HYDROLOGICAL PROCESSES
Hydrol. Process. (2012)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9589

Assessing SWAT's performance in the Kaskaskia River watershed as influenced by the number of calibration stations used

Li-Chi Chiang,[1] Yongping Yuan,[1]* Megan Mehaffey,[1] Michael Jackson[1] and Indrajeet Chaubey[2]

[1] *USEPA NERL, Landscape Ecology Branch, Environmental Sciences Division, 944 East Harmon Avenue, Las Vegas, Nevada 89119, USA*
[2] *Departments of Agricultural and Biological Engineering, and Earth and Atmospheric Sciences, Division of Environmental and Ecological Engineering, Purdue University, West Lafayette, Indiana 47907, USA*

Abstract:

The Future Midwestern Landscapes (FML) project is part of the U.S. Environmental Protection Agency's Ecosystem Services Research Program. The goal of the FML project is to quantify changes in ecosystem services across the Midwestern region as a result of the growing demand for biofuels. Watershed models are an efficient way to quantify ecosystem services of water quality and quantity. By calibrating models, we can better capture watershed characteristics before they are applied to make predictions. The Kaskaskia River watershed in Illinois was selected to investigate the effectiveness of different calibration strategies (single-site and multi-site calibrations) for streamflow, total suspended sediment (TSS) and total nitrogen (TN) loadings using the Soil and Water Assessment Tool. Four USGS gauges were evaluated in this study. Single-site calibration was performed from a downstream site to an upstream site, and multi-site calibration was performed and fine-tuned based on the single-site calibration results. Generally, simulated streamflow and TSS were not much affected by different calibration strategies. However, when single-site calibration was performed at the most downstream site, the Nash–Sutcliffe efficiency (NSE) values for TN ranged between −0.09 and 0.53 at the other sites; and when single-site calibration was performed at the most upstream site, the NSE values ranged between −8.38 and −0.07 for the other sites. The NSE values for TN were improved to 0.5 – 0.59 for all four sites when multi-site calibration was performed. The results of the multi-site calibration and validation showed an improvement on model performance on TN and highlighted that multi-site calibrations are needed to assess the hydrological and water quality processes at various spatial scales. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS Future Midwestern Landscapes study; SWAT; single-site calibration; multi-site calibration; model validation

Received 17 December 2011; Accepted 3 October 2012

INTRODUCTION

The benefits that people gain from the natural ecosystems (termed ecosystem services) include provisioning services, regulating services, cultural services and supporting services (MA, 2004). The measurement of ecosystem services is a new strategic focus for the United States Environmental Protection Agency's (EPA) Ecosystem Services Research Program (ESRP). The ESRP's Future Midwestern Landscapes study aims to quantify the current magnitude of ecosystem services in the Midwestern U.S. and examine how those services would change over the next decade given the growing demand for biofuels. Although an increase in corn production can have a positive energy benefit for biofuel production, it may also have many negative impacts on environmental quality such as increased nutrient and pesticide losses to water bodies. In the United States, the increased nitrogen has contributed to the degradation of the ecosystem, especially the estuaries (NRC, 2000). The Gulf of Mexico is one of the examples that increased nutrient fluxes have been linked to the increased occurrence of seasonal hypoxia in the northern Gulf of Mexico (CENR, 2000; Alexander *et al.*, 2008). Moreover, more than 70% of the nitrogen and phosphorus delivery to the Gulf of Mexico are from the agricultural lands in the Mississippi River Basin (Alexander *et al.*, 2008).

Frequently watershed models are used to simulate responses under various land use/management scenarios in order to make management recommendations for improving water quality. For example, the Soil and Water Assessment Tool (SWAT) has been widely applied to evaluate alternative land use, best management practices and other factors on pollutant losses to streams within a watershed (Vache *et al.*, 2002; Chaplot *et al.*, 2004; Santhi *et al.*, 2006; Chaubey *et al.*, 2010; Chiang *et al.*, 2010). Before evaluating alternative scenarios on watershed responses, a model is usually calibrated by comparing the simulated data to available measured data. The availability and the location of observed data are some of the factors driving choices about the calibration and validation strategies in terms of temporal span and

*Correspondence to: Yongping Yuan, USEPA NERL, Landscape Ecology Branch, Environmental Sciences Division, 944 East Harmon Avenue, Las Vegas, Nevada 89119, USA.
E-mail: yuan.yongping@epa.gov

Notice: Although this work was reviewed by USEPA and approved for publication, this scientific paper may not necessarily reflect official Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

spatial scales. For watersheds that have only one gauge available, the calibration and validation results have typically been applied to an entire watershed and the spatial variability of parameters cannot be accounted for across the watershed (Jha *et al.*, 2006; Srivastava *et al.*, 2006; Stewart *et al.*, 2006; Green and van Griensven, 2008). The gauging station for these types of studies was located either close to or at the watershed outlet. The watershed drainage areas range from 0.04 to 447 500 km^2. Although the authors used the coefficient of determination (R^2) and Nash–Sutcliffe efficiency (NSE) to evaluate model's performance and concluded that the model performed satisfactorily (R^2 ranged from 0.57 to 0.96 and NSE ranged from 0.54 to 0.94 for monthly streamflow), the calibration and validation of hydrologic processes at the watershed outlet do not imply a full understanding of the complexity of hydrologic patterns within the watershed. The satisfactory performance of the model at the watershed outlet may cover up the overestimations or underestimations at the subwatershed outlets. It is because the overestimations or underestimations can be averaged out at the watershed outlet, especially for a large watershed with spatially complex hydrological and water quality processes.

It is found that the calibration results for small-scale watersheds are generally better than that for large-scale watersheds. For example, a study performed by Green and van Griensven (2008) found that R^2 values ranged between 0.87 and 0.96 and NSE values ranged between 0.82 and 0.94 at the monthly time step for six small-scale watersheds (4 – 8.4 ha) in Texas. Therefore, many studies have conducted multi-site calibration and validation for different variables (Qi and Grunwald, 2005; White and Chaubey, 2005; Cao *et al.*, 2006; Zhang *et al.*, 2008) and further developed their watershed management plans (Kirsch *et al.*, 2002). White and Chaubey (2005) performed calibration and validation at three different sites for streamflow, sediment, total phosphorus and (NO^3-N) plus nitrite-nitrogen (NO^2-N). Qi and Grunwald (2005) conducted a spatially distributed calibration and validation of surface, groundwater and total flow concurrently at four subwatersheds and the watershed outlet. Similarly, Cao *et al.* (2006) evaluated the performance of model calibration and validation on hydrological processes at six subcatchments and the catchment. Above studies concluded that multi-site and multi-variable method can identify the areas where hydrological process needs more calibration effort, thus improving model performance compared to single-site calibration.

Many studies have shown that the equifinality of model parameters where multiple combinations of parameters may yield the same model outputs is one of the concerns in hydrological modeling (Beven and Binley, 1992; Wagener and Kollat, 2007). Therefore, a systematical assessment of selecting optimal combinations of parameters for multiple sites is needed for calibrating a model which truly represents the diverse characteristics of a watershed. Moreover, relatively fewer studies have been done to evaluate the impact of multi-site calibration and validation on nutrient loadings (White and Chaubey, 2005; Santhi *et al.*, 2006). The calibration results for nitrate and nitrite ranged from 0.01 to 0.84 (R^2) and from −2.35 to 0.29 (NSE) (White and Chaubey, 2005), while the calibration results for mineral N ($R^2 = 0.64$ and 0.72) and organic N ($R^2 = 0.61$ and 0.60) were similar (Santhi *et al.*, 2006). For subwatersheds having similar land use distribution, model performance tends to be similar at each subwatershed outlet when common calibrated nutrient-related values are applied. In contrast, various calibration results at multiple sites indicate the impacts of complex landscape and in-stream processes on nutrient loadings at different subwatersheds. The objectives of this study were to: (1) assess the model performance in the Kaskaskia River watershed of single-site and multi-site calibration on streamflow, total suspended sediment (TSS) and total nitrogen (TN) for different sizes of subwatersheds; and (2) develop a multi-site calibration strategy that can truly calibrate the model with a calibrated parameter set to best capture the hydrological and water quality processes at various subwatershed outlets.

