ABSTRACT

Title of Document: EXAMINATION OF A GIS-BASED WATER QUALITY MODEL USING USGS GAGED WATERSHEDS IN MARYLAND

Dorianne E. Shivers, Master of Science, 2007

Directed By: Professor Glenn E. Moglen, Civil and Environmental Engineering

Water quality models are important tools used by the Maryland Department of the Environment (MDE) in developing Total Maximum Daily Loads (TMDLs), which serve as water quality standards. The MDE tool, which spatially interpolates output from the Chesapeake Bay Program Watershed Model (WSM), is often used because it requires little time, data, or training. In contrast, the WSM requires extensive time, data, and training to run.

This study examines if the MDE tool provides accurate estimates of pollutant loads and whether the mid-level complexity model AVGWLF provides comparatively more accurate estimates. The accuracy of the models was assessed based on qualitative comparisons, t-tests, and Nash-Sutcliffe coefficients. The MDE tool was found to more accurately predict total nitrogen and total sediment loads and the AVGWLF model was found to more accurately predict total phosphorus loads. The study also found that a consistent method for calculating observed loads needs to be developed.

EXAMINATION OF A GIS-BASED WATER QUALITY MODEL USING USGS GAGED WATERSHEDS IN MARYLAND.

By

Dorianne E. Shivers

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2007

Advisory Committee:
Professor Glenn E. Moglen, Chair
Professor Richard H. McCuen
Professor Kaye Brubaker

Acknowledgements

I would like to thank my advisor Dr. Glenn E. Moglen for his assistance and guidance during this research. I would also like to express appreciation to my committee, Dr. Kaye Brubaker and Dr. Richard H. McCuen, for their time and their comments.

I would also like to thank my family and friends for their moral support and my office mates Anubha, Alfonso, David, Jeff, Karthik, Melih, Sandra, and Sarah for their help and for making the office an enjoyable place to work. I would especially like to thank Justin for his support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Concern for increasing water pollution in the 1970's led to the Clean Water Act (CWA) of 1972. The CWA and subsequent amendments require that all waters in the nation meet specific pollution reduction goals that guarantee that the water is "fishable and swimmable." Section 303(d) of the CWA requires that states, territories, and tribes identify impaired or damaged water bodies and establish Total Maximum Daily Loads (TMDLs) for each pollutant. These TMDLs serve as water quality standards (USEPA, 1999).

The United States Environmental Protection Agency (USEPA) reports that excessive amounts of nitrogen and phosphorus are the cause of most surface water impairments in the United States. The USEPA estimates that 50 percent of lakes and 60 percent of rivers are impaired by too much nitrogen and phosphorus, which causes eutrophication or nutrient pollution (Carpenter et al., 1998).

Eutrophication can impair water bodies in several ways. For example, eutrophication results in increased algae and plant growth. The amount of algae can increase so much that it blocks sunlight, which is needed by submerged aquatic vegetation (SAV) in order to photosynthesize. If enough sunlight is obscured then the SAV will die (CBP, 2005). Bacterial decomposition of the SAV and the algae, once it dies, uses oxygen and can result in severely depleted levels of dissolved oxygen. This causes stress for the aquatic life in the water body, which needs oxygen for respiration. In addition, low dissolved oxygen can result in the release of certain toxins from the sediment. Also, the increase in bacterial decomposition results in an increase of un-ionized ammonia, which can have negative impacts on aquatic life (USEPA, 1999).

Another major source of surface water impairment is excessive sediment, which is "caused by erosion from agriculture, logging and construction" (Carpenter et al., 1998). In agriculturally-dominated watersheds, excessive sediment is responsible for even more surface water impairments than excessive nitrogen and phosphorus (Mattikalli et al., 1996). Excessive sediment is harmful because it increases the turbidity of the water, blocking sunlight from SAV. Sediment also carries with it other materials, such as, phosphorus or pathogens, which can be harmful to aquatic life in the water body. In addition, it can bury bottom-dwelling plants and animals (USGS, 2005b).

1.2 Chesapeake Bay

The Chesapeake Bay is one of the largest and most diverse estuaries in the world. Its watershed includes 6 states and more than 3,600 species of plants and animals live within its watershed (CBP, 2004). The Chesapeake Bay is also considered an impaired water body based on the criteria of the CWA. According to the USGS, the Bay is impaired because of low dissolved oxygen caused by eutrophication (2007c) and low clarity caused by excessive sediment and nutrients (2007d).

In 1976, a six-year study of water quality in the Chesapeake Bay concluded that eutrophication and turbidity were causing a decrease in the number and diversity of aquatic life and SAV. In order to solve this issue the Chesapeake Bay Agreement was adopted in 1987. The program is an agreement between Maryland, Pennsylvania, Virginia, Washington, D.C., the Chesapeake Bay Commission, and the USEPA to regulate nutrients entering the Bay and reduce them by 40 percent (from 1985 levels) by 2000 (MCE, 2006).

1.2.1 Sources of Nutrients and Sediment

There are many sources for nutrients and sediment in the Chesapeake Bay watershed. These sources can be either point sources or non-point sources. Point sources are sources that have an identifiable physical location. Examples include wastewater treatment plants, factories, and industries. Point sources are easy to regulate because they have a known location and the government mandates an allowable load through a permitting process (FWS, 2007).

Non-point sources (NPS) do not have a specific physical location and therefore are not easily regulated. Pollution from NPS results when rainwater runs off the land and carries pollutants with it. These pollutants can end up in local waterways or groundwater depending on where the rainwater runoff goes (FWS, 2007). NPS can be regulated through the use of voluntary and mandatory best management practices (BMPs).

1.3 Calculating Pollutant Loads

In order to ensure that TMDLs are met, the current pollutant loads for a water body must be known. The USEPA describes two techniques for estimating pollutant loads. The first is to use in-stream monitoring data to estimate pollutant loads. The second is to use a water-quality model to estimate pollutant loads (USEPA, 2005).

1.3.1 In-Stream Monitoring

In-stream monitoring data measures actual in-stream pollutant concentrations. The pollutant load can be calculated by multiplying the pollutant concentration by the instantaneous discharge. The advantage of in-stream monitoring is that the estimated pollutant loads represent actual data. The disadvantages are that the pollutant loads cannot be attributed to a particular area or source and future loadings cannot be predicted (USEPA, 2005).

Another disadvantage of in-stream monitoring is that samples are typically taken periodically (e.g., weekly or monthly) and sometimes are sporadic with large gaps in data. The USEPA suggests that rating curves between pollutant loads and discharge be developed so that discharge values can be used to estimate pollutant loads when water quality data are unavailable. This method of developing rating curves is referred to as the "EPA method" in this study.

The EPA method of developing rating curves is statistically incorrect because the relationship is developed between pollutant load, which is found by multiplying pollutant concentration and discharge, and discharge. Therefore the rating curve represents the relationship between discharge and discharge times a variable. This results in inflated goodness-of-fit statistics and possibly incorrect regression coefficients (McCuen and Surbeck, 2007). The statistically correct rating curve would represent the relationship between pollutant concentration and discharge. This method of developing a rating curve will be referred to as the "concentration-derived method" in this study.

1.3.2 Water Quality Modeling

A water-quality model is "a representation of an environmental system through the use of mathematical equations or relationships" (USEPA, 2005). The advantages of using models to predict pollutant loads are that loads can be attributed to a certain source or area. In addition, models can predict future loadings and estimate loading changes due to changes in the watershed (USEPA, 2005). The disadvantages of models are that they may require a lot of data and therefore time and effort to run. In addition, the model is only a representation of a real system and it is important to be aware of the shortcomings of the model.

There are two classes of models: stochastic and deterministic. Most water quality models are deterministic, meaning that the same set of inputs will always return the same set of outputs. A stochastic model, such as, the SPARROW model (Schwarz et al., 2006), accounts for random occurrences and will give different outputs for different trials, even if the input remains constant (Tim and Crumpton, 2003; Lowrance, 2003).

The two classes of models can be further divided into lumped parameter or distributed parameter. A lumped parameter model treats the model area as homogeneous and has the same inputs for the whole area. In contrast, a distributed parameter model partitions the model area by land use/cover, soil type, slope, etc. The different partitions have different inputs (Lowrance, 2003). A distributed parameter model is a more realistic representation of the real world and allows the user to see the effects of changing certain aspects of the real world, such as, land use/cover.

Models can also be continuous or event-based. A continuous model represents the full range of hydrologic conditions, including processes during and between storms. An event-based model represents processes that occur during a precipitation event, such as, a hurricane, or snow melt (Lowrance, 2003).

Water-quality models can also have different levels of complexity or sophistication. An example of a simple model is the P-load model (USEPA, 2001b). P-load calculates pollutant loads by multiplying the area of a particular land use by the pollutant loading rate of that particular land use (USEPA, 2001b). A more sophisticated model attempts to represent physical processes, such as infiltration, evaporation, etc. In theory, a more complex model would be more accurate. In practice, this is not always the case because increasing complexity can lead to irrationalities. In addition, complex models require more data, time and skill to operate (Tim and Crumpton, 2003).

1.3.3 HSPF

The Maryland Department of the Environment (MDE) is in charge of ensuring that Maryland's water bodies are in compliance with the CWA. MDE uses water quality models to estimate current pollutant loads and the resulting future pollutant loads due to potential future changes in the watershed. MDE recently decided to use the Chesapeake Bay Program Model (WSM) (Hopkins et al., 2000) in order to be consistent with the Chesapeake Bay Program (CBP). The WSM model is based on the Hydrologic Simulation Program – FORTRAN (HSPF) model (Bicknell et al., 1997; Shoemaker et al., 2005). However, this model is very complex and requires extensive training, time, and data to run. Therefore the MDE developed a simplified approach that spatially interpolates the WSM output (Moglen, pers. comm., 2007). This approach is referred to as the "MDE tool" and is described in section 2.3.

In order to ensure compliance with the CWA, the MDE tool must give accurate results. Otherwise, changes made in the watershed could result in increased pollutant loads even though the MDE tool simulated loads did not increase and vice versa. It is even more essential that MDE tool give accurate results because almost the entire state of Maryland is within the Chesapeake Bay watershed. Given that this watershed is so important and is already impaired due to nutrients and sediment it is important to reduce the amounts of those pollutants entering the Bay.

1.4 Goals and Objectives

The goal of this study is to evaluate two water quality models for watersheds in Maryland by comparing the simulated loads to the observed loads calculated using several different rating curves. In order to accomplish this goal the following objectives were developed:

1. Develop rating curves for total nitrogen, total phosphorus, and suspended sediment using the EPA and concentration-derived methods. Use daily discharge data and rating curves to estimate daily observed pollutant loads.
2. Determine if the pollutant loads simulated by the MDE tool are accurate as compared to the EPA and concentration-derived observed loads calculated using the rating curves developed in objective one.
3. Perform a literature review of the various types of water quality models and select one to compare to the MDE tool.
4. Collect data, develop input layers, and run the selected water quality model for a set of USGS gaged watersheds in Maryland.
5. Determine if the pollutant loads simulated by the selected model are accurate as compared to the EPA and concentration-derived loads.
6. Determine if the pollutant loads simulated by the selected model are more accurate than those simulated by the MDE tool.

Chapter 2: Literature Review

2.1 TMDLs

The CWA requires that all water bodies meet specific pollutant criteria that ensure that the water is “fishable and swimmable”. If a water body does not meet these criteria it is considered impaired and section 303(d) of the CWA requires that impaired water bodies be identified and TMDLs developed for each pollutant that is above allowable limits (USEPA, 1999). TMDLs represent the maximum amount of a pollutant that can enter a watershed before the watershed becomes impaired. They are used as a policy tool to maintain water quality standards and can be represented by Equation 2-1 (USEPA, 1999) below:

$$TMDL = LC = \sum WLA + \sum LA + MOS \qquad (2\text{-}1)$$

where LC is the maximum loading a water body can receive without violating water quality standards, WLA is the allocated loading for point sources, LA is the allocated loading for non-point sources and MOS is the margin of safety. TMDLs are typically expressed in terms of mass per time, however, toxicity or other appropriate measures can be used (USEPA, 1999).

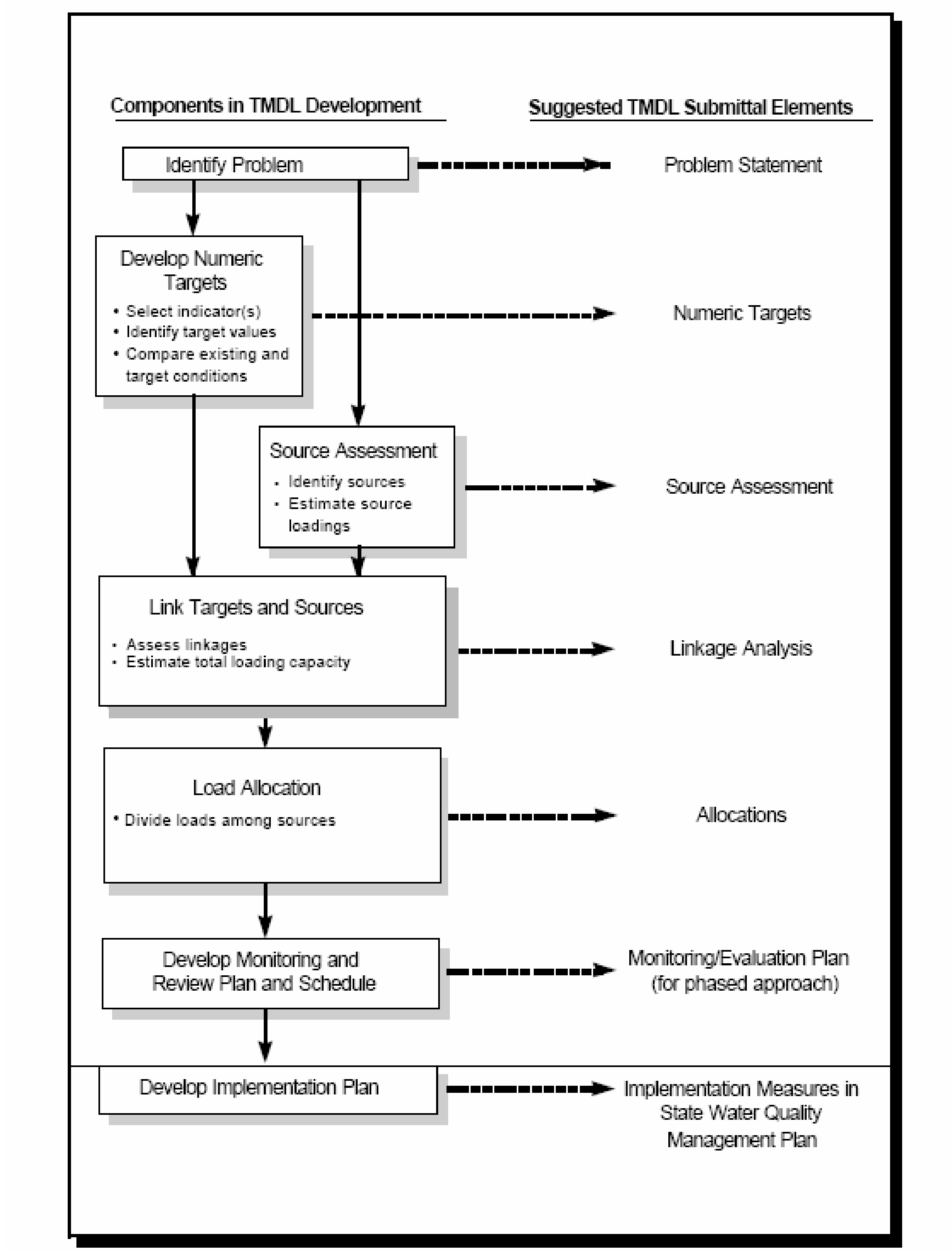

Figure 1: Steps of TMDL Development Process (USEPA, 1999)

The USEPA has identified seven steps for developing TMDLs. Figure 1 shows a flow chart of the steps in the TMDL development process. The first step is problem identification, which consists of identifying the pollutants that are causing the impairment, identifying the scale of the problem and describing any other information that will be helpful in the development of the TMDL. The second step is identifying water quality indicators or targets. This step involves identifying numeric targets to ensure that the water body meets quality standards. If numeric targets are not available, then narrative water quality standards can be interpreted in order to develop measurable or numeric targets. The third step is source assessment and consists of identifying the type, magnitude and location of all the pollutant sources in the watershed. The fourth step is defining a link between water quality targets and pollutant sources. This link establishes the relationship between the pollutant sources in the watershed and the pollutant concentrations in the water body. The USEPA recommends that the link be established using monitoring data; however, it is generally established using water quality modeling. This step also allows the total loading capacity of the watershed to be estimated, where the total loading capacity is defined as the maximum loading a water body can receive without violating water quality standards. The fifth step is allocations and consists of allocating pollutant loads among the sources identified in the third step so that the total pollutant loads do not exceed the total loading capacity of the watershed. If pollutant loadings need to be reduced, the regulatory agency can use NPDES permits to require reductions in point source loadings. Reduction of non-point source loadings is more difficult because there are no permit requirements. However, implementation of BMPs can be encouraged through incentive programs. The margin of safety is usually identified during this step to account for uncertainty about the link between pollutant sources and resulting water quality. The sixth step is monitoring and evaluation. The purpose of this step is to determine whether the water quality standards are being met. The seventh and final step is to develop an implementation plan. This step involves compiling the components of the TMDL that are required by regulation and submitting them for review (USEPA, 1999).

The Maryland Department of the Environment (MDE) is responsible for identifying impaired watersheds in Maryland and developing TMDLs. A list of impaired watersheds in Maryland can be found at the MDE website http://www.mde.state.md.us/Programs/WaterPrograms/TMDL/Sumittals/index.asp. The State of Maryland has developed water quality standards based on designated uses. Each major stream segment in Maryland is assigned a use and each use has an associated numeric and/or narrative quality standard designed to protect that use. Therefore if a stream segment is impaired the water quality target is known because the use is known (MDE, 2007). Table 1 below shows the eight designated uses in Maryland.

Table 1: Designated Uses for Stream Segments in the State of Maryland used as Water Quality Standards (MDE, 2007).

Name	Description
Use I	Water Contact Recreation, and Protection of Nontidal Warmwater Aquatic Life
Use I-P	Water Contact Recreation, Protection of Aquatic Life, and Public Water Supply
Use II	Support of Estuarine and Marine Aquatic Life and Shellfish Harvesting
Use II-P	Support of Estuarine and Marine Aquatic Life and Shellfish Harvesting and Public Water Supply
Use III	Nontidal Cold Water
Use III-P	Nontidal Cold Water and Public Water Supply
Use IV	Recreational Trout Waters
Use IV-P	Recreational Trout Waters and Public Water Supply

The MDE often uses water quality modeling in the development of TMDLs. Modeling can be used in the third step to identify magnitudes of non-point source loads based on land use and in the fourth step to determine the loading capacity of the watershed. It can also be used in the fifth step to determine whether certain BMPs will reduce non-point sources enough to meet water quality standards. There are many water quality models available for use in the TMDL development process; however, the MDE has recently decided to use the Chesapeake Bay Program Model (WSM) for consistency with the Chesapeake Bay Program (CPB) (Moglen, pers. comm., 2007). A description of this model can be found in the following section.

2.2 Chesapeake Bay Program Model

The Chesapeake Bay Watershed Model (WSM) was first developed in 1982. Since then it has been updated and improved several times (Linker et al., 2002). Phase 4.2 has been calibrated and run for 89 watershed model segments. The model segments are made up of 11-digit hydrologic units (HUCs) (CPB, 2000). The WSM output is used as a management tool to see the effects of certain actions on the nutrient and sediment loads that end up in the Chesapeake Bay (Shenk et al., 2001).

The WSM is based on the HSPF modeling environment (Linker et al., 2002). The HSPF model is a continuous, lumped parameter model. It is process-based and is considered to be of high complexity. It requires continuous precipitation and potential evaporation, soil properties, point source loadings, and other inputs. Its outputs include runoff, sediment, nutrients, toxics/pesticides, metals, BOD, and bacteria (Shoemaker et al., 2005); however, only runoff, sediment, and nutrient output are used for the WSM.

An HSPF model can provide accurate estimates for urban, rural and agricultural land uses. It is most accurate on a watershed scale. The model time step is user-defined and can range from one minute to one day (Shoemaker et al., 2005).

The HSPF modeling environment has been extensively reviewed and applied to numerous projects. It has been incorporated into the BASINS (USEPA, 2001a) program and is recommended by the EPA for complex TMDL development. However, the HSPF modeling environment and therefore the WSM model require a lot of training, calibration and time (more than 6 months for entire model application) to run (Shoemaker et al., 2005). For this reason, the MDE has chosen to develop a simplified approach that interpolates the WSM model output. This approach is described in the following section.

2.3 MDE Tool

The MDE recently decided to use the WSM in order to be consistent with the CBP. However, the WSM is very complex and requires extensive amounts of training, time, and data (Shoemaker et al., 2005). Therefore the MDE developed a tool in GISHydro2000 (Moglen, 2006) that would interpolate the WSM output. This tool uses WSM phase 4.2 loading coefficients and 2002 land cover to estimate mean annual nitrogen, phosphorus, and sediment loadings. The tool can be used for quick analyses to determine the effect of changing land use on nutrient and sediment loads (Moglen, pers. comm., 2007).

The advantages of this tool are that it requires little training, time or data to run and it is representative of the WSM model, which has been thoroughly evaluated and is considered to be an accurate model (Shoemaker et al., 2005). The disadvantages of this tool are that it can only show the effect of changing land use and it has not been tested to determine whether the output is accurate.

One of the objectives of this study is to compare the simulated nutrient and sediment loads from the MDE tool with another water quality model. Due to time limitations only one model could be compared to the MDE tool and therefore I conducted a literature review to determine which model to select. The literature review is summarized in the following section.

2.4 Model Selection

Ten water quality modeling systems were considered. Characteristics of those ten models are shown in Table 2. I decided that the selected model should be of medium to low complexity; otherwise, it would not be effective as a planning tool because it would take too much time and training to run. The model should be continuous, as opposed to event-based, because "for NPS modeling, the only feasible option is to incorporate a continuous approach" (Deliman et al., 1999). The model should give reasonable simulations for urban, rural, and agricultural land uses because all three land uses exist in Maryland watersheds. The model should also have a time step that is less than annual because many nutrient loads are seasonal due to agricultural practices, such as manure application. The output from the model should include, at a minimum, simulated runoff (to be used for calibration), nitrogen, phosphorus, and sediment.

The BASINS and SWMM (Rossman, 2005) models were excluded because they are very complex models and require moderate to substantial training to run. They also require three to six months in order to complete the entire application of the model. The AGNPS model was excluded because it is an event-based model and therefore would not be useful for modeling non-point sources from a planning perspective (Shoemaker et al., 2005).

Table 2: Characteristics of Water Quality Models Considered in this Study (Shoemaker et al., 2005)

Name	Full Name	Complexity	Type	Land Use	Hydrology Represented
AGNPS	Agricultural Nonpoint Source Pollution Model	medium	Distributed Event	Rural Agriculture	surface
AnnAGNPS	Annualized Agricultural Nonpoint Source Pollution Model	medium	Distributed Continuous	Rural Agriculture	surface
BASINS	Better Assessment Science Integrating point and Nonpoint Sources	high	lumped event and continuous	Urban Rural Agriculture	surface and groundwater
GLEAMS	Groundwater Loading Effects of Agricultural Management Systems	low	lumped continuous	Rural Agriculture	surface
GWLF	Generalized Watershed Loading Functions	low	semi-distributed continuous	Urban Rural Agriculture	surface and groundwater
PLOAD	Pollutant Loading Application [3]	low [3]	lumped continuous[3]	Urban Rural Agriculture[3]	surface [3]
SPARROW	SPAtially Referenced Regression On Watershed Attributes	medium	distributed[1] continuous[1]	Urban Rural Agriculture	surface
STORM	Storage, Treatment, Overflow, Runoff Model	medium	lumped [2] continuous	urban	surface
SWAT	Soil and Water Assessment Tool	medium	continuous	Urban Rural Agriculture	surface and groundwater
SWMM	Storm Water Management Model	high	lumped [2] event and continuous	Urban	surface and groundwater

[1]Schwarz et al., 2006
[2]Shamsi, 2002
[3]USEPA, 2001

Table 2 (cont): Characteristics of Water Quality Models (Shoemaker et al., 2005)

Name	Land Management	Temporal Scale	Spatial Scale	Input
AGNPS	Field Practices BMP	event	watershed (up to 80mi^2)	weather data, land characteristics, management Information
AnnAGNPS	Field Practices BMP	daily	watershed	weather data, land characteristics, management Information
BASINS	Field Practices BMP	varies depending on model used	watershed	varies depending on model used (see HSPF, SWAT, PLOAD)
GLEAMS	Field Practices BMP	daily	field	weather data, soil properties, agricultural practices
GWLF	Field Practices	daily input monthly output	sub-watershed	weather data, transport parameters, chemical parameters, septic systems, point sources
PLOAD	BMP [3]	annual	watershed [3]	weather data, land use, export coefficients [3]
SPARROW	none	annual	large watersheds	stream reaches, physical coefficients, bilogical and chemical reaction rates, land use, soil
STORM	BMP	hourly	watershed	runoff coefficients, SCS parameters, hourly precipitation
SWAT	Field Practices BMP	daily	watershed	weather data, soil, dem, point sources, crop and management databases, discharge data, watershed quality data
SWMM	BMP	user-defined minutes to hours	sub-watershed	weather data, land use, soil properties, stormwater system characterisitcs

[3]USEPA, 2001

Table 2 (cont): Characteristics of Water Quality Models (Shoemaker et al., 2005)

Name	Output	Credibility
AGNPS	Runoff Nutrients (N, P) Sediment Toxics/Pesticides	Widely used for watershed studies. Has been evaluated by a few studies using measured data. One study found simulated runoff to be systematically underestimated.
AnnAGNPS	Runoff Nutrients (N, P) Sediment Toxics/Pesticides	Studies have found that it predicts monthly nutrient loads with moderate accuracy and sediment loads with high accuracy.
BASINS	Runoff Nutrients (N, P) Sediment Toxics/Pesticides Metals BOD Bacteria	Applied to many TMDLs across the USA. Underlying models SWAT and HSPF have been used extensively in studies. There are many peer-reviewed publications about the model.
GLEAMS	Runoff Nutrients (N, P) Sediment Toxics/Pesticides	Improvement of CREAMS to include groundwater. Many peer reviewed articles on both.
GWLF	Runoff Nutrients (N, P) Sediment	Used extensively in northeast and mid-Atlantic regions for TMDL development. Used to estimate pollutant loads to New York City drinking water supply reservoirs.[4]
PLOAD	Nutrients (N, P) Sediment BOD 3	Incorporated in the BASINS model.
SPARROW	Runoff Nutrients (N, P) Sediment Toxics/Pesticides	Is used in the Chesapeake Bay region.[4]
STORM	Runoff Nutrients (N, P) Sediment Bacteria	Used extensively in the 1970s and 1980s. Used for San Francisco master drainage plan.
SWAT	Runoff Nutrients (N, P) Sediment Toxics/Pesticides Metals	Has been used widely to study TMDLs since the 1990s
SWMM	Runoff Nutrients (N, P) Sediment Toxics/Pesticides Metals BOD Bacteria	Applied extensively to US, Canadian, European, and Australian cities.

[3]USEPA, 2001
[4]AGU, 2001

The AnnAGNPS (Bingner et al., 1998) and GLEAMS (Leonard et al., 1987) models were excluded because they were calibrated for agricultural and rural watersheds and therefore contain assumptions specific to non-urban watersheds and do not perform well on urban watersheds. The STORM (USACE-HEC, 1977) model was excluded because it was calibrated for urban watersheds and performs poorly for rural watersheds (Shoemaker et al., 2005).

The PLOAD and SPARROW models were excluded because they simulate output on an annual basis (Shoemaker et al., 2005). Since nutrient and sediment loads can vary seasonally due to agricultural practices and changes in precipitation, I felt it was important for the selected model to have a time step smaller than an annual time step. In addition, critical conditions for TMDL analysis due to eutrophication typically exist during the growing season (MDE, 2005). If the model gives only annual output, it could not be determined whether target values were exceeded during the growing season.

The only models not excluded were the GWLF (Haith and Shoemaker, 1987) and SWAT (Neitsch et al., 2005) models. The GWLF model was selected for two reasons. First, the SWAT model requires more training and one to three months to run, whereas the GWLF model requires less than one month to run. The second reason is that an improved version of GWLF that uses an ArcView interface has been developed and is used for TMDL development in Pennsylvania and West Virginia (Shoemaker et al., 2005). The improved version is called ArcView Generalized Watershed Loading Function or AVGWLF (Evans et al., 2002).

2.5 AVGWLF

The AVGWLF model is based on the GWLF model, which was developed by Haith and Shoemaker in 1987. The GWLF model was developed to be a compromise between the simplicity of export coefficients and the "complexity of chemical simulation models" (Haith and Shoemaker, 1987). It is a continuous simulation model and is considered to be semi-distributed because, for surface loading, it is distributed in terms of land use, but lumped in terms of other parameters considered by the model. For sub-surface loading the model is considered a lumped parameter model (Evans et al., 2006).

GWLF simulates daily runoff and monthly total sediment and nutrient (total nitrogen and total phosphorus) loads. It calculates runoff using curve numbers (CN) (SCS, 1986) and daily precipitation and temperature data. It calculates erosion using the Universal Soil Loss Equation (USLE) and calculates sediment loads using a sediment delivery ratio, a transport capacity and the calculated erosion from the USLE (Evans et al., 2006). It calculates nutrient loads using an average nutrient concentration based on land use and the simulated runoff. GWLF calculates sediment and nutrient loads on a daily basis, however, it reports them on a monthly basis because simplifications in the modeling result in inaccurate daily values (Shoemaker et al., 2005). The GWLF model also considers point sources, manured areas, and septic system loads (Evans et al., 2006).

The GWLF model has three input files that include information relating to transport, nutrient, and weather data (Evans et al., 2002). Figure 2 shows the information contained within each input file and the sources used to derive that information.

WEATHER.DAT file	Historical weather data from National Weather Service monitoring stations
TRANSPORT.DAT file	
Basin size	GIS/derived from basin boundaries
Land use/cover distribution	GIS/derived from land use/cover map
Curve numbers by source area	GIS/derived from land cover and soil maps
USLE (KLSCP) factors by source area	GIS/derived from soil, DEM, and land cover
ET cover coefficients	GIS/derived from land cover
Erosivity coefficients	GIS/ derived from physiography map
Daylight hrs. by month	Computed automatically for state
Growing season months	Input by user
Initial saturated storage	Default value of 10 cm
Initial unsaturated storage	Default value of 0 cm
Recession coefficient	Default value of 0.1
Seepage coefficient	Default value of 0
Initial snow amount (cm water)	Default value of 0
Sediment delivery ratio	GIS/based on basin size
Soil water (available water capacity)	GIS/derived from soil map
NUTRIENT.DAT file	
Dissolved N in runoff by land cover type	Default values/adjusted using AEU density
Dissolved P in runoff by land cover type	Default values/adjusted using AEU density
N/P concentrations in manure runoff	Default values/adjusted using AEU density
N/P buildup in urban areas	Default values (from GWLF Manual)
N and P point source loads	GIS/derived from NPDES point coverage
Background N/P concentrations in GW	GIS/derived from new background N map
Background P concentrations in soil	GIS/derived from soil P loading map
Background N concentrations in soil	Based on map in GWLF Manual
Months of manure spreading	Input by user
Population on septic systems	GIS/derived from census tract map
Per capita septic system loads (N/P)	Default values (from GWLF Manual)

Figure 2: Information Stored in GWLF Input Files (Evans et al., 2002)

The AVGWLF model is primarily an ArcView GIS interface that derives the inputs to the GWLF model. However, some improvements were made to the GWLF model. The developer of the AVGWLF model modified the water balance to include water withdrawals from surface and groundwater. AVGWLF also includes a streambank erosion routine, in which a lateral erosion rate is calculated based on attributes of the watershed. The total streambank erosion is then calculated by multiplying the lateral erosion rate by the length of the streams in the watershed, the average streambank height, and the average soil bulk density (Evans et al., 2006).

The AVGWLF model derives the needed inputs for GWLF from GIS data files and other user supplied non-spatial information. AVGWLF requires 6 data layers to derive the needed input. There are also 11 optional data layers that can be included to calculate certain parameters. If these layers are not included default values for the parameters will be assigned (Evans et al., 2006). Section 3.3 and Appendix A of this paper provide more detail on the AVGWLF inputs.

Chapter 3: Methodology

3.1 Watershed Selection

For this study watersheds were selected based on the availability of water quality data (specifically related to nitrogen, phosphorus, and sediment concentrations). An effort was made to represent multiple land uses, sizes, and geographic areas.

