GIS-BASED HYDROLOGIC MODELING IN THE SANDUSKY WATERSHED USING SWAT

C. Qi, S. Grunwald

ABSTRACT. *Typically, simulation output generated by computer simulation models is calibrated and validated at the drainage outlet of a watershed. Shortcomings of such an approach include the validation at only one geographic location without explicitly accounting for variable hydrologic patterns within the watershed. Our objective was to conduct a spatially distributed calibration and validation of water flow using the Soil and Water Assessment Tool (SWAT). We simulated surface, groundwater, and total flow in the Sandusky watershed in Ohio, located in the Lake Erie watershed and the Great Lakes basin. The calibration and validation of simulated water flow was conducted concurrently at four subwatersheds and the drainage outlet of the Sandusky watershed. We used measured streamflow data at five U.S. Geological Survey (USGS) gauge stations from water years 1998 and 1999 for calibration and from water years 2000 and 2001 for validation. The surface water simulations at all monitoring stations were better than the groundwater simulations. The validation of total water flow showed a range in mean error of 0.03 to 4.00 $m^3\ s^{-1}$, a root mean square error of 0.06 to 2.56 $m^3\ s^{-1}$, correlation coefficients of 0.70 to 0.90, and Nash−Sutcliffe coefficients of 0.40 to 0.73. In conclusion, spatially distributed calibration and validation accounted for hydrologic patterns in the subwatersheds nested within the relatively large Sandusky watershed. Overall, simulations of water flow in the Sandusky watershed and subwatersheds were satisfactory except for winter rainfall−runoff events. This study showed the importance of spatially distributed calibration and validation.*

Keywords. *GIS, Hydrology, Modeling, Sandusky watershed, SWAT.*

Computer−based hydrological modeling has been carried out for more than three decades. The traditional deterministic hydrological models are of lumped conceptual (lumped parameter) type. Due to the nature of lumped models, all homogenous parameters and variables represent average values over the entire watershed. Whether the homogeneity can be justified depends on the scale and the uniform nature of many spatially variable parameters. Nevertheless, the description of the hydrological processes cannot be based directly on the equations supposed valid for the individual simulation units; hence, the equations become semi−empirical (Abbott and Refsgaard, 1996; Vieux, 2001). Variable−source hydrology is widely recognized by hydrologists to show saturated areas that expand and contract seasonally and in space (Dunne and Black, 1970; Dunne, 1978; Frankenberger et al., 1999). In order to predict the spatial pattern of variable source areas, a spatially distributed approach is needed. Freeze and Harlan (Freeze and Harlan, 1969) initially outlined the needs and concept for a distributed physically based watershed hydrologic modeling approach, which inspired the development of the first distributed hydrological model, SHE (Système Hydrologique Européen). Since then, a large number of distributed models have been developed (Abbott and Refsgaard, 1996).

Distributed, physically based watershed hydrologic models exploit the spatial and temporal characters of topography, land use, soils, meteorological variables, and precipitation in the simulation of hydrologic processes. In fact, one of the key differences between lumped and distributed models is the fundamentally different ways in which they are taking spatial variability into account. Hydrologic systems, e.g., watersheds, can be disaggregated in geographic space and their processes in time. Correspondingly, the analyses and remedial measures must also be described in a spatially distributed, temporally dependent way. Thus, the distributed, physically based watershed hydrologic models become indispensable. Singh (1995), in his classical textbook *Computer Models of Watershed Hydrology,* emphasized the need for spatially distributed modeling. Technology has facilitated the transformation of hydrologic modeling from lumped to distributed representations with the advent of high−performance computers and improved hydrologic and meteorological data collection systems (Abbott and Refsgaard, 1996; Vieux, 2001).

Watershed−scale hydrologic models have been developed to simulate the surface runoff from a watershed over various time periods. Michaud and Sorooshian (1994) compared the spatially distributed model KINEROS, a simple distributed model, and a simple lumped model (SCS method) in a midsized watershed using 24 severe rainfall−runoff events where the complex distributed model was more accurate than the simple distributed model. Short−term hydrologic models include TR−20 (USDA, 1965) and HEC−1 (Hydrologic

Article was submitted for review in April 2004; approved for publication by the Soil & Water of ASAE in December 2004. Presented at the 2003 ASAE Annual Meeting as Paper No. 032058.

This research was supported by the Florida Agricultural Experiment Station and approved for publication as Journal Series No. R−10157.

The authors are **Chen Qi,** Graduate Student, Department of Civil Engineering, and **Sabine Grunwald, ASAE Member,** Assistant Professor, Department of Soil and Water Science, University of Florida, Gainesville, Florida. **Corresponding author:** Sabine Grunwald, Department of Soil and Water Science, University of Florida, 2169 McCarty Hall, P.O. Box 110290, Gainesville, FL 32611; phone: 352−392−1951, ext. 204; fax: 352−392−3902; e−mail: sgrunwald@ifas.ufl.edu.

Engineering Center, 1998). Based on the HEC−1 model, the HEC−HMS (Hydrologic Modeling System) model is the latest release from the Hydrologic Engineering Center for hydrological analysis of a watershed (Scharffenberg, 2001). Some short−term hydrological models have been enhanced to simulate sediment yield and nutrients. Beasley et al. (1980) developed the distributed parameter, event−based ANSWERS (Areal Nonpoint−Source Watershed Environment Response Simulation) model in the late 1970s. Its current version, ANSWERS−2000, evolved into a continuous model capable of simulating nonpoint sources of nitrogen and phosphorus (Bouraoui and Dillaha, 1996).

AGNPS (AGricultural NonPoint Source) (Young et al., 1989) is another short−term, distributed−parameter model developed by the USDA Agricultural Research Service (ARS). AnnAGNPS (ANNualized AGricultural NonPoint Source) (Bingner and Theurer, 2003), HSPF (Hydrological Simulation Program− Fortran) (Bicknell et al., 1997), and SWAT (Soil and Water Assessment Tool) (Arnold et al., 1998) were developed for long−term continuous simulations (Borah, 2002). These models are based on similar mathematical routines to describe continuous processes such as movement of water and constituents in streams. AnnAGNPS emerged from AGNPS as a continuous model. HSPF, first developed in the early 1960s as the Stanford Watershed Model, simulates for extended periods of time the hydrologic cycle, and associated water quality, using continuous rainfall and other meteorological records. The model is able to represent the hydrologic regimes of a wide variety of streams and rivers with reasonable accuracy.

SWAT was developed at the USDA−ARS Grassland, Soil and Water Research Laboratory. The model is capable of predicting and assessing the impact of management on water, sediment, and nutrients at watershed scale. SWAT (Arnold et al., 1998) is a continuous−time, quasi−physically based, distributed model designed to simulate water, sediment, and agricultural chemical transport on a river−basin scale. SWAT was developed from earlier hydrologic and nutrient assessment models, including SWRRB, CREAMS, GLEAMS, EPIC, and ROTO (Borah and Bera, 2003), for predicting the long−term impact of different management scenarios and/or climate changes on watershed−scale hydrology and nonpoint−source pollution. It represents the hydrological cycle by interception, evapotranspiration, surface runoff, infiltration, soil percolation, lateral flow, groundwater flow, and channel routing processes.

Numerous SWAT applications have been documented stressing the success of long−term hydrologic simulations. Binger (1996) used SWAT in the Goodwin Creek watershed in Mississippi with a drainage area of 21.31 km^2. The model was calibrated for 10 years, and each of the subbasins was simulated separately. Only total runoff was compared with the observed total volume at each of the 14 streamflow stations. Although the spatial variability of the watershed characteristics was effectively simulated, simulations showed deviations from measured streamflow through time. Srinivasan et al. (1998) reported on the application of SWAT to the Richland and Chambers Creek watersheds in the Upper Trinity River basin in Texas. The model was calibrated and validated satisfactorily using flow data from two USGS stream gauge stations for the period 1965 to 1984; the calibration and validation were carried out separately for the two stations. Furthermore, the authors experienced difficulties simulating streamflow during spring and summer months when the spatial variability of rainfall was high. Santhi et al. (2001) calibrated and validated the SWAT model for flow in the Bosque River watershed in Texas at two streamflow gauge stations. Since one station was located upstream of the other one, which was close to the drainage outlet, the model accounted for nested hydrologic patterns within the Bosque River watershed.