RESEARCH METHOD AND PROCEDURES

Study site

Our study was conducted in the 14 152 km^2 Kaskaskia River watershed within the Upper Mississippi River Basin located in southern Illinois (Figure 1). Land covers in the Kaskaskia River watershed mainly comprise cropland (approximately 67% of the watershed area), forest (17.6%), pasture (6.7%) and urban (8.8%). Corn, soybean and wheat are the three major crops, and corn and soybean rotation accounts for more than 44% of the watershed (Table I). The average hillslope of the watershed is 2% with a maximum slope of 58%. The predominant soil associations in the watershed include the Bluford –Ava-Hickory (IL038, 15%), Hosmer-Stoy-Hickory (IL037, 14%), Cisne-Hoyleton-Darmstadt (IL006, 13%), Flanagan-Drummer-Catlin (IL010, 13%) and Cowden-Oconee-Darmstadt (IL005, 13%).

Soil associations IL038, IL037, IL005 and IL006 underlie the southern part of the watershed accounting for 55% of the watershed. Bluford soils (silt loam) tend to be anywhere from completely flat to gently sloping and are frequently poorly drained. Hickory soils (loam) are moderately steep to very steep and moderately well drained. The Cisne soil series (silt loam) are poorly drained with very slow permeability. The Darmstadt series (silt loam) also consists of poorly drained soils. The major hydrologic groups in the Darmstadt series are C, D and C/D, indicating soils having slow infiltration rates and high runoff potential. The Flanagan soils (silt loam) dominate the northern part of the watershed where most croplands are located. This soil series has moderate soil erodibility ranging from 0.28 to 0.43, indicating moderate susceptibility to detachment and moderate rate of runoff (Walker and Pope, 1983; Hamilton, 1993).

DOI: 10.1002/hyp

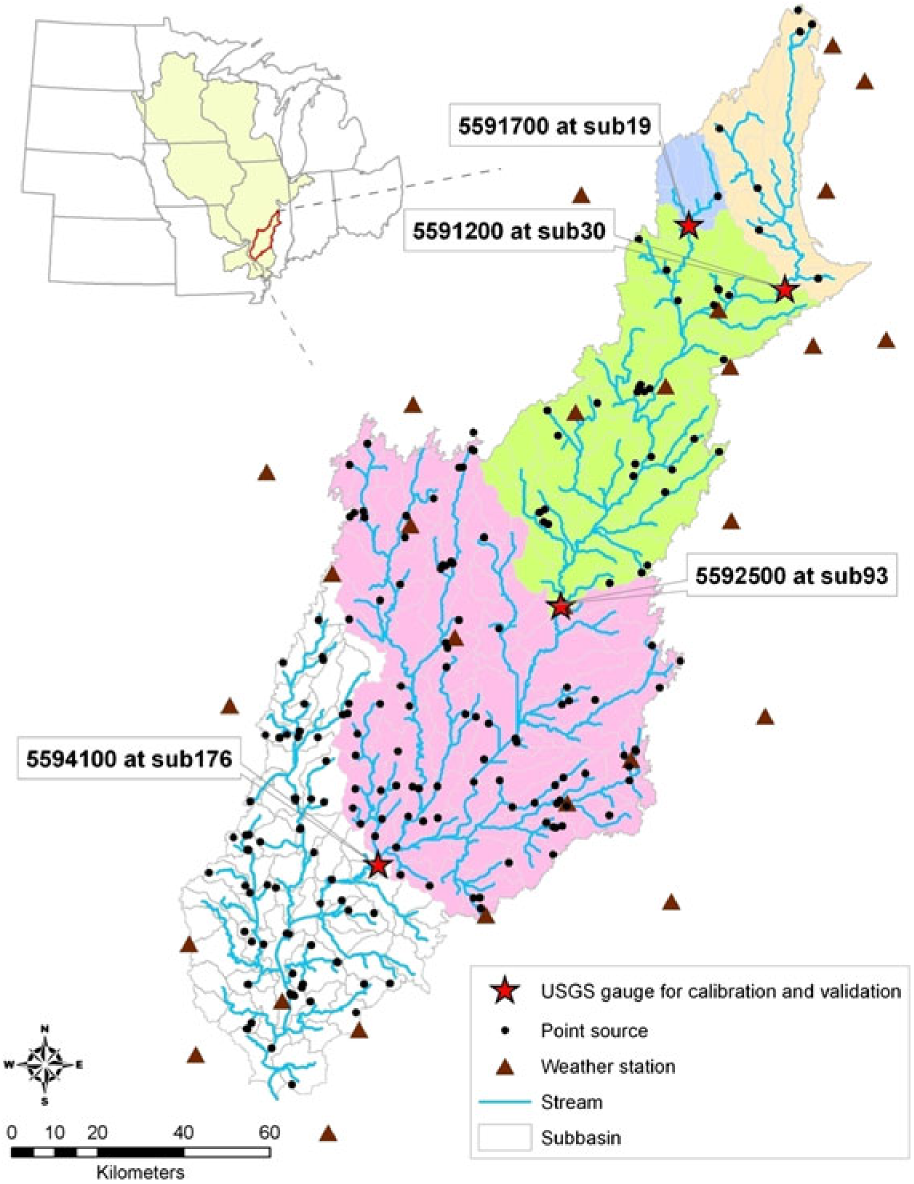

Figure 1. Location of point sources, weather stations and USGS gauges for model evaluation in the Kaskaskia River watershed

Table I. Land use distribution at subwatersheds and entire watershed

(% of the subwatershed)	sub19	sub30	sub93	sub176	Kaskaskia
Urban	5.61	9.10	7.61	8.24	8.76
Pasture	1.17	2.51	5.06	6.31	6.73
Forest	0.83	1.57	14.25	17.28	17.61
Soybean	0.21	0.24	0.22	0.33	0.32
Corn	6.99	5.93	6.55	5.64	5.20
Corn-Soybean	82.68	74.93	55.55	46.92	44.03
Soybean-Other	2.08	4.83	9.65	13.92	15.95
Corn-Other	0.43	0.90	1.12	1.36	1.36
Water	0.00	0.00	0.00	0.00	0.03
Total area (ha)	26 638	112 949	487 501	1 100 095	1 470 704

Based on the availability and locations of the monitoring data, we selected four gauges for model calibration and validation. These four gauges are USGS5591700 at subwatershed 19 delineated by the SWAT model, USGS5581200 at subwatershed 30, USGS5592500 at subwatershed 93 and USGS5594100 at subwatershed 176 (Figure 1). Subwatershed 19 and subwatershed 30 are stand alone, but they are nested in subwatershed 93

DOI: 10.1002/hyp

and subwatershed 93 is nested in subwatershed 176 (Figure 1). Most of the croplands are located in the northern part of the watershed. Both subwatershed 19 and 30 have more than 85% of the subwatershed area occupied by croplands, while subwatershed 93 and 176 have less croplands and more forest lands. The annual amount of N fertilizer and P fertilizer applied on the corn lands in the watershed ranged from 67 903 to 85 049 t (metric tons) and from 24 993 to 35 607 t, respectively, during 1990–2009. The soybean lands in the watershed received an average annual amount of 1225 t of N fertilizer and 5171 t of P fertilizer from 1990 to 2009 (ESMIS, 1990).

