HYDROLOGICAL PROCESSES
Hydrol. Process. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.7152

Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model

Xuesong Zhang,[1,2]* Raghavan Srinivasan,[1] Kaiguang Zhao[1] and Mike Van Liew[3]

[1] *Spatial Sciences Laboratory, Department of Ecosystem Sciences and Management, Texas A&M University, College Station, TX 77843, USA*
[2] *1500 Research Parkway, Suite B223, Collgege Station, TX 77843, USA*
[3] *Montana Dept Environmental Quality, Helena, MT 59620, USA*

Abstract:

With the popularity of complex hydrologic models, the time taken to run these models is increasing substantially. Comparing and evaluating the efficacy of different optimization algorithms for calibrating computationally intensive hydrologic models is becoming a nontrivial issue. In this study, five global optimization algorithms (genetic algorithms, shuffled complex evolution, particle swarm optimization, differential evolution, and artificial immune system) were tested for automatic parameter calibration of a complex hydrologic model, Soil and Water Assessment Tool (SWAT), in four watersheds. The results show that genetic algorithms (GA) outperform the other four algorithms given model evaluation numbers larger than 2000, while particle swarm optimization (PSO) can obtain better parameter solutions than other algorithms given fewer number of model runs (less than 2000). Given limited computational time, the PSO algorithm is preferred, while GA should be chosen given plenty of computational resources. When applying GA and PSO for parameter optimization of SWAT, small population size should be chosen. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS global optimization algorithm; calibration; hydrologic model; SWAT; computational intensive

Received 6 February 2008; Accepted 20 August 2008

INTRODUCTION

Hydrologic models are more and more widely applied by hydrologists and resources managers as a tool to understand and manage natural and human activities that affect watershed systems. The successful application of a hydrologic model depends on how well the model is calibrated (Duan *et al*., 1992). Hydrologic models, even those physically-based models, often contain parameters that cannot be measured directly due to measurement limits and scale issues (Beven, 2000). These parameters need to be estimated through an inverse method by calibration so that observed and predicted output values are in agreement. Before the widespread availability of high speed computers, hydrologic practitioners utilized knowledge of the watershed and experience with the model to adjust the parameters through a manual trial and error procedure (Gupta *et al*., 1999). This approach to calibration is subjective and labour intensive. Automatic calibration methods, which are objective and relatively easy to implement with high speed computers, have become more popular in recent years (Vrugt *et al*., 2003). Global optimization algorithms can efficiently and effectively search optimum parameter solutions that can minimize (or maximize) objective functions which represent the agreement between observations and model simulations. They have been successively applied in the research field of automatic calibration of hydrologic methods. For example, Duan *et al*. (1992) developed the shuffled complex evolution algorithm (SCE-UA), which has been widely used in hydrologic modelling (Sorooshian *et al*., 1993) and proved to be consistent and efficient for searching global optimum parameter values of hydrologic models (Vrugt *et al*., 2003). Other optimization algorithms (i.e. genetic algorithms (GA), simulated annealing (SA), and Levenberg–Marquardt) are also popular methods for automatic calibration of parameters in hydrologic models.

With the popularity of sophisticated physically-based watershed models, the complexity of the calibration problem has increased substantially (Gupta *et al*., 1998). Although the speed and capacity of computers have increased multi-fold in the past several decades, the time consumed running hydrologic models (especially those complex, physically based, distributed hydrologic models) is still a concern for hydrologic practitioners. As to which of the available optimization methods can effectively and efficiently identify good parameter sets is a topic of considerable interest. Several studies have been conducted to evaluate the performance of different algorithms. For example, Cooper *et al*., (1997) evaluated SCE-UA, GA and SA methods for optimization of the Tank model; Kuczera (1997) compared four search algorithms, SCE-UA, GA, and multiple random start using either simplex or quasi-Newton local

* Correspondence to: Xuesong Zhang, 1500 Research Parkway, Suite B223, College Station, TX 77843, USA.
E-mail: xuesongzhang2004@gmail.com

searches for parameter optimization of catchment models; Chen *et al.* (2005) compared the performance of multi-start Powell and SCE-UA methods for calibrating the Tank model; Jha *et al.*, (2006) compared the traditional (Levenberg–Marquardt and Gauss–Newton) and nontraditional (GA) techniques for determining well parameters. The results obtained by the above comparison studies showed that the evolutionary algorithms (SCE-UA and GA) could provide equal or better performance than other methods (Cooper *et al.*, 1997; Kuczera, 1997; Chen *et al.*, 2005; Jha *et al.*, 2006). With the robustness for searching global optimum and ease of implementation, evolutionary algorithms have been widely used in hydrologic modeling. Besides the SCE-UA and GA, the particle swarm optimization (PSO) has also been used to optimize the arrangement of hydraulic devices in a pipeline system (Jung and Karney, 2006), and train artificial neural networks for river stage prediction (Chau, 2006). Other evolutionary algorithms, such as differential evaluation (DE) (Storn and Price, 1997) and artificial immune systems (AIS) (de Castro and Von Zuben, 2002a, 2002b), although rarely used in hydrologic model calibration, showed promising ability for global optimization of complex systems.

There are many physically-based watershed models that have been successfully applied in practical hydrologic modelling problems. However, since running these models is time intensive, it is nearly impossible to test the optimization algorithms for the complex models. In this study one complex distributed hydrologic model—Soil and Water Assessment Tool (SWAT) (Arnold *et al.*, 1998)—was selected to test the effectiveness and efficiency of different optimization algorithms. The SWAT model has been applied worldwide for hydrologic and water quality modelling. For example, the SWAT model has been incorporated into the US Environmental Protection Agency (USEPA) Better Assessment Science Integrating Point & Nonpoint Sources (BASINS) software package (Di Luzio *et al.*, 2004), and is being applied by the United States Department of Agriculture (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 research that includes SWAT (Gassman *et al.*, 2007). The objective of this paper was therefore to evaluate the efficacy of five evolutionary algorithms (SCE-UA, GA, PSO, DE, and AIS) for parameter optimization of SWAT. As the time and computational resources did not allow for a vast number of model runs with SWAT, the performance of the five optimization algorithms were only tested for a limited number of evaluations of the model. The results of this paper should provide hydrologic practitioners with valuable information for assessing the efficiency and effectiveness of automatic algorithms for calibrating SWAT.

MATERIALS AND METHODS

Study area description

The efficacy of optimization algorithms is dependent on the characteristics of the objective function response surface of the hydrologic model (Duan *et al.*, 1992), which is related to the watershed characteristics. In order to evaluate the general performance of different optimization algorithms, the SWAT model was applied to four watersheds with different climatic and hydrologic characteristics. The four watersheds included the Yellow River headwaters watershed (YRHW), Reynolds Creek Experimental Watershed (RCEW), Little River Experimental Watershed (LREW), and Mahantango Creek Experimental Watershed (MCEW). The locations of the four watersheds are shown in Figures 1 and 2. Among the four watersheds, the YRHW is located in China, and the other three watersheds are located in the USA. The three watersheds in USA are US Department of Agriculture Agricultural Research Service (USDA ARS) experimental watersheds, and have been used by Van Liew *et al.* (2007) for testing the suitability of SWAT for the CEAP. In YRHW and RCEW, the streamflows are significantly affected by snow fall and snow melt processes, while the LREW and MCEW are located in temperate and subtropical regions respectively, where snow related processes are not significant. The basic characteristics of the four test watersheds are described below.

