surface and subsurface flow; erosion and sedimentation, including both overland and channel sediment processes; crop growth for customized agricultural management practices; and water quality, including various species of nitrogen and phosphorus, among others. The model operates on either a daily or sub-daily time scale. Spatially, the model subdivides a watershed in to subbasins and, potentially, further delineates subbasins into hydrologic response units (HRUs), based on physical characteristics of the watershed (i.e. topography, soil, and land use).

SWAT simulates major hydrologic components and their interactions as simply and yet realistically as possible (Arnold and Allen, 1996). In addition to its crop growth and water quality components, the hydrologic routines within SWAT simulate surface and subsurface runoff processes, accounting for snow fall and snow melt, vadose zone processes (i.e. infiltration, evaporation, plant uptake, lateral flows and percolation), and ground water flows. Runoff volume is estimated using the Curve Number technique, and peak runoff rate is calculated using the Modified Rational Formula (Williams, 1975). Sediment yield from each subbasin is generated using the Modified Universal Soil Loss Equation (MUSLE). The model updates the C factor of the MUSLE equation on a daily basis using information from the crop growth module, hence accounting for variation in plant cover during its growth cycle and its effect on erosion.

3. The demonstration watershed and data

Big Creek watershed, shown in Fig. 1, is used throughout this study for the demonstration of the methodologies and the models developed in this study. Located in southern Illinois, this 133 km^2 basin not only contributes significant amounts of water to the Lower Cache River, but also carries a higher sediment load than other tributaries located in the area (Demissie et al., 2001). Application of SWAT to a basin such as Big Creek requires topographic, soil, land use, and climate data, as well as streamflow and sediment data for calibration, verification, and analysis of uncertainty. Data obtained included a 10 m resolution Digital Elevation Model (DEM) from the Natural Resources Conservation Service (NRCS), 30 m-pixel land use maps for the years 1999 and 2000 from the National Agricultural Statistics Service (NASS), and a 30 m-resolution soil map from the Southern Illinois District of the NRCS. Daily historical data related to precipitation, maximum and minimum temperatures, wind speed, humidity, solar radiation, and potential evapotranspiration from January 1990 to August 2001 were obtained from the Midwest Climate Center (MCC) for nearby climate stations. Fifteen minute interval precipitation data for the period January 1971–April 2002 was also obtained from the MCC for a station located at nearby Murphysboro, IL and was used to derive a monthly maximum half an hour rainfall for all the months of a year. Finally, daily streamflow and sediment concentration data were obtained from the Illinois State Water Survey (ISWS) for Perks Road Station (PRS), a gauging station that drains approximately 65% of the watershed, and for Church Road Station (CRS), a station that drains about 18% of the watershed (see Fig. 2). Daily streamflow and daily sediment concentration data spanned from June 25, 1999 to August 26,
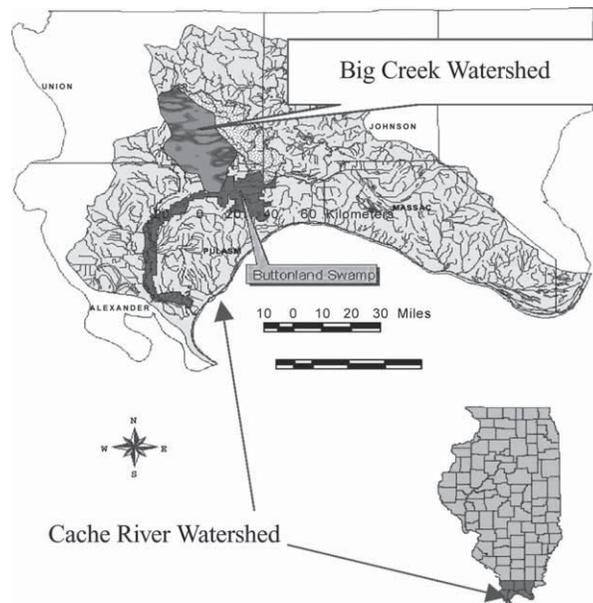
Fig. 1. Location of Big Creek watershed.

2001 for PRS and April 20, 2000 to August 26, 2001 for CRS. The sediment concentration record was intermittent, and over the spans, a total of only 682 and 413 daily records for PRS and CRS, respectively, were available. The original soils map obtained for this study was preprocessed in order to match classifications with SWAT's *soil5id* system, which was used for this study.

4. Parameter identification

As a distributed model, SWAT allows for subdivision of a watershed into smaller subwatersheds, the smallest spatial scale considered in this study. Based on a detailed spatial sensitivity and model feasibility analysis (Muleta, 2003), the watershed was divided into 78 subbasins. Each of these subbasins is represented by a number of parameters that could be derived by calibration. Determining parameter values that are both realistic and 'optimal' for such a large number of parameters is not feasible, calling for a necessary reduction of the number of calibrable parameters. In this study, three hierarchical methods are applied to achieve this reduction.

4.1. Screening

Screening, as applied in this study, refers to identification of model parameters that could be estimated with reasonable accuracy based on field data alone. A detailed investigation of the literature related to SWAT has assisted in identifying the 35 parameters in Table 1 that are integrally related to the model's streamflow and sediment yield prediction; whose estimation from readily available data alone may pose significant uncertainty; and for which there exists insufficient information from which
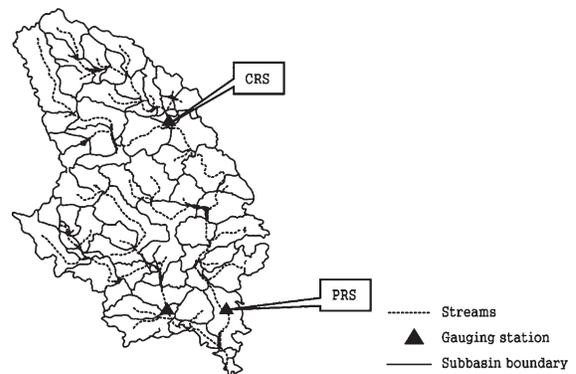

Fig. 2. Location of streamflow and sediment gauging stations.

