Journal of Hydrology (2007) **337**, 68–86

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

# Cannonsville Reservoir Watershed SWAT2000 model development, calibration and validation

**Bryan A. Tolson [a,*], Christine A. Shoemaker [b]**

[a] *Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ont., Canada N2L 3G1*
[b] *School of Civil and Environmental Engineering, Cornell University, 210 Hollister Hall, Ithaca, NY 14853, USA*

Received 21 October 2005; received in revised form 21 September 2006; accepted 11 January 2007

**KEYWORDS**
Distributed watershed model;
Hydrologic modelling;
Nonpoint source pollution;
Phosphorus;
Calibration;
Validation

**Summary** The Soil and Water Assessment Tool version 2000 (SWAT2000) watershed model was utilized to simulate the transport of flow, sediments and phosphorus to the Cannonsville Reservoir in Upstate, New York. The available datasets for model development, particularly the phosphorus input and water quality calibration data, in this case study are unique because of the large amount of watershed specific, spatially and temporally varying data that are available for model development. Relative to the default SWAT inputs, alternative model input generation methodologies were tested and shown to produce more representative inputs that generate substantially different simulation results. The successful application of SWAT2000 in this case study required two critical model modifications regarding excess soil water movement in frozen soils and soil erosion predictions under snow cover. The Nash–Suttcliffe coefficient of efficiency ($E_{NS}$) for daily flows at the main flow station in the watershed was at least 0.80 in both the seven-year calibration period and the one year and four year validation periods. Average monthly total phosphorus loads were predicted within 15% of the corresponding measured data and the monthly $E_{NS}$ coefficients for total phosphorus were at least 0.63 in the calibration and validation periods. The results of this study are important for future SWAT modelling studies in gauged and ungauged watersheds, especially those in regions like the Northeast US that are subject to freezing temperatures in winter.


## Introduction

As the capability of watershed models to simulate spatially distributed fluxes of water, sediment and nutrients continues to grow, so too does the spatially distributed inputs needed to drive the model predictions. Similarly, as the

* Corresponding author. Tel.: +1 519 8884567x33377.
*E-mail addresses:* btolson@uwaterloo.ca (B.A. Tolson), cas12@cornell.edu (C.A. Shoemaker).

0022-1694/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2007.01.017

number of anthropogenic activities that models ostensibly claim to represent increases (especially nonpoint source related agricultural management practices) so too does the list of activities that need to be described in model inputs. In order to use these spatially distributed models in a responsible manner to help guide or prioritize spatially variable management decisions, a substantial effort can be required to ultimately identify, transform and convert inputs to the scale of interest.

The ultimate goal of this study is to develop a Soil and Water Assessment Tool version 2000 (SWAT2000) model (Neitsch et al., 2001a) of the Cannonsville Watershed (see Fig. 1) to simulate flow, sediment and phosphorus (P) transport to the Cannonsville Reservoir. SWAT2000 is a spatially distributed watershed model capable of representing a variety of agricultural practices. This study outlines the model input development, calibration and then multi-period validation of the model. Novel input development methodologies are introduced and compared with common approaches in SWAT and significant SWAT model limitations are identified and partially addressed with the addition of alternative model forms that substantially improve our case study results. The quantity and quality of the spatially variable model inputs as well as the sediment and phosphorus water quality monitoring data (more than nine years of daily loading or flux data) create a unique opportunity in which to carefully test the efficacy of the SWAT2000 model to represent and simulate spatially variable watershed processes on a medium to large scale watershed.

The Cannonsville Reservoir is one of New York City's (NYC's) largest drinking water reservoirs and is located in Delaware County in the Catskill region of Upstate NY. The Cannonsville Reservoir has historically experienced serious water quality problems due to excessive phosphorus loading. If water quality levels in NYC's reservoirs are not satisfactory, NYC may need to build a water filtration plant with initial construction costs on the order of billions of dollars. Upstate NY counties in the NYC reservoir watersheds are subject to multiple water quality regulations including a recent Total Maximum Daily Load (TMDL) assessment for phosphorus (Kane, 1999). Additional regulations are imposed on the upstate counties through a Memorandum of Agreement (MOA) with NYC (http://www.cwconline.org/pubs/moa.html) if reservoir phosphorus concentrations become too high. In the recent past, these MOA regulations have triggered 'phosphorus restrictions' in Delaware County, which restricts future economic growth in the basin when the growth directly or indirectly increases phosphorus delivery to surface waters.

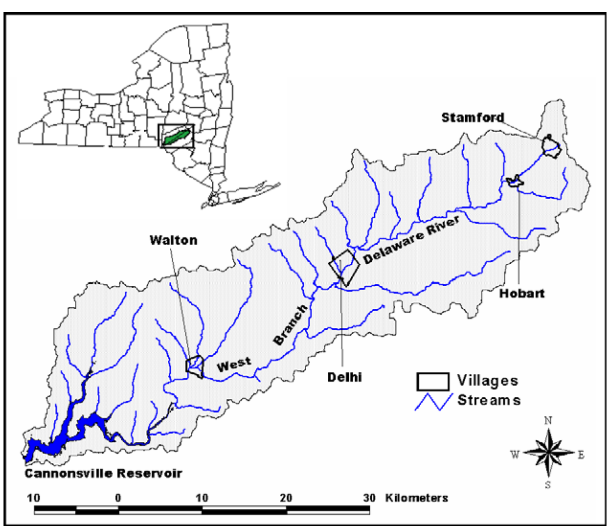


**Figure 1** The Cannonsville Reservoir Watershed in New York State.

SWAT2000 was selected for this study because the ultimate purpose of the modelling exercise is to explicitly simulate various future agricultural management options, which may vary in space and time, in order to help local decision-makers improve water quality entering the Cannonsville Reservoir. Schneiderman et al. (2002) model the Cannonsville Watershed using the lumped parameter Generalized Watershed Loading Function (GWLF) model (Haith et al., 1992). The GWLF model only simulates watershed responses at the outlet and does not take full advantage of the available spatially distributed input data sets. Another limitation of the GWLF application to the Cannonsville Watershed is that, unlike SWAT, GWLF does not simulate soil nutrient levels as state variables and is thus unable to simulate dynamic soil nutrient levels. As a result, dissolved nutrient concentrations in GWLF surface runoff are land use specific input constants that can be calibrated.

The SWAT model has been widely applied across the US (Arnold et al., 1998; Bingner, 1996; FitzHugh and Mackay, 2000; Fontaine and Jacomino, 1997; Manguerra and Engel, 1998; Peterson and Hamlett, 1998) and other parts of the world (Eckhardt et al., 2002; Grizzetti et al., 2003). The majority of previous SWAT applications focus on flow or, less frequently, flow and sediment simulation. SWAT nutrient modelling applications are much less common in the literature but some recent studies do exist (e.g. Grizzetti et al., 2003; Kirsch et al., 2002; Muleta and Nicklow, 2005; Santhi et al., 2001). A few studies have reported on SWAT modelling problems and provided model improvements. Peterson and Hamlett (1998) report problems simulating flows in a Pennsylvania watershed with conditions very similar conditions to the Cannonsville Watershed. Eckhardt et al. (2002) modify SWAT version 99.2 so that streamflow in low mountain ranges of Germany can be accurately simulated.

An important subset of the previous SWAT modelling studies focuses on how alternative levels of watershed spatial discretization impact model predictions of flow and sediment yield (e.g. Bingner et al., 1997; FitzHugh and Mackay, 2000; Chen and Mackay, 2004). In SWAT, the watershed is discretized into sub-basins and these sub-basins are typically further discretized into Hydrologic Response Units (HRUs). Chen and Mackay (2004) highlight the inconsistency in SWAT2000 caused by the integration of the HRU concept with the Modified Universal Soil Loss Equation (MUSLE) for predicting sediment erosion on HRUs. They recommend modelling with small, single HRU sub-basins. In contrast, our study investigates the importance of refined spatial data input representation for a case study where the watershed discretization level was already largely determined by other factors.

The proper inclusion and representation of available watershed specific spatial data is crucial in this case study given the importance of the modelling analyses to both Delaware County and NYC. Alternative SWAT2000 input development methodologies in this study included estimating HRU specific slopes, spatially variable initial soil P levels, spatially variable groundwater soluble P inputs and spatially variable daily manure application rates. In addition, the accuracy of soil property inputs are improved by utilizing higher resolution soils data (SSURGO). Most of these alternative methodologies use as inputs national (US) or regional databases and are therefore particularly relevant for future SWAT model applications. The model input development considerations outlined in this work also apply to other spatially distributed watershed models that are applied to agricultural watersheds.

## Methodology

### Case study: Cannonsville Watershed

The major land uses in the 1178 km$^2$ Cannonsville Watershed are forests (59% of the land area), agricultural grasses (26% of land area) and successional farmland (10%). Corn crops account for 1% of the land area while urban areas comprise less than 0.5% of the watershed. Agricultural lands are dominantly dairy farms. Mean annual precipitation at the Walton and Delhi climate stations is about 1100 mm/yr. The elevation of the watershed ranges from approximately 300 m above mean sea level in the lowland areas to approximately 1000 m in the uplands while the average land-surface slope is 19%.

### SWAT2000 model description

SWAT2000 (Neitsch et al., 2001a) is a distributed-parameter model designed to compute long-term runoff and nutrient export from rural watersheds, especially those dominated by agriculture (Arnold et al., 1998). The model is maintained by the Agricultural Research Service of the United States Department of Agriculture (USDA). Although the model simulates on a daily time-step when the SCS curve number method is used to calculate surface runoff, SWAT was designed as a long-term yield model and is not designed to accurately simulate detailed, single-event flood routing (Neitsch et al., 2001a). The base model inputs for SWAT2000 are developed with the aid of the SWAT2000 Arcview© GIS Interface (AVSWAT2000) program (DiLuzio et al., 2001) that automatically assigns default model parameter values and creates input files based on various GIS map layers provided to the interface. During model calibration it was apparent that modifications to the original SWAT2000 model were required to adequately simulate flow and water quality in winter months. These modifications are discussed further in Section 'Modifications to SWAT2000 equations'.

### Model input development

The SWAT2000 model of the Cannonsville Reservoir Watershed described in this study was derived from a preliminary model version for flow and sediment transport (no phosphorus) described in Benaman et al. (2005), which is also applied in Benaman and Shoemaker (2004, 2005). Details of the substantial input development effort for this version of the model are minimized in the following sections but interested readers should refer to the modelling report by Tolson and Shoemaker (2006) for a thorough description of all input development steps. Unless otherwise noted below, all model inputs are assumed constant over the modelling simulation time period considered (1987–2001).

#### Watershed configuration

SWAT divides the entire watershed into sub-basins and the sub-basins can be further subdivided into HRUs. A digital elevation map (DEM), soil, land use and stream network coverage were input to AVSWAT2000 in order to create base SWAT2000 model inputs. The sub-basins defined for this study followed those designated by the New York City Department of Environmental Protection (NYCDEP). Thirteen additional sub-basins were defined along the mainstem of the West Branch Delaware River (WBDR) in order to isolate the urban areas and also to match more closely the available flow and water quality monitoring stations. A total 43 sub-basins (Fig. 2) were delineated in AVSWAT2000 using a 25 m DEM provided by NYCDEP and US Census TIGER files for the Cannonsville Basin stream network definition.

