University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural
Research Service, Lincoln, Nebraska

2007

Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA Agricultural Research Service Watersheds

Michael W. Van Liew
Win Butler University, vanliew@windstream.net

Tamie L. Veith
USDA-ARS

David D. Bosch
USDA-ARS

Jeffrey G. Arnold
USDA-ARS, jeff.arnold@ars.usda.gov

Follow this and additional works at: https://digitalcommons.unl.edu/usdaarsfacpub

Part of the Agricultural Science Commons

Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA Agricultural Research Service Watersheds

Michael W. Van Liew, M.ASCE[1]; Tamie L. Veith[2]; David D. Bosch[3]; and Jeffrey G. Arnold[4]

Abstract: Recent interest in tracking environmental benefits of conservation practices on agricultural watersheds throughout the United States has led to the development of the U.S. Department of Agriculture's (USDA) Conservation Effects Assessment Project (CEAP). The purpose of CEAP is to assess environmental benefits derived from implementing various USDA conservation programs for cultivated, range, and irrigated lands. Watershed scale, hydrologic simulation models such as the Soil and Water Assessment Tool (SWAT) will be used to relate principal source areas of contaminants to transport paths and processes under a range in climatic, soils, topographic, and land use conditions on agricultural watersheds. To better understand SWAT's strengths and weaknesses in simulating streamflow for anticipated applications related to CEAP, we conducted a study to evaluate the model's performance under a range of climatic, topographic, soils, and land use conditions. Hydrologic responses were simulated on five USDA Agricultural Research Service watersheds that included Mahantango Creek Experimental Watershed in Pennsylvania and Reynolds Creek Experimental Watershed in Idaho in the northern part of the United States, and Little River Experimental Watershed in Georgia, Little Washita River Experimental Watershed in Oklahoma, and Walnut Gulch Experimental Watershed in Arizona in the south. Model simulations were performed on a total of 30 calibration and validation data sets that were obtained from a long record of multigauge climatic and streamflow data on each of the watersheds. A newly developed autocalibration tool for the SWAT model was employed to calibrate eleven parameters that govern surface and subsurface response for the three southern watersheds, and an additional five parameters that govern the accumulation of snow and snowmelt runoff processes for the two northern watersheds. Based on a comparison of measured versus simulated average annual streamflow, SWAT exhibits an element of robustness in estimating hydrologic responses across a range in topographic, soils, and land use conditions. Differences in model performance, however, are noticeable on a climatic basis in that SWAT will generally perform better on watersheds in more humid climates than in desert or semidesert climates. The model may therefore be better suited for CEAP investigations in wetter regions of the eastern part of the United States that are predominantly cultivated than the dryer regions of the West that are more characteristically rangeland.

DOI: 10.1061/(ASCE)1084-0699(2007)12:2(173)

CE Database subject headings: Calibration; Validation; Optimization; Watershed management; Hydrology; Simulation; Conservation.

Introduction

The need to evaluate environmental benefits of conservation practices on farms and ranches throughout the United States during the past few decades has led to the implementation of various policies and programs that are designed to protect the nation's soil

[1]Hydrologist, Montana DEQ, 1520 E. 6th Ave., Helena, MT 59620. E-mail: mvanliew@mt.gov

[2]Agricultural Engineer, USDA ARS PSWM, Building 3702, Curtin Rd., University Park, PA 16802. E-mail: tveith@psu.edu

[3]Civil Engineer, USDA ARS SEWRL, P.O. Box 946, Tifton, GA 31794. E-mail: dbosch@tifton.usda.gov

[4]Agricultural Engineer, USDA ARS GSWRL, 808 E. Blackland Rd., Temple, TX 76502. E-mail: jgarnold@spa.ars.usda.gov

and water resources. One such program that has been recently implemented is a multiagency effort referred to as the Conservation Effects Assessment Project (CEAP). Headed by the U.S. Dept. of Agriculture (USDA) Natural Resource Conservation Service, the focus of CEAP is to produce an assessment of environmental benefits derived from implementing various USDA conservation programs. Although it is well recognized that various conservation programs have been designed to protect range and agricultural lands from soil erosion, conserve water use, enhance water quality, and promote wildlife habitat restoration, the environmental benefits derived from these programs at a watershed scale have not been well quantified. Scientists involved in the CEAP effort will study and model USDA Agricultural Research Service (ARS) experimental watersheds to quantify the net cumulative effects of conservation practices within a watershed. One objective of these watershed assessments will be to quantify the net cumulative effects of conservation practices within a watershed. Quantitative relationships between implemented conservation practices and changes in water quality, soil quality, and water conservation will be developed.

One of the mathematical models that will be used to perform these watershed scale CEAP assessments is the Soil and Water

Assessment Tool (SWAT) (Arnold et al. 1998). This model simulates complex hydrologic processes on agricultural watersheds and is useful as an analytical tool for estimating the effects of conservation practices at various spatial and temporal scales on the watershed. It can also be used to evaluate total maximum daily load standards and select suitable land use and best management practice scenarios that help reduce damaging effects of storm-water runoff on water bodies and the landscape (Borah and Bera 2003). With regard to the nationwide CEAP effort, it is important to quantify and document how well SWAT performs across a variety of climatic, physiographic, and hydrologic conditions. Previous investigations by Arnold and Williams (1987), Wilcox et al. (1990), and Borah and Bera (2003) have demonstrated the importance of examining model performance under a range of such conditions. Arnold and Williams (1987) demonstrated that Simulator for Water Resources in Rural Basins (SWRRB), the precursor to SWAT, realistically simulated water and sediment yields for various physiographic and climatic conditions on 11 USDA ARS watersheds. Borah and Bera (2003) compiled simulation results from 17 SWAT applications for watersheds located throughout the United States. They reported that the model was found suitable for predicting yearly flow volumes, sediment and nutrient loads. They also reported that monthly predictions were generally good, except for months having extreme storm events and hydrologic conditions.

To date, very few modeling studies that employ SWAT have been undertaken to compare parameter values or simulation results obtained with the same calibration procedure for distinctly different watershed applications throughout the United States. A comparison of model results across distinct regions would provide valuable guidance to modelers in region-specific model parameterization and calibration. In particular, it would aid CEAP researchers to more efficiently simulate region-specific impacts of conservation practices on hydrologic responses for cropland, rangeland, and irrigated agriculture throughout the United States.

A recent enhancement of the AVSWAT 2003 version of the Soil and Water Assessment Tool consists of an autocalibration tool developed by Van Griensven and Bauwens (2003). The autocalibration procedure utilizes the shuffled complex evolution algorithm (SCE-UA) that allows for the calibration of model parameters based on the minimization of a single function (Duan et al. 1992). This global search algorithm consists of a controlled random search (Nelder and Mead 1965), a systematic evolution of points in the direction of global improvement, competitive evolution (Holland 1975), and the concept of complex shuffling. This automated approach to model calibration holds promise as a labor saving tool for practitioners who employ SWAT under a range of climatic, soils, topographic, and land use conditions on agricultural watersheds.

To better understand SWAT's strengths and weaknesses under conditions such as those anticipated for the CEAP effort, we conducted an investigation to assess the performance of the model on a variety of watersheds located throughout the United States. Although output from SWAT includes streamflow, sediment loads, and a host of water quality parameters, only the streamflow component was compared in this study. Model tests to evaluate output results from SWAT based on automated calibration were conducted on five USDA ARS experimental watersheds. Unique climatic, physiographic, and land management conditions characterize each of these five watersheds, and they were considered ideally suited for this study because of the availability of long term climatic and runoff data, extensive coverage of precipitation gauges, and the quality of data available for monitoring streamflow. Watershed locations included Mahantango Creek Experimental Watershed in Pennsylvania, Little Washita River Experimental Watershed in Oklahoma, Little River Experimental Watershed in Georgia, Walnut Gulch Experimental Watershed in Arizona, and Reynolds Creek Experimental Watershed in Idaho. The specific objectives of the study were to (1) identify SWAT's strengths and weaknesses in simulating streamflow responses at monthly and daily time scales for agricultural watersheds; and to (2) determine how climatic and physiographic conditions influence these model simulations.

Methods and Materials

Test Watersheds

A site map of the United States that shows the location of the five experimental watersheds is presented in Fig. 1. Each watershed was established in compliance with U.S. Senate Document 59 to depict downstream, off-site impacts of watershed practices. The watersheds selected for this study represent a range in climatic, physiographic, and land use characteristics that are suitable for testing the robustness of the SWAT model in simulating hydrologic responses.

Mahantango Creek Experimental Watershed in Pennsylvania

Subwatersheds FD-36 (0.4 km^2) and WE-38 (7 km^2) are subwatersheds of Mahantango Creek Experimental Watershed, a tributary of the Susquehanna River in central Pennsylvania. Located about 50 km northeast of Harrisburg, the Mahantango Watershed is typical of upland agricultural watersheds within the nonglaciated, folded, and faulted Appalachian Valley and Ridge Physiographic Province (Veith et al. 2005). Climate in the region is temperate and humid, with a long-term average annual precipitation of 1,100 mm. The watershed is characterized by shallow, fragipan soils in near-stream areas, and deep, well-drained soils in the uplands. Land use types consist of pasture (38%), forest (34%), mixed croplands (26%), and farmsteads (2%).

Little River Experimental Watershed in Georgia

The 334 km^2 Little River Experimental Watershed is located in south central Georgia just north of Tifton. Climate in the region is characterized as humid subtropical with long, warm summers and short, mild winters, with an average annual precipitation of approximately 1,167 mm based on data collected by USDA ARS from 1971 to 2000. The Little River Watershed landscape is dominated by a dense dendritic network of stream channels bordered by riparian forest wetlands and uplands areas devoted mostly to agricultural uses. The region has low topographic relief and is characterized by broad, flat alluvial floodplains, river terraces, and gently sloping uplands (Sheridan 1997). Soils on the watershed are predominantly sands and sandy loams with high infiltration rates. As surface soils are underlain by shallow, relatively impermeable subsurface horizons, deep seepage and recharge to regional groundwater systems are impeded (Sheridan 1997). Land use types include forest (65%), cropland (30%), rangeland and pasture (2%), wetland (2%), and miscellaneous (1%).
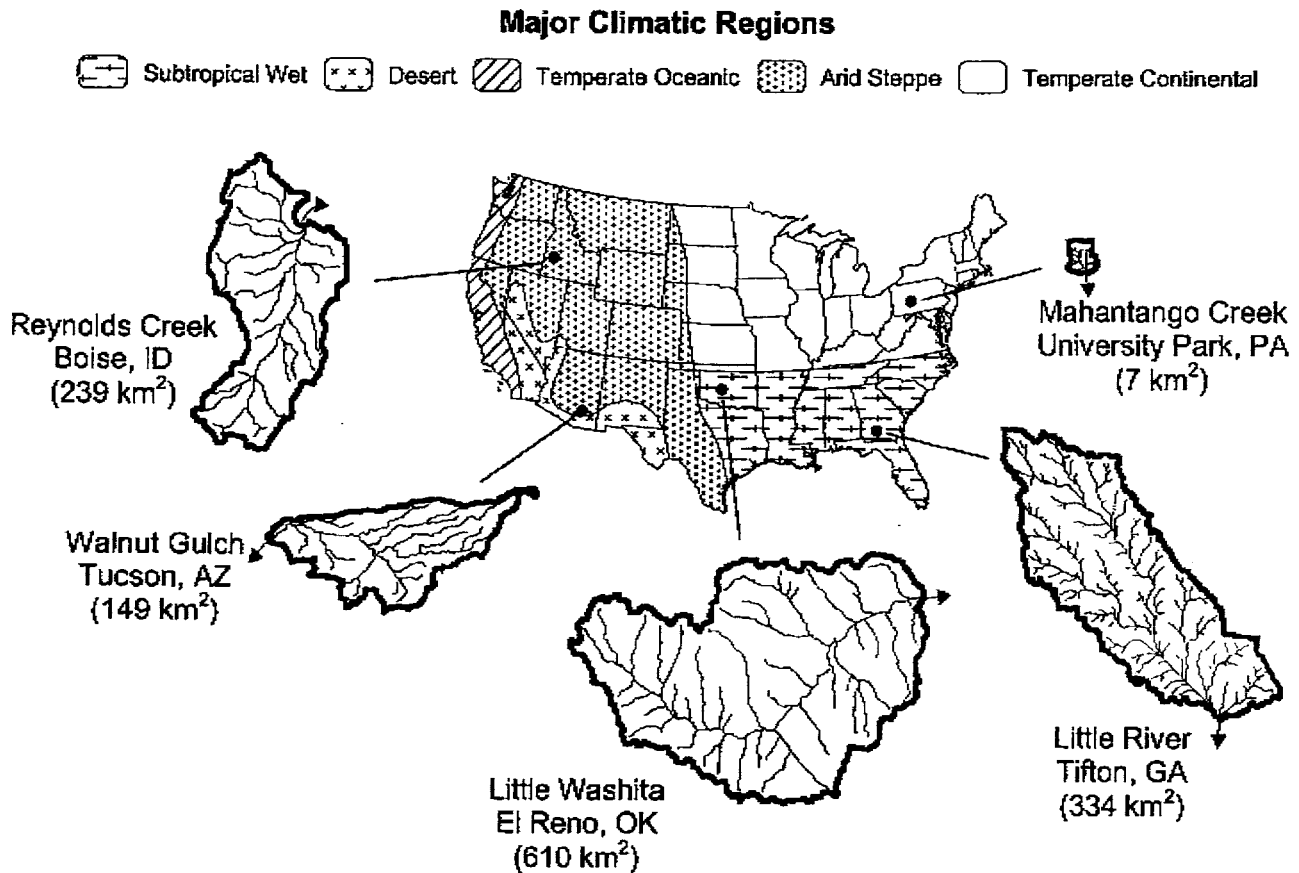
Major Climatic Regions