Since a regression equation will be developed to predict daily observed loads there must be sufficient water quality data to develop that relationship. Equation 3-1, shown below, can be used to solve for the needed sample size of water quality data:

$$n = \left(\frac{t_{\alpha/2}}{H_n} \right)^2 \tag{3-1}$$

where n is the sample size, $t_{\alpha/2}$ is the critical t value with $\nu = n - 2$ degrees of freedom and a confidence level of α which is assumed to be 5%, and H_n is the normalized half-width, defined to be the half width of the confidence interval divided by the variance and is assumed to be 0.5. Both the α and H_n values were chosen based on convention.

Since the t value is dependent on n, Equation 3-1 must be solved iteratively. Starting with an assumed sample size of 10 and solving iteratively yields a required sample size of 13. Equations 3-2 through 3-4 below show the calculations.

$$n = \left(\frac{1.833}{0.5} \right)^2 = 13.4 \qquad \text{use } 14 \tag{3-2}$$

$$n = \left(\frac{1.771}{0.5} \right)^2 = 12.5 \qquad \text{use } 13 \tag{3-3}$$

$$n = \left(\frac{1.782}{0.5} \right)^2 = 12.7 \qquad \text{use } 13 \tag{3-4}$$

The National Water Quality Assessment (NAWQA) Program was implemented in 1991 by the USGS to develop long-term information related to water-quality in surface and ground water. As part of this program, routine samples of nitrogen, phosphorus, and suspended sediment concentrations are taken for particular watersheds. Table 3 shows all the USGS gaged watersheds in Maryland that have more than 13 sample points and the number of sample points they have (USGS, 2007b).

Table 3: Number of Sample Points for Specific Water Quality Parameters for USGS Gaged Watersheds in Maryland. Gages Shown in Italics were excluded, for Reasons Discussed in the Text.

Gage No.	Gage Name	Discharge Instant.	Suspended Sed.	Nitrogen	Phosphorus
1485000	*Willards*	*32*	*18*	*32*	*32*
1485500	Snowhill	87	87	79	79
1493112	Crumpton	113	75	108	109
1493500	Kennedy	80	77	93	80
1578310	*Conowingo*	*172*	*169*	*170*	*171*
1614500	Fairview	743	455	621	612
1639000	Bridgeport	273	33	265	265
1643020	*Fredrick*	*18*	*20*	*24*	*24*

Of the eight watersheds listed in Table 3, three were excluded from the study dataset (excluded watersheds are shown in italics). Gage 01643020 was excluded because it does not have surface water data and therefore daily water quality data could not be calculated. Gage 01578310 was excluded because the drainage area is 27,100 mi^2 and would require massive datasets to analyze. Gage 01485000 was excluded because the GIS-delineated drainage area was 19.8% less than the USGS-reported drainage area.

Figure 3 below shows the location of the five watersheds used in this study. Table 4 shows characteristics of the five watersheds, including drainage area (mi^2), percent urban, percent agriculture, and percent forest.

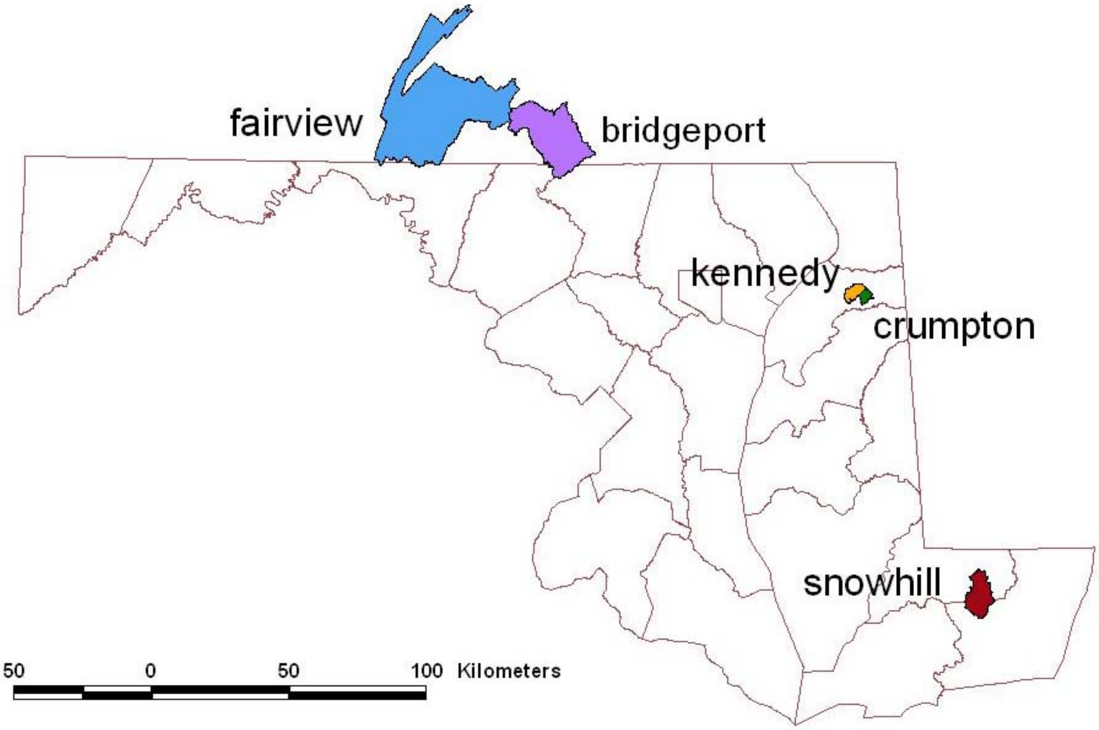

Figure 3: Location of the USGS Gaged Watersheds used in this Study.

Table 4: Watershed Characteristics of USGS Gaged Watersheds in this Study.

Basin	Area (mi^2)	Percent Urban	Percent Agriculture	Percent Forest
Kennedy	11.97	1.704	88.27	9.512
Crumpton	6.503	0.8019	89.86	8.912
Fairview	502.4	4.119	59.95	35.71
Bridgeport	173.2	2.455	77.61	19.73
Snowhill	45.03	3.646	18.53	76.25

3.2 Calculation of Daily Observed Loads

The water quality samples collected by NAWQA are few and sporadic. Due to the lack of collected water quality samples a rating curve was developed for each of the five watersheds so that the daily pollutant loads could be predicted based on the daily discharge. Since the USGS records daily discharge for the watersheds used in this study, daily pollutant loads can be predicted through the use of a rating curve. This allows for a more complete comparison between observed and simulated nutrient loads.

The EPA suggests developing a regression relationship or rating curve between the pollutant load and discharge. This allows the pollutant loads to be estimated "on days when flow is available but water quality data are not" (USEPA, 2005). Since water quality samples are measured in terms of concentration, the pollutant concentration data must be multiplied by the corresponding instantaneous discharge data in order to determine the pollutant load. A relationship is then developed between the pollutant load and discharge data. Daily discharge values for the period of 4/1/1990 to 3/31/2000 were downloaded from the USGS website and used in the pollutant load versus discharge relationship to estimate daily observed loads.

The EPA's method for developing a rating curve is statistically incorrect because discharge is an element of both the x and y-axis quantities. Because load equals the product of discharge and concentration, the rating curve developed is actually representing the relationship between discharge and discharge times a variable (i.e. concentration). This results in goodness-of-fit statistics that indicate that the predictive capabilities of the rating curve are better than they actually are. It can also result in calibrated coefficients that are not accurate representations of the true coefficients (McCuen and Surbeck, 2007).

The statistically correct way to develop the rating curves would be to develop a relationship between pollutant concentration and discharge. The relationship can then be used with average daily downloaded discharge values to determine average daily concentration. The average daily concentration can then be converted to an average daily load by multiplying by the average daily discharge.

The EPA's method of developing a load versus discharge relationship is the generally accepted way of developing a rating curve. Even though it is incorrect, I was interested to see the effect of the rating curve on the estimated pollutant loads. For this reason, the estimated observed pollutant loads in this study were calculated two different ways. A rating curve was developed between pollutant load and discharge. The estimated observed loads as determined from this rating curve will here after be referred to as the EPA observed loads. A second rating curve was developed between pollutant concentration and discharge. The estimated observed loads as determined from this rating curve will here after be referred to as the concentration-derived loads.

Rating curves were developed for each of the five watersheds for total nitrogen, total phosphorus and total sediment because these are the three pollutants simulated by the AVGWLF model and the MDE tool. For simplicity, these three pollutants will generally be referred to this paper without the preceding "total" (i.e. nitrogen, phosphorus and sediment). However, the preceding "total" will be used occasionally for clarity and emphasis.

Please note that NAWQA collects filtered and unfiltered samples of nitrogen and phosphorus. When available the unfiltered sample was used instead of the filtered because it better represents the total nitrogen or phosphorus, which is what the MDE tool and AVGWLF estimate. However, the difference between unfiltered and filtered is small and falls within the error of sample measurements and therefore does not greatly affect the accuracy of the rating curves (Davis, pers. comm., 2006).

3.2.1 EPA Observed Loads

The EPA observed loads were calculated by developing a rating curve between measured pollutant loads and discharge data. Daily discharge values were then used to calculate daily observed loads. A rating curve was developed for each pollutant of interest (nitrogen, phosphorus, and sediment). The process was then repeated for each of the five watersheds. This results in a total of 15 rating curves that were used to estimate EPA observed loads. The EPA rating curves can be seen in Table 5 below. Goodness-of-fit statistics for the EPA and concentration-derived rating curves are presented in Appendix B. The development of the rating curves for the Kennedy watershed will be shown as an example.

Table 5: EPA Rating Curves used to Calculate EPA Observed Loads

	Nitrogen	Phosphorus	Sediment
Kennedy	y = 9.905x - 13.58	y = 0.2351x - 2.405	$y = 886.2x^{0.8755}$
Crumpton	y = 5.171x + 114.9	y = 0.5331x - 7.130	$y = 2969x^{0.8259}$
Fairview	y = 10.23x + 1736	y = 1.065x - 606.3	$y = 500.7x^{0.9939}$
Bridgeport	y = 4.837x + 944.9	y = 0.8594x - 130.6	$y = 72.75x^{1.139}$
Snowhill	y = 3.461x + 15.92	y = 0.1364x - 5.787	$y = 5.303x^{1.389}$

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

3.2.1.1 Nitrogen

The first step in developing the EPA observed nitrogen load is to determine the total nitrogen concentration. The total nitrogen concentration is found by summing the ammonia (NH_3), organic nitrogen (orgN), nitrite (NO_2), and nitrate (NO_3) concentrations in mg N/L. The total nitrogen concentration (mg/L) is then multiplied by the corresponding instantaneous discharge in ft^3/s and a conversion factor of 2.447 to obtain total nitrogen load in kg/day. Table 6 shows the calculations involved in determining the total nitrogen load. The table also shows the parameter short name and code. Please note that the total nitrogen concentration was found by adding parameters 625 and 631. Parameter 608 was used in place of parameter 625, if it was unavailable and parameter 613 was used in place of parameter 631, if it was unavailable.

Table 6: Calculations for Nitrogen Load for the Kennedy Watershed

Parameter Short Name	NH_3+orgN _ wu	Ammonia _ wf	NO_2+NO_3 _ wf	Nitrite_ wf	Total Nitrogen	Discharge _ instant.	Load
Parameter Code	625	608	631	613	-	61	-
Report Units	mg/L as N	mg/L as N	mg/L as N	mg/L as N	mg N/L	ft^3/s	Kg N/day
	0.95	0.304	2.763	0.029	3.713	11.7	106.3
	1.546	0.917	3.368	0.023	4.914	5.1	61.31
	1.252	0.637	3.523	0.032	4.775	5.5	64.25
	0.989	0.291	2.767	0.039	3.756	7.1	65.24
	1.737	0.397	2.198	0.035	3.935	20	192.5
	1.234	0.343	2.294	0.048	3.528	5.8	50.06
	0.903	0.242	2.495	0.054	3.398	5.8	48.22
	1.387	0.395	2.327	0.065	3.714	5.5	49.98
	1.23	0.441	1.987	0.11	3.217	4.7	36.99
	1.457	0.591	2.659	0.138	4.116	5.8	58.41
	0.695	0.138	2.447	0.064	3.142	4	30.75
	1.113	0.226	2.514	0.075	3.627	6.2	55.02
	0.705	0.136	2.629	0.079	3.334	4.3	35.07
	0.759	0.124	2.161	0.092	2.92	3.7	26.43
	0.609	0.08	2.311	0.074	2.92	3.4	24.29
	0.572	0.085	2.115	0.033	2.687	2.6	17.09
	0.509	0.042	2.181	0.026	2.69	2	13.16
	0.603	0.053	2.14	0.026	2.743	4.7	31.54
	0.545	0.074	2.283	0.019	2.828	1.4	9.686

Note: wu stands for unfiltered sample and wf stands for filtered sample

Most samples had all four types of nitrogen concentration data (i.e. parameters 625 and 631). There were a couple of instances when nitrate was not available. In such cases parameter 613 was used instead of parameter 631. The Kennedy watershed, however, had 22 samples that did not include organic nitrogen so parameter 608 was used instead of parameter 625. The organic nitrogen for the other 57 samples contributed approximately 28 percent to the total nitrogen. Since organic nitrogen was such a large part of the total nitrogen in this watershed, the sample points that did not include organic nitrogen were excluded.

A linear rating curve is then developed between the total nitrogen load in kg/day and the corresponding instantaneous discharge in ft^3/s using least-squares regression. This rating curve is shown in Figure 4. The rating curve is then used to estimate daily observed loads in kg/day. For example, the mean daily discharge for 4/1/1990 for the Kennedy watershed is 13 ft^3/s. Using the rating curve and the mean daily gives a daily load of 9.905*13 - 13.58 or 115.2 kg/day. These daily loads can be summed to produce monthly loads and yearly loads for comparison to the AVGWLF and the MDE tool simulated loads, respectively.

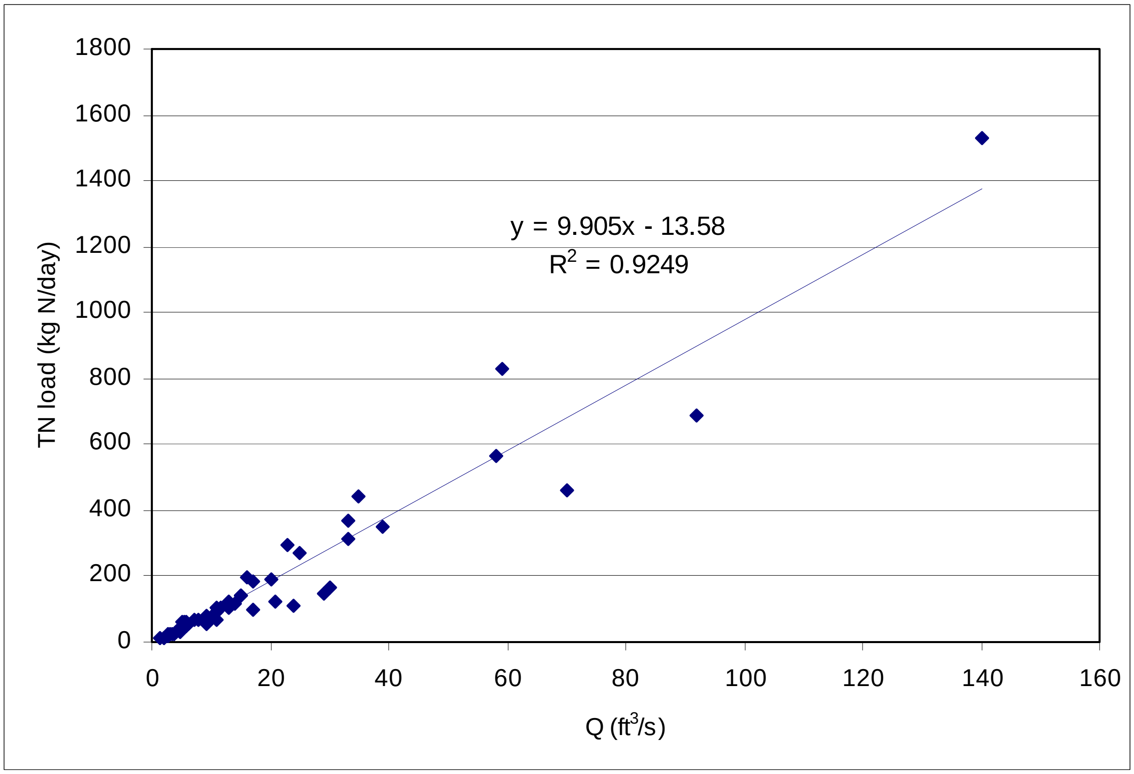

Figure 4: EPA Nitrogen Rating Curve for the Kennedy Watershed

3.2.1.2 Phosphorus

The EPA observed total phosphorus load is developed the same way as the total nitrogen load. The unfiltered phosphorus (parameter 665) samples are assumed to equal total phosphorus. The total phosphorus in mg/L is multiplied by the corresponding instantaneous discharge in ft^3/s and a conversion factor of 2.447 to get total phosphorus load in kg/day. A least-squares regression is then used to develop a linear rating curve. Figure 5 shows the rating curve for total phosphorus for the Kennedy watershed. The rating curve and daily discharge data can then be used to calculate daily loads, which can then be summed to get monthly and yearly loads.

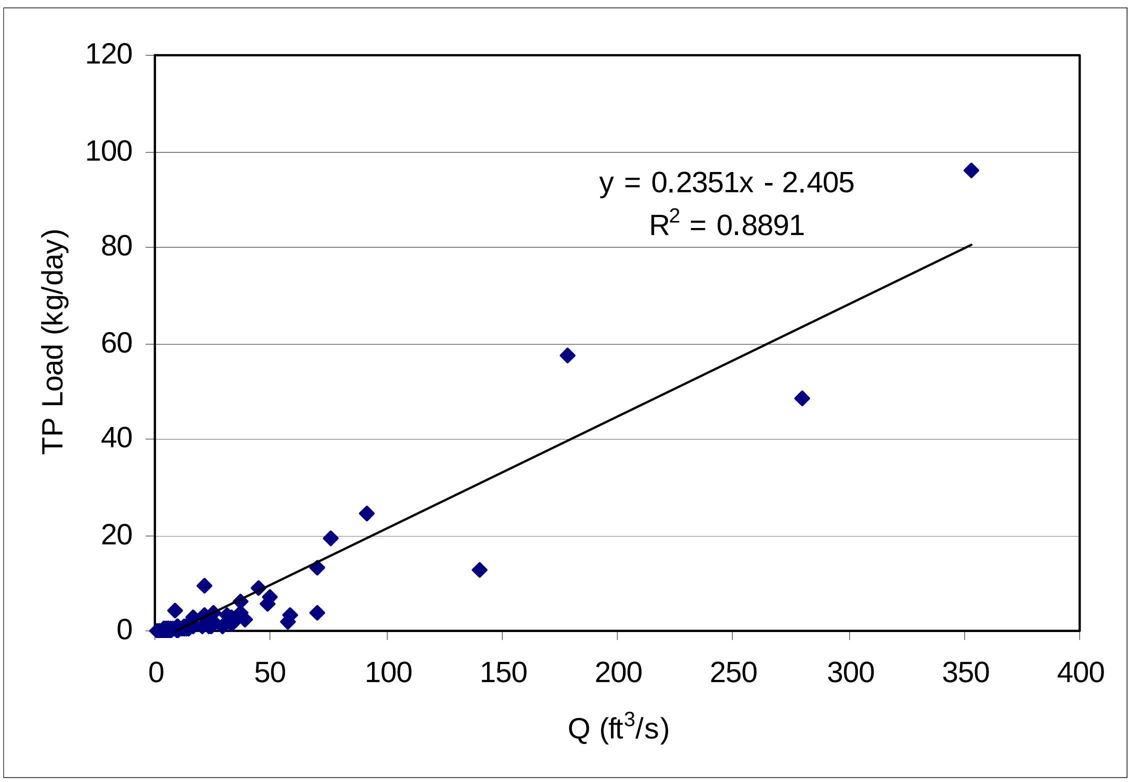

Figure 5: EPA Phosphorus Rating Curve for Kennedy Watershed

3.2.1.3 Phosphorus Estimation using a Zero Intercept Rating Curve

The EPA rating curves shown in Table 5 for phosphorus all have negative intercepts. This causes a problem when phosphorus daily loads are calculated because some of the discharges are not high enough to counteract the negative intercept and therefore many calculated daily loads are negative. Since a negative load is impossible the load is assumed to be zero. However, since so many loads were zero, the EPA phosphorus rating curves were calculated again; this time the intercept was fixed at zero so that negative loads would not occur. Figure 6 shows the rating curve for phosphorus when the intercept is set to zero. Daily discharge values and the rating curves shown in Table 7 can be used to calculate daily phosphorus loads and then summed to obtain monthly and annual loads.

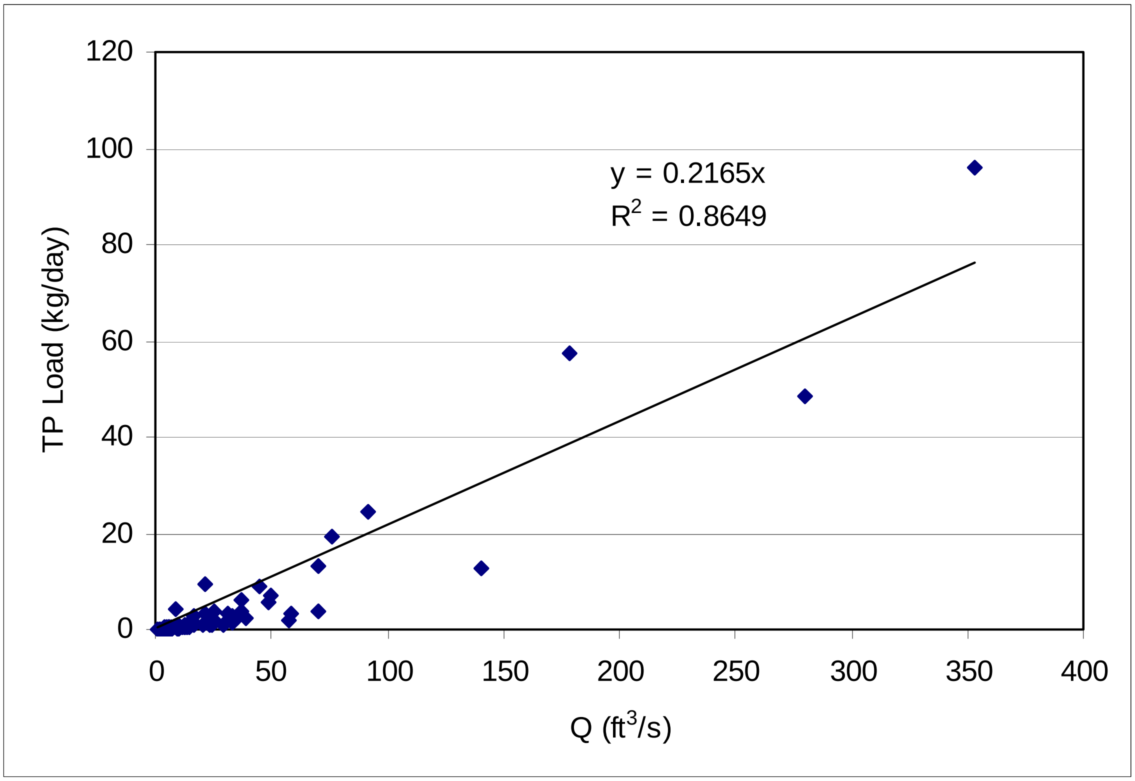

Figure 6: EPA Phosphorus Calculated using a Zero Intercept Rating Curve for the Kennedy Watershed

Table 7: EPA Phosphorus Rating Curves with a Zero Intercept

Watershed	Phosphorus (with 0 intercept)
Kennedy	y = 0.2165x
Crumpton	y = 0.5289x
Fairview	y = 0.9622x
Bridgeport	y = 0.8430x
Snowhill	y = 0.1248x

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

3.2.1.4 Sediment

Sediment is storm-driven, meaning that the sediment load increases exponentially with increasing discharge. For this reason, sediment tends to be linear when plotted on a log scale. Using a least-squares regression to develop a linear relationship in log-log space leads to error and large biases (USEPA, 2005). In order to reduce this error and unbias the rating curve the numerical optimization program NUMOPT (McCuen, 1993) was used.

First the suspended sediment concentration (parameter 80154) is converted to a load by multiplying by the corresponding instantaneous discharge and a conversion factor. Second, the NUMOPT program is used to develop an unbiased power model between suspended sediment load and discharge. The power model, shown in Figure 7, can then be used to calculate daily suspended sediment loads using daily discharge values. The daily loads can then be summed to get monthly and yearly loads.

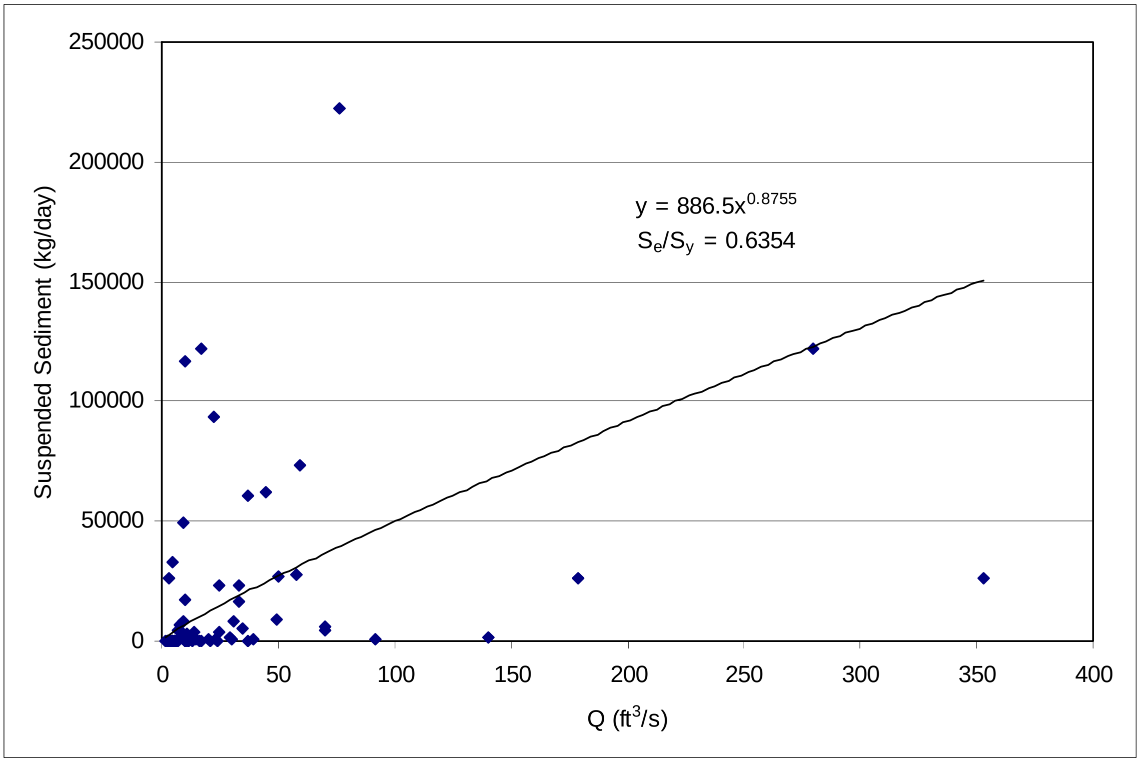

Figure 7: EPA Suspended Sediment Rating Curve for Kennedy Watershed

3.2.2 Concentration-derived Loads

The concentration-derived loads were calculated by developing a rating curve between the measured pollutant instantaneous concentration in mg/L and the corresponding instantaneous discharge in ft^3/s. Similarly to the EPA rating curves, least-squares regression was used to develop linear rating curves between nitrogen concentration and discharge and phosphorus concentration and discharge. The NUMOPT program was used to develop an unbiased power model between suspended sediment concentration and discharge. The concentration-derived rating curves are shown in Table 8 below.

The correlation coefficients (*R*) for the concentration-derived rating curves for nitrogen and phosphorus were much lower than the correlation coefficients for the EPA rating curves. Therefore, a t-test was used to determine whether the concentration-derived *R* values are statistically different than 0. If the t-test showed statistical significance at the 5 percent level ($t > t_{crit,\,5\%}$) then the concentration-derived rating curve was used to calculate loads, otherwise the average concentration, shown in Table 9, was used. Table 10 below shows the calculations for the t-test. The *t* value was calculated using equation 3-5 below:

$$t = \frac{R}{\left(\frac{1-R^2}{n-2}\right)^{0.5}} \qquad (3\text{-}5)$$

where *R* is the correlation coefficient and *n* is the sample size. Table 10 also shows the decision made and the rejection probability. If the decision is to reject, then the rating curve was used to estimate daily concentrations. If the decision was to accept, then the average concentration was used.

Table 8: Rating Curves used for Calculation of Concentration-derived Observed Loads

	Nitrogen	Phosphorus	Sediment
Kennedy	y = 0.006782x + 3.412	y = 0.0002969x + 0.02495	$y = 40.53x^{0.3796}$
Crumpton	y = -0.001831x + 5.959	y = 0.00007613x + 0.04347	$y = 17.43x^{0.3912}$
Fairview	y = -0.00006098x + 4.795	y = 0.00002060x + 0.2242	$y = 7.194x^{0.3917}$
Bridgeport	y = 0.00001776x + 1.961	y = 0.00002015x + 0.1743	$y = 85.43x^{-0.0031}$
Snowhill	y = 0.0005337x + 1.197	y = 0.00002298x + 0.02638	$y = 4.492x^{0.2654}$

Note: y is the pollutant concentration in mg/L and x is the discharge in ft^3/s

Table 9: Average Observed Pollutant Concentration (mg/L)

	Nitrogen	Phosphorus	Sediment
Kennedy	3.538	0.03072	123.3
Crumpton	5.846	0.05037	46.81
Fairview	4.627	0.2805	138.2
Bridgeport	2.016	0.2368	84.15
Snowhill	1.267	0.02942	13.89

Table 10: Calculations for t-Test on the Significance of the Concentration-derived Rating Curves

Watershed	Nutrient	*R*	*n*	*t*	$t_{crit,\,5\%}$	Decision	Rej. Prob.
Kennedy*	Nitrogen	0.2015	57	1.526	2.004	accept	13.80%
	Phosphorus	0.4654	81	4.674	1.991	reject	<0.01%
Crumpton	Nitrogen	0.3226	105	3.460	1.983	reject	0.08%
	Phosphorus	0.3302	106	3.567	1.983	reject	0.05%
Fairview	Nitrogen	0.1584	618	3.982	1.964	reject	<0.01%
	Phosphorus	0.2298	610	5.822	1.964	reject	<0.01%
Bridgeport	Nitrogen	0.0837	265	1.362	1.969	accept	17.50%
	Phosphorus	0.4080	265	7.248	1.969	reject	<0.01%
Snowhill	Nitrogen	0.2112	79	1.896	1.991	accept	6.17%
	Phosphorus	0.2100	79	1.885	1.991	accept	6.32%

* does not include water quality measurements with no organic nitrogen

Since the rating curve developed for sediment concentration is not linear, an *R* value cannot be determined and a t-test cannot be performed to see if the rating curve is statistically better than the average for prediction. Table 11 below shows the relative standard error (S_e/S_y) and relative bias (e/y) for the sediment rating curves in Table 8. By definition, if $S_e/S_y > 1$, then the rating curve does not explain any of the variation in the data. Therefore, if the $S_e/S_y > 1$, the average sediment concentration shown in Table 9 is used as the daily concentration. If $S_e/S_y < 1$, the rating curve shown in Table 8 is used with daily discharges to calculate daily concentrations. The daily concentration can then be converted to daily loads by multiplying by the corresponding discharge and a conversion factor. The daily loads can then be summed to get monthly and yearly loads for comparison to the AVGWLF and the MDE tool output.

Table 11: Goodness-of-fit Statistics for Concentration-derived Sediment Rating Curves

Watershed	b_0	b_1	S_e/S_y	e/y
Kennedy	40.53	0.3796	0.9288	0.0000
Crumpton	17.43	0.3912	0.6865	0.0000
Fairview	7.195	0.3917	0.9211	0.0000
Bridgeport	85.43	-0.0031	1.016	0.0000
Snowhill	4.492	0.2654	0.9158	0.0000

3.3 AVGWLF Inputs

AVGWLF is an interface that allows GIS to calculate the input parameters for the GWLF model. AVGWLF uses up to 17 GIS data layers to derive the needed inputs for GWLF. Table 12 below shows the AVGWLF input data layers, a brief description of the layers, and whether or not they are required.

