SCIENTIFIC BRIEFING

HP TODAY

HYDROLOGICAL PROCESSES
Hydrol. Process. **25**, 2313–2320 (2011)
Published online 18 March 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/hyp.8058

Simultaneous calibration of surface flow and baseflow simulations: a revisit of the SWAT model calibration framework

Xuesong Zhang,[1]*
Raghavan Srinivasan,[2]
Jeff Arnold,[3]
R. Cesar Izaurralde[1]
and David Bosch[4]

[1] *Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD 20740, USA*
[2] *Spatial Sciences Laboratory, Texas A&M University, College Station, TX 77843, USA*
[3] *Grassland, Soil & Water Research Laboratory USDA-ARS, 808 East Blackland Road, Temple, TX 76502, USA*
[4] *Southeast Watershed Research Laboratory, USDA-ARS, Tifton, GA 31793, USA*

*Correspondence to:
Xuesong Zhang, Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD 20740, USA.
E-mail: xuesong.zhang@pnl.gov

Received 23 April 2010
Accepted 11 February 2011

Abstract

Accurate analysis of water flow pathways from rainfall to streams is critical for simulating water use, climate change impact, and contaminants transport. In this study, we developed a new scheme to simultaneously calibrate surface flow (SF) and baseflow (BF) simulations of soil and water assessment tool (SWAT) by combing evolutionary multi-objective optimization (EMO) and BF separation techniques. The application of this scheme demonstrated pronounced trade-off of SWAT's performance on SF and BF simulations. The simulated major water fluxes and storages variables (e.g. soil moisture, evapotranspiration, and groundwater) using the multiple parameters from EMO span wide ranges. Uncertainty analysis was conducted by Bayesian model averaging of the Pareto optimal solutions. The 90% confidence interval (CI) estimated using all streamflows substantially overestimate the uncertainty of low flows on BF days while underestimating the uncertainty of high flows on SF days. Despite using statistical criteria calculated based on streamflow for model selection, it is important to conduct diagnostic analysis of the agreement of SWAT behaviour and actual watershed dynamics. The new calibration technique can serve as a useful tool to explore the trade-off between SF and BF simulations and provide candidates for further diagnostic assessment and model identification. Copyright © 2011 John Wiley & Sons, Ltd.

Key Words multi-objective optimization; SWAT; uncertainty analysis; physically oriented calibration

Introduction

Accurate analysis of water flow pathways from rainfall to streams is critical for the optimal protection of surface and groundwater (GW) resources, assessing the impacts of changes on the hydrological response and predicting contaminant transport (e.g. Wenninger *et al.*, 2004; Kannan *et al.*, 2007; Kim *et al.*, 2008; Gonzales *et al.*, 2009). Soil and water assessment tool (SWAT) (Arnold *et al.*, 1998) has been used worldwide to assist in water resources management (Gassman *et al.*, 2007). The importance of simultaneous calibration of surface flow (SF) and baseflow (BF) simulations has been emphasized in previous applications of SWAT (Arnold and Allen, 1999). Santhi *et al.* (2001) developed the general procedures for calibrating SWAT (Figure 1a). This calibration scheme requires the SF and BF be separated. Then, SF, BF, sediment, and nutrients are calibrated in sequence until each of these variables meets the statistical criteria set by the modellers. It is worth noting that the parameter calibration processes of SF and BF are interdependent with each other. Modellers need to check the performance of the SWAT on SF when they are calibrating BF and vice versa. This recursive manual calibration methodology is very time consuming. Although there are numerous automatic calibration programs for calibrating the SWAT model parameters (e.g. Abbaspour, 2008; Van Griensven, 2008; Zhang *et al.*, 2009b), no program exists for simultaneous calibration of SF and BF simulations of SWAT. In addition, few previous applications of SWAT conducted diagnostic analysis that has been shown to play critical role in model identification (e.g. Wagener, 2003; Wagener *et al.*, 2003; Wagener and McIntyre, 2005; Gupta *et al.*, 2008). Therefore, the major objective of

