Watershed Modeling 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 1.0

Technical Report
School of Civil and Environmental Engineering
Cornell University

Authors:
Bryan A. Tolson and Christine A. Shoemaker
Department of Civil and Environmental Engineering
Cornell University, Ithaca, NY, 14853
(bat9@cornell.edu and cas12@cornell.edu)

Prepared for:
Delaware County Board of Supervisors

Project Administered by:
Christine Shoemaker; Ripley Professor of Engineering,
Keith Porter; Director of NYS Water Resources Institute,

Sponsored and Funded by:
1) The Delaware County Board of Supervisors with funds provided by the US Environmental Protection Agency through the New York State Department of Environmental Conservation and 2) funds from Cornell University.

The appropriate reference for this report is:
Tolson, B. A. and Shoemaker, C. A. (2004). Watershed modeling 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 1.0. Technical Report, School of Civil and Environmental Engineering, Cornell University. 159 pp.
http://techreports.library.cornell.edu:8081/Dienst/UI/1.0/Display/cul.wat/2004-2.

Abstract

This report describes the calibration and validation of a spatially distributed watershed model of the Cannonsville Reservoir Basin. The Soil and Water Assessment Tool 2000 (SWAT2000) was selected as the watershed model. A set of SWAT2000 inputs representative of the watershed conditions was derived from a wide array of data sources. Important methods were developed for converting available information to SWAT2000 inputs for groundwater soluble phosphorus concentrations, initial soil phosphorus levels and daily manure application.

The Cannonsville Reservoir is a New York City water supply reservoir located in upstate New York that has historically experienced water quality problems associated with phosphorus loading. As a result, the watershed has been subjected to multiple water quality regulations including a recent Total Maximum Daily Load (TMDL) assessment for phosphorus. The reservoir watershed covers an 1178 km^2 area and is dominated by agriculture, particularly dairy farming. The SWAT2000 model of the Cannonsville Reservoir Watershed is a valuable tool that can be used to help identify and quantitatively evaluate the long-term effects of various phosphorus management options for mitigating loading to the reservoir.

SWAT2000 was developed by the Agricultural Research Service of the United States Department of Agriculture. SWAT2000 simulates through time the daily soil water balance, growth of plants, build-up and subsequent transport of soil nutrients to surface waters in response to agricultural management practices. The simulated mass balance of soil phosphorus in SWAT2000 is an important aspect of any watershed model that is to be used for regulatory purposes. The authors modified a few of the SWAT model equations to better simulate measured flows, sediment loading and phosphorus loading during the winter.

The model was calibrated and validated for the prediction of dissolved and particulate phosphorus transport, and therefore also flow and sediment transport, against a large set of monitoring data. Extensive continuous flow and water quality data over a 10-year period from multiple locations within the basin were used for model calibration and validation. Sensitive model parameters were adjusted within their feasible ranges during calibration to minimize model prediction errors for daily flows and monthly sediment and phosphorus loading. At the main flow gauging station in the basin (Walton), draining almost 80% of the watershed, daily calibration resulted in model predictions of average flow within 1.0% of the measured average flow while the daily Nash-Sutcliffe (NS) measure was 0.79. Daily validation results at Walton showed the model predicted average flow within 4.5% of the measured average flow with a NS of 0.78. At the main water quality gauging station in the basin (Beerston), just downstream of Walton, the calibration results showed the model predicted the average monthly sediment and total phosphorus loading within 3% and 6% of their respective measured average monthly loadings. The monthly calibration NS values at Beerston for sediment and total phosphorus loading were 0.66 and 0.68, respectively. Validation results at Beerston

showed the model predicted the average monthly sediment and total phosphorus loading within 27% and 9% of their respective measured average monthly loadings. The monthly validation NS values at Beerston for sediment and total phosphorus loading were 0.51 and 0.61, respectively. The largest errors in model predictions for phosphorus and sediment loading were always associated with peak flow prediction errors. Model predictions were also shown to qualitatively replicate bi-weekly sampling of total phosphorus concentrations taken from 10 different locations across the watershed.

Model simulation results over the calibration and validation period (1990-2000) highlighted a number of useful findings. The model predicted that 68% of the total phosphorus loading to surface waters in the watershed originates from active agricultural lands. Corn land use was simulated as the major source of agricultural phosphorus loading even though it covered only 1.2% of the watershed area. Areas North and East of the Town of Delhi tended to have the largest phosphorus loading rates per unit area. Areas immediately surrounding the Cannonsville Reservoir that are not monitored were simulated to have substantially lower non-point source phosphorus unit area loading rates than the monitored portion of the watershed.

Acknowledgements

Many organizations and people contributed to this modeling effort. This project was 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. Christine Shoemaker and Keith Porter supervised project work. In addition, the work was supported by the US Army Corps of Engineers, graduate fellowships administered through Cornell, and volunteer efforts by students. The authors are especially thankful for the Delaware County Board of Supervisors for giving us the opportunity to work under DCAP on this project.

Jennifer Benaman initiated the modeling work on this project and developed model versions for flow and sediment predictions as part of her Cornell PhD research, which was funded by an EPA Star Fellowship. The authors are indebted to her for the great amount of work that she put forth to find, organize and then transform data to create the initial version of the Cannonsville SWAT model for flow and sediment transport. Keith Porter and Steve Pacenka from the Water Resources Institute at Cornell University were instrumental in this work by providing data, guidance and advice throughout the modeling effort. Pat Bishop of NYSDEC contributed also in multiple ways. She provided NYSDEC continuous monitoring data for Town Brook and Beerston water quality stations as well as data on the point sources in the basin. In addition, her comments on the modeling work throughout the process and particularly in the review of an earlier draft of this document were very helpful. Dean Frazier was crucial to this effort as he was our coordinating link to the stakeholders whom this modeling effort was intended to benefit.

Paul Cerosaletti of Cornell Cooperative Extension put forth considerable effort to provide us with critical data and valuable comments on dairy cattle phosphorus inputs. His work was also supported through DCAP funding. In addition, Dale Dewing (Cornell Cooperative Extension Watershed Agricultural Program), with respect to farm information, and Art DeGaetano (Director of Northeast Regional Climate Center) with respect to climatological understanding, provided input data and spent a significant amount of time to discussing and clarifying model inputs.

Andrew Sharpley, Ray Bryant and Bil Gburek of Penn State USDA Agricultural Research Service met with us on multiple occasions and provided valuable feedback on the modeling work. In addition, Bil Gburek provided insightful review comments on an earlier draft of this document.

Contacts from the NYCDEP were also instrumental in this modeling effort as NYCDEP provided land use maps that formed the basis for model inputs. Elliot Schneiderman, G. Mendoza, D. Thongs and M. Zion (NYCDEP Terrestrial Modeling staff) provided insightful review comments on an earlier version of this document while W. Stasiuk and

C. Stepczuk provided spatially distributed phosphorus monitoring samples for model development.

Special thanks to all members of the Delaware County Scientific Support Group who are not listed above as they helped guide and improve this work by supplying feedback and data during model development. Thanks to Karl Czymmek, Quirine Ketterings, Gary Lamont and Tom Tylutki who provided information for the modeling through personal communications with the authors. The authors would also like to thank Jeff Arnold, SWAT model developer, for comments on the model source code changes. The programming and modeling work by Cornell students Ziyin Shen, Mike Sorenson and Marcelo Cerucci helped with defining and entering various model inputs.

Finally, we would like to acknowledge and sincerely thank all the reviewers of an earlier draft of this document for their willingness to review the document and for their truly insightful and helpful commentaries. Pat Bishop, Elliot Schneiderman, G. Mendoza, D. Thongs, M. Zion and Bil Gburek, David Wells (Office of Water, USEPA), Betty McQuaid and Kerry Robinson (both from Natural Resources Conservation Service - United States Department of Agriculture) each reviewed the earlier draft. The over 20 total pages of detailed review comments from the reviewers led to very substantial improvements, not only in the model report, but also the model performance for flow, sediment and phosphorus prediction.

*Readers of the electronic version of this document should note that it contains active cross-references or links. All references in text to numbered Sections, Figures or Tables (including the Table of Contents, Figures and Tables) are linked to the original location in the document and can be immediately viewed by one left mouse click on the **number** (a pointing finger icon appears at the linked number).*

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms used in Report:

Table I. List of institutional acronyms.

Acronym	Full name
ARS	Agricultural Research Service
ASAE	American Society of Agricultural Engineers
DCAP	Delaware County Action Plan
NRCS	Natural Resource Conservation Service
NYCDEP	New York City Department of Environmental Protection
NYSDEC	New York State Department of Environmental Conservation
SCS	Soil Conservation Service
USDA	United States Department of Agriculture
USGS	United States Geological Society
WAP	Watershed Agriculture Program for NYC watersheds
WRI	NY State Water Resources Institute at Cornell University

Table II. Additional acronyms used in report.

Acronym	Full name
AP2	Labile P in the soil in mg labile P/kg soil
AVSWAT	SWAT2000 Arcview Interface
AU	Animal Units
BS	Base Saturation as a percentage
D	Percent Difference
DEM	Digital Elevation Map
DMR	National Pollutant Discharge Elimination System Discharge Monitoring Report
E_{NS}	Nash-Suttcliffe Coefficient
GIS	Geographic Information System
HRU	Hydrologic Response Unit
MUSLE	Modified Universal Soil Loss Equation
N	Nitrogen
P	Phosphorus
PP	Particulate Phosphorus
PSP	Soil Phosphorus Sorption Coefficient in SWAT model
QAPP	Quality Assurance Project Plan
r^2	Coefficient of Determination
Sno_i	Snow water equivalent depth (mm) of an HRU snowpack on day i
SRP	Soluble Reactive Phosphorus
TSS	Total Suspended Sediments
SSURGO	Soil Survey Geographic Database
STATSGO	State Soils Geographic Database
$SurQ_i$	Surface runoff depth (mm) from an HRU on day i
SWAT	Soil and Water Assessment Tool
TDP	Total Dissolved Phosphorus
USLE	Universal Soil Loss Equation
WBDR	West Branch Delaware River
WQ	Water Quality
WWTP	Wastewater Treatment Plant

List of Personal Communication References and Affiliations

Table III. List of people cited as personal communication references.

Dr. Jeff Arnold	USDA-ARS Grassland Soil and Water Research Laboratory Developer of SWAT
Pat Bishop	NYS Department of Environmental Conservation Division of Water
Dr. Paul Cerosaletti	Cornell Cooperative Extension - Delaware County Dairy and Field Crops Agent
Karl Czymmek	Cornell University Department of Animal Science Sr. Extension Associate
Dr. Art DeGaetano	Cornell University Department of Earth and Atmospheric Sciences Associate Professor and Director of Northeast Regional Climate Center
Dale Dewing	Cornell Cooperative Extension - Delaware County Watershed Agricultural Program Field Crop Specialist / Whole Farm Planner
Dr. Quirine Ketterings	Cornell University Department of Crop and Soil Sciences Assistant Professor/Cornell Nutrient Analysis Lab Advisor
Gary Lamont	NYC Watershed Agricultural Program Watershed Coordinator
Dr. Andrew Sharpley	USDA-ARS Pasture Systems and Watershed Management Research Unit
Tom Tylutki	Cornell University Animal Science Department Senior Research Associate

Outline of Report

This document is written to explain a model analysis of the Cannonsville Basin for describing the sources and transport of phosphorus to the Cannonsville Reservoir. The methodology described can be useful in other nutrient transport modeling studies. The Executive Summary (Section 1) and Case Study Introduction (Section 2) are general and are intended to be accessible to policy makers as well as to hydrologic scientists and modelers. Readers may refer to the Executive Summary of the report (Section 1) for a summary of model performance. Policy makers may be specifically interested in the example modeling applications listed in Section 1.1 that demonstrate the understanding and benefits gained by developing a watershed model. The remainder of the report details the hydrologic science driving the model predictions, model inputs and model output comparisons with measured data.

The report sections after Section 2 are written mainly for hydrologic scientists and modelers. Section 3 introduces the watershed model used in this work (SWAT) and highlights previous SWAT applications in the US. The extensive input data descriptions in Section 4 are written to be detailed enough to be useful for other modelers developing SWAT model inputs for this basin. As such, some modelers may need to refer the SWAT model documentation (see SWAT references in Section 3) in order to gain a complete understanding of specific model inputs. Section 4, outlines the development of model inputs in a way that should be helpful to modelers developing similar types of model inputs for the Cannonsville Basin or other NYC Watersheds. Furthermore, agricultural stakeholders in the basin should find interest in the material in the agricultural input sections (Sections 4.7, 4.9.1). Section 5 outlines all modifications to the original watershed model source code that were mainly required to improve the representation of physical processes simulated by the model. Section 6 provides an extensive array of spatially and temporally distributed model calibration and then validation results. Section 7 compares model performance to other SWAT modeling applications, critically evaluates some of the model shortcomings and then briefly summarizes future modeling work. Section 8 is the report conclusions and Section 9 lists all references.

The Appendix to this document (Section 10) contains more detailed descriptions of some of the material in this report as well as other supplementary information. Tables I and II, which immediately precede this outline, define the various institutional acronyms and other general acronyms used throughout this report. All personal communication references cited in this report, along with their affiliations, are listed in Table III, which follows Table II. The affiliations for each reference are also provided to show their qualifications and demonstrate that they are a credible reference.

1 Executive Summary

This report describes a spatially distributed watershed model of the Cannonsville Basin that has been developed using SWAT2000 (Neitsch et al. 2001a) to describe phosphorus transport to the Cannonsville Reservoir in upstate New York. SWAT2000 was developed by the Agricultural Research Service of the United States Department of Agriculture and is distributed by the US Environmental Protection Agency for watershed management.

The Cannonsville Reservoir is a New York City water supply reservoir that has historically experienced severe water quality problems due to excessive phosphorus loading. Effective phosphorus management is crucial in the Cannonsville Basin because economic activity in Delaware County (which contains nearly all of the watershed area) is 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 County through a Memorandum of Agreement (MOA) with New York City if reservoir phosphorus concentrations become too high. These MOA regulations can restrict future economic growth in the basin when the growth directly or indirectly increases phosphorus loading to surface waters in the basin.

Among many other processes represented in SWAT2000 (see Section 3), SWAT2000 simulates through time the daily soil water balance, growth of plants, build-up of nutrients in the soil due to agricultural management practices and subsequent erosion and transport of nutrients to streams and rivers. The soil nutrient mass balance in SWAT2000 results in temporally varying levels of soil nutrients due to either nutrient depletion or excessive nutrient application rates. The rate of phosphorus export from soils to rivers increases as soil phosphorus levels increase. The simulated mass balance of soil phosphorus in SWAT2000 is an improvement in physical process representation over simplified watershed loading function models like GWLF (Haith et al. 1992) that assume constant soil nutrient levels over the period of time simulated by the model. The spatially distributed nature of SWAT2000 means that the above processes are represented independently in different regions across the basin. There are 758 different regions (called HRUs) in this application. A set of SWAT2000 inputs representative of the basin has been developed and successfully tested against measured flow and water quality data. The model will be used to spatially and temporally evaluate the performance of a wide range of phosphorus management options in order to help local decision-makers allocate resources and efficiently reduce phosphorus loading to the reservoir.

The overall goal of the application of SWAT to the Cannonsville Watershed has three parts:

a. To understand the phosphorus budget, its inputs, outputs to watercourses, and storage within the watershed;
b. To identify and evaluate quantitatively the effects of management options on phosphorus loading to the Cannonsville Reservoir;

c. To assist in evaluating the results of the adoption and implementation of management options in improving water quality in the Basin.

In addition to overall goal above, this modeling work and report has other general benefits such as:

- This comprehensive report completely documents the modeling work for Delaware County and will serve as a guide to transferring all data files associated with this modeling effort for those who would like to use the model in the future.
- Model phosphorus inputs are summarized to be useful to local decision-makers even without utilization of the model. This will help facilitate understanding and communication between all stakeholders.
- The basis for phosphorus input information is outlined in sufficient detail for future modelers in the basin to repeat the input development when new input data become available or if a future model of the basin (related to or independent of SWAT) is developed.
- Spatially distributed modeling approach can also be utilized to help direct future monitoring efforts in the basin.
- Provides a means for investigating assumptions made in previous modeling and monitoring work that currently form the basis of the regulatory framework in the basin.
- SWAT basin modeling approach provides the scientific framework for compiling and integrating the Delaware County Action Plan (Delaware County Board of Supervisors 1999).The modeling effort will hopefully help to improve the general SWAT2000 watershed model.

Model Input

Model development first consisted of developing extensive model inputs. Most input development focused on input information about phosphorus that was derived mainly from basin specific data. For example, model phosphorus inputs (see Section 4) that were assigned based on available basin specific data included wastewater treatment plant phosphorus discharges, spatially distributed groundwater soluble phosphorus concentrations, cattle population estimates, manure production factors, manure phosphorus characteristics and initial soil phosphorus levels. Agricultural management practices are also specified based on generalized management practices representative of Delaware County.

SWAT2000 Model Modification

In addition to standard model parameter adjustments, a number of improvements were made to the default SWAT characterization of the basin (e.g. default model input or parameter settings outlined in Section 6.3.2). Furthermore, the SWAT2000 model was enhanced in this modeling effort by a few modifications to the model equations that were required to better represent processes in the Northeast US (see Section 5). The final stage of model development involved selecting a relatively small number of model parameters for calibration of the model predictions relative to the monitored data (see Section 6).

Calibration and Validation Results

The model is calibrated and validated against river water quality monitoring data over a number of years and at multiple locations within the basin. For streamflow, the model is calibrated against measured USGS flow data from 1994 to 2000 at six different locations in the basin and validated at Walton (which drains about 80% of the basin area) from 1990 to 1993. Suspended sediment, total dissolved phosphorus and particulate phosphorus loading are calibrated against measured data from the New York State Department of Environmental Conservation (NYSDEC 2001) from 1994 to 2000 at two locations in the basin (Beerston and Town Brook) and validated at Beerston from 1991 to 1993. In addition, the spatially distributed performance of the model for total phosphorus is qualitatively checked against periodic water quality sampling data distributed across the basin from 1990 to 1999 and provided by New York City Department of Environmental Protection (NYCDEP Unpublished data).

Model calibration involves the determination of model parameter values that result in model predictions that best reproduce a set of measured data (e.g. daily flows, monthly total phosphorus loads). Model validation is the process of testing model performance of the calibrated model parameter set against an independent set of measured data. The measured validation and calibration data sets cover different time periods or involve separate spatial locations. Good validation results support the usefulness of the model to predict future conditions (i.e. phosphorus loading) under alternative management scenarios and future climates. Since validation assesses the performance of the calibrated model parameter set against a set of independent measured data, it is typically more difficult to get good validation performance in comparison to calibration.

Table 1.0.1 summarizes the model predictive performance measures for flow at Walton and sediment, total dissolved phosphorus and total phosphorus at Beerston over the model calibration and validation period. The percent difference (D), coefficient of determination (r^2) and Nash-Suttcliffe simulation efficiency (E_{NS}) (Nash and Suttcliffe 1970) are the three performance measures used to assess model predictions relative to measured data. Briefly, D measures how well the model predicts the average of the measured data while the r^2 and E_{NS} are alternative measures of how well the model predicts the observed trends through time in the measured data. A value of 0% is best for D while a value of 1.0 is the maximum or best value for both r^2 and E_{NS}. All three performance measures are defined mathematically in Section 10.3.

Model performance with respect to daily flow predictions during calibration is quite good while model performance in validation is very similar to the calibration performance. At the main flow gauging in the basin (Walton) the calibration results showed the model predicted average flow within 1.0% of the measured average flow while the r^2 and E_{NS} values were 0.80 and 0.79, respectively. The r^2 and E_{NS} values approaching 1.0 indicate good results. Validation results at Walton showed the model predicted average flow within 4.5% of the measured average flow the r^2 and E_{NS} values were 0.79 and 0.78,

respectively. Note that the monthly E_{NS} for prediction at Walton was 0.89 for calibration and 0.94 for validation.

Table 1.0.1. Summary of SWAT2000 model performance measures for the calibration and validation periods.

Modeled Quantity (simulation period)	Percent (%) Difference[1]	r^2	E_{NS}
Daily Flow (Calibration)	1.0	0.80	0.79
Daily Flow (Validation)	4.5	0.79	0.78
Monthly Sediment (Calibration)	-2.2	0.71	0.66
Monthly Sediment (Validation)	-26.5	0.66	0.51
Monthly Total Dissolved Phosphorus (Calibration)	16.1	0.77	0.73
Monthly Total Dissolved Phosphorus (Validation)	13.4	0.93	0.90
Monthly Total Phosphorus (Calibration)	-6.1	0.72	0.68
Monthly Total Phosphorus (Validation)	-9.2	0.70	0.61

1. A positive percent difference means the model over-predicted the average while a negative percent difference means the model under-predicted the average.

The measured sediment and phosphorus loading data are estimated from samples that are not taken every day. Therefore, the model was calibrated based on a monthly, rather than daily, time step. Model performance with respect to monthly sediment predictions during calibration and validation is reasonable based on comparisons with performance levels in other studies (see Section 7.1.1). At the main water quality station in the watershed (Beerston), located approximately 8 km downstream of Walton, the calibration results showed the model predicted the average monthly sediment load within 2.2% of the measured average monthly sediment load while the r^2 and E_{NS} values were 0.71 and 0.66, respectively. Validation results at Beerston showed the model predicted the average monthly sediment load within 26.5% of the measured average monthly sediment load while the r^2 and E_{NS} values were 0.66 and 0.51, respectively.

Model performance with respect to monthly phosphorus loading predictions during calibration is generally good while model performance in validation is better than calibration performance for total dissolved phosphorus. The calibration and validation performance measures for total dissolved phosphorus are compared in Table 1.0.1. The calibration and validation total phosphorus results at Beerston show that total phosphorus loading predictions are very good. At Beerston, the calibration results showed the model predicted the average monthly total phosphorus load within 6.1% of the measured average monthly total phosphorus load while the r^2 and E_{NS} values were 0.72 and 0.68, respectively. Validation results at Beerston showed the model predicted the average monthly total phosphorus load within 9.2% of the measured average total phosphorus load while the r^2 and E_{NS} values were 0.70 and 0.61, respectively. Particulate phosphorus results closely resemble the suspended sediment results since particulate phosphorus is carried in the stream as suspended sediment.

Sections 6.4 and 6.5 examine the performance of the model in more detail than the above performance summary. Relative to similar published modeling studies with available

performance statistics, the performance of the Cannonsville Basin model is satisfactorily comparable (see Section 7.1). Furthermore, some of the observed model limitations and inaccuracies are discussed in Sections 7.2 and 7.3. Overall, the model predictive performance is observed to be good. Therefore, the model can be used with some confidence to evaluate phosphorus loading to the Cannonsville Reservoir under future climate and/or management conditions.

1.1 Model Applications and Analyses

This section focuses on providing some example applications and analyses with the Cannonsville Basin model. These demonstrate the utility of the SWAT model for evaluating future phosphorus management options.

1.1.1 Application 1: Comparing Measured and Modeled Annual Phosphorus Loads

The current NYSDEC water quality monitoring station at Beerston provides phosphorus concentrations that are used by NYSDEC to calculate annual phosphorus loading estimates from areas upstream of Beerston. The Beerston water quality station monitors the water quality from approximately 80% of the Cannonsville Basin. The measured phosphorus loading data from Beerston are compared with phosphorus loading at Beerston that is predicted by the SWAT model in Figure 1.1.1. The results in Figure 1.1.1 are presented on an annual water year basin from 1992 through 2000. A water year is from October 1 through September 30.

The comparison in Figure 1.1.1 demonstrates a few important points. First of all, the SWAT model predicts phosphorus loads that approximate the temporal pattern of measured phosphorus loads. For example, the model correctly predicts that 1996 has the highest phosphorus loading and 1995 has the lowest phosphorus loading. The ability of the model to predict historic phosphorus loading reasonably well is required in order to have confidence that the model can be used to predict future phosphorus loading. Secondly, there are some years in which the model predictions are not that close to measured phosphorus loading, most notably 1996. Although the model correctly predicts 1996 to have the highest phosphorus loading of any water year between 1992 and 2000, the model predicted load is only 43% of the measured load. During this year the devastating flood of Jan. 19-20, 1996 occurred. Since SWAT is not designed to represent such extreme flooding it would be unsound to attempt to calibrate the model for such an extreme event. See Section 6.4.2 for a brief description of this flood and further explanation of why the model under-predicted phosphorus loading.

A purpose in this study is to develop a model that can be used to evaluate future phosphorus loading relative to a regulatory phosphorus loading standard (called the Total Maximum Daily Load, or TMDL, allowable load) under various phosphorus management strategies. The model accurately predicts that annual phosphorus loading from the extreme Jan. 1996 event will result in phosphorus loading to the reservoir that greatly exceeds the TMDL allowable load of 46,944 kg. The last important point presented in Figure 1.1.1 is that when the average annual phosphorus loading rates from 1992-2000,

not including Jan. 1996, are compared, the SWAT model is observed to under-predict measured phosphorus loading by about 7%. This small bias should nevertheless be accounted for when the model is used to evaluate whether future phosphorus management scenarios result in TMDL compliance. In other words, SWAT estimated total phosphorus loading to the Cannonsville Reservoir should be scaled up to adjust for this under-prediction. Failure to adjust for this model tendency could lead to overly optimistic predictions of future phosphorus management scenario effectiveness.

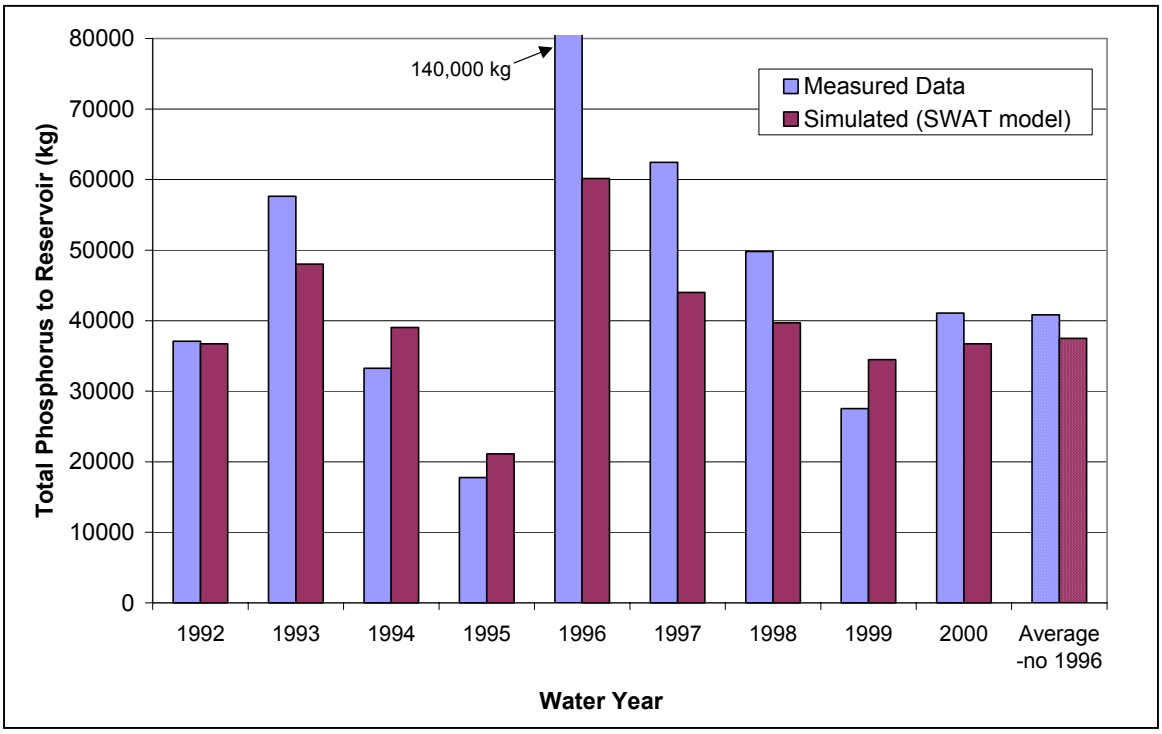

Figure 1.1.1. Measured and SWAT predicted annual total phosphorus loading at Beerston by water year (Oct. 1 – Sept 30) from 1992 to 2000. 1996 contained an extreme event as discussed in the text above.

Finally, it should be mentioned that the measured total phosphorus loads at Beerston are actually estimated loads. Actual total phosphorus (and sediment) loads are difficult if not impossible to measure with certainty as discussed in Section 6.1. Therefore, we cannot assume that the measured phosphorus loads are always accurate.

1.1.2 Application 2: Estimating Phosphorus Loading from Different Land Uses

The model provides estimates of sediment and phosphorus non-point source (NPS) loading rates to streams for each land use. Although other factors impact model predictions (e.g. land slope and soil type), it can be useful to determine the relative importance of land use in the basin with respect to sediment and phosphorus loading to surface waters. As an example, model output over the period from 1994 to 2000 is analyzed. Table 1.1.1 shows the loading rates of sediment and total phosphorus to streams from various land uses in the basin. For a description of the land uses in Table

1.1.1 see Section 4.2. Clearly, the model predicts that corn sediment and phosphorus loading rates are generally more than an order of magnitude higher than loading rates from other land uses. Although the corn sediment loading rate is high, it is close to results in previous modeling by the NYCDEP (Schneiderman et al. 1998) where corn sediment loading rates in the Cannonsville Basin were modeled at about 5.2 mt/ha. However, the SWAT modeled average corn sediment loading rate is less than the average annual sediment loading rate for continuous corn measured at 13 mt/ha in a long-term field study (Ghidey and Alberts 1998). The soil type and management in this field study were very similar to the conditions simulated for corn in this SWAT model application. Forested lands generate the lowest sediment and phosphorus loads of all land uses.

Table 1.1.1. Basin-wide area-weighted average annual SWAT predicted loading rates to the surface waters in the Cannonsville Basin.

Land use	Sediment load (mt[1]/ha)	Phosphorus load[2] (kg/ha)
Pasture	0.16	0.59
Hay	0.08	0.36
Corn	7.52	16.66
Grass-shrub	0.07	0.20
Idle Agriculture	0.10	0.25
Deciduous forest	0.04	0.04
Coniferous forest	0.04	0.03
Urban	1.73	1.04

1. mt is metric tons, which is equivalent to 1000 kg.
2. The phosphorus load is total phosphorus and includes groundwater phosphorus delivered to surface waters from each land use.

The data in Table 1.1.1 are informative when adjusted to reflect the areas of the land uses such that relative total loading rates can be evaluated. Figure 1.1.2 shows the results of this adjustment and compares the relative magnitudes of land area, sediment and phosphorus loading across the eight land uses. Figure 1.1.2 shows that even though corn only covers about 1% of the basin area, the extremely high sediment and phosphorus loading rates per hectare in Table 1.1.1 result in corn predicted as accounting for 58% of the basin-wide NPS sediment and phosphorus load to surface waters. Conversely, although deciduous forests cover 53% of the basin area, the model predicts that deciduous forests account for only 15% of the basin sediment load and 7% of the basin phosphorus load. This is due to the very small unit area loading rates for forests in Table 1.1.1.

The previous phosphorus loading summaries by land use are aggregated together in more general categories and then compared with the other sources of phosphorus in the SWAT model as shown in Figure 1.1.3 for 1994-2000. Active agricultural land in Figure 1.1.3 is composed of corn, hay and pasture land use. The forest category includes both deciduous and coniferous forests. The idle and successional agricultural land use category in Figure 1.1.3 combines the grass-shrub land use and idle agricultural land use categories in Figure 1.1.2. Note that in contrast with the previous NPS phosphorus loading comparisons in Table 1.1.1 and Figure 1.1.2, the groundwater phosphorus in Figure 1.1.3

is now disassociated from each land use and the groundwater load is simply summed together across all land uses in the basin. Thus, phosphorus loading from other land uses in Figure 1.1.3 do not include groundwater phosphorus.

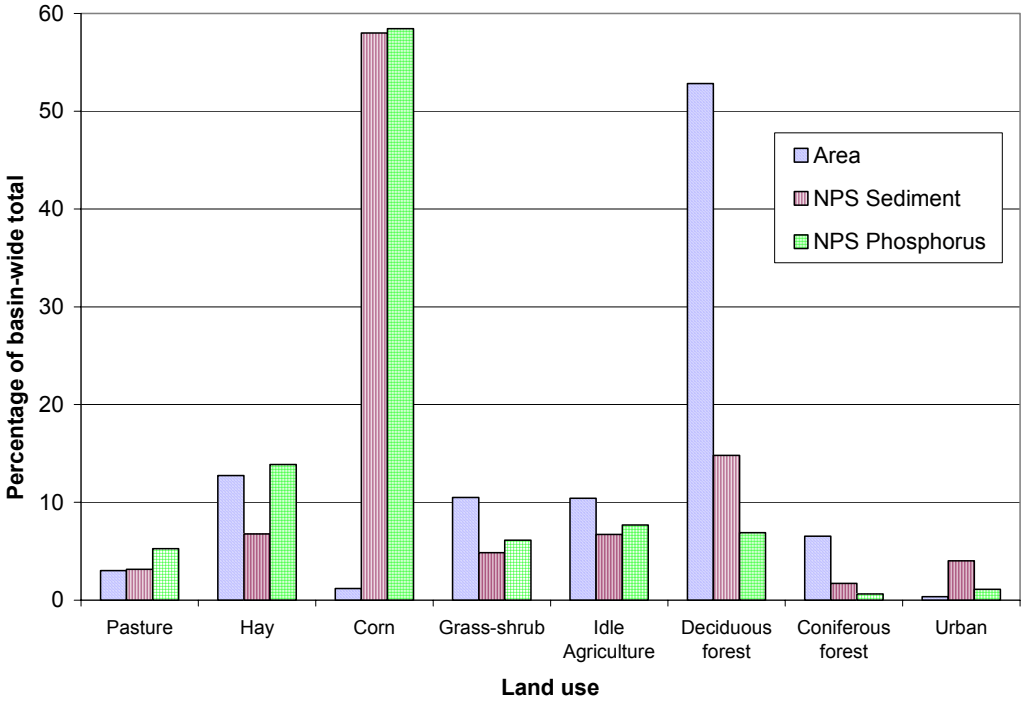

Figure 1.1.2. SWAT simulated percentage of total land area, NPS sediment and NPS phosphorus load to streams for each land use category from 1994-2000.

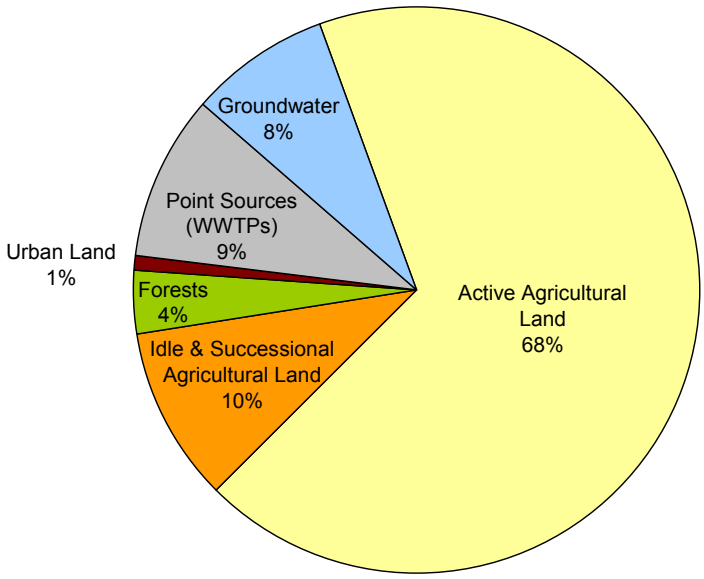

Figure 1.1.3. Relative contributions from all phosphorus sources to the total phosphorus loading to streams and rivers in the Cannonsville Basin as simulated by SWAT from 1994-2000.

Figure 1.1.3 shows the model predictions of the relative contributions from all phosphorus sources to the total phosphorus loading to basin rivers. Groundwater phosphorus accounts for 8% of the basin-wide total phosphorus loading to rivers in the Cannonsville Basin. Surface runoff from active agricultural land is responsible for 68% of the total phosphorus loading to rivers in the Cannonsville Basin. The point sources in the SWAT model (see Section 4.5) include wastewater treatment plants (WWTPs) and account for 9% of the phosphorus loading to rivers. Note that due to the tightened water quality standards imposed on WWTPs, the point sources currently account for significantly less than 9% of the total phosphorus loading to rivers. Surface runoff from urban lands accounts for only 1% of the total phosphorus load to rivers because the urban land area in the basin is very small. Forests generate very little surface runoff in the basin. Therefore, only 4% of the total phosphorus load to rivers is derived from forest surface runoff. Note that most of the groundwater phosphorus comes from groundwater generated underneath forests.

There are other much less significant sources of phosphorus in the Cannonsville Basin that are not explicitly modeled in the current SWAT model of the basin. These include septic systems and highway runoff phosphorus. Although septic system phosphorus loading is not explicitly modeled, groundwater phosphorus loading implicitly includes at least a portion the septic system phosphorus loaded to basin surface waters based on the methodology for setting groundwater phosphorus concentrations in the model (see Section 4.6). The phosphorus loading to surface waters from urban land includes the phosphorus loading in urban stormwater and at least a portion of the phosphorus derived from highway runoff. However, since most of the highways in the basin are not identified as urban land in the spatial data, SWAT does not explicitly model the phosphorus loading from highways. The SWAT model uses a simplified approach for estimating urban phosphorus loading. In addition, no data were available to evaluate the accuracy of SWAT urban land use loading estimates. Preliminary loading estimates before modeling with SWAT showed that urban stormwater and highway runoff phosphorus were quite small relative to other estimated sources. The magnitude of these preliminary estimates suggests that urban stormwater might account for 2% of the total reservoir phosphorus load while highway runoff phosphorus could account for, at the very most, 7% of the total reservoir phosphorus load. It is however, important to note that highways can act as a conduit that carries phosphorus otherwise destined for the basin streams by natural flows (e.g. carrying runoff from adjacent forested land), rather than act purely as a source of phosphorus. The current modeling treatment for septic systems, urban land and highways is deemed appropriate for this modeling study given the lack of available data and the relatively low estimated phosphorus loading rates from these sources.

Future modeling of the Cannonsville Basin with SWAT could explicitly incorporate septic systems, highways and a more detailed model of urban lands if this is deemed worthwhile.

1.1.3 Application 3: Basin-wide Phosphorus Loading Estimates to Reservoir

This application shows that the SWAT model can be used to predict phosphorus loads for the unmonitored area (measured loads are unavailable) of the Cannonsville Basin (see Figure 1.1.4) that are considered more appropriate than an alternative method based only on extrapolating monitoring data. In fact, the SWAT model predicts substantially lower phosphorus loads entering the reservoir from these unmonitored areas than estimates derived from the alternative method. The analysis is outlined in detail in the following paragraphs.

The current NYSDEC water quality monitoring station at Beerston provides annual phosphorus loading estimates to the reservoir from areas upstream of Beerston. However, the Beerston drainage area is just under 80% of the total drainage area of the basin (see Figure 1.1.4). Thus, loading for the other 20% of the basin must be estimated. Figure 1.1.4 outlines the areas of the basin that are measured (or monitored) and unmonitored for phosphorus loading. In the absence of actual monitoring data, probably the most reliable way to estimate the loading from unmonitored areas is to use a spatially distributed model such as SWAT to predict the loading. This approach takes into account the intrinsic differences in land uses and phosphorus sources between the measured and unmonitored areas shown in Figure 1.1.4.

Evaluation of TMDL compliance requires estimating total annual loads to the reservoir. The current phosphorus TMDL for the basin using a reservoir concentration of 20 µg/L is 53,650 kg per year while the available or allowable load to the reservoir (which is the TMDL minus a margin of safety) is 46,944 kg/yr (see Kane 1999). To comply with the TMDL, total loading to the reservoir must be less than the *allowable* load.

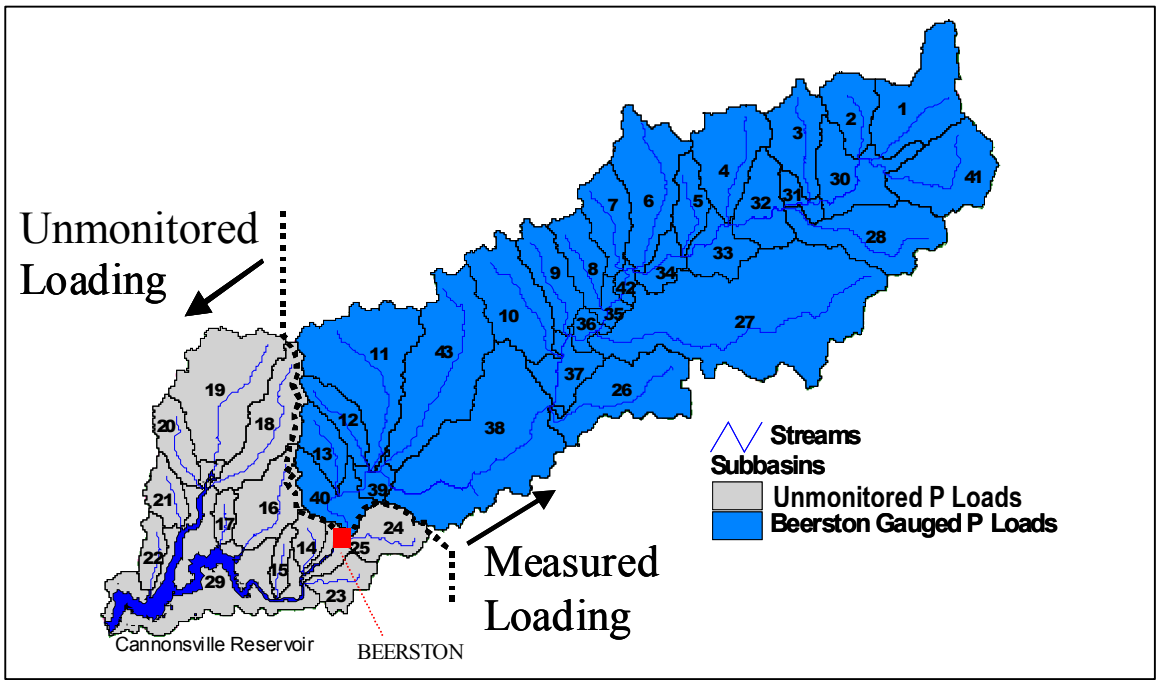

Figure 1.1.4. Measured and unmonitored portions of the Cannonsville Basin.

Given the available monitoring data at Beerston and the developed SWAT model there are two relatively simple methods for estimating basin-wide total phosphorus loading to the Cannonsville Reservoir in previous years. The first method involves utilizing only the NYSDEC monitored phosphorus loading at Beerston and then extrapolating the estimated NPS phosphorus load at Beerston to the unmonitored subbasins by assuming the NPS phosphorus loading per unit area of the unmonitored subbasins is assumed equal to the measured NPS phosphorus loading per unit area of the Beerston drainage area. This method will be referred to as the data extrapolation approach. The data extrapolation approach has been previously used to estimate basin-wide total phosphorus loading to the Cannonsville Reservoir (Delaware County Board of Supervisors 1999). In the absence of any watershed model, the data extrapolation approach based completely on monitoring data is probably the best approach. A second method is possible when a spatially distributed watershed model like SWAT is used to model phosphorus loading from the unmonitored subbasins. In this case, when both Beerston monitoring data and model output from SWAT are available, the measured phosphorus load at Beerston and the SWAT simulated phosphorus load from the unmonitored subbasins (e.g. downstream of Beerston) can simply be added together to estimate the total phosphorus load to the Cannonsville Reservoir. This approach will be referred to as the data and model approach. Of course, in the absence of any monitoring data at Beerston, a third method to estimate basin-wide phosphorus loading to the reservoir would be to use the SWAT model to predict phosphorus loading from the entire reservoir basin.

Figure 1.1.5 compares the TMDL allowable load to the estimated total phosphorus loads to the Cannonsville Reservoir derived from the data and model approach and the data extrapolation approach for water years 1992 through 2000. The data extrapolation approach estimates reservoir phosphorus loading levels that are always higher than the phosphorus loading estimated when the measured data are combined with the SWAT model phosphorus loading for the unmonitored subbasins. On average, from 1992-2000, the data extrapolation approach estimates the unmonitored subbasin phosphorus load to be 11,600 kg/yr while the data and model approach estimates only 3,500 kg/yr. Thus, the data extrapolation approach estimates 8,100 kg/yr more phosphorus (or 230%) than the SWAT model for the unmonitored subbasins. This finding is particularly important because an estimate of the phosphorus loading from the unmonitored subbasins is required to determine if there is a TMDL violation. For example, consider the reservoir phosphorus loading estimates for the year 2000 in Figure 1.1.5. The measured data extrapolation approach suggests the TMDL allowable load was violated by approximately 3,500 kg while the measured data and SWAT approach suggests that the loading was *below* the TMDL allowable load by approximately 1900 kg.

A closer examination of the available spatial input data disaggregated by measured (Beerston drainage area) and unmonitored subbasins demonstrates why the measured data extrapolation approach generates phosphorus loading estimates to the reservoir that appear to be much too high. First of all, land uses in the two regions are significantly different. For example, forested land makes up 72% of the area in the unmonitored area while forested land in the Beerston drainage area accounts for only 57% of the area. The increased forested proportion of area in the unmonitored subbasins replaces agricultural

land. Available cattle population data (see Section 4.7.1.1) show a substantially lower density of cattle, and thus reduced manure application, in the unmonitored subbasins. Also, no urban land use occurs in the unmonitored subbasins. Table 1.1.1 shows that significantly different land uses can lead to significantly different loading rates. These intrinsic differences in land use and phosphorus inputs between the Beerston drainage area and unmonitored subbasins are incorporated in SWAT model predictions but are not incorporated in the data extrapolation approach. These results show that predictions of the phosphorus loading from the unmonitored subbasins must take into account the specific land use and phosphorus inputs found in the unmonitored subbasins.

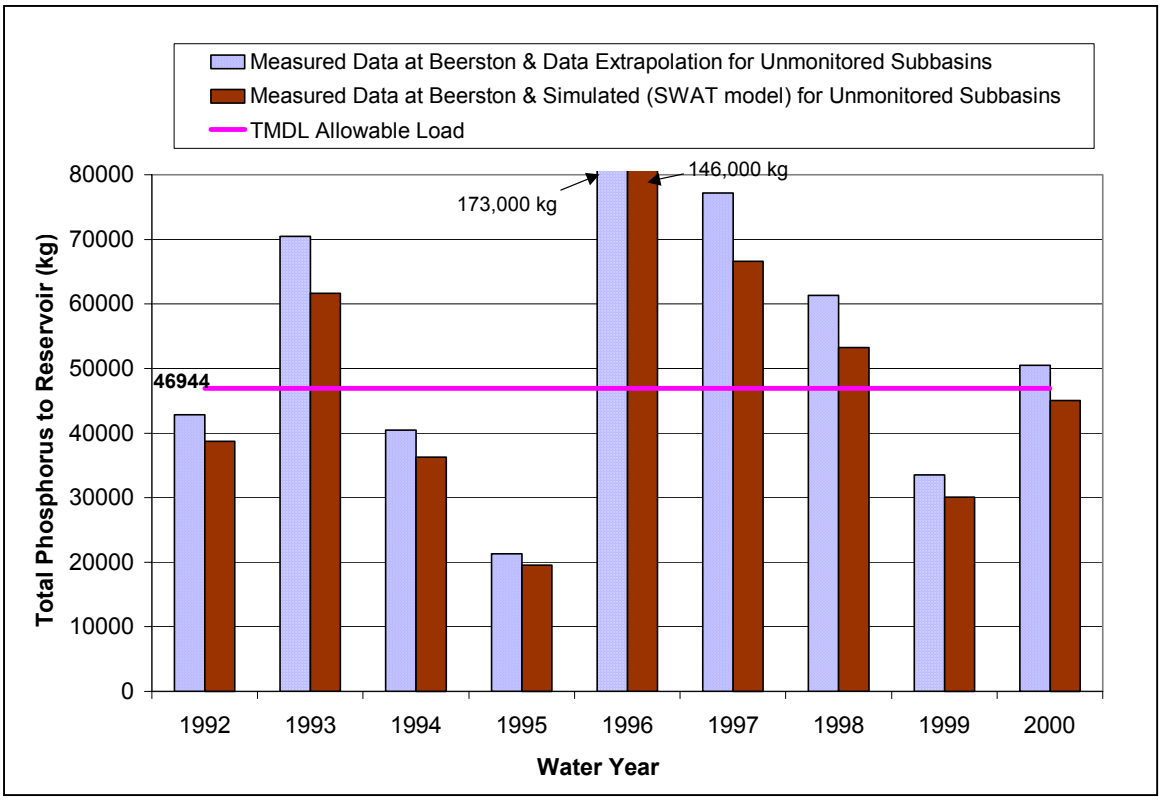

Figure 1.1.5. Estimated total phosphorus loads to the Cannonsville Reservoir from monitoring data and a combination of monitoring data and SWAT model predictions for water years (Oct. 1 – Sept 30) from 1992 to 2000.