Subtropical Wet · Desert · Temperate Oceanic · Arid Steppe · Temperate Continental

Reynolds Creek
Boise, ID
(239 km²)

Mahantango Creek
University Park, PA
(7 km²)

Walnut Gulch
Tucson, AZ
(149 km²)

Little River
Tifton, GA
(334 km²)

Little Washita
El Reno, OK
(610 km²)

Fig. 1. Location of the five USDA ARS experimental watersheds

Little Washita River Experimental Watershed in Oklahoma

The 610 km² Little Washita River Experimental Watershed is located about 80 km southwest of Oklahoma City. The climate in the region is subhumid to semiarid, with an average annual precipitation of approximately 795 mm, based on data collected by USDA ARS from 1961 to 2000.

Topography of the watershed is characterized by gently to moderately rolling hills, and the soil types primarily consist of silt loams, loams, fine sandy loams, and sandy loams. Land use types include rangeland and pasture (66%), cropland (18%), forest (9%), and miscellaneous (7%—urban, abandoned oil fields, farmsteads, ponds) (Allen and Naney 1991). The watershed has numerous farm ponds located primarily in the lower portions of the watershed, and 45 NRCS flood-retarding structures (FRSs) constructed from 1969 to 1982. These FRSs control drainages on the Little Washita that range in size from 137 to 2,860 ha and consist of storage capacities ranging in size from 1.58×10^5 to 2.97×10^6 m³. These structures delay and reduce peak surface flows and modify subsurface flows. They also lead to small increases in average annual evaporation due to a larger percentage of the watershed existing as a free water surface.

Reynolds Creek Experimental Watershed in Idaho

The 239 km² Reynolds Creek Experimental Watershed is located approximately 80 km southwest of Boise and exhibits a considerable degree of spatial heterogeneity. The topography of the watershed ranges from a broad, flat alluvial valley to steep, rugged mountain slopes, with a range in elevation from 1,101 to 2,241 m (Seyfried et al. 2000). As a result of orographic effects, average annual precipitation increases almost fourfold on Reynolds Creek,

ranging from approximately 250 mm near the outlet to more than 1,100 mm along the mountain ridges at the upper end of the watershed. Perennial streamflow is generated at the highest elevations in the southern parts of Reynolds Creek where deep, late lying snowpacks are the source for most water (Seyfried et al. 2000). Although much of the watershed has steep, shallow, rocky soils, there are areas of deep, loamy soils that are rock free. Land cover on Reynolds Creek consists of rangeland and forest communities of sagebrush, greasewood, aspen, conifers (94%), and irrigated cropland (6%).

Walnut Gulch Experimental Watershed in Arizona

Located about 115 km southeast of Tucson, the Walnut Gulch Experimental Watershed comprises 149 km² and surrounds the historical western town of Tombstone. With a mean annual precipitation of about 325 mm, the climate of the region is considered semiarid. Walnut Gulch soils are generally well-drained, calcareous, gravelly loams with large percentages of rock and gravel at the surface (Gelderman 1970). Watershed relief ranges from 1,250 to 1,585 m, and the topography consists of low undulating hills to mountains. Streamflow that results from high intensity rainstorms primarily during the summer months is subject to substantial transmission losses in the channel beds and banks. Shrub canopy cover ranges from 30 to 40%, and grass canopy cover ranges from 10 to 80%. Land cover on Walnut Gulch is comprised of 83% rangeland, 12% forest, and 5% miscellaneous. Cattle grazing is the primary land use, with mining, limited urbanization, and recreation making up the remaining uses (Renard et al. 1993).

Model Description

The SWAT model was originally developed by the USDA ARS to predict the impact of land management practices on water, sediment, and agricultural chemical yields in large ungauged basins (Arnold et al. 1998). SWAT incorporates features of several ARS models and is a direct outgrowth of the SWRRB model (Williams et al. 1985). Specific models that contributed to the development of SWAT include Chemicals, Runoff, and Erosion from Agricultural Management Systems (CREAMS) (Knisel 1980), Groundwater Loading Effects on Agricultural Management Systems (GLEAMS) (Leonard et al. 1987), and Erosion-Productivity Impact Calculator (EPIC) (Williams et al. 1984). The SCS runoff curve number is used to estimate surface runoff from daily precipitation (USDA SCS 1986). The curve number is adjusted according to moisture conditions in the watershed (Arnold et al. 1993). SWAT can also be run on a subdaily time step basis using the Green and Ampt (1911) infiltration method. Other hydrologic processes simulated by the model include evapotranspiration, infiltration, percolation losses, channel transmission losses, channel routing, and surface, lateral, shallow aquifer, and deep aquifer flow (Arnold and Allen 1996). The runoff curve number option (Neitsch et al. 2002) is adopted in this study. Evapotranspiration (ET) in SWAT is computed using the Priestly and Taylor (1972), Penman-Monteith (Allen et al. 1989) or Hargreaves (1975) method. For all watersheds the Hargreaves (1975) method was used to estimate potential ET, as extraterrestrial radiation and air temperature were the only two measured variables required for computing daily potential ET values with this method. Channel routing in SWAT is accomplished by either the variable storage or Muskingum routing methods. For this study, the variable storage method was used to route flows in SWAT.

SWAT is a distributed parameter model that partitions a watershed into a number of subbasins. Each subbasin delineated within the model is simulated as a homogeneous area in terms of climatic conditions, but with additional subdivisions within each subbasin to represent various soils and land use types. Each of these subdivisions is referred to as a hydrologic response unit (HRU) and is assumed to be spatially uniform in terms of soils, land use, topographic, and climatic data.

AVSWAT 2003 was the version of the model used in this study, which incorporates an ArcView GIS interface for expediting model input and output (Di Luzio et al. 2002). The ArcView GIS raster based system consists of a modular structure that contains a tool for optimizing the definition and segmentation of a watershed and network based on topography. It also consists of a tool for defining the HRUs over the watershed and an integrated user-friendly interface. The GIS interface not only allows users to segment a watershed, but to import and format the supporting data necessary for the specific application and calibration of the model.

AVSWAT 2003 also includes a multiobjective, automated calibration procedure that was developed by Van Griensven and Bauwens (2003). The calibration procedure is based on a SCE-UA (Duan et al. 1992) and a single objective function. In a first step, the SCE-UA selects an initial population of parameters by random sampling throughout the feasible parameter space for "p" parameters to be optimized, based on given parameter ranges. The population is partitioned into several communities, each consisting of "$2p+1$" points. Each community is made to evolve based on a statistical "reproduction process" that uses the simplex method, an algorithm that evaluates the objective function in a systematic way with regard to the progress of the search in pre-

vious iterations (Nelder and Mead 1965). At periodic stages in the evolution, the entire population is shuffled and points are reassigned to communities to ensure information sharing. As the search progresses, the entire population tends to converge toward the neighborhood of global optimization, provided the initial population size is sufficiently large (Duan et al. 1992). The SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing, and land surface modeling, and has generally been found to be robust, effective, and efficient (Duan 2003).

In the optimization scheme developed for SWAT 2003, parameters in the model that affect hydrology or water quality can be changed in either a lumped (over the entire watershed) or distributed (for selected subbasins or HRUs) way. In addition, the parameters can be modified by replacement, by addition of an absolute change or by a multiplication of a relative change. In addition to weight assignments for output variables that can be made in multiobjective calibrations (e.g., 50% streamflow, 30% sediment, and 20% nutrients), the user can specify a particular objective function that is minimized. The objective function is an indicator of the deviation between a measured and a simulated series (Van Griensven and Bauwens 2003). An approach often selected as an objective function is the sum of squares of residuals method

$$ \text{SSQ} = (1/n) \sum_{i=1,n}^{n} (Q_{i,\text{obs}} - Q_{i,\text{sim}})^2 \qquad (1) $$

where SSQ=sum of squares of the residuals; n=number of pairs of measured and simulated variables; $Q_{i,\text{obs}}$=observed variable at a daily time scale; and $Q_{i,\text{sim}}$=simulated variable at a daily time scale.

Eq. (1) represents the classical mean square error method that aims at matching a simulated time series to a measured series.

Calibration Parameters

For this investigation, eleven calibration parameters that govern rainfall runoff processes in SWAT were selected for the three southern watersheds (Little River, Little Washita River, and Walnut Gulch), and an additional five parameters that govern the accumulation of snow and snowmelt runoff processes were selected for the two northern watersheds (Mahantango Creek and Reynolds Creek). Model parameters were grouped into three categories (Table 1), which were considered to predominantly govern surface, subsurface, and basin response. Table 1 lists parameters, descriptions, and units that were calibrated with the autocalibration tool for the five watersheds. The following is a brief description of each parameter.

Parameters Governing Surface Response
Calibration parameters governing the surface-water response in SWAT include the runoff curve number, the soil evaporation compensation factor, and the available soil water capacity. The runoff curve number (CN2) is used to compute runoff depth from total rainfall depth. It is a function of watershed properties that include soil type, land use and treatment, ground surface condition, and antecedent moisture condition. The soil evaporation compensation factor adjusts the depth distribution for evaporation from the soil to account for the effect of capillary action, crusting, and cracks. The available soil water capacity (SOL_AWC) is the volume of water that is available to plants if the soil was at field

Table 1. Parameters That Were Calibrated in SWAT

Parameter	Description	Units
Parameters governing surface water response		
CN2	SCS runoff curve number	none
ESCO	Soil evaporation compensation factor	none
SOL_AWC	Available soil water capacity	mm/mm
Parameters governing subsurface water response		
GW_REVAP	Groundwater "revap" coefficient	none
REVAPMN	Minimum threshold depth of water in the shallow aquifer for "revap" to occur	mm
GWQMN	Minium threshold depth of water in the shallow aquifer required for return flow to occur	mm
GW_DELAY	Groundwater delay	days
ALPHA_BF	Baseflow alpha factor or recession constant	days
RCHRG_DP	Deep aquifer percolation fraction	fraction
Parameters governing basin response		
SURLAG	Surface runoff lag time	days
CH_K2	Channel hydraulic conductivity	mm/h
TIMP	Snow pack temperature lag factor	none
SFTMP	Snowfall temperature	°C
SMTMP	Snowmelt base temperature	°C
SMFMX	Melt factor for snow on June 21	mm/°C day
SMFMN	Melt factor for snow on December 21	mm/°C day

capacity. It is estimated by determining the amount of water released between in situ field capacity and the permanent wilting point.

