MULTI-SITE CALIBRATION OF THE SWAT
MODEL FOR HYDROLOGIC MODELING

X. Zhang, R. Srinivasan, M. Van Liew

ABSTRACT. *The growing popularity of applying complex, semi-physically based distributed hydrologic models to solve water resource problems poses important issues that must be addressed related to the use of spatial data to calibrate and validate such models. In this study, a single-objective optimization method (GA) and a multi-objective optimization algorithm (SPEA2) were applied to optimize the parameters of the Soil and Water Assessment Tool (SWAT) using observed streamflow data at three monitoring sites within the Reynolds Creek Experimental Watershed, Idaho. Results indicated that different optimization schemes can lead to substantially different objective function values, parameter solutions, and corresponding simulated hydrographs. Thus, the selection of an optimization scheme can potentially impact modeled streamflow. Parameters estimated by optimizing the objective function at three monitoring sites consistently produced better goodness-of-fit than those obtained by optimization at a single monitoring site. This stresses the importance of collecting detailed, spatially distributed data to conduct simultaneous multi-site calibrations. When applied with multi-site data, the single-objective (GA) method better identified parameter solutions in the calibration period, but the multi-objective (SPEA2) method performed better in the validation period. Overall, the application of different optimization schemes in the Reynolds Creek Experimental Watershed demonstrated that the single-objective (GA) and the multi-objective (SPEA2) optimization methods can provide promising results for multi-site calibration and validation of the SWAT model. These results are expected to help the users of SWAT and other distributed hydrologic models understand the sensitivity of distributed hydrologic simulation to different calibration methods and to demonstrate the advantages and disadvantages of single-objective and multi-objective parameter estimation methods.*

Keywords. *Automatic calibration, Distributed hydrologic model, Multi-objective optimization, Multi-site calibration, SWAT.*

In recent years, hydrologic models have been increasingly used by hydrologists and water resource managers to understand and manage natural and human activities that affect watershed systems. These hydrologic models can contain parameters that cannot be measured directly due to measurement limitations and scaling issues (Beven, 2000). For practical applications in solving water resources problems, model parameters are calibrated to produce model predictions that are as close as possible to observed values. When calibrating a hydrologic model, one or more objectives are often used to measure the agreement between observed and simulated values. The objectives to be optimized can be the combination of multiple goodness-of-fit estimators (e.g., relative error, coefficient of determination), multiple variables (e.g., water, energy, sediment, and nutrients), and multiple sites (Yapo et al., 1998; Gupta et al., 1998; Santhi et al., 2001b; Van Liew and Garbrecht, 2003; White and Chau-

bey, 2005; Demarty et al., 2005; Cao et al., 2006; Engeland et al., 2006; Bekele and Nicklow, 2007). With the recent development of distributed hydrologic models that can spatially simulate hydrologic variables, the use of multi-site observed data to evaluate model performance is becoming more common.

The Soil and Water Assessment Tool (SWAT) model (Arnold et al., 1998) is a continuous, long term, distributed-parameter model that can simulate surface and subsurface flow, soil erosion and sediment deposition, and nutrient fate and movement through watersheds. SWAT has been applied worldwide for hydrologic and water quality simulation. For example, the SWAT model has been incorporated into the U.S. Environmental Protection Agency (USEPA) Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) software package, and is being applied by the USDA for the Conservation Effects Assessment Project (CEAP) (Gassman et al., 2007). Over 250 peer-reviewed published articles have reported SWAT applications, reviews of SWAT components, or other SWAT-related research (Gassman et al., 2007). The SWAT model has been extensively tested for hydrologic modeling at different spatial scales. For example, Gollamudi et al. (2007) evaluated SWAT in two agricultural fields in southern Quebec; Spruill et al. (2000) and Chu and Shirmohammadi (2004) successfully simulated monthly flow in a 5.5 km² watershed in Kentucky and a 3.4 km² watershed in the Piedmont physiographic region of Maryland, respectively; Santhi et al. (2001a) successfully simulated monthly flow in the Bosque River watershed with a drainage area of 4,277 km²; Zhang et al. (2007) applied

The authors are **Xuesong Zhang, ASABE Member Engineer,** Postdoctoral Research Associate, Department of Ecosystem Sciences and Management, Texas A&M University, College Station, Texas; **Raghavan Srinivasan, ASABE Member,** Professor, Departments of Biological and Agricultural Engineering and Ecosystem Sciences and Management, Texas A&M University, College Station, Texas; and **Michael Van Liew, ASABE Member,** Hydrologist, Montana Department of Environmental Quality, Helena, Montana. **Corresponding author:** Raghavan Srinivasan, 1500 Research Parkway, Suite B223, College Station, TX 77845; phone: 979-845-5069; fax: 979-862-2607; e-mail: r-srinivasan@tamu.edu.

SWAT for daily and monthly streamflow simulation in the 5,239 km² Luohe River watershed, China; and Srinivasan et al. (1998) and Arnold et al. (1999) evaluated the SWAT model for hydrologic modeling of the conterminous U.S.

In the application of SWAT, multi-site data have been used to calibrate parameter values (Santhi et al., 2001b; Van Liew and Garbrecht, 2003; White and Chaubey, 2005; Cao et al., 2006; Bekele and Nicklow, 2007). For simultaneous multi-site automatic calibration of SWAT, two types of calibration methods are usually implemented. The first calibration method aggregates the different objective function values calculated at each monitoring site into one integrated value, and then applies the single-objective optimization algorithms for parameter estimation (e.g., van Griensven and Bauwens, 2003). The second calibration method uses multi-objective evolutionary algorithms to optimize the different objective functions calculated at multiple sites simultaneously, and finds a set of multiple Pareto optimal solutions (e.g., Bekele and Nicklow, 2007). Currently, the Shuffled Complex Evolution (SCE) algorithm (Duan et al., 1992) is incorporated into SWAT for automatic parameter estimation using a single objective (one objective function or integrated multiple objective functions) (van Griensven and Bauwens, 2003).

In many SWAT applications, the model has been calibrated using objective functions at a single site or for integrated multi-sites objective functions, but multi-objective evolutionary algorithms were seldom applied for multi-site calibration. Therefore, the objective of this study was to compare and evaluate the effect of single and multi-objective optimization schemes on the calibrated parameter values and simulated hydrographs from SWAT. In order to accomplish this objective, a program for parameter optimization of SWAT using single and multi-objective evolutionary algorithms was developed. Based on previous studies (Tang et al., 2006; Zhang et al., 2008a), the single-objective and multi-objective optimization algorithms applied in this study were Genetic Algorithm (GA) and Strength Pareto Evolutionary Algorithm 2 (SPEA2), respectively. These two optimization algorithms were implemented to estimate the parameters in SWAT for the Reynolds Creek Experimental Watershed in Idaho with observed streamflow data at three monitoring sites. The differences between estimated parameter values and simulated hydrographs are explored and discussed. The results of this study are expected to help the users of SWAT and other distributed hydrologic models understand the sensitivity of distributed hydrologic simulation to different calibration methods and to demonstrate the advantages and disadvantages of single-objective and multi-objective parameter estimation methods.