Within each sub-basin, HRUs in AVSWAT2000 are formed as unique soil and land use combinations that are not necessarily contiguous land parcels. AVSWAT2000 thresholds for defining HRUs within a sub-basin were set at 5% for soils and 1% for land use. The original AVSWAT2000 generated HRUs were augmented with additional HRUs to ensure that none of the corn land use in the basin was reclassified as other land uses. This was deemed important because corn silage lands were thought to have the highest erosion rates and it was observed that the threshold application for HRU delineation defined only 86% of the total NYCDEP corn silage area.

#### Land use

NYCDEP provided the land use data (25 m grid) that were derived by supervised classification of 1992 and 1993 thematic mapper satellite imagery. Besides the corn land use, the only other agricultural land use category of significant size in the NYCDEP spatial data that was retained for modelling was grass. The most common agricultural land uses in the basin are known to be continuous hay, pasture, hay in rotation with corn silage and abandoned/idle agricultural land. In order to represent these four types of grass land uses independently in the model, data from the 1992 and 1997 US Census of Agriculture, as well as available local land information, were utilized to guide the subdivision of the initial NYCDEP grass land use HRUs into continuous and rotated hay, pasture and idle agricultural grass HRUs.

According to local agricultural planners, the majority of watershed corn is rotated with hay. For modelling purposes, it was assumed that all rotated corn in the watershed is rotated with hay HRUs on the same soil type as the corn. Local agricultural planners report the typical rotation on upland HRUs as 2 years corn and 8 years hay and the typical rotation on lowland HRUs (i.e. valleys) as 3 years corn and 6 years hay. In order to model a constant area of corn for all years

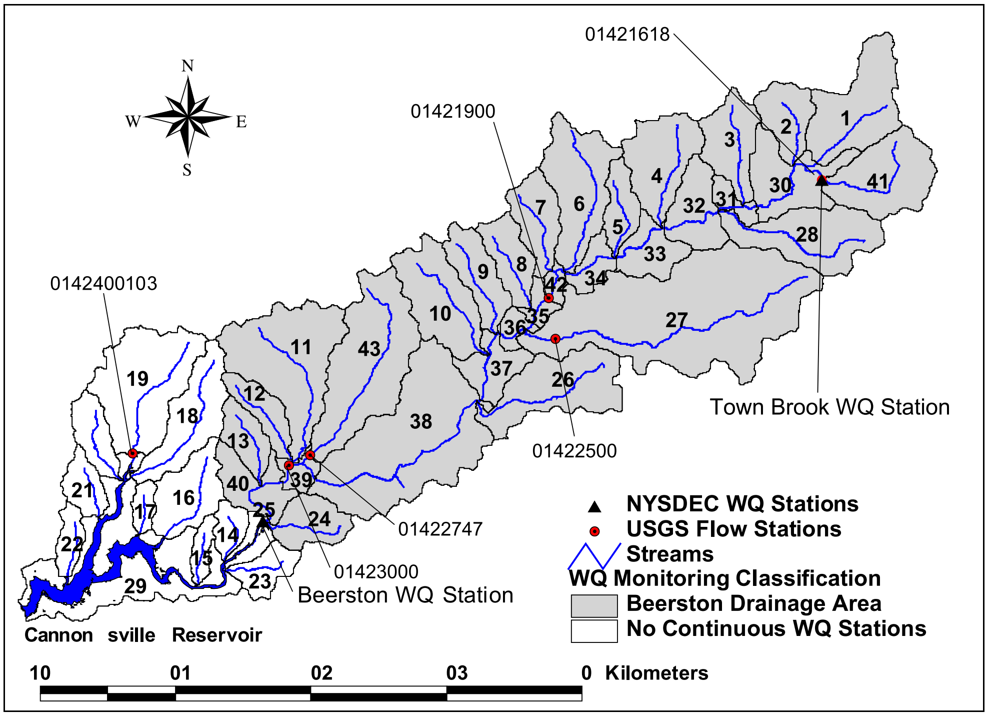


**Figure 2** Sub-basins and continuous flow and water quality (WQ) monitoring stations in the Cannonsville Watershed.

in the model simulations, additional HRUs for rotation were defined by reclassifying portions of the NYCDEP grass HRUs based on the years in rotation. For example, if an upland corn HRU occurred within a sub-basin, then four additional HRUs, each equal to the size of the original corn HRU, were defined and removed from the NYCDEP grass HRU on the same soil type as the original corn HRU. After defining these rotation HRUs, the remaining NYCDEP grass HRUs were reclassified as continuous hay, pasture or idle agricultural land to best approximate the available local information and US Census of Agriculture data. In total, 758 HRUs represent the Cannonsville Watershed.

### Soil property inputs

SWAT2000 by default uses the State Soils Geographic Database (STATSGO) to describe the physical characteristics of soils. STATSGO spatial data (1:250,000) from the USDA - Natural Resource Conservation Service (NRCS) were utilized to map the 12 New York STATSGO soils occurring within the basin. The default AVSWAT2000 approach assigns soil properties to model HRUs based on the most common soil component (or soil series) within each STATSGO map unit. This approach resulted in soil properties that were not representative of the true area-weighted average soil properties across the STATSGO map unit. For example, the most common soil series in each STATSGO map unit usually accounted for only 10—20% of the total area defined by each STATSGO map unit and the soil properties of the disregarded soil series were often vastly different than the most common series.

In order to assign more representative Cannonsville Watershed specific soil properties while minimizing the number of HRUs modelled, the areas of each SSURGO soil map unit within each STATSGO map unit were tabulated and used to compute area-weighted average soil properties for each STATSGO map unit. The result of this approach was to estimate soil properties for the STATSGO map units that were consistent with average watershed soil properties that would be derived independently using only the more detailed SSURGO soils data.

SSURGO soil survey spatial data (version 2) and tabular soil property data (version 1) for Delaware County, New York (NY025) from the USDA-NRCS (http://soildatamart.nrcs.usda.gov/) were used for this analysis. First, base SWAT2000 soil input properties for each SSURGO map unit were computed. Spatial and depth weighted average soil properties were then computed for each STATSGO map unit from the soil properties of the constituent SSURGO map units. The depth averaging of soil properties required idealized soil layer boundaries to be defined. All soils in the watershed were represented with four layers since approximately 50% of all SSURGO soil series in the watershed contained four soil layers. Based on the ranges of the tabular SSURGO data, lower and upper bound estimates for all soil properties except profile depth were calculated. This analysis enabled the estimation of more spatially representative area-weighted averages for soil P levels and surface runoff curve numbers since these HRU-based parameters could each be associated with SSURGO map units.

### Climate data

Daily climate inputs utilized in this SWAT2000 model application were minimum and maximum temperature, precipitation, solar radiation and relative humidity. All four of these inputs were based completely on measured data within or close to the Cannonsville Basin. Data sources were either the Northeast Regional Climate Center at Cornell University or the National Climatic Data Center, with precedence given to data from the former when both sources were available for the same day. Temperature data from two stations (Walton and Delhi) within the basin were available while precipitation data from three stations (Walton, Delhi and Stamford) within the basin were available. Solar radiation and relative humidity inputs were based on climate station data from Binghamton, NY, approximately 60 km east of the reservoir/watershed outlet. In this application, orographic temperature adjustments were made for elevation differences between the climate stations and the sub-basins.

### HRU slope inputs

By default, AVSWAT2000 calculates average sub-basin slopes and then assigns each HRU within a sub-basin the average sub-basin slope. In the Cannonsville Watershed, the lowland valley bottoms are largely agricultural land while the steep upland hill slopes tend to be covered in forests. Averaging the valley bottom slopes with the upland hill slopes within a sub-basin artificially increases the slopes of agricultural land and decreases the slopes of forested land. SWAT2000 erosion predictions are calculated with the MUSLE equation and are a function of HRU slopes. Therefore, HRU specific slopes were estimated separate from AVSWAT2000 and utilized in this study.

### Agricultural management representation

All agricultural management and activities input to the model were derived from information provided by multiple local farm planners and agricultural researchers. Cattle farming (90% dairy and 10% beef) dominate all other types of agriculture in the watershed and thus are the focus of generalized agricultural model inputs. Agricultural management inputs included growing corn and hay in rotation, dates for harvest and tillage of corn fields and specifying harvested biomass rates for hay and pasture such that the total watershed-wide harvested biomass approximated the estimated watershed-wide cattle population dry matter intake derived from locally grown crops.

Manure production from the basin cattle was by far the biggest source of P to account for in the model inputs. Manure production and distribution to HRUs was approached from a mass balance perspective within each sub-basin. A local NRCS agricultural survey in Cannonsville Watershed conducted from 1992 to 1994 was used as the basis for estimating spatially (by sub-basin) and temporally (by year) variable cattle populations. NRCS survey results are assumed to represent 1992 model inputs. The US Census of Agriculture survey for Delaware County (containing the Cannonsville Watershed) shows the dairy cow animal count decreased by 43% between 1987 and 2002. Therefore, it was deemed important to represent this temporal trend and the static NRCS derived watershed cattle population data for 1992 were extrapolated to 1987, 1997 and 2002 based on the US Census of Agriculture reported cattle populations for Delaware County. Annual cattle populations were then linearly interpolated for all remaining years in the simulation period.

Within each sub-basin, based on the cattle population in a given year and various Cannonsville Watershed specific manure production factors (animal weight, manure production per day, manure N and P content by season), daily sub-basin specific manure loads were calculated and then allocated to HRUs. Since very few farms in the watershed had manure storage capacity during the model simulation period, manure in the model was applied daily to HRUs. Pasture HRUs received daily manure inputs during the grazing season. All remaining manure was allocated on a daily basis to corn, hay and sometimes pasture HRUs according to a set of seasonally specific manure spreading rules that generalize the behavior of area farmers.

Based on information from local farm planners and nutrient specialists, starter inorganic fertilizer was applied each year to all corn silage HRUs on the same day that corn is planted. Inorganic fertilizer application rates were 39.3 kg/ha of N and 17.1 kg/ha of P and 99% of these inputs were incorporated to the soil layer under the top 10 mm of soil.

### Additional nutrient inputs

Model input development focused mainly on P since P transport was the ultimate modelling goal of this study. P inputs were loaded to watershed soils or surface waters in the model from point source P discharges, atmospheric P deposition, inorganic and organic agricultural fertilizer application and groundwater. Watershed specific data on organic and mineral N levels in cattle manure were incorporated into model inputs.

Soil test P data from across the watershed and collected by the Watershed Agriculture Program for the NYC Watersheds and then analyzed by the Cornell Nutrient Analysis Laboratory formed the basis of initial soil P (labile or soluble P) concentration estimates for agricultural land covers. Each of these 2800 entries of soil test P (Morgan's P) had their spatial references removed but included a corresponding date, crop cover and SSURGO soil series. The crop cover and soil series information enabled averaging across HRU types and a total of 31 different initial soil P levels were assigned to agricultural HRUs. Based on soil nutrient expert Dr. Andrew Sharpley's recommendations (Personal Communication), Morgan's P was converted to SWAT labile P inputs assuming 1 mg labile P/kg soil equals 0.4 mg Morgan's P/kg soil. Relatively limited watershed forest soil sampling data were used to specify initial forest and grass-shrub land use labile soil P levels that were assumed constant across all watershed soil types. For all land uses, organic P levels were determined as a function of HRU labile P concentration and assumed organic P fractions of total P.

Constant groundwater soluble P concentrations are an input for each HRU in SWAT. NYCDEP bi-weekly water quality grab samples for dissolved P from across the watershed were analyzed when streamflow was dominated by baseflow and, after accounting for point sources of P, assumed to represent spatially variable groundwater soluble P inputs. Eleven spatially variable groundwater soluble P values were estimated from this analysis and the resulting area-weighted watershed average groundwater soluble P concentration was 12 $\mu$g/L.

