

3. METHODOLOGY AND DATA DEVELOPMENT

3.1 *The Bull Run Watershed*

The 25.8 mi² Bull Run watershed in Northern Virginia (Figure 3-1) was selected for this study because it had an available historical record of monitoring data, it had experienced a land use shift from agriculture to rural residential during the monitoring period, and it was small enough to allow modeling with a small cell size within the constraints of the AGNPS 5.0 model. The Bull Run watershed described in this study does not include the total drainage of Bull Run, but was defined by its monitoring station (ST60) at the intersection of Bull Run with Route 705 near Catharpin, Virginia (OWML, 1994). The Bull Run watershed includes portions of Loudoun and Prince William counties in northern Virginia, and is bounded by the Bull Run Mountains on the west. Bull Run is part of the larger Occoquan watershed, which contributes flow to the Occoquan Reservoir. The Occoquan Reservoir is located about 25 miles southwest of Washington, D.C., and is the primary water supply for D.C.'s Virginia suburbs.

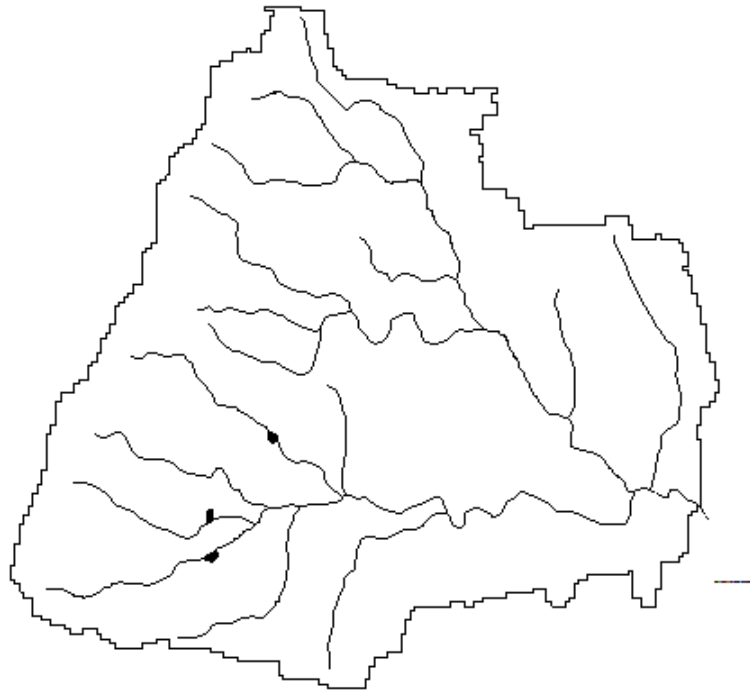


Figure 3-1. Bull Run Watershed at Route 705

Because of its proximity to a metropolitan area, the Occoquan watershed began experiencing large increases in its rural and suburban population starting in the early 1960's. Declining water quality in the reservoir was noted by the late 1960's, and was partially blamed on urban stormwater from increasing areas of suburban land uses. Because of its importance as a water supply for a large populace, steps were taken to protect the watershed and water quality in the

reservoir. One component of the resulting policy was the installation of a monitoring network covering the entire 570 mi² of the Occoquan watershed. The Occoquan Watershed Monitoring Laboratory (OWML) was established to maintain the monitoring network and was the source of the monitoring data used in this study. From 1963 to 1986, the U.S. Geological Survey also maintained a surface runoff gauge at the same site. This watershed lies in the Triassic region of the Piedmont Lowland. Land use statistics from the Northern Virginia Planning District Commission (NVPDC, 1991) show an increase in non-agricultural and non-forested land uses in the Occoquan from 7.3% in 1977 to 29.0% in 1989. In 1990, Bull Run land use included approximately 40% forest, 46% agriculture (mainly hayland), and 12½% rural residential, as categorized by the EPA Chesapeake Bay Program's staff in their 1990 digitized map of the entire Chesapeake Bay drainage area (Neumiller et al., 1994).

3.2 Framework for Comparison

In this study, runoff and loads calculated from monitored data were compared with modeled runoff and loads from the AGNPS 5.0 model on a select number of storms from the period 1977-1992. Runoff and loads were modeled in a spatially distributed manner throughout the watershed, although the comparison with monitored runoff and loads was only at one point, the watershed outlet.

Initially, monitored runoff and loads were calculated on a daily basis, while the AGNPS model simulated runoff and loads on an event basis. Two procedures were used to assign alternative parameter values in the AGNPS input files producing two modeled data sets with an event time basis, and two procedures were used to simulate monthly runoff and loads producing two modeled data sets on a monthly time basis. Overall then, four sets of modeled runoff and loads were produced by the auxiliary procedures in this study - two on an event basis and two on a monthly basis - for comparison with observed data.

3.2.1 Time Periods for Comparison

A daily time interval was used initially as the basis for comparing modeled events with observed data. Flow-weighted composite sampling, on which the observed data were based, was usually performed over a multi-day period. Average concentrations and total flow during this interval were used to calculate total loads and runoff volume for the runoff event(s). Daily loads and runoff were extracted from these totals on a proportional basis. Runoff and loads from modeled events were assigned to the day of the storm.

When comparing modeled and observed runoff and loads, a problematic trend was observed in the data. This trend showed that modeled runoff and loads generally exceeded observed runoff and loads on the day of the storm, with observed runoff and loads exceeding modeled runoff and loads on one or more days following the storm. This apparent mis-alignment of observed and modeled output led to a re-assessment of the daily time interval as the basis for comparing modeled and monitored event data.

A closer look revealed the following problems with using the daily time basis:

- Modeled runoff and loads from storm events are not limited to the day of the storm. Indeed, rainfall frequently gives rise to runoff and its associated loads over a period of days.
- Average values within the composite sampling period mask the hydrograph's distribution of runoff and loads within a runoff event. Therefore, a time proportional allocation of runoff and loads within the composite period will distort the data distribution.

A more equitable time period for comparison, therefore, was the time period corresponding with individual composite sample intervals, since distributional information within the interval was not available, and this interval generally corresponded with that of a runoff event, the same as that resulting from an AGNPS model run. This time interval is referred to as the composite period in this study, and was used instead of the daily time interval for comparing modeled and observed event data. Event data for the most part corresponded with composite periods. However, modeled output from sequential events were added together, where one runoff event could not be clearly distinguished from the next, or where more than one rainfall event occurred during a composite period.

3.2.2 Auxiliary Modeling Procedures

Two alternative parameterization procedures were programmed to assign parameter values for various soil, land use or parcel attribute data. The first procedure used typical annual average values for parameters, while the second procedure assigned seasonally-variable parameter values on a daily, bi-weekly or monthly basis. A geographic information system (GIS) was used to facilitate linking parameter values by soil type, land use, or parcel to create model input files for modeling with AGNPS. Modeled output from these alternative procedures were compared with monitored data on a composite period basis. Event modeled output from these procedures were added together, where necessary, to correspond with monitored composite periods.

Two alternative monthly simulation procedures were programmed to calculate monthly runoff and loads from the time-variable composite period data. The first monthly procedure aggregated event output from all modeled events within any given month, while its alternative supplemented the aggregated monthly output with baseflow and modeled monthly septic system loads. A storm sequencing procedure was developed to facilitate running AGNPS in batch mode, while updating all of the necessary parameter values from one event to the next. Modeled runoff and loads from the two monthly simulation procedures were then compared with monthly monitored runoff and loads. Since the AGNPS model predicts flow and loads only from surface runoff, the latter monthly simulation procedure was created to simulate additional flow and loads included in the monitored data.

3.2.3 Data Sets for Comparison

The following data sets were created during this study from the OWML monitored data, and from AGNPS model runs with each of the event parameterization and monthly simulation procedures:

Monitored Data Sets

- OWcp: runoff and loads calculated from monitored OWML flow and concentrations over selected individual composite periods.
- OWmn: runoff and loads calculated from monitored OWML flow and concentrations from weekly baseflow and storm samples over each month.

Modeled Data Sets

- AG0cp: modeled runoff and loads summed over individual composite periods, using typical average annual parameter values.
- AG1cp: modeled runoff and loads summed over individual composite periods, using time-variable parameter values.
- AG1mn: modeled runoff and loads summed for all AG1cp events within each month.
- AG2mn: AG1mn, supplemented with monthly baseflow and septic system loads.

3.2.4 Selection of Rain Gauge Site

At the beginning of this study, daily rainfall was used from a National Weather Service rain gauge at Chantilly, Virginia. However, a comparison with the OWML runoff record indicated many mismatches with the Chantilly runoff. Three National Weather Service stations are located in the vicinity of the Bull Run watershed. Therefore, the complete record of daily rainfall available at that time, 1964-1992, was obtained from OWML for all three stations. When compared with the OWML runoff record, the rainfall from The Plains gauge provided the best match with the OWML runoff. A follow-up Thiessen weighting was performed on these gauges, which showed that about 90% of the watershed was influenced by the gauge located at The Plains in Fauquier County. The Plains rainfall record, therefore, was used for comparison with monitored OWML runoff in this study. The Plains rain gauge was located 3.75 miles from the nearest point in the watershed, and 9.75 miles from the watershed outlet.

3.2.5 Selection of Period of Record for Comparing Rainfall and Runoff

A complete database of the monitored data for Bull Run watershed at station ST60 was obtained from the Occoquan Watershed Monitoring Laboratory (OWML) in Manassas, Virginia for the years 1975 through the middle of 1994. In reviewing the database, persistent errors and omissions were noted in the sampling and analysis records for key parameters during the monitoring startup period, 1975-1976, so these years of data were not considered. The time period common to both the rainfall database and the runoff database after startup, 1977-1992, was then used for developing subsequent modeling and monitored data sets for comparison. The OWML database for this period included 292 composite sampled events and 3 events which were sampled at discrete intervals.

3.2.6 Selection of Daily Rainfall Records for Modeling

The following procedure was used to identify appropriate storms for modeling, and to achieve good correspondence between the rainfall and monitored runoff records. A program was written to compare The Plains rainfall database with the runoff records contained in the OWML database. This program selected only storms which fell within any of the composite periods

from the OWML database. Storms corresponding with discrete samples were manually added to the list of selected storms. A total of 779 days of measurable rainfall was identified during the 1997-1992 period. Small storms were excluded from modeling, as they were expected to produce little, if any modeled runoff and loads. Small storms in this study were defined as all storms less than 0.30", and those storms less than 0.50" with average and dry antecedent moisture conditions. These storms were eliminated from the list of storms to be modeled, resulting in 273 daily rainfall events to be modeled.

As modeling and data analysis progressed, large variability was noted in unit area loads for many of the smaller runoff-producing storms. The nitrogen and phosphorus loads appeared to be sediment driven, as runoff was minimal. A sensitivity analysis of the AGNPS model using the TR-55 and geomorphic options was performed by incrementally increasing daily rainfall and associated EI, and checking the resultant sediment yield for each AMC condition. The results (Figure 3-2) showed that the model predicted in an unreasonable fashion below different thresholds depending on the AMC condition. From the graph, the following thresholds were defined: 1.50" (AMC 1); 0.75" (AMC 2); and 0.30" (AMC 3). The set of daily events to be modeled was then reduced to eliminate storms below these thresholds. This reduced the number of daily events for modeling to 131.

3.2.7 Selection of Composite Period Data for Model Comparison

Within a composite period, more than one storm might contribute to resultant runoff and loads. If any of the storms within a composite period was rejected due to one of the TR-55 thresholds, those composite periods were considered to be incomplete and were not used for comparison. Any other storms which fell within an incomplete composite period were also removed from the list of storms to be modeled. This further reduced the list of selected daily rainfalls to 112, and to this point constituted 89 composite periods for comparing modeled and monitored output. These daily events accounted for a total of 153.1 inches of rainfall and 62.8 inches of runoff. These figures represent 24% of all rainfall and 31% of all runoff during the 1977-1992 period, and 74% of the total rainfall during that period from daily rainfalls greater than 1.50 inches. A list of the selected storms can be found in Appendix A, and a monthly distribution of rainfall is shown in Figure 3-3.

Further analysis of the selected storms and the composite periods showed that 4 daily rainfalls overlapped composite sampling periods. Since it could not be determined how much of the daily rainfall contributed to each of the adjoining periods, the adjoining periods were combined to put the comparison on a more equitable basis. This reduced the number of composited events to 85. Runoff and loads calculated from the observed sample data and generated by the various modeling procedures are reported on a composite period basis in Appendix B.