YR headwaters watershed. The YRHW is an 114 345 km^2 mountainous river basin, which is located in the north-east part of the Tibetan plateau. This area is the important source of water generation for the Yellow River Basin (Liu, 2004). The average elevation is about

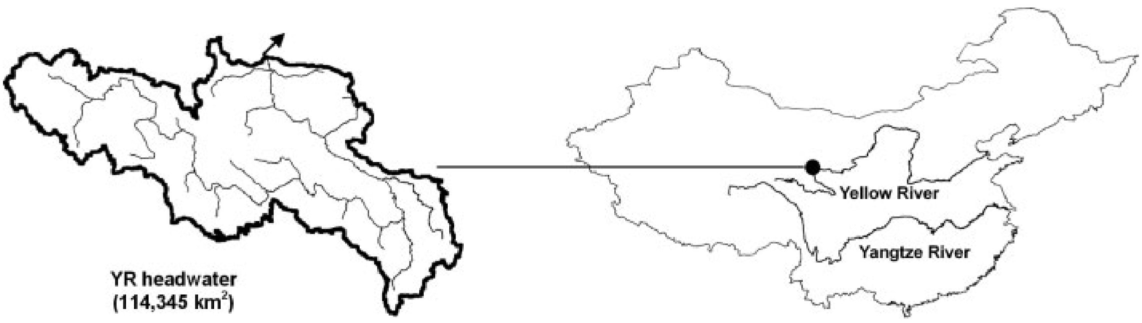

Figure 1. Location of the headwaters region of the Yellow River

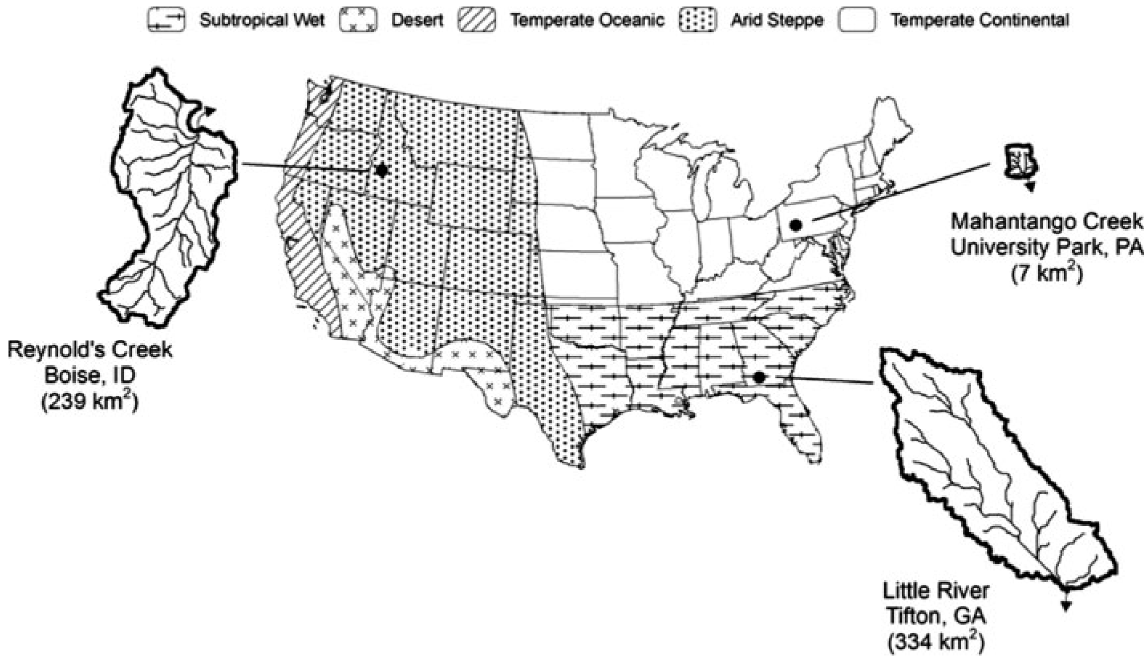

Figure 2. Locations of three USDA ARS experimental watersheds (modified from Van Liew *et al.*, 2007)

4217 m, and ranges between 2600 and 6266 m. The area slopes downward from west to east, ranging from a combined landform of low-mountains and wide valleys with lakes to smooth plateaus (Wang *et al.*, 2003). The headwaters area has a typical continental alpine cold and dry climate. The annual precipitation is around 600 mm and the average annual temperature for the YR headwater is near 0 °C. In winter the average temperature is below 0 °C at most of the weather stations, while in summer the average temperature is above 0 °C. This seasonal temperature variation makes snowmelt a significant process in this area (Zhang *et al.*, 2008). This watershed is characterized by gently sloping upland and river bed, and swamp and wetland. The major types of soils in this area are clay and loam with relatively low infiltration rate. The major land cover in the study area is grassland, which accounts for approximately 90% of the total area. Other land use/land cover (forest land, rangeland, agriculture land, and bare area) account for the remaining 10% of the area.

Mahantango Creek Experimental Watershed. The MCEW is a tributary of the Susquehanna River in Central Pennsylvania. The MCEW is typical of upland agricultural watersheds within the nonglaciated, folded and faulted, Appalachian Valley and Ridge Physiographic Province (Veith *et al.*, 2005). Climate in the region is temperate and humid, with a long-term average annual precipitation of 1100 mm. The watershed is characterized by shallow, fragipan soils in near-stream areas, and deep, well-drained soils in the uplands (Van Liew *et al.*, 2007). Land use types consist of pasture (38%), forest (34%), mixed croplands (26%), and farmsteads (2%).

Little River Experimental Watershed. The LREW is the upper 334 km^2 of the USDA-ARS and cooperators (Sheridan, 1997). The LREW is located in the Tifton Upland physiographic region, which is characterized by intensive agriculture in relatively small fields in upland areas and riparian forests along stream channels. The region has low topographic relief and is characterized by broad, flat alluvial floodplains, river terraces, and gently sloping uplands (Sheridan, 1997).

Climate in this region is characterized as humid subtropical with an average annual precipitation of about 1167 mm based on data collected by USDA ARS from 1971 to 2000. Soils on the watershed are predominantly sands and sandy loams with high infiltration rates. Since surface soils are underlain by shallow, relatively impermeable subsurface horizons, deep seepage and recharge to regional ground water systems are impeded (Sheridan, 1997). Land use types include forest (65%), cropland (30%), rangeland and pasture (2%), wetland (2%), and miscellaneous (1%).

Reynolds Creek Experimental Watershed. The RCEW, with drainage area of 239 km^2, is located about 80 km south-west 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 watershed. Perennial streamflow is generated at the highest elevations in the southern part of Reynolds Creek 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

DOI: 10.1002/hyp

soils. Land cover on Reynolds Creek consists of rangeland and forest communities of sagebrush, greasewood, aspen, and conifers (94%) and irrigated cropland (6%).