Five point sources discharging to surface waters in the watershed were considered in this study. Monthly data from the New York State Department of Environmental Conservation (NYSDEC) and the State Pollutant Discharge Elimination System Discharge Monitoring Reports were utilized to define monthly loading rates for flow, sediment and P from the point sources. Point source P was partitioned to mineral and organic P fractions according to limited NYSDEC data and Longabucco and Rafferty (1998). Point source total P loads account for approximately 19% of the monitored total P load for the watershed.

## Modifications to SWAT2000 equations

In the initial stages of model calibration, it became apparent that some alternative model forms would be necessary in this specific case study to improve representation of watershed characteristics and adequately simulate the measured data. The two most critical changes are discussed in detail in the following two sections and are both relevant for the Northeastern US and other regions with freezing winter temperatures. All other relatively minor modifications to SWAT2000 for this case study are detailed in Tolson and Shoemaker (2006).

### Soil water excess in frozen soils

In SWAT2000, for any nonfrozen soil layer, when the soil water content exceeds the field capacity of the soil, the excess soil water is partitioned between lateral flow and percolation to the next soil layer. Percolation from the last soil layer goes to groundwater. When a soil layer is calculated to have a temperature below freezing, the original SWAT2000 approach assumes: (1) no lateral flow occurs, (2) the soil layer can hold excess water above field capacity until saturation, (3) in frozen saturated soil layers, water draining from above passes through the frozen saturated layer and drains into the next soil layer.

As a result, when the soil is frozen but not saturated, all infiltrated water up to saturation is held in the profile and thus does not reach the channel via groundwater return flow or lateral flow until after the soil temperature increases above freezing.

After a multitude of attempts at varying the prescribed model flow parameters during the model calibration of baseflow and streamflow, it was observed that the SWAT2000 approach described above was responsible for holding too much infiltrated water in the soil in the winter and thus eventually transferring too much of the winter precipitation/snowmelt to groundwater storage under all reasonable flow parameter estimates. The resultant delay of lateral flow and groundwater delivery addition to streamflow caused consistent cumulative winter flow under-predictions and corresponding baseflow over-predictions late into the summer. Winter streamflows could only be simulated adequately when streamflows included significant volumes of lateral and/or groundwater flow (which was not possible with the original SWAT2000 approach). For example, simply increasing the surface runoff volumes to reduce infiltration volumes in the winter (via an increase in the SCS curve numbers) produced unreasonably high daily peak flows relative to the measured data. Therefore, an alternative approach to soil water excess partitioning in frozen soils was used in this study that simply involved the simulation of percolation and lateral flow in frozen soils in the same way as unfrozen soils (i.e. percolation and lateral flow are unaffected by soil temperature). Although results will show this approach improved predictions in this case study, it is not likely applicable in all frozen soils since the infiltration rate of frozen soils depends on other factors such as soil texture and structure.

### Soil erosion under snow cover

Initial sediment calibration attempts showed that the default SWAT2000 equation for the reduction of soil erosion under snow cover (Eq. (13.3.1) in Neitsch et al., 2001a) was unsuitable for prediction of HRU sediment erosion under snow cover in the Cannonsville Watershed. The SWAT equation in question reduces simulated erosion based only on the depth of snow water equivalent on the soil surface. The amount of surface runoff (i.e. 1 mm versus 50 mm) has no impact on the effectiveness of the snow cover to reduce erosion. Therefore, an alternative approach to winter soil erosion prediction that accounts for both snow cover depth and surface runoff volume was developed for this study. Eq. (13.3.1) in Neitsch et al. (2001a), given again here as Eq. (1) below, was replaced with Eq. (2) below for the prediction of daily soil erosion in each HRU:

$$sedyld = sed \Big/ \exp\left[\frac{3Sno}{25.4}\right] \tag{1}$$

$$sedyld = sed(SurQ/[SurQ + F(Sno)]) \tag{2}$$

where *sedyld* is the sediment yield (metric tonnes) eroded from the HRU for the day, *sed* is the predicted sediment yield (metric tonnes) before the effects of snow cover are accounted for, *SurQ* is the depth of surface runoff (mm) for the day, *Sno* is the snow cover in water equivalent depth (mm) for the day and *F* is a positive calibration factor that controls the influence of snow cover depth on sediment yield.

## Flow and water quality data

The locations of the measured flow and continuous water quality stations supplying calibration and validation data are provided in Fig. 2. Up to a total of 11 years of daily flow data from six United States Geologic Survey (USGS) stations were utilized for flow calibration and validation. The Walton gauge (01423000) monitors the largest drainage area in the basin (860 $km^2$) and was therefore the main focus of the flow calibration. Up to a total of nine years of daily water quality loading data for total suspended sediment (TSS), total dissolved phosphorus (TDP) and particulate phosphorus (PP) were utilized for water quality calibration and validation. The Beerston water quality station drains most of the watershed (913 $km^2$) and was the main focus of water quality calibration. The daily water quality loading data derived from event-based water quality sampling (see Longabucco and Rafferty, 1998) was provided by NYSDEC. In addition, the NYCDEP provided a spatially distributed water quality dataset of bi-weekly P concentrations covering a 10-year period.

## Model calibration and validation

Model calibration is the adjustment of model parameters, within recommended ranges, to optimize the agreement between measured data and model simulation results. The term 'validation' is used for clarity and historical reasons only. Validation is taken to mean 'model testing' and by no means do we consider our validated model to be a perfect predictor. Rather, good validation (model testing) results are simply stronger evidence that the calibrated model is a good simulator of the measured data and does not over-fit the measured data in the calibration period. The start date for all model simulations was January 1, 1987. Model outputs were not printed until January 1, 1994 (calibration run) or January 1, 1990 (validation run). The three year initialization or start-up period was used so the impacts of uncertain initial conditions in the model were minimized. The calibration period was January 1994 to September 2000. Two distinct time periods that enclose the calibration period were used for model validation. The first validation period spans January 1990 to December 1993 while the second validation period was water year (WY) 2001.

Calibration efforts focused on improving model predictions at the main watershed monitoring stations (Beerston/Walton). Model predictions at other flow and water quality stations were not calibrated independently. Flow calibration was focused on daily predictions. Although daily P and TSS load estimates were available at Beerston, their uncertainty was higher than the measured flow uncertainty due to reduced sampling frequency. Therefore, water quality calibration focused mainly on monthly loading predictions rather than daily predictions. For each constituent of interest, model performance is qualitatively evaluated with time series plots and quantitatively evaluated using three model performance statistics. The coefficient of determination ($r^2$) and the Nash–Suttcliffe coefficient ($E_{NS}$) (Nash and Suttcliffe, 1970) were used to quantitatively assess the ability of the model to replicate temporal trends (daily and monthly) in measured data. The %Bias is defined as the relative percentage difference between the average simulation and measured data time series over $n$ time steps and is given in Eq. (3):

$$\%\text{Bias} = \frac{100 \times \left(\sum_{j=1}^{n}\text{Simulated}_j - \sum_{j=1}^{n}\text{Measured}_j\right)}{\sum_{j=1}^{n}\text{Measured}_j} \qquad (3)$$

The calibration objective for each constituent of interest was to maximize the $E_{NS}$ coefficient while simultaneously attempting to reduce the absolute value of %Bias to values ideally less than 10%. In general, the model was calibrated first for flow, then summer TSS, winter TSS (using the $F$ factor in Eq. (2)), total P and finally, TDP and PP. Total P is the sum of TDP and PP. The baseflow separation technique in Arnold and Allen (1999) was used to estimate the baseflow component of measured and calibrated model flows at Walton (the main flow gauging station). Higher absolute %Bias values for TDP and PP were deemed acceptable when total P absolute %Bias values were relatively low. Measured TDP was compared to the SWAT mineral Pa output while measured PP was compared to the SWAT organic P output.

In addition to the model changes described in Section 'Modifications to SWAT2000 equations', model calibration involved the selection of various process representation options. The Priestly–Taylor option for estimating potential evapotranspiration was selected because it resulted in the best late summer early fall low flow predictions of the three available options. Surface runoff was simulated using the daily runoff curve number option. Channel sediment erosion parameters were set to values that allowed for small amounts (<5% of total sediment load) of stream channel bed and bank erosion. In-stream nutrient processes were not simulated (optional in SWAT2000) because attempts to simulate these resulted in very large and unexpected increases of total P loading at the Beerston water quality station under default in-stream reaction parameters and adjusting the in-stream reaction parameters did not result in acceptable simulation results for P. As a result, the calibrated model routes nutrient loadings downstream without simulating in-stream organic P mineralization to soluble P, organic P settling to the streambed or the conversion between organic P and soluble P due to algal activity.

The set of final model parameters modified from their default values during calibration was determined in two general stages. The first involved the adjustment of various 'data-driven' parameters (or model inputs) to better represent known or assumed conditions or physical properties of the watershed. These adjustments were made independent of their impacts on the predictive accuracy of simulated flow and water quality. Examples of data-driven model parameter settings include assigning the P contents of corn silage and harvested hay based on watershed crop nutrient analyses and fixing all initial soil property inputs to the midpoints of their ranges as determined from the analysis described in Section 'Soil property inputs'. In the second stage, a subset of model parameters was identified as 'performance optimization' parameters and their values were iteratively adjusted by trial-and-error to best match measured flow, sediment and P data. The performance optimization parameters were selected based on SWAT2000 calibration recommendations in Neitsch et al. (2001b) and parameter sensitivity information reported in Benaman (2002).

The majority of the performance optimization parameters were flow related. The most important parameter adjustments in this group were the uniform reduction of all base runoff curve numbers by 20% of their base value and the reduction of the surface runoff lag coefficient (SURLAG) to 1.0. These two modifications respectively functioned to reduce surface runoff volumes and increase the delay in the delivery of surface runoff to the channel. Tolson and Shoemaker (2006) provide a complete list of all data-driven and performance optimization parameters.

## Results

The results are presented here in four sections. Section 'Calibration' compares simulated results and measured flow and water quality data over the 1994–2000 calibration period. Then, the sensitivity of calibrated model performance to some of the alternative inputs and model changes outlined in Sections 'Model input development' and 'Modifications to SWAT2000 equations' are evaluated in Section 'Sensitivity analyses'. Next, simulated results and measured flow and water quality data over the 1990–1993 and then WY 2001 validation periods are compared in Section 'Validation'. Lastly, the relationship between daily flow prediction errors and daily total P load prediction errors over the entire model simulation period (1990—2001) is assessed in Section 'Total phosphorus prediction error analysis'.

## Calibration

Daily simulated and measured flows at the main USGS gauging station (Walton) are compared in Fig. 3 for two WYs in the calibration period. Fig. 3 shows the range in annual predictive flow performance with WY 1996 having the worst daily $r^2$ for flows (0.73) while WY 1997 had the best daily $r^2$ (0.94) for flows. Fig. 3 results are representative of the general findings that although the simulated flows usually matched the trend in the measured flows quite well, late summer and fall flows were often over-predicted. Baseflow calibration results at Walton also showed good agreement. The fraction of streamflow that was baseflow calculated with the approach in Arnold and Allen (1999) was between 0.42 and 0.66 for the measured flows and between 0.40 and 0.67 for the simulated flows.