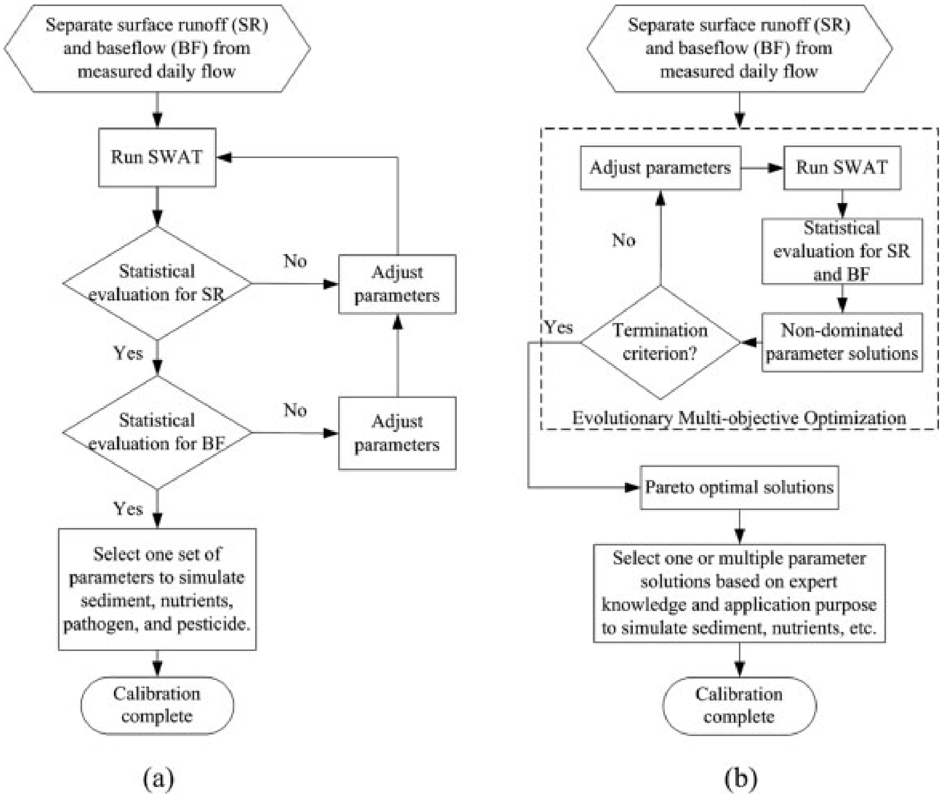

Figure 1. Two schemes for SWAT calibration (a) is adapted from Santhi *et al.* (2001) and (b) is the new scheme that employs multi-objective optimization

this study is to develop a new calibration scheme that can simultaneously calibrate SF and BF by combining multi-objective optimization and BF separation techniques and provide candidate parameter sets for further diagnostic assessment and model selection. Previous studies (e.g. Gupta *et al.*, 1998; Yapo *et al.*, 1998; Boyle *et al.*, 2000; Wagener *et al.*, 2001) have shown that evolutionary multi-objective optimization (EMO) algorithms can serve as an effective means to calibrate hydrological models with non-commensurate objectives. In this study, A Multi-ALgorithm Genetically Adaptive Method (AMALGAM; Vrugt and Robinson, 2007) adapted by Zhang *et al.* (2010) is used to simultaneously calibrate SF and BF simulations of SWAT and explore the trade-off. Another objective of this study is to examine the uncertainty analysis for SF and BF using Bayesian model averaging (BMA) by combining the multiple parameter solutions produced by EMO. The results of this research are expected to provide insights into SWAT performance on SF and BF simulations and robust model selection and uncertainty analysis when applying SWAT for watershed planning.

Materials and Methods

Study area description

The Little River Experimental Watershed (LREW) is one of the US Department of Agriculture—Agricultural Research Service experimental watersheds (Bosch *et al.*, 2007). The LREW (Figure 2) 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 LREW is the upper 334 km^2 of the Little River, 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. Soils on the watershed are predominantly sands and sandy loams with high infiltration rates. Land use types include forest (65%), cropland (30%), rangeland and pasture (2%), wetland (2%), and miscellaneous (1%).

SWAT description

The SWAT model 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 (Arnold *et al.*, 1998). SWAT subdivides a watershed into sub-watersheds connected by a stream network and further delineates hydrologic response units consisting of unique combinations of land cover and soils in each sub-watershed. The hydrologic routines within SWAT account for snow fall and melt, surface runoff (SR), vadose zone processes [i.e. infiltration, evaporation, plant uptake, lateral flows, and