SWAT model description

SWAT is a continuous-time, long-term, distributed-parameter model (Arnold *et al*., 1998). SWAT subdivides a watershed into subbasins connected by a stream network, and further delineates hydrologic response units (HRUs) consisting of unique combinations of land cover and soils in each subbasin. It is assumed that there is no interaction between HRUs, that is, the HRUs are non-spatially distributed. HRU delineation can minimize computational costs of simulations by lumping similar soil and land use areas into a single unit (Neitsch *et al*., 2002a). SWAT is able to simulate surface and subsurface flow, sediment generation and deposit, and nutrient fate and movement through the watershed system. For this study, only the components of SWAT concerned with runoff simulation were considered. The hydrologic routines within SWAT account for snow fall and melt, vadose zone processes (i.e. infiltration, evaporation, plant uptake, lateral flows, and percolation), and groundwater flows. Surface runoff volume is estimated using a modified version of the Soil Conservation Service (SCS) Curve Number (CN) method (Kannan *et al*., 2008). 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). The Muskingum method is used for channel flood routing. Outflow from a channel is adjusted for transmission losses, evaporation, diversions, and return flow.

Global optimization algorithms

Five global optimization algorithms (GA, SCE, PSO, AIS, and DE) were investigated in this study. The parameter solution is referred to as 'Chromosome' in GA, 'point' in SCE, 'particle' in PSO, 'Antibody' in AIS, and 'Individual' in DE. All the five algorithms are population based; the Latin Hypercube algorithm is used to initialize the first population of parameter solutions in this study.

Genetic algorithms. Genetic algorithms are stochastic search procedures inspired by evolutionary biology of natural selection and genetics (Holland, 1975; Goldberg, 1989), such as inheritance, mutation, selection, and crossover. The implementation of GA starts with initializing a population of candidate solutions (called chromosomes) which are randomly sampled from the feasible parameter space. In each generation, the individual chromosomes are selected through a fitness-based process, where the more fit chromosomes in the population are preferentially selected to reproduce new promising offspring. Next, a new generation population of chromosomes is generated from these selected ones using crossover and mutation operations. The crossover operator chooses 'parent' solutions and exchange important building blocks of the two parent chromosomes to generate new 'offspring' solutions. The 'offspring' solutions are then randomly muted to increase the diversity of the new population. Through a steady-state-delete-worst plan (Reca and Martinez, 2006), the fitter chromosomes among the old and new population are input into the next generation for evolution. This generational evolution of the parameter solutions is repeated until a maximum number of model evaluations are reached. 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), including hydrologic modelling. For example, Kuo and Liu (2003) applied GAs for optimizing a model for irrigation planning and management; Chang *et al*. (2005) showed that the GA provided an adequate, effective and robust way for searching the reservoir operating rule curves. Srivastava (2002) and Arabi *et al*. (2006) used GAs for optimizing allocation watershed management practices. The parameters that control the GA's evolution were determined according to Schaffer *et al*. (1989) and Reca and Martinez (2006).

Shuffled complex evolution (SCE)

The SCE algorithm developed by Duan *et al*. (1992) merges the strengths of the downhill simplex procedure (Nelder and Mead, 1965) with the concepts of controlled random search, competitive evolution (Holland, 1975), and complex shuffling. In a first step of implementation of SCE, an initial population of parameter solutions is randomly sampled for 'p' parameters to be optimized. The population is partitioned into several communities, each consisting '$2p + 1$' points. Each community is made to evolve based on a statistical 'reproduction process' that uses the simplex method, an algorithm that evaluates the objective function in a systematic way with regard to the progress of the search in previous iterations (Nelder and Mead, 1965). At periodic stages in the evolution, the entire population is shuffled and parameter solutions are reassigned to communities to ensure information sharing. As the search progresses, the entire population tends to converge towards the neighbourhood of global optimization. SCE searches the entire parameter space and finds the global optimum efficiently and effectively (Sorooshian *et al*., 1993). SCE has been successfully used for calibration of SWAT (van Griensven and Bauwens, 2003; Eckhardt *et al*., 2005; Van Liew *et al*., 2005; van Griensven and Bauwens 2005; Zhang *et al*., 2007). In order to apply SCE efficiently, the control parameters, except for population size, were selected according to Duan *et al*. (1994). For further information on SCE, refer to Duan *et al*. (1992).

Particle swarm optimization

Particle swarm optimization is a population based stochastic optimization technique inspired by social behaviour of bird flocking or fish schooling (Kennedy and Eberhart, 2001). During the optimization process, in order to find the global optimum, each particle in the population adjusts its 'flying' according to its own flying experience and its companions' flying experience (Eberhart and Shi, 1998). The basic PSO algorithm consists of three steps: (1) generate the positions of particles (coordinate in the parameter space) and their velocities ('flying' direction and speed); (2) update the velocity of each particle using the information from the best solution it has achieved so far (personal best) and another particle with the best fitness value that has been obtained so far by all the particles in the population (global best); (3) finally, the new position of each particle is calculated by adding the updated velocity to the current position. PSO has been successfully applied to optimize artificial neural networks for river stage prediction (Chau, 2006) and parameter estimation of hydrologic models (Gill *et al*., 2006). The control parameters for implementing PSO were selected based on previous studies (Shi and Eberhart, 1998; Parsopoulos and Vrahatis, 2002).

Differential evolution. The differential evolution algorithm is a simple and powerful evolutionary algorithm developed by Storn and Price (1997) for global optimization. The basic procedure of DE is very similar to that of GA. DE also uses three operators (i.e. mutation, crossover, and selection) to evolve the population of parameter solutions towards the global optimum. But the actual implementation of the three operators is different from that of GA. First, the mutation operator is implemented. For each individual (target vector) in the population, three other individuals in the same population are randomly selected, and the weighted difference of two of the vectors is added to the third to generate a new individual (called the donor vector). Second, the crossover operator is then used to generate a trial vector from the elements of the target vector and the donor vector. Finally, the target vector and trial vector are compared and the fitter one is selected into the next generation. There are several variants of DE (Storn and Price, 1997; Krishna, 2007). One variant of DE, noted as DE/rand/1/bin according to Storn and Price (1997), was applied in this study. This variant of DE has been most often used in practice (Brest *et al*., 2006; Krishna, 2007). The control parameters for DE were set following suggestions from Storn and Price (1997) and Brest *et al.* (2006).

Artificial immune system. Artificial immune system is a type of optimization algorithm inspired by the principles and processes of the vertebrate immune system. In this study, the CLONALG (de Castro and Von Zuben, 2002a), a classical AIS algorithm was introduced and applied for parameter optimization of SWAT. In CLONALG, the antibodies with higher objective function value and lower similarity (expressed as a parameter solution's Euclidean distance to other parameter solutions in the population) are selected to reproduce the next generation of candidate antibodies using Gaussian mutations, while the parameter solutions with lower objective function value and higher similarity are replaced with new randomly generated parameter solutions. Although AIS has seldom been used for optimizing hydrologic models, it has been successfully applied for optimizing complex systems, like the radial basis function (de Castro and Von Zuben, 2002b), neural networks (Byrski and Kisiel-Dorohinicki, 2005), economic dispatch in power systems (Rahman *et al*., 2006), and several constrained global optimization problems (Cruz-Cortés *et al*., 2005). The control parameters were selected following de Castro and Von Zuben (2002a), which provides a detailed discussion on the application of CLONALG algorithms for parameter optimization.