Flow predictions are quantitatively compared to measured flows in Table 1 for Walton and the other five USGS gauging stations utilized in this study. The daily $E_{NS}$ coefficient for Walton flows of 0.80 shows good general agreement between simulated and measured flows. Walton flows were only slightly over-predicted (%Bias of 1.7%). Although the model parameters were not independently calibrated for any of the five smaller USGS drainage areas, predictions were acceptable as the daily $E_{NS}$ coefficients ranged from 0.59 to 0.78 and the absolute value of the %Bias for all but one location was less than 6.4%. Monthly flow $E_{NS}$ coefficients at all USGS locations were all between 0.80 and 0.89 except for East Brook (0.70). Performance statistics were very similar when calculated for only the winter (November—April) and summer months (May—October).

Simulated monthly total suspended sediment (TSS) and phosphorus (P) loads at the Beerston water quality station are compared with measured data in Fig. 4 for the calibration period. Fig. 4 compares different P fractions (TDP, PP and total P). For TSS, TDP, PP and total P, the general trends in the measured data were replicated fairly well in the simulated results. Simulated peak TSS loads almost always substantially under-predicted the measured data. All attempts to increase simulated peak monthly TSS loads during calibration increased %Bias to unacceptable levels. Potential causes of these TSS prediction errors are outlined in the 'Discussion' Section. PP results are very similar to TSS results since PP is attached to sediments. The model more accurately simulated the trend in and peak TDP loads in comparison with TSS, PP and total P. Total P results in Fig. 4 show reasonably good agreement in most months except for the sizable under-predictions in Jan-96, Nov-96, Jan-98, Jul-98 and Feb-00 which are further analyzed in Section 'Total phosphorus prediction error analysis'.

On January 19th and 20th of 1996, large rainfall depths combined with a deep snow pack, unseasonably warm temperatures and high winds produced incredible amounts of rainfall and snowmelt induced runoff. The corresponding flood event was the largest streamflow on record for many locations in the watershed and had an estimated return period of 70—100 years (Lumia, 1998). Widespread flooding initiated transport of huge amounts of floodplain P and sediment. Longabucco and Rafferty (1998) report that the two-day event accounted for approximately 75% of the TSS and PP loads recorded for the entire year. Since SWAT is not designed to simulate such an extreme event, the calibration effort focused on improving model performance statistics that did not include measured and simulated TSS and P loads during January 1996. Accordingly, unless otherwise noted, all TSS and P calibration performance statistics reported in this paper do not include measured and simulated data for January 1996.

Simulated monthly TSS and P loads at the Beerston water quality station are quantitatively compared with measured data in Table 2. Although the performance statistics dis-

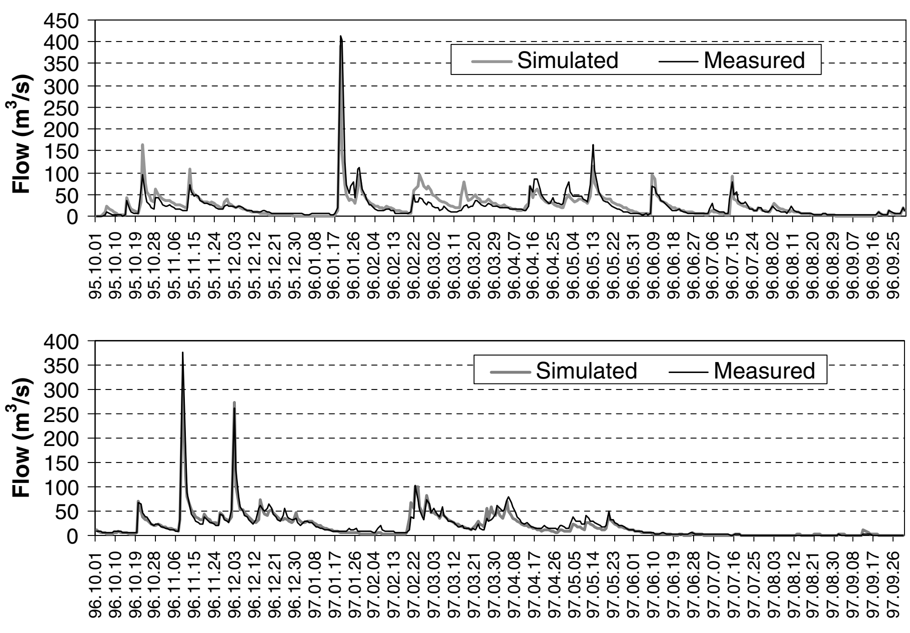


**Figure 3** Measured and simulated daily flows at the Walton USGS station (01423000) for the water years with the lowest (1996, $r^2$ = 0.73) and highest (1997, $r^2$ = 0.94) daily $r^2$ values in the calibration period.

**Table 1** Daily flow calibration results at six USGS gauge stations for the period 1994–2000

| Location (USGS gauge) | Period | Mean measured data ($m^3/s$) | Mean simulated results[a] ($m^3/s$) | %Bias[b] | $r^2$ | $E_{NS}$ |
|---|---|---|---|---|---|---|
| WBDR Walton (01423000) | Jan. 1994–Sept. 2000 | 17.8 | 18.1 | 1.7 | 0.80 | 0.80 |
| WBDR Delhi (01421900) | Dec. 1996–Sept. 2000 | 7.1 | 6.9 | −2.2 | 0.75 | 0.75 |
| Little Delaware (01422500) | Jan. 1997–Sept. 2000 | 2.8 | 2.6 | −6.1 | 0.79 | 0.78 |
| East Brook (01422747) | Oct. 1998–Sept. 2000 | 1.1 | 1.3 | 17.5 | 0.64 | 0.59 |
| Trout Creek (0142400103) | Dec. 1996–Sept. 2000 | 1.1 | 1.0 | −1.3 | 0.67 | 0.67 |
| Town Brook (01421618) | Oct. 1997–Sept. 2000 | 0.8 | 0.8 | −6.4 | 0.64 | 0.63 |

See Fig. 2 for USGS gauge locations.
[a] Calculated based only on the days with measured data for each station.
[b] Calculated with higher precision than the reported means.

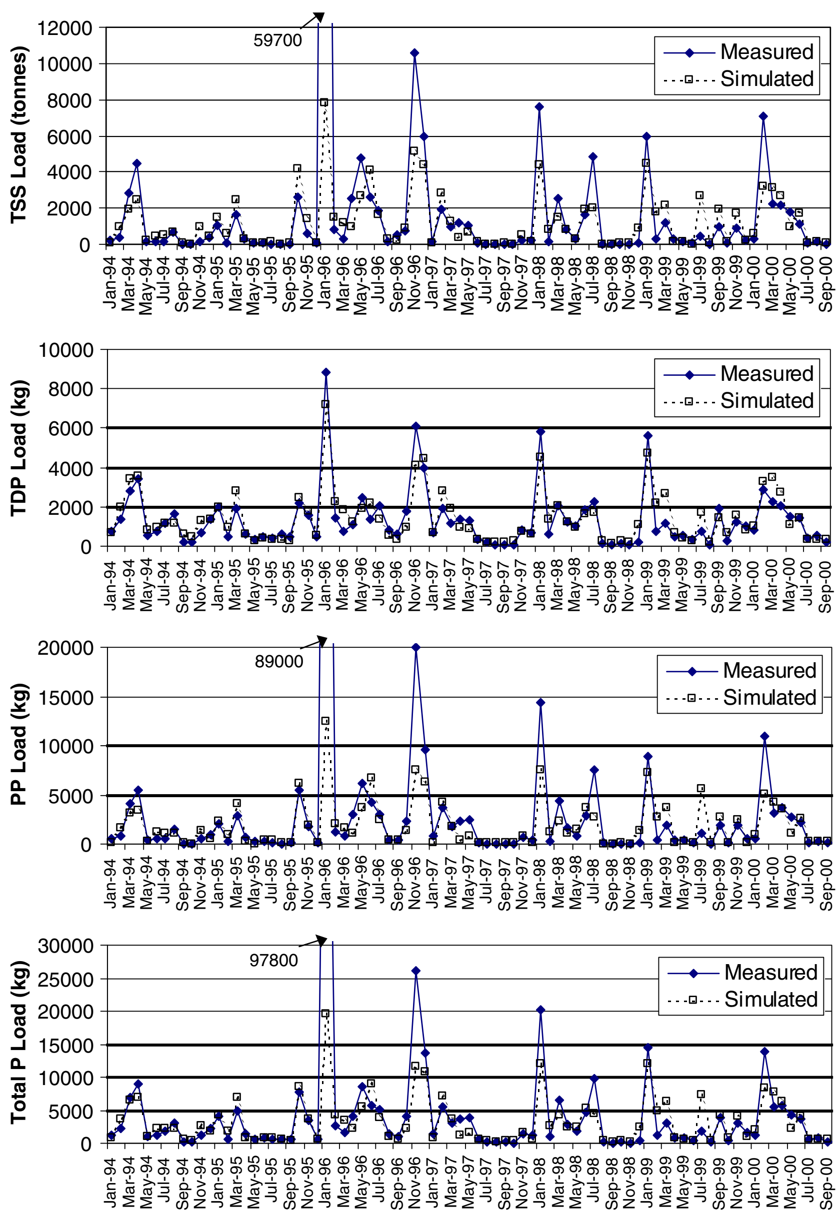


**Figure 4** Measured and simulated monthly loads of total suspended sediment (TSS), total dissolved phosphorus (TDP), particulate phosphorus (PP) and total P (TDP + PP) at Beerston water quality station over the calibration period.

**Table 2** Monthly sediment (TSS) and phosphorus (P) performance statistics for the calibration period (Jan 1994–Sept 2000) at Beerston with and without[a] Jan 1996

| Water quality constituent | Mean monthly measured data | Mean monthly simulated results | %Bias | $r^2$ | $E_{NS}$ |
|---|---|---|---|---|---|
| TSS (tonnes) | 1204 | 1121 | −6.9 (−37.5)[a] | 0.70 (0.47) | 0.67 (0.24) |
| TDP (kg) | 1244 | 1355 | 16.1 (6.7) | 0.79 (0.84) | 0.78 (0.84) |
| PP (kg) | 2094 | 1742 | −19.3 (−40.8) | 0.67 (0.50) | 0.61 (0.26) |
| Total P (kg) | 3338 | 3097 | −6.1 (−26.7) | 0.73 (0.58) | 0.70 (0.37) |

[a] All parentheses enclose statistic that includes January 1996.

cussed in this paragraph do not include January 1996, the statistics in Table 2 are given with and without January 1996. TSS and total P predictions were within 7% of the measured loads. Although TDP and PP %Bias measures were not as accurate as flow, the discrepancies were deemed reasonable given the good results for total P. The monthly $E_{NS}$ coefficients for total P and TSS were 0.70 and 0.67, respectively. Performance statistics were very similar when calculated for only the winter (November–April) and summer months (May–October). Although calibration was focused on monthly $E_{NS}$ values, the daily $E_{NS}$ coefficients achieved for sediment and total P, 0.57 and 0.56, respectively, were also acceptable.

### Spatial phosphorus calibration

Regular, bi-weekly grab samples for the concentration of total P taken by the NYCDEP were available for many locations in the watershed during the calibration period. Since these grab samples were nearly always taken on the same day across the watershed, their arithmetic average across the calibration period is a fairly reliable measure of the relative variation in magnitudes of total P concentrations in surface waters across the watershed.