Table 1
Model parameters involved in the calibration process

Name	Description	Range of values	
		Min.	Max.
SMFMX	Maximum melt rate for snow during the year (mm/°C-day)	1.4	7.5
SMFMN	Minimum melt rate for snow during the year (mm/°C-day)	1.4	7.5
TIMP	Snow pack temperature lag factor	0.5	1
SMTMP	Snow melt base temperature (°C)	−2	20
SNO50COV	Snow water equivalent corresponding to 50 percent snow cover (mm)	0.2	0.8
SFTMP	Snowfall temperature (°C)	−10	5
SPCON	A linear parameter used in channel sediment routing	0.001	0.01
SPEXP	An exponent parameter used in channel sediment routing	1	2
SURLAG	Surface runoff lag time (days)	0.001	15
APM	Adjustment factor for sediment routing in tributary channels	0.5	2
PRF	Adjustment factor for sediment routing in the main channel	0.001	2
ESCO	Soil evaporation compensation factor	0.001	1
EPCO	Plant uptake compensation factor	0.001	1
MSK_CO1	Muskingum routing coefficient 1	0.005	10
MSK_CO2	Muskingum routing coefficient 2	0.005	10
EVRCH	Reach evaporation adjustment factor	0.5	1
ALPHA_BF	Base flow alpha factor (days)	0.001	1
GW_REVAP	Groundwater *revap* coefficient	0.02	0.2
REVAPMN	Threshold depth of water in the shallow aquifer for 'revap' to occur (mm)	0	100
OV_NP	Manning's n value for pasture land for overland flow	0.05	0.3
OV_NF	Manning's n value for forest land for overland flow	0.2	0.8
LAT_TTIME	Lateral flow travel time	0	1
CANMXP	Maximum canopy storage for pasture land	1	3
CN2PA	SCS runoff curve number for moisture condition II	39	68
CH_COV	Channel cover factor	−0.001	1
CH_EROD	Channel erodibility factor	−0.05	0.6
ALPHA_BNK	Base flow alpha factor for bank storage	0.05	1
CH_N2UP	Manning's n value for the main channels in uplands	0.01	0.15
CH_N2LW	Manning's n value for the main channels in lowlands	0.01	0.15
CH_N1	Manning's n value for the tributary channels	0.01	0.15
CANMXF	Maximum canopy storage for forest land	2	6.5
USLE_CP	Minimum value of MUSLE C factor applicable to the land cover	0.001	0.5
GW_DELAYE	Groundwater delay	0.001	100
RCHRG_DP	Deep aquifer percolation fraction.	0.01	0.75
GWQMIN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	0	100

the parameters could be directly estimated. The first 16 of these 35 parameters assume uniform values over the watershed, while values for the remaining 19 could differ between the 78 subbasins, depending on soil type, land use, and/or topographic features. Thus, if screening is the only parameter reduction mechanism used, the parameter estimation algorithm is left to identify best-fit values for over one thousand parameters, which is still quite a daunting task.

In an attempt to further reduce the number of calibrable parameters, some spatially varying inputs (e.g. ground water flow parameters) are forced to assume uniform values over the watershed and some

others (e.g. Manning's coefficient for channels) are broadly grouped. For several other parameters, including the Curve Number, Manning's roughness coefficient for overland flow, and maximum water holding capacity of a canopy, a concept referred to in this study as parameterization was applied.

4.2. Parameterization

Parameterization is a technique for transferring model parameters of a given spatial unit to other spatial units in the watershed. For this study, a 'representative subbasin' is selected, upon which

the model assumes homogeneity of parameters and variables. A relationship between required parameters of this representative modeling unit and corresponding parameters of other homogeneous units (e.g. subbasins) is developed using available information about the parameters. In this way, the definition of variables in the representative subbasin enables determination of a corresponding parameter in other subbasins. For parameterization of the Curve Number, CN, for example, a hypothetical subbasin covered with pasture lands, the major land use in the study watershed, that is grown under treatment conditions common for pasture lands in the watershed and soil group 'A' was considered to be representative. Then, the relationship between the CN of the representative virtual subbasin and other subbasins that have the same land use and treatment conditions as the representative subbasin, but belonging to a different soil group, were derived based on CN values recommended in the literature. Likewise, the relationship between CN of a subbasin having a given soil grouping and land use and CN value of a subbasin that is covered by pasture land and the same soil grouping were developed. Similar approaches were applied for evaluation of Manning's roughness coefficient, n, for overland flow and the maximum water holding capacity of a canopy, *canmax*.

A combination of the screening and parameterization reduced the number of parameters that need to be calibrated to 35 (i.e. 19 that were originally allowed to vary spatially and 16 that assume uniform values over the watershed). Yet, it may still not be necessary or wise to apply a search algorithm to all remaining 35 parameters. Particularly for watersheds like Big Creek that lack long years of recorded data, it is essential to reduce the number of calibrable parameters as much as possible. Fortunately, model outputs are not equally sensitive to all parameters of a model. If an output is not appreciably sensitive to certain parameters, it would be reasonable to assign nominal estimates for those parameters and consider only the parameters to which the model is sensitive during the calibration effort, calling for a parameter sensitivity analysis.

4.3. Parameter sensitivity analysis

Parameter sensitivity analysis (SA) is applied to identify parameters of SWAT model that contribute most to the variability of streamflow and sediment yield, and thus, those that should be calibrated. While there are a number of techniques available for conducting SA (Saltelli, 2000), all can be broadly grouped as local and global approaches (Saltelli et al., 1999). In local techniques, output responses are determined by sequentially varying each of the input factors and by fixing all other factors to constant nominal values. The further the perturbation moves away from the nominal value, the less reliable the analysis results become (Helton, 1993). Also, the more nonlinear the relationship between inputs and output variables, which is typical in hydrologic models, the more difficult and unreliable it is to employ local techniques. Furthermore, since sampling is performed for one input at a time by fixing all other inputs at constant values, local approaches do not account for any interaction between inputs, if any exists. Unlike the local techniques, global SA methods explore the entire range of input factors, and all input factors can be simultaneously varied, allowing investigation of output variation as a result of all inputs and their possible interaction (i.e. output uncertainty is averaged over all input factors). Monte Carlo analysis, also known as a sampling-based method, the response surface methodology, and the Fourier amplitude method are common global SA techniques.

The global SA method used in this study belongs to the Monte Carlo family of methods. A large computational demand is typically a concern of these SAs and is a result of the random and unsystematic generation of inputs from specified distributions. However, the use of more strategic, efficient, and effective sampling approaches, such as importance sampling and Latin hypercube sampling, can significantly reduce computational demand (McKay et al., 1979; Iman and Conover, 1980). Therefore, this research further integrates Latin hypercube sampling into the SA.

In Monte Carlo analysis, once a 'sufficient' number of input–output pairs are sampled using any preferred sampling technique, further analysis needs to be performed to explore the input–output mapping and to provide a qualitative or quantitative measure of output uncertainty caused by each input. Simple scatterplot analysis, regression analysis, correlation and partial correlation analysis, and stepwise regression analysis,

the latter of which is used in this study, are the common approaches for accomplishing this task. Based on the literature, no previous study in watershed modeling has applied stepwise regression for sensitivity analysis. In fact, most previous sampling based SA efforts in water resources related studies used scatter plot analysis, which is only satisfactory when few parameters are involved.

4.3.1. Input ranges and distributions, and data sampling

Assignment of input ranges and specification of associated probability distribution functions (PDFs) is the most difficult and subjective stage in application of Monte Carlo analysis to hydrologic studies. The reason is simply that many model parameters are not directly measurable, and even if measurable, it would be cost prohibitive to collect numerous, random samples of inputs to determine their 'true' PDFs and ranges. Any available knowledge about the watershed and its parameters, including information from model documentation, previous studies, and other literature should be explored to make an 'educated guess' at this stage. Haan et al. (1998) and Helton (1993) indicate that proper assignment of input ranges is more influential on SA results than knowledge of actual PDFs. Furthermore, both studies suggest that simple distributions (e.g. uniform or triangular) would suffice for exploratory SA studies. Accordingly, for this study, all of the remaining 35 calibrable SWAT parameters were assumed to follow a uniform distribution. Here it should be noted that choice of uniform distribution over triangular is entirely due to its simplicity. The authors, however, suspect that triangular distribution may better represent some of the parameters. Ranges, as shown in Table 1, were assigned for each input and were based on a combination of a detailed literature review, SWAT user documentation, previous studies that applied SWAT, and knowledge of Big Creek watershed.