3.2.8 Selection of Complete Monthly Data for Indexing

Complete months of corresponding rainfall/runoff data were required for calculation of the NPSP index. Complete months of data were considered to be those months for which no storms

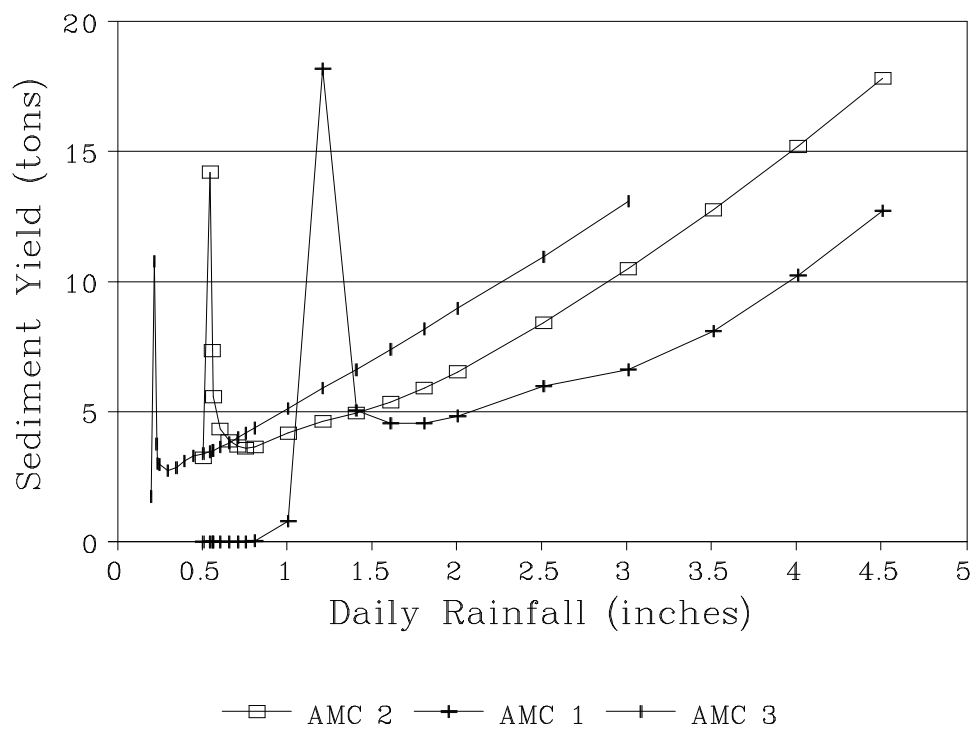


Figure 3-2. AGNPS TR-55/Geomorphic Modeling Thresholds

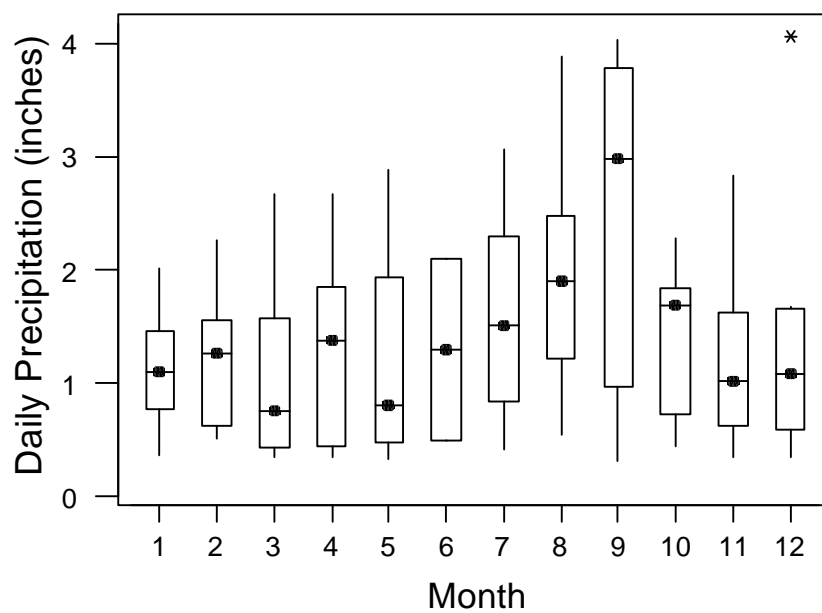


Figure 3-3. Monthly Rainfall Distribution of Daily Modeled Events with Interquartile Ranges (boxes), Median Values (black dots), and Outliers (*)

were rejected using the TR-55 thresholds mentioned above. Using this criterion, 24 complete months were identified. Runoff and loads calculated from the observed sample data and generated by the various modeling procedures are reported on a monthly basis in Appendix C.

3.3 Calculation of Monitored Loads

3.3.1 Corrections to the Monitored Data Records

The historical ST60 monitoring data record for the years 1977-1992 on Bull Run contained water quality samples collected with two different methods. Discrete sequential and composite sampling methods were used during different periods within the data record. Discrete sequential samples were collected in the earlier part of the data record through December 1976, and again, briefly, from 8/11/79 through 9/6/79, but the majority of the samples in the 1977-1992 period were taken as composite samples. These data were reviewed both for errors and for missing data, and corrected. The following corrections were applied to the database.

- During the period of study, many of the daily baseflow rates (FLO) were recorded as either blank, 0.00, or ERR. Stage-flow rating relationships were developed using baseflow records with complete stage and flow measurements encompassing the range of stages with missing flow values, and flow was calculated from the following regression equations, with stage given in feet, and FLO in cfs:

<u>Stage (ft)</u>	<u>Regression Equation</u>	<u>R²</u>
0.80 - 1.05	FLO = 0.05	
1.06 - 1.20	FLO = 1.6667*stage - 1.7	0.95
1.21 - 1.30	FLO = 10*stage - 11.7	0.83

- Concentrations reported as being below the detection limit were flagged with a “-” and the detection limit. These values were replaced with the positive detection limit.
- Where concentrations of a given pollutant were listed as ERR, indicating a processing or sampling error, the runoff data records were sorted in ascending order for that pollutant, and the concentration was calculated as an average of its adjacent neighbors. Where consecutive samples were reported as ERR, the average was calculated from the nearest two neighbors on either side of the missing values. Where many consecutive values of total suspended solids (TSS) were missing, and where many values were reported for the same flow rate, the average was calculated from all values corresponding to the neighboring FLO values.
- Where TYPE was missing, “B” indicating baseflow, was assigned where a single date and time was given, and an “R” indicating runoff, was assigned where two dates and times were given, indicating a composite sample. No missing TYPE values occurred during the discrete sampling period.
- Concentrations for several parameters from 2 composite periods were questionable. As monitored data were provided on an as-is basis, the validity of the samples was not corroborated by OWML. One of the questionable values was a total phosphorus (TP) concentration of 10 mg/L for the composite period 6/28-30/83. This was the highest TP

value in the database, and, unlike all other data entries, had neither a decimal point nor two decimal places in the entry. This entry was most likely a typographical error. The second questionable set of values, oddly enough, was the next consecutively sampled runoff event, four months later. Interim baseflow samples had been taken, but no other runoff samples. The second highest value of TP (6.26 mg/L) was recorded for this 10/23-24/83 sample. The highest values in the database for ortho-phosphorus (5.40 mg/L), ammonia nitrogen (5.84 mg/L), and total Kjeldahl nitrogen (56.33 mg/L) were also recorded for this sample. These values were all 4-6 times higher than the next highest concentration for each parameter. These high readings could have resulted from poor maintenance of the sampling equipment, especially with the long time between runoff samples. Output for total nitrogen and total phosphorus from initial model runs using composite period data, including the questionable data, are illustrated in Figures 3-4 and 3-5. Neither of these questionable records were used in the analysis, but are included in the data summaries in the Appendices.

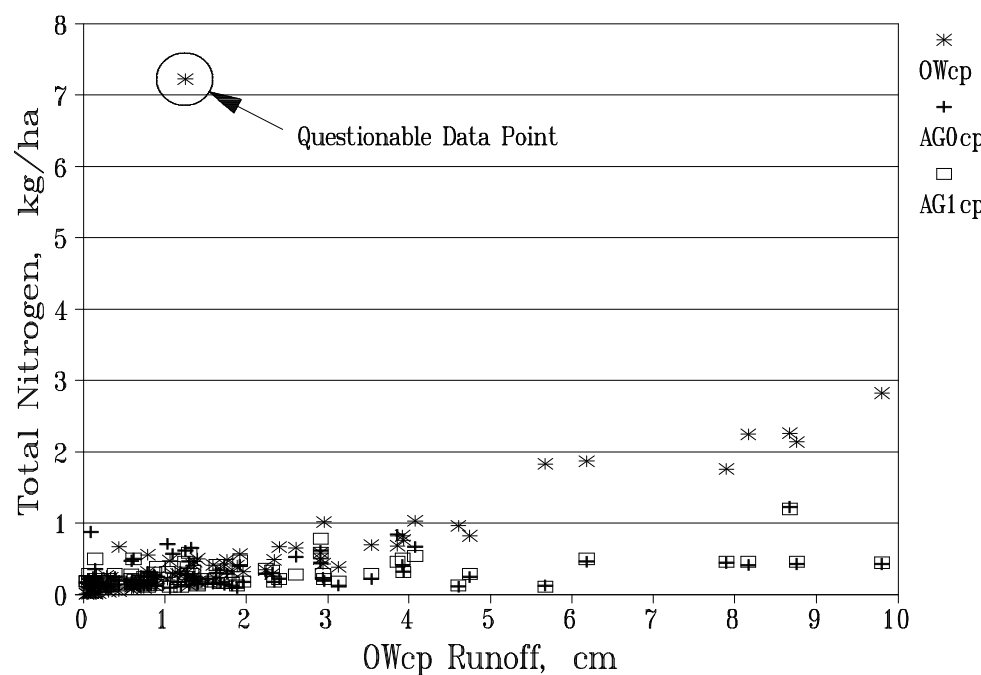


Figure 3-4. Composite Period Total Nitrogen with Questionable Data

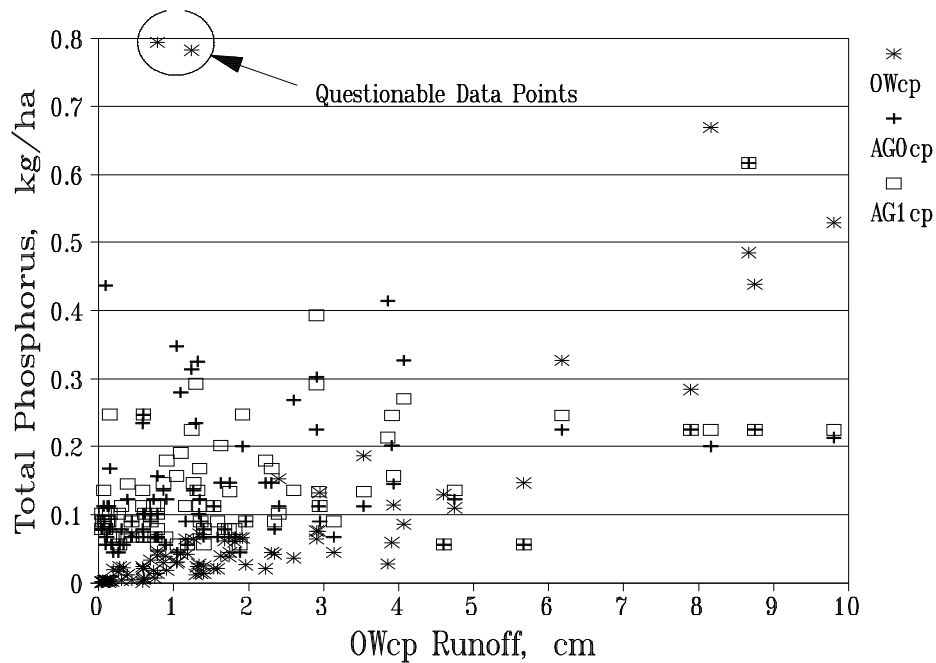


Figure 3-5. Composite Period Total Phosphorus with Questionable Data

3.3.2 Calculation of Integrated Loads and Runoff

Daily runoff and loads were calculated from OWML monitored flow rates and concentrations, and then aggregated on either a composite period (OWcp), or monthly (OWmn) basis.

Two important changes in sampling protocol existed within the OWML monitored database: a change from recording instantaneous or average flow rate (cfs) to a calculated total flow (cf), and a change from discrete to composite sampling. These were accommodated in the calculations as described below.