Table 12: AVGWLF Required and Optional Data Layers (Evans and Corradini, 2006)

File Names	Short Description	Required
Shape Files		
Weather stations	Weather station locations (points)	Y
Point Sources	Point source discharge locations (points)	N
Water Extraction	Water withdrawal locations (points)	N
Tile Drain	Locations of tile-drained areas (polygons)	N
Basins	Basin boundary used for modeling (polygons)	Y
Streams	Map of stream network (lines)	Y
Unpaved Roads	Map of unpaved roads (lines)	N
Roads	Road map (lines)	N
Counties	County boundaries - for USLE data (polygons)	N
Septic Systems	Septic system numbers and types (polygons)	N
Animal Density	Animal density (in AEUs per acre) (polygons)	N
Soils	Contains various soil-related data (polygons)	Y
Physiographic Provinces	Contains hydrologic parameter data (polygons)	N
Grid Files		
Land Use/Cover	Map of land use/cover (16 classes)	Y
Elevation	Elevation grid	Y
Groundwater-N	Background estimate of N in mg/l	N
Soil-P	Estimate of soil P in mg/kg (total or soil test P)	N

In this study all of the data layers were developed for the State of Maryland, except for the Tile Drain, Unpaved Roads, Roads, Groundwater-N, and Soil-P layers. The Tile Drain layer was not developed because I was unable to find comprehensive maps of the location of Tile Drains in the state of Maryland. The percent of land by area that is occupied by Tile Drains is likely small and therefore the amount of error introduced by not including this layer is small. The Unpaved Road layer was not included because the amount of unpaved roads in the five study watersheds is small. The

Roads layer was not included because it serves only as a “background layer” and does not affect the output from GWLF. The Groundwater-N and Soil-P grids were not included because they are data intensive and I felt that the improvements in accuracy were not worth the time needed to create these layers. Appendix A provides more information on the development of AVGWLF data layers.

AVGWLF also prompts the user to give other, non-spatial information, which will be used to derive GWLF input parameters. Table 13 shows the information used in this study for the non-spatial information. All of the non-spatial information is constant for the five watersheds, except for the months of manure application. Table 14 shows the months of manure application by watershed.

Table 13: Non-spatial Data Inputs to AVGWLF

First year of weather data	1990
Last year of weather data	2000
ET Calculation Method	Hammon Method
First month of growing season	April
Last month of growing season	November
Fraction of irrigation water to return to surface/subsurface flow	0.4
Nutrient retention by lakes and wetlands	Not considered

Table 14: Months of Manure Application

Watershed	Months
Kennedy	March, April, May, June
Crumpton	March, April, May, June
Fairview	April, May, September, October
Brideport	April, May, October, November
Snowhill	February, March, April, October

3.4 Calibration

In water quality modeling it is important to estimate water quantity with accuracy. If the water quantity is over or under predicted then confidence in the water quality results is low because they are directly dependent on discharge.

The most important factors that affect streamflow in the GWLF model are curve number, recession and seepage coefficients, and *ET* cover coefficients (K_{et}) (Lee et al., 2000; Schneidernam et al., 1998). According to the developer of AVGWLF, it would be difficult to improve upon the literature values of curve numbers and recession and seepage coefficients (Evans, 2002). For this reason, I decided to fine-tune the simulated discharge by adjusting the K_{et} coefficients.

The K_{et} coefficients control the amount of precipitation that is evaporated or transpired. From continuity, it is known that water coming into a watershed must equal the water leaving the watershed. Using this concept, we arrive at equation 3-6 shown below:

$$P = ET + Q + \Delta S \tag{3-6}$$

where P is the precipitation, ET is the evapotranspiration, Q is the discharge and ΔS is the change in storage.

The change in storage is negligible over a long period of time. Since the simulation length is 10 years, it can be assumed that the change in storage equals 0. Therefore equation 3-6 can be rewritten as shown in equation 3-7.

$$P = ET + Q \tag{3-7}$$

The K_{et} coefficients control the amount of evapotranspiration. Increasing the K_{et} coefficients increases the evapotranspiration and therefore decreases the discharge.

Decreasing the K_{et} coefficients decreases the evapotranspiration and therefore increases the discharge. There are 12 K_{et} coefficients, one for each month. In order to make the calibration simple, they were all multiplied by the same adjustment factor in order to reduce or increase the evapotranspiration. Table 15 below shows the observed mean annual discharge, the initially simulated mean annual discharge, the K_{et} adjustment factors and the resulting simulated mean annual discharge after K_{et} adjustment. Please note that the mean annual discharges were calculated using discharge data from the entire 10 year simulation period, with the exception of the Crumpton watershed. Figure 8 shows the hydrograph for the Snowhill watershed before the K_{et} adjustment and Figure 9 shows the hydrograph after K_{et} adjustment.

Table 15: K_{et} Adjustment Factors for Water Quantity Calibration

Basin	Obs. Flow (cm)	Sim. Flow (cm)	K_{et} Adjustment Factor	Sim. Flow after K_{et} Adjustment (cm)
Kennedy	36.93	12.56	0.48	37.24
Crumpton	156.4*	173.5*	0.79	156.9*
Fairview	48.84	47.33	0.96	48.74
Bridgeport	49.67	54.64	1.17	49.55
Snowhill	41.03	56.65	1.47	41.07

* Sum of flow from 7/1/1996 to 3/31/2000 (instead of from 4/1/1990 to 3/31/2000)

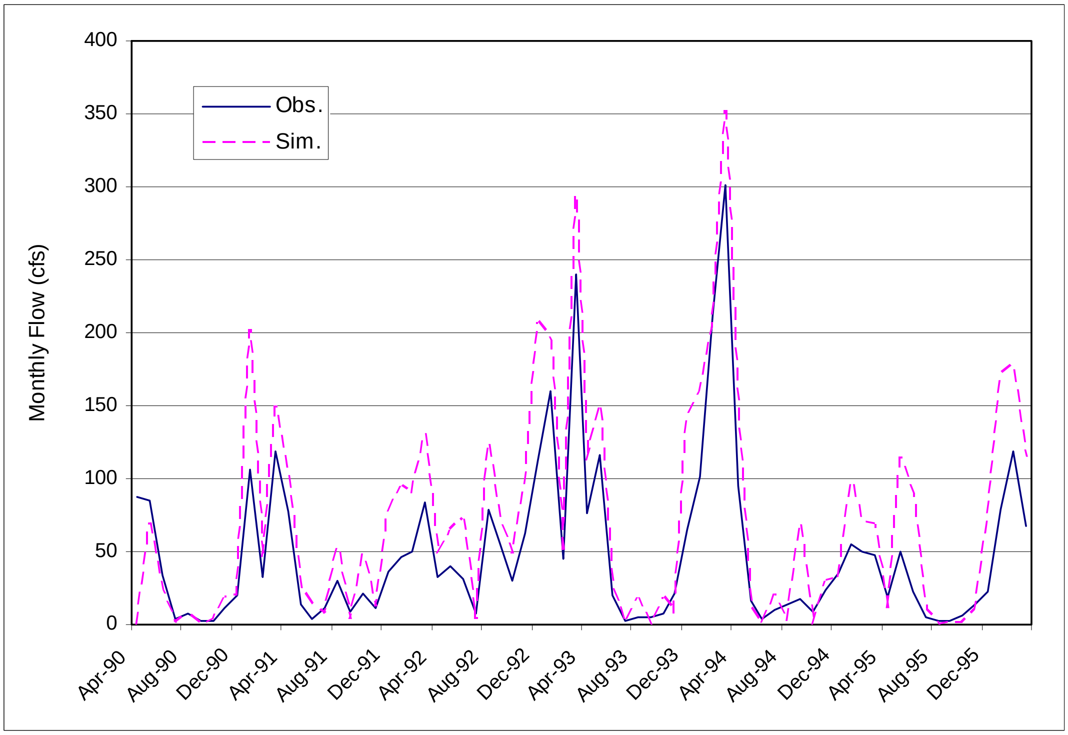

Figure 8: Hydrograph for Snowhill before K_{et} Adjustment

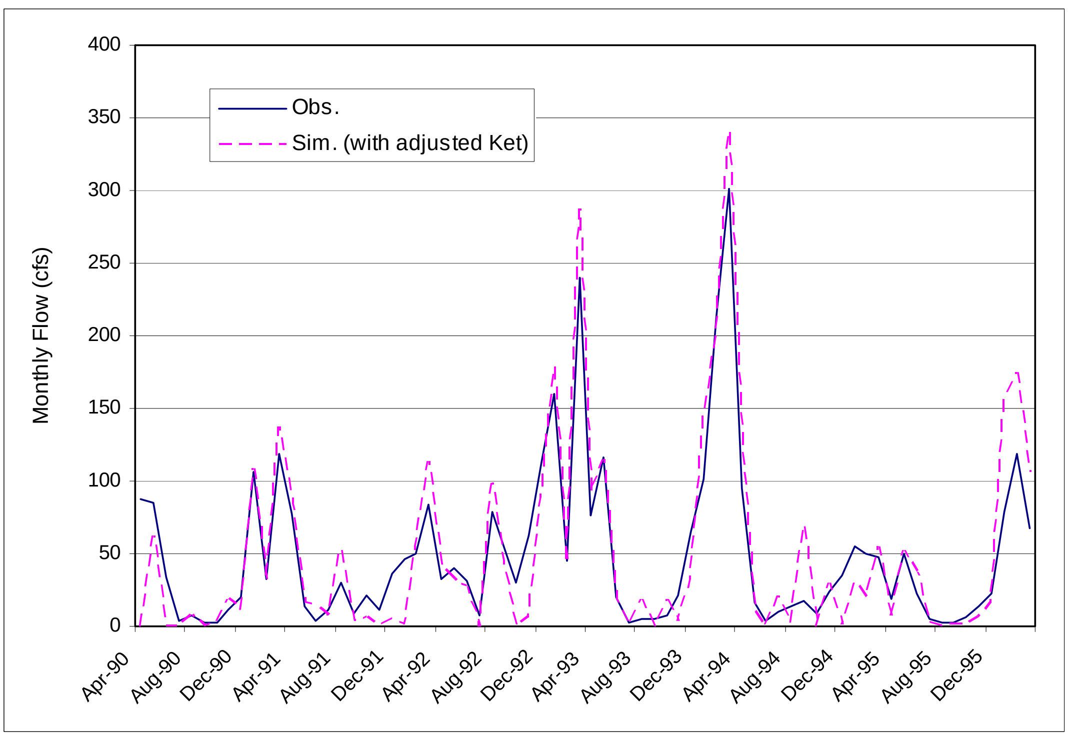

Figure 9: Hydrograph for Snowhill after K_{et} Adjustment

The purpose of the K_{et} adjustment is to ensure that over and under estimates of the streamflow balance so that AVGWLF is not systematically over or under predicting. It is evident from Table 15, Figure 8 and Figure 9 that AVGWLF correctly predicts water quantity after K_{et} adjustment and while it does over predict sometimes and under predict other times, the overall simulated volume and timing of streamflow is accurate.

Other steps can be taken to calibrate AVGWLF's nitrogen, phosphorus, and sediment load models. In this study these calibrations were not done because it is unlikely that someone using AVGWLF in the State of Maryland would have enough water-quality data to calibrate the pollutant load models. In order to get a realistic idea of the predictive capabilities of AVGWLF the pollutant load models were not calibrated (i.e. model default values were used).

Chapter 4: Results and Discussion

4.1 Comparison of Concentration-derived and EPA Observed Loads

The water quality data available for watersheds in Maryland are limited. Therefore a rating curve was developed between the available water quality data and corresponding discharge so that daily discharges could be used to predict water quality on days when water quality data are not available. The USEPA suggests that a rating curve be developed between pollutant load and discharge. However, as discussed in Chapter 3, since load is the product of concentration and discharge, the rating curve developed actually represents the relationship between discharge and discharge times a variable (i.e. concentration). This results in inflated goodness-of-fit statistics and calibrated coefficients that do not accurately represent the true coefficients (McCuen and Surbeck, 2007).

In order to correct this issue, a rating curve was developed between pollutant concentration and discharge. Daily discharge values were used to calculate daily pollutant concentrations, which were then converted to pollutant loads by multiplying by the discharge and a conversion factor. The pollutant loads calculated using this method are referred to as the concentration-derived observed loads and the rating curves used to derived them are shown in Table 16. The pollutant loads calculated using the method described by the USEPA are referred to as the EPA observed loads and the rating curves used to derive them are shown in Table 17. A detailed comparison of the results of the two different methods follows.

Table 16: Concentration-derived Rating Curves.

	Nitrogen	Phosphorus	Sediment
Kennedy	$y = 8.656x$	$y = 0.0007340x^2 + 0.06116x$	$y = 99.17x^{1.380}$
Crumpton	$y = -0.004404x^2 + 14.58x$	$y = 0.0001957x^2 + 0.1064x$	$y = 42.64x^{1.391}$
Fairview	$y = -0.0001468x^2 + 11.74x$	$y = 0.00004893x^2 + 0.5485x$	$y = 17.60x^{1.392}$
Bridgeport	$y = 4.932x$	$y = 0.00004893x^2 + 0.4264x$	$y = 206.0x$
Snowhill	$y = 3.100x$	$y = 0.07193x$	$y = 10.99x^{1.265}$

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 17: EPA Rating Curves.

Watershed	Nitrogen	Phosphorus	Sediment
Kennedy	$y = 9.905x - 13.58$	$y = 0.2351x - 2.405$	$y = 886.2x^{0.8755}$
Crumpton	$y = 5.171x + 114.9$	$y = 0.5331x - 7.130$	$y = 2969x^{0.8259}$
Fairview	$y = 10.23x + 1736$	$y = 1.065x - 606.3$	$y = 500.7x^{0.9939}$
Bridgeport	$y = 4.837x + 944.9$	$y = 0.8594x - 130.6$	$y = 72.75x^{1.139}$
Snowhill	$y = 3.461x + 15.92$	$y = 0.1364x - 5.787$	$y = 5.303x^{1.389}$

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

4.1.1 Nitrogen

The two methods described above were used to develop rating curves for the pollutant total nitrogen. Table 16 shows the concentration-derived rating curves and Table 17 shows the EPA rating curves for nitrogen. In general, the nitrogen loads determined using the EPA rating curves are higher. This can be seen in Figure 10, which shows a box and whisker plot for annual total nitrogen loads in kg/mi^2 of watershed area for the concentration-derived and EPA methods. Please note that the box and whisker plots represent 10 annual pollutant loads for the 10 year simulation period, with the exception of the Crumpton watershed. The box and whisker plot for the Crumpton watershed represents 4 annual loads because daily discharge data for the Crumpton watershed was only available from 7/1/1996 to 3/31/2000.

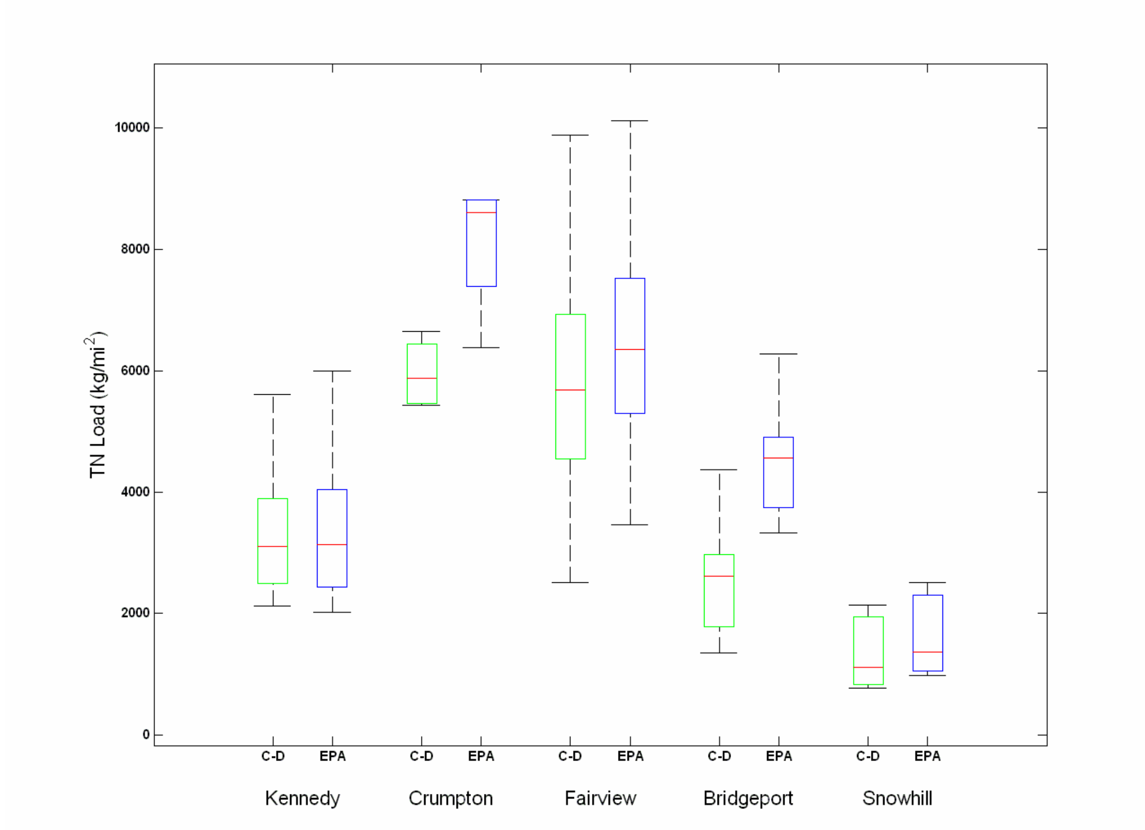

Figure 10: Box and Whisker Plot of Concentration-derived (C-D) and EPA Annual Observed Nitrogen

The EPA nitrogen annual loads are clearly higher for the Crumpton and Bridgeport watersheds because the 1st and 3rd quartiles do not overlap. For the Fairview and Snowhill watersheds the EPA annual loads are closer to the concentration-derived annual loads, but are still higher. The exception is the Kennedy watershed. The EPA median nitrogen load is still higher than the concentration-derived nitrogen load, but only by 1 percent. A possible reason for this exception is that sample measurements that were missing organic nitrogen were excluded from the dataset used to develop the total nitrogen rating curves for the Kennedy watershed.

The EPA method is not always higher as indicated by Figure 10. The EPA method tends to be higher during low flow conditions and lower during peaks, as compared to the concentration-derived loads. Figure 11 below shows the EPA and concentration-derived observed daily loads versus discharge for the Crumpton watershed. As can be seen by the figure, the EPA observed loads are higher for discharges below approximately 13 ft^3/s and lower for discharges above approximately 13 ft^3/s. Since the average daily discharge between 4-01-1990 and 3-31-2000 is 8.04 ft^3/s, the majority of days have a higher load using the EPA rating curve, as compared to the concentration-derived rating curve. Figure 12 below shows the EPA and concentration-derived daily loads by date for the Crumpton watershed. The EPA method gives higher baseflows and lower peaks. The EPA method also has less variation than the concentration-derived method.

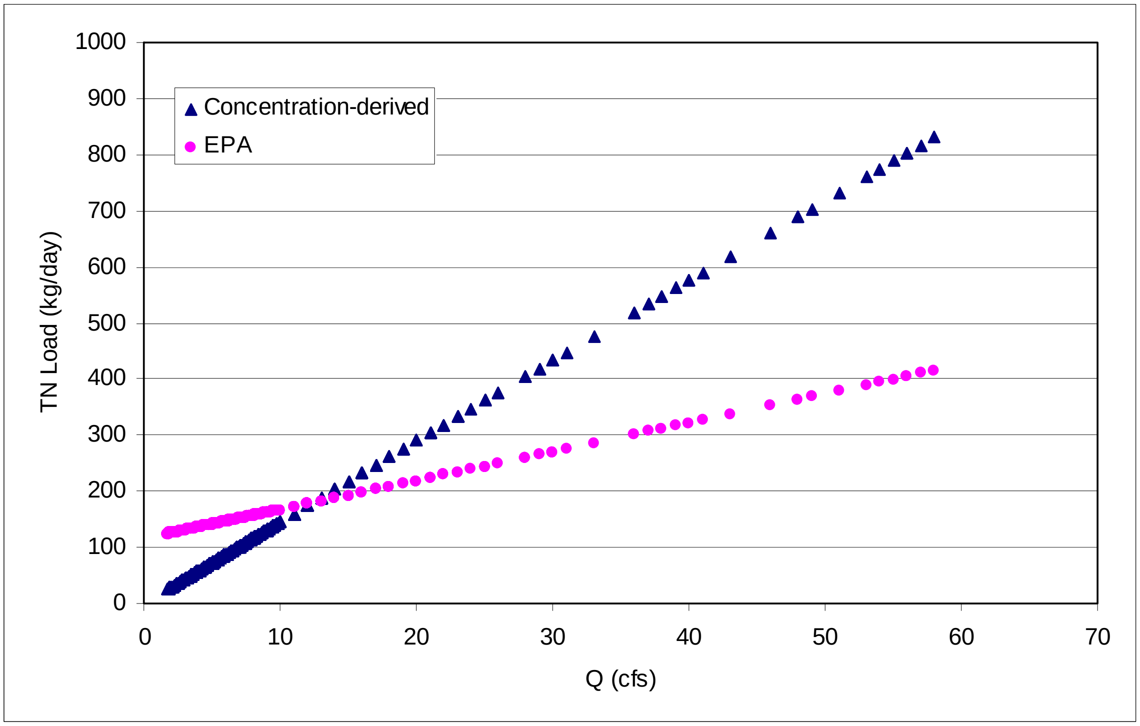

Figure 11: Concentration-derived and EPA Daily Observed Total Nitrogen (TN) Loads versus Discharge for the Crumpton Watershed.

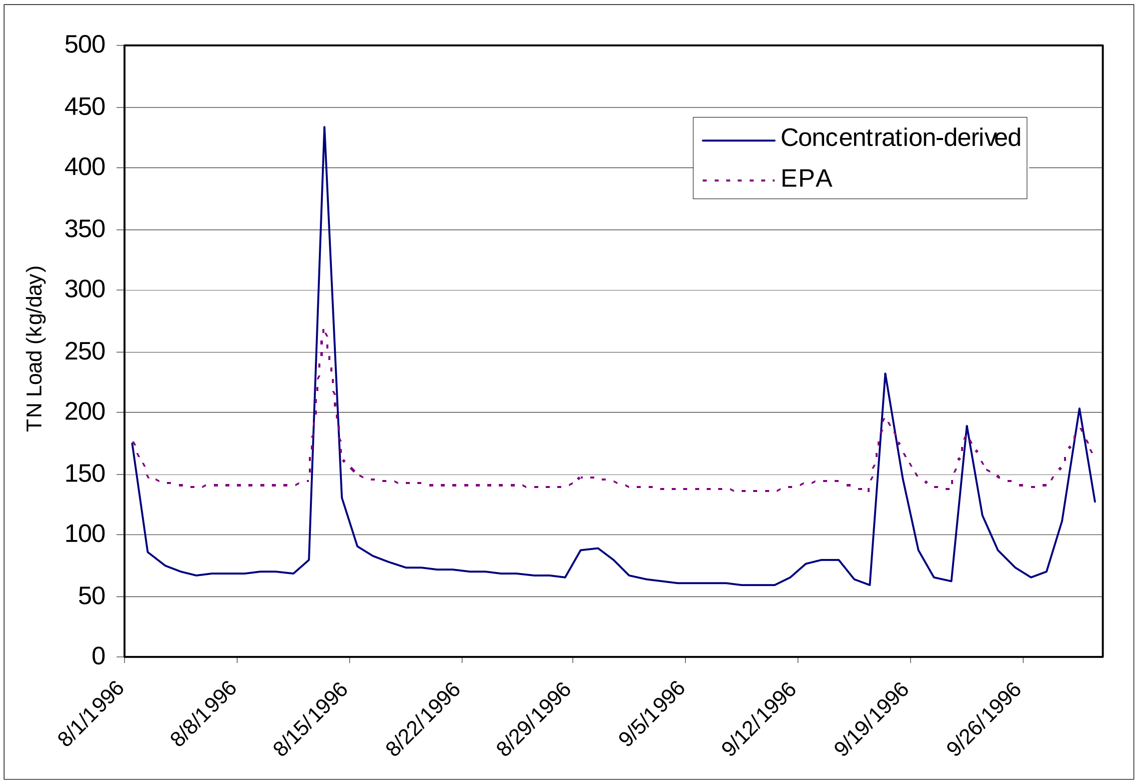

Figure 12: Concentration-derived and EPA Total Nitrogen (TN) Loads by Date for the Crumpton Watershed.

The relationship between the two rating curves and discharge for the Fairview and Bridgeport watersheds is similar to that of the Crumpton watershed. The EPA method predicts higher baseflows and lower peaks. In each case, the average flow is significantly lower than the discharge for which the concentration-derived method produces a higher load and therefore the EPA method generally results in a larger load. For the Snowhill watershed the EPA method predicts higher values for both low and high flows as can be seen in Figure 13.

Figure 14 shows the nitrogen loads predicted by the two different methods versus discharge for the Kennedy watershed. The relationship between the EPA and concentration-derived loads for the Kennedy watershed is the opposite of the relationship for the Crumpton, Fairview, and Bridegport watersheds. The concentration-derived loads are higher for baseflows and the EPA loads are higher for the peaks. The reason that the EPA method still produces nitrogen loads that are higher on an annual basis as seen in Figure 10 is that the average daily discharge between 4-01-90 and 3-31-00 is 12.8 ft^3/s. Since the EPA method produces higher loads for discharges above approximately 11 ft^3/s, the majority of days have a higher load using the EPA rating curve, as compared to the concentration-derived rating curve. However, the two methods predict loads that are very close in value for low discharges (below approximately 20 ft^3/s) and therefore the annual values predicted by the two methods are very similar for the Kennedy watershed.

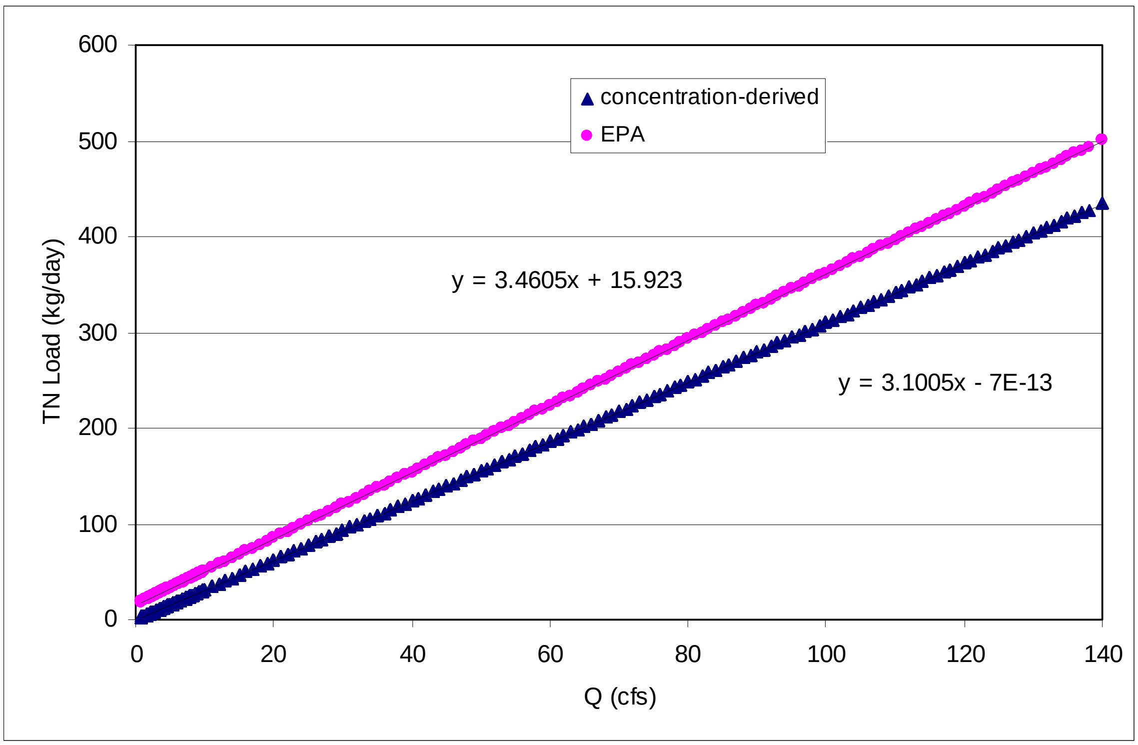

Figure 13: Concentration-derived and EPA Daily Observed Total Nitrogen (TN) Loads versus Discharge for the Snowhill Watershed.

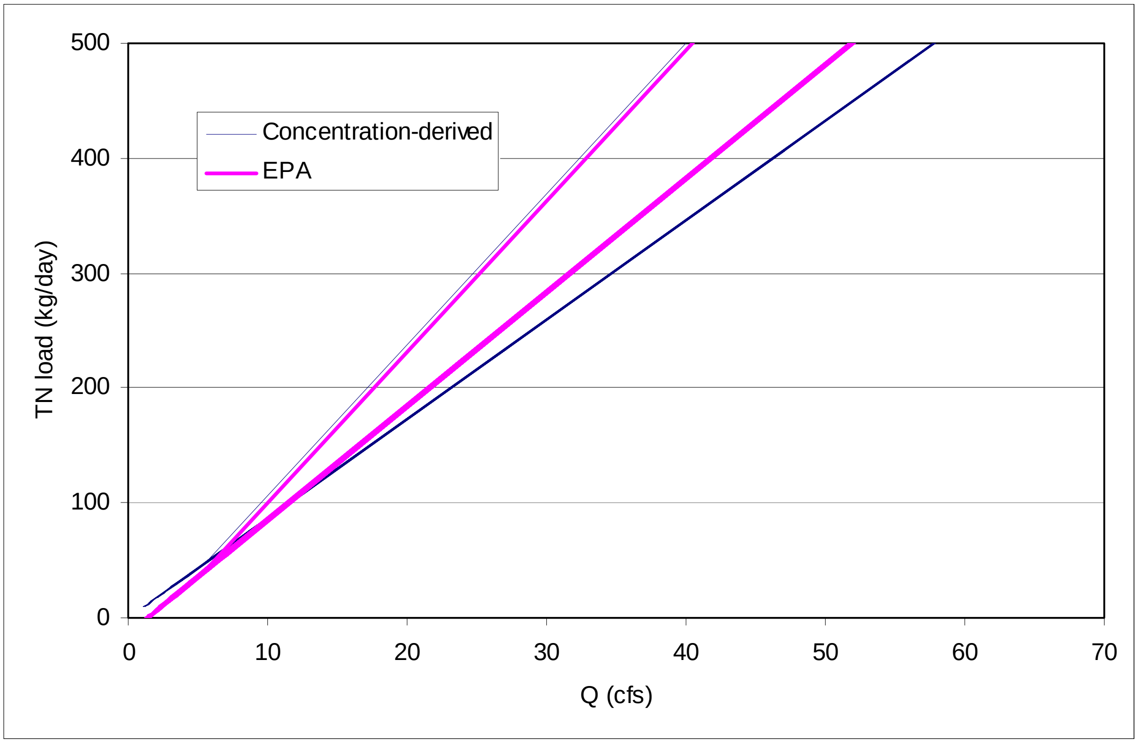

Figure 14: Concentration-derived and EPA Daily Observed Total Nitrogen (TN) Loads versus Discharge for the Kennedy Watershed.

4.1.2 Phosphorus

Figure 15 below shows a box and whisker plot comparing the concentration-derived and EPA total phosphorus annual loads. According to the figure the concentration-derived observed loads are higher for every watershed except Bridgeport. However, even though the concentration-derived loads are higher on average, the two methods produce total phosphorus loads that are very similar.