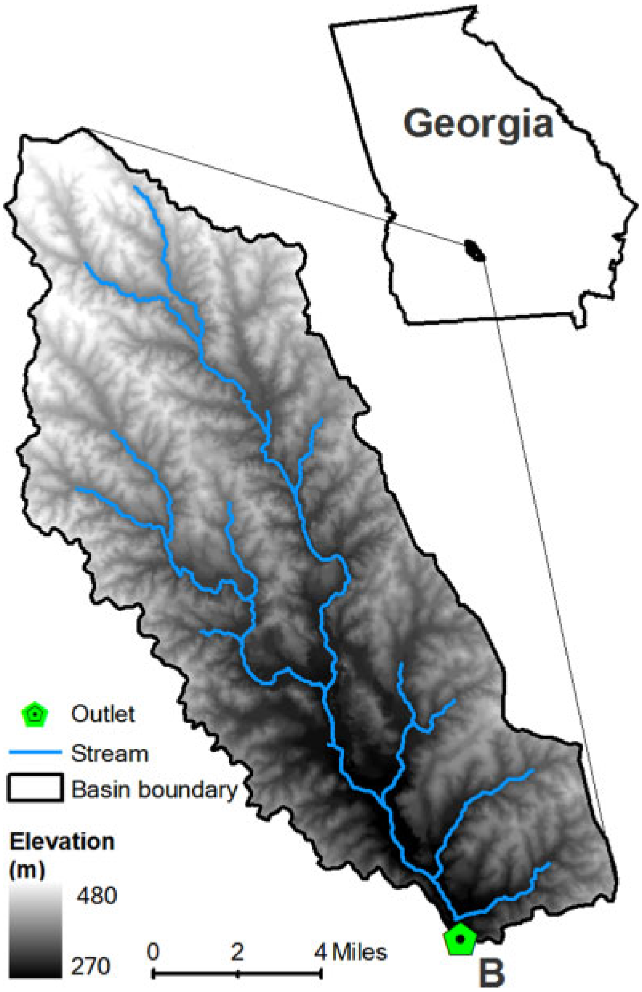

Figure 2. The location of the Little River Experimental Watershed, Georgia

Table I. Parameters for calibration of the SWAT model

Code	Parameter	Description	Range
1	CN2	Curve number II	±20%
2	ESCO	Soil evaporation compensation factor	0–1
3	SOL_AWC	Available soil water capacity	±25%
4	GW_REVAP	Groundwater 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	Baseflow recession constant	0–1
9	RCHRG_DP	Deep aquifer percolation fraction	0–1
10	CH_K2	Effective hydraulic conductivity in main channel alluvium (mm h^{-1})	0·01–150
11	SURLAG	Surface runoff lag coefficient (day)	0·5–10

percolation (PERC)], ground water flows, and river routing. In this study, 11 parameters and the corresponding parameter ranges are selected for calibration (Table I).

Modified multi-objective calibration scheme

The general multi-objective optimization problem can be defined as follows: find the parameter solution $\mathbf{x}$ that will optimize the objective function vector $\mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})]$, where $f_i(\mathbf{x})$ is the ith objective function and m the number of objective functions. An objective function vector $\mathbf{F}(\mathbf{x}') = [f_1(\mathbf{x}'), f_2(\mathbf{x}'), \ldots, f_m(\mathbf{x}')]$ is said to dominate another objective function vector $\mathbf{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})]$ (denoted by $\mathbf{F}(\mathbf{x}') \succ \mathbf{F}(\mathbf{x})$), if $\forall i \in \{1, 2, \ldots, m\}$, $f_i(\mathbf{x}') \geq f_i(\mathbf{x}) \wedge \exists i \in \{1, 2, \ldots, m\}$, $f_i(\mathbf{x}') > f_i(\mathbf{x})$ (Zitzler and Thiele, 1999). If the objective function vector

F($\mathbf{x}^*$) of a parameter solution $\mathbf{x}^* \in \Omega$ 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.

We proposed to incorporate EMO for simultaneous calibration of SF and BF simulations of SWAT (Figure 1b). This new scheme includes several major procedures: (1) separate SF and BF from observed streamflow data using BF filer, (2) employ EMO to automatically adjust the parameters of SWAT to search non-dominated parameters solutions, and (3) generate multiple Pareto optimal parameter solutions allowing users to select parameters based on expert knowledge and application purpose. In comparison to the scheme depicted in Figure 1a, the major strength of this new scheme consists in its capacity to calibrate SR and BF simultaneously and provide multiple solutions to explore trade-offs between SR and BF simulations. The major components of the new scheme are described as follows.

Multi-objective optimization algorithm. AMALGAM adaptively and simultaneously employs multiple EMO algorithms to ensure a fast, reliable, and computationally efficient solution to multi-objective optimization problems (Vrugt and Robinson, 2007). In this study, four candidate EMO algorithms, including strength Pareto evolutionary algorithm 2, particle swarm optimization, differential evolution, and adaptive metropolis sampler, were incorporated into AMALGAM following Zhang *et al.* (2010).

BF separator. The digital BF filter (Arnold and Allen, 1999) that has performed well in comparison with measured field estimates in multiple watersheds (Arnold *et al.*, 1999) is used to separate BF from total streamflow.