MATERIAL AND METHODS

STUDY AREA DESCRIPTION

The Reynolds Creek Experimental Watershed (RCEW), with drainage area of 239 km², is located about 80 km southwest of Boise, Idaho, and exhibits a considerable degree of spatial heterogeneity. The topography of the watershed ranges from a broad, flat alluvial valley to steep, rugged mountain slopes, with a range in elevation from 1101 to 2241 m (Seyfried et al., 2000). Because of orographic effects, the average annual precipitation ranges from about 250 mm at the outlet to more than 1100 mm at the upper end of the wa-

tershed. Perennial streamflow is generated at the highest elevations in the southern part of the RCEW where deep, late-lying snowpacks are the source for most water (Seyfried et al., 2000). Although much of the watershed has steep, shallow, rocky soils, there are areas of deep, loamy rock-free soils. Land cover on the RCEW consists of rangeland and forest communities of sagebrush, greasewood, aspen, and conifers (94%) and irrigated cropland (6%). The average slopes of the three subwatersheds are about 0.04 m m⁻¹. The lengths of stream segments are about 10, 11, and 26 km for the Salmon, Tolgate, and Outlet subwatersheds.

The locations of the RCEW and three streamflow monitoring gauges (Salmon, Tolgate, and Outlet) are shown in figure. 1. For modeling purposes, the RCEW was partitioned into subwatersheds connected by a stream network and then into hydrologic response units (HRUs) consisting of unique combinations of land cover and soils in each subwatershed. The delineation of subwatersheds and HRUs allows SWAT to consider the spatial heterogeneity of land uses and soils in the watershed. It was assumed that there were no interaction between HRUs, that is, the HRUs are non-spatially distributed. Through HRU delineation, the computational costs of simulations can be minimized by lumping similar soil and land use areas into a single unit (Neitsch et al., 2005). The number of subwatersheds, number of HRUs, and drainage area of each monitoring site are listed in table 1.

APPLICATION OF SWAT

The hydrologic routines within SWAT account for snowfall and melt, vadose zone processes (i.e., infiltration, evaporation, plant uptake, lateral flow, and percolation), and
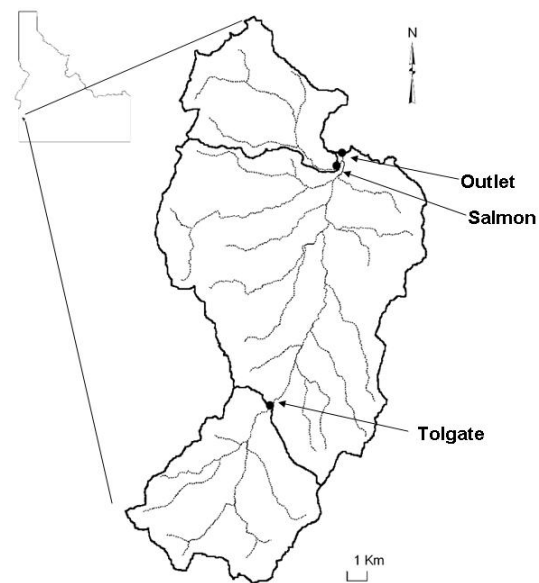

Figure 1. Reynolds Creek Experimental Watershed and three streamflow monitoring gauges.

Table 1. Basic characteristics of Reynolds Creek Experimental Watershed.

Watershed	Subwatershed	No. of Subbasins	No. of HRUs	Area (km²)
Reynolds Creek	Tolgate	16	47	55
	Salmon	10	19	36
	Outlet	60	162	239

Table 2. Parameters for calibration of the SWAT model.

Code	Parameter	Description	Range
Parameters governing surface water response			
1	CN2	Curve number II	±20%
2	ESCO	Soil evaporation compensation factor	0-1
3	SOL_AWC	Available soil water capacity	±20%
Parameters governing subsurface water response			
4	GW_REVAP	Groundwater reevaporation coefficient	0.02-0.2
5	REVAPMN	Threshold depth of water in the shallow aquifer for reevaporation to occur (mm)	0-500
6	GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	0-5000
7	GW_DELAY	Groundwater delay (days)	0-50
8	ALPHA_BF	Base flow recession constant	0-1
9	RCHRG_DP	Deep aquifer percolation fraction	0-1
Parameters governing basin response			
10	CH_K2	Effective hydraulic conductivity in main channel alluvium (mm h^{-1})	0.01-150
11	TIMP	Snow pack temperature lag factor	0-1
12	SURLAG	Surface runoff lag coefficient (day)	0-10
13	SFTMP	Snow melt base temperature (°C)	0-5
14	SMTMP	Snowfall temperature (°C)	0-5
15	SMFMX	Maximum snowmelt factor for June 21 (mm H_2O/°C-day)	0-10
16	SMFMN	Minimum snowmelt factor for Dec. 21 (mm H_2O/°C-day)	0-10

groundwater flow. Surface runoff volume is estimated using a modified version of the Soil Conservation Service (SCS) curve number (CN) method (USDA-SCS, 1972). A kinematic storage model (Sloan et al., 1983) is used to predict lateral flow, whereas return flow is simulated by creating a shallow aquifer (Arnold et al., 1998). Either the Muskingum or Variable Storage method is used for channel flood routing. Outflow from a channel is adjusted for transmission losses, evaporation, diversions, and return flow. The SWAT model contains a large number of parameters that are used to describe the spatially distributed water movement through the watershed system. Many of these parameters, such as CN and surface runoff lag coefficient (SURLAG), cannot be measured directly and thus must be estimated through calibration. In this study, 16 parameters that govern the surface water response, subsurface water response, and basin response of the SWAT model were used in calibration. A general description of the 16 parameters is shown in table 2 (Van Liew et al., 2007). The default parameters were determined by the methods introduced by Neitsch et al. (2005). The adjusting range of each parameter was determined based on recommendations from previous literature (e.g., Neitsch et al., 2005; van Griensven et al., 2006; Van Liew et al., 2007). The parameter calibration was conducted simultaneously for the three subwatersheds because the calibration results of the three stations are interrelated. There are two reasons for the simultaneously adjusting of parameters. First, the Salmon and Tolgate stations feed into the Outlet before exiting the watershed (fig. 1), which means that the calibration of Salmon and Tolgate can influence the calibration of the Outlet. Second, some parameters (e.g., surface lag and snow melt rate) can only be adjusted for the entire watershed for SWAT. For further discussion of multi-site calibration of SWAT, refer to Migliaccio and Chaubey (2007).

The period from water year (WY) 1966 to 1967 was used as the model parameter initialization period, which allowed the model to cycle multiple times to minimize the effects of the user's estimates of initial state variables, such as soil water content and surface residue. Daily streamflow data from 1968 to 1969 were used to calibrate SWAT, and daily streamflow data from 1970 to 1972 were used for validation.

OPTIMIZATION ALGORITHMS

Before describing the optimization algorithms, several common variables are introduced here: D is the number of optimized parameters; T is the maximum number of generations; t is the current generation number; P_t is the population of parameter solutions at generation number t; N is the number of parameter solutions in a population; $\mathbf{x}$ is the vector of decision variables, which consists of hydrologic parameters in this study; $\mathbf{x}_i$ is the ith parameter solution in the population, which is a D-dimensional vector $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iD})$, and x_{id} is the dth dimension of the ith parameter solution. The parameter solution is denoted as a chromosome in the GA and SPEA2.