SWAT model description

The SWAT model can predict long-term impacts of land use and management on water, sediment and agricultural chemical yields at various temporal and spatial scales in a watershed (Arnold *et al*., 1998). More than 600 peer-reviewed journal articles have been published demonstrating the SWAT applications on sensitivity analyses, model calibration and validation, hydrologic analyses, pollutant load assessment and evaluation of conservation practices (Gassman *et al*., 2007). In this study, surface runoff was estimated using the SCS curve number (CN) procedure (SCS, 1972), and potential evapotranspiration was estimated using the Penman–Monteith method (Monteith, 1965; Allen, 1986; Allen *et al*., 1989). Erosion caused by rainfall and runoff is computed with the Modified Universal Soil Loss Equation (Williams, 1995). Once the sediment yield is estimated, sediment transport in the channel network is simulated as a function of two processes, deposition and degradation (Neitsch *et al*., 2002).

Nitrogen cycle is a dynamic system that includes atmosphere, soil and water. SWAT monitors five different pools of nitrogen in the soil: two pools are inorganic forms of nitrogen (NH_4^+ and NO_3^-), and three pools are organic forms of nitrogen (active, stable and fresh organic N). The amount of nitrogen loss from land areas depends on various factors, such as mineralization, nitrification, denitrification and fixation by legumes. SWAT models the transport of different forms of nitrogen from land areas into streams. Nitrate may be moved with surface runoff, lateral flow or percolation, while organic N attached to soil particles may be transported by surface runoff. After nitrogen is transported into main streams, SWAT simulates in-stream nutrient processes by incorporating the QUAL2E model (Brown and Barnwell, 1987).

The key GIS input files to SWAT for this study included a 30 m digital elevation model (DEM) downloaded from the National Elevation Dataset at the resolution of 1 arc-second from the USGS (weblink: http://ned.usgs.gov), (USGS; weblink: http://ned.usgs.gov/), an enhanced land cover/land use based on the 2001 National Land Cover Database (NLCD) and State Soil Geographic Database (STATSGO) from the USDA-NRCS (USDA-NRCS; weblink: http://soildatamart.nrcs.usda.gov/). Although the Soil Survey Geographic (SSURGO) database provides higher spatial resolution soil data, the GIS-enabled SWAT input interfaces were developed only for processing the STATSGO database, and significantly additional effort is needed to prepare SSURGO database for model use for such a large watershed (Sheshukov et al., 2011). Using the DEM and the outlets selected within the watershed, the watershed was delineated into several subwatersheds. In this study, the watershed delineation was matched to the boundaries of the 12-digit National Hydrography dataset, and additional subwatersheds were delineated at USGS gauge stations to facilitate comparisons of model outputs to the measured data. Thereby, a total of 218 subwatersheds were delineated. Subsequently, the subwatersheds were partitioned into homogeneous units (hydrologic response units), which share the same land use and soil type. In this study, we used 1% threshold for both land use and soil type. Besides the GIS layers, other input files, such as weather data, agricultural management schedule and operations, fertilizer application rates and point sources, were prepared and further discussed in the following sections.

Model inputs for the Kaskaskia River Watershed

Landcover dataset. We used the 2001 NLCD as our base year landcover dataset for the Midwest. However, the 2001 NLCD did not have detailed information on crop type and rotational practices. In particular, the NLCD does not differentiate varieties of crops, such as corn, soybeans and wheat, and has only a single time period rendering rotational practices. Therefore, for our modeling needs, we developed an aggregate landcover classification by combining the NLCD 2001 with the USDA National Agriculture Statistical Survey (NASS) Cropland Data Layer for the years 2004–2007. Our method expanded the 'Single cultivated crops' land use within the NLCD into 18 classes of agriculture including monoculture cropping and rotational cropping types (Mehaffey *et al.* 2011).

Weather data. Weather data (daily precipitation, minimum and maximum temperature) used in this study were the National Weather Service Cooperative Observer daily observations from the National Climatic Data Center (NCDC). A total of 28 weather stations were found within a radius of 25 miles of the watershed. Daily observations during the period of 1960 – 2009 were downloaded from the NCDC website. Missing records of daily observations at a given station were interpolated using the weather data from neighborhood weather stations using the method developed by Di Luzio *et al.* (2008).

Point source. The SWAT model can read point source data in various time steps, such as hourly, daily, monthly and yearly, and also at constant concentrations. The model also incorporates the point source into the routing process through the watershed. A total of 265 point sources were observed in the watershed for flow, ammonia and total suspended solids during 1998–2007. The point source data were collected once a month and are available at the EPA's Envirofacts Website (http://www.epa.gov/enviro/html/pcs/adhoc.html).

DOI: 10.1002/hyp

Thus, the daily input was used, and the collected point source data once a month were assumed to be constant for that month; i.e. the once a month point source data were the daily input for each day in that month. SWAT subbasins (and/or streams) where those point sources are located were first identified, the effluent from point sources along with flow and pollutants generated from that subbasin were put into nearly streams during SWAT simulations. In cases where a point source is located on a stream, the effluent is directly put into that stream. The water and pollutant loadings from point sources were relatively small comparing with nonpoint source loadings in this watershed.

Fertilizer input. Management schedules for the various crop rotations were obtained from the crop management template in the Revised Universal Soil Loss Equation 2 developed by the USDA Natural Resource Conservation Service (NRCS). The crop management templates were grouped based on the NRCS crop management zones (CMZ). The CMZ data were further processed using the Annualized Agricultural Nonpoint Source Pollutant Loading model Input Editor and then prepared for the SWAT crop management files. The schedules and detailed information of operations, such as planting, tillage and harvesting, were documented in the CMZ templates, except for the fertilizer application. Fertilizer was assumed to be applied one day before planting, and fertilizer application rates for corn, soybean and wheat were estimated based on data obtained from USDA Economics, Statistics and Market Information System (ESMIS) and the Census of Agriculture of USDA NASS (Table II).

Table II. The average values of N and P fertilizer application rates for corn, soybean and wheat during 1990–2009 in the Kaskaskia River watershed and Illinois state average in 1990

	Fertilizer type	Average rate (kg/ha)	Illinois state application rate in 1990 (kg/ha)
Corn	N	187.6	183.8
	P	78.8	82.6
Soybean	N	3.2	4.5
	P	13.7	17.5
Wheat	N	98.1	94.4
	P	66.7	73.7

Model calibration and performance evaluation

The calibration and validation of the SWAT model were performed manually for streamflow, TSS and TN at four selected subwatershed outlets using a monthly time step. The simulation period was from 1960 to 2009, where the first 3 years were used as the model warming-up years. The streamflow had the longest observation period, which could approximately cover the simulation period (Table III). There were only two gauges having the observed TSS concentrations, and USGS5594100 has the longest period of nitrogen data from 1974 to 1997. Based on the availability of streamflow and water quality observations, the calibration and validation were performed from 1962 to 1985 and from 1986 to 2009, respectively. Since the water quality data were not collected continuously, a load estimation tool (Runkel *et al.*, 2004) was used to estimate the constituent loads at a continuous monthly time step. Estimated water quality loads for days when measurements were taken were compared with measured water quality loads and good relationships between daily estimated and measured were achieved with relative errors from −6.8% to 28.2% and R-square values from 0.61 to 0.95 (Table IV). For monthly load estimation, the lower and upper limit of load estimation was provided in Table IV. In addition, the relationships of monthly estimated water quality loads with measured flow were also provided in Table IV (R-square values from 0.65 to 0.98).