Statistical criteria for evaluating model performance. In this study, the Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), a normalized form of square of residuals, was selected as the model calibration objection function. In addition, two complementary evaluation coefficients, percent bias (PBIAS; Gupta *et al.*, 1999) and coefficient of determination (R^2; Legates and McCabe, 1999), are also used to assess the performance of model predictions.

Uncertainty analysis of SF and BF simulations

The multiple Pareto optimal solutions from AMALGAM can be used to further explore the uncertainty estimation of SF and BF simulations. In this study, the BMA technique described by Zhang *et al.* (2009a) is employed for uncertainty analysis. Given the errors associated with input data, boundary conditions, model structures, parameters, and observed variables, the model predictions are not a certain value, and should be represented with a confidence range (Beven, 2006). The 90% CI is examined. Two coefficients were used to evaluate the robustness uncertainty intervals: (1) the percentage of coverage (POC) of observations in the uncertainty interval and (2) the average width (AW) of the CI.

Results and Discussion

Model performance on SF and BF simulations

The BF filter was used to separate SF and BF from the observed daily streamflow from 1991 to 2002. The BF ratio obtained in this study using BF filter is 54%, which is consistent with analysed BF ratio between 50 and 55% in LREW (D. D. Boscch, personal communication). The entire evaluation period was split into SF days and BF days for calibration. A day is assumed to be a SF day if the SF contribution (SF divided by the total streamflow) of that day was larger than 0·5, otherwise, it is taken as a BF day. Among the 4383 days of the entire evaluation period, 3583 days were classified as BF days and 800 days were belonging to SF days. We divided the entire samples into two sets, viz. calibration period from 1991 to 1996 and validation period from 1996 to 2002.

For model calibration, we designed three objectives to evaluate the performance of SWAT, which are NSE values calculated using streamflows of all days (NSE_{all}), streamflows of SF days (NSE_{SF}), and streamflows of BF days (NSE_{BF}). Thirty Pareto optimal solutions were found by AMALGAM (Figure 3), which show pronounced trade-off between model performance on SF and BF simulations. The correlation between NSE_{SF} and NSE_{all} is 0·99, while the correlation between NSE_{all} and NSE_{BF} is negative (Figure 3). In most previous applications of SWAT, streamflows of all days are used for calibration, which may lead to good performance on SF days but poor performance on BF days. The ranges of NSE_{all}, NSE_{SF}, and NSE_{BF} are (0·65, 0·79), (0·52, 0·78), and (0·62, 0·78), respectively. Accordingly, Figure 4 shows the spread of the 30 parameter sets that has been normalized to a value between 0 and 1 using the corresponding ends of its range. Three parameter solutions with best NSE values on all days (Best_all), SF days (Best_SF), and BF days (Best_BF) are selected to illustrate the performance of SWAT. As Best_all and Best_SF are the same, only two parameter sets and the corresponding evaluation coefficients are listed in Table II. The preferences of the two parameter solutions to different objective functions lead to different parameter values. For example, SOL_AWC values of Best_surface and Best_base are −9 and −25%, respectively. The spread of objective function and parameter values reveal the high trade-off of SWAT's performances on SF and BF simulations, leading to the important topic on model selection.

Uncertainty analysis for SF and BF simulations

The Pareto optimal solutions found by the AMALGAM were used to provide the basis for uncertainty estimation

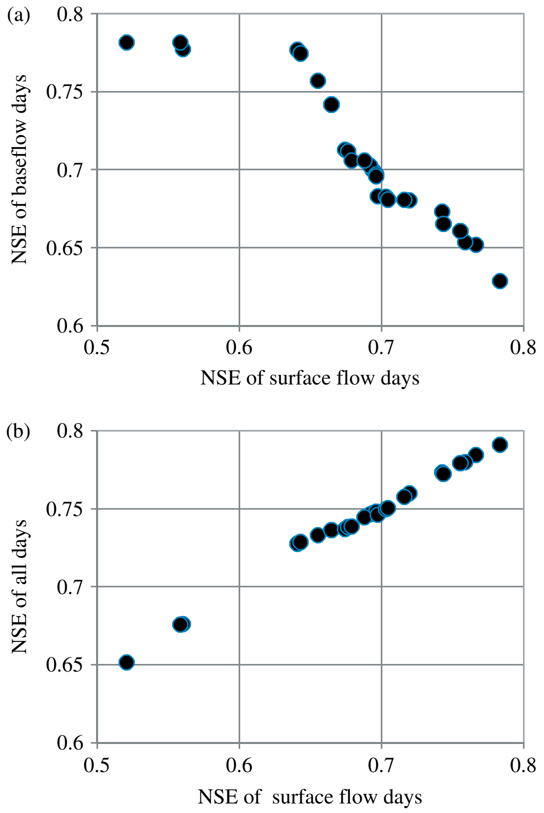

Figure 3. Thirty Pareto optimal solutions displayed in a two-dimensional objective function space: (a) NSE_{SF} *versus* NSE_{BF} and (b) NSE_{SF} *versus* NSE_{all}