Single-Objective Optimization

For single-objective optimization, there is only one objective function that needs to be optimized. Given the objective function $f : \Omega \subseteq R^D, \Omega \neq \Theta$, for $\mathbf{x} \in \Omega$, the value $f^* = f(\mathbf{x}^*) < \infty$ is called a global maximum if and only if $\forall \mathbf{x} \in \Omega :$ $f(\mathbf{x}^*) \geq f(\mathbf{x})$, where $\mathbf{x}^*$ is the parameter solution for global maximum and the set Ω is the feasible parameter space.

There are many automatic calibration algorithms that can be used to implement the single-objective optimization. For example, Duan et al. (1992) developed the SCE-UA algorithm, which has been incorporated into the 2005 version of SWAT and successfully applied for parameter estimation of SWAT (e.g., van Griensven and Bauwens, 2003; Eckhardt et al., 2005; Van Liew et al., 2005; van Griensven and Bauwens, 2005; Van Liew et al., 2007; Zhang et al., 2007; Zhang et al., 2008b). Besides SCE-UA, Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Artificial Immune Systems (AIS), and Differential Evaluation (DE) also have been successfully applied for parameter estimation in hydrologic models or other complex problems. Zhang et al. (2008a) compared the efficacy of these five global optimization algorithms for calibrating SWAT, and found GA is a promising single-objective optimization method.

GA involves stochastic search procedures inspired by evolutionary processes in biology of natural selection and genetics (Holland, 1975; Goldberg, 1989), such as inheritance,

mutation, selection, and crossover. With flexibility and robustness, GA has been successfully applied to solve complex nonlinear programming problems in many science and engineering branches (Reca and Martinez, 2006). The general procedure for applying GA (fig. 2) involves three major operators: selection, crossover, and mutation.

Selection operator. The fittest chromosomes in a population are preferred to be selected to reproduce promising new offspring. A roulette wheel algorithm is applied to select chromosomes for the crossover and mutation operations. The probability of a chromosome being selected as a parent is proportional to its fitness. In order to overcome shortcomings of using the original fitness value in the roulette wheel algorithm, a relative fitness value (F') for each chromosome was calculated using the following equation (Reca and Martinez, 2006):

$$F'(r) = (q+1)(N+1) - 2rq$$

where q is the selection pressure ($0 \leq q \leq 1$), and r is the rank assigned to the chromosomes based on their fitness (Reca and Martinez, 2006).

Crossover operator. The purpose of crossover is exchanging important building blocks of two parent chromosomes to generate new offspring. The probability of crossing two chromosomes is determined by an input parameter (P_c). There are three main crossover methods: one-point crossover, two-point crossover, and uniform crossover (Goldberg, 1989). In this study, the uniform crossover operator was applied. In uniform crossover, for each dimension of the parent chromosomes, two parent chromosomes swap the parameter values with a probability of 0.5 to generate new offspring.

Mutation operator. For each dimension of the offspring chromosome, a random number between [0,1] is generated. If this random number is less than the mutation probability (P_m), then a newly generated parameter value will replace the old parameter value of the specific dimension.

After the newly bred chromosomes are generated through the selection, crossover, and mutation operators, they will be incorporated into the population using a steady-state-delete-worst plan (Reca and Martinez, 2006), in which the least fit member of the parent population is eliminated and replaced by the offspring. Several control parameters of the GA were determined according to Schaffer et al. (1989) and Reca and Martinez (2006): P_m was set to 1/D, P_c was set equal to 0.5, and q was set to 1.

Multi-Objective Optimization

For multi-objective optimization problems, a series of objective functions need to be taken into account simultaneously. The general multi-objective optimization problem can be defined as: find the parameter solution $\mathbf{x}^*$ that will optimize the objective function vector $\mathbf{f(x)} = [f_1(\mathbf{x}), f_2(\mathbf{x}), \mathrm{K}, f_m(\mathbf{x})]$ where m is the number of objective functions. As there are multiple objective functions that need to be optimized simultaneously, and different objective functions prefer different parameter solutions, it is difficult to find a single global optimum parameter solution. The Pareto optimality concept is defined to evaluate whether a parameter set is "optimal" or not. An objective function vector $\mathbf{f(x')} = [f_1(\mathbf{x'}), f_2(\mathbf{x'}), \mathrm{K}, f_m(\mathbf{x'})]$ is said to dominate another objective function vector $\mathbf{f(x)} = [f_1(\mathbf{x}), f_2(\mathbf{x}), \mathrm{K}, f_m(\mathbf{x})]$ (denoted by $\mathbf{f(x')} \mathrm{f} \mathbf{f(x)}$), if $\forall i \in \{1, \ldots m\}$, $f_i(\mathbf{x'}) \geq f_i(\mathbf{x}) \wedge \exists i \in \{1, \ldots m\} : f_i(\mathbf{x'}) > f_i(\mathbf{x})$ (Zitzler and Thiele, 1999). If the ob–
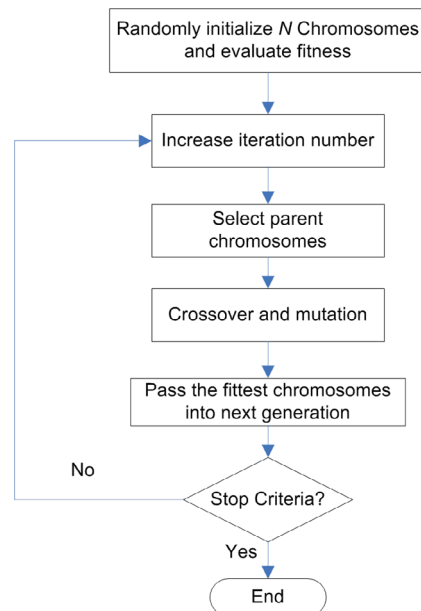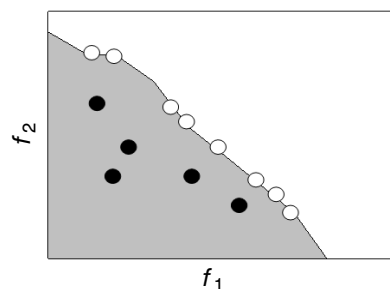

Figure 2. Flowchart of GA.

Figure 3. Graphical illustration of the *PF** and dominated objective function vectors for a maximization problem. Solid circles denote dominated objective function vectors, and empty circles consist of the *PF**.

jective function vector $\mathbf{f(x^*)}$ of a point $\mathbf{x}^* \in \Omega$ is not dominated by all the other objective function vectors of the parameter solutions in the feasible parameter space, then $\mathbf{x}^*$ is taken as a Pareto optimal parameter solution. The Pareto optimal set (P^*) is defined by the set of parameter solutions that are not dominated by other parameter solutions. The objective function vectors corresponding to the Pareto optimal set comprise the Pareto front (PF^*). An illustration of the PF^* and dominated objective function vectors for a maximization problem with two objectives f_1 and f_2 is shown in figure 3. The purpose of multi-objective optimization is to search the feasible parameter space and find those parameter solutions that are Pareto optimal.

