

Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

Open-File Report 2007-1040

Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

By Gerard McMahon, Larinda Tervelt, and William Donehoo

Open-File Report 2007–1040

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2007

For product and ordering information:

World Wide Web: <http://www.usgs.gov/pubprod>
Telephone: 1-888-ASK-USGS

For more information on the USGS--the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:

World Wide Web: <http://www.usgs.gov>
Telephone: 1-888-ASK-USGS

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

McMahon, Gerard, Tervelt, Larinda, and Donehoo, William, 2007, Methods for estimating annual wastewater nutrient loads in the southeastern United States: U.S. Geological Survey Open-File Report 2007-1040, 81 p.

Contents

Abstract.....	1
Introduction.....	1
Purpose and Scope	3
Use of Point-Source Nutrient Load Estimates in SPARROW.....	3
Challenges in Developing Point-Source Nutrient Load Estimates	3
Acknowledgments	4
Methods of Investigation.....	4
Developing a Site File by Using the Permit Compliance System and State Databases.....	4
Retrieving Monitoring Data from the Permit Compliance System and State Databases	6
Importing Data into a Common Database.....	6
Preparing Data for Load Estimation.....	6
Estimating Annual Nutrient Load in Wastewater Discharge.....	8
Quality Assuring Estimates of Nutrient Load in Wastewater Discharges.....	9
Summary	11
References Cited.....	12
Attachment 1: PCS Data Retrieval Routine.....	15
Attachment 2: PCS Data Import Program	17
Attachment 3: North Carolina Data Import Program	27
Attachment 4: Data Preparation Program	39
Attachment 5: Typical Pollutant-Concentration Program	51
Attachment 6: Load Calculation Program	57
Attachment 7: Load Summary Program	65
Attachment 8: Flow Magnitude Checking Program	79

Figures

1. Map showing Major River Basin-2 study area and point-source dischargers.....	2
2–4. Graphs showing—	
2. Summary of monthly effluent flow observations, by wastewater-flow classification and state, 2000–2002	7
3. Summary, by state, of sources of total nitrogen concentration information used in load calculations.....	8
4. Percentage of dischargers using load calculation approaches based on data available for all 12 months, 3 or 4 quarters, or less than 3 quarters.....	10

Tables

1. Standard Industrial Classification codes.....	5
2. Point-source wastewater dischargers by state and flow classification, 2000	9
3. Total nitrogen and total phosphorus loads from National Pollutant Discharge Elimination System permitted dischargers, by state and flow classification, 2000.....	9
4. Annual total nitrogen and total phosphorus loads from point-source wastewater load dischargers in the southeastern United States, by Standard Industrial Classification (SIC) code, for 2000.....	11

Conversion Factors

Multiply	By	To obtain
Length		
mile (mi)	1.609	kilometer (km)
Flow rate		
million gallons per day (MGD)	0.04381	cubic meters per second (m ³ /s)

Note: The abbreviation MGD will be used in this report as an abbreviation for “million gallons per day” for consistency between the text in the body of the report and language used in the programs included in the attachments.

Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

By Gerard McMahon¹, Larinda Tervelt², and William Donehoo²

Abstract

This report describes an approach for estimating annual total nitrogen and total phosphorus loads from point-source dischargers in the southeastern United States. Nutrient load estimates for 2002 were used in the calibration and application of a regional nutrient model, referred to as the SPARROW (SPAtially Referenced Regression On Watershed attributes) watershed model. Loads from dischargers permitted under the National Pollutant Discharge Elimination System were calculated using data from the U.S. Environmental Protection Agency Permit Compliance System database and individual state databases. Site information from both state and U.S. Environmental Protection Agency databases, including latitude and longitude and monitored effluent data, was compiled into a project database. For sites with a complete effluent-monitoring record, effluent-flow and nutrient-concentration data were used to develop estimates of annual point-source nitrogen and phosphorus loads. When flow data were available but nutrient-concentration data were missing or incomplete, typical pollutant-concentration values of total nitrogen and total phosphorus were used to estimate load. In developing typical pollutant-concentration values, the major factors assumed to influence wastewater nutrient-concentration variability were the size of the discharger (the amount of flow), the season during which discharge occurred, and the Standard Industrial Classification code of the discharger. One insight gained from this study is that in order to gain access to flow, concentration, and location data, close communication and collaboration are required with the agencies that collect and manage the data. In addition, the accuracy and usefulness of the load estimates depend on the willingness of the states and the U.S. Environmental Protection Agency to provide guidance and review for at least a subset of the load estimates that may be problematic.

Introduction

One of the greatest challenges in developing regional-scale nutrient water-quality models is the estimation of nutrient inputs used in the model calibration and application (McMahon and others, 2003). Nonpoint-source nutrient input estimates can be mass-based (for example, total annual nutrient mass associated with fertilizers or atmospheric deposition) or area-based (for example, areas of agricultural or urban land). Nonpoint-nutrient input estimates typically are developed for a spatial area of interest, such as a stream reach catchment, a watershed, a county, or for some larger regional unit. Point-source nutrient input estimates, expressed as annual nutrient loads, depend on knowledge of effluent discharge, or flow, and nutrient-concentration data from point-source dischargers in the study area. In the case of both point and non-point sources of nutrients, investigators may find it difficult to develop input data estimates that are accurate, that cover the entire study area, and that are compiled using comparable methods across an entire regional study area.

U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program and U.S. Environmental Protection Agency (USEPA) researchers have estimated annual total nitrogen (TN) and total phosphorus (TP) loads associated with approximately 3,000 point-source wastewater dischargers in the southeastern United States. These nutrient load estimates are for use in the calibration and application of a regional nutrient model, referred to as the SPARROW (SPAtially Referenced Regression On Watershed attributes; Smith and others, 1997) watershed model. The study area, referred to as Major River Basin-2 (MRB2), includes river basins draining to the South Atlantic and Gulf of Mexico (fig. 1).

¹ U.S. Geological Survey, Raleigh, North Carolina.

² U.S. Environmental Protection Agency, Atlanta, Georgia.

2 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States



LOCATION OF THE STUDY AREA IN THE SOUTHEASTERN UNITED STATES

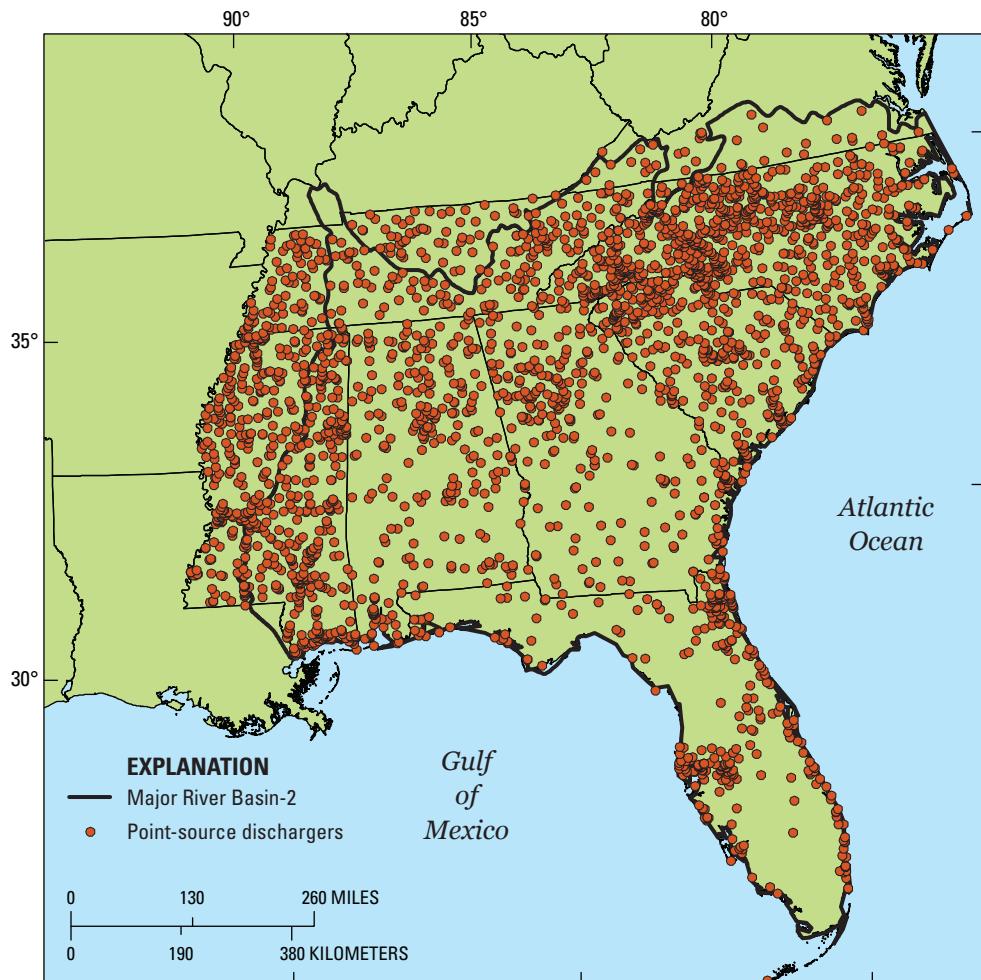


Figure 1. Major River Basin-2 study area and point-source dischargers.

Purpose and Scope

This report presents a replicable method, based on explicit assumptions, for estimating the annual loads of total nitrogen and total phosphorus in wastewater discharges from National Pollutant Discharge Elimination System (NPDES) regulated facilities in the southeastern United States. This methodology is intended to be applicable in studies in other parts of the country. The spatial scope of the study included individual point-source dischargers in Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee (fig. 1). The temporal scale of the project included the development of annual point-source nutrient estimates for 2000–2002, a time period coincident with the SPARROW model calibration period.

Use of Point-Source Nutrient Load Estimates in SPARROW

The SPARROW model, which is calibrated using nonlinear regression methods, uses a mass-balance conceptual approach to account for the introduction, transport, and processing of nutrients at the scale of individual reaches in a stream network. The model enables reach-specific predictions of nutrient load and concentration, as well as the apportioning of shares of this load among the nutrient sources (Smith and others, 1997). Geographic information system (GIS) data-management capabilities are used to manage information about nutrient sources, in-stream nutrient flux measured at monitoring sites, and landscape characteristics at the scale of individual stream reaches and reach catchments. The GIS uses the topological characteristics of the stream network (for example, Reach File 1 (RF1; U.S. Geological Survey, 2003) or National Hydrography Dataset (NHD; U.S. Geological Survey, 2006), such as connectivity and flow direction, to account for location and connectivity of individual stream reaches, thus enabling the introduction, transport, and processing of nutrients within a watershed to be accounted for in a spatially explicit manner. SPARROW specifies that in-stream nutrient flux is a function of a nonlinear relation between reach-specific nutrient inputs (for example, point sources, atmospheric deposition, agricultural inputs), and landscape and in-stream nitrogen processing. The statistical basis of SPARROW provides an objective means of specifying a relation between nutrient flux and the sources and losses of nutrients that occur within the network of stream reaches and reach catchments. The SPARROW modeling approach uses information derived from the stream-reach network about the spatial relation among nutrient fluxes, sources, landscape characteristics, and stream characteristics. The processing and delivery of nutrients to downstream water bodies are estimated as a function of the location, magnitude, and interactions of nutrient sources with the terrestrial and aquatic properties of the river basin. Estimates of the magnitude and proportional share of various point and nonpoint sources can be made for an individual stream reach or for the entire basin.

Challenges in Developing Point-Source Nutrient Load Estimates

Implementation of a SPARROW model requires information about nutrient sources, including point-source nutrient loads, at the stream-reach scale. Researchers would like to have data for each point-source discharger in the model study area, including regular and accurate entries for flow and nutrient concentration. Ideally, these data would be reported frequently enough to represent variability in facility operating conditions and in factors, such as seasonality and size and type of discharger, that influence nutrient discharge concentrations. Unfortunately, the ideal point-source data set typically does not exist for all dischargers in a region the size of this study area.

Point-source dischargers are required to monitor and report information about pollutants as defined in facility-specific NPDES discharge permits that are required by the Clean Water Act (United States Code, Amended 2002). These monitoring data are required to be stored in the USEPA Permit Compliance System (PCS) database. Typically, these data are collected as part of a self-monitoring program defined in the discharger's NPDES permit. The data are transferred to a state database by means of a facility-specific Discharge Monitoring Report (DMR), from which the data are transferred to regional USEPA PCS databases.

The reasons are varied why the PCS and state databases may not provide complete, up-to-date information for all facilities that may be discharging nutrients. Although the data structure of the USEPA PCS database is consistent nationwide, reporting requirements under the Clean Water Act vary for individual dischargers. Permit-specific discharge-concentration limits and reporting requirements for a constituent discharged by a facility, such as nitrogen or phosphorus, may vary among dischargers in a region as will the data entered into the PCS database. For example, not all dischargers may be required to monitor for total nitrogen during a particular time period; thus, data may be available for flow but not for total nitrogen concentration (or the concentrations of nitrogen constituents that would be combined to estimate total nitrogen), even if nitrogen is in the effluent. Generally, if a state does not have water-quality standards for a specific constituent in the watershed where a facility is located, the facility's permit may not include a limit for that constituent, and the monitoring data from that facility generally will not include information for the constituent.