These arithmetic averages for the measured total P concentrations along the mainstem of the WBDR are compared with simulated flow-weighted total P concentration averages in Fig. 5 for both the calibration and validation period. With respect to calibration period results, simulated flow-weighted total P averages were always higher than measured total P. Based on the different nature of each average this behavior was expected. NYCDEP water quality data are bi-weekly and therefore tend to miss sampling during most of the high flow/high total P concentration events (lowering the average). In contrast, the model simulates the entire

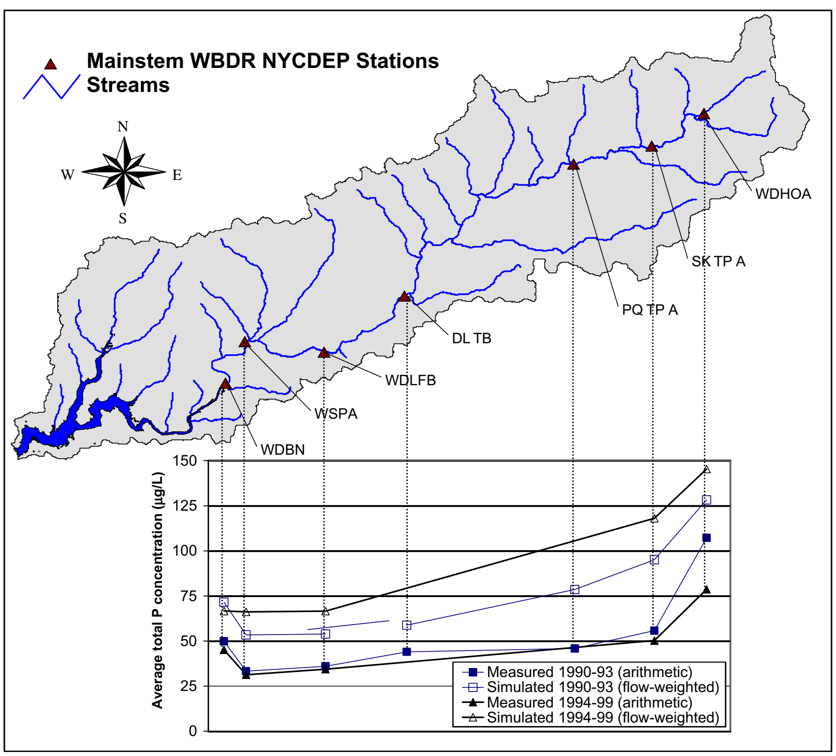


**Figure 5** Simulated and NYCDEP measured average total P concentrations for selected NYCDEP mainstem WBDR water quality stations during the calibration period (Jan. 1994–Dec. 1999 only) and validation period (Jan. 1990–Dec. 1993).

period, and therefore the flow-weighted averages over the simulation period encapsulate all simulated high flow/high concentration events. More importantly, Fig. 5 shows that during the calibration period, the trend of the measured total P concentrations moving downstream was simulated quite well in the model. In addition to the mainstem WBDR NYCDEP water quality stations in Fig. 5, adequate data were available for a few headwater WBDR tributaries. Considering these headwater and mainstem NYCDEP water quality stations, sub-basin 18 showed the lowest average total P concentration during the calibration period. The model correctly identified sub-basin 18 as the NYCDEP monitoring location with the lowest average calibration period total P concentration.

## Sensitivity analyses

Model calibration involved a series of analyses to derive inputs, iteratively adjust parameter values and modify the SWAT2000 model equations. The sensitivity of model predictions and performance to the most critical model input derivation methods (Sections 'Soil property inputs' and 'HRU slope inputs') and model equation modifications (Section 'Modifications to SWAT2000 equations') are highlighted in this section. For each analysis below that involves model simulation results, the calibrated model predictions (from Section 'Calibration') are compared to model predictions under the exact same set of model inputs except that the default SWAT2000 inputs or equations for the issue being analyzed was utilized. In other words, the model was not recalibrated under the default SWAT inputs or equations.

### Impact of alternative model input derivations

*Slope input sensitivity.* The HRU slope input options considered in this case study are discussed in detail in Section 'HRU slope inputs'. Table 3 compares the model predictions under the default SWAT2000 slope input approach with calibrated predictions using HRU specific slopes. Simulated sediment yield was nearly doubled when sub-basin average slopes rather than HRU specific slopes were used. Not surprisingly, the $E_{NS}$ coefficients for TSS at Beerston were also degraded when sub-basin slopes were used. Table 3 results demonstrate the large increase (67%) in nonpoint source (NPS) total P loading in surface runoff when sub-basin average slopes were utilized. Nearly all of this increase in total P loading (93%) was attributable to agricultural lands.

*Soil input sensitivity.* As described in Section 'Soil property inputs', the soil inputs for this application were determined using averaged SSURGO tabular data within each STATSGO map unit. This approach is compared here to the default AVSWAT2000 approach that assigns soil properties to each STATSGO map unit based on the most common soil series found in the STATSGO map unit. The following analysis compares the soil property input values assigned by each approach for only a subset of soil properties.

The derivation of average STATSGO map unit soil properties from the Delaware County SSURGO soils tabular data resulted in lower and upper bound estimates for soil properties since the database provides ranges for almost all soil properties. For example, Table 4 demonstrates the calculated bounds for the rock fragment content of layer 1 (a SWAT soil property input) and for the soil profile field capacity (which is calculated as a function of other soil

**Table 3** Comparison of model predictions under the default SWAT2000 slope input approach with the calibrated predictions using HRU specific slopes

| Performance measure or simulated quantity | HRU slope input source | |
|---|---|---|
| | Sub-basin average slopes (default) | HRU specific slopes |
| Basin-wide sediment yield (tonnes/ha/yr) | 0.267 | 0.145 |
| Monthly and daily $E_{NS}$ for TSS at Beerston | 0.58 and 0.42 | 0.67 and 0.57 |
| Basin-wide surface runoff total P load (kg/yr) | 58,600 | 35,000 |

**Table 4** Comparison of default SWAT2000 and averaged SSURGO within STATSGO map unit soil characteristics for a variety of STATSGO map units within the Cannonsville Watershed

| | Soil profile field capacity (mm) as calculated from soil properties[a] | | | Layer 1 rock fragments (%) | | |
|---|---|---|---|---|---|---|
| | NY132 (58% of basin) | NY129 (24% of basin) | NY136 (2% of basin) | NY132 (58% of basin) | NY129 (24% of basin) | NY136 (2% of basin) |
| Default SWAT2000 | 126 | 237 | 457 | 36 | 9 | 6 |
| Lower bound average SSURGO | 89 | 92 | 106 | 16 | 13 | 11 |
| Upper bound average SSURGO | 304 | 307 | 348 | 60 | 58 | 58 |

[a] Soil profile field capacity calculated in SWAT2000 as a function of the input soil depth, available water content, bulk density and clay content.

property inputs). These explicitly calculated SWAT soil property input ranges provide a sound basis for bounding soil property inputs that are adjusted in calibration. Table 4 shows that the default SWAT2000 soil inputs for some STATSGO map units were not even within the ranges of the more representative average SSURGO soil inputs. Therefore, for a number of soil properties, the default SWAT2000 soil inputs are inaccurate point estimates in this case study.

**Impact of SWAT2000 equation modifications**

*Frozen soils*. The model modification regarding the handling of soil water excess in frozen soils proved to be critical in achieving the high seasonal (summer and winter) and overall model performance statistics for daily flows over the calibration period. Table 5 compares the model predictions under the default SWAT2000 simulation of soil water excess in frozen soils with calibrated predictions under the modified approach (where frozen soils do not change lateral flow or percolation calculations). Under the default model simulation approach, the model predictions for basin-wide surface runoff volume more were more than double the calibrated results. As a result, basin-wide sediment yields nearly doubled. For each of the other five flow calibration locations, the reduction in daily flow $E_{NS}$ under the default SWAT2000 simulation approach relative to the calibrated results was even more severe than the Walton $E_{NS}$ reduction of 0.25 in Table 5. Prior to this model change for frozen soils, all other SWAT parameter modifications during model calibration failed to generate performance statistics close to the final calibrated results.

*MUSLE snow cover adjustment*. The numerical results for the default SWAT2000 MUSLE snow cover adjustment equation, Eq. (1), are shown in comparison to the modified snow cover adjustment equation, Eq. (2), for various surface runoff depths in Fig. 6. Unlike the original SWAT2000 equation, at any given snow cover depth, as surface runoff increases, Eq. (2) calculates a decreasing amount of snow cover induced erosion protection. Fig. 7 compares the simulated TSS results under Eq. (1) and Eq. (2), with $F$ calibrated to 0.15, to the measured TSS loads during the winter months (December–April) across the calibration period. Fig. 7 shows that Eq. (1) caused the grossly under-predicted (near 0) TSS loads during Mar-94, Jan-96 (the most extreme flood event on record), Jan-99 and Feb-00. Prior to replacing Eq. (1) with Eq. (2), all other SWAT sediment parameter modifications during model calibration failed to improve these

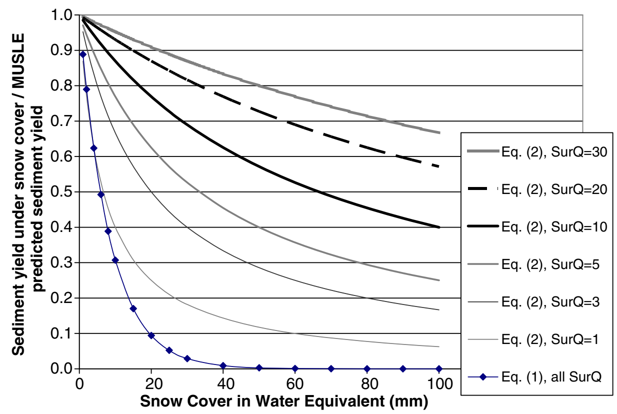


**Figure 6** Comparison of HRU sediment yield adjustments under snow cover between the original SWAT2000 equation, Eq. (1), and modified equation, Eq. (2), with $F = 0.15$, for various surface runoff volumes (SurQ).

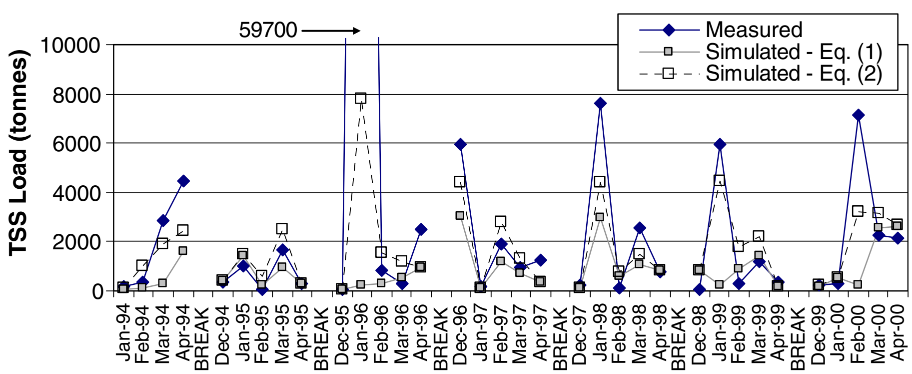


**Figure 7** Comparison of calibration period total suspended sediment (TSS) predictions at Beerston in the months of December through April between the original SWAT2000 equation, Eq. (1), and modified equation, Eq. (2), with $F = 0.15$, utilized in this study.

**Table 5** Comparison of SWAT2000 default and modified approaches to simulation of soil water excess in frozen soils

| Performance measure or simulated quantity | Approach for soil water excess in frozen soils | |
|---|---|---|
| | SWAT2000 default | Calibrated results |
| Daily $E_{NS}$ for Walton flow (1994–2000) | 0.55 | 0.80 |
| %Bias in summer and winter Walton flows | 55 and −18 | −2 and 3 |
| Basin-wide surface runoff (% of total water yield) | 36 | 15 |
| Basin-wide sediment yield (tonnes/ha) | 0.246 | 0.145 |

gross under-predictions. In addition to the improvement of TSS predictions in these four months, Eq. (2) improved winter TSS predictions in the majority of all other months in the calibration period. The monthly $r^2$ value for the winter months (not including Jan-96) improved to 0.77 from 0.47 when Eq. (1) was replaced with Eq. (2).