The majority of flow records were recorded as a flow rate. Therefore, whenever total flow was recorded, it was converted to an average flow rate by dividing by the time covered by the composited sample in seconds. This allowed consistency to be maintained in the calculations of runoff and pollutant loads. The OWML monitored database contained discrete and composite runoff samples and baseflow samples. Discrete runoff and baseflow samples were taken at a specific date and time. Composite runoff samples were composed of a number of flow-weighted samples taken over a period of time, with average values recorded. Composite samples were recorded with both a beginning and ending date and time. Flow-weighted runoff in mm, and pollutant loads in kg/ha, were calculated for all time intervals, Δt , between the various types of samples, and within the composite sample interval. Unit area loads in kg/ha were calculated for each nitrogen, phosphorus and sediment parameter in the OWML database as:

$$\text{load} = \left(\frac{q_1 * c_1 + q_2 * c_2}{2} \right) * \Delta t * k1 \quad (3.1)$$

$$(\text{kg/ha}) \quad (\text{cfs} * \text{mg/L}) \quad (\text{min}) \quad \left(\frac{\text{kg} / \text{ha}}{\text{cfs} * \text{mg} / \text{L} * \text{min}} \right)$$

where subscripts 1 and 2 refer to the flow rates and concentrations at the beginning and ending of a given interval, respectively. The conversion factor, k1, is calculated as:

$$k1 = \left(\frac{0.6463 \text{MG} / \text{day}}{1 \text{cfs}} \right) * \left(\frac{8.34 \text{lb} / \text{day}}{1 \text{mg} / \text{L}} \right) * \left(\frac{1 \text{kg}}{2.2046 \text{lg}} \right) * \left(\frac{1 \text{day}}{1440 \text{min}} \right) * \left(\frac{1}{\text{wsarea}} \right) = \frac{0.0001699}{\text{wsarea}} \quad (3.2)$$

where *wsarea* is the watershed area in hectares (ha). Runoff was calculated as:

$$\text{runoff} = \left(\frac{q_1 + q_2}{2} \right) * \Delta t * k2 \quad (3.3)$$

$$(\text{mm}) \quad (\text{cfs}) \quad (\text{min}) \quad \left(\frac{\text{mm}}{\text{cfs} * \text{min}} \right)$$

with the conversion factor, k2, calculated as:

$$k2 = \frac{60 \text{sec}}{\text{min}} * \frac{1 \text{ac} - \text{in}}{3630 \text{cf}} * \frac{1 \text{ha}}{2.471 \text{ac}} * \frac{25.4 \text{mm}}{\text{in}} * \frac{1}{\text{wsarea}} = \frac{0.1699}{\text{wsarea}} \quad (3.4)$$

Where no baseflow samples were taken in between two runoff events, and a time period existed with unassigned flow and concentrations, default concentrations and flow rate were used to estimate the load over that period. The following median values for each parameter were identified from all baseflow records during the 1977-92 period and used as default values: FLO = 6.30 cfs; ortho-P = 0.01 mg/L; total soluble P = 0.02 mg/L; total P = 0.03 mg/L; ammonia-N = 0.02 mg/L; total Kjeldahl N = 0.41 mg/L; combined nitrite and nitrate N = 0.16 mg/L; and total suspended solids = 2.8 mg/L.

A FORTRAN program was written to perform the calculations for each of the 5813 days included in the 191 months of the monitored period. This program was based on the algorithms above and on the rules in Table 3-1 for assignment of flow and concentration values within any given time interval. A summary of composite period runoff, total nitrogen, total phosphorus and total suspended solids at the watershed outlet is in Appendix B, while a summary of monthly runoff and loads is in Appendix C for the complete months.

Table 3-1. Rules for Calculating Monitored Runoff and Loads Between Sample Types

Beginning Sample Type *	Ending Sample Type *	Rule
B	R _d	Use values for the baseflow sample, where runoff > baseflow, else use average values.
B	R _c	
R _c	B	
B	B	Use linear interpolation between samples.
R _d	R _d	
R _d	B	Use linear interpolation for 2 days, then use baseflow values.
R _d	R _c	Use average values equal to 1/2 of beginning sample values. After 2 days use default values.
R _c	R _d	
within R _c		Use average values.

* Sample Types: B = baseflow; R_d = discrete; and R_c = composite.

3.4 Calculation of Modeled Loads

3.4.1 AGNPS 5.0 File Builder and PC-VirGIS

AGNPS 5.00 File Builder (Yagow and Shanholtz, 1995) is a GIS utility for creating input files to the Agricultural NonPoint Source (AGNPS) model versions 4.03 and 5.00 (Young et al., 1987, 1994; USDA-ARS-MWA, 1995). File Builder was developed as a component within the existing PC-VirGIS geographic information system (MapTech, Inc., 1994), developed with support from the Virginia Department of Conservation and Recreation's Division of Soil and Water Conservation (DCR-DSWC). File Builder facilitates user input of required data through a menu and prompt-driven interface. While AGNPS comes with a user-friendly spreadsheet editor interface to facilitate user input, a minimum of 22 parameters must be specified for each cell, a daunting and time-consuming task for large watersheds. By accessing GIS data layers and relational attribute tables, repetitive user input was minimized with File Builder, while spatial variability was enhanced and consistently characterized. In a cover letter to the most recent 5.00 version, the model developers acknowledged that some errors still exist in the spreadsheet editor, and recommended that the spreadsheet editor not be used for parameterizing large watersheds. While developing File Builder, trial runs revealed considerable improvement in execution time when AGNPS is run, as in File Builder, outside of the spreadsheet editor. AGNPS options, non-distributed coefficients, and model default parameters used in modeling for this study are provided in Table 3-2.

The user is required to provide a number of GIS data layers and attribute tables as input to File Builder. File Builder prompts the user to input these data layer and attribute file names, to derive

additional data layers required by AGNPS, to access and edit default model parameters, to create and run the AGNPS model, and to create user-selected output parameter data layers. A sample user response file with line-by-line explanations is shown in Table 3-3. File Builder does not include provisions for utilizing the pesticide component of AGNPS.

All programs within File Builder operate from data supplied in the user response and parameter values files. Only the program which builds output data layers from the AGNPS GIS output file requires a separate response file. This response file is built internally from user selections in the last screen of File Builder, which are not incorporated into the user response file.

3.4.2 Data Collection and Preparation

Digital spatial data and attribute data were collected from a variety of sources in various formats. The majority of data available was in a 1/9 ha cell size. The Bull Run watershed consists of 16,057 acres, which would be represented by 58,484 cells at 1/9 ha. The maximum number of cells which can be handled by the AGNPS model has not been defined precisely, though a watershed parameterized with only the basic 22 parameters per cell has been tested successfully with 16,384 cells (Witte et al., 1995). Theoretically, the model code should be able to handle a total number of cells somewhere near 32,767, the size of a 2-byte integer. The available data were aggregated in order to reduce the number of cells representing the watershed to a level which could be accommodated by AGNPS. Grid cells are square, so aggregation began with a square number of cells, which were then reduced to a single square cell equal in size to the original cells. The 1/9 ha data were grouped in squares of 2 cells x 2 cells, and aggregated to a single cell, resulting in a cell size of 4/9 ha (66.67 m x 66.67 m), and a watershed with 14,621 cells.

3.4.2.1 VirGIS Database

Digital elevation, hydrography, land use and soils spatial data and soils attribute data were obtained from the DCR-DSWC in Richmond, as developed by the Information Support Systems Laboratory at Virginia Tech. The spatial data were encoded at a cell size of 1/9 ha in run-length compressed (rlc) format, with data boundaries conforming to county boundaries. These data were developed as part of a statewide effort to develop targeting tools for agricultural NPS pollution at the local Soil and Water Conservation District (SWCD) level. VirGIS numeric soil codes are unique only within each county. Since this watershed straddles two counties, Loudoun county soil codes were used as assigned, but Prince William county soil codes were all incremented by 200 to make each soil code unique within the watershed. The assigned soil codes and soils attribute information are listed in Appendix D.

3.4.2.2 Parcel

Digital parcel data were obtained in the form of Arc/INFO export files - 18 files for the Loudoun portion of the watershed, and 12 files for the Prince William portion. The Loudoun data were exported from a UNIX platform in uncompressed ASCII format at 1600 bpi and a record length of 8000. The Prince William data were received on a 4-mm tape in UNIX tar format.

Table 3-2. AGNPS Modeling Options, Non-Distributed Coefficients and Default Parameters

<u>Parameter</u>	<u>Selected or Default</u>		<u>Appropriate</u>	<u>Units</u>
	<u>Values</u>	<u>Optional Values</u>	<u>Range</u>	
Peak Flow Indicator	0	0=TR-55; 1=AGNPS	0 / 1	
Geomorphic Option Indicator	1	0=No; 1=Yes	0 / 1	
Hydrograph Shape Indicator	0	0=%Runoff; 1=K-Coeff.	0 / 1	
%Runoff	37.50		0.00-100.00	
K-Coefficient	484.00			
Event EI	0.0	Calculated internally	0.0-1000.0	
Event Duration	6.0	Not used in File Builder	0.0-99.0	
Storm Type	II	Not used in File Builder	I/II/IIa/III	
N Conc. in Rainfall (ppm)	0.50		0.10-2.00	ppm
P factor	1.0		0.0 - 1.0	
Pesticide Ind.	0	Defaults to "0" in File Builder	0 / 1	
Soil N (lb N/lb soil)	0.0010		0.0005-0.0030	lb N/lb soil
Soil P (lb P/lb soil)	0.0005		0.0000-0.0020	lb P/lb soil
Pore water N (ppm)	5.00		0.1-25.0	ppm
Pore water P (ppm)	2.00			ppm
Runoff N Extr. Coeff.	0.050		0.01-0.40	
Runoff P Extr. Coeff.	0.025			
Leaching N Extr. Coeff.	0.250			
Leaching P Extr. Coeff.	0.250			
Soil OM (%)		5		0-100 %
N (lb N/lb soil)	0.0010		0.005-0.0030	lb N/lb soil
P (lb P/lb soil)	0.0005		0.0000-0.0020	lb P/lb soil
Diam. (in.)	6		0-99	in
Infiltration (in/hr)	0.05	high clay	0.00-20.00	in/hr
	0.40	silt		
	0.70	high sand		
	1.50	peat		
Channel width (ft)	0	Not used for TR55/Geo option		ft
Channel width coeff.	3.4250			
Channel width exp.	0.3151			
Channel depth (ft)	0	Not used for TR55/Geo option		ft
Channel depth coeff.	0.4537			
Channel depth exp.	0.2192			
Channel length coeff.	153.000			
Channel length exp.	0.600			
Channel side slope (%)	10.0		0.0-100.0	%
Channel "n"	0.040		0.001-0.990	
AGNPS decay function	1	0=no; 1=yes	0 / 1	
N % decay	0	Generated by AGNPS	0-100	%
P % decay	0	Generated by AGNPS	0-100	%
COD % decay	0	Generated by AGNPS	0-100	%
Scouring by Clay	1	0=no; 1=yes	0 / 1	
Scouring by Silt	1	0=no; 1=yes	0 / 1	
Scouring by SAGG	1	0=no; 1=yes	0 / 1	
Scouring by LAGG	1	0=no; 1=yes	0 / 1	
Scouring by Sand	1	0=no; 1=yes	0 / 1	

Table 3-3. Sample File Builder User Response File (usrrsp.urf)

Sample Data	Data Description
c:\gis\pcvmenus\agnps	sub-directory for internal File Builder files
blrn	4-digit watershed code
blrnsd4.rlc	watershed map (.rlc)
blrnfl4.rlc	elevation map (.rlc)
blrnms4c.rlc	soils map (.rlc)
blrnms.tbl	soils attribute table (.tbl)
0	% OM flag
blrnsl4.rlc	slope map (.rlc)
blrnsl4.rlc	slope length map (.rlc)
blrnbl4c.rlc	streams map (.rlc)
2 4 9	streams/pond/waterbodies codes
0	land use map (.rlc)
lu-emap4.tbl	crop/land use attribute table (.tbl)
blrnpr4c.rlc	property map (.rlc)
blrnpr77.tbl	property attribute table (.tbl)
0	feedlot/point source map (.rlc)
0	feedlot attribute table (.tbl)
blrnfd4r.rlc	AGNPS flow direction map
blrnno.rlc	AGNPS cell number map
blrnsl.rlc	AGNPS slope shape factor map
blrnchs.rlc	channel slope map
blrnchl.rlc	channel length/cell map
blrnrc.rlc	AGNPS receiving cell map
bl07777.dat	model run file name (.dat)
Bull Run Watershed	watershed name
Dissertation Runs	run description
77	rainfall date (Julian)
1977	rainfall year
2.31	daily rainfall (inches)
1	AMC
175	annual R-factor
180	no. of frost-free days
blrnvp2.tbl	default values table (.tbl)

Considerable trial-and-error was involved in identifying exact file formats and procedures needed to access the data. A series of different computers was required to access the proper tape readers, and to transfer the large individual files involved to a SUN workstation, where Arc/INFO was used to convert the Arc export files into Arc coverages. The following steps were then taken to create the raster parcel data layer:

- The first step after importing the Arc export files into Arc coverages within Arc/INFO was to project the data into the common coordinate system as used by the VirGIS data - Universal Transverse Mercator (UTM), zone 18, North American Datum (NAD)27, in units of meters.