Single-site calibration was performed first. For single-site calibration, the model was calibrated from downstream to upstream. The purpose of this calibration strategy was to evaluate how the model performed at different sites when the constant calibrated values derived from a single site were applied to the entire watershed. Thus, the outlet at subwatershed 176, downstream of the Kaskaskia River was first selected to calibrate the model and apply the calibrated parameter values to the entire watershed. In the same way, models were also calibrated using the observed data from subwatershed 93, 30 and 19, and calibrated parameters from each calibration were applied to the entire watershed. Based on the experience of single-site calibration and how the constant calibrated parameter values affected the model performance at each subwatershed outlet, multi-site calibration was performed.

Table III. List of available periods of measured streamflow, total suspended sediment (TSS) and total nitrogen (TN) at 4 USGS gauges. The entire monitoring period was split into calibration and validation periods

USGS gauge	Subbasin	USGS-drainage area (ha)		Streamflow	TSS	TN
05591200	30	122 506	Calibration	1970/10 – 1985/12	1979/1 – 1985/12	1980/1 – 1985/12
			Validation	1986/1 – 2009/12	1986/1 – 1992/12	1986/1 – 1997/12
05591700	19	29 008	Calibration	1980/3 – 1985/12	–	1980/3 – 1985/12
			Validation	1986/1 – 2009/12	–	1986/1 – 1997/12
05592500	93	502 457	Calibration	1963/1 – 1985/12	–	1977/1 – 1985/12
			Validation	1986/1 – 2009/12	–	1986/1 – 1997/12
05594100	176	1 137 781	Calibration	1969/10 – 1985/12	1975/1 – 1985/12	1974/1 – 1985/12
			Validation	1986/1 – 2009/12	1986/1 – 1992/12	1989/1 – 1997/12

Table IV. Comparison of estimated daily TSS and TN loads with measured daily TSS and TN loads; and estimated monthly TSS and TN loads and their relationships with measured monthly flow. (Note: number in the parenthesis denotes the total number of measured daily data points during the entire monitoring period)

			Daily				Monthly LOADEST			
								95% Confidence Intervals		
	Subbasin	Period	USGS measured (kg)	LOADEST (kg)	RE (%)	R-square	Mean Load (t)	Lower (t)	Upper (t)	R-square with measured flow
TSS	30	1979–1992 (218)	508 182.4	651 494.8	28.2	0.61	4277.6	2200.6	7592.6	0.65
	176	1975–1992 (213)	2 094 666.8	2 631 793.9	25.6	0.55	50 447.1	28 725.3	82 655.0	0.88
TN	19	1980–1997 (113)	3282.6	3894.6	18.6	0.95	119.2	91.0	153.7	0.98
	30	1980–1997 (137)	11 931.7	12 499.8	4.8	0.95	361.8	286.8	450.8	0.94
	93	1979–1997 (144)	23 227.6	21 641.3	−6.8	0.72	641.5	478.9	842.7	0.94
	176	1974–1997 (196)	23 842.7	25 323.7	6.2	0.85	796.4	619.7	1008.3	0.96

Four statistical measurements were used to evaluate the model performance (White and Chaubey, 2005; Engel *et al.*, 2007; Moriasi *et al.*, 2007). They were the coefficient of determination (R^2), NSE, RMSE-observations standard deviation ratio (RSR) and percent bias (PBIAS). The process of calibration was repeated by adjusting the parameters and computing the R^2, NSE, RSR and PBIAS between observed and predicted data. Model validation was performed using the optimal calibrated parameters, and the predicted data were evaluated by calculating the values of R^2, NSE, RSR and PBIAS.

$$R^2 = \left[\frac{\sum_{i=1}^{n} \left(O_i - O_{avg}\right)\left(P_i - P_{avg}\right)}{\left[\sum_{i=1}^{n} \left(O_i - O_{avg}\right)^2 \sum_{i=1}^{n} \left(P_i - P_{avg}\right)^2\right]^{0.5}} \right]^2 \quad (1)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} \left(O_i - O_{avg}\right)^2} \right] \quad (2)$$

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^{n} (O_i - P_i)^2}}{\sqrt{\sum_{i=1}^{n} \left(O_i - O_{avg}\right)^2}} \quad (3)$$

$$PBIAS = \frac{\sum_{i=1}^{n} (\,O_i - P_i) * 100}{\sum_{i=1}^{n} (O_i)} \quad (4)$$

Where O_i is the ith observation for the constituent being evaluated; O_{avg} is the average of observations during the period of concern; P_i is the ith simulated value for the constituent being evaluated; P_{avg} is the average of simulations during the period of concern.

These four criteria of model performance reflect different aspect of goodness of fit. The R^2 value describes the variance in measured data explained by the model. Williams (2003) considered four levels of prediction accuracy based on the R^2 value: (1) a R^2 value between 0.5 and 0.65 indicates that more than 50% of the variance in measured data is explained by the model; (2) a R^2 value between 0.66 and 0.81 indicates approximately quantitative predictions; (3) a R^2 value between 0.82 and 0.9 reveals good predictions; and (4) a calibrated model having a R^2 value greater than 0.91 is considered to be excellent. In many studies on hydrological modeling, a model having a R^2 value greater than 0.5 (Santhi *et al.*, 2001; Chung *et al.*, 2002; Van Liew *et al.*, 2003; Green *et al.*, 2006) or 0.6 (Ramanarayanan *et al.*, 1997) is considered acceptable. The NSE is a normalized statistic indicating how well the observed and predicted data fit the 1:1 line (Nash and Sutcliffe, 1970); NSE values range between $-\infty$ and 1 with 1 being the optimal value. Moriasi *et al.* (2007) summarized the evaluation results of model performance from various studies and suggested that the NSE value greater than 0.5 indicates a satisfactory model performance on a monthly time step. Some studies even suggested a lower NSE value to determine if the model results were satisfactory. For example, a value of 0.4 (Green *et al.*, 2006; Green and van Griensven, 2008) or 0.3 (Chung *et al.*, 2002) was suggested to determine if the model results were satisfactory. The RSR is calculated as the ratio of root mean square error (RMSE) to standard deviation (SD) of measured data (Singh *et al.*, 2004). The lower the RMSE value, the better the model performance is. Although the RMSE is commonly used as an error index statistics, it could not be used to compare various constituents. Therefore, the RSR with a normalization factor (the SD of measured data) was preferable. The RSR values range from 0 to a large positive value and a RSR value less than 0.7 is regarded as a satisfactory model performance (Moriasi *et al.*, 2007). A similar criterion, called the ratio of prediction to deviation (RPD), is the reverse of RSR (Saeys *et al.*, 2005). Saeys *et al.* (2005) suggested that the calibration was not usable when a model had a RPD value below 1.5, which can be interpreted as that a model having a RSR value greater than 0.7 is unacceptable. The PBIAS value indicates the average tendency of the simulated data to be larger or smaller than the observed data (Gupta *et al.*, 1998). The positive value indicates model underestimation bias, while the negative value indicates model overestimation bias. Moriasi *et al.* (2007) suggested various ranges of model satisfactory for streamflow, sediment and nutrients, which were $\pm 15 \leq \text{PBIAS} \leq \pm 25$, $\pm 30 \leq \text{PBIAS} \leq \pm 55$ and $\pm 40 \leq \text{PBIAS} \leq \pm 70$, respectively. These various

DOI: 10.1002/hyp

ranges were consistent with the conclusions from Harmel and Smith (2007) that greater uncertainties were found in nutrient samples due to errors in streamflow measurement and sample collection, storage and analysis.