Completeness, accuracy, and currency of the monitoring data may be affected by errors in data entry and data transfer. Data entry errors occur in two ways. Data may be entered incorrectly; for example, a monthly average flow of 0.10 million gallons per day (MGD) may be entered as 10 MGD. Also, data reported from the laboratory inadvertently may not be entered into the database. Another possibility is that updated data may not be transferred to the PCS database. For example, improved site-location information may be collected as part of a permit-renewal process by using global positioning system instruments to develop more accurate facility latitude and longitude data. Although these data may be entered in state-maintained site files, the data may not be transferred to the PCS database.

Acknowledgments

The authors thank the following individuals for assistance in compiling data and providing advice on preparing load estimates: Mike Templeton and Jeanne Phillips of the North Carolina Division of Water Quality; Dale Stoudemire, Larry Turner, Angela Murray, and Mike Montebello of the South Carolina Department of Health and Environmental Control; Bill Noell of the Georgia Department of Natural Resources; Wayne Magley of the Florida Division of Water Resource Management; Lynn Sisk of the Alabama Department of Environmental Management; Dusty Myers of the Mississippi Department of Environmental Quality; Bruce Evans of the Tennessee Division of Water Pollution Control; and Doug Moyers of the USGS Virginia Water Science Center. Thanks also are extended to Anne Hoos and Tammy Ivahnenko of the USGS and Steve Rubin of the USEPA for providing editorial and scientific reviews that substantially improved the quality of this report.

Methods of Investigation

The method for estimating point-source loads presented in this report extends the procedures developed at the USEPA Gulf of Mexico Program for studies of the Mississippi River basin during the late 1990s (U.S. Environmental Protection Agency, 1998a, b). Parallels and points of divergence can be found in the USEPA methods and the methods presented here.

A census approach was used in both methods to estimate loads from individual dischargers rather than using statistical approaches to estimate aggregate nutrient loads for the region of interest. For this study, the census approach was used because of the need to develop accurate, reach-scale estimates of nutrient inputs for calibration and application of SPARROW models.

Both methods are based on multiple sources of data, primarily the USEPA PCS database and state databases maintained independently of the PCS. The USEPA method also relied on treatment level information from the USEPA Clean Watersheds Needs Survey database and was able to deploy field investigators to examine individual discharger NPDES permit renewal applications and DMRs. In the method developed for this study, preference was given to using state data over PCS data when duplicate records were available for the same discharger, outfall, and date. It was assumed that data updates were made in the state databases first, with no assurance that the data updates were transferred to the PCS.

Both methods include calculated nutrient loads only for facilities that reported flow data. No effort was made to synthesize or estimate flow data for dischargers that did not report flow values; in other words, it was assumed that if no flow data were reported by a facility in the PCS for a given period, no discharge occurred from that facility for the given reporting period.

Both estimation efforts subset the data in the PCS or state databases to focus on certain dischargers thought to have

nutrients in their effluent, using the 1987 Standard Industrial Classification SIC codes (U.S. Census Bureau, 2005). The current project included 54 SIC codes to identify dischargers that regularly had nutrients reported in their effluent, as indicated by the data (table 1).

The development of both methods for estimating point-source loads used typical pollutant concentrations (TPC) to populate monitoring records that include flow data but are missing nutrient data. The TPCs used in the USEPA studies (U.S. Environmental Protection Agency, 1998a, b) were drawn from studies conducted by the National Oceanic and Atmospheric Administration (NOAA; Percy Pacheco, National Oceanic and Atmospheric Administration, written commun., February 2006).

To develop the TPC data used in this study, nutrient concentration data for MRB2 dischargers were summarized by using a combination of factors thought to influence variability in discharge nutrient concentrations, including SIC codes, the magnitude of facility discharges, and seasons. The TPC data were matched to discharger records with missing concentration data. Additional national-scale TPC data, based on SIC codes, were obtained from a USEPA database and combined with regional data to populate remaining missing concentration values.

The method developed to estimate annual point-source nutrient loads is a six-step procedure:

1. Develop a site file by using PCS and state databases.
2. Retrieve monitoring data from PCS and state databases.
3. Import data into a common database.
4. Prepare data for load estimation.
5. Estimate annual nutrient load in wastewater discharge.
6. Quality assure estimates of nutrient load in wastewater discharges.

Each of these steps is described in detail below.

Developing a Site File by Using the Permit Compliance System and State Databases

For each discharger, the permit number, name, SIC code, and location (latitude and longitude) were retrieved from the PCS by the USEPA Region 4 database manager. This information was supplemented with data obtained from the States of Alabama, Mississippi, North and South Carolina, and Tennessee containing the most recent available location information. SIC codes associated with effluents containing nutrients were identified based on the USEPA methods (U.S. Environmental Protection Agency, 1998a, b) and a review of nutrient-concentration data for permitted dischargers in the study area. These data were merged to produce a site-file listing of approximately 5,000 dischargers in the study area with a reasonable probability of having nutrients in their effluents.

Table 1. Standard Industrial Classification (SIC) codes (U.S. Census Bureau, 2005).

SIC code	SIC description
211	Beef Cattle Feedlots
213	Hogs
251	Broiler, Fryers, and Roaster Chickens
252	Chicken Eggs
253	Turkey and Turkey Eggs
254	Poultry Hatcheries
259	Poultry and Eggs, NEC
279	Animal Specialties, NEC
1475	Phosphate Rock
1541	General Contractors-Industrial Buildings and Warehouses
2011	Meat Packing Plants
2015	Poultry Slaughtering and Processing
2046	Wet Corn Milling
2085	Distilled and Blended Liquors
2611	Pulp Mills
2621	Paper Mills
2631	Paperboard Mills
2812	Alkalies and Chlorine
2819	Industrial Inorganic Chemicals, NEC
2821	Plastics Material and Synthetic Resins, and Nonvulcanizable Elastomers
2823	Cellulosic Manmade Fibers
2824	Manmade Organic Fibers, Except Cellulosic
2851	Paints, Varnishes, Lacquers, Enamels, and Allied Products
2865	Cyclic Organic Crudes and Intermediates, and Organic Dyes and Pigments
2869	Industrial Organic Chemicals, NEC
2873	Nitrogenous Fertilizers
2874	Phosphatic Fertilizers
2899	Chemicals and Chemical Preparations, NEC
2911	Petroleum Refining
3334	Primary Production of Aluminum
4941	Water Supply
4952	Sewerage Systems
4953	Refuse Systems
4959	Sanitary Services, NEC
5171	Petroleum Bulk Stations and Terminals
6513	Operators of Apartment Buildings
6515	Operators of Residential Mobile Home Sites
6552	Land Subdividers and Developers, Except Cemeteries
7011	Hotels and Motels
7032	Sporting and Recreational Camps
7033	Recreational Vehicle Parks and Campsites
7999	Amusement and Recreation Services, NEC
8062	General Medical and Surgical Hospitals
8211	Elementary and Secondary Schools
8221	Colleges, Universities, and Professional Schools
8222	Junior Colleges and Technical Institutes
8231	Libraries
8249	Vocational Schools, NEC
8351	Child Day Care Services
8361	Residential Care
8422	Arboreta and Botanical or Zoological Gardens
9223	Correctional Institutions
9621	Regulation and Administration of Transportation Programs
9711	National Security

Retrieving Monitoring Data from the Permit Compliance System and State Databases

Permit Compliance System and state databases were used to develop a database of discharger flow and nutrient concentration data for the period 2000–2002. Each record in the databases represents a specific combination of a unique discharge identifier (designated by a permit number), a discharger outfall identifier (designating a discharge pipe), and date (typically data were reported as a summary of monthly mean conditions). A PCS retrieval resulted in four database files containing flow, concentration (file called DMR), facility, and base codes and unit information (file called LIMITS) for all dischargers in the study region (Attachment 1). All states in the region were contacted about the availability of state databases to use in conjunction with the PCS data to develop state-specific files of flow and concentration data. Only North and South Carolina were able to provide relatively complete state data files.

Importing Data into a Common Database

The next step was to develop databases using a statistical program that could be used in subsequent steps to calculate annual nutrient loads for each discharger. A computer program (pcs_import_093005.sas; Attachment 2) was used to combine and analyze the data tables from PCS containing concentration data, facility information, flow data, and data on permit limits and data units (program lines 16–53; Attachment 2). Concentration and flow records with remark codes and coding errors were identified (lines 56–156; 160–243). Censored values were set to the censoring value for flow and concentration. Data files were created with a single record for each date, with variables defined for flow (lines 329–400) and concentrations of TN (lines 249–275) and TP (lines 277–314). All values represent mean monthly conditions. These files then were merged by discharger, outfall, and date. When subsequent load analysis detected errors in the data, the errors were corrected in the data-import program (e.g., lines 295–305; lines 333–391), a revised data file was created, and subsequent programs estimating nutrient load were rerun.

State data were processed similarly. The processing program for North Carolina nitrogen data is presented in Attachment 3. Nitrogen data, reported by river basin, were imported and combined with flow data (lines 20–164; Attachment 3). A data set with variables for individual nitrogen species was developed (lines 168–345). Several data characterization routines were run (lines 348–540), and a final data set was created (lines 543–581).

Preparing Data for Load Estimation

The next step was to prepare state-specific data sets having a single record for each discharger, outfall, and date (typically, one record per month) containing flow and TN- and TP-concentration values. A data-preparation program (Attachment 4) was run for each state, using PCS data and state data, if available. In the example given in Attachment 4, data from the State of North Carolina were used to create a file that summarizes flow and concentration for each discharger and outfall by month (lines 31–102). Data from the state database were combined with PCS data (lines 109–187) to create a single record for each unique combination of permit/outfall/year/month. PCS values for flow (q1flow), TN (c2p600), and TP (c2p665) were based on the PCS variables Q1 (flow) and C2 (nutrient concentration). The PCS file containing information on flow and concentration units did not contain information for most discharger records, so an assumption regarding units was necessary. Flow and concentration values in the databases were assumed to be monthly mean values, and values for flow and concentration were assumed to represent million gallons per day and milligrams per liter, respectively. All flow and concentration data were screened for outlier values.

A single flow value was defined for each discharger in North Carolina for each month for which data were available during 2000–2002 (lines 189–236; Attachment 4) and five flow-class values were defined for each facility's discharge in any month (lines 229–233). Preference was given to using state data, if available, for reasons described above. The number of flow values contained in the dataset, by state, for the period 2000–2002 ranged from 25,806 in North Carolina to 7,768 in Alabama. The majority of the flow observations were from minor dischargers; that is, permit holders with discharges of less than 1 MGD (fig. 2).

Guidelines for the calculation of nutrient load in the next processing step were that (1) only records with flow data were used in load calculations, and (2) all records with flow data were used in load calculations if a suitable TN or TP concentration could be estimated. Two approaches were used to calculate the nutrient-concentration data for as many records, which included flow data, as possible and are discussed in order of preference.

Monthly TN and TP concentration values from PCS or state databases were used if they were available for a discharger in a given year and month. Single monthly TN and TP concentration values were defined for each discharger and outfall (lines 253–301; Attachment 4), with preference given to state data. If a few concentration values were missing for a given discharger during a year, median seasonal values for TN and TP were estimated using data for the entire 2000–2002 period for that

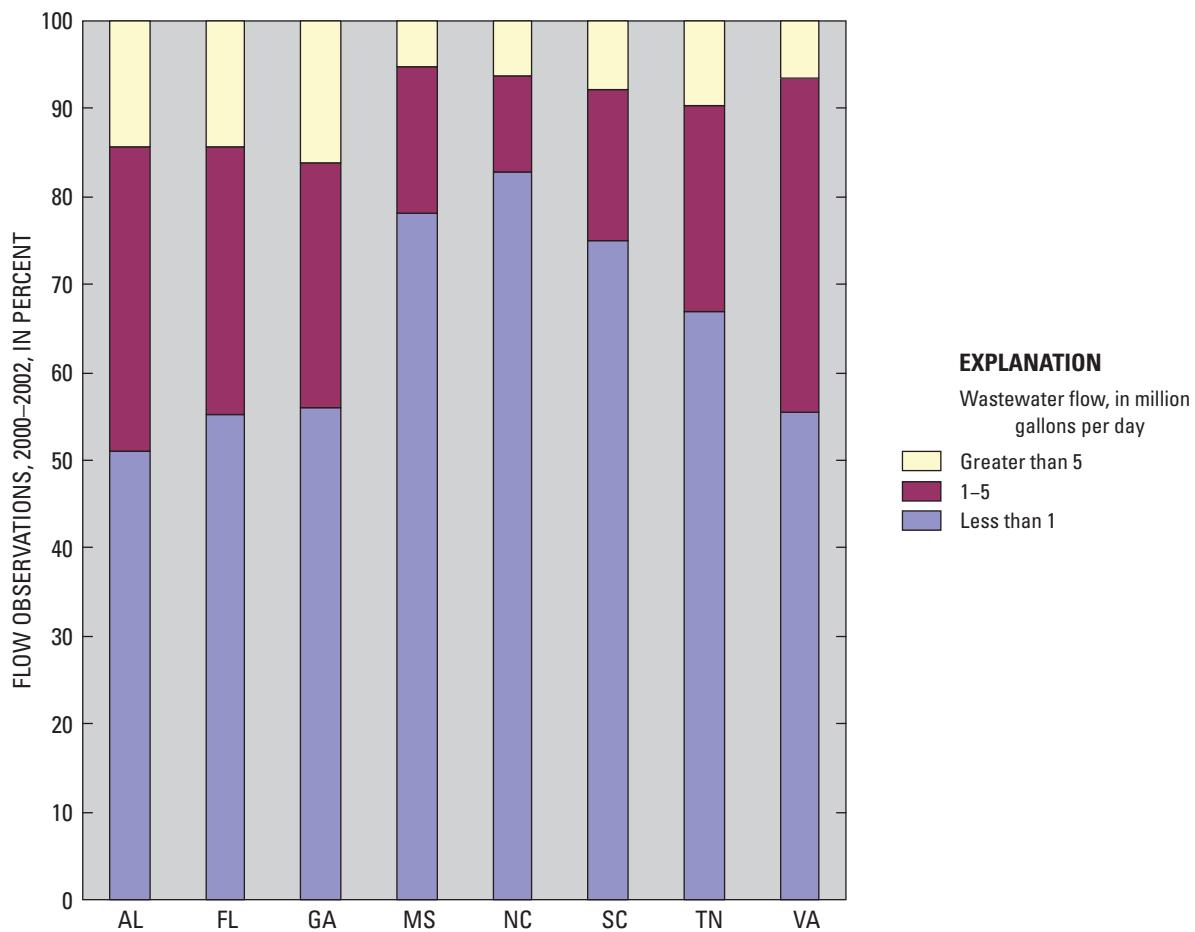


Figure 2. Summary of monthly effluent flow observations, by wastewater-flow classification and state, 2000–2002.

discharger; the resulting median concentration values were used for months in which flow data were available but concentration values were not (lines 330–370). Discharger-specific TN and TP concentration data were available most extensively in the States of Florida and North Carolina (fig. 3).