Parameters Governing Subsurface Response

Six calibration parameters govern the subsurface water response in SWAT. One of these parameters is referred to as the groundwater "revap" coefficient (GW_REVAP), which controls the amount of water that will move from the shallow aquifer to the root zone as a result of soil moisture depletion and the amount of direct groundwater uptake from deep-rooted trees and shrubs. Another parameter that governs the subsurface response is the threshold depth of water in the shallow aquifer for revap to occur (REVAPMN). Movement of water from the shallow aquifer to the root zone or to plants is allowed only if the depth of water in the shallow aquifer is equal to or greater than the minimum revap. A third parameter is the threshold depth of water in the shallow aquifer required for return flow to occur to the stream (GWQMN). Two other parameters that govern watershed response include the baseflow alpha factor and groundwater delay. The baseflow alpha factor (ALPHA_BF), or recession constant,

characterizes the groundwater recession curve. This factor approaches one for flat recessions and approaches zero for steep recessions. The groundwater delay (GW_DELAY) is the time required for water leaving the bottom of the root zone to reach the shallow aquifer. A sixth factor is the deep aquifer percolation fraction which governs the fraction of percolation from the root zone to the deep aquifer (RCHRG_DP).

Parameters Governing Basin Response

Two parameters that govern basin response in SWAT were calibrated in this study on each of the five watersheds. These included channel hydraulic conductivity (CH_K2) that governs the movement of water from the streambed to the subsurface for ephemeral or transient streams, and the surface runoff lag time (SURLAG) that provides a storage factor in the model to allow runoff to take longer than one day to reach a subbasin outlet. Five other basin parameters that govern snowfall and snowmelt in SWAT were calibrated on the two northern watersheds, Mahantango and Reynolds Creek. One parameter is the snowfall temperature (SFTMP) which is the mean air temperature at which precipitation is equally likely to be rain as snow or freezing rain. A second parameter is the snowmelt base temperature (SMTMP) that defines the snow pack temperature above which snowmelt will occur. SMFMX and SMFMN are melt factors for snow on June 21 and December 21, respectively, in the Northern Hemisphere that allow the rate of snowmelt to vary through the year as a function of snow pack density. A fifth parameter is the snow pack temperature lag factor (TIMP) that controls the impact of the current day's air temperature on the snow pack temperature.

Watershed Delineation and Model Calibration

Elevation, land use, and soil characteristics were obtained from GIS data layers for each of the five watersheds. The elevation layer was developed from USGS 30 m DEM quads for Little River, Little Washita River, and Reynolds Creek, and 10 m DEM quads for Walnut Gulch and Mahantango Creek. The land use layer for Mahantango Creek was obtained from a digitized aerial photograph, and from Landsat images for the other four watersheds. The soils layer was obtained from SURGO soils information for Mahantango Creek and Walnut Gulch, and the STATSGO database for the other three watersheds.

For this investigation, subbasins were delineated in each watershed to account for variations in rainfall based on the spatial distribution of precipitation gauges in each watershed and to account for the impact of the FRSs on hydrologic response in the Little Washita. A reservoir file was also created for each of the FRSs in the Little Washita to consider the effects of storage and release of water from rainstorms on that watershed. Data input for the FRSs was based on a previous study by Van Liew et al. (2003). The number of HRUs in the delineation of the respective watersheds was constrained by a threshold based on a land use and soil type covering an area of at least 5 and 20%, respectively, within any given subbasin. Although smaller percentage thresholds for land use and soil type could have been selected, the additional computational time that would have been required was not warranted. This constraint in the number of HRUs was the same one employed in a previous study involving autocalibration of SWAT on the Little River and Little Washita River Watersheds (Van Liew et al. 2005). Cursory testing on one of the nested subwatersheds within the Little Washita assuming a threshold based on a land use and soil type covering an area of at least 20 and 33%, respectively, within any given subbasin revealed no

Table 2. Number of Subbasins, Number of Hydrologic Response Units, and Drainage Areas for the Five USDA ARS Experimental Watersheds

Watershed	Subwatershed	Number of subbasins	Number of HRUs[a]	Area (km²)
Mahantango Cr.	FD-36	1	5	0.4
Mahantango Cr.	WE-38	7	44	7
Little R.	F	12	51	114
Little R.	B	40	161	330
Little Washita R.	526	22	138	160
Little Washita R.	550	73	486	600
Little Washita R.	522	66	413	538
Reynolds Cr.	Tolgate	16	47	55
Reynolds Cr.	Salmon	10	19	36
Reynolds Cr.	Outlet	60	162	239
Walnut Gl.	Flume 9	10	17	24
Walnut Gl.	Flume 2	43	82	114
Walnut Gl.	Flume 1	55	102	149

[a]Hydrologic response units.

appreciable differences in the simulation of streamflow response with SWAT in comparison to the 5% and 20% constraints used in this study. Because specific USDA CEAP investigations involve various sized watersheds, differences in the sizes and characteristics of land owner tracts, and particular water quality issues, it is anticipated that watershed projects will be delineated with various resolutions for land use and soil type thresholds to address the pertinent needs of each project. The flexibility that exists to delineate HRUs in the model will therefore enable practitioners to hone in on potential areas that are targeted for the implementation of best management practices.

Table 2 lists the respective number of subbasins and HRUs delineated for each of the simulated subwatersheds, along with the respective drainage areas. Although there have been minor changes in land use types over time on Little River and Little Washita River Watersheds, records are not available on either of these watersheds to accurately denote year to year changes that have occurred. Moreover, preliminary testing conducted with SWAT on the Little Washita River showed that changes in land use as indicated by available records resulted in only very minor changes in streamflow. Few changes in land use types have occurred on the other three ARS watersheds during the past few decades. For this study it was therefore assumed that the respective land use types on each of the watersheds remained constant for the period of record simulated. To assess SWAT's capability of simulating surface and subsurface hydrologic processes, a baseflow method developed by Arnold and Allen (1999) was used to divide total measured streamflow into surface flow and baseflow. These values were assumed to represent measured surface flow and baseflow for this study and were compared to surface flow and baseflow contributions computed, by the baseflow method, from total streamflow simulated by SWAT for each calibration period on the respective watersheds.

To account for spatial variability in topographic, soil, and land use factors between or among subwatersheds within a given watershed, parameters in SWAT were calibrated in a distributed fashion using the automated calibration procedure, where observed and simulated outputs were compared at the same outlet points of the watersheds. Therefore, with the completion of a given optimization, two sets of calibrated parameters were computed for the Mahantango Subwatersheds FD-36 and WE-38, two sets for Little River that corresponded to Subwatersheds F and B, two sets for

the Little Washita subwatersheds 526 and 550, three sets for the Walnut Gulch Subwatersheds Flumes 9, 2, and 1, and three sets for the Reynolds Creek Subwatersheds referred to as Tolgate, Salmon, and Outlet. With the exception of the runoff curve number and the surface runoff lag time, default values suggested by Van Griensven (2002) were selected as the initial lower and upper ranges for the respective model parameters. To maintain a range in the calibrated values of the curve number similar to those reported in the literature (USDA SCS 1986), the initial lower and upper values for this parameter were restricted from −10 to +10%, respectively. To properly account for the release of surface runoff from a subbasin to the main channel, the range in values of the surface runoff lag time was restricted from 0.5 to 10.0 days.

In addition to range restrictions placed on CN2 and SURLAG, preliminary testing showed that better agreement between measured and simulated streamflow responses was obtained for Mahantango Creek and Little River when the initial lower and upper range in the groundwater delay parameter (GW_DELAY) was restricted from 0 to 5 days. Preliminary testing also revealed that improved model performance was obtained on Walnut Gulch when initial values of the groundwater delay parameter (GW_DELAY) were restricted from 0 to 5 days and values of the baseflow alpha factor (ALPHA_BF) were restricted from 0.9 to 1.0. For this study, initial values of these parameters as mentioned herein were selected for autocalibration on the Mahantango, Little River, and Walnut Gulch Watersheds.

Evaluation Criteria

Six evaluation criteria were used to assess monthly and daily streamflow simulated by SWAT. The first three criteria were quantitative statistics that measured the agreement between simulated and observed values, and the second three criteria were a visual comparison of plots of simulated and observed values. The first evaluation criterion used was the percent bias (PBIAS), which is a measure of the average tendency of the simulated flows to be larger or smaller than their observed values. The optimal PBIAS value is 0.0; a positive value indicates a model bias toward underestimation, whereas a negative value indicates a bias toward overestimation (Gupta et al. 1999). PBIAS may be expressed as

$$\text{PBIAS} = \sum_{k=1,n}^{n} (Q_{k\text{obs}} - Q_{k\text{sim}})(100) \bigg/ \sum_{k=1,n}^{n} (Q_{k\text{obs}}) \qquad (2)$$

where PBIAS=deviation of streamflow discharge (expressed as a percent); $Q_{k\text{obs}}$=observed streamflow [m³ s⁻¹ (cm s)]; and $Q_{k\text{sim}}$=simulated streamflow (cm s).

Donigian et al. (1983) considered HSPF model performance "very good" if the absolute percent error is <10%, "good" if the error is between 10 and <15%, and "fair" if the error is between 15 and <25% for calibration and validation. Measurement errors associated with streamflow as recommended by Harmel et al. (2006) follow the same standard. This standard was therefore adopted for the PBIAS evaluation criterion used in this study, with PBIAS values ≥25% considered as unsatisfactory.

The second evaluation criterion was the model coefficient of efficiency [Nash Sutcliffe coefficient of efficiency (NSE) (Nash and Sutcliffe 1970)], which Sevat and Dezetter (1991) found to be the best objective function for reflecting the overall fit of a hydrograph. NSE expresses the fraction of the measured streamflow variance that is reproduced by the model.

$$NSE = 1 - \left[\sum_{k=1,n}^{n} (Q_{kobs} - Q_{ksim})^2 \Bigg/ \sum_{k=1,n}^{n} (Q_{kobs} - Q_{mean})^2 \right]$$

(3)

where NSE=Nash Sutcliffe coefficient of efficiency; and Q_{mean}=mean observed streamflow during the evaluation period (cm s).

NSE values were computed for both monthly and daily streamflow. Simulation results were considered to be good for values of NSE >0.75, whereas for values of NSE between 0.75 and 0.36, the simulation results are considered to be satisfactory (Motovilov et al. 1999). For this study NSE values <0.36 were considered to be unsatisfactory.

The third evaluation criterion compared the measured and simulated fraction of baseflow to total flow that was computed for each of the calibration periods for the respective subwatersheds. This fraction was estimated using the baseflow separation technique developed by Arnold and Allen (1999). The fourth criterion compared average monthly measured and simulated streamflow for the calibration period on each of the 12 subwatersheds. The fifth criterion compared simulated monthly and daily hydrographs to observed values. At the daily time scale, particular attention was given to the timing and magnitude of peak flows and the shape of recession curves. The sixth criterion compared measured versus simulated flow duration curves to determine how well the model predicted the range in magnitude of daily flows throughout the calibration or validation periods.