Among many multi-objective optimization algorithms that have been presented and successfully applied, SPEA2 (Zitzler et al., 2001) was adopted here to conduct parameter estimation in the SWAT model according to the comparison of several state-of-the-art multi-objective optimization algorithms (Tang et al., 2006). In order to give a clear description of SPEA2, several new symbols are defined: $\overline{P}_t$ is the external archive at generation t, which is used to store the parameter solutions with high fitness values; $\overline{N}$ is the external archive size; and A is the Pareto optimal set. SPEA2 applies different procedures than the single-objective GA to calculate the fit-

ness of each parameter solution and keep the diversity of candidate parameters. The basic procedures for implementing SPEA2 are illustrated in figure 4 and the following sections (Zitzler et al., 2001):

Initialization. Generate an initial population P_0 and create the empty archive (external repository) $\overline{P}_0 = \Theta$. Set $t = 0$.

Fitness assignment. In order to calculate fitness values of chromosomes in P_t and $\overline{P}_t$, three major steps need to be implemented. First, each chromosome i in the archive $\overline{P}_t$ and the population P_t is assigned a strength value $S(i)$ representing the number of solutions it dominates. Second, on the basis of S values, the raw fitness $R(i)$ of a chromosome i is calculated as:

$$R(i) = \sum_{j \in P_t + \overline{P}_t, \, j \dagger i} S(j)$$

where $R(i) = 0$ represents a non-dominated chromosome, while a high $R(i)$ value means that i is dominated by many chromosomes. Third, the raw fitness $R(i)$ needs to be adjusted to incorporate the density of each chromosome, in case many chromosomes have the same raw fitness when most chromosomes do not dominate each other. The density of each chromosome is calculated using the kth nearest neighbor method, which defines the density of a chromosome as a function of its distance to the kth nearest neighbors in the objective space ($\sigma_i{}^k$) (Zitzler et al., 2001). In SPEA2, k is set equal to $\sqrt{N + \overline{N}}$. Then the density of each chromosome i is defined as:

$$D(i) = \frac{1}{\sigma_i^k + 2}$$

where the number "2" is added to the denominator to ensure that $D(i)$ is less than 1 (Zitzler et al., 2001). Finally, adding $D(i)$ to the raw fitness value $R(i)$ yields each chromosome's fitness $F(i) = R(i) + D(i)$.

Environmental selection. Copy all Pareto optimal chromosomes in P_t and $\overline{P}_t$ to $\overline{P}_{t+1}$. If the size of $\overline{P}_{t+1}$ exceeds $\overline{N}$, then reduce $\overline{P}_{t+1}$ by means of truncating the non-dominated chromosomes with less fitness $F(i)$, otherwise if the size of $\overline{P}_{t+1}$ is less than $\overline{N}$, then fill $\overline{P}_{t+1}$ with best dominated chromosomes in P_t and $\overline{P}_t$.

Termination: If $t > T$ or another stopping criterion is satisfied, then set A to the set of parameter vectors represented by the non-dominated chromosomes in $\overline{P}_{t+1}$.

Mating selection and variation. Perform tournament selection (Goldberg, 1989) with replacement on $\overline{P}_{t+1}$ to fill the mating pool. Apply crossover and mutation operators to the mating pool, and set P_{t+1} to the resulting population. Increment the generation counter ($t = t + 1$), and go to the fitness assignment step.

EXPERIMENTAL TEST DESIGN

The optimization objective functions are indicators of agreement between the measured and simulated series of the variable of interest. The sum of squares of residuals (SSR) is an often applied objective function in calibrating hydrologic models (Van Liew et al., 2007). In this study, the Nash-Sutcliffe efficiency (E_{ns}), a normalized form of SSR, was se-
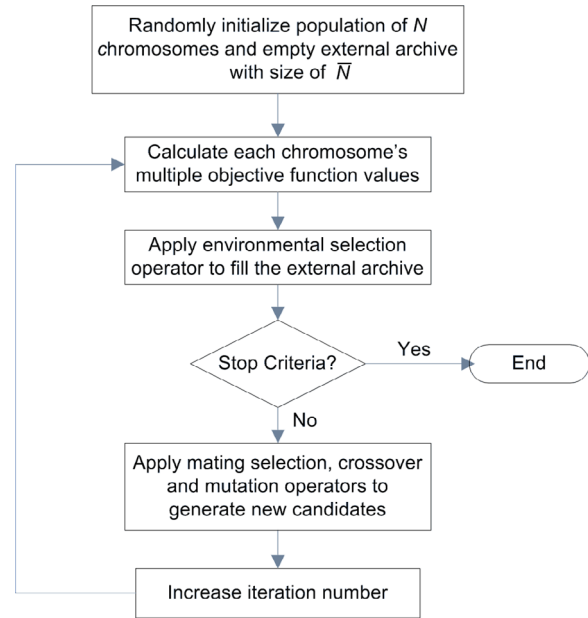

Figure 4. Flowchart of SPEA2.

lected. The formula to calculate E_{ns} is (Nash and Sutcliffe, 1970):

$$E_{ns} = 1.0 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{1}$$

where P_i is the model simulated value, O_i is the observed data, $\overline{O}$ is the mean for the observed data, and $i = 1, 2, ..., n$, where n is the total number of pairs of simulated and observed data. E_{ns} indicates how well the plot of the observed value versus the simulated value fits the 1:1 line, and ranges from $-\infty$ to 1 (Legates and McCabe, 1999).

In order to understand the effect of single-objective and multi-objective parameter optimization schemes on the resulting calibrated parameter values and simulated hydrographs, several GA optimization schemes and a multi-objective optimization scheme (SPEA2) were applied. Specifically, four GA and four SPEA2 optimization cases were applied with the E_{ns} objective function: GA applied at Salmon (GA-sal), GA applied at Tolgate (GA-tol), GA applied at the outlet (GA-out), and GA applied for the sum of all (GA-sum). SPEA2 can optimize the E_{ns} objective functions at Salmon, Tolgate, and Outlet simultaneously and find a set of Pareto optimal solutions; therefore, the objective function optimized by SPEA2 is a vector, that is [E_{ns}-Salmon, E_{ns}-Tolgate, E_{ns}-Outlet, and E_{ns}-Sum]. Several representative Pareto optimal solutions were selected for analysis and comparison: SPEA2-sal, SPEA2-tol, SPEA2-out, and SPEA2-sum denote the parameter solutions with the best objective function values at Salmon, Tolgate, Outlet, and the sum of all three monitoring stations, respectively. The comparison between the different optimization schemes was mainly based on the eight representative parameter sets (GA-sal, GA-tol, GA-out, GA-sum, SPEA2-sal, SPEA2-tol, SPEA2-out, and SPEA2-sum). The GA and SPEA2 algorithms are stochastically based. Based on previous studies of
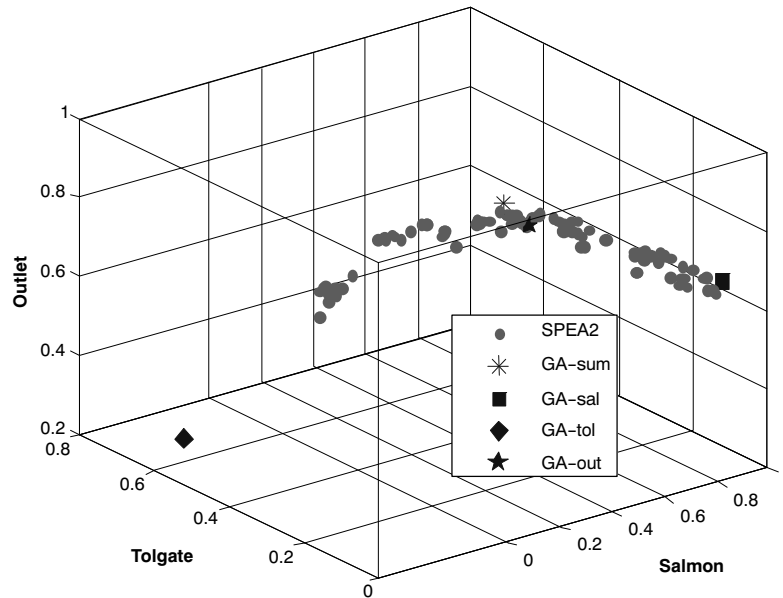
Figure 5. The best or Pareto optimal objective function values at Salmon, Tolgate, and Outlet obtained by GA-sal, GA-tol, GA-out, GA-sum, and SPEA2.