In application of Latin hypercube sampling to sample data from the specified distributions and ranges, a number of samples are first generated. For each sample, the hydrologic simulation model is executed using the topographic, soil, land use, and climate data previously described for the study watershed. Time series of daily streamflow and daily sediment concentration estimates at the calibration site

were extracted for the days on which observed data was available. The simulated and observed values of the corresponding outputs were compared, and the sum of the square of the residuals, a measure used as an output variable in the SA, was evaluated. Three hundred of such Monte Carlo realizations (i.e. input–output pairs) were generated for streamflow and sediment yield. These input–output data pairs were further analyzed, as described in the following sections, in order to apportion output uncertainty to each of the input factors and determine the most influential parameters for each of the two outputs.

4.3.2. Determination of measure of sensitivity

Stepwise regression analysis (Helton and Davis, 2000) is used herein to determine a measure of sensitivity that is capable of describing the effect of individual input factors on outputs. As the name implies, stepwise regression involves a step-by-step construction of multiple regressions, starting with a simple regression model and building upon it. At each step, another input that is highly correlated to output uncertainty unexplained by the regression model is considered, until a stage appears at which inclusion of another input factor does not appreciably improve performance of the regression model. In brief, stepwise regression analysis involves the following tasks:

  i. Perform a correlation analysis between each of the input factors and an output variable, and select the input factor most related to the output.
 ii. Construct a simple regression model between the output variable and an input factor highly correlated to the output variable.
iii. Determine output uncertainty not yet explained by the current regression model, perform a correlation analysis between output uncertainty not yet explained by the regression model and all input factors, excluding those factors selected in previous steps, and identify the highly correlated input factor.
 iv. Construct a multiple regression model between the output variable and the input factor(s) selected to this point.
  v. Test the performance of the regression model and the stopping criteria. Also, determine a measure of sensitivity for each of the input factors included

and test the significance of the individual input factors. If any of these input parameters are not any more significant, which could occur if the input factors are naturally correlated, reject the factor and construct another regression model using only the remaining input factors. Subsequently, evaluate the performance, stopping criteria, measure of sensitivity for the factors, and significance of each of the factors in the new model.

vi. If the stopping criteria are not satisfied, go to step *iii*.

The SA model used in this study extensively involves construction of regression models and correlation analysis. Myers (1990) may be referred for description of these two statistical methods. Since regression and correlation analyses are based on developing linear relationships between input and output variables, they often perform poorly when the relationships are nonlinear (Iman and Conover, 1979; Conover and Iman, 1981; Helton and Davis, 2000). The remedy to this problem is often to use the rank of the individual data rather than the actual data, a concept known as rank transformation (Iman and Conover, 1979; Conover and Iman, 1981). The use of rank-transformed data results in an analysis based on the strength of monotonic relationships rather than on the strength of linear relationships (Helton and Davis, 2000). Convinced by the nonlinearity of the hydrologic processes being investigated, rank transformation has, therefore, been implemented.

The coefficient of multiple determinations (R^2) and p-value, also known as the α-value, are the statistics used as stopping criteria. R^2 is an indicator of the extent to which the regression model explains the output uncertainty. R^2 values lie between zero and one; a value close to one indicates that the regression model is accounting for most of the uncertainty in the observed output. The p-value is the probability that a regression coefficient, b_i, with an absolute value as large as or larger than the one constructed in the analysis would be obtained if, in reality, there was no relationship between the input and the output variable, and, as a result, the apparent relationship that led to the constructed regression coefficient was due entirely to chance (Helton and Davis, 2000). A small p-value indicates that the regression coefficient is less likely to

have occurred simply by chance and suggests acceptance of the hypothesis that the regression coefficient is nonzero.

The R^2 value is evaluated for every regression model developed, and the difference in R^2 between two successive regression models is used as one of the stopping criteria. If the difference is less than a user-defined threshold, the stepwise regression procedure would be terminated, indicating that added consideration of the latest input factor did not significantly improve model performance. The F-statistic (Myers, 1990) was applied to test the hypothesis that all coefficients of the regression model (i.e. b_i for $i=1,...,L$, where L is the number of input factors included in the regression model being tested) are not significantly different from zero (i.e. $b=0$) and, hence, to determine the p-value.

The probability of exceeding an F-statistic value of F calculated with (v_1, v_2) degrees of freedom, $Q_F(\tilde{F} > F/v_1, v_2)$, can be estimated by (Helton and Davis, 2000)

$$Q_F(\tilde{F} > F/v_1, v_2) = I_K\left(\frac{1}{2}v_2, \frac{1}{2}v_1\right) \tag{1}$$

and

$$K = \frac{v_2}{v_2 + v_1 F} \tag{2}$$

where $I_K(a_1, a_2)$ denotes the incomplete beta function (Press et al., 1992), and M is the total number of samples. The probability $Q_F(\tilde{F} > F/L, M - L - 1)$ is the same as a p-value corresponding to the assumption that $b=0$. If the p-value under this assumption exceeds the user-defined threshold, which was one percent in this study, the hypothesis that $b=0$ would be accepted and the stepwise regression is terminated, implying that addition of the latest input factor did not significantly improve performance of the model. Therefore, the stepwise regression model is halted if either criterion, a p-value corresponding to $b=0$ exceeds a threshold value, or the difference in R^2 of successive regression models is lower than a threshold, is satisfied.

A test was also performed for the individual regression coefficients (i.e. b_i for $i=1,...,L$) included in the regression model to ensure that the hypothesis $b_i=0$ is rejected, otherwise the input factor would be removed and another regression model with $L-1$ input

factors would be constructed. This was accomplished using the t-statistic (Myers, 1990). The probability of obtaining a value $\tilde{t}$ from the t-distribution for which $|\tilde{t}| > |t|$, or $Q_t(|\tilde{t}| > |t|/M - L - 1)$, is given by (Helton and Davis, 2000)

$$Q_F(|\tilde{t}| > |t|/M - L - 1) = 1 - I_O\left[\frac{1}{2}(M - L - 1),\frac{1}{2}\right]$$

(3)

where

$$O = \frac{M - L - 1}{M - L - 1 + t^2}$$

(4)

and $I_O[a_1,a_2]$ denotes the incomplete beta function (Press et al., 1992). The probability $Q_t(|\tilde{t}| > |t|/M - L - 1)$ is the same as the p-value corresponding to $b_i = 0$. If the p-value of an individual input factor exceeds a user-defined probability, defined here as 2.5%, the input factor is removed from the regression model. The following aspects of stepwise regression analysis are used as qualitative and/or quantitative indicators of each input variable's importance, or measure of sensitivity.

- The order in which the variables are selected and enter into the regression model is a qualitative indicator of their importance, with the most important variable being selected first.
- The difference in R^2 values of subsequent regression models provides a measure of importance for the latest variable by indicating the amount of output accounted for by including the input factor into the regression model. Provided that the input factors are uncorrelated, the differences in the R^2 values for the regression models constructed at successive steps equals the fraction of the total variability in the output variable that can be accounted for by the individual input variables being added at each step (Helton, 1993).
- The absolute values of the standardized regression coefficients (SRC) in the regression models are quantitative indicators of an input factor's importance. SRC provides a measure of importance based on the effect of moving each variable away from its expected value by a fixed fraction of its standard deviation, while retaining all other variables at their expected values. The sign of a standardized

regression coefficient (SRC) indicates whether the input and output variable tends to increase, or decrease, together (i.e. positive sign), or tend to inversely related (i.e. negative sign).