1.1.4 Application 4: Spatial Variation in Phosphorus Loading

The current NYSDEC water quality monitoring station at Beerston monitors the cumulative phosphorus loading of all areas upstream of Beerston. However, this monitoring does not give any insight into the relative magnitudes of the phosphorus loading rates in the various locations or subbasins within the Beerston drainage area. With a spatially distributed model like SWAT, independent model predictions of phosphorus loading are made for subregions (called subbasins) within the watershed being modeled. Therefore, an area that has a relatively high percentage of agricultural land will be predicted to generate more phosphorus loading to streams than an equivalent

area of forests. Being able to identify critical areas of the basin where phosphorus loading is highest allows decision-makers to focus on the most significant problem areas and therefore make the most efficient management decisions.
Figure 1.1.6 shows the NPS total phosphorus load from each subbasin per unit area of each subbasin. Figure 1.1.6 demonstrates that upstream subbasins (especially subbasins 1, 4, 6 and 30) tend to have the highest NPS total phosphorus loading rates per unit area while many of the smaller subbasins around the reservoir have the lowest NPS total phosphorus loading rates per unit area. The spatial output displayed in Figure 1.1.6 is a sample of the spatial information that is output from the SWAT model. The effectiveness of management practices can also be evaluated spatially. The spatial output from SWAT can be used by Delaware County to locate high intensity phosphorus loading areas and also determine the areas in the basin that are predicted to respond best to various management practices.

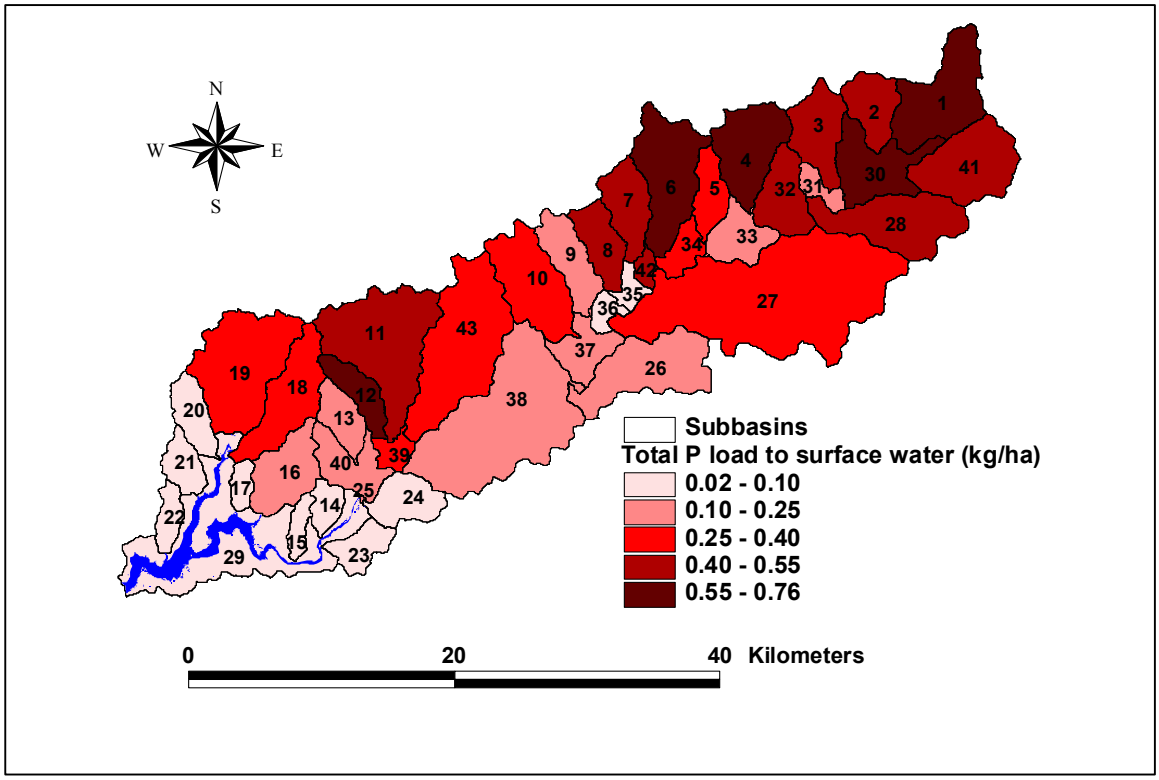

Figure 1.1.6. Average annual non-point source (NPS) total phosphorus loading rates per hectare of subbasin for the Cannonsville Basin from 1994-1999.

1.1.5 Future Modeling Analyses

The preceding analyses serve as informative examples to both modelers/scientists and decision-makers. There are a number of other useful modeling analyses (see Section 7.4) that can be conducted. For example, a critical examination of model uncertainties may be important to take into account when evaluating model predictions for management decisions. A later report will discuss and compare forecasts of future phosphorus loading under a variety of potential management strategies.

1.2 Summary

In summary, the validation results are encouraging. SWAT also gives a useful account of the spatial distribution of the loadings of phosphorus within the basin. The previous model applications demonstrate a few of the useful and insightful findings that were only made possible by the SWAT modeling approach employed in this study.

2 Case Study Introduction

The purpose of this report is to provide a comprehensive account of the current watershed modeling effort of the Cannonsville Basin initiated under the Delaware County Action Plan (DCAP) (Delaware County Board of Supervisors 1999). More specifically, this document reports the status of the modeling efforts completed as of September 2003. This version of the modeling report is the final version of a draft report that was peer reviewed by 5 agencies in the summer of 2002. The intended audience of this current report is experienced watershed modelers and scientists and those with prior knowledge of input or monitoring data specific to the Cannonsville Basin. However, Sections 1 and 2 are written for a more general audience.

The Cannonsville Reservoir is one of NY City's largest drinking water reservoirs and is located in Delaware County in the Catskill region of Upstate NY. Figure 1.2.1 shows a general map of the Cannonsville Basin. The 1178 km^2 Cannonsville Reservoir Basin is primarily forest and agricultural while less than 0.5% of the basin is urban. The major villages in the basin are Walton, Delhi, Hobart and Stamford. The elevation of the basin varies from 285 m above mean sea level in the lowland areas to around 995 m for the hilltops and the average slope of basin lands is approximately 19%.

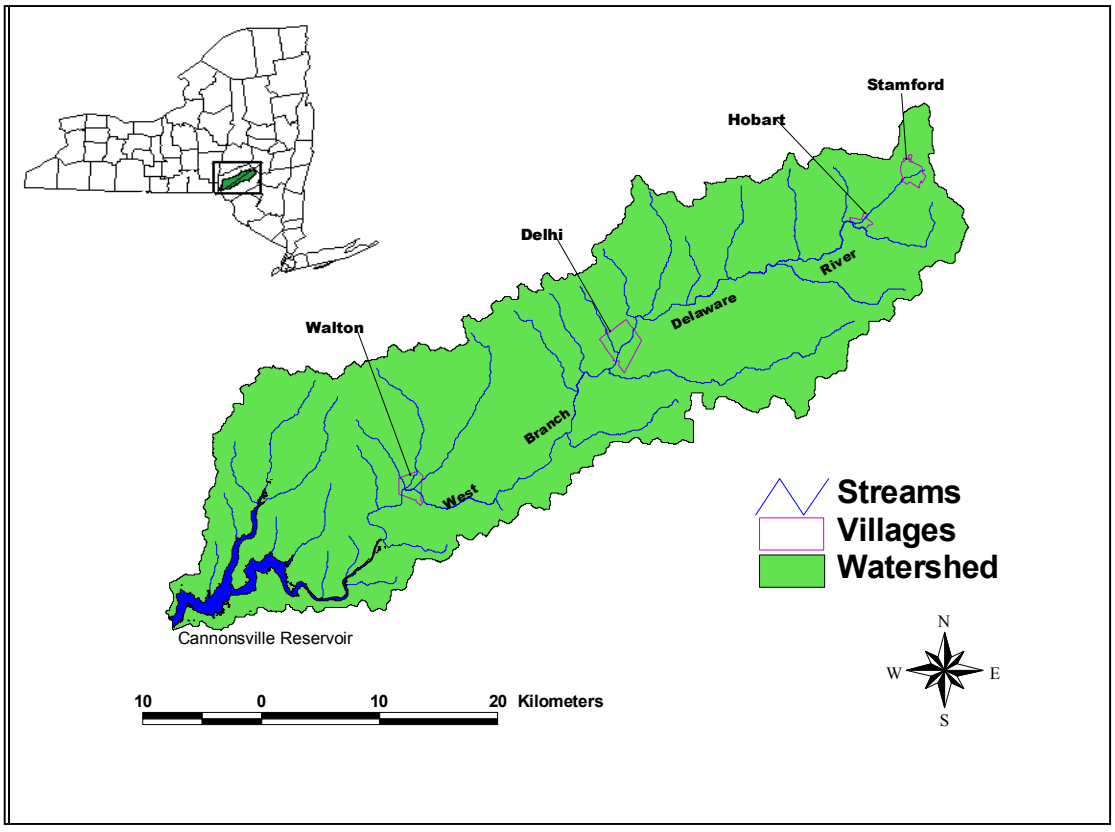

Figure 1.2.1. The Cannonsville Basin in NY State.

2.1 Purpose of Modeling

The Cannonsville Reservoir has historically experienced severe water quality problems due to excessive phosphorus loading. In addition to improving drinking water quality for New York City residents, effective phosphorus management is crucial in the Cannonsville Basin for two other reasons. First, Delaware County (containing nearly all of the watershed area), is 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 County through a Memorandum of Agreement (MOA) with New York City (http:// www.cwconline.org/about/moa/moaindex.htm) if reservoir phosphorus concentrations become too high. These MOA regulations have in the recent past triggered a 'phosphorus restriction' in Delaware County, which restricts future economic growth in the basin when the growth directly or indirectly increases phosphorus loadings. Secondly, if water quality levels in New York City's reservoirs are not satisfactory, New York City may need to construct a water filtration plant with costs approaching 10 billion dollars.

Whole-basin scale modeling of phosphorus will assist in identification, adoption and evaluation of management measures proposed in the Delaware County Action Plan (DCAP) (Delaware County Board of Supervisors 1999). The DCAP is available at http://wri.eas.cornell.edu/projects/nycwshed/delaware/. Results from the model will help evaluate and guide the phosphorus reduction program. In addition, the modeling work will predict reductions in phosphorus loadings that can otherwise only be identified over a considerably longer time frame by monitoring. This modeling work has been done with reference to previous TMDL modeling work of the Cannonsville Basin by New York City Department of Environmental Protection (NYCDEP) (Schneiderman et al. 1998, Schneiderman et al. 2002) and in conjunction with the New York State Department of Environmental Conservation (NYSDEC) tributary monitoring and loading analyses.

The main goal of this modeling is to describe and explain the cause and effects of existing and prospective land uses on loading of phosphorus delivered to the reservoir. Meeting this goal initially requires the development and validation against measured data of a suitable watershed model. Therefore, the main purpose of this report is to outline the development of the SWAT Cannonsville Basin Model and then demonstrate that model performance is reasonably accurate to evaluate various phosphorus management options in the basin.

2.2 Quality Assurance

The project principals filed a Quality Assurance Project Plan (QAPP) with NYSDEC and US EPA Region 2 (NYSWRI and DCPD 2001). The QAPP is available at http://wri.eas.cornell.edu/projects/nycwshed/delaware/.

The QAPP indicates that the primary success criterion for watershed modeling will be the acceptance of the calibrated model for strategic planning purposes, such as allocation of implementation funds across source categories. The acceptance must come from the

Scientific Support Group, NYSDEC and other state regulators, NYCDEP, and US EPA. The county agencies and municipal elected officials must also be comfortable that the model gives a reasonable representation of local conditions. The work leading to this report is intended to provide a scientifically sound model application and sufficient understanding within constituent agencies to put the work to productive use within the various watershed management programs.

The QAPP committed to several aspects of model application and consultation, which have been followed closely:

- The basin model has been calibrated using standard techniques against hydrologic data collected by USGS (United States Geological Society) at six gauge stations, and water quality data collected by NYSDEC at two stations. Bi-weekly phosphorus sampling results from many NYCDEP water quality stations have also been taken into account in developing model inputs and qualitatively assessing model performance.
- The calibration process used graphical representations (maps and time series plots) and standard mathematical measures of goodness of fit between measured and simulated time series.
- Co-operator reviews have been initiated, via this report and meetings.
- Discussions about software and data installation at four selected co-operator sites have been initiated.
- Consultations with the Delaware County Phosphorus Study Committee and the Scientific Support Group have continued, most recently in June 2002 when management scenarios were discussed.

Going beyond the published QA plan (which only covered through March 2002), the project proceeded to perform standard types of model "validation" which involves validation of the calibrated model against measured data not included in the calibration process. This is analogous to NYCDEP's GWLF calibration/validation approach (Schneiderman et al. 1998) using different time periods. NYCDEP's work has been accepted for major regulatory purposes, including the phosphorus TMDL process.

3 Soil and Water Assessment Tool (SWAT) Version 2000

The Soil and Water Assessment Tool version 2000 (SWAT2000) is a continuous time, physically based, distributed watershed model developed by the Agricultural Research Service of the United States Department of Agriculture (USDA). SWAT, which was preceded by the Simulator for Water Resources in Rural Basins, requires a significant amount of data and parameters for development and calibration. The principle purpose of SWAT is computation of runoff and loadings from rural watersheds, especially those dominated by agriculture (Williams and Arnold 1993, Arnold et al. 1998). Although the model time step is daily, 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).

SWAT was selected for this modeling because model predictions are distributed spatially. A simpler alternative to a spatially distributed model is a lumped watershed model. Lumped models make a single prediction that only applies at the watershed outlet and therefore provide no additional spatial information regarding the upstream sources of the modeled quantities. For example, the GWLF watershed model, which was previously applied to the Cannonsville Basin (Schneiderman et al. 1998, Schneiderman et al. 2002), is a lumped watershed model. Another reason for selecting the SWAT model was that it has been widely applied throughout the United States (see Section 3.1). The United States Environmental Protection Agency (USEPA) considers the model to be an acceptable tool for development of Total Maximum Daily Loads (TMDLs). More recently, the model was used for TMDL development in Texas (Santhi et al. 2001). SWAT2000 has been integrated into the latest version of the USEPA's publicly available environmental analysis system software package, BASINS 3.0, which can be found at http://www.epa.gov/waterscience/BASINS/.

SWAT2000 allows for two different watershed discretizations. In this application, two levels of discretizations are used. First, the watershed is subdivided into a number of subbasins. This is the spatial scale at which the model routes constituents through the channel network. These subbasins are further divided into hydrologic response units (HRUs), which are units of unique intersections of land use and soils. For example, all the corn land on soil type A in a certain subbasin is one HRU. HRUs are not necessarily contiguous land areas and no channel routing is simulated between HRUs. Most computations in SWAT occur at the HRU spatial scale where the state variables in each HRU, such as soil water, plant biomass and soil phosphorus concentrations, are updated daily.

The model inputs for SWAT2000 are developed with the aid of the SWAT2000 Arcview© Interface Program (DiLuzio et al. 2001) that automatically assigns default model parameter values and creates input files based on various Geographic Information System (GIS) map layers provided to the interface. In addition, the SWAT2000 Arcview© Interface Program (AVSWAT) can be used to run the model and map the results in within Arcview©.

Some of the major watershed processes are simulated in this SWAT2000 application as follows:

- Surface water runoff is calculated using the Soil Conservation Service (SCS) curve number equation (SCS 1972).
- Potential evapotranspiration is estimated using the Priestly-Taylor equation and is corrected for land cover, based on simulated plant growth, to give actual evapotranspiration (Neitsch et al. 2001a).
- Percolation occurs when soil water content exceeds field capacity and determines the amount of water moving from one soil layer to the next using a storage routing method (Neitsch et al. 2001a).
- SWAT simulates two groundwater aquifers in each subbasin: an unconfined, shallow aquifer that contributes to stream flow and a deep aquifer that does not add to stream flow within the modeled watershed (Arnold et al. 1993).
- Lateral subsurface flow is simulated using a kinematic storage model for subsurface flow (Neitsch et al. 2001a).
- Sediment erosion from each HRU is simulated using the Modified Universal Soil Loss Equation (MUSLE) (Williams and Berndt 1977). This equation replaces the traditional Universal Soil Loss Equation's (USLE) rainfall factor with a runoff factor in order to estimate event-based sediment yield estimates. MUSLE predicts sediment erosion for each day there is surface water runoff and reduces the erosion estimates when there is snow cover.
- Plant growth and nutrient uptake.
- Mineral phosphorus and organic phosphorus cycling in the soil between six different pools of soil phosphorus.

For more details on the entire suite of processes modeled in SWAT, as well as model usage instructions, the reader is referenced to the following documents:

- SWAT2000 Theoretical Documentation (Neitsch et al. 2001a).
- SWAT2000 User's Manual (Neitsch et al. 2001b).
- SWAT2000 Arcview Interface Guide (DiLuzio et al. 2001).

All of the above documents and associated programs can be downloaded from http://www.brc.tamus.edu/swat. In general, for the remainder of this document, the acronym SWAT will refer to the SWAT2000 model unless otherwise stated.

3.1 Previous Applications of SWAT

Most of the applications of SWAT in the current literature are based on versions of SWAT that precede version 2000. Various versions of SWAT have been applied throughout the United States, primarily in Midwest and Southwest regions, including the Mississippi River basin, Illinois, Indiana, Texas, Oklahoma, and Wisconsin (Arnold and Allen 1996, Bingner 1996, Bingner et al. 1997, Mamillapalli 1998, Manguerra and Engel 1998, Srinivasan et al. 1998, Arnold et al. 1999, Arnold et al. 2000, FitzHugh and Mackay 2000, Santhi et al. 2001, Kirsch et al. 2002). SWAT has also successfully been

applied in Europe (Shepherd et al. 1999) and has been successfully adapted and modified for European conditions (Eckhardt and Arnold 2001). Most of the previous research has only focused on hydrology simulation in these regions, with only a few recent publications on additional constituents such as sediment or nutrients (Bingner 1996, Bingner et al. 1997, Srinivasan et al. 1998, FitzHugh and Mackay 2000, Santhi et al. 2001, Kirsch et al. 2002). The few studies applying SWAT in Northeastern climates have centered primarily in northeast Pennsylvania (Cho et al. 1995, Peterson and Hamlett 1998). SWAT was applied to a 250 km^2 subwatershed of the Delaware River in Northeast Pennsylvania (the ultimate receiving water for the Cannonsville Reservoir outflow) in order to simulate hydrology and nitrogen (Cho et al. 1995). Peterson and Hamlett (1998) focused their work on modeling a catchment in Pennsylvania that was dominated by fragipan soils and experienced 'severe' snowmelt events throughout the winter and spring.

Of the SWAT studies mentioned previously, the one that is perhaps most comparable with this one in terms of monitoring data quality and quantity for nutrients and sediment loads, in addition to hydrology, is the study by Santhi et al. (2001). Although the other previous studies provide reasonable general references for SWAT applications, the study by Santhi et al. (2001) provides the most useful information and approaches for model development. For example, the area modeled by Santhi et al. (2001) is approximately the same size as the Cannonsville Basin and the area is dominated by dairy farming. Furthermore, based on the data availability and the accurate model performance in Santhi et al. (2001), their paper provides the most useful comparison case for evaluating the performance of SWAT on the Cannonsville Basin relative to any other application of SWAT currently in the literature. Therefore, the study by Santhi et al. (2001) is closely compared to this one in terms of model predictive performance. The study by Kirsch et al. (2002) is also briefly discussed. Comparisons to these two studies can be found in Section 7.1.

4 Model Inputs

Benaman (2002) initially developed the Cannonsville Basin SWAT model for hydrology and sediment using SWAT version 99.2. The model was initially calibrated to hydrology and sediment data only (Benaman and Shoemaker 2003a, Benaman et al. 2003b). The current report is based on upgrading the version 99.2 model to version 2000, re-calibrating hydrology and sediment and then calibrating the model for phosphorus. In addition, the current model/report was updated and improved based on a peer review of a July, 2002 draft of the report.

This section of the report outlines the development of the model inputs. Specifically, this section outlines the raw data sources utilized, translation of these raw data inputs to original SWAT2000 ASCII input files, deviations from default model settings, initial conditions in the model and any optional physical processes selected for simulation in SWAT. A considerable effort was made to generate the most representative model inputs as possible given the vast data available for the Cannonsville Basin. Much of this derived input data will be beneficial to any future watershed models of the Cannonsville Basin.

Model simulations covered the period from Jan. 1990 through Sept. 2000 and were made up of separate model runs over two time periods. Inputs had to be specified for both the model calibration period (Jan. 1994 to Sept. 2000) and a model validation period (Jan. 1990 to Dec. 1993) as defined in Section 6. Although model inputs for both periods were the same in most cases, there are instances where model inputs were different. Each of the following input sections clearly outlines to which period the inputs apply.

4.1 Model Input File Generation using the SWAT2000 Arcview Interface (AVSWAT)

AVSWAT (DiLuzio et al. 2001) was used to create the initial model input files. AVSWAT processes mapped land use (Section 4.2) and soils (Section 4.3) data as well as a Digital Elevation Map (DEM) to create a set of default model input files. The DEM can be utilized by AVSWAT to delineate basin (or watershed) and subbasin boundaries, calculate subbasin average slopes and delineate the stream network. NYCDEP supplied a 30-m DEM, derived from 1:24,000-scale USGS quadrangle sheets, used in this application. The DEM was utilized to delineate watershed boundaries and calculate subbasin average slopes. However, the subbasin average slopes were replaced by HRU specific slopes in this application (see Section 6.3.2.2). The stream network was defined using the AVSWAT 'burn in' option based on US Census TIGER files for the Cannonsville Basin. Modeled subbasins were delineated based mainly on those designated by the NYCDEP that cover the major tributaries entering the WBDR and Cannonsville Reservoir. Additional subbasins were created using AVSWAT in order to subdivide the mainstem of the WBDR into multiple subbasins to enable more accurate routing and climate inputs along the mainstem. Additionally, some subbasins were also re-defined to coincide with various monitoring locations. A total of 43 subbasins were used to represent the Cannonsville Basin and are displayed in Figure 4.1.1.

Within each subbasin, HRUs are created by AVSWAT. HRU creation in AVSWAT requires land use and soil threshold inputs (DiLuzio et al. 2001) in order to define the level of spatial detail to include in the model. These thresholds are applied to each subbasin and function to control the size and number of HRUs created. For example, if a subbasin has an area of deciduous forest in the original land use data that covered less than the land use threshold, the AVSWAT processing would convert the deciduous forest to the other land uses in the basin, proportional to their relative sizes, that were above the land use threshold. For a more detailed description of the AVSWAT land use and soils threshold application refer to the AVSWAT manual (DiLuzio et al. 2001). The land use and soils threshold values used in this application were 1% and 5%, respectively, and were selected in order to keep the number of HRUs to a reasonable number while modeling most of the important agricultural cover types. Application of these thresholds eliminated a few of the land uses and soils that covered relatively small areas in the basin and created a total of 482 HRUs. HRUs are not necessarily contiguous parcels of land. The model calculated surface runoff and nutrient transport is unique to each HRU (e.g. influenced by the land use, soil and subbasin).

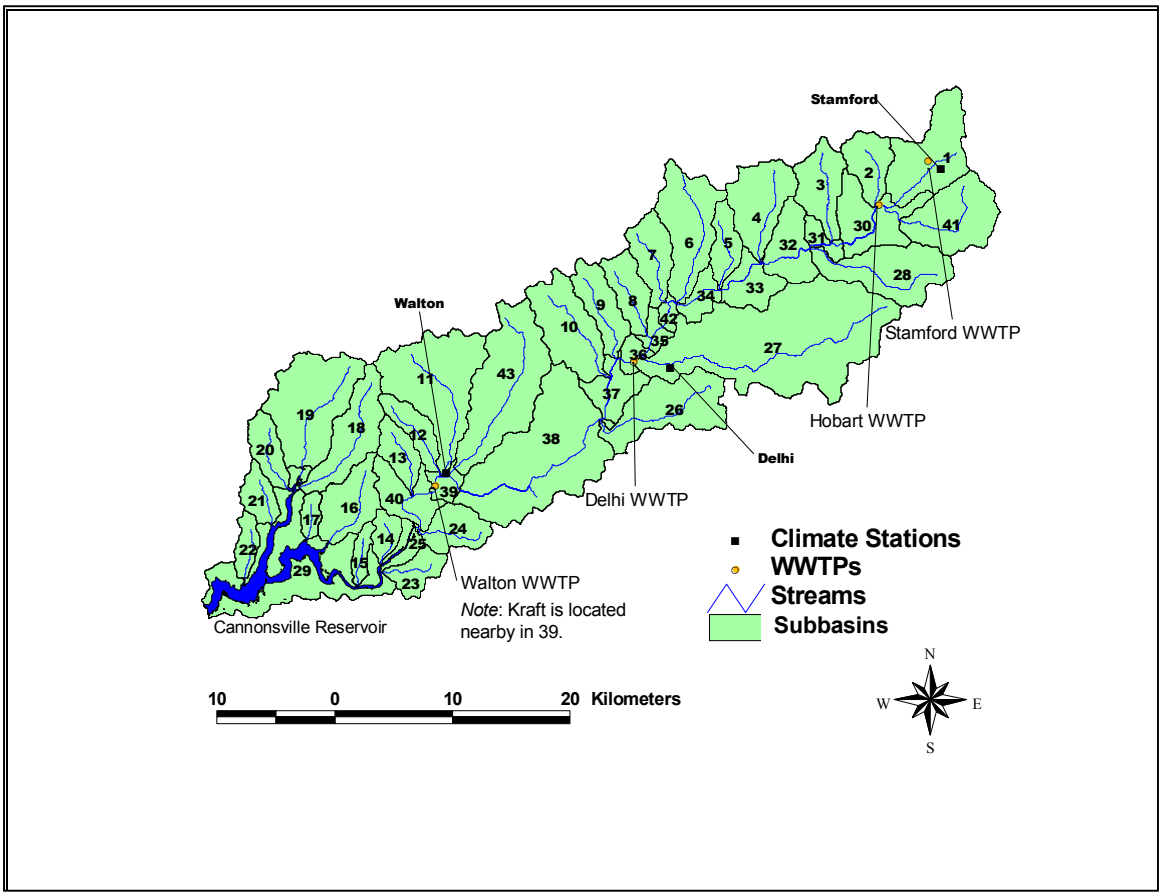

Figure 4.1.1. Cannonsville Basin subbasin delineation, climate station, and phosphorus point source locations.

4.2 Land uses

NYCDEP provided a 25-m grid land use coverage derived by supervised classification of 1992 and 1993 thematic mapper satellite imagery (see Figure 4.2.1). Forests, mainly deciduous, cover 59% of the Cannonsville Basin. Land associated with agriculture (grass, corn and alfalfa) cover 27% of the basin while successional agricultural land (referred to as grass_shrub in the original NYCDEP data) covers 11% of the basin. The remaining areas in the basin are covered by water (3%) and urban land (<1%). The NYCDEP data were processed by the SWAT2000 Arcview Interface (AVSWAT) to create an initial set of HRUs (see Section 4.1). AVSWAT processing slightly modified the original land use data by completely eliminating alfalfa (< 0.05% of the basin) and mixed forest (< 0.05% of the basin) as a land use in the model. The resultant agricultural HRUs in the model were either corn or grass.

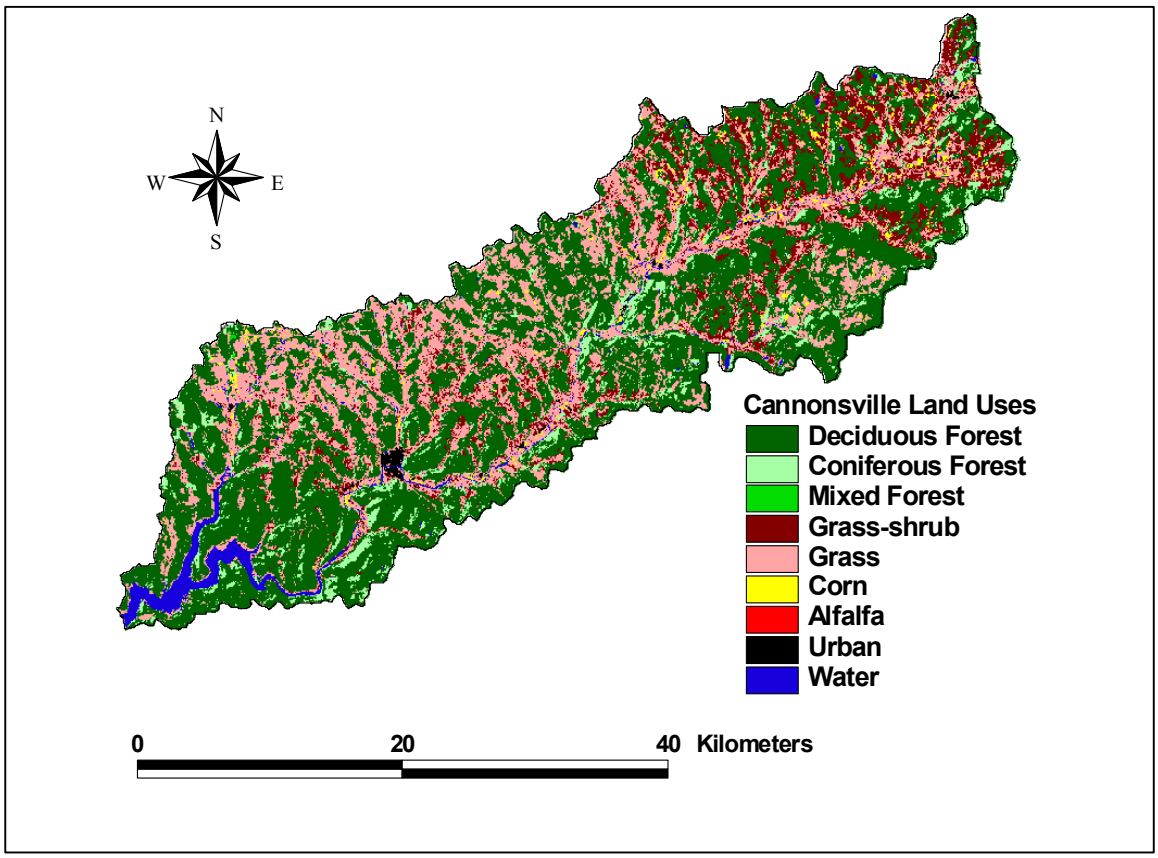

Figure 4.2.1. Land uses in the Cannonsville Basin as categorized by NYCDEP.

The initial set of 482 HRUs lacked corn HRUs in ten subbasins that contained corn as a land use before AVSWAT processing. In fact, the application of the AVSWAT land use threshold eliminated approximately 16% of the original corn area in the basin. Since corn receives the highest manure application rates of all agricultural lands, it was deemed important to ensure that AVSWAT retained 100% of the original NYCDEP corn land use. This is also critical because the unit area sediment and P loading rates to surface

waters from corn are an order of magnitude higher than other land uses in the basin (e.g. see Section 1.1.2). As a result, not modeling 16% of the total corn area would potentially force the model to misappropriate a large amount of P to other land uses. Therefore, to avoid losing corn area due to AVSWAT processing of land use data, the NYCDEP GIS data were reprocessed outside of AVSWAT and 22 new HRUs representing the originally eliminated corn HRUs were added to the model resulting in a total of 504 HRUs. This approach minimized the number of HRUs in the model since the 1% AVSWAT threshold for land use was still applied to all other land uses.

4.2.1 Additional Modifications to Land Use Classification

Land use classification of HRUs determines the type of plant that grows on the HRU and, if the HRU is agricultural, how the crop is managed. As such, default model parameters for plant growth, runoff and sediment erosion depend on the specified land use. Since land use classification is such a critical input to the model the available NYCDEP land use data were closely scrutinized before the land use inputs in the model were finalized. This process showed that two general categories of land use input modifications were necessary to better estimate the actual land uses in the Cannonsville Basin.

4.2.1.1 Trout Creek Corn Area Modification

Trout Creek (subbasin 19) is not monitored by NYSDEC and empties directly into the Cannonsville Reservoir. It was determined during initial model runs that subbasin 19 was generating an inordinately high concentration of total phosphorus in the flow leaving the subbasin. This simulated result was in strong disagreement with the available data from NYCDEP (see Section 6.4.4.1). A detailed investigation showed that the corn area in subbasin 19 was responsible for nearly 80% of the total phosphorus load. Furthermore, subbasin 19 contained 11% of the total corn area identified in the NYCDEP land use data even though it is only 5% of the total land area in the Cannonsville Basin. Dale Dewing (Personal Communication) noted that according to Watershed Agricultural Program records, there were no more than 36 ha of corn in subbasin 19 in the recent past. NYCDEP spatial land use data showed 158 ha of corn in subbasin 19. Therefore, given the simulated model result for Trout Creek was in complete disagreement with available water quality data and available farm records suggest that the corn area in Trout Creek is greatly over-estimated by the NYCDEP land use data, it was deemed necessary to reduce the area of corn modeled in subbasin 19 from 160 ha to 36 ha. The three original NYCDEP corn HRUs were all reduced to 23% of their original size and the removed area was redistributed evenly to NYCDEP grass and grass-shrub land uses.

4.2.1.2 Subdividing NYCDEP Grass Land Use

Unfortunately, the NYCDEP data do not distinguish all types of agricultural land use within the Cannonsville Basin. The typical farm in the basin has both pasture (grasses grazed by cattle) and hay (grass cut 2 or 3 times a year for cattle forage) and the hay is either continuous hay or rotated with corn. Hay and pasture are managed very differently with respect to the timing and application rates of manure. Rotated and non-rotated lands must be distinguished in order to model and evaluate the impact of crop rotations. In

order to independently represent pasture, rotated hay and non-rotated hay land use areas in SWAT2000 it was necessary to further edit and reclassify some of the HRUs derived from the NYCDEP land use data.

Local agricultural stakeholders and the 1992 US Census of Agriculture data for Delaware County were used to guide the reclassification of the initial grass HRUs derived from the NYCDEP land use. The overall reclassification was accomplished in the following steps:

1. The 1992 US Census of Agriculture was used to estimate the areas of pasture and hay located in the Cannonsville Basin and it was assumed that all hay and pasture in the basin were included within the NYCDEP grass land use category.
2. Grass HRUs were partitioned to create new HRUs that could be modeled as rotated corn-hay land use so that 100% of the corn HRUs could be rotated with hay using typical rotation lengths while keeping the total area of corn constant in each year of the model simulation. These new HRUs, along with the original corn HRUs, represent the area of rotated corn-hay land use.
3. The remaining grass HRUs not yet reclassified were defined as either continuous hay, pasture or idle agricultural land so as to approximate the basin-wide areas of hay and pasture determined in step 1.

STEP1. A summary of the available data used to define the SWAT2000 modeled areas of agriculture land uses (active and successional) is given in Table 4.2.1. Table 4.2.1 compares estimated agricultural land use areas for the Cannonsville Basin from the 1992 US Census of Agriculture for Delaware County and from the original NYCDEP GIS data. Although corn area estimates from both sources are reasonably close, the estimated hay+pasture area from the 1992 US Census of Agriculture is only 68% of the area identified as grass in the NYCDEP data. The NYCDEP grass data probably include other types of agriculture. However, data from the 1992 Census of Agriculture suggest that areas of other types of agriculture (Oats, vegetables harvested and berries) are only about 200 ha in the Cannonsville Basin and therefore cannot account for this discrepancy. This discrepancy can be partly explained by the fact that in any given year there are a significant number of fields of idle or unmanaged grasses, that are not necessarily successional agricultural land, but are also not reported as hay or pasture in US Census of Agriculture (Personal Communication, Dale Dewing). Another cause for the discrepancy is that NYCDEP data likely include grasses that are not managed for agriculture such as front lawns and parks. In fact, it is more likely that the hay+pasture area in the 1992 US Census of Agriculture is closer to 60% of the area identified as grass in the NYCDEP data considering that the 1992 US Census of Agriculture data for pasture includes areas of brush that presumably would have been categorized as grass-shrub in the NYCDEP data. Therefore, the assumption that 100% of the hay and pasture in the basin is located completely within the NYCDEP grass land use is reasonable. Further, it is reasonable to assume that a fourth subtype of agricultural land use within NYCDEP grass is idle agricultural land. Consequently, the NYCDEP land use category of grass-shrub was assumed to be 100% successional agricultural land that is not managed.

Table 4.2.1. Comparison of estimated land use areas for the Cannonsville Basin based on the 1992 US Census of Agriculture for Delaware County and the 1992 NYCDEP GIS land use map for the Cannonsville Basin.

Generic Land use Name	Equivalent US Census of Agriculture Name and Description	Area from 1992 US Census of Agriculture for Delaware County (ha)	Cannonsville Basin Area (ha)		
			Upper bound *estimate* based on 1992 US Census of Agriculture for Delaware County[1]	NYCDEP GIS data for 1992 (equivalent NYCDEP land use name)	Area modeled in SWAT2000[2]
Corn	Corn for silage or green chop + corn for grain or seed	4300[3]	1927	1519 (corn)	1387
Hay	Hay - alfalfa, other tame, small grain, wild, grass silage, green chop etc.	24196	10840	See Hay+Pasture below	15020
Pasture	Cropland pasture + permanent pasture or rangeland	21257[4]	9523	See Hay+Pasture below	3542
Hay + Pasture	See above descriptions	49753	20462	30305 (grass[5])	18562
Successional Agriculture	NA	NA	NA	12534 (grass-shrub[6])	12269

1. Estimated at 45% of Del. County 1992 US Census of Agr. based on ratio of total farms in Cannonsville Basin (Delaware County Board of Supervisors 1999) to farms in 1992 US Census of Agr. for Delaware County = 321/716. A lower bound estimate would be based on the ratio of the areas of Cannonsville Basin to Delaware County (~30%).
2. SWAT2000 areas are the result of AVSWAT processing and other land use reclassification outlined in Sections 4.1 and 4.2.
3. Corn for silage or green chop (~90% of total) + corn for grain or seed (~10%).
4. Cropland pasture (~50%) + permanent pasture or rangeland (~50%). Permanent pasture or rangeland includes brush.
5. NYCDEP description of grass land use: Includes turf, pasture and grass hay.
6. NYCDEP description of grass-shrub land use: Successional land composed of grasses, forbs and woody shrubs.

STEP 2. The second step of the grass HRU reclassification involved the creation of a set of HRUs that could be used to specify corn-hay crop rotations in the basin. This step was accomplished based on the following assumptions and modifications:

- Total area of modeled corn in SWAT should be constant in all simulation years and should be equal to the NYCDEP corn land use area after the Trout Creek corn area reduction outlined in Section 4.2.1.1.
- Rotations occur within a subbasin.
- HRUs should be created so that 100% of the modeled corn in the basin can be simulated under rotation with hay.
- HRUs within a rotation all have the same soil type as dictated by the soil type of the original corn HRU.
- Upland and lowland soils (see Figure 4.7.3) were assumed to have rotations that were 2 years corn / 8 years hay and 3 years corn / 6 years hay, respectively.
- In the six cases where there was not enough area in the grass HRUs to define multiple hay HRUs that would satisfy the assumptions above, either A) the corn HRU area was lumped into another corn HRU with a different soil type (1 ha of corn land had reassigned soil type) or B) the required grass HRU areas were increased by reducing the size of grass-shrub HRUs on the same soil type (152 ha of new grass HRUs created).

STEP 3. The remaining grass HRUs that were not reclassified as potential rotated corn-hay HRUs had to be reclassified as pasture, continuous hay or idle agricultural land. The pasture land use area in the Cannonsville Basin estimated at about 9500 ha from the US Census of Agriculture overestimates the area of actively grazed pasture in the basin (Personal Communication, Paul Cerosaletti). As noted above, this is at least partly due to the fact that some of the pasture area in the 1992 US Census of Agriculture includes brush. Assuming an average area of pastured cropland of 14 ha/farm for Delaware County (derived from Watershed Agriculture Program data as summarized in Delaware County Board of Supervisors (1999)) and 226 beef and dairy farms in the Cannonsville Basin (see Section 4.7.1.1), a better estimate of pasture (as grass) in the basin would be about 3200 ha. Therefore, reclassification of the remaining NYCDEP grass HRUs focused on generating a total area of pasture in the basin that was closer to 3200 ha instead of 9500 ha.

A reclassification algorithm was developed to complete Step 3 and operated on each subbasin independently. After Step 2, there were between one and four remaining grass HRUs in each subbasin to be reclassified as pasture, continuous hay or idle agricultural land. The algorithm was guided by the areas of the grass HRUs and the total area of corn HRUs within the subbasin. The resulting reclassified HRUs in each subbasin always had the following characteristics:

- All subbasins with corn HRUs had grass HRUs reclassified such that there was at least one pasture and one continuous hay HRU.
- If there were no corn HRUs in the subbasin, in which case there were also no cattle located in the subbasin (see Section 4.7.1.1), then all remaining NYCDEP grass HRUs were reclassified as idle agricultural land.
- The smallest grass HRUs tended to be reclassified as pasture.

- Grass HRUs reclassified as idle agricultural HRUs, were within an order of magnitude of the total area of continuous hay HRUs within the subbasin.

In summary, the reclassification outlined in Steps 1, 2 and 3 above functioned to generate a reasonable set of HRUs to independently represent areas of continuous and rotated hay, idle agriculture and pasture. The reclassified basin-wide land use areas modeled in SWAT are reasonable because they appear to be in general agreement with the available data (see Table 4.2.1) and are an attempt to independently represent each of the major agricultural land uses that are known to exist in the Cannonsville Basin. The reclassification of NYCDEP grass HRUs did not improve the spatial accuracy of the four grass land use subcategories. However, the impacts of this spatial inaccuracy were deemed acceptable in order to represent the four subcategories of grass land use in SWAT2000. This representation of grass land use is an improvement over a lumped approach which models the NYCDEP grass land use with one set of model inputs and parameters as in previous Cannonsville Basin modeling (Schneiderman et al. 1998).

The finalized land uses and their corresponding areas used as inputs in the SWAT2000 model of the Cannonsville Basin resulting from the cumulative land use modifications described above are summarized for the agricultural land uses in Table 4.2.1 and for all land uses in Table 4.2.2. In total, 758 HRUs are used to represent the entire Cannonsville basin. The final area of hay and pasture modeled in SWAT as shown in Table 4.2.1 is now much closer to the hay and pasture area estimated by the 1992 US Census of Agriculture for Delaware County. The data in Table 4.2.2 show that only about 17% of the basin is modeled in SWAT as managed agriculture (hay, pasture and corn). Land use data were assumed constant for both the calibration and validation periods.

Table 4.2.2. Final SWAT2000 land use input summary for the Cannonsville Basin

Modeled Land Use	**Area in km (percent of total)**
Deciduous forest	622 (52.8)
Coniferous forest	77 (6.5)
Grass-Shrub	124 (10.5)
Idle Agriculture	123 (10.4)
Hay[1]	150 (12.7)
Pasture	35 (3.0)
Corn	14 (1.2)
Urban	4 (0.4)
Water	29 (2.4)

1. All hay in this version of the model is continuous and the total area modeled is equal to the sum of the areas of hay HRUs created for rotation and those that were classified as continuous (see following paragraph for further explanation).

The HRUs in the current version of the model were defined to accommodate crop rotations. However, the current model input files do not specify crop rotations such that all corn and hay are simulated as continuous crops. Crop rotations can be incorporated at a later date by modifying the HRU management file inputs.

4.3 Soils

SWAT uses the State Soils Geographic Database (STATSGO) to describe the physical characteristics of soils. STATSGO spatial data (current as of 1994) downloaded from the USDA/Natural Resource Conservation Service (NRCS) website (http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/index.html) were utilized to map the various NY STATSGO soils (see Figure 4.3.1). STATSGO spatial data are available at a 1:250,000 resolution. The STATSGO data files formatted for SWAT for each soil type were downloaded from the SWAT homepage (http://www.brc.tamus.edu/swat).

The STATSGO soil properties assigned to the initial set of model HRUs (482) were 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 to 20 percent of the total area defined by each map unit. This default SWAT approach was judged to be a major deficiency in this study and required correction.

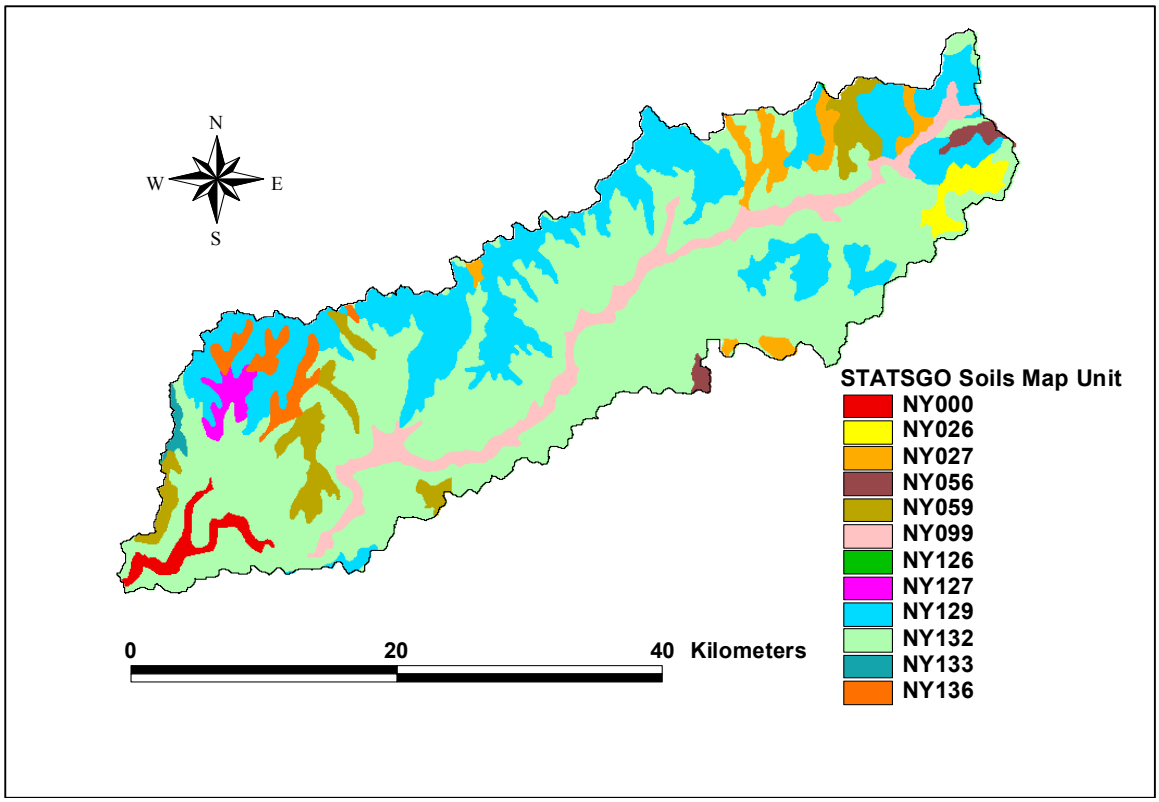

Figure 4.3.1. STATSGO soil classes in the Cannonsville Basin.

One alternative to overcome the STATSGO soil deficiency is to replace STATSGO soils with the more detailed Soil Survey Geographic Database (SSURGO). SSURGO soils

data maps out each soil series while STATSGO data groups together up to 21 similar soil series in each STATSGO map unit. A method for incorporating SSURGO data into SWAT is available (http://waterhome.tamu.edu/NRCSdata/SWAT_SSURGO/). However, the large scale of this modeling effort did not justify such an increase in soil detail since the study outlined on the above website recommends incorporating SSURGO soil data for small-scale modeling efforts only. Furthermore, since there are more than 10 times as many SSURGO soil map units than STATSGO soil map units in the Cannonsville Basin, using SSURGO soils to define HRUs would have increased the number of input files and model execution time by roughly one order of magnitude.

In order to assign more representative soil properties while minimizing the number of HRUs modeled in SWAT, the STATSGO soil map units were retained to represent the spatial extents of various soil types but the soil properties in each map unit were derived from the specific SSURGO soil series that occurred within each STATSGO map unit. SSURGO spatial data at a 1:24,000 resolution and data tables used in this study were downloaded from http://www.ncgc.nrcs.usda.gov/branch/ssb/products/ssurgo/. A GIS analysis was performed to intersect the STATSGO soil map units with SSURGO soil map units to determine the Cannonsville Basin specific soil series and their corresponding areas that occur within each STATSGO map unit. This intersection generated the necessary information for calculating more representative area-weighted soil parameters. Since SWAT also models up to 10 soil layers, the SSURGO data had to be analyzed to determine an appropriate number of soil layers to represent within each STATSGO map unit. Since there was a maximum of four soil layers in the SSURGO soils found in the basin, each STATSGO map unit was represented with four soil layers. Area-weighted average soil layer depths were first derived for each STATSGO map unit. Then, area- and depth-weighted average soil properties for each STATSGO map unit were computed for each soil layer. A more complete description of this methodology is given in the Appendix (Section 10.1). The result of this approach was to estimate soil properties for the STATSGO map units that better approximated the cumulative basin-wide soil properties that could be derived independently using only the detailed SSURGO soil data.