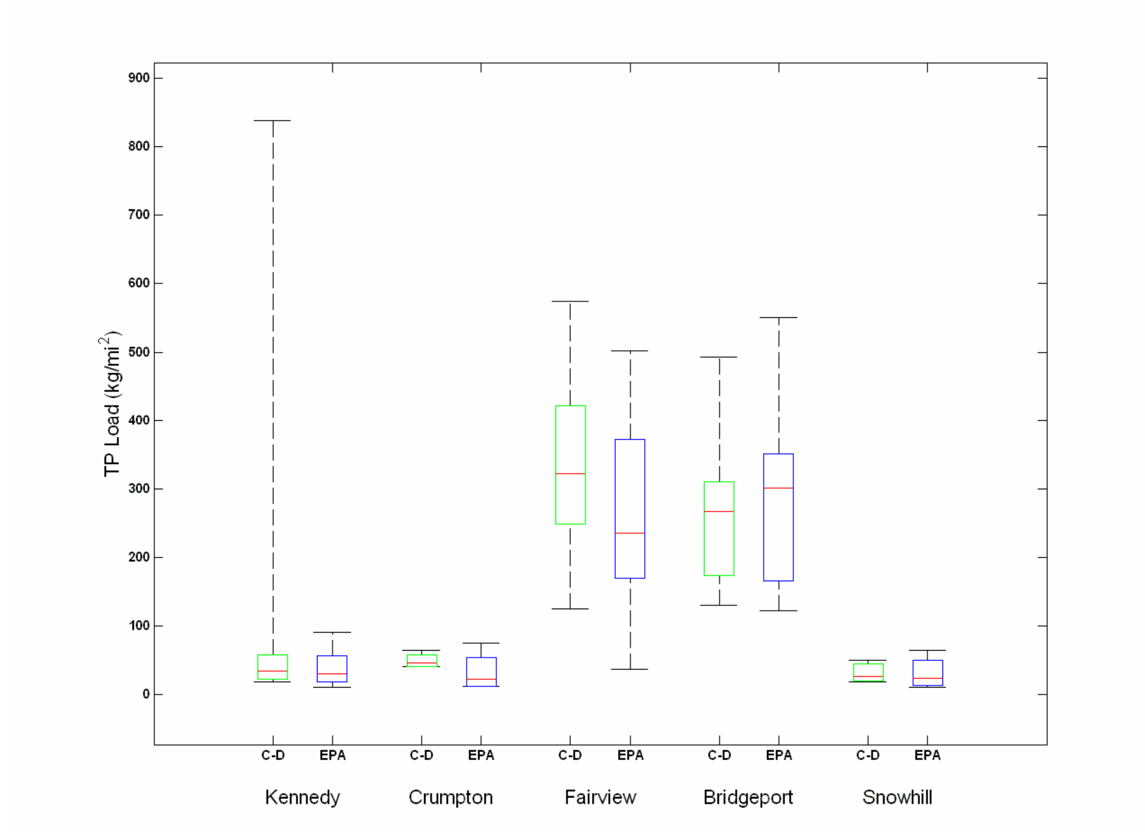

Figure 15: Box and Whisker Plot of Concentration-derived (C-D) and EPA Observed Phosphorus

Figure 15 is a plot of annual phosphorus loads and therefore masks a problem with the EPA daily phosphorus loads. Figure 16 below shows the box and whisker plot of daily phosphorus loads for each of the study years and for each method for the Kennedy watershed. From the figure it is evident that the majority of loads are lower when calculated using the EPA rating curve, which is consistent with Figure 15. However, it also shows that for every year except 1990, 1996, and 1997, the first quartile, median, and third quartile EPA load is 0. This is because the EPA rating curve, shown in Table 17, has a negative intercept and since the Kennedy watershed is only 12.7 square miles the discharges are not large enough to counteract the negative intercept. Therefore many of the daily load values are negative. Since a negative load is physically impossible the load is recorded as 0.

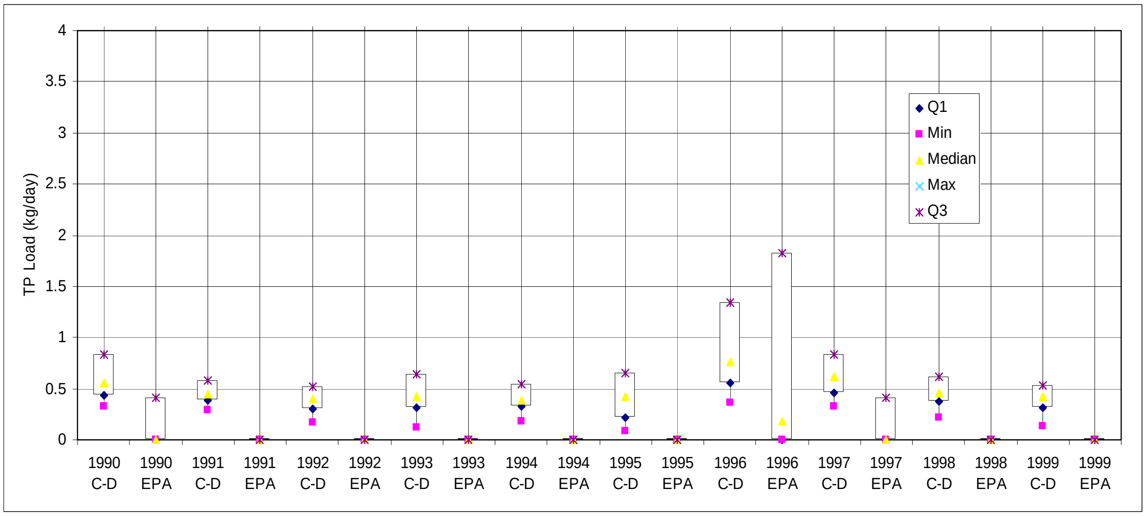

Figure 16: Box and Whisker Plot of Daily Total Phosphorus (TP) Loads for the Kennedy Watershed

The same behavior is seen in the other watersheds; however, it is more pronounced in the Kennedy and Crumpton watersheds because they are small and therefore do not typically have discharges that are large enough to counteract the negative intercept. Since the negative intercept in the EPA total phosphorus rating curves causes the irrational predictions evident in Figure 16, another set of total phosphorus EPA rating curves were developed with the intercept set to zero. A comparison of the concentration-derived and EPA total phosphorus calculated using a zero intercept rating curve follows.

4.1.3 Phosphorus with Zero Intercept Rating Curve

The EPA total phosphorus rating curves with a zero intercept are shown in Table 18. Figure 17 shows a box and whisker plot of annual total phosphorus loads. It compares the concentration-derived loads with the EPA loads as calculated using a rating curve with a zero intercept. By comparing Figure 15 and Figure 17 it can be seen that the EPA phosphorus loads increase when they are calculated using a zero intercept rating curve. Figure 17 shows that the EPA loads calculated using a zero intercept rating curve are higher that the concentration-derived loads for every watershed. Figure 18 shows a box and whisker plot of the daily phosphorus loads for the Kennedy watershed. The irrationality seen in Figure 16 of first, second, and third quartiles being zero is not seen in Figure 18.

Table 18: EPA Rating Curves for Phosphorus with a Zero Intercept

Watershed	Phosphorus (with 0 intercept)
Kennedy	y = 0.2165x
Crumpton	y = 0.5289x
Fairview	y = 0.9622x
Bridgeport	y = 0.8430x
Snowhill	y = 0.1248x

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

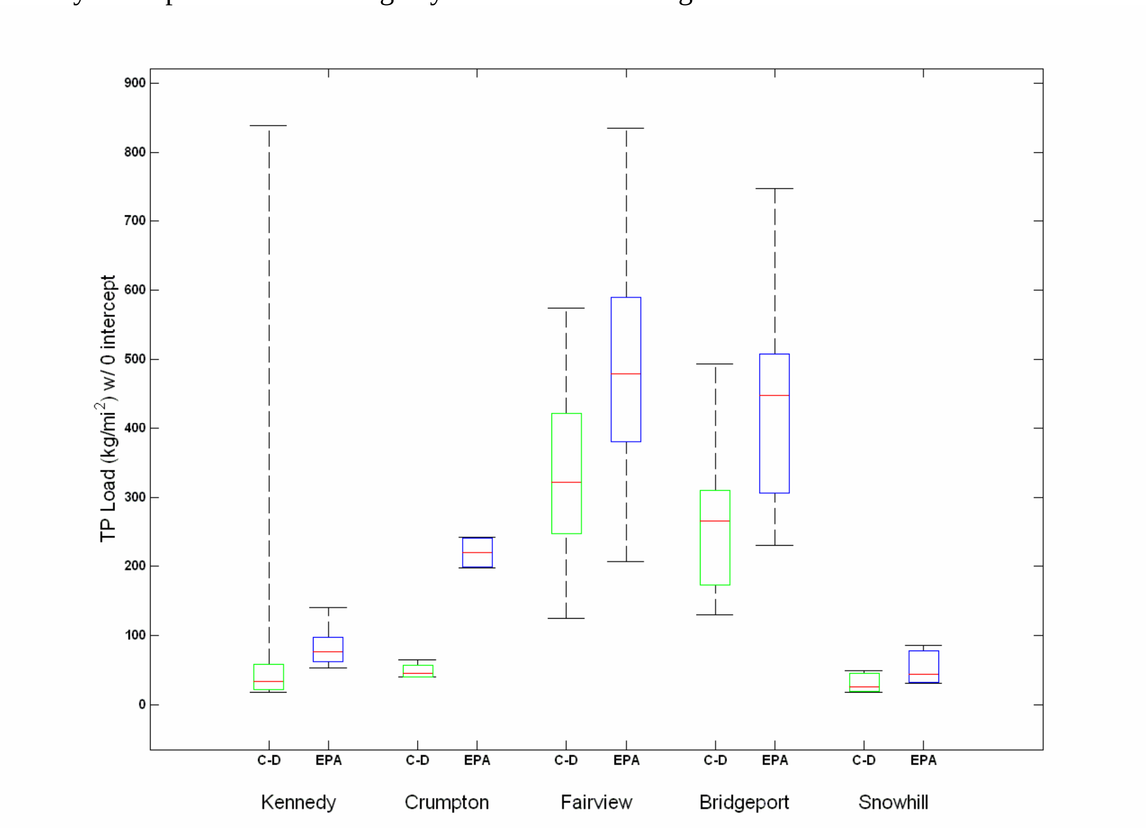

Figure 17: Box and Whisker Plot of Concentration-derived (C-D) and EPA Annual Observed Phosphorus (EPA Phosphorus is Calculated using a Rating Curve with a Zero Intercept)

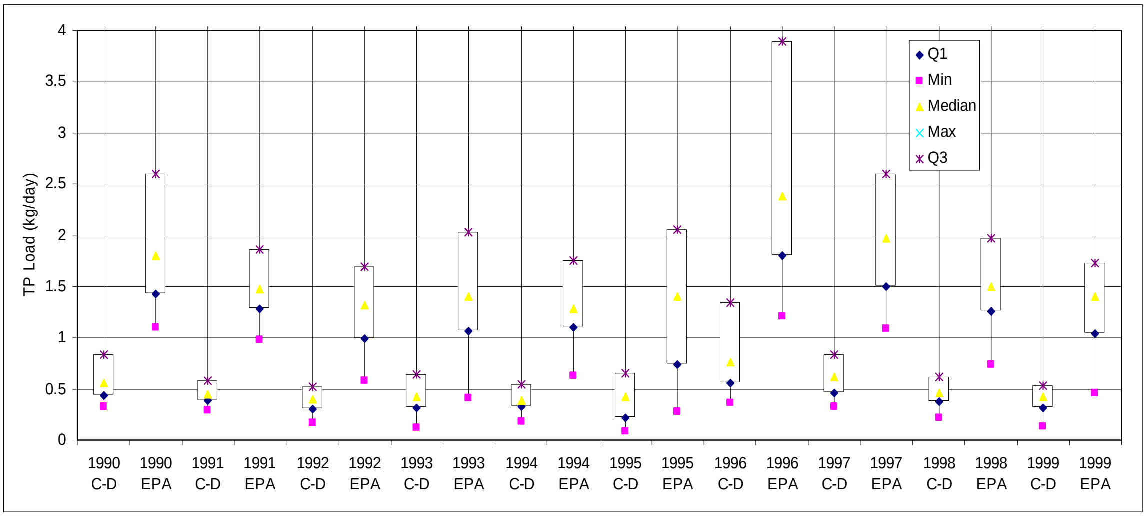

Figure 18: Box and Whisker Plot of Daily Total Phosphorus (TP) Loads Calculated using a Zero Intercept Rating Curve for the Kennedy Watershed

4.1.4 Sediment

Figure 19 shows a box and whisker plot comparing the concentration-derived and EPA annual suspended sediment loads for the study watersheds. From the figure it is evident that the concentration-derived suspended sediment loads are slightly higher for the Bridgeport and Snowhill watersheds and lower for the other three watersheds.

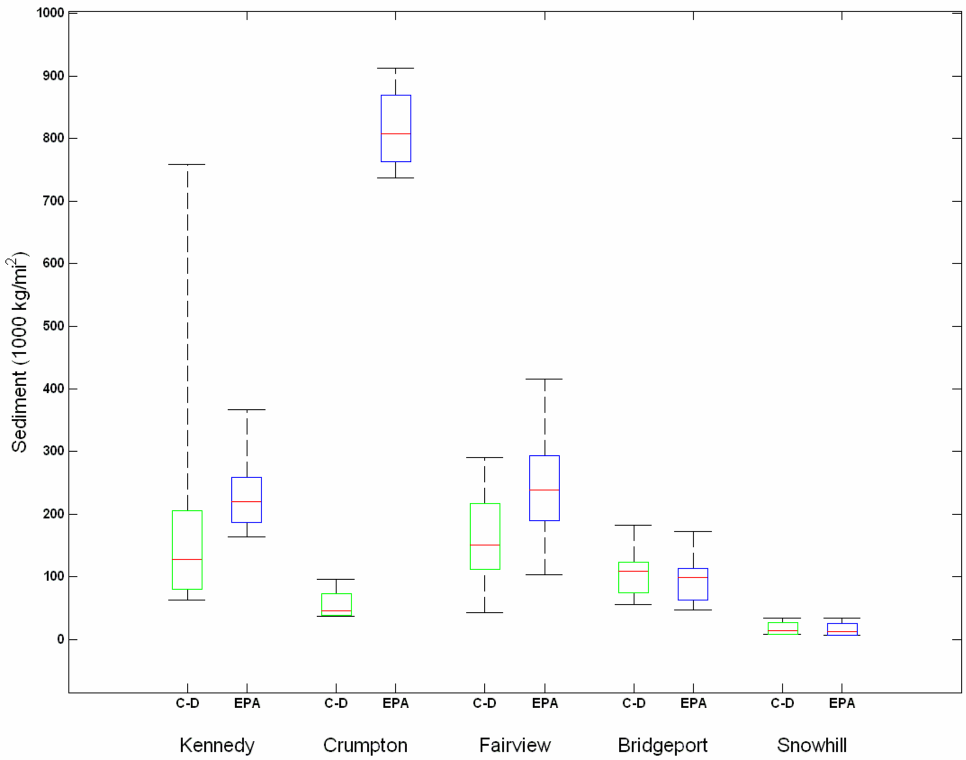

Figure 19: Box and Whisker Plot of Concentration-derived (C-D) and EPA Observed Sediment

As was the case with the total nitrogen loads, the relative values of the concentration-derived and EPA loads depends on the discharge. For the Kennedy and Fairview watersheds, the concentration-derived sediment loads are higher for peak discharges and lower for baseflow. The concentration-derived sediment loads are always lower for the Crumpton watershed regardless of discharge. The Bridgeport and Snowhill watersheds are the opposite of the Kennedy and Fairview watersheds because the concentration-derived loads are higher for baseflows and lower for peaks. In each case the average flow for the watershed is low enough that the method that produces higher loads during baseflows produces the higher load for the majority of the days.

4.2 Evaluation of Simulated Annual Loads

Two different water-quality models were used to simulate total nitrogen, total phosphorus, and total sediment for five gaged watersheds in Maryland. The MDE tool, which uses CBP nutrient loads and MDP 2002 land use, was used to simulate the mean annual pollutant load for the simulation period of 4/1/1990 to 3/31/2000. The AVGWLF model was used to simulate monthly pollutant loads, which were then summed to determine annual loads.

The MDE tool is currently used by the MDE and Department of Natural Resources (DNR) to simulate pollutant loadings in Maryland. One of the main objectives of this study is to determine whether the pollutant loads simulated by AVGWLF are more accurate than those simulated by the MDE tool.

Since the MDE tool can only simulate a mean annual load, a comparison was made between the observed loads calculated by the two different methods and the simulated loads as determined by the two different models on an annual basis.

4.2.1 Comparison of Simulated Loads to Concentration-derived Observed Loads

The simulated pollutant loads and concentration-derived loads were compared using three different methods. First, box and whisker plots were developed to represent the observed and AVGWLF simulated loads in order to show the variation over the 10 year study period. A horizontal line, on the plots, represents the MDE tool load because only an average load for the 10 year simulation length is simulated. These plots are used to make a qualitative assessment of the accuracy of the MDE tool and AVGWLF simulated loads in relation to each other. The model is considered to be moderately accurate if the 1st or 3rd quartiles of the AVGWLF simulated annual pollutant loads or the

MDE mean annual simulated pollutant load is between the 1st and 3rd quartiles of the observed loads. The model is considered to over predict if the median AVGWLF annual pollutant load or the MDE mean annual pollutant load is above the median observed load. Similarly, the model is considered to under predict if the median AVGWLF annual pollutant load or the MDE mean annual pollutant load is below the median observed load.

In order to make a quantitative assessment of the accuracy of the AVGWLF and the MDE tool simulated loads a t-test was used to determine whether the mean annual simulated loads were statistically different than the mean annual observed loads. The t-test was used with the following test statistic:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_p \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{0.5}} \tag{4-1}$$

where $\overline{X}_1$ and $\overline{X}_2$ are the means of the observed loads and simulated loads, respectively, n_1 and n_2 are the sample sizes of the observed and simulated loads, respectively, and S_p is the square root of the pooled variance which is given by the equation:

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \tag{4-2}$$

where S_1^2 and S_2^2 are the variances of the observed and simulated loads, respectively.

A two-tailed test was used because the goal is to determine whether the means are different, not whether one is greater or less than the other. The mean annual simulated and observed loads are statistically different if the *t* value, as calculated using equation 4-1, is less than or greater than the critical *t* value. The critical *t* value depends on α and *n*. In this study an α value of 5 percent was used. Since the *n* value varies depending on which model simulation is being compared to the observed (the AVGWLF simulation has 10 annual values and the MDE tool has 1 annual value), the critical *t* values also vary. The critical *t* values for the Crumpton watershed are also different because daily observed discharge was only available for 1996 until 2000 and therefore the *n* value for both the observed and the AVGWLF simulated annual loads is 4. Table 19 shows the critical *t* values by model and Table 20 shows the critical *t* values used for the Crumpton watershed by model.

Table 19: Critical *t* Values

Model	*n*	*t* critical
AVGWLF	10	2.101
MDE tool	1	2.262

Table 20: Critical *t* Values used for the Crumpton Watershed

Model	*n*	*t* critical
AVGWLF	4	2.447
MDE tool	1	3.182

Table 21 through Table 34 show the standard deviation of the observed pollutant loads (S_{obs}), the standard deviation of the simulated pollutant loads (S_{MDE} or S_{avgwlf} depending on model used), the S_p calculated using equation 4-2, the *t* value calculated using equation 4-1, and the decision made on whether the simulated mean annual load is statistically the same as the observed load.

Nash-Sutcliffe coefficients were also calculated in order to determine whether the simulated loads were more accurate than using an average observed load. The Nash-Sutcliffe (*N-S*) coefficient is given by equation 4-3 (Nash and Sutcliffe, 1970) below:

$$\text{N - S} = 1 - \left[\frac{(Q_o - Q_p)^2}{(Q_o - Q_a)^2} \right] \tag{4-3}$$

where Q_o is the observed pollutant load, Q_p is the predicted pollutant load, and Q_a is the average pollutant load. An *N-S* coefficient of 1 indicates that the model simulations match the observed values perfectly. An *N-S* coefficient of 0 indicates that the model simulations are no better than using an average observed value to predict pollutant loads and a negative *N-S* coefficient indicates that using an average load would provide more accurate estimates than the model simulations.

4.2.1.1 Nitrogen

Figure 20 shows the concentration-derived annual total nitrogen loads, the AVGWLF annual total nitrogen loads, and the MDE tool mean annual total nitrogen load for the period of 4/1/1990 to 3/31/2000. AVGWLF under predicts nitrogen in all five watersheds, however, it provides a moderate estimate for the Bridgeport and Snowhill watersheds because the boxes representing the 1st and 3rd quartiles overlap.

Figure 20 also shows that the MDE tool over predicts for three out of the five watersheds. For the Kennedy and Snowhill watersheds, the MDE tool simulated nitrogen load is larger than the maximum observed value and therefore provides a very poor estimate of nitrogen loads. For the Bridgeport watershed the MDE tool simulated load is larger than the third quartile of the observed load. For both the Crumpton and Fairview watersheds the MDE tool simulated load is below the concentration-derived observed load; however, the MDE tool provides a moderate estimate for the Crumpton watershed because the simulated load is between the median and 3rd quartile of the observed. The AVGWLF simulated load provides a more accurate estimate for the Kennedy, Bridgeport and Snowhill watersheds and the MDE tool simulated load provides a more accurate estimate for the Crumpton and Fairview watersheds. Based on Figure 20 I would conclude that AVGWLF is a better model for predicting total nitrogen because it more accurately simulates the total nitrogen observed loads for three out of the five watersheds.

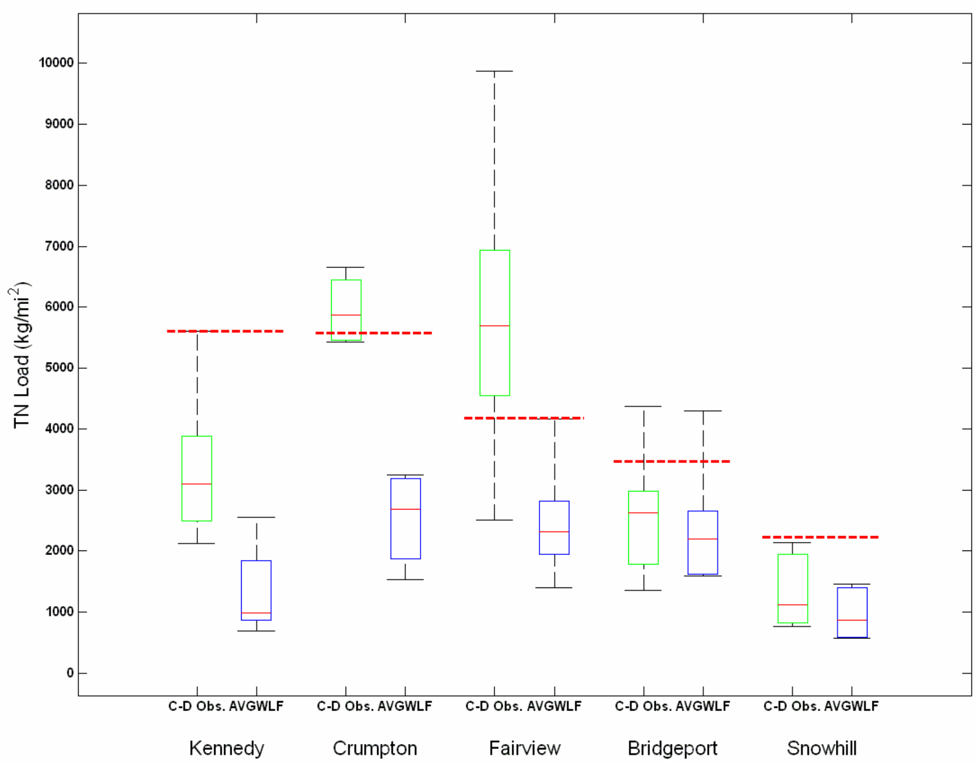

Figure 20: Box and Whisker Plot of Concentration-derived (C-D) Observed Annual Nitrogen (TN) Loads and AVGWLF and MDE Tool Simulated Annual Nitrogen Loads

Table 21 and Table 22 show the t-test statistics for the MDE tool and AVGWLF simulated nitrogen loads, respectively. Table 21 shows that for the pollutant nitrogen the MDE tool mean annual simulated loads are not statistically different than the concentration-derived observed mean annual loads. Table 22 shows that for the Bridgeport and Snowhill watersheds the AVGWLF simulated mean annual load is not statistically different than the observed load. However, at the 5 percent level the AVGWLF simulated mean annual loads for the Kennedy, Crumpton, and Fairview watersheds are statistically different than the observed loads. This conclusion can be inferred from Figure 20 because the boxes for the observed and AVGWLF simulated loads overlap for the Bridgeport and Snowhill watersheds, indicating that the AVGWLF simulated loads are more accurate for these two watersheds. Based on Table 21 and Table 22 I would conclude that the MDE tool is a better model because the mean annual simulated nitrogen loads are statistically the same as the concentration-derived observed loads for all five watersheds.

Table 21: Hypothesis Test Statistics for Concentration-derived Observed and MDE Tool Simulated Total Nitrogen Loads

Watershed	S_{obs} (kg/yr)	S_{MDE} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	1.394E+04	0	1.945E+08	-1.821	same
Crumpton	3850	0	1.482E+07	1.170	same
Fairview	1.083E+06	0	1.172E+12	0.7598	same
Bridgeport	1.791E+05	0	3.207E+10	-0.8032	same
Snowhill	2.607E+04	0	6.799E+08	-1.465	same

Table 22: Hypothesis Test Statistics for Concentration-derived Observed and AVGWLF Simulated Total Nitrogen Loads

Watershed	S_{obs} (kg/yr)	S_{AVGWLF} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	1.394E+04	8165	1.306E+08	4.813	different
Crumpton	3850	5271	2.130E+07	7.712	different
Fairview	1.083E+06	3.932E+05	6.633E+11	4.745	different
Bridgeport	1.791E+05	1.515E+05	2.751E+10	0.4715	same
Snowhill	2.607E+04	1.728E+04	4.893E+08	1.756	same

The mean annual total nitrogen loads per square mile watershed area were calculated for each watershed and Figure 21 shows the AVGWLF and MDE tool simulated mean annual loads per square mile versus the observed mean annual loads per square mile.

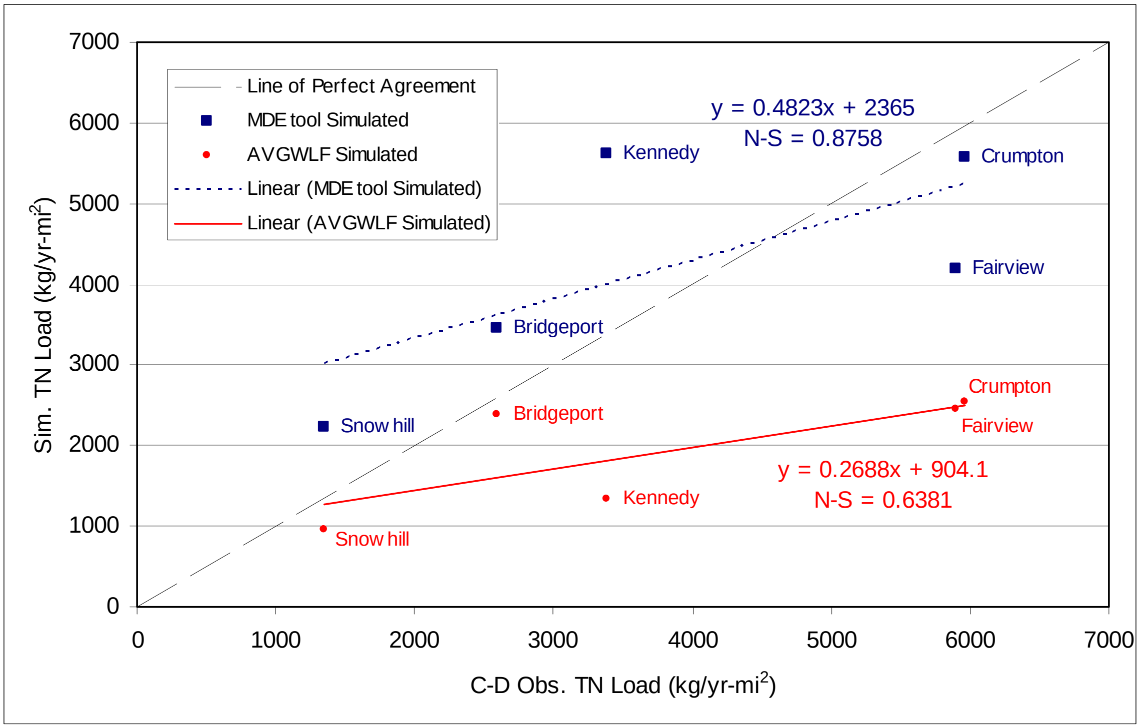

Figure 21: AVGWLF and MDE Tool Simulated Total Nitrogen (TN) Loads versus Concentration-derived (C-D) Observed Loads

The *N-S* coefficient for the MDE tool simulated mean annual loads is 0.88 and the *N-S* coefficient for the AVGWLF simulated mean annual loads is 0.64. Both coefficients are positive indicating that both models are better for estimating annual loads than using an average observed load. However, the *N-S* coefficient is higher for the MDE tool simulated loads, which indicates that the MDE tool is better than the AVGWLF model at predicting total nitrogen load per unit area on an annual time scale.

4.2.1.2 Phosphorus

Figure 22 shows the concentration-derived annual phosphorus loads, the AVGWLF simulated annual loads, and the MDE tool simulated mean annual loads for the period of 4/01/1990 to 3/31/2000. AVGWLF over predicts phosphorus loads for the

Crumpton watershed and under predicts for the Fairview watershed. AVGWLF provides a moderately accurate estimate for the Kennedy, Bridgeport, and Snowhill watersheds.

Figure 22 also shows that the MDE tool over predicts for every watersheds except Fairview, for which it under predicts. Based on Figure 22, I would conclude that AVGWLF is a better model because the simulated phosphorus loads are more accurate for four out of the five watersheds (the exception being the Fairview watershed).

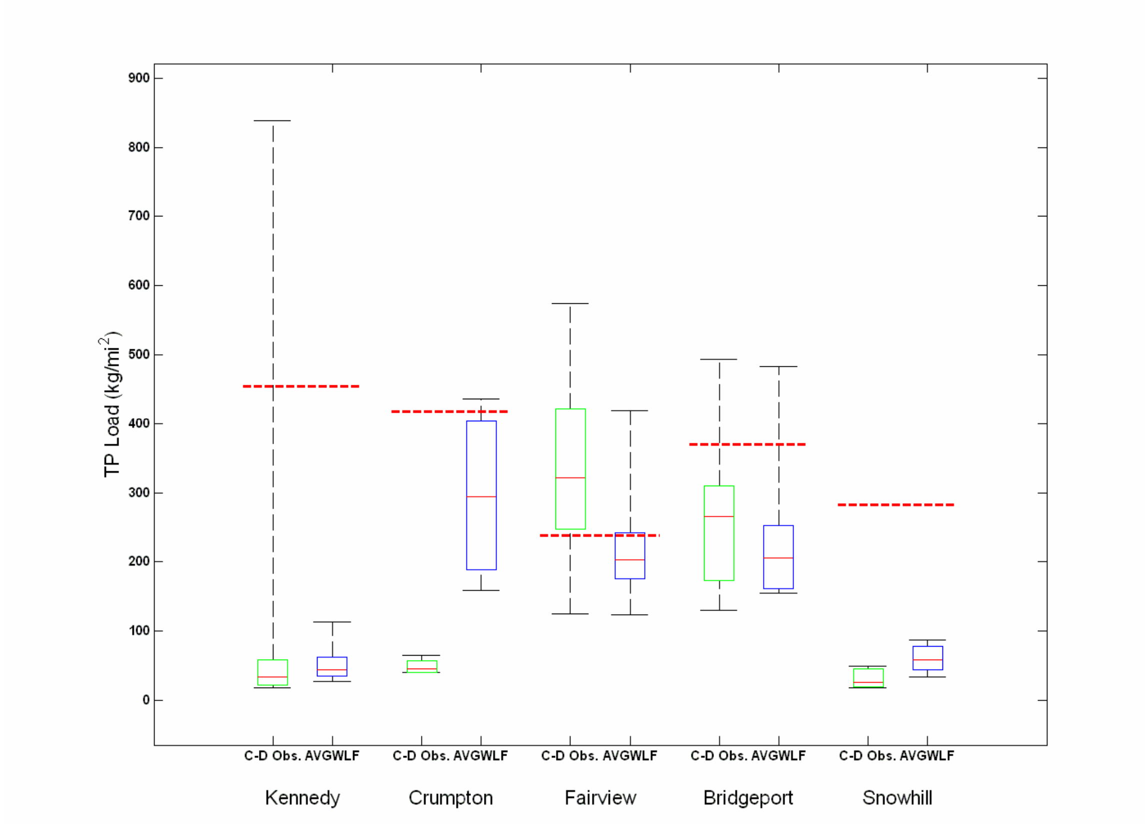

Figure 22: Box and Whisker Plot of Concentration-derived (C-D) Observed Annual Phosphorus Loads and MDE Tool and AVGWLF Simulated Annual Phosphorus (TP) Loads

Table 23 and Table 24 show the t-test statistics for the MDE tool and AVGWLF simulated total phosphorus loads, respectively. Table 23 shows that the MDE tool simulated load is statistically the same as the observed mean annual load for the Kennedy, Fairview, and Bridgeport watersheds. Table 24 shows that the AVGWLF simulated mean annual loads are statistically the same as the observed loads for the Kennedy and Bridgeport watersheds. Based on Table 23 and Table 24 I would conclude that the MDE tool is a better model because the simulated mean annual load is statistically the same as the observed for three out of five watersheds, as opposed to the AVGWLF model, which is only the same for two out of five watersheds.