4.2.2. Application results and discussion

The SA model was developed and incorporated as a submodule within the source code of SWAT. Input data collected for the demonstration watershed were used for application of the new SA model. The land use data obtained was for years 1999 and 2000, and land uses for the remaining simulation years (i.e. 1998 and 2001) were derived from the available land uses. Urbanized subbasins and other subbasins that were covered primarily by pasture, forest, and hay were assumed to have the same land use for the whole simulation period. Corn and soybean were assumed to rotate on an annual basis (i.e. farm land covered by corn in 1999 was assumed to have been covered by soybean in 1998). Based on interviews conducted with personnel from the Southern Illinois District office of the NRCS, pasture, hay, corn, and soybean were assumed to have been grown without tillage. Furthermore, other farm management operations, such as planting and harvesting/cutting dates of agricultural crops, were assumed to be consistent with dates commonly used in the demonstration watershed.

Three hundred Latin hypercube samples were generated from the input ranges shown in Table 1 and the assumed uniform distribution. For spatially varying parameters, the sampled values were assumed to be valid for the representative subbasin previously described with regard to parameterization. The relationships developed to link parameter values of the representative subbasin to other subbasins were used to derive corresponding values for other subbasins. Once model parameters for all subbasins were known, the computational subroutines of SWAT were executed and the output variable used in the SA for streamflow (i.e. sum of the square of residuals between observed and estimated values) was extracted. After obtaining input–output pairs for all of the 300 samples, rank-transformation was performed, and the resulting rank-transformed data were fed to subroutines of the SA model for stepwise regression. Generation of Latin hypercube samples,

model execution and extraction of output variable, rank-transformation, and application of the stepwise regression tasks were also accomplished for sediment yield. Results of the SA model for sediment yield are given in Table 2. The Table provides the input factors selected at each step of the stepwise regression model, R^2 of the regression model constructed using the input factor(s) selected at each stage, and the SRC and p-value of each input factor. The threshold values used as a stopping/parameter removal criteria for the difference in R^2 of successive regression models, the p-value based on the test that considers all input factors included to that point (i.e. $b=0$), and the p-value used

Table 2
Parameter sensitivity results for sediment yield

Step no.	Parameter	SRC	P-value	R^2
1	PRF	0.62	0	0.3842
2	PRF	0.60	0	0.5131
	CH_N2LW	−0.36	0	
3	PRF	0.60	0	0.5842
	CH_N2LW	−0.34	0	
	CH_EROD	0.27	0	
4	PRF	0.60	0	0.6514
	CH_N2LW	−0.34	0	
	CH_EROD	0.28	0	
	SPCON	0.26	0	
5	PRF	0.61	0	0.676
	CH_N2LW	−0.34	0	
	CH_EROD	0.28	0	
	SPCON	0.27	0	
	CH_COV	0.16	0	
6	PRF	0.61	0	0.684
	CH_N2LW	−0.35	0	
	CH_EROD	0.27	0	
	SPCON	0.27	0	
	CH_COV	0.16	0	
	SPEXP	−0.09	0.0067	
7	PRF	0.61	0	0.691
	CH_N2LW	−0.35	0	
	CH_EROD	0.28	0	
	SPCON	0.26	0	
	CH_COV	0.16	0	
	SPEXP	−0.09	0.0077	
	APM	0.08	0.0104	
8	PRF	0.61	0	0.6966
	CH_N2LW	−0.35	0	
	CH_EROD	0.28	0	
	SPCON	0.26	0	
	CH_COV	0.16	0	
	SPEXP	−0.09	0.0058	
	APM	0.08	0.0172	
	CN2PA	0.08	0.0207	

to test significance of the individual input factors (i.e. $b_i=0$) were 0.1, 1, and 2.5%, respectively. For these criteria, the SA identified 12 parameters that play a significant role in explaining the uncertainty of streamflow and eight inputs from the perspective of sediment yield. The relative importance of each of these input factors could be judged using the order in which the parameters were selected, improvement in R^2 that was achieved due to inclusion of the input factor into the regression model and the SRC coefficient of the input factor, both of which are listed in the Table. Interestingly, most of the parameters recommended by the original developers of SWAT for inclusion into calibration efforts, a recommendation that may not be valid for every study watershed, are among the input factors obtained for both streamflow and sediment yield. Note that a similar table for streamflow is available in Muleta (2003).

In conclusion of the parameter identification stage of this study, a combination of the three techniques (i.e. screening of which model parameters to estimate based on field data alone and which to determine based on calibration; parameterization of physically immeasurable parameters; and sensitivity analysis) has reduced the number of calibrable SWAT parameters to 20 (i.e. 12 for streamflow and 8 for sediment yield). These numbers can be better managed in the subsequent stage of calibration, parameter estimation.

5. Parameter estimation

Parameter estimation follows the decision of which parameters of the simulation model to calibrate. In addition to parameter identification, data quantity and quality, proper selection of a performance measure (i.e. objective function), and application of a proper search mechanism are the most decisive factors in successful automatic model calibration. In this study, a GA is used to systematically and extensively search through combinations of parameters to achieve the set that is 'best' in terms of satisfying the criterion of accuracy. The criterion used here is the sum of the square of differences between corresponding simulated and observed values at a given time scale.

From the perspective of data quantity and quality, though, there is a general agreement that the amount

of information contained in the data is more important than the amount of data. Nevertheless, there have been some recommendations regarding minimum data length. For example, Klemes (1986) argued on the necessity of 3–5 years of data, and Sorooshian and Gupta (1995) argued that the length of data (i.e. number of observations) should be at least 20 times the number of parameters to be estimated (e.g. 200 to calibrate 10 parameters). For the current application, the length of available data at PRS is about 793 samples for streamflow and 682 samples for sediment yield, respectively. Since the number of calibrable parameters is 20, the available data length is sufficient with respect to the recommendation by Sorooshian and Gupta (1995).

For selection of an objective function, some studies (e.g. Gupta et al., 1998; Sorooshian and Gupta, 1995) have indicated that calibration success is very dependent on the objective function used as a selection criterion. The objective function used should be consistent with the anticipated application of the model. In this study, the overall objective is to achieve a reasonable simulation of sediment yield. Since sediment yield cannot be simulated without correct simulation of the streamflow, both streamflow and sediment yield are being calibrated. However, sediment yield is a seasonally fluctuating watershed response, and the majority of sediment is usually generated and transported during flood events. In Big Creek, for example, Demissie et al. (1990) have reported that 96.3% of the sediment moves during five percent of the time, based on data collected during the late 1980s. Therefore, it is essential that flood peaks are simulated correctly. Accordingly, the root of the mean of square of errors, or residuals (RMSE) between observed and simulated values of streamflow and sediment concentration, an objective function biased towards peak flows, is used in this study.