applying evolutionary optimization algorithms for SWAT (e.g., van Griensven, and Bauwens. 2003; Tolson and Shoemaker, 2007; Zhang et al., 2008a), all the optimization schemes were run as 10 trials (10,000 model evaluations for each trial) to obtain the optimized parameter solutions and objective function values. The best results obtained by the 10 trials were used as the final parameter solutions for each optimization scheme.

RESULTS AND DISCUSSION

OPTIMIZED OBJECTIVE FUNCTION VALUES, PARAMETER SETS, AND HYDROGRAPHS

The objective function values at Salmon, Tolgate, and Outlet obtained by the GA-sal, GA-tol, GA-out, GA-sum, and SPEA2 schemes are shown in figure 5. For each single-objective optimization scheme, a best objective function value was obtained, while 96 Pareto optimal objective function vectors were obtained for the multi-objective optimization scheme. For each objective function, the values obtained by the different optimization schemes are listed in table 3. Simulations demonstrated that the single-objective optimization schemes can identify better values for each separate objective function than the multi-objective optimization scheme. This means that the objective function vectors found by each single optimization scheme are not dominated by other objective function vectors obtained by SPEA2, and can be added to the Pareto front found by SPEA2. Although the single-objective optimization schemes can identify better results for each separate objective function, they need to be run several times separately. On the other hand, with one trial, SPEA2 can find multiple objective function vectors that perform as well as the parameter solutions obtained by GA. These results are in agreement with the "no free lunch theorem," which states that "for any optimization algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class" (Wolpert and Macready, 1997). For example, GA-sum achieved better objec-

tive values for E_{ns}-Tolgate and E_{ns}-Sum, while SPEA2-Sum obtained better results for E_{ns}-Salmon, and E_{ns}-Outlet. For parameter calibration, the single-objective and multi-objective optimization schemes can find parameter solutions that are not inferior to each other.

Test results demonstrated that a substantial difference exists between the objective function values at different monitoring sites obtained by each single-objective optimization scheme. For example, GA-sal obtained an E_{ns} value of 0.854 at Salmon, while GA-tol obtained an E_{ns} value of –0.361 at Salmon. Optimizing the objective function value at one site can lead to a serious bias of objective function values at other sites. Similarly, the multi-objective optimization scheme can also obtain a wide range of objective function values at each monitoring site. For example, the range of objective function values at Salmon, Tolgate, Outlet are [0.097, 0.827], [0.02, 0.599], and [0.471, 0.763], respectively. Given the substantial variation of the optimized objective function values, the corresponding parameter values were expected to scatter within the feasible space. The normalized parameter values obtained by different optimization schemes are shown in figure 6. All of the parameter values were normalized between their lower and upper bounds so that they ranged between 0 and 1. As expected, the value of each parameter varied substantially depending on the selected optimization scheme.

Table 3. Nash-Sutcliffe values obtained by different parameter optimization schemes at Salmon, Tolgate, and Outlet for the calibration period.

	Salmon	Tolgate	Outlet	Sum
GA-sal	0.85	0.02	0.69	1.57
GA-tol	-0.36	0.62	0.25	0.51
GA-out	0.50	0.29	0.78	1.57
GA-sum	0.62	0.44	0.74	1.80
SPEA2-sal	0.83	0.03	0.677	1.57
SPEA2-tol	0.13	0.60	0.47	1.20
SPEA2-out	0.55	0.18	0.76	1.49
SPEA2-sum	0.70	0.31	0.74	1.75
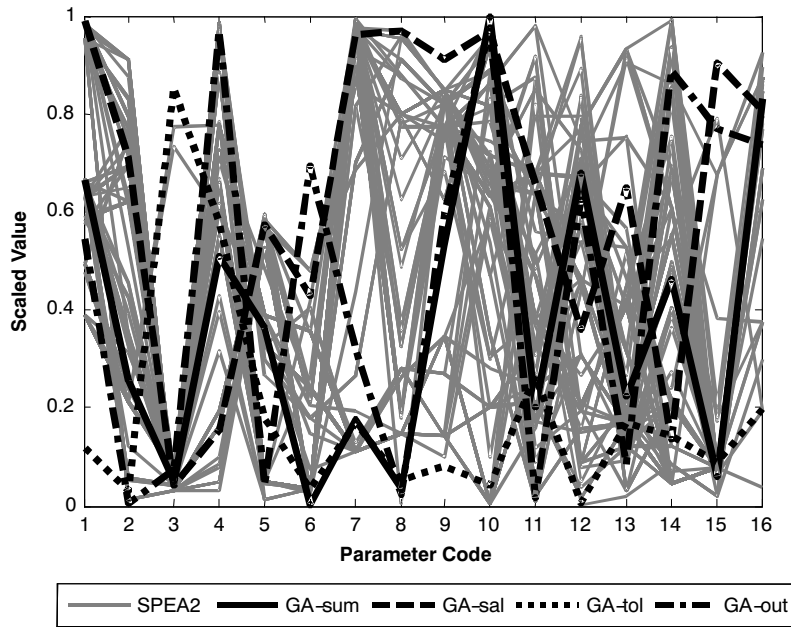
Figure 6. Normalized parameter values obtained by SPEA2 and different GA-based single-objective optimization schemes (parameter codes are listed in table 2).

Figure 7. Simulated hydrographs using parameter sets calibrated by different optimization schemes (thick solid lines are the observed hydrograph; thin solid lines are the simulated hydrographs using eight representative parameter sets calibrated by GA and SPEA2 optimization schemes).

The range of the initial CN was 45 to 60. For the single-objective optimization schemes, the optimized parameter values varied from each other. For example, the normalized CN values are 0.99, 0.12, 0.54, and 0.67 for GA-sal, GA-tol, GA-out, and GA-sum, respectively. For SPEA2, the range of CN values obtained by the 96 Pareto optimal parameter sets is [0.38, 0.99]. These differences between the optimized parameter values reveal that the relationships between streamflow and topography, land use, and precipitation are different for each subwatershed, which results in a specific parameter solution for a given subwatershed.