It should be noted that the conclusions made based on the t-tests are dependent on the α value selected. If the α value were 0.5 percent, as opposed to 5 percent, the AVGWLF mean annual loads for the Crumpton and Fairview watersheds would be statistically the same as the observed mean annual loads. However, the MDE tool simulated loads for the Crumpton and Snowhill watersheds would still be statistically different than the observed loads because the *t* values are so large (-29.01 and -17.80, respectively). Therefore, if the α value were 0.5 percent the AVGWLF model would be the more accurate model because the mean annual simulated loads would be statistically the same as the observed loads for four out of five watersheds, as opposed to the MDE tool, which is the same for three out of five.

Table 23: Hypothesis Test Statistics for Concentration-derived Observed and MDE Tool Simulated Total Phosphorus Loads

Watershed	S_{obs} (kg/yr)	S_{MDE} (kg/yr)	S_p^2 (kg/yr)^2	*t*	Decision
Kennedy	3042	0	9.253E+06	-1.263	same
Crumpton	73.41	0	5389	-29.01	different
Fairview	7.159E+04	0	5.125E+09	0.6692	same
Bridgeport	2.183E+04	0	4.766E+08	-0.7408	same
Snowhill	604.9	0	3.659E+05	-17.80	different

Table 24: Hypothesis Test Statistics for Concentration-derived Observed and AVGWLF Simulated Total Phosphorus Loads

Watershed	S_{obs} (kg/yr)	S_{AVGWLF} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	3042	346.6	4.687E+06	0.7634	same
Crumpton	73.41	840.3	3.558E+05	-3.761	different
Fairview	7.159E+04	4.051E+04	3.383E+09	2.244	different
Bridgeport	2.183E+04	1.692E+04	3.815E+08	0.8115	same
Snowhill	604.9	863.9	5.562E+05	-3.947	different

Figure 23 shows the AVGWLF and MDE tool simulated mean annual total phosphorus load per unit area versus the observed mean annual load per unit area for each study watershed. The calculated *N-S* coefficient for the MDE tool is -0.71. Since the coefficient is below 0 it indicates that using an average observed load per unit area would give better predictions than the MDE tool for mean annual total phosphorus. The *N-S* coefficient for the AVGWLF model is 0.58, indicating that AVGWLF predicts better than using an average observed load. Since the *N-S* coefficient for the AVGWLF model (0.58) is higher than the *N-S* coefficient for the MDE tool (-0.71), it can be concluded that the AVGWLF model is more accurate at predicting total phosphorus on an annual time scale.

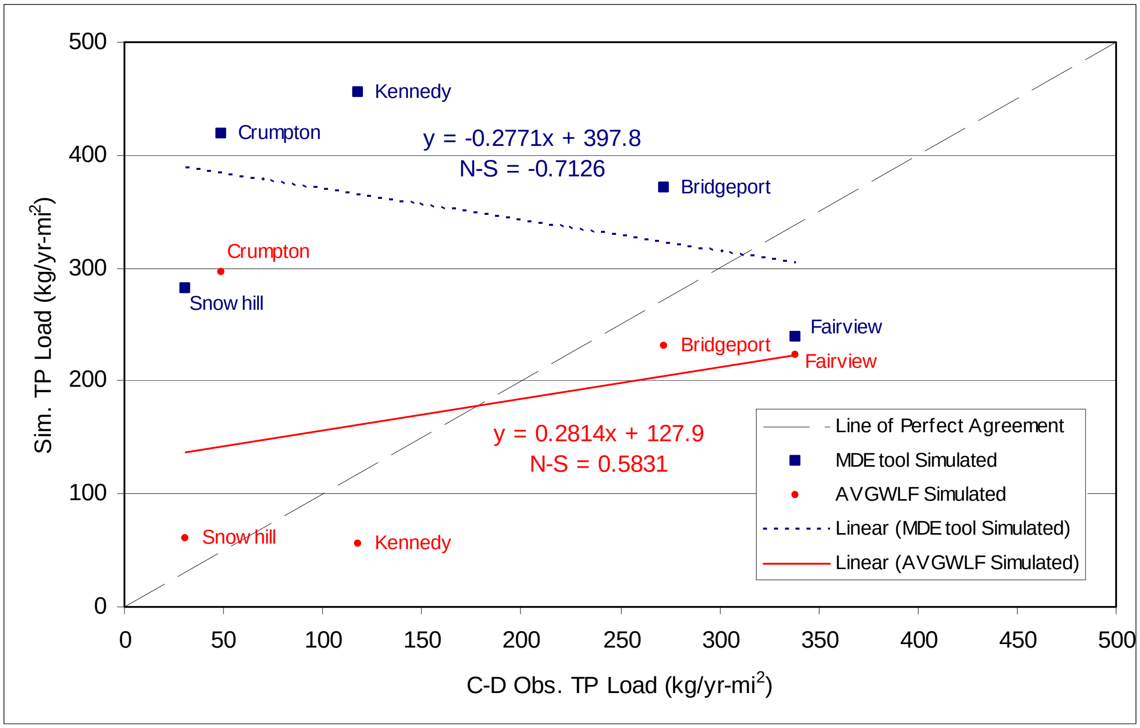

Figure 23: AVGWLF and MDE Tool Simulated Total Phosphorus (TP) Loads versus Concentration-derived (C-D) Observed Loads

4.2.1.3 Sediment

Figure 24 shows the concentration-derived annual suspended sediment loads, the AVGWLF annual simulated total sediment loads, and the MDE tool mean annual total sediment load for the period of 4/1/1990 to 3/31/2000. The comparison between observed and simulated sediment loads is difficult to make because the observed is suspended sediment and the simulated is total sediment, which includes bed load and suspended sediment (USGS, 1994). However, since bed load is a very small percentage of suspended sediment load (Pizzuto, pers. comm., 2007), I assumed that the total sediment equals the suspended sediment and made the comparison.

From Figure 24 it can be seen that AVGWLF over predicts sediment for every watershed except the Kennedy watershed. It can also be seen that the MDE tool over predicts the sediment load for every watershed (i.e. the MDE mean annual sediment load is above the median observed annual sediment load); however, the MDE tool simulated sediment load is below the median AVGWLF simulated load for every watershed except Kennedy, meaning that the MDE tool provides a more accurate estimate for annual sediment load for every watershed except Kennedy.

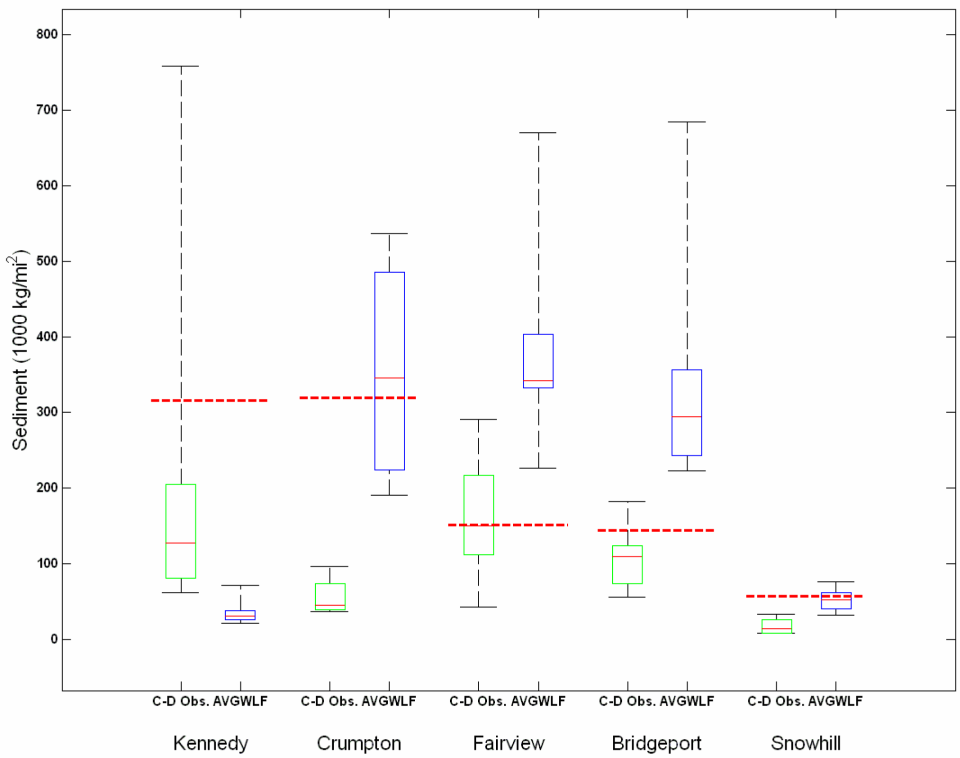

Figure 24: Box and Whisker Plot of Concentration-derived (C-D) Observed Annual Sediment Loads and MDE Tool and AVGWLF Simulated Annual Sediment Loads

Table 25 below shows the t-test statistics for the MDE tool simulated sediment loads. According to the test statistic the MDE tool predicts mean annual loads that were statistically the same for the Kennedy, Fairview, and Bridgeport watersheds.

Table 26 shows the t-test statistics for the AVGWLF simulated mean annual sediment loads. The t-test found every predicted AVGWLF load to be statistically different than the observed load. Therefore, according to Table 25 and Table 26, the MDE tool is more accurate because the simulated sediment loads are statistically the same for three out of five watersheds.

Table 25: Hypothesis Test Statistics for Concentration-derived Observed and MDE Tool Simulated Total Sediment Loads

Watershed	S_{obs} (Mg/yr)	S_{MDE} (Mg/yr)	S_p^2 (Mg/yr)^2	t	Decision
Kennedy	2511	0	6.307E+06	-0.5351	same
Crumpton	177.0	0	3.132E+04	-8.493	different
Fairview	4.171E+04	0	1.740E+09	0.1337	same
Bridgeport	7475	0	5.588E+07	-0.8043	same
Snowhill	453.9	0	2.061E+05	-3.660	different

Table 26: Hypothesis Test Statistics for Concentration-derived Observed and AVGWLF Simulated Total Sediment Loads

Watershed	S_{obs} (Mg/yr)	S_{AVGWLF} (Mg/yr)	S_p^2 (Mg/yr)^2	t	Decision
Kennedy	2511	195.4	3.173E+06	2.403	different
Crumpton	177.0	1033	5.494E+05	-3.749	different
Fairview	4.171E+04	5.932E+04	2.629E+09	-4.677	different
Bridgeport	7475	2.347E+04	3.035E+08	-4.930	different
Snowhill	453.9	640.0	3.078E+05	-6.284	different

Figure 25 shows the AVGWLF and MDE tool simulated mean annual sediment load per unit area versus the observed mean annual load per unit area for each study watershed. The *N-S* coefficient for the MDE tool is -0.72, which indicates that using the average observed load per unit area would be better for predicting sediment. The *N-S* coefficient for the AVGWLF model is -3.22. Since the coefficient is negative, better sediment predictions would be obtained by using an average mean annual load per unit area.

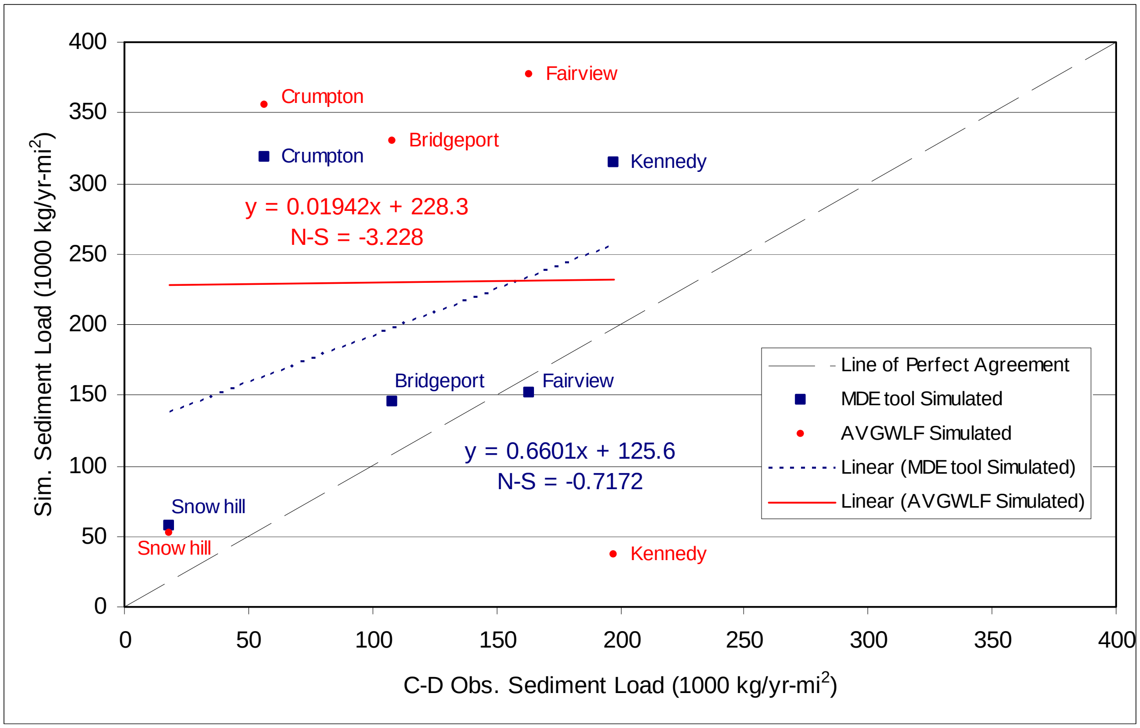

Figure 25: AVGWLF and MDE Tool Simulated Total Sediment Loads versus Concentration-derived (C-D) Observed Loads

4.2.2 Comparison of Simulated Loads to EPA Observed Loads

Similarly to the last section, simulated pollutant loads and EPA loads were compared using three different methods. Box and whisker plots were developed to provide a qualitative comparison between AVGWLF and MDE tool simulated loads and EPA observed loads. A t-test was performed to determine whether the simulated pollutant loads were statistically different than the EPA observed loads at the $\alpha = 5$ percent level. *N-S* coefficients were also calculated to assess whether the model provides better predictions than simply using an average observed load.

4.2.2.1 Nitrogen

Figure 26 shows the EPA observed annual total nitrogen loads, the AVGWLF simulated annual total nitrogen loads, and the MDE tool mean annual total nitrogen load for the period of 4/1/1990 to 3/31/2000. AVGWLF under predicts total nitrogen for all five watersheds, although it simulates loads with moderate accuracy for the Snowhill watershed. The MDE tool over predicts for the Kennedy watershed, provides a moderate estimate for the Snowhill watershed and under predicts for the Crumpton, Fairview, and Bridgeport watersheds. The AVGWLF simulated loads provide a more accurate estimate for the Kennedy and Snowhill watersheds, while the MDE tool provides a more accurate estimate for the Crumpton, Fairview and Bridgeport watersheds. Therefore, according to Figure 26, the MDE tool is a better model.

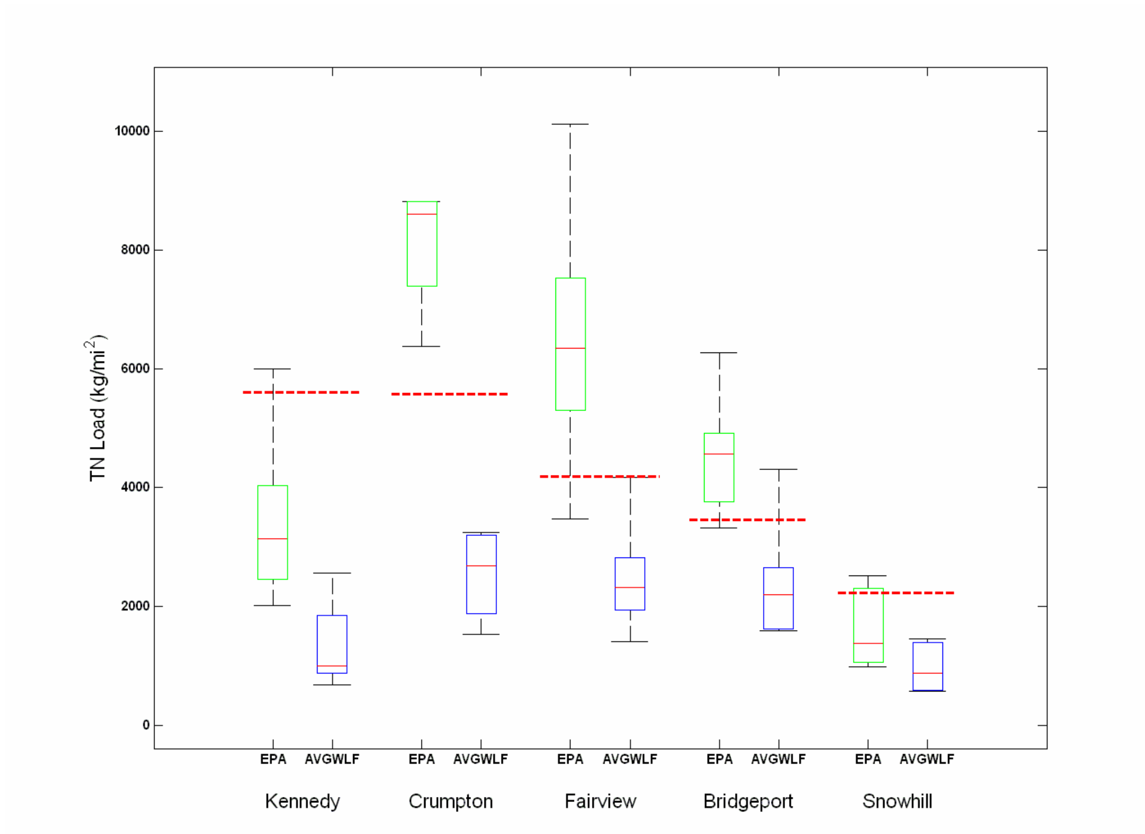

Figure 26: Box and Whisker Plot of EPA Observed Annual Nitrogen (TN) Loads and AVGWLF and MDE Tool Simulated Annual Nitrogen Loads

Table 27 and Table 28 show the t-test statistics for the MDE tool and AVGWLF simulated total nitrogen loads, respectively. Table 27 shows that for the pollutant total nitrogen the MDE tool mean annual simulated loads are not statistically different than the EPA observed mean annual loads. Table 28 shows that the AVGWLF simulated mean annual loads are statistically different than the EPA observed mean loads for all five watersheds. Based on Table 27 and Table 28, I would conclude that the MDE tool is a better model because the mean annual simulated total nitrogen loads are statistically the same as the EPA observed loads for all five watersheds.

Table 27: Hypothesis Test Statistics for EPA Observed and MDE Tool Simulated Total Nitrogen Loads

Watershed	S_{obs} (kg/yr)	S_{MDE} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	1.596E+04	0	2.546E+08	-1.539	same
Crumpton	7563	0	5.720E+07	2.356	same
Fairview	9.902E+05	0	9.805E+11	1.144	same
Bridgeport	1.754E+05	0	3.077E+10	1.009	same
Snowhill	2.910E+04	0	8.467E+08	-0.8917	same

Table 28: Hypothesis Test Statistics for EPA Observed and AVGWLF Simulated Total Nitrogen Loads

Watershed	S_{obs} (kg/yr)	S_{AVGWLF} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	1.596E+04	8165	1.606E+08	4.495	different
Crumpton	7563	5271	4.249E+07	8.691	different
Fairview	9.902E+05	3.932E+05	5.675E+11	6.097	different
Bridgeport	1.754E+05	1.515E+05	2.686E+10	7.618	different
Snowhill	2.910E+04	1.728E+04	5.727E+08	2.824	different

Figure 27 shows the MDE tool and AVGWLF simulated mean annual total nitrogen load per unit area versus the EPA observed mean annual total nitrogen load per unit area. The *N-S* coefficients for the MDE tool and AVGWLF mean annual loads are 0.86 and 0.56, respectively. Both are positive and therefore provide better estimates of the EPA observed than using an average observed value. The *N-S* coefficient is larger for the MDE tool, indicating that the MDE tool provides better estimates than the AVGWLF model.

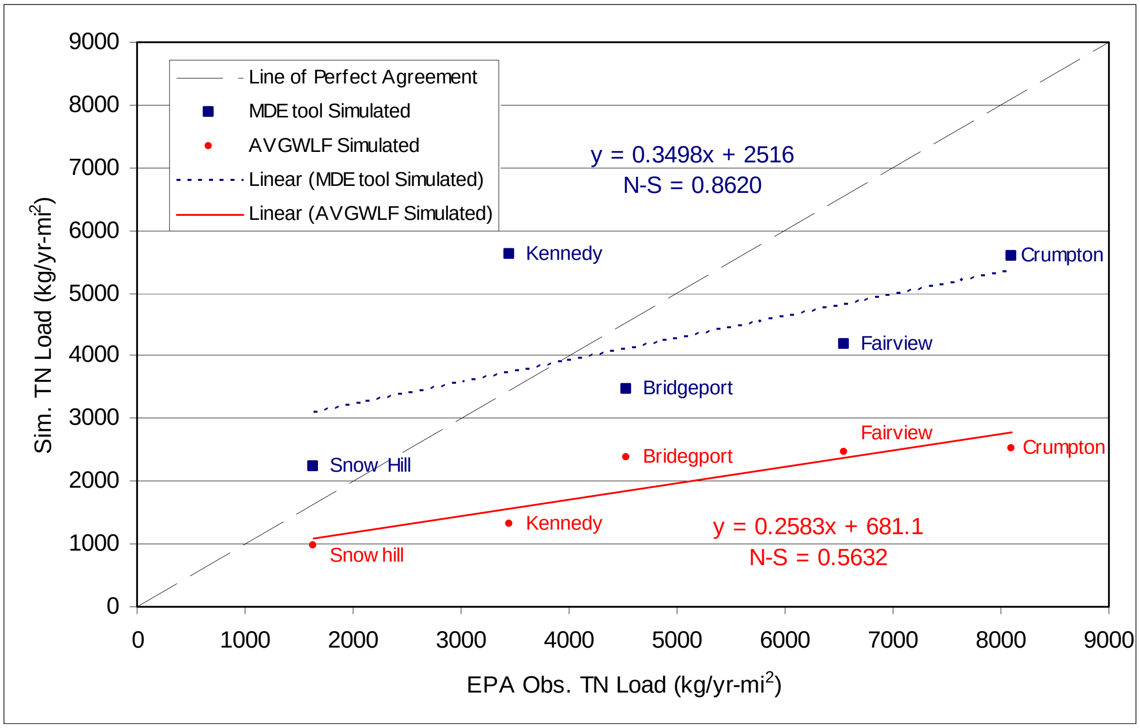

Figure 27: AVGWLF and MDE Tool Simulated Total Nitrogen (TN) Loads versus EPA Observed Loads

4.2.2.2 Phosphorus

Figure 28 shows the EPA total phosphorus annual loads, the AVGWLF and the MDE tool simulated annual loads for the period 4/01/1990 to 3/31/2000. The figure shows that AVGWLF over predicts for the Crumpton watershed and provides a moderate estimate of annual phosphorus loads for the other four watersheds. It also shows that the MDE tool provides a moderate estimate of phosphorus load for the Fairview watershed and over predicts annual phosphorus for the other four watersheds. The AVGWLF model provides a better estimate of annual phosphorus loads for the Kennedy, Crumpton, and Snowhill watersheds. The MDE tool provides a better estimate of annual phosphorus loads for the Fairview and Bridgeport watersheds. Based on Figure 28, AVGWLF is a slightly more accurate model because it provides more accurate estimates for three out of the five watersheds.

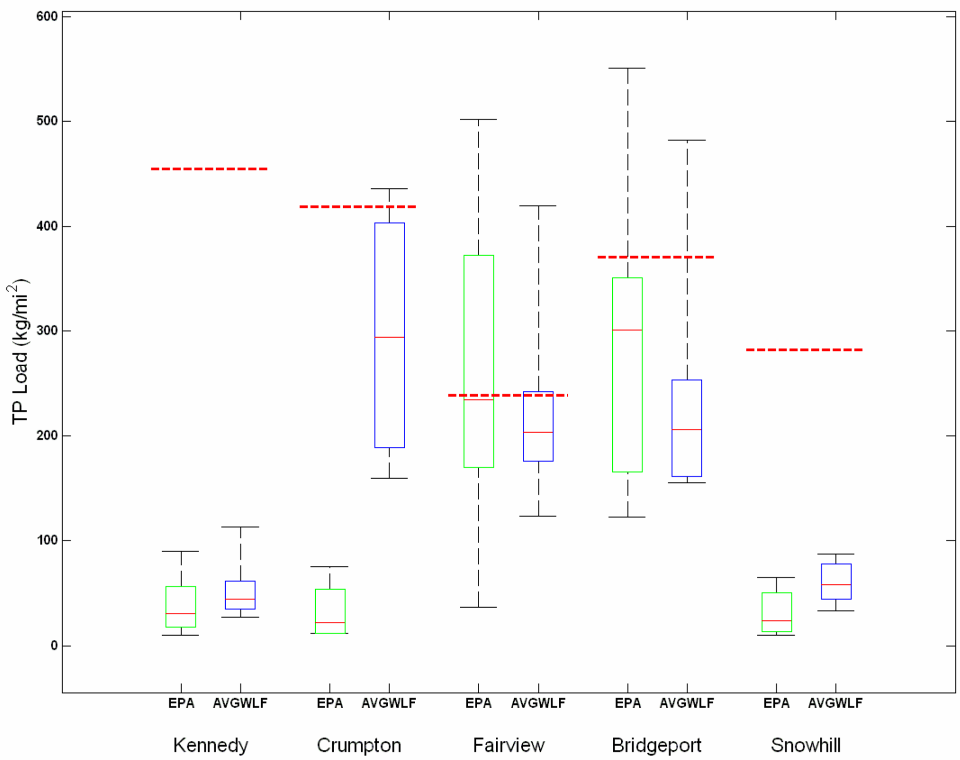

Figure 28: Box and Whisker Plot of EPA Observed Annual Phosphorus Loads and AVGWLF and MDE Tool Simulated Annual Phosphorus Loads

Table 29 and Table 30 show the t-test statistics for the MDE tool and the AVGWLF simulated total phosphorus loads, respectively. Table 29 shows that the MDE tool simulated mean annual loads are statistically the same as the observed for the Fairview and Bridgeport watersheds. Table 30 shows that the AVGWLF simulated mean annual phosphorus loads are statistically the same as the observed for the Kennedy, Fairview, and Bridgeport watersheds. According the Table 29 and Table 30, the AVGWLF model performs slightly better because it correctly predicts phosphorus loads for three out of five watersheds, as opposed to two out of five watersheds for the MDE tool.

Table 29: Hypothesis Test Statistics for EPA Observed and MDE Tool Simulated Total Phosphorus Loads

Watershed	S_{obs} (kg/yr)	S_{MDE} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	346.5	0	1.201E+05	-13.68	different
Crumpton	192.6	0	3.708E+04	-11.57	different
Fairview	7.905E+04	0	6.249E+09	0.1448	same
Bridgeport	2.632E+04	0	6.927E+08	-0.4463	same
Snowhill	957.7	0	9.171E+05	-11.21	different

Table 30: Hypothesis Test Statistics for EPA Observed and AVGWLF Simulated Total Phosphorus Loads

Watershed	S_{obs} (kg/yr)	S_{AVGWLF} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	346.5	346.6	1.201E+05	-1.320	same
Crumpton	192.6	840.3	3.716E+05	-3.938	different
Fairview	7.905E+04	4.051E+04	3.945E+09	0.717	same
Bridgeport	2.632E+04	1.692E+04	4.896E+08	1.185	same
Snowhill	957.7	863.9	8.318E+05	-3.134	different

Figure 29 shows the MDE tool and the AVGWLF simulated mean annual total phosphorus loads per unit area versus the EPA observed mean annual loads per unit area for each watershed. The *N-S* coefficient for the MDE tool mean annual loads is –1.302. Since the *N-S* coefficient is less than zero, using an average observed phosphorus load would provide a better estimate than the MDE tool. The *N-S* coefficient for the AVGWLF mean annual loads is 0.55, which means that the AVGWLF model produces better results than using an average load.

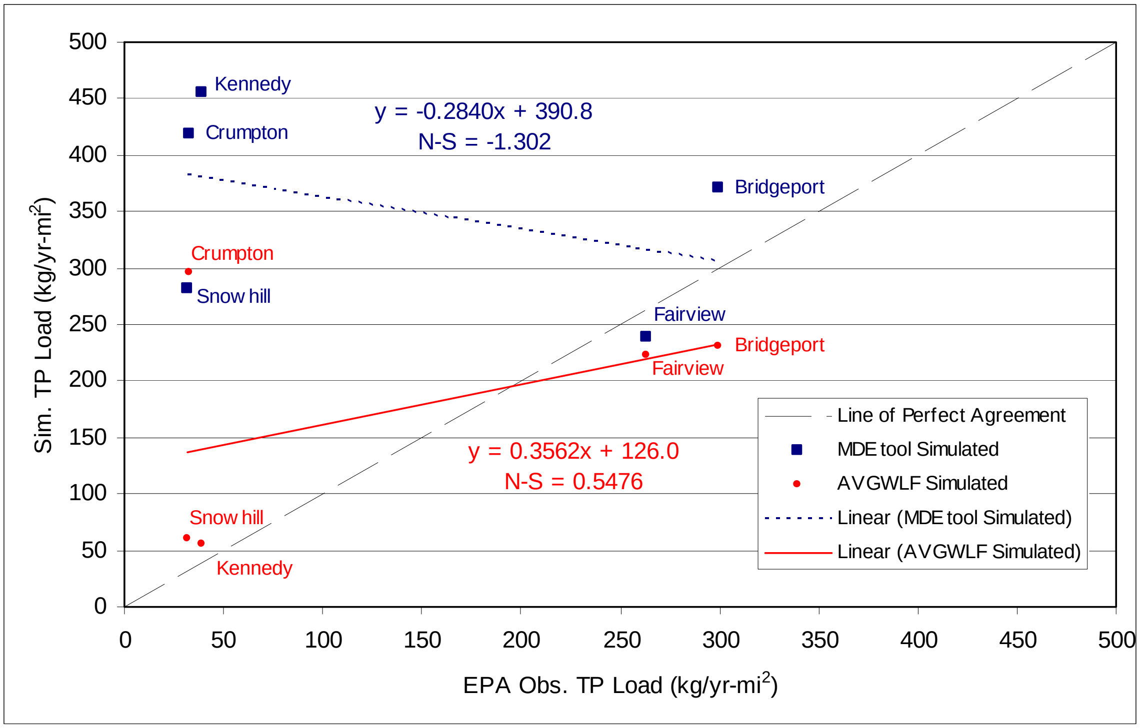

Figure 29: AVGWLF and MDE Tool Simulated Total Phosphorus Loads versus EPA Observed Loads

4.2.2.3 Phosphorus Calculated using a Zero Intercept Rating Curve

Figure 30 shows the EPA observed annual total phosphorus loads calculated using a rating curve with a zero intercept, the AVGWLF simulated annual total phosphorus loads, and the MDE tool mean annual phosphorus load for the period of 4/1/1990 to 3/31/2000. The AVGWLF model under predicts total phosphorus for the Kennedy, Fairview, and Bridgeport watersheds. Note that when phosphorus is calculated using a rating curve without a zero intercept, AVGWLF provides a moderate prediction of phosphorus loads for the Kennedy, Fairview, and Bridgeport watersheds. The AVGWLF model predicts moderately well for the Crumpton and Snowhill watersheds, although the simulated median load is higher than the observed median load for both watersheds. The MDE tool over predicts annual phosphorus for the Kennedy, Crumpton, and Snowhill watersheds. It provides a moderate estimate for the Bridgeport watershed, although the simulated mean annual load is below the median observed load, and it under predicts for the Fairview watershed. The AVGWLF model provides a more accurate estimate of annual phosphorus loads for three out of the five watersheds and therefore, according to Figure 30, is the better model.