It should be noted that the current methodology for averaging soil properties preserves only the basin-wide area-weighted SSURGO values. Average soil properties within a subbasin are not necessarily the average values that would be derived from SSURGO soils data within the subbasin. Therefore, in the future, the spatial accuracy of the soil properties input to SWAT could be further improved if the methodology developed here was employed in each subbasin independently.

4.4 Climate Inputs

In addition to precipitation inputs, climate inputs are utilized in SWAT for a number of calculations including the prediction of crop growth, evapotranspiration and snowmelt. Daily climate data inputs required for this SWAT model application were minimum and maximum temperature, precipitation depths, solar radiation and relative humidity. All four of these inputs were based completely on measured data within or close to the

Cannonsville Basin. Care was taken to ensure that the model climate inputs were based purely on measured data (or filled in data from nearby stations) instead of allowing the model to randomly generate missing data. The two sources of the climate data were the Northeast Regional Climate Center (NRCC) at Cornell University and the National Climatic Data Center (NCDC). When both data sources had the same type of data available on the same day precedence was given to the NRCC data.

Four climate stations provided climate inputs that covered the model calibration and validation periods. All available climate data from within the Cannonsville basin were utilized. Precipitation data were taken from the Walton, Delhi and Stamford precipitation stations while temperature data were taken from the Walton and Delhi temperature stations. The location of the Walton, Delhi and Stamford climate stations are given in Figure 4.1.1. Precipitation and temperature data at the Deposit climate station, which is just outside of the basin and a few kilometres west of the reservoir, was used to supply missing data and to check data quality at the Walton station. The climate station at Binghamton, NY, which is approximately 60 km due east of the reservoir/watershed outlet, provided solar radiation and relative humidity inputs.

SWAT2000 assigns climate inputs to each subbasin in the model based on the closest climate station to the centroid of the subbasin. Thus, solar radiation and relative humidity input data for all subbasins in the model is from the Binghamton station. Subbasins 38 and 43, and all subbasins to the west of these, receive temperature inputs from the Walton climate station. All subbasins east of subbasins 38 and 43 receive temperature inputs from the Delhi climate station. Figure 4.4.1 shows the source of the precipitation input for all subbasins.

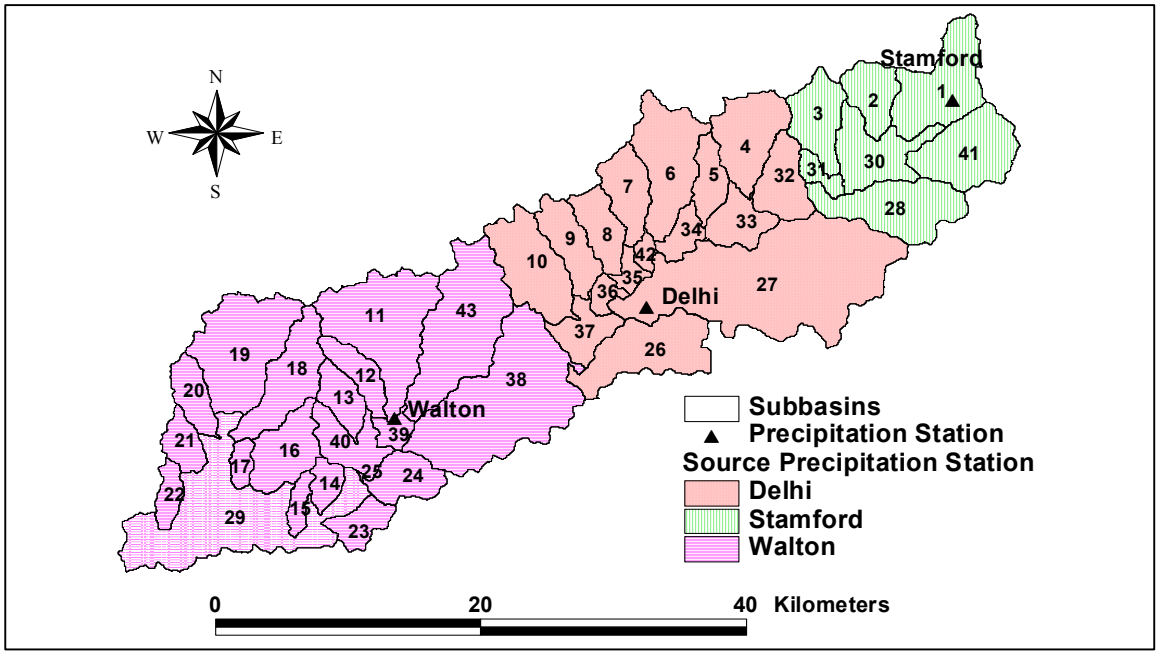

Figure 4.4.1. Source of subbasin precipitation inputs.

SWAT2000 allows users the option to model the orographic changes (variations with elevation) in precipitation and temperature across the basin. Since there are significant elevation differences between the location of the climate stations and the subbasins they represent (sometimes over 200 m), orographic changes in temperature are modeled in this application. Orographic changes in precipitation are not modeled. Modeling orographic temperature changes requires SWAT inputs for weather station elevations, latitude and the definition of elevation bands for each subbasin. Ten elevation bands were created in each subbasin by subdividing the subbasin elevation range into ten equally spaced elevation bands. The elevation change within each elevation band was $1/10^{th}$ the entire elevation range of the subbasin. The subbasin elevation ranges and fraction of subbasin area within each elevation band was determined from the AVSWAT elevation report.

The simulation of orographic changes in minimum and maximum daily temperatures involves adjusting subbasin temperatures based on the temperature lapse rate and the difference in elevation between the subbasin elevation bands and the elevation of the recording temperature station. The SWAT default temperature (or environmental) lapse rate of –6 °C/km is used in this application. The only processes in SWAT2000 modeled separately for each individual elevation band are the accumulation, sublimation and melting of snow.

The total Walton climate record is made up of Walton (Pre-1997) and Walton2 (after 1996) climate stations. Climate station locations are essentially the same spatially but the elevations of Walton and Walton2 are 378 m and 451 m, respectively. Thus, the average elevation of 414.5 m was input to the model as the elevation of this station. The elevation of the Stamford precipitation station is 558 m while the elevation of the Delhi climate station (measuring both precipitation and temperature) is 438 m. Modeling orographic temperature change is important in this application because 89% of the Cannonsville Basin has an elevation greater than the Delhi climate station. In fact, the median elevation in the basin is 580 m. This means that, based on the fixed SWAT default of –6 °C/km for the temperature lapse rate, 50% of the basin has a temperature that is at least 0.9 °C colder than what was measured at Delhi.

4.4.1 Adjustments and Corrections to Raw Temperature and Precipitation Data

The precipitation and temperature inputs for the model derived from the original data sources were found to require multiple adjustments and corrections to be used reliably in this application. In all, there were four different categories of modifications made to the original data. The first modifications involved filling in measured data for days when temperature or precipitation data were missing from both data sources. There were relatively few days in the calibration and validation period where missing data were encountered. Climate data recorded from nearby climate stations were utilized to fill in the missing data. When missing temperature data were filled in, the data were adjusted for the elevation differences between the climate stations based on a temperature lapse rate of –6 °C. After filling in missing data, the temperature and precipitation input data time series required three additional types of modifications.

4.4.1.1 Walton Precipitation Data 'Observer Shifting' Correction

Precipitation data from the Walton2 precipitation station was observed to be one day out of phase with both the Delhi and Deposit precipitation station observations for 84% of the period of record between January 1, 1998 and Sept 30, 2000. Walton2 is found nearly equidistant between the Deposit and Delhi stations and therefore should be reporting most rainfall events the same day as both of these stations. This problem was confirmed with the climatologist in charge of archiving the NRCC data (Personal Communication, Dr. Art DeGaetano) and is caused by 'Observer Shifting' of the precipitation data. In other words, the precipitation was recorded by the station observer based on his/her judgement as to what calendar day the precipitation fell on instead of based on NRCC standards that are followed at the other climate stations. The Walton2 precipitation data were therefore corrected or 'unshifted' such that the Walton2 data precipitation data were in phase with the Delhi and Deposit data. In addition, a few precipitation events at the original Walton station were identified to have this problem and were also corrected in a similar fashion. This analysis and correction procedure is outlined in the Appendix (Section 10.2.1).

4.4.1.2 Temperature Adjustments for Correct Precipitation Type Classification

The SWAT model classifies the precipitation depth input (in water equivalent) for each subbasin as either snow or rain based on the SWAT snowfall temperature parameter SFTMP and the average temperature for the day. When temperatures are above SFTMP, the precipitation is simulated by SWAT to be rain; otherwise it is simulated as snow. This model approach is reasonable with limited climate data. However, for much of the period of record, Delhi and Walton have measurements of the depth of snow that fell each day in addition to the water equivalent precipitation depth. This meant that for a large number of days, it was relatively clear as to whether the precipitation fell mainly as snow or rain. Furthermore, it was also observed that for a significant fraction of these days when the type of precipitation was known with relative certainty, the SWAT model would misclassify the precipitation type for any calibrated value of the SFTMP parameter. Therefore, it was deemed necessary to find a method for forcing the model to correctly assign the precipitation type on days when the main type of precipitation was known from the available climate data.

The methodology applied in this study to force the correct classification of the precipitation type in SWAT is outlined in detail in the Appendix (Section 10.2.2) and only briefly outlined below. Rather than modify the model source code it was determined that the SFTMP parameter could be fixed at 1° C (SWAT default value) and then small adjustments could be made to the temperatures input to the model in order to force the model to correctly classify the precipitation type at Walton and Delhi. Temperature adjustments were bounded so that they were relatively minor. In addition, adjustments were only made on days at Walton and/or Delhi when there was at least 2.5 mm of water equivalent precipitation recorded and when the bulk of the precipitation for the day could be classified with confidence as either rain or snow. This approach resulted in minor temperature adjustments on 67 days at the Walton stations and 69 days at the Delhi station over the calibration and validation period (January 1, 1990 to September 30,

2000). These minor temperature adjustments from the measured data were judged to have a smaller impact on model predictions than the alternative option that was to allow the model to misclassify the precipitation type.

4.4.1.3 Climate Data Adjustments for Large Precipitation Events

During the calibration process, there were a number of storm events for which daily simulated flows were in serious disagreement with available measured flows. For these events, the temperature and precipitation inputs were closely scrutinized. The event inputs were closely scrutinized against three data sources to make sure precipitation and climate inputs were representative. These three data sources were:

- Nearby climate stations for relative consistency.
- Descriptions of large storms events from the NCDC Storm Event database for New York (http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwevent~storms).
- Hourly precipitation data from a nearby climate station (Sydney, New York).

Results of the comparison showed that the temperature or precipitation inputs for six large precipitation events at either Walton, Delhi and/or Stamford warranted some adjustment to better approximate the information in one or more of the three data sources above. These modifications were generally required because this application of SWAT uses a daily time step and these events actually occurred within a period of time significantly shorter than 24 hours or within a 24-hr period that was split between two days in the climate records. For example, a large rainfall event at Sydney, New York was observed to most of the precipitation over a 14 hr period, however, the same event was recorded at Stamford with the total storm precipitation split almost evenly across two days. In this case, in was deemed necessary to change the Stamford precipitation so that most of the precipitation fell on one day instead of two. The original and modified climate input data, as well as a justification for the input data change, for each of these adjusted events is given in the Appendix (Section 10.2.3).

4.4.1.4 Summary of Climate Data Adjustments and Corrections

The development of the above methodology for adjusting and correcting the climate data was initiated due to observed model errors encountered during the model calibration process. However, the methodology and assumptions behind all of the above changes to the climate data, with the exception of those outlined in Section 4.4.1.2, were determined using only the available climate data as opposed to iteratively refining the methodology by referring to model predictive performance for flows across the basin. In other words, problems in the climate data were first identified and then corrected for in a systematic fashion, and calibration of the model continued by varying the model parameters in Section 6.3.3. The approach to correctly classify the precipitation type was only slightly influenced by model performance before being fixed and was then followed by the model parameter calibration in Section 6.3.3. All of the above climate adjustments and corrections resulted in significantly improved model predictions of daily and monthly hydrology.

4.5 Point Sources of P

Point source inputs in SWAT can vary daily, monthly or annually or be constant and require average loading rates of water, suspended sediments and nutrients. Since most point source measured loading rates are available at the monthly time scale, point source inputs in the model were defined monthly. Therefore, point source data varied between the calibration and validation periods. Monthly data from NYSDEC and the State Pollutant Discharge Elimination System Discharge Monitoring Reports (DMRs), as provided in a computer spreadsheet file by Pat Bishop (Personal Communication), were utilized to define monthly loading rates for the point sources. Some of the point source data in this computer file has been previously described and used in Longabucco and Rafferty (Longabucco and Rafferty 1998). In addition, the recommendations for Cannonsville Basin modeling point source inputs of P as outlined in Table 6 of (NYSWRI 2002b) were also adopted. Note that the P point source paper NYSWRI (2002b) is available at the New York State WRI public website http://wri.eas.cornell.edu/projects/nycwshed/delaware/. DMRs for the basin, can be found on the EPA's Surf Your Watershed website for Delaware County http://cfpub.epa.gov/surf/huc.cfm?huc_code=02040101.

Walton, Delhi, Stamford and Hobart wastewater treatment plants (WWTPs), as well as Kraft cooling water discharge, are included in the model and were the only significant point sources in the basin. The point source locations are given in Figure 4.1.1. Total P loads at each point source are defined in SWAT as mineral P and organic P and were assumed to represent, respectively, measured total dissolved phosphorus (TDP) and particulate phosphorus (PP). Walton WWTP P was assumed to have a TDP fraction of 0.6 while the remaining WWTPs were assumed to have a TDP fraction of 0.92 (Longabucco and Rafferty 1998). Kraft P load was assumed to be entirely TDP.

NYSDEC and DMR data as provided by Pat Bishop (Personal Communication) were both available to specify Walton WWTP inputs. The DMR data are the facility's self monitoring data. NYSWRI (2002b) outlines all differences in these data sets. Since NYSDEC WWTP sampling was most frequent, the NYSDEC data for the WWTPs was used when available. For months where NYSDEC Walton data were missing, DMR data were used. For months where no monthly data were available from either source, the annual average loading rates in NYSWRI (2002b) were used to derive monthly loading rates. Finally, for months where no monthly or annual data are available, the average annual loading rate for the period containing the missing data (before or after the Walton P treatment upgrade) was used to replace the missing values.

Loading rates for the other WWTPs were defined with either NYCDEP or DMR data (NYSWRI 2002b) and missing data were handled in a similar fashion as Walton missing data. Annual loading rates from Kraft, as estimated by NYCDEP, and the DMRs for the facility were used to derive monthly loading rates instead of NYSDEC data based on the recommendations in NYSWRI (2002b).

Figure 4.5.1 shows the monthly time series of total P loadings used as inputs for the point sources. A straight line generally indicates that monthly loading rates were assumed

constant for that point source and were estimated from an annual loading rate. Historically, Walton WWTP was the largest point source of P. Currently, Walton and Delhi discharge roughly the same total P load to the WBDR. Also note that Kraft's cooling water discharge is now assumed to contain no added phosphorus, based on their discharge monitoring reports. There is some pass-through phosphorus from their intake water. Table 4.5.1 summarizes the data in Figure 4.5.1 by water year over the calibration and validation periods and shows that P loading from point sources in the late 1990s was less than half the P loading from these sources in the early 1990s.

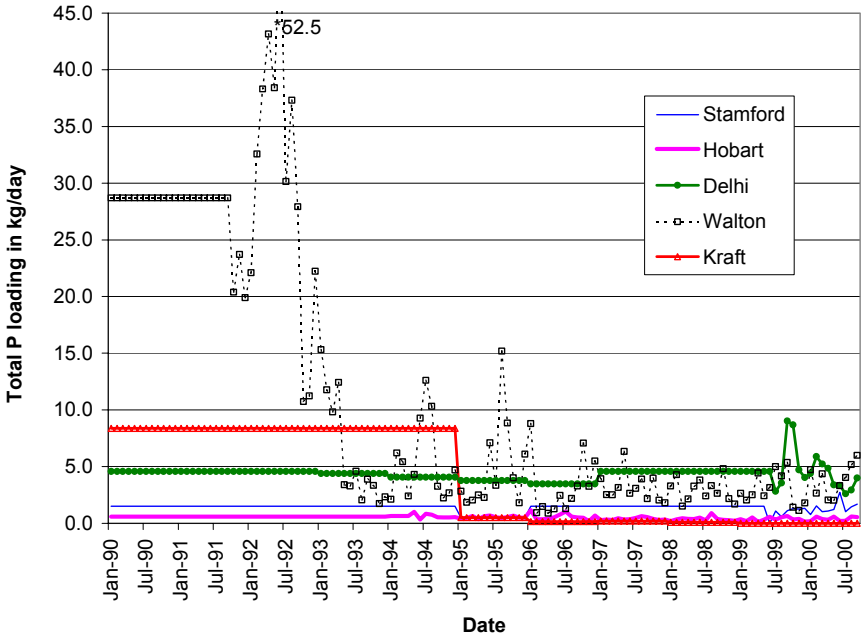

Figure 4.5.1. Time series of monthly point source total P loadings input to SWAT.

Table 4.5.1. Average annual water year total P loading (kg) from all modeled point sources within the Cannonsville Basin from 1990 through 2000.

Water Year	Stamford WWTP	Hobart WWTP	Delhi WWTP	Walton WWTP	Kraft Cooling Water Discharge	Point Source Total P
1990	550	215	1672	10478	3058	15973
1991	550	215	1672	10478	3058	15973
1992	550	215	1672	11753	3058	17249
1993	550	216	1621	3369	3058	8813
1994	550	237	1514	1927	3058	7286
1995	274	193	1403	1690	3058	6619
1996	458	225	1297	1049	184	3212
1997	550	154	1572	1404	55	3734
1998	550	147	1672	1050	75	3495
1999	451	134	1723	1230	33	3571
2000	499	138	1644	1178	2	3461

4.6 Groundwater Phosphorus

SWAT requires constant groundwater soluble P concentrations as inputs for each HRU. All groundwater that is predicted by SWAT to enter the stream is then assigned to carry this constant concentration of soluble P. NYCDEP bi-weekly water quality grab sampling data, as provided in a computer spreadsheet file (NYCDEP Unpublished data), in conjunction with USGS flow data were utilized to estimate spatially distributed groundwater soluble P concentrations. Groundwater soluble P concentrations are assumed constant within each subbasin.

Little groundwater P sampling data are available for the Cannonsville Basin. Therefore, groundwater soluble P concentrations had to be estimated from surface water quality TDP samples (or TDP estimated from another form of P sampled) that were taken across the basin at times when the flow was largely due to groundwater baseflow. Flow data at the Walton USGS gauge was analyzed with a baseflow separation program (Arnold and Allen 1999) to determine periods of low flow where the majority of flow is derived from groundwater. Here, flow was assumed to be essentially baseflow if results showed that the estimated baseflow percentage of total streamflow was greater than 80% for at least 6 days. P Water quality data that were sampled on dates falling within these baseflow periods were then further analyzed to derive groundwater soluble P concentrations.

In most of the NYCDEP water quality samples TDP concentrations are not directly measured. Therefore, the continuous monitoring data summary provided in Longabucco and Rafferty (1998) was used to derive conversion factors for estimating TDP from either total P or soluble reactive phosphorus (SRP) concentrations during baseflow conditions. The baseflow P data in Longabucco and Rafferty (1998) shows that the ratio of average SRP over average TDP concentration is 0.5 while the ratio of the average TDP over total P concentration is 0.8. These factors were used in the following methodology to estimate a TDP concentration from some form of NYCDEP sampled P:

- If an SRP concentration was sampled and greater than the detection limit, TDP was estimated as SRP/0.5.
- If SRP/0.5 was greater than or equal to the measured total P concentration, then TDP was instead estimated as 0.8*total P.
- If no SRP concentration was measured, TDP was estimated as 0.8*total P.

This approach generated estimated ‘sample’ TDP concentrations even when TDP was not measured or when SRP was below the detection limit of 3 µg/L. When NYCDEP samples were available during baseflow dominated flow conditions (and therefore utilized in this analysis), all TDP or total P samples were above their respective detection limits.

Although TDP concentrations could be estimated from NYCDEP water quality data at over 20 locations across the basin not all of the water quality data could be utilized in this analysis. For some of the NYCDEP water quality sampling locations, the sampled TDP concentrations approximately represented the groundwater soluble P concentrations. Such sampling locations were not downstream of any significant point sources of P and

are referred to as Group 1 subbasins. Conversely, some of the NYCDEP sampling locations were not suitable for this analysis because their samples either did not represent a specific subbasin or group of subbasins or the measured P concentrations could not reasonably be separated from upstream point source loadings. However, at some of the mainstem WBDR NYCDEP water quality stations, additional flow analyses and resultant mass loading estimates of TDP were used to derive groundwater soluble P estimates for additional subbasins. These subbasins are referred to as Group 2 subbasins and are subdivided into smaller groups based on the different sampling locations used to estimate their groundwater soluble P concentration.

Figure 4.6.1 shows the locations of the water quality stations that were utilized for this analysis and presents the subbasins as groups one through four. The subbasins are grouped based on the stations and/or methodology for estimating groundwater soluble P concentrations. A single iteration through the steps above generates one calculated TDP concentration, based on P sampling data, for a group 2 subgroup. A groundwater soluble P concentration is assigned to the subbasins in that subgroup equal to the arithmetic average of the subgroups calculated TDP concentrations. The three arithmetic averages of the group 2 subgroups are based on 2-6 calculated TDP concentrations.

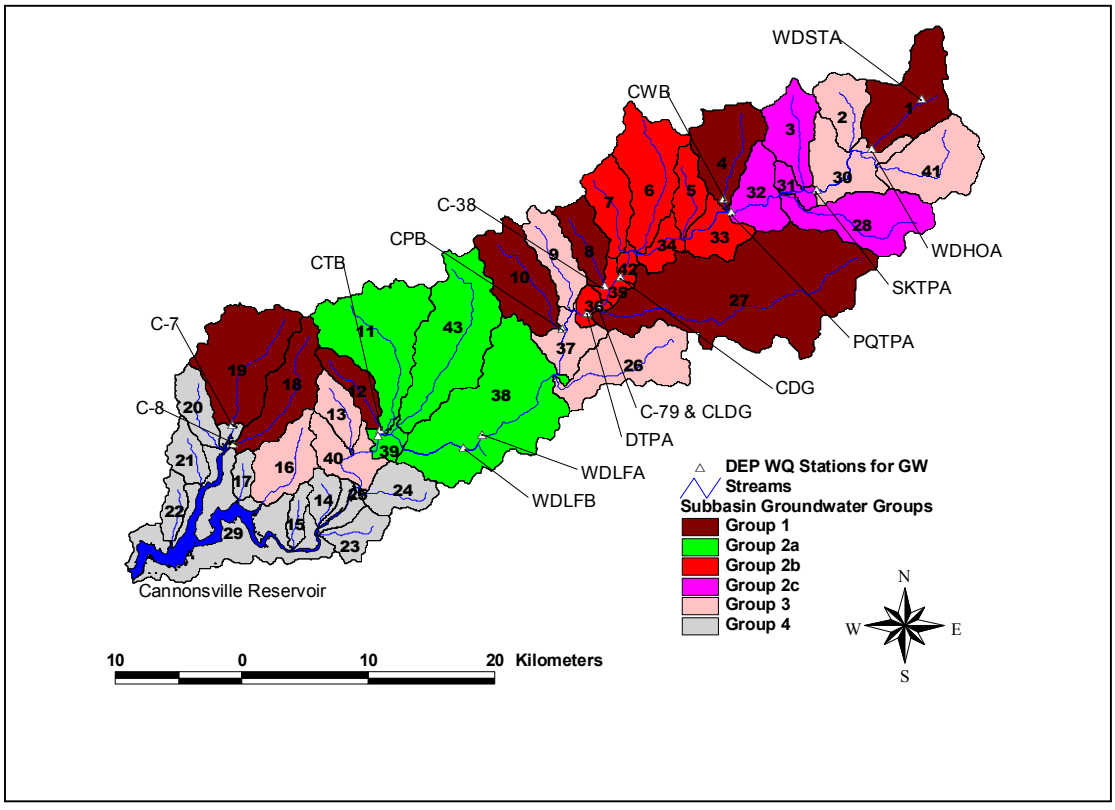

Figure 4.6.1. NYCDEP water quality (WQ) stations and corresponding subbasin groupings for estimating groundwater dissolved phosphorus concentrations (see text for definition of groups)

Table 4.6.1 outlines the NYCDEP water quality stations, as well as their associated subbasins, and the groundwater soluble P concentration value derived from this analysis. The methodology for estimating groundwater soluble P concentrations must be explained independently for each subbasin group.

Table 4.6.1. Groundwater soluble P concentrations assigned to subbasins in both the calibration and validation period.

Subbasin Group Name[1]	**Subbasin Number**	**Corresponding NYCDEP water quality stations (Figure 4.6.1) P concentration derived from**	**Soluble groundwater P Conc. (µg/L)**
Group 1	1	WDSTA[2]	21.6
Group 1	4	CWB	24.5
Group 1	8	C-38	17.2
Group 1	10	CPB	13.7
Group 1	12	CTB	13.6
Group 1	18	C-8	12.4
Group 1	19	C-7	9.9
Group 1	27	C-79 & CLDG[3]	12.4
Group 2a	11, 38, 39, 43	WDLFA, WDLFB, WSPA, CTB	12.7
Group 2b	5-7, 33-36, 42	DTPA, PQTPA, CDG, CWB, C-38, C-79 & CLDG	7.9
Group 2c	3, 28, 31, 32	SKTPA, PQTPA	7.7
Group 3	2, 9, 13, 16, 26, 30, 37, 40, 41	Area-weighted subbasin average	12.0
Group 4	14, 15, 17, 20-25, 29	-	3.0

1. Locations of subbasin groups are shown Figure 4.6.1.
2. It is assumed that WDSTA represents all of subbasin 1 groundwater soluble P levels even though it drains only about 1/3 of the subbasin.
3. Two water quality stations close to one another measured Little Delaware total P over the period of record.

GROUP 1 SUBBASINS. The arithmetic averages of sampled TDP from each NYCDEP water quality station monitoring a headwater subbasin that was free of interference from significant upstream point sources of P was assumed to directly represent the groundwater soluble P concentration in the respective subbasin. The number of baseflow TDP estimates used to generate the average TDP concentrations ranged from 11 to 17. For example, at the CWB water quality station, the average of the sampled TDP concentrations from 15 baseflow periods that with water quality samples was 24.5 µg/L.

GROUP 2 SUBBASINS. Estimates of groundwater soluble P concentrations from group 2 subbasins required a pairing of mainstem WBDR NYCDEP water quality stations. Furthermore, no significant point sources discharge to the WBDR between group 2

subbasins and any group 1 subbasins draining to the WBDR between a pair of group 2 subbasins must be accounted for. Group 2 is subdivided into groups 2a, 2b, and 2c based on the pair of water quality stations utilized.

The methodology for estimating the groundwater soluble P concentration for each group 2 subgroup is the same and requires the calculation of approximate TDP concentrations in each baseflow period as follows:

- Steady flow conditions are assumed and all water quality sampling locations are verified to have samples taken on the same day.
- A flow value is estimated for all water quality stations involved in the TDP concentration calculation based on a unit area flow extrapolation using the average of the daily flows over the period as measured at the nearest representative USGS gauge.
- For each water quality station, estimated flows and TDP concentrations are multiplied together to estimate an average TDP mass loading.
- Based on mass balance considerations, the TDP mass and volume of water originating between the paired water quality stations can be calculated provided that any group 1 subbasins emptying to the WBDR between the paired group 2 water quality stations are accounted for in the TDP and water mass balance.
- Provided the calculated mass load was greater than zero (in some cases it was not), a TDP concentration representative of the groundwater in the subbasins between the paired water quality stations (except for any intermediate group 1 subbasins) could be calculated based on the TDP mass and flow volume.

A single iteration through the steps above generates one calculated TDP concentration, based on P sampling data, for a group 2 subgroup. A groundwater soluble P concentration is assigned to the subbasins in that subgroup equal to the arithmetic average of the subgroups calculated TDP concentrations. The three arithmetic averages of the group 2 subgroups are based on 2-6 calculated TDP concentrations.

GROUP 3 and 4 SUBBASINS. Group 3 and 4 subbasins do not have suitable or available data that can be used as above for group 1 or 2 subbasins to estimate groundwater soluble P concentrations. Instead, the groundwater soluble P concentrations of groups 3 and 4 subbasins are assumed based on the groundwater soluble P concentrations determined for groups 1 and 2. Group 3 subbasins are those that have manure spread on the land surface while group 4 subbasins are those around the reservoir that do not have manure spread on the land. Since group 4 subbasins also have a relatively smaller concentration of septic systems than the rest of the subbasins, group 4 subbasins are assumed to have the lowest groundwater soluble P concentration in the basin at a value of 3.0 µg/L. Since group 3 subbasins are generally the same as groups 1 and 2 in terms of P inputs to the land, group 3 subbasins are assigned a groundwater soluble P concentration equal to the area-weighted average of the groups 1 and 2 groundwater soluble P concentrations (12 µg/L).

Overall, the area-weighted basin-wide average soluble P concentration in groundwater determined by this analysis was 10.9 µg/L. For the Beerston water quality station

drainage area, the area-weighted average soluble P concentration in groundwater was estimated as 12.0 µg/L. This value compares well with the soluble P concentration in groundwater (13 µg/L) assumed for the Beerston drainage area in the modeling work by Schneiderman et al. (1998).

The baseflow periods identified in this analysis occurred in both the model calibration and validation time periods. However, since this analysis yielded a relatively small number of samples representing groundwater soluble P, no attempt was made to estimate groundwater soluble P concentrations that differed between the calibration and validation period. Instead, the groundwater soluble P concentrations derived in this analysis were assumed to be equal for both the calibration and validation period.

4.7 Agricultural P Sources and Management Practices

Agricultural sources of P and agricultural management option inputs must be provided to SWAT in order to simulate reasonably representative agricultural practices. Agricultural management practices simulated in this SWAT model are manure application, tillage, starter corn fertilization and hay cutting. All agricultural management practices were selected to represent the average conditions or typical management behaviour of farmers in the basin.

Since manure is the largest potential source of P to the reservoir, much of the modeling work was focused on the realistic representation of manure production and distribution across the watershed. SWAT manure inputs in each subbasin were generated based on the five following general steps:

1. Estimate the cattle counts and convert to total live animal mass.
2. Identify basin manure characteristics such as production rate and manure solids and P content.
3. Based on 1 and 2, calculate the total manure solids load produced in the basin.
4. Allocate the manure solid load in 3 to be either directly deposited by cattle in actively grazed pastures or be collected for spreading.
5. Based on the behaviour of the typical farmer in Delaware County, spread the collected manure to the appropriate HRUs according to season, land use priority and a maximum application rate.

Each of these five steps is described in detail in the following sections. The mass balance approach described above ensures that only the available manure is distributed to the land surface as opposed to simply assuming a manure spreading rate associated with each agricultural land use regardless of the presence of a cattle population nearby required to produce the manure.

4.7.1 Manure Production Estimates

Spatially variable manure production rates were estimated based on a surveyed/estimated cattle population for the watershed and a number of manure production factors. Whenever possible, manure production factors were based on factors derived from

Cannonsville Basin agricultural data, otherwise standard values in the agricultural literature were used.

4.7.1.1 Cannonsville Cattle Population

A NRCS agricultural survey in Cannonsville Basin conducted from 1992 to 1994 (NRCS Unpublished data) was used as the basis for estimating subbasin specific cattle populations suitable for the model simulation period. The researcher in charge of conducting the survey estimated that the survey included at least 90% of the cattle farms in the basin (Personal Communication, Gary Lamont). Survey results summarized each farm's primary operation and provided ranges for the livestock population. Beef and dairy cattle were counted independently if a farm had both types of animal. Only beef and dairy arms were considered here since they make up the majority of farms in the basin (Delaware County Board of Supervisors 1999). The survey identified 226 total farms for beef or dairy (141 dairy and 85 beef) in the basin. Survey results were available spatially (see Figure 4.7.1) such that farm locations could be attributed to a particular subbasin. It was assumed that the cattle and resultant manure were confined to the subbasin that each farm was located in.

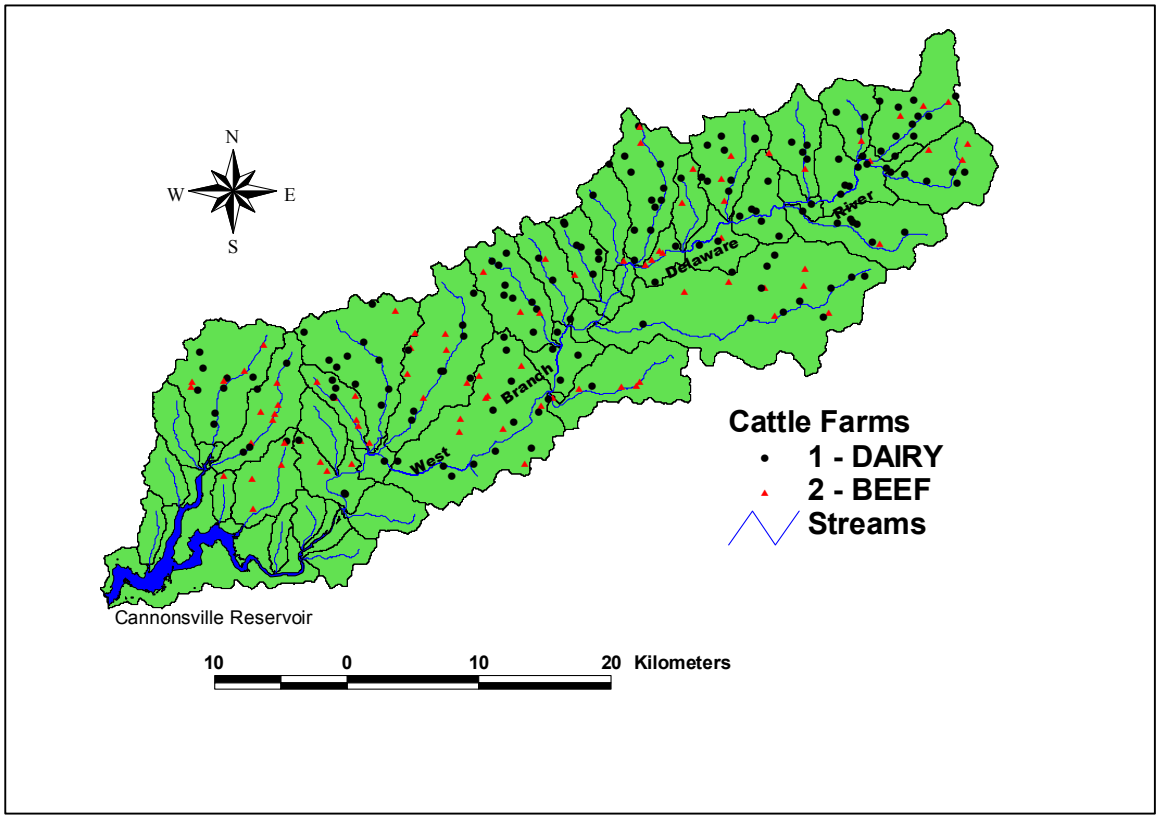

Figure 4.7.1. 1992 Farm locations in the Cannonsville Basin (NRCS Unpublished data)

The cattle populations on each farm were sub-characterized as cows, heifers or calves and reported as ranges instead of point estimates. The survey snapshot of these animal types

was assumed to be representative of the cattle herd at all times of the year such that the number of calves surveyed would be the effective number of calves present across an entire year. Generally, the midpoints of each animal type population range were used to represent the estimated number of animals. The upper range in the survey for dairy cows was open ended at greater than 100 animals per farm. A point estimate of animals in this category was assumed to be 110 animals per farm. This approach was reasonably representative of the survey results except for two outlier dairy farms that, although included in the survey, had a much higher population of dairy cattle (Personal Communication, Gary Lamont). Based on recommendations from Gary Lamont, dairy population inputs were adjusted to account for these outlier farms as described below.

The true size of the combined dairy herd on these two outlier farms was known to be about 1600 mature cows and 800 heifers and calves during the validation period (Personal Communication, Gary Lamont). Accurate representation of these two outliers is important considering that these two farms account for roughly 10% of the total dairy population in the basin in 1992. One of these two farms ceased operations in 1998 or 1999, which is about ¾ of the calibration period. To represent this change while keeping the dairy population constant during the calibration period, the combined dairy herd size at these two farms was assumed to be 1200 mature cows and 500 heifers and calves during the calibration period. However, due to privacy concerns, the exact location of these farms could not be identified. As a result, survey results were analyzed to determine all possible farm locations of these two outliers. The eight largest dairy farms were identified from the survey to have more than 100 mature cows and more than 50 heifers and calves. It was assumed that the two outlier farms were two of these eight farms. Accordingly, the estimated calibration and validation period dairy populations from these two outliers were distributed evenly across these eight farms. These eight farms were located in subbasins 1, 6, 11, 30, 34 and 43. As a result, the dairy population will be greatly under-estimated in one or two of these six subbasins and moderately over-estimated in four or five of these six subbasins.

Cattle populations are often summarized in terms of the number of animal units (AU) where 1 AU is 1000 lbs of live animal weight. Summarizing the population in AUs was necessary to compare population estimates with other sources and estimate manure production rates. Representative weights assumed for each type of animal are given in Table 4.7.1. Based on this, it was estimated that the total number of dairy cattle in the Cannonsville Basin was 20,100 AU for 1992. Based on a count of the dairy animals, the estimated 1992 population was 10,400 cows and 9100 heifers and calves. The initial resulting beef cattle AU for the basin was deemed much too small relative to the dairy numbers based on the 11:1 dairy to beef ratio estimated in DCAP (Delaware County Board of Supervisors 1999). Therefore, numbers somewhat higher than the midpoints of the beef population ranges in the NRCS Cannonsville Basin survey were used to derive beef population estimates per subbasin. This adjustment yielded a total of 1000 beef AU for 1992. Further upwards adjustments to the beef survey results were not made since they were deemed to misrepresent the survey results.

Table 4.7.1. Average assumed beef and dairy cattle animal weights.

Animal Type	Animal Mass (lbs)[1]
Dairy Cow (milking)	1400
Dairy Cow (Non-milking)	1300
Dairy Heifer & Calf	700
Beef Cow	1000
Beef Heifer & Calf	600

1. Source of estimates: Paul Cerosaletti (Personal Communication).

The simulation period for calibration and validation is from 1990 to 2000. Therefore, a representative cattle population should be used to represent each year in this period. The estimated 1992 beef and dairy cattle populations derived in the previous paragraphs are assumed to adequately approximate the true cattle population over the model validation period (i.e. constant cattle population from Jan. 1990 - Dec. 1993). Comparison of the 1992 and 1997 US Agriculture Survey numbers for Delaware County (USDA 1994, 1999) showed that the cattle population in the County, and therefore the Cannonsville Basin, were not constant between these periods. For example, the number of dairy farms and dairy cows both decreased by 21% from 1992 to 1997. Conversely, the number of beef farms and beef cows both increased by 33% over the same period. Therefore, the estimated 1992 beef and dairy populations in the Cannonsville Basin are adjusted to 1997 based on these observed changes in the US Agriculture Survey Delaware County cattle data from 1992 to 1997. This adjustment was applied equally across all 1992 estimated subbasin beef and cattle populations.

The cattle population estimated for 1997 was a total of 15,300 dairy AU and 1300 beef AU. Based on a count of the dairy animals, the estimated 1997 population was 7900 cows and 7000 heifers and calves. The estimated 1997 cattle population estimate is assumed representative during the entire model calibration period (i.e. constant cattle population from Jan. 1994 to Sept. 2000). The 2002 basin-wide dairy cow population was estimated by an independent source to be between 7000 and 8000 animals (Cerosaletti 2002). This encompasses the 1997 estimate of 7900 dairy cows for the basin. These dairy cow population estimates are in good agreement and suggest the dairy cow population has remained relatively static or has decreased only slightly in the basin from 1997 to 2002. Therefore, the assumption of a constant dairy population for 1997 to 2000 is demonstrated to be reasonable.

4.7.1.2 Manure Production Factors

Manure production rates and characteristics were needed to derive manure loading to the subbasins from the subbasin cattle populations. These rates and factors were estimated from Cannonsville Basin data whenever possible. Manure production rates and factors for beef and cattle were estimated independently. In some cases, production factors also vary between grazing (May 1 to Oct 31) and non-grazing season. The various rates, factors and sources for each are presented in Table 4.7.2 and are applied to both the calibration and validation period.

Data taken from ASAE (1998) represent the averages of a number of surveys nationwide. However, basin specific data for some manure characteristics are available from a 1998 survey of manure at six dairy farms in the Cannonsville Basin (Dewing 1998). Samples of manure at each farm were taken in the summer (grazing) and winter (non-grazing) to assess characteristics in each season. The study provided seasonal values for manure solids content, total P content and total N content. These were averaged across the six study farms and assumed to represent the manure characteristics of all dairy farms in the basin. Basin specific data were used whenever available. This approach yielded an estimated annual manure P production rate for the basin of 224,000 kg from dairy cattle and 20,000 kg from beef cattle during the calibration period (1994-2000). During the validation period (1990-1993), the annual manure P production rate for the basin was estimated as 295,000 kg from dairy cattle and 15,000 kg from beef cattle.

Table 4.7.2. Dairy and beef manure production rates, characteristics and sources.

Quantity	**Units**	**Dairy Values for Summer (grazing)**	**Dairy Values for Winter (non-grazing)**	**Beef Values (all year)**
Manure Production Rate	kg of total manure produced per day per 1000 kg of live animal mass	90[1&2]	90[1&2]	58[1]
Manure Solids	Fraction of total manure	0.141[3]	0.114[3]	0.147[1]
Manure P Content	Fraction of solid manure that is P	0.0070[3] [0.0080][4]	0.0086[3] [0.0080][4]	0.0108[1]
Mineral P content	Fraction of total P that is mineral P	0.75[5] [0.625][4]	0.75[5] [0.625][4]	0.330[1]
Organic P content	Fraction of total P that is organic P	0.25[5] [0.375][4]	0.25[5] [0.375][4]	0.670[1]
Mineral N content	Fraction of solid manure that is mineral N	0.013[3] [0.007][4]	0.019[3] [0.007][4]	0.010[1]
Organic N content	Fraction of solid manure that is organic N	0.017[3] [0.031][4]	0.025[3] [0.031][4]	0.030[1]
Cows Milking	Fraction of dairy cow herd being milked	0.85[6]	0.85[6]	-

1. Based on national average rates for beef and dairy (ASAE 1998) as provided in the SWAT manual (Neitsch et al. 2001b).
2. Includes manure from animals at 86 kg/day (ASAE 1998)+ 4 kg/day to represent animal bedding that is also collected and spread with the manure (Personal Communication, Tom Tylutki).
3. From Cannonsville Basin data (Dewing 1998).
4. Default SWAT2000 values for these inputs.
5. Based on Cannonsville Basin data (Kleinman 1999).
6. Estimated for Cannonsville Basin (Personal Communication, Paul Cerosaletti).

The resulting beef and dairy manure production rates estimated for each subbasin are summed and managed together. This approach allows realistic manure production rates to be estimated even though the different management styles of beef and dairy farms are not simulated. The concentration of mineral and organic P in the manure should be a function of the relative amounts of dairy and beef manure. Such an approach would require defining in the SWAT model a different manure type for each subbasin in each season (close to 100 manure types). In order to simplify this cumbersome approach, only 3 types of cattle manure are defined in the model: beef, winter dairy and summer dairy. If a subbasin had >50% dairy manure, then dairy manure (winter or summer depending on the season) was spread in the model, otherwise beef manure was spread. Although this simplified approach created a minor error in the amount of total P applied to the basin in the model, the error was acceptably small because most subbasin manure loads are >90% dairy. During the model calibration period, the manure P distributed to the HRUs in the SWAT model is 242,000 kg (1.1% less than the projected annual P load from cattle manure calculated in the previous paragraph). During the model validation period, the manure P distributed to the HRUs in the SWAT model is 307,000 kg (1.0% less than the projected annual P load from cattle manure calculated in the previous paragraph).

Since the HRU concept combines similar crops from both beef and dairy farms, simulating beef and dairy farms differently is not currently possible in the model and also not warranted given the minor contribution of beef to the total manure load in the basin. Spreading only beef manure on only beef pastures would require excessive initial GIS processing outside of AVSWAT and would increase the number of HRUs by approximately 50%.

4.7.1.3 Manure Application to HRUs

SWAT model inputs for manure application are dry manure application rates per hectare and mineral and organic P fractions in the dry manure. Each of these can be determined from data in Table 4.7.2, the areas of the HRU types in each subbasin receiving manure and the allocation of the manure between agricultural land uses. Total manure loads in each subbasin are allocated to the subbasin agricultural lands in a way that attempts to replicate the behaviour of the typical farmer from 1990 to 2000. Under the guidance of Paul Cerosaletti and Dale Dewing (Personal Communication), agricultural practices in the basin representative of practices prior to the implementation of whole farm plans were identified. Based on these practices, as well as SWAT model considerations, the following assumptions were made in order to generally characterize the application of manure to agricultural HRUs in the SWAT model:

- Manure is not stored and spreading occurs 365 days a year. This was considered to be reasonable since an upper estimate of farms with manure storage currently is around 25% and many of those would have had storage facilities constructed after the year 2000.
- Weather does not influence the timing of manure spreading. Although this is not necessarily true in agricultural practice, preprocessing the weather data and determining a representative management response to all weather conditions was deemed too time consuming.

- All manure produced was either directly deposited by cattle in pasture HRUs or collected and available for spreading to any agricultural HRU.
- The grazing season for beef and dairy cattle are assumed the same because beef and dairy manure are managed together in the model. The grazing season defined in this application is from May 1 to Oct 31.
- During grazing, beef and dairy cattle were assumed to directly deposit 60% and 45%, respectively, of the daily manure produced in pasture HRUs. This difference reflects the fact that beef cattle spend more time in pasture than dairy cattle. The remaining manure was collected and subject to further management.
- All manure is surface applied.

Furthermore, the timing and distribution of *collected* manure to agricultural HRUs must be defined. Since the manure availability differs between subbasins, a general set of rules was developed with the aid of Dale Dewing and Paul Cerosaletti (Personal Communication) that attempts to replicate the typical behaviour of farmers. These rules are evaluated in order to distribute or spread the total collected manure loads over agricultural land in each subbasin as follows:

1. Annual manure season begins October 1 and each year is divided into four manure management periods. In each manure management period:
    I. First priority is given to allocate manure to corn silage HRUs (at a constant daily spreading rate) until a maximum of 7000 kg/ha of dry manure (24.5 English tons/acre of wet manure) is spread for the year.
    II. Second priority is given to allocate manure to hay HRUs (at a constant daily spreading rate) until a maximum of 10,000 kg/ha of dry manure (35 English tons/acre of wet manure) is spread for the year.
    III. Then, the remaining manure, if any, is applied to pasture HRUs or idle agriculture HRUs if pasture HRUs are not available for spreading.
2. These priorities are followed except under the following conditions:
    - When a specific land cover is not available in a subbasin.
    - Corn is never manured from May to September (corn growing season).
    - Collected manure is not distributed to pasture when the cows are grazing.

These rules depend on the growing season for corn (see Section 4.7.3) and grazing season for cattle. Since these seasons do not match exactly, there were three periods per year with distinct manure production and application rates. Since SWAT schedules management practices on an annual calendar, the winter period was split into two periods making a total of four manure management periods per year in which SWAT inputs must be defined. For each management period, subbasin and HRU, a daily manure application rate, along with the appropriate manure type, was input to SWAT.

A compact summary of the resulting manure application pattern during the calibration period is provided in Table 4.7.3 and Figure 4.7.2. Table 4.7.3 shows that corn receives manure at the highest average rate and that the application rate varies little between subbasins (because the maximum and average application rates are very close). Table 4.7.3 also shows that although hay and pasture manure application rates are lower compared to corn, the application rate varies more between subbasins. In the calibration

period, hay areas are extensive enough in each subbasin to receive all remaining collected manure after distribution to corn HRUs. Therefore, collected manure is not applied to pasture or idle agriculture HRUs in the calibration period. The same holds true during the validation period. Figure 4.7.2 shows that the bulk of total manure goes to hay, then corn and then pasture HRUs. It also shows that hay is mainly fertilized from January to September while 2/3 of the basin-wide manure applied to corn is applied from October through December. Manure spreading is similar for the validation period except that manure application rates are elevated since a higher manure production rate was estimated.

Table 4.7.3. Summary of model annual application rates across land use types for the model calibration period in English tons of wet manure/acre[1] and (kg dry manure/ha).

Land Use	Area-Weighted Average Manure Application Rate	Maximum Manure Application Rate
Pasture	6.4 (1824)	22.8 (6516)
Idle Agriculture	0.0	0.0
Corn	23.6 (6745)	24.5 (7000)
Hay	3.5 (1006)	15.3 (4358)

1. These units are used since they are common in the field. All necessary conversions are made before data input to SWAT.