RESULTS AND DISCUSSION

Model performance for baseline scenario

Observed data and simulated results (Uncalib) without calibration at various subwatersheds (sub19, sub30, sub93 and sub176) are summarized in Table V. The model simulated the streamflow well at subwatersheds 19, 30 and 93 as indicated by acceptable NSE, R^2, RSR and PBIAS values. However, the simulated streamflow at subwatershed 176 was lower than the measured streamflow with the corresponding values of NSE and R^2 smaller than 0.5. The simulated TSS losses were lower than the measured TSS losses at subwatersheds 30 and 176, indicating the model simulated a lower sediment delivery. Due to the different spatial distribution of soil and land use in the upper subwatersheds (sub19 and sub30) and the lower subwatersheds (sub93 and sub176), the average monthly measured TN losses ranged from 3.43 to 5.10 kg/ha at the upper subwatersheds and from 0.78 to 1.30 kg/ha at the lower subwatersheds. We found that the model underestimated TN losses at the upper subwatershed (sub19) and overestimated TN losses at the lower subwatersheds (sub93 and sub176) in terms of positive and negative number of PBIAS, respectively. Although the simulated TN losses at subwatershed 19 were lower, the NSE, R^2, RSR and PBIAS values indicated a satisfactory model performance, and no further calibration was needed for this subwatershed. The difference in model performance on TN losses at the subwatershed scale indicated that the model should be calibrated with variable N-related SWAT parameters in order to better represent the complex nitrogen processes in a large watershed.

Single-site calibration

Calibration at subwatershed 176 (Calib1). Before calibration, the average monthly simulated flow at subwatershed 176 was slightly overestimated with unacceptable statistical values ($R^2 = 0.49$ and NSE = 0.46),

Table V. Model performance of uncalibration (Uncalib), single-site calibration (Calib1-4) and multi-site calibration (Calib_MS) for streamflow, total suspended sediment (TSS) and total nitrogen (TN) at subwatershed 19, 30, 93 and 176. (Note: Mea. Avg. denotes the measured monthly average, and Sim. Avg. denotes the simulated monthly average for calibration period. Those units of averages are cms for flow and kg/ha for TSS and TN)

		Flow				TSS				TN	
		sub19	sub30	sub93	sub176	sub30	sub176	sub19	sub30	sub93	sub176
Uncalib	Mea. Avg.	119.36	413.01	1435.52	3427.85	41.98	54.82	5.10	3.43	1.30	0.78
	Sim. Avg.	83.17	367.27	1472.85	3503.15	27.67	36.55	3.76	3.75	2.68	2.04
	NSE	0.54	0.70	0.58	0.46	0.44	0.36	0.59	0.41	−2.62	−4.00
	R^2	0.65	0.72	0.60	0.49	0.53	0.46	0.64	0.58	0.55	0.51
	RSR	0.68	0.55	0.65	0.74	0.75	0.80	0.64	0.77	1.90	2.24
	PBIAS	30.32	11.08	−2.60	−2.20	34.10	33.33	26.29	−9.26	−106.12	−160.80
Calib1	Sim. Avg.	83.06	366.39	1455.44	3444.37	30.52	42.00	1.25	1.28	0.99	0.85
	NSE	0.54	0.70	0.64	0.56	0.49	0.30	−0.09	0.10	0.53	0.46
	R^2	0.66	0.73	0.64	0.56	0.53	0.49	0.51	0.52	0.59	0.54
	RSR	0.68	0.54	0.60	0.66	0.71	0.84	1.04	0.95	0.69	0.73
	PBIAS	30.41	11.29	−1.39	−0.48	27.29	23.39	75.52	62.88	23.84	−8.32
Calib2	Sim. Avg.	83.18	367.31	1473.06	3503.67	32.51	51.58	1.50	1.52	1.15	0.94
	NSE	0.54	0.70	0.58	0.46	0.50	0.11	0.02	0.24	0.56	0.39
	R^2	0.65	0.72	0.60	0.49	0.52	0.43	0.57	0.57	0.61	0.56
	RSR	0.68	0.55	0.65	0.74	0.71	0.94	0.99	0.88	0.67	0.78
	PBIAS	30.31	11.07	−2.62	−2.21	22.57	5.92	70.68	55.87	11.72	−20.01
Calib3	Sim. Avg.	83.17	367.28	1472.89	3503.30	32.51	51.57	3.05	3.09	2.30	1.78
	NSE	0.54	0.70	0.58	0.46	0.50	0.11	0.50	0.54	−1.40	−2.63
	R^2	0.65	0.72	0.60	0.49	0.52	0.43	0.64	0.59	0.56	0.52
	RSR	0.68	0.55	0.65	0.74	0.71	0.94	0.71	0.68	1.55	1.90
	PBIAS	30.32	11.07	−2.60	−2.20	22.57	5.93	40.13	10.11	−76.75	−128.43
Calib4	Sim. Avg.	83.17	367.25	1472.70	3502.75	32.50	51.56	4.55	4.56	3.30	2.50
	NSE	0.54	0.70	0.58	0.46	0.50	0.11	0.59	−0.07	−6.14	−8.38
	R^2	0.65	0.72	0.60	0.49	0.52	0.43	0.62	0.56	0.49	0.45
	RSR	0.68	0.55	0.65	0.74	0.71	0.94	0.64	1.03	2.67	3.06
	PBIAS	30.32	11.08	−2.59	−2.19	22.58	5.94	10.82	−32.84	−153.63	−220.53
Calib_MS	Sim. Avg.	83.17	367.28	1465.86	3485.57	32.51	49.15	3.76	3.09	1.26	0.82
	NSE	0.54	0.70	0.61	0.50	0.50	0.17	0.59	0.54	0.55	0.50
	R^2	0.65	0.72	0.62	0.51	0.52	0.46	0.64	0.59	0.62	0.55
	RSR	0.68	0.55	0.63	0.71	0.71	0.91	0.64	0.68	0.67	0.71
	PBIAS	30.32	11.07	−2.11	−1.68	22.57	10.34	26.29	10.11	3.12	−5.14

DOI: 10.1002/hyp

and the simulated monthly TSS loss at subwatershed 176 (36.55 kg/ha) was underestimated compared to the measured one (54.82 kg/ha) as shown in Table V. The average monthly simulated TN loss at subwatershed 176 was much overestimated (Table V). Thus, we first adjusted the parameters which were the most sensitive to flow and TSS losses, and further adjusted sensitive parameters to TN losses. The CN was reduced by 30% for all land uses to reduce the simulated flow; and the exponent parameter for calculating the channel sediment routing (SPEXP) was increased from 1 (default) to 2 (the maximum value suggested in the SWAT manual, Neitsch *et al.*, 2002) to increase the simulated TSS losses. Many studies have found that the most sensitive parameter to simulated nitrogen (N) is the organic N enrichment ratio (ERORGN) (Green and van Griensven, 2008; Jha *et al.*, 2010; Meng *et al.*, 2010). ERORGN is the concentration of organic N transported with the sediment to the concentration of organic N in the soil surface layer. The greater the ERORGN value, the more organic N can be transported with sediment. Thus, to reduce simulated TN at the subwatershed 176, ERORGN was reduced to 0.3. When ERORGN was reduced to 0.3, the simulated TN at subwatershed 176 was reduced and became closer to the measured TN (Table VI).