Results

Following autocalibration on each watershed, results of the parameter search and calibrated streamflow output were retrieved from files generated by the autocalibration tool. The parasolout.out file listed detailed output for each optimization loop and uncertainty outputs, sceparobj.out included parameter values for all simulation runs and global optimization criterion, bestpar.out listed the optimal parameter values that were calibrated, and the autocal.out files provided simulated values of streamflow for the respective calibration points for each subwatershed. As noted in simulation results, output variables obtained from the respective autocal.out files correspond to those listed in the output.rch file. Examination of parameter values listed in the sceparobj.out file demonstrated that for each parameter, the entire range in parameter values specified as initial default values was sampled. This check ensured that a global optimum parameter search had been conducted for each autocalibration performed.

Table 3 shows initial values that were input to the autocalibration tool in SWAT and the resultant values that were obtained through calibration for each of the subwatersheds. As noted in Table 3, values of the SCS runoff curve number CN2 and available soil water capacity SOL_AWC are expressed as a percent change from the default values. As these two parameters were calibrated for each HRU, the calibrated data set consists of a multitude of values for these parameters for each watershed. For brevity, a complete listing of these individual values of CN2 and SOL_AWC is not reported in the study. For illustrative purposes, however, default and calibrated values of SOL_AWC for each soil type on the Little River Watershed are presented in Table 4. Default and calibrated values of CN2 for the various land cover types on the Tifton soil are also shown in Table 4.

The wide range in parameter values reported in Tables 3 and 4

reflects both differences that exist within watersheds as well as among watersheds. In some cases autocalibration of a given parameter resulted in the selection of the lower bound for one watershed and the upper bound for another. No attempt was made to relate optimized values reported in the table to watershed characteristics, although such an analysis could prove to be helpful in identifying subtle hydrologic responses of the model to surface or subsurface watershed features.

Average annual values of precipitation as well as measured and simulated streamflow for the five watersheds are presented in Table 5. Table 5 also lists PBIAS (Criterion 1), NSE on a monthly and daily basis (Criterion 2), and the fraction of baseflow to total flow for the respective calibration periods (Criterion 3). Fig. 2 compares average monthly measured and simulated streamflow for the calibration periods (Criterion 4). Values reported in Fig. 2 represent 4–7 year averages of monthly streamflow, depending on the length of the respective calibration period selected for a given watershed. Figs. 3 and 4 illustrate the comparison of monthly and daily hydrographs (Criterion 5), and Fig. 5 illustrates the duration of daily flow curves for four of the watersheds (Criterion 6).

Mahantango Creek Experimental Watershed

SWAT's performance on Mahantango FD-36 was among the best of any of the subwatersheds simulated in the study, with NSE values of 0.84 and 0.69 on a monthly and daily basis for the calibration period, respectively. For most of the available period of record, very good agreement was obtained between measured and simulated daily flows. A PBIAS of 23.5% is primarily attributed to SWAT's underestimation of the lower range in flows on the subwatershed, as reflected by the comparison of measured (0.50) versus simulated (0.28) fraction of baseflow to total streamflow for the calibration period (Table 5).

On a monthly basis, model performance on Mahantango WE-38 was similar to that obtained on FD-36, as evidenced by the comparison of average monthly measured and simulated streamflow for the calibration period on the two watersheds (Fig. 2). Fig. 3 also shows very good agreement between measured and simulated monthly hydrographs at the monthly time scale. Even though the model calibration on WE-38 yielded good agreement between measured and simulated daily hydrograph peaks and time to peaks, there was moderate underestimation of the recession side of the hydrograph for many events, as illustrated in Fig. 4 for the period from January 1, 1998 to June 30, 1998.

To assess SWAT's performance under varying climatic conditions, the available climatological and streamflow record was divided into three periods: from 1975 to 1979 ("wet climatic condition"), from 1980 to 1988 ("dry climatic condition"), and from 1989 to 1993 ("average climatic condition"). Simulation of the daily hydrographs for the three validation periods on WE-38 generally followed the same trend as that of the calibration period. A PBIAS of 15.8% was considered fair for the calibration period, whereas values of 33.8, 43.6, and 33.1% were considered unsatisfactory for the 1975–1979 (wet climatic condition), 1980–1988 (dry climatic condition), and 1989–1993 (average climatic condition) validation periods, respectively.

SWAT simulated a baseflow to total streamflow fraction of 0.31 for the calibration period for WE-38 compared to a measured fraction of 0.56. These results are similar to those reported for FD-36. Moderate underestimation of approximately 80% of the simulated duration of daily flow curve for the calibration period on WE-38 (Fig. 5) reflects the dominating effect of the sum of squares of residuals optimization function [Eq. (1)] in minimizing

Table 3. Parameter Values Calibrated in SWAT Using the Autocalibration Tool

| Parameter | Initial default values | | Mahantango | | Little River | | Little Washita | | Reynolds Creek | | | Walnut Gulch | | |
	Lower bound	Upper bound	FD-36	WE-38	F	B	526	550	Tollgate	Salmon	Outlet	Flume 9	Flume 2	Flume 1
CN2 (%)	−10	10	−10	−10	−9.92	−9.83	−9	10	7.28	9.85	10	8.29	−8.73	−10
ESCO	0	1	0	0	0.495	0.714	0.17	1	0.3	0.793	0.391	0	0.404	0.165
SOL_AWC (%)	−50	50	50	50	45.3	44.1	3.8	−38.8	−10.3	35.9	45	49.9	50	−45.8
GW_REVAP	0.02	0.2	0.02	0.02	0.157	0.044	0.026	0.06	0.176	0.081	0.033	0.198	0.033	0.096
REVAPMN	0	500	500	11.7	390.1	128.4	124.2	500	500	199.5	406.7	471	35.8	200
GWQMN	0	5000	0	0	15.8	0	29.4	3,364	491	2,887	4,367	2,167	1,150	5,000
GW_DELAY	0	500	3	5	2	2	106	102	16	5	24	3	0	2
ALPHA_BF	0	1	0.864	0.911	0.475	0.458	0.64	0.72	0.288	0.906	0.551	0.9	0.945	0.997
RCHRG_DP	0	1	0.582	0	0.043	0.959	0.065	0.916	0.859	0.108	0.516	0.079	0.953	0.037
SURLAG	0.50	10	0.50	0.5	0.5	0.5	0.57	0.57	0.5	0.5	0.5	9.13	9.13	9.130
CH_K2	0	150	90.9	0	149.8	75.3	10.5	103	6.7	14.1	17.9	2.3	141.5	103.5
TIMP	0.01	1	0.01	0.01	1	1	1	1	0.99	0.99	0.99	1	1	1
SFTMP	0	5	0	0	1	1	1	1	0.39	0.39	0.39	1	1	1
SMTMP	0	5	0.51	0.51	0.5	0.5	0.5	0.5	0.03	0.03	0.03	0.5	0.5	0.5
SMFMX	0	10	0	0	4.5	4.5	4.5	4.5	9.83	9.83	9.83	4.5	4.5	4.5
SMFMN	0	10	7.39	7.39	4.5	4.5	4.5	4.5	9.23	9.23	9.23	4.5	4.5	4.5

Note: CN and SOL_AWC parameter values expressed as percent change from default values; GW_DELAY initial lower and upper bounds restricted to 0.0 and 5.0, respectively, for Walnut Gulch, Mahantango Creek, and Little River; and ALPHA_BF initial lower and upper bounds restricted to 0.9 and 1.0, respectively, for Walnut Gulch.

Table 4. Calibrated Values of SOL_AWC for Each Soil Type and CN2 for Various Land Cover Types on the Tifton Soil for the Little River Watershed

Soil type	Default value of SOL_AWC	Calibrated value of SOL_AWC	Land cover type on Tifton soil	Default value of CN2	Calibrated value of CN2
Tifton	217	313	Pasture	69	62.2
Osier	111	159	Forested wetlands	66	59.5
Pelham	186	271	Deciduous forest	66	59.5
Troup	143	211	Evergreen forest	55	49.6
			Mixed forest	60	54.1
			Agricultural crops	77	69.4

differences between measured and simulated peak flows at the expense of the majority of the flows in the remainder of the data set. Optimization in this manner therefore contributed to the model's bias in underestimating streamflows for each of the Mahantango data sets.

Little River Experimental Watershed

Of the five ARS experimental watersheds, SWAT performed best on Little River, based strictly on the PBIAS and NSE error statistics. For all of the calibration and validation data sets on the two subwatersheds, NSE values were considered good at the monthly time scale and adequate at the daily time scale (Table 5). PBIAS was considered very good for the calibration and validation data sets on Subwatershed F (4 and −6.3%), good for the calibration data set (−12.2%), and fair for the valibration data set (−19%) on Subwatershed B. Visual inspection of the average monthly streamflow, monthly and daily hydrographs, and duration of daily flow curves, however, provides a better understanding of model performance than that provided solely by the two error

Table 5. Average Annual Measured Precipitation and Streamflow and Simulated Streamflow Percent Bias (PBIAS) and Monthly and Daily Coefficient of Efficiency (NSE)

Run type[a]	Watershed name	Subwatershed	Time series	Measured precipitation (mm)	Measured streamflow (mm)	Simulated streamflow (mm)	PBIAS (%)	NSE Monthly	NSE Daily	Fraction of baseflow to total flow Meas	Fraction of baseflow to total flow Sim
cal	Mahantango Cr.	FD-36	1997–2000	1,021	392	300	23.5	0.84	0.69	0.50	0.28
cal	Mahantango Cr.	WE-38	1997–2000	1,021	361	304	15.8	0.88	0.46	0.56	0.31
val	Mahantango Cr.	WE-38	1975–1979	1,203	685	453	33.8	0.61	0.54		
val	Mahantango Cr.	WE-38	1980–1988	971	473	267	43.6	0.46	0.35		
val	Mahantango Cr.	WE-38	1989–1993	1,097	557	373	33.1	0.75	0.43		
cal	Little R.	F	1997–2002	1,124	310	298	4.0	0.83	0.64	0.55	0.67
val	Little R.	F	1972–1996	1,260	374	398	−6.3	0.88	0.66		
cal	Little R.	B	1997–2002	997	271	304	−12.2	0.90	0.71	0.52	0.66
val	Little R.	B	1972–1996	1,178	348	414	−19.0	0.89	0.68		
cat	Little Washita R.	526	1993–1999	852	153	148	2.9	0.68	0.54	0.70	0.84
val	Little Washita R.	526	1980–1985	833	103	118	−14.9	0.60	0.56		
cal	Little Washita R.	550	1993–1999	832	118	98	16.9	0.76	0.63	0.66	0.59
val	Little Washita R.	522	1964–1968	697	29	47	−61.2	−0.36	0.13		
val	Little Washita R.	522	1975–1979	765	65	95	−46.5	0.29	0.20		
val	Little Washita R.	522	1981–1985	888	97	143	−47.6	−0.11	0.29		
cal	Reynolds Cr.	Tolgate	1968–1972	676	283	208	26.5	0.64	0.53	0.76	0.73
val	Reynolds Cr.	Tolgate	1973–1975	689	272	190	30	0.71	0.47		
val	Reynolds Cr.	Tolgate	1976–1990	773	250	199	20.2	0.73	0.62		
cal	Reynolds Cr.	Salmon	1968–1972	464	84	76	9.4	0.52	0.51	0.67	0.62
val	Reynolds Cr.	Salmon	1973–1975	411	77	71	8.2	0.40	0.08		
val	Reynolds Cr.	Salmon	1976–1990	471	72	98	−35.3	0.21	−0.17		
cal	Reynolds Cr.	Outlet	1968–1972	485	90	94	−4.8	0.79	0.73	0.70	0.72
val	Reynolds Cr.	Outlet	1973–1975	434	86	84	2.7	0.74	0.50		
val	Reynolds Cr.	Outlet	1976–1990	423	74	91	−23.5	0.63	0.48		
cal	Walnut Gl.	Flume 9	1968–1972	329	9.3	11	−18.7	0.86	0.76	0.05	0.22
val	Walnut Gl.	Flume 9	1973–1982	312	6.7	13	−94.6	−2.29	−1.81		
cal	Walnut Gl.	Flume 2	1968–1972	309	3.9	4.4	−14.0	0.48	0.30	0.07	0.39
val	Walnut Gl.	Flume 2	1973–1982	345	3.2	7.1	−121.1	−0.62	−1.02		
cal	Walnut Gl.	Flume 1	1968–1972	309	3.1	6.0	−91.7	0.59	0.54	0.06	0.40
val	Walnut Gl.	Flume 1	1973–1982	299	2.5	6.3	−155.6	−2.50	−1.77		