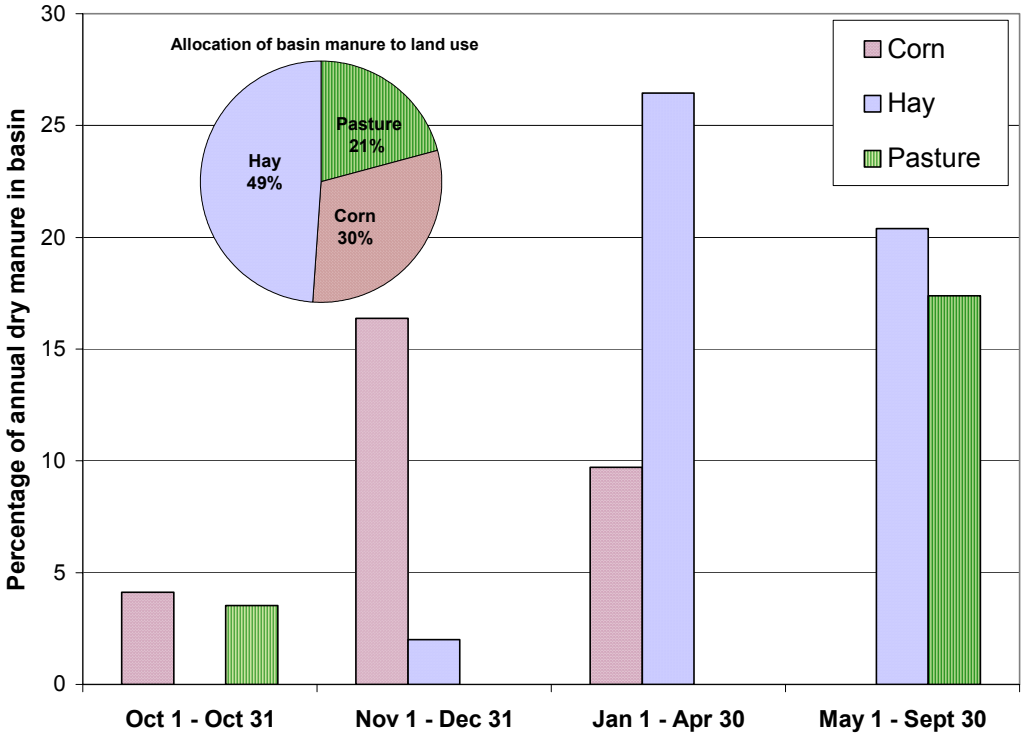

Figure 4.7.2. Pattern of basin-wide total manure application in SWAT for the calibration period.

4.7.2 Starter Corn Fertilization

Farmers in the basin generally apply starter corn fertilization (Personal Communication, Karl Czymmek and Dale Dewing). To simulate this in SWAT, it was assumed that starter fertilizer was applied each year on the same day that corn is planted. Based on recommendations from Karl Czymmek and Dale Dewing (Personal Communication), 39.3 kg/ha of Nitrogen (N) and 17.1 kg/ha of P is applied to corn HRUs and 99% of N and P is incorporated to the soil layer under the top 10 mm of soil. These corn fertilization inputs were assumed constant between the calibration and validation periods. The application of starter corn fertilizer adds a total of 23,700 kg of P or approximately 25% of the total P added to corn silage HRUs from manure and starter corn fertilizer.

4.7.3 Tillage, Crop Growth and Harvest Scheduling

According to local agricultural planners some farm management inputs vary depending on whether the farm is located in a lowland valley or on upland slopes. Farms in the basin are therefore classified as either lowland or upland farms. In order to represent the differences in management practices between these farm types, the HRUs in the SWAT model must be classified as either upland or lowland. It was assumed that the HRUs on the NY099, NY127 and NY136 STATSGO soil types were lowland since they were located in the valleys while the remaining HRUs were upland. The lowland and upland areas in the model are shown in Figure 4.7.3. Although this representation of upland and lowland areas is imperfect, it is deemed appropriate for this scale of watershed model.

The available spatial data provide the precise spatial location of corn land use in the basin. Therefore, all corn silage HRUs can be classified as either upland or lowland with relative certainty. In contrast, the precise spatial location of pasture and hay land use in the basin is not known. The definition of hay HRUs in Section 4.2.1.2 allows for easy incorporation of crop rotations into the SWAT model in the future. This approach created hay HRUs that should be in rotation as well as hay HRUs that should be continuous hay in future model versions. As explained in Section 4.2.1.2, all hay HRUs in the current version of the model are assumed to be continuous hay. The definition of these hay HRUs that could be rotated with corn was based on the corn and hay having like STATSGO soil types. Therefore, these 'rotation' hay HRUs are also classified as either upland or lowland HRUs. The approach outlined in Section 4.2.1.2 for disaggregating the remaining grass land use category (after the definition of 'rotation' hay HRUs) into pasture, continuous hay and idle agriculture HRUs arbitrarily assigns these HRUs to be located on either upland or lowland soils. Therefore, it was deemed more representative of this imprecise knowledge to characterize these pasture and hay HRUs as lumped upland and lowland HRUs with management inputs that represent a compromise or average of the inputs from upland and lowland HRUs. Hay HRUs were therefore categorized as upland, lowland or lumped upland/lowland while pasture HRUs were all categorized as lumped upland/lowland.

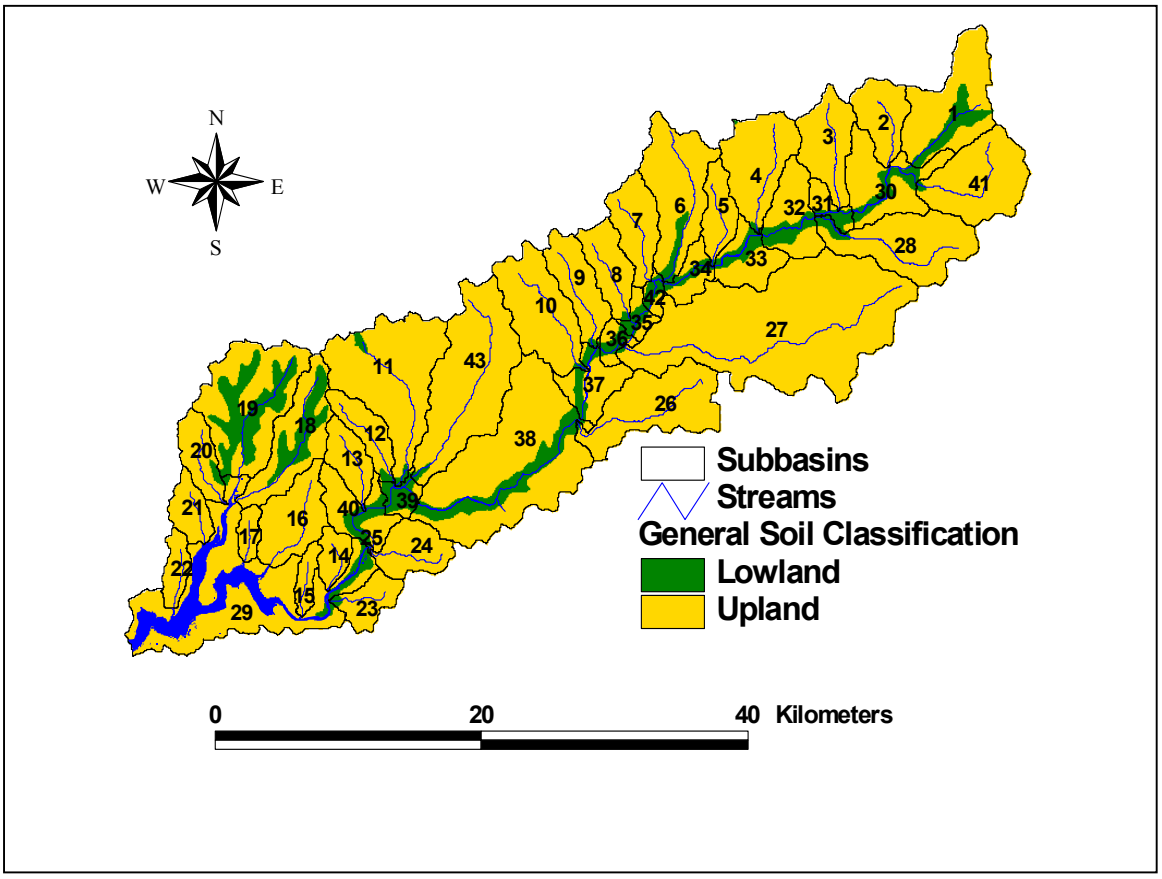

Figure 4.7.3. Definition of upland and lowland areas in Cannonsville Basin.

Based on information from various agricultural experts familiar with local management practices (Dale Dewing and Karl Czymmek, Personal Communication) management inputs for each agricultural HRU type were assigned and are outlined in Table 4.7.4. The shorter growing season on upland HRUs is reflected in the later planting and tillage dates for upland corn silage HRUs and in the reduced rates of hay harvest for upland hay HRUs. Since all hay in this model is grown continuously, there is no tillage on hay HRUs. The harvest index values for hay and pasture were selected such that the total basin-wide harvested biomass represented a reasonable proportion of the estimated dry matter intake rate of the basin-wide cattle population (see Section 10.4.2 for more discussion on this issue).

Table 4.7.4. Summary of non-fertilization agricultural management inputs for SWAT model during calibration and validation.

Modeled Land Use	SWAT Crop/Plant Cover Modeled	Upland or lowland HRU	Planting Date	Harvest Dates (Harvest Index[1])	Tillage Dates
Corn	Corn Silage	Lowland	May 5	Sept. 25	Apr. 23[2] May 3[2]
Corn	Corn Silage	Upland	May 20	Sept. 25	May 8[2] May 18[2]
Hay	Timothy	Lowland	-[3]	May 20 (0.5) July 1 (0.4) Aug. 15 (0.4)	-
Hay	Timothy	Upland	-[3]	June 10 (0.5) July 30 (0.4)	-
Hay	Timothy	Lowland / Upland	-[3]	May 20 (0.3) July 1 (0.5) Aug. 15 (0.2)	-
Pasture	Timothy[4]	Lowland / Upland	-[3]	May 20 (0.45)[5] July 1 (0.55)[5] Aug. 15 (0.45)[5] Sept. 30 (0.35)[5]	-

1. The harvest index or HIOVR parameter in SWAT is used to specify the fraction of above ground biomass removed in each hay cutting operation.
2. The corn silage tillage operations for both upland and lowland farms, in order of occurrence, are tillage with a moldboard plow to 150mm in depth and then tillage with a tandem disk to 75mm in depth.
3. Timothy is a perennial plant and starts growth once spring temperatures warm enough.
4. Some timothy crop parameters for pasture were set differently than timothy crop parameters for hay (see Section 6.3.2.1 for details).
5. Pasture grazing management inputs (e.g. biomass removal) represented in this SWAT model by harvesting pasture.

4.8 Atmospheric N and P Deposition

Available data show that atmospheric N and P deposition rates to basin soils are significant. Atmospheric N and P deposition rates for the various locations are monitored as part of the National Atmospheric Deposition Program (NADP) and are available through http://nadp.sws.uiuc.edu/. The monitoring station closest to the Cannonsville Basin is site NY68 in Ulster County, NY.

Atmospheric N deposition is input to SWAT as the RCN parameter, which is the constant concentration of nitrate (NO_3) in rainfall falling across the basin. NADP data are available for nitrate deposition rates as kg/ha and nitrate precipitation-weighted

concentrations in rainfall as mg of nitrate per litre. NADP data are also available for ammonium (NH_4) deposition rates as kg/ha. The average annual deposition rates for nitrate and ammonium at NY68 from 1983-2002 are approximately 18 kg/ha and 2.5 kg/ha, respectively. Converting these to N deposition rates yields average annual deposition of 4.1 kg/ha of nitrate-N and 1.6 kg/ha of ammonium-N. The average annual precipitation-weighted concentration of nitrate at NY068 from 1983-2002 is approximately 1.5 mg/L. Since ammonium-N accounts for about 1/3 of the atmospheric N deposition measured at this station, it was deemed important to account for this N addition to basin soils. Therefore, the nitrate N concentration was adjusted by 1.5/0.67=2.25 mg/L and input to SWAT as the value of the RCN parameter. This adjustment forced the SWAT input of atmospheric N to represent the total measured atmospheric N deposition at NY68 at the expense of ignoring the fact that about 33% of the N is from ammonium rather than nitrate.

NADP atmospheric P data were much less abundant for NY68. In fact, P data are not available directly on the NADP website and are only available by contacting the NADP program office through their website. The available P data are deposition rates in kg/ha of PO_4. Unfortunately, 94% of the weekly samples of atmospheric P O_4 deposition rates were below the detection limit. The NADP estimated the average annual PO_4 deposition rates by substituting ½ the detection limit for samples with PO_4 under the detection limit. The resulting annual average deposition rate of P from PO_4 is calculated as 0.024 kg P/ha based on estimated annual averages from 1990-2001. Based on the high proportion of non-detects in this data set, another source of atmospheric P deposition rates is needed.

Independent of the NADP atmospheric P deposition data at NY68, a P budget study of Northeast US forests by Yanai (1992) provides a second estimate for atmospheric P deposition. Yanai (1992) estimates that atmospheric P is deposited at an average annual rate of 0.04 kg P/ha. This is relatively close to the estimate using NADP data. To be conservative in terms of the simulated model inputs of P, it is assumed that atmospheric P is deposited on all basin soils at an annual rate of 0.04 kg/ha. Therefore, a total of approximately 4,700 kg of P/yr is added to the basin via atmospheric P deposition. This P deposition is added to model soils by specifying a fertilization operation on all basin soils with 0.04 kg of pure P fertilizer on December 31 each year. Since model predictions of P loading during calibration were not sensitive to this minor addition, this one-time addition of P to soils in the model is a reasonable way to approximate the year-round deposition of atmospheric P to basin soils.

4.9 Model Initial Conditions

Any SWAT model run should include a warm-up period in which the model is allowed to initialize and then approach reasonable starting values for model state variables. For example, the warm-up period allows the model to deposit sediment in the river network and fill the soil partially with soil water before simulation results are considered realistic. Model predictions should only be compared to measured data after the simulation warm-up period. After some initial tests, it was deemed necessary to have at least a two-year warm-up period in the model for reasonable initial channel sediment levels. Furthermore,

due to the uncertainty associated with exact initial phosphorous levels the warm-up period was extended to be 3 years. The longer the model warm-up period, the less impact highly uncertain initial model inputs have on simulation results.

At the beginning date of both the model calibration and validation simulations (e.g. the first day of the model warm-up period) all model state variables are left at their SWAT default initial values except for the state variables listed in Table 4.9.1 below and the P state variables discussed in Section 4.9.1. State variables set in Table 4.9.1 were set to initial values deemed more reasonable than the model default values. Calibration and validation simulations begin on January 1 with no crops growing, and plant growth on all land covers begins in the spring of the first simulated year. Model predictions are not evaluated in accordance with the 3-year warm-up period until 3 full years have been simulated.

Table 4.9.1. Initial state variable values specified differently than SWAT defaults.

SWAT Variable Name	Variable Description	Initial Variable Value	Model Default Value
RSD_IN	Initial mass of biomass residue in top 1 cm soil layer (kg/ha)	5000 for forests & 3000 for all other land uses	0 for all land covers
FFCB	Initial soil water storage expressed as a fraction of field capacity	0.5	0.0

4.9.1 Initial Soil Phosphorus Levels

Considerable effort was made to specify initial soil P levels that were representative of the spatially variable soil P conditions in the Cannonsville Basin. Unpublished soil test P data collected by the Watershed Agriculture Program (WAP) for the New York City Watersheds and analyzed by the Cornell Nutrient Analysis Laboratory, Department of Soil, Crop, and Atmospheric Sciences, Ithaca NY was utilized for the purposes of specifying initial soil P conditions for agricultural land covers. The database for the Cannonsville Basin was provided with all spatial references removed to protect privacy of farmers. The database contains over 2800 entries of soil test Morgan's P levels in lbs/acre and each entry has a corresponding date, crop cover and soil type associated with it. Forest sampling data (mg Morgan's P /kg) taken from two farm-scale research sites in the basin (NYSWRI 2002a) were used to specify initial forest and grass-shrub soil P levels. Only 14 forest sites in the NYSWRI (2002a) sampling effort were deemed representative of unmanaged forested land. Other land cover types in the model (urban land and water make up less than 3% of the basin) are initialized with model default SWAT soil P levels.

4.9.1.1 Labile Soil P Levels for Agricultural Lands

Since over 95% of the Delaware County soil test database entries represent soil conditions between 1995 and 1998, no values were available for direct specification of soil P levels at the beginning of either the calibration or validation period. Instead, it was assumed that the average soil test P levels in the database are representative of the initial soil P levels at the start of both the calibration (Jan. 1991) and validation period (Jan. 1987) simulations. Based on the crop cover and soil type, groups of soil test P values were formed and their averages were assumed representative of each crop cover/soil combination. Since SWAT uses STATSGO soils (which combine the more specific soil types referenced in the Delaware County soil test database), these averages had to be converted to represent the STATSGO soils in Delaware County. This was achieved by taking a weighted average of the crop cover/SSURGO soil averages, based on the SSURGO soil type composition percentages for each of the STATSGO soil types. The result of the analysis is unique soil test P levels for each STATSGO soil on all agricultural HRUs (e.g. corn silage, hay and pasture). For example, all NY056/corn silage HRUs were assigned the same initial soil P level to begin the simulation regardless of the subbasin the HRUs are located in.

The Morgan's P samples are assumed representative of a 15 cm plough layer. Thus, the soil test database is only used to initialize soil P levels in the top two soil layers of agricultural soils. These top two soil layers in SWAT are represented with a surface layer 1.0 cm in depth and a second soil layer ranging from 14.4 to 17.6 cm in thickness across the basin. Note that initially, the properties of the top two layers are identical for agricultural lands since SWAT actually generates this top layer of soil based on the top layer of soil given in the STATSGO database. The top 1 cm of soil is the only soil layer in which the model assumes interacts with surface water runoff. For all soil layers below second layer, SWAT default values are used to initialize soil P levels.

SWAT models six pools of soil P throughout the simulation. These include the mineral soil P pools of labile (or soluble) P, active mineral P and stable mineral P and the organic P pools of fresh organic P and the active and stable organic P pools associated with soil humus. Morgan's P samples do not represent the sum of all P in the soil as represented in SWAT. However, Morgan's P soil tests can be converted to estimate the labile P concentration in the soil. First, the Morgan's P levels in lbs/acre are converted to mg Morgan's P /kg soil by the common soil test laboratory conversion factor:

1 mg Morgan's P/kg soil = 2 lbs Morgan's P/acre

(Quirine Ketterings, Personal Communication). Then, Morgan's P levels in mg/kg soil are converted to labile P concentrations in the soil by the conversion:

1 mg labile P/kg soil = 0.4 mg Morgan's P/kg soil

(Andrew Sharpley, Personal Communication). Although these conversion factors can vary between sample sites and lab, they were assumed reasonable for the purposes of estimating initial labile P levels that are averages under similar land cover/soil type

conditions instead of a single sample associated with a specific sampling location. The resulting input area-weighted averages (across soil STATSGO soils) and ranges used as initial labile P concentrations for all agricultural land uses are summarized in Table 4.9.2.

4.9.1.2 Labile Soil P Levels for Forests, Grass-shrub and Idle Agriculture

Data collected at two forested sites in the Cannonsville basin (NYSWRI 2002a) were assumed to be representative of all forested land across the basin even though the sampling sites are representative of only upland forest soil P conditions. Most of the forest sampling results are assumed to represent deciduous and coniferous forests while some of the samples are more representative of grass-shrub land use rather than forested land. Although the forest sampling effort distinguished between the humus layer (mainly organic soil from leaf litter) and first mineral soil layer, the humus layer P samples were not used here. Instead, the mineral soil P levels sampled were assumed to represent the soil P levels in the two upper layers of soils modeled by SWAT. This is reasonable because the equations for P transport in SWAT are developed for mineral, not humus, soils and the measured soil P levels in the humus layer were greater than the measured soil P levels for corn. This approach was advocated by soil P expert Andrew Sharpley (Personal Communication). The same assumptions from Section 4.9.1.1 are also used here to estimate initial labile P concentrations for forest and grass-shrub soils in both the calibration and validation period. Forest and grass-shrub soil labile P concentrations are also summarized in Table 4.9.2 below.

Soil P samples were not available for idle agricultural land and had to be estimated. The idle agricultural land was assumed to have previously been pasture or hay since this land use is a subcategory of the NYCDEP grass land use (see Section 4.2.1.2). Therefore, the initial soil P levels are assumed to be closely related to the levels measured on hay and pasture on the same STATSGO soil map unit. In addition, it is assumed that idle agriculture land should have initial soil P levels that are slightly higher than the measured grass-shrub levels. Therefore, initial labile P concentrations for idle agriculture ($Sol\ P_{idle\ agr,\ i}$) on each STATSGO soil type i were calculated based on the following equation:

$$Sol\ P_{idle\ agr,\ i} = \{Sol\ P_{grass\text{-}shrub} + (Sol\ P_{hay,\ i} + Sol\ P_{pasture,\ i})/2\}/2$$

The above equation averages the soil type hay and pasture values for labile P because the idle agricultural land could have previously been hay or pasture.

Table 4.9.2 shows that the relative magnitudes of the labile P levels for forests and agricultural lands (corn, hay, and pasture) appear reasonable since soluble P levels of the forests, which do not receive P inputs, are approximately ¼ of the average levels in agricultural lands, which have P applied to them.

Table 4.9.2. Model calibration and validation initial labile (soluble) P concentrations in mg P/kg soil for the top two soil layers averaged (area-weighted) across all soils.

Land Use	Area-weighted Average	Minimum	Maximum
Forest	5.6	5.6	5.6
Grass-shrub	7.4	7.4	7.4
Corn, hay, and pasture[1]	18.4	8.6	65.1
Corn	43.6	22.3	65.1
Hay	22.2	9.8	31.2
Pasture	13.9	10.5	19.1
Idle agriculture	12.0	8.6	16.3
All land covers[2]	9.4	5.6	65.1

1. Corn, hay and pasture areas represent all active agricultural lands across the basin.
2. SWAT Default P levels assigned to urban land are not considered in this total. However, their impact would be negligible on this estimate.

4.9.1.3 Total Mineral Soil P Levels for all Land Uses

The initial total mineral P levels in SWAT (active and stable mineral P in addition to labile P) are calculated by the model depending on 1) the initial labile P concentrations as determined in the previous section and 2) a soil P sorption coefficient (PSP) which is a model parameter that is applied at a constant value across all basin soils. Using some basin specific soil characteristics, reasonable bounds can be determined for PSP as described below.

The soil phosphorus cycling algorithms in SWAT are incorporated from the EPIC model (Williams et al. 1984). The PSP coefficient, which is defined as the fraction of fertilizer P remaining in the labile pool after the initial rapid phase of P sorption is complete, can be predicted from a set of equations related to general soil characteristics. A reasonable equation for predicting PSP in slightly weathered, non-calcareous soils that are representative of the Cannonsville Basin is given in Sharpley et al. (1984) as:

$$PSP = 0.0043\ BS + 0.0034\ AP2 + 0.11\ PH - 0.7$$

where BS is the base saturation as a percentage, AP2 is the labile P in the soil in mg labile P/kg soil and PH is the soil PH. Using the basin wide area-weighted average concentration of 9.4 mg labile P/kg soil from Table 4.9.2 above, assuming an average soil pH of 6.0 based on Delaware County agricultural soil data collected by Kleinman et al. (1999) and assuming a base saturation of 40 to 100 percent, the PSP value predicted by the equation above ranges from 0.16 to 0.42. In order to calculate the initial total mineral P concentrations in the soil the final PSP value for this model, which is assumed determined in Section 6.3.2.2 to be 0.25, is required.

Based on the PSP value of 0.25 and the initial labile P concentrations in Section 4.9.1.2, the area-weighted averages and ranges of initial total mineral P concentrations are given in Table 4.9.3 for all land uses. Again, as in the Table 4.9.2 above, the mineral P levels

in Table 4.9.3 below seem reasonable with respect to forest relative to agricultural land mineral soil P concentrations.

Table 4.9.3. Model calibration and validation initial mineral P concentrations (labile P + active mineral P + stable mineral P) in mg P/kg soil for the top two soil layers averaged (area-weighted) across all soils and using PSP=0.25.

Land Use	Area-weighted Averages	Minimum	Maximum
Forest	90	90	90
Grass-shrub	118	118	118
Corn, hay, and pasture[1]	294	137	1042
Corn	698	358	1042
Hay	355	158	498
Pasture	223	169	306
Idle agriculture	192	137	260
All land covers[2]	150	90	1042

1. Corn, hay and pasture areas represent all active agricultural lands across the basin.
2. SWAT Default P levels assigned to urban land are not considered in this total. However, their impact would be negligible on this estimate.

4.9.1.4 Organic Soil P Levels for all Land Uses

There is little basin-specific data available for setting initial soil organic P levels in the model. However, rather than having SWAT set default initial organic P levels that may be inconsistent with the total mineral P derived above, it was deemed more appropriate to set initial organic P levels that were related to the initial total mineral P levels. As above, organic P levels are only derived for the first two soil layers (approximately the top 15 cm of soil) in SWAT and model default values are used to initialize organic P levels in soil below the second layer.

It is assumed that the organic P content of soils, relative to total soil P, is highest for forest soils and lowest for cropped land like corn. The organic P fraction of total soil P for forest, grass-shrub, idle agricultural, hay, pasture and corn silage are assumed to be 0.7, 0.7, 0.6, 0.5, 0.5 and 0.25, respectively. These organic P content assumptions are generally consistent with the basin-specific soil P data collected for corn and forest by Kleinman (1999).

4.9.1.5 Total Soil P Levels for all Land Uses

The SWAT model only requires inputs for initial soil P levels for labile P and active+stable organic P. These inputs are derived above. Table 4.9.4 gives the average initial soil total P levels each land use in this study. Corn land use is modeled with the highest initial total P concentration while forests are modeled with the smallest. Agricultural land uses have a higher initial total P concentration than non-agricultural land uses. The average, area-weighted soil total P level of approximately the top 15 cm of basin soils was about 400 mg P/kg soil. Note that Table 4.9.4 does not include a small amount (no more than 10 mg P/kg) of fresh organic P in the initial residue on the surface

layer of the soil. These basin specific, data-driven soil P levels are significantly higher than the SWAT model default soil P levels.

Table 4.9.4. Model calibration and validation initial total soil P concentrations in mg P/kg soil for the top two soil layers averaged (area-weighted) across all soils.

Land Use	Area-weighted Averages	Minimum	Maximum
Forest	300	300	300
Grass-shrub	393	393	393
Corn, hay, and pasture[1]	603	256	1389
Corn	931	477	1389
Hay	709	315	997
Pasture	446	337	612
Idle agriculture	481	343	650
All land covers[2]	395	209	1389

1. Corn, hay and pasture areas represent all active agricultural lands across the basin.
2. SWAT Default P levels assigned to urban land are not considered in this total. However, their impact would be negligible on this estimate.

5 SWAT2000 Source Code Modifications/Corrections

The SWAT2000 model source code is distributed with the download of the model executable program. This means that any changes to the model equations or calculation methods deemed necessary were possible to implement. This also means that the model can be recompiled so that it runs using a minimum amount of computer processing time and so that it can be linked with supplementary programs such as those to be used for auto-calibration or uncertainty analysis of the SWAT2000 model. Although the purpose of the calibration and validation was not to modify the equations in the SWAT2000 model, it became apparent that seven source code modifications were necessary for either convenience, improved model performance or because an error in the original SWAT2000 source code required correction.

5.1 Modification 1: Implement Efficient Daily Manure Spreading

In order to efficiently specify daily spreading of manure the model source code related to implementing the grazing management option in the SWAT2000 model was slightly modified. After this modification, the grazing management option could be used to *only* spread manure on all crop types (e.g. WMANURE input variable > 0) and the removal of plant biomass by cattle was not simulated (BMEAT input variable = 0) at all times of the year. Although grazing of cattle does currently occur in the basin, it is assumed that simulating grazing with the model was unnecessary since general grazing management practices within the basin result in only small changes in erosion (Karl Czymmek, Personal Communication). Furthermore, the grazing management option in SWAT2000 is new and relatively untested against observed grazing data (Jeff Arnold, Personal Communication).

5.2 Modification 2: Ensure Plant Growth Simulated in all Years

Initial model runs showed that plant growth in a number of corn and forest HRUs was not occurring in at least one year of the simulation. This non-growth produced highly excessive sediment and P loads from affected HRUs due to the loss of plant cover protection from erosion. After consulting with model developers, this model behavior was deemed unique and should not be occurring (Jeff Arnold, Personal Communication). Furthermore, it would be unreasonable relative to actual conditions in the basin to simulate one or more years where corn HRUs were fallow and forests do not grow. The only solution to this observed problem was to make a minor modification to the source code of the model nutrient stress evaluation routine.

5.3 Modification 3: Winter Surface Runoff Adjustment on Frozen Soils

The default SWAT2000 approach is to increase predicted surface runoff volumes when the second soil layer (not the top 1 cm soil layer) is frozen. This approach was thought to be the cause of flow over-predictions observed for multiple, late fall and spring events in

the calibration period. These events occurred in November and April and therefore occurred when the soil temperatures were very close to freezing near the soil surface. Since SWAT2000 simulates the increase in surface runoff based on soil a temperature threshold being crossed, model predictions change drastically for simulated soil temperatures just above freezing and just below freezing. Therefore, it was deemed appropriate to change the model such that increases in surface runoff were only triggered when it was relatively certain that most of the soil profile was frozen. This change involved calculating a depth-weighted average soil profile temperature and then comparing this temperature to the freezing mark in order to determine if surface runoff volumes would be triggered to increase.

5.4 Modification 4: Soil Water Above Field Capacity in Frozen Soils

In SWAT2000, for any non-frozen 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 default SWAT2000 approach is to:

1. Assume no lateral flow occurs.
2. Assume soil layer can hold excess water above field capacity until soil saturation is reached (presumably based on assumption that all water freezes in the soil).
3. Assume that for a frozen saturated soil layer, additional percolated water from above passes through the layer to become percolate to the next soil layer.

The end result of the above approach is that rainfall or 55 simulated as percolate (as opposed to surface runoff) on a frozen and saturated soil profile goes entirely to groundwater and the transfer of this percolate to the stream is greatly delayed. In fact, it was observed that the default SWAT2000 approach described above was responsible for transferring too much of the winter precipitation to streamflow in the early summer.

The most efficient model modification to make the handling of percolate into a frozen soil layer more reasonable was to allow the model to predict lateral flow in frozen soils. In other words, the model partitioning of water above field capacity into percolate and lateral flow is not affected by soil layer temperature. This change in the SWAT2000 code produced more accurate flow predictions relative to the measured data. Model performance improvement, relative to measured flows, associated specifically with this change is evaluated explicitly at the end of Section 6.4.1.

5.5 Modification 5: Orographic Adjustment of Average Monthly Subbasin Temperatures

Average monthly maximum and minimum temperatures in the SWAT2000 weather generator input files (or ".wgn" files) are used by the model for soil temperature calculations. However, the default SWAT2000 approach does not adjust these for elevation differences between the weather generator climate station and the average

subbasin elevation to which they are assigned. In order to be consistent with the simulation of orographic impacts on temperature, the source code was modified to adjust average monthly maximum and minimum temperatures for elevation differences between the weather generator climate station and the average subbasin elevation.

5.6 Modification 6: Correct Snowmelt Logic when Elevation Bands Simulated

It was discovered that the default SWAT2000 model logic for estimating snowmelt is inconsistent depending on whether elevation bands are simulated. Specifically, when elevation bands are not simulated, the model allows snowfall and snowmelt to occur on the same day depending on the model parameters for these processes (SFTMP and SMTMP). However, when elevation bands are simulated, regardless of the SFTMP and SMTMP parameters, the default SWAT2000 code does not allow snowmelt to occur on days when the precipitation type is classified as snow. Since snowmelt is computed in SWAT2000 as a function of both average and maximum daily temperature, there are days when the average temperature suggests snowfall and the maximum temperature suggests that snowmelt also occurred. To allow simulation of this reasonable phenomenon, the source code for snowfall and snowmelt computation in elevation bands was modified to be replicate the SWAT2000 logic for snowfall and snowmelt computation when elevation bands are not simulated.

5.7 Modification 7: Change the MUSLE Snow Cover Adjustment Equation

During model calibration (Section 6.3) it became apparent that the MUSLE equation adjustment for erosion prediction under snow cover (equation 13.3.1 in Neitsch et al. (2001a)) was unsuitable for prediction of snowmelt erosion in the Cannonsville Basin based on comparison of the predicted and measured sediment loads for snowmelt months and consideration of the nature of the MUSLE equation adjustment. For example, during the most extreme sediment loading month on record (January 1996) the initial model predictions showed almost no significant sediment load (relative to other months) at Beerston but does predict a significantly larger than average flow for the month. Although the model is not expected to perform accurately under this extreme month, it should predict at least some increase sediment load when flows are significantly higher than the average.

The failure of the model to predict any sediment load increase during this month was directly attributable to the MUSLE snow cover adjustment equation. For example, consider the behavior of the SWAT2000 snow cover adjustment equation plotted in Figure 5.7.1. At a snow cover depth of 50 mm of snow water equivalent, the MUSLE predicted erosion is reduced by 99.7% in the original SWAT2000 model regardless of the surface runoff depth for the day. While this may be reasonable for relatively small snowmelt events, this relationship becomes more unreasonable as the surface runoff depth increases relative to the snow cover water equivalent. For example, during the January 1996 event, many HRUs were predicted to have approximately 50 mm of surface runoff in a single day where on the order of ½ the snow water equivalent melted. Surely, with this volume of water draining over the soil surface, significant erosion should be

occurring – even if the soil was initially frozen. Similar situations were observed in other simulated snowmelt periods.

Clearly, the larger than average snowmelt events should produce some significant sediment erosion. In fact, other studies in the literature suggest that MUSLE predictions of erosion under snow cover should not necessarily be reduced (Johnson et al. 1985, McConkey et al. 1997). Perhaps the best argument for modifying the SWAT2000 model erosion predictions under snow cover is the total inability of the model to predict higher than average sediment loads for snowmelt months where the measured sediment loads were the highest recorded. Therefore, various attempts were made to find a more representative adjustment to MUSLE sediment erosion under snow cover. Modifying the default SWAT2000 MUSLE adjustment equation for snow cover is reasonable because this equation appears to be empirically (rather than physically) based. The model adjustment was selected to account for the idea that as the fraction of the snowpack that melts increases, so too should the sediment erosion that occurs on that day. Furthermore, it was assumed that a snowpack of any depth provides some minimum level of protection from erosion. The new adjustment to MUSLE sediment erosion estimates under snow cover is based on the depth of surface runoff relative to the depth of the snow cover and is given by the following equation:

$$MUSLE_i` = MUSLE_i * \min (0.95, [SurQ_i / Sno_i]^{0.25})$$

where $MUSLE_i`$ is the final snow cover adjusted sediment yield for a given HRU on day i, $MUSLE_i$ is the sediment yield predicted with the original MULSE equation (Williams and Berndt 1977) on day i, $SurQ_i$ is the surface runoff depth (mm) on day i and Sno_i is the snow water equivalent depth (mm) of the snowpack on day i. The constants in the equation above were iteratively refined in the calibration process until monthly simulated winter sediment loads more closely replicated measured winter sediment loads.

The new equation for adjustment to the MUSLE sediment erosion estimates under snow cover is plotted for various surface runoff depths in Figure 5.7.1. For example, with reference to the conditions simulated for the January 1996 event discussed above, for an HRU with 100 mm of snow cover water equivalent and a surface runoff depth of 40 mm, the new equation above reduces the MUSLE predicted sediment load by only 20% instead of by 99.7% as predicted in the original SWAT2000 model code.

This change in the SWAT2000 code did produce more realistic sediment load predictions relative to the measured data. Model performance improvement, relative to measured sediment, associated specifically with this change is evaluated explicitly in Section 6.4.2.

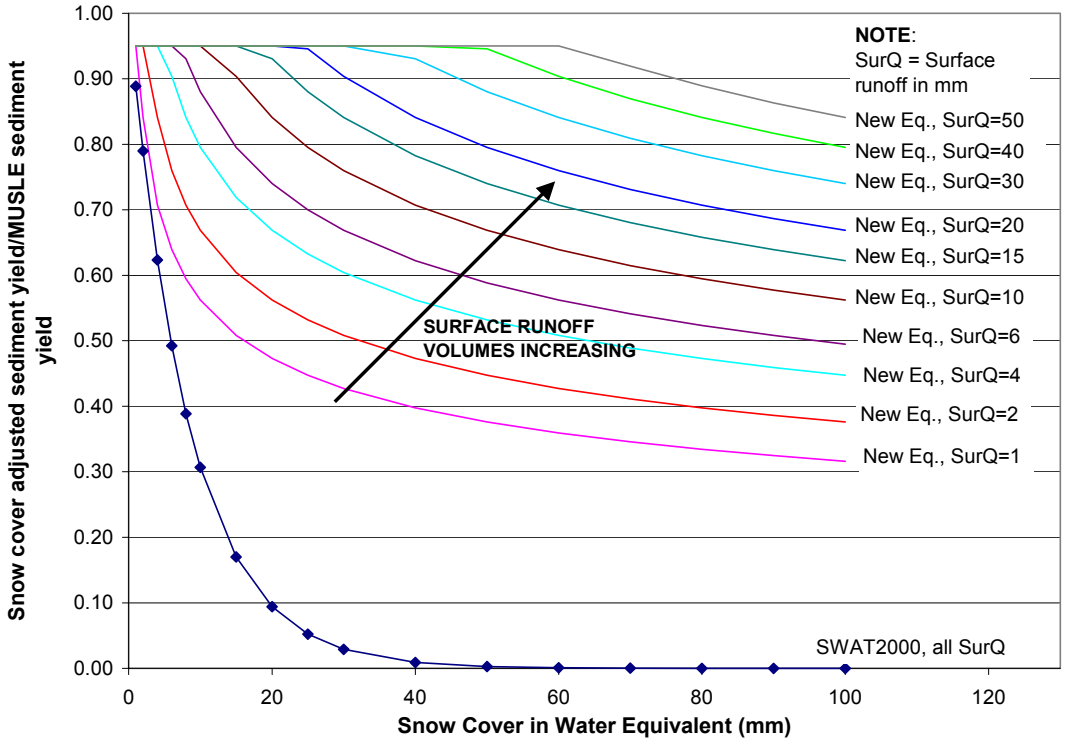

Figure 5.7.1. Comparison of sediment yield adjustments under snow cover for various surface runoff volumes (SurQ) between the original SWAT2000 model code and the modified model code (New Eq.) proposed for the Cannonsville Basin.

5.8 SWAT2000 Model Modification Summary

All of the above source code modifications were thoroughly checked to ensure that the desired results were achieved and so that no other simulated processes were adversely affected. Except for the modification related to daily manure spreading through grazing, the above modifications should be viewed as improvements over the alternative of using the original version of SWAT2000. The changes above have been discussed with the lead model developer (Dr. Jeff Arnold) who did not disagree with any of the modifications. Thus, the changes to the SWAT2000 model listed above were assumed reasonable for the purposes of this study. All other changes to the model source code only involved printing internal model variables to output files for the purposes of checking model simulation results that are not available in standard model output files.

6 Calibration and Validation

Watershed models contain many parameters, some of which cannot be measured. In order to utilize any predictive watershed model for estimating the effectiveness of future potential management practices the model must be first calibrated to measured data and should then be tested (without further parameter adjustment) against an independent set of measured data. This testing of a model on an independent data set is commonly referred to as model validation. Model calibration determines the best, or at least a reasonable, parameter set while validation ensures that the calibrated parameter set performs reasonably well under an independent data set. Provided the model predictive capability is demonstrated as being reasonable in both the calibration and validation phases, the model can be used with some confidence for future predictions under somewhat different management scenarios.

6.1 Monitored Flow and Water Quality Data

A large amount of data, both spatial and temporal, exists to aid in the calibration and validation of the Cannonsville SWAT2000 watershed model. This calibration effort focused on utilizing all available monitoring data for model calibration and validation. Stream monitoring data were available from a number sources and are all listed in Table 6.1.1.

USGS streamflow data and NYSDEC continuous water quality monitoring data (NYSDEC Unpublished data) on total suspended sediments (TSS), total dissolved phosphorus (TDP) and particulate phosphorus (PP) were used in model calibration and validation. Portions of the NYSDEC data set have previously been published in Longabucco and Rafferty (Longabucco and Rafferty 1998). As discussed in Section 4.6, the NYCDEP water quality grab sample data (NYCDEP Unpublished data) were used to specify spatially distributed groundwater TDP concentrations across the basin. The same NYCDEP water quality data were also used to qualitatively test the spatially distributed performance of the model during calibration and validation for total P predictions. Figure 6.1.1 shows the location of available USGS flow stations and the NYSDEC water quality monitoring stations while the locations of the NYCDEP water quality monitoring stations used for spatially distributed calibration and validation of total P are provided in Figure 6.4.7.

Table 6.1.1. Summary of stream monitoring data used in model calibration and validation.

Data	Locations (drainage area)	Period of Record	Supplying Agency
Continuous TSS, TDP and PP Monitoring	Beerston (913 km^2) Town Brook (37 km^2)	1991-pres 1998-pres.	New York Department of Environmental Conservation (NYSDEC)[1]
Bi-weekly Phosphorus Monitoring	Over 20 locations of bi-weekly grab-samples taken across the basin	1988-1999	New York Department of Environmental Protection (NYCDEP)[2]
Streamflow Monitoring	WBDR @ Walton, 01423000[3] (860 km^2)[4] WBDR upstream of Delhi, 01421900 (347 km^2) Little Delaware, 01422500 (135 km^2) East Brook, 01422747 (64 km^2) Trout Creek, 0142400103 (54 km^2) Town Brook, 01421618 (37 km^2)	1950-present 1997-present 1997-present 1998-present 1997-present 1997-present	USGS[4]

1. NYSDEC (Unpublished data).
2. NYCDEP (Unpublished data).
3. USGS gauge number.
4. Available at http://water.usgs.gov.

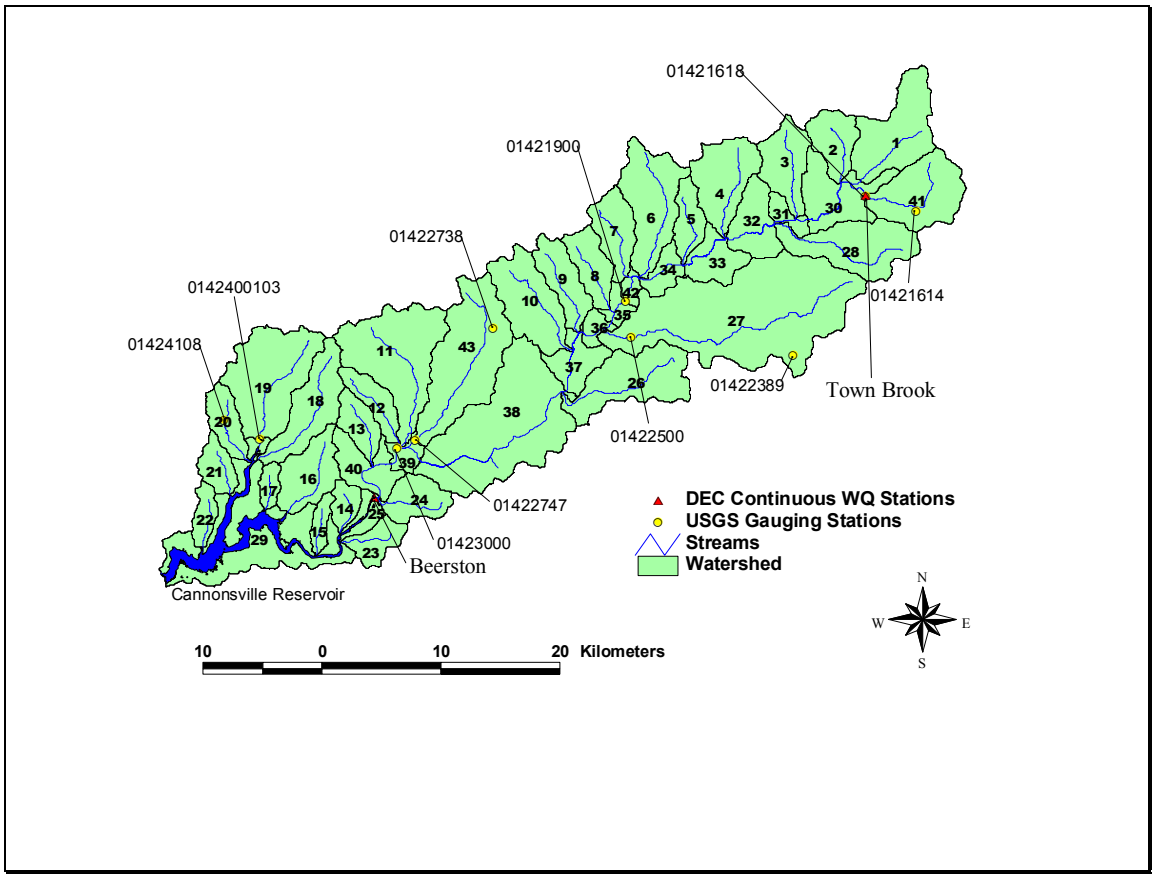

Figure 6.1.1. Continuous flow and water quality monitoring stations in the Cannonsville Basin as listed in Table 6.1.1.

6.1.1 'Measured' TSS and P Loading

NYSDEC continuous water quality monitoring data (NYSDEC 2001) are *calculated* monthly loading *estimates* of TSS, PP and TDP. NYSDEC does not measure monthly sediment or nutrient loading explicitly. Therefore, the NYSDEC data are not strictly *measured* loading although in the remainder of the report they are referred to as 'measured loadings' in order to clearly distinguish these quantities from the simulated model results. NYSDEC calculates all loading estimates from measured USGS flow data and instantaneous water quality sampling data. Loading estimates are essentially the product of flow and a concentration. The NYSDEC calculation and sampling methodology for load estimation is outlined in detail in Longabucco and Rafferty (Longabucco and Rafferty 1998). It should be noted that although the NYSDEC water quality sampling methodology has recently been changed (automatic samplers are usually now used), nearly all of the NYSDEC estimated loadings for TSS and P used in this report were based on the sampling methodology as outlined in Longabucco and Rafferty (Longabucco and Rafferty 1998).

The main issue with respect to measured loadings is that they are in fact estimated or calculated loadings and are subject to some estimation error. The NYSDEC event-based

continuous monitoring strategy and loading analyses described in Longabucco and Rafferty (Longabucco and Rafferty 1998) is a great improvement over loading estimates derived from fixed interval sampling (i.e. weekly or bi-weekly sampling). However, the NYSDEC calculated loadings are still estimates that contain some error. For example, Longabucco and Rafferty (Longabucco and Rafferty 1998) report that NYSDEC water quality sampling at Beerston consisted of collecting a sample at approximately mid-channel by immersing a 4-L plastic jug in the water. This approach yields approximate estimates of the true channel cross-sectional TSS or P concentrations. Furthermore, although flow data are measured every 15 or 60 minutes by the USGS, TSS and P concentrations are sampled much less frequently. For example, over the entire calibration and validation period, the average NYSDEC water quality sampling frequency is 1 sample every two days. It should be noted that the NYSDEC sampling frequency during days classified as flow-event days is much higher at about 2 samples per flow-event day. Errors in water quality concentration samples combined with errors in flow result in NYSDEC TSS and P loading estimates that are subject to larger errors than the monitoring errors for flow. Therefore, differences between simulated and measured TSS and P loading are expected to be larger in relative magnitudes than the differences in simulated and measured flows.

Of course, on a practical level, measuring an integrated cross sectional water quality sample and/or sampling water quality continuously at 15 min or 60 minute intervals are not feasible approaches for the NYSDEC continuous monitoring program of the WBDR and its tributaries. The purpose of this discussion is not to suggest that improvements in the sampling methodology are required. Instead the purpose is to explain some of the reasons why differences between model results and measured loadings are expected to be less for flows than for TSS and P (true in all watershed modeling applications).

6.2 Equivalence of SWAT Outputs with Monitoring Data

Comparing simulated model results with monitoring data collected in the basin requires a careful examination as to the equivalence of the modeled quantity and the measured quantity. In the case of streamflow, monitored streamflow and modeled streamflow are equivalent and recorded on the same time scales and thus can be directly compared. However, special considerations are required when comparing sediment and phosphorus monitoring data with model predictions of these quantities.

6.2.1 Sediment

The total sediment load (particulate organic and inorganic matter) transported by rivers is made up of total suspended sediment (TSS) and bedload sediment. TSS is composed of organic and inorganic particulate matter that is transported in the water column while bedload is larger sediment particles that are transported along the river bottom. Although bedload can be greater than the total suspended sediment transport in smaller mountain or glacier fed streams, given the different nature and size of the WBDR, it is assumed that the TSS load accounts for at least 90% of the WBDR total sediment load.