As shown in Table VI (Calib1), the simulated streamflow at all four subwatersheds were acceptable in terms of all the statistical values after this calibration. The values of NSE, R^2, RSR and PBIAS ranged between 0.54 – 0.7, 0.56 – 0.73, 0.54– 0.68 and −0.48 – 30.41, respectively. The simulated monthly average TSS losses at subwatershed 176 increased from 36.55 kg/ha to 42 kg/ha with improved R^2 and PBIAS values. Meanwhile, the model performance of TSS at subwatershed 30 was also improved with an increased NSE value. The calibrated ERORGN (0.3) resulted in great improvements in simulated TN losses at subwatersheds 93 and 176, but degraded much of the simulated TN losses at subwatersheds 19 and 30. The main goal for adjusting ERORGN value was to reduce the overestimated TN at subwatershed 176. As these calibrated values were applied to the entire watershed, the TN losses at upper subwatersheds (sub19 and sub30) were reduced which resulted in further underestimation of TN at subwatershed 19. Applying common calibrated parameter values on subwatersheds with different characteristics resulted in variable model performances at subbwatershed levels. For example, NSE values at subwatersheds 93 and 176 were improved (0.53 and 0.46, respectively) after calibration, while NSE values at subwatersheds 19 and 30 were degraded (−0.09 and 0.1, respectively). Similarly, PBIAS values at subwatersheds 93 and 176 were much improved (from −106.12 to 23.84 and from −160.8 to −8.32, respectively), while PBIAS values at subwatersheds 19 and 30 were much degraded (from 26.29 to 75.52 and from −9.26 to 62.88, respectively).

Calibration at subwatershed 93 (Calib2). Before calibration, the model simulated streamflow well at subwatershed 93 with NSE and R^2 values greater than 0.5. Thus, no further calibration was needed for streamflow. Although there was no measured TSS data at subwatershed 93, we applied the calibrated SPEXP value (=2) as we did to improve the TSS simulation at subwatershed 176 in Calib1 scenario. It was found that when ERORGN value was reduced to 0.5 the simulated average TN became closer to the measured one. Four model performance indicators for TN simulation were improved. Especially, the NSE and PBIAS values were improved from −2.62 to 0.56 and from −106.12 to 11.72, respectively (Table V). Similar as the calibration 1, the calibrated ERORGN (0.5) resulted in great improvements in simulated TN losses at subwatersheds 93 and 176, but degraded the TN simulations at subwatersheds 19 and 30 compared to the uncalibrated scenario (Table V) because of the similar reasons as discussed above. The TN simulation at subwatershed 176 was still overestimated although the PBIAS value was improved (from −160.8 to −20.01).

Calibration at subwatershed 30 (Calib3). The simulated flow were acceptable at subwatershed 30 before the

Table VI. List of default and calibrated values of SWAT parameters used for single-site and multi-site calibrations

Output Parameters Short Description		Flow CN[1] Initial SCS runoff curve number	TSS SPEXP[2] Exponent coefficient for sediment routing	TN ERORGN[3] Organic nitrogen enrichment ratio
Uncalibration	Uncalib	range: 30–98	1	2.5–3
Single-Site Calibration	Calib1	−30%	2	0.3
	Calib2	Default	2	0.5
	Calib3	Default	2	2
	Calib4	Default	2	4
Multi-Site Calibration	Calib_MS	−30%: subwatersheds greater than 10 000 ha	2	Default: sub19 2: sub30 0.01: sub93 and 176

Note:
1. CN range: 0.7*default value – 1.3*default value. (Meng *et al.*, 2010)
2. SPEXP range: 1 – 2. (Neitsch *et al.*, 2002)
3. ERORGN range: 0 – 5. (Meng *et al.*, 2010)

model was calibrated, thus no further adjustment on streamflow-related parameter was needed. Before model calibration, both TSS losses at subwatersheds 176 and 30 were underestimated. Therefore, the SPEXP value was increased to 2 to improve the TSS simulation at subwatershed 30. For TN simulation, we did not need to decrease the ERORGN value by much compared to the one used in the previous single-site calibrations at the downstream subwatersheds (Calib 1 and Calib 2). The default ERORGN value used for the first simulation (Uncalib) was calculated for each storm event, and the value varies between 2.5 and 3. Since the TN losses were a little overestimated at subwatershed 30 (Table V), the ERORGN value was reduced to 2. When the ERORGN value was reduced to 2, TN simulations at subwatersheds 30, 93 and 176 were improved. However, subwatershed 19 had the NSE value decreased from 0.59 to 0.5 resulting in a greater underestimation in TN losses.

Calibration at subwatershed 19 (Calib4). Similar to the calibration procedure at subwatershed 30 (Calib3), we retained the default CN value and increased the SPEXP value to 2. When we reduced the ERORGN value during the first three calibrations, we noticed that the TN simulation at subwatershed 19 became worse in terms of a lower simulated average. Thus, In order to improve the TN performance at subwatershed 19, we had to increase the ERORGN value to 4 (Calib4). In that case (Calib4), simulated TN at subwatersheds 30, 93 and 176 were more overestimated which further degraded the model performance at those subwatersheds.

Multi-site calibration and validation

Many studies have indicated that as the size of watershed increases, there are greater uncertainty in SWAT simulations (Heathman and Larose, 2007; Thampi *et al.*, 2010). Therefore, variable values for different subwatersheds should be applied to capture the varied characteristics of a large heterogeneous watershed. Based on the calibration results from single-site calibrations, multi-site calibration was performed from upstream to downstream. As shown in Table VI, variable values for different subwatersheds were applied for selected parameters. First, the CN value for all subwatersheds greater than 10 000 hectare was reduced by 30%. Second, a constant value of SPEXP (2) was applied to the entire watershed in order to improve TSS simulations at both upstream and downstream subwatersheds. Third, variable ERORGN values were applied across the watershed based on the calibration results of the single-site calibration. We kept default value of ERORGN at subwatershed 19 and a value of two was used at subwatershed 30. After several ERORGN values were tested at the rest subwatersheds during the calibration process, the ERORGN value of 0.01 was found to be the most suitable one to improve TN simulations at subwatersheds 93 and 176. Because subwatersheds 19 and 30 are nested in subwatershed 93 and subwatershed 93 is nested in subwatershed 176, the ERORGN value of 0.01 was applied to the area excluding subwatersheds 19 and 30. The calibration results show that a single ERORGN value was not sufficient to represent the different enrichment ratio of organic nitrogen in a large watershed.

TN losses at the upstream gauges were greater than those at the downstream gauges. This may be due to the relatively greater cropland percentage and greater soil organic N in the upstream subwatersheds than in the downstream subwatersheds (Table I). In addition, in-stream processes may be different at different spatial scales. Thus, the model may need improvement in the in-stream processes to better represent large-scale complex nitrogen processes. Moreover, the percent of fine particles in sediment increases because the slower velocity of runoff promotes settlement of the larger sediment particles. The greater the content of fine particles in sediment, the more organic nitrogen is attached and thus the enrichment ratio rises.

The overall model performance was improved after multi-site calibration as shown in Table V. The multi-site calibration results (Calib_MS) for streamflow, TSS and TN are all satisfactory based on the statistical criteria, except for the TSS simulation at subwatershed 176. The validation results for streamflow showed the calibrated model performed well at the four subwatershed outlets (Table VII). The TSS losses at subwatershed 30 were slightly underestimated with values of NSE and R^2 of 0.28 and 0.32, respectively, while the TSS losses at subwatershed 176 were overestimated with an R^2 value of 0.61. Overall, the model performed well for TN

Table VII. Multi-site validation for monthly flow (cms), total suspended sediment (TSS, kg/ha) and total nitrogen (TN, kg/ha). (Note: Mea. Avg. denotes the measured monthly average, and Sim. Avg. denotes the simulated monthly average for validation period)

	Flow				TSS		TN			
	sub19	sub30	sub93	sub176	sub30	sub176	sub19	sub30	sub93	sub176
Mea. Avg.	89.58	366.62	1343.98	2901.18	27.86	27.87	3.63	2.71	1.18	0.62
Sim. Avg.	65.45	312.92	1352.09	3073.06	23.67	32.54	2.93	2.61	1.05	0.70
NSE	0.64	0.65	0.67	0.64	0.28	−0.18	0.73	0.48	0.60	0.52
R^2	0.73	0.67	0.68	0.65	0.32	0.61	0.76	0.63	0.63	0.57
RSR	0.60	0.60	0.57	0.60	0.85	1.09	0.52	0.72	0.64	0.69
PBIAS	26.94	14.65	−0.60	−5.93	15.02	−16.79	19.17	3.90	10.99	−12.85

DOI: 10.1002/hyp

simulations for all four subwatershed outlets during the validation period with NSE, R^2, RSR and PBIAS values falling within acceptable ranges.