In the study conducted by Vaché et al. (2002), SWAT was calibrated to evaluate the effect of management practices on surface water discharges, sediment, and nitrate in two watersheds in Iowa. Model calibration was performed at the drainage outlet of each respective watershed, which was relatively small in size.

Our objectives were to calibrate and validate hydrologic processes at the drainage outlet of the Sandusky watershed as well as for four subwatersheds concurrently using SWAT. We expected that spatially distributed calibration and validation of water flow would enhance the reliability of simulations for this relatively large watershed. Such a spatially distributed calibration and validation would improve our knowledge about internal hydrologic patterns nested within the whole watershed. Since watershed characteristics such as soils, geology, land use and land cover, topography, and the interactions among them influence hydrology, it is advantageous to control hydrologic patterns at both the subwatershed and watershed scales.

STUDY AREA

The Sandusky watershed is located within the Great Lakes basin (fig. 1), which drains into Lake Erie with a drainage area at Fremont of 3,240 km^2. In this study, we focused on four subwatersheds: Honey Creek (388.2 km^2), Rock Creek (90.3 km^2), Tymochtee (607.4 km^2), and Bucyrus (223.8 km^2). Bedrock underlying the Sandusky watershed is primarily Silurian dolostone and Devonian limestone. In the eastern portion of the watershed, Devonian shale and Mississippian sandstone are present. Surface features of the Sandusky watershed reflect the effects of the Wisconsinian glaciers, which retreated approximately 13,000 years ago. This resulted in two physiographic regions in the watershed: the Lake Plains in the northern portion, and the Till Plains in the central and southern portions. The landscape of the Lake Plains is an extremely flat plain of fine clay sediments, formed by wave action of glacial meltwater lakes that preceded Lake Erie. The Till Plains consists of flat to gently rolling plains with heavy till soils. Most of the relief within the Till Plains is located in three end moraines that lie in an east−west orientation. The majority of the Till Plaines consists of flatter, ground moraines that lie between the end moraines. Besides glacial till, lacustrine sediments and alluvial deposits occur along the drainage system of the Sandusky River.

SOILS

Dominant soils are Hapludalfs, Ochraqualfs, Fragiaqualfs, Medisaprists, Fluvaquents, and Argiaquolls (USDA Natural Resource Conservation Service, Ohio). Table 1 summarizes the types of soil in the Bucyrus, Honey Creek, Rock Creek, and Tymochtee subwatersheds within the Sandusky watershed. The Blount−Pewamo−Glynwood asso−

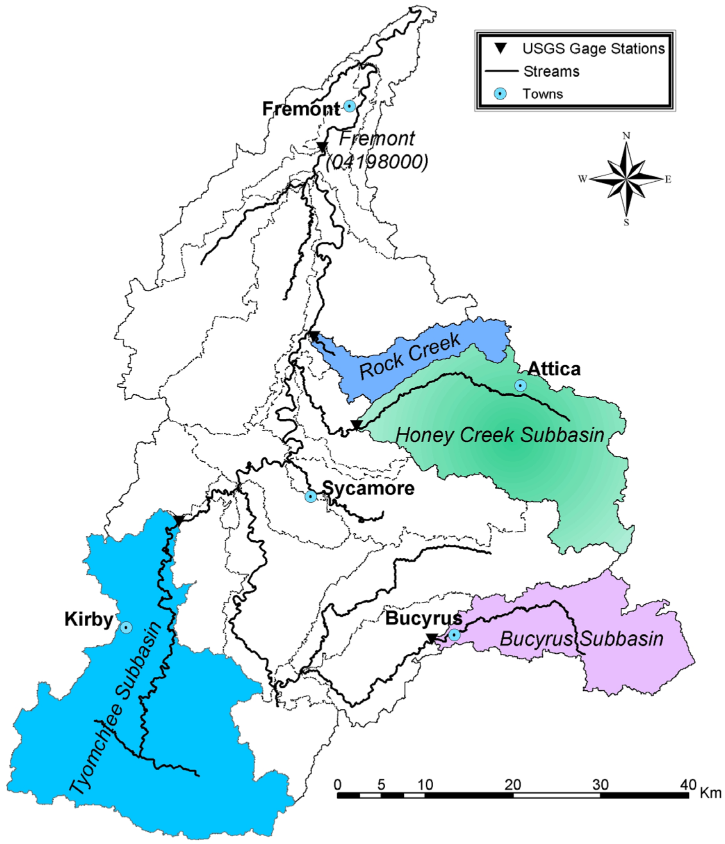

Figure 1. Subbasins within the Sandusky watershed. The Fremont station represents the drainage outlet of the Sandusky watershed.

Table 1. Soil types (%) in the Sandusky watershed and four major subwatersheds.

Soil Associations	Sandusky Watershed	Subwatersheds: Bucyrus	Honey	Rock	Tymochtee
Lenawee−Colwood−Lenawee Variant (OH001)	0.5	0.0	4.5	0.0	0.0
Milford−Del Rey−Shinrock (OH004)	7.4	0.0	0.0	0.0	11.4
Hoytville−Nappanee−Blount (OH006)	6.9	0.0	0.0	0.0	0.0
Latty−Fulton−Nappanee (OH009)	0.6	0.0	0.0	0.0	3.4
Paulding−Latty−Fulton (OH010)	2.7	0.0	0.0	0.0	10.3
Lenawee−Del Rey−Kibbie (OH011)	2.0	0.0	2.9	0.0	0.0
Milford−Luray−Tiro (OH012)	1.6	5.0	0.0	0.0	0.0
Belmore−Spinks−Kibbie (OH019)	2.8	0.0	2.7	11.8	0.0
Toledo−Fulton−Nappanee (OH020)	0.4	0.0	0.0	0.0	0.0
Blount−Glynwood−Morley (OH021)	5.5	4.1	0.0	18.5	5.5
Blount−Pewamo−Glynwood (OH022)	32.8	4.8	13.5	69.1	65.4
Castalia−Millsdale−Milton (OH024)	0.2	0.0	0.0	0.0	0.0
Milton−Millsdale−Randolph (OH036)	1.0	0.0	0.0	0.0	0.0
Tiro−Pandora−Bennington (OH041)	11.2	15.6	49.6	0.0	0.0
Bennington−Cardington−Orrville (OH063)	4.2	6.2	5.5	0.0	0.0
Pewamo−Bennington−Medway (OH065)	1.0	15.8	0.0	0.0	0.0
Bennington−Pewamo−Cardington (OH066)	3.5	17.8	8.0	0.0	0.0
Bennington−Condit−Cardington (OH079)	3.7	24.0	13.3	0.0	0.0
Rittman−Wadsworth−Orrville (OH084)	0.0	0.1	0.0	0.0	0.0
Cardington−Bennington−Sloan (OH127)	0.9	6.6	0.0	0.0	0.0
Kibbie−Colwood−Bixler (OH129)	7.0	0.0	0.0	0.5	0.0
Glynwood−Blount−Genesee (OH134)	2.4	0.0	0.0	0.0	3.9
Tiro−Fitchville−Glenford (OH160)	2.1	0.0	0.0	0.0	0.0

ciation is dominant in the Sandusky watershed (32.8% coverage), Tymochtee (65.4%), and Rock Creek (69.1% coverage), followed by the Tiro−Pandora−Bennington association. The Blount Soil Series is a fine, illitic, mesic Aeric Epiaqualfs; the Pewamo Soil Series a fine, mixed, active, mesic Typic Argiaquolls; and the Glynwood Soil Series a fine, illitic, mesic Aquic Hapludalfs. Pewamo soils are very poorly drained soils with clay loam texture in the A horizon, silty clay texture in the Btg horizons, and silty clay loam texture in the Cg horizon. The Glynwood soils are moderately well drained soils with silt loam texture in the A and E horizons; silty clay loam texture in the BE horizon; silty clay, clay, and silty clay loam in the Bt and BC horizons; and clay loam in the C horizon.