The relationship between the model parameters identified and the output variables being calibrated is too complicated to derive simple (i.e. smooth and unimodal) mathematical functions and, hence, their derivative information. This implies that gradient-based optimization methods have little to offer for complex problems like calibration of distributed models. As an alternative, heuristic search algorithms, such as GAs, that require no derivative information about the objective function or constraints, have been practically proven to work well on nonlinear, nonconvex, and multimodal problems (Schwefel, 1995). GAs are a technique that applies Darwinian theory optimization problems. Though not ultimately guaranteed to locate global optima (Cieniawski et al., 1995), GAs search a wide portion of the solution space and, thus, have a better capability of locating optimal solutions. In fact, the majority of GA literature consistently demonstrates an ability to identify global or very near global optima for a range of complicated problems (Nicklow, 2000).

The parameter estimation effort was first conducted for streamflow and subsequently for sediment yield. Best-fit values obtained for the parameters during calibration of streamflow were used during calibration of sediment yield. For calibration of each of these two watershed responses, 1500 random samples, generated from the ranges and the uniform distribution assigned for the most influential parameters, collectively known as a population, were the initial solution candidates supplied to the GA. Each sample, known as chromosome, is defined by a sequence of parameter values, known as genes. A gene is essentially a value for one of the influential parameters selected for the output variable (i.e. 12 for streamflow and 8 for sediment yield) being calibrated. Since the values of genes could assume any real number, a continuous GA, also known as a real coded GA, was used. Once a chromosome is selected, the computational subroutines of SWAT were executed to provide the corresponding RMSE using observed and simulated data at PRS, the station with relatively longer periods of historical data. Sample generation and evaluation of the corresponding objective function (RMSE) was repeated until a user-defined number of initial chromosomes had been reached. The objective was to determine a set of parameter values yielding a simulation result that best resembles the observed data. Therefore, a solution alternative that has the lowest RMSE was the fittest set of parameters.

The initial solution candidates were ranked in ascending order, and parent alternatives were selected using multiple tournament selection. Here, five solution candidates among those ranked within the top half of the population were randomly selected, and the fittest of these five chromosomes was selected as a mate. The procedure was repeated, and two

successively selected mates were paired to form parents that ultimately produce two offspring alternatives through the process of uniform crossover. The process of choosing mates, forming pairs, and creating offspring were continued until a user-defined number of new candidates (i.e. half of the current population) had been obtained.

Mutation was performed on 15% of the solution candidates that were selected randomly. For each of mutant candidates, the specific gene to be manipulated was also decided by random selection. Once the solution candidate and the specific gene to be mutated were known, a value randomly selected from the range and distribution assigned for the parameter replaced the existing value of the gene. This mutation process minimizes the chance of premature convergence by introducing new traits or characteristics into the search process. Fitness values were then determined for the mutated solution candidates and for the newly born alternatives.

Ranking solutions according to the ascending order of their fitness values, selecting mating pairs, creating offspring by crossover, injecting new traits by mutation, and evaluating fitness values for the mutated and new candidates was repeated for 75 generations, or iterations. The number of solution candidates was, however, reduced from 1500 to 150 from the third generation forwards due to issues related to computational time. The genes corresponding to the fittest chromosome at the end of 75 generations were considered to be 'optimal' parameter values, or values that yield the 'best' match between observed and simulated data. For these GA parameters, the search process required about 50 h of CPU time on a 1.69 GHz, Pentium IV processor.

Calibration results for streamflow are given in Table 3 and Fig. 4. Results for sediment yield are given in Table 4 and Fig. 4. The Tables provide values of the selected parameters and values of E_f, the coefficient of efficiency (Nash and Sutcliffe, 1970) used to test goodness-of-fit, for the watershed responses simulated using the best-fit parameter values, as well as default values of the model parameters without calibration. The value of E_f can vary from negative infinity to unity, which corresponds to the ideal condition when measured and estimated values perfectly match. In addition, a summary of the annual outputs is presented in the Tables observed, and calibrated and noncalibrated

Table 3
Summary of streamflow calibration results

Parameter	Calibrated Value	E_f (calibrated)	E_f (default)
SURLAG	1.9512	0.744	−0.38
MSK_CO2	0.005		
EPCO	1		
ESCO	0.0882		
CANMXP	2.8488		
CH_N2UP	0.0824		
CN2PA	52.5877		
CH_N2LW	0.01		
GWQMN	37.3652		
REVAPMN	28.2478		
RCHRG_DP	0.582		
GWDELAYE	47.3122		
Year	Observed	Calibrated	Default simulation
Annual Streamflow (mm)			
1999 (June 25–Dec.31)	31.082	24.569	516.9 mm/yr
2000	262.743	191.897	
2001	131.919	88.865	

Table 4
Summary of calibration results for sediment concentration

| Parameter	Calibrated value	E_f (Calibrated)	E_f (default)
CN2PA	59.9097	0.461	−4.1
CH_N2LW	0.0345		
PRF	0.1963		
CH_EROD	0.1549		
SPCON	0.0031		
CH_COV	0.6129		
SPEXP	1.496		
APM	1.0247		
Year	Observed	Calibrated	Default simulation
Average daily sediment concentration (mg/l)			
1999 (June 25–Dec. 31)	25.95	23.06	196.77 mg/l
2000	65.49	60.40	
2001	54.77	50.52	
Annual streamflow obtained using these parameter values (mm)			
1999 (June 25–Dec. 31)	31.082	29.585	516.9 mm/yr
2000	262.743	220.138	
2001	131.919	93.686	
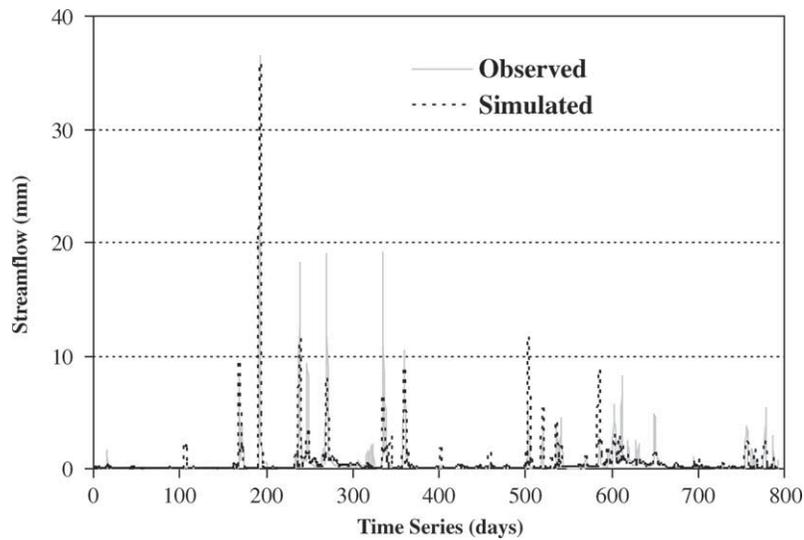
Fig. 3. Comparison of observed and calibrated streamflow.

simulations. Figs. 3 and 4 provides a graphical comparison of the observed data and calibrated values on a daily basis for streamflow and sediment yield, respectively.