It is generally accepted that the calibration of hydrological models should be treated as a multiple-criteria objective problem (Gupta *et al*., 1998). However, there have been no absolute criteria established for judging model performance. In most studies, the criteria may vary depending on the modeling time steps (annual, monthly or daily). It is suggested that criteria can be appropriately relaxing and tightening of the standard value on the monthly time step when the model is performed on the daily and annual time steps, respectively (Moriasi *et al*., 2007). Moreover, many studies on statistical evaluation of model performance usually described whether the criteria value is close to the ideal (minimum or maximum) criteria ranges instead of setting up a critical value for judging whether the model performance is good or poor (Beldring, 2002; Vazquez *et al*., 2008; White and Chaubey, 2005). In this study, we did not only evaluate the calibration results using the criteria suggested by other studies but also evaluated the improvement that could be achieved with multi-site calibration comparing with single-site calibration.

To summarize the model's performance on TN during calibration, NSE values ranged between −4 and 0.59 before calibration (Uncalib). With single-site calibration, NSE values ranged between −0.9 and 0.53 after calibration at subwatershed 176 (Calib1), between 0.02 and 0.56 after calibration at subwatershed 93 (Calib2), between −2.63 and 0.54 after calibration at subwatershed 30 (Calib3); and between −8.38 and 0.59 after calibration at subwatershed 19 (Calib4). However, with multi-site calibration, the model performance was much improved with NSE values greater than 0.5 for all subwatersheds (Table V). The R^2 values did not vary much before and after single-site and multi-site calibrations because the R^2 value describes the variance in measured data explained by the model. For the criterion of RSR, RSR values did not vary much before calibration (0.64 – 2.24) and after various single-site calibrations (0.64 – 3.60). However, with multi-site calibration, the model performance was much improved with RSR values from 0.64 to 0.71. For the criterion of PBIAS, only when model calibration was performed at multiple sites the PBIAS values were within acceptable ranges suggested by Moriasi *et al*. (2007). The improvement of multiple criteria values with multi-site calibration denotes that manual calibration on multiple sites with various sets of parameters could improve the model performance at different sizes of subwatersheds.

SUMMARY AND CONCLUSIONS

Hydrologic models are efficient ways to evaluate the impact of alternative conservation practice scenarios on water quality. Before using a model, calibration is needed. By calibrating a model, the characteristics within the watershed can be better represented and more reliable watershed responses can be generated. In this study, we evaluated model performance using two different calibration strategies: single-site and multi-site calibrations. Before calibration, the model performed well for streamflow at three out of the four subwatershed outlets. The simulated TSS losses at subwatersheds 30 and 176 were underestimated, and thus the SPEXP value was elevated to increase the sediment delivery. The various observed TN values at subwatersheds were reflected by their different soil and land use distribution, and in-stream nitrogen processes. The downstream subwatersheds (sub93 and sub176) have lower observed TN losses comparing with the upstream subwatersheds (sub19 and sub30) due to the relatively smaller croplands percentage in the downstream subwatersheds and the scale effect. Based on the calibration results of the single-site calibration at the most downstream subwatershed (sub176) for TN, the model performed poorly at the most upstream subwatershed (sub19). The results indicated a variable ERORGN value would improve SWAT model's performance. Based on the results from the single-site calibration, multi-site calibration was performed. The CN value was reduced by 30% for the larger subwatersheds (greater than 10 000 ha) in order to improve the simulated streamflow at subwatershed 176. A constant SPEXP value of two was applied to the entire watershed because the uncalibrated model tended to underestimate TSS losses at both upstream and downstream subwatersheds. For TN simulation, different ERORGN values were used for different subwatersheds because the uncalibrated model tended to underestimate TN losses at the upstream watersheds and overestimate TN losses at the downstream subwatersheds. After multi-site calibration, all four statistical criteria for TN simulation indicated a satisfactory model performance. It indicated that the process of organic N transported by surface runoff in SWAT would be better understood with multi-site calibration. However, further evaluation of how different N species responding to other N-related parameters is needed to get a full picture of nitrogen simulation in a large watershed. Compared to the single-site calibration, multi-site calibration could provide sounder model performance at different sizes of subwatersheds. For future SWAT model development, it is suggested that model needs improvement to better capture the complex hydrological and nutrient processes at different scales.

REFERENCES

Alexander RB, Smith RA, Schwarz GE, Boyer EW, Nolan JV, Brakebill JW. 2008. Differences in phosphorus and nitrogen delivery to the gulf of Mexico from the Mississippi river basin. *Environmental Science & Technology* **42**: 822–830 DOI 10.1021/es0716103.

Allen RG. 1986. A Penman for all seasons. *Journal of Irrigation and Drainage Engineering-Asce* **112**: 348–368.

Allen RG, Jensen ME, Wright JL, Burman RD. 1989. Operational estimates of reference evapotranspiration. *Agronomy Journal* **81**: 650–662.