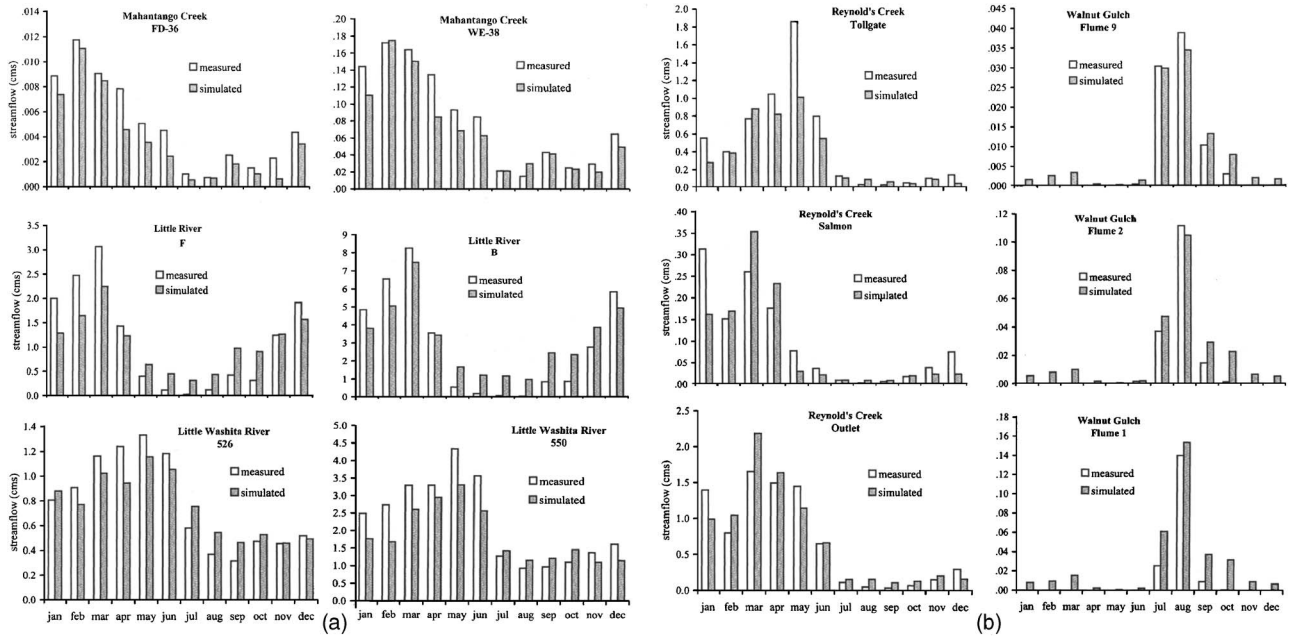[a]cal=calibration period; and val=validation period.

Fig. 2. Average monthly measured and simulated streamflow for the calibration periods on each of the subwatersheds
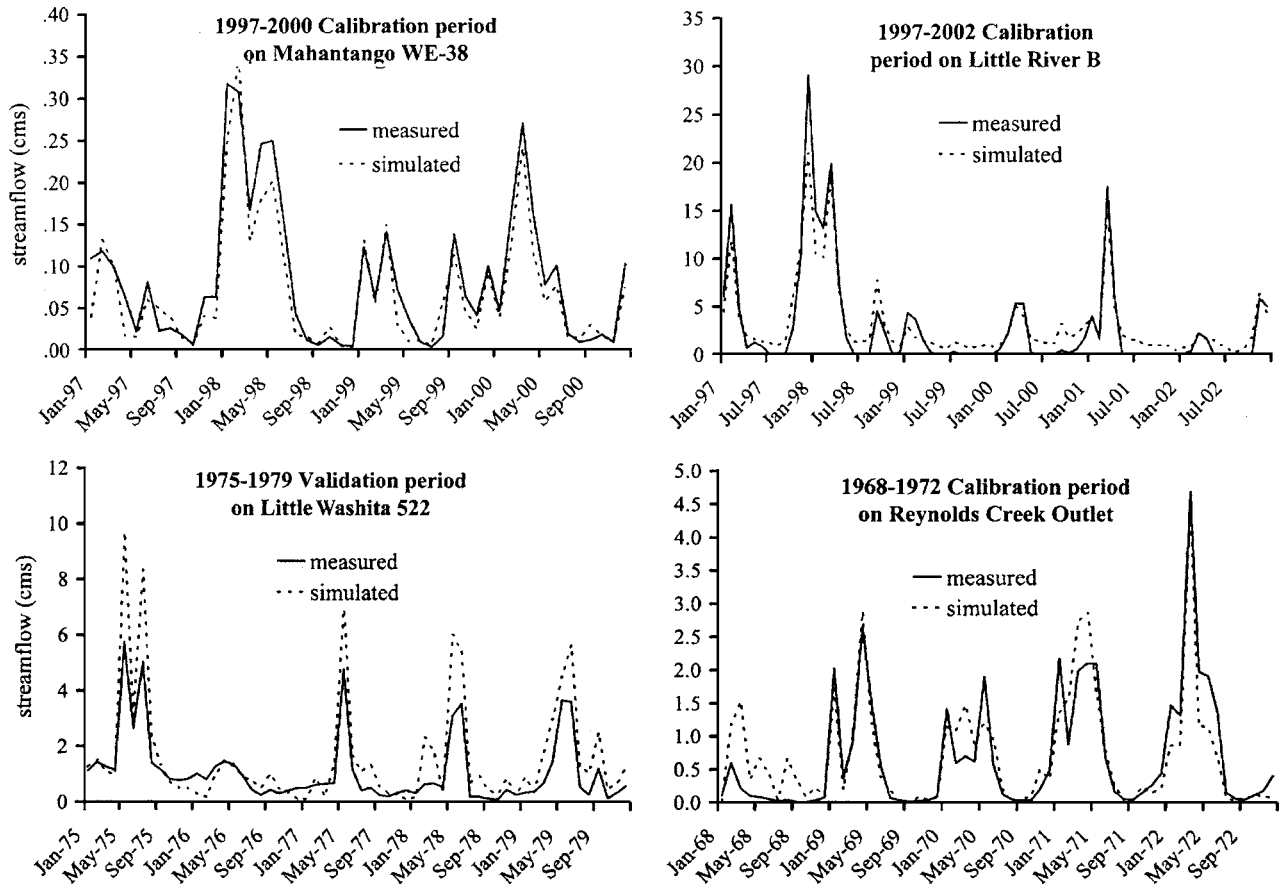

Fig. 3. Monthly measured and simulated hydrographs for selected subwatersheds
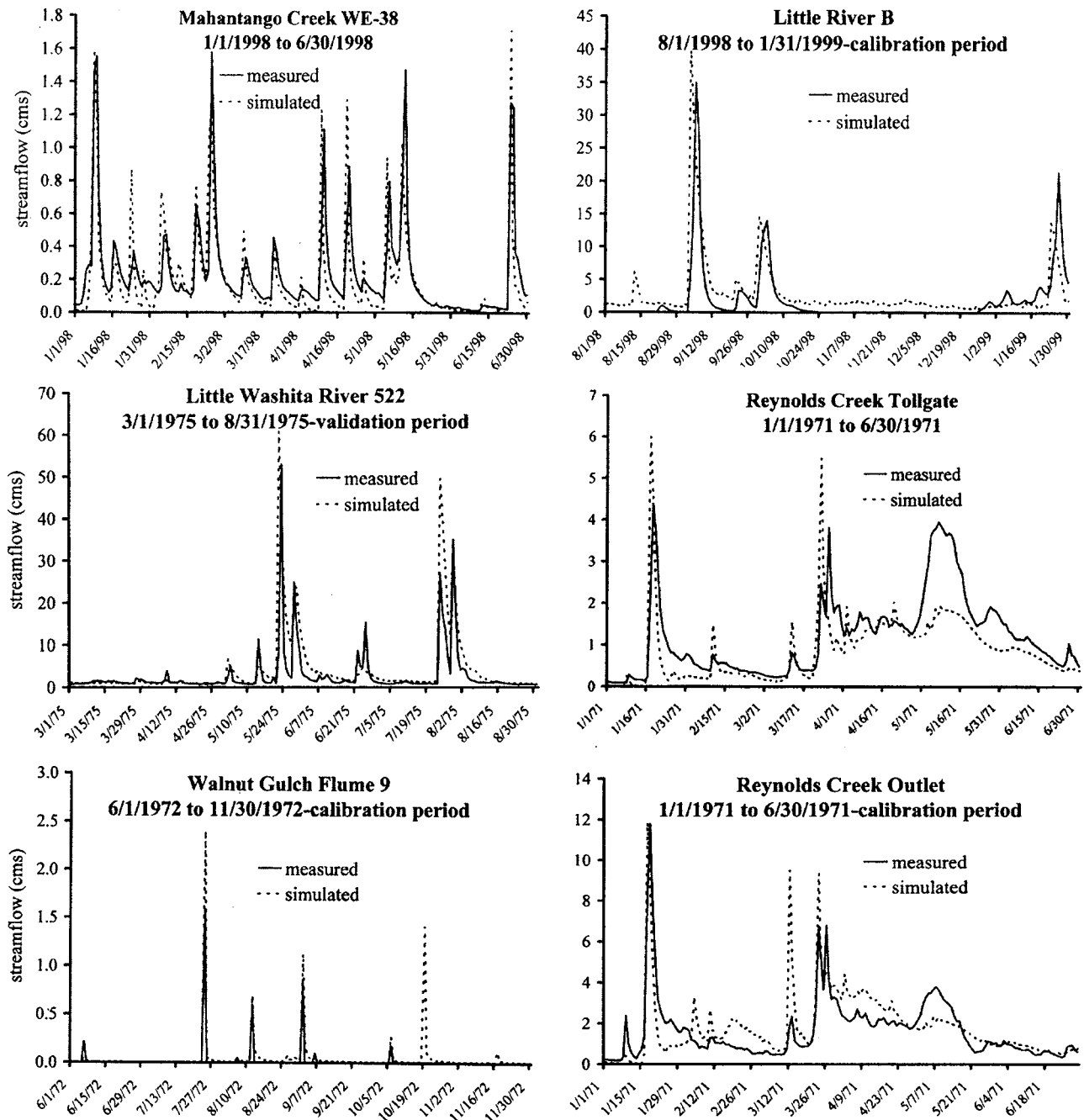
Fig. 4. Daily measured and simulated hydrographs for selected subwatersheds

statistics. Examination of Fig. 2 reveals that for both subwatersheds, SWAT underestimated high flows during the winter months and overestimated low flows during the summer months. This is the same kind of response that a precursor version of SWAT (the SWRRB model) exhibited in a study conducted by Arnold and Williams (1987) on Subwatershed B.