This required knowing details about the original coordinate systems and transformations used with each data set. Both Loudoun and Prince William data sets were based on the Virginia State Plane coordinate system, North Zone (5551), in units of feet. The Loudoun data were based on the 1927 datum (NAD27), with tiles measuring 4000 feet by 6000 feet, while the Prince William data were based on the 1983 datum (NAD83) in tiles measuring 10,000 feet by 10,000 feet, along with a coordinate shift. Coordinates in the x-direction were shifted by 11,000,000 feet, and coordinates in the y-direction were shifted by 6,500,000 feet. Both sets of data were converted to the UTM coordinate system referenced above.

- The next series of steps was used to translate the Arc coverages into DLG-3 formatted files. DLG files were created from Arc polygon files and INFO attribute files. DLG attributes were coded as a series of Major and Minor codes, where Major codes indicated the category of data, such as parcel number, tax map number, and parcel sub-categories, and the Minor codes represented the numeric value of its related Major category. Up to 5 pairs of Major/Minor attribute pairs can be written for each parcel polygon. Conversion was complicated by the fact that the two counties involved used different categories and formats for their attribute data.
 - * For each parcel coverage, attribute data were extracted from INFO, transferred to the PC, edited and formatted into Major and Minor coded data sets, comma-delimited in an attribute file, and transferred back to the SUN workstation.
 - * Arc was then used to join the attribute files with individual coverages into polygon attribute table (PAT) files.
 - * Back on the PC, PC-VirGIS utilities were used to translate the corner State Plane coordinates for each Arc coverage into geographic and UTM coordinates required by the DLG format. These coordinates were then used to build header files for each of the individual DLG files corresponding to the 30 original parcel files.
 - * Finally, Arc/INFO was used to combine the polygon, attribute and header files into the desired DLG-3 formatted files.
- The DLG files were then transferred back to the PC for conversion into raster files. A PC-VirGIS utility (dlga1a2a) was used to perform internal conversion of coordinates. The DLG files were further edited to insert the correct UTM's in the DLG header which were inexplicably changed during ARC's conversion process. All files were converted to 1/9 ha raster cell size using PC-VirGIS utilities.
- A watershed parcel layer was then created by inserting the individual raster parcel files into a master file. Before combining the files, however, parcel numbers required special attention, as they are only unique within an individual county. In order to merge parcel numbers from these two adjacent counties, and to keep them unique, a number larger than any of the Prince William parcel numbers (10,000) was added to the Loudoun county parcel numbers, before combining.

- This file was covered with the USGS streams data layer to verify correct coordinate conversion. Initial conversions for Loudoun county from NAD83 to NAD27 were incorrect, and showed up as a shift in the data in the Loudoun county portion of the watershed. The Loudoun data were reprocessed with the correct conversions and verified.
- The parcel data layer was aggregated to a 4/9 ha cell size, consistent with the database for this study, using the majority rule. The 4/9 ha data layer was finally extracted to the common coordinate window and clipped to the watershed boundary.

3.4.2.3 Land Use

Two digital land use data layers were available for use with this modeling effort: the 1981 VirGIS data at a 33.33 meter cell width with 3 land use categories, and the Chesapeake Bay Program's 1990 land use (CBPLU) data at a 25.8 meter cell width with 9 categories of land use (Neumiller et al., 1994). The 1981 data were based on National High Altitude Photography (NHAP) color infrared aerial photography, and the 1990 data were based on EPA's EMAP data, as modified with NOAA's C-CAP and USGS's GIRAS databases. The 1990 data layer had been created by EPA's Chesapeake Bay Program for use with its Phase III watershed modeling. The data categories are listed in Table 3-4.

Table 3-4. Land Use Categories for Source and Study Data

Code	1981 VirGIS Categories	Code	1990 CBPLU Categories	Code	Current Study Categories
1	Non-agricultural	11	High intensity urban	12	Low intensity urban
2	Cropland	12	Low intensity urban	13	Herbaceous urban
3	Pasture land	13	Herbaceous urban	14	Woody urban
		14	Woody urban	20	Unspecified herbaceous
		20	Herbaceous	22	Cropland
		30	Woody	23	Pasture land
		40	Exposed	30	Woody
		60	Water	60	Water
		70	Herbaceous wetlands	70	Herbaceous wetlands

A window of the CBPLU data covering the study watershed was requested from EPA's Chesapeake Bay Liaison Office, and later transferred to, and downloaded from, the EPA Region III office's *ftp* site (cbpsrv1.ann.epa.gov). These data were received in a GRASS format and converted to run-length compressed (rlc) raster using a PC-VirGIS utility (grs2rlc). Raster editing was necessary along map seams to correct spurious data values. The raster data (25.8m) were then converted to vector format, re-sampled at a 33.33m cell width, and aggregated to a 66.66m (4/9 ha) cell size. The following modifications were then made to the 1990 CBPLU land use layer. Herbaceous areas (20) which coincided with 1981 cropland and pasture land areas were relabeled as categories 22 and 23, respectively, and the remaining area in category 20 was interpreted as unspecified herbaceous.

The 1981 VirGIS land use data were modified and relabeled to use the same land use categories as the CBPLU data. Cropland and pasture land were reassigned categories 22 and 23, respectively. Non-agricultural land, category 1, included all other land uses, and was distributed amongst the 1990 CBPLU non-agricultural categories as follows: all areas forested in 1990 were assumed forested in 1981, and all 1990 areas of unspecified herbaceous, water and herbaceous wetlands were assumed to have been the same in 1981. A map was made of the remaining 1981 non-agricultural land areas to define areas needing more specific interpretation. Plots of individual vector parcel boundaries were generated with corresponding raster plots of uncategorized non-agricultural areas. Location of the scattered cells was facilitated by the plots, and corresponding areas identified from NHAP photographs for manual interpretation of land use in those areas. Re-assigned categories were noted on the plots and the spatial digital layer updated using the PC-VirGIS raster editor (redit). CBPLU land use categories not shown in the study categories above were not encountered in the study watershed. A listing of crop / land use attributes used with the AGNPS model are given in Table 3-5.

Annual changes in agricultural cropping history were expected to be available through the local ASCS offices, so that annual agricultural land use could be recorded as a parcel attribute. An attempt was made to identify agricultural land use and land use changes between 1981 and 1990 through the use of county ASCS records. An evaluation of the records, however, showed that much of the agricultural acreage was not covered by an ASCS plan, and, where coverage existed, the plans only covered a couple years. Furthermore, the data which were available were listed by field, whereas the parcel layer only included farm boundaries. So, unless all the acreage was planted to the same crop or in pasture, information was lost during the assignment of annual land use to a parcel. In consultation with local NRCS conservationists, however, predominant rotations were identified by county and time period, as shown in Table 3-6.

In the absence of detailed annual land use data between 1981 and 1990, a computer program was written to interpolate land uses between the two available land use scenes from 1981 and 1990, and to extrapolate to 1977 and to 1992. If 1981 and 1990 land uses were the same for a given cell and classified as cropland, predominant rotations were assigned by county and year. Rotations prior to 1983 were assumed to have a lower level of residue management. A random starting year within the rotation was chosen on a cell-by-cell basis. All other land uses were kept constant over the 16-year period, if the 1981 and 1990 land uses were the same.

If 1981 and 1990 land uses were different, and the 1990 land use was coded as rural residential, the septic system date of installation from the septic system database (described in the next section) was used to designate the year of change from one land use to the other. Where no record of septic system installation existed, or where neither the 1981 nor 1990 land use was residential, the date of change was chosen randomly between 1982 and 1989 on a parcel by parcel basis. An annual summary of the number of grid cells by land use and year is shown in Table 3-7.

Table 3-5. Crop/Land Use Attributes Used with AGNPS Model

LUDESC	LUCODE		OVN	SCC	COD		CF	PF	N1DATE		N2DATE		PRATE		PEAK	
	CNCODE				TTYPER				N1RATE		N2RATE		SPC		HRVST	
rural residential	12	341	0.250	0.15	80	11	0.013	0	60	22	259	44	44	0.00057		0
herbaceous urban	13	712	0.250	0.22	80	11	0.006	0						0.00057		0
woody urban	14	352	0.400	0.29	65	11	0.039	0						0.00047		0
corn 112bu 30" NE	1002	13	0.100	0.05	170	5	0	0	107	25	137	60	60	0.00048	149	245
red cl-tim-oat	1011	141	0.080	0.29	20	5	0	0	107	25	137	60	60	0.00048	149	245
red cl-tim y2 regrow	1013	141	0.130	0.29	20	5	0	0	107	25	137	60	60	0.00048	149	245
red cl-tim y3 regrow	1015	141	0.200	0.29	20	5	0	0	107	25	137	60	60	0.00048	149	245
red cl-tim y3 senesc	1017	141	0.080	0.22	20	5	0	0	304				60	0.00048	365	0
wheat winter w/ legume	1023	112	0.150	0.29	80	5	0	0	259	20	61	80	30	0.00048	319	107
winter sm. gr. cover	1024	112	0.100	0.29	80	5	0	0	259	50			30	0.00048	319	107
wheat winter 45bu C	1025	112	0.080	0.29	80	1	0	0	289	20	61	80	30	0.00048	350	156
wheat w - southeast	1026	112	0.080	0.29	80	1	0	0	305	20	61	80	30	0.00048	1	172
soybeans 19" 35bu so	1027	13	0.080	0.05	117	1	0	0	171				25	0.00048	213	305
soybeans 30" 35bu so	1028	13	0.080	0.05	117	10	0	0	181				25	0.00048	218	294
unspecified agricultural	20	401	0.080	0.05	170	2	0.350	0						0.00048		0
VirGIS pasture - good	1031	353	0.130	0.39	60	11	0.010	0	223	50			80	0.00040		0
VirGIS pasture - fair	1032	211	0.080	0.22	60	11	0.060	0	223	50			80	0.00040		0
VirGIS pasture - poor	1033	212	0.060	0.15	60	11	0.100	0					80	0.00040		0
woody (forest)	30	342	0.800	0.59	65	11	0.002	0						0.00030		0
barren (quarries/pits)	40	620	0.250	0.22	115	11	0.345	0						0.00034		0
water	60	910	0.990	0.00	0	11	0.000	0						0.00000		0
herbaceous wetlands	70	924	0.990	0.00	25	11	0.000	0						0.00000		0
transportation	80	840	0.020	0.01	0	11	0.010	0						0.00000		0

Field Name	Description	Field Name	Description
LUDESC	land use description	N1DATE	date of first N application (Julian)
LUCODE	land use map index code	N1RATE	rate of first N application
CNCODE	curve number table code	N2DATE	date of second N application (Julian)
OVN	Manning's "n" for overland flow	N2RATE	rate of second N application
SCC	surface condition constant	PRATE	rate of P application
COD	chemical oxygen demand	SPC	soil phosphorus content
TTYPER	tillage type code	PEAK	peak growth date (Julian)
CF	USLE C-factor	HRVST	harvest date (Julian)
PF	USLE P-factor		

Table 3-6. Predominant Cropping Rotations Used to Create Agricultural Land Use History

County	Year	Rotation Code	Predominant Rotation
Loudoun	1977-1983	12	Corn, Corn, Small Grain(MT), Hay, Hay, Hay
Loudoun	1984-1992	13	Corn(NT), Corn(NT), Hay, Hay, Hay
Prince William	1977-1983	14	Corn(MT), Small Grain(MT), Soybeans(MT)
Prince William	1984-1992	15	Corn(NT), Small Grain(MT), Soybeans(NT)

Table 3-7. Annual Land Use Summary - Bull Run Watershed (Total No. of Cells = 14,621)

Land Use Code	No. of Cells By Land Use and Year															
	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
12	403	403	403	403	408	416	442	511	570	602	653	742	782	825	825	825
13	363	363	363	363	363	364	400	430	461	488	528	558	589	604	604	604
14	401	401	401	401	401	401	407	413	413	414	420	419	422	427	427	427
20	1353	1353	1353	1353	1353	1353	1353	1355	1359	1360	1361	1358	1358	1360	1360	1360
1201	591	765	603	665	699	0	0	0	0	0	0	0	0	0	0	0
1202	0	591	765	603	665	698	0	0	0	0	0	0	0	0	0	0
1203	699	0	591	765	603	662	0	0	0	0	0	0	0	0	0	0
1204	665	699	0	591	765	602	0	0	0	0	0	0	0	0	0	0
1205	603	665	699	0	591	765	0	0	0	0	0	0	0	0	0	0
1206	765	603	665	699	0	588	0	0	0	0	0	0	0	0	0	0
1301	0	0	0	0	0	0	602	649	583	573	763	525	541	577	461	646
1302	0	0	0	0	0	0	765	602	581	583	566	679	525	541	577	461
1303	0	0	0	0	0	0	578	765	602	544	579	560	646	525	541	577
1304	0	0	0	0	0	0	661	578	764	575	542	577	528	646	525	541
1305	0	0	0	0	0	0	662	586	573	764	533	541	577	461	646	525
1401	687	740	687	740	684	738	0	0	0	0	0	0	0	0	0	0
1402	740	687	740	687	684	684	0	0	0	0	0	0	0	0	0	0
1501	0	0	0	0	0	0	684	703	623	702	543	643	529	563	529	563
1502	0	0	0	0	0	0	716	654	703	565	698	529	611	529	563	529
2000	2086	2086	2086	2086	2086	2085	2021	1936	1869	1822	1754	1678	1596	1537	1537	1537
30	5249	5249	5249	5249	5249	5249	5314	5423	5503	5612	5664	5794	5899	6008	6008	6008
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7
70	10	10	10	10	10	10	10	10	11	11	11	11	11	11	11	11

3.4.2.4 Septic Systems

Septic system data were obtained from the Northern Virginia Planning District Commission's (NVPDC) septic system inventory of the Occoquan watershed in the form of a Final Report (NVPDC, 1990) and an R:Base relational database (Microrim Inc., 1986). R:Base was used to access the database and to extract the data relevant to Loudoun and Prince William counties. Septic system data were reported by two geographic categories, TAXMAP and LOTID, an alphanumeric combination of parcel map identifiers. In order to use these data in a spatial manner, each septic system record was related to a specific land parcel and to a representative cell within that parcel.