SWAT simulates the loading of sediment to streams from the MUSLE equation. The MUSLE equation simulates total sediment erosion from land. In addition, SWAT simulates channel bed and bank sediment erosion and deposition. SWAT sediment routing equations simulate the transport of the total sediment loads (MUSLE equation estimated sediment loading and channel bed and bank derived sediments) through the stream network. Therefore, the model estimates of sediment loading are for total sediment loading. In comparison, the NYSDEC sediment monitoring data measures only total suspended sediment (TSS). However, since bedload sediment transport is estimated to be a small portion of the total sediment load, the model simulated sediment loading is assumed to be directly comparable with measured TSS loads. Since SWAT simulated sediment loading is assumed to be at least 90% TSS, simulated sediment loading is usually referred to as simulated TSS in the remainder of the report.

6.2.2 Phosphorus

With respect to P, NYSDEC water quality samples were analyzed for total P, soluble reactive P (SRP) and total dissolved P (TDP). Filtration of samples through a 0.45 µm membrane filter isolated the TDP fraction prior to analysis and particulate P (PP) was computed as the difference between total P and TDP (Longabucco and Rafferty 1998). SRP data are not used in calibration or validation of the SWAT model.

SWAT simulates six forms of P in the soil (see Table 6.2.1) and assumes groundwater soluble P concentrations for each HRU are constant through time. These seven forms of P are loaded to the stream network from the HRUs and are aggregated into two general forms of P for summary in the main model output files (*.rch and *.bsb files). These general P forms are called mineral P and organic P in SWAT model documentation and output. Mineral P loading to a subbasin stream from an HRU is the sum of P loading from GWsolP, P_{sol} and $MinP_{act}$ while organic P is the sum of $OrgP_{act}$, $OrgP_{sta}$, $OrgP_{frsh}$ and $MinP_{sta}$. Based on the components of SWAT mineral and organic P, it is clear that SWAT organic P is a misnomer since it includes $MinP_{sta}$ (Stable mineral P). SWAT assumes that, except for solution P, all other forms of P in the soil are attached to sediment particles upon entering the stream.

Table 6.2.1. Forms of P in SWAT.

Name of P Form	SWAT Variable Acronym
Active mineral P[1]	$MinP_{act}$
Stable mineral P[1]	$MinP_{sta}$
Solution P[1]	P_{sol}
Active organic P[1]	$OrgP_{act}$
Stable organic P[1]	$OrgP_{sta}$
Fresh organic P[1]	$OrgP_{frsh}$
Groundwater soluble P	GWsolP
Mineral P[2]	MINP_out
Organic P[2]	ORGP_out

1. Forms of P in the soil.
2. Forms of P routed in channel network.

SWAT simulated total P (mineral + organic P forms) and monitored data for total P are assumed equivalent. However, SWAT simulated loading predictions of the P components in the stream (SWAT mineral and organic P) may not exactly represent the component P forms monitored by NYSDEC. SWAT simulated mineral P includes active mineral P that is attached to sediment particles transported in surface runoff to the channel. If SWAT mineral P is to be directly compared to monitored TDP, then it must be assumed that 100% of active mineral P is completely desorbed from soil particles once surface runoff reaches the channel. Dr. Andrew Sharpley (Personal communication) commented that the amount of active mineral P desorbing from soil particles in the stream is probably significantly less than 100% of the available active mineral P. Assuming this is true and a significant amount of active mineral P remains bound to sediments transported in the channel, then the SWAT model mineral P form is likely to over-estimate monitored TDP and consequently under-estimate monitored PP.

As a result, the calibration to phosphorus data focuses mainly on total P. However, comparison of the simulated and measured phosphorus components should still show good agreement with respect to similar trends in the data. Significant deviations in simulated organic and mineral P loads from the corresponding PP and TDP monitoring data will be judged to be acceptable provided the simulated total P closely replicates the monitored total P load.

For brevity, simulated P results are mainly referred to as simulated TDP and simulated PP in the remainder of the report.

6.3 Calibration

6.3.1 General Approach

Benaman (2002) initially calibrated the model for flow and sediment. The authors of this report substantially modified the initial flow and sediment calibration by Benaman (2002) and also calibrated the model to phosphorus data. This section outlines the overall calibration approach and the following sections describe the calibration approach (i.e. model parameter changes from default values) in more detail. The final calibrated model accuracy with respect to the measured data are summarized using multiple numerical performance indicators and time series plots in Section 6.4.

Model inputs and parameters were adjusted in multiple stages. First, independent of numerical calibration (e.g. adjusting model parameters to optimize the model performance measures), model inputs and parameters, as outlined in Section 4, were updated. Second, in addition to the changes outlined in Section 4, there were a number of other model inputs and parameters adjusted to better represent known conditions in the watershed. These additional adjustments are described in Section 6.3.2. Model input and parameter changes listed in Section 4 and 6.3.2 are referred to in the remainder of this document as *data-driven* parameters. Data-driven parameters are grouped together because their final model default value was governed by available data or known conditions in the watershed rather than model predictive performance of measured flow

or water quality data. Lastly, a relatively small group of model parameters were selected as *performance optimization* parameters and their values were iteratively adjusted to best match measured flow, sediment and phosphorus data. Performance optimization parameters were selected because their values are more uncertain relative to other parameters and they were previously identified in the literature as sensitive model parameters (Neitsch et al. 2001b, Benaman 2002). Section 6.3.3 lists the performance optimization parameters. In a few special cases, parameters were classified as both a data-driven and performance optimization parameter. Therefore, the final values for these data-driven/performance optimization parameters were determined by considering both known information regarding their ranges and the predictive performance of the model. All data-driven and performance optimization parameters discussed below are constant in the calibration and validation periods. All SWAT2000 model parameters that were not identified as data-driven or performance optimization parameters were left at their model default values and were either observed or assumed to have relatively small impacts on model predictions.

In the process of fine-tuning the performance optimization parameters, it was observed that certain model prediction deficiencies could not be overcome by parameter adjustment alone. Further investigation into these deficiencies resulted in the identification of the model source code modifications outlined in Section 5.

The final values of the performance optimization parameters in this work were fine-tuned by a trial and error approach as opposed to any automatic calibration procedure. As a result, the number of performance optimization parameters modified during the calibration was kept to a minimum relative to the total number of SWAT parameters available for calibration (e.g., non data-driven parameters). Performance optimization parameters were modified iteratively, within SWAT2000 ranges, and generally one at a time, until no further significant improvement in model predictive performance was evident. The performance optimization parameters in the model were generally calibrated to measured hydrology, sediment and phosphorus data in that order. Model predictive performance relative to available measured data are evaluated for each constituent by calculating the percent difference (D) between average measured and simulated time series, the coefficient of determination (r^2) and the Nash-Suttcliffe simulation efficiency (E_{NS}) (Nash and Suttcliffe 1970). These three measures of model performance are defined in the appendix (Section 10.3) and were selected because together they measure how well the model predicts the average and temporal trends in the measured data. The goal of the calibration was to find values of performance optimization parameters that, for flow, sediment and phosphorus, resulted in a percent difference between average simulation and average measured data over the calibration period that was reasonably small (generally less than about 10%) and a r^2 value that was a maximum. Although model performance is mainly fine-tuned with respect to the performance measures for the entire year, seasonal performance measures were also assessed.

Hydrology calibration was focused on reproducing measured daily flows while sediment and phosphorus calibration was focused on replicating measured monthly loads. The

monthly time step is used for water quality calibration because of the reduced sampling frequency for water quality, relative to flow measurements, discussed in Section 6.1.1. The calibration effort generally focused on results at the Beerston/Walton monitoring stations since these most closely represented the sediment and P loading to the Cannonsville Reservoir and provided the longest time series of data. Therefore, performance optimization parameters were generally only adjusted to either improve model accuracy at Beerston/Walton or to improve the overall model performance at most or all of the monitoring locations. In other words, model performance at the monitoring stations other than Beerston/Walton were not optimized independently.

In addition to simulating measured daily flows, model calibration for hydrology considered baseflow predictions. Total streamflow is composed of baseflow and storm flow. To compare model predicted baseflow with measured baseflow, estimates of baseflow volumes from both the simulated and measured flow data are required. Based on the approach outlined by SWAT developers in Santhi et al. (2001), a baseflow analysis program by Arnold and Allen (1999) was used to analyze both the measured and simulated flows at the Walton USGS station and then compare the baseflows estimated from the measured and simulated data. Model prediction of baseflow was evaluated on an average annual and average monthly basis.

Sediment calibration was completed in two general steps. First, the model was calibrated to simulate measured monthly sediment loads in the summer. Second, the model predictions of monthly winter sediment loads were adjusted to better match the measured loads by adjusting the coefficients in the new MUSLE equation adjustment for erosion prediction under snow cover given in Section 5.7.

Phosphorus calibration was also completed in two general steps. First the model was calibrated to simulate measured monthly total P loads. Secondly, other performance optimization parameters were adjusted to try and improve model predictions of TDP and PP without significantly degrading model predictive performance for total P.

6.3.2 Data-Driven Input/Parameter Modifications

All SWAT2000 inputs and parameters listed in Section 4 are considered data-driven and will not be repeated in this Section. The remaining data-driven parameters are separated into three groups based on the primary model process affected by each parameter. All data-driven inputs and parameters listed below are assigned constant values over the calibration and validation periods.

6.3.2.1 Data-Driven Land use Crop/Plant Covers and Related Parameters

SWAT2000 simulates the growth of specific plants on each HRU. The plants that were simulated on each land use for the Cannonsville Basin are summarized in Table 6.3.1 and were assigned based on consultations with local agricultural planners/experts. The plants growing on each land use were selected from the original SWAT2000 crop database. Hay and pasture plants were modeled as timothy while corn was modeled as corn silage. Coniferous and deciduous forests were modeled as evergreen and deciduous forests,

respectively. Idle agricultural HRUs were modeled as timothy and grass-shrub HRUs were also modeled as timothy because no other plant in the SWAT crop database could be identified that represented this land use. In addition, it was assumed that idle agricultural and grass-shrub land uses have greater protection from erosion than harvested or grazed agricultural grasses due to more extensive biomass ground cover. Thus, the SWAT minimum USLE C factor (USLE_C) parameters for timothy grown on idle agricultural and grass-shrub HRUs were increased from their defaults. The grass-shrub USLE_C parameter was set as the average of the default SWAT USLE_C values for deciduous forest and timothy while the idle agricultural cover factor was set as the average of the grass-shrub USLE_C and the default SWAT USLE_C value for timothy. The final values of these USLE_C parameters are given in Table 6.3.2.

Table 6.3.1. SWAT2000 land uses and corresponding crop cover type in the Cannonsville Basin.

Modeled Land Use	**SWAT Crop/Plant Covers Modeled (4 letter code)**
Deciduous forest	Deciduous forest (FRSD)
Coniferous forest	Coniferous forest (FRSE)
Grass-shrub	Timothy A[1] (GRSH)
Idle agriculture	Timothy A (TIMU)
Hay[2]	Timothy B[1] (TIMC)
Pasture	Timothy B (TIMP)
Corn	Corn silage (CSIL)
Urban	High density residential areas and Altai wildrye grass (RYEA)
Water	-

1. Timothy A represents unmanaged grasses and shrubs while Timothy B represents managed grasses. Differences in their growth and erosion parameters are outlined in Table 6.3.2.

Urban areas in the Cannonsville Basin vary from relatively dense areas of commercial/industrial buildings to relatively sparse areas of residential land. Since the available NYCDEP land use data classified all urban land as one type and the total area of urban land was only 0.4% of the entire basin, only one type of cover was modeled for urban land. It was assumed that the SWAT urban land use category called Residential-High Density was representative of all urban land uses lumped together. The fraction of this area considered impervious is 0.6 (default value) and the pervious areas were modeled as a perennial grass (Altai wildrye). Impervious areas were simulated with the build-up and wash off process option in SWAT2000 for urban sediment and nutrient loading. SWAT2000 default parameters for this urban land use were used without modification.

Each plant cover in the SWAT2000 crop database has default growth parameters associated with it. After consultations with the model developer and local agricultural experts it was determined that some of the crop growth parameters should be modified from their defaults. Table 6.3.2 outlines the crop growth parameter changes. The

SWAT2000 model developer (Dr. Jeff Arnold, Personal Communication) suggested modifications to the SWAT temperature-based growth parameters while a report by Cerosaletti (2002) outlines the lab analysis results for corn silage and hay forage for dairy cattle on four farms within the basin. The average forage P contents of corn silage (44 samples) and hay (135 samples) from these farms were assumed to best represent the P content of all corn silage and timothy grown in the Cannonsville Basin.

Table 6.3.2. Data-driven crop parameters changed from default SWAT2000 values.

SWAT Parameter (short variable name)	**Description of Change**	**SWAT Crop Cover: Modified Value (Default Value)**
Minimum USLE C factor (USLE_C)	Decreased to reduce erosion from idle and successional agricultural land relative to erosion from hay or pasture	FRSD: no change (0.001) TIMC & TIMP: no change (0.003) GRSH: 0.0020 (none) TIMU: 0.0025 (none)
Optimal temperature for plant growth in °C (T_opt)	Changed to account for Northeast climate	FRSD: 25 (30) FRSE: 25 (30) TIM*[1]: 20 (25)
Minimum (base) temperature for plant growth in °C (T_base)	Changed to account for Northeast climate	FRSD: 6 (10) FRSE: 4 (0) TIM*[1]: 5 (8)
Normal fraction of phosphorus in yield in kg P/kg yield (CPYLD)	Increased based on Cannonsville Basin data for locally grown corn silage and hay (Cerosaletti 2002)	CSIL: 0.0025 (0.0016) TIMC & TIMP: 0.0040 (0.0033)
Phosphorus uptake parameter #1 (BP1)	Increased by same relative amount as CPYLD increase above	CSIL: 0.0075 (0.0048) TIMC & TIMP: 0.0046 (0.0038)
Phosphorus uptake parameter #2 (BP2)	Increased by same relative amount as CPYLD increase above	CSIL: 0.0028 (0.0018) TIMC & TIMP: 0.0030 (0.0025)
Phosphorus uptake parameter #3 (BP3)	Increased by same relative amount as CPYLD increase above	CSIL: 0.0022 (0.0014) TIMC & TIMP: 0.0023 (0.0019)

1. TIM* is all Timothy – GRSH, TIMC, TIMU and TIMP.

6.3.2.2 Data-Driven Flow, Sediment and Phosphorus Parameters

Three flow related, four sediment related and one phosphorus related data-driven parameters were not fully described in Section 4. These parameters and their final or reduced parameter ranges are summarized in Table 6.3.3 and the basis for their modifications are discussed in more detail in the paragraphs below.

Table 6.3.3. Data-driven flow, sediment and phosphorus related parameters.

Parameter, Units (SWAT variable name)	Basis for Modification	SWAT2000 Default Value	Final Value (Range)
Snowfall temperature, °C (SFTMP)	Precipitation, precipitation type and temperature data @ Delhi and Walton Climate Stations	1.0	1.0 (0.0-2.2)
Main and Tributary Channel Manning's n (CH_(2) & CH_(1))	Table 6.2 in Neitsch et al. (2001a)	0.014	0.05 (0.025-0.065)
Base or Initial SCS Curve Numbers for all land covers[1] (CN2)	Default CN2s for idle agricultural and grass-shrub land use needed and corn silage CN2s refined for each STATSGO map unit	Varies based on Table 6.2 in Neitsch et al. (2001a)	See Table 6.3.4
HRU slope fraction[2] (SLOPE)	Calculate actual HRU slopes from DEM	Average subbasin slope	Varies by HRU
Corn silage biological mixing coefficient (BIOMIX)	Tilled agricultural land has less biological activity than land with no-till	0.20	0.05 (NA)
USLE equation support practice factor (USLE_P)	Represent erosion control practices on corn silage and hay	1.0	Varies by HRU according to slope. See this Section text.
Sediment concentration in lateral and groundwater flow, mg/L (LAT_SED)	NYCDEC monitoring data for during non-events (Longabucco and Rafferty 1998)	0.0	See Section 6.3.3 (0.0-5.3)
Fraction of fertilizer P remaining in labile pool after initial rapid phase of P sorption complete (PSP)	Multiple Sources (see Section 4.9.1.3)	0.40	0.25 (0.16-0.42)

1. The base SCS curve numbers are modified further as outlined in Section 6.3.3.1.
2. The SWAT parameters SLSOIL and SLSUBBSN are reassigned based on the new HRU slopes (see paragraphs below for details).

The snowfall temperature parameter in SWAT2000 (SFTMP) determines whether precipitation falls as snow or rain. If the temperature of a subbasin is greater than SFTMP then precipitation in the model is classified as rain, otherwise it is classified as snow. Available weather data were analyzed to determine that the probable SFTMP

range was between 0 and 2.2 °C. Refer to appendix (Section 10.4.1) for details of this analysis. Since the default SFTMP value of 1.0 °C falls close to the middle of the probable range, the default of 1.0 °C was used for this parameter. Fixing the SFTMP parameter simplified the approach outlined in Section 4.4.1.2 for forcing the model to correctly classify precipitation type when it is known with relative certainty based on available data.

Default SWAT2000 Manning's n values for all basin stream channels were set at 0.014. Table 6.2 in Neitsch et al. (2001a) shows that for natural streams with few trees, stones or brush a nominal Manning's n value for channel flow is 0.05. Therefore, Manning's n values for all main and tributary channels in the model were assigned a Manning's n value of 0.05.

The Soil Conservation Service (SCS) Curve Number (CN2) parameter in SWAT influences the amount of surface runoff calculated for each rainfall or snowmelt event. CN2 varies by land use and soil type in SWAT. The base AVSWAT generated values for hay, corn silage and pervious urban lands were accepted as reasonable base values and are given in Table 6.3.4 for each soil hydrologic group. CN2 values in Table 6.3.4 are referred to as base values because they are further modified in Section 6.3.3 by a constant multiplicative factor that is considered a performance optimization parameter. Based on Table 20.2 in Neitsch et al. (2001a), new base CN2 values for pasture, idle agriculture, grass-shrub and forest were identified to replace to default AVSWAT CN2 values. All forested land was assumed to be in good hydrologic condition while all pasture land was assumed to be in fair to good (average) hydrologic condition. These assumptions for forest and pasture result in the base CN2s for the land uses given in Table 6.3.4. Idle agriculture and grass-shrub CN2s were calculated by averaging CN2s for other land uses as shown in Table 6.3.4. All HRUs with water as a land use are assigned AVSWAT default CN2s of 92.

Corn silage base CN2s in Table 6.3.4 for soil hydrologic groups A, B, C and D were used to estimate more accurate base CN2s that were unique to each STATSGO map unit. More accurate CN2s for corn silage were deemed important because initial simulation results showed corn silage HRUs have surface runoff and erosion rates that were the much higher (an order of magnitude for sediment erosion) than any of the other land uses. Corn silage base CN2s for each soil hydrologic group in Table 6.3.4 were assigned to the SSURGO soils and then the STATSGO map unit CN2s were calculated by an area-weighted average of the SSURGO soil CN2s within each STATSGO map unit.

HRU specific slopes were estimated for the 482 initial HRUs and the 22 manually added corn HRUs (see Section 4.2). New, subdivided grass HRUs derived from this initial set of HRUs were all assigned slopes equal to the HRU they were derived from. Default AVSWAT slope estimates were based on average subbasin slopes such that steep forested valley walls and agricultural lands in the relatively flatter valley bottoms were assigned the same slopes. Therefore, the average subbasin slopes are overestimates for the agricultural HRUs that are often concentrated along rivers and tributaries. Hydrologic, sediment and therefore P model predictions are sensitive to the HRU slope

estimates. Since other sediment erosion parameters are assigned at the HRU level, the subbasin average slopes were replaced with HRU specific slopes calculated from additional GIS processing. Analysis of the HRU specific slopes verified the significant difference between forest and agricultural HRU slopes. For example, the average forest HRU slopes are approximately twice as steep as the average corn silage, hay and pasture HRU slopes.

Table 6.3.4. Base CN2 values assigned to all HRUs.

	Land Use						
Soil Hydrologic Group[1]	**Pasture**	**Hay**	**Corn Silage[2]**	**Idle Agr.[3]**	**Grass-Shrub[4]**	**All forest**	**Urban[5] (Pervious)**
B	65.0	59.0	77.0	59.5	57.0	55.0	59.0
C	76.0	72.0	83.0	72.5	71.0	70.0	72.0

1. The most common hydrologic group derived for each STATSGO soil map unit as described in Section 4.3 was either group B or C. Soils in STATSGO map unit NY099 are classed as hydrologic group B and all others as hydrologic group C.
2. To refine corn silage base CN2s further, base CN2 for soil hydrologic group A (67.0) and group D (83.0) were required. The refined base CN2s assigned to corn silage HRUs on STATSGO map units NY026, NY027, NY056, NY059, NY099, NY127, NY129, NY132, NY133 and NY136 were 82.2, 80.6, 82.9, 82.4, 73.4, 73.9, 82.4, 81.1, 83.0 and 79.7, respectively.
3. Idle agriculture CN2 set equal to:
   ½[grass-shrub CN2 + ½(hay CN2 + pasture CN2)].
4. Grass-shrub CN2 set equal to: ½[All forest CN2 + hay CN2].
5. Urban CN2 applied to pervious grass areas only (impervious area CN2 fixed in model at 98).

The SWAT parameters SLSOIL and SLSUBBSN are adjusted for the new HRU specific slopes. The default SWAT2000 approach functions to assign SLSOIL and SLSUBBSN values based on the default slope values assigned by AVSWAT. Therefore, the original AVSWAT default values for these parameters are replaced with values that correspond to the new HRU specific slopes.

The biological mixing parameter (BIOMIX) for corn silage was reduced relative to BIOMIX for all other land uses (all of which were assigned the default BIOMIX value of 0.20) since corn silage is the only land use that is tilled in the model. Tilled land has less biological activity, and thus less biological mixing, compared with land that is not tilled.

The USLE equation support practice factor (USLE_P) in SWAT is used to represent the impact of various farming management practices on MUSLE estimates of HRU sediment erosion. USLE_P values are reduced from the SWAT default of 1.0 for corn silage and hay HRUs to represent the fact that crops in the Cannonsville Basin, although not contoured, are also not managed as straight up-down slope farming (Karl Czymmek, Personal Communication). The USLE_P factor values for contoured crops on various slopes can be found in Table 20.4 of Neitsch et al. (2001b). To account for some erosion

control, the USLE_P factors for corn silage and hay HRUs were set to values between 1.0 (which represents no erosion control) and the value for contoured fields of the specific HRU slope. The corn silage USLE_P factors were set as the sum of the value from Table 20.4 of Neitsch et al. (2001b) and 50% of the range between this value and 1. For example, if the value in Table 20.4 of Neitsch et al. (2001b) was 0.6 for a corn silage HRU, corresponding to a slope of between 9 and 12 percent, then the USLE_P value for that HRU would be input as 0.6 + (1-0.6)*0.5 = 0.80. Similarly, the USLE_P factor for hay was set as the sum of the value from Table 20.4 of Neitsch et al. (2001b) and 75% of the range between this value and 1. Corn silage and hay USLE_P factors are adjusted differently to account for the fact that corn silage fields are more intensively managed than hay fields. In other words, more erosion control practices are followed on corn silage fields.

The concentration of sediment (or TSS) in lateral and groundwater flow (LAT_SED) is input as a constant parameter in the SWAT2000 model. Data in Longabucco and Rafferty (1998) show that in periods when the flow at Beerston would have been composed almost entirely of lateral and groundwater flow (i.e. non-event period) the flow-weighted average TSS concentration in the early 1990s was 5.3 mg/L. A significant portion of this average measured TSS concentration at Beerston would have been derived from sources other than groundwater or lateral flow. In particular, sources would have included point sources and channel bed and bank erosion. Therefore, a value of 5.3 mg/L is taken as an upper bound for the LAT_SED parameter and a calibrated LAT_SED value is determined in Section 6.3.3.

The phosphorus availability index (PSP) parameter in SWAT governs the equilibration of soil phosphorus between the solution and active mineral pool and also functions to initialize mineral P levels in the basin soils. The probable range of this parameter value for the Cannonsville Basin was derived from data in the basin as 0.16 to 0.42 in Section 4.9.1.3. Given that this parameter controls initial soil P levels and influences the flux of phosphorus between soil P pools, calibrating this parameter to achieve desirable results would be particularly difficult. Therefore, a value of 0.25 was assumed because it resulted in reasonable levels of total soil P and because it is close to the midpoint of the estimated range.

6.3.2.3 Miscellaneous Data-Driven Parameters

After the original AVSWAT model default ASCII input files were created it was determined that a few of the default inputs were not representative of the conditions or physical characteristics of the basin and therefore required modifications. These modifications were generally completed manually outside of AVSWAT. The modifications are as follows:

1. The default AVSWAT routing file required a new routing file to be created in order to route the tributaries into the mainstem WBDR in the correct locations.
2. Subbasin elevations were modified. AVSWAT subbasin elevation defaults were assigned as the elevation of the subbasin centroid. Given the variations in subbasin elevations it was assumed that average subbasin elevations were more representative inputs.

3. Average historical monthly minimum and maximum temperatures for Walton and Delhi were substituted for the default temperatures in the weather generator input files. These monthly temperature inputs impact soil temperature calculations.
4. The harvest rates for timothy and pasture in Table 4.7.4 were assigned to approximate estimated basin-wide local feed for cattle. Calculations outlined in the Appendix (Section 10.4.2) show how these local cattle feed consumption rates were determined.

6.3.3 Performance Optimization Parameter Modifications

The performance optimization parameters in the model were generally calibrated to measured hydrology, sediment and phosphorus data in that order. However, since many of these parameters impacted all three of these modeled quantities, the latter stages of the calibration involved modifying parameters that often influenced all three quantities. The performance optimization parameters are summarized in Table 6.3.5, Table 6.3.6 and Table 6.3.7 on the following pages and are combined based on the modeled quantity the parameters affected most directly. All performance optimization parameters will generally be referred to by their short variable names for the remainder of this document. However, their full descriptive names are given in Table 6.3.5, Table 6.3.6 and Table 6.3.7. All performance optimization parameters are modified within their acceptable ranges as found in either the model user's manual (Neitsch et al. 2001b) or in the literature (see Benaman (2002) for a summary).

It is noteworthy, that in some cases, it would be reasonable in future modeling to consider some of the data-driven parameters as performance optimization parameters with a reduced range that encompasses the point estimates of the data-driven parameters given Section 6.3.2 and even some of those in Section 4. In fact, this approach would likely result in at least a small improvement in model performance. However, given the high number of parameters available for calibration in SWAT2000, reclassifying some of the data-driven parameters as optimization parameters should only be attempted if some type of automatic calibration algorithm is used to optimize model parameters.

6.3.3.1 Hydrology Performance Optimization Parameters

Hydrologic performance optimization parameters are defined and their final calibrated values are given in Table 6.3.5. Parameter modifications were either a constant value input across the basin or, for parameters whose base or initial value varied spatially, the base parameter values were modified by a multiplicative factor that was constant across the basin. This second type of parameter adjustment retained the relative differences in values of the parameter across the basin.

Calibration to estimated baseflow of measured Walton flows required the modification of four groundwater parameters. The groundwater parameters in Table 6.3.5 (BF_ALPHA, GW_DELAY, GWQMN and REVAPMN) all proved to be sensitive parameters with respect to baseflow. However, even with the calibration of these four parameters, baseflow predictions were still poor until implementing the source code modification for soil water excess in frozen soils (Section 5.4).

Table 6.3.5. Performance optimization calibrated hydrology parameters.

Parameter (SWAT variable name)	Change	Final Value (Default Value)
SCS Curve Number (CN2) multiplicative adjustment factor	Reduced base CN2s outlined in Section 6.3.2.2 by 20% (relative to their base value)	Varies by HRU
Manning's n for overland flow (OV_N)	Constant value for all land uses due to extremely short model simulated time of concentrations for majority of HRUs	0.20 (varied by land use)
Snowpack temperature at which snowmelt begins (SMTMP)	Increased in order to 'delay' snowmelt until warmer temperatures persisted	1.75 °C (0.5)
Melt factor for snow on June 21 (SMFMX)	Decreased to reduce snowmelt rate	3.0 mm H2O/ºC-day (4.5)
Melt factor for snow on December 21 (SMFMN)	Decreased to reduce snowmelt rate and so that winter snowmelt rate < spring melt rate	2.9 mm H2O/ºC-day (4.5)
Surface runoff lag coefficient (SURLAG)	Reduced so that some portion of surface runoff is lagged one day before reaching the channel	1.0 (4.0)
Lateral flow travel time (LAT_TIME)	Fixed for all HRUs so that lateral flow lag time was greater than surface runoff lag time	4.0 days (calculated by model and varied by HRU)
Available water capacity for all soils (AWC)[1]	Increased base values (data-driven) by 70% for layer 1 inputs[2] & 30% for all other layers so soil could hold more water; change related to decrease in rock fragment volume in Table 6.3.6	Varies by HRU soil type
Soil evaporation compensation factor (ESCO)[1]	Reduced to allow more evaporation from lower soil layers	0.70 (0.95)
Maximum potential leaf area index (BLAI)[1]	Increased to simulate more plant uptake of soil water	All forest: 7 (5) TIM*[3]: 6 (4) CSIL: 6 (4)
Potential Heat Units for selected land uses (PHU)[1]	Increased to ensure plant uptake of soil water into the fall	All forest: 3000 (1580) GRSH: 3000 (1600) TIMU: 3000 (1600)
Maximum Stomatal conductance (GSI)[1]	Increased to simulate more plant uptake of water during crop growth	CSIL: 0.0071 (0.0070) TIM*[3]: 0.0055 (0.0050) All units are m/s
Baseflow alpha factor (BF_ALPHA)[1]	Increased to simulate steeper hydrograph recession	0.15 (0.048)
Groundwater delay time (GW_DELAY)[1]	Decreased so that groundwater return flow occurs more quickly	10 days (31 days)
Threshold depth of water in shallow aquifer for return flow to occur (GWQMN)[1]	Increased to create groundwater storage capacity	90 mm (0 mm)
Threshold depth of water in shallow aquifer for 'revap' to occur (REVAPMN)[1]	Increased over GWQMN so that groundwater return flow occurs before 'revap' (transfer of groundwater to upper soil layers)	100 mm (1.0 mm)

1. Parameters modified to improve model predictions of groundwater and summer low flows.
2. Layer 1 soil inputs are applied to approximately the top 15 cm of soil, including the 1 cm surface layer.
3. TIM* is all timothy – GRSH, TIMC, TIMU and TIMP.

The surface runoff parameters adjusted to replicate daily flows at Walton were OV_N, CN2 multiplicative factor, SURLAG, LAT_TIME and the snowmelt parameters SMTMP, SMFMX and SMFMN. After fixing the OV_N parameter in SWAT at a constant value across the basin, the CN2 multiplicative factor, SURLAG and LAT_TIME parameters were adjusted so that peak flows were reasonably well predicted. Although adjustments of these three factors were made to improve daily flow predictions year round, more importance was attached to prediction of late spring, summer and fall flows because the snowmelt parameters could be adjusted to correct winter and late spring flow predictions. The base CN2 values in Table 6.3.4 proved to be much too high for the basin as peak daily flows for spring, summer and fall were almost always over-predicted by 100% or more and the hydrograph recessions were much too steep. Even after reducing CN2 values by 20% for all HRUs, which was deemed to be the maximum reduction due to impacts on winter peak flows, the SURLAG parameter had to be reduced to its minimum value of 1.0 in order to force the model to simulate a small lag in the delivery of surface runoff to the channel. The resulting area-weighted average CN2 value for the entire basin after base CN2 values were reduced by 20% is 57.0. The resulting basin-wide area-weighted average fraction of surface runoff reaching the channel on the day it was generated using a SURLAG value of 1.0 is 0.61. More than 50% of the basin is simulated to have a fraction of surface runoff reaching the channel on the day it was generated that is greater than 0.73. Simulating a lag in delivery to the channel for less than half of the surface runoff is deemed reasonable considering that rainfall events that occur or extend late into the evening would no doubt generate some surface runoff that was not delivered to the channel until the next calendar day. The default SWAT approach is to calculate LAT_TIME for each HRU rather than assigning a constant value. However, this approach, combined with a SURLAG parameter setting of 1.0, resulted in lateral flow reaching the channel more quickly than surface runoff for approximately 1% of the basin area. To correct this physically improbable result, and improve daily flow prediction at Walton, a constant value of LAT_TIME equal to 4.0 was used. When compared with the default LAT_TIME approach, a value of 4.0 for LAT_TIME functioned to slightly reduce the area-weighted average lag time for lateral flow delivery to the stream.

Late summer and fall flows tended to be over-predicted under the above parameter settings. To correct this, it was determined that more soil storage of water and more end of season evapotranspiration of soil water was required. The parameters soil water storage parameter AWC and the evapotranspiration parameters ESCO, BLAI, PHU and GSI were adjusted as described in Table 6.3.5 to reduce late summer and fall flows. In addition, out of the three available potential evapotranspiration calculation methods in SWAT2000, the Priestly-Taylor method was selected because it resulted in the highest late summer/early fall evapotranspiration rates.

Finally, the snowmelt parameters in Table 6.3.5 were adjusted to improve winter flow predictions. The SMTMP parameter change improved the timing of simulated snowmelt. The SMFMX and SMFMN parameter changes improved the peak flow predictions. During the calibration of winter flows, the four winter flow related source code modifications presented in Sections 5.3, 5.4, 5.5 and 5.6 were identified and the model

deficiency with respect to correctly classifying precipitation type (Section 4.4.1.2) was also identified. The most significant source code modification was the change for soil water excess in frozen soils (Section 5.4).

In addition to the parameter changes listed above, the iterative analysis of model prediction errors throughout the flow calibration process suggested a detailed re-examination of climate input data quality. This re-examination resulted in the corrections and modifications to the climate inputs outlined in Section 4.4.1. It is important to note that climate data input changes were only implemented when data or information other than measured flows across the basin suggested such changes were warranted. In other words, in the absence of supporting data, climate inputs were not modified to simply improve prediction of flows at Walton or any other USGS flow station.

6.3.3.2 Sediment Performance Optimization Parameters

SWAT sediment performance optimization parameters are related to erosion from the land surface on each HRU, the TSS concentration in lateral/groundwater flow or channel sediment erosion and deposition processes. Table 6.3.6 outlines and defines all modified sediment performance optimization parameters. TSS calibration involved a number of steps. First, the HRU sediment generation parameters APM and ROCK were adjusted to predict late spring, summer and early fall TSS loads. In the second TSS calibration step, the winter TSS predictions were increased to better replicate measured TSS loads by iteratively adjusting the coefficients A and B in the new MUSLE snow cover adjustment equation described in Section 5.7 and given again in Table 6.3.6. The third step in TSS calibration involved assigning values to the LAT_SED and channel erosion parameters (CH_EROD and CH_COV) such that reasonably small amounts of TSS were derived from these sources. Lastly, the APM factor was refined so that the average simulated TSS load approximated the average measured TSS load.

Under default values in Table 6.3.6, TSS loads were greatly over-predicted. Therefore, the APM parameter was reduced significantly to reduce TSS loads from all HRUs. In addition, the ROCK parameter representing the percentage rock by volume in the first and second soil layer was reduced basin-wide by approximately 50% relative to the base ROCK values determined for each HRU as described in Section 4.3. The change in ROCK was assumed related to the AWC parameter change in Table 6.3.5. Note that the AWC parameter is for the entire soil layer including rock fragments. AWC and ROCK within a soil layer are inversely correlated since they both occupy soil volume. Therefore, increasing AWC of soil layer 1 should be accompanied by a decrease in rock fragment volume (ROCK). The base rock fragment parameter values were often close to 40% for soil layer 1 in many soils. Although an average rock fragment content of 40% in the plough layer (approximately the top 15 cm) may be reasonable, the rock fragment content in the 1 cm surface layer of soil should be reduced given the increased organic matter content in the surface layer. Therefore, independent of the AWC adjustment, the base values for rock fragment volume of the first soil layer were deemed to require some significant reduction since they were also applied to the 1 cm surface layer of soil. Base rock fragment volumes input for soil layer 2 and below were not changed because they have no impact on model calculations.

Table 6.3.6. Performance optimization calibrated sediment parameters.

Parameter (SWAT variable name)	Change	Final Value (Default Value)
Peak rate adjustment factor for sediment routing (APM)[1]	Decreased in order to reduce total sediment loading from HRUs	0.55 (1.0)
Soil layer 1[2] rock fragments as a % of soil layer volume (ROCK)[1]	Decreased base values (data-driven) by approximately 50% in association with AWC increase in layer 1 in Table 6.3.5	Varies by HRU soil type
MUSLE equation erosion predictions[1]	Estimate coefficients A and B in the equation in Section 5.7: $MUSLE_i = MUSLE_i` * \min (A, [SurQ_i / Sno_i]^B)$	A = 0.95 B = 0.25
Sediment concentration in lateral and groundwater flow (LAT_SED)	Increased within derived bounds in Table 6.3.3 so that lateral flow and groundwater contain low levels of TSS	2.5 mg/L for all HRUs (0.0)
Channel erodibility factor (CH_EROD)[3]	Increased to allow for small amount of stream channel erosion	0.0003 (0.0)
Channel cover factor (CH_COV)[3]	Increased to allow for small amount of stream channel erosion	1.0 (0.0)

1. Parameters related to HRU sediment erosion.
2. Layer 1 soil inputs are applied to approximately the top 15 cm of soil, including the 1 cm surface layer.
3. Parameters related to stream channel erosion.

Quantitative data were unavailable to estimate the significance of channel bed and bank erosion. However, Pat Bishop (Personal Communication) reported that some erosion of the stream channel does occur in the WBDR and estimated that it most likely accounts for no more than 10% of the total annual sediment load. Therefore, reasonable channel sediment parameters were deemed to be those that erode the stream channel (as opposed to net sediment deposition in the channel) such that no more than 10% of the average annual total annual sediment load was produced from stream channel erosion. The calibration effort showed that good model performance could be achieved by modifying only the two channel erosion parameters that are outlined in Table 6.3.6. The product of the parameters CH_EROD and CH_COV (which is between 0 and 1) determines how protected the stream channel is from erosion. A value of 0 indicates that the stream channel is never eroded while a value of one indicates that there is no channel protection. Since the product of these parameters controls channel erosion, an infinite number of CH_EROD and CH_COV parameter combinations exist that would produce exactly the same simulated TSS loads. The product of CH_EROD (set at 0.0003) and CH_COV (set at 1.0) equal to 0.0003 was determined to produce good model performance while simulating reasonable amounts of stream channel erosion. Other channel erosion parameters that were considered as performance optimization parameters but were

ultimately not modified from their default values because their modification did not result in significantly better model predictions were the SPCON, SPEXP and PRF parameters.

A value of 2.5 mg/L for the LAT_SED parameter was selected since it was approximately the midpoint of the bounds determined for LAT_SED in Section 6.3.2.2.

6.3.3.3 Phosphorus Performance Optimization Parameters

Phosphorus performance optimization parameters modified from their default values are outlined in Table 6.3.7. Adjustment of the CMN parameter resulted in a small improvement in model predictions of total, total dissolved and particulate phosphorus. The ERORGP default is to calculate the phosphorus enrichment ratio for each storm. However, a constant ERORGP value of 2.0 improved model P predictions.

The SWAT2000 option to simulate instream nutrient reactions was investigated thoroughly in the calibration process. However, the simulation of instream reactions under various instream reaction parameter sets all resulted in large increases in total P loading at Beerston. The simulation of instream reactions should function to transform P between dissolved and particulate forms without significant changes in total P loading. Since this model behaviour was unexpected and could not be explained, instream reactions were not used in this study.

Three additional phosphorus parameters were considered as performance optimization parameters. Adjustments to the SWAT parameters PHOSKD and PPERCO, which relate to the calculation of dissolved phosphorus concentrations in surface runoff and percolate, respectively, and the UBP parameter, which defines the depth in which plant roots extract the phosphorus from soils, were investigated but were ultimately not modified from their default values because their modification did not result in significantly better model predictions.

Table 6.3.7. Performance optimization calibrated phosphorus parameters.

Parameter (SWAT variable name)	**Change**	**Final Value (Default Value)**
Rate factor for humus mineralization of active organic nutrients (CMN)	Increased to increase rate of P mineralization in the soil	0.001 (0.0003)
Phosphorus enrichment ratio for loading with sediment (ERORGP)	Changed so that the model enrichment ratio did not change with storm magnitude	2.0 (0.0)

6.3.4 Model Calibration Summary

Any future modeling effort with SWAT2000 attempting to replicate some or all of the model inputs and parameters used in this report should refer to the inputs described in Section 4, the data-driven and performance optimization parameters outlined above in Section 6.3 and the changes to the model source code outlined in Section 5.

6.4 Calibration Results

Unless otherwise noted, all calibration (and validation) performance statistics are calculated for the entire period of time over which measured data were available. In other words, performance is not always assessed separately for season or for each year unless noted below. In some cases below, when seasonal performance statistics are reported, performance is analyzed for what is referred to as the summer and winter seasons. The summer season refers to the period between May 1 and Oct. 31 while winter season refers to the period between Nov. 1 and Apr. 30.

6.4.1 Hydrology Calibration Results

Baseflow predictions were evaluated on a seasonal and annual basis for the Walton USGS station only. The baseflow filter program by Arnold and Allen (1999) generates a range of predicted baseflow volumes. On an annual basis, the measured flow at Walton is estimated as 42% to 66% baseflow over the calibration period. In comparison, the simulated flow at Walton is estimated as 42% to 68% baseflow over the calibration period. Therefore, the calibrated model was deemed to generate acceptable predictions of baseflow on an annual basis.

The baseflow estimates from the Arnold and Allen (1999) program were averaged for each month over the calibration to determine seasonal performance and the average baseflow component of streamflow for each month was calculated. Again, since a range of baseflow estimates are estimated from the baseflow filter program, the seasonal baseflow analysis compared the range of the average monthly baseflow component of streamflow. This seasonal comparison is presented in Figure 6.4.1 and shows good seasonal agreement of estimated baseflows between the simulated and measured flows.

Calibrated model predictive performance for daily flows is summarized in Table 6.4.1 under the calibration parameter values outlined in Section 6.3 and the changes to the model source code in outlined in Section 5. Predictive performance is assessed at all six USGS gauging stations although the calibration focused mainly on improving predicted flows at Walton.

Simulated results were matched to the six USGS gauge stations by the closest SWAT subbasin outlet location. Since not all subbasin outlets were located at the exact location of the USGS gauging stations, some measured flows had to be slightly prorated to the nearest subbasin outlet based on the change in drainage area between the USGS gauge and the subbasin outlet. Thus, measured flows for the WBDR at Walton, WBDR at Delhi, Little Delaware and Trout Creek, were multiplied by a factor of 1.003, 1.005, 1.052 and 1.025, respectively, in order to compare with simulated flows. All reported measured flow statistics in the calibration and validation apply to the prorated measured flows.

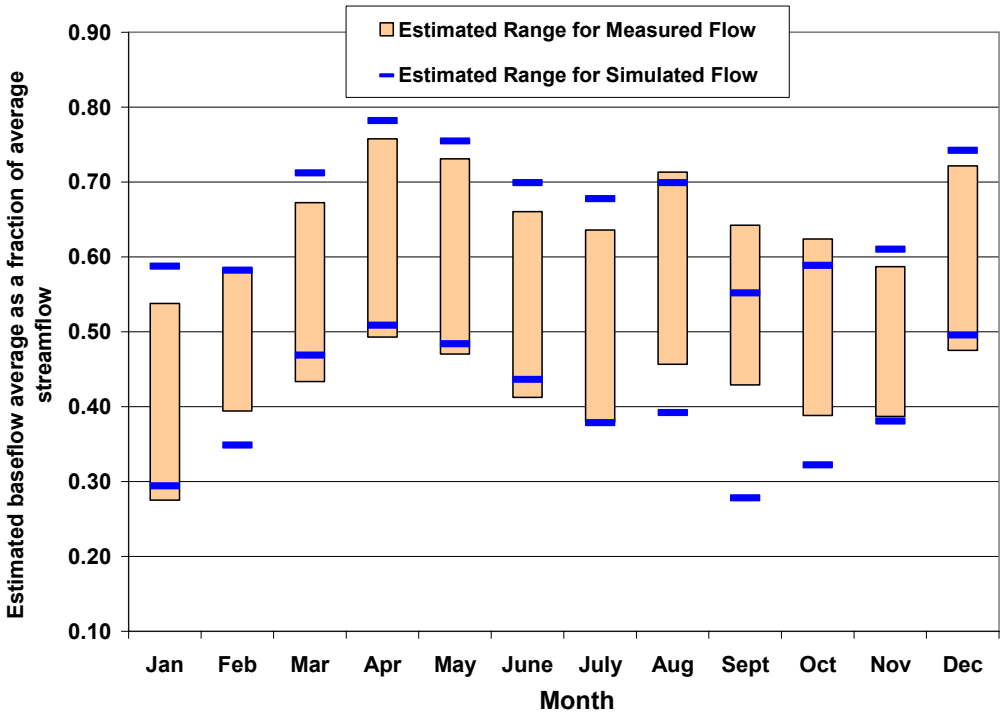

Figure 6.4.1. Comparison of average monthly baseflows as a fraction of streamflow between simulated and measured flows at Walton for period Jan. 1994 to Sept. 2000.

Table 6.4.1. Daily hydrology calibration results at six USGS gauge stations for the period 1994 to 2000 (see Figure 6.1.1 for gauge locations).

Location (USGS gauge #)	**Days of Measured Data**	**Mean Measured Data[A] (m^3/s)**	**Mean Simulated Results[A] (m^3/s)**	**Percent Difference between Measured & Simulated[B]**	**r^2**	**E_{NS}**
WBDR at Walton (01423000)	2465	17.8	18.0	1.0	0.80	0.79
WBDR at Delhi (01421900)	1396	7.1	6.9	-2.7	0.72	0.72
Little Delaware (01422500)	1349	2.8	2.6	-7.3	0.78	0.78
East Brook (01422747)	731	1.1	1.3	15.7	0.62	0.57
Trout Creek (0142400103)	1396	1.1	1.0	-2.3	0.64	0.63
Town Brook (01421618)	1096	0.8	0.8	-5.8	0.59	0.59

A) Calculated based only on the days with measured data for each station.
B) Calculated with higher precision than the reported mean measured data and mean simulated results.

The calibration results demonstrate that the model predicts daily flows across the basin quite well. Flow statistics are best at the three largest USGS stations (Walton, Delhi and Little Delaware). Flow predictions at Town Brook and East Brook are least accurate. However, this is probably due to their small size and could also be a result of the shorter period of record (by at least a year) when compared with the other stations. Given that the calibration focused on Walton and no attempt to independently calibrate to flow data at other USGS stations was attempted, the model performance at all locations is deemed acceptable.

A visual comparison of the daily simulated and measured flow time series at Walton is presented in Figure 6.4.2 for select water years. Water years 1995, 1997 and 1998 are presented in Figure 6.4.2 because they demonstrate the range of annual model performance and typical model inaccuracies. The annual model performance is worst during water year 1995 and best in water year 1997. Time series plots for flow at Walton during the other water years in the calibration period are shown in the Appendix (Figure 10.5.2). Table 6.4.2 provides various performance statistics for the water years during the calibration period.

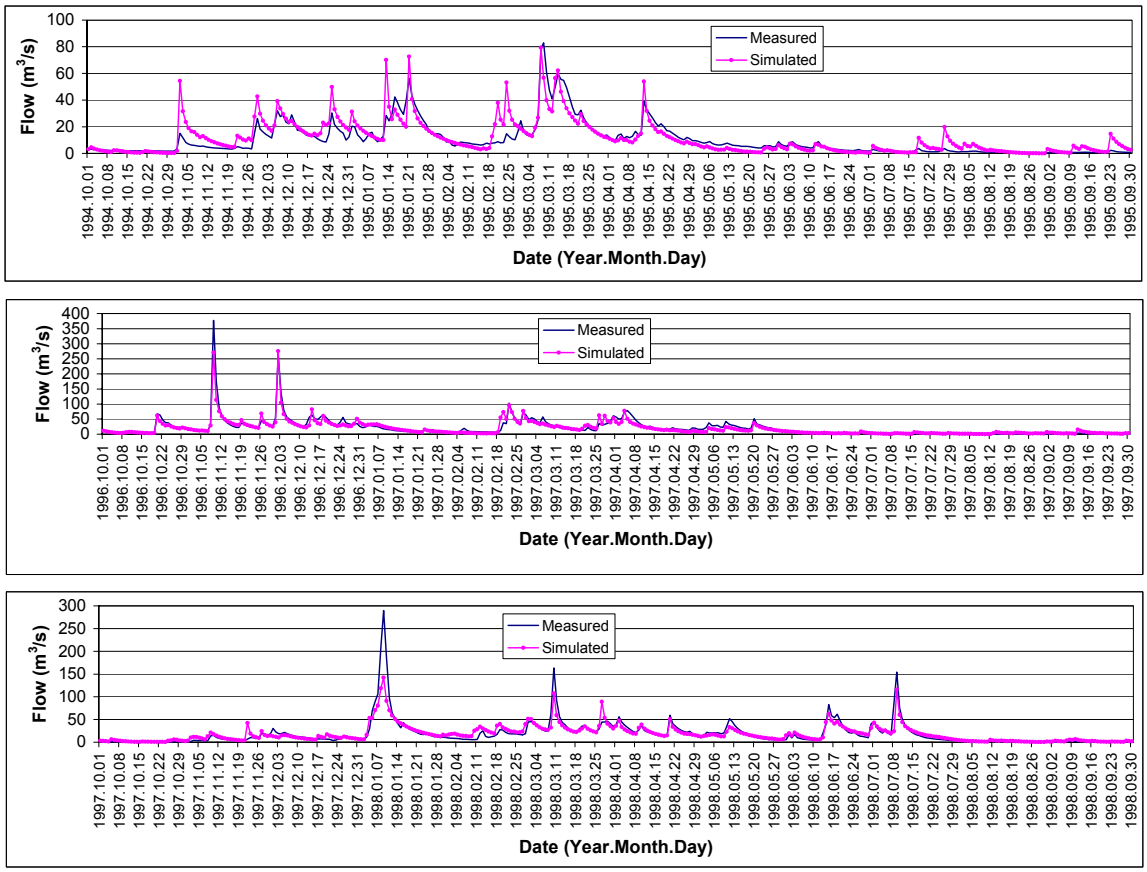

Figure 6.4.2. Time series of measured and simulated daily hydrology calibration results at Walton for water years 1995, 1997 and 1998.