For the remaining records that contained flow data but no TN and TP concentration data, TPC values, developed in the program flowsicsea_060906.sas (Attachment 5), were substituted for missing concentration values. Three different types of TPC values were used in load calculations in declining order of preference. First, discharger data from all southeastern states were pooled and median seasonal concentration values were calculated by flow class, SIC code, and season (lines 37–121; Attachment 5). The rationale for preferring this method for estimating TPC is that flow class, SIC code, and season have an important effect on the variability of nutrient concentrations in effluent. Data on other factors that might affect variability, such as treatment level, were not uniformly available for all dischargers in the PCS and state databases. TPC values developed by using this method were merged with the main database by flow class, SIC code, and season (lines 402–425; Attachment 4).

To develop TN and TP concentration values for the remaining records with no concentration data (that is, monthly TN or TP data from either PCS or the state were not available and the TPC values based on flow class, SIC code, and season could not be matched to dischargers having flow data but no concentration data), the pooled data from all southeastern states were used to estimate median concentrations for TN and TP by flow class and SIC code (lines 124–161; Attachment 5). The resulting TPC values were merged with the main database and used to define missing TN and TP concentration values (lines 455–479; Attachment 4). Most concentration data for TN (fig. 3) and TP were derived either from facility-specific sources (that is, the monthly TN and TP concentration values for a discharger from PCS or the state data) or from TPC developed using discharger data specific to the MRB2 region.

Finally, if a TPC could not be drawn from the regional TPC database and matched with missing TN- or TP-concentration records, regional and national concentration data were used to develop TPC values by SIC code (lines 163–285; Attachment 5). National data were derived from a USEPA database (Steven Rubin, U.S. Environmental Protection Agency, written commun., February 2006). SIC-based TPC were used most

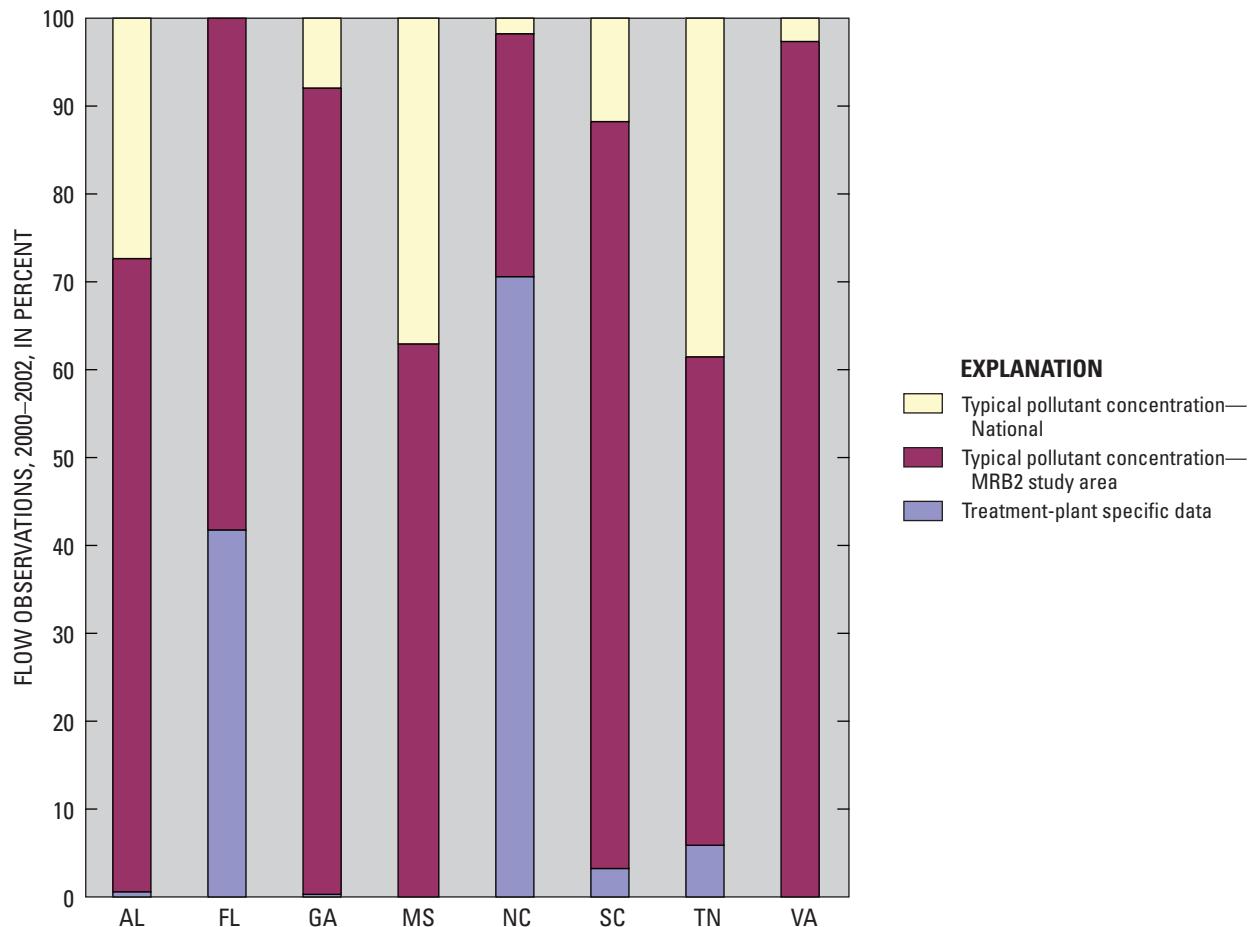


Figure 3. Summary, by state, of sources of total nitrogen concentration information used in load calculations.

extensively in the States of Alabama, Mississippi, and Tennessee (lines 508–534, Attachment 4; fig. 3).

TN and TP concentrations for the entire state data set were assessed to identify anomalous values for each discharger (Attachment 4, lines 607–669). Concentration greater than 10 times the median concentration value for a specific discharger during 2000–2002 and greater than the 95th percentile value of all concentration values over the entire data set were identified and replaced with the median concentration value for the specific discharger.

As a final step in the development of a data set for calculating nutrient loads, a value was calculated for each discharger, by outfall and year, indicating the number of months and the number of quarters with flow data (Attachment 4, lines 565–604). These values were used to determine the type of load calculation routine to be used for a discharger in a given year during the period 2000–2002, as described below. Summary data were used to prepare a GIS data set of facilities with annual nutrient load data; these GIS data were used to locate dischargers relative to streams within the MRB2 study area boundary.

Estimating Annual Nutrient Load in Wastewater Discharge

The objective of this step was to calculate an annual nutrient load for each discharger in the database. In general, loads were calculated by multiplying flow and concentration values (typically representing mean conditions for a month) for a given time period and summing these load values for the period. Three load calculation scenarios were used. (Attachment 6 provides an example of a state-specific load calculation program.) If flow values were available for each month of the year for a discharger, annual load was calculated as: concentration multiplied by flow multiplied by the number of days in the month, summed over all months for the year (Attachment 6, lines 22–72). If monthly flow values existed for less than 12 months but were available in 3 or 4 quarters of the year, it was assumed that flow (and load) occurred in all 12 months of the year. In this case, annual load was calculated on a seasonal basis, equal to the discharger's seasonal median concentration (for 2000–2002) multiplied by its seasonal

median flow multiplied by the number of days in the season, with seasonal load values summed to produce an annual load value (lines 77–131). If monthly flow values were available for less than 3 quarters of the year, load was calculated only for the months with flow data; monthly load values were summed to obtain an annual load estimate (lines 133–183). Data sets for individual years were created (Attachment 6, lines 206–250) and merged with site file information (lines 254–278).

The final objective of this step was to summarize information developed by the load preparation and load calculation programs (Attachment 7). Examples of data summaries include the number of dischargers by flow classification and state (table 2), the distribution of point-source load by treatment-plant size (table 3), the allocation of load calculation among the three load calculation methods (fig. 4), and the share of TN and TP produced by industries associated with the major SIC codes (table 4).

Quality Assuring Estimates of Nutrient Load in Wastewater Discharges

The objective of this step was to identify and correct erroneous load values. In some cases, the sources of errors and appropriate corrective actions were obvious. For instance, if the median flow value for a discharger was 0.122 MGD and the reported flow value for a given month was 130 MGD, it was assumed that an error was made in data entry, and a corrected value of 0.130 MGD was entered in place of 130 MGD. Following this initial round of data corrections, states that supplied data were asked to examine estimated load values, identify potentially problematic values, and suggest possible corrections.

Two data-checking routines were used to identify potentially erroneous data. Initial inspection of load values for individual dischargers indicated that many of the load values that appeared to be erroneous were associated with very large flow values relative to the other flow values for a particular discharger rather than inordinately large TN or TP concentration values. A flow-magnitude checking program (Attachment 8) identified several flow categories of potential concern. The focus was to identify potential problems with the data from dischargers of flow greater than 1 MGD, referred to as major dischargers, as these dischargers contributed the large proportion of nutrient load in the MRB2 study area. Additional criteria for identifying data for further checking included mean flow values greater than 100 MGD, monthly flow values exceeding 10 times the median flow value for a particular discharger during the period 2000–2002, and flow values exceeding 100 times the median flow value.

A second data-checking routine, located in the load-calculation program (Attachment 6, lines 287–322), calculates interannual differences in TN and TP load for each discharger, and identifies large positive (greater than the 95th percentile) and negative (less than the 5th percentile) differences. An assumption was made that data problems, such as extreme

Table 2. Point-source wastewater dischargers by state and flow classification, 2000.

[Flow from minor dischargers averages less than 1 million gallons per day (MGD); major dischargers greater than or equal to 1 MGD]

State	Number of dischargers, 2000		
	Minor	Major	Total
Alabama	54	93	147
Florida	136	103	239
Georgia	152	98	250
Mississippi	473	64	537
North Carolina	670	123	793
South Carolina	224	83	307
Tennessee	187	84	271
Virginia	21	16	37
Total	1,917	664	2,581

Table 3. Total nitrogen and total phosphorus loads from National Pollutant Discharge Elimination System permitted dischargers, by state and flow classification, 2000.

[Flow classification units are million gallons per day. Percentages refer to the percentage of metric tons from all flow classes]

State	Flow classification			Metric tons, all flow classes
	Less than 1	1–5	Greater than 5	
Total nitrogen				
Alabama	5%	19%	76%	7,462
Florida	3%	8%	89%	6,768
Georgia	5%	15%	79%	11,440
Mississippi	17%	39%	44%	4,710
North Carolina	9%	27%	63%	11,607
South Carolina	9%	24%	67%	7,032
Tennessee	8%	18%	74%	9,602
Virginia	11%	36%	53%	982
Total phosphorus				
Alabama	7%	24%	69%	1,117
Florida	7%	22%	71%	1,157
Georgia	6%	27%	67%	1,312
Mississippi	17%	51%	32%	737
North Carolina	9%	31%	60%	2,384
South Carolina	11%	31%	58%	942
Tennessee	10%	23%	67%	1,239
Virginia	11%	39%	50%	149