There are contrasting soils in the Honey Creek watershed. The Tiro−Pandora−Bennington association is dominant (49.6% coverage), which are fine−silty, mixed (illitic), mesic Aeric (Typic) Epiaqualfs. Epiaqualfs are soils that are formed by episaturation, which means that there is perching of water over a more slowly permeable layer of horizon in the soil. These soils have silt loam in the upper horizon (Ap), silt loam and silty clay loam in the B horizons (BE, Bt, and Btg), and silty clay loam, loam, and clay loam texture in the lower horizons (BC, BCg, and C).

Soils in the Bucyrus watershed are more diverse when compared with the other watersheds including Tiro, Pandora, Bennington, Pewamo, Medway (fine−loamy, mixed, superactive, mesic Fluvaquentic Hapludolls), Condit (fine, illitic, mesic Typic Epiaqualfs), and Cardington (fine, illitic, mesic Aquic Hapludalfs) soils. These soils show very different drainage behavior, ranging from very poorly drained soils (e.g., Condit) to moderately well drained soils (e.g., Medway and Cardington).

Land Use

An analysis of 1994 LANDSAT data indicated that 84.0% of the land was used for agriculture, 12.6% was wooded, 1.2% was urban, and 1.1% was non−forested wetlands. Land use mapping performed in 2000−2001 by the University of Toledo (Czajkowski, 2001) identified the major land use as agriculture (61.2%) and woody vegetation (27.9%) (fig. 2). The major crops were cabbage, soybean, and corn. Major crops based on county−level estimates in 1985 were corn with 35.6% of cropland acreage, soybeans with 44.9%, and wheat with 19.5%. Crop production was similar in 1995 with 32.1% in corn, 49.1% in soybeans, and 18.7% in wheat. Tillage practices shifted from 86.0% conventional management in 1985 to 50.5% in 1995, as farmers replaced conventional with conservation tillage practices. Tile drainage is used extensively throughout the watershed. Urban areas within the Sandusky watershed are Bucyrus, Fremont, Tiffin, and Upper Sandusky, and numerous smaller communities. Average annual precipitation ranges from 881 mm at Fremont to 964 mm at Bucyrus. Annual mean discharge is 29.1 $m^3 s^{-1}$ for the Sandusky watershed at Fremont, 3.8 $m^3 s^{-1}$ for Honey Creek at Melmore, 5.3 $m^3 s^{-1}$ for Tymochtee at Crawford, and 0.88 $m^3 s^{-1}$ for Rock Creek at Tiffin (Baker and Ostrand, 2000).

SWAT Model Input

In SWAT, the watershed of interest is divided into subbasins, which gives the model the strength to better represent the properties of land uses and/or soils of each subbasin that have a significant effect on its hydrology. Two options are available to discretize the watershed into simulation elements: (1) a single hydrological response unit (HRU) based on dominant land use and soil classes within a subbasin, or (2) multiple HRUs for each subbasin considering the percentages of the land use and the soil class within the subbasin. Subbasins are grouped based on HRUs and other watershed characteristics. Using a daily time step, SWAT is designed to simulate long−term, continuous processes of flow, sediment yields, and nutrient transport in ungauged watersheds of diverse hydrologic, geologic, and climate conditions (Borah and Bera, 2003). The strength of SWAT is its ability to simulate continuous phenomena, with less emphasis on event−based, short−term rainfall−runoff simu−

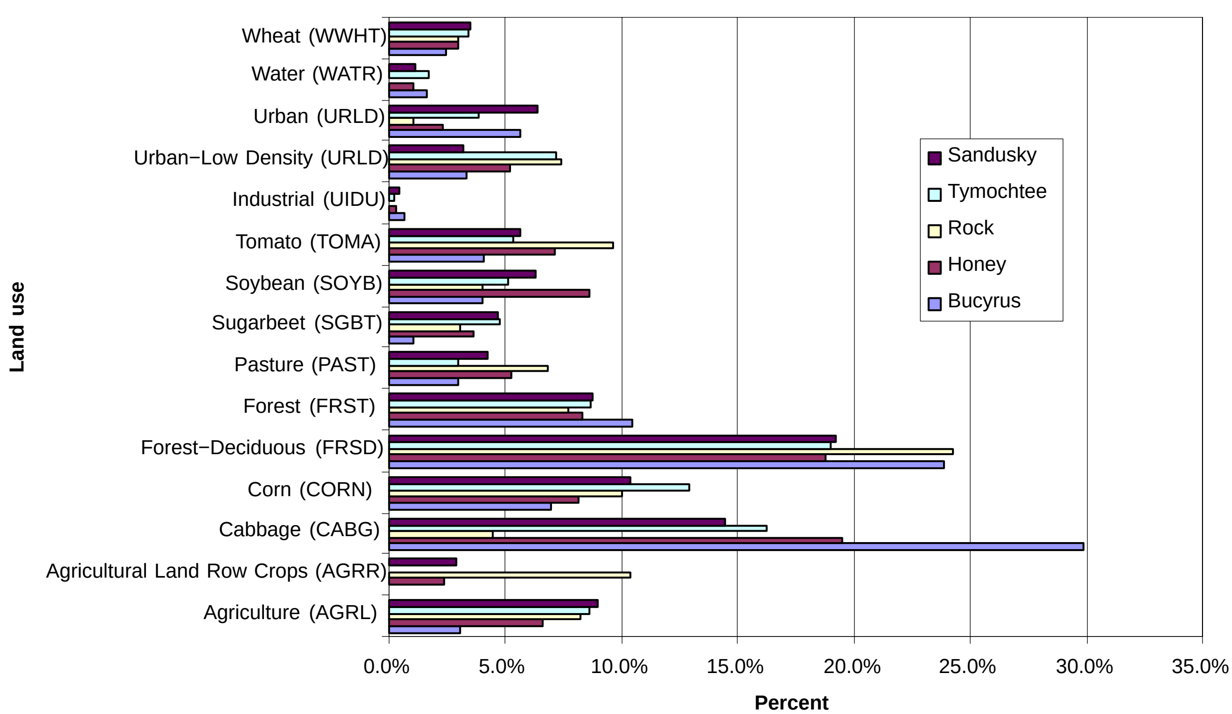

Figure 2. Land use and land cover distribution in the Sandusky watershed and subwatersheds (data source: Landsat ETM+ imagery 2000, University of Toledo).

lations. SWAT is a complex quasi−physically based water quality model relying on numerous input files and parameters to represent hydrologic, climatic, water quality, management, plant, and soil processes within a watershed. This poses a great challenge when attempting to set up the model manually. The ArcView GIS interface for SWAT (AVSWAT) (Di Luzio et al., 2002) was used for this study, providing an easy−to−use graphical user interface for organizing all required input for the SWAT model. The interface includes a set of modular tools for delineating the watershed boundaries, defining the HRUs, generating SWAT input files, creating agricultural management scenarios, executing SWAT simulations, and reading and charting results. A detailed description of AVSWAT was presented by Di Luzio et al. (2002). We acquired a digital elevation model (DEM), hydrography, soil maps, land use and land cover, management, climate, water use, and pond data for the Sandusky watershed. We used the National Elevation Dataset (NED) with cell size of 30 × 30 m from USGS. Digital land cover and land use maps derived from Landsat TM/ETM+ imagery for the years 1992 and 2000 was provided by the Department of Geography and Planning, University of Toledo. The land use types in the Sandusky watershed and subwatersheds are summarized in figure 2. We used a digital soil map from the State Soil Geographic Database (STATSGO) developed by the USDA−NRCS.

We delineated watershed boundaries using AVSWAT. The drainage area threshold, which is the minimum area to maintain a stream, was set to 1,000 ha based on a previous sensitivity study conducted in the Honey Creek subwatershed. Multiple HRUs were delineated to discretize the Sandusky watershed into simulation elements. The threshold values of land use and soil were both set at 10%, i.e., land use that covered less than 10% of the subwatershed area and soil that accounted for less than 10% of the subwatershed were not considered in the simulation.