From observation of these results, the reader should acknowledge the improvement achieved in the model simulations by the calibration effort for both streamflow as well as sediment yield. Contrary to the default simulation, which highly overestimates watershed responses, the calibrated model seems to 'slightly'

underestimate both streamflow and sediment yield. These figures alone demonstrate the importance of some sort of model calibration. Results for streamflow simulations, with an E_f of 0.74, are very good when compared to that of previous calibration efforts attempted for SWAT (e.g. Santhi et al., 2001; Eckhardt and Arnold, 2001). The results obtained herein for sediment yield may not be considered impressive, but considerable improvement is achieved when compared to the default simulation. Note that no previous work
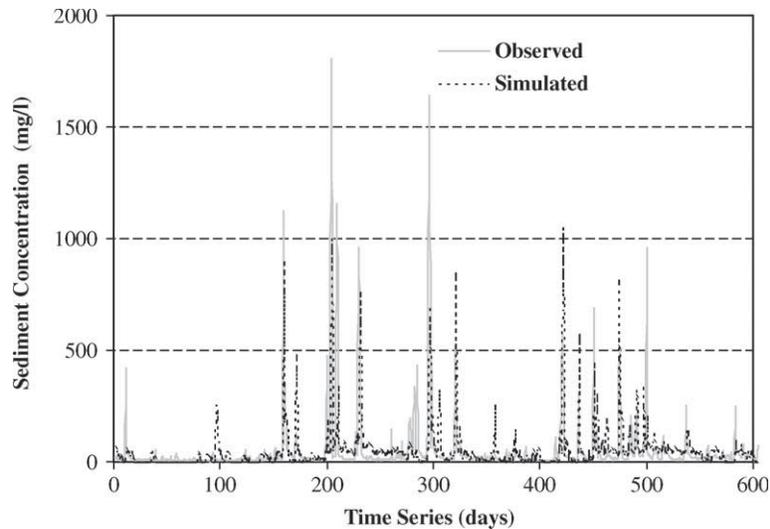

Fig. 4. Comparison of observed and calibrated sediment yield.
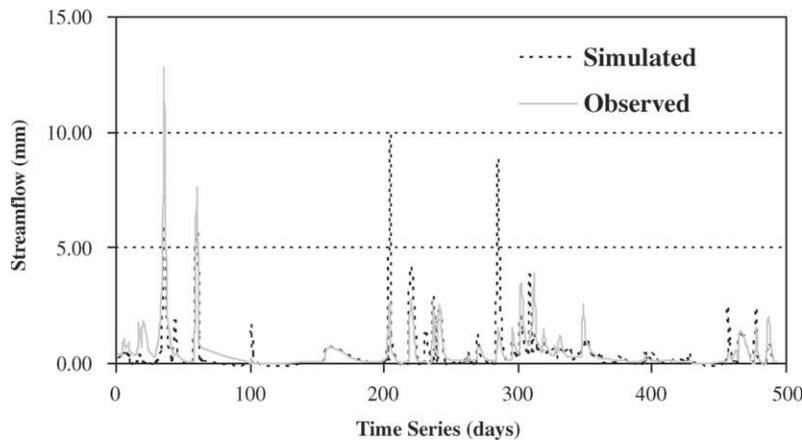
Fig. 5. Verification results for streamflow.

has been conducted in relation to calibrating SWAT for sediment yield on a daily time scale. The sediment calibration effort has achieved a significant improvement for streamflow (i.e. Table 4) when compared to the results given in Table 3.

6. Model verification

Once calibration has been conducted to estimate the best values for model parameters, the outcome needs to be verified to determine if the results provide adequate information for answering the questions that face decision-makers. For distributed models, it is important to determine if spatially distributed predictions are satisfactory so that model results from subbasins can be determined with reasonable confidence. This requires multiple gauging stations in the watershed to measure model outputs of interest. For Big Creek watershed, as described previously, there exists two flow and sediment gauging stations, PRS and CRS. Owing to availability of relatively longer periods of data at the PRS station, and since it drains most of the watershed, data from PRS was used for calibration. The calibrated model was then tested for its capability of simulating a reasonable response at the gauging station not used in calibration, or CRS. Verification results at CRS are graphically presented in Figs. 5 and 6 for streamflow and sediment yield, respectively.
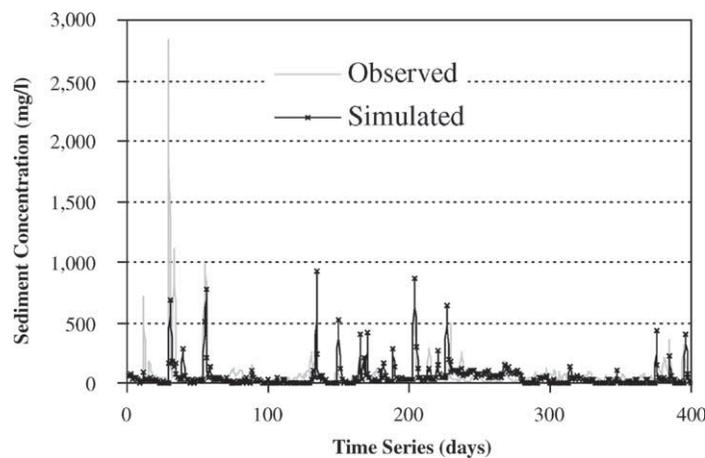

Fig. 6. Verification result for sediment concentration.

The verification result is not particularly encouraging, but indicates a significant improvement over the default simulation. An E_f value of 0.233 is obtained for streamflow at CRS, as compared to -4.14 for the default simulation, and an E_f of -0.005 is obtained for sediment concentration, as compared to -0.26 obtained using the default simulation. The study by Senarath et al. (2000) attempted the type of verification implemented in this research (i.e. calibrating using data at one gauging station and conducting verification using data from other gauging station(s) in the watershed) for streamflow, but no similar work has been completed for sediment yield. Using a physically based, two-dimensional, distributed parameter model known as CASC2D (Julien et al., 1995), a model specifically designed to accurately model runoff alone, Senarath et al. (2000) obtained an absolute error (i.e. absolute deviation between observed and estimated values) for runoff volume ranging from 28 to 42% for four internal gauging stations used for verification. For this study, the absolute deviation between the observed data and simulated estimates for streamflow and sediment yield at CRS station is found to be 33 and 26 percent, respectively; a result quite comparable with that of Senarath et al. (2000). The poor performance of the verification effort could be attributed to a relatively short duration of data (i.e. less than three years) used in the model calibration effort. Application of the model to a watershed with longer data records may better reveal the capability of the simulation model.

7. Uncertainty analysis

Even after calibration, there are many reasons for the modeler to be suspicious of model results. Uncertainty arises due to incomplete information (e.g. quantity and quality of input data) used for calibration; simplifications and approximations introduced into the modeling exercise; and parameter estimates (e.g. capability of the search algorithm). It is, therefore, essential that every modeling effort be accompanied by an uncertainty analysis (UA), a technique of determining reliability of model predictions, accounting for various sources of uncertainty. The mean-value, first-order, second-moment method (Melching, 1995), Rosenblueth's point estimation

method (Rosenblueth, 1975), Harr's point estimation method (Harr, 1989), and Monte Carlo based approaches have been used to investigate reliability of hydrologic model outputs.