A plot of the calibration period on Subwatershed B at the monthly time scale shows significant overestimation of the baseflow portion of streamflow during periods of the record when negligible flow was reported (from September to November 1999; from June to August 2000; from August to November 2001; from June to September 2002) (Fig. 3). As shown in Fig. 4 for the period from October 15, 1997 to April 15, 1998 on Subwatershed B, this overestimation is further substantiated at the daily time

scale, with overestimation of baseflow that occurs between rainstorms on the watershed. Similar results shown in the figure were obtained throughout the calibration and validation periods on the two subwatersheds. Large discrepancies in the lower flow portion of the duration of daily flow curves are also evident in Fig. 5 for the calibration period on Subwatershed B. Results of the automated approach to calibration on Little River suggest one of two possibilities: either that (1) autocalibration may not be able to determine a suitable set of parameters to account for subsurface flow contributions to streamflow; or (2) a model refinement is needed in SWAT to adequately account for the wide variations in seasonal streamflow response that occurred on the watershed during the study period. Additional research is therefore needed to address this discrepancy between measured and simulated results.
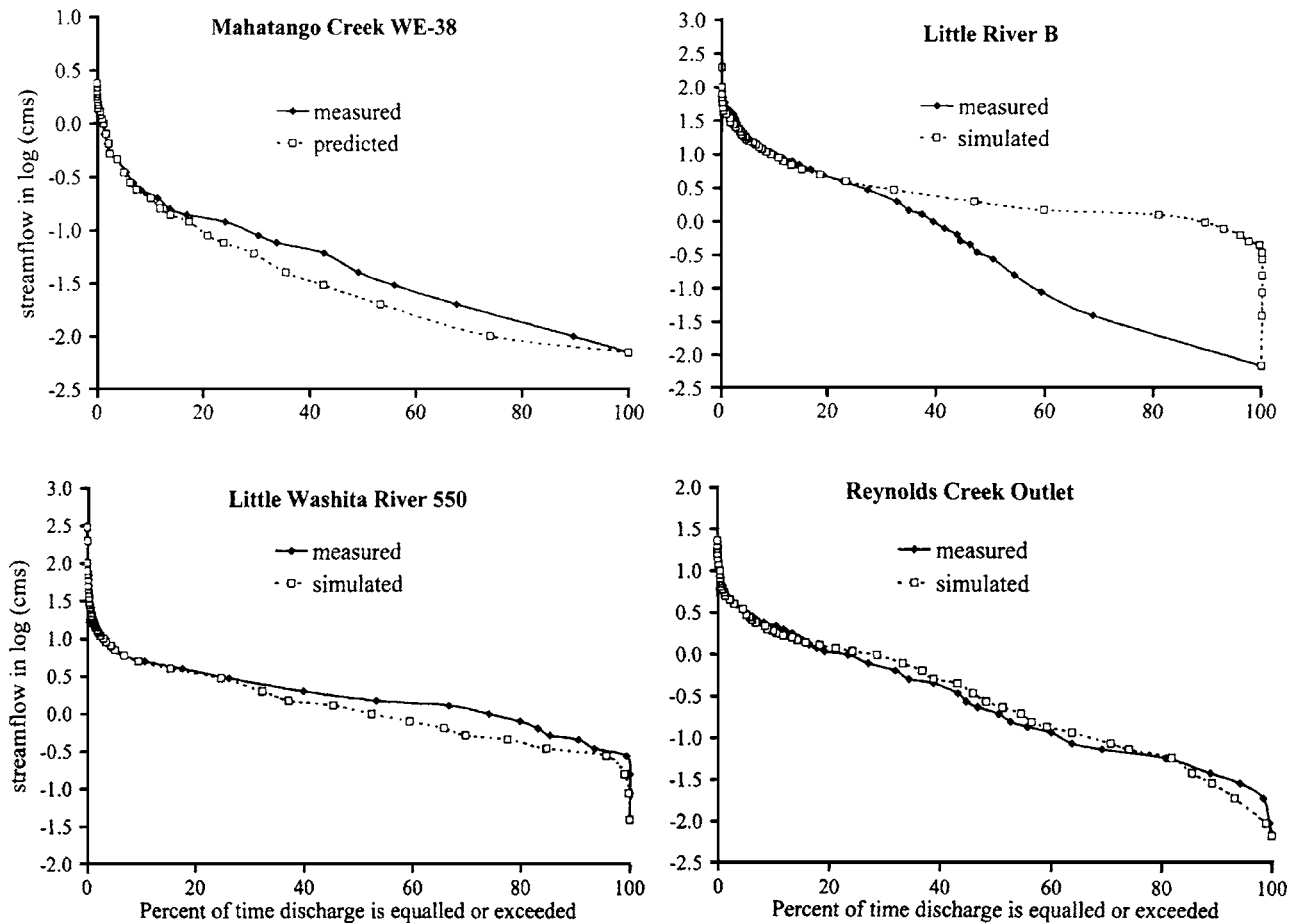
Fig. 5. Measured and simulated duration of daily flow curves for calibration periods on selected subwatersheds

Little Washita River Experimental Watershed

Because of a change in stream gauge location on the Little Washita, SWAT was calibrated on Subwatersheds 526 and 550, and then validated on Subwatersheds 526 and 522. The model performed well based on the NSE error statistic determined for the calibration period on Little Washita Subwatersheds 526 and 550. Results were considered good to adequate, with monthly values of NSE ranging from 0.68 to 0.76, and daily values ranging from 0.54 to 0.63 for Subwatersheds 526 and 550, respectively. Inconsistencies in model performance were noted in the comparison of average monthly streamflow for the calibration periods on the two subwatersheds, in that the model tended to underestimate flows during the winter and spring months and overestimate flows during the summer and fall (Fig. 2). However, SWAT accurately estimated the range in magnitude of flows as illustrated in Fig. 5 for the calibration period on Subwatershed 550, with very close agreement for the upper range of flows and some underestimation of the baseflow portion of the flows. The comparison of duration of daily flow curves obtained on Subwatershed 526 was similar to that reported for Subwatershed 550. A comparison of measured and simulated baseflow to total flow shows adequate agreement for both Subwatersheds 526 and 550. For Subwatershed 526, measured and simulated fractions were 0.70 and 0.84, respectively (Table 5). For Subwatershed 550, the respective measured and simulated fractions were 0.66 and 0.59.

Although simulation results obtained for the validation period on Subwatershed 526 were nearly as good as those obtained for the calibration period (Table 5), test results were considered un-acceptable based on values of PBIAS and NSE for each of the three validation periods on Subwatershed 522. A comparison of measured and simulated monthly streamflow for the validation period from 1975 to 1979 (average climatic condition) shows that the model tended to overestimate streamflow during periods of seasonally high rainfall (Fig. 3). At the daily time scale, a comparison of measured and simulated streamflow shows that SWAT tended to overestimate both the storm peaks as well as recessions following precipitation (Fig. 4). Results obtained from the other two validation periods displayed similar behavior. For the largest storm on record that occurred from October 19 to 20, 1983, SWAT's simulation of the storm peak was only 14% too high, but the overestimation of streamflow amount was nearly 110%. For each of the validation periods on Subwatershed 522, a comparison of duration of daily flow curves showed that SWAT underestimated lower flows and overestimated higher flows. This response follows the same pattern reported by Van Liew and Garbrecht (2003) for Subwatershed 522. It is not likely that differences in hydrologic response that were noted between the calibrated period on Subwatershed 550 and the validated period on Subwatershed 522 were due to the presence of the flood retarding structures on the watershed, as these spatial and temporal impacts were included in the project delineation of the Little Washita. Differences in hydrologic response were more likely attributed to subtle differences in land cover, channel, soil, and subsurface aquifer characteristics between the two subwatersheds.

Reynolds Creek Experimental Watershed

The combination of statistical and graphical criteria used in this study suggest that SWAT's overall performance on the mountainous rangeland Reynolds Creek Watershed was considered best among the five watersheds. Based on the PBIAS error statistic, results were considered very good for the 1968 to 1972 calibration and 1973 to 1975 validation periods on the Salmon and Outlet Subwatersheds, fair for the 1976 to 1990 validation periods on the Tolgate and Outlet Subwatersheds, and unsatisfactory for all other time series that were tested. Based on daily NSE values, model performance was considered adequate for all of the time series with the exception of the two validation periods on Salmon Creek. Comparison of average monthly measured and simulated streamflow for the calibration period on the three subwatersheds shows reasonably consistent results (Fig. 2). In general, model performance was not markedly different between the 1973 to 1975 and the 1976 to 1990 validation data sets, where the latter consisted of a lower precipitation gauge density than the former.

For the most part, Fig. 3 shows good agreement between monthly measured and simulated hydrographs for the calibration period from 1968 to 1972 on the Outlet Subwatershed. At the daily time scale, Fig. 4 indicates that SWAT performed well in estimating streamflow resulting from rainfall on frozen soils and/or snow covered conditions for the storm of January 17, 1971, but tended to overestimate flows resulting from storms that occurred on March 12 and 23, 1971 for the Tolgate and Outlet Subwatersheds. Fig. 4 also suggests that snowmelt processes occurring during April, May, and June of 1971 on the watershed were reasonably well represented by the model. As illustrated in Fig. 5, agreement between measured and simulated duration of daily flow curves for the calibration period on the Reynolds Creek Subwatershed Outlet was the best that was obtained among the five ARS watersheds. A comparison of the measured and simulated fraction of baseflow to total flow for each of the three Reynolds Creek Subwatersheds also showed the best agreement among the five watersheds (Table 5). For the calibration periods on the Tolgate, Salmon, and Outlet Subwatersheds, the measured fractions of baseflow to total flow were 0.76, 0.67, and 0.70, respectively. These compare very well to simulated fractions of 0.73, 0.62, and 0.72. Based on the statistical and graphical comparisons for Reynolds Creek, model simulations conducted in this study suggest that SWAT shows promising results for estimating streamflow on mountainous watersheds that are governed by snowmelt processes.

Walnut Gulch Experimental Watershed

SWAT's performance on the desert rangeland Walnut Gulch Experimental Watershed was considered only marginal compared to results obtained on the other four ARS watersheds. To improve estimations of the subsurface portion of the streamflow response, significant restrictions were placed on the initial lower and upper bounds for calibrating groundwater delay and the alpha baseflow factor in the model (Table 3). Based on average monthly comparisons for each of the three subwatersheds (Fig. 2), test results showed that SWAT simulated streamflow responses to summer rainstorms reasonably well during this time of year. For the nonsummer months of the year, however, SWAT consistently overestimated baseflow contributions to streamflow. This type of hydrologic response is also substantiated by the comparison of the measured and simulated fraction of baseflow to total flow for each of the three subwatersheds. For Flumes 9, 2, and 1, the measured

fractions of baseflow to total flow were only 0.05, 0.07, and 0.06, respectively. The simulated fractions for these three subwatersheds were 0.22, 0.39, and 0.40, respectively (Table 5).

Based on monthly and daily NSE values, simulation results were considered good for only the calibration period on Subwatershed Flume 9. Although not readily apparent in Fig. 4, SWAT tended to both over- and underestimate short duration, ephemeral flows resulting from high intensity rainstorms on Flume 9. Comparisons of daily hydrographs for other subwatersheds and time series on Walnut Gulch generally revealed that the model had trouble reproducing these very short-term events. These discrepancies were likely due to a combination of factors, such as the need for a shorter computational time step than one day and the fact that the model was unable to accurately account for antecedent soil moisture conditions immediately prior to events on the watershed, which in turn affected the partitioning of precipitation and runoff with the SCS curve number method.