Optimization test cases design

Two important factors that affect the complexity of the optimization problem are the parameter dimension that needs to be adjusted and the parameter ranges. van Griensven *et al.* (2006) conducted detailed global sensitivity analysis of the parameters in SWAT, and results showed that ten parameters are sensitive to the hydrologic simulation of SWAT. Van Liew *et al*. (2007) tested the suitability of SWAT for the CEAP in USDA agricultural research service watersheds. In their study 16 parameters, which include the ten parameters identified by van Griensven *et al*. (2006), were adjusted to calibrate the SWAT model for hydrologic simulation. The 16 parameters identified by Van Liew *et al*. (2007) were applied in this study. A general description of the 16 parameters is shown in Table I. The parameter ranges were determined according to van Griensven *et al*. (2006) and Neitsch *et al*. (2002b). Among these 16 parameters, nine govern surface and subsurface water response in SWAT, and the other seven parameters govern basin response.

The performance of the five optimization algorithms is related with the control parameters. For most of the control parameters, previous literature provides suggestions on how to choose appropriate settings. Among the control parameters, population size is an important factor that determines the performance of different algorithms. In this study, most of the control parameters of the five optimization algorithms were set according to recommendations from previous studies, while the effect of population size on the performance of different algorithms was further examined with one relatively small population size and one relatively large population size for each optimization algorithm. There are no common criteria for evaluating whether a population size is large or small for different algorithms. Small and large population sizes are different for the five algorithms, and were chosen according to population sizes that have been tested in previous empirical studies that applied these optimization techniques. The small population sizes were 66 (two complexes), 50, 30, 10, and 50 for SCE, GA, PSO, AIS and DE, respectively. The large population sizes were 165 (five complexes), 200, 100, 50, and 160 for SCE, GA, PSO, AIS and DE, respectively. There are two optimization cases that were defined for each watershed: (1) small population size scenario; and (2) large population size

Table I. Parameters for calibration in SWAT model

Parameter		Description	Range
Parameters governing surface water response			
1	*CN2*	Curve number	±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*	Ground water re-evaporation coefficient	0·02–0·2
5	*REVAPMN*	Threshold depth of water in the shallow aquifer for re-evaporation 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 21 June (mm $H_2O\,°C^{-1}day^{-1}$)	0–10
16	*SMFMN*	Minimum snowmelt factor for 21 Dec (mm $H_2O\,°C^{-1}$-day^{-1})	0–10

scenario. Hence, there were a total of eight optimization cases for each optimization algorithm that were applied in this study. The definition of the optimization case was denoted using the combination of watershed name and population size, i.e. 'watershed name + population size'. For example, 'RCEW + Small' denotes that the optimization algorithms were tested on the RCEW with small population size. In this study, the SWAT model was set up for daily flow simulation at the outlets at different watersheds. The calibration periods consists of 10 years (1976–1985) in the YR headwater watershed, 6 years (1995–2000) in MCEW, 4 years (1995–1998) in LREW, and 7 years (1966–1972) in RCEW.

Evaluating performance of different algorithms

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

$$E_{ns} = 1{\cdot}0 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2} \qquad (1)$$

where P is the model simulated value, O is the observed data, the overbar is the mean for the entire time period of the evaluation, and $i = 1, 2, \ldots, N$, 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.

The five algorithms involve random sampling of the parameter values, so the results obtained by one trial are stochastic and cannot be used to accurately evaluate the algorithm's performance. The average behaviour of multiple trials of each algorithm was used to compare the performance of different algorithms, which is a popular performance comparison method reported in the literature (Ali *et al.*, 2005). Ideally, the optimization algorithm that can find the best E_{ns} value with the lowest number of model runs is preferred. In Duan *et al.*'s (1994) work, the number of successes (NS) and the average number of function evaluations (AFE) of the successful runs were used to evaluate the efficacy of SCE algorithms. As our test is to evaluate the efficacy of the optimization algorithms with a limited number of runs of the computationally intensive model, two other similar coefficients were applied in this study. The average E_{ns} values are used to evaluate the ability of each algorithm to find good objective values at different numbers of model evaluations. A new variable, RE_{ns}, which was defined as the ratio between the E_{ns} value obtained for different numbers of model evaluations and the best E_{ns} value obtained at the maximum number of model evaluations for each algorithm, was adopted here. In this study, it was assumed that a RE_{ns} value of 0·99 represents a close approximation of the objective values obtained after the maximum number of model evaluations.

On a computer with Pentium IV 3 GHZ and 1GB RAM, the time consumed by one SWAT model evaluation was 30 s for YRHW, 18 s for MCEW, 56 s for LREW, and 1 min and 8 s for RCEW. As time and computer resources are limited, it was not possible to run the SWAT model for a very long simulation period or for an unlimited number of model evaluations. The five algorithms were compared based on the average performance of 10 trials within a limited and affordable number of model evaluations. According to previous

DOI: 10.1002/hyp

calibration studies of SWAT, usually less than 10 000 model evaluations were implemented (van Griensven and Bauwens, 2003; Tolson and Shoemaker, 2007). Considering the time and computer resources available, the maximum number of model evaluations was limited to 10 000 for the four test watersheds. The time consumed by one trial was 84 h for YRHW, 50 h for MCEW, 155 h and 190 h for LREW and RCEW, respectively.

RESULTS AND DISCUSSION

The curves of average objective function values against model evaluation numbers obtained by different algorithms with large and small population size are shown in Figure 3 for the four test watersheds. The average objective values and relative performance ranks of different algorithms at different model evaluation numbers are listed in Table II. Based on Figure 3 and Table II, the performances of the different algorithms in the four test watersheds were analysed and results are presented in the following sections. For most cases, AIS performed worst among the five algorithms. The analysis was focused mainly on SCE, GA, PSO and DE.

Performance of the different algorithms in YRHW

The selected optimization algorithms exhibited various performance levels at different model evaluation numbers (Figure 3a and Table II). From Table II, after 10 000