Table 4. Kolmogorov-Smirnov test results between the hydrographs simulated using different parameter solutions
obtained by different optimization schemes at Salmon, Tolgate, and Outlet for the calibration period.

		GA_sal	GA_tol	GA_out	GA_sum	SPEA2-sal	SPEA2-tol	SPEA2-out	SPEA2-sum
Salmon	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	0	--				
	SPEA2-sal	0	1	1	1	--			
	SPEA2-tol	1	1	1	1	1	--		
	SPEA2-out	1	1	0	0	1	1	--	
	SPEA2-sum	1	1	0	1	0	1	0	--
Tolgate	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	0	--				
	SPEA2-sal	0	1	1	1	--			
	SPEA2-tol	1	0	1	1	1	--		
	SPEA2-out	1	1	1	1	1	1	--	
	SPEA2-sum	1	1	1	1	1	1	1	--
Outlet	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	0	--				
	SPEA2-sal	0	1	1	1	--			
	SPEA2-tol	1	0	1	1	1	--		
	SPEA2-out	1	1	0	0	1	1	--	
	SPEA2-sum	1	1	0	1	1	1	1	--

Note: 1 represents that there is a significant difference between the two simulated hydrographs at a significant level of 0.05, while 0 represents a lack of significance (α = 0.05).

The eight representative parameter sets obtained by both single and multi-objective optimization schemes were used to simulate the hydrographs at Salmon, Tolgate, and Outlet (fig. 7). Considerable variation among the optimization schemes is apparent in the simulated hydrograph for each of these stations (fig. 7). The two-sample Kolmogorov-Smirnov test (Massey, 1951) was used to test whether there is a statistically significant difference between the simulated hydrographs using the parameter sets obtained by different optimization schemes (table 4). Of the total 28 comparisons at each monitoring site, 21, 25, and 22 pairs of hydrographs are significantly different from each other. This indicates that the selection of parameter optimization schemes can lead to significantly different simulated hydrographs, which may have important implications for water resources management investigations.

MODEL VALIDATION USING PARAMETERS OBTAINED BY GA AND SPEA2 OPTIMIZATION SCHEMES

After calibration, the optimized parameter sets need to be validated using another independent set of observed data. Observed daily streamflow data (1970-1972) at Salmon, Tolgate, and Outlet was used to validate the optimized parameter sets obtained by different optimization schemes. Figure 8 shows the simulated and observed hydrographs at Salmon, Tolgate, and Outlet. The hydrographs simulated by different parameter sets exhibit moderate variation from each other for the validation period. Based on the two-sample Kolmogorov-Smirnov test, 20, 26, and 24 pairs of hydrographs are significantly different from each other (table 5), which indicates that the selection of different parameter sets has significant influence on the simulation results for water resources management investigations.

The evaluation coefficients (table 6) demonstrated that GA-sum and SPEA2-sum performed better for the validation than the other parameter calibration schemes. If only one specific objective function was emphasized in the calibration process, the calibrated parameter sets tended to achieve relatively better performance for that specific objective function at the cost of the performance of the other objective functions. For example, GA-tol and SPEA2-tol achieved a value of E_{ns}-Tolgate larger than 0.25, while E_{ns}-Salmon was less than –0.35. In addition, the emphasis on one specific objective function tends to calibrate the model parameters so that they fit that specific objective function, while ignoring important information contained in other objective functions. On the contrary, GA-sum and SPEA2-sum consider all objective functions and search for a compromise among these different objective functions. Given the uncertainties associated with the observed data and model structure, the overfit to one specific objective function leads to worse performance during the validation period than the calibration period. For example, GA-sum and SPEA2-sum achieved better E_{ns} values at all three monitoring sites than other optimization schemes that only emphasized one specific objective function during calibration. Hence, the better performance of GA-sum and SPEA2-sum compared to other optimization schemes stresses the importance of collecting more detailed spatially distributed data to calibrate a distributed hydrologic model.

Often hydrologic conditions of a validation period are different from those of a calibration period, which may lead to differences in performance of the parameter solutions for the respective periods. For example, GA-sum achieved the highest E_{ns} values based on the sum of the three sites for the calibration period, but SPEA2-sum performed better than GA-sum for the validation period. The 92 Pareto optimal pa–
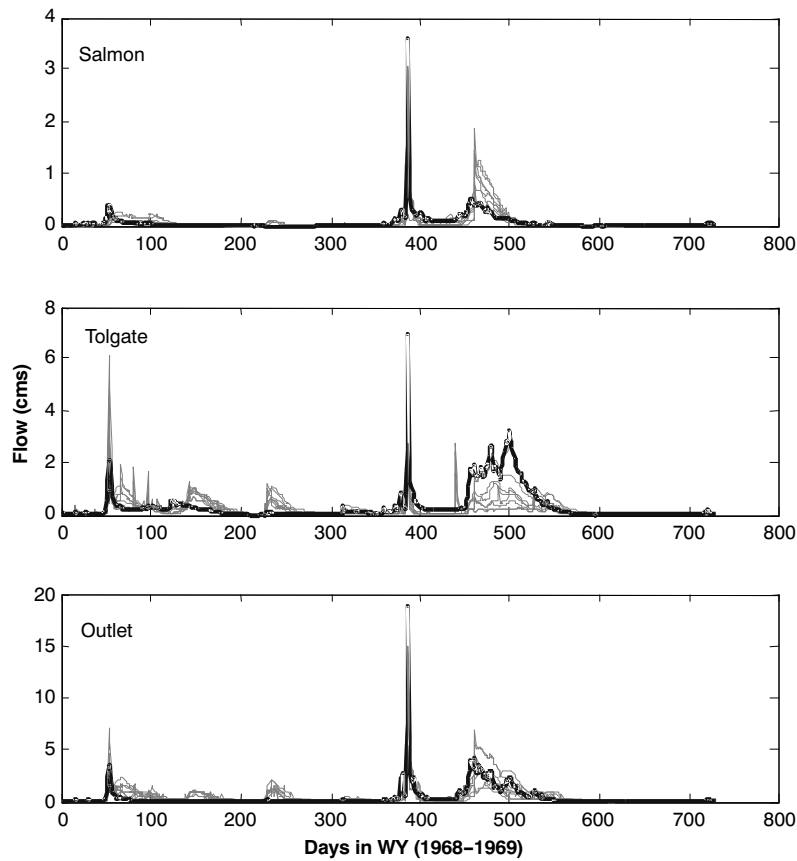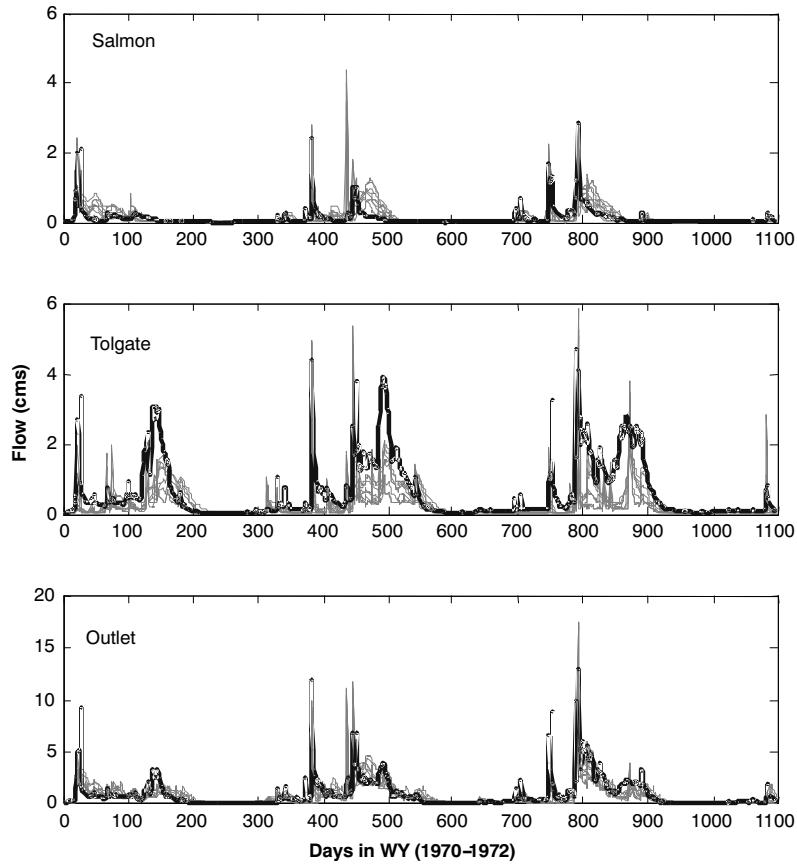
Figure 8. Simulated hydrographs using parameter sets calibrated by different optimization schemes (thick solid line is the observed hydrograph; thin solid lines are the simulated hydrographs using eight representative parameter sets calibrated by GA and SPEA2 optimization schemes).