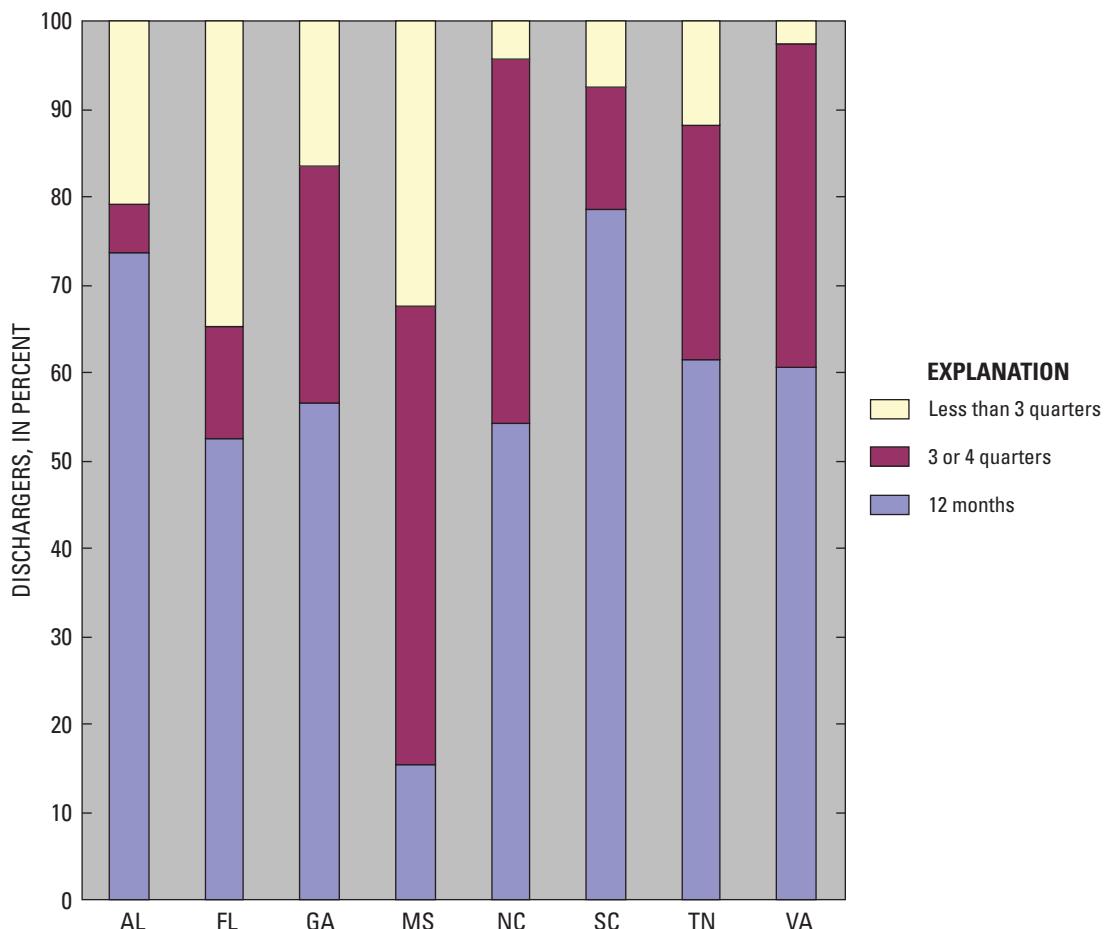


Figure 4. Percentage of dischargers using load calculation approaches based on data available for all 12 months, 3 or 4 quarters, or less than 3 quarters.

concentration or flow values attributed to laboratory errors or miscoding, would be revealed in larger interyear load estimates for a discharger. Annual load values flagged under these difference criteria, loads that were much larger than all other loads in the state (greater than the 95th percentile of all annual load estimates in the state), and data records with unusually high flow values, as discussed above, were compiled in a spreadsheet for further review by the states in the MRB2 study area.

Finally, data from 829 dischargers that had 12 months of flow data in 2002 and discharger-specific concentration data were used to compare loads generated using both original data and appropriate TPC SIC-based TN- and TP-concentration values (Attachment 7, lines 501–603). SIC-based con-

centration values were assumed most likely to differ from the discharger-specific concentration values. The Wilcoxon sign-rank test was used to test the hypothesis that the median difference between annual TN and TP loads calculated using the discharger-specific concentration data and TPC-based concentration data was zero (Helsel and Hirsch, 1992). No significant difference was noted in the TN loads calculated using the discharger and TPC concentration data. The TP loads calculated using the discharger-specific data were larger than loads calculated using the TPC data, a finding that indicates that TP loads calculated using SIC-based concentration values will be more conservative, or smaller, than TP loads calculated using discharger-specific concentration data.

Table 4. Annual total nitrogen and total phosphorus loads from point-source wastewater load dischargers in the southeastern United States, by Standard Industrial Classification (SIC) code, for 2000.

[Loads are expressed as the percentage share of total predicted point-source loads]

SIC code	SIC description	Share of total load
Nitrogen		
4952	Sewerage systems	78%
2611	Pulp mills	4%
2621	Paper mills	4%
2631	Paperboard mills	3%
2823	Cellulosic manmade fibers	2%
2011	Meat packing plants	2%
2015	Poultry slaughtering and processing	2%
2869	Industrial organic chemicals, NEC	1%
2824	Manmade organic fibers, except cellulosic	1%
9711	National security	1%
Phosphorus		
4952	Sewerage systems	73%
2621	Paper mills	6%
2611	Pulp mills	5%
2874	Phosphatic fertilizers	4%
2011	Meat packing plants	3%
2631	Paperboard mills	3%
2015	Poultry slaughtering and processing	2%
1475	Phosphate rock	2%

Summary

Several lessons are evident from the development and application of the methods presented in this report for estimating annual nutrient loads from permitted point-source dischargers. First, access to flow, concentration, and location data requires close communication and collaboration with the agencies that collect and manage these data. Data needed to estimate loads at a regional scale cannot be retrieved from the Internet. Once load calculations have been completed, the accuracy and usefulness of the load estimates depend on the willingness of the states and the USEPA to provide guidance and review for at least a subset of the load values that appear to be problematic. States are in the best position to provide the most up-to-date location information for discharge facilities. Accurate locational information is critical for appropriate allocation of discharger nutrient-load estimates relative to the stream-reach data-management framework used in the SPARROW model.

Second, the number of discharger records with flow data alone far exceeds the number of records having both flow and

nutrient-concentration data. In this situation, the accuracy of load estimates depends on the reasonableness of the TPC values. For this study, the major factors assumed in determining concentration variability were the magnitude of the discharge, the season when the discharge was measured, and the SIC code of the discharger. These factors were considered when assigning TPC to dischargers that did not have nutrient-concentration data available.

Finally, nutrient loads were calculated only for facilities that reported flow data. Although the majority of discharge observations across the study area were from minor dischargers (daily flow less than 1.0 MGD), a large proportion of nutrient load was generated by dischargers with flow greater than 1.0 MGD. From the standpoint of characterizing point-source nutrient loads for the region as a whole, it probably would be adequate to restrict load estimation to major dischargers. Because of the need in the SPARROW model to have a mass balance calculated using reach-scale data, nutrient load data from all dischargers are necessary in addition to facility location information, in order to locate discharges relative to stream reaches in the stream network.

References Cited

- Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Amsterdam, Elsevier, 529 p.
- McMahon, G., Alexander, R.B., and Qian, S., 2003, Support of TMDL programs using spatially referenced regression models: ASCE Journal of Water Resources Planning and Management, v. 129, p. 315–329.
- Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data: Water Resources Research, v. 33, no. 12, p. 2781–2798.
- U.S. Census Bureau, 2005, 1997 NAICS and 1987 SIC Correspondence Tables; accessed February 25, 2005, at <http://www.census.gov/epcd/naics/naicstb2.csv>
- U.S. Environmental Protection Agency, 1998a, Documentation of Phase I and Phase II activities in support of point-source nutrient loading analysis in the Mississippi River system, Draft report, July 1998, USEPA Gulf of Mexico Program, Stennis Space Center, MS; accessed July 19, 2006, at <http://www.epa.gov/msbasin/taskforce/phases.htm>
- U.S. Environmental Protection Agency, 1998b, Protocol for point source nutrient loading analysis in the Mississippi River system, April 1998, USEPA Gulf of Mexico Program, Stennis Space Center, MS; accessed July 19, 2006, at <http://www.epa.gov/msbasin/taskforce/protocol.htm>
- U.S. Geological Survey, 2003, ERF1_2 — Enhanced River Reach File 2.0; accessed July 17, 2006, at http://water.usgs.gov/GIS/metadata/usgsrd/XML/erf1_2.xml
- U.S. Geological Survey, 2006, National hydrography dataset; accessed July 17, 2006, at <http://nhd.usgs.gov/index.html>
- United States Code, amended in 2002, Federal Water Pollution Control Act (a.k.a. Clean Water Act): Washington, D.C., U.S. Government Printing Office, 33 U.S.C. Secs. 101-607; accessed in June 2006, at <http://www.epa.gov/region5/water/pdf/ecwa.pdf>

Attachments

Attachment 1: PCS Data Retrieval Routine

This retrieval extracts information describing the facility, including design flow:

```
00 RMT=HOLD TIME=99M PRTY=0 MSGCLASS=A MSGLEVEL=1,1
01 04 FACILITIES INFORMATION
02 JERRY MCMAHON
10 STTE NE KY
20 QF CARD=NO DSN=FACILITY DISP=MOD
30 NPID
60 NPID FNMS MADI FLOW FLAT FLON FHBC CNTN
```

This retrieval extracts effluent limits and monitoring requirements for flow, BOD, nitrogen, and phosphorus:

```
00 RMT=HOLD TIME=99M PRTY=0 MSGCLASS=A MSGLEVEL=1,1
01 04 LIMITS INFORMATION
02 JERRY MCMAHON
10 STTE NE KY
20 QF CARD=NO DSN=LIMITS DISP=MOD GHOST=YES RESTRICT=YES
30 NPID PDSG PRAM MLOC SEAN MODN LIPQ LNTP
60 NPID PDSG LIPQ LNTP PRAM MLOC SEAN MODN ELSD ELED
60 LQUCD LQAV LQAS LQMX LQXS LCUCD LCMN LCMS LCAV LCAS LCMX LCXS
60 WITH PRAM AL 00310
60 WITH PRAM AL 00600
60 WITH PRAM AL 00610
60 WITH PRAM AL 00625
60 WITH PRAM AL 00630
60 WITH PRAM AL 00665
60 WITH PRAM AL 50050
60 WITH PRAM AL 80082
```

This retrieval extracts actual flow measurements between January 1994 and December 2004:

```
00 RMT=HOLD TIME=99M PRTY=0 MSGCLASS=A MSGLEVEL=1,1
01 04 DMR'S FOR FLOW
02 JERRY MCMAHON
10 STTE NE KY
20 QF CARD=NO DSN=FLOW DISP=MOD ARCH=YES RESTRICT=YES
30 NPID VDSG VPRM VMLO MVDT
60 NPID VDSG VIPQ VLIM VPRM VMLO VSEA VMOD MVDT MQAV MQMX MCMN MCAV MCMX
60 WITH VPRM EQ 50050
60 WITH MVDT GT 010194
60 WITH MVDT LT 010195
```

This retrieval extracts BOD, nitrogen, and phosphorus measurements for 1994. One such retrieval was run for each year through 2004:

```
00 RMT=HOLD TIME=99M PRTY=0 MSGCLASS=A MSGLEVEL=1,1
01 04 DMR'S FOR SELECTED PARAMETERS - 1994
02 JERRY MCMAHON
10 STTE NE KY
20 QF CARD=NO DSN=DMR94 DISP=MOD ARCH=YES RESTRICT=YES
30 NPID VDSG VPRM VMLO MVDT
60 NPID VDSG VIPQ VLIM VPRM VMLO VSEA VMOD MVDT MQAV MQMX MCMN MCAV MCMX
60 WITH VPRM AL 00310
```

16 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

60 WITH VPRM AL 00600
60 WITH VPRM AL 00610
60 WITH VPRM AL 00625
60 WITH VPRM AL 00630
60 WITH VPRM AL 00665
60 WITH VPRM AL 80082
60 WITH MVDT GT 010194
60 WITH MVDT LT 010195

1 Attachment 2: PCS Data Import Program

```
2
3 /* ****
4
5 program: pcs_import_032106.sas
6 date: 3/20-21/06 7/6/2006(GA) 7/7-14/06 (FL,AL, MS, and TN))
7 from: pcs_import_093005.sas (3-20-06)
8
9 **** */
10
11 libname test 'z:\nutrients\sparrow\point_source\PCS_donahoo_0805';
12
13 options ps=54 ls=80;
14
15 /* import data developed by Mike Donahoo, August 2005
16 and create tabular PCS data sets, */
17
18 /* import DMR, facility, flow, and limits data */
19 /*
20
21 PROC IMPORT OUT= TEST.DMR
22   DATATYPE= "DMR"
23   DBMS=ACCESS2000 REPLACE;
24 DATABASE="z:\nutrients\sparrow\point_source\PCS_donahoo_0805\mcmahon2.mdb";
25 RUN;
26
27 proc contents; run;
28
29 PROC IMPORT OUT= TEST.facility
30   DATATYPE= "Facility"
31   DBMS=ACCESS2000 REPLACE;
32 DATABASE="z:\nutrients\sparrow\point_source\PCS_donahoo_0805\mcmahon2.mdb";
33 RUN;
34
35
36 proc contents; run;
37
38 PROC IMPORT OUT= TEST.flow
39   DATATYPE= "Flow"
40   DBMS=ACCESS2000 REPLACE;
41 DATABASE="z:\nutrients\sparrow\point_source\PCS_donahoo_0805\mcmahon2.mdb";
42 RUN;
43
44 proc contents; run;
45
46 PROC IMPORT OUT= TEST.limits
47   DATATYPE= "Limits"
48   DBMS=ACCESS2000 REPLACE;
49 DATABASE="z:\nutrients\sparrow\point_source\PCS_donahoo_0805\mcmahon2.mdb";
50 RUN;
51
52 proc contents; run;
```