of streamflow simulation of all days, SF days, and BF days, respectively. The CIs estimated using streamflow of all days, SF days, and BF days are denoted as CI-all, CI-SF, and CI-BF, respectively. CI-all includes both SF days and BF days, while CI-SF and CI-BF only

Table II. Parameter values and evaluation coefficients of three parameter solutions with best performances on all days, SF days, and BF days, respectively

Parameter solutions Parameter values and evaluation coefficients		Best_BF	Best_SF/Best_all
Parameter values	CN	−18·4%	−14·9%
	ESCO	0·90	0·98
	Surlag	0·58	0·97
	ALPHA_BF	0·87	0·93
	SOL_AWC	−25%	−9%
	CH_K2	116·7	146·2
	GW_REVAP	0·14	0·17
	GW_DELAY	3·02	3·95
	RCHRG_DP	0·03	0·04
	GWQMN	146·5	99·2
	REVAPMN	64·33	28·2
NSE	SF days	0·52	0·78
	BF days	0·78	0·62
R^2	SF days	0·76	0·79
	BF days	0·78	0·73
PBIAS	SF days	−47%	−8%
	BF days	0·5%	−7%

include SF days and BF days, respectively. Table III lists the evaluation coefficients of the 90% CIs estimated by BMA. If we use simulated streamflow of all days, the estimated POC values correspond well to the expected coverage (90%) for the observed streamflow of all days. However, it is worth noting that CI-all does not work well for SF days and BF days separately. POC value of CI-all decreases to 68% for SF days while increases to 94·8% for BF days. For CI-all, we split it into SF days and BF days for comparison purpose. On SF days, AW of CI-SF is about twice that of CI-all, while AW

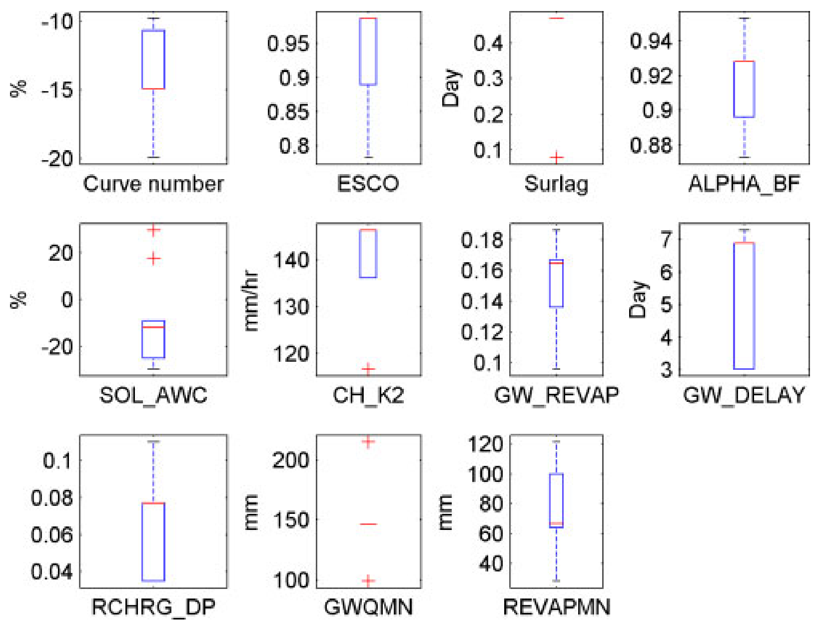

Figure 4. Normalized parameter values of the thirty Pareto optimal solutions

Table III. Properties of the 90% CIs estimated using streamflow of all days, SF days, and BF days, respectively

Evaluation coefficients		Calibration period		Validation period	
Variables		POC (%)	AW (m^3 s^{-1})	POC (%)	AW (m^3 s^{-1})
All days	CI-all	89·6	10·0	93·9	9·9
SF days	CI-all	60·4	11·4	79·6	11·7
	CI-SF	87·2	20·5	95·2	21·1
BF days	CI-all	94·8	9·6	95·8	9·5
	CI-BF	87·8	5·4	94·8	5·3