Table 5. Kolmogorov-Smirnov test results between the hydrographs simulated using different parameter solutions obtained by different optimization schemes at Salmon, Tolgate, and Outlet for the validation period.

		GA_sal	GA_tol	GA_out	GA_sum	SPEA2-sal	SPEA2-tol	SPEA2-out	SPEA2-sum
Salmon	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	1	--				
	SPEA2-sal	1	1	0	0	--			
	SPEA2-tol	1	0	1	1	1	--		
	SPEA2-out	1	1	0	1	1	0	--	
	SPEA2-sum	1	1	0	0	0	1	1	--
Tolgate	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	1	--				
	SPEA2-sal	1	1	1	1	--			
	SPEA2-tol	1	0	1	1	1	--		
	SPEA2-out	1	1	1	1	1	1	--	
	SPEA2-sum	1	1	1	1	0	1	1	--
Outlet	GA_sal	--							
	GA_tol	1	--						
	GA_out	1	1	--					
	GA_sum	1	1	1	--				
	SPEA2-sal	1	1	1	1	--			
	SPEA2-tol	1	0	1	1	0	--		
	SPEA2-out	1	1	1	1	1	1	--	
	SPEA2-sum	1	1	1	1	0	0	1	--

Table 6. Objective values obtained by different parameter optimization schemes at Salmon, Tolgate, and Outlet for the validation period.

	Salmon	Tolgate	Outlet	Sum
GA-sal	0.30	0.01	0.56	0.88
GA-tol	-0.48	0.35	0.43	0.30
GA-out	0.08	0.12	0.62	0.82
GA-sum	0.32	0.51	0.73	1.56
SPEA2-sal	0.29	0.01	0.59	0.89
SPEA2-tol	-0.37	0.26	0.45	0.33
SPEA2-out	0.20	0.24	0.63	1.07
SPEA2-sum	0.44	0.49	0.73	1.65

rameter sets (except for SPEA2-sal, SPEA2-tol, SPEA2-out, and SPEA2-sum) achieved by SPEA2 were also evaluated for the validation period. Some of the parameter sets outperformed the eight representative parameter solutions obtained during calibration. The best validation objective function values are 0.46 at Salmon, 0.52 for Tolgate, 0.73 for Outlet, and 1.68 for the sum, respectively. These values are better than those listed in table 5. Among the 96 Pareto optimal parameter sets, 22 of them achieved the sum of three objective function values larger than 1.55, and two of them achieved better results than GA-sum for all three objective functions. These validation results demonstrate that the multiple Pareto optimal parameter sets obtained by SPEA2 may contain some useful information that GA did not identify. The multiple Pareto optimal parameter sets can also allow practitioners to use different ways to select reasonable parameter values. If only one "most likely" parameter set is used to forecast or simulate streamflow, then graphical visualization techniques (Gupta et al., 1998) and expert knowledge (Khu and Madsen, 2005) can be applied to assist the parameter selection.

SUMMARY

With the increasing availability of spatial hydrologic data and the growing popularity of complex, semi-physically based distributed hydrologic models, use of the spatial data to calibrate and validate hydrologic models is becoming an increasingly important issue. In this study, different optimization schemes were applied to optimize a distributed hydrologic model, SWAT, using observed streamflow data at three monitoring sites within the Reynolds Creek Experimental Watershed in Idaho. The results demonstrated that different optimization schemes can lead to substantially different objective function values, parameter solutions, and corresponding simulated hydrographs. This in turn indicates that the selection of optimization schemes can significantly impact how well hydrologic models simulate actual streamflow. Parameters estimated by optimizing the objective function at three monitoring sites consistently produced better goodness-of-fit than those obtained through optimizing the objective function at a single monitoring site, which stresses the importance of having spatially distributed data to conduct such simultaneous multi-site calibration.

When applied with multi-site data, the single-objective (GA) method can identify better parameter solutions in the calibration period, but the multi-objective (SPEA2) method performed better in the validation period. The multi-objective optimization method, however, can identify multiple Pareto optimal parameter solutions, which allows hydrologic practitioners to use expert knowledge and visual graphic analysis to select one preferred solution. The multi-objective optimization method also eliminates the need for multiple runs by determining the optimal values simultaneously. Overall, the application of different optimization schemes in the Reynolds Creek Experimental Watershed demonstrated that the single-objective (GA) and multi-objective (SPEA2) optimization methods both produce reasonable results for multi-site calibration and validation of the SWAT model. Results also supported the "no free lunch theorem" (Wolpert and Macready, 1997) regarding optimization algorithms. Each optimization scheme has its strengths and weaknesses and may perform better under one set of hydrologic conditions as compared to another; therefore, a method to combine the strengths of different optimization schemes deserves further research in the future.

REFERENCES

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. 1998. Large-area hydrologic modeling and assessment: Part I. Model development. *J. American Water Resources Assoc.* 34(1): 73-89.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and P. M. Allen. 1999. Continental-scale simulation of the hydrologic balance. *J. American Water Resources Assoc.* 35(5): 1037-1052.

Bekele, E. G., and J. W. Nicklow. 2007. Multi-objective automatic calibration of SWAT using NSGA-II. *J. Hydrol.* 341(3-4): 165-176.

Beven, K. J. 2000. *Rainfall-Runoff Modelling: The Primer.* New York, N.Y.: John Wiley and Sons.

Cao, W., B. W. Bowden, T. Davie, and A. Fenemor. 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrol. Process.* 20(5): 1057-1073.