Table 6.4.2. Daily hydrology calibration results by water year for the Walton USGS station for the period 1994 to 2000.

Water Year	Percent Difference between Measured & Simulated	r^2
1994[1]	5.6	0.81
1995	10.0	0.74
1996	2.6	0.75
1997	-8.4	0.91
1998	-2.7	0.84
1999	8.2	0.74
2000	0.2	0.76

1. A partial water year: January 1 to September 30, 1994.

Results in Figure 6.4.2 show very good general agreement between the measured and simulated trends in Walton flows. There are some notable periods and events in which model predictions are not in good agreement with the measured flows. Summer peak flows tend to be over-predicted by the model as shown for water year 1995 at Walton in Figure 6.4.2. Note that for summer flows in 1997 and late summer flows in 1998 are also over-predicted even though it is not obvious in Figure 6.4.2 because the flow (y-axis) scale is compressed in these water years. Considering all summer seasons over the calibration period together, the model over-predicted Walton flow by 4.2% and the daily r^2 coefficient was 0.82.

An examination of the entire calibration period (e.g. see Figure 10.5.2 also) shows that the model usually under-predicts the largest flow events. This result is easier to visualize with reference to the scatter plot of measured versus simulated flows at Walton over the calibration period in the Appendix (see Figure 10.5.1). For example, Figure 10.5.1 shows that of the 18 daily flows over 130 m^3/s, 17 were under-predicted and of these, 16 were under-predicted by at least 18%. The majority of the large flow event under-predictions are winter related flow events such as the events in January 1996, November 1996, January and March 1998, January 1999 and March 2000. Considering all winter seasons over the calibration period all together, the model under-predicted Walton flow by 0.4% and the daily r^2 coefficient was 0.77.

Many studies often present hydrology calibration on a monthly time step. Therefore, model performance statistics are also calculated for all six USGS flow stations on a monthly time step and are provided in Table 6.4.3. In addition, plots of monthly average measured and simulated flows at all USGS stations are provided in the Appendix in Figure 10.5.3 and Figure 10.5.4.

Over the calibration period, the simulated basin-wide water balance components on an annual average basis are as follows:

- 1194 mm of precipitation (250 mm of which was snow).
- 538 mm of evapotranspiration.
- 653 mm of water yield (i.e. streamflow leaving the basin) made up of:
    - 97 mm of surface runoff (15% of water yield).

- 306 mm of lateral flow (47% of water yield).
- 253 mm of groundwater flow (39% of water yield).

Not included in the above simulated water balance are the very minimal losses of water to deep aquifer percolation, sublimation and channel transmission, which total less than 1% of the annual precipitation.

Table 6.4.3. Monthly hydrology calibration results at six USGS gauge stations for the period 1994 to 2000 (see Figure 6.1.1 for gauge locations).

Location (USGS gauge #)	Months of Measured Data	r^2	E_{NS}
WBDR at Walton (01423000)	81	0.89	0.89
WBDR at Delhi (01421900)	45	0.87	0.86
Little Delaware (01422500)	44	0.90	0.89
East Brook (01422747)	24	0.82	0.76
Trout Creek (0142400103)	45	0.81	0.81
Town Brook (01421618)	36	0.81	0.81

The model source code change for handling soil water excess in frozen soils outlined in Section 5.4 proved to be critical in achieving the above model calibration results. Model predictions under the exact same model conditions as the calibrated model, except for the model source code change in Section 5.4, showed dramatically different simulated flows and major reductions in model performance statistics. These included:

- Seasonal simulated baseflows that were much worse in comparison with simulated seasonal baseflows in Figure 6.4.1.
- Seasonal flows were much worse as summer flows were over-predicted by 47% while winter flows were under-predicted by 16%.
- The surface runoff component of basin-wide water yield increased from 15% in the calibrated model to 41% of water yield.
- The r^2 and E_{NS} model performance statistics for Walton daily flows were reduced to 0.59 and 0.50, respectively.

6.4.2 Sediment Calibration Results

The January 19-20, 1996 flood event was the largest streamflow on record and the estimated return period of the event is 70 years for the Walton USGS gauge (Lumia 1998). In addition the estimated return period for the event as measured at the Delhi and Little Delaware USGS gauges were both more than 100 years (Lumia 1998). As a result, this event caused widespread overbank flow and initiated transport of huge amounts of floodplain phosphorus and sediment. Longabucco and Rafferty (1998) report that the two-day event accounted for approximately 75% of the TSS and particulate phosphorus loads recorded for the entire 1996 year. Since SWAT is not designed to simulate such an extreme event the calibration effort focused on the improving model performance measures that did not include measured and simulated TSS and phosphorus loads during January 1996. This approach is consistent with Schneiderman et al. (1998) (although not

as extreme) whose approach to handling this event involved eliminating comparisons between model and data for January, February and March of 1996. Model calibration performance measures for TSS and phosphorus are reported in the following sections with and without January 1996 results.

Section 5.7 describes the change made to the MUSLE snow cover erosion prediction adjustment. Before evaluating the cumulative model performance under the final calibrated parameter set it is useful to show the improvement in model performance achieved directly from the model modification described in Section 5.7. Figure 6.4.3 shows the impact of this change by comparing model simulation results against measured data at Beerston for two simulation runs in which the only difference was the change to the MUSLE snow cover adjustment. It should be noted that although the simulation runs plotted in Figure 6.4.3 do not use exactly the same calibrated parameter values as outlined in Section 6.3, the parameter values are quite close to the final set of calibrated parameters. The purpose of Figure 6.4.3 is to point out the model improvement during most winter months with a high TSS load measured at Beerston. For example, that model accuracy was most significantly improved in Mar-94, Apr-94, Jan-96, Dec-96, Jan-98, Mar-98, Jan-99 and Feb-00. In contrast, model accuracy was significantly decreased in only a few months (e.g. Feb-94, Feb-96, Mar-96, Mar-99). The model inaccuracy in these four months is at least partially due to the fact that the highest and/or second highest daily peak flows in these months are over-predicted by at least 40%. Therefore, the over-predictions in these four months are in fact expected and can only be partially attributed to the new MUSLE snow cover adjustment equation. The modification to the MUSLE snow cover adjustment equation was deemed a good approach to improve SWAT sediment load predictions and was utilized for all model simulations.

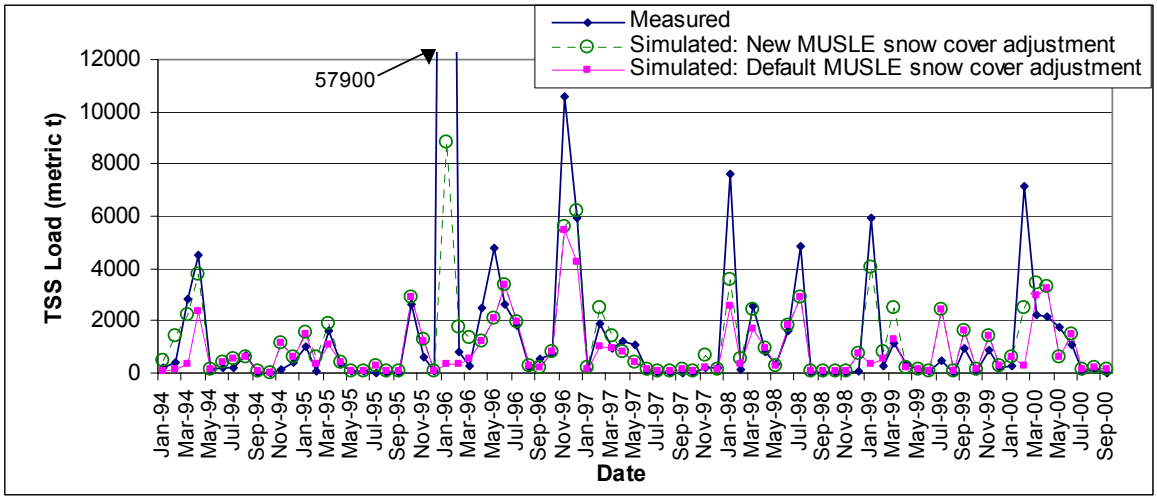

Figure 6.4.3. Change in TSS model performance at Beerston when the MUSLE snow cover erosion prediction adjustment is modified to better represent months with snowmelt.

Monthly model predictions for TSS under the calibration parameter values outlined in Section 6.3 are summarized in Table 6.4.4 and plotted in Figure 6.4.4. Measured TSS taken at Beerston and Town Brook corresponded exactly to SWAT subbasin outlets and therefore simulated TSS and measured TSS could be compared without further adjustments. The model performance measures for Beerston in Table 6.4.4 are given with and without January 1996 to show the impact this event has on the values of the performance measures.

Table 6.4.4. Monthly sediment calibration results at continuous water quality monitoring stations for the period 1994 to 2000 (see Figure 6.1.1 for locations).

Location	**Months of Measured Data**	**Mean Measured Data[A] (metric t)**	**Mean Simulated Results[A] (metric t)**	**Percent Difference between Measured & Simulated**	**r^2**	**E_{NS}**
Beerston: *without* Jan. 1996	80	1204	1177	-2.2	0.71	0.66
Beerston *including* Jan. 1996	81	1926	1248	-35.2	0.42	0.22
Town Brook	24	194	93	-52.2	0.43	0.27

A) Calculated based only on the days with measured data for each station.

Figure 6.4.4 shows the monthly time series of simulated and measured TSS at the Beerston and Town Brook NYSDEC continuous water quality monitoring stations. Although the model predicted sediment loads are less accurate than the corresponding flow predictions at Beerston and Town Brook, the model still replicates the trend in measured data fairly well. This is especially true at Beerston and is also evident at Town Brook if the TSS loading in Sept-99 is ignored. Results in Table 6.4.4 suggest that the model performance at Town Brook is degraded relative to Beerston performance. Even when model performance at Town Brook and Beerston are compared over the same time period (Oct-98 through Sept-00), the r^2 at Beerston of 0.54 is still significantly higher than the r^2 at Town Brook (0.43) and Beerston TSS loads are predicted within 10% of measured TSS loads. Degraded model performance for TSS and phosphorus is expected at Town Brook given that daily flow predictions at Town Brook were substantially worse than Beerston daily flow predictions. Town Brook TSS predictions could be improved in the future by independently calibrating the Town Brook subbasin to the measured data.

A comparison of Town Brook and Beerston TSS loading for July through September of 1999 in Figure 6.4.4 suggests the measured data appear inconsistent. These three months are the only months over the Oct-98 through Sept-00 period that Town Brook TSS loading is measured greater than Beerston TSS loading. This result is unexpected because the Town Brook drainage area is only 4% of the Beerston drainage area. In the remaining 21 months with Town Brook monitoring data, measured Town Brook TSS loading accounts for 8% of the measured Beerston TSS loading. In comparison, for July

through September of 1999, measured Town Brook TSS loading is 197% of the measured Beerston TSS loading. This apparent anomaly suggests that either A) most of the TSS load from Town Brook during July through September of 1999 period did not reach Beerston and was deposited in the mainstem WBDR before Beerston or B) the measured data over this period at Beerston and/or Town Brook are inaccurate. If case A above is true, the model is unable to simulate the proper magnitude of sediment deposition in the stream channel. Alternatively, if case B above is true, the sizable TSS loading prediction errors at Town Brook and/or Beerston for July and September of 1999 are actually not as severe as the current measured TSS data suggests. In the absence of additional data, it is impossible to determine if case A or B above is true. Therefore, the TSS loading prediction errors at Town Brook and Beerston for July and September of 1999 cannot currently be wholly attributed to the model and may in fact be due to errors in the measured data. The discussion in this paragraph also applies to Town Brook and Beerston total phosphorus loading data for July and September of 1999.

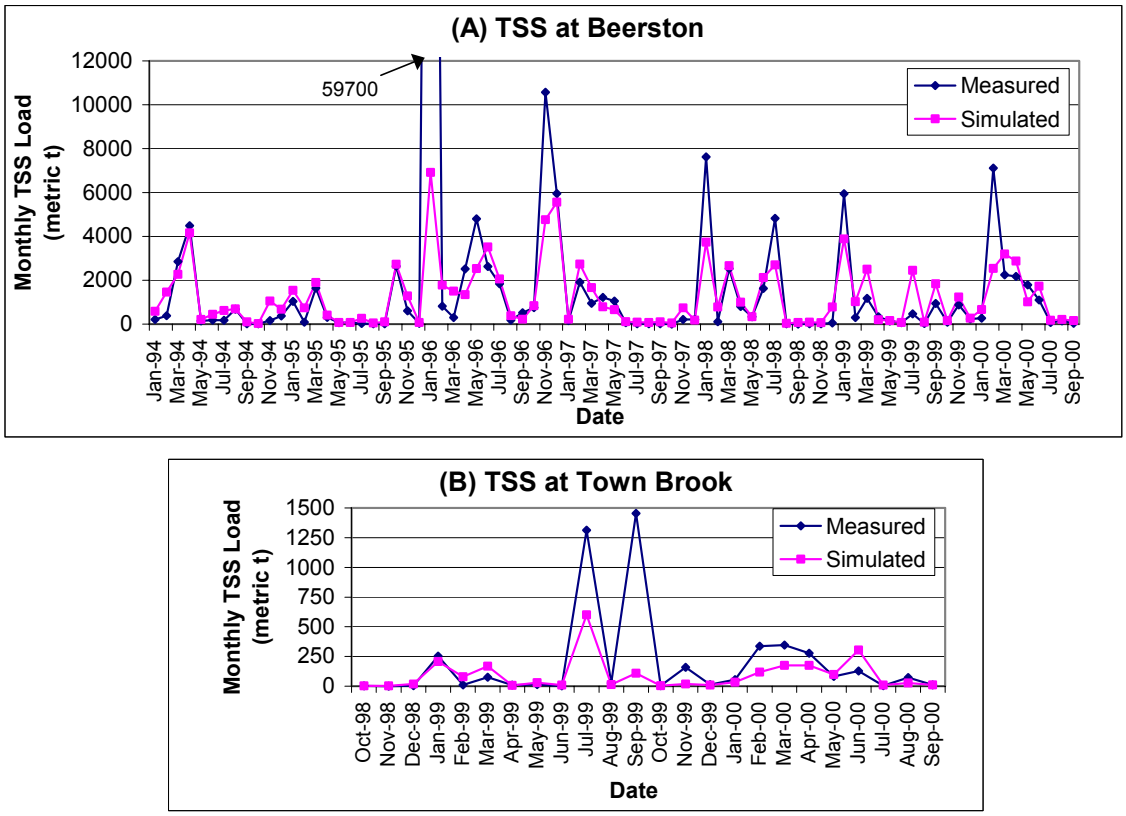

Figure 6.4.4. Time series of monthly measured and simulated sediment calibration results at A) Beerston and B) Town Brook water quality stations for the period 1994 to 2000. Please see the discussion in Section 6.4.2 regarding the extreme Jan. 96 event.

Consistent with hydrology results, Figure 6.4.4 demonstrates that at Beerston the model tends to under-predict TSS loading in months associated with snowmelt. The most severe errors in predicted TSS loads (e.g. Jan-96, May-96, Nov-96, Jan-98, July-98, Jan-99, July-99 and Feb-00) all occur in months where there are large predictive errors in the

peak daily flows. In fact, as discussed further in Section 7.3.3, the TSS predictive errors in each of these months are positively correlated with peak daily flow predictive errors of ±20% or worse. Therefore, errors in TSS predictions are largely a result of errors in peak flow prediction.

Considering all summer seasons over the calibration period together, the model over-predicted Beerston TSS by 6.7% and the monthly r^2 coefficient was 0.69. Considering all winter seasons over the calibration period all together, the model under-predicted Beerston TSS by 5.8% and the monthly r^2 coefficient was 0.71. Therefore, model performance levels were approximately the same in each season considered.

Although model TSS calibration performance focused on maximizing monthly r^2 and E_{NS} coefficients, the calibrated model TSS predictions compared well with measured TSS data on a daily time scale. Daily model performance measures for Beerston TSS over the calibration period (excluding January 1996) are r^2 and E_{NS} coefficients of 0.54 and 0.53, respectively. Similarly, the daily TSS r^2 and E_{NS} coefficients at Town Brook for the Oct. 1998 through Sept. 2000 period are 0.49 and 0.37, respectively.

The simulated TSS load is composed of sediments in lateral flow and groundwater, channel bed and bank derived sediments, point sources and soil erosion by surface runoff. Over the calibration period, the basin-wide TSS load is composed of:

- 10% from lateral flow and groundwater.
- 3% from channel bed and bank erosion.
- 1% from point sources.
- 86% from erosion of soils by surface runoff.

6.4.3 Temporal Phosphorus Calibration Results

Monthly model predictions for total dissolved, particulate and total phosphorus (TDP, PP, and total P, respectively) loads at Beerston and Town Brook under the calibration parameter values outlined in Section 6.3 are compared against measured monthly data in Figure 6.4.5 for Beerston and Figure 6.4.6 for Town Brook on the following pages. In addition, Table 6.4.5 summarizes the model performance measures for phosphorus at Beerston and Town Brook. Unless otherwise noted, results in Table 6.4.5 do not include January 1996.

Figure 6.4.5 shows the monthly time series of model predictions and measured data for total dissolved, particulate and total phosphorus at the Beerston NYSDEC continuous water quality monitoring station. Although some effort was made to calibrate to TDP and PP, based on Section 6.2.2 the main concern of the calibration effort was to match total P loads at Beerston. For example, while Beerston TDP loads are over-predicted by about 16% and PP loads are under-predicted by about 19%, the total P loads are predicted to within 6%. These relatively higher percent differences between measured and simulated TDP and PP loads are acceptable based on the discussion in Section 6.2.2 and the small percent difference between the measured and simulated total P load. Furthermore, no parameter set could be identified that resulted in further improvements

to the percent difference measures for TDP and PP. Overall, the trend in total P loading at Beerston is predicted fairly well by the model.

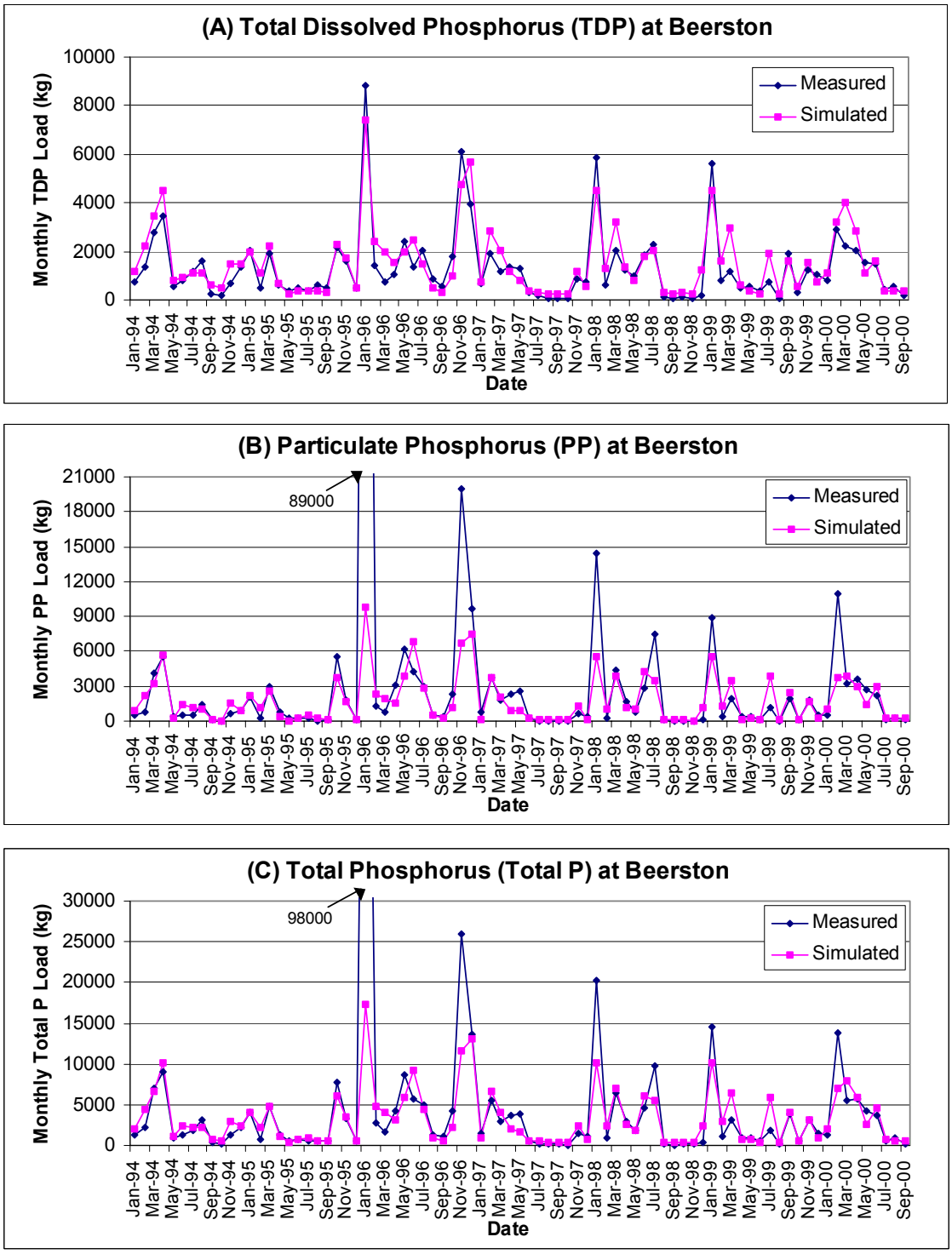

Figure 6.4.5. Time series of monthly measured and simulated A) TDP B) PP C) total P calibration results at Beerston water quality station for the period 1994 to 2000. Please see the discussion in Section 6.4.2 regarding the extreme Jan. 96 event.

Simulated total P loads in Figure 6.4.5 are generally under-predicted during peak snowmelt months. The under-predictions of total P during snowmelt months are consistent with the under-predictions of hydrology and sediment in those same months. The dependence of TDP and PP loads on each other and on flow (surface runoff) and TSS make interpreting all TDP and PP prediction inaccuracies difficult. Consistent with the largest observed TSS predictive errors, the largest errors in PP are positively correlated with peak daily flow predictive errors of ±20% or worse. Therefore, errors in PP predictions are largely a result of errors in peak flow prediction. To a lesser extent, the largest errors in TDP predictions also seem to be associated with errors in peak flow predictions. However, based on the discussion in Section 6.2.2, it is better to compare errors total P prediction to errors in peak daily flow predictions. For the most part, the largest errors in total P prediction occur in the same months and show the same pattern as the largest monthly TSS prediction errors identified in Section 6.4.2. The largest TSS and total P errors observed in the calibration are directly compared with peak daily flow errors in Section 7.3.3.

Results for Town Brook P predictions in Figure 6.4.6 are significantly worse in comparison to Beerston P predictions in Figure 6.4.5. These differences in performance are evident in the comparison of numerical model performance at each location in Table 6.4.5. Since the daily flow predictions at Town Brook are substantially less accurate than the daily flow predictions at Beerston, the model predictions for total P at Town Brook are expected to be worse than Beerston total P predictions. Although Figure 6.4.6 shows the trend in the measured total P data are somewhat replicated by the model, especially after Sept. 1999, the model seriously under-predicts the measured peak monthly total P loads in January, July and September of 1999. The biggest discrepancy between simulated and measured TDP, PP and total P occurs in September of 1999. As a result, over the entire period of record for Town Brook, total P loads are under-predicted by 49%.

It appears that P loads taken earlier in the monitoring history of the Town Brook water quality station may have been measured with significantly less accuracy than the later P loads – in which case model performance at Town Brook would be considered fairly good. In fact Pat Bishop (Personal Communication) confirmed that early in the Town Brook water quality sampling program, the sampling coverage of events was sometimes inadequate. For water year 1999, there are approximately 160 NYSDEC water quality samples used to estimate TSS and phosphorus loading at Town Brook. In comparison, for water year 2000, a total of almost 330 water quality samples (approximately 70% of which are NYSDEC samples and 30% USGS samples) are used to estimate TSS and phosphorus loading. Since the average sampling frequency used by NYSDEC for load estimation from October 1999 through September 2000 is more than double that of the previous water year, it is reasonable to expect that NYSDEC measured (or estimated) monthly loads for water year 2000 are significantly more accurate than those for water year 1999. Therefore, the improved model performance for water year 2000 (monthly r^2 coefficient of 0.61 and total P load under-predicted by 23%) over water year 1999 (e.g. compare two water years in Figure 6.4.6) may be an indication that overall Town Brook model predictions are acceptable.

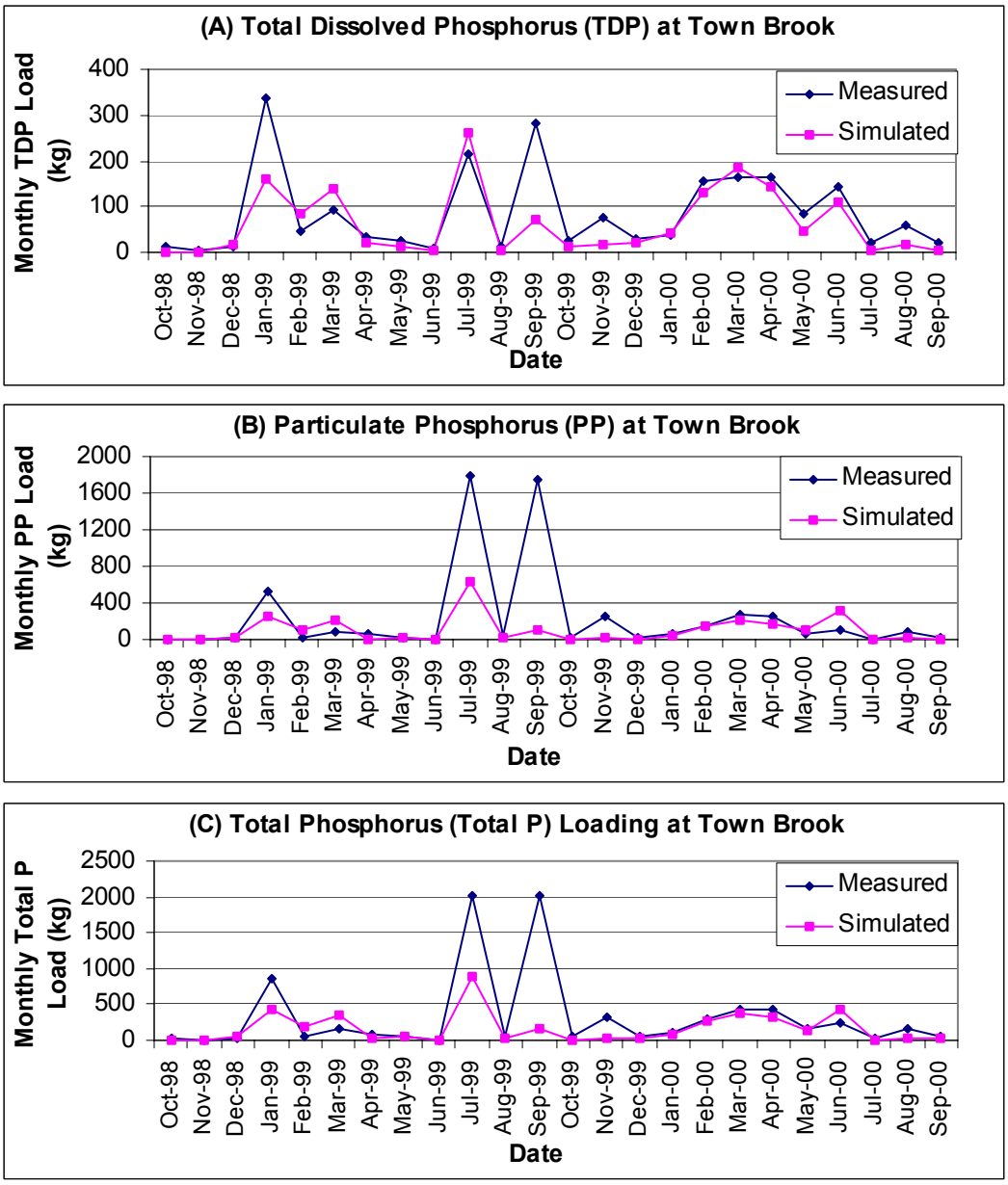

Figure 6.4.6. Time series of monthly measured and simulated A) TDP B) PP C) total P calibration results at Town Brook water quality station for the period Oct. 1998 to Sept. 2000.

Similar to measured TSS data, measured total P loading in July of 1999 for Town Brook and Beerston appear to be inconsistent. In July 1999, Town Brook total P loading is 105% of the Beerston total P loading. In comparison, for the other 23 months of record at Town Brook, Town Brook total P loading is only 9% of the Beerston total P loading. Therefore, based on the discussion and conclusion in Section 6.4.2 regarding the same data inconsistency for TSS, the total P loading prediction errors at Town Brook and Beerston for July of 1999 cannot currently be wholly attributed to the model and may in fact be due to errors in the measured data.

Model performance measures for Town Brook and Beerston are summarized in Table 6.4.5. The Beerston performance measures are given with and without the extreme phosphorus loading for January 1996. Based on the r^2 and E_{NS} coefficients at both locations, model predictions for TDP more closely replicate the trend in the data as compared to PP. This result is expected given that PP predictions are directly related to both flow and TSS predictions while TDP is mainly dependent on flow predictions. At Beerston, model performance measures for TDP prediction are actually better when January 1996 is included in the performance measure computation. This is because the model only slightly under-predicts TDP loading during January 1996. Given the tendency of the model to under-predict daily peak flows, the 6% under-prediction of total P loading at Beerston is quite reasonable. The total P r^2 and E_{NS} coefficients for Beerston of 0.72 and 0.68, respectively, indicate good model performance.

Table 6.4.5. Monthly phosphorus calibration results at continuous water quality monitoring stations for the period 1994 to 2000 (see Figure 6.1.1 for locations).

Phosphorus Fraction, Location	**Months of Measured Data**	**Mean Measured Data[A] (kg)**	**Mean Simulated Results[A] (kg)**	**Percent Difference between Measured & Simulated**	**r^2**	**E_{NS}**
TDP, Beerston	80	1244	1444	16.1 (13.5) [B]	0.77 (0.82)[B]	0.73 (0.81)[B]
PP, Beerston	80	2094	1690	-19.3 (-43.4)[B]	0.65 (0.43)[B]	0.57 (0.20)[B]
Total P, Beerston	80	3338	3135	-6.1 (-26.5)[B]	0.72 (0.52)[B]	0.68 (0.33)[B]
TDP, Town Brook	24	86	63	-26.8	0.59	0.52
PP, Town Brook	24	235	101	-57.0	0.43	0.22
Total P, Town Brook	24	321	164	-49.0	0.46	0.29

A. Calculated based only on the days with measured data for each station.
B. Performance measure includes January 1996.

Model performance measures for predicting total P were similar in the summer and winter seasons. Considering all summer seasons over the calibration period together, the model under-predicted Beerston total P by 3.1% and the monthly r^2 coefficient was 0.70. Considering all winter seasons over the calibration period all together, the model under-predicted Beerston total P by 7.4% and the monthly r^2 coefficient was 0.72.

Although model total P calibration performance focused on maximizing monthly r^2 and E_{NS} coefficients, the calibrated model total P predictions compared well with measured total P data on a daily time scale. Daily model performance measures for Beerston total P over the calibration period (excluding January 1996) show an r^2 and E_{NS} coefficient

both equal to 0.53. Similarly, the daily total P r^2 and E_{NS} coefficients at Town Brook for the Oct. 1998 through Sept. 2000 period are 0.49 and 0.38, respectively.

Further efforts to calibrate the model for phosphorus at Town Brook should only be attempted when another year of water quality monitoring data is available. As with TSS, phosphorus predictions at Town Brook could be improved by independently calibrating the subbasin parameters to the measured data. The Town Brook results for total P are less important relative to Beerston results since the latter represent most of the phosphorus delivered to the Cannonsville Reservoir.

The total simulated basin-wide phosphorus load during the calibration period is broken down by land use and sources in Section 1.1.2.

6.4.4 Spatially Distributed Analysis of Total Phosphorus Predictions

Model calibration focused mainly on results at Walton/Beerston. The calibration results observed at Town Brook for phosphorus loads were generally a byproduct of the calibration effort at Beerston. Due to this, and the fact that no continuous monitoring data at Town Brook is available for the validation period, there is little evidence that the spatially distributed model predictions for phosphorus are reasonable. Therefore, NYCDEP bi-weekly water quality grab sampling data, as provided in a computer spreadsheet file NYCDEP (Unpublished data) were utilized to perform a qualitative check on the spatially distributed model predictions for relative total phosphorus levels in the basin tributaries and along the WBDR mainstem.

NYCDEP data were available at over 20 locations across the basin. Of these, suitable sampling locations were selected to evaluate spatially distributed performance based on data availability and data suitability. Only water quality stations that sampled over approximately the same period of time, and therefore have approximately the same number of samples, are compared in this analysis. For the paired NYCDEP water quality stations that are immediately upstream and downstream of point sources, the downstream station is not used in this analysis to avoid samples where the stream cross-section may not be fully mixed. In addition, only NYCDEP water quality stations that could be reasonably associated with corresponding model subbasin outlets were selected. The NYCDEP water quality stations selected for this analysis were categorized as either a headwater water quality station (no upstream subbasin modeled) or a mainstem WBDR water quality station.

The bi-weekly and single sample nature of this spatially distributed data precluded its use for direct comparison with simulated daily phosphorus loading results. Instead the data were analyzed to generate long-term arithmetic averages for total P concentrations over the periods of interest. These averages can be used for relative comparisons with simulation results over the same periods given that the sampling data were collected bi-weekly, generally on the same day for all water quality stations such that a storm event was usually either sampled or not sampled at all NYCDEP stations in the basin. Namely, the average *relative* total P concentrations measured at each location can be compared to

average *relative* model simulated total P concentrations at each location. In other words, this assessment functions to make sure the highest and lowest relative total P concentrations measured in the basin are reasonably replicated by the model predictions for total P concentrations. The NYCDEP data for total P concentrations were arithmetically averaged while the simulation results from SWAT were only available as flow-weighted concentrations. Although the nature of the averages differ, since they were constant within each data set (measured and simulated) the relative nature of the concentration of total P is retained and the comparison is valid.

Spatially distributed phosphorus calibration results are presented in Section 6.4.4.1 for the calibration period and Section 6.5.4 for the validation period.

6.4.4.1 Spatially Distributed Phosphorus Calibration Results

Water quality data for a subset of NYCDEP stations was available from the start of the calibration period (January 1994) to December of 1999. Therefore, model calibration simulation results are only analyzed through the end of 1999 in this comparison. A total of ten NYCDEP water quality stations had suitable data over the Jan. 1994 through Dec. 1999 time period to be included in this analysis. One of these selected water quality station (WDHOA - close to the subbasin 1 drainage outlet) is considered here to be both a mainstem WBDR and headwater water quality station. Of the remaining nine selected locations, five of the NYCDEP monitoring locations are headwater water quality stations and four of the locations are mainstem WBDR water quality stations. The locations of the NYCDEP water quality stations used in this analysis are shown in Figure 6.4.7.

Table 6.4.6 summarizes the analysis for headwater water quality stations over the calibration comparison period. A comparison between the absolute magnitudes of the measured and simulated total P average concentrations at each location shows that modeled total P is always higher than measured total P. Based on the different nature of each average this behavior is expected. NYCDEP water quality data do not cover all high flow events and will therefore tend to miss sampling many of the higher concentration flow events. In contrast, the model simulates the entire period, and therefore most of the high flow/high concentration events. As a result, the simulated average total P concentrations will tend to include the highest concentrations while the measured average total P concentrations will not include most of the high concentrations.

Of all ten water quality stations used in the calibration comparison period, the highest average measured total P concentration occurs at WDHOA (79 µg/L) and the lowest occurs at C-8 (14 µg/L) and model performance at both of these locations is evaluated in Table 6.4.6. The last two columns in Table 6.4.6 rank the average total P concentrations for the measured data and simulated results. Comparing the ranks is one way to qualitatively evaluate the model performance. Results show excellent agreement between the ranks of the measured data and simulated results as all locations are ranked the same for both sets of average total P concentrations.

Table 6.4.6. Comparison of measured and simulated average total P concentrations for headwater subbasins for the Jan. 1994 through Dec. 1999 period.

NYCDEP Water Quality Station Name	SWAT Subbasin Outlet #[1]	Average[2] Measured Water Quality Station total P (µg/L)	Average[3] Simulated Subbasin Outlet total P (µg/L)	Relative Rank[4] of Measured total P conc.	Relative Rank[4] of Simulated total P conc.
WDHOA	1	79	133	1	1
CWB	4	35	98	2	2
C-38	8	33	76	3	3
C-8	18	14	49	7	7
C-7	19	16	55	6	6
C-79+CLDG[5]	27	20	55	5	5
WSPA[6]	38+43+11+12	31	68	4	4

1. Corresponds to adjacent NYCDEP water quality station location.
2. Arithmetic average of 147 to 151 samples (except 137 and 173 samples for C-38 and C-79+CLDG, respectively).
3. Flow-weighted average.
4. Rank of 1 corresponds to the highest total P concentration location.
5. Two water quality stations close by each other measured Little Delaware total P over period of record – both used in computation of average total P.
6. For comparison purposes, the mainstem WBDR location with the lowest average measured total P concentration is included. This station is upstream of the Walton WWTP and is therefore compared to SWAT results that do not include the WWTP.

Initial simulation results from the Jan. 1994 through Dec. 1999 period showed that the model predicted average flow-weighted total P concentrations for subbasin 19 (Trout Creek subbasin, water quality station C-7) that were in total disagreement with the measured average total P concentrations in Table 6.4.6. These initial simulation results predicted that the average total P concentration for subbasin 19 was the second highest of all ten locations used in the comparison and was nearly equal to the highest simulated average total P concentration for subbasin 1 (WDHOA). In comparison with the available data, this result was deemed to be a serious discrepancy and prompted further investigation to try to determine the cause. The investigation included a re-examination of the spatial land use in subbasin 19 and found that independent farm records (Dale Dewing, Personal Communication) strongly suggested the actual corn land use area in subbasin 19 is greatly over-estimated in the original NYCDEP land use data (by over 4 times). Based on the independent data on corn land use provided by Dale Dewing (Personal Communication) and the serious model prediction errors observed for subbasin 19, the corn land use areas input to the model for subbasin 19 were deemed incorrect and were reduced. The magnitude of the reduction in corn land use area was based completely on the data provided by Dale Dewing (Personal Communication) and is outlined in Section 4.2.1.1. The corrected corn land use areas input to subbasin 19 resulted in the much-improved simulated average total P concentration for subbasin 19 in Table 6.4.6.

The measured and simulated average total P concentrations at five NYCDEP mainstem WBDR water quality stations over the calibration comparison period are compared spatially and graphically in Figure 6.4.7. Figure 6.4.7 also shows the results for mainstem WBDR water quality stations in the validation period that are discussed further in Section 6.5.4. Only the results in the calibration period chart in Figure 6.4.7 will be discussed in this section. The measured (arithmetic) data for the calibration period in Figure 6.4.7 show that there is a significant dilution of total P moving downstream until a the increase at WDBN. Measured (arithmetic) total P concentrations at WSPA and WDLFB are less than 50% of those at WDHOA. This dilution of total P in the downstream direction is also apparent in the simulated average total P concentrations at the five locations. For example, simulated total P concentrations at WSPA and WDLFB are approximately 50% of those at WDHOA. However, the simulated data show no significant difference between the concentrations at WSPA and WDBN where as the measured (arithmetic) average total P concentration increases at WDBN by almost 50% over the WSPA total P concentration.

An alternative average can be computed for the measured data by weighting the bi-weekly samples of total P by the measured USGS flow at the Walton flow-gauging station (very close to WSPA) for the corresponding day. This alternative average is computed and plotted for the calibration period in Figure 6.4.7 for the WDLFB, WSPA and WDBN stations. Note that the Walton USGS measured flows are assumed to be representative weighting factors for these three stations because of their proximity to the Walton USGS station. Calculating weighted average total P concentrations based on Walton USGS flows at other NYCDEP water quality stations was deemed inappropriate given their distance from the Walton USGS station. The flow-weighted average total P concentrations for the measured data at WDLFB, WSPA and WDBN show that there is only a very minimal increase in total P concentration between WSPA and WDBN. The trend in the simulated total P results between WDLFB and WDBN closely resembles the trend in the flow-weighted measured total P concentrations. It is not clear whether the arithmetic or flow-weighted average of measured total P concentrations at WDLFB, WSPA and WDBN are most representative of the trend in actual total P concentrations between these locations. Therefore, since model predictions resemble the trend observed between WDLFB, WSPA and WDBN when flow-weighted total P concentrations are considered, the simulated results along the mainstem WBDR were deemed to be acceptable.

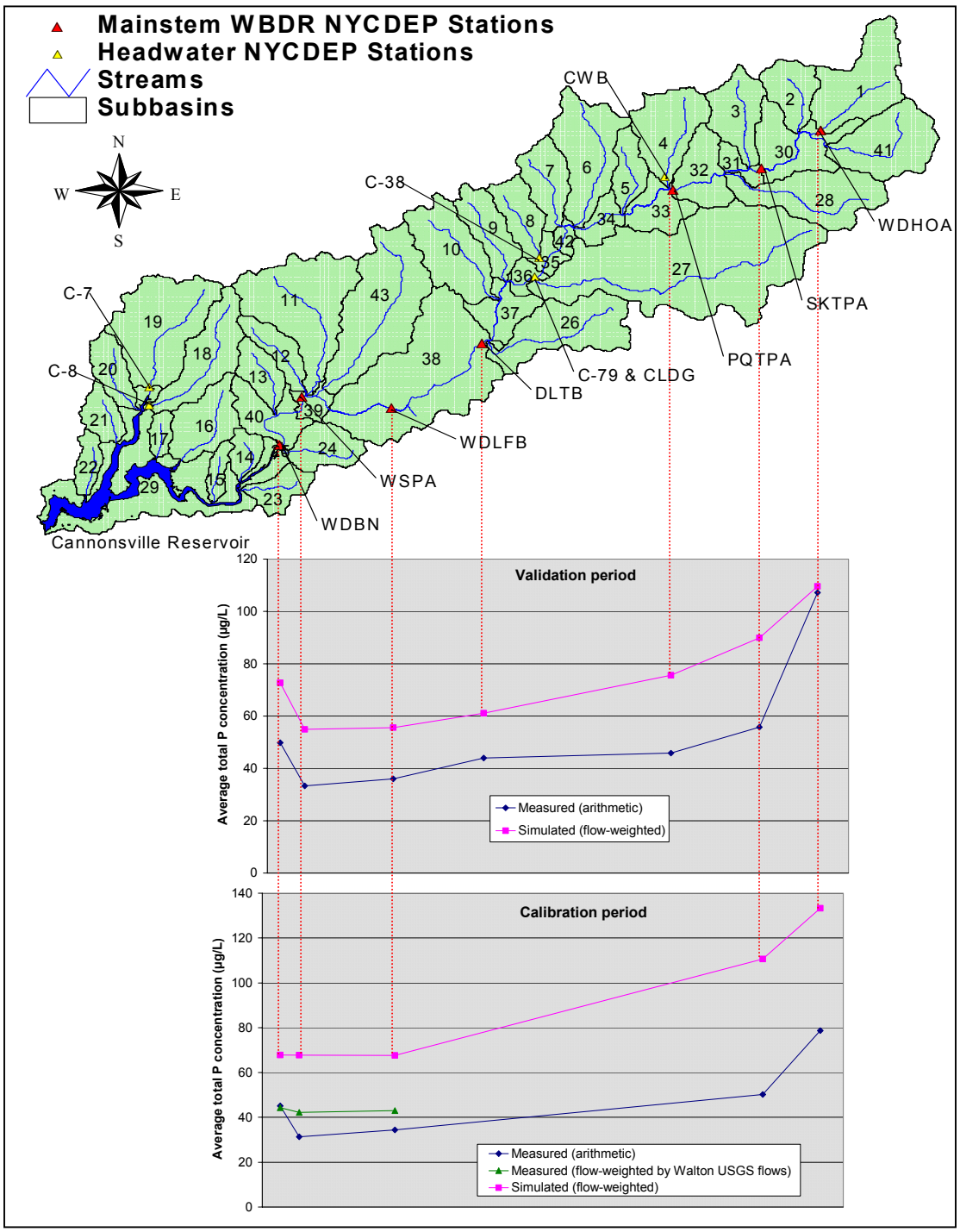

Figure 6.4.7. Measured and simulated average total P concentrations for selected NYCDEP mainstem WBDR water quality stations during the calibration period (Jan. 1994 – Dec. 1999 only) and validation period (Jan. 1990 – Dec. 1993).

6.5 Validation Results

Validation is the process of testing model performance of the calibrated model parameter set against an independent set of measured data. Validation was somewhat limited by data availability in both time and space. Validation results against continuous monitoring data for all constituents can only evaluated at the Walton/Beerston stations. The validation period selected was from January 1990 through December 1993 (48 months). Flow data were available for this entire period while TSS and phosphorus data were only available from October 1991 to December 1993 (27 months). Measured NYCDEP data were also available over the entire validation period to qualitatively validate spatially distributed total P predictions. Note that although flow data are available before 1990, a four year validation period for flow was assumed acceptable.

Model validation was conducted using the same warm-up period length (3 years), initial conditions, inputs and parameter values used or determined during model calibration except for the following:

- Appropriate climate data from 1987 to 1993 are used for the validation period.
- 1990 monthly point source loads are also used to represent the loads from January 1987 to December 1989.
- Manure generation rates per subbasin from 1990 to 1993 were assumed represented by the manure generation rates estimated from the 1992 cattle population in the basin (see Section 4.7.1.1). As a result, the total manure P applied to basin soils in validation is 27% higher than the manure P applied in the calibration period.

6.5.1 Hydrology Validation Results

Model validation performance measures for flow at Walton are presented for a daily and monthly time step in Table 6.5.1. The results show good model performance. The daily performance measures are just slightly lower than the corresponding performance measures in the calibration period. The monthly r^2 and E_{NS} coefficients for validation (0.94) are both somewhat higher than the corresponding coefficients in the calibration period (0.89).

Table 6.5.1. Hydrology validation results at Walton (USGS gauge 01423000) over the period Jan. 1990 to Dec. 1993.

Time Step for Model Performance Measures	Mean Measured Data (m^3/s)	Mean Simulated Results (m^3/s)	Percent Difference between Measured & Simulated	r^2	E_{NS}
Daily	16.0	16.7	4.5	0.79	0.78
Monthly	-	-	-	0.94	0.94

Daily measured and simulated flows at Walton are plotted for each calendar year in the validation period in Figure 6.5.1. Model performance measures by calendar year for the

validation flow predictions range from 0.70 to 0.83 for the r^2 coefficient and –1.5% to 13.4% for the percent difference in average daily flows. Late summer flows (September and October) tend to be over-predicted by the model, especially in 1991 and 1992. In fact, considering all summer seasons (May-Oct) over the validation period together, the model over-predicted Walton flows by 24.5% and the daily r^2 coefficient was 0.78. Performance across all winter seasons (Nov-Apr) in the validation period is better as flows are predicted within 1% and the daily r^2 coefficient is 0.74.