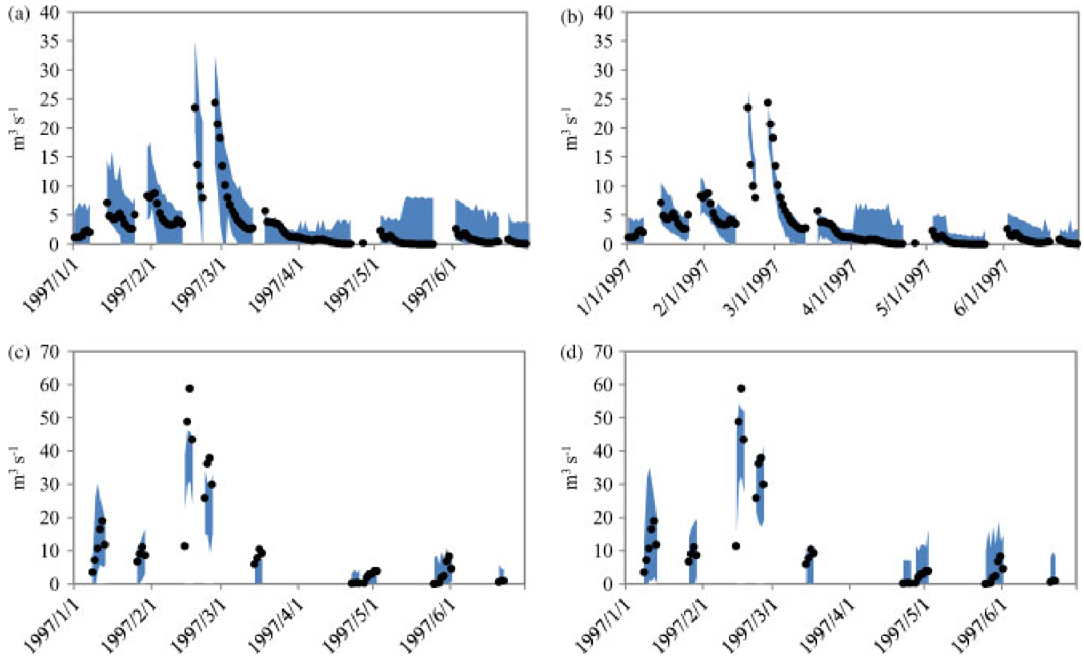

Figure 5. Estimated 90% CIs in year 1997 for (a) BF days derived using streamflow of all days, (b) BF days using streamflow of BF days, (c) SF days derived using streamflow of all days, and (d) SF days derived using streamflow of SF days

of CI-BF is only half that of CI-all on BF days. Similar results were obtained for the validation period. We use year 1997 to visually illustrate the difference between estimated CI-all, CI-SF, and CI-BF at 90% confidence level (Figure 5). In general, CI-all overestimates the uncertainty of model simulations on BF days, while underestimates this uncertainty on SF days.

Discussion

The above results reveal the unbalanced performance of SWAT for SF and BF simulations, which is similar to previous applications of multi-objective optimization algorithms for calibrating conceptual models with objective functions tailored for different ranges of flows (e.g. Gupta *et al.*, 1998; Yapo *et al.*, 1998; Boyle *et al.*, 2000; Wagener *et al.*, 2001). The very different performances of the Pareto optimal solutions lead to another important topic of model selection. Which parameter solution should be selected for further watershed analysis? By combining the criteria suggested by Santhi *et al.* (2001) and Moriasi *et al.* (2007), model simulation can be judged as satisfactory if NSE >0·50, R^2 > 0·60, and PBIAS <25% for streamflow. According to these criteria, Best_BF should not be adopted because of its large PBIAS on SF days (Table II). Among the 30 Pareto optimal solutions, 19 meet these requirements for calibration period. For the validation period, five solutions are satisfactory. If we increase the statistical criteria standards, less model simulations will be selected. Finally, only one model simulation with the best statistical performance coefficients can be selected. However, this scheme neglects whether the hydrologic model can reasonably approximate the rainfall–runoff processes essential for practical water resources management. The ranges of evapotranspiration (ET), soil water (SW), PERC, SR, GW, overland sediment load, and BF ratio simulated by different parameter solutions are shown in Table IV. These values are averaged over multiple years and across the entire watershed. For the 30 Pareto optimal solutions, wide ranges of simulated variables are observed. For example, SW ranges between 141 and 263 mm and sediment load ranges from 0·8 to 3·2 t/ha. Even when we

Table IV. Ranges of simulated variables of interest by different parameter solutions