18 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
53
54
55 /* preparation of dmr data
56
57 identification of records where concentration and quantity have "<" ">" or "-" value
58 or other value that cannot be translated into a numeric value. Identification done
59 by multiplying each of the concentration and quantity variables by one, looking at the
60 error messages, and making changes in code creating test.dmr2. Assumption is made that sets
61 conc and flow values that have a "<" or ">" remark code equal to the censored value.
62 */
63
64 /*
65 data test.dmr1;
66 format state $2. season $6.;;
67 set test.dmr;
68 month=month(datepart(date));
69 if month = 12 or month = 1 or month = 2 then season = 'winter';
70 if month = 5 or month = 3 or month = 4 then season = 'spring';
71 if month = 6 or month = 7 or month = 8 then season = 'summer';
72 if month = 11 or month = 9 or month = 10 then season = 'fall';
73 state=substr(npdes,1,2);
74 year=year(datepart(date));
75 outfall=substr(outfall,1,3);
76 run;
77
78
79 data test.dmr2;
80 set test.dmr1;
81 remark_c1 = 0;
82 remark_c2 = 0;
83 remark_c3 = 0;
84 remark_q1 = 0;
85 remark_q2 = 0;
86 if substr(c1,1,1)='<' then remark_c1 = 1;
87 if substr(c1,1,1)='>' then remark_c1 = 2;
88 if substr(c1,1,1)='-' then remark_c1 = 3;
89 if substr(c1,1,1)= '*' then remark_c1 = 3;
90 if substr(c2,1,1)='<' then remark_c2 = 1;
91 if substr(c2,1,1)='>' then remark_c2 = 2;
92 if substr(c2,1,1)='-' then remark_c2 = 3;
93 if substr(c2,6,1)='-' then remark_c2 = 3;
94 if substr(c2,1,1)= '*' then remark_c2 = 3;
95 if c2 = '10 24' then remark_c2 = 3;
96 if substr(c2,3,1)='-' then remark_c2 = 3;
97 if c3 = '1 5' then remark_c3 = 3;
98 if c3 = '1 9' then remark_c3 = 3;
99 if c3 = 'T' then remark_c3 = 3;
100 if c3 = '9 224' then remark_c3 = 3;
101 if c3 = '1-7' then remark_c3 = 3;
102 if c3 = '4.0-' then remark_c3 = 3;
103 if c3 = '4.4-' then remark_c3 = 3;
104 if c3 = '3.4-' then remark_c3 = 3;
105 if substr(c3,1,1)='<' then remark_c3 = 1;
106 if substr(c3,1,1)='>' then remark_c3 = 2;
107 if substr(c3,1,1)='-' then remark_c3 = 3;
108 if substr(q1,1,1)='<' then remark_q1 = 1;
```

```

109 if substr(q1,1,1)='>' then remark_q1 = 2;
110 if substr(q1,1,1)='-' then remark_q1 = 3;
111 if q1 = '9 18' then remark_q1 = 3;
112 if q1 = '11 .42' then remark_q1 = 3;
113 if q1 = '300 626' then remark_q1 = 3;
114 if q1 = '5.54-' then remark_q1 = 3;
115 if q1 = '14-38' then remark_q1 = 3;
116 if q1 = '13.4-' then remark_q1 = 3;
117 if q1 = '28 .0' then remark_q1 = 3;
118 if q1 = '5 592.6' then remark_q1 = 3;
119 if q1 = '6 433' then remark_q1 = 3;
120 if q1 = '*.2' then remark_q1 = 3;
121 if substr(q2,1,1)='<' then remark_q2 = 1;
122 if substr(q2,1,1)='>' then remark_q2 = 2;
123 if substr(q2,1,1)='-' then remark_q2 = 3;
124 if substr(q2,1,1)='*' then remark_q2 = 3;
125 if q2 = '50.4-' then remark_q2 = 3;
126 if q2 = '30.04-' then remark_q2 = 3;
127 if q2 = '*.47' then remark_q2 = 3;
128 if q2 = '30.8 3' then remark_q2 = 3;
129 run;
130
131 data test.temp;
132 set test.dmr2;
133 if remark_c3 eq 0;
134 conc3=c3*1;
135 run;
136
137
138
139 data test.dmr3;
140 set test.dmr2;
141 if remark_c1 = 1 or remark_c1 = 2 then conc1=substr(c1,2,7)*1;
142 if remark_c1 = 3 then conc1 = .;
143 if remark_c1 = 0 then conc1 = c1*1;
144 if remark_c2 = 1 or remark_c2 = 2 then conc2=substr(c2,2,7)*1;
145 if remark_c2 = 3 then conc2 = .;
146 if remark_c2 = 0 then conc2 = c2*1;
147 if remark_c3 = 1 or remark_c3 = 2 then conc3=substr(c3,2,7)*1;
148 if remark_c3 = 3 then conc3 = .;
149 if remark_c3 = 0 then conc3 = c3*1;
150 if remark_q1 = 1 or remark_q1 = 2 then quan1=substr(q1,2,7)*1;
151 if remark_q1 = 3 then quan1 = .;
152 if remark_q1 = 0 then quan1 = q1*1;
153 if remark_q2 = 1 or remark_q2 = 2 then quan2=substr(q2,2,7)*1;
154 if remark_q2 = 3 then quan2 = .;
155 if remark_q2 = 0 then quan2 = q2*1;
156 run;
157
158
159
160 /* preparation of flow data
161
162 including deletion of records with monitoring location ne 1 (effluent gross value)
163 identification of records where concentration and quantity have "<" ">" or "-" value
164 Identification done by multiplying each of the concentration and quantity variables by one,

```

20 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
165 looking at the
166 error messages, and making changes in code creating test.flow2. Assumption is made that sets
167 conc and flow values that have a "<" or ">" remark code equal to the censored value.
168 */
169 /*
170 /*
171 data test.flow1;
172 format state $2. season $6.;
173 set test.flow;
174 month=month(datepart(date));
175 if month = 12 or month = 1 or month = 2 then season = 'winter';
176 if month = 5 or month = 3 or month = 4 then season = 'spring';
177 if month = 6 or month = 7 or month = 8 then season = 'summer';
178 if month = 11 or month = 9 or month = 10 then season = 'fall';
179 state=substr(npdes,1,2);
180 year=year(datepart(date));
181 outfall=substr(outfall,1,3);
182 run;
183
184 data test.flow2;
185 set test.flow1;
186 remark_c1 = 0;
187 remark_c2 = 0;
188 remark_c3 = 0;
189 remark_q1 = 0;
190 remark_q2 = 0;
191 if substr(c1,1,1)='<' then remark_c1 = 1;
192 if substr(c1,1,1)='>' then remark_c1 = 2;
193 if substr(c1,1,1)='-' then remark_c1 = 3;
194 if substr(c1,1,1)='*' then remark_c1 = 3;
195 if substr(c2,1,1)='<' then remark_c2 = 1;
196 if substr(c2,1,1)='>' then remark_c2 = 2;
197 if substr(c2,1,1)='-' then remark_c2 = 3;
198 if substr(c2,6,1)='-' then remark_c2 = 3;
199 if substr(c2,1,1)='*' then remark_c2 = 3;
200 if substr(c3,1,1)='<' then remark_c3 = 1;
201 if substr(c3,1,1)='>' then remark_c3 = 2;
202 if substr(c3,1,1)='-' then remark_c3 = 3;
203 if substr(q1,1,1)='<' then remark_q1 = 1;
204 if substr(q1,1,1)='>' then remark_q1 = 2;
205 if substr(q1,1,1)='-' then remark_q1 = 3;
206 if q1 = '. 798' then remark_q1 = 3;
207 if q1 = '1.58 1' then remark_q1 = 3;
208 if q1 = '.09 6' then remark_q1 = 3;
209 if q1 = '.03 0' then remark_q1 = 3;
210 if q1 = '0.1805 0' then remark_q1 = 3;
211 if substr(q2,1,1)='<' then remark_q2 = 1;
212 if substr(q2,1,1)='>' then remark_q2 = 2;
213 if substr(q2,1,1)='-' then remark_q2 = 3;
214 if substr(q2,1,1)='T' then remark_q2 = 3;
215 if q2 = '.0 0030' then remark_q2 = 3;
216 if q2 = '0.0 72' then remark_q2 = 3;
217 if q2 = '0. 007' then remark_q2 = 3;
218
219 run;
220
```

```

221 data test.temp;
222 set test.flow2;
223 if remark_c3 eq 0;
224 conc3=c3*1;
225 run;
226
227 data test.flow3;
228 set test.flow2;
229 if remark_c1 = 1 or remark_c1 = 2 then conc1=substr(c1,2,7)*1;
230 if remark_c1 = 3 then conc1 = .;
231 if remark_c1 = 0 then conc1 = c1*1;
232 if remark_c2 = 1 or remark_c2 = 2 then conc2=substr(c2,2,7)*1;
233 if remark_c2 = 3 then conc2 = .;
234 if remark_c2 = 0 then conc2 = c2*1;
235 if remark_c3 = 1 or remark_c3 = 2 then conc3=substr(c3,2,7)*1;
236 if remark_c3 = 3 then conc3 = .;
237 if remark_c3 = 0 then conc3 = c3*1;
238 if remark_q1 = 1 or remark_q1 = 2 then quan1=substr(q1,2,7)*1;
239 if remark_q1 = 3 then quan1 = .;
240 if remark_q1 eq 0 then quan1=q1*1;
241 if remark_q2 = 1 or remark_q2 = 2 then quan2=substr(q2,2,7)*1;
242 if remark_q2 = 3 then quan2 = .;
243 if remark_q2 eq 0 then quan2=q2*1;
244 run;
245
246 /* create and merge files for individual parameters — note that there
247 are a number of edits related to checking limits file information
248 and reviewing initial load estimates and associated data */
249 /*
250 data temp600;
251 set test.dmr3;
252 if parameter = '00600';
253 if mon_loc='1' or mon_loc = 'E';
254 if npdes eq 'NC0006033' and year = 2002 and (month = 11 or month = 12 ); conc2=conc2/10;
255
256 run;
257 proc sort; by year month; run;
258
259 proc print; var npdes outfall year month conc2 ;run;
260
261
262
263 data test.p600;
264 set test.dmr3;
265 if parameter = '00600';
266 if mon_loc='1' or mon_loc = 'E';
267 if npdes eq 'NC0006033' and year = 2002 and (month = 11 or month = 12 ) then conc2=conc2/10;
268 c1p600=conc1;
269 c2p600=conc2;
270 c3p600=conc3;
271 q1p600=quan1;
272 q2p600=quan2;
273 keep npdes outfall date year state c1p600 c2p600 c3p600 q1p600 q2p600;
274 run;
275
276 proc sort; by npdes outfall date; run;

```

22 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
277  
278 data temp665;  
279 set test.dmr3;  
280 if parameter = '00665';  
281 if mon_loc='1' or mon_loc = 'E';  
282 if npdes eq 'GA0000973' and outfall = '020' and year=2000 and month=7 ; conc2=conc2/100;  
283  
284 run;  
285 proc sort; by year month; run;  
286  
287  
288 proc print; var npdes outfall year month conc2 ;run;  
289  
290  
291  
292 data test.p665;  
293 set test.dmr3;  
294 if parameter = '00665';  
295 if mon_loc='1' or mon_loc = 'E';  
296 if npdes = 'FL0043443' and ( date ge "30APR2002:00:00:00"dt and  
297 date le "30JUN2004:00:00:00"dt )then conc2=conc1/1000;  
298 if npdes = 'FL0168581' and date gt "31JUL2002:00:00:00"dt then conc2=conc1/1000;  
299 if state = 'FL' and ( conc1 gt 0 or conc3 gt 0 ) and conc2 eq '' then conc2=(conc1+conc3)/2;  
300 if npdes eq 'AL0061671' and year=2002 and month=2 then conc2=conc2/100;  
301 if npdes eq 'GA0000973' and outfall = '010' and year=2000 and month=7 then conc2=conc2/100;  
302 if npdes eq 'GA0000973' and outfall = '020' and year=2000 and month=7 then conc2=conc2/100;  
303 if npdes eq 'GA0030791' and year=2001 and month=11 then conc2=conc2/100;  
304 if npdes eq 'GA0046035' and outfall = '0B0' and year=2000 and month=7 then conc2=conc2/10;  
305 if npdes eq 'GA0046035' and outfall = '0B0' and year=2001 and month=6 then conc2=conc2/10;  
306 if npdes eq 'GA0047236' and outfall = '0B0' and year=2000 and month=5 then conc2=conc2/100;  
307 c1p665=conc1;  
308 c2p665=conc2;  
309 c3p665=conc3;  
310 q1p665=quan1;  
311 q2p665=quan2;  
312 keep npdes outfall date year state conc1 conc2 conc3 c1p665 c2p665 c3p665 q1p665 q2p665;  
313 run;  
314  
315 proc sort; by npdes outfall date; run;  
316 */  
317  
318 data tempflow;  
319 set test.flow3;  
320 if parameter = '50050';  
321 if mon_loc='1' or mon_loc = 'E';  
322 flag=1;  
323 if npdes = 'TN0065081' and year =2002;  
324 run;  
325 proc sort; by year npdes outfall month; run;  
326  
327  
328 proc print; var npdes outfall year month quan1 flag;run;  
329  
330  
331 data test.p50050;  
332 set test.flow3;
```