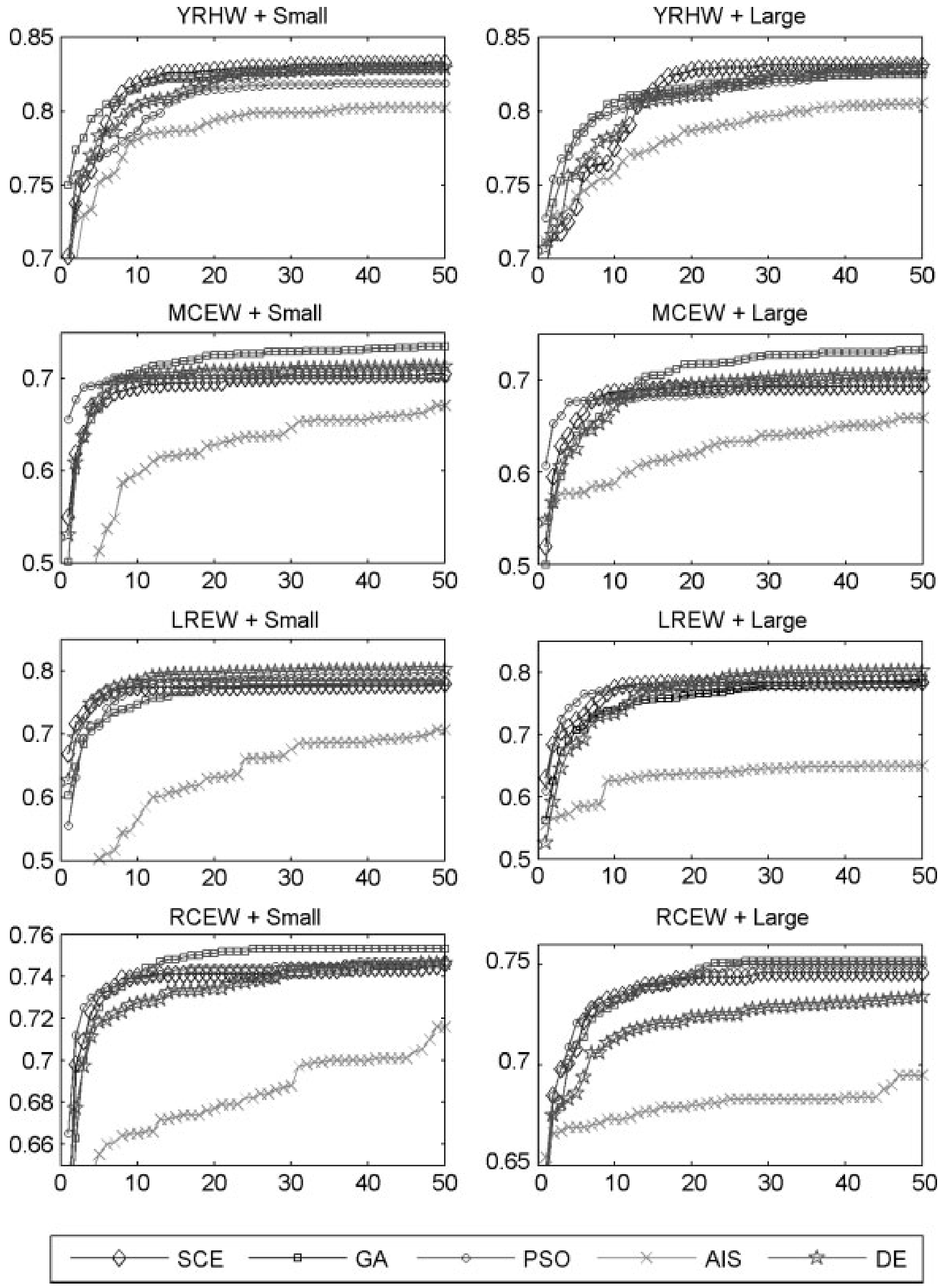

Figure 3. Performances of the different optimization algorithms versus number of evaluations in the four test watersheds (x axis represents the number of model evaluations × 200, y axis represents E_{ns} value)

Table II. E_{ns} values obtained by different optimization algorithms at different number of model runs in the four test watersheds

		YRHW					MCEW				
		SCE	GA	PSO	AIS	DE	SCE	GA	PSO	AIS	DE
500	Small	0·751	0·782	0·759	0·73	0·758	0·64	0·638	0·691	0·474	0·637
	Large	0·718	0·753	0·768	0·73	0·726	0·628	0·595	0·662	0·576	0·606
1000	Small	0·772	0·799	0·769	0·752	0·783	0·673	0·667	0·694	0·513	0·676
	Large	0·735	0·785	0·78	0·743	0·757	0·656	0·639	0·677	0·577	0·625
2000	Small	0·819	0·815	0·787	0·781	0·803	0·691	0·708	0·7	0·596	0·701
	Large	0·775	0·806	0·801	0·758	0·783	0·687	0·681	0·68	0·588	0·671
3000	Small	0·826	0·822	0·808	0·787	0·81	0·695	0·718	0·7	0·616	0·704
	Large	0·814	0·813	0·811	0·775	0·807	0·69	0·705	0·682	0·612	0·689
4000	Small	0·828	0·824	0·815	0·794	0·819	0·696	0·726	0·7	0·628	0·708
	Large	0·828	0·816	0·814	0·787	0·811	0·692	0·717	0·684	0·619	0·695
5000	Small	0·83	0·826	0·817	0·799	0·825	0·701	0·727	0·7	0·637	0·709
	Large	0·83	0·82	0·816	0·791	0·818	0·693	0·719	0·689	0·633	0·698
10 000	Small	0·833	0·829	0·819	0·803	0·829	0·704	0·735	0·7	0·671	0·713
	Large	0·831	0·825	0·827	0·806	0·83	0·696	0·733	0·703	0·659	0·707

		LREW					RCEW				
		SCE	GA	PSO	AIS	DE	SCE	GA	PSO	AIS	DE
500	Small	0·731	0·685	0·694	0·487	0·715	0·709	0·7	0·725	0·64	0·697
	Large	0·704	0·673	0·725	0·569	0·645	0·698	0·679	0·698	0·667	0·681
1000	Small	0·757	0·717	0·713	0·503	0·759	0·73	0·725	0·732	0·655	0·718
	Large	0·723	0·707	0·752	0·583	0·685	0·708	0·71	0·721	0·669	0·686
2000	Small	0·774	0·747	0·774	0·565	0·783	0·74	0·74	0·739	0·665	0·727
	Large	0·775	0·739	0·771	0·626	0·733	0·734	0·73	0·733	0·673	0·713
3000	Small	0·774	0·766	0·783	0·609	0·794	0·741	0·748	0·743	0·673	0·733
	Large	0·782	0·757	0·776	0·634	0·768	0·74	0·739	0·741	0·677	0·72
4000	Small	0·774	0·775	0·783	0·632	0·795	0·741	0·751	0·744	0·677	0·734
	Large	0·783	0·764	0·778	0·637	0·782	0·744	0·745	0·744	0·68	0·724
5000	Small	0·775	0·779	0·784	0·662	0·797	0·741	0·753	0·744	0·682	0·738
	Large	0·783	0·771	0·779	0·639	0·793	0·744	0·751	0·75	0·683	0·725
10 000	Small	0·78	0·782	0·784	0·707	0·802	0·746	0·753	0·747	0·716	0·746
	Large	0·784	0·784	0·781	0·65	0·802	0·746	0·752	0·751	0·695	0·734

model runs, the best average E_{ns} values obtained by the selected algorithms were 0·833 (SCE with large population), 0·829 (GA with small population), 0·827 (PSO with large population), and 0·830 (DE with small population). Although the four algorithms obtained close objective values with a large number of model runs, they exhibited very different performance levels with small numbers of model evaluations (Figure 3a and Table II). One algorithm may be preferred for a small number of model evaluations while another algorithm may be preferred for a large number of model runs. For example, GA found better objective values with a small number of model runs (500 and 1000), while SCE obtained better results given a large number of model evaluations (more than 2000). The differences between the best average E_{ns} values obtained by the different algorithms at a small number of model runs are larger than those obtained with a larger number of model evaluations. For example, the maximum difference between the best final average E_{ns} values obtained by SCE, GA, PSO and DE was 0·006, while this difference was 0·031 given 500 model runs.