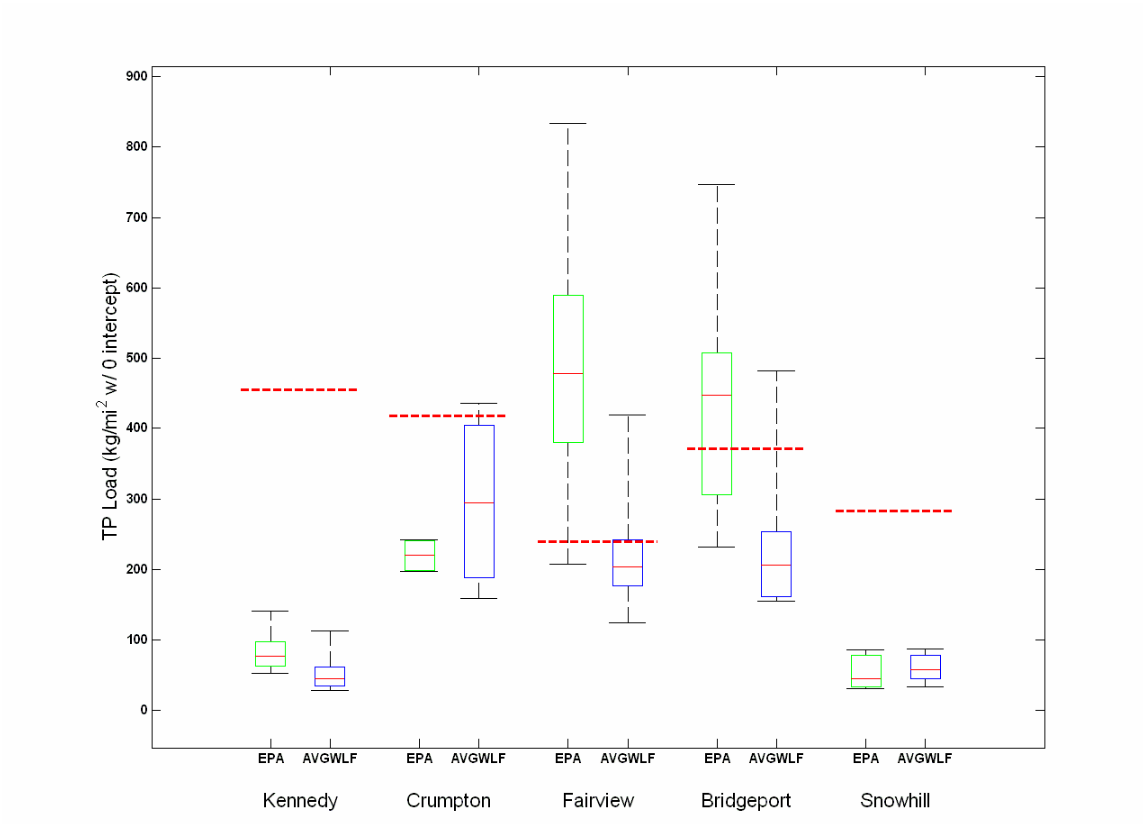

Figure 30: Box and Whisker Plot of EPA Observed Annual Phosphorus Loads Calculated using a Zero Intercept Rating Curve and AVGWLF and MDE Tool Simulated Annual Phosphorus Loads

Table 31 and Table 32 show the t-test statistics for the MDE tool and AVGWLF simulated total nitrogen loads, respectively. Table 31 shows that the MDE tool simulated mean annual loads are statistically the same as the EPA observed loads for the Fairview and Bridgeport watersheds. Table 32 shows that the AVGWLF simulated mean annual loads are statistically the same as the observed for the Crumpton and Snowhill watersheds.

Table 31: Hypothesis Test Statistics for EPA Observed and MDE Tool Simulated Total Phosphorus Loads Calculated using a Zero Intercept Rating Curve

Watershed	S_{obs} (kg/yr)	S_{MDE} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	348.8	0	1.216E+05	-12.11	different
Crumpton	158.8	0	2.523E+04	-6.730	different
Fairview	9.320E+04	0	8.686E+09	1.329	same
Bridgeport	3.061E+04	0	9.368E+08	0.3935	same
Snowhill	1050	0	1.102E+06	-9.321	different

Table 32: Hypothesis Test Statistics for EPA Observed and AVGWLF Simulated Total Phosphorus Loads Calculated using a Zero Intercept Rating Curve

Watershed	S_{obs} (kg/yr)	S_{AVGWLF} (kg/yr)	S_p^2 (kg/yr)^2	t	Decision
Kennedy	348.8	346.6	1.209E+05	2.173	different
Crumpton	158.8	840.3	3.657E+05	-0.9368	same
Fairview	9.320E+04	4.051E+04	5.164E+09	4.296	different
Bridgeport	3.061E+04	1.692E+04	6.116E+08	3.316	different
Snowhill	1050	863.9	9.240E+05	-0.6575	same

Figure 31 shows the MDE tool and AVGWLF simulated mean annual total phosphorus load per unit area versus the EPA observed mean annual total phosphorus load per unit area calculated using a rating curve with a zero intercept. The *N-S* coefficient for the MDE tool is 0.43 and the *N-S* coefficient for the AVGWLF model is 0.76. Both values are positive; therefore both models provide better estimates than using an average observed load. The *N-S* coefficient for the AVGWLF model is higher indicating that the AVGWLF model is more accurate.

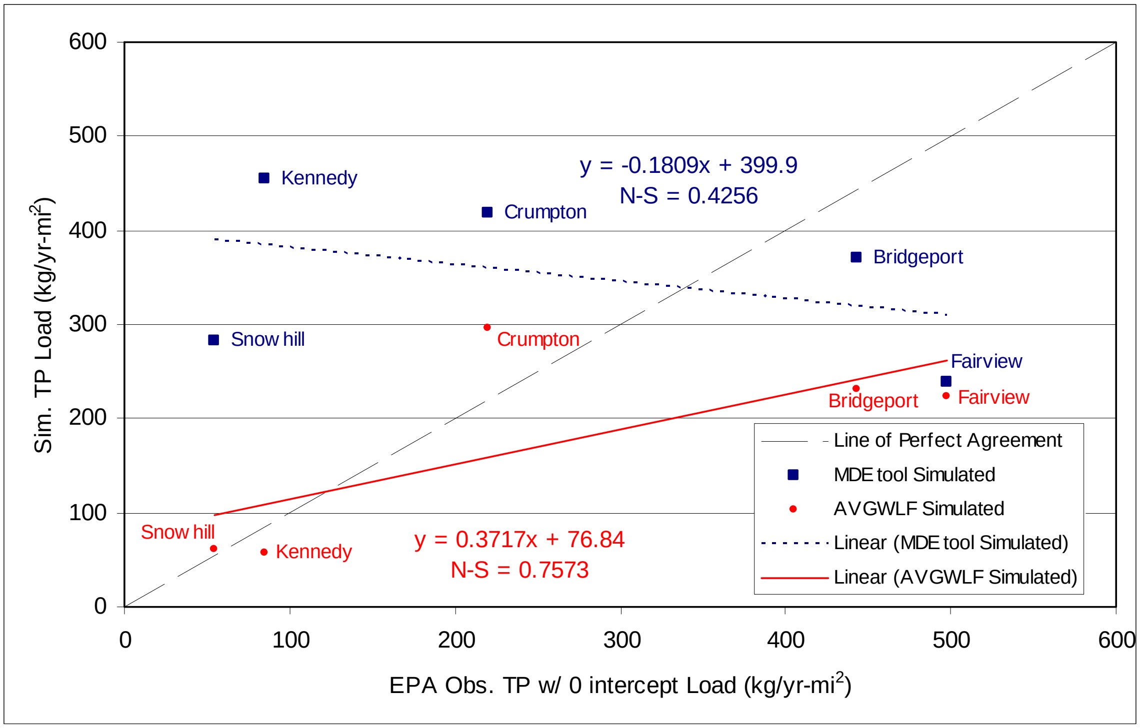

Figure 31: AVGWLF and MDE Tool Simulated Total Phosphorus Loads versus EPA Observed Phosphorus Loads Calculated using a Zero Intercept Rating Curve

4.2.2.4 Sediment

Figure 32 shows the EPA observed annual suspended sediment loads, the AVGWLF simulated annual total sediment loads, and the MDE tool mean annual total sediment load for the period of 4/1/1990 to 3/31/2000. The figure shows that the AVGWLF model under predicts annual sediment loads for the Kennedy and Crumpton watersheds and over predicts for the Fairview, Bridgeport, and Snowhill watersheds. The figure also shows that the MDE tool over predicts annual sediment loads for the Kennedy, Bridgeport, and Snowhill watersheds, under predicts for the Fairview watershed, and provides a moderate estimate for the Crumpton watershed. According to Figure 32 the MDE tool is a better model because it simulates more accurate annual sediment loads for three (Kennedy, Fairview, and Bridgeport) out of the five watersheds.

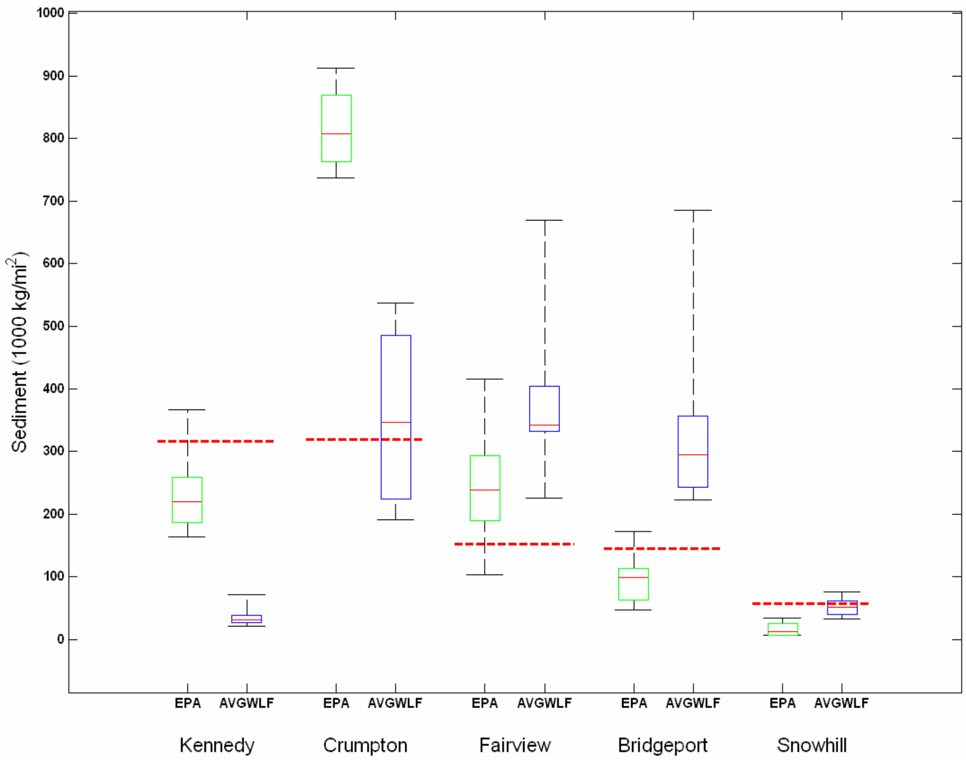

Figure 32: Box and Whisker Plot of EPA Observed Annual Sediment Loads and AVGWLF and MDE Tool Simulated Annual Sediment Loads

Table 33 and Table 34 show the t-test statistics for the MDE tool and AVGWLF simulated total sediment loads, respectively. Table 33 shows that the MDE tool mean annual simulated sediment loads are statistically the same as the observed loads for the Kennedy, Fairview, and Bridgeport watersheds. Table 34 shows that the AVGWLF mean annual sediment loads are statistically different than the observed for all five watersheds. Based on Table 33 and Table 34, I would conclude that the MDE tool is a more accurate model because the MDE simulated sediment loads are statistically the same for three out of the five watersheds.

Table 33: Hypothesis Test Statistics for EPA Observed and MDE tool Simulated Total Sediment Loads

Watershed	S_{obs} (Mg/yr)	S_{MDE} (Mg/yr)	S_p^2 (Mg/yr)^2	t	Decision
Kennedy	727.2	0	5.288E+05	-1.324	same
Crumpton	482.0	0	2.323E+05	6.660	different
Fairview	4.616E+04	0	2.131E+09	0.9980	same
Bridgeport	7709	0	5.942E+07	-0.9800	same
Snowhill	471.8	0	2.226E+05	-3.635	different

Table 34: Hypothesis Test Statistics for EPA Observed and AVGWLF tool Simulated Total Sediment Loads

Watershed	S_{obs} (Mg/yr)	S_{AVGWLF} (Mg/yr)	S_p^2 (Mg/yr)^2	t	Decision
Kennedy	727	195.4	2.835E+05	9.719	different
Crumpton	482	1033	6.500E+05	5.797	different
Fairview	4.616E+04	5.932E+04	2.825E+09	-2.726	different
Bridgeport	7709	2.347E+04	3.052E+08	-5.122	different
Snowhill	472	640.0	3.161E+05	-6.425	different

Figure 33 shows the MDE tool and AVGWLF simulated mean annual total sediment load per unit area versus the EPA observed mean annual total sediment load per unit area. The *N-S* coefficient for the MDE tool is 0.64 and the *N-S* coefficient for the AVGWLF model is 0.57. Both values are positive; therefore both models provide better estimates than using an average observed load. The *N-S* coefficient for the MDE tool is higher indicating that the MDE tool is more accurate.

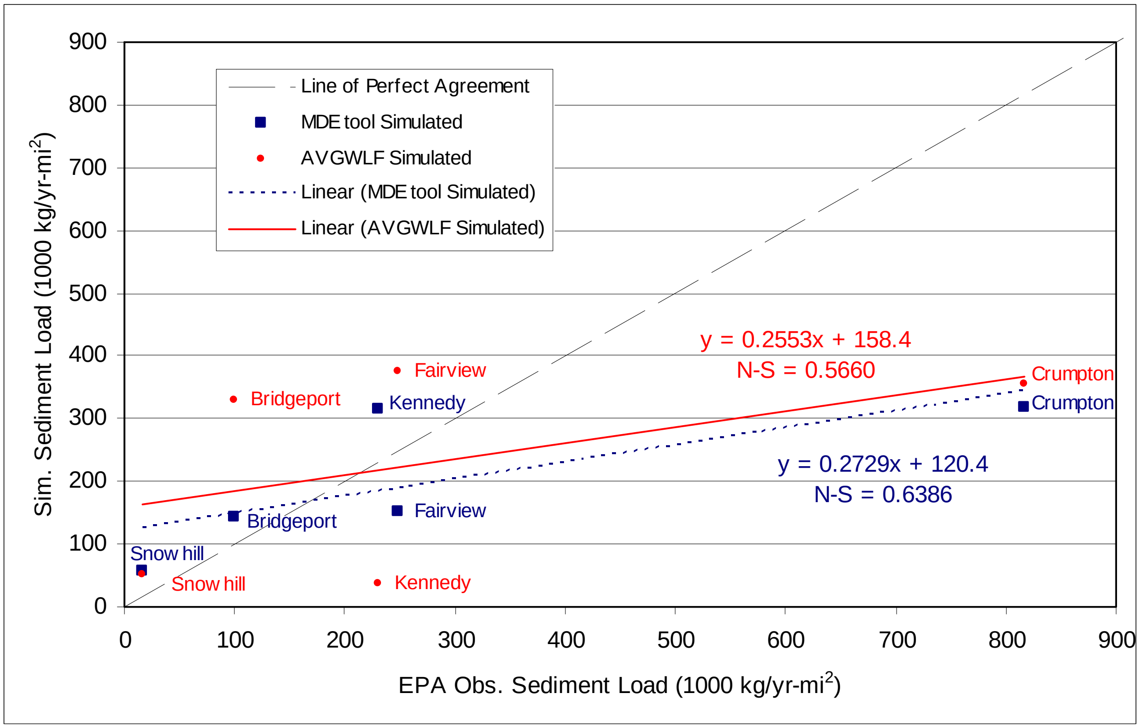

Figure 33: AVGWLF and MDE Tool Simulated Total Sediment Loads versus EPA Observed Loads

4.3 Evaluation of Monthly Loads

The AVGWLF water quality model simulates pollutant loads on a finer time scale than the MDE tool. AVGWLF simulates daily pollutant loads and reports monthly loads to the user. Accurately simulated monthly loads would be more valuable than simulated mean annual loads from a planning perspective because mean annual loads mask seasonal fluctuations. Pollutant loads could be above regulatory limits at certain times of the year, but if the yearly average is below the regulatory limit then planners do not know that remedial action is needed.

N-S coefficients were calculated to assess the goodness-of-fit of the AVGWLF model simulations on a monthly basis. A qualitative comparison was also made by tabulating the number of simulated monthly loads between the 1st and 3rd quartiles of the monthly observed loads.

4.3.1 Comparison of AVGWLF to Concentration-derived Observed

Figure 34 shows the AVGWLF simulated monthly total nitrogen loads versus the concentration-derived observed loads for the Snowhill watershed. Figure 35 shows the same plot for monthly total phosphorus loads and Figure 36 shows the same plot for total sediment loads. Goodness-of-fit statistics for similar figures for the other 4 study watersheds can be found in Appendix C.

Figure 34 through Figure 36 show that AVGWLF under predicts total nitrogen and over predicts total phosphorus and total sediment. The figures also show the *N-S* coefficients, which indicate that AVGWLF predicts total nitrogen well, but predicts total phosphorus and total sediment poorly. In fact, AVGWLF provides such poor estimates for total phosphorus and total sediment that using the average pollutant loads would be a better estimate of monthly loads.

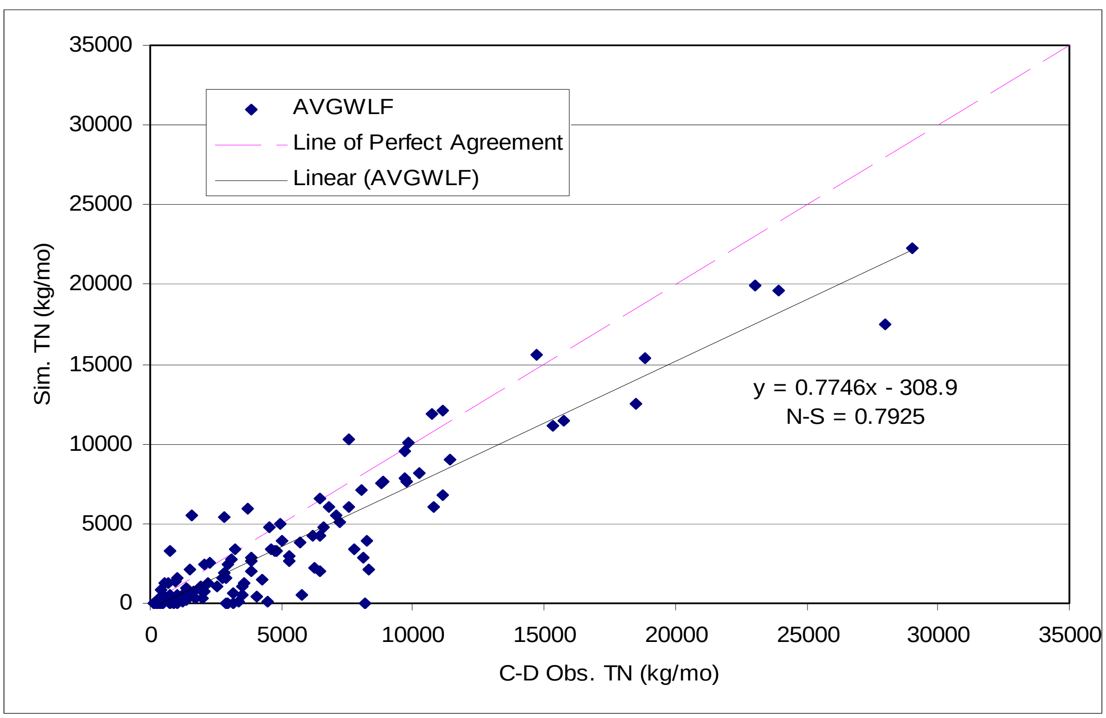

Figure 34: AVGWLF Simulated Monthly Total Nitrogen (TN) Loads versus Concentration-derived Observed Monthly Loads for the Snowhill Watershed

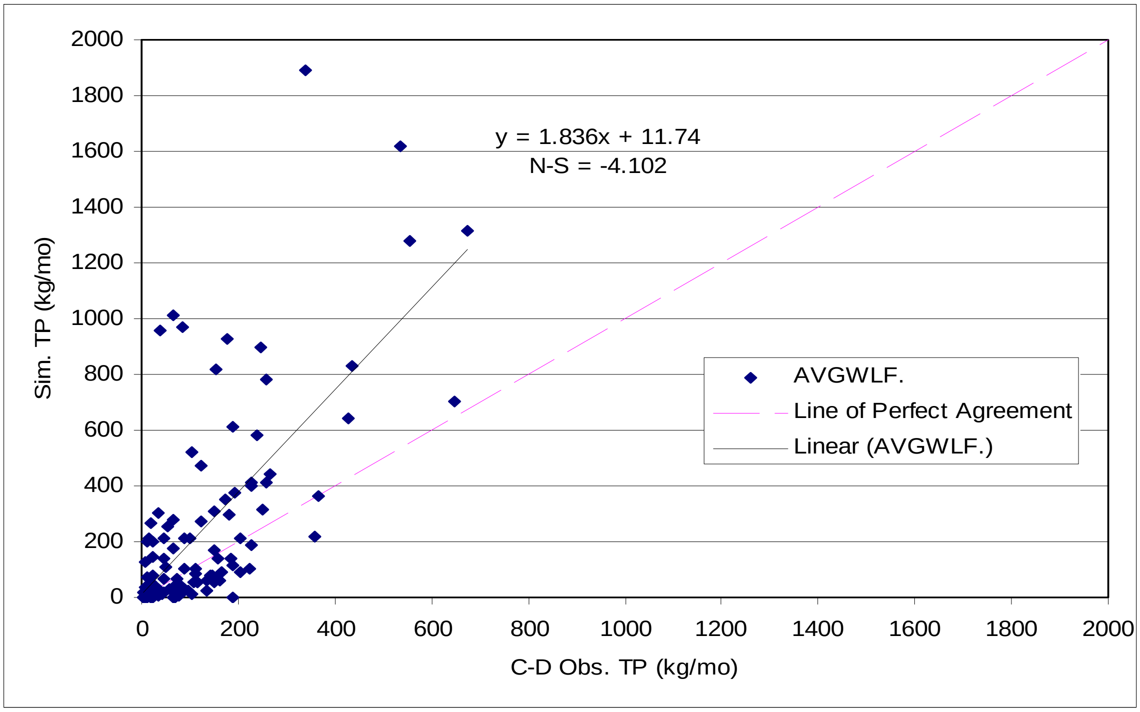

Figure 35: AVGWLF Simulated Monthly Total Phosphorus (TP) Loads versus Concentration-derived Observed Monthly Loads for the Snowhill Watershed

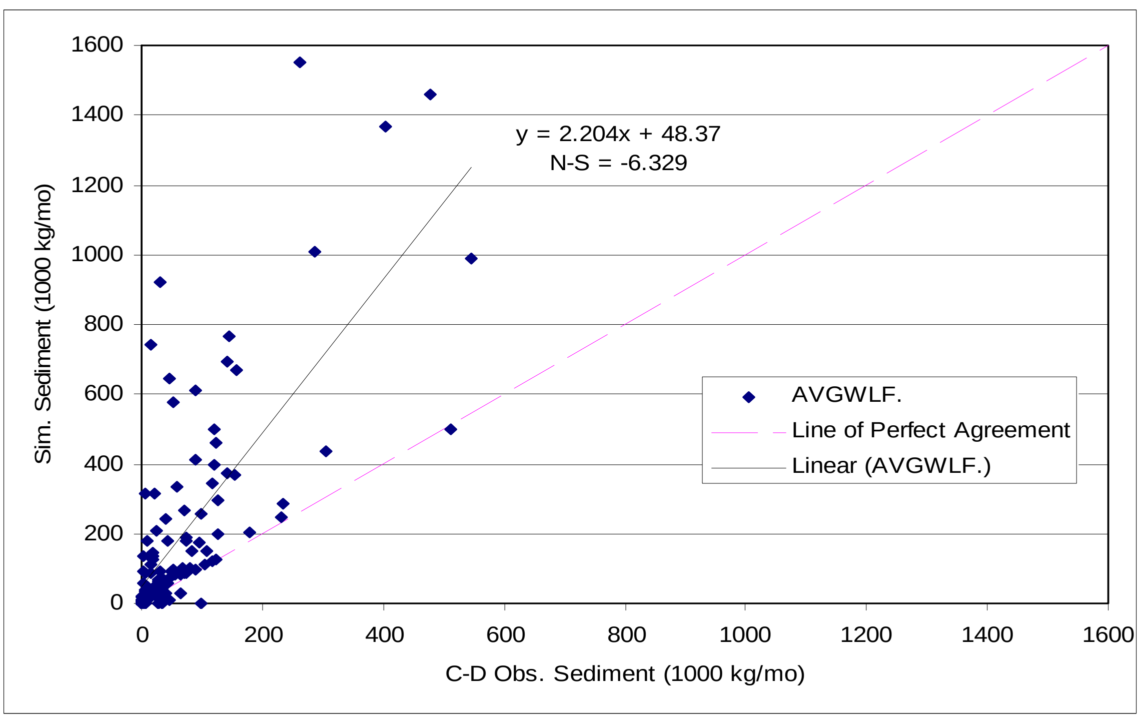

Figure 36: AVGWLF Simulated Monthly Sediment Loads versus Concentration-derived Observed Monthly Loads for the Snowhill Watershed

Table 35 summarizes the monthly and yearly *N-S* coefficients for the five study watersheds and the three pollutants: total nitrogen (N), total phosphorus (P), and total sediment (S). AVGWLF predicts total nitrogen moderately well for every watershed except Crumpton. AVGWLF gives poor estimates for total phosphorus and total sediment because the *N-S* coefficients are below zero (with the exception of annual phosphorus loads for the Bridgeport watershed, which AVGWLF predicts moderately). The estimates are so poor that for every watershed except Kennedy a better estimate would be achieved by using the pollutant load average. Table 35 also shows that with few exceptions AVGWLF simulates better monthly pollutant loads than yearly loads.

Table 35: Nash-Sutcliffe (*N-S*) Coefficients for Month and Year for Concentration-derived Observed

Watershed	N-mo	N-yr	P-mo	P-yr	S-mo	S-yr
Kennedy	0.2689	-2.740	0.1379	0.05813	0.1034	-0.5134
Crumpton	-2.019	-46.63	-190.9	-756.3	-58.56	-185.9
Fairview	0.1617	-2.253	-0.08177	-0.1352	-3.674	-6.989
Bridgeport	0.5719	0.8409	-0.05897	0.5924	-19.62	-34.10
Snowhill	0.7925	0.3435	-4.102	-5.057	-6.329	-12.94
median	0.2689	-2.253	-0.08177	-0.1352	-6.329	-12.94

Table 36 below shows the number of months (out of a total of 120) that the AVGWLF simulated monthly load fell within the first and third quartiles of the concentration-derived observed load. According to the table AVGWLF does not perform very well because on average only a quarter of the simulated monthly loads fall within the first and third quarters of the concentration-derived load.

Table 36: Number of Months that the AVGWLF Simulated Monthly Load was between the First and Third Quartile of the Concentration-derived Load

Watershed	Nitrogen	Phosphorus	Sediment
Kennedy	21	37	24
Crumpton	4*	4*	3*
Fairview	38	27	38
Bridgeport	54	30	42
Snowhill	52	40	44

* Out of 44 months (instead of 120)

4.3.2 Comparison of AVGWLF to EPA Observed Loads

Figure 37 shows the AVGWLF simulated monthly total nitrogen loads versus the EPA observed loads for the Snowhill watershed. Figure 38 shows the same plot for monthly total phosphorus loads, Figure 39 shows the same plot for monthly total phosphorus loads calculated using a zero intercept rating curve, and Figure 40 shows the same plot for total sediment loads. Figure 37 through Figure 40 show that AVGWLF under predicts monthly nitrogen and over predicts monthly phosphorus and sediment.

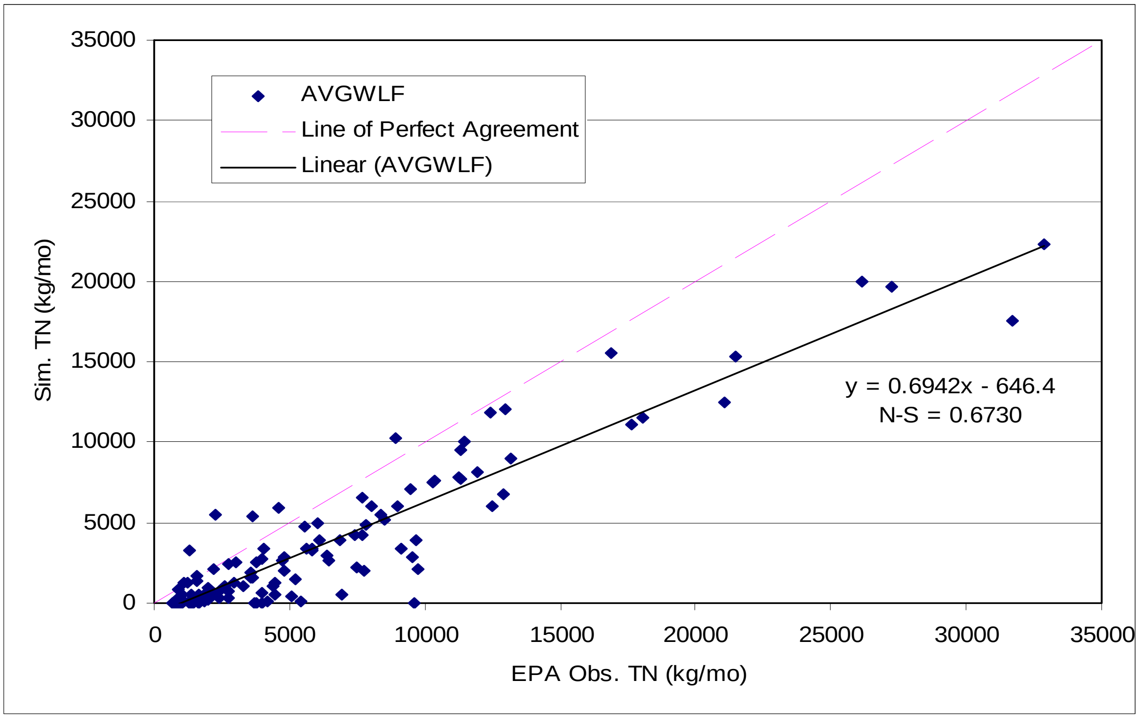

Figure 37: AVGWLF Simulated Monthly Total Nitrogen (TN) Loads versus EPA Observed Monthly Loads for the Snowhill Watershed

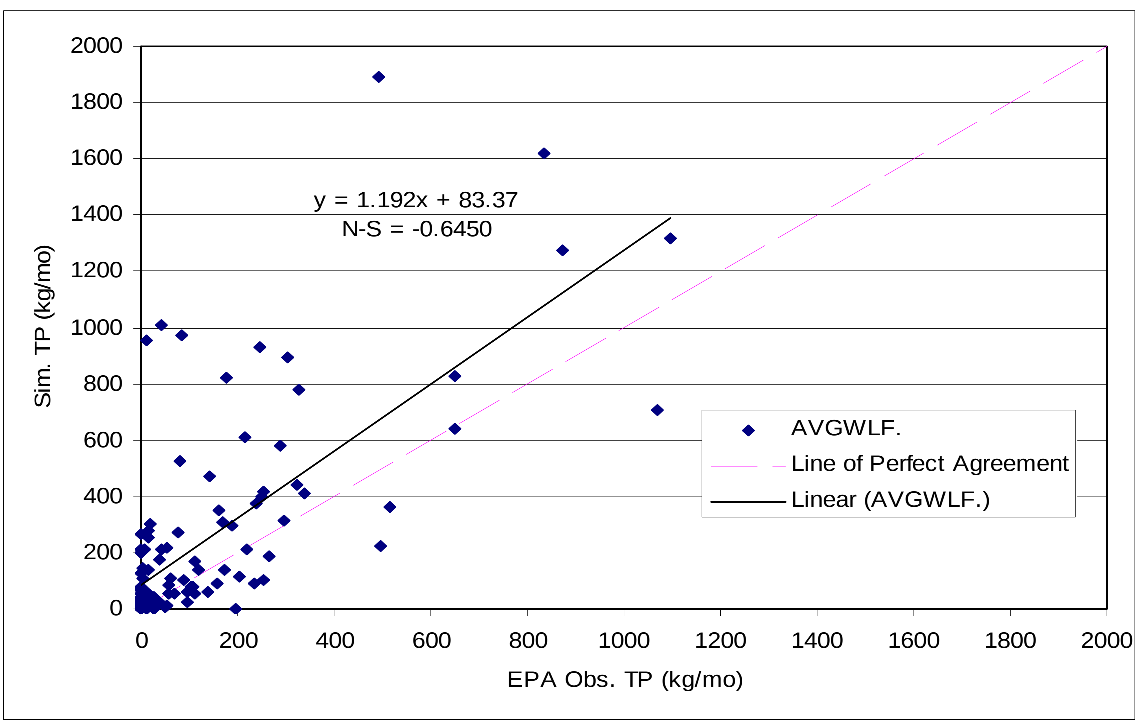

Figure 38: AVGWLF Simulated Monthly Total Phosphorus (TP) Loads versus EPA Observed Monthly Loads for the Snowhill Watershed