```

333 if parameter = '50050';
334 if mon_loc='1' or mon_loc = 'E';
335 flag=1;
336 if state = 'FL' and quan1 eq ' ' and quan2 gt 0 and quan2 le 8.1 then quan1=quan2;
337 if not (npdes eq 'FL0000256') ;
338 if not (npdes eq 'FL0000230') ;
339 if npdes eq 'FL0000523' and not( outfall='001' or outfall='003') then flag=0 ;
340 if npdes eq 'FL0000655' and not( outfall='102' or outfall='104') then flag=0 ;
341 if npdes eq 'FL0001911' and quan1 gt 1 then quan1=quan1/1000;
342 if npdes eq 'FL0002488' and not(outfall='002') then flag=0;
343 if npdes eq 'FL0002488' and outfall='002' and year=2001 and month=2 then quan1=quan1/10;
344 if npdes eq 'FL0002607' and year=2001 and month=4 then quan1=quan1/1000000;
345 if npdes eq 'FL0002631' and year=2001 and month=10 then quan1=quan1/100;
346 if npdes eq 'FL0023922' and year=2002 and month=8 then quan1=quan1/1000;
347 if not (npdes eq 'FL0032441') ;
348 if npdes eq 'FL0037940' and quan1 gt 10 then quan1=quan1/1000;
349 if npdes eq 'FL0041785' and year=2002 then flag=0;
350 if npdes eq 'FL0042315' and year=2001 and month=8 then quan1=quan1/1000;
351 if npdes eq 'FL0043443' and year=2000 and month = 7 then quan1=quan1/1000;
352 if npdes eq 'FL0043770' and not(year=2002) then flag=0;
353 if npdes eq 'FL0044245' and ((year=2000 and month=4) or (year=2001 and month=11)) then
354     quan1=quan1/10;
355 if npdes eq 'FL0140023' and not(year=2002) then flag=0 ;
356 if npdes eq 'FL0166511' and year=2002 and month=4 then quan1=quan1/1000;
357 if not(npdes eq 'FL0267538') ;
358 if state = 'AL' and ( quan1 eq ' ' or quan1 eq 0 ) and quan2 gt 0 and quan2 le .77 then quan1=quan2;
359 if not(npdes eq 'AL0002780');
360 if npdes ='AL0020672' and year=2000 and month=9 then quan1=quan2;
361 if npdes ='AL0020869' and year=2000 and month=1 then quan1=quan1/1000;
362 if npdes ='AL0020869' and year=2000 and month=3 then quan1=quan1/1000;
363 if npdes ='AL0020869' and year=2001 and month=2 then quan1=quan1/1000;
364 if npdes ='AL0022713' and year=2001 and month=11 then quan1=quan1/1000;
365 if npdes ='AL0024724' and year=2001 and month=3 then quan1=quan1/1000;
366 if npdes ='AL0024724' and year=2001 and month=7 then quan1=quan1/1000;
367 if npdes ='AL0044105' and year=2002 and month=12 then quan1=quan1/1000;
368 if npdes ='AL0048763' and year=2000 and month=9 then quan1=quan1/1000;
369 if npdes ='AL0054631' and year=2002 and month=5 then quan1=quan1/1000;
370 if npdes ='AL0055841' and year=2002 and month=12 and outfall='001' then quan1=quan1/1000;
371 if npdes ='AL0055841' and year=2001 and month=3 and outfall='001' then quan1=quan1/1000;
372 if npdes ='AL0059218' and year=2001 and month=8 then quan1=quan1/1000;
373 if npdes ='AL0059218' and year=2001 and month=5 then quan1=quan1/1000;
374 if npdes ='AL0059218' and year=2000 and month=8 then quan1=quan1/1000;
375 if npdes= 'AL0020206' and year=2000 and month=4 then quan1=7.5;
376 if not( npdes eq 'AL0003093');
377 if not( npdes eq 'AL0003301');
378 if not( npdes eq 'AL0025968');
379 if not( npdes eq 'AL0002666' and year=2002 and ( outfall = '003' or outfall ='004')) ;
380 if npdes eq 'AL0020885' and year=2000 and ( month=1 or month=2 or month=3 or month=4 or month=5
381     or month=9 or month=10) then quan1=quan1/100;
382 if npdes eq 'AL0022314' and outfall='002' and year=2000 and ( month=1 or month=2 or month=3
383     or month=4 or month=6 or month=8) then quan1=quan1/1000;
384 if npdes eq 'AL0022314' and outfall='002' and year=2000 and month=5 then quan1=quan1/1000000;
385 if npdes eq 'AL0023922' and year=2002 then quan1=.27;
386 if npdes eq 'AL0052019' and year=2002 then flag=0 ;
387 if npdes eq 'AL0052850' and year=2002 then quan1=.02;
388 if not(npdes eq 'AL0053091');

```

24 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
389 if npdes eq 'AL0000116' and ( outfall = '002' or outfall = '003') then flag=0 ;
390 if npdes eq 'AL0003026' and not(outfall='001') then flag=0;
391 if not(npdes eq 'AL0003158');
392 if npdes = 'SC0047821' and year=2001 and month=10 then quan1=quan1/100;
393 if state = 'MS' and quan1 eq '' and quan2 gt 0 and quan2 le .05 then quan1=quan2;
394 if npdes = 'MS0033961' and year=2001 and month=4 then quan1=quan1/1000;
395 if npdes = 'MS0033961' and year=2002 and month=5 then quan1=quan1/1000;
396 if npdes='MS0000213' and year=2000 and month=12 then quan1=quan1/100;
397 if npdes='MS0023345' and year=2000 and month=1 then quan1=quan1/1000;
398 if npdes='MS0024627' and year=2000 and month=1 then quan1=quan1/1000;
399 if npdes='MS0029513' and year=2000 and month=9 then quan1=quan1/1000;
400 if npdes = 'MS0000213' then sic_code='2611';
401 if not(npdes eq 'MS0046655');
402 if not(npdes eq 'MS0052949');
403 if not(npdes eq 'MS0021601');
404 if not(npdes eq 'MS0038849');
405 if not(npdes eq 'MS0044296');
406 if not(npdes eq 'MS0044679');
407 if not(npdes eq 'MS0039934');
408 if not(npdes eq 'MS0046639');
409 if not(npdes eq 'MS0047147');
410 if not(npdes eq 'MS0049107');
411 if not(npdes eq 'MS0055387');
412 if npdes eq 'MS0001520' and not(outfall='001') then flag=0 ;
413 if npdes eq 'MS0001686' and outfall='001' and year=2000 and month=9 then flag=0;
414 if npdes eq 'MS0020206' and year=2001 and month=9 then quan1=quan1/1000;
415 if npdes eq 'MS0020362' and outfall='003' and year=2001 and month=3 then quan1=quan1/1000;
416 if npdes eq 'MS0025526' and year=2002 and month=7 then quan1=quan1/100;
417 if npdes eq 'MS0028029' and year=2002 and month=6 then quan1=quan1/1000;
418 if npdes eq 'MS0028185' and year=2000 and month=6 then quan1=quan1/10;
419 if not(npdes eq 'MS0034878');
420 if npdes eq 'MS0036331' and year=2000 and month=12 then quan1=quan1/100 ;
421 if npdes eq 'MS0039144' and year=2002 and month=12 then flag=0;
422 if npdes eq 'MS0045551' and year=2002 and month =9 and outfall='001' then quan1=quan1/10 ;
423 if npdes eq 'MS0045772' and year=2001 and month=9 and outfall='001' then quan1=quan1/1000;
424 if npdes eq 'MS0045772' and year=2002 and month=12 and outfall='002' then quan1=quan1/1000;
425 if npdes eq 'MS0053589' and year=2001 and month=3 then quan1=quan1/1000;
426 if npdes eq 'MS0055379' and year=2002 and (month=4 or month=5 or month=6 or month=7 or month=8
427 or month=9 or month=11) then quan1=quan1/1000;
428 if npdes eq 'GA0000973' and outfall='020' and year=2000 and month=10 then quan1=quan1/100 ;
429 if npdes eq 'GA0020478' and outfall='0B0' and year=2000 and month=5 then quan1=quan1/1000 ;
430 if npdes eq 'GA0031046' and outfall='0B1' and year=2002 and month=8 then quan1=quan1/1000 ;
431 if npdes eq 'GA0032514' and year=2002 and month=7 then quan1=quan1/1000 ;
432 if npdes='GA0000311' and year=2001 and month = 5 then quan1=.067;
433 if not(npdes eq 'GA0003671');
434 if not(npdes eq 'GA0038318');
435 if not(npdes eq 'GA0003590');
436 if npdes eq 'GA0001279' and year = 2000 and month = 2 then quan1=.201;
437 if npdes eq 'GA0002798' and year = 2000 and month=10 then quan1=quan1/10;
438 if npdes eq 'GA0002798' and year = 2002 and month=1 and outfall = '001' then quan1=quan1/10;
439 if npdes eq 'GA0003620' and year = 2000 and month = 2 then quan1=quan1/10;
440 if npdes eq 'GA0021032' and quan1 ge 100 then quan1=quan1/1000;
441 if npdes eq 'GA0021032' and ( quan1 ge 1.65 and quan1 lt 10) then quan1=quan1/10;
442 if npdes eq 'GA0021512' and quan1 gt 1 then quan1=quan1/1000;
443 if npdes eq 'GA0022900' and year = 2002 and month=10 then quan1=quan1/100;
444 if npdes eq 'GA0025674' and year = 2002 and month=2 then quan1=quan1/10;
```

```
445 if npdes eq 'GA0034819' and year=2000 and month=4 then quan1=quan1/1000;
446 if npdes eq 'GA0049166' and year=2001 and month=1 then quan1=.058;
447 if npdes eq 'TN0002135' and year=2000 and month=5 then quan1=quan1/1000;
448 if npdes eq 'TN0002356' and year=2000 and month=7 and outfall='001' then quan1=quan1/100;
449 if npdes eq 'TN0003433' and year=2000 and month=9 then quan1=quan1/1000 ;
450 if npdes eq 'TN0020494' and year=2002 and (month=2 or month=5 or month=6) then quan1=quan1/1000 ;
451 if npdes eq 'TN0020613' and year=2001 and month=8 then quan1=quan1/1000 ;
452 if npdes eq 'TN0020672' and year=2000 and month=7 then quan1=quan1/1000 ;
453 if npdes eq 'TN0020702' and year=2000 and month=7 then quan1=quan1/1000 ;
454 if npdes eq 'TN0021164' and year=2001 and month=6 then quan1=.028 ;
455 if npdes eq 'TN0023001' and year=2000 and month=12 then quan1=quan1/1000 ;
456 if npdes eq 'TN0023353' and year=2000 and month=10 then quan1=quan1/10000 ;
457 if npdes eq 'TN0023469' and year=2002 and month=12 then quan1=quan1/1000 ;
458 if npdes eq 'TN0023477' and year=2000 and month=7 then quan1=quan1/10;
459 if npdes eq 'TN0024996' and year=2001 and month=10 then quan1=quan1/1000 ;
460 if npdes eq 'TN0025038' and year=2002 and month=5 then quan1=2.5 ;
461 if npdes eq 'TN0026247' and year=2000 and month=10 then quan1=quan1/1000 ;
462 if npdes eq 'TN0026506' and year=2000 and month ne 3 then quan1=quan1/1000 ;
463 if npdes eq 'TN0026506' and year=2001 and (month=1 or month=2 or month=3 or month=11) then
464 quan1=quan1/1000 ;
465 if npdes eq 'TN0026506' and year=2002 and (month=3 or month=8 ) then quan1=quan1/1000 ;
466 if npdes eq 'TN0026573' and year=2001 and (month=5 or month=10 ) then quan1=quan1/1000 ;
467 if npdes eq 'TN0026638' and year=2001 and month=10 then quan1=quan1/1000 ;
468 if npdes eq 'TN0028622' and year=2001 and month=2 then quan1=quan1/1000 ;
469 if npdes eq 'TN0058181' and year=2001 and month=10 then quan1=quan1/1000 ;
470 if npdes eq 'TN0060186' and year=2002 and (month=12 or month=1 or month=2) then quan1=.103 ;
471 if npdes eq 'TN0060186' and year=2002 and (month=3 or month=4 or month=5) then quan1=.0855 ;
472 if npdes eq 'TN0060186' and year=2002 and (month=6 or month=7 or month=8) then quan1=.67 ;
473 if npdes eq 'TN0060186' and year=2002 and (month=9 or month=10 or month=11) then quan1=.095 ;
474 if npdes eq 'TN0062057' and year=2002 and month=1 then quan1=quan1/1000;
475 if npdes eq 'TN0062294' and year=2002 and (month=10 or month=11) then quan1=quan1/1000;
476 if npdes eq 'TN0062499' and year=2001 and month=9 then quan1=quan1/1000;
477 if npdes eq 'TN0064912' and year=2001 and month=4 then quan1=quan1/1000;
478 if npdes eq 'TN0065358' and year=2001 and month=2 then quan1=quan1/1000;
479 if npdes eq 'TN0065501' and year=2001 and month=2 then quan1=quan1/100000;
480 if npdes eq 'TN0067423' and year=2000 and month=10 then quan1=quan1/1000;
481 if not(npdes eq 'TN0002356');
482 if not(npdes eq 'TN0002411');
483 if not(npdes eq 'TN0000205');
484 if not(npdes eq 'TN0002640');
485 if not(npdes eq 'TN0002950');
486 if not(npdes eq 'TN0003671');
487 if not(npdes eq 'TN0022519');
488 if npdes eq 'TN0026506' and year=2001 and month=2 then flag=0;
489 if not(npdes eq 'TN0041939');
490 if not(npdes eq 'TN0057487');
491 if npdes eq 'TN0061387' and year=2002 and month=7 and quan1=50000 then flag=0;
492 if not(npdes eq 'TN0062120');
493 if not(npdes eq 'TN0074730');
494 if npdes eq 'TN0000060' and year=2002 and month=10 and outfall = '009' then quan1=quan1/1000;
495 if npdes eq 'TN0000060' and not(outfall='009') then flag=0;
496 if not(npdes eq 'TN0003751');
497 if npdes eq 'TN0020753' and quan1 gt 1 then quan1=quan1/1000 ;
498 if npdes eq 'TN0021644' and year=2002 and month=9 then quan1=quan1/1000 ;
499 if npdes eq 'TN0021717' and year=2000 and month=10 then quan1=quan1/100000;
500 if npdes eq 'TN0022560' and quan1 ge 1 then quan1=quan1/10;
```