It was also found that the objective values change relatively quickly for the initial 1000 model evaluations, and then changed relatively slowly after that. The RE_{ns} values at small numbers of model evaluations represent the capacity of each algorithm to approach the objective values that can be obtained by each scheme with 10 000 model evaluations. In the YRHW, all the algorithms reached RE_{ns} values larger than 0·86 and 0·88 for 500 and 1000 model runs, respectively. Based on the RE_{ns} values obtained with limited model runs, the objective values obtained by each algorithm with a large number of model evaluations can be roughly estimated. In general, each scheme needs less than 5000 model evaluations to reach a RE_{ns} value of 0·99 and to thus approximate the best objective value that can be obtained by each algorithm with 10 000 model evaluations.

The effect of population size on average E_{ns} values obtained by the optimization algorithms was relatively stronger for the initial 5000 model evaluations than for the model runs after 5000. For example, the difference between average objective values obtained by SCE with small and large population sizes reached 0·044 at 2000 model evaluations, while these differences were within 0·008 for all optimization algorithms after 5000 model evaluations.

DOI: 10.1002/hyp

Performances of different algorithms in MCEW

Figure 3b shows that the performance of the GA is significantly better than the other algorithms after the initial 3000 model evaluations. The GA exhibited an average E_{ns} value larger than 0·72 with 5000 model evaluations, while the other algorithms did not reach E_{ns} values of 0·72 even after 10 000 model evaluations. Within 1000 model runs, PSO performed considerably better than other algorithms (Table II). The maximum difference between the final average E_{ns} values obtained by SCE, GA, PSO and DE was 0·032 after 10 000 model evaluations. This shows that different optimization algorithms can obtain substantially different objective values even after a large number of model runs. In the MCEW, RE_{ns} values reached 0·99 for all optimization algorithms within 5500 model evaluations, except for the AIS (Table II). With 500 model runs, SCE and PSO obtained RE_{ns} values larger than 0·9, and GA and DE obtained RE_{ns} values larger than 0·8. For 1000 model runs, SCE and PSO obtained RE_{ns} values larger than 0·95, and GA and DE obtained RE_{ns} values larger than 0·85.

For a smaller number of model runs, the difference between the objective values obtained by the each algorithm with large or small population sizes was relatively larger than that for large numbers of model runs. For instance, the differences between the final average objective values obtained by SCE, GA, PSO and DE with small and large population sizes were less than 0·008, while this difference obtained by GA with small and large population sizes reached 0·043 at 500 model runs.

Performance of different algorithms in LREW

With 10 000 model evaluations, the final average E_{ns} values obtained by the selected algorithms were 0·784 (SCE with large population), 0·784 (GA with small population), 0·784 (PSO with small population), and 0·802 (DE with small population), respectively (Table II). SCE obtained better objective values than other algorithms with 500 model runs, and DE obtained better objective values after 1000 model runs. To reach RE_{ns} values of 0·99, SCE and PSO need 2000 model evaluations, and GA and DE need 5000 model evaluations. With 500 model runs, SCE and PSO obtained RE_{ns} values larger than 0·89, and GA and DE obtained RE_{ns} values larger than 0·8. For 1000 model runs, SCE, GA, and PSO obtained RE_{ns} values larger than 0·9, and DE obtained RE_{ns} values larger than 0·85.

The differences between final average E_{ns} values obtained by different algorithms with small or large population sizes were within 0·004, which shows that the E_{ns} values are not sensitive to population size after a large number of model runs in the LREW. But this difference was 0·06 for DE with small and large populations at 500 model runs.

Performance of different algorithms in RCEW

After 10 000 model evaluations, the GA with small population size exhibited the best objective function value (0·753), followed by PSO with large population size (0·751), SCE (0·746), and DE with small population size (0·746), respectively (Table II). For the initial 500 or 1000 model runs, PSO obtained better results than other algorithms. The maximum difference between the best final average E_{ns} values obtained by SCE, GA, PSO and DE was 0·007, and this difference is 0·027 at 500 model runs. In the RCEW, to reach RE_{ns} values of 0·99, 2600 model evaluations were required for the SCE and PSO, and 4500 for the GA and DE. All the algorithms reached RE_{ns} values larger than 0·90 for 500 model runs, and 0·93 for 1000 model evaluations.

Similar to previous results obtained in the three test watersheds, the differences between final average E_{ns} values obtained by different algorithms with small or large population sizes were relatively small (within 0·012), and these differences were relatively large for the initial model runs.

Discussion

The results obtained in previous sections show that no one optimization algorithm can consistently perform better than the other algorithms for the selected test watersheds. To some extent, this indicates the complexity and difficulty of parameter optimization for the SWAT model. Although all the test cases used SWAT as the model for parameter calibration, it appears as though the properties of the four optimization cases are different from each other, which leads to evidently different performances of the selected algorithms. The overall performances of the five optimization algorithms, and the influence of model evaluation number and population size, are discussed in the following sections.

Using the best final average E_{ns} values obtained by each of the selected algorithms as the indicator of performance, the performance ranks of the algorithms in the four test watersheds are shown in Table III. The GA performed best for the RCEW and MCEW, DE performed best for the LREW, and SCE performed best for the YRHW. Using the cumulative rank as the indicator of the comprehensive performance in the four test watersheds (Figure 4), GA performed the best in terms of finding good objective values with a large number of model runs, followed by DE, SCE, PSO and AIS.

For computationally intensive models, the number of model evaluations needed to obtain acceptable objective values was an important factor for selecting the optimization algorithm. The SWAT model of detailed characterization of a large river basin can take hours or days to implement once. In this case, it is difficult to run a large number of model runs. The algorithms that can find better objective values within a limited number of model evaluations (e.g. less than 1000) are preferred. The performance ranks of the different algorithms evaluated with the best average E_{ns} values obtained at different numbers of model evaluations are listed in Table III. As the AIS cannot obtain results comparable with the other optimization algorithms, it is not discussed here. It is apparent

DOI: 10.1002/hyp

Table III. Performance ranks of different optimization algorithms for different numbers of model evaluations in the four test watersheds

		SCE	GA	PSO	DE			SCE	GA	PSO	DE
YRHW	500	4	1	2	3	LREW	500	1	4	2	3
	1000	4	1	3	2		1000	2	3	4	1
	2000	1	2	4	3		2000	2	4	3	1
	3000	1	2	3	4		3000	3	4	2	1
	4000	1	2	4	3		4000	2	4	2	1
	5000	1	2	4	3		5000	3	4	2	1
	10 000	1	3	4	2		10 000	2	2	2	1
MCEW	500	2	3	1	4	RCEW	500	2	3	1	4
	1000	3	4	1	2		1000	2	3	1	4
	2000	4	1	3	2		2000	1	1	3	4
	3000	4	1	3	2		3000	3	1	2	4
	4000	4	1	3	2		4000	2	1	2	4
	5000	3	1	4	2		5000	3	1	2	4
	10 000	3	1	4	2		10 000	3	1	2	4