Initially, an attempt was made to create parcel labels by vectorizing the 4/9 ha raster parcel data layer and extracting the polygon label coordinates for each parcel. The procedure used by the conversion program to create polygon label coordinates resulted in an unnatural pattern of label points. It also became apparent that many parcel numbers were no longer represented in the parcel data layer, due to the aggregation of the original 1/9 ha data into 4/9 ha cells. Therefore, a procedure was developed to derive the parcel label coordinates from the original 30 DLG files. Individual files were edited, and parcel numbers, number of arcs, and label coordinates extracted for all parcels. These data were linked with parcel attribute data extracted earlier from INFO.

The parcel numbers, coordinate, and attribute data were then combined by county and imported into Quattro (Borland Inc., 1989) spreadsheets. In order to match parcels with septic system data, three fields of data were required for the match with Loudoun data, and four fields of data with the Prince William county data. The spreadsheets were used to prepare the parcel data for matching. The Loudoun county TAX-NUM data were in the form “*aaabbccddddeef*”, where *a* = tax map, *b* = alphabetic subdivision, *c* = numeric subdivision, *d* = not used with current data, *e* = lot identifier, and *f* = alphabetic lot subdivision. All non-value places within the field were represented by a “/”, as in the tax number 105/A/1/////9/. These numbers were parsed into three data fields, corresponding to fields *a*, *b* and *c* combined, field *d*, and fields *e* and *f* combined, and labeled MAP, SUBDIV, and LOTID. Prince William county data included MAP, SUBDIV and LOTNUM fields within its TAXMAP# attribute. The LOTNUM field was then separated into LOTID and SBLot fields using Quattro commands.

Because of the complexity of the alphanumeric numbering scheme for the various fields, it was difficult to program the sorting and ordering needed for all possible combinations. Therefore both parcel and septic system data sets were sorted manually for specific combinations of numeric and alphabetic characters in individual fields and separated into several data subsets for matching. Separate FORTRAN programs were created for matching each county’s data, which performed a match using three fields with Loudoun data, and using four fields with Prince William data. This procedure resulted in 228 matches for Loudoun County and 1033 for Prince William County. The matched data were then added back into the septic system spreadsheets, LOTS107 and LOTS153, to link the parcel numbers with septic system records. The linked parcel-septic system data were then extracted from both files and combined into another spreadsheet file (SSDBASE). The septic system attributes needed to determine septic system type and the number of systems (PARCEL, NUMSYS, YRINSTAL, and REMEDY) were extracted by parcel.

Parcel coordinate data were extracted from the files created earlier containing parcel, coordinate, and attribute data. Row and column of raster data were then calculated for each parcel label coordinate point within the common data window. The data were sorted first by row, then by column and processed through two additional FORTRAN programs. The first of these programs eliminated all parcels whose cells were outside the watershed boundary, while the second chose between two or more duplicate coordinate sets per parcel. This selection was based on the number of arcs used in the DLG file to create the partial polygon parcel, where a parcel overlaps one or more DLG file boundaries. The result was a singular set of 2051 parcel numbers with row and column coordinates within the watershed (177 for Loudoun County and 1874 for Prince William County). This list was compared with the matched parcel-septic system records from SSDBASE using the SSTBL program to see which of these records pertain to parcels within the watershed. A total of 588 parcels with septic system data were identified within the watershed (65 in Loudoun County and 479 in Prince William County). This FORTRAN program also assigned a septic system type for each parcel and linked these septic system attributes to an AGNPS watershed cell number corresponding with each parcel’s row and column coordinates. The program allowed for multiple parcels occupying the same AGNPS cell number, in order not

to lose septic system contributions. The resulting output file of septic system attributes, sorted by AGNPS cell number, is used by the MONTHLY program in calculating monthly septic system N and P pollutant loads, and is referenced in Appendix E.

3.4.3 GIS Data Layer Derivation

The basic spatial data layers of elevation, land use, surface water, and soils were further manipulated with various geographic information system routines to derive additional spatial data layers needed as input for the AGNPS model.

3.4.3.1 Watershed Delineation

A watershed delineation procedure developed at the U.S. Geological Survey (USGS) was modified to accept input and output of rlc-formatted spatial data layers (Jensen and Dominigue, 1988). Three spatial data layers were derived for use with AGNPS using the USGS procedure: a corrected elevation file, a flow directions file and a watershed boundary file. The flow directions file was relabeled to correspond with AGNPS direction codes as in Table 3-8, and is illustrated within the delineated watershed boundary in Figure 3-6.

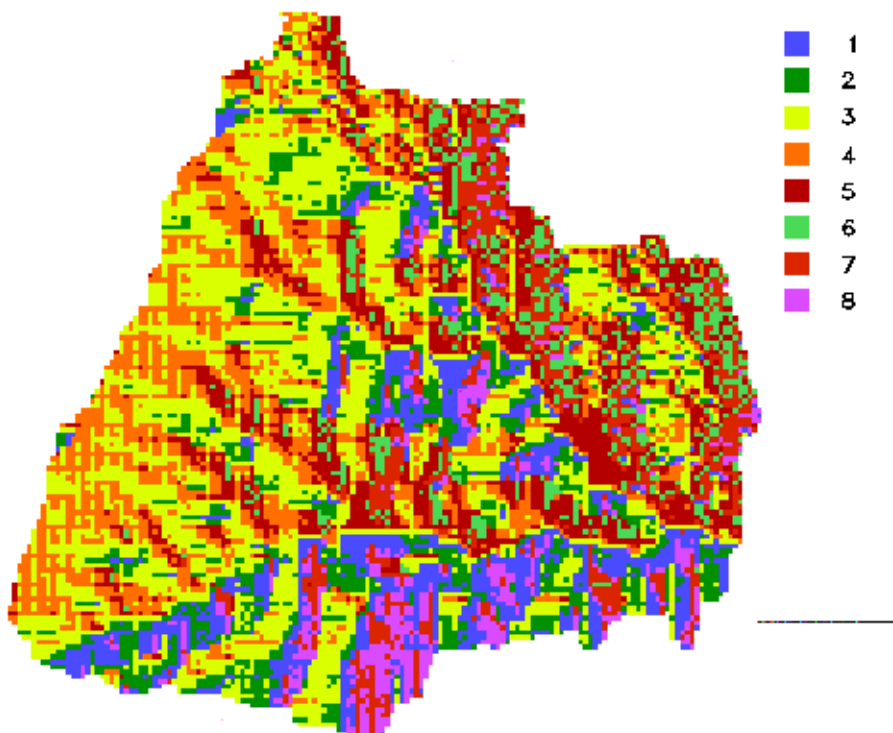


Figure 3-6. USGS Procedure-Generated Watershed Boundary and Flow Directions Map for Bull Run at Route 705

Table 3-8. USGS and AGNPS Flow Direction Codes

USGS Code	AGNPS Code	Cardinal Direction	USGS Code	AGNPS Code	Cardinal Direction
1	2	NE	16	6	SW
2	3	E	32	7	W
4	4	SE	64	8	NW
8	5	S	128	1	N

3.4.3.2 AGNPS Cell Number

This data layer was created from the identified watershed data layer by consecutively numbering all cells within the watershed from left-to-right, top-to-bottom, starting with number one. The maximum cell number allowed in this program is 31,999.

3.4.3.3 Channel Slope and Slope Length/Cell

These two data layers were both generated by the same program and derived from the cell number, elevation, and streams data layers. Channel slope, in percent, was calculated for each stream cell based on the maximum elevation difference between any two stream cells in the outer band of a 5x5 matrix, centered on the stream cell. In the headwater cells, where only one stream cell will be encountered in the outer band, the maximum difference was calculated between the stream cell and the outer cell. Channel length/cell was then calculated as the average distance/cell between the two cells used to calculate channel slope. This program also identified the watershed outlet by cell number and placed it in a file called “outlet”, for later use by the File Builder utility. This program further created a file called AGN_BLNO.RLC, a map layer of stream cells numbered with corresponding AGNPS cell numbers, for use with the monthly septic system and baseflow programs described in a later section.

3.4.3.4 Flow Receiving Cell

This data layer was derived from the cell number and the AGNPS flow direction data layers. Each cell contains the number of the cell receiving flow from the current cell. The receiving cell of the watershed outlet is outside the watershed bounded area. This receiving cell has a standard coding of 32,000 in the receiving cell data layer, but is undefined in the cell number data layer.

3.4.3.5 Slope Shape Factor

This data layer calculated the AGNPS slope shape factor (*ssf*) based on the slope between each cell and the two adjacent cells aligned with the AGNPS flow direction. Values of *ssf* were defined as: 1 = uniform, 2 = convex, and 3 = concave. Because only one adjacent cell was available to all boundary cells, boundary cells in the extraction window all have a default *ssf* value of 1.

3.4.4 Common Parameterization Methods

The following EI and AMC methods were automated identically for both of the event parameterization procedures and are included in the File Builder utilities for creating AGNPS input files.

3.4.4.1 Storm EI

An algorithm developed by Richardson et al. (1983) was used to estimate the daily erosion index (EI). This algorithm was developed as a non-linear statistical relationship between daily rainfall amount in inches, PRECIP, and daily EI, on a monthly basis. The relationship was developed at eleven locations in the eastern half of the United States, with an average of twelve years of data at each station. The relationship, converted to English units, is described by the following equation:

$$EI = 10.063 * a * PRECIP^b \quad (3.5)$$

where monthly values of “a” were used for Blacksburg, Virginia, ranging from 0.14 to 1.02, and $b = 1.59$.

3.4.4.2 Antecedent Moisture Condition

A program was written to calculate the preceding 5-day rainfall total for each day with a daily rainfall total equal to or greater than 0.1” from an input file of sequential daily precipitation. AMC categories were assigned, based on whether the storm occurred in a dormant or growing season, and corresponding categories of AMC by preceding 5-day rainfall totals. The three AMC categories are: 1 = dry; 2 = average; and 3 = wet. The program requires names for input and output files; beginning and ending frost-free dates, which define the dormant and growing seasons; and the 5-day rainfall totals used to separate AMC into 3 classes for both dormant and growing seasons. The AMCs calculated for each modeled event are included in Appendix A. The frost-free dates were obtained from summaries in the Handbook for Agronomy (VCES, 1987). The AMC categories used for this calculation were defined as the 5-day antecedent rainfall breakpoints used in SCS’s TR-55 procedure (USDA-SCS, 1986).

3.4.5 Typical Parameterization

The following variations were programmed into AGNPS File Builder procedures to simulate typical parameterization. In this procedure, spatial variability was maintained, while average annual values were used for the selected parameters. Agricultural land use was modeled as the predominant crop (LUCODE) in each rotation-year (ROTYR) combination with a fixed average annual C-factor. The predominant crop code is listed by rotation year in Table 3-9, and the annual average C-factor by crop code is listed in Table 3-10. The annual K-factor was modeled by soil type as listed in Appendix D. Fertilizer rates were the applied rates as listed in Table 3-5 by LUCODE. Curve numbers (CN) were assigned by HSG and AMC using the same CNCODEs listed by LUCODE in Table 3-5.