Variables Parameter solutions	ET (mm)	SW (mm)	PERC (mm)	SR (mm)	GW (mm)	Sediment (t/ha)	BF ratio
Thirty Pareto optimal solutions	(636, 778)	(141, 263)	(272, 431)	(124, 208)	(156, 257)	(0·8, 3·2)	(0·43, 0·65)
Nineteen solutions with satisfactory performance in calibration period	(636, 751)	(146, 243)	(303, 431)	(140, 208)	(158, 257)	(1, 3·2)	(0·43, 0·65)
Five solutions with satisfactory performance in calibration and validation period	(654, 751)	(146, 239)	(303, 406)	(150, 173)	(171, 231)	(2·2, 2·6)	(0·50, 0·58)

reduce the number of satisfactory solutions to five, pronounced variability still exists. For example, the range of SW values simulated by the five parameter solutions is between 146 and 239 mm. Auxiliary information of these water fluxes and storage variables and sediment load is required to conduct physically oriented calibration and evaluation of the performance of different parameter solutions on approximating watershed dynamics. Practically, limited data availability may hinder physically oriented calibration and evaluation. In addition, the limitations of model structure make it very difficult to achieve good approximation of all the rainfall–runoff processes (Wagener and McIntyre, 2005). Therefore, purpose-oriented calibration, which calibrates hydrologic models to support a specific management decision process, may be a practical choice. For example, the model's performance on SF simulation should be emphasized if the major purpose of model application is to evaluate sediment load, because SF is the major driver of soil erosion. Recent research (e.g. Herbst *et al.*, 2009a, 2009b) showed that self-organizing map can be used to classify modelled streamflow time series and analyse diagnostic signatures of model behaviour, which could potentially serve as tools for decision-makers select model realizations according to the purpose of the model application. However, it is worth noting that the purpose-driven model calibration may sacrifice the hydrological realism (Wagener and McIntyre, 2005). Overall, the results obtained in this study illustrate the weakness of relying on statistical criteria and streamflow data to calibrate SWAT model. Diagnostic approach to model evaluation and identification that has been extensively presented and discussed in previous research (e.g. Wagener, 2003; Wagener *et al.*, 2003; Wagener and McIntyre, 2005; Gupta *et al.*, 2008) should be explored in the future application of SWAT. The new calibration scheme designed in this study is expected to serve as a useful tool to facilitate the implementation of the general framework described by Gupta *et al.* (2008) towards better diagnostic analysis of SWAT.

For uncertainty analysis, it is also important to differentiate the performances of SWAT on SF and BF simulations. The major reason for the unbalanced performances of CI-all on SF days and BF days is the difference between the residual characteristics associated with SF days and BF days. For example, during calibration period, the average absolute residual values from the thirty Pareto optimal solutions range between 4·57 and 5·75 $m^3\ s^{-1}$ for SF days while between 1·09 and 1·26 $m^3\ s^{-1}$ for BF days. The mechanisms that control SF and BF are very different in SWAT. However, means to represent the residual characteristics and model mechanisms that are associated with BF and SF simulations into Markov Chain Monte Carlo algorithms and meaningful probabilistic function forms still deserves further research. Dynamic Bayesian-based model averaging methods (e.g. Marshall *et al.*, 2007) that can differentiate model performance over different temporal periods (e.g. BF or SF dominated periods) could potentially serve as tools to improve uncertainty analysis of hydrologic modelling.

Summary

A new calibration scheme by incorporating EMO was developed to simultaneously calibrate SF and BF simulations of SWAT. The application of this scheme demonstrated pronounced trade-off between SWAT's performance on SF and BF simulation. The wide ranges of simulated major water fluxes and storage variables associated with different Pareto optimal solutions indicate the importance of using auxiliary information on the physical water cycle to assist in model selection. Uncertainty analysis without differentiating SF and BF simulations may lead to unreasonable uncertainty estimation with respect to SF and BF. The BMA analysis results show that, if all streamflow data are used, uncertainties associated with low flows on BF days and high flows on SF days are overestimated and underestimated, respectively. When applying SWAT for water resources management, it is important to conduct diagnostic analysis of the agreement of

SWAT behaviour and actual watershed dynamics, despite using statistical criteria calculated based on streamflow for model selection. The combination of multi-objective optimization and BF separation method can serve as a useful tool to explore the trade-offs between SF and BF simulations and provide candidates for model identification and uncertainty analysis.

Acknowledgements

We are grateful to the associate editor and two anonymous reviewers for their constructive suggestions that have improved this manuscript significantly.