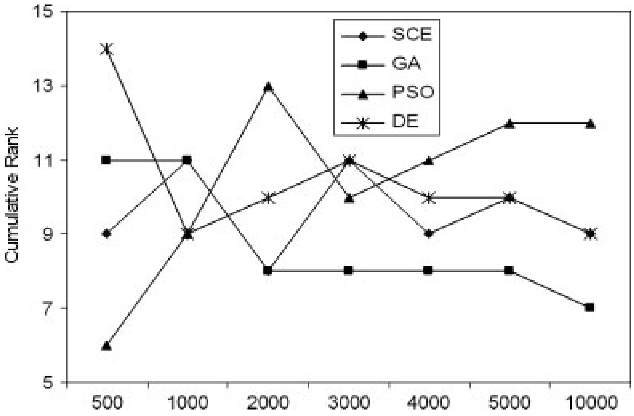

Figure 4. Cumulative performance ranks of four optimization algorithms for different numbers of model evaluations in the four test watersheds

that the performance of the selected optimization algorithms changes significantly with the number of model evaluation and watershed characteristics. The cumulative performance ranks of the four optimization algorithms for different numbers of model evaluations in the four test watersheds are shown in Figure 4. It is seen that PSO performed best with 500 model runs, PSO and DE performed best with 1000 model runs, SCE and GA performed best with 2000 model runs, and GA performed best with more than 2000 model runs. PSO is the preferred choice for less than 1000 model evaluations. For most optimization cases, PSO can obtain RE_{ns} values larger than 90% with 500 or 1000 model runs, which is a fairly good approximation of the best values obtained by PSO after 10 000 model runs. In general, results show that SCE and PSO converge faster than GA and DE. The numbers of model evaluations required by various optimization algorithms to obtain a RE_{ns} value of 0·99 are summarized with a conservative consideration of the convergence numbers of the four optimization techniques with small and large population sizes in the four test watersheds. Overall, SCE, GA, PSO and DE need no more than 3200, 5400, 4400, and 4800 model runs, respectively.

It should be noted that the population size could influence the performance of the various algorithms, as the difference between the E_{ns} values obtained by each optimization technique when using small or large population size at a small number of model evaluations is less than that at a small number of model evaluations. The selection of population size is based mainly on the performances of different algorithms at a fewer number of model evaluations (less than 5000). From Table II, for most cases, small population size provided better objective function values than large population size for SCE, GA, DE, and PSO with fewer model evaluations. In the future application of these algorithms for optimizing SWAT, small population size is preferred.

The results discussed above, to some extent, agree with the popular no free lunch (NFL) theorem 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). Meaning that GA performed better than the other algorithms in terms of finding good average E_{ns} values, on the other hand, PSO need less model runs to find acceptable objective values than other algorithms. Although AIS performed worst in terms of both finding best E_{ns} values and efficient convergence to good objective values, it can search multiple local optimums simultaneously. This could be an advantage for hydrologic model calibration and further analysis in the future should be considered. Similar results were also obtained by other numerical evaluations of different global optimization algorithms. For instance, based on the comparison of five stochastic global optimization algorithms, Ali *et al.* (2005) concluded 'one algorithm may be preferred if a small number of function evaluations is allowed but a different algorithm may be favored if a large number of function evaluations is permitted'. Although the GA algorithm exhibits the best comprehensive rank in terms of finding good average E_{ns} values, it is not possible to infer that this algorithm will always provide the best performance on parameter calibration of the SWAT model. The GA can be the first choice when modellers are interested in finding global optimum. The PSO may be a better

DOI: 10.1002/hyp

choice when modellers are interested in obtaining acceptable good calibration results within limited computation budget.

CONCLUSIONS

Efficient and effective algorithms for optimization of computationally intensive hydrologic models like SWAT are becoming increasingly more important because of limited time and computational resources. The purpose of this study was to evaluate the performance of five optimization algorithms for parameter calibration of SWAT within the context of limited model evaluations. In this study, five global optimization algorithms (SCE, GA, PSO, AIS and DE) were tested for parameter calibration of SWAT in four watersheds. For future application of SWAT across the USA and other watersheds worldwide, several empirical recommendations on selecting optimization algorithms for SWAT are provided based on the overall performances of the optimization algorithms in the four test watersheds. The GA outperforms the other four algorithms given model evaluation numbers larger than 2000, while PSO can obtain better parameter solutions than other algorithms given fewer model runs (less than 2000). Given limited computational time, the PSO algorithm is preferred, while GA should be chosen when sufficient computational resources are available. If GA is chosen to optimize SWAT with a large number of model evaluations, the performances of GA is not improved beyond 5400 model runs. When applying PSO and GA to calibrate SWAT parameters, a small population size is preferred. Also, different optimization algorithms exhibited various preferred properties and incorporating the strength of different algorithms into one powerful algorithm will be investigated in future studies.

ACKNOWLEDGEMENTS

The authors would like to thank Alan Verser for his help in preparing the watershed location figures. We appreciate the valuable comments and suggestions provided by the anonymous reviewers and associate editor, which greatly enhanced the quality of the manuscript.

REFERENCES

Ali MM, Khompatraporn C, Zabinsky ZB. 2005. A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. *Journal of Global Optimization* **31**: 635–672.

Arabi M, Govindaraju RS, Hantush MM. 2006. Cost-effective allocation of watershed management practices using a genetic algorithm. *Water Resources Research* **42**: W10429, DOI: 10·1029/2006WR004931. Available Online: http://www.agu.org/pubs/crossref/2006/2006WR004931.shtml.

Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large area hydrologic modeling and assessment part I: model development. *Journal of American Water Resource Association* **34**(1): 73–89.

Beven KJ. 2000. *Rainfall-Runoff Modeling: the Primer*. John Wiley & Sons: New York.

Brest J, Greiner S, Bošković B, Mernik M, Žumer V. 2006. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computing* **10**(6): 646–656.

Byrski A, Kisiel-Dorohinicki M. 2005. Immune-based optimization of predicting neural networks. *Computational Science—ICCS* **3516**: 703–710.

Chang FJ, Chen L, Chang LC. 2005. Optimizing the reservoir operating rule curves by genetic algorithms. *Hydrological Processes* **19**: 2277–2289.

Chau KW. 2006. Particle Swarm Optimization training algorithm for ANNs in stage prediction of Shing Mun River. *Journal of Hydrology* **329**: 363–367.

Chen R, Pi L, Hsieh C. 2005. Application of parameter optimization method for calibrating tank model. *Journal of American Water Resources Association* **41**(2): 389–402.

Cooper VA, Nguyen VTV, Nicell JA. 1997. Evaluation of global optimization methods for conceptual rainfall-runoff model calibration. *Water Science Technology* **36**(5): 53–60.

Cruz-Cortés N, Trejo-Pérez D, Coello Coello CA. 2005. Handling constraints in global optimization using an artificial immune system. *Lecture Notes in Computer Science* **3627**: 234–247.

de Castro LN, Von Zuben FJ. 2002a. Learning and optimization using the clonal selection principle. *IEEE Transaction on Evolutionary Computing* **6**(3): 239–251.

de Castro LN, Von Zuben FJ. 2002b. Automatic determination of radial basis functions: an immunity-based approach. *International Journal of Neurological Systems* **11**(6): 523–535.

Di Luzio M, Srinivasan R, Arnold JG. 2004. A GIS-coupled hydrological model system for the watershed assessment of agricultural nonpoint and point sources of pollution. *Transactions GIS* **8**(1): 113–136.