The UA method used in this study, known as Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992; Beven and Freer, 2001) is a Monte Carlo based approach. However, it is grounded on a conception that conflicts with model calibration procedures, such as the one applied in this study, that strive to derive a single best optimal parameter set to satisfy a user-defined fitness function. The GLUE procedure argues that given the inherent uncertainties associated with distributed watershed models and modeling exercises, "…*there is no reason to expect that any one set of parameter values will represent a true parameter set (within a given model structure) to be found by calibration procedures*" (Beven and Binley, 1992). The method is based on the premise that, once the unavoidability of the uncertainties is acknowledged, it is only possible to make an assessment of the likelihood of a particular parameter set being an acceptable simulator of the watershed behavior. Alternatively stated, the methodology considers the optimal parameter set derived by the calibration procedure simply as parameter values that are more likely to represent the watershed being modeled under the given conditions (i.e. data and model used, and calibration procedure applied); it argues that for a different set of data (e.g. better quality data) or even for a different simulation period, this seemingly optimal parameter set may perform poorly and many other parameter sets may perform better. The approach thus recognizes that many model parameters within a given model structure and input data could predict the watershed behavior reasonably well, a concept known as the *equifinality* problem (Beven and Freer, 2001), and applies this conception to predict bounds of uncertainty for the model outputs.

Among the major advantages of using the GLUE methodology for UA of a distributed watershed model are that the method considers global uncertainty; it accounts for the effects of uncertainties due to input data, model structure, and parameters on reliability of the model output (Beven and Freer, 2001); the method is computationally tolerable if systematic sampling methods such as Latin hypercube sampling are used; and the method is conceptually simple. The major

drawback of the GLUE methodology is subjectivity of the likelihood level assignment that groups the parameter sets into the acceptable and nonacceptable categories. For details on the GLUE methodology, the reader is refereed to Beven and Binley (1992) and Beven and Freer (2001).

7.1. Application results and discussion

The GLUE methodology has similarities to the SA model developed in this study in the sense that GLUE also requires generation of numerous input samples from ranges and distributions assigned for the input factors involved in the uncertainty estimation; it requires derivation of model simulations for the generated samples; it involves determining a user-defined measure of fitness, or likelihood measure; and it applies input–output information to determine measures of uncertainty. As in the SA model, Latin hypercube sampling is used to generate input samples for the UA model, hence improving the computational efficiency of the methodology. Therefore, the programmatic codes developed for the SA model to accomplish tasks related to Latin hyper sampling of input factors, parameterization of spatial varying input factors, and extraction of model predictions at PRS are directly used for the UA as well, and a discussion of these tasks is not repeated here.

All of the parameters used during the parameter estimation stage (see Table 3) were considered in the UA. All of the input factors were assumed to follow a uniform distribution, and ranges of the factors were preserved to the assignment made during the SA. Using the input ranges and distributions assigned to the eight parameters, 5000 Latin hypercube samples were generated, and SWAT was executed to predict sediment yield and streamflow values at PRS for each sample. Values of E_f, used as a measure of fitness, were evaluated for both sediment yield and streamflow. E_f values of 0.45 and 0.25 were used as threshold likelihood values to consider a parameter set that is a behavioral, or acceptable, simulator of the system for streamflow and sediment yield, respectively. The E_f values obtained for streamflow and sediment yield during the calibration procedure aided in the choice of these threshold values. Again, it is important to emphasize that there is a considerable degree of subjectivity involved in the choice of these

threshold values, a major shortcoming of the GLUE methodology. Of the 5000 parameter sets, only 887 behavioral simulations were accepted according to the threshold values and were involved in the subsequent steps of the UA.

Following the approach used by Brazier et al. (2000) to derive a combined measure of likelihood (CML_i) that simultaneously serves as a goodness of fit for both streamflow and sediment yield, E_f values obtained for the two variables were combined as follows for each of the accepted behavioral simulations

$$CML_i = E_{f,Fi} \times E_{f,Yi} \tag{5}$$

where $E_{f,Fi}$ and $E_{f,Si}$ are coefficients of efficiencies of streamflow and sediment yield, respectively, for simulation i. CML_i values are then rescaled from zero to one, such that the rescaled likelihood measures ($SCML_i$) sum to unity and, hence, yield a probability distribution function for the parameter sets. The function ultimately serves as a probabilistic weighting function for model outputs in order to derive measures of uncertainty

$$SCML_i = \frac{CML_i}{\sum_{i=1}^{N} CML_i} \tag{6}$$

In order to derive measures of uncertainty, simulations results from all of the accepted behavioral simulations were ranked in their ascending order for each time step (i.e. daily, monthly, and annually), and the distribution function of the predictions is calculated using the $SCML_i$ of the corresponding model run. Cumulative distribution of the SCML is derived and used to determine various statistics such as
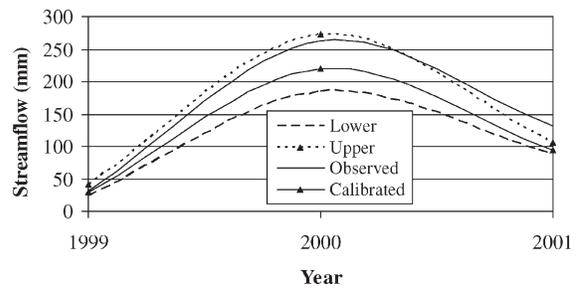

Fig. 7. 95% confidence limit with observed and calibrated streamflow.
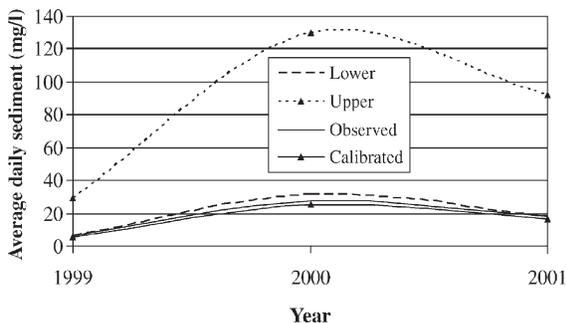
Fig. 8. 95% confidence limit with observed and calibrated sediment concentration.

the uncertainty bounds (e.g. lower and upper bounds of the 95% confidence limits).

Figs. 7 and 8 provide uncertainty bounds corresponding to the 95% confidence limit, observed data and calibrated model simulation data for annual streamflow and average daily sediment concentration, respectively. The bounds given in Fig. 7 indicate that the model's streamflow prediction is fairly consistent in the sense that the uncertainty bounds are 'narrow', having a ratio between the upper and lower bounds of less than a factor of two. In addition, the bounds obtained have embraced the observed data and the calibrated model simulations for the year 1999 and 2000. Unlike streamflow simulation, the 95% confidence limits for sediment concentration (i.e. Fig. 8) are very wide. The upper bound exceeds the lower bound by up to a factor of six, indicating a great deal of uncertainty associated with the model's sediment prediction. Fig. 8 clearly shows the over-predicting behavior of the model for sediment concentration, as well as its inability to envelop the observed data for 2000 and 2001. Once again, this work is conducted using a watershed with limited recorded sediment and streamflow data, and the result given here should not be a generalized characteristic of the SWAT model. Application of the UA model to other watersheds that have a relatively longer data record could yield additional insight regarding the behavior of SWAT.