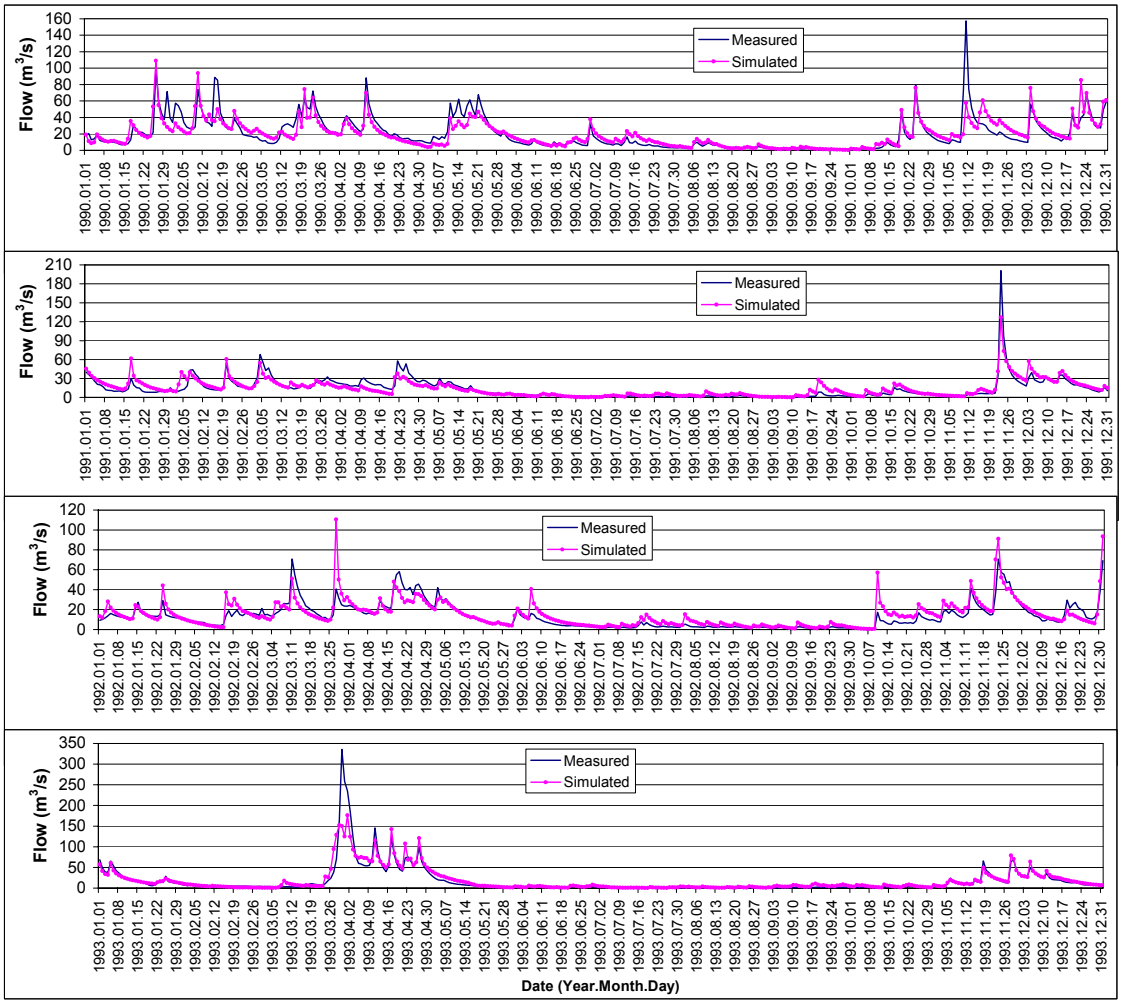

Figure 6.5.1. Time series of measured and simulated daily hydrology validation results at Walton for the period 1990-1993.

Figure 6.5.1 shows that in all calendar years except for 1993, the two largest peak daily flows in each year are significantly under-predicted by the model. The tendency of the model to under-predict peak flows is consistent with the calibration results. The scatter plot in the Appendix (Figure 10.5.1) of measured versus simulated daily flows at Walton over the validation period highlights these under-predictions more clearly. For example, of the eight daily flows in the validation period that were greater than 130 m^3/s, all were under-predicted and seven of these were under-predicted by more than 21%. As in the

calibration, many of these under-predicted peak flows were associated with snowmelt events.

Figure 6.5.1 also shows that the largest peak flow event in the validation period in late March of 1993 is severely under-predicted (by more than 50%) by the model. This 1993 event is associated with a melting snowpack that was measured at 37 inches of snow at the Walton climate station and 28 inches of snow at the Delhi climate station on March 14 and decreased steadily at both locations to a depth under 2 inches by the end of March. This under-prediction is consistent with the calibration results, that also showed the model tends to under-predict peak flows of large snowmelt related flow events. This March 1993 peak flow under-prediction does not appear to be due to errors in precipitation inputs since the cumulative measured and simulated flows at Walton between March 20, 1993 and April 30, 1993 are within 5% and within 1% between March 20, 1993 and May 15, 1993.

6.5.2 Sediment Validation Results

Validation results for TSS are outlined in Table 6.5.2 and plotted against measured TSS in Figure 6.5.2. Although the temporal variation in the measured TSS loading during the validation period (see Figure 6.5.2) is somewhat reduced in comparison to the calibration time series at Beerston in Figure 6.4.4, predictions for TSS generally display the same characteristics as in model calibration. Validation results show that model performance is fairly good with monthly r^2 and E_{NS} coefficients of 0.66 and 0.51, respectively. Over the validation period, the model under-predicts monthly TSS loads by about 26.5% on average. This under-prediction is mainly due to the model inability to predict the extreme sediment load in March 1993. The large under-prediction in March 1993 is largely due to the peak flow under-prediction for this month as discussed in Section 6.5.1. Figure 6.5.2 shows that, with the exception March 1993, the simulated TSS trend closely matches the trend in the measured TSS data. The measured load for TSS in Nov. 1991 is highly uncertain relative to other monthly measured loads due to poor sampling coverage of high flow events (Pat Bishop, Personal Communication). Therefore, the disagreement between the simulated and measured TSS load in Nov. 1991 is acceptable.

Table 6.5.2. Monthly sediment validation results at Beerston over the period Oct. 1991 to Dec. 1993.

Time Step for Model Performance Measures	Months of Measured Data	Mean Measured Data (metric t)	Mean Simulated Results[A] (metric t)	Percent Difference between Measured & Simulated	r^2	E_{NS}
Monthly	27	1266	931	-26.5	0.66	0.51
Daily	27	-	-	-	0.35	0.29

A) Calculated based only on the months with measured data.

As in the calibration, the largest monthly TSS prediction errors are positively correlated with the errors in peak daily flows for the corresponding month. For example, when TSS

loads are under-predicted in Nov-91 and Mar-93, peak daily flows in these months are also significantly under-predicted. Furthermore, when TSS loads are over-predicted in Jan-92 and Oct-92 through Dec-92, peak flows are significantly over-predicted. This correlation is more closely examined in Section 7.3.3.

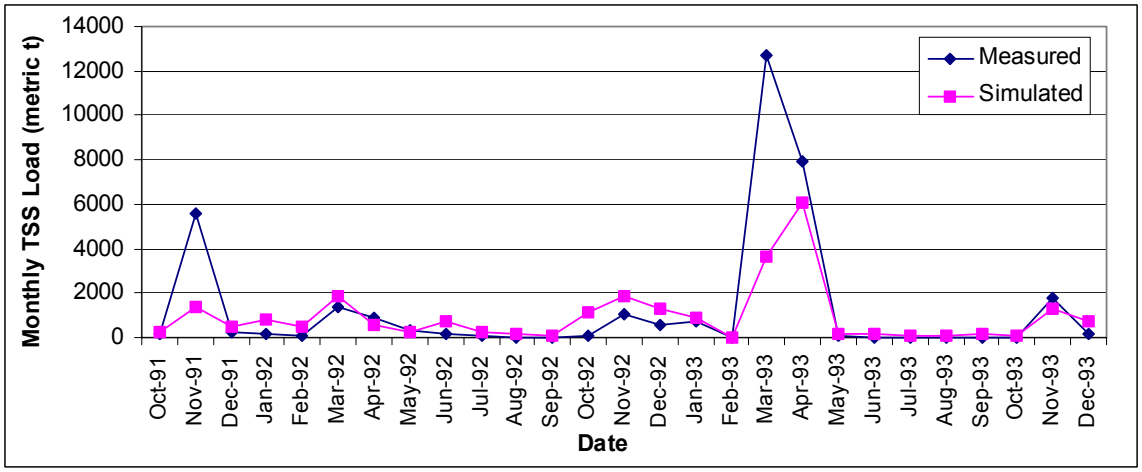

Figure 6.5.2. Time series of monthly measured and simulated sediment validation results at Beerston for the period Oct. 1991 to Dec. 1993.

6.5.3 Temporal Phosphorus Validation Results

Validation results for phosphorus are outlined in Table 6.5.3 and plotted against measured data in Figure 6.5.3. Table 6.5.3 shows that, consistent with calibration results, the model is over-predicting TDP (13.4% on average) and under-predicting PP (27.1% on average) while still matching average total P loads to within 10%. Again, since the model tends to under-predict peak daily flows, the 9% under-prediction of average total P loading is expected. When the validation period r^2 and E_{NS} values in Table 6.5.3 are considered, model performance for predicting P was approximately the same as or better than the calibration period performance. For example, although model performance for total P prediction in each period was similar (r^2 from 0.72 in calibration to 0.70 validation, E_{NS} from 0.68 to 0.61), TDP results improved significantly (r^2 from 0.77 to 0.93, E_{NS} from 0.73 to 0.90).

Figure 6.5.3 plots the measured and simulated TDP, PP and total P for the validation period. The simulated TDP loads very closely match the trends in the measured TDP data and the model performs well in the two most extreme months (March and April 1993). Consistent with the severe TSS under-prediction in March 1993, simulated PP in this month is also severely under-predicted. Figure 6.5.3 shows that the simulated total P trend follows the trend in the total P data closely. However, consistent with the peak daily flow and monthly TSS load, model predictions for total P during March 1993 are severely under-predicted.

Table 6.5.3. Phosphorus validation results at the Beerston monitoring station for the period Oct. 1991 to Dec. 1993.

Phosphorus Fraction (Time Step for Model Performance Measures)	Months of Measured Data	Mean Measured Data (kg)	Mean Simulated Results[A] (kg)	Percent Difference between Measured & Simulated	r^2	E_{NS}
TDP (monthly)	27	1658	1880	13.4	0.93	0.90
PP (monthly)	27	2091	1525	-27.1	0.56	0.43
Total P (monthly)	27	3749	3405	-9.2	0.70	0.61
TDP (daily)	27	-	-	-	0.52	0.44
PP (daily)	27	-	-	-	0.30	0.26
Total P (daily)	27	-	-	-	0.36	0.35

A) Calculated based only on the months with measured data.

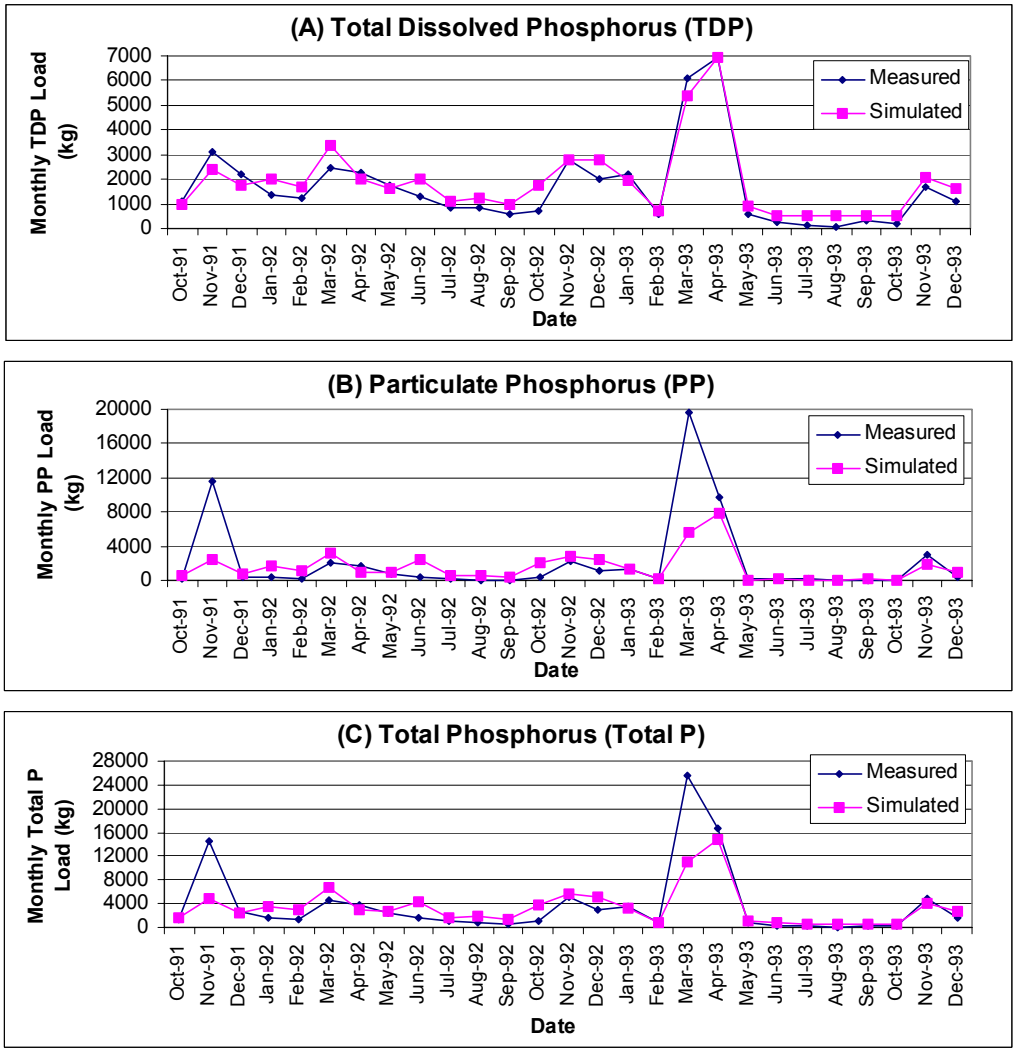

Figure 6.5.3. Time series of monthly measured and simulated A) TDP B) PP C) total P validation results at Beerston for the period Oct. 1991 to Dec. 1993.

6.5.4 Spatially Distributed Phosphorus Validation Results

The analysis described in Section 6.4.4 is repeated here for the validation period. Water quality data for a subset of NYCDEP stations was available for the entire validation period (Jan. 1990 through Dec. 1994). A total of eleven NYCDEP water quality stations had suitable data over the validation period to be included in this analysis. In addition to the WDHOA water quality station, four of the eleven selected NYCDEP monitoring locations are headwater water quality stations and six of the locations are mainstem WBDR water quality stations. The locations of the NYCDEP water quality stations used in this analysis are shown in Figure 6.4.7.

Table 6.5.4 summarizes the analysis for headwater subbasins over the validation period. The results in Table 6.5.4 show average simulated total P concentrations are always greater than average measured total P concentrations. As discussed previously in Section 6.4.4.1, these results are expected and are an indication of good model performance.

Table 6.5.4. Comparison of measured and simulated average total P concentrations for headwater subbasins for the validation period (Jan. 1990 to Dec. 1993).

NYCDEP Water Quality Station Name	**SWAT Subbasin Outlet #[1]**	**Average[2] Measured Water Quality Station total P (µg/L)**	**Average[3] Simulated Subbasin Outlet total P (µg/L)**	**Relative Rank[4] of Measured total P conc.**	**Relative Rank[4] of Simulated total P conc.**
WDHOA	1	107	110	1	1
C-38	8	36	68	2	2
C-8	18	15	37	6	6
C-7	19	17	37	5	5
C-79	27	21	46	4	4
WSPA[5]	38+43+11+12	33	55	3	3

1. Corresponds to adjacent NYCDEP water quality station location.
2. Arithmetic average of 98 to 102 samples for all water quality stations.
3. Flow-weighted average.
4. Rank of 1 corresponds to the highest total P concentration location.
5. For comparison purposes, the mainstem WBDR location with the lowest average measured total P concentration is included. This station is upstream of the Walton WWTP and is therefore compared to SWAT results that do not include the WWTP.

Of all eleven water quality stations used in the validation period, the highest average measured total P concentration occurs at WDHOA (107 µg/L) and the lowest occurs at C-8 (15 µg/L). These are the same locations identified in the calibration period with the highest and lowest average measured total P concentrations. As in the calibration period, simulated results in Table 6.5.4 show that the model also correctly predicts these locations to have the most extreme average total P concentrations.

The last two columns in Table 6.5.4 rank the average total P concentrations for the measured data and simulated results at all headwater water quality station locations. Results show excellent agreement between the ranks of the measured data and simulated results as all locations are ranked the same for both sets of average total P concentrations.

One of the most important results in Table 6.5.4 is that the corn land use reduction for Trout Creek (subbasin 19) resulted in excellent agreement between measured and simulated average total P concentrations over the validation period for this subbasin (water quality station C-7). As in the calibration results, the measured data suggest that relative to other locations, subbasin 19 has a very low average total P concentration. The relative ranks for this location are both 5 for the measured total P and simulated total P concentrations indicating that subbasin 19 is correctly simulated to have relatively low average total P concentrations. Also important, the simulated average total P concentrations for C-7 and C-8 water quality station locations are approximately the same magnitude, which is consistent with the relationship for these locations in the measured average total P concentrations. These results are good additional evidence that the land use change in Trout Creek (outlined first in Section 4.2.1.1 and then Section 6.4.4.1) is warranted.

The measured and simulated average total P concentrations at seven NYCDEP mainstem WBDR water quality stations over the calibration comparison period are compared spatially and graphically in Figure 6.4.7. Figure 6.4.7 also shows the results for mainstem WBDR water quality stations in the calibration period. Unless otherwise noted, all results discussed below refer to the measured and/or simulated average total P concentrations plotted in Figure 6.4.7 for the validation period. Similar to measured data in the calibration period, measured data over the validation shows there is a significant dilution of total P moving downstream until the increase at WDBN. In the validation period, the increase in both the measured and simulated average total P concentration at WDBN can be attributed to the high total P loading levels in the effluent from the Kraft cooling water discharge and Walton WWTP that are both discharged between the WSPA and WDBN locations. Note that the total P loading levels from these two point sources are greatly reduced in the calibration period (see Figure 4.5.1). Figure 6.4.7 shows that for the validation period the simulated average total P concentrations closely match the trend in measured average total P concentrations along the mainstem WBDR.

The only location in Figure 6.4.7 showing a minor disagreement between the measured and simulated average total P concentrations is the WDHOA water quality station. Based on the expectation that average simulated total P concentrations should be somewhat higher than average measured total P concentrations at the same location (discussed previously in Section 6.4.4.1), the simulated total P concentration at WDHOA (110 µg/L) should likely be even higher than it is now relative to the measured concentration (107 µg/L). One potential reason for this is that Stamford WWTP (which discharges into subbasin 1) total P loading levels for the validation period are not measured and had to be assumed in Section 4.5. The assumed total P load from the Stamford WWTP accounts for nearly 25% of the total P loading over the validation period. Therefore, if the actual total P loading rates from the Stamford WWTP were higher than the assumed total P

loading rates over the validation period then the actual total P loading rates input to the model would result in a significantly higher simulated average total P concentration at WDHOA. It is noteworthy that this problem does not occur at WDHOA during the calibration period - for which measured data are used to estimate much of the Stamford WWTP total P loading rates over the period. As a result, the minor disagreement between the measured and simulated average total P concentrations at the WDHOA water quality station in the validation is acceptable.

6.5.5 Validation Summary

Model performance levels for phosphorus, TSS and flow at Beerston/Walton in the validation period are approximately the same as model performance levels during the calibration period. The improved TDP predictions in the validation period are promising. Overall, spatially distributed phosphorus validation results are also very good. The validation results above are good evidence that the current set of model inputs and parameter values are fairly robust with respect to temporal and model predictions at Walton/Beerston as well as spatially distributed model predictions.

7 Discussion

Two general aspects that must be considered when evaluating the performance of the SWAT2000 Cannonsville Basin model are the nature of the loading data and the model performance statistics relative to similar published modeling studies. The NYSDEC TSS, TDP, and PP loading data are not directly measured and are instead calculated. This distinction is discussed in detail in Section 6.1.1. The Cannonsville Basin model is compared to model performance statistics in the following sections.

7.1 Comparison of Model Performance with Related Modeling Application

Since all modeling studies predict results with some error, it is useful to compare the numerical results obtained in Section 6 with previous published modeling studies in order to evaluate the generally accepted levels of model performance. Although there are numerous studies comparing monthly flow predictions with measured monthly data, there are very few that compare monthly sediment loads with measured monthly data. There are even fewer studies that compare monthly P load predictions with measured monthly data. The main reason for this is the lack of continuously monitored sediment and P loading data.

The main study selected for comparison here is a study by (Santhi et al. 2001) since they use SWAT and calibrate the model for flow, sediment and P. The study by Williams and Berndt (Williams and Berndt 1977), which provided some of the original testing of the MUSLE equation predictive accuracy for watershed monthly sediment loading, is also discussed.

A study by Kirsch et al. (2002) uses a modified version of SWAT 98.1 to simulate runoff, sediment and phosphorus loading in a mainly agricultural watershed in eastern Wisconsin. Kirsch et al. (2002) calibrated their model against five years of annual P loading data and achieved annual r^2 and E_{NS} coefficients of 0.95 and 0.07, respectively, and on an average over-prediction of total P of 22%. For comparison purposes, the annualized calibration r^2 and E_{NS} values over a period of six years (1994-1999), *including* the January 1996 flood event, achieved in this study at Beerston were 0.93 and 0.46, respectively. In addition, over this same period, the Cannonsville model under-predicted total P by 28%. Overall, annual model performance for total P loading is better in this study based on the large difference in the annual E_{NS} coefficients.

A detailed comparison with the NYCDEP developed GWLF model of the Cannonsville Basin (Schneiderman et al. 1998) would be insightful. However, suitable performance statistics for model comparison are not available in the published NYCDEP report (Schneiderman et al. 1998). The only reported statistics in their report are for validation and are given as the coefficient of model efficiency. Since a comparison based on this single validation statistic for hydrology, sediment and phosphorus is not complete, the results of the two Cannonsville models are not compared here.

7.1.1 Monthly Sediment Predictions in Williams and Berndt (Williams and Berndt 1977)

The MUSLE equation is the foundation of sediment prediction in SWAT and is used to predict the amount of sediment reaching the stream channels. Williams and Berndt (Williams and Berndt 1977) set out to test the predictive accuracy of the equation when a hydrologic model was used to estimate runoff volumes. Although the hydrologic model utilized in conjunction with the MUSLE equation differed from SWAT and the MUSLE equation was applied in a lumped fashion as opposed to a distributed approach (as in SWAT), the comparison with the Cannonsville SWAT2000 model is sensible because the MUSLE equation is an integral part of SWAT sediment loading predictions. More importantly, the MUSLE equation is a widely accepted approach for estimating temporally variable sediment loads. Many applications of MUSLE are common in the literature (Johnson et al. 1985, McConkey et al. 1997). Therefore, the performance accuracy of the MUSLE equation as outlined in Williams and Berndt (Williams and Berndt 1977) can be interpreted as an acceptable level of predictive capability with respect to temporally varying sediment loads.

Williams and Berndt (Williams and Berndt 1977) evaluated the performance of the MUSLE equation against data from 27 watersheds. They found that the average annual MUSLE predicted sediment yield was from 57% less than or 45% more than the average annual measured sediment yield. Monthly r^2 values were only available from nine watersheds and ranged from 0.95 to 0.32 with 1/3 of them having an r^2 less than 0.51. The Cannonsville SWAT2000 model performance measures fall well within the ranges of MUSLE equation performance statistics from Williams and Berndt (Williams and Berndt 1977). For example, average sediment loads predicted by the model at Beerston during calibration and validation were within 35% of the measured loads including January 1996 or 27% if January 1996 was not included. The r^2 values for Beerston during calibration and validation were 0.71 (no January 1996) and 0.66, respectively. The comparable performance statistics to those in Williams and Berndt (Williams and Berndt 1977) show that sediment predictions are quite reasonable given the general acceptance of the MUSLE equation under similar performance levels.

7.1.2 Monthly Flow, Sediment and P Predictions in Santhi et al. (2001)

A recent modeling effort in Texas by the developers of the SWAT model (Santhi et al. 2001) is the most similar SWAT application to this study in the literature. The Hico Basin drainage area as reported in Santhi et al. (2001) is 926 km^2 (nearly the same size as the Beerston drainage area) and has a comparable data record length used in calibration. Furthermore, they had good continuous data on flow, TSS and P as in this study. Lastly, their study area was a major dairy producing area in Texas and thus involved similar management practices. Some of the main differences between the two studies are that New York and Texas climate and soils are very different, Santhi et al. (2001) used an older version of SWAT (SWAT98.1) and their validation period is only 12 months in length.

Calibration and validation results from Santhi et al. (2001) for the Hico Basin are summarized in Table 7.1.1. Model performance statistics in Santhi et al (2001) were based on loading rates per unit area and it is assumed that the quantities they reported (sediment, organic P and mineral P) correspond to the quantities TSS, PP and TDP as defined in this study. Comparing calibration performance measures, the Cannonsville model results for flow and TDP and better than the Hico model. However, the Hico model outperforms the Cannonsville model (with respect to r^2 and E_{NS} only) for TSS and PP predictions over their respective calibration periods. Similar relative performance levels are observed when the model performance levels are compared over the validation periods. The observation that the Cannonsville model TDP validation results are significantly better than the Hico results is noteworthy considering the fact that the validation results in Santhi et al. (2001) are biased towards good results with respect to r^2 and E_{NS} since only 1 month out of the 12 month period they selected had significantly different measured quantities for flow, TSS and P.

Hydrology, TSS, TDP and PP percent differences in Table 7.1.1 are all significantly smaller in magnitude for the Cannonsville model validation results compared to Hico. This suggests that, in some respects at least, the Cannonsville model is performing more accurately for prediction of the model outputs considered. Model prediction accuracy in validation is particularly important since validation attempts to assess model performance under new conditions and new independent measured data. When model accuracy is preserved in validation, confidence in the accuracy of additional model forecasting (e.g. management scenarios) increases.

Table 7.1.1. Monthly calibration and verification results from Santhi et al (2001) for the Hico Watershed in Texas compared with monthly results for the Cannonsville SWAT2000 model in this study (in brackets).

Constituent (period)	**Months of Measured Data**	**Percent Difference between Measured & Simulated**	**r^2**	**E_{NS}**
Flow (calibration)	60 (81)[A]	-2.5 (1.0)[A]	0.80 (0.89)[A]	0.79 (0.89)[A]
TSS (calibration)	60 (80)	-15.6 (-2.2)	0.81 (0.71)	0.80 (0.66)
PP (calibration)	60 (80)	-13.9 (-19.3)	0.71 (0.65)	0.70 (0.57)
TDP (calibration)	60 (80)	-19.2 (16.1)	0.60 (0.77)	0.59 (0.73)
Flow (validation)	12 (48)	-22.4 (4.5)	0.92 (0.94)	0.87 (0.94)
TSS (validation)	12 (27)	-44.4 (-26.5)	0.98 (0.66)	0.70 (0.51)
PP (validation)	12 (27)	-46.5 (-27.1)	0.95 (0.56)	0.72 (0.43)
TDP (validation)	12 (27)	-50.0 (13.4)	0.83 (0.93)	0.53 (0.90)

A) For all rows in columns 2 through 5, the first number is for the Hico Watershed model and the number in brackets is for the Cannonsville Watershed model.

These results indicate that the SWAT2000 model has good performance relative to the Hico Basin study. Cannonsville model performance is more impressive given that there are a number of reasons why the Hico Basin application of SWAT would be expected to

generate better numerical performance in comparison with the Cannonsville Basin SWAT model. These are as follows:

- SWAT was originally developed in Texas and thus applications there are likely to be more accurate than locations in states with significantly different conditions.
- For the Cannonsville Basin, SWAT is forced to simulate snowmelt processes. From the hydrology calibration and validation results is clear that much of the poor performance by SWAT is related to snowmelt processes.
- The original developers of SWAT are authors of the paper. They are therefore more experienced SWAT modelers and should be expected to generally produce more accurate results than a relatively less experienced SWAT modeler.
- There are 5-7 temperature and precipitation stations are used for climate inputs in the Hico Basin, while only 3 precipitation stations and 2 temperature stations are available for the Cannonsville Basin model. In other words, since the drainage areas of the models are nearly the same, the spatial resolution of the climate inputs in the Hico model is nearly double that in the Cannonsville model.

Overall, the SWAT application comparison showed that relative to a high quality, similar SWAT modeling study (Santhi et al. 2001) the SWAT2000 model of the Cannonsville Basin generates results of comparable quality.

7.2 Potentially Important Processes and Phosphorus Sources Excluded from Model

There are some potentially important processes and phosphorus sources that are not included or represented explicitly in the model of the Cannonsville Basin that warrant additional discussion. These were omitted from the basin model based on either time or model constraints. The following processes and phosphorus sources were not modeled or represented in the SWAT model of the Cannonsville Basin:

- Crop rotations.
- Concentrated source areas of P on farms.
- Septic systems.
- Instream nutrient reactions.

Crop rotations are an important management practice that reduces soil erosion and nutrient buildup on agricultural fields. Steps were taken to incorporate corn-hay crop rotations in the model. However, due to time constraints the current version of the model does not simulate crop rotations. The current model HRUs are defined in such a way that in future versions of the model, corn-hay rotations can be simulated without modifying the number or size of the HRUs while simultaneously holding the modeled areas and corn and hay constant throughout the simulations. It is probable that the future inclusion of crop rotations will reduce the current sediment and P loading rates simulated to originate from corn silage HRUs. The model currently simulates nearly 60% of the non-point source sediment and total P loading as originating from corn silage HRUs. This issue is important to consider when the current model is used to evaluate phosphorus management options that are applied to corn silage.

Concentrated source areas of P on farms (e.g. milk house wastes and barnyards, respectively) are not modeled currently because of the additional model inputs required by each. Point sources could be included given inputs such as average point source P discharge rates while barnyards could be modeled as HRUs given other estimates of barnyard areas per subbasin and fractions of the total manure produced that are not collected from the barnyard. Given the mass balance approach to manure, alternative allocations of manure P could be incorporated into the model. However, the current model inputs are assumed reasonable since all estimated manure produced in a subbasin is distributed onto the land surface somewhere in the subbasin.

SWAT2000 does not explicitly model nutrient inputs due to septic systems. However, the data based subbasin groundwater soluble P inputs to the model do represent, at least partially, the impact of septic system P on groundwater. An alternative septic system model that can be linked with SWAT is currently being considered that would estimate subbasin septic P loads to the river based on the septic system characterization in the subbasin. This model output would be incorporated into SWAT as a point source input. However, it is not yet clear how the corresponding groundwater P concentrations input to the model should then be reduced.

The option for simulating instream nutrient reactions in SWAT2000 was investigated in this study. However, suitable parameters controlling the conversion of P between particulate and dissolved forms could not be identified and the current instream nutrient equations in SWAT2000 were deemed unsuitable for the Cannonsville Basin. As discussed in Section 6.2.2, a related issue for improving model predictions of the forms of phosphorus in the stream (measured PP versus SWAT simulated organic P and measured TDP versus SWAT simulated mineral P) is the questionable model assumption that 100% of active mineral P is completely desorbed from eroded soil particles once surface runoff reaches the channel. Any future work focused on improving the model predictions for the fractionation of phosphorus between PP and TDP should evaluate this modeling assumption closely before or in conjunction with attempting to simulate instream nutrient reactions.

7.3 Model Limitations

Although model performance is generally quite good, there are some significant limitations that are evident from the model calibration and validation results. It is important to try to account for the model limitations for two reasons. First, assuming their correction is to be considered, the cause of the limitation requires determination. Secondly, since this model is to be used for management option evaluations, more confidence in future predictions can be attained if the significant model limitations can be partly explained by conditions that will not be present in the future management option simulations.

7.3.1 Hydrology Predictions

Model predictions of the largest daily flows (i.e. greater than 130 m^3/s) over the calibration and validation period are almost always much too low. In contrast, the largest peak summer flows are usually over-predicted. These limitations are a function of at least two sources.

Model prediction accuracy of winter snowmelt events appears to be limited. This can be partly attributed to the manner in which SWAT applies various parameters. For example, it was clear from the measured climate data analysis (see Section 10.4.1) and from the calibration effort (see Section 6.3) that optimal settings for some of the snowmelt parameters differed between the Northeastern-most and Southwestern-most basin flow gauges. These parameters included the temperature that defines whether precipitation fell as snow or rain (SFTMP) and the temperature at which the snowpack begins to melt (SMTMP). However, since these parameters are applied basin-wide in SWAT, there is currently only one value used to represent each parameter. This approach should be considered in the future.

The final calibrated CN2 values in the model (SWAT parameter controlling surface runoff generation) were identified by balancing the over-predictions of peak daily flows in the summer and fall with the under-predictions of peak daily flows in the winter and spring. If the seasonal values of the CN2 parameter were allowed to vary in the model, model predictions could be significantly improved in all seasons. Although the SCS curve number approach used in SWAT adjusts surface runoff predictions for season and antecedent soil moisture conditions, the previous observation suggests that the current SWAT approach for seasonal or soil moisture adjustments to surface runoff predictions requires modification if peak flows across all seasons are to be more accurately predicted.

7.3.2 Suspended Sediment Predictions

7.3.2.1 Management Representation in Model

The static scheduling of agricultural management options in SWAT is meant to represent the 'average' management behavior on a typical basin farm. However, this does not capture the year-to-year variations in management, especially the timing of management operations, which the typical farm would undergo simply due to climate variation. In general, tillage, planting and harvesting of crops are all dependent on antecedent climate conditions. Thus, the model representation of agricultural management practices likely deviates significantly from the actual management practices in at least a few months of the model simulations. As a result, a few of the most significant model inaccuracies could be at least partially due to this problem. For example, consider the following:

- Although very infrequent, in years where weather conditions are perfect, up to 10% of fields in the basin could be tilled in the fall according to Dale Dewing (Personal Communication). This is not represented in model inputs. This could account for some of the severe under-predictions of sediment loading measured in November 1991 and 1996. Tillage in the fall removes erosion protecting crop

residue from the top of the soil and thus exposes soil to more erosion until the next growing season.

- Tillage is scheduled in the model at the end of April and beginning of May of each year. However, in reality, the actual dates for tillage vary based on climate. Therefore, some discrepancies in April and May sediment load predictions are possible. Consider that in April-96 sediment loads were slightly over-predicted while in May-96, sediment loads were significantly under-predicted. This could be due to the combination of significant rainstorms observed in the climate data at the end of April-96 and the model scheduled tillage at the end of that month which potentially did not happen because of the rain. If all tillage that year was pushed into May, then the model under-prediction of May sediment loading is at least partially due to the erroneous model scheduling of a tillage operation in April.

In agricultural practice there is a relationship between climate (or even forecasted climate) and timing of management operations on the farm. Failure to represent this in model development almost certainly accounts for some of the model problems. Failure to account for this relationship in the evaluation of management scenarios could also generate unreasonable results. Therefore, work is required to evaluate the sensitivity of model results to management timing around large storms and then consider if management operations should be implemented as a function of antecedent climate conditions. In other words, model simulations should perhaps also mimic the average farmer's response to variable climate conditions.

7.3.2.2 Process Representation in the Model

The general under-prediction of sediment loads during winter was partially corrected by the change in the MUSLE snow cover erosion adjustment equation (see Sections 5.7 and 6.4.2). Another cause of sediment load under-predictions in winter is the under-prediction of peak daily flows in winter. In addition, a number of other potential causes for this under-prediction are given as follows:

- SWAT is unable to simulate floodplain erosion during overbank flow.
- SWAT does not simulate the erosion of soils in colder climates due to freeze-thaw effects.
- SWAT does not distinguish between manure spread directly on snow versus manure spread directly on the soil surface.

The change made to the model code to increase sediment erosion in months affected by snowmelt is a reasonable empirical correction for the erosion of soils due to freeze-thaw affects. However, this correction is not entirely suitable for correcting the model for the other issues.

Failure to predict overbank flow erosion should not be considered a model shortcoming because the SWAT2000 model purpose clearly states that it is not designed to do this. Furthermore, based on the enormous 1996 flood event measured PP load (75% of the total PP for the year), a watershed model is not needed to predict that sediment and nutrient loading during an extreme overbank flow event will be extremely high. In other

words, given that a flow event is predicted to so dramatically exceed the banks of the WBDR, it is quite reasonable to assume that there will be water quality problems (i.e. TMDL violation) despite the application of the best management options available to Delaware County.

Failure of the model to track the amount of manure spread directly on snow is a shortcoming since it is easy to envision that sediment and nutrients from manure incorporated in a snowpack is readily available for transport to the stream during snowmelt without requiring energy to detach it from the soil surface. SWAT2000 currently incorporates manure into the first soil layer (1 cm in depth) regardless of the season or depth of snow on the soil surface at the time of manure application. Based on the calibration period climate station data for Delhi, on average, there are 60 days each winter season where the snowpack was one or more inches deep. Based on the manure production assumptions outlined in Section 4.7.1 and daily spreading of manure over 60 days, an estimate of the manure spread directly onto snow is approximately 4000 mt of dry manure for an average winter season. Assuming more than 50% of this snowpack applied manure was transported to streams, this would account for a significant part of the TSS under-estimation in many of the problematic months identified in Figure 6.4.3.

Clearly, there are multiple processes contributing to the sediment (and therefore PP) loading of the reservoir. Although the cumulative representation of sediment loads at Beerston is currently reasonable, further work may be necessary to ensure the model physically represents the application of manure on snow and the subsequent transport of this manure sediment to the stream during snowmelt. This is a particularly important consideration if the model is to be used to evaluate future management scenarios that focus on changing manure spreading practices during the winter and early spring periods.

An alternative to modifying the MUSLE snow cover erosion adjustment equation is to investigate a change in the SWAT2000 source code so that the spreading of manure on snow is accounted for and modeled as a physical process. This would require tracking manure in the snowpack as a new state variable and predicting the resulting fraction of snowpack manure transported in snowmelt and runoff. This would potentially replace the current model code change to the MUSLE snow cover adjustment equation (see Section 5.7). However, such a drastic change in model coding is best left to the SWAT model developers.

7.3.3 Phosphorus, Sediment and Hydrology Interdependencies

Simulation results for sediment and phosphorus are largely dependent on flow. Figure 7.3.1 below shows a general schematic that represents the most significant general relationships between the quantities of interest in this study simulated by SWAT. The boxes represent the quantity modeled while the directional arrows identify which quantities are dependant on one another. For example, a change in the surface runoff volume predicted in any HRU will cause a direct change in the predicted sediment and dissolved phosphorus load leaving the HRU in surface runoff. In addition, since sediment predictions increased, so too will particulate phosphorus predictions. Figure

7.3.1 does not show all possible interactions between simulated quantities. One such relatively minor relationship would be that changes in soil P levels could impact hydrology results because of the resulting changes in plant growth and thus changes in the water balance of the HRU. The purpose of Figure 7.3.1 is to simply highlight the most significant relationships that can be inferred by evaluation of the SWAT2000 model equations. Model performance can be evaluated more rigorously when consideration is given to the relationships in Figure 7.3.1.

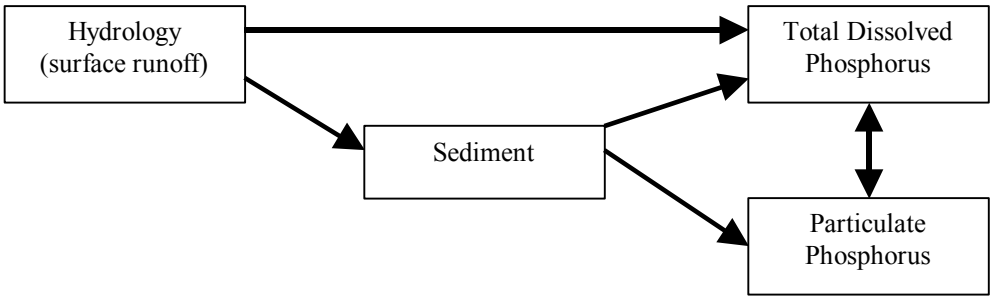

Figure 7.3.1. General relationships between the simulated export of quantities from HRUs in SWAT2000.

Figure 7.3.1 shows that any errors in hydrology propagate to create errors in all other modeled constituents. In fact, predictions for both forms of P are impacted by hydrology and sediment prediction errors. While the prediction error propagation form (e.g. linear or exponential) is not evaluated exactly here, it is clear that P prediction errors are strongly dependent on hydrology and sediment prediction errors.

The simulated results and measured data for flow, TSS and total P demonstrate that the above statement is generally true. To show this is the case, the months in both the calibration and validation period that have the largest absolute deviations between simulated and measured total P loads at Beerston are analyzed more closely. An absolute value of 2000 kg was arbitrarily chosen to distinguish a subset of large deviations and resulted in 15 out of 81 months in calibration and 5 out of 27 months in validation being selected for closer analysis. These 20 selected months and their respective simulated and measured total P load deviations and percentage errors are given in the first three columns of Table 7.3.1. The largest absolute deviations were used as a criterion for selecting these months (as opposed to percent error) because improving model predictions in months with the largest absolute deviations in simulated and measured total P would make the greatest improvement in model performance measures. For example, if the model could somehow predict the exact measured total P loadings in the 15 months from the calibration period selected for this analysis, then the calibration performance measures for total P would improve drastically – from 0.52 to 0.996 for the r^2 coefficient, 0.33 to 0.97 for the E_{NS} coefficient and –26.5% to 2.4% for the percent difference when January 1996 is included.

For each of the selected months in Table 7.3.1, the deviation in simulated and measured monthly TSS loading, the percent error of the monthly TSS deviation and the percent error in the simulated peak flow for the month, relative to the measured peak flow, are

also presented. The results in Table 7.3.1 are sorted so the largest monthly under-predictions of total P (i.e. the largest negative deviation of simulated and measured total P) appear at the top of table and the largest monthly over-predictions appear at the bottom of the table. All negative percent errors for total P occur in months when there are large negative percent errors for TSS and large negative errors for peak daily flows. Similarly, all positive percent errors for total P occur in months when there are large positive percent errors for TSS and large positive errors for peak daily flows. These observations qualitatively demonstrate the positive correlations between the model predictive errors of total P, TSS, and peak daily flow.

Table 7.3.1. Monthly sediment load and peak daily flow prediction errors for all months over the calibration and validation period in which the total phosphorus load was not predicted by the model within ±2000 kg.

Month-Year (C or V)[1]	Total Phosphorus load for Month		Sediment (TSS) load for Month		% Error in Peak Measured Daily Flow for Month[4]
	Sim – Meas[2] (kg)	% Error[3]	Sim – Meas[2] (metric t)	% Error[3]	
Jan-1996 (C)	-80576	-82	-52812	-88	-65 *[5]
Mar-1993 (V)	-14628	-57	-9036	-71	-45
Nov-1996 (C)	-14528	-56	-5810	-55	-30
Jan-1998 (C)	-10152	-50	-3886	-51	-50
Nov-1991 (V)	-9922	-68	-4147	-75	-35
Feb-2000 (C)	-6902	-50	-4576	-64	-45
Jan-1999 (C)	-4480	-31	-2065	-35	-20
July-1998 (C)	-4246	-43	-2131	-44	-25
May-1996 (C)	-2711	-32	-2261	-47	-35
May-1997 (C)	-2136	-55	-400	-38	-30
Dec-1992 (V)	2097	69	769	143	35
Dec-1998 (C)	2120	624	717	1320	420 *
Feb-1994 (C)	2271	105	1073	281	230
Mar-1996 (C)	2393	152	1210	406	150 *
Mar-2000 (C)	2491	46	958	43	60
June-1992 (V)	2716	159	626	441	155
Oct-1992 (V)	2855	279	1087	1351	235
Mar-1999 (C)	3316	106	1337	115	30
June-1996 (C)	3605	64	897	34	30 *
July-1999 (C)	3908	204	1993	429	120

1. C indicates calibration period, V indicates validation period.
2. Sim – Meas = simulated monthly load – measured monthly load.
3. % Error = 100 x (Sim – Meas) / Meas.
4. Computed as in (3) above except that the peak daily simulated (Sim) and measured flows (Meas) do not always occur on the same day. Precision reduced to reflect this.
5. The asterisk (*) beside the value indicates that the error applies to the 2nd largest peak daily measured flow in the month (e.g. largest peak flow accurately simulated).

A quantitative measure of the correlation between these errors can be assessed by constructing scatter plots and computing the r^2 coefficients between the percent errors for

all pairs of the total P, TSS and errors in peak daily flow values given in Table 7.3.1. Figure 7.3.2 is a scatter plot comparing the percent errors between monthly total P and daily peak flows and clearly shows the strong positive correlation between these quantities (an r^2 coefficient of 0.89). Similar scatter plots between the percent errors in TSS and total P as well as TSS and errors in peak daily flow both show strong positive correlations.

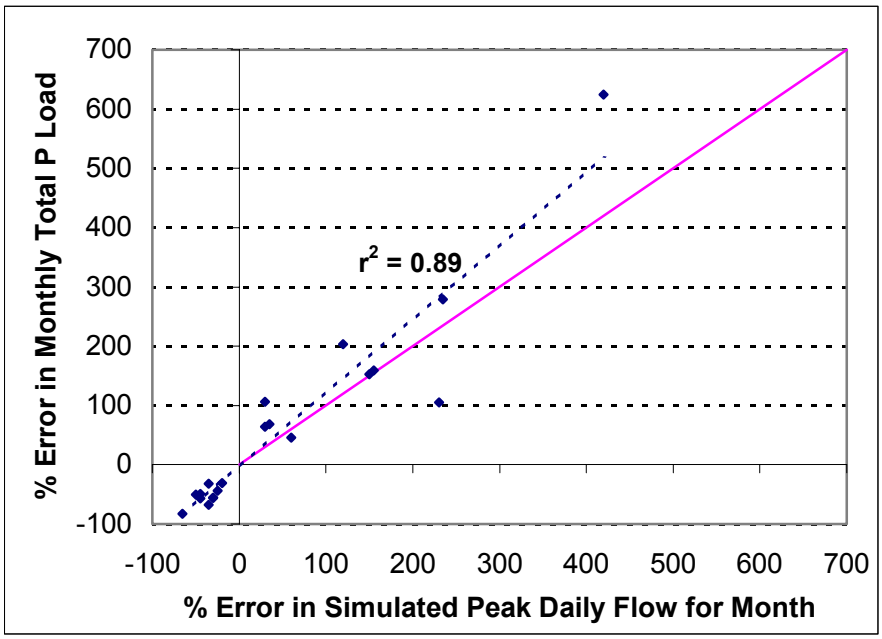

Figure 7.3.2. Scatter plot comparing the percent error of total P monthly load predictions with the percent error of peak daily flow predictions for the months in Table 7.3.1.

The demonstrated relationships between the largest errors in simulated monthly total P loads (and TSS) and the errors in simulated peak daily flows show that model predictive errors in TSS and total P are largely due to errors in peak flow predictions. This finding is intuitive because most of the total P and TSS load is transported in the peak flow events. Therefore, if further work is conducted on this model to improve total P or TSS predictions, it should be first focused on improving the simulated daily peak flow accuracy for the months identified in Table 7.3.1.

7.4 Summary of Future Model Improvements and Modeling Analyses

Any future work to improve the SWAT2000 model should refer to the processes and phosphorus sources that are not currently modeled (as outlined in Section 7.2) as well as the potential approaches outlined in Section 7.3 to address the current model limitations. Future phosphorus management scenarios for the Cannonsville Basin that have been evaluated with the calibrated SWAT2000 model include temporal changes in the cattle population, reduced inorganic phosphorus fertilizer spreading to corn silage, manure export from the basin, WWTP upgrades and precision feeding of dairy cattle among others.

8 Conclusions

The current version of the SWAT2000 Cannonsville Basin Model has been calibrated and validated successfully. The model has been shown to reasonably represent the temporal and spatial nature of the measured flow and water quality data at multiple locations in the basin. In addition, model performance was comparable and better in some cases to the performance reported in a recent application of SWAT to another watershed by the SWAT model developers (Santhi et al. 2001). Therefore, the model in its current form could be used to evaluate potential management strategies for reducing P loading to the Cannonsville Reservoir.

A number of important benefits to Delaware County stakeholders as well as other users of the general SWAT2000 model are the result of this comprehensive modeling effort. They are as follows:

- The spatially distributed modeling approach used in SWAT allows evaluation of management options that vary spatially across the basin. This is particularly important given that Delaware County has a finite amount of resources to manage phosphorus in the basin and would therefore be interested in the most cost effective management strategies.
- Model phosphorus inputs are summarized in a way that is potentially useful to local decision-makers in the absence of a basin model and will help facilitate understanding and communication between all stakeholders.
- The SWAT basin modeling approach provides the scientific framework for compiling and integrating the Delaware County Action Plan (Delaware County Board of Supervisors 1999).
- The basis for phosphorus input information is outlined in sufficient detail for future modelers in the basin to repeat the input development when new input data become available or if a future model of the basin (with or without SWAT) is developed.
- The spatially distributed modeling approach can also be utilized to help direct future water quality monitoring efforts in the basin.
- The modeling work described here provides a means for investigating assumptions made in previous modeling and monitoring work that currently form the basis of the management restrictions in the basin.
- This model could provide improved inputs to the more complex reservoir model currently being investigated by the NYCDEP. In other words, this advancement parallels the advancement being made over the relatively simple NYCDEP Phase II TMDL reservoir model (Kane 1999).
- This comprehensive report completely documents the model development work for Delaware County and will serve as a guide to transferring all data files associated with this modeling effort.
- The results of the modeling effort will hopefully help to improve the general SWAT2000 watershed model.