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

Duan Q, Sorooshian S, Gupta VK. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of Hydrology* **158**: 265–284.

Eberhart RC, Shi Y. 1998. Comparison between genetic algorithms and particle swarm optimization. In *Proceedings of the Seventh Annual Conference on Evolutionary Programming*. Springer: New York; 611–616.

Eckhardt K, Fohrer N, Frede HG. 2005. Automatic model calibration *Hydrological Processes* **19**: 651–658.

Gassman PW, Reyes M, Green CH, Arnold JG. 2007. The soil and water assessment tool: historical development, applications, and future directions. *Transactions ASABE* **50**(4): 1212–1250.

Gill MK, Kaheil YH, Khalil A, McKee M, Bastidas L. 2006. Multiobjective particle swarm optimization for parameter estimation in hydrology. *Water Resources Research* **42**: W07417, DOI: 10·1029/2005WR004528. Available Online: http://www.agu.org/pubs/crossref/2006/2005WR004528.shtml.

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

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

Gupta HV, Sorooshian S, Yapo PO. 1999. Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *Journal of Hydrological Engineering* **4**(2): 135–143.

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

Jha MK, Kumar A, Nanda G, Bhatt G. 2006. Evaluation of traditional and nontraditional optimization techniques for determining well parameters from step-drawdown test data. *Journal of Hydrological Engineering* **11**(6): 617–630.

Jung BS, Karney BW. 2006. Hydraulic optimization of Transient protection devices using GA and PSO approaches. *Journal of Water Resources Planning and Management* **132**(1): 44–52.

Kannan N, Santhi C, Williams JR, Arnold JG. 2008. Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods. *Hydrological Processes* **22**: 2114–2121.

Kennedy J, Eberhart RC. 2001. *Swarm Intelligence*. Morgan Kaufmann: San Mateo; CA.

Krishna AG. 2007. Optimization of surface grinding operations using a differential evolution approach. *Journal of Material Processing Technology* **183**(2–3): 202–209.

Kuczera G. 1997. Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research* **33**(1): 177–185.

Kuo SF, Liu CW. 2003. Simulation and optimization model for irrigation planning and management. *Hydrological Processes* **17**: 3141–3159.

Liu C. 2004. Study of some problems in water cycle changes of the Yellow River basin. *Advances in Water Science* (in Chinese) **15**(5): 608–614.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, Part I: a discussion of principles. *Journal of Hydrology* **10**(3): 282–290.

Neitsch SL, Arnold JG, Kiniry JR, King KW, Williams JR. 2002a. Soil and Water Assessment Tool (SWAT) theoretical documentation. Blackland Research Center, Texas Agricultural Experiment Station; Temple, TX, BRC Report 02–05.

Neitsch SL, Arnold JG, Kiniry JR, Srinivasan R, Williams JR. 2002b. Soil and Water Assessment Tool (SWAT) users manual. Blackland Research Center, Texas Agricultural Experiment Station, Temple, Texas, BRC Report 02–06.

Nelder JA, Mead R. 1965. A simplex method for function minimization. *Computing Journal* **7**: 308–313.

Parsopoulos KE, Vrahatis MN. 2002. Recent approaches to global optimization problems through particle swarm optimization. *Natural Computing* **1**: 235–306.

Rahman TK, Suliman A, Musirin SI. 2006. Artificial immune-based optimization technique for solving economic dispatch in power system. *Neural Nets* **3931**: 338–345.

Reca J, Martínez J. 2006. Genetic algorithms for the design of looped irrigation water distribution networks. *Water Resources Research* **42**: W05416, DOI: 10·1029/2005WR004383. Available Online: http://www.agu.org/pubs/crossref/2006/2005WR004383.shtml.

Schaffer JD, Caruana RA, Eshelman LJ, Das R. 1989. A study of control parameters affecting online performance of genetic algorithms for function optimization. In *Proceedings of the Third International Conference on Genetic algorithms*, Schaffer JD (eds). Morgan Kaufmann: San Mateo, CA; 51–60.

Seyfried MS, Harris RC, Marks D, Jacob B. 2000. A geographic database for watershed research. Reynolds Creek Experimental Watershed, Idaho, USA. USDA ARS Technical Bulletin NWRC-2000-3.

Sheridan JM. 1997. Rainfall-streamflow relations for coastal plain watersheds. *Transactions ASAE* **13**(3): 333–344.

Shi Y, Eberhart RC. 1998. Parameter selection in particle swarm optimization. In *Evolutionary Programming VII*, Porto VW, Saravanan N, Waagen D, Eiben AE (eds). Springer: Berlin; 611–616.

Sorooshian S, Duan Q, Gupta VK. 1993. Calibration of rainfall runoff models: application of global optimization to the Sacramento soil moisture accounting model. *Water Resources Research* **29**: 1185–1194.

Srivastava P. 2002. Watershed optimization of best management practices using AnnAGNPS and a genetic algorithm. *Water Resources Research* **38**(3): 1021, DOI: 10·1029/2001WR000365.

Storn R, Price K. 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**: 341–359.

Tolson BA, Shoemaker CA. 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research,* **43**: W01413, DOI: 10·1029/2005WR004723. Available Online: http://www.agu.org/pubs/crossref/2007/2005WR004723.shtml.

van Griensven A, Bauwens W. 2003. Multi-objective autocalibration for semi-distributed water quality models. *Water Resources Research* **39**(12): 1348, DOI: 10·1029/2003WR002284.

van Griensven A, Bauwens W. 2005. Application and evaluation of ESWAT on the Dender basin and the Wister Lake basin. *Hydrological Processes* **19**: 827–838.

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

Van Liew MW, Arnold JG, Bosch DD. 2005. Problems and potential of autocalibrating a hydrologic model. *Transactions of the ASAE* **48**(3): 1025–1040.

Van Liew MW, Veith TL, Bosch DD, Arnold JG. 2007. Suitability of SWAT for the conservation effects assessment project: a comparison on USDA ARS watersheds. *Journal of Hydrological Engineering* **12**(2): 173–189.

Veith TL, Sharpley AN, Weld JL, Grurek WJ. 2005. Comparison of measured and simulated phosphorous losses with index site vulnerability. *Transactions of the ASAE* **48**(2): 557–565.

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* **39**(8): 1201, DOI: 10·1029/2002WR001642.

Wang G, Guo X, Shen Y, Cheng G. 2003. Evolving landscapes in the headwaters area of the Yellow River (China) and their ecological implications. *Landscape Ecology* **18**: 363–375.

Wolpert DH, Macready WG. 1997. No free lunch theorems for optimization. *IEEE Transaction of Evolutionary Computing* **1**(1): 67–82.

Zhang X, Srinivasan R, Hao F. 2007. Predicting hydrologic response to climate change in the Luohe River Basin using the SWAT model. *Transactions of the ASABE* **50**(3): 901–910.

Zhang X, Srinivasan R, Bekele D, Hao F. 2008. Runoff simulation of the headwaters of the Yellow River using SWAT model with three snowmelt algorithms. *Journal of the American Water Resources Association* **44**(1): 48–61. DOI: 10·1111/j.1752–1688·2007·00137.x.