## Validation

Model validation involves the comparison of model output to measured data for which there has been no adjustment of parameters (because previously calibrated parameter values are used). This is in contrast to calibration, where the parameter values are modified so that predictions best match the measured calibration data. The data set for validation is independent of the calibration data set. Validation performance statistics are typically worse than calibration performance statistics.

Two disjoint time periods that enclose the calibration period were used for model validation. In the first validation period, covering January 1990 through December 1993, simulated and measured flows and water quality predictions at Walton and Beerston are compared. In the second validation period, covering WY 2001, in addition to Walton and Beerston, measured and simulated flows at the other 5 USGS stations in Fig. 2 are also compared because more data were available.

Flow validation results for the 1990–1993 period at Walton are summarized in Table 6 and show that the model performance statistics were very similar and even slightly better than those achieved during calibration. Table 6 also summarizes the model performance statistics for WY 2001 flows at all six USGS stations. Again, performance statistics were generally similar or improved relative to the calibration statistics at the corresponding stations. Two exceptions to this were the relatively lower $E_{NS}$ coefficients for daily flow at two of the three smallest USGS stations (East Brook and Trout Creek). Daily simulated and measured flows at Walton for WY 2001 are compared in Fig. 8. Results for WY 2001 showed simulations replicated the trends in measured flows fairly well although there were a few significant prediction errors between December and March.

Validation results for TSS and P were limited in the first validation period to October 1991 through January 1993 because water quality data were not available before this. Simulated monthly TSS and P loads at the Beerston water quality station are compared with measured data in Fig. 9 for both validation periods. Similar to the results during the calibration period, the model tended to under-predict peak TSS loads in the validation period. PP results were very similar to TSS results. TDP prediction errors were relatively smaller in comparison to PP and TSS during the first validation period. However, simulated TDP loads in Dec-00 and Apr-01 were badly over-predicted. These TDP over-predictions are more closely examined in the 'Discussion' Section. Despite these two TDP over-predictions, the total P loads for these two months were predicted fairly well since the TDP over-predictions were offset by PP under-predictions, especially for April 2001.

Sediment (TSS) and P performance statistics for both validation periods are detailed for Beerston in Table 7. Performance statistics for TSS in both periods were slightly

**Table 6** Daily and monthly flow validation results at six USGS gauge stations for the periods Jan. 1990–Dec. 1993 and WY 2001

| Location (USGS gauge), for WY 2001 unless noted | Mean daily measured data (m$^3$/s) | Mean daily simulated results (m$^3$/s) | %Bias, Eq. (2) | Daily $r^2$ (monthly $r^2$) | Daily $E_{NS}$ (monthly $E_{NS}$) |
|---|---|---|---|---|---|
| WBDR Walton (01423000) 1990–1993 | 16.0 | 16.7 | 4.6 | 0.83 (0.95) | 0.83 (0.94) |
| WBDR Walton (01423000) | 12.3 | 11.8 | −3.9 | 0.86 (0.97) | 0.86 (0.97) |
| WBDR Delhi (01421900) | 5.0 | 4.8 | −5.7 | 0.76 (0.95) | 0.75 (0.95) |
| Little Delaware (01422500) | 1.9 | 1.7 | −11.9 | 0.88 (0.97) | 0.88 (0.96) |
| East Brook (01422747) | 0.9 | 1.0 | 5.3 | 0.69 (0.89) | 0.44 (0.89) |
| Trout Creek (0142400103) | 0.7 | 0.8 | 10.2 | 0.70 (0.88) | 0.43 (0.88) |
| Town Brook (01421618) | 0.6 | 0.5 | −16.5 | 0.75 (0.91) | 0.75 (0.91) |

See Fig. 2 for USGS gauge locations.

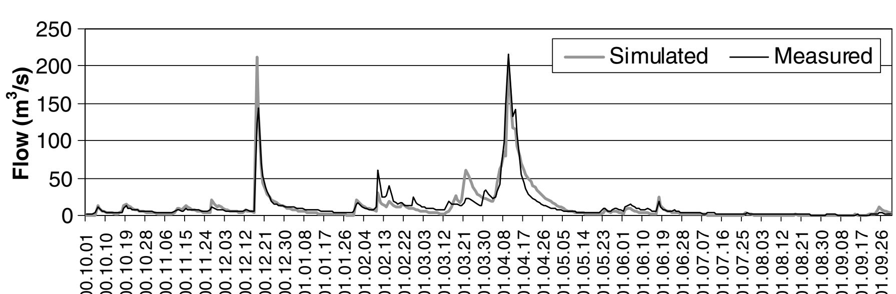


**Figure 8** Measured and simulated daily flows at Walton USGS station (01423000) for WY 2001 of the validation period.

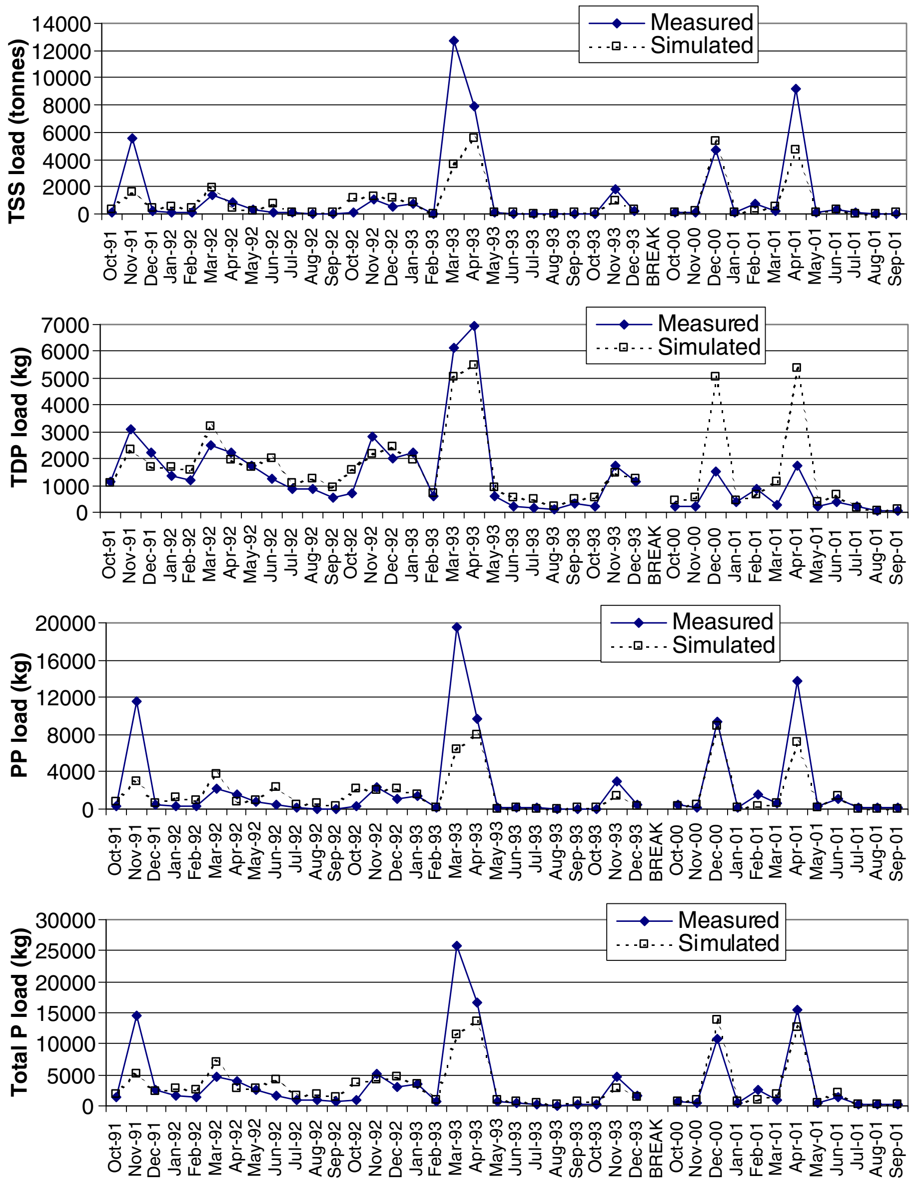


**Figure 9** Measured and simulated monthly loads of total suspended sediment (TSS) total dissolved phosphorus (TDP), particulate phosphorus (PP) and total P (TDP + PP) at the Beerston water quality station over both validation periods.

degraded in comparison to the calibration data. In particular, the TSS %Bias errors were more severe than in calibration. TDP performance statistics for the first validation period were substantially better than calibration results while statistics for WY 2001 were severely degraded for $E_{NS}$ and %Bias. The total P performance statistics in the first

**Table 7** Sediment (TSS) and phosphorus (P) performance statistics in both validation periods (Oct. 1991–Dec. 1993 and WY 2001)

| Water quality constituent (Period) | Mean monthly measured data | Mean monthly simulated results | %Bias | Monthly $r^2$ (daily $r^2$) | Monthly $E_{NS}$ (daily $E_{NS}$) |
|---|---|---|---|---|---|
| TSS (Oct. 1991–Dec. 1993) | 1266 tonnes | 827 tonnes | −34.6 | 0.72 (0.42) | 0.52 (0.33) |
| TDP (Oct. 1991–Dec. 1993) | 1658 kg | 1693 kg | 2.1 | 0.93 (0.62) | 0.89 (0.61) |
| PP (Oct. 1991–Dec. 1993) | 2090 kg | 1508 kg | −27.9 | 0.63 (0.37) | 0.48 (0.32) |
| Total P (Oct. 1991–Dec. 1993) | 3749 kg | 3201 kg | −14.6 | 0.75 (0.43) | 0.63 (0.40) |
| TSS (WY 2001) | 1317 tonnes | 981 tonnes | −25.5 | 0.83 (0.83) | 0.76 (0.83) |
| TDP (WY 2001) | 506 kg | 1250 kg | 146.9 | 0.89 (0.71) | −6.50 (−5.34) |
| PP (WY 2001) | 2328 kg | 1639 kg | −29.6 | 0.88 (0.85) | 0.79 (0.85) |
| Total P (WY 2001) | 2834 kg | 2889 kg | 1.8 | 0.92 (0.87) | 0.92 (0.78) |

period were only slightly worse than the calibration period while the statistics in WY 2001 were notably better than calibration results.

The spatially distributed total P validation results are summarized in Fig. 5. The trend in average measured total P concentrations (January 1990 through December 1993) moving downstream in the WBDR was very closely replicated in the simulated results. In fact, Fig. 5 shows that the simulated validation results in this period even more closely replicated the trend in measured data than during the calibration period (i.e. between WSPA and WDBN water quality stations). In addition to the mainstem WBDR water quality stations in Fig. 5, adequate validation period data were available for a few headwater WBDR tributaries. Considering these headwater and mainstem NYCDEP water quality stations, the model correctly identified sub-basin 18 as the NYCDEP monitoring location with the lowest average validation period (1990–1993) total P concentration.