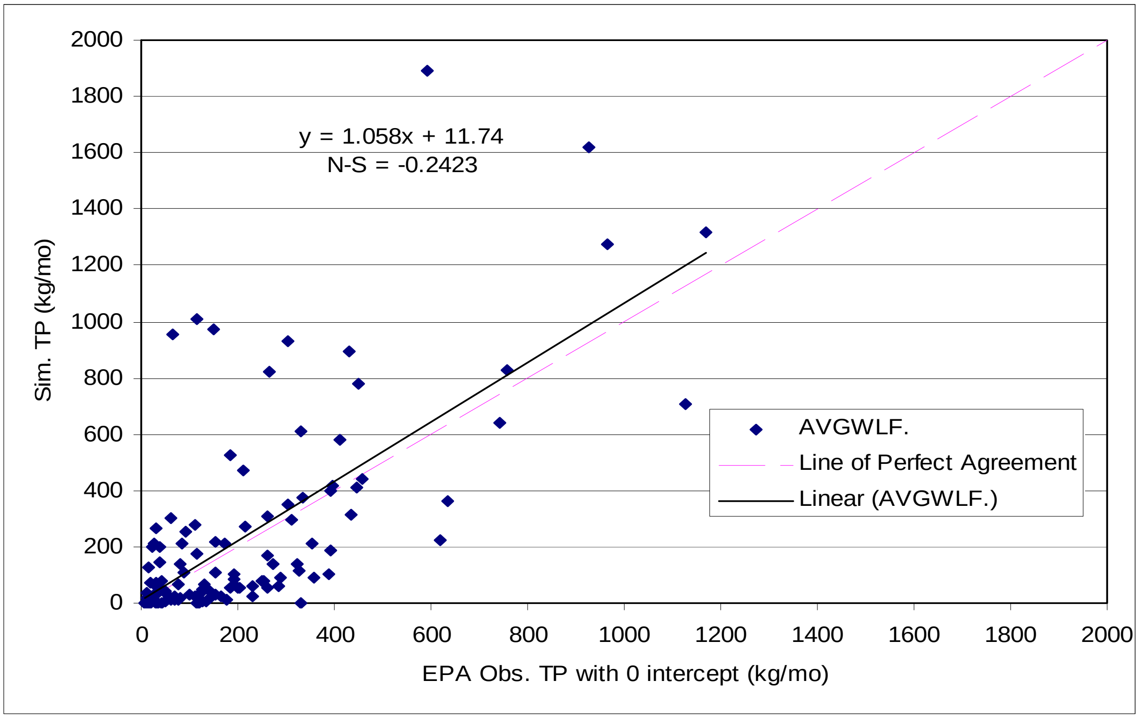

Figure 39: AVGWLF Simulated Monthly Total Phosphorus (TP) Loads versus EPA Observed Monthly Loads Calculated using a Zero Intercept Rating Curve for the Snowhill Watershed

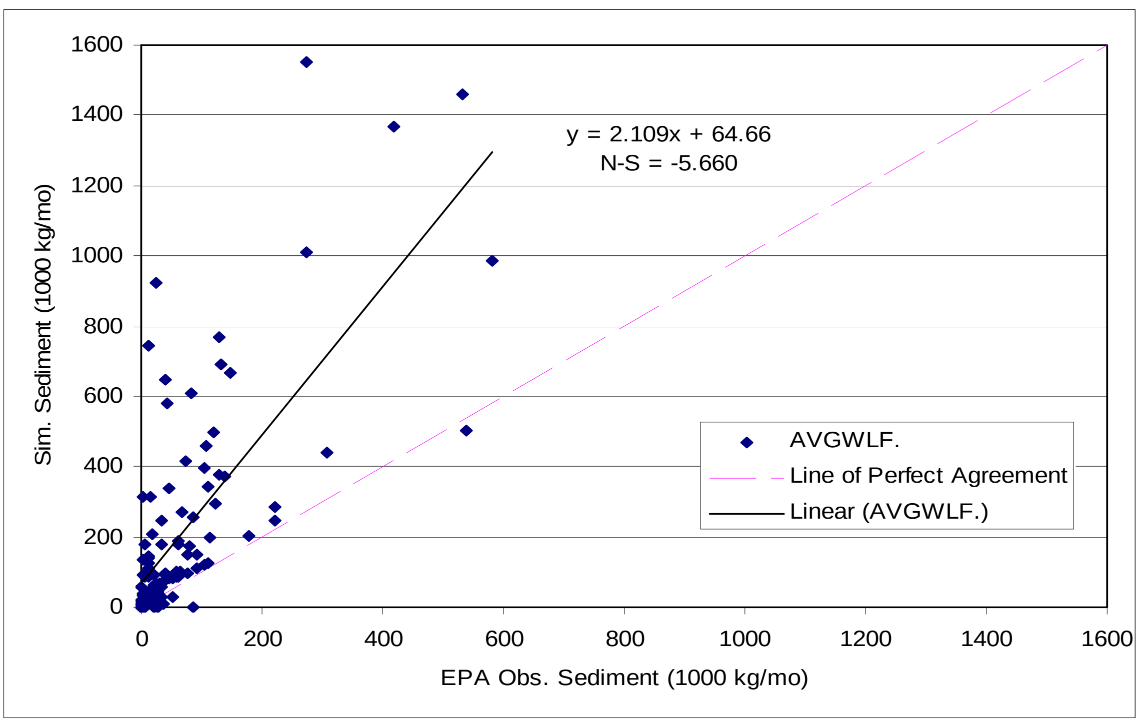

Figure 40: AVGWLF Simulated Monthly Total Sediment Loads versus EPA Observed Monthly Loads for the Snowhill Watershed

Table 37 summarizes the monthly and yearly *N-S* coefficients for the 5 study watersheds and the 3 pollutants, total nitrogen (N), total phosphorus (P), total phosphorus calculated using a zero intercept rating curve (P 0) and total sediment (S). AVGWLF predicts moderately well for total nitrogen, with the exception of the Crumpton and Fairview watersheds, and predicts moderately well for total phosphorus, with the exception of the Crumpton and Snowhill watersheds. AVGWLF gives poor estimates for total sediment. The estimates are so poor that a better estimate would be achieved by using the pollutant load average.

Table 37: Nash-Sutcliffe (*N-S*) Coefficients by Month and Year for EPA Observed

Watershed	N-mo	N-yr	P-mo	P-yr	P 0-mo	P 0-yr	S-mo	S-yr
Kennedy	0.2770	-2.168	0.6035	0.4348	0.5118	-0.1331	-0.8850	-10.91
Crumpton	-18.25	-31.84	-35.68	-118.8	-20.10	-38.63	-6.351	-57.36
Fairview	-0.08105	-4.164	0.1438	0.4697	-0.1430	-1.880	-2.613	-1.515
Bridegport	0.01766	-4.095	0.1983	0.3922	0.1097	-0.9788	-18.60	-34.35
Snowhill	0.6730	-0.4083	-0.6450	-1.464	-0.2423	0.5360	-5.660	-12.88
median	0.01766	-4.095	0.1438	0.3922	-0.1430	-0.9788	-5.660	-12.88

Table 38 shows the number of months (out of a total of 120) that the AVGWLF simulated monthly load fell within the first and third quartiles of the EPA load. According to the table, AVGWLF does not perform very well (with the exception of phosphorus) because on average only a fifth of the simulated monthly loads are within the first and third quarters of the EPA load. AVGWLF predicts phosphorus with moderate accuracy when compared to EPA observed loads calculated using a rating curve without a zero intercept, with approximately 65 percent of months falling between the first and third quartiles of the observed.

Table 38: Number of Months that the AVGWLF Simulated Monthly Load was between the First and Third Quartile of the EPA Observed Load

Watershed	Nitrogen	Phosphorus	Phosphorus with 0 intercept	Sediment
Kennedy	25	84	15	4
Crumpton	0*	22*	0*	2*
Fairview	23	99	26	48
Bridgeport	17	68	26	47
Snowhill	47	76	45	37

* Out of 44 months (instead of 120)

4.4 Interpretation

The main objectives of this study are to determine whether the MDE tool provides accurate estimates and whether the AVGWLF model provides more accurate estimates for total nitrogen, total phosphorus, and total sediment for watersheds in Maryland. The accuracy of these two models were assessed on an annual time scale based on qualitative comparisons, t-tests, and *N-S* coefficients. A summary of the *N-S* coefficients is provided in Table 39. The AVGWLF model was also assessed on a monthly time scale based on *N-S* coefficients and the number of simulated monthly loads that fell between the 1st and 3rd quartiles of the observed loads.

The observed loads were calculated using two different methods: the concentration-derived method and the EPA method. The simulated loads from the two models were then compared to the observed loads calculated by the two methods.

Table 39: Comparison of EPA and Concentration-derived *N-S* Coefficients for Annual Pollutant Loads

Method	Model	Nitrogen	Phosphorus	Phosphorus with 0 intercept	Sediment
	MDE tool	0.8758	-0.7126	-	-0.7172
Concentration-derived	AVGWLF	0.6381	0.5831	-	-3.228
	MDE tool	0.8620	-1.302	0.4256	0.6386
EPA	AVGWLF	0.5632	0.5476	0.7573	0.5660

Overall, both the MDE tool and the AVGWLF model simulated moderately accurate annual nitrogen loads. The MDE tool simulated more accurate annual nitrogen loads as evidenced by the higher *N-S* coefficients in Table 39 and the t-test statistics shown in Table 21 and Table 27. The two methods of calculating observed loads did not have much effect on the conclusion of the accuracy of the two models. The *N-S* coefficients and the t-test statistics changed slightly, but the overall conclusion of the accuracy of the models for predicting annual total nitrogen was not affected.

The AVGWLF model provides moderately accurate estimates of annual total phosphorus, as evidenced by the *N-S* coefficients in Table 39 and the t-test statistics in Table 24, Table 30, and Table 32. According to the *N-S* coefficients in Table 39 the MDE tool poorly predicts annual phosphorus loads unless the simulated loads are being compared to observed loads calculated using a EPA rating curve with a zero intercept. However, according to the t-test statistics shown in Table 23, Table 29, and Table 31, the MDE tool predicts annual phosphorus loads with moderate accuracy, correctly predicting mean annual loads that are statistically the same as the observed loads for two to three watersheds, depending on the method used to calculate the observed loads. I would conclude that the AVGWLF model provides more accurate estimates of annual total phosphorus loads because the *N-S* coefficients indicate that it is a more accurate model than the MDE tool, whereas the t-test statistics indicate that the two models exhibit the same level of accuracy. Again I would conclude that the method of calculating observed loads does not have a large effect on the conclusion of which model is more accurate.

The MDE tool provides moderately accurate estimates of total annual sediment loads, as evidenced by the t-test statistics shown in Table 25 and Table 33. The *N-S* coefficients shown in Table 39 indicate that the MDE tool provides moderately accurate estimates when compared to the EPA observed annual sediment loads, but not when compared to the concentration-derived loads. The AVGWLF model does not provide accurate estimates for annual sediment loads according to the t-test statistics shown in Table 26 and Table 34 and the qualitative comparisons shown in Figure 24 and Figure 32. It can therefore be concluded that the MDE tool provides more accurate estimates for annual total sediment than the AVGWLF model. It can also be concluded that the method used to calculate observed sediment loads does affect the conclusion of the accuracy of the model.

The AVGWLF model was compared to both the concentration-derived and EPA observed loads on a monthly time scale. Table 35 and Table 37 show the monthly *N-S* coefficients by watershed and pollutant. Table 36 and Table 38 shows the number of simulated months that fell between the 1st and 3rd quartiles of the observed. From these tables it can be concluded that AVGWLF poorly predicts monthly pollutant loads. There are a few exceptions, such as, nitrogen for the Snowhill watershed and phosphorus for the Kennedy watershed, where AVGWLF provides moderate to good estimates. However, overall AVGWLF does not provide accurate estimates on a monthly time scale.

4.5 Sources of Errors and Uncertainty

As is the nature of water-quality modeling there are many sources of error and uncertainty in this study. This section will address some of these sources.

4.5.1 Observed Loads

Pollutant concentration measurements for the study watersheds were downloaded from the USGS NAWQA website. These measurements were then used to develop rating curves so that daily discharge values could be used to predict pollutant loads on days when actual measurements were not available.

Table 40: Relative Standard Error of Concentration-derived Rating Curves

Watershed	Nitrogen	Phosphorus	Sediment
	Se/Sy	*Se/Sy*	*Se/Sy*
Kennedy	0.3081	0.4272	0.9288
Crumpton	3.880	0.1982	0.6865
Fairview	0.3270	0.6227	0.9211
Bridgeport	12744	0.5924	1.016
Snowhill	972.7	500.4	0.9158

Table 41: Relative Standard Error of EPA Rating Curves

Watershed	Nitrogen	Phosphorus	Sediment
	S_e/S_y	S_e/S_y	S_e/S_y
Kennedy	0.7950	0.4024	0.6354
Crumpton	0.0826	0.1588	0.2894
Fairview	0.3202	0.6522	0.7604
Bridgeport	0.4025	0.5459	0.2805
Snowhill	0.2416	0.5313	0.4725

The development of rating curves introduces three sources of error. First there is the error in the actual measurements, due to lack of accuracy in the instrumentation. Second there is the error introduced by the use of a rating curve because it does not explain all of the variation. The percent unexplained variation or the relative standard error for the concentration-derived rating curves is shown in Table 40 and the relative standard error for the EPA rating curves is shown in Table 41. The third source of error is introduced by using average daily discharges and the rating curves developed using instantaneous discharges to calculate concentration-derived observed loads. Equation 4-4 shows the average daily load calculated using instantaneous discharges and equation 4-5 shows the average daily load calculated using average discharges. Equation 4-4 represents the true average daily load, while equation 4-5 represents the average daily load calculated using the concentration-derived rating curves.

$$\text{Average Daily Load} = \frac{\int_0^{24} Q(t)C(t)dt}{24} \tag{4-4}$$

where Q is the discharge, C is the concentration and t is the time measured in hours.

$$\text{Average Daily Load} = \overline{Q} * \overline{C} \tag{4-5}$$

where $\overline{Q}$ and $\overline{C}$ represent the average discharge and concentration, respectively.

If the assumptions shown in equation 4-6 are made then equation 4-4 becomes equation 4-7, which can be simplified as shown in equation 4-8. Note that the cross terms $\overline{Q}\Delta C(t)$ and $\overline{C}\Delta Q(t)$ are equal to zero because the sum of either $\Delta C(t)$ or $\Delta Q(t)$ is zero over the period of one day. Therefore the sum of $\overline{Q}\Delta C(t)$ and $\overline{C}\Delta Q(t)$ are also zero because $\overline{Q}$ and $\overline{C}$ are constants.

$$\begin{aligned} Q(t) &= \overline{Q} + \Delta Q(t) \\ C(t) &= \overline{C} + \Delta C(t) \end{aligned} \tag{4-6}$$

where $\Delta Q(t)$ and $\Delta C(t)$ represent the difference between the instantaneous and daily average discharges and concentrations, respectively.

$$\text{Average Daily Load} = \frac{\int_0^{24} \left(\overline{Q} + \Delta Q(t)\right)\left(\overline{C} + \Delta C(t)\right)dt}{24} \tag{4-7}$$

$$\text{Average Daily Load} = \frac{\int_0^{24} \left(\overline{Q} * \overline{C} + \Delta Q(t) \Delta C(t) \right) dt}{24} \tag{4-8}$$

The difference between the average daily load calculated using instantaneous discharges (equation 4-8) and average discharges (equation 4-5) depends on the magnitude and sign of the term $\Delta Q(t)\Delta C(t)$. If discharge and concentration are positively correlated then the term $\Delta Q(t)\Delta C(t)$ will be positive and the average load calculated using instantaneous discharges will be larger. If discharge and concentration are negatively correlated then the term $\Delta Q(t)\Delta C(t)$ will be negative and the average load calculated using instantaneous discharges will be smaller. The difference between the concentration-derived loads calculated using instantaneous discharges and using average discharges will tend to be greater for smaller watersheds. This is because there is more variation in daily discharge for a small watershed and therefore the $\Delta Q(t)$ term will be larger.

The use of instantaneous discharges to calculate concentration-derived observed loads would result in different conclusions about the accuracy of the two models. For instance, the AVGWLF model under predicts annual nitrogen and the MDE tool generally over predicts annual nitrogen when compared to the concentration-derived observed. Since the discharge and nitrogen concentration are positively correlated for the watersheds in this study the use of instantaneous discharges would increase the concentration-derived observed loads. Therefore the accuracy of the MDE tool would increase, while the accuracy of the AVGWLF model would decrease.

It should also be noted that water quality samples are rarely taken during extremely high flows. Therefore it would be expected that a rating curve would have much greater uncertainty when predicting concentrations during high flows. However, in this study, most of the rating curves were developed for a range of flows that included both high and low flows. Table 42 shows the range, mean, and standard deviation (stdev) of the discharges used in developing the rating curves and Table 43 shows the range, mean, and standard deviation of the discharges that are recorded during the simulation period. Table 42 and Table 43 show that the range of recorded discharges for the simulation period falls within the range of discharges for the rating curves for every watershed except Snowhill and Kennedy. The rating curves for the Snowhill watershed predict pollutant loads during high flows well because the highest recorded flow during the simulation period is 2130 ft^3/s, which is in the same order of magnitude as the highest flow represented by the rating curve (1300 ft^3/s). However, the rating curves developed for the Kennedy watershed do not predict pollutant loads during high flows well because the highest recorded flow (3600 ft^3/s) is much higher than the highest flow represented by the rating curves (353 ft^3/s).

Table 42: Range, Mean, and Standard Deviation of Daily Discharges in ft^3/s for Rating Curves

Water-shed	Nitrogen			Phosphorus			Sediment		
	range	mean	stdev	range	mean	stdev	range	mean	stdev
Kennedy	1.4-353	18.55	24.23	1.4 - 353	29.84	54.66	1.4-353	31.23	56.16
Crumpton	2.5-2940	68.39	338.2	0.53-2940	67.71	336.7	2.5-2940	61.84	354.8
Fairview	43-16700	2767	2939	43-16700	2780	2955	43-14500	2666	2654
Bridgeport	7.96-14200	3088	3873	7.96-14200	3099	3875	13.74-9190	625.4	1656
Snowhill	1.5-1300	132.3	222.0	1.5-1300	132.3	222.0	1.5-1300	132.4	212.7

Table 43: Range, Mean, and Standard Deviation of Daily Discharges in ft^3/s for the Simulation Period 4/1/1990 to 3/31/2000

Watershed	All Pollutants		
	range	mean	stdev
Kennedy	1.3-3600	12.81	63.59
Crumpton	1.7-722	8.040	20.39
Fairview	66-14600	711.1	1024
Bridgeport	0.43-13700	249.3	636.2
Snowhill	0.83-2130	53.53	103.7

4.5.2 Modeling Assumptions

The AVGWLF model uses up to 17 GIS layers to derive inputs to the GWLF model. Assumptions were made in the development of these layers that could lead to errors in the AVGWLF pollutant load simulations. For example, in order to develop the water extraction layer, which identifies the location, amount, and seasonality of water withdrawals, I had to assume the seasonality. The seasonality of withdrawals refers to the time of year during which the withdrawals are made. There are four categories for seasonality; drinking water or commercial water withdrawals represent year-round use, agricultural withdrawals for irrigation are only made from May-September, withdrawals for snow-making are made from November-March, and withdrawals for golf course irrigation are made from April-October. Since I did not know the seasonality of the water withdrawals, I overlaid a Maryland Department of Planning (MDP) land use layer and assumed the extraction seasonality based on the land use category that each extraction point was located in. Table 44 shows the seasonality that I assumed based on which land use category the extraction point was located in. The snow-making seasonality was not used because there are no ski resorts in the study watersheds. The seasonality of water extractions affects the simulated discharge and since water-quality is directly related to discharge it also affects the simulated pollutant loads.

Table 44: Assumed Seasonality of Water Extractions Based on Land Use

Land Use Code	Land Use Category	Assumed Seasonality
21	Cropland	Agricultural May-September
22	Pasture	
23	Orchards	
25	Row Crops	
191	Large Lot Agriculture	
241	Feeding Operations	
242	Agricultural Buildings	
16	Institutional	Golf Course April-October
all others		year-round

AVGWLF requires a particular land use coding scheme and therefore I had to make assumptions in order to convert the MDP land use codes to AVGWLF land use codes. Table 49 in Appendix A shows the MDP land use codes and categories and the AVGWLF land use codes and categories that I converted them to. The assumptions that introduce the most error are the conversion of residential medium density to low density development and the conversion of industrial and commercial to high density development. The conversion of institutional to low-density development could also result in error because sometimes institutional land has a large percent impervious area and would be better modeled as high-density development.

Error can also be introduced by the used of default values. Default values were used for the cover and management factor (C) and the support management factor (P). These values were used in the USLE equation to calculate soil erosion. Default values were also used for the cold and warm weather rainfall erosivity coefficients and the groundwater recession coefficient.

Error is also introduced because the tile drains, unpaved roads, groundwater nitrogen, and soil phosphorus data layers were not developed. However, the development of these layers was not thought to greatly improve the accuracy of the AVGWLF simulations. A more detailed explanation of these layers and why they will not greatly improve the accuracy of AVGWLF will be presented in Appendix A.

4.5.3 Data Source Limits

There are several instances where coarse-level or missing data could have caused inaccuracies in the AVGWLF simulations. The soil and animal density layers had very spatially coarse data. This could lead to inaccuracies, especially in small watersheds, because the information contained within these layer is an average over a large area and the actual values may be different. For instance, the animal density layer contains information on the number of animals in a county and I assumed that the animals are evenly spaced throughout that county. In reality the animals are likely contained within a few small areas on farms. A watershed within that county may or may not contain those farms and therefore the animal density is either overestimated if there are, in reality, no farms in the watershed or underestimated if all the animals are actually in the watershed.

Inaccuracies in the AVGWLF simulated discharge and therefore the simulated pollutant loads could also have been caused by missing data in the weather files. Table 45 in Appendix A shows the number of months and days that weather information was missing (the total number of missing days can be obtained by summing the number of missing months and missing days). Weather stations that were missing more than 5 percent or 184 total days during the 10 year simulation period were excluded. Missing data in the other weather stations were estimated and therefore do not represent actual data.

The weather database developed for this study includes 13 weather stations. All of the acceptable weather stations that were located near the study watersheds were used. However, there are not many acceptable stations and therefore only one watershed had a weather station located within it and the nearest weather station to many of the watersheds was several miles away. The lack of acceptable weather data may lead to inaccuracies in the AVGWLF simulation of water quantity, especially in the summer months which are prone to small thunderstorms, because the rain gauges may record rain that did not fall on the watershed or not record rain that did.

Chapter 5: Conclusions and Recommendations

The Clean Water Act (CWA) of 1972 requires that all water bodies in the U.S. meet specific standards that ensure that the water is “fishable and swimmable”. Section 303(d) of the CWA requires that Total Maximum Daily Loads (TMDLs) be developed for all water bodies that do not meet quality standards and are therefore considered impaired (USEPA, 1999).

The Maryland Department of the Environment (MDE) is responsible for developing TMDLs for impaired water bodies in the state of Maryland. Water quality models are important tools that the MDE uses in the source assessment, linkage analysis, and allocation steps of TMDL development (MDE, 2007).

The MDE uses various water quality models, but has recently decided to use the Chesapeake Bay Program Watershed model (WSM) for consistency with the Chesapeake Bay Program (CBP). However, since the WSM model, which is based on the Hydrologic Simulation Program – FORTRAN (HSPF) model, is complex and requires extensive training, time, and data, the MDE developed a tool that interpolates the output from the WSM. The MDE tool uses WSM phase 4.2 loading coefficients and 2002 land cover to estimate mean annual total nitrogen, phosphorus, and sediment loads. Since the MDE tool requires minimal training, data and time to run, it can be used for quick analyses to see the effects of changing land use on nutrient and sediment loads (Moglen, pers. comm., 2007).

The MDE tool is based on models that have been extensively tested and are considered accurate. However, the MDE tool has never been tested to determine its accuracy. It is very important that the model used by the MDE to develop TMDLs be accurate because otherwise the TMDL will not accurately represent the needed reductions in pollutant loads.

The main objectives of this study were to 1) determine if the total nitrogen, phosphorus, and sediment loads simulated by the MDE tool are accurate and 2) determine if the total nitrogen, phosphorus, and sediment loads simulated by ArcView Generalized Watershed Loading Function (AVGWLF) are more accurate than those simulated by the MDE tool. The accuracy of the two models was assessed on an annual time scale based on qualitative comparisons, t-tests, and Nash-Sutcliffe (*N-S*) coefficients. Since the AVGWLF model simulates pollutants on a finer time scale, it was also assessed on a monthly time scale based on *N-S* coefficients and the number of simulated monthly loads that are between the 1st and 3rd quartiles of the observed monthly loads.

The accuracy of the two models was determined by comparing the simulated pollutant loads to observed pollutant loads. However, the available water quality data in Maryland is sporadic and therefore rating curves were developed so that daily discharge values could be used to estimate daily pollutant loads. Two sets of rating curves were developed: one between pollutant load and discharge and the other between pollutant concentration and discharge. Developing a regression-based rating curve between pollutant load and discharge is statistically incorrect because it represents the relationship between discharge times a variable (i.e. concentration) and discharge. This results in inflated goodness-of-fit statistics and possibly incorrect regression coefficients (McCuen and Surbeck, 2007). However, this method is endorsed by the U. S. Environmental Protection Agency (USEPA) and has been used in various studies (USEPA, 1999; Evans, 2002; Mattikalli et al. 1996). Therefore the observed loads were calculated using two different methods. The EPA method used rating curves developed between pollutant load and discharge, and the concentration-derived method used rating curves developed between pollutant concentration and discharge. The concentration-derived method is the statistically correct method.

The EPA method for calculating observed total nitrogen loads resulted in higher observed loads for all five watersheds; however, the two methods produce similar annual loads for the Kennedy, Fairview, and Snowhill watersheds. The EPA method for calculating total phosphorus using a rating curve with a zero intercept produced the highest observed total phosphorus loads for all five watersheds. The EPA method using a rating curve without a zero intercept and the concentration-derived method produced total phosphorus loads with similar magnitudes for all five watersheds. The two methods of calculating total sediment produce observed loads that are similar in value for every watershed except the Crumpton watershed. The EPA observed loads for the Crumpton watershed are almost 20 times the concentration-derived loads.

Both the MDE tool and the AVGWLF model simulate moderately accurate annual total nitrogen loads, as evidenced by *N-S* coefficients and t-tests. The *N-S* coefficients for the MDE tool (0.88 for concentration-derived observed and 0.86 for EPA observed) are higher than the *N-S* coefficients for the AVGWLF model (0.64 for concentration-derived observed and 0.56 for EPA observed) indicating that the MDE tool is better at predicting annual total nitrogen than the AVGWLF model.

AVGWLF predicts annual total phosphorus with moderate accuracy based on *N-S* coefficients and t-tests. The MDE tool provides moderately accurate estimates of annual total phosphorus according to t-tests, but according to *N-S* coefficients, the MDE tool provides poor estimates of annual total phosphorus, unless the observed annual phosphorus loads are calculated using a EPA rating curve with a zero intercept, in which case, the MDE tool provides moderately accurate estimates.

The MDE tool simulates annual total sediment loads with moderate accuracy according to t-tests and the AVGWLF model simulates annual total sediment loads with poor accuracy according to t-tests. The conclusion drawn for both models from *N-S* coefficients depends on the method used to calculate observed loads. Both models provide moderately accurate estimates of annual sediment loads when the observed sediment loads are calculated using the EPA method, and poor estimates when the observed loads are calculated using the concentration-derived method.

The AVGWLF model was compared to both the concentration-derived and EPA observed loads on a monthly time scale. Based on *N-S* coefficients it can be concluded that the AVGWLF model provides poor estimates of pollutant loads on a monthly time scale. There were a few exceptions, such as, nitrogen for the Snowhill watershed and phosphorus for the Kennedy watershed, where AVGWLF provides moderate to good monthly estimates. AVGWLF was also assessed based on the number of months that simulated loads fell between the 1^{st} and 3^{rd} quartiles of the observed loads. Based on this assessment, AVGWLF provides poor estimates, with only a quarter of the simulated loads falling between the 1^{st} and 3^{rd} quartiles of the concentration-derived loads and a fifth falling between the 1^{st} and 3^{rd} quartiles of the EPA observed loads. The exception is when observed phosphorus is calculated using an EPA rating curve with a negative intercept (i.e. not a zero intercept). In this case the AVGWLF model predicts phosphorus loads with moderate accuracy, with 65 percent of the simulated loads falling between the 1st and 3rd quartiles of the observed loads.

Overall, I think that the MDE tool is a good model for quick analyses to obtain an estimate of how pollutant loads will change as a result of changing land use. However, the MDE tool provides only moderately accurate estimates of total nitrogen and sediment, at best, and poor estimates for total phosphorus. Therefore I recommend that a more sophisticated model that has been analyzed and shown to be accurate be used in the development of TMDLs.

The AVGWLF model, which is a more sophisticated model than the MDE tool, did not perform much better than the MDE tool. It provides moderately accurate estimates of annual total nitrogen and phosphorus and poor to moderately accurate estimates of annual total sediment, depending on the method used to calculate the observed load. The AVGWLF model also simulates monthly loads with poor accuracy. However, while the AVGWLF model does not provide better estimates (except for annual total phosphorus loads), it does require more data and therefore more time to run. It also involves some training because the water quantity has to be calibrated. Therefore, I do not recommend that the AVGWLF model be used instead of the MDE tool, because it has the disadvantages (more training, time, and data) of a more sophisticated model without the advantage of improved results.

A consistent method for making rating curves needs to be developed. The method used to calculate the observed pollutant loads affected the conclusion of whether the models were accurately predicting annual total sediment loads (i.e. the models provided moderately accurate estimates when compared to EPA observed loads and poor estimates when compared to concentration-derived loads). The method used to develop rating curves could also affect the decision of whether to classify a water body as impaired. For instance, the EPA method produced nitrogen loads that were higher than the concentration-derived method for the watersheds in this study. Therefore, if the EPA method is used to estimate observed nitrogen loads, the watershed is more likely to be classified as impaired, whereas if the concentration-derived method is used the watershed is less likely to be classified as impaired.

A method for converting an instantaneous rating curve to an average daily rating curve is also needed. The concentration-derived method for developing rating curves results in the under or over prediction of observed loads because the rating curve is developed using instantaneous data, but is then used with the average daily discharge to estimate an average daily load. Since it is unreasonable to collect instantaneous discharges to be used with the instantaneous rating curve, a method should be developed that corrects for the systematic under or over prediction of the rating curve.

The use of actual in-stream monitoring data as the observed data would be preferred. However, the cost and time needed to collect in-stream monitoring data is generally prohibitive and therefore rating curves will continue to be used to estimate observed data. In this study two simple methods were used to develop rating curves with only one predictor variable (discharge). However, any number of rating curves could be developed using a number of other predictor variables, such as, land use, time of year, precipitation, etc. Since observed loads are used to calibrate water quality models and determine if water bodies are impaired it is important that a consistent, accurate method be developed.

Appendices

Appendix A

AVGWLF uses up to 17 GIS layers to derive the inputs for GWLF. These data layers have been developed for the state of Pennsylvania and can be downloaded from the AVGWLF website www.avgwlf.psu.edu. Also available at the AVGWLF website is a Format Guide (Evans and Corradini, 2006), which describes how to develop the data layers. This guide was used to help create the data layers for Maryland used in this study. This appendix will briefly describe how the data layers used in this study were developed for Maryland (if a watershed extended into Pennsylvania the data layers available on the AVGWLF website were used). For details on how to develop your own data layers for AVGWLF please refer to "A Guide to Creating Software-compatible Data Sets" (Evans and Corradini, 2006).

Basins

The basins layer is a polygon that outlines the boundary of the watershed. GISHydro2000 (Moglen, 2006) was used to delineate the watershed, which I then converted to a watershed boundary layer and imported into AVGWLF.

Streams

The streams layer represents the streams found in the watershed. The streams layer for Maryland was downloaded from the National Hydrography Dataset (NHD) website (USGS, 2005a). It has a resolution of 1:100,000.