As a cautionary note, this model should not be expected to be able to simulate the exact impact of all possible management alternatives for phosphorus in the basin. Any model is a simplification of reality and its forecasts are based on available information, which is always limited. Furthermore, there may be management options that can be easily assessed outside of the model. However, the model does provide the most thorough characterization of the watershed to date and therefore forms the foundation required to evaluate a variety of phosphorus management scenarios.

9 References

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

Arnold, J. G., and P. M. Allen. 1996. Estimating Hydrologic Budgets for Three Illinois Watersheds. Journal of Hydrology **176**:57-77.

Arnold, J. G., P. M. Allen, and G. Bernhardt. 1993. A Comprehensive Surface-Groundwater Flow Model. Journal of Hydrology **142**:47-69.

Arnold, J. G., R. S. Muttiah, R. Srinivasan, and P. M. Allen. 2000. Regional Estimation of Base Flow and Groundwater Recharge in the Upper Mississippi River Basin. Journal of Hydrology **227**:21-40.

Arnold, J. G., R. Srinivasan, R. R. Muttiah, and J. R. Williams. 1998. Large Area Hydrologic Modeling and Assessment Part I : Model Development. Journal of the American Water Resources Association **34**:73-89.

Arnold, J. G., R. Srinivasan, T. S. Ramanarayanan, and M. DiLuzio. 1999. Water Resources of the Texas Gulf Basin. Wate Science and Technology **39**:121-133.

ASAE. 1998. Manure production and characteristics. Pages 646-648 *in* ASAE Standards 1998, 45th edition, Section D384.1, St. Joseph.

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

Benaman, J. A., and C. A. Shoemaker. 2003a. An analysis of high-flow sediment event data for evaluating model performance. *Accepted* in Hydrological Processes.

Benaman, J. A., C. A. Shoemaker, and D. A. Haith. 2003b. Calibration and validation of a distributed watershed model for basin-wide management. *Accepted* in ASCE Journal of Hydrologic Engineering.

Bingner, R. L. 1996. Runoff Simulated from Goodwin Creek Watershed using SWAT. Transactions of the American Society of Agricultural Engineers **39**:85-90.

Bingner, R. L., J. Garbrecht, J. G. Arnold, and R. Srinivasan. 1997. Effect of Watershed Subdivision on Simulation Runoff and Fine Sediment Yield. Transactions of the American Society of Agricultural Engineers **40**:1329-1335.

Cerosaletti, P. E. 2002. The Delaware County phosphorus reduction through precision animal feeding project: Final technical report. Cornell Cooperative Extension of Delaware County.

Cho, S., G. D. Jennings, C. Stallings, and H. A. Devine. 1995. GIS-Based Water Quality Model Calibration in the Delaware River Basin. American Society of Agricultural Engineers Meeting Presentation:Microfiche No. 95-2404.

Dewing, D. 1998. Unpublished Report. 1998 Watershed Agricultural Program Manure Sampling Project. Watershed Agricultural Program, Cornell Cooperative Extension Delaware County.

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

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

FitzHugh, T. W., and D. S. Mackay. 2000. Impacts of Input Parameter Spatial Aggregation on an Agricultural Nonpoint Source Pollution Model. Journal of Hydrology **236**:35-53.
Ghidey, F., and E. E. Alberts. 1998. Runoff and soil losses as affected by corn and soybean tillage systems. Journal of Soil and Water Conservation **53**:64-70.
Johnson, C. W., N. D. Gordon, and C. L. Hanson. 1985. Northwest rangeland sediment yield analysis by the MUSLE. Transactions of the American Society of Agricultural Engineers **28**:1889-1895.
Kane, K. 1999. Proposed Phase II Phosphorus TMDL Calculations for Cannonsville Reservoir. New York City Department of Environmental Protection.
Kirsch, K., A. Kirsch, and J. G. Arnold. 2002. Predicting sediment and phosphorus loads in the Rock River basin using SWAT. Transactions of the ASAE **45**:1757-1769.
Kleinman, P. J. 1999. Examination of Phosphorus in Agricultural Soils of New York's Delaware River Watershed. PhD. Cornell University, Ithaca.
Kleinman, P. J. A., R. B. Bryant, and W. S. Reid. 1999. Development of pedotransfer functions to quantify phosphorus saturation of agricultural soils. Journal of Environmental Quality **28**:2026-2030.
Longabucco, P., and M. R. Rafferty. 1998. Analysis of Material Loading to Cannonsville Reservoir: Advantages of Event-Based Sampling. Journal of Lakes and Reservoir Management **14**:197-212.
Lumia, R. 1998. Flood of January 19-20, 1996 in New York State. Water-Resources Investigations Report 97-4252, U.S. Geological Society, Albany, NY.
Mamillapalli, S. 1998. Effect of Spatial Variability on Stream Flow Modeling. Ph.D. Thesis. Purdue University.
Manguerra, H. B., and B. A. Engel. 1998. Hydrologic Parameterization of Watersheds for Runoff Prediction Using SWAT. Journal of the American Water Resources Association **34**:1149-1162.
McConkey, B. G., W. Nicholaichuk, H. Steppuhn, and C. D. Reimer. 1997. Sediment yeild and seasonal soil erodibility for semiarid cropland in western Canada. Canadian Journal of Soil Science **77**:33-40.
Nash, J. E., and J. V. Suttcliffe. 1970. River flow forecasting through conceptual models, Part I. A discussion of principles. Journal of Hydrology **10**:282-290.
Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams. 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., J. G. Arnold, J. R. Kiniry, and J. R. Williams. 2001b. Soil and Water Assessment Tool User's Manual Version 2000: Draft-April 2001. US Department of Agriculture - Agricultural Research Service, Temple, Texas.
NRCS. (Unpublished Data). "Computer file of Natural Resource Conservation Service Agricultural Survey in Cannonsville Basin conducted from 1992 to 1994." Provided by Gary Lamont of the NYC Watershed Agricultural Program.
NYCDEP. (Unpublished Data). "Computer file of water quality sampling data covering period from 1988 to 1999. Data retrievals from cumulative stream sampling database." Provided by W. Stasiuk and C. Stepczuk of the New York City Department of Environmental Protection.

NYSDEC. (Unpublished Data). "Computer files of continuous water quality sampling data and corresponding daily sediment and nutrient loading estimates from 1991 to 2000." Provided by Pat Bishop of the NYSDEC.
NYSWRI. 2002a. Forest Soil Sampling at Department of Conservation Research Sites. Delaware County Action Plan for Phosphorus Reduction, Phase IIc., New York State Water Resources Institute, Ithaca, NY.
NYSWRI. 2002b. Refinement of Wastewater Treatment Plant Total Phosphorus Loadings for Cannonsville Basin Modeling - 3rd Draft. Delaware County Action Plan for Phosphorus Reduction, Phase IIb., New York State Water Resources Institute, Ithaca, NY.
NYSWRI and DCPD. (2001). "Quality Assurance Plan for Environmental Monitoring Projects: Delaware County Action Plan for Phosphorus Reduction, Phase IIc: Systematic Monitoring, Modeling and Evaluation for Management in the Cannonsville Reservoir Basin." New York State Water Resources Institute (NYSWRI).
Peterson, J. R., and J. M. Hamlett. 1998. Hydrologic Calibration of the SWAT Model in a Watershed Containing Fragipan Soils. Journal of the American Water Resources Association **34**:531-544.
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. Journal of American Water Resources Association **37**:1169-1188.
Schneiderman, E. M., D. G. Lounsbury, and R. Lohre. 1998. Calibrate and Verify GWLF Models for Remaining Catskill/Delaware Reservoirs. Internal Report New York Department of Environmental Protection, (NYCDEP), Kingston, NY.
Schneiderman, E. M., D. C. Pierson, D. G. Lounsbury, and M. S. Zion. 2002. Modeling the hydrochemistry of the Cannonsville Watershed with Generalized Watershed Loading Functions (GWLF). Journal of the American Water Resources Association **38**:1323-1347.
SCS. 1972. Section 4: Hydrology. *in* National Engineering Handbook, Soil Conservation Service.
Sharpley, A. N., C. A. Jones, C. Gray, and C. V. Cole. 1984. A simplified soil and plant phosphorus model: II. Prediction of labile, organic and sorbed phosphorus. Soil Sci. Soc. Am. J. **48**:805-809.
Shepherd, B., D. Harper, and A. Millington. 1999. Modelling catchment-scale nutrient transport to watercourses in the UK. Hydrobiologia **396**:227-237.
Srinivasan, R., T. S. Ramanarayanan, J. G. Arnold, and S. T. Bednarz. 1998. Large Area Hydrologic Modeling and Assessment Part II : Model Application. Journal of the American Water Resources Association **34**:91-101.
Delaware County Board of Supervisors. 1999. Delaware County Comprehensive Strategy for Phosphorus Reductions. Delhi, NY. http://wri.eas.cornell.edu/projects/nycwshed/delaware/
USDA. 1994. 1992 Census of Agriculture: New York State and County Data. AC92-A-32, United States Department of Agriculture: National Agricultural Statistics Service.
USDA. 1999. 1997 Census of Agriculture: New York State and County Data. AC97-A-32, United States Department of Agriculture: National Agricultural Statistics Service, Washington, D.C.

Williams, J., and J. Arnold. 1993. A System of Hydrologic Models. *in*. Proceedings of the Federal Interagency Workshop on Hydrologic Modeling Demands for the 90's, Water-Resources Investigations Report 93-4018, U.S. Geological Survey.
Williams, J. R., and H. D. Berndt. 1977. Sediment Yield Prediction Based on Watershed Hydrology. Transactions of the ASAE **20**:1100-1104.
Williams, J. R., C. A. Jones, and P. T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity. Transactions of the American Society of Agricultural Engineers **27**:129-144.
Yanai, R. D. 1992. Phosphorus budget of a 70-year-old northern hardwood forest. Biogeochemistry **17**:1-22.

10 Appendices

Discussions about data sharing are currently underway. Therefore, not all input data derived for the model are provided in this report. To obtain the latest information regarding further data/model input availability, contact Bryan Tolson (bat9@cornell.edu).

10.1 Soil Property Derivation from SSURGO and STATSGO Intersection

The methodology used in the calculation of area- and depth-weighted soil properties for STATSGO soil map units from the SSURGO soils database is explained as follows with reference to Figure 10.1.1:

- All soils in the model are represented with 4 soil layers.
- The soil layer boundaries for each STATSGO map unit were calculated as area-weighted averages of the Cannonsville Basin SSURGO soil series within the STATSGO map unit.
- SSURGO soil layers that occurred outside of the idealized, area-weighted average soil layer depths (i.e. boundaries in Figure 10.1.1) did not influence the average soil properties computed for that layer.
- Figure 10.1.1 shows the averaged representation for STATSGO soil map unit NY056. As an example, consider the SSURGO soil series ElE (11^{th} from left) in Figure 10.1.1. The properties in the third layer in ElE do not impact the average properties calculated for layer 3 of the NY056 map unit – instead layer 2 properties of ElE are used in the computation of average properties for layer 3 of NY056. Similarly, the ElE layer 2 and 3 properties are depth-weighted (2/5 for layer 2, 3/5 for layer 3) in order to calculate soil properties for ElE in the idealized 4^{th} layer of NY056. Then the depth-weighted average ElE properties for the idealized 4^{th} layer are used in the area-weighted computation of average soil properties for layer 4 in NY056.
- Bedrock occurs as the last soil layer in 20% of the SSURGO soil data available for the basin. In these cases, the depth to bedrock was less than 2 m below the soil surface. The only available soil properties in the SSURGO database for the bedrock layers are the layer depth and the saturated hydraulic conductivity. Therefore, these bedrock layers were not included in the calculation of the average soil layer properties.

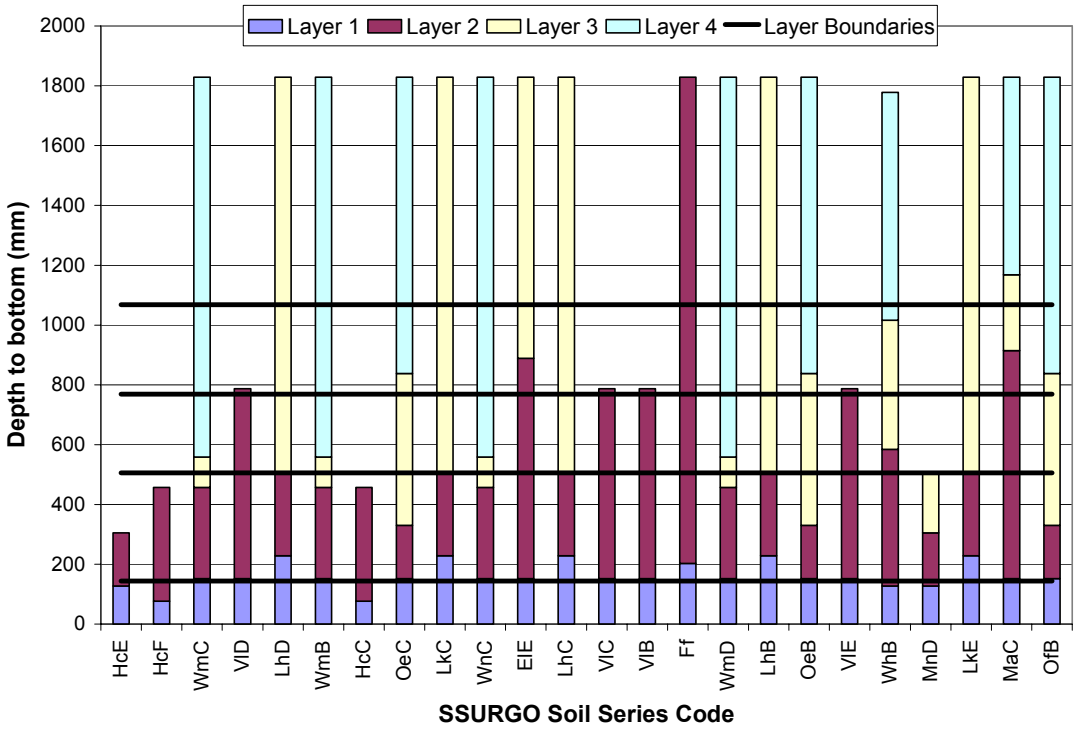

Figure 10.1.1. Averaged representation of the soil profile for STATSGO map unit NY056.

10.2 Precipitation and Temperature Data Adjustments and Corrections

The details provided in Sections 10.2.1, 10.2.2 and 10.2.3 below expand on the climate data adjustments introduced briefly in Sections 4.4.1.1, 4.4.1.2 and 4.4.1.3, respectively.

10.2.1 Details on Walton Precipitation Data 'Observer Shifting' Correction

The observer shifting problem at the Walton2 precipitation station is not consistent from January 1998 through September 2000. Some portions of the precipitation record over this period do not exhibit this problem. Therefore, this problem could not be solved by simply shifting all precipitation recorded at Walton2 after January 1998 by one day. Instead, the precipitation over this period was closely examined to identify periods in time where the recorded data required 'unshifting' (moving the data one day forward) so that it was in phase with the data at Delhi and Deposit. In total, 84% of the days at Walton2 had recorded precipitation out of phase with Delhi and Deposit.

Table 10.2.1 shows the calculated coefficients of determination (r^2) between various precipitation depths recorded at pairs of precipitation stations over different time periods. With the original precipitation data, from 1990-1997, all three pairings of the precipitation stations (Walton, Delhi and Deposit) show highly correlated daily precipitation depths with r^2 values over 0.59. From 1998-2000, the original precipitation data at Delhi and Deposit are similarly correlated. However, the original precipitation data at Walton2 from 1998-2000 show almost no correlation with the Delhi and Deposit stations, as the r^2 coefficient is 0.06 for both pairings. These low r^2 coefficients in comparison with the high values for the 1990-1997 period, as well as the high values for Delhi versus deposit from 1998-2000, clearly demonstrate that a problem exists at Walton2. After correction of the observer shifting problem at Walton2, Table 10.2.1 shows that much higher r^2 coefficients (≥ 0.60) are observed between Walton2 and the other stations. The higher r^2 coefficients are a good indication that the corrective procedure implemented here worked as desired.

Table 10.2.1. Correlation analysis of precipitation depths at Walton, Delhi and Deposit for the original data and the corrected data at Walton after January 1998.

Original Precipitation Data			**Precipitation Data After Correction**		
Stations Correlated	**Period of Comparison**	**r^2**	**Stations Correlated**	**Period of Comparison**	**r^2**
Walton[A] vs Delhi	1990-1997	0.77			
Walton[A] vs Deposit	1990-1997	0.63			
Delhi vs Deposit	1990-1997	0.59			
Walton2 vs Delhi	1998-2000	0.06	Walton2 vs Delhi	1998-2000	0.63
Walton2 vs Deposit	1998-2000	0.06	Walton2 vs Deposit	1998-2000	0.60
Delhi vs Deposit	1998-2000	0.70			

A) In this period, Walton precipitation data are from the original Walton and Walton2 precipitation stations.

In addition to this chronic problem after January 1998, this issue appeared sporadically before 1998 at the Walton precipitation station. There were seven significant precipitation events at the Walton precipitation station that were recorded out of phase with the Delhi and Deposit stations and were deemed important to correct. These events were identified for correction because they each had the following characteristics:

- Walton peak storm precipitation was greater than 10 mm.
- The Walton peak precipitation was at least twice as large as the precipitation recorded on the preceding and following days at Walton. This criterion avoided excessive modifications to the Walton precipitation data during multi-day precipitation events with a relatively constant daily rainfall rate.
- Walton precipitation occurred one day before both the Delhi and Deposit peak precipitation.

Table 10.2.2 below outlines the original precipitation depths recorded at Delhi, Deposit and Walton and shows how the Walton precipitation depths were modified for the seven events at Walton prior to January 1, 1998. The adjustment to the precipitation data at Walton only functioned to shift the out of phase precipitation recorded at Walton to a different day and did not change the total depth of precipitation recorded at Walton. The seven precipitation events are distinguished from one another in Table 10.2.2 by the horizontal lines. The Walton precipitation was adjusted (Walton – FIX column in Table 10.2.2) so that each storm event at Walton looked like the average storm event that occurred at Delhi and Deposit.

Table 10.2.2. Precipitation adjustments at Walton prior to January 1, 1998.

		Original precip data in mm			
Year	Julian day	Deposit	Delhi	Walton	Walton - FIX (mm)
1995	181	0.0	0.0	0.0	0.0
1995	182	0.0	0.1	19.8	0.0
1995	183	3.8	26.4	7.1	26.9
1995	184	0.0	0.0	0.0	0.0
1996	301	0.0	0.0	0.5	0.0
1996	302	1.3	1.5	10.9	0.5
1996	303	9.1	6.9	0.0	10.9
1997	34	0.0	1.3	0.0	0.0
1997	35	0.0	0.0	15.7	0.0
1997	36	14.0	18.5	0.5	15.7
1997	37	1.0	0.3	0.0	0.5
1997	128	0.0	0.0	1.5	0.0
1997	129	3.6	2.5	11.7	2.8
1997	130	8.4	22.6	2.0	11.0
1997	131	1.8	4.3	0.8	2.2
1997	138	0.0	2.3	1.8	1.1
1997	139	20.1	9.7	36.3	13.2
1997	140	23.9	29.0	0.0	23.8
1997	201	0.0	0.0	0.0	0.0
1997	202	0.0	0.3	14.2	0.0
1997	203	8.4	10.2	0.0	14.2
1997	207	0.0	0	0	0
1997	208	0.0	0.0	17.0	0.0
1997	209	9.4	1.5	0.1	17.1
1997	210	0.0	0.0	0.0	0.0

10.2.2 Details on Temperature Adjustments for Correct Precipitation Type Classification

The methodology applied in this study to force the correct classification of the precipitation type in SWAT is outlined here. Rather than modify the model source code it was determined that the SWAT SFTMP parameter could be fixed at 1° C (SWAT default value) and then small adjustments could be made to the temperatures input to the model in order to force the model to correctly classify the precipitation type at Walton and Delhi. Based on the fixed SFTMP value and the time series of precipitation minimum and maximum daily temperatures at either Delhi or Walton, the type of precipitation the model would simulate was assessed (outside of the model) and then compared to available measured data on snowfall depths to determine if the model classification would be incorrect. This comparison and subsequent adjustment methodology were implemented as follows:

- Adjustments to temperature inputs were only made on days at Walton and/or Delhi when there was at least 2.5 mm of water equivalent precipitation recorded. The assumption here is that misclassification of the precipitation type in small precipitation events does not warrant changing the measured temperatures.
- When considering a change to force model to classify precipitation as rain, the measured snowfall depth must have been < 2 times the precipitation water equivalent depth. In other words, if a typical snowfall density was assumed to be 10% (e.g. 1 inch deep snow = 0.1 inch melted water), then 80% or more of the precipitation for the day would have had to fall as rain to trigger an adjustment to the minimum and maximum observed temperature data.
- When considering a change to force model to classify precipitation as snow, the measured snowfall depth must have been > 6 times the precipitation water equivalent depth. In other words, if a typical snowfall density was assumed to be 10% (e.g. 1 inch deep snow = 0.1 inch melted water), then 60% or more of the precipitation for the day would have had to fall as snow to trigger an adjustment to the minimum and maximum observed temperature data.
- When the above criteria were met and temperatures were to be adjusted, the temperature station average temperature was fixed at 0.1 °C higher or lower than SFTMP and the adjusted minimum and maximum daily temperatures were calculated as follows:
    - When forcing precipitation to fall as rain, the daily average temperature was fixed at 1.1 °C, the adjusted maximum daily temperature was set as the max (the measured maximum for the day, 2.2 °C) and the adjusted minimum daily temperature was calculated from the adjusted average and maximum temperatures. This approach limited the increase in the maximum temperature so that the impact of this temperature adjustment on snowmelt was minimized.
    - When forcing precipitation to fall as snow, the daily average temperature was fixed at 0.9 °C, the maximum daily temperature was not adjusted and the minimum daily temperature was calculated from the adjusted average and unadjusted maximum temperatures.

10.2.3 Details on Climate Data Adjustments for Large Precipitation Events

Adjustments to climate data outlined in this section were mainly identified based on supplementary information from descriptions of large storms events from the NCDC Storm Event database (http://www4.ncdc.noaa.gov/cgi-win/wwcgi.dll?wwevent~storms) for New York unless otherwise noted below. In some cases hourly precipitation data from a nearby climate station (Sydney, New York) also highlighted the need to modify the climate data. In other cases the adjustment was guided by nearby climate stations within the Cannonsville Basin. All of these cases where climate adjustment was required for large storm events are described below in Table 10.2.3.

Table 10.2.3. Climate input data adjustments and justification for large precipitation events.

Climate Station	Date (Yr-M-D)	Original Input Data			Adjusted Input Data			Justification for Adjustment of Climate Inputs
		Max Temp (°F)	Min Temp (°F)	Precip. (in.)	Max Temp (°F)	Min Temp (°F)	Precip. (in.)	
Walton	1994-8-18	-	-	1.15	-	-	1.64	Flash flood recorded starting at 10 am Aug 18. Walton peak precip. not on same day as Delhi and Deposit. Make Walton precip. occur like avg. storm recorded at Delhi and Deposit. 57.5% of Delhi & 75.8% of Deposit event precip. on 18th, rest on 19th. Keep event precip. total at Walton same but make 66.7% event precip. on 18th , rest on 19th for Walton.
Walton	1994-8-19	-	-	1.31	-	-	0.82	
Walton	1995-11-15	38	31	1.15	38.0	29.4	-	Heavy snowfall event: assume Walton precip. fell as snow, adjust temperatures according to procedure in Section 10.2.2.
Delhi	1995-11-15	37	30	0.86	37.0	30.4	-	Same as above for Delhi.
Walton	1996-1-19	57	27	0.26	-	-	2.41	1 in 75 year flood event, high rainfall. All precip. from 19th and 20th moved to occur on 19th for Delhi and Walton so that observed temperatures make it fall as rain instead of snow. Stamford precip. good since mainly fell on 19th.
Delhi	1996-1-19	55	44	0.28	-	-	2.33	
Walton	1996-1-20	27	13	2.16	-	-	0.0	
Delhi	1996-1-20	56	13	2.05	-	-	0.0	
Walton	1996-1-27	46	20	0.62	45.7	28	-	27th and 28th are heavy rain from Spring-like storm. Adjust temperatures at Walton & Delhi such that due to orographic effects, all locations in the basin with elevations less than the average basin elevation will receive rainfall on these days. Locations higher than avg. elevation receives snow.
Delhi	1996-1-27	45	19	0.65	45	28	-	
Walton	1996-1-28	36	17	0.90	51	22.7	-	Assume most rain fell early in day - based on flood time, Walton temperature is 10 hrs behind precip. reporting time so replace Walton max temp. with Delhi max temp.
Delhi	1996-1-28	51	16	0.62	51	22	-	Same as 27th above for Delhi.
Stamford	1999-1-16	-	-	2.61	-	-	0.67	Hourly precip. from East Sidney shows 80% of storm precip. there fell over 14 hour period. Also, % of storm precip on peak day (17th) for Delhi = 78% and Walton2=80% compared to Stamford (64%). Make peak precip. at Stamford higher and look like avg. of Delhi & Walton: 79% of 15th–17th rain (4.08 in) moved to 17th. Total Stamford precip. unchanged.
Stamford	1999-1-17	-	-	1.28	-	-	3.22	
Delhi	1999-3-22	43	30	1.53	43.0	24.4		Heavy snowfall event: assume Delhi precip. fell as snow, adjust temperatures according to procedure in Section 10.2.2.
Walton2	1999-3-22	44	20	1.33	44.0	23.4		Same as above for Walton.
Walton2	1999-3-23	42	27	0.66	42.0	25.4		Same as above for Walton (note Delhi climates good this day).
Delhi	2000-4-9	73	22	1.34	-	-	0.09	Precip. started as rain but quickly went to snow and then very heavy snow with a max of 15 in. in Del County for event. Temperatures on 9th too high to modify so make bulk of precip. fall on the 10th such that at a snow density of 10%, 15 inches of snow = 1.5 in of water equivalent. 1.5 in of precip. across basin on the 10th. Total event precip. unchanged.
Walton2	2000-4-9	74	23	1.28	-	-	0.12	
Stamford	2000-4-9	-	-	3.18	-	-	1.7	
Delhi	2000-4-10	31	22	0.25	-	-	1.5	
Walton2	2000-4-10	33	22	0.34	-	-	1.5	
Stamford	2000-4-10	-	-	0.02	-	-	1.5	

10.3 Model Performance Statistics

The three numerical model performance measures used in this study are the percent difference (D), coefficient of determination (r^2 coefficient) and the Nash-Suttcliffe simulation efficiency (E_{NS}) (Nash and Suttcliffe 1970).

The percent difference measures the average difference between the simulated and measured values for a given quantity over a specified period (usually the entire calibration or validation period in this study). The percent difference for a quantity (D) over a specified period with n total days is calculated from the measured and simulated values of the quantity in each model time step as:

$$D = \frac{100 \times \left(\sum_{j=1}^{n} Simulated_j - \sum_{j=1}^{n} Measured_j \right)}{\sum_{j=1}^{n} Measured_j}$$

A value close to 0% is best for D. However, higher values for D are acceptable if the accuracy in which the measured data were gathered is relatively poor.

The r^2 coefficient and E_{NS} simulation efficiency measure how well the trends in the measured data are reproduced by the simulated results over a specified time period and for a specified time step. For example, in this study, these measures were computed for a daily, monthly and sometimes annual time step.

The r^2 coefficient for n time steps is calculated as:

$$r^2 = \frac{\left[\sum_{i=1}^{n} (simulated_i - simulated_{avg})(measured_i - measured_{avg}) \right]^2}{\sum_{i=1}^{n} (simulated_i - simulated_{avg})^2 \sum_{i=1}^{n} (measured_i - measured_{avg})^2}$$

The range of values for r^2 is 1.0 (best) to 0.0. The r^2 coefficient measures the fraction of the variation in the measured data that is replicated in the simulated model results. A value of 0.0 for r^2 means that none of the variance in the measured data is replicated by the model predictions. On the other hand, a value of 1.0 indicates that all of the variance in the measured data is replicated by the model predictions.

The E_{NS} simulation efficiency for n time steps is calculated as:

$$E_{NS} = 1 - \frac{\sum_{i=1}^{n} (measured_i - simulated_i)^2}{\sum_{i=1}^{n} \left[measured_i - 1/n \sum_{i=1}^{n} measured_i \right]^2}$$

E_{NS} values range from 1.0 (best) to negative infinity. E_{NS} is a more stringent test of performance than r^2 and is never larger than r^2. E_{NS} measures how well the simulated results predict the measured data relative to simply predicting the quantity of interest by

using the average of the measured data over the period of comparison. A value of 0.0 for E_{NS} means that the model predictions are just as accurate as using the measured data average to predict the measured data. E_{NS} values less than 0.0 indicate the measured data average is a better predictor of the measured data than the model predictions while a value greater than 0.0 indicates the model is a better predictor of the measured data than the measured data average.

10.4 Additional Details on Derivation of Select Data-Driven Parameter Values or Ranges

10.4.1 SFTMP Parameter

SFTMP is the mean air temperature at which precipitation is equally likely (e.g. probability of 0.5) to be rain as snow/freezing rain. Thus, if a climate station has temperature and precipitation depths and type (as snow or rain) the frequency of snow or rainfall can be estimated for each discrete temperature measurement. Analysis of climate data at Walton and Delhi enabled the identification of the range of values for SFTMP specific to the Cannonsville Basin. The results of this analysis are summarized in Table 10.4.1 below. For example, for the Walton station from 1956 to 1990, of the 130 days when the temperature was recorded at 1.7 °C, the precipitation fell mainly as rain in 57 days. Thus, the frequency of rain at 1.7 °C is then 57/130=0.44. Since only one value of SFTMP can be input to SWAT the results of the analysis at both stations must be aggregated. The highlighted cells at each station yield the possible values of SFTMP where the observed frequency of rain is approximately equal to the observed frequency of snow. This gives a possible SFTMP range of approximately 0 to 2.2 °C.

A) Walton

	Record: 1956 - 1999		Record: 1990 - 1999	
Avg Daily Temp (deg C)	# Days with Precip as Rain	Fraction Precip days as Rain	# Days with Precip as Rain	Fraction Precip days as Rain
-1.7	16	0.13	6	0.19
-1.1	10	0.08	6	0.18
-0.6	19	0.18	6	0.21
0	26	0.27	10	0.31
0.6	39	0.27	18	0.44
1.1	45	0.34	15	0.43
1.7	57	0.44	18	0.56
2.2	56	0.50	14	0.70
2.8	63	0.59	23	0.74

B) Delhi

	Record: 1926 - 2000		Record: 1990 - 2000	
Avg Daily Temp (deg C)	# Days with Precip as Rain	Fraction Precip days as Rain	# Days with Precip as Rain	Fraction Precip days as Rain
-1.7	34	0.31	6	0.33
-1.1	52	0.36	6	0.27
-0.6	51	0.40	11	0.38
0	55	0.45	12	0.57
0.6	65	0.52	11	0.42
1.1	101	0.66	23	0.70
1.7	106	0.74	22	0.85

Table 10.4.1. Analysis of (A) Walton and (B) Delhi climate station data for range of possible SFTMP values.

10.4.2 Locally Derived Cattle Feed Estimate

Based on discussions with Co-operative Extension scientist Paul Cerosaletti, factors were identified to estimate the amount of dry biomass consumed by cattle in the Cannonsville Basin. This consumption rate was then used to guide the setting the SWAT2000 model inputs controlling the amount of harvested and grazed dry biomass. The purpose of this comparison was to ensure that the annual average simulated amount of harvested plus grazed dry biomass was reasonably close to the amount of dry biomass that the cattle in the basin are estimated to consume each year.

The assumed factors used to estimate the total locally grown dry matter consumed by the basin cattle population are:

- Dry matter consumed/day as a % of animal mass by milking dairy cows = 3.3.
- Dry matter consumed/day as a % of animal mass by other cattle = 2.5.
- Dairy and beef population estimates derived in Section 4.7.1.1.
- Animal masses given in Table 4.7.1.

- 85% of dairy cows are milking.
- 50% of total dry matter intake for basin cattle derived from locally grown forage.

Based on the above factors, in 1997, the total mass of the milking dairy cow herd was estimated as $4.2x10^6$ kg and the total mass of the rest of the dairy and beef cattle herds was estimated as $3.5x10^6$ kg. The resulting total dry matter consumption rate for the basin was estimated at $8.1x10^7$ kg/yr, 50% of which ($4.0x10^7$ kg/yr) was derived from locally grown forage.

The final values of the harvest indices for pasture and hay HRUs given in Table 4.7.4 were set with reference to the estimated $4.0x10^7$ kg/yr of locally grown forage over the calibration period. The calibrated model simulated a total of $5.2x10^7$ kg/yr of harvested and grazed dry biomass removed from basin soils. Of this simulated dry biomass total, 56% is from harvested hay, 25% is from grazed pasture and 19% is from harvested corn silage.

The simulated dry biomass removal rate of $5.2x10^7$ kg/yr from basin soils would account for approximately 65% of the estimated basin-wide cattle population dry matter consumption. This is somewhat higher than the model independent estimated locally grown dry matter consumption of 50% (or $4.0x10^7$ kg/yr) of the total dry matter intake of the cattle population. This model result was deemed reasonable for two reasons. First, the harvest indices in Table 4.7.4 are already set to relatively low values so reducing them further was considered unreasonable. Secondly, the manure production rate in Table 4.7.2 includes animal bedding (hay) that is not fed to animals but is harvested from local fields. Therefore, the simulated biomass removal rate of $5.2x10^7$ kg/yr from basin soils includes some biomass that is not fed to the local cattle population.

The harvest indices in Table 4.7.4 were used in both the calibration and validation period. Therefore, during the validation period, although dry biomass was removed from basin soils at approximately the same rate as in the calibration period, the locally grown forage content of the basin-wide dry matter intake of cattle was reduced relative to the calibration period since the cattle population was higher in validation. During the validation period (1990-1993), the simulated dry biomass removal rate accounted for approximately 55% of the estimated basin-wide cattle population dry matter consumption (compared to 65% in the calibration period). In other words, this approach simulates a higher proportion of imported cattle feed in the basin cattle diet from 1990-1993 compared to 1994-2000.

10.5 Additional Time Series Plots of Measured and Simulated Quantities

10.5.1 Daily Measured and Simulated Flows at Walton

The calibration and validation scatter plots of the daily Walton flows are shown in Figure 10.5.1. Time series of the simulated and measured daily flows by water year at Walton for the entire calibration period are given in Figure 10.5.2

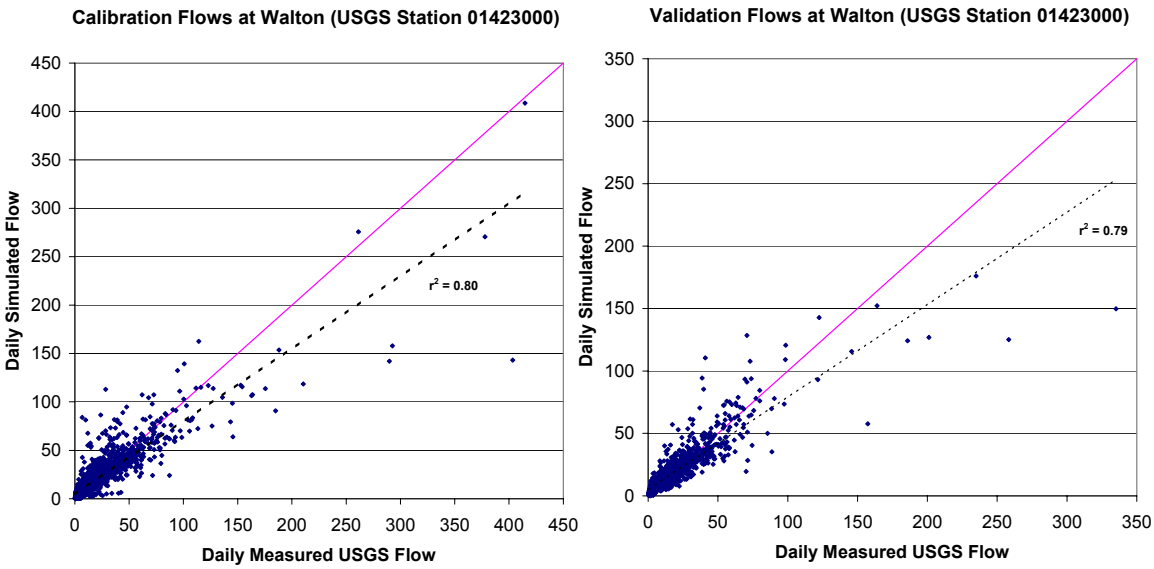

Figure 10.5.1. Scatter plots measured and simulated of Walton flows for the calibration and validation period.

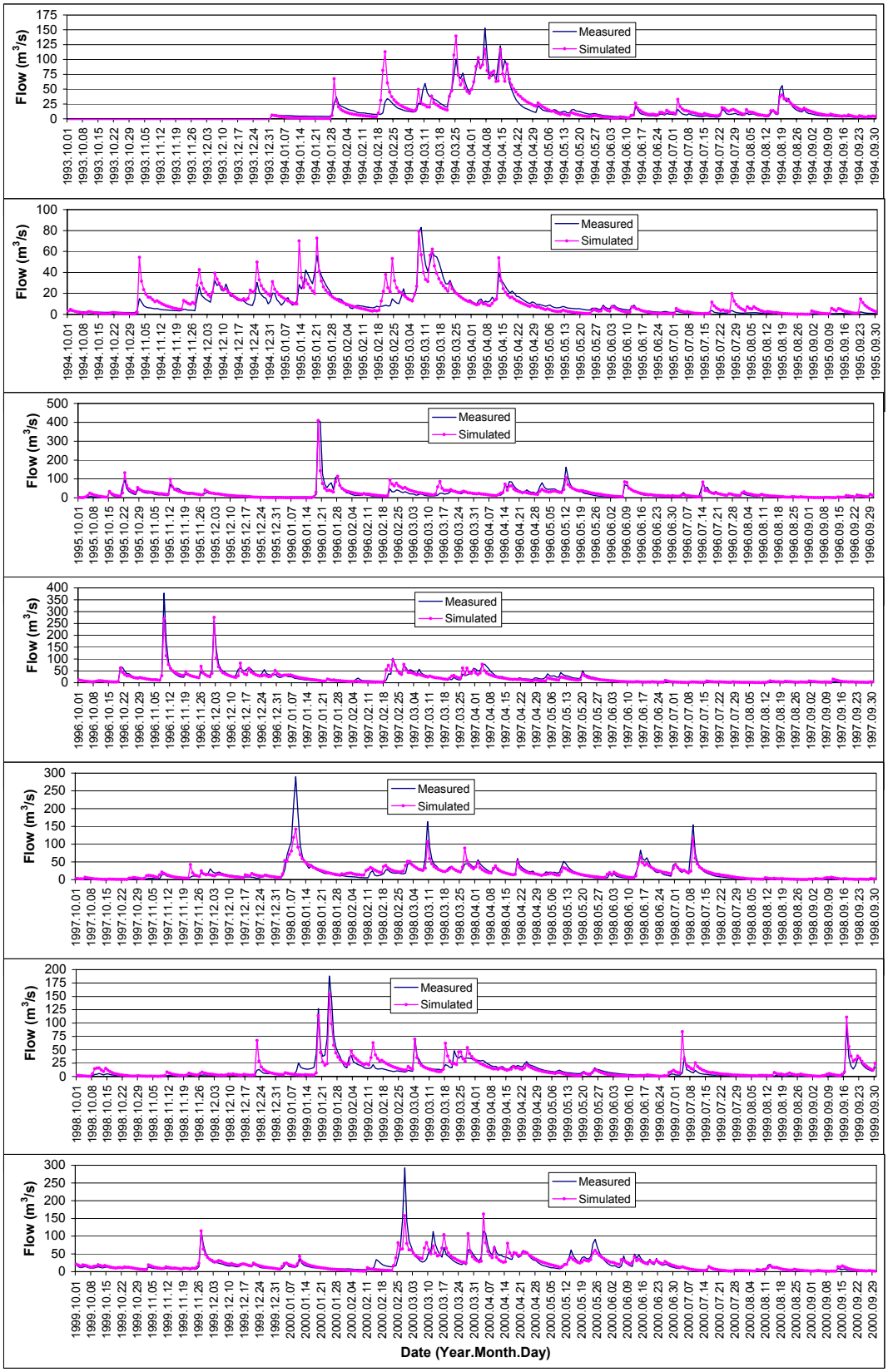

Figure 10.5.2. Time series of measured and simulated daily flows for calibration period.

10.5.2 Monthly Measured and Simulated Average Flows

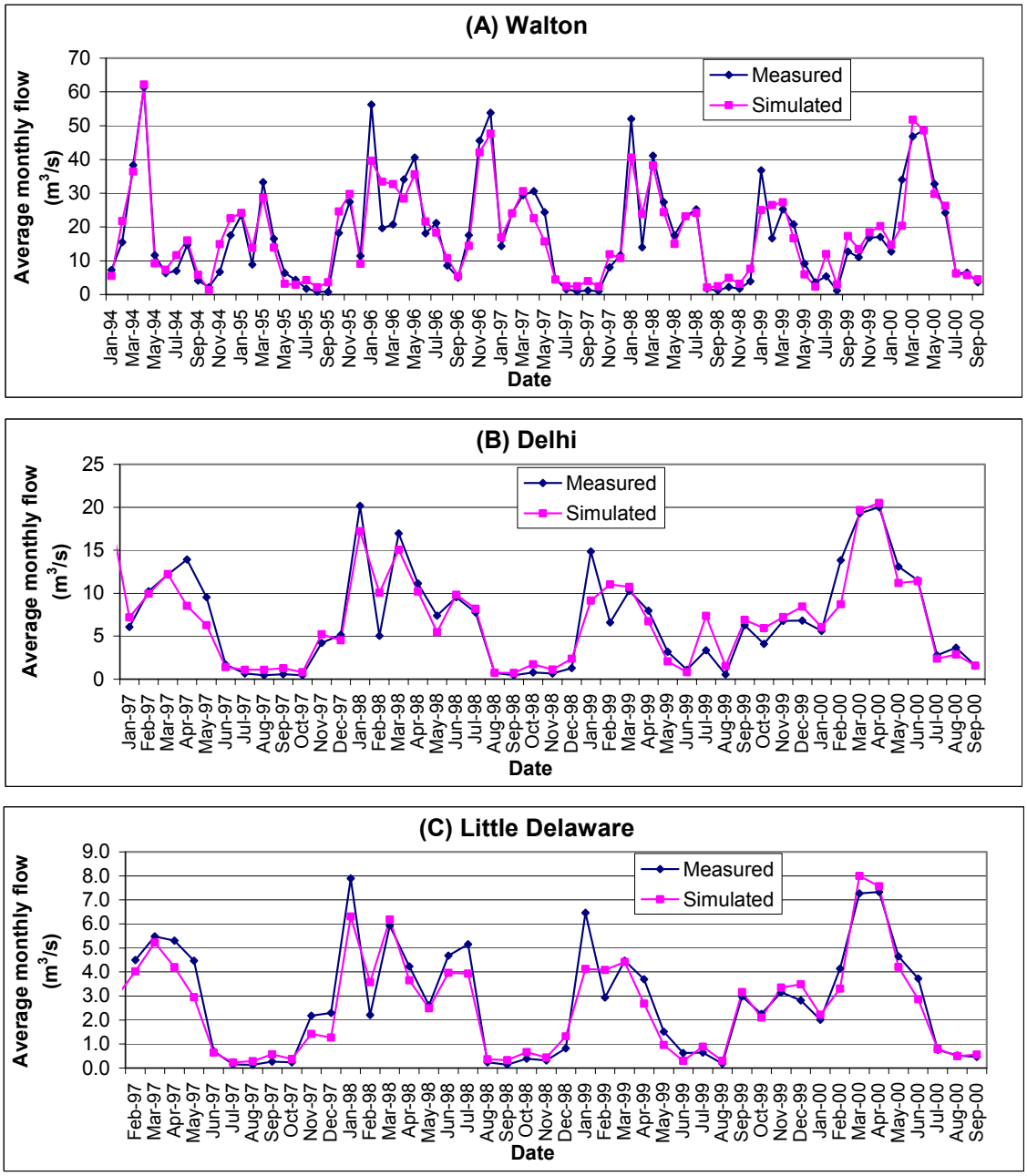

Figure 10.5.3. Time series of monthly measured and simulated hydrology calibration results at A) Walton B) Delhi and C) Little Delaware for the period 1994 to 2000.

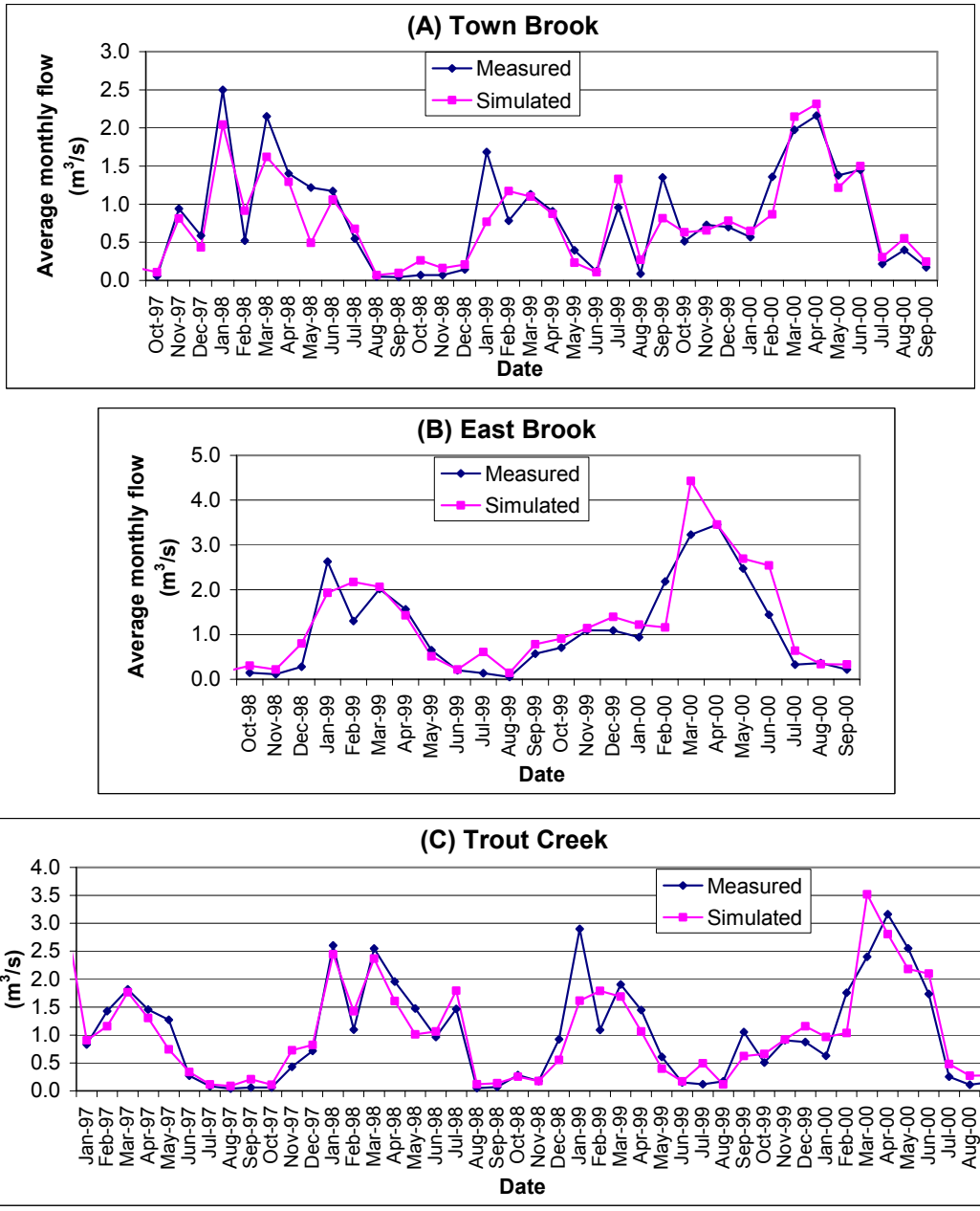

Figure 10.5.4. Time series of monthly measured and simulated hydrology calibration results at A) Town Brook B) East Brook and C) Trout Creek for the period 1994 to 2000.