### Total phosphorus prediction error analysis

Although calibration and validation results were generally quite good, there were a notable number of months where total P loads were substantially in error. Therefore, the daily total P prediction errors (simulated – measured) at Beerston over the water quality calibration and validation periods (1991–2001) were further analyzed. These total P prediction errors were sorted in terms of absolute error magnitudes to focus the analysis on the most severe total P prediction errors. Ignoring the two days of the January 1996 extreme flow event, days on which the total P prediction error was greater in absolute value than 1.5 tonnes were retained for this analysis. The errors on these 31 days (less than 1% of a total of more than 3500 days) accounted for 87% of total calibration and validation period sum of squared errors for daily total P prediction. The 31 corresponding daily flow prediction errors (simulated – measured) at the Walton USGS station (94% of the Beerston drainage area) were calculated and compared in Fig. 10 to the daily total P prediction errors at Beerston.

Fig. 10 shows the strong positive correlation between total P and flow prediction errors when the absolute value of the total P error is greater than 1.5 tonnes. The data points in Fig. 10 show that all total P over-predictions correspond to a flow over-prediction and all but one of the total P under-predictions correspond to flow under-predictions. Daily total P errors all showed absolute %Bias values of 41% or greater while all but one of the corresponding flow errors were also notably large with absolute %Bias values of 12% or greater (most more than 40%).

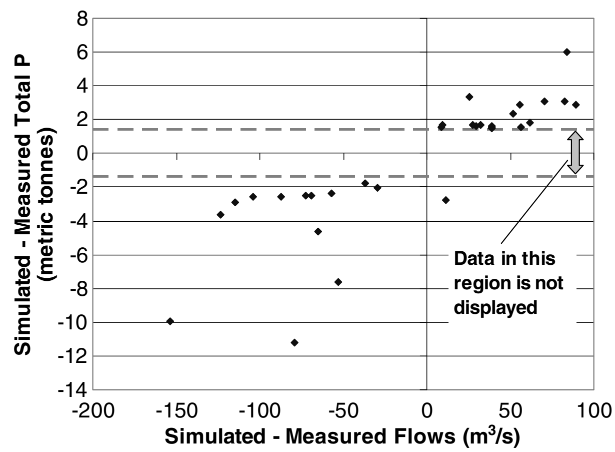


**Figure 10** Daily total P (TDP + PP) prediction errors versus the corresponding daily flow prediction errors for all days between Oct. 1991 and Sept. 2001 when the absolute value of the total P prediction error was greater than 1.5 tonnes.

## Discussion

### Agricultural input representation

The methods of representing agricultural land uses and management in this case study highlighted some important modelling considerations. The mass balance approach to manure production and distribution within each sub-basin produced realistic, spatially varying manure application rates that were governed by the estimated cattle population. From a watershed-wide nutrient mass balance perspective, this approach eliminated mass balance errors that would be introduced by assuming constant and approximate land use specific manure application rates and then applying those to imperfect spatial land use data. Our manure mass balance approach allows for more straightforward and reliable simulation of source-control nutrient management options. Another important part of the agricultural management representation in this application was the way in which corn-hay crop rotations were simulated. Model simulations showed that corn silage land use was the source of more than 50% of the basin-wide NPS sediment and P loads despite the fact that it accounted for only 1.2% of the watershed area. Simulating corn-hay crop rotations on the original, unequally sized AVSWAT2000 HRUs within each sub-basin would have resulted in considerable artificial annual variations in the sub-basin, as well as watershed-wide, total area of corn silage crops. Local agricultural management practices in the watershed result in a relatively constant area of corn silage land use. Artificial corn land use area variations through time would have caused substantial artificial annual fluctuations in simulated sediment and P loads at Beerston. Under such artificial fluctuations, it is unlikely that the calibration and validation performance levels reported in the previous sections could have been achieved. Therefore, our refined rotation HRUs which were defined in a way that allowed crop rotations to be simulated while simultaneously holding annual land use areas constant was a critical part of the model development. Kirsch et al. (2002) also take special measures outside of AVSWAT2000 to define HRUs involved in crop rotations.

### Calibration

The extensive efforts to derive spatially representative model inputs, adjust SWAT model equations when absolutely necessary and adjust model calibration parameters combined to generate the good temporal and spatial calibration results presented in subsection 'Calibration' of Section 'Results'. Daily or monthly $E_{NS}$ coefficients for flow at the main USGS station in the watershed (Walton) were approximately equal or better than a number of other SWAT modelling studies (Eckhardt and Arnold, 2001; Grizzetti et al., 2003; Muleta and Nicklow, 2005; Santhi et al., 2001). Sediment and P calibration results were also good in comparison with previous SWAT studies. Muleta and Nicklow (2005) achieve a daily $E_{NS}$ coefficient of 0.46 for sediment with the aid of an optimization algorithm for automatic model calibration while our manual calibration effort resulted in a daily $E_{NS}$ coefficient of 0.57. Grizzetti et al. (2003) achieve a daily $E_{NS}$ coefficient of 0.74 for total P over a relatively short calibration period (just over two years) while our study achieved a daily $E_{NS}$ coefficient of 0.56 for calibration over a seven-year period.

Peak monthly TSS loads show a distinct tendency to under-predict measured TSS loads (see Fig. 4). The top six largest under-predictions of monthly TSS could be almost entirely attributed to severe TSS under-prediction for a single day in each month. These six severe daily TSS under-predictions correspond with daily flow under-predictions errors of 21, 22, 39, 43, 75 and 86%. Therefore, under-predicted daily flows appear to be largely responsible for the peak monthly TSS under-predictions. Although not investigated, another potentially significant contributor to the peak TSS load under-predictions could be the fact that, as identified by Chen and Mackay (2004), as SWAT HRU areas get smaller (our HRUs are on average 1.6 $km^2$), the erosion rates generated by MUSLE decrease nonlinearly.

## Soil and slope inputs: implications for ungauged watersheds

The calibration step in the application of SWAT to a watershed with measured flow and water quality data will often be able to mask or eliminate some of the prediction inaccuracies due to incorrect or inaccurate model inputs such as the slope and soil inputs discussed in Section 'Impact of alternative model input derivations'. Parameters can be adjusted to at least approximate the averages in the measured data. This is especially true when calibration is focused on a single well-monitored location where spatially distributed predictions are integrated to a single output time series. In contrast, when SWAT is applied to an ungauged watershed, any such adjustments of parameters to correct prediction inaccuracies are not possible because no measured data exist to guide the parameter adjustments. It is therefore critical that representative model inputs are derived for ungauged watershed applications of SWAT.

The analyses in Section 'Impact of alternative model input derivations' are indications of how important our improved HRU slope and soil property input methodologies are with respect to modelling ungauged watersheds. For example, if the HRU slopes in the Cannonsville Watershed were input based on the default SWAT2000 (sub-basin average slopes) and there were no measured data available, the simulated sediment and P export rates from the basin would incorrectly increase by 67% or more (see Table 3) with almost all of the increase attributed to agricultural lands. Any water quality guidelines or regulations implemented based on these uncalibrated model predictions would risk being misguided if the unrepresentative default SWAT2000 approach for HRU slope inputs was used. As an alternative to computing individual HRU slopes, which can be difficult if the AVSWAT2000 land use or soil thresholds function to reclassify large areas of land use or soils, estimates could be improved by calculating the average slopes of each land use within a sub-basin.

A similar argument could be made regarding the use of the inaccurate STATSGO soils as opposed to the approach applied here that derived average SSURGO properties within each STATSGO map unit. For a subset of STATSGO NY map units, Table 4 showed the capacity of soils to hold water and the level of soil erosion protection from rocks assigned by each soil property derivation method can vary greatly. Although STATSGO inaccuracies are not consistent watershed-wide, they would be an especially important issue to address in modelling studies that evaluate spatially varying management practices. A more subtle reason for deriving soil parameters directly from the SSURGO data is that upper and lower bounds for all properties except soil depth can be calculated. These bounds are a direct reflection of the soil survey data uncertainty.

## SWAT2000 equation modifications

### Frozen soils

The winter streamflow simulation problems that precipitated the model change to soil water excess in frozen soils were not unique to this case study. Peterson and Hamlett (1998) report for a Pennsylvania SWAT application with soils similar to those in this case study that SWAT could not adequately simulate streamflow in snowmelt periods. Our modification to SWAT for soil water excess in frozen soils evaluated in Section 'Impact of SWAT2000 equation modifications' is not unlike the assumptions in the HSPF model (Bicknell et al., 2001). For example, lateral flow (referred to as interflow in HSPF) is neither stopped nor restricted in HSPF when soils are frozen. Instead, interflow in HSPF actually increases when soils are frozen. Our modification to SWAT2000 functioned to simply simulate lateral flow in frozen soils in comparison with the original SWAT2000 approach which does not simulate lateral flow in frozen soils.

### MUSLE snow cover adjustment

The original simple empirical SWAT2000 adjustment given in Eq. (1) was replaced with Eq. (2) which is another empirical equation of comparable simplicity that allows for surface runoff depth to mitigate the protective effect of snow cover. The $F$ factor in Eq. (2) required calibration to observed winter TSS loads. The new equation was designed to improve calibration predictions — namely so that SWAT could simulate at least some notable erosion for events with substantial runoff depths occurring on HRUs with 30 mm or more of water equivalent snow cover. Other studies indicate that substantial sediment yields under snow cover can be generated. For example, McConkey et al. (1997) report the greatest amount of soil erodibility in months where the soil was partially frozen while Johnson et al. (1985) report little differences between sediment yields from summer rainfall and snowmelt driven flow events of similar magnitude. Although the general applicability of Eq. (2) is limited by the fact it was only tested for the Cannonsville, this discussion along with Fig. 6 importantly highlight that Eq. (1) may be problematic for other modellers subject to similar winter climates.

Although the two alternative SWAT model forms above greatly improved our predictions, they were not rigorously tested and as such should only be called upon and tested by modellers when needed. These findings do highlight two areas of SWAT2000 that model developers can more thoroughly investigate if there becomes a need to modify SWAT hydrologic and sediment simulation results in regions with similar freezing temperatures and a significant amount of lateral flow.

## Validation

Model validation results were generally encouraging and the performance statistics in this study compare quite favourably to previous watershed modelling studies. The daily flow validation $E_{NS}$ coefficients at the Walton USGS station (0.83 for 1990–1993 and 0.86 for WY 2001) were better than the calibration result (0.80). Our validation results are substantially better than the daily flow validation $E_{NS}$ coefficients that Muleta and Nicklow (2005) and Grizzetti et al. (2003) report (0.23 and 0.54, respectively). TSS and P validation performance statistics were typically slightly degraded in the 1991–1993 water quality validation period relative to calibration. In contrast, statistics for the second, shorter validation period were generally improved. As discussed in detail in the next paragraph, one exception to this was the TDP results. The $E_{NS}$ coefficients for daily TSS validation (0.33 for 1991–1993 and 0.83 for WY 2001) were substantially higher than the daily sediment validation $E_{NS}$ of −0.01 that Muleta and Nicklow (2005) report. The $E_{NS}$ coefficients for daily total P validation (0.40 for 1991–1993 and 0.78 for WY 2001) compare well with the daily total P validation $E_{NS}$ of 0.54 that Grizzetti et al. (2003) report. Schneiderman et al. (2002) conduct a previous modelling study of the Cannonsville Watershed using the simpler GWLF model. Schneiderman et al. (2002) report a validation period monthly $E_{NS}$ value of 0.62 for TDP and 0.72 for total P for WYs 1993–1996, excluding January 1996. For the same time period and same location (Beerston), our simulations yield an $E_{NS}$ value of 0.78 for TDP and 0.71 for total P. Schneiderman et al. (2002) do not evaluate the spatial accuracy of their predictions. In our study, it was shown that the model properly simulated the observed spatial variation of measured total P concentrations in both the calibration and validation periods (see Fig. 5).