Weather Stations

The weather stations data layer is a point file that identifies the locations of the weather stations for which daily precipitation and maximum and minimum temperature data exists. The Maryland weather station data layer was developed from weather data provided by the National Climatic Data Center (NOAA, 2007).

Weather stations near the study watersheds that had daily weather data, specifically precipitation and maximum and minimum temperatures, during the study time period of 1990 until 2000 are shown in Table 45. Table 45 also shows the amount of weather data that was missing from the weather stations. If a large amount of weather data were missing than the weather station was excluded to prevent large errors in the simulation; stations which were excluded are shown in bold. A station was excluded if it did not have data during the entire range from 1990 to 2000, if it did not have all three needed weather elements (precipitation, maximum temperature, and minimum temperature), or if it had more than 200 missing days. Please note that the total number of missing days can be obtained by adding the number of missing months and the number of missing days.

The Pennsylvania weather station 361354 should have been excluded because it has 8 months and 16 days or approximately 256 missing days (8*30 +16 = 256), which is more than the exclusion criteria of 200 missing days. However, station 361354 is located inside the Fairview watershed and therefore I felt it was important to include it. Weather data from nearby stations were used to estimate the weather data for the missing 8 months. Precipitation data were taken from station 368308, which is also within the Fairview watershed, and daily maximum and minimum temperature data from 1990 until

9/30/1998 were taken from station 184030. Daily maximum and minimum temperature data for 10/1/1998 until 10/2000 were taken from station 183980.

Table 45: Weather Stations in Maryland and the amount of missing data. Stations shown in bold were excluded because they were missing too much data.

State	Gage	Range	No. of Elements	No. Months Missing	No. of Days missing
DE	72730	1990-2005	3	0	2
DE	**73570**	**1990-1997**	**3**	**0**	**0**
DE	75320	1990-2005	3	1	9
DE	73595	1990-2005	3	1	9
MD	**180015**	**1990-2005**	**3**	**0**	**2123**
MD	**180335**	**1990-2005**	**3**	**6**	**112**
MD	181750	1990-2005	3	0	2
MD	**182770**	**1990-1995**	**1**	**0**	**0**
MD	182906	1990-2005	3	0	13
MD	**183975**	**1990-1993**	**3**	**0**	**173**
MD	**183980**	**1997-2005**	**3**	**0**	**59**
MD	**184030**	**1990-1998**	**3**	**2**	**237**
MD	**185985**	**1990-2005**	**3**	**19**	**559**
MD	188000	1990-2005	3	2	20
MD	188005	1990-2005	3	0	6
MD	**188207**	**1998-2005**	**3**	**0**	**22**
MD	188380	1990-2005	3	0	0
PA	360656	1990-2005	3	0	1
PA	360763	1990-2005	3	3	62
PA	361354	1990-2005	3	8	16+
PA	362537	1990-2005	3	0	178
PA	**363665**	**1993-2005**	**3**	**1**	**9**
PA	**366955**	**1990-2005**	**1**	**0**	**52**
PA	368073	1990-2005	3	0	1
PA	**368308**	**1990-2005**	**1**	**0**	**56**

Weather data were also estimated for the other stations, which were included, but had missing days of weather data. If the precipitation was missing it was assumed to be 0. If the maximum or minimum daily temperature was missing it was assumed to be the average maximum or minimum monthly temperature shown in Table 46. These assumptions introduce error; however, it is likely small because the number of missing days is small (less than 5 percent of the total simulation length of 10 years).

Table 46: Average Maximum and Minimum Monthly Temperatures for Maryland (TWC, 2007)

Month	Average Max (°F)	Average Min (°F)
January	42	25
February	46	28
March	55	35
April	66	44
May	75	54
June	84	64
July	89	69
August	87	67
September	80	60
October	69	47
November	57	38
December	47	30

Soils

The soils layer contains specific information relating to soil properties, including the available water-holding capacity (awc), the soil erodibility (K), organic matter content and dominant soil hydrologic group.

The soils layer for Maryland used for this study was downloaded from the Maryland Department of Planning (MDP) website (MDP, 2007). The properties of the MDP soil groups included awc, K, and dominant hydrologic group. It did not include information on the organic soil content. However, the organic soil content is not used in the current version of AVGWLF (6.3.7) and it is suggested that a default value of 2.5 percent be used if the actual value is not known (Evans and Corradini, 2006). Therefore the organic matter content was assumed to be 2.5 percent.

The awc used in AVGWLF is in units of depth and the MDP awc is in units of depth per depth. In order to convert the unitless awc to units of depth it needs to be multiplied by the mean rooting depth (Haith et al., 1992). Since the mean rooting depth is not known, it was assumed to be 3 feet. This value was assumed because it is the rooting depth of corn and small grains, which are the major row crops grown in Maryland (NDSU, 1997 and Delgado et al., 2001)

Point Sources

The point source data layer identifies the point source locations and mean annual total nitrogen and phosphorus point source discharges. The locations and discharges of the point sources in Maryland were obtained from the Chesapeake Bay Program (CBP) Nutrient Point Source Database (CBP, 1998). The mean annual total phosphorus and total nitrogen point source discharges were calculated by downloading the annual discharges from 1992 until 1997 and averaging them.

Water Extraction

The water extraction layer identifies the location of water extractions and the volume of water extracted. The location of well permits and the volume of water permitted to be extracted were obtained from the MDP. Some error is introduced into the AVGWLF simulation because the water extraction layer contains allowable water extractions, which is different than actual water extractions. The actual volume of water extracted is probably less that the permitted amount.

Tile Drains

The tile drains layer identifies the locations where agricultural tile drainage is used. I was unable to find maps that identify where tile drainage is used in Maryland and therefore this data layer was not included. The amount of tile drainage used in the study watersheds is likely small and therefore does not introduce much error.

Unpaved Roads

The unpaved roads layer identifies the location of unpaved roads. The area which is unpaved is treated as a "non-vegetated" surface in the AVGWLF simulations. This data layer was not developed for this study because the percentage of the watersheds that were covered with unpaved roads was so small that it was unlikely to affect the results.

Roads

The roads layer is a vector that identifies the location of paved roads in the watersheds. This data layer was not developed because the roads shapefile is a "background" layer and has no effect on the simulation results.

County Boundaries

The county boundaries shapefile contains information on the C and P values used in the USLE. The AVGWLF Format Guide (Evans and Corradini, 2006) uses the following values for the entire state of Pennsylvania:

C_crop = 0.42
C_past = 0.03
C_wood = 0.002
P1 = 0.52
P2 = 0.45
P3 = 0.52
P4 = 0.66
P5 = 0.74

These values can be modified to reflect the local cropping practices and geography. However, I felt that modifying the default values would not significantly increase the accuracy of the simulations and therefore I used the same default values used for Pennsylvania.

Septic Systems

The septic systems layer is used to identify how many people use septic systems, public sewers, or other waste disposal systems. In order to obtain this information for Maryland the 1990 federal census was downloaded from the MDP website. The 1990 census was used, as opposed to the 2000 census, because the 1990 census includes information on sewage disposal. Unfortunately, the 1990 census includes the number of housing units, instead of the number of people, that use each sewage disposal system (septic, sewer, other). In order to find the number of people who use each sewage disposal system I multiplied the number of housing units by the average number of people per housing unit. I found the average number of people per housing unit by dividing the population by the total housing units.

Animal Density

The animal density layer contains information on the animal equivalent units (AEUs) per acre, where an AEU is defined as 1000 pounds of weight. This layer is used to estimate nutrient runoff from animal manure. The Census of Agriculture (USDA, 2006) was used to develop the animal density data layer for the state of Maryland. When data was available the 1997 census was used because the study time period is from 4/1/1990 to 3/31/2000. If data from 1997 were not available, then data from the 2002 census were used.

The Census of Agriculture provides information on the number of agricultural animals per county in the United States. I downloaded the number of cows, hogs, sheep, horses, chickens, and turkeys for each county in Maryland from the Census of Agriculture website (USDA, 2006). This gave me the total number of agricultural animals per county in Maryland. In order to convert the total number of animals to AEU I multiplied each animal by an AEU coefficient, where the AEU coefficient is defined by equation A-1:

$$AEU = \frac{\overline{w}}{1000} \qquad \text{(A-1)}$$

where $\overline{w}$ is the average weight of the animal in pounds. Table 47 shows the average weight of the agricultural animals and the AEU coefficients used to determine the total AEU for each county in Maryland.

Once the total AEU had been determined for each county in Maryland, I found the area of each county in acres. Then I determined the AEU per acre for each county in Maryland by dividing the AEU for the county by the area of the county in acres.

Table 47: Average Weight and AEU Coefficient for Agricultural Animals in Maryland

Animal	Average Weight (lbs.)	AEU Coefficient
Cattle	900	0.9
Hogs	200	0.2
Sheep	125	0.125
Horses	1000	1
Chickens	3.5	0.0035
Broilers (meat chickens)	2.9	0.0029
Turkeys	12	0.012

Physiographic Province

The physiographic province layer contains information on the rainfall intensity during warm and cool seasons and the groundwater recession coefficient for different provinces. Table B-14 and Figure B-1 in the GWLF User's Manual (Haith et al., 1997) were used to determine the warm and cool season rainfall intensities in Maryland provinces. The groundwater recession coefficient was assumed to be the AVGWLF default value of 0.1. However, in the smaller watersheds using a groundwater recession coefficient of 0.1 resulted in many days when the simulated streamflow was zero. Since a streamflow of zero is not likely the groundwater recession coefficient was adjusted. A groundwater coefficient of 0.01 was used for the two smaller watersheds.

Land Use/Cover

The land use/cover layer contains information on the surface use or cover of land. It is very important and used to derive several GWLF model parameters. The AVGWLF land use layer has specific land use codes, which are shown in Table 48. In order to develop this data layer for Maryland the 2000 MDP land use, when available, was obtained from GISHydro (Moglen, 2006). Then the MDP land use codes were converted to the AVGWLF land use codes. Table 49 shows which AVGWLF land use codes the MDP land use codes were converted to. For example, the MDP land use code 14, which represents commercial, was converted to the AVGWLF land use code 3, which represents high-density development.

Table 48: AVGWLF Land Use/cover Codes (Evans and Corradini, 2006)

Category	Cell Value
Water	1
Low-Density Development	2
High-Density Development	3
Hay/Pasture[1]	4
Row Crops	5 or 6
Coniferous Forest[2]	7
Mixed Forest[2]	8
Deciduous[2]	9
Woody Wetland[3]	10
Emergent Wetland[3]	11
Quarries[4]	12
Coal Mines[4]	13
Beaches	14
Transitional[4]	15
Turfgrass/Golf Course[5]	16

Table 49: MDP and Corresponding AVGWLF Land Use Codes

MDP Land Use		AVGWLF Land Use	
Code	Category Name	Code	Category Name
11	res. low density	2	low-density development
12	res. medium density	2	low-density development
13	res. high density	3	high-density development
14	commercial	3	high-density development
15	industrial	3	high-density development
16	institutional	2	low-density development
17	extractive	15	transitional
18	open urban land	16	turf grass/golf course
21	cropland	5	row crops
22	pasture	4	hay/pasture
23	orchards	9	deciduous
25	row crops	5	row crops
41	deciduous	9	deciduous
42	evergreen	7	coniferous forest
43	mixed	8	mixed forest
44	brush	8	mixed forest
50	water	1	water
60	wetlands	10	woody wetlands
70	barren	15	transitional
71	beaches	14	beaches
72	bare exposed rock	15	transitional
73	bare ground	15	transitional
80	transportation	3	high-density development
191	large lot agricultural	5	row crops
192	large lot forest	8	mixed forest
241	feeding operations	15	transitional
242	agricultural buildings	4	hay/pasture

If the MDP land use was not available for the entire watershed, which was the case for the Fairview and Bridgeport watersheds, then GIRAS (USGS, 2007b) land use from the 1970's was used. Table 50 shows which AVGWLF land use codes the Anderson land use codes (Anderson et al., 1976) of the GIRAS data were converted to.

Table 50: GIRAS and Corresponding AVGWLF Land Use Codes

GIRAS Land Use		AVGWLF Land Use	
Code	Category Name	Code	Category Name
11	residential	2	low-density development
12	commercial and services	3	high-density development
13	industrial	3	high-density development
14	transportation, communications, and utilities	3	high-density development
15	industrial and commercial complexes	3	high-density development
16	mixed urban or built-up land	3	high-density development
17	other urban or built-up land	3	high-density development
21	cropland and pasture	5	row crops
22	orchards, groves, vineyards	9	deciduous
23	confined feeding operations	15	transitional
24	other agricultural land	5	row crops
31	herbaceous rangeland	4	hay/pasture
32	shrub and brush rangeland	4	hay/pasture
33	mixed rangeland	4	hay/pasture
41	deciduous forest land	9	deciduous
42	evergreen forest land	7	coniferous forest
43	mixed forest land	8	mixed forest
51	streams and canals	1	water
52	lakes	1	water
53	reservoirs	1	water
54	bays and estuaries	1	water
61	forrested wetland	10	woody wetland
62	nonforested wetland	11	emergent wetland
71	dry salt flats	15	transitional
72	beaches	14	beaches
73	sandy areas other than beaches	14	beaches
74	bare exposed rock	15	transitional
75	strip mines, quarries, and gravel pits	15	transitional
76	transitional areas	15	transitional
77	mixed barren land	15	transitional

Surface Elevation

The surface elevation layer contains information on the elevation of the land surface. This data layer was developed for Maryland using digital elevation model (DEM) data obtained from GISHydro (Moglen, 2006). The resolution of the DEM data layer is 30 meters.

Groundwater Nitrogen

The groundwater nitrogen layer provides estimates of initial background concentrations of nitrogen in groundwater and affects the amount of simulated nitrogen in streams. This data layer was not developed for Maryland because it only provides an initial estimate and therefore only affects the simulated nitrogen concentration for the first year, at most.

Soil Phosphorus

The soil phosphorus layer contains information on the concentration of soil phosphorus in the soil. It can represent either soil phosphorus as measured by the Bray, Olsen, or Mehlich tests (Evans and Corradini, 2006), or total phosphorus, which includes organic, inorganic, dissolved and solid phosphorus. The soil phosphorus layer is developed by using surface interpolation routines within the Spatial Analyst extension of ArcView (cite ESRI) and known soil phosphorus concentrations at specific test locations. This data layer was not developed for Maryland because the improvements in accuracy seemed unlikely to warrant the time and energy to develop it. However, if the AVGWLF model is found to systematically underestimate total phosphorus, this data layer will be developed and the AVGWLF model run again to see if total phosphorus estimates improve.

Appendix B

Table 51: Goodness-of-fit Statistics for Concentration-derived Nitrogen Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	y = 8.656x	-0.0417	0.3081
Crumpton	$y = -0.004404x^2 + 14.58x$	-2.156	3.880
Fairview	$y = -0.0001468x^2 + 11.74x$	0.0274	0.3270
Bridgeport	y = 4.932x	7603	12744
Snowhill	y = 3.100x	432.0	972.7

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 52: Goodness-of-fit Statistics for Concentration-derived Phosphorus Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	$y = 0.0007340x^2 + 0.06116x$	-0.0205	0.4272
Crumpton	$y = 0.0001957x^2 + 0.1064x$	0.0556	0.1982
Fairview	$y = 0.00004893x^2 + 0.5485x$	-0.0029	0.6227
Bridgeport	$y = 0.00004893x^2 + 0.4264x$	0.0067	0.5924
Snowhill	y = 0.07193x	387.1	500.4

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 53: Goodness-of-fit Statistics for Concentration-derived Sediment Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	$y = 99.17x^{1.380}$	0.0000	0.9288
Crumpton	$y = 42.64x^{1.391}$	0.0000	0.6865
Fairview	$y = 17.60x^{1.392}$	0.0000	0.9211
Bridgeport	$y = 206.0x^2$	0.0000	1.016
Snowhill	$y = 10.99x^{1.265}$	0.0000	0.9158

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 54: Goodness-of-fit Statistics for EPA Nitrogen Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	y = 9.905x - 13.58	-0.0209	0.7950
Crumpton	y = 5.171x + 114.9	0.0013	0.0826
Fairview	y = 10.23x + 1736	0.0002	0.3202
Bridgeport	y = 4.837x + 944.9	0.0001	0.4025
Snowhill	y = 3.461x + 15.92	0.0001	0.2416

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 55: Goodness-of-fit Statistics for EPA Phosphorus Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	y = 0.2351x - 2.405	-0.0067	0.4024
Crumpton	y = 0.5331x - 7.130	0.0001	0.1588
Fairview	y = 1.065x - 606.3	0.0004	0.6522
Bridgeport	y = 0.8594x - 130.6	0.0000	0.5459
Snowhill	y = 0.1364x - 5.787	0.0001	0.5313

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 56: Goodness-of-fit Statistics for EPA Phosphorus Rating Curves with a Zero Intercept

	Equation	*e/y*	S_e/S_y
Kennedy	y = 0.2165x	0.4009	0.4437
Crumpton	y = 0.5289x	0.2364	0.1635
Fairview	y = 0.9622x	0.1412	0.6637
Bridgeport	y = 0.8430x	0.0315	0.5465
Snowhill	y = 0.1248x	0.3470	0.5497

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 57: Goodness-of-fit Statistics for EPA Sediment Rating Curves

	Equation	*e/y*	S_e/S_y
Kennedy	$y = 886.2x^{0.8755}$	0.0000	0.6354
Crumpton	$y = 2969x^{0.8259}$	0.0000	0.2894
Fairview	$y = 500.7x^{0.9939}$	0.0000	0.7604
Bridgeport	$y = 72.75x^{1.139}$	0.0000	0.2805
Snowhill	$y = 5.303x^{1.389}$	0.0000	0.4725

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Appendix C

Table 58: Goodness-of-fit Statistics for Monthly AVGWLF Nitrogen Loads Compared to Concentration-derived Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.4079x - 51.26	0.2689	-0.6073	0.8550
Crumpton	y = 0.1646x + 790.4	-2.019	-0.6091	1.738
Fairview	y = 0.4187x - 370.4	0.1617	-0.5828	0.9156
Bridgeport	y = 0.8338x + 3307	0.5719	-0.07787	0.6543
Snowhill	y = 0.7746x - 308.9	0.7925	-0.2865	0.4555

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 59: Goodness-of-fit Statistics for Monthly AVGWLF Phosphorus Loads Compared to Concentration-derived Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.07801x + 47.05	0.1379	-0.5227	0.9285
Crumpton	y = 11.45x - 166.8	-190.9	4.655	13.85
Fairview	y = 0.5541x + 1462	-0.08177	-0.3428	1.040
Bridgeport	y = 0.7862x + 249.8	-0.05897	-0.1502	1.029
Snowhill	y = 1.836x + 11.74	-4.102	0.9360	2.259

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 60: Goodness-of-fit Statistics for Monthly AVGWLF Sediment Loads Compared to Concentration-derived Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.07960x + 21.76	0.1034	-0.8099	0.9469
Crumpton	y = 6.479x - 14.60	-58.56	5.020	7.717
Fairview	y = 1.427x + 6017	-3.674	1.307	2.162
Bridgeport	y = 2.935x + 176.9	-19.62	2.048	4.541
Snowhill	y = 2.204x + 48.37	-6.329	1.917	2.707

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 61: Goodness-of-fit Statistics for Monthly AVGWLF Nitrogen Loads Compared to EPA Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.3563x + 96.56	0.2770	-0.6157	0.8503
Crumpton	y = 0.3824x - 453.2	-18.25	-0.7127	4.387
Fairview	y = 0.4551x - 21720	-0.08105	-0.6241	1.040
Bridgeport	y = 0.8492x - 21080	0.01766	-0.4728	0.9911
Snowhill	y = 0.6942x - 646.4	0.6730	-0.4113	0.5718

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 62: Goodness-of-fit Statistics for Monthly AVGWLF Phosphorus Loads Compared to EPA Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.9124x + 20.49	0.6035	0.4351	0.6297
Crumpton	y = 5.234x + 61.62	-35.68	7.391	6.057
Fairview	y = 0.4712x + 4140	0.1438	-0.1525	0.9253
Bridgeport	y = 0.5815x + 829.6	0.1983	-0.2264	0.8954
Snowhill	y = 1.192x + 83.37	-0.6450	0.8847	1.283

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 63: Goodness-of-fit Statistics for Monthly AVGWLF Phosphorus Loads Compared to EPA Loads Calculated using a Rating Curve with a Zero Intercept

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.9489x - 286.3	0.5118	-0.3337	0.6987
Crumpton	y = 4.240x - 384.7	-20.10	0.2624	4.593
Fairview	y = 0.4361x + 240.9	-0.1430	-0.5523	1.069
Bridgeport	y = 0.5237x - 9.365	0.1097	-0.4778	0.9435
Snowhill	y = 1.058x + 11.74	-0.2423	0.1158	1.115

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Table 64: Goodness-of-fit Statistics for Monthly AVGWLF Sediment Loads Compared to EPA Loads

Watershed	Equation	*N-S*	*e/y*	S_e/S_y
Kennedy	y = 0.3929x - 53.04	-0.8850	-0.8374	1.373
Crumpton	y = 1.896x - 708.6	-6.351	-0.5837	2.711
Fairview	y = 1.401x + 1242	-2.613	0.5202	1.901
Bridgeport	y = 2.878x + 653.4	-18.60	2.336	4.427
Snowhill	y = 2.109x + 64.66	-5.660	2.134	2.581

Note: y is the pollutant load in kg/day and x is the discharge in ft^3/s

Bibliography

American Geophysical Union (AGU). 2001. Proceedings of the American Geophysical Union Fall Meeting. San Francisco, CA. December 10-14, 2001.

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer. 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S. Geological Survey Professional Paper 964. 28 pp.

Bicknell, B.R., J.C. Imhoff, J.L. Kittle Jr., A.S. Donigian Jr., and R.C. Johanson. 1997. Hydrological Simulation Program--Fortran, User's manual for version 11. EPA-600-R-97-080. National Exposure Research Laboratory. U.S. Environmental Protection Agency, Athens, Ga. 755 pp.

Bingner, R. L., F.D. Theurer, R.G. Cronshey, R.W. Darden. 2006. AGNPS 2006 Web Site. <http://www.ars.usda.gov/Research/docs.htm?docid=5199> May, 2007.

Bingner, R.L., R.W. Darden, F.D. Theurer, C.V. Alonso, and P. Smith. 1998. AnnAGNPS input parameter editor interface. Proceedings of the First Federal Interagency Hydrologic Modeling Conference. Las Vegas, Nevada. April 19-23, 1998. pp. 8-15 to 8-18.

Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley and V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, Vol. 8, No. 3, pp. 559-568.

Chesapeake Bay Program (CBP). 1998. CBP Nutrient Point Source Database. <http://www.chesapeakebay.net/data/index.htm> October, 2006.

Chesapeake Bay Program (CBP). 2004. The State of the Chesapeake Bay and its Watershed. CBP-TRS-273-05. EPA-903-R-04-009. USEPA Chesapeake Bay Program Office, Annapolis, MD. 23 pp.

Chesapeake Bay Program (CBP). 2005. Nutrient Pollution. <http://www.chesapeakebay.net/nutr1.htm> March, 2007.

Delgado, J.A., R.R. Riggenbach, R.T. Sparks, M.A. Dillon, L.M. Kawanabe, and R.J. Ristau. 2001. Evaluation of Nitrate-Nitrogen Transport in a Potato–Barley Rotation. *Soil Science Society of America*, Vol. 65, pp. 878-883.

Deliman, P.N., R.H. Glick, and C.E. Ruiz. 1999. Review of Water Quality Models. Technical Report W-99-1. U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Evans, B.M. 2002. Development of an Automated GIS-Based Modeling Approach to Support Regional Watershed Assessments. Ph.D. Dissertation in Soil Science, The Pennsylvania State University. 231 pp.

Evans, B.M., D.W. Lehning, K.J. Corradini, G.W. Petersen, E. Nizeyimana, J.M. Hamlett, P.D. Robillard, and R.L. Day. 2002. A Comprehensive GIS-Based Modeling Approach for Predicting Nutrient Loads in Watersheds. *Journal of Spatial Hydrology*, Vol. 2, No. 2, 18 pp.

Evans, B.M., and K.J. Corradini. 2006. AVGWLF, Version 6.0: A Guide to Creating Software-compatible Data Sets. Penn State Institutes of the Environment. 34 pp.

Evans, B.M., S.A. Sheeder, and K.J. Corradini. 2006. AVGWLF, Version 6.3: Users Guide. Penn State Institutes of the Environment. 79 pp.

Haith, D.A. and L.L. Shoemaker. 1987. Generalized Loading Functions for Stream Flow Nutrients. *Water Resources Bulletin*, Vol. 23, No. 3, pp. 471-478.

Hopkins, K., B. Brown, L.C. Linker, R.L. Mader, Jr.. 2000. Chesapeake Bay Watershed Model Land Use and Model Linkages to the Airshed and Estuarine Models. A report of the Modeling Subcommitte. USEPA Chesapeake Bay Program Office, Annapolis, MD. 76 pp.

Lee, K.Y., D.E. Weller, T.R. Fisher, T.E. Jordan, and D.L. Correl. 2000. Modeling the Hydrochemistry of the Choptank River Basin using GWLF and Arc/Info. *Biogeochemistry*, Vol. 49, No. 2, pp. 143-173.

Leonard, R.A., W.G. Knisel, and D.A. Still. 1987. GLEAMS: Groundwater Loading Effects of Agricultural Management Systems. *Transactions ASAE*, Vol. 30, pp. 1403-1418.

Linker, L.C., G.W. Shenk, and K. Hopkins. 2002. Watershed Model Phase 4.3 Calibration Rules. USEPA Chesapeake Bay Program Office. Annapolis, MD. 71 pp.

Lowrance, R.R. 2003. Modeling water quality. Encyclopedia of Water Science. Ed. Stewart, B.A., and T.A. Howell. New York: Marcel Dekker, 2003.

Maryland Cooperative Extension (MCE). 2006. How the MCE Agricultural Nutrient Management Program Got Started. Agricultural Nutrient Management Program. <http://www.agnr.umd.edu/users/agron/nutrient/Bkgrnd/History_ANMP.pdf> March, 2007.

Maryland Department of the Environment (MDE). 2005. Final: Total Maximum Daily Loads of Nitrogen, Phosphorus and Biochemical Oxygen Demand for Breton Bay in St. Mary's County, Maryland. Maryland Department of the Environment, Baltimore, MD. 35 pp.

Maryland Department of the Environment (MDE). 2007. Water Quality Standards. <http://www.mde.state.md.us/researchcenter/data/waterqualitystandards/index.asp> March, 2007.

Maryland Department of Planning (MDP). 2007. GIS Data Download. Natural Soil Groups. <http://www.mdp.state.md.us/zip_downloads_accept.htm> July, 2006

Mattikalli, N.M. and K.S. Richards. 1996. Estimation of Surface Water Quality Changes in Response to Land Use Change: Application of the Export Coefficient Model Using Remote Sensing and Geographical Information System. *Journal of Environmental Management*, Vol. 48, pp. 262-282.

McCuen, R.H. 1993. Microcomputer Applications in Statistical Hydrology. Englewood Cliffs, NJ: Prentice-Hall.

McCuen, R.H. and C.Q. Surbeck. 2007 An Alternative to Specious Linearization of Environmental Models. *Environment Science and Technology* (under review).

Moglen, G.E. 2006. Instructions for Using the Web-Based Version of GISHydro2000. <http://www.gishydro.umd.edu/documents/mdsha_reports/gishydro_web_version.pdf> November, 2006.

Nash, J.E. and J.V. Sutcliffe. 1970. River Flow Forecasting Through Conceptual Models Part I – A Discussion of Principles. *Journal of Hydrology*, Vol. 10, pp. 282-290.

National Oceanic and Atmospheric Administration (NOAA). 2007. National Climatic Data Center (NCDC) Climate Data Online. <http://cdo.ncdc.noaa.gov/CDO/dataproduct> September, 2006.

Neitsch, S.L, J.G. Arnold, J.R. Kiniry, and J.R. Williams. 2005. Soil and Water Assessment Tool Theoretical Documentation, Version 2005. U.S. Department of Agriculture, Agricultural Research Service, Temple, TX.

North Dakota State University (NDSU). May 1997. Corn Production Guide: Irrigation Management. <http://www.ag.ndsu.edu/pubs/plantsci/rowcrops/a1130-2.htm> November, 2006.

Rossman, L.A. 2005. Storm Water Management Model User's Manual Version 5.0. EPA 600-R-05-040. Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH. 249 pp.

Schneiderman, E., D.G. Lounsbury, B.J. Dibeler, D.J. Thongs, J.W. Tone, and R. Danboise-Lohre. 1998. Application of GWLF to Non-Point Source Loading Model to the NYC Catskill and Delaware System Watersheds. Study Report to EPA. 89 pp.

Schwarz, G.E., A.B. Hoos, R.B. Alexander, and R.A. Smith. 2006. The SPARROW Surface Water-Quality Model: Theory, Application and User Documentation. U.S. Geological Survey Techniques and Methods 6-B3. 248 pp.

Shamsi, U. 2002. GIS Tools for Water, Wastewater, and Stormwater Systems. Reston, VA:ASCE Publications.

Shenk, G.W., and L.C. Linker. 2001. Simulating the Chesapeake Bay Watershed with Time-Varying Land Use and Management Action. USEPA Chesapeake Bay Program Office, Annapolis, MD. 13 pp.

Shoemaker, L., T. Dai, and J. Koenig. 2005. TMDL Model Evaluation and Research Needs. EPA 600-R-05-149. Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, OH. 109 pp.

Soil Conservation Service (SCS). 1986. Urban Hydrology for Small Watershed. Technical Report 55. (TR-55), Washington, D.C.

Tim, U.S. and B. Crumpton. 2003. Nutrient Loading Models. Presentation at Evaluating Wetland Condition for Nutrient Criteria: An Outreach Workshop. Wetlands Work Group. Atlantic City, NJ. March 25-27, 2003.

U.S. Army Corps of Engineers, Hydrologic Engineering Center (USACE-HEC). 1977. Storage, Treatment, Overflow, Runoff Model, STORM, Generalized Computer Program 723-58-L77520. USACE-HEC, Davis, California.

U.S. Environmental Protection Agency (USEPA). 1999. Protocol for Developing Nutrient TMDLs. EPA 841-B-99-007. Office of Water (4503F). U.S. Environmental Protection Agency, Washington D.C. 135 pp.

U.S. Environmental Protection Agency (USEPA). 2001a. Better Assessment Science Integrating Point and Nonpoint Sources, BASINS Version 3.0. EPA 823-B-01-001. Office of Water (4305F). U.S. Environmental Protection Agency, Washington D.C. 337 pp.

U.S. Environmental Protection Agency (USEPA). 2001b. PLOAD version 3.0: An ArcView GIS Tool to Calculate Nonpoint Sources of Pollution in Watershed and Stormwater Projects: User’s Manual. U.S. Environmental Protection Agency, Washington D.C. 48 pp.

U.S. Environmental Protection Agency (USEPA). 2005. Draft Handbook for Developing Watershed Plans to Restore and Protect Our Waters. EPA 841-B-05-005. Office of Water: Nonpoint Source Control Branch. U.S. Environmental Protection Agency, Washington, D.C. 414 pp.

U.S. Department of Agriculture (USDA). 2007. National Agricultural Statistics Service. <http://www.nass.usda.gov/index.asp> October, 2006.

U.S. Geological Survey (USGS). 1994. USGS Water Resources Data. Definition of Terms. <http://ca.water.usgs.gov/archive/waterdata/text/definition.html> March, 2007.

U.S. Geographical Survey (USGS). 2005a. National Hydrography Dataset. <http://nhd.usgs.gov/data.html> August, 2006.

U.S. Geological Survey (USGS). 2005b. The Impact of Sediment on the Chesapeake Bay and its Watershed. <http://chesapeake.usgs.gov/SedimentBay605.pdf> March, 2007.

U.S. Geographical Survey (USGS). 2007a. Geographic Data Download: GIRAS <http://edc2.usgs.gov/geodata/index.php> September, 2006.

U.S. Geological Survey (USGS). 2007b. National Water Quality Assessment Data Warehouse. <http://infotrek.er.usgs.gov/traverse/f?p=NAWQA:HOME:755096144276101> September, 2006.

U.S. Geological Survey (USGS). 2007c. USGS Chesapeake Bay Activities: Water – Nutrients. <http://chesapeake.usgs.gov/nutrients.html> March, 2007.

U.S. Geological Survey (USGS). 2007d. USGS Chesapeake Bay Activities: Water – Sediments <http://chesapeake.usgs.gov/sediments.html> March, 2007.

U.S. Fish and Wildlife Service (FWS). 2007. Nutrient Pollution. <http://www.fws.gov/chesapeakebay/nutrient.htm> March, 2007.

The Weather Channel (TWC). 2007. Monthly Averages for College Park, MD. <http://www.weather.com/> September, 2006.