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

Beldring S. 2002. Multi-criteria validation of a precipitation-runoff model, *Journal of Hydrology* **257**(1–4): 189–211.
Beven K, Binley A. 1992. The future of distributed models - model calibration and uncertainty prediction. *Hydrological Processes* **6**: 279–298 DOI 10.1002/hyp.3360060305.
Brown LC, Barnwell TO. 1987. The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: documentation and user manual. Env. Res. Laboratory. US EPA, EPA /600/3-87/007, Athens, GA. 189 pp.
Cao WZ, Bowden WB, Davie T, Fenemor A. 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. *Hydrological Processes* **20**: 1057–1073 DOI 10.1002/hyp.5933.
CENR. 2000. Integrated Assessment of Hypoxia in the Northern Gulf of Mexico; National Science and Technology Council Committee on Environment and Natural Resources: Washington, DC.
Chaplot V, Saleh A, Jaynes DB, Arnold J. 2004. Predicting water, sediment and NO3-N loads under scenarios of land-use and management practices in a flat watershed. *Water, Air, and Soil Pollution* **154**: 271–293.
Chaubey I, Chiang L, Gitau MW, Mohamed S. 2010. Effectiveness of best management practices in improving water quality in a pasture-dominated watershed. *Journal of Soil and Water Conservation* **65**: 424–437 DOI 10.2489/jswc.65.6.424.
Chung SW, Gassman PW, Gu R, Kanwar RS. 2002. Evaluation of epic for assessing tile flow and nitrogen losses for alternative agricultural management systems. *Transactions of ASAE* **45**: 1135–1146.
Chiang L, Chaubey I, Gitau MW, Arnold JG. 2010. Differentiating impacts of land use changes from pasture management in a CEAP watershed using the SWAT model. *Transactions of the ASABE* **53**: 1569–1584.
Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**: 475–497 DOI 10.1175/2007jamc1356.1
Engel B, Storm D, White M, Arnold J, Arabi M. 2007. A hydrologic/water quality model application protocol. *Journal of the American Water Resources Association* **43**: 1223–1236 DOI 10.1111/j.1752-1688.2007.00105.x
Gassman PW, Reyes MR, Green CH, Arnold JG. 2007. The soil and water assessment tool: Historical development, applications, and future research directions. *Transactions of the ASABE* **50**: 1211–1250.
Green CH, Tomer MD, Di Luzio M, Arnold JG. 2006. Hydrologic evaluation of The Soil and Water Assessment Tool for a large tile-drained watershed in Iowa. *Transactions of the ASABE* **49**(2): 413–422.
Green CH, van Griensven A. 2008. Autocalibration in hydrologic modeling: Using SWAT2005 in small-scale watersheds. *Environmental Modelling & Software* **23**: 422–434 DOI 10.1016/j.envsoft.2007.06.002.
Gupta HV, Sorooshian S, Yapo PO. 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research* **34**: 751–763.
Hamilton G. 1993. Soil survey of Coles County, Illinois. *Illinois Agricultural Experiment Station soil report* 138.
Harmel RD, Smith PK. 2007. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology* **337**: 326–336 DOI 10.1016/j.jhydrol.2007.01.043.
Heathman GC, Larose M. 2007. Influence of scale on SWAT model calibration for streamflow. In *MODSIM 2007 International Congress on Modelling and Simulation*, Oxley L, Kulasiri D (eds). Modelling and Simulation Society of Australia and New Zealand, Christchurch, NZ. December 2007; 2747–2753. Available at http://mssanz.org.au/MODSIM07/papers/49_s11/InfluenceOfScales11_Heathman_.pdf.
Jha M, Arnold JG, Gassman PW, Giorgi F, Gu RR. 2006. Climate change sensitivity assessment on Upper Mississippi River Basin streamflows using SWAT. *Journal of the American Water Resources Association* **42**: 997–1015.
Jha MK, Schilling KE, Gassman PW, Wolter CF. 2010. Targeting land-use change for nitrate-nitrogen load reductions in an agricultural watershed. *Journal of Soil and Water Conservation* **65**: 342–352 DOI 10.2489/jswc.65.6.342.
Kirsch K, Kirsch A, Arnold JG. 2002. Predicting sediment and phosphorus loads in the Rock River basin using SWAT. *Transactions of ASAE* **45**: 1757–1769.
Mehaffey M, Van Remortel R, Smith E, Bruins R. 2011. Developing a dataset to assess ecosystem services in the Midwest United States. *International Journal of Geographical Information Science* **25**: 681–695 DOI 10.1080/13658816.2010.497148.
Meng H, Sexton AM, Maddox MC, Sood A, Brown CW, Ferraro RR, Murtugudde R. 2010. Modeling Rappahannock River Basin using SWAT - Pilot for Chesapeake bay watershed. *Applied Engineering in Agriculture* **26**: 795–805.
Millennium Ecosystem Assessment (MA). 2004. http://www.millenniumassessment.org/.
Monteith JL. 1965. Evaporation and the environment. P. 205–234. The state and movement of water in living organisms. XIXth Symposium. Soc. For Exp. Biol., Swansea, Cambridge University Press.
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**: 885–900.
Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models: Part 1. A discussion of principles. *Journal of Hydrology* **10**(3): 282–290.
Neitsch SL, Arnold JG, Kiniry JR, Williams JR, King KW. 2002. Soil and Water Assessment Tool Theoretical Documentation: Version 2000. TWRI Report TR-191, Texas Water Resources Institute, College Station, Texas, 506 pp.
National Research Council (NRC). 2000. Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution. National Academy Press: Washington, DC.
Qi C, Grunwald S. 2005. GIS-based hydrologic modeling in the sandusky watershed using SWAT. *Transactions of ASAE* **48**: 169–180.
Ramanarayanan TS, Williams JR, Dugas WQ, Hauck LM, McFarland AMS. 1997. Using APEX to Identify Alternative Practices for Animal Waste Management: Part II. Model Application, ASAE Paper 97–2209 ASAE, St. Joseph, MI.
Runkel RL, Crawford CG, Cohn TA. 2004. Load Estimator (LOADEST): a FORTRAN program for estimating constituent loads in streams and rivers: U.S. *Geological Survey Techniques and Methods Book 4*, Chapter A5, 69 p. Reston, VA.
Saeys W, Mouazen AM, Ramon H. 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering* **91**(4): 393–402.
Santhi C, Arnold JG, Williams JR, Dugas WA, Srinivasan R, Hauck LM. 2001. Validation of tbe swat model on a large river basin with point and nonpoint sources. *Journal of the American Water Resources Association* **37**: 1169–1188.
Santhi C, Srinivasan R, Arnold JG, Williams JR. 2006. A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas. *Environmental Modelling & Software* **21**: 1141–1157 DOI 10.1016/j.envsoft.2005.05.013.
Soil Conservation Service (SCS). 1972. *Section 4: Hydrology in National Engineering Handbook*. Washington, DC.
Sheshukov AY, Daggupati P, Douglas-Mankin KR, Lee M-C. 2011. High spatial resolution soil data for watershed modeling: 1. Development of a SSURGO-ArcSWAT utility. *Journal of Natural and Environmental Sciences* **2**(2): 15–24.
Singh J, Knapp HV, Demissie M. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004–08. Champaign, III.: Illionois State Water Survey. Available at: www.sws.uiuc.edu/pubdoc/CR/.
Srivastava P, McVair JN, Johnson TE. 2006. Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed. *Journal of the American Water Resources Association* **42**: 545–563.
Stewart GR, Munster CL, Vietor DM, Arnold JG, McFarland AMS, White R, Provin T. 2006. Simulating water quality improvements in the Upper North Bosque River watershed due to phosphorus export through turfgrass sod. *Transactions of the ASABE* **49**: 357–366.
Thampi SG, Raneesh KY, Surya TV. 2010. Influence of scale on SWAT model calibration for streamflow in a river basin in the humid Tropics. *Water Resources Management* **24**: 4567–4578 DOI 10.1007/s11269.010.9676.y.
USDA, Economics, Statistics, and Market Information System (USDA, ESMIS). 1990. Available at: http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1560 (last access in February 2010).
Vache KB, Eilers JM, Santelmann MV. 2002. Water quality modeling of alternative agricultural scenarios in the US corn belt. *Journal of the American Water Resources Association* **38**: 773–787
Van Liew MW, Arnold JG, Garbrecht JD. 2003. Hydrologic simulation on agricultural watersheds: Choosing between two models. *Transactions of ASAE* **46**: 1539–1551.

Vazquez-Amabile G, Engel B. 2008. Fitting of time series models to forecast streamflow and groundwater using simulated data from SWAT. *J. Hydrol. Eng.*, **13**: 554–562 DOI 10.1061/(ASCE)1084-0699(2008)13:7(554).

Wagener T, Kollat J. 2007. Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling & Software* **22**: 1021–1033 DOI 10.1016/j.envsoft.2006.06.017.

Walker RD, Pope RA. 1983. Estimating your soil erosion losses with the universal soil loss equation (USLE). Urbana, Ill.: University of Illinois at Urbana-Champaign, College of Agriculture, Cooperative Extension Service: 1220.

White KL, Chaubey I. 2005. Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *Journal of the American Water Resources Association* **41**: 1077–1089.

Williams JR. 1995. Chapter 25: The EPIC model. P. 909–1000. In *Computer models of watershed hydrology*, VP Singh (ed). Water Resources Publications, Highlands Ranch, CO.

Williams P. 2003. Near-infrared Technology–Getting the Best Out of Light. PDK Grain. Nanaimo, Canada.

Zhang X, Srinivasan R, Van Liew M. 2008. Multi-site calibration of the SWAT model for hydrologic modeling. *Transactions of the ASABE* **51**: 2039–2049.