Discussion

Assessing Model Strengths and Weaknesses

A significant advantage of the newly developed autocalibration tool in AVSWAT 2003 that became apparent in this study was that very little manual labor was required to prepare the model's input files that were necessary for calibrating the respective subwatersheds. Generally, only a few hours were required to manually prepare the input files to set up an autocalibration run; completion of the automated parameter search in most cases required at least one week of computational time. The primary disadvantage that was realized in using the autocalibration tool was in some cases, this approach did not lead to acceptable agreement between the measured and simulated water balance for a given calibration period, nor was it possible to represent the range in magnitude of daily flows. Use of the sum of squares of residuals optimization scheme [Eq. (1)] generally resulted in good matches between daily peak flows at the expense of lower flows in a given time series used for calibration. This was not surprising, as the sum of squares (SSQ) approach is biased toward the use of the NSE statistic. For hydrologic studies that involve low flow simulations, other optimization schemes may need to be employed to provide a better match over the range in magnitude of flows. The total mass balance controller (TMC) and the sum of squares of the difference of the measured and simulated variables after ranking (SSQR) represent alternative objective functions that can be minimized in the optimization scheme developed for AVSWAT 2003 (Van Griensven and Bauwens 2003). The former objective function minimizes the error on the model bias by calculating the deviation of the simulated mass from the measured bias. The latter function represents the fitting of the frequency distributions of the observed and simulated series. Previous work by Van Liew et al. (2005) showed that the SSQR objective function gave somewhat better results in matching measured and simulated duration of daily flow curves for the Little Washita River, but no appreciable improvement for the Little River. Nevertheless, further tests are warranted to investigate the impact of using the TMC, SSQR, or another objective function in SWAT to evaluate model performance under a range of climatic, soils, topographic, and land use conditions, especially where low flow simulations are critical for water quantity or water quality assessments.

The SSQ optimization scheme employed in this study was strictly a quantitative approach that involved minimizing the dif-

ference between measured and simulated values in the time series selected for calibration on a given subwatershed. Had a manual approach been utilized for model calibration, it could have involved using both quantitative and qualitative comparisons that consist of mass balance computations and graphical comparisons of hydrographs or flow frequency curves. A manual approach to calibration could have therefore been superior to the automated approach in providing a better match for the duration of daily flow curves at the expense of matching peak flows. Such a tradeoff may be advantageous in achieving improved model performance for watersheds like Mahantango Creek or Little River. Hence, the governing approach taken for watershed calibration depends on the particular needs that must be addressed for the modeling application (Van Liew et al. 2005).

Results obtained from this investigation reflect SWAT's strengths and weaknesses when applying the model to a range in climatic, physiographic, soil, and topographic conditions. Based on the error statistics, it would appear that the model performed reasonably well in simulating streamflow for both the calibration and validation periods on both subwatersheds of Little River. Examination of the duration of daily flow curves, however, revealed that the range in magnitude was not well represented. Moreover, inconsistencies in model performance as indicated by plots of the average monthly measured and simulated streamflow indicated that surface and subsurface contributions to total flow were not being properly accounted for on Little River. Simplifications that exist in the model that describe subsurface flow and channel transmission losses are likely reasons for discrepancies that exist between measured and simulated output (Van Liew and Garbrecht 2003). For the most part, SWAT performed well in simulating streamflow on the three subwatersheds of Reynolds Creek, based on not only the quantitative error statistics but the matches between measured and simulated hydrographs and duration of daily flow curves as well. This suggests that algorithms governing snow accumulation and snowmelt in the model provide reasonable computations for describing these processes.

SWAT's marginal performance on Walnut Gulch may in part be due to at least two reasons. First of all, a daily time step is likely too long for accurate simulations, due to rapid infiltration and runoff response that occurs on the watershed. Subdaily time step is an option available in SWAT, but it was not utilized for model simulations in this study. Second, a more sophisticated approach for computing ET could yield improved streamflow response, since ET is the dominant hydrologic process on this watershed. However, cursory testing of ET estimation on Walnut Gulch with the Hargreaves, Priestly Taylor, and Penman Monteith methods revealed that Penman Monteith provided only slight improvements in streamflow simulation. Although uncertainties in the measurement of streamflow on the Walnut Gulch Subwatersheds may be a possible cause for discrepancies between measured and simulated streamflow, this possibility is not likely in view of the quality of monitoring that has been in place there for more than four decades. Rating relationships for the Walnut Gulch supercritical flume that has been in use on the larger channels in the watershed are in excellent agreement with model tests conducted at Colorado State University in Ft. Collins, Colo., and the ARS Stillwater Hydraulic Laboratory in Stillwater, Okla. (Smith et al. 1982).

Selection of Initial Default Parameter Values

As noted in Table 3, there were several cases in which the optimal value of a parameter calibrated for a particular subwatershed cor-

responded to either the lower or upper bound initial default value for that parameter. Although the entire parameter space was sampled during a given autocalibration based on output obtained from the sceparobj.out file, these results suggested that model performance could be improved if one of the default boundaries was widened. To test this hypothesis, the range in initial default values for the curve number CN2 and the surface runoff lag time SURLAG was widened to ±50% and 0.0–10, respectively. These changes led to improved model performance on each of the five watersheds.

For Mahantango Creek, Little Washita, and Reynolds Creek, recalibrating the model with unrestricted ranges in CN2 and SURLAG resulted in marked improvement for nearly all of the validation periods on these three watersheds, with good or satisfactory results based on the PBIAS and NSE error statistics. However, percent departures in CN2 of −36% on Little Washita 526 to +34% on Reynolds Salmon, and a value of SURLAG as low as 0.14 on Mahantango FD-36 and WE-38 are unrealistic, and likely reflect model deficiencies related to HRU runoff partitioning and subbasin surface water storage. Although selection of a lower limit for SURLAG is somewhat arbitrary, we believe parameter values less than 0.5 do not properly account for the release of surface runoff from a subbasin to the main channel. In the case of parameters in SWAT such as groundwater delay and the alpha baseflow factor that govern groundwater response in the model, restricting the range in initial default values for these parameters may in some cases actually improve model performance. An understanding of subsurface aquifer properties and groundwater movement on a particular watershed is therefore tantamount to the selection of appropriate values that reflect watershed conditions.

Options for Improving Model Performance

Besides the possibility of adjusting the lower and upper initial bounds for parameters that are calibrated in SWAT, other options are available that can be implemented to enhance model performance. Fontaine et al. (2001) report additional options that they exercised to improve SWAT's performance using an automated approach to calibration for simulations on a mountainous watershed in South Dakota. These options included consultation with the model developer, the collection of additional field data, and a sensitivity analysis. Results of this study suggest that site specific climatological data including wind speed, solar radiation, and relative humidity data could provide improved estimates of ET for each of the watersheds, especially Walnut Gulch. The collection and analysis of soil moisture and/or groundwater flow data could help to provide better initial estimates of the parameters that govern the movement of water in these respective zones. Although not performed in this study, a sensitivity analysis could also provide another avenue for achieving improvements in model performance. Such an analysis would entail a ranking of the 16 parameters used in this study in addition to other parameters that govern streamflow response in the model, such as the Manning's roughness coefficient for the main channel (CH_N), the temperature lapse rate (TLAPS), and the plant uptake compensation factor (EPCO).

Overall Evaluation of Model Performance

To evaluate SWAT's performance for a range in climatic conditions, monthly and daily NSE values were plotted against average annual precipitation for each subwatershed time series. These
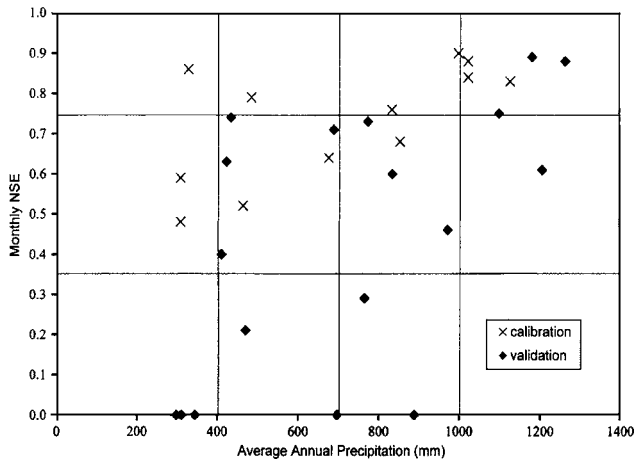
Fig. 6. Monthly NSE model performance as a function of average annual precipitation

Fig. 8. Comparison of average annual measured versus simulated streamflow for the calibration and validation data sets on the five watersheds

plots are shown in Figs. 6 and 7 for monthly and daily NSE values, respectively. To facilitate presentation of the data, NSE values less than zero are plotted as zero values. Subdivisions are also delineated in the figure, according to ranges in NSE values (>0.75—good; 0.36–0.75—satisfactory; <0.36—unsatisfactory) and precipitation amounts (<400 mm, 401–700 mm, 701–1000 mm, and >1,000 mm). As expected, Figs. 6 and 7 shows that the calibration data sets gave consistently higher NSE values than those that were obtained for the validation data sets. More importantly, results show that as average annual precipitation increases, scatter in the values of both monthly and daily NSE decreases and NSE values tend to increase. Based on monthly NSE values, 86, 25, 11, and 17% of the data sets are considered good that correspond to average annual precipitation amounts of >1000, 701–1000, 401–700, and <400 mm, respectively. Based on daily NSE values, 100, 63, 56, and 17% of the data sets are considered adequate that correspond to average annual precipitation amounts of >1,000, 701–1000, 401–700, and <400 mm, respectively. Results of this analysis suggest that SWAT will generally perform better on watersheds in more humid climates than in desert or semidesert climates where evapotranspiration is the dominant hydrologic process.
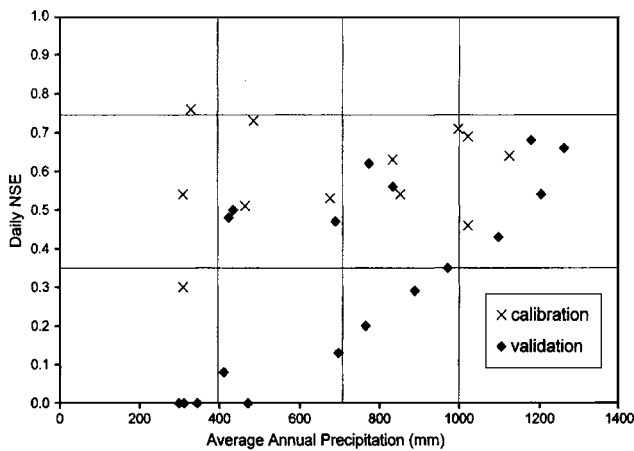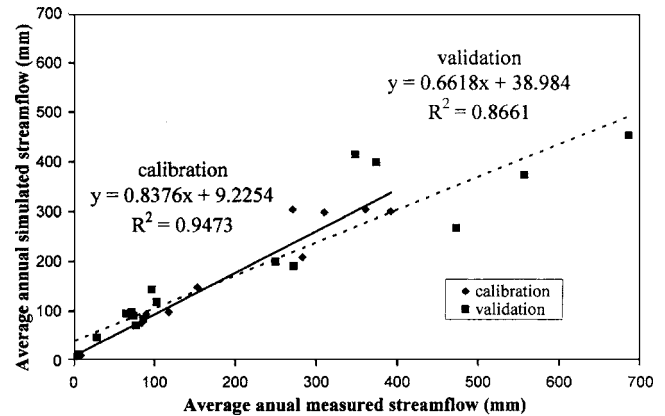
To evaluate SWAT's overall performance in simulating average annual streamflow based on the criteria established for autocalibration in this study, a linear regression analysis was employed to compare measured versus simulated values as reported in mm (Table 5). Results of the statistical analysis showed that for the calibration data sets, the coefficient of determination (r^2) was 0.95, with values of the slope of the regression line and y intercept equal to 0.84 and 9.2, respectively (Fig. 8). For the validation data sets, the coefficient of determination (r^2) was 0.87, with slope and y intercept values equal to 0.66 and 39.0, respectively. The Student's T-test was used to evaluate the hypotheses that the slope of the regression line equals unity and the intercept equals zero at the 95% confidence level. Test results showed that the slope was not significantly different from unity, nor was the intercept significantly different from zero for either of the two data sets. These results demonstrate that for the most part, SWAT exhibits an element of robustness in simulating average annual streamflow amounts under a range in climatic and physiographic conditions.