The meteorological data of daily precipitation and temperature time series were obtained from seven rain gauge stations and ten climate stations located within and around the Sandusky watershed. Solar radiation, wind speed, and relative humidity data were unavailable. Four reservoirs including St. John's dam (OH02392), Bacon low−head dam (OH00801), a second low−head dam (OH00802), and Ballville dam (OH00809) were treated as reservoirs in the model, while all other 26 reservoirs were considered as large ponds in the simulation. Observed daily flows at five USGS gauge stations for water years 1998 and 1999 were used for calibration of hydrology in the SWAT model, and water years 2000 and 2001 were used to validate the water flow of the model. The model was run for water year 1997, which was not included in the analysis and interpretation. This was done to avoid any bias on simulations from assumed initial watershed conditions. Parameters were adjusted during the calibration. The model proved to be sensitive while adjusting the SCS runoff curve number for moisture condition II and a number of groundwater parameters. Table 2 lists CN numbers that gave the best results. Calibrated curve numbers were consistent with the range of curve numbers developed by the USDA Soil and Water Conservation Service (SCS). Surface runoff lag time was modified to two days because the Sandusky watershed is large and only a portion of the surface runoff can reach the drainage outlet in less than a day.

Table 2. USDA−SCS runoff curve numbers (CN), type II for calibration of the model.

Land use	CN
Agricultural − generic land	77
Agricultural − land−row crops	78
Urban/residential, low density	66
Mixed forest	77
Deciduous forest	77
Sugar beet	83
Soybean	85
Tomato	83
Cabbage	65
Corn	70
Wheat	81
Pasture	69
Industrial	88
Water	92

The monthly precipitation values and snowfall depths are shown in figure 3. Since the watershed is located in Ohio, where snow is a common phenomenon of precipitation in the winter, we also adjusted parameters related to snow accumulation and melting processes. Snowfall and snow−melting temperatures were 1°C and 0.5°C, respectively; maximum and minimum snow melting rates were adjusted to 10 mm (°C day)$^{-1}$ and 4.5 mm (°C day)$^{-1}$. The threshold depth of water in the shallow aquifer for re−evaporation (REVAPMN) was set at 500 mm, while the threshold depth of water in the shallow aquifer required for base flow to occur (GWQMN) was set at 0.

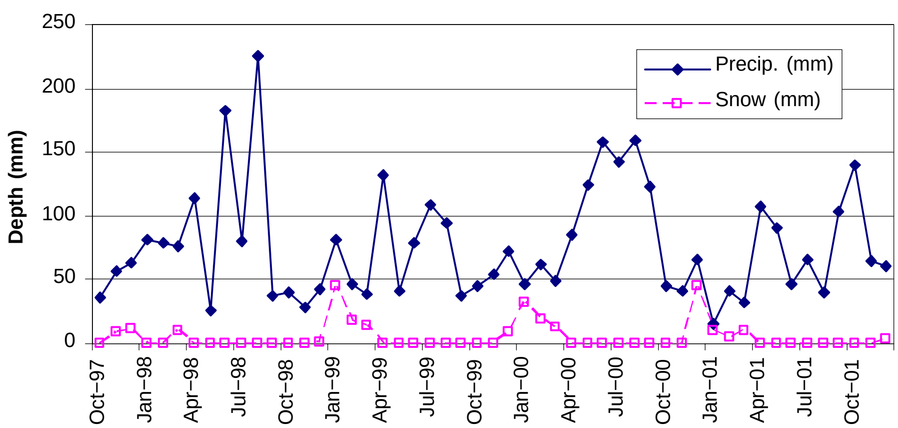

Figure 3. Monthly total precipitation versus snowfall depth (in equivalent water depth) at Tiffin.

Rainfall distribution was calculated by the skewed normal method. Potential evapotranspiration was estimated by the Hargreaves method because only precipitations and temperatures during the simulation period were obtained. The variable−storage method was used to route water in the channel. Although a daily time step was used for the simulations, we present aggregated flow simulations in figures 4 through 14 to summarize our results.

RESULTS

SURFACE AND GROUNDWATER CALIBRATION AND VALIDATION

The model was calibrated against observed monthly runoff at five USGS gauge stations for water years 1998 and 1999. We used multiple metrics to evaluate the simulation success. Overall, the monthly surface water calibration was satisfactory (table 3). Most correlation coefficients of observed and simulated surface flow were greater than 0.70, with a minimum correlation of 0.67 and a maximum of 0.82. The Nash−Sutcliffe coefficient ranges between 0 and 1, with 1 indicating a perfect match between observed and simulated value. Negative Nash−Sutcliffe values indicate poor predictions. Most Nash−Sutcliffe coefficients for surface water calibration were acceptable, with most values greater than 0.4. The Nash−Sutcliffe coefficient is sensitive to outliers, e.g., over− or underestimated rainfall−runoff events, which generated some poorer coefficients below 0.4. The statistics of simulated monthly surface water flow results from calibration are summarized in table 3 and figures 4 to 8.

The lower correlation coefficients and Nash−Sutcliffe coefficients in the calibration suggested that certain processes were not sufficiently modeled. Further investigations revealed that the model had difficulties dealing with precipitation−runoff simulation during the winter, when heavy snowfalls accounted for most of the annual total precipitation in the watershed (e.g., January 1999, fig. 3). Other uncertainties in the observed data can also impact the calibration, including spatial variability in precipitation, land use, soil types, and channel flow measurements.

After the model was calibrated, the SWAT model was validated using the same set of parameters for water years 2000 and 2001. The statistics of simulated monthly surface water flow from the validation are presented in table 3. Overall, most correlation coefficients were as good as those from the calibration, while the Nash−Sutcliffe coefficients were slightly worse than those from the calibration, with one value around zero. One explanation is that the model was given only one year for adjusting (water year 1997); although the set of parameters appeared to be the best choice, they actually contained errors and finally propagated into the validation process. Correlations were relatively high, whereas some of the Nash−Sutcliffe coefficients were below 0.25 due to the impact from few rainfall−runoff events, which were not modeled very well. The Bucyrus station showed the poorest statistical validation values (e.g., negative Nash−Sutcliffe Coefficient of −0.04), which was caused by winter rainfall−runoff events that were overestimated in January 2000 (fig. 5). In contrast, the surface runoff simulations matched closely the observed values at the Fremont station at the drainage outlet of the Sandusky watershed (fig. 4).

Figures 4 through 8 show the comparisons of the simulated and observed monthly surface water flow at Fremont, Bucyrus, Honey Creek, Tymochtee, and Rock Creek, in which dashed lines indicate observed data. Seasonal trends were simulated very well, especially during dry seasons. The SWAT model had difficulties simulating high surface water values in winter, which might be caused by the potential evapotranspiration estimation method and/or by the simplified thawing routine implemented in SWAT. As illustrated in figure 3, during winter, snowfall accounts for a significant percentage of the total monthly precipitation. Other reasons for poorer simulations of rainfall−runoff events might be related to the daily time step used for simulations and/or the relatively simple curve number method used to simulate infiltration and surface runoff. Interesting to note are the underestimations of surface flow during the calibration period at all stations, indicated by the negative mean error values in table 3. In contrast, during the validation period, the simulation model overestimated surface flow consistently at four out of five stations.

Statistics for simulated groundwater flow are presented in table 4, and simulated and observed groundwater flow at the Fremont station are shown in figure 9. Results for surface water simulations were better when compared to groundwater simulations. The former is more important for transport of nutrients and pesticides within the watershed. For the calibration, the mean error was highest (−0.78) at the Fremont station and lowest (−0.13) at the Bucyrus station. Similar trends were found using the RMSE, which showed the highest value (0.91) at the Fremont station and the lowest values at the Rock Creek (0.08) and Bucyrus (0.10) stations. The correlation coefficient was highest (0.82) at the drainage

Table 3. Statistics of monthly surface water flow.