Table 3-9. Predominant Crop Code (LUCODE) by Rotation-Year (ROTYR)

ROTYR	LUCODE	ROTYR	LUCODE	ROTYR	LUCODE
1201	1015	1301	1002	1401	1002
1202	1015	1302	1002	1402	1027
1203	1005	1303	1011	1501	1007
1204	1005	1304	1013	1502	1026
1205	1011	1305	1017	2000	1032
1206	1013				

Table 3-10. Average Annual C-Factors (CF) By Agricultural Crop Type (LUCODE)

LUCODE	CF	LUCODE	CF
1002	0.236	1015	0.004
1005	0.212	1017	0.010
1007	0.214	1026	0.064
1011	0.072	1027	0.085
1013	0.005	1032	0.080

3.4.6 Time-Variable Parameterization

The time-variable parameterization procedure incorporated the following methods within the File Builder utilities for creating AGNPS input files.

3.4.6.1 Seasonal Nitrogen Fertilizer Availability

Nitrogen fertilizer applications are allowed at either one or two dates throughout a calendar year for each crop through File Builder. Phosphorus application was assumed to take place after primary tillage, and the phosphorus to be instantly bound to the soil matrix. The nitrogen fertilizer availability rate was calculated for a specified storm date by incorporating a decay function to simulate nutrient uptake by the crop(s) during the growing season, after an initial germination period of non-growth. This availability rate is separate from the initial fertilizer availability, which is a function of incorporation by tillage operation. The algorithms were based on the following assumptions. If the first fertilizer application was in the spring, crop uptake was assumed to be linear, after emergence, to the total applied N fertilizer, whether applied in a single or split applications. If the first fertilizer application was in the fall, crop uptake was assumed to be linear, after emergence, to each application amount individually, with a dormancy period in between. Fall fertilization was defined as any fertilization which takes place after August 1st. The dormancy period runs from December 15th to the spring application of fertilizer. Graphically, the two relationships are illustrated in Figures 3-7 and 3-8.

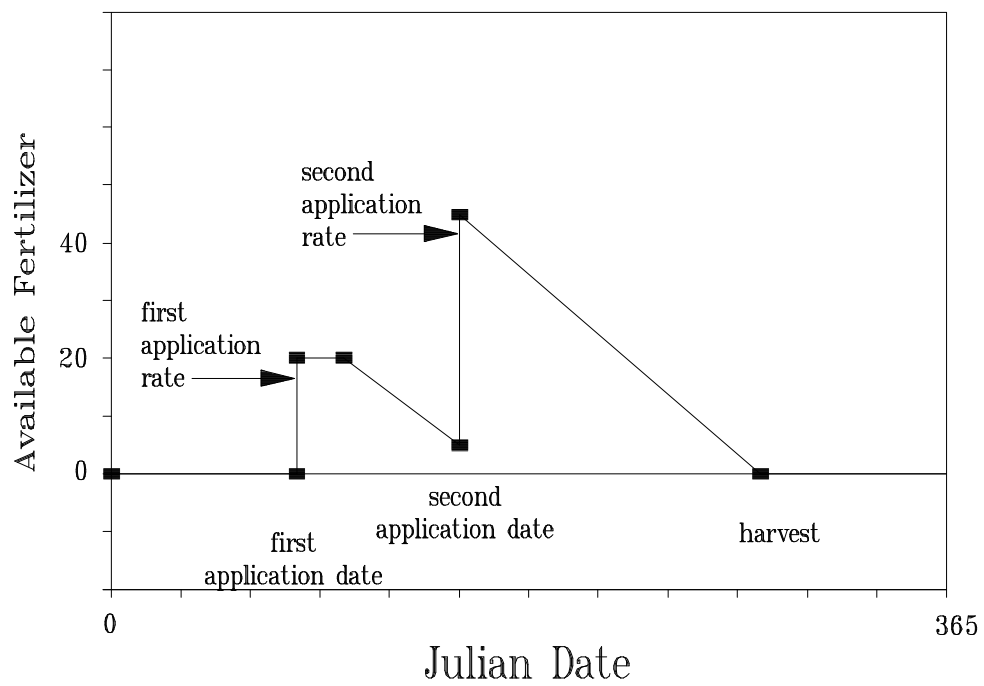


Figure 3-7. Seasonally-Variable Fertilizer Availability with First Application in the Spring

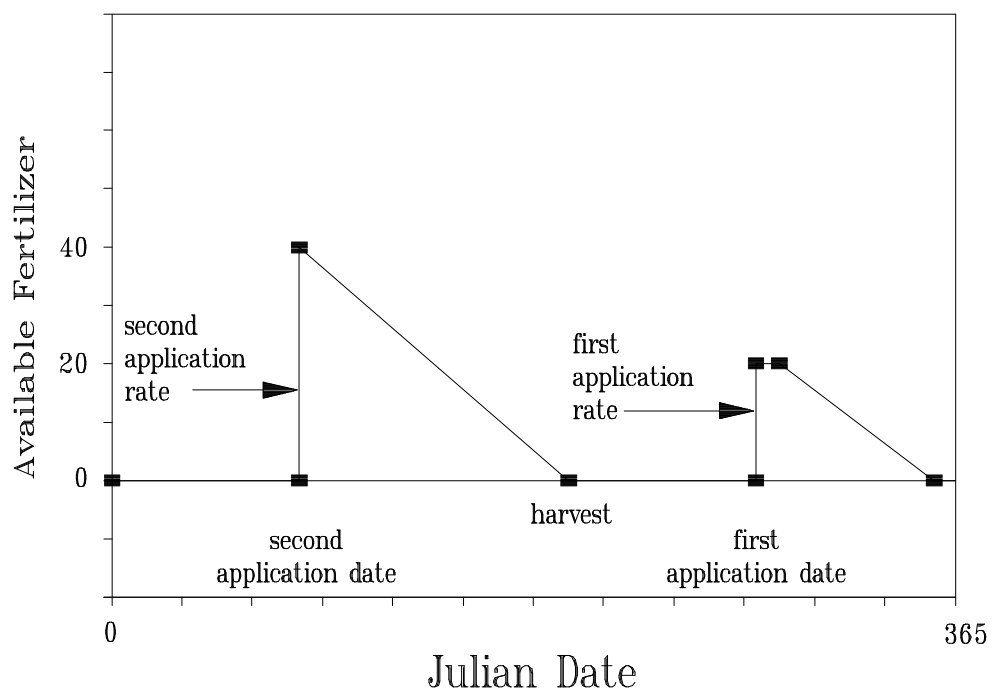


Figure 3-8. Seasonally-Variable Fertilizer Availability with First Application in the Fall

3.4.6.2 Seasonally-Variable C-Factor

Variations in surface cover for agricultural land uses were modeled by calculating bi-weekly C-factors using the RUSLE model (USDA-ARS/SCS, 1989). Typical crop rotations, management practices, field operations and associated dates were identified for the watershed with assistance from local FSA, SWCD and NRCS field personnel. All crop codes and associated beginning dates constituting each crop rotation/year combination are given in Table 3-11. Individual crop codes are referenced in the crop / land use table, Table 3-5, and crop rotations are listed in Table 3-6. The cropping data were input into RUSLE and a listing of bi-weekly C-factors was generated for each year within each rotation. This output was formatted for use with the AGNPS File Builder utilities, and is listed in Appendix F.

Table 3-11. Crop Rotation Table: Crop Codes and Associated Beginning Julian Dates by Rotation and Year within Rotation

ROTYR	PF	Crop0	JD1	Crop1	JD2	Crop2	JD3	Crop3	JD4	Crop4
1201		1013	61	1015						0
1202		1015	61	1015						0
1203		1015	106	1005	259	1024				0
1204		1024	106	1005	259	1023				0
1205		1023	61	1011						0
1206		1011	61	1013						0
1301		1017	116	1002						0
1302		1002	116	1002	306	1011				0
1303		1011								0
1304		1011	61	1013						0
1305		1013	61	1017						0
1401		1027	111	1002	289	1025				0
1402		1025	167	1027						0
1501		1028	121	1007	306	1026				0
1502		1026	182	1028						0

ROTYR = rotation/year combination, where the first two digits represent the rotation code, and the second two digits represent the year within the rotation,

PF = the USLE supporting practice P-factor,

Crop0 = the crop code for the crop in the field at the beginning of the year,

JDx = the Julian date when a new crop is planted, x is the number of crop changes in any given year, and

CropX = the corresponding crop code for each successive new crop.

3.4.6.3 Seasonally-Variable K-Factor

File Builder incorporated the time-variant K-factor calculations developed for the RUSLE model for conditions where monthly average temperatures are greater than 27°F. These algorithms require the USLE annual R-factor (R), the annual K-factor (KF), the number of frost-free days (FF), and the storm date given as a Julian date. The following equations define TMAX, the time of year at which the soil erodibility factor is at a maximum and assumed to coincide with the beginning of the growing season, and TMIN, the time of year at which the soil erodibility factor is at a minimum (the end of the growing season), both expressed as a Julian date.

$$TMAX = 214 - 0.44 * R \quad (3.6)$$

$$\text{If } TMAX \geq 60: \quad TMAX = TMAX - 60, \quad (3.7)$$

$$\text{and if } TMAX < 60: \quad TMAX = TMAX + 60. \quad (3.8)$$

$$TMIN = TMAX + FF \quad (3.9)$$

The minimum K-factor, KMIN, and the maximum K-factor, KMAX are defined as follows:

$$KMAX = (3.0 - 0.005 * R) * KF \quad (3.10)$$

$$KMIN = KMAX / (8.6 - 0.019 * R) \quad (3.11)$$

And the seasonally variable K-factor, SVK, is defined as follows for the appropriate season:

$$\text{Growing Season: } SVK = KMAX * \left(\frac{KMIN}{KMAX} \right)^{(a/FF)} \quad (3.12)$$

$$\text{Dormant Season: } SVK = KMIN * e^{(0.009*b)} \quad (3.13)$$

where a = the number of days since the beginning of the growing season, and
b = the number of days since the end of the growing season.

3.4.6.4 Seasonally-Variable Curve Numbers (CN) for Cropland

Seasonally variable CNs were calculated for cropland, based on the following relationships (USDA-SCS, 1986):

$$CN_{peak_growth} = 2 * CN_{ave} - CN_{fallow} \quad (3.14)$$

This relationship built on stages of growth for individual crops. Two fallow conditions were defined for this procedure: one applied to crops using residue management, while the other applied to crops using clean tillage practices. Figure 3-9 illustrates how CN varied for the two different fallow conditions throughout the year, based on dates of plowing, planting, peak growth, and harvest for an individual crop.

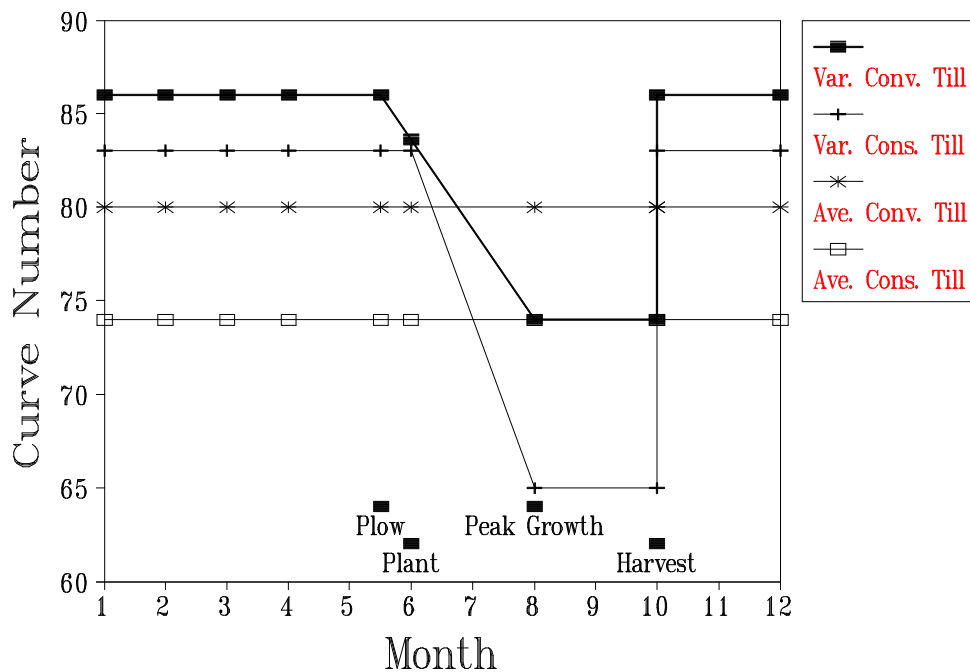


Figure 3-9. Example Seasonally-Variable vs. Average Annual Curve Numbers for Cropland by Tillage Type

Seasonal CN variation for pasture was based on the hydrologic condition of the field: poor, fair, or good. Pasture condition was assumed to vary from season to season as follows: Poor - July through October; Fair - January through March; and Good - all the rest. The Handbook of Agronomy (VCES, 1987) was used to estimate crop harvest dates, along with the RUSLE crop database obtained from the State NRCS office. Peak growth rates were assumed to occur at a fixed number of days after planting based on conversations with a state extension specialist (Brann, 1995) as follows: corn (42), soybeans (30), and wheat and grasses (60).