8. Summary and conclusions

An automatic calibration model is developed in order to improve streamflow and sediment yield estimated by the USDA's distributed watershed model known as SWAT. The automatic technique presented allows one to avoid the limitations of existing 'trial-and-error' calibration techniques. The model is demonstrated using a watershed located in southern Illinois. The parameter specification (i.e. identification) stage of model calibration is accomplished using hierarchy of three techniques: parameter screening, spatial parameterization, and a detailed parameter sensitivity analysis. The parameter sensitivity analysis was conducted using Monte Carlo based approach that uses Latin hypercube sampling strategy along with stepwise regression analysis of rank-transformed input–output data pairs. The parameter identification exercise effectively reduced the number of calibrable parameters of SWAT's streamflow and sediment concentration variables to a number that can be easily managed at the parameter estimation phase of model calibration.

Parameter estimation was performed using a genetic algorithm, a search method that has become popular in recent years for solving realistic, complex control problems. Application results indicate significant improvement over default simulations and results of previous works that have reported on calibration of SWAT's streamflow and/or sediment concentration estimates. The verification exercise indicated poor performance of the calibrated model for both streamflow and sediment concentration, a behavior shared by other distributed hydrologic models. In addition to the calibration and verification study, a model that uses the Generalized Likelihood Uncertainty Estimate (GLUE) is developed and incorporated into the watershed simulation model to analyze uncertainty of streamflow and sediment yield estimates. The model accounts for the major sources of uncertainty including those arising from input data, model algorithms and assumptions, and parameters estimates. Application of the uncertainty estimation model indicates that the model's streamflow prediction is fairly consistent in that the uncertainty bounds are 'narrow'. In contrast, sediment yield prediction of the SWAT model was found to involve a great deal of uncertainty. The study did, however, rely upon a relatively short duration of recorded data due to the lack of sufficient data for the study watershed. The capability of the automatic calibration algorithm and the uncertainty analysis methodology, as well as

the behavior of the simulation model could be more effectively tested if the calibration, the verification, and the uncertainty analysis efforts could be applied to a 'data rich' watershed.

Acknowledgements

References

Arnold, J.G., Allen, P.M., 1996. Estimating hydrologic budgets for three Illinois watersheds. Journal of Hydrology 176, 55–77.

Arnold, J.G., Srinivasan, R., Muttah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. Journal of the American Water Resources Association 34 (1), 73–89.

ASCE, 1999.Anon., 1999. GIS Modules and Distributed Models of Watersheds. American Society of Civil Engineers, Reston, VA.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processing 6, 279–298.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249, 11–29.

Brazier, R.E., Beven, K., Freer, J., Rowan, J.S., 2000. Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP—the water erosion prediction project—for sites in the UK and USA. Earth Surface Processes and Landforms 25, 825–845.

Cieniawski, S.E., Eheart, J.W., Ranjithan, S., 1995. Using genetic algorithms to solve a multiobjective groundwater monitoring problem. Water Resources and Research 31 (2), 399–409.

Conover, W.J., Iman, R.L., 1981. Rank transformations as a bridge between parametric and non-parametric statistics. The American Statistician 35, 124–129.

Demissie, M., Soong, T., Allgire, R., Keefer, L., Makowski, P., 1990. Cache River Basin: Hydrology, Hydraulics and Sediment Transport. Vol. 1: Background, Data Collection, and Analysis, Contract Report 484. Illinois State Water Survey, Champaign, IL.

Demissie, M., Knapp, V.H., Parmer, P., Kriesant, D.J., 2001. Hydrology of the Big Creek Watershed and its Influence on the Lower Cache River, Contract Report 2001-06. Illinois State Water Survey, Champaign, IL.

Duan, Q., Sorooshian, S., Gupta, V.K., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resources and Research 28 (4), 1015–1031.

Eckhardt, K., Arnold, J.G., 2001. Automatic calibration of distributed catchment model. Journal of Hydrology 251, 103–109.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resources and Research 34 (4), 751–763.

Haan, C.T., Storm, D.E., Al-Issa, T., Prabhu, S., Sabbagh, G.J., Edwards, D.R., 1998. Effect of parameter distribution on uncertainty analysis of hydrologic models. Transactions of ASAE 41 (1), 65–70.

Harr, M.E., 1989. Probability estimates for multivariate analyses. Applied Mathematical Modelling 13, 313–318.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering and System Safety 42, 327–367.

Helton, J.C., Davis, F.J., 2000. Sampling-based methods, in: Saltelli, A., Chan, K., Scott, E.M. (Eds.), Sensitivity Analysis. Wiley, New York.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, MI.

Iman, R.L., Conover, W.J., 1979. The use of rank transform in regression. Technometrics 21, 499–509.

Iman, R.L., Conover, W.J., 1980. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. Communications in Statistics: Theory and Methods A 9, 1749–1874.

Julien, P.Y., Saghafian, B., Ogden, F.L., 1995. Raster based hydrologic modeling of spatially-varied surface runoff. Water Resources Bulletin 31 (5), 523–536.

Klemes, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal 31 (1), 13–24.

McKay, M.D., Conover, W.J., Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Melching, C.S., 1995. Reliability Estimation, in: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, CO.

Muleta, M.K., 2003. A decision support system for the management of non-point source pollution from watersheds. PhD Thesis, College of Engineering, Southern Illinois University at Carbondale, Carbondale, IL.

Myers, R.H., 1990. Classical and Modern Regression with Applications. PWS-KENT Publishing Company, Boston, MA.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I—a discussion of principles. Journal of Hydrology 125, 277–291.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2001. Soil and Water Assessment Tool Theoretical Documentation, Version 2000. Grassland, soil and water research service, Temple, TX.

Nicklow, J.W., 2000. Discrete-time optimal control for water resources engineering and management. Water International 25 (1), 89–95.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in Fortran. Cambridge University Press, New York, NY.

Refsgaard, J.C., 1997. Parameterization, calibration and validation of distributed hydrologic models. Journal of Hydrology 198, 69–97.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrologic models. Water Resources and Research 32 (7), 2189–2202.

Rosenblueth, E., 1975. Point estimates for probability moments. Proceedings of the National Academy of Sciences USA 72 (10), 3812–3814.

Saltelli, A., 2000. What is sensitivity analysis?, in: Saltelli, A., Chan, K., Scott, E.M. (Eds.), Sensitivity Analysis. Wiley, New York.

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

Santhi, C., Arnold, J.G., Williams, J.R., Srinivasan, R., Hauck, L.M., 2001. Validation of the SWAT model on a large river basin with point and non point sources. Journal of the American Water Resources Association 37 (5), 1169–1188.

Schwefel, H.-P., 1995. Evolution and Optimum Seeking. Wiley, New York.

Senarath, S.U.S., Ogden, F.L., Downer, C.W., Sharif, H.O., 2000. On the calibration and verification of two-dimensional, distributed, hortonian, continuous watershed models. Water Resources and Research 36 (6), 1495–1510.

Sorooshian, S., Gupta, V.K., 1995. Model calibartaion, in: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publication, Highlands Ranch, CO.

Williams, J.R., 1975. Sediment routing for agricultural watersheds. Water Resources Bulletin 11 (5), 965–974.