Station	Simulated ($m^3\ s^{-1}$)	Observed ($m^3\ s^{-1}$)	Mean Error ($m^3\ s^{-1}$)	RMSE ($m^3\ s^{-1}$)	Correlation	Nash−Sutcliffe Coefficient
Calibration						
Bucyrus	1.10	1.79	−0.69	0.35	0.70	0.31
Fremont	17.16	19.05	−1.89	2.44	0.82	0.65
Honey	2.30	2.58	−0.28	0.34	0.82	0.64
Rock	0.55	0.69	−0.14	0.11	0.77	0.54
Tymochtee	2.88	3.05	−0.16	0.57	0.67	0.43
Validation						
Bucyrus	1.71	1.08	0.63	0.24	0.67	−0.04
Fremont	23.03	15.49	7.54	2.61	0.79	0.23
Honey	2.94	1.97	0.98	0.28	0.87	0.45
Rock	0.46	0.46	0.00	0.05	0.87	0.75
Tymochtee	3.09	2.45	0.65	0.48	0.62	0.33

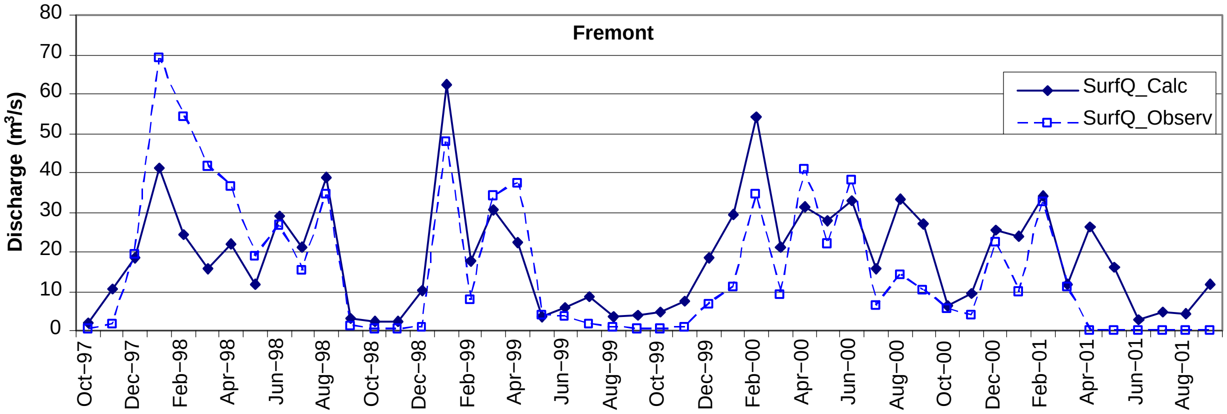

Figure 4. Simulated and observed monthly surface water flow at Fremont.

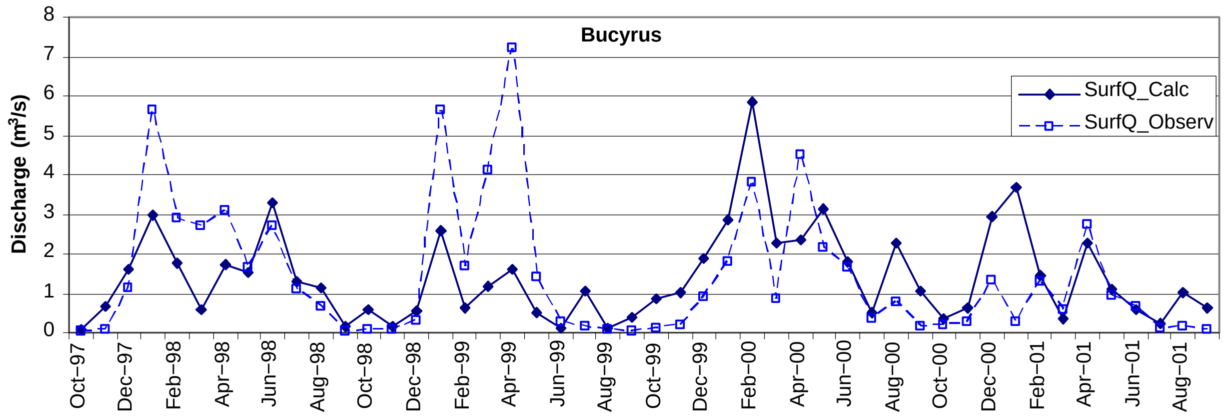

Figure 5. Simulated and observed monthly surface water flow at Bucyrus.

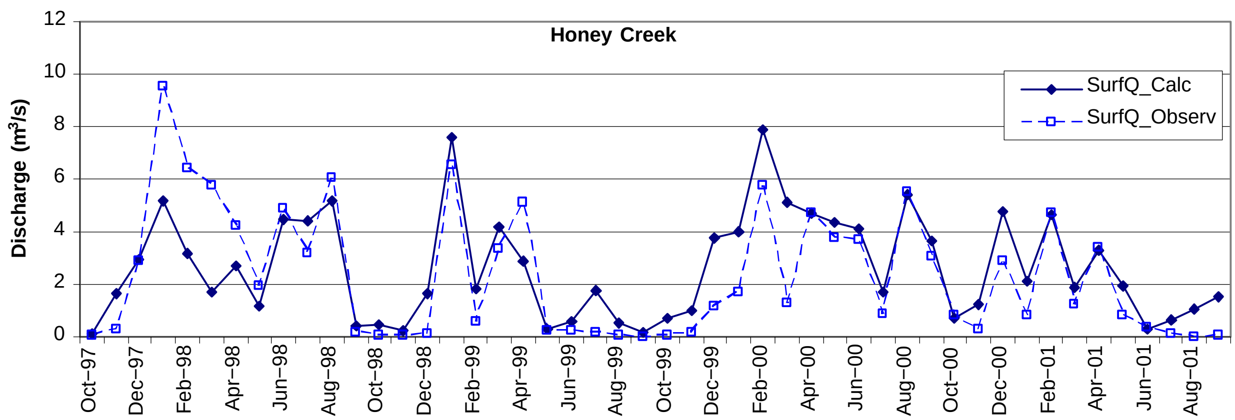

Figure 6. Simulated and observed monthly surface water flow at Honey Creek.

outlet (Fremont station) and lowest (0.49) at the Bucyrus station. Likewise, correlation coefficients were high at the Rock Creek (0.79) and Tymochtee (0.79) stations, suggesting that the simulations closely matched the observed values. The calibration results indicated that groundwater flow was modeled relatively well at the drainage outlet, although some of the subwatersheds showed high discrepancies between measured and observed groundwater flow.

For the validation, the correlation coefficients for groundwater flow ranged from 0.31 at the Honey Creek station to 0.55 at the Tymochtee station. At the Fremont station, the mean error value of −3.51 $m^3 s^{-1}$ indicated an underestimation of groundwater flow, which was consistent with all other subbasin stations except for Rock Creek. Differences in the groundwater flow calibration and validation within the Sandusky watershed and subwatersheds can be attributed to the different geology and physiography, with the Till Plains in the central and southern portion covering the Tymochtee and Bucyrus subwatersheds. In contrast, Rock Creek and Honey Creek are at the transition to the Lake Plains.

Figures 10 to 14 present comparisons of the simulated and observed monthly total flow at the five gauge stations, in

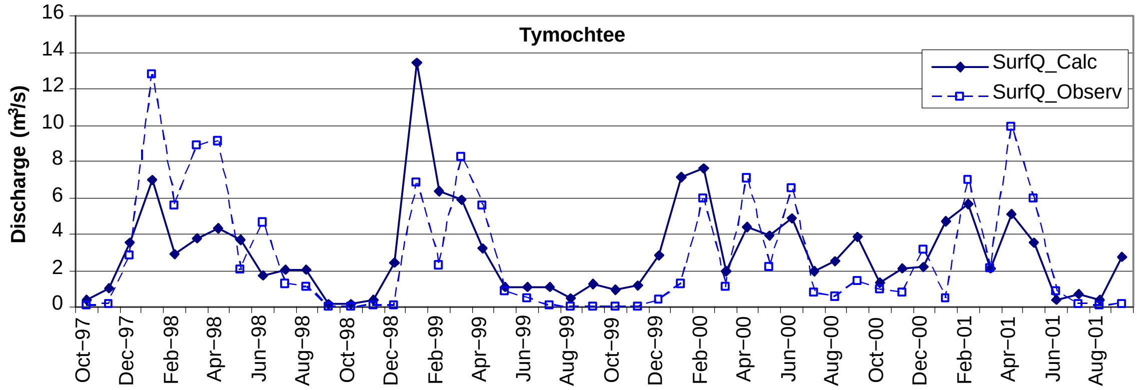

Figure 7. Simulated and observed monthly surface water flow at Tymochtee.