26 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
501 if npdes eq 'TN0022586' and year=2000 and month=5 then quan1=quan1/10;
502 if npdes eq 'TN0024171' and year=2002 and month=5 then quan1=quan1/10000;
503 if npdes eq 'TN0024210' and not(outfall='001') then flag=0;
504 if npdes eq 'TN0024953' and year=2001 and (month=7 or month=8) then quan1=quan1/100 ;
505 if npdes eq 'TN0024970' and year=2000 and (month=10 or month=11 or month=12)then flag=0;
506 if npdes = 'TN0025488' and year=2000 and month=11 then quan1=quan1/10;
507 if npdes = 'TN0026166' and year=2001 and month=12 then quan1=quan1/10;
508 if npdes = 'TN0026573' and quan1 gt 10 then quan1=quan1/1000;
509 if npdes = 'TN0058181' and quan1 gt 1000 then quan1=quan1/1000;
510 if npdes = 'TN0059226' and year=2002 and month=10 and outfall='004' then quan1=quan1/10;
511 if npdes = 'TN0059226' and year=2002 and month=7 and outfall='004' then quan1=quan1/1000;
512 if npdes = 'TN0060186' and year=2001 and month=12 then quan1=quan1/1000;
513 if npdes = 'TN0064785' and quan1 gt 10 then quan1=quan1/100;
514 if flag=1;
515 c1flow=conc1;
516 c2flow=conc2;
517 c3flow=conc3;
518 q1flow=quan1;
519 q2flow=quan2;
520 keep npdes outfall date year state c1flow c2flow c3flow q1flow q2flow;
521 run;
522
523 proc sort; by npdes outfall date; run;
524
525 data test.dmr_flow_093005;
526 merge test.p600 test.p665 test.p50050;
527 by npdes outfall date;
528 run;
529
530
531 proc sort data=test.dmr_flow_093005; by state year; run;
532
533 proc univariate noprint;
534 var c1p600 c2p600 c3p600 q1p600 q2p600
535 c1p665 c2p665 c3p665 q1p665 q2p665
536 c1flow c2flow c3flow q1flow q2flow;
537 by state year;
538 output out=stat median=medc1p600 medc2p600 medc3p600 medq1p600 medq2p600
539 medc1p665 medc2p665 medc3p665 medq1p665 medq2p665
540 medc1flow medc2flow medc3flow medq1flow
541 medq2flow
542 n=n_c1p600 n_c2p600 n_c3p600 n_q1p600 n_q2p600
543 n_c1p665 n_c2p665 n_c3p665 n_q1p665 n_q2p665
544 n_c1flow n_c2flow n_c3flow n_q1flow n_q2flow;
545 run;
546
547 proc print; run;
548
549 PROC EXPORT DATA= work.stat
550 OUTFILE= "Z:\Nutrients\SPARROW\point_source\PCS_donahoo_0805
551 \dmr_flow_stat_010406t.xls"
552 DBMS=EXCEL REPLACE;
553 SHEET="stats";
554 RUN;
555
556
557
558 quit;
```

Attachment 3: North Carolina Data Import Program

```
1  /*
2 3
4  ****
5
6  program: nc_n_031305.sas
7  date: 3/9-14/05 10-11-05
8  based on: point92_data.sas (3-7/8-01)
9
10 ****
11
12 libname test 'z:\Nutrients\sparrow\point_source\nc\nitrogen';
13 libname test1 'z:\Nutrients\sparrow\point_source\nc\flow';
14 libname test2 'Z:\Nutrients\SPARROW\point_source\site_1005';
15 options ps=54 ls=80;
16
17
18
19
20 /* read in nitrogen data files and combine */
21
22 /*
23 PROC IMPORT OUT= TEST.SUB0301n
24 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\SUB0301n.xls"
25 DBMS=EXCEL2000 REPLACE;
26 GETNAMES=YES;
27 RUN;
28
29 proc contents; run;
30
31 PROC IMPORT OUT= TEST.Sub0302n
32 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub0302n.xls"
33 DBMS=EXCEL2000 REPLACE;
34 GETNAMES=YES;
35 RUN;
36
37 proc contents; run;
38
39
40 PROC IMPORT OUT= TEST.SUB0303n
41 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\SUB0303n.xls"
42 DBMS=EXCEL2000 REPLACE;
43 GETNAMES=YES;
44 RUN;
45
46 proc contents; run;
47
48 PROC IMPORT OUT= TEST.Sub0304n
49 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub0304n.xls"
50 DBMS=EXCEL2000 REPLACE;
51 GETNAMES=YES;
52 RUN;
53
54 proc contents; run;
55
```

28 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
56  
57 PROC IMPORT OUT= TEST.SUB0305n  
58 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\SUB0305n.xls"  
59 DBMS=EXCEL2000 REPLACE;  
60 GETNAMES=YES;  
61 RUN;  
62  
63 proc contents; run;  
64  
65  
66 PROC IMPORT OUT= TEST.Sub0306n  
67 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub0306n.xls"  
68 DBMS=EXCEL2000 REPLACE;  
69 GETNAMES=YES;  
70 RUN;  
71 proc contents; run;  
72  
73  
74 PROC IMPORT OUT= TEST.SUB0307n  
75 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\SUB0307n.xls"  
76 DBMS=EXCEL2000 REPLACE;  
77 GETNAMES=YES;  
78 RUN;  
79 proc contents; run;  
80  
81  
82 PROC IMPORT OUT= TEST.Sub0308n  
83 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub0308n.xls"  
84 DBMS=EXCEL2000 REPLACE;  
85 GETNAMES=YES;  
86 RUN;  
87 proc contents; run;  
88  
89  
90 PROC IMPORT OUT= TEST.Sub0313n  
91 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub0313n.xls"  
92 DBMS=EXCEL2000 REPLACE;  
93 GETNAMES=YES;  
94 RUN;  
95 proc contents; run;  
96  
97  
98 PROC IMPORT OUT= TEST.SUB04n  
99 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\SUB04n.xls"  
100 DBMS=EXCEL2000 REPLACE;  
101 GETNAMES=YES;  
102 RUN;  
103 proc contents; run;  
104  
105  
106 PROC IMPORT OUT= TEST.Sub05n  
107 DATAFILE= "z:\Nutrients\SPARROW\point_source\nc\nitrogen\Sub05n.xls"  
108 DBMS=EXCEL2000 REPLACE;  
109 GETNAMES=YES;  
110 RUN;  
111 proc contents; run;
```

```
112  
113  
114 data test.nc_nitrogen_030805;  
115 set test.sub0301n test.sub0302n test.sub0303n test.sub0304n  
116 test.sub0305n test.sub0306n test.sub0307n test.sub0308n  
117 test.sub0313n test.sub04n test.sub05n;  
118 run;  
119  
120 /*  
121  
122  
123  
124  
125 /* combine flow and nitrogen data */  
126 /*  
127 proc contents data=test.nc_nitrogen_030805; run;  
128 proc contents data=test1.nc_flow_030805; run;  
129  
130 data test.flow1;  
131 set test1.nc_flow_030805;  
132 flow_mgd=value;  
133 keep permit outfall date flow_mgd;  
134 run;  
135  
136 data test.nitrogen;  
137 set test.nc_nitrogen_030805;  
138 units=UoM;  
139 keep facility permit outfall region subbasin parm value date month year units ;  
140 run;  
141  
142 proc sort data=test.nitrogen;  
143 by permit outfall date;  
144 run;  
145  
146 proc sort data=test.flow1;  
147 by permit outfall date;  
148 run;  
149  
150 data test.tmp1;  
151 merge test.flow1 test.nitrogen;  
152 by permit outfall date;  
153 run;  
154  
155 proc print data=test.tmp1(obs=500); run;  
156  
157 data test.flow_n_030905;  
158 set test.tmp1;  
159 if value ne '.';  
160 run;  
161  
162 proc contents; run;  
163  
164 proc print data=test.flow_n_030905(obs=500); run;  
165  
166  
167
```

30 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
168 /* create tabular data set, 1 record per date */
169 /*
170 /*
171 proc sort data=test.flow_n_030905; by parm units; run;
172
173 proc univariate data=test.flow_n_030905 noprint;
174   var value;
175   by parm units;
176   output out=temp1 n=nobs_parm mean=mean_val;
177 run;
178
179 proc print; run;
180
181 data test.p600a;
182   set test.flow_n_030905;
183   if parm eq 'P00600';
184   if units eq 'mg/l';
185   p600_mgl=value;
186   drop units value parm;
187 run;
188
189 proc sort data=test.p600a; by permit outfall date; run;
190
191 data test.p600b;
192   set test.flow_n_030905;
193   if parm eq 'P00600';
194   if units eq 'lbs/day';
195   p600lbdy=value;
196   drop units value parm;
197 run;
198 proc sort data=test.p600b; by permit outfall date; run;
199
200 data test.p600c;
201   set test.flow_n_030905;
202   if parm eq 'P00600';
203   if units eq 'lbs/yr';
204   p600lbyr=value;
205   drop units value parm;
206 run;
207 proc sort data=test.p600c; by permit outfall date; run;
208
209 data test.p610a;
210   set test.flow_n_030905;
211   if parm eq 'P00610';
212   if units eq 'mg/l';
213   p610_mgl=value;
214   drop units value parm;
215 run;
216
217 proc sort data=test.p610a; by permit outfall date; run;
218
219 data test.p610b;
220   set test.flow_n_030905;
221   if parm eq 'P00610';
222   if units eq 'lbs/day';
223   p610lbdy=value;
```

```
224 drop units value parm;
225 run;
226 proc sort data=test.p610b; by permit outfall date; run;
227
228 data test.p610c;
229 set test.flow_n_030905;
230 if parm eq 'P00610';
231 if units eq 'lbs/yr';
232 p610lbyr=value;
233 drop units value parm;
234 run;
235 proc sort data=test.p610c; by permit outfall date; run;
236
237 data test.p625a;
238 set test.flow_n_030905;
239 if parm eq 'P00625';
240 if units eq 'mg/l';
241 p625_mgl=value;
242 drop units value parm;
243 run;
244 proc sort data=test.p625a; by permit outfall date; run;
245
246 data test.p630a;
247 set test.flow_n_030905;
248 if parm eq 'P00630';
249 if units eq 'mg/l';
250 p630_mgl=value;
251 drop units value parm;
252 run;
253 proc sort data=test.p630a; by permit outfall date; run;
254
255 data test.nc_nitrogen;
256 merge test.p600a test.p600b test.p600c test.p610a test.p610b test.p610c test.p625a test.p630a;
257 by permit outfall date;
258 run;
259
260
261 data test.nc_nitrogen;
262 set test.nc_nitrogen;
263 year=year(datepart(date));
264 quarter=qtr(datepart(date));
265 month=month(datepart(date));
266 if p600_mgl gt 0 then screen = 'p600_mgl';
267 if( screen ne 'p600_mgl' and p600lbdy gt 0 ) then screen = 'p600lbdy';
268 if( screen = ' ' and p600lbyr gt 0 ) then screen = 'p600lbyr';
269 if( screen = ' ' and p610_mgl gt 0 ) then screen = 'p610_mgl';
270 if( screen = ' ' and p610lbdy gt 0 ) then screen = 'p610lbdy';
271 if( screen = ' ' and p610lbyr gt 0 ) then screen = 'p610lbyr';
272 if( screen = ' ' and p625_mgl gt 0 ) then screen = 'p625_mgl';
273 if( screen = ' ' and p630_mgl gt 0 ) then screen = 'p630_mgl';
274 if( screen = 'p625_mgl' and p630_mgl gt 0 ) then screen = 'p625,30';
275 run;
276
277 data test.nc_nitrogen;
278 set test.nc_nitrogen;
279 if screen ne ' ';
```

32 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
280 run;
281 /* convert p625 + p630 to p600 mg/l */
283 /*
284 data temp625_30;
285 set test.nc_nitrogen;
286 if screen = 'p625,30';
287 p600_mgl = p625_mgl + p630;
288 keep permit outfall date p600_mgl ;
289 run;
290
291 proc sort; by permit outfall date; run;
292
293 /* convert p600 lb/day to p600 mg/l */
294 /*
295 data temp600a;
296 set test.nc_nitrogen;
297 if screen = 'p600lbdy';
298 p600_mgl =(p600lbdy/flow_mgd)*.264*.000001*453*1000;
299 keep permit outfall date p600_mgl p600lbdy;
300 run;
301
302 proc sort; by permit outfall date; run;
303
304 /* convert p600 lb/year to p600 mg/l */
305 /*
306 data temp600b;
307 set test.nc_nitrogen;
308 if screen = 'p600lbyr';
309 p600_mgl =((p600lbyr/365)/flow_mgd)*.264*.000001*453*1000;
310 keep permit outfall date p600_mgl p600lbyr;
311 run;
312
313 proc sort; by permit outfall date; run;
314
315 /* convert p610lbdy to p610 mg/l */
316 /*
317 data temp610a;
318 set test.nc_nitrogen;
319 if screen = 'p610lbdy';
320 p610_mgl =(p610lbdy/flow_mgd)*.264*.000001*453*1000;
321 keep permit outfall date p600_mgl p600lbdy;
322 run;
323
324 proc sort; by permit outfall date; run;
325
326 /* convert p610lbyr to p610 mg/l */
327 /*
328 data temp610b;
329 set test.nc_nitrogen;
330 if screen = 'p610lbyr';
331 p610_mgl =((p610lbyr/365)/flow_mgd)*.264*.000001*453*1000;
332 keep permit outfall date p600_mgl p600lbyr;
333 run;
334
335 proc sort; by permit outfall date; run;
```

```
336  
337 data test.nc_nitrogen;  
338 merge test.nc_nitrogen temp625_30 temp600a temp600b temp610a temp610b;  
339 by permit outfall date;  
340 run;  
341  
342  
343  
344 proc contents data=test.nc_nitrogen; run;  
345 proc print data=test.nc_nitrogen(obs=25); run;  
346  
347  
348 /* median value and count for each parameter, by permit, outfall, year */  
349 /*  
350 data ckry94;  
351 set test.nc_nitrogen;  
352 if year = 1994;  
353 run;  
354 proc sort data=ckyr94; by permit outfall year; run;  
355  
356 proc univariate data=ckyr94 noprint;  
357 var p600_mgl p610_mgl p625_mgl p630_mgl;  
358 by permit outfall year;  
359 output out=temp94 median=med600_mgl med610_mgl med625_mgl med630_mgl  
360 n=n600_mgl n610_mgl n625_mgl n630_mgl;;  
361 run;  
362  
363 data ckry95;  
364 set test.nc_nitrogen;  
365 if year = 1995;  
366 run;  
367 proc sort data=ckyr95; by permit outfall year; run;  
368  
369 proc univariate data=ckyr95 noprint;  
370 var p600_mgl p610_mgl p625_mgl p630_mgl;  
371 by permit outfall year;  
372 output out=temp95 median=med600_mgl med610_mgl med625_mgl med630_mgl  
373 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
374 run;  
375  
376 data ckry96;  
377 set test.nc_nitrogen;  
378 if year = 1996;  
379 run;  
380 proc sort data=ckyr96; by permit outfall year; run;  
381  
382 proc univariate data=ckyr96 noprint;  
383 var p600_mgl p610_mgl p625_mgl p630_mgl;  
384 by permit outfall year;  
385 output out=temp96 median=med600_mgl med610_mgl med625_mgl med630_mgl  
386 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
387 run;  
388  
389  
390 data ckry97;  
391 set test.nc_nitrogen;
```