3.4.6.5 Impervious Area Buildup / Washoff

As development of the modeling procedures progressed, impervious areas were to be simulated using a set of buildup / washoff functions, using the AGNPS “Non-Feedlot Point Source” (NFPS) parameters. Testing of this option revealed a flaw in the operation of this function. Instead of NFPS flow rates and concentrations being added to a single cell as intended, they were also added subsequently to each downstream cell from the original cell, in an additive fashion, so that each NFPS input produced incredibly large downstream nutrient loads. Since no consideration was given to changing AGNPS source code and recompiling, this option for incorporating impervious loads had to be abandoned. Although curve numbers were modified for impervious areas to reflect increased runoff from these areas, the simulation of imperviousness was much weaker without the ability to simulate the increased loading from dry deposition in these areas.

3.4.7 Monthly Simulation Procedures

Both of the monthly simulation procedures were incorporated within the storm sequencing procedure described in the next section to maintain consistency and reduce manual input in changes needed in parameterization between storms. The first monthly simulation procedure basically aggregated spatial data layer output from all storms within any given month. The second monthly simulation procedure added monthly baseflow and monthly modeled septic system loads using the following methods. These two procedures were run back-to-back as the second procedure builds on the first.

3.4.7.1 Septic Systems

Septic system loads of nitrogen and phosphorus were generated on a monthly basis as a function of population and system type. Septic system records were obtained from a 1990 septic system inventory conducted by the NVPDC (1990) and linked with tax parcels. The procedure of Mandel (1993) was used to classify septic system types and to calculate loads for each type. Monthly groundwater discharge distribution was assumed to be equal to the long-term monthly stream discharge distribution. Groundwater discharge for individual months was then calculated as a fraction of annual average discharge. These fractional values are shown in Table 3-12. Values for daily nutrient loads in septic tank effluent and for daily plant uptake during the growing season were taken from the GWLF model (Haith et al., 1992).

Table 3-12. Fractional Monthly Groundwater Discharge Based on Surface Runoff Distribution

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
.053	.069	.109	.108	.143	.126	.121	.069	.111	.013	.034	.043

Mandel defines four septic system types: normal, short-circuited, ponded, and direct discharge. Normal systems are those which have been properly sized and located, with adequate drain fields and proper maintenance. Systems were considered normal in the absence of data to the contrary. Normal systems (type 1) were assumed to contribute no phosphorus loads to streams, while nitrogen loads were calculated as:

$$N_i = 0.001 * a * GR_m * (e_i - u_{ij}) * 365 \quad (3.15)$$

where e_i = per capita daily nutrient load in septic tank effluent for nutrient i , (g/day),

where nutrient 1 = nitrogen and 2 = phosphorus,

u_{ij} = daily nutrient uptake by plants for nutrient i in season j , (g/day),

where seasons are defined as 1=dormant and 2=growing,

a = population served,

GR_m = monthly fraction of annual groundwater discharge, and

N_i = nutrient i load in kg.

Short-circuited systems (type 2) were defined as systems located within 15 m of streams so that negligible adsorption takes place. With the chosen cell size of 66.66 meters for this study, any system coincident with stream cells was classified as short-circuited. Loads for short-circuited systems were calculated as:

$$N_i = 0.001 * a * d_m * (e_i - u_{ij}) \quad (3.16)$$

for both nitrogen and phosphorus, where d_m is days in month m . Poned systems (type 3) were assumed to lack sufficient infiltration capacity during frozen conditions. If the septic systems were installed prior to 1970, were located on soils with a hydrologic soil group D, or had a history of reported failures, they were classified as type 3. Loads for ponded systems were represented by:

$$N_i = 0.001 * a * d_m * e_i \quad (3.17)$$

Direct discharge systems (type 4) were defined as populated sites without septic systems. Because the Occoquan watershed has been closely monitored as a water supply for a major metropolitan area, no direct discharge systems were encountered.

An average household population of three persons was assigned to each septic system. According to Mandel's procedure, septic system loads were estimated as monthly loads delivered to the closest downstream channel cell from each site. A new data layer was created using the watershed delineation procedure, which defined sub-watersheds for each stream channel cell, simplifying the summation of septic N and P loads by stream cell.

3.4.7.2 Baseflow

Monthly baseflow was calculated for each stream cell based on its direct contributing surface drainage area, and a monthly unit area groundwater (UAG) flow in cfs/sq.mi. (cfsm) as obtained from USGS Water Year data in STORET and from baseflow data gathered by the Occoquan Watershed Monitoring Lab (OWML). The USGS daily flow record for the ST60 station ended in March 1987. OWML data were used to extend the flow record through 1992. Monthly baseflow was calculated for each month and year within the period of record. Monthly baseflow was calculated from USGS data as the average of monthly minimum and mean daily flows, and from OWML data as the average of sampled weekly baseflows, with a default median baseflow value used for non-sampled weeks. Values for months with no recorded monthly samples were estimated as the average of baseflow from the preceding and following months in the current year, and between the same month in the preceding and following year. Annual mean UAG for the years 1977-1992, converted into watershed-inches, ranged from 0.151 to 1.567, while monthly means ranged from 0.098 to 1.042, with an overall monthly mean of 0.546 inches. Data for individual months are given in Table 3-13. Associated baseflow loads were calculated based on the area of each hydrologic soil group (HSG) within each stream cell's direct contributing surface drainage area. The nitrate concentration for a drainage area was calculated as the area-weighted concentration by HSG, where national median nitrate groundwater concentrations were derived by HSG as follows (Mueller et al., 1995):

HSG:	A	B	C	D
[NO3-N] mg/L:	4.7	3.7	2.2	0.17

No categorical differences were identified for total phosphorus, so the median in the USGS national retrospective data set, 0.05 mg/L, was used as a constant concentration.

Table 3-13. Monthly Unit Area Groundwater Runoff Based on the Monthly Average of Annual USGS Mean and Minimum Daily Flows (inches)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1977	0.130	0.224	1.032	0.733	0.127	0.059	0.018	0.050	0.016	0.117	1.114	1.966
1978	2.788	0.727	1.787	0.566	1.390	0.145	0.156	1.153	0.089	0.042	0.123	0.715
1979	3.128	1.897	1.765	0.995	1.095	1.866	0.295	0.900	1.983	3.173	1.059	0.648
1980	1.408	0.466	1.915	1.535	1.005	0.132	0.117	0.017	0.003	0.016	0.122	0.141
1981	0.063	0.775	0.221	0.259	0.474	0.141	0.114	0.034	0.056	0.041	0.054	0.112
1982	0.391	2.381	1.207	0.811	0.382	1.477	0.071	0.346	0.029	0.069	0.489	0.713
1983	0.454	1.594	2.838	4.476	1.050	0.625	0.049	0.006	0.002	0.339	1.373	2.547
1984	0.849	2.550	3.173	2.551	1.519	0.182	0.059	1.978	0.154	0.109	0.402	0.532
1985	0.407	1.554	0.465	0.225	0.139	0.059	0.002	0.005	0.011	0.250	1.274	0.534
1986	0.313	1.312	1.070	0.709	0.141	0.013	0.000	0.147	0.026	0.001	0.069	1.186
1987	0.108	0.087	0.127	0.476	0.158	0.368	0.020	0.000	0.017	0.246	0.439	0.065
1988	0.207	0.156	0.155	0.380	0.431	0.123	0.078	0.064	0.041	0.048	0.174	0.185
1989	0.087	0.105	0.087	0.389	0.297	0.296	0.212	0.065	0.029	0.157	0.736	0.472
1990	0.080	0.232	0.252	0.473	0.150	0.166	0.320	0.110	0.088	0.176	0.258	0.202
1991	0.555	0.143	0.279	0.400	0.102	0.032	0.034	0.000	0.000	0.000	0.039	0.226
1992	0.085	0.078	0.301	0.267	0.581	0.136	0.028	0.155	0.119	0.219	0.329	0.202
AVG:	0.691	0.892	1.042	0.953	0.565	0.364	0.098	0.314	0.166	0.313	0.503	0.653
STD:	0.896	0.813	0.937	1.043	0.456	0.504	0.094	0.524	0.457	0.723	0.434	0.659

3.4.7.3 In-Stream Decay of Nutrients

Both the baseflow and septic system load estimation procedures produce loads to specific stream channel cells. To be consistent with AGNPS procedures used with overland runoff loads, the septic system loads of dissolved pollutants were allowed to decay as they flowed toward the watershed outlet. The following geomorphic and TR-55 (USDA-SCS,1986) relationships used in AGNPS 5.0 were used to route these monthly loads. For TR-55 calculations, a rectangular channel shape was assumed with the following maximum dimensions:

$$max_depth = 0.4537 * area^{0.2192} \quad (3.18)$$

$$width = 3.4250 * area^{0.3151} \quad (3.19)$$

where *area* is total watershed area draining into any given stream cell in acres. *Depth* and *width* are in feet. The average depth of flow was based on Manning's equation as:

$$depth = \left(\frac{runoff * mann}{width * 1.49 * \left(\frac{chs}{100} \right)^{0.5}} \right)^{0.6} \quad (3.20)$$

where *runoff* is in cfs, *mann* = Manning's roughness coefficient, and *chs* is channel slope in percent. The velocity (*vel*) in each channel segment then becomes:

$$vel = \frac{runoff}{width * depth} \quad (3.21)$$

The travel time (*time*) through the channel segment, in hours, is:

$$time = \frac{channel_length}{vel * 3600} \quad (3.22)$$

First-order decay functions were used instead of the AGNPS decay functions, because the AGNPS algorithms were not directly convertible to the procedure used. The first-order nitrogen decay function was based on a conservative decay rate of 1.5 %/day (0.0625 %/hr), while the phosphorus decay coefficient was set using the same proportion with nitrogen as used in the AGNPS functions. The functions used were:

$$N_dk = 0.0625 * time \quad (3.23)$$

$$P_dk = 0.0446 * time \quad (3.24)$$

with a maximum decay of 75% for any channel segment. The load exiting any channel segment was then calculated as:

$$nlo = nin * (1.0 - N_dk) \quad (3.25)$$

$$plo = pin * (1.0 - P_dk) \quad (3.26)$$

where *nin* and *pin* represent the combined N and P loads, respectively, from upstream cell(s) and from septic systems and baseflow loadings to the current cell. Processing proceeded from each primary stream cell (the upstream end stream cells with no inflow from other stream cells) to its receiving stream cell, and continued to successive receiving cells until a stream branch was encountered. If the encountered branch had not yet been processed, the program would then search for the next primary stream cell. If processing had already occurred on a branch, its load

would be added to that of the current cell, and processing continued to the next branch. Processing terminated when all branches were processed and the watershed outlet reached.

3.4.8 Storm Sequencing Procedure for Monthly Input/Output

The storm sequencing procedure encompassed both variations of event parameterization, the File Builder utilities for creating event input files for AGNPS, and both monthly simulation procedures. The storm sequencing programs operated from information in the File Builder response file, 'usrrsp.urf', and a list of daily storms. The sequencing procedure used component programs within File Builder, rather than using the PC-VirGIS interface to automate and expedite runs of hundreds of storms at a time. These programs process one storm, then loop and look for the next sequential storm, and check to see if it is in the same month. A diagram of the steps involved is shown in Figure 3-10.

The sequencing procedure updated the 16 response files needed to process each storm and/or month. The event parameters updated in each loop included date, daily rainfall amount, AMC, C-factor, K-factor, EI, available fertilizer, and CN. When a new year was encountered, the annual land use for all cells was updated. After all event parameters and response files were updated, File Builder programs were accessed to build the AGNPS input file for a given event, and the AGNPS model was run. After each AGNPS model run, the runoff and load data layers were extracted from the AGNPS GIS file, and accumulated with other runoff and loads from the same month. At the end of each month, septic system loads and baseflow were calculated for the month and added to the total runoff and loads for the month. The nonpoint source pollution index was then calculated based on the monthly totals. The program then looped back to the start and looked for the next sequential storm until the last storm was processed. After all storms in the list were modeled, a list of output parameters at the watershed outlet were extracted from the individual data layers summarized by storm and by month. Daily storm loads were subsequently aggregated to correspond with monitored composite sample intervals.