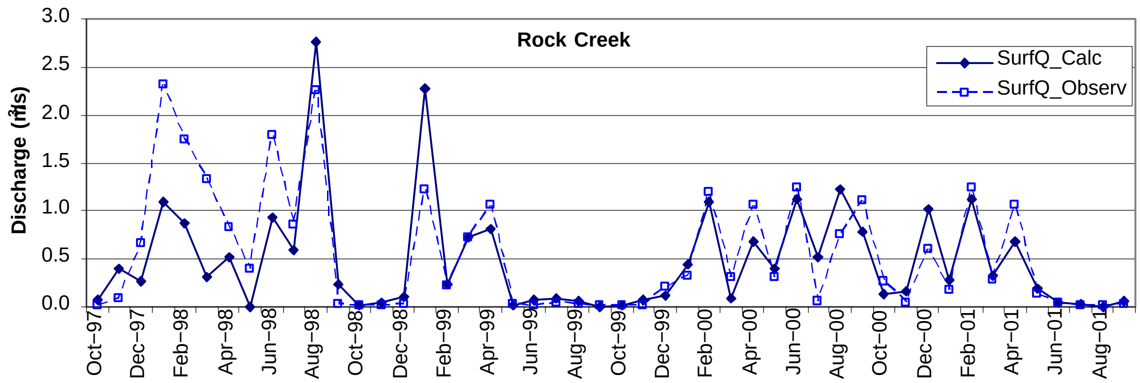

Figure 8. Simulated and observed monthly surface water flow at Rock Creek.

Table 4. Statistics of monthly groundwater flow.

Station	Simulated (m^3 s^{-1})	Observed (m^3 s^{-1})	Mean Error (m^3 s^{-1})	RMSE (m^3 s^{-1})	Correlation	Nash−Sutcliffe Coefficient
Calibration						
Bucyrus	0.42	0.55	−0.13	0.10	0.49	−0.24
Fremont	7.26	8.04	−0.78	0.91	0.82	0.60
Honey	0.54	0.78	−0.24	0.14	0.57	0.15
Rock	0.40	0.15	0.25	0.08	0.79	−9.13
Tymochtee	1.12	0.96	0.15	0.16	0.79	0.48
Validation						
Bucyrus	0.26	0.48	−0.22	0.09	0.41	−0.16
Fremont	3.78	7.29	−3.51	1.33	0.32	0.01
Honey	0.38	0.67	−0.29	0.12	0.31	−0.33
Rock	0.15	0.12	0.03	0.02	0.41	−0.57
Tymochtee	0.67	0.90	−0.23	0.16	0.55	0.22

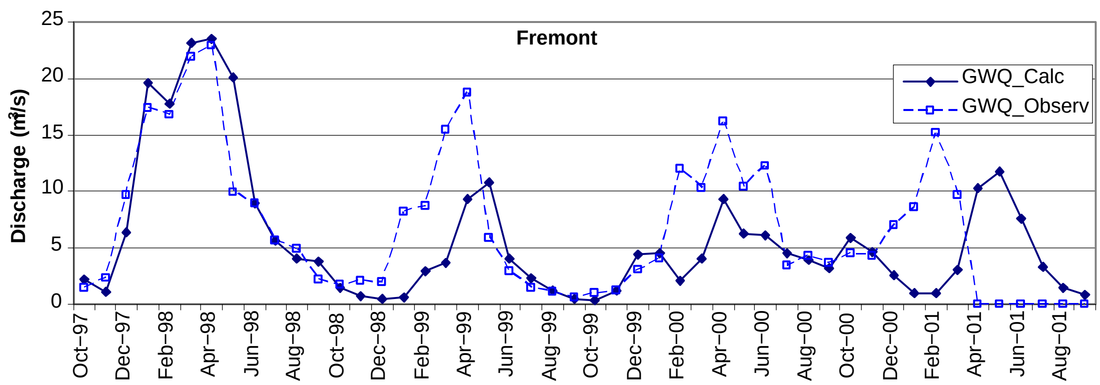

Figure 9. Simulated and observed monthly groundwater flow at Fremont.

Table 5. Statistics of total water flow.

Station	Simulated ($m^3 s^{-1}$)	Observed ($m^3 s^{-1}$)	Mean Error ($m^3 s^{-1}$)	RMSE ($m^3 s^{-1}$)	Correlation	Nash−Sutcliffe Coefficient
Calibration						
Bucyrus	1.52	2.33	−0.82	0.42	0.68	0.31
Fremont	24.42	27.09	−2.67	2.46	0.93	0.79
Honey	2.84	3.36	−0.52	0.41	0.87	0.66
Rock	0.95	0.84	0.11	0.08	0.91	0.81
Tymochtee	4.00	4.01	−0.01	0.57	0.80	0.64
Validation						
Bucyrus	1.97	1.56	0.41	0.24	0.70	0.40
Fremont	26.81	22.78	4.03	2.56	0.80	0.58
Honey	3.32	2.64	0.68	0.28	0.90	0.65
Rock	0.62	0.58	0.03	0.06	0.86	0.73
Tymochtee	3.76	3.34	0.42	0.24	0.70	0.40

which dashed lines correspond to observed data. The simulated monthly total flows at all five gauge stations were fairly close to the observed values. All statistical properties (table 5) indicated that the simulations of water flow in the Sandusky watershed were acceptable. For calibration at the drainage outlet, the mean error of total water flow was −2.67 $m^3 s^{-1}$, which was caused by an underestimation of measured surface and groundwater flow. In contrast, at the Rock Creek station, the mean error was 0.11 $m^3 s^{-1}$, which was based on an underestimation of surface water flow (mean error: −0.14 $m^3 s^{-1}$) and an overestimation of groundwater flow of 0.25 $m^3 s^{-1}$. At the Tymochtee station, the mean error value was 0.01 $m^3 s^{-1}$, and underestimations of surface water flow were compensated by an overestimation of groundwater flow. For the calibration, the highest RMSE was calculated for the Fremont station (2.46 $m^3 s^{-1}$) and the lowest for the Rock Creek station (0.08 $m^3 s^{-1}$). Correlation coefficients were high for all stations (> 0.68), with the highest (0.93) at the Fremont station, suggesting a close fit between observed and simulated total flow. This was confirmed by the Nash−Sutcliffe coefficient, which was highest (0.79) at the Fremont station.

The validation results of mean error suggested that total flow was overestimated at all stations, with the highest mean

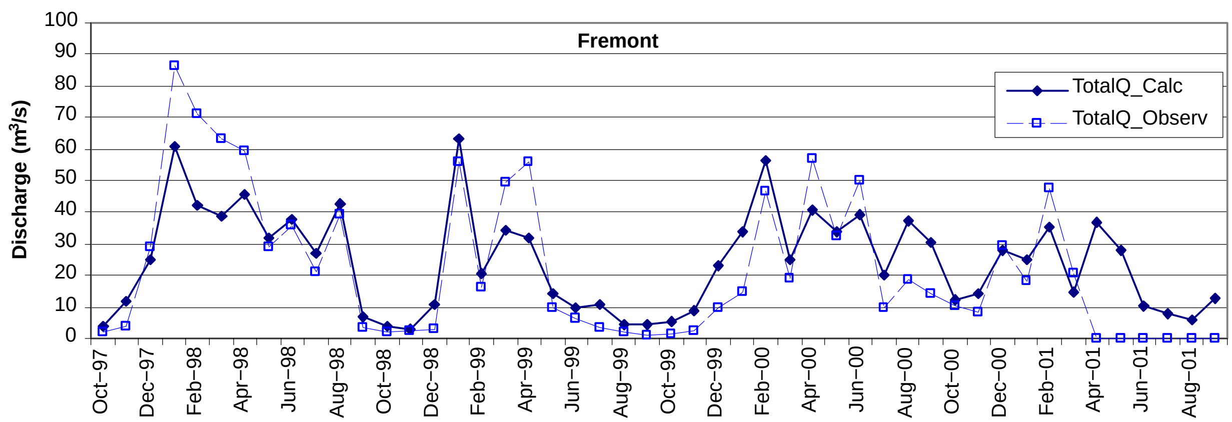

Figure 10. Simulated and observed monthly total water flow at Fremont.