Conclusions and Recommendations

In this study an investigation was conducted to evaluate the performance of the 2003 version of AVSWAT to better understand the model's strengths and weaknesses in simulating streamflow for anticipated applications related to the Conservation Effects Assessment Project. Model performance was tested on five USDA ARS experimental watersheds representing a wide range in climatic, physiographic, and land use conditions that exist in the United States. A long record of multigauge climatic and streamflow data on each of the watersheds was used for model calibration and validation. Eleven parameters that govern surface and subsurface response in the model were calibrated on the three southern watersheds, and an additional five parameters that govern the accumulation of snow, snowmelt runoff, and the impact of frozen soils on runoff were calibrated on the two northern watersheds. The sum of squares of residuals method that matches a simulated time series to a measured series was used as the optimization scheme for model calibrations.

The newly developed autocalibration tool in AVSWAT (Van Griensven and Bauwens 2003) was used to calibrate streamflow response in the model. Test results demonstrate the uniqueness of

Fig. 7. Daily NSE model performance as a function of average annual precipitation

each calibrated parameter set and corresponding hydrologic response. Simulation results reflect the fact that even though model performance based on statistics such as percent bias and the coefficient of efficiency may be considered satisfactory, plots of average monthly hydrographs or duration of daily flow curves may not necessarily compare well for a particular watershed. Findings of this investigation therefore substantiate the importance of using both quantitative and qualitative criteria in evaluating model performance.

This study accentuates the need for exercising caution in the use of an automated approach to model calibration. Practitioners who employ the autocalibration tool in SWAT must exercise good engineering judgment in selecting appropriate initial lower and upper bounds for certain model parameters such as the curve number, surface runoff lag time, and groundwater delay factor so that distortions in model processes are minimized. Available field observations should therefore be employed whenever possible to define the range of initial default values for parameters calibrated in the model.

Results of this investigation also suggest that the sum of squares of residuals optimization scheme selected for autocalibration tends to work better on some watersheds (Little Washita and Reynolds Creek) than on others (Little River and Mahantango Creek). It is therefore recommended for investigations such as the Conservation Effects Assessment Project, an initial phase of model calibration using an automated approach should consist of comparing results from other optimization schemes such as the sum of squares of residuals after ranking or the total mass balance controller. Such a selection process should provide valuable insight in identifying important aspects of watershed response as it relates to solving specific problems in water quality, water availability, low flow investigations, or climate change.

As suggested in previous investigations by Gupta et al. (1999) and Van Liew et al. (2005), employing both automated and manual approaches to calibration holds promise for achieving optimal model performance. Although it was not the purpose of this study to compare or combine these two calibration approaches, such a strategy appears to have strong merit for overcoming some of the shortcomings in model simulation discussed in this paper. Combining manual and automated approaches also appears to be particularly useful in tailoring a model's simulation capabilities to specific project needs. To implement such a strategy, initial calibration runs could be conducted manually to verify that appropriate surface, lateral, and subsurface contributions to total streamflow and values of plant biomass are being simulated for a particular watershed. Next, the autocalibration tool could be employed to determine values of the designated parameter set and associated uncertainties, given a selected initial range in model parameters and desired optimization functions. Finally, manual adjustments to the optimized parameter values could be conducted to achieve improvements in mass balance or the representation of the range in magnitude of the output variables.

Results of this investigation suggest that for the most part, SWAT exhibits an element of robustness in simulating streamflow responses for a range in topographic, soils, and land use conditions. Differences in model performance, however, are noticeable on a climatic basis in that, except when explicitly calibrated, SWAT will generally perform better on watersheds in more humid climates than in desert or semidesert climates. Findings from this study therefore suggest that it may be better suited for CEAP investigations in wetter regions of the eastern part of the United States that are predominantly cultivated than the dryer regions of the West that are more characteristically rangeland. As future

CEAP assessments will entail procedures to estimate uncertainty in model predictions, studies are currently underway that employ SWAT's autocalibration tool to generate output for describing parameter and model uncertainties. Results of these analyses will be especially valuable in CEAP investigations that involve quantifying probability distribution functions related to downstream hydrologic response as a result of conservation practice implementation.

References

Allen, P. B., and Naney, J. W. (1991). "Hydrology of the Little Washita River Watershed, Oklahoma." *Rep. No. ARS-90*, U.S.D.A., Agricultural Research Service.

Allen, R. G., Jensen, M. E., Wright, J. L., and Burman, R. D. (1989). "Operational estimates of evapotranspiration." *Agron. J.*, 81(4), 650–662.

Arnold, J. G., and Allen, P. M. (1996). "Estimating hydrologic budgets for three Illinois watersheds." *J. Hydrol.*, 176, 57–77.

Arnold, J. G., and Allen, P. M. (1999). "Automated methods for estimating baseflow and groundwater recharge from stream flow records." *J. Am. Water Resour. Assoc.*, 35(2), 411–424.

Arnold, J. G., Allen, P. M., and Bernhardt, G. (1993). "A comprehensive surface-groundwater flow model." *J. Hydrol.*, 142, 47–69.

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

Arnold, J. G., and Williams, J. R. (1987). "Validation of SWRRB—simulator for water resources in rural basins." *J. Water Resour. Plann. Manage.*, 113(2), 243–256.

Borah, D. K., and Bera, M. (2003). "Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases." *Trans. ASAE*, 46(6), 1553–1566.

Di Luzio, M., Srinivasan, R., and Arnold, J. G. (2002). "Integration of watershed Tools and SWAT model into BASINS." *J. Am. Water Resour. Assoc.*, 38(4), 1127–1141.

Donigian, A. S., Imhoff, J. C., and Bicknell, B. R. (1983). "Predicting water quality resulting from agricultural nonpoint source pollution via simulation—HSPF." *Agricultural management and water quality*, Iowa State University Press, Ames, Iowa, 200–249.

Duan, Q. D. (2003). "Global optimization for watershed model calibration." *Calibration of watershed models*, Water Science Applied Series, Vol. 6, Q. Duan, et al., eds., AGU, Washington, D.C., 89–104.

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

Fontaine, T. A., Klassen, J. F., Cruickshank, T. S., and Hotchkiss, F. H. (2001). "Hydrological response to climate change in the Black Hills of South Dakota, USA." *Hydrol. Sci. J.*, 46(1), 27–40.

Gelderman, F. W. (1970). "Soil survey, Walnut Gulch Experimental Watershed, Arizona." *Special Rep.*, USDA SCS, USDA ARS, and Arizona Agricultural Experiment Station.

Green, W. H., and Ampt, G. A. (1911). "Studies on soil physics, 1. The flow of air and water through soils." *J. Agric. Sci.*, 4, 11–24.

Gupta, H. V., Sorooshian, S., and Yapo, P. O. (1999). "Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration." *J. Hydrol. Eng.*, 4(2), 135–143.

Hargreaves, G. H. (1975). "Moisture availability and crop production." *Trans. ASAE*, 18, 980–984.

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., and Arnold, J. G. (2006). "Cumulative uncertainty in measured streamflow and water quality data for small watersheds." *Trans. ASABE*, 49(3), 689–701.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, Mich.

Knisel, W. G. (1980). "CREAMS, a field scale model for chemicals,

runoff and erosion from agricultural management systems." *USDA Conservation Research Rep. No. 26*.

Leonard, R. A., Knisel, W. G., and Still, D. A. (1987). "GLEAMS: Groundwater loading effects on agricultural management systems." *Trans. ASAE*, 30(5), 1403–1428.

Motovilov, Y. G., Gottschalk, L., Engeland, K., and Rodhe, A. (1999). "Validation of distributed hydrological model against spatial observations." *Agric. Forest Meteorol.*, 98, 257–277.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models. Part 1—A discussion of principles." *J. Hydrol.*, 10(3), 282–290.

Neitsch, S. L., Arnold, A. G., Kiniry, J. R., Srinivasan, J. R., and Williams, J. R. (2002). *Soil and water assessment tool user's manual: Version 2000, GSWRL Rep. No. 02-02, BRC Rep. No. 02-06*, Texas Water Resources Institute TR-192, College Station, Tex.

Nelder, J. A., and Mead, R. A. (1965). "Simplex method for function minimization." *Comput. J.*, 7, 308–313.

Priestly, C. H. B., and Taylor, R. J. (1972). "On the assessment of surface heat flux and evaporation using large-scale measurements." *Mon. Weather Rev.*, 100(2), 81–92.

Renard, K. G., Lane, L. J., Simanton, J. R., Emmerich, W. E., Stone, J. J., Weltz, M. A., Goodrich, D. C., and Yakowitz, D. S. (1993). "Agricultural impacts in an arid environment: Walnut Gulch studies." *Hydrol. Sci. Technol.*, 9(1–4), 145–190.

Sevat, E., and Dezetter, A. (1991). "Selection of calibration of objective functions in the context of rainfall-runoff modeling in a Sudanese savannah area." *Hydrol. Sci. J.*, 36, 307–330.

Seyfried, M. S., Harris, R. C., Marks, D., and Jacob, B. (2000). "A geographic database for watershed research, Reynolds Creek Experimental Watershed, Idaho, USA." *USDA ARS Technical Bulletin No. NWRC-2000-3*.

Sheridan, J. M. (1997). "Rainfall-steamflow relations for coastal plain watersheds." *Trans. ASAE*, 13(3), 333–344.

Smith, R. E., Chery, D. L., Jr., Renard, K. G., and Gwinn, W. R. (1982). "Supercriticalflow flumes for measuring sediment-laden flow." *USDA-ARS Technical Bulletin No. 1655*, 70.

U.S. Dept. of Agriculture Soil Conservation Service (USDA SCS). (1986). "Urban hydrology for small watersheds." *Technical Release No. 55 (TR-55)*, Washington, D.C.

Van Griensven, A. (2002). "Developments towards integrated water quality modeling for river basins." *Publication No. 40*, Dept. of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel.

Van Griensven, A., and Bauwens, W. (2003). "Multiobjective autocalibration for semidistributed water quality models." *Water Resour. Res.*, 39(12), 1348.

Van Liew, M. W., Arnold, J. G., and Bosch, D. D. (2005). "Problems and potential of autocalibrating a hydrologic model." *Trans. ASAE*, 48(3), 1025–1040.

Van Liew, M. W., and Garbrecht, J. (2003). "Hydrologic simulation of the Little Washita River Experimental Watershed using SWAT." *J. Am. Water Resour. Assoc.*, 39(2), 413–426.

Van Liew, M. W., Garbrecht, J. D., and Arnold, J. G. (2003). "Simulation of the impacts of flood retarding structures on streamflow for a watershed in southwestern Oklahoma under dry, average, and wet climatic conditions." *J. Soil Water Conservat.*, 58(6), 340–348.

Veith, T. L., Sharpley, A. N., Weld, J. L., and Grurek, W. J. (2005). "Comparison of measured and simulated phosphorus losses with index site vulnerability." *Trans. ASAE*, 48(2), 557–565.

Wilcox, B. P., Rawls, W. J., Brkensiek, D. L., and Wight, J. R. (1990). "Predicting runoff from rangeland catchments: A comparison of two models." *Water Resour. Res.*, 26(10), 2401–2410.

Williams, J. R., Jones, C. A., and Dyke, P. T. (1984). "A modeling approach to determining the relationship between erosion and soil productivity." *Trans. ASAE*, 27(1), 129–144.

Williams, J. R., Nicks, A. D., and Arnold, J. G. (1985). "Simulator for water resources in rural basins." *J. Hydraul. Eng.*, 111(6), 970–986.