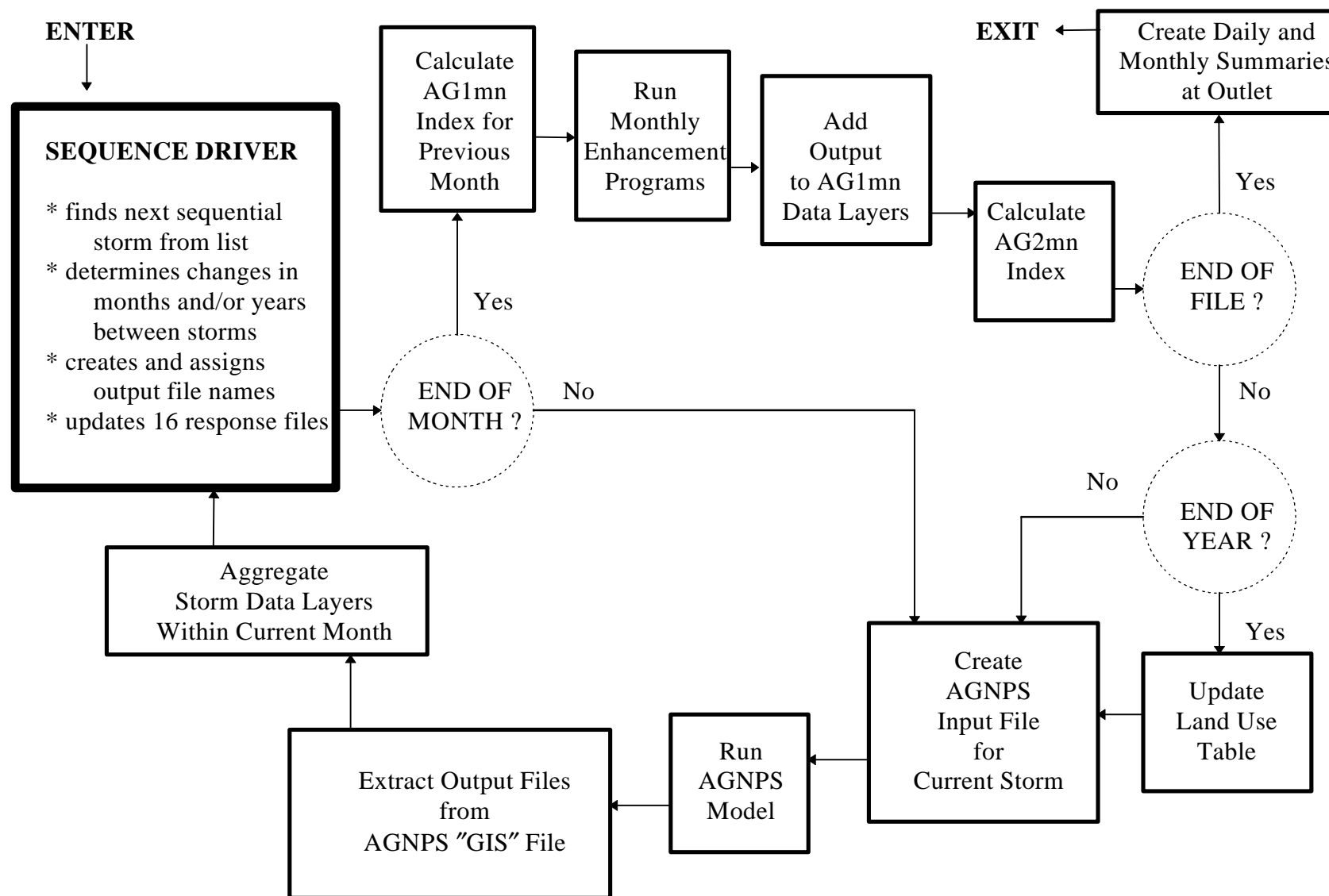


Figure 3-10. Storm Sequencing Procedure using the AGNPS Model

3.5 Data Distribution

The Ryan-Joiner normality test, a variant of the Shapiro-Wilk test within the statistical software package MINITAB (Minitab Inc., 1996), was used to test the normality of the composite period runoff and loads. The resultant probability that the composite period data were normally distributed was < 0.01 for each of the observed composite period data sets - runoff, total nitrogen, total phosphorus and suspended sediment. The probability plots for each parameter were similar to the runoff normal probability plot shown in Figure 3-11.

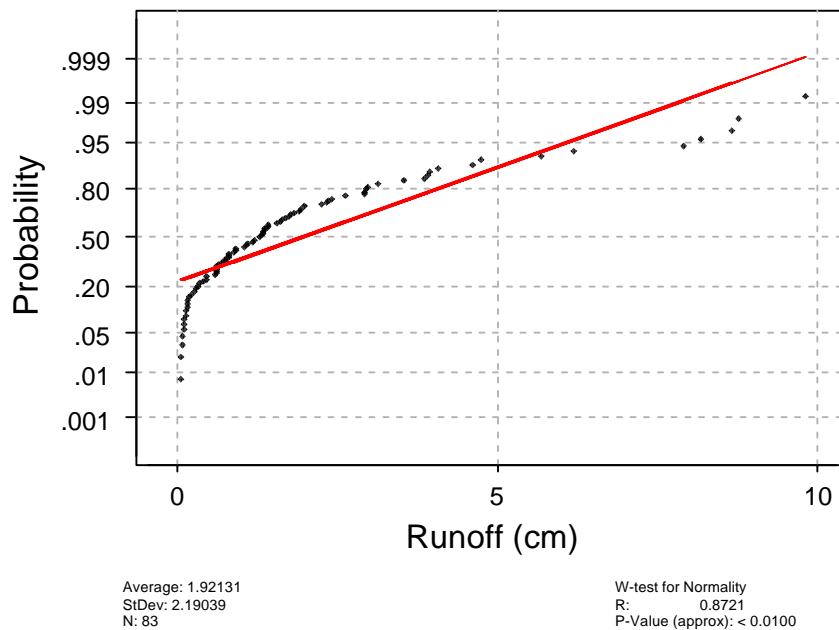


Figure 3-11. Normal Probability Plot for Runoff Calculated from Monitored Flow

Data distribution and summary statistics are provided within Figures 3-12 and 3-13 for composite period runoff and total nitrogen, respectively. These figures include distribution plots of the data, the mean, order statistics based on the median, and confidence intervals for the mean and the median.

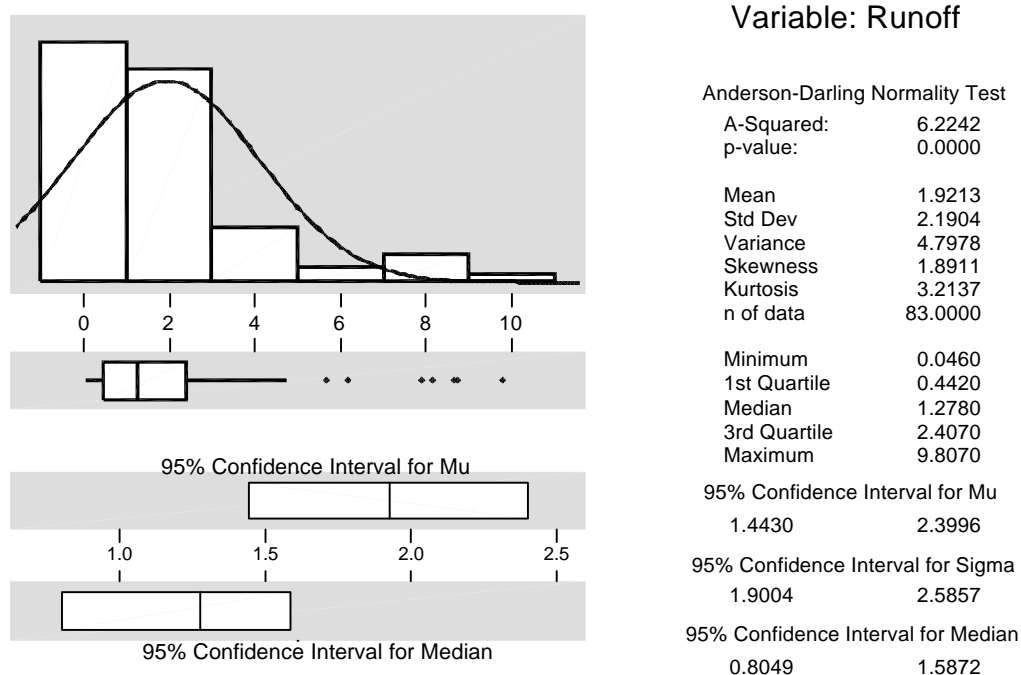


Figure 3-12. Data Distribution and Summary Statistics for OWcp Runoff (cm)

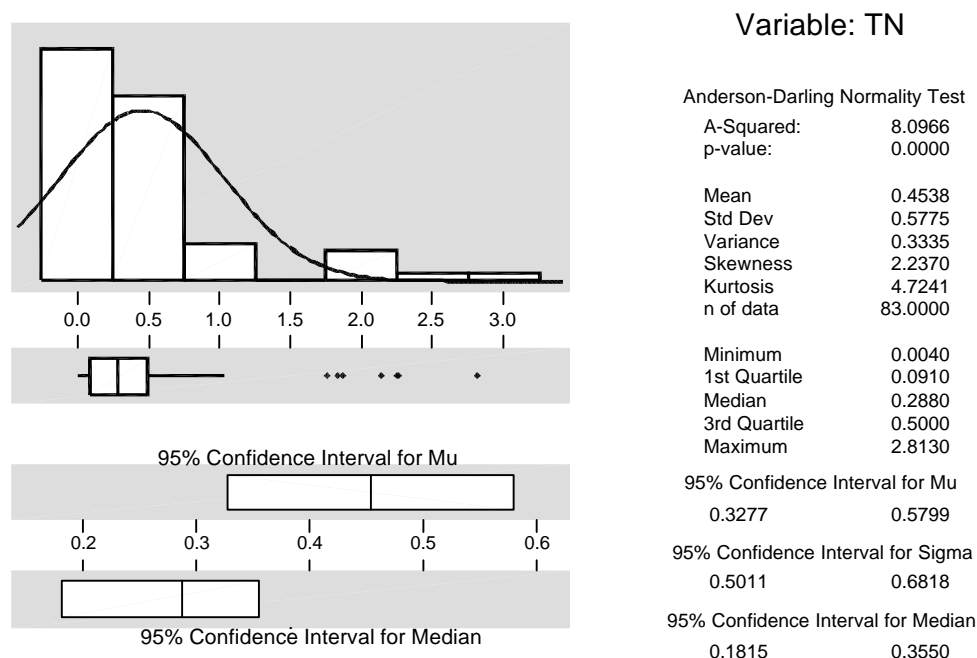


Figure 3-13. Data Distribution and Summary Statistics for OWcp Total Nitrogen (kg/ha)

These plots and associated statistics indicate that the monitoring data are not normally distributed. Reckhow et al. (1990) would recommend transforming the data to a normal distribution, and then proceeding with parametric analysis. They recommend a logarithmic transform as being appropriate for most water quality data. The monthly data set from this study, however, only contained 23 observations, which was not large enough for meaningful use with parametric analysis. Nonparametric analysis, therefore, was required with the monthly data set. The composite data set was not limited by the number of observations, and could have been evaluated with parametric analysis. However, in order to keep the analyses comparable between the data sets for the two time periods, the same type of analysis, nonparametric, was used for model evaluation of all data sets. Nonparametric analysis had the added advantage of providing numbers, rather than logarithms, for comparison with customary values and expected ranges. The MINITAB statistical software was also used for the nonparametric analysis.

3.6 Model Evaluation

A wide range of measures was included to evaluate the alternative procedures used for model input preparation with the AGNPS model. The chosen measures were intended for use not only for evaluation of the alternative procedures for this study, but also for comparison with future model validation studies, and to assist in development of a future model validation protocol. The evaluation measures included correlation, graphs, goodness-of-fit measures and hypothesis tests.

Correlation was performed with the non-parametric Spearman's rank correlation coefficient. Graphs included standard plots of observed vs. predicted loads, plots of cumulative loads, and sequential plots of monthly paired differences for each of the monthly procedures. Hypothesis tests were performed on paired observations using the Wilcoxon signed rank test.

The following nonparametric goodness-of-fit measures were calculated based on the median: the normalized median absolute error (MdAE), the robust coefficient of determination (CD*), and robust modeling efficiency (EF*). These are defined as follows:

$$\text{MdAE} = \text{median} |O_i - P_i|_{i=1}^n * \left(\frac{100}{\tilde{O}} \right) \quad (3.27)$$

$$\text{CD}^* = \frac{\text{median} |O_i - \tilde{O}|_{i=1}^n}{\text{median} |P_i - \tilde{O}|_{i=1}^n} \quad (3.28)$$

$$\text{EF}^* = \frac{\text{median} |O_i - \tilde{O}|_{i=1}^n - \text{median} |O_i - P_i|_{i=1}^n}{\text{median} |O_i - \tilde{O}|_{i=1}^n} \quad (3.29)$$