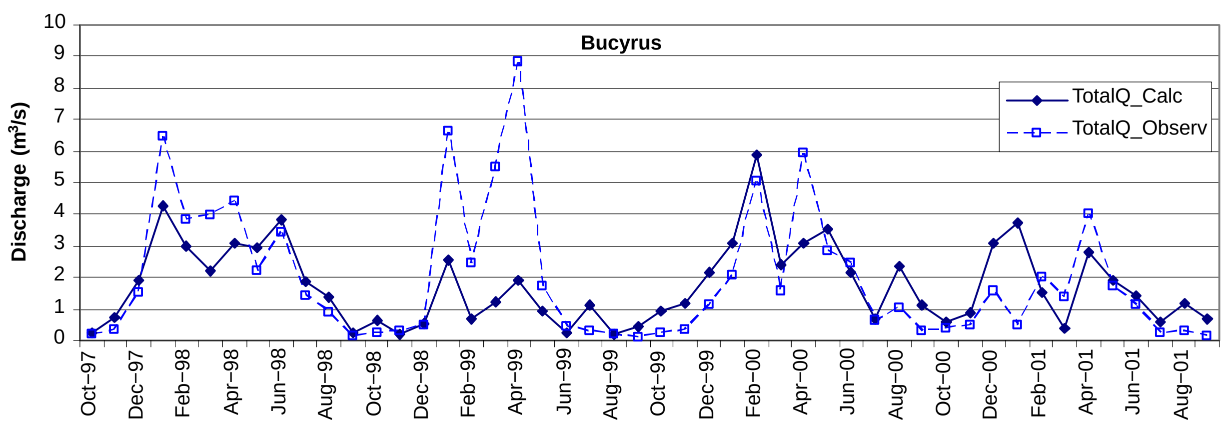

Figure 11. Simulated and observed monthly total water flow at Bucyrus.

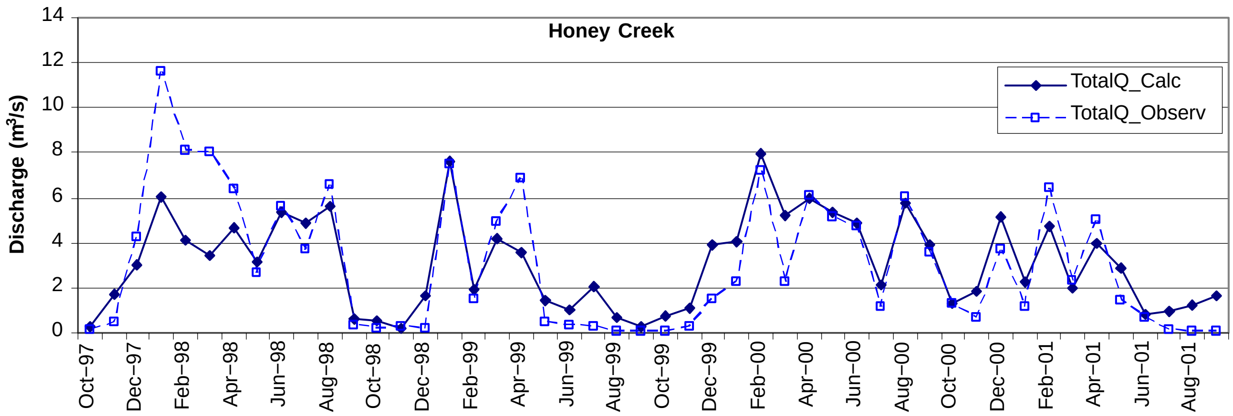

Figure 12. Simulated and observed monthly total water flow at Honey Creek.

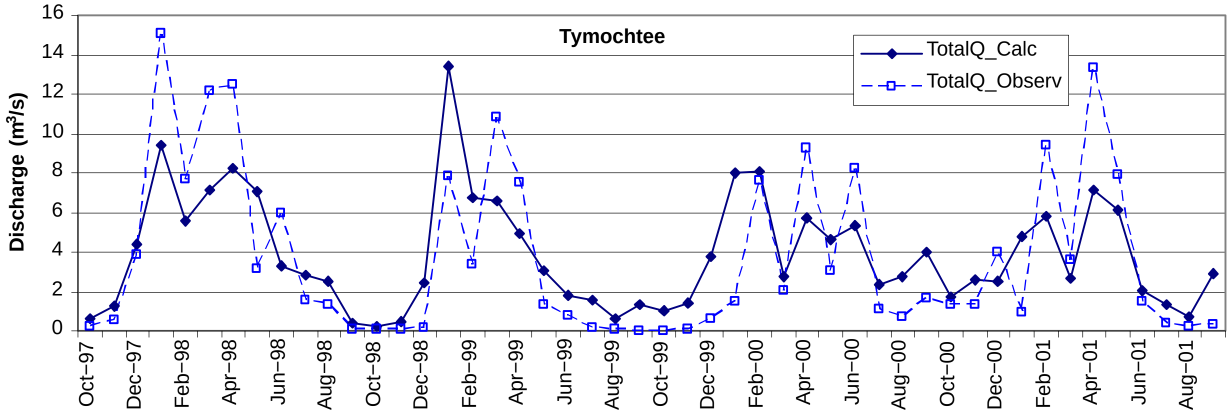

Figure 13. Simulated and observed monthly total water flow at Tymochtee.

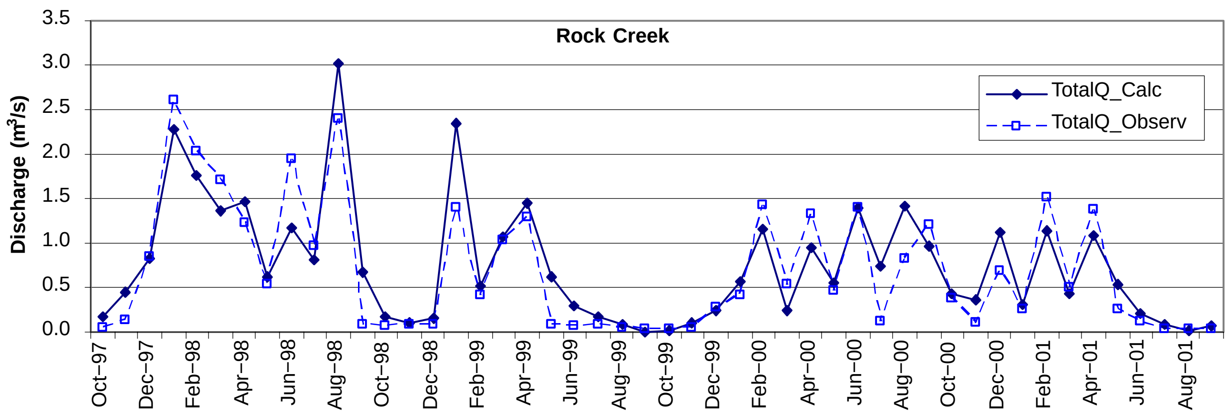

Figure 14. Simulated and observed monthly total water flow at Rock Creek.

error (4.03 $m^3\ s^{-1}$) at the Fremont station and the lowest (0.03 $m^3\ s^{-1}$) at the Rock Creek station. The RMSE showed the same trends as the mean error for all stations. Overall, the correlation coefficients for total flow at all stations were high, ranging from 0.70 at the Bucyrus and Tymochtee stations to 0.9 at the Honey Creek station. At the drainage outlet, the correlation coefficient was 0.80 for total flow. The impact on total flow validation from the relatively poor groundwater validation at the Fremont station was relatively small. The Nash−Sutcliffe coefficients ranged from 0.40 at the Tymochtee and Bucyrus stations to 0.73 at the Rock Creek station.