Potential causes of the four most significant monthly validation period P prediction errors (November 1991, March 1993, December 2000, April 2001) were investigated and identified. During the first model validation period, TSS and total P loads were substantially under-predicted in November 1991 and March 1993 (see Fig. 9). According to the NYSDEC, their calculated November 1991 TSS and P loads were more uncertain than other months because the largest flow event in November 1991 was not adequately sampled. The March 1993 under-prediction corresponded to a large approximately one week long snowmelt runoff event where the peak flow was seriously under-predicted. The two large TDP over-predictions in the WY 2001 validation period (December 2000 and April 2001) were another notable problem. However, an analysis of the measured P data showed that the percentage of monthly total P that was TDP for the two months in question (11% and 14%) were the smallest two percentages out of the 119 months in both the calibration and validation periods (excluding January 1996). The next smallest TDP percentage was 21% and the average TDP percentage of these 119 months was 54%. This suggests that some of the disagreement between the model and data could possibly be due to TDP measurement error — especially considering the total P loads in December 2000 and April 2001 were fairly accurate. Another potential explanation involves improved agricultural management practices in the watershed that were not represented in the model. For example, phased implementation of various agricultural best management practices is known to have begun at some locations within the watershed in the year 2000. Comparing model predictions to data beyond WY 2001 when it becomes available may shed some light on the cause of the TDP disagreements in WY 2001.

### Model performance summary and future research

Despite the two large monthly TDP over-predictions in WY 2001, the total P predictions in these two months were adequate. Therefore, assuming SWAT2000 modelling assumptions are largely or at least partially responsible for these P prediction errors, it seems likely a major cause of this problem in this case study was the partitioning of P between the dissolved and particulate phases during channel transport. Correcting this P partitioning P should simultaneously consider (1) improvement to the SWAT2000 in-stream reaction equations and (2) a re-evaluation of the assumption that 100% of the active mineral P pool (transported on sediments) dissolves upon entering the stream.

In addition, Fig. 10 shows that a large part of the worst total P prediction errors in this study was attributable to peak daily flow prediction errors. Therefore, future work on improving SWAT water quality predictions in the Cannonsville Watershed should also include an effort to improve daily peak flow predictions. On the basis of our extensive calibration attempts, it seems unlikely that significant improvements in daily peak flow predictions will be achieved through refined parameter optimization. This SWAT2000 modelling application has identified and demonstrated the importance of a number of potential improvements and listed future research directions that could enhance the SWAT2000 model. The alternative methods for model input development and case-study specific model modifications, in conjunction with SWAT modelling recommendations in Chen and Mackay (2004), provide valuable guidance for future SWAT2000 model application studies. The model improvements and future research directions identified here work towards addressing some of the future SWAT model improvements recently summarized in Arnold and Fohrer (2005).

## Conclusions

This paper described the successful development of a SWAT2000 model of the Cannonsville Watershed for the prediction of flow, sediment and phosphorus (P) transport into Cannonsville Reservoir. Extensive datasets were derived for P inputs that varied spatially and temporally. Compared with a number of previous P modelling studies, our study reports comparable or better temporal and spatial model performance statistics in calibration and validation. The good spatial and temporal validation results indicate the potential value of the model as an NPS P management tool. The manure mass balance approach and definition of corn-hay crop rotations with a constant area over time were important steps in the SWAT2000 model application to the Cannonsville Watershed and consideration should be given to incorporating these approaches in future versions of SWAT.

Alternative model input generation methodologies relative to default SWAT modelling approaches were tested and shown to produce more representative model inputs that generate substantially different predictions. These included estimating HRU specific slopes and combining SSURGO and STATSGO soils data. The comparison of input generation methodologies highlights why it is so important for SWAT modellers of ungauged watersheds to derive and utilize the most representative slopes and soils input data.

The successful application of SWAT2000 in this case study required that two alternate model forms be implemented. The winter HRU erosion estimates under snow cover were much too small and a modification to allow the HRU erosion protection level to vary with surface runoff volume as well as snow cover depth was required. In addition, model predictions of no lateral flow in frozen soils were determined to be the cause of major seasonal flow prediction errors and an alternative approach was needed. Future work to improve peak flow predictions should also be considered since the majority of the most serious total P prediction errors were positively correlated with errors in daily peak flow predictions. The Cannonsville Reservoir Watershed application of SWAT2000 identifies modifications, other limitations and demonstrates improved approaches that should be considered in future SWAT development and model applications.

## Acknowledgements

Support for writing this paper and the later stages of research was provided by NSF grant BES-0229176 to Prof. Shoemaker. This paper is partially based on the Tolson and Shoemaker (2006) modelling report, the first version of which (see http://hdl.handle.net/1813/2710) was originally sponsored by the Delaware County Board of Supervisors under the Delaware County Action Plan (DCAP) and co-funded by the US Environmental Protection Agency, through the New York State Department of Environmental Conservation, and Cornell University. The authors are especially thankful for the Delaware County Board of Supervisors for giving us the opportunity to work under DCAP on this project. Thanks to Dr. Jennifer Benaman who initiated the SWAT modelling work on the Cannonsville watershed as well as Keith Porter and Steve Pacenka from the Water Resources Institute at Cornell University who provided data, guidance and advice throughout the modelling effort. Other people that provided critical data, assistance or guidance included Pat Bishop of NYSDEC, Dean Frazier and many other members of the Delaware Country Scientific Support Group. We are also grateful to the reviewers of this manuscript who helped us improve upon the original submission.

## References

Arnold, J., Allen, P.M., 1999. Automated methods for estimating baseflow and groundwater recharge from streamflow records. Journal of American Water Resources Association 35 (2), 411–424.

Arnold, J.G., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling. Hydrological Processes 19, 563–572.

Arnold, J.G., Srinivasan, R., Muttiah, R.R., Williams, J.R., 1998. Large area hydrologic modelling and assessment. Part I: Model development. Journal of the American Water Resources Association 34 (1), 73–89.

Benaman, J., 2002. A systematic approach to uncertainty analysis for a distributed watershed model. Ph.D. Thesis, Cornell University, Ithaca, NY.

Benaman, J., Shoemaker, C.A., 2004. Methodology for analyzing ranges of uncertain model parameters and their impact on the Total Maximum Daily Load process. ASCE Journal of Environmental Engineering 130 (6), 648–656.

Benaman, J., Shoemaker, C.A., 2005. An analysis of high-flow sediment event data for evaluating model performance. Hydrologic Processes 19, 605–620.

Benaman, J., Shoemaker, C.A., Haith, D.L., 2005. Calibration and validation of soil and Water Assessment Tool on an agricultural watershed in Upstate New York. Journal of Hydrologic Engineering 10 (5), 363–374.

Bicknell, B.R., Imoff, J.C., Kittle, J.L., Jobes, T.H., Donigan, A.S., 2001. Hydrological Simulation Program — FORTRAN. Version 12. User's Manual, AQUA TERRA Consultants. United States Environmental Protection Agency.

Bingner, R.L., 1996. Runoff simulated from Goodwin Creek Watershed using SWAT. Transactions of the American Society of Agricultural Engineers 39 (1), 85–90.

Bingner, R.L., Garbrecht, J., Arnold, J.G., Srinivasan, R., 1997. Effect of watershed subdivision on simulation runoff and fine sediment yield. Transactions of the American Society of Agricultural Engineers 40 (5), 1329–1335.

Chen, E., Mackay, D.S., 2004. Effects of distribution-based parameter aggregation on a spatially distributed agricultural nonpoint source pollution model. Journal of Hydrology 295, 211–224.

DiLuzio, M., Srinivasan, R. and Arnold, J.G., 2001. Arcview interface for SWAT2000 User's Guide, US Department of Agriculture — Agricultural Research Service, Temple, TX.

Eckhardt, K., Arnold, J.G., 2001. Automatic calibration of a distributed catchment model. Journal of Hydrology 251, 103–109.

Eckhardt, K., Haverkamp, S., Fohrer, N., Frede, H.G., 2002. SWAT-G, a version of SWAT99.2 modified for application to low mountain range catchments. Physics and Chemistry of The Earth 27 (9–10), 641–644.

FitzHugh, T.W., Mackay, D.S., 2000. Impacts of input parameter spatial aggregation on an agricultural nonpoint source pollution model. Journal of Hydrology 236, 35–53.

Fontaine, T.A., Jacomino, V.M.F., 1997. Sensitivity analysis of simulated contaminated sediment transport. Journal of the American Water Resources Association 33 (2), 313–326.

Grizzetti, B., Bouraoui, F., Granlund, K., Rekolainen, S., Bidoglio, G., 2003. Modelling diffuse emission and retention of nutrients in the Vantaanjoki watershed (Finland) using the SWAT model. Ecological Modelling 169, 25–38.

Haith, D.A., Mandel, R., Wu, R.S., 1992. Generalized Watershed Loading Functions: Version 2.0 User's Manual, Department of Agricultural and Biological Engineering Cornell University.

Johnson, C.W., Gordon, N.D., Hanson, C.L., 1985. Northwest rangeland sediment yield analysis by the MUSLE. Transactions of the American Society of Agricultural Engineers 28 (6), 1889—1895.

Kane, K., 1999. Proposed Phase II phosphorus TMDL calculations for Cannonsville Reservoir, New York City Department of Environmental Protection.

Kirsch, K., Kirsch, A., Arnold, J.G., 2002. Predicting sediment and phosphorus loads in the Rock River basin using SWAT. Transactions of the ASAE 45 (6), 1757—1769.

Longabucco, P., Rafferty, M.R., 1998. Analysis of material loading to Cannonsville Reservoir: advantages of event-based sampling. Journal of Lakes and Reservoir Management 14 (2—3), 197—212.

Lumia, R., 1998. Flood of January 19—20, 1996 in New York State. Water-Resources Investigations Report 97-4252, US Geological Society, Albany, NY.

Manguerra, H.B., Engel, B.A., 1998. Hydrologic parameterization of watersheds for runoff prediction using SWAT. Journal of the American Water Resources Association 34 (5), 1149—1162.

McConkey, B.G., Nicholaichuk, W., Steppuhn, H., Reimer, C.D., 1997. Sediment yield and seasonal soil erodibility for semiarid cropland in western Canada. Canadian Journal of Soil Science 77, 33—40.

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of Hydrology 306, 127—145.

Nash, J.E., Suttcliffe, J.V., 1970. River flow forecasting through conceptual models, Part I. A discussion of principles. Journal of Hydrology 10 (3), 282—290.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2001a. Soil and Water Assessment Tool theoretical documentation version 2000: Draft-April 2001, US Department of Agriculture — Agricultural Research Service, Temple, Texas.

Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2001b. Soil and Water Assessment Tool user's manual version 2000: Draft-April 2001, US Department of Agriculture — Agricultural Research Service, Temple, Texas.

Peterson, J.R., Hamlett, J.M., 1998. Hydrologic calibration of the SWAT model in a watershed containing fragipan soils. Journal of the American Water Resources Association 34 (3), 531—544.

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

Schneiderman, E.M., Pierson, D.C., Lounsbury, D.G., Zion, M.S., 2002. Modelling the hydrochemistry of the Cannonsville Watershed with Generalized Watershed Loading Functions (GWLF). Journal of the American Water Resources Association 38 (5), 1323—1347.

Tolson, B.A. and Shoemaker, C.A., 2006. Watershed modelling of the Cannonsville Basin using SWAT2000: model development, calibration and validation for the prediction of flow, sediment and phosphorus transport to the Cannonsville Reservoir — Version 2.0. Cornell University, School of Civil and Environmental Engineering Technical Report, Ithaca, NY.