Chu, T. W., and A. Shirmohammadi. 2004. Evaluation of the SWAT model's hydrology component in the Piedmont physiographic region of Maryland. *Trans. ASAE* 47(4): 1057-1073.

Demarty, J., C. Ottlé, I. Braud, A. Olioso, J. P. Frangi, H. V. Gupta, and L. A. Bastidas. 2005. Constraining a physically based soil-vegetation-atmosphere transfer model with surface water content and thermal infrared brightness temperature measurements using a multiobjective approach. *Water Resources Res.* 41: W01011, doi: 10.1029/2004WR003695.

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

Eckhardt, K., N. Fohrer, and H. G. Frede. 2005. Automatic model calibration. *Hydrol. Process.* 19(3): 651-658.

Engeland, K., I. Braud, L. Gottschalk, and E. Leblois. 2006. Multi-objective regional modelling. *J. Hydrol.* 327(3-4): 339-351

Gassman, P. W., M. Reyes, C. H. Green, and J. G. Arnold. 2007. The Soil and Water Assessment Tool: Historical development, applications, and future directions. *Trans. ASABE* 50(4): 1212-1250.

Goldberg, D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning.* Reading, Mass.: Addison-Wesley.

Gollamudi, A., C. A. Madramootoo, and P. Enright. 2007. Water quality modeling of two agricultural fields in southern Quebec using SWAT. *Trans. ASABE* 50(3): 1973-1980.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurate measures of information. *Water Resources Res.* 34(4): 751-763.

Holland, J. 1975. *Adaptation in Natural and Artificial Systems.* Ann Arbor, Mich.: University of Michigan Press.

Khu, S. T., and H. Madsen. 2005. Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Res.* 41: W03004, doi: 10.1029/2004WR003041.

Legates, D. R., and G. J. McCabe. 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Res.* 35(1): 233-241.

Massey, F. J., Jr. 1951. The Kolmogorov-Smirnov test of goodness of fit. *J. American Stat. Assoc.* 46(253): 68-78.

Migliaccio, K. W., and I. Chaubey. 2007. Comment on W. Cao, B. W. Bowden, T. Davie, and A. Fenemor. 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrol. Process.* 21(4): 3326-3328.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models: Part I. A discussion of principles. *J. Hydrol.* 10(3): 282-290.

Neitsch, S. L., A. G. Arnold, J. R. Kiniry, J. R. Srinivasan, and J. R. Williams. 2005. *Soil and Water Assessment Tool User's Manual: Version 2005.* TR-192. College Station, Tex.: Texas Water Resources Institute.

Reca, J., and J. Martínez. 2006. Genetic algorithms for the design of looped irrigation water distribution networks. *Water Resources Res.* 42: W05416, doi: 10.1029/2005WR004383.

Santhi, C., J. G. Arnold, J. R. Williams, L. M. Hauck, and W. A. Dugas. 2001a. Application of a watershed model to evaluate management effects on point and nonpoint source pollution. *Trans. ASAE* 44(6): 1559-1570.

Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, and L. Hauck. 2001b. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. American Water Resources Assoc.* 37(5): 1169-1188.

Schaffer, J. D., R. A. Caruana, L. J. Eshelman, and R. Das. 1989. A study of control parameters affecting online performance of genetic algorithms for function optimization. In *Proc. 3rd Intl. Conf. on Genetic Algorithms*, 51-60. J. D. Schaffer, ed. San Mateo, Cal.: Morgan Kaufmann.

Seyfried, M. S., R. C. Harris, D. Marks, and B. Jacob. 2000. A geographic database for watershed research, Reynolds Creek Experimental Watershed, Idaho, USA. Tech. Bulletin NWRC-2000-3. Washington, D.C.: USDA-ARS.

Sloan, P. G., I. D. Morre, G. B. Coltharp, and J. D. Eigel. 1983. Modeling surface and subsurface stormflow on steeply sloping forested watersheds. Water Resources Institute Report 142. Lexington, Ky.: University of Kentucky.

Spruill, C. A., S. R. Workman, and J. L. Taraba. 2000. Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. *Trans. ASAE* 43(6): 1431-1439.

Srinivasan, R., J. G. Arnold, and C. A. Jones. 1998. Hydrologic modeling of the United States with the soil and water assessment tool. *Water Resources Development* 14(3): 315-325.

Tang, Y., P. Reed, and T. Wagener. 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration? *Hydrol. Earth Syst. Sci.* 10(2): 289-307.

Tolson, B. A., and C. A. Shoemaker. 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Res. 43: W01413, doi: 10.1029/2005WR004723.

USDA-SCS. 1972. Chapter 4-10, Section 4: Hydrology. In *National Engineering Handbook.* Washington, D.C.: USDA-SCS.

van Griensven, A., and W. Bauwens. 2003. Multiobjective autocalibration for semidistributed water quality models. *Water Resources Res.* 39(12): 1348, doi: 10.1029/2003WR002284.

van Griensven, A., and W. Bauwens. 2005. Application and evaluation of ESWAT on the Dender basin and Wister Lake basin. *Hydrol. Process.* 19(3): 827-838.

van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Di luzio, and R. Srinivasan. 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. *J. Hydrol.* 324: 10-23.

Van Liew, M. W., and J. Garbrecht. 2003. Hydrologic simulation of the Little Washita River experimental watershed using SWAT. *J. American Water Resources Assoc.* 39(2): 413-426.

Van Liew, M. W., J. G. Arnold, and D. D. Bosch, 2005. Problems and potential of autocalibrating a hydrologic model. *Trans. ASAE* 50(3): 1025-1040.

Van Liew, M. W., T. L. Veith, D. D. Bosch, and J. G. Arnold. 2007. Suitability of SWAT for the Conservation Effects Assessment Project: A comparison on USDA-ARS watersheds. *J. Hydrol. Eng.* 12(2): 173-189.

White, L. K., and I. Chaubey, 2005. Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *J. American Water Resources Assoc.* 41(5): 1077-1089.

Wolpert, D. H., and W. G. Macready. 1997. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* 1(1): 67-82.

Yapo, P. O., H. V. Gupta, and S. Sorooshian. 1998. Multi-objective global optimization for hydrologic models. *J. Hydrol.* 204: 83-97.

Zhang, X., R. Srinivasan, and F. Hao. 2007. Predicting hydrologic response to climate change in the Luohe River basin using the SWAT model. *Trans. ASABE* 50(3): 901-910.

Zhang, X., R. Srinivasan, B. Debele, and F. Hao. 2008a. Runoff simulation of the headwaters of the Yellow River using the SWAT model with three snowmelt algorithms. *J. American Water Resources Assoc.* 44(1): 48-61.

Zhang, X., R. Srinivasan, K. Zhao, and M. Van Liew. 2008b. Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrol. Process.* (in press).

Zitzler, E., and L. Thiele. 1999. Multiobjective evolutionary algorithms: A comparative case study and the Strength Pareto approach. *IEEE Trans. Evol. Comput.* 3(4): 257-271.

Zitzler, E., M. Laumanns, and L. Thiele. 2001. SPEA2: Improving the performance of the strength Pareto evolutionary algorithm. Tech. Report 103. Zurich, Switzerland: Swiss Federal Institute of Technology (ETH), Computer Engineering and Communication Networks Lab (TIK).