We suspect that land use differences had a major impact on the quality of the hydrologic simulations in the investigated watersheds. The following summary focuses on the dominant land uses in the different watersheds. In the Bucyrus watershed, deciduous forest (35%) and cabbage (30%) were dominant, whereas in the Rock Creek watershed, deciduous forest accounted for 32%, with a high percentage of agricultural land (18%) in addition to tomatoes (9%) and pasture (8%). In the Tymochtee watershed, deciduous forest was 27%, and cabbage accounted for 16%, corn for 14%, and agricultural land for 8%. Urban areas were higher in the Tymochtee and Bucyrus watersheds when compared to the other watersheds. In the Honey Creek watershed, deciduous forest was 26%, cabbage 19%, agriculture and row crops 11%, and soybean 8%. In the Sandusky watershed, deciduous forest was 27%, cabbage 19%, agriculture and row crops 11%, corn 11%, and soybean 8%. The land use determined the CN numbers, which described the process of infiltration and surface runoff.

CONCLUSIONS

We used SWAT to simulate the hydrology in the Sandusky watershed in Ohio. Spatially distributed model simulations improved our understanding of the hydrologic patterns across the Sandusky watershed and its subwatersheds. We used a spatially distributed calibration and validation procedure considering the surface, ground, and total flow at the drainage outlet of the Fremont station and four subwatershed stations (Honey Creek, Rock Creek, Tymochtee, and Bucyrus). The model was first calibrated for hydrology using observed streamflow data from five USGS gauge stations for water years 1998 and 1999. Then the model was validated for surface and groundwater using two water years (2000 and 2001) of recorded streamflow data from the same USGS stream gauges.

Overall, the simulated trends matched moderately well with the observed data. However, the model had problems dealing with snow accumulation and melting processes, as shown in the comparison of simulated and observed water yields in the winter periods. Due to the unavailability of solar radiation, wind speed, and relative humidity data, the potential evapotranspiration was estimated by the Hargreaves method, which might have contributed to the errors in simulating water yields. Other errors in observed data might also have impacted simulation, including spatial variability in precipitation, land use, soil types, and channel flow measurements. The land use maps derived from remote sensing showed a high percentage of cabbage and tomato, which were overestimated when compared to the county statistics for the Sandusky watershed.

Our study showed that calibration and validation of water flow at the drainage outlet do not imply an understanding of hydrologic patterns within the watershed. Watershed characteristics within and between subwatersheds differed in term of land use and land cover, topography, soils, and geology. These watershed characteristics generated variable hydrologic patterns across the Sandusky watershed. The linkage between subwatersheds also influenced hydrologic patterns, with upstream water flow impacting downstream water flow. In some cases, over− and underestimations of water flow at the outlet of subwatersheds averaged out at the drainage outlet. In other cases, over− and underestimations accumulated, causing major discrepancies between observed and simulated values at the drainage outlet. Only a spatially variable calibration and validation procedure has the power to address these issues. The more calibration and validation stations are available to adjust a model such as SWAT to local and regional watershed characteristics, the better will be the prospects for reliable simulations of water flow. Correct simulations of hydrologic patterns are a prerequisite for simulating transport processes of nitrogen, phosphorus, pesticides, and suspended sediment. Thus, the quality and reliability of solid and solutes simulations is directly linked to hydrologic simulations. Errors propagate from the hydrology routine into the routines of transport processes, which can cause major discrepancies in model predictions.

ACKNOWLEDGEMENTS

We thank T. F. A. Bishop for editing and assembling the SWAT input datasets. The project was funded by the Ohio Lake Erie Commission.

REFERENCES

Abbott, M. B., and J. C. Refsgaard, eds. 1996. *Distributed Hydrological Modelling*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Baker, D. B., and M. Ostrand. 2000. *Sandusky River Watershed Resource Inventory.* Fremont, Ohio: Sandusky River Watershed Coalition.

Beasley, D. B., L. F. Huggins, and B. J. Monke. 1980. ANSWERS: A model for watershed planning. *Trans. ASAE* 23(4): 938−944.

Bicknell, B. R., J. C. Imhoff, J. L. Kittle Jr., A. S. Donigian Jr., and R. C. Johanson. 1997. Hydrological simulation program−−Fortran: User's manual for version 11. EPA/600/R−97/080. Athens, Ga.: U.S. Environmental Protection Agency, National Exposure Research Laboratory.

Bingner, R. L. 1996. Runoff simulated from Goodwin Creek watershed using SWAT. *Trans. ASAE*. 39(1): 85−90.

Bingner, R. L., and F. D. Theurer. 2003. AnnAGNPS technical processes: Documentation version 3.3. Available at: http://msa.ars.usda.gov/ms/oxford/nsl/agnps/PLModel/document/Tech_Doc.PDF. Oxford, Miss., and Gaithersburg, Md.: USDA−NRCS.

Borah, D. K. 2002. Watershed−scale nonpoint−source pollution models: Mathematical bases. In *Proc. 2002 ASAE Annual International Meeeting/CIGR XVth World Congress*, CD−ROM. St Joseph, Mich.: ASAE.

Borah, D. K., and M. Bera. 2003. SWAT model background and application reviews. In *Proc. 2002 ASAE Annual International Meeting*, CD−ROM. St. Joseph, Mich.: ASAE.

Bouraoui, F., and T. A. Dillaha, 1996. ANSWERS−2000: Runoff and sediment transport model. *J. Environ. Eng., ASCE* 122(6): 493−502.

Czajkowski, K. 2001. Land use / land cover map derived from Landsat ETM imagery. Toledo, Ohio: University of Toledo, Department of Geography and Planning.

Di Luzio, M., R. Srinivasan, J. Arnold, and S. L. Neitsch. 2002. ArcView interface for SWAT 2000 user's guide. Temple, Texas: Blackland Research Center, Texas Agricultural Experiment Station.

Dunne, T. 1978. Field studies of hillslope flow processes. In *Hillslope Hydrology*, 227−289. M. J. Kirkly, ed. New York, N.Y.: John Wiley and Sons.

Dunne, T., and R. Black. 1970. An experimental investigation of runoff production in permeable soils. *Water Resources Research* 6(1): 478−499.

Frankenberger, J. R., E. S. Brooks, M. T. Walter, M. F. Walter, and T. S Steenhuis. 1999. A GIS−based variable source area hydrology model. *Hydrological Processes* 13(6): 805−822.

Freeze, R. A., and R. L. Harlan. 1969. Blueprint for a physically−based digitally−simulated hydrological response model. *J. Hydrology* 9(3): 237−258.

Hydrologic Engineering Center. 1998. HEC−1 flood hydrology package user's manual. Washington, D.C.: U.S. Army Corps of Engineers.

Michaud, J., and S. Sorooshian. 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resources Research* 30(3): 593−605.

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

Scharffenberg, W. A. 2001. Hydrologic modeling system HEC−HMS user's manual. Washington, D.C.: U.S. Army Corps of Engineers.

Singh, V. P., ed. 1995. *Computer Models of Watershed Hydrology.* Highlands Ranch, Colo.: Water Resources Publications.

Srinivasan, R., T. S. Ramanarayanan, J. G. Arnold, and S. T. Bednarz. 1998. Large−area hydrological modeling and assessment: Part II. Model application. *J. American Water Resources Assoc.* 34(1): 91−101.

USDA. 1965. Computer program for project formulation hydrology. Washington, D.C.: USDA Soil Conservation Service.

Vaché, K. B., J. M. Eilers, and M. V. Santelmann. 2002. Water quality modeling of alternative agricultural scenarios in the U.S. Corn Belt. *J. American Water Resources Assoc.* 38 (3): 773−787.

Vieux, B. E. 2001. *Distributed Hydrologic Modelling Using GIS.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Young, R. A., C. A. Onstad, D. B. Bosch, and W. P. Anderson. 1989. AGNPS: A nonpoint−source pollution model for evaluating agricultural watersheds. *J. Soil and Water Conservation* 44(2): 168−173.