References

Abbaspour KC. 2008. *SWAT-CUP2: SWAT Calibration and Uncertainty Programs—A User Manual*. Department of Systems Analysis, Integrated Assessment and Modeling (SIAM), Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland, 95pp.

Arnold JG, Allen PM. 1999. Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association* **35**(2): 411–424.

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

Beven KJ. 2006. A manifesto for the enquiringly thesis. *Journal of Hydrology* **320**: 18–36.

Bosch DD, Sheridan JM, Lowrance RR, Hubbard RK, Strickland TC, Feyereisen GW, Sullivan DG. 2007. Little River Experimental Watershed database. *Water Resources Research* **43**(9): W09470. DOI: 10.1029/2006WR005844.

Boyle DP, Gupta HV, Sorooshian S. 2000. Toward improved calibration of hydrologic models. Combining the strengths of manual and automatic methods. *Water Resources Research* **36**: 3663–3674.

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

Gonzales AL, Nonner J, Heijkers J, Uhlenbrook S. 2009. Comparison of different base flow separation methods in a lowland catchment. *Hydrol. Earth Syst. Sci.* **6**(2): 3483–3515.

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

Gupta HV, Wagener T, Liu Y. 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes* **22**: 3802–3813. doi: 10.1002/hyp.6989.

Herbst M, Casper MC, Grundmann J, Buchholz O. 2009a. Comparative analysis of model behaviour for flood prediction purposes using Self-Organizing Maps. *Natural Hazards and Earth System Sciences* **9**: 373–392.

Herbst M, Gupta HV, Casper MC. 2009b. Mapping model behaviour using Self-Organizing Maps. *Hydrology and Earth System Sciences* **13**: 395–409.

Kannan N, White SM, Worrall F, Whelan MJ. 2007. Hydrological modeling of a small catchment using SWAT-2000—ensuring correct flow partitioning for contaminant modeling. *Journal of Hydrology* **334**(1–2): 64–72.

Kim NW, Chung IM, Won YS, Arnold JG. 2008. Development and application of the integrated SWAT–MODFLOW model. *Journal of Hydrology* **356**: 1–16.

Legates DR, McCabe GJ. 1999. Evaluating the use of "goodness of fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35**(1): 233–241.

Marshall L, Nott D, Sharma A. 2007. Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes* **21**: 847–861.

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50**(3): 885–900.

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

Santhi C, Arnold JG, Williams JR, Dugas WA, Hauck L. 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. *Journal of the American Water Resources Association* **37**(5): 1169–1188.

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

Van Griensven A, Meixner T, Srinivasan R, Grunwals S. 2008. Fit-for-purpose analysis of uncertainty using split-sampling evaluations. *Hydrological Sciences Journal* **53**(5): 1090–1103.

Vrugt JA, Robinson BA. 2007. Improved evolutionary optimization from genetically adaptive multimethod search. *PNAS* **104**: 708–711.

Wagener T. 2003. Evaluation of catchment models. *Hydrological Processes* **17**: 3375–3378.

Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001. A framework for development and application of hydrological models. *Hydrology and Earth System Sciences* **5**(1): 13–26.

Wagener T, McIntyre N. 2005. Identification of hydrologic models for operational purposes. *Hydrological Sciences Journal* **50**(5): 1–18.

Wagener T, Wheater HS, Gupta HV. 2003. Identification and evaluation of watershed models. Water Science Application. Series, vol. 6, In *Calibration of Watershed Models*, Duan Q, Sorooshian S, Gupta HV, Rousseau A, Turcotte R (eds). AGU, Washington, D. C. 29–47.

Wenninger J, Uhlenbrook S, Tilch N, Leibundgut C. 2004. Experimental evidence of fast groundwater responses in a hillslope/floodplain area in the Black Forest Mountains, Germany. *Hydrological Processes* **18**(17): 3305–3322.

Yapo PO, Gupta HV, Sorooshian S. 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* **204**: 83–97.

Zhang X, Srinivasan R, Bosch D. 2009a. Calibration and uncertainty analysis of the SWAT model using genetic algorithms and Bayesian model averaging. *Journal of Hydrology* **374**: 307–317.

Zhang X, Srinivasan R, Zhao K, Van Liew M. 2009b. Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. *Hydrological Processes* **23**: 430–441.

Zhang X, Srinivasan R, Van Liew M. 2010. On the use of a multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model. *Hydrological Processes* **24**: 955–969.

Zitzler E, Thiele L. 1999. Multiobjective evolutionary algorithms: a comparative case study and the Strength Pareto approach. *IEEE Transactions on Evolutionary Computation* **3**(4): 257–271.