34 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
392 if year = 1997;  
393 run;  
394 proc sort data=ckyr97; by permit outfall year; run;  
395  
396 proc univariate data=ckyr97 noprint;  
397 var p600_mgl p610_mgl p625_mgl p630_mgl;  
398 by permit outfall year;  
399 output out=temp97 median=med600_mgl med610_mgl med625_mgl med630_mgl  
400 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
401 run;  
402  
403  
404 data ckyr98;  
405 set test.nc_nitrogen;  
406 if year = 1998;  
407 run;  
408 proc sort data=ckyr98; by permit outfall year; run;  
409  
410 proc univariate data=ckyr98 noprint;  
411 var p600_mgl p610_mgl p625_mgl p630_mgl;  
412 by permit outfall year;  
413 output out=temp98 median=med600_mgl med610_mgl med625_mgl med630_mgl  
414 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
415 run;  
416  
417  
418 data ckyr99;  
419 set test.nc_nitrogen;  
420 if year = 1999;  
421 run;  
422 proc sort data=ckyr99; by permit outfall year; run;  
423  
424 proc univariate data=ckyr99 noprint;  
425 var p600_mgl p610_mgl p625_mgl p630_mgl;  
426 by permit outfall year;  
427 output out=temp99 median=med600_mgl med610_mgl med625_mgl med630_mgl  
428 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
429 run;  
430  
431  
432 data ckyr00;  
433 set test.nc_nitrogen;  
434 if year = 2000;  
435 run;  
436 proc sort data=ckyr00; by permit outfall year; run;  
437  
438 proc univariate data=ckyr00 noprint;  
439 var p600_mgl p610_mgl p625_mgl p630_mgl;  
440 by permit outfall year;  
441 output out=temp00 median=med600_mgl med610_mgl med625_mgl med630_mgl  
442 n=n600_mgl n610_mgl n625_mgl n630_mgl;  
443 run;  
444  
445  
446 data ckyr01;  
447 set test.nc_nitrogen;
```

```
448 if year = 2001;
449 run;
450 proc sort data=ckyr01; by permit outfall year; run;
451
452 proc univariate data=ckyr01 noprint;
453 var p600_mgl p610_mgl p625_mgl p630_mgl;
454 by permit outfall year;
455 output out=temp01 median=med600_mgl med610_mgl med625_mgl med630_mgl
456 n=n600_mgl n610_mgl n625_mgl n630_mgl;
457 run;
458
459
460 data ckryr02;
461 set test.nc_nitrogen;
462 if year = 2002;
463 run;
464 proc sort data=ckryr02; by permit outfall year; run;
465
466 proc univariate data=ckryr02 noprint;
467 var p600_mgl p610_mgl p625_mgl p630_mgl;
468 by permit outfall year;
469 output out=temp02 median=med600_mgl med610_mgl med625_mgl med630_mgl
470 n=n600_mgl n610_mgl n625_mgl n630_mgl;
471 run;
472
473
474 data ckryr03;
475 set test.nc_nitrogen;
476 if year = 2003;
477 run;
478 proc sort data=ckryr03; by permit outfall year; run;
479
480 proc univariate data=ckryr03 noprint;
481 var p600_mgl p610_mgl p625_mgl p630_mgl;
482 by permit outfall year;
483 output out=temp03 median=med600_mgl med610_mgl med625_mgl med630_mgl
484 n=n600_mgl n610_mgl n625_mgl n630_mgl;
485 run;
486
487
488 data test.n_check_yr;
489 set temp94 temp95 temp96 temp97 temp98 temp99 temp00 temp01
490 temp02 temp03 ;
491 run;
492
493 proc sort; by permit outfall year; run;
494
495 PROC EXPORT DATA= TEST.N_CHECK_yr
496 OUTFILE= "Z:\Nutrients\SPARROW\point_source\nc\Nitrogen\n_ch
497 eck_yrtmp.xls"
498 DBMS=EXCEL2000 REPLACE;
499 RUN;
500
501
502
503 /* count number of quarters with data for each permit, by year and parameter */
```

36 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
504 /*
505 proc sort data=test.nc_nitrogen; by permit outfall year quarter; run;
506
507 proc univariate noprint;
508 var p600_mgl p610_mgl p625_mgl p630_mgl;
509 by permit outfall year quarter;
510 output out=test.q1 n=q600_mgl q610_mgl q625_mgl q630_mgl;
511 run;
512
513 proc print data=test.q1(obs=100); var permit outfall year quarter q600_mgl
514 q610_mgl q625_mgl q630_mgl; run;
515
516 data test.q2;
517 set test.q1;
518 if q600_mgl > 0 then q600_mgl = 1;
519 if q610_mgl > 0 then q610_mgl = 1;
520 if q625_mgl > 0 then q625_mgl = 1;
521 if q630_mgl > 0 then q630_mgl = 1;
522 keep permit outfall year quarter q600_mgl q610_mgl q625_mgl q630_mgl;
523 run;
524
525 proc sort; by permit outfall year; run;
526
527 proc univariate noprint;
528 var q600_mgl q610_mgl q625_mgl q630_mgl;
529 by permit outfall year;
530 output out=test.q3 sum=qs600_mgl qs610_mgl qs625_mgl qs630_mgl;
531 run;
532
533 data test.q4;
534 set test.q3;
535 keep permit outfall year qs600_mgl qs610_mgl qs625_mgl qs630_mgl;
536 run;
537
538 proc sort; by permit outfall year; run;
539
540 proc print data=test.q4(obs=100); run;
541
542
543 /* create final NC nitrogen point source data set; include site information
544 updated in Oct 2005, that includes identification of SIC codes used
545 in EPA Mississippi R. PS loads study */
546
547 proc sort data=test.nc_nitrogen; by permit outfall year; run;
548 proc sort data=test.q4; by permit outfall year; run;
549 proc sort data=test.n_check_yr; by permit outfall year; run;
550
551 data temppcs;
552 format permit $9.;
553 set test2.pcs_site_100505;
554 permit=npdes;
555 lat_pcs1=latitude;
556 lat_pcs2=lat_0805;
557 long_pcs1=longitude;
558 long_pcs2=long_0805;
559 if state = 'NC';
```

```
560 keep permit npdes epa_noaa flow lat_pcs1 lat_pcs2 long_pcs1 long_pcs2 sic_code sic_desc ;
561 run;
562 proc sort; by permit; run;
563 proc contents; run;
564
565 data test.n_final_031405;
566 merge test.nc_nitrogen test.q4 test.n_check_yr;
567 by permit outfall year;
568 run;
569
570 proc sort; by permit; run;
571
572 data test.n_final_031405;
573 merge temppcs test.n_final_031405 ;
574 by permit;
575 run;
576
577 proc print data=test.n_final_031405(obs=50); var permit outfall lat_pcs sic_pcs permit_flo; run;
578
579 proc contents; run;
580
581 quit;
```


1 Attachment 4: Data Preparation Program

```

2 /*
3  program: nc_loadprep_021606.sas
4  date: 10/13-20/05 11/9/05 1/5/06 1/12/06 1/25/06 2/6/06 2/16/06 2/23/06
5  5/5/06 6/12/06 6/19/06 6/23/06
6
7
8
9
10
11 ****
12
13 libname test 'Z:\Nutrients\SPARROW\point_source\load_1005\NC';
14 libname ncn 'z:\Nutrients\sparrow\point_source\nc\nitrogen';
15 libname ncp 'z:\Nutrients\sparrow\point_source\nc\phosphorus';
16 libname sc 'z:\Nutrients\sparrow\point_source\sc\Flow';
17 libname pcs 'z:\nutrients\sparrow\point_source\PCS_donahoo_0805';
18 libname site 'Z:\Nutrients\SPARROW\point_source\site_1005';
19 libname ncsite 'Z:\Nutrients\SPARROW\point_source\nc\site';
20 libname summary 'Z:\Nutrients\SPARROW\point_source\load_1005\summary';
21 libname noaa 'Z:\GIS\PointSources\NOAA_1991';
22 libname tpc 'Z:\Nutrients\SPARROW\point_source\load_1005\Rubin_0206';
23
24 options ps=54 ls=80;
25
26
27 /*
28 proc contents data=ncn.nc_nitrogen; run;
29 proc contents data=ncp.nc_phosphorus;run;
30 proc contents data=site.pcs_site_101205;run;*/
31
32 /* NC data — input flow, 600 and 665 data from NC DEHNR, summarize values by month to deal
33 with cases where there is more than one observation per month */
34 /*
35 data ncn;
36 format npdes $9.;
37 set ncn.nc_nitrogen;
38 npdes=permit;
39 month=month(datepart(date));
40 year=year(datepart(date));
41 quarter=qtr(datepart(date));
42 keep facility permit npdes outfall year quarter month p600_mgl flow_mgd;
43 run;
44 proc sort; by facility npdes permit outfall year quarter month ; run;
45
46
47 proc univariate noprint;
48 var p600_mgl flow_mgd;
49 by facility npdes permit outfall year quarter month ;
50 output out=ncn1 median=med600 medflow;
51 run;
52
53 data ncn2;
54 set ncn1;
55 p600_mgl=med600;

```

40 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
56 flow_mgd=medflow;
57 drop med600 medflow;
58 run;
59
60 proc sort; by npdes outfall year month; run;
61
62 data ncp;
63 format npdes $9.;
64 set ncp.nc_phosphorus;
65 npdes=permit;
66 month=month(datepart(date));
67 year=year(datepart(date));
68 quarter=qtr(datepart(date));
69 keep facility permit npdes outfall year quarter month p665_mgl flow_mgd;
70 run;
71
72 proc sort; by facility npdes permit outfall year quarter month ; run;
73
74 proc univariate noprint;
75 var p665_mgl flow_mgd;
76 by facility npdes permit outfall year quarter month ;
77 output out=ncp1 median=med665 medflow;
78 run;
79
80 data ncp2;
81 set ncp1;
82 p665_mgl=med665;
83 flow_mgd=medflow;
84 drop med665 medflow;
85 run;
86
87 proc sort; by npdes outfall year month; run;
88
89 data test.ncdmr_101305;
90 merge ncn2 ncp2;
91 by npdes outfall year month;
92 run;
93
94 data test.ncdmr_101305;
95 format outfall $4. date_nc datetime20.;
96 set test.ncdmr_101305;
97 if month = 12 or month = 1 or month = 2 then season = 'winter';
98 if month = 5 or month = 3 or month = 4 then season = 'spring';
99 if month = 6 or month = 7 or month = 8 then season = 'summer';
100 if month = 11 or month = 9 or month = 10 then season = 'fall';
101 nc_data=1;
102 run;
103
104 proc sort data=test.ncdmr_101305; by npdes outfall year month;; run;
105
106 proc print data=test.ncdmr_101305(obs=500); run;
107
108
109 /* combine NC and PCS data and sitefile data and create
110 season class variable */
111 /*
```

```
112  
113 data test.nc_pcs1;  
114 format date datetime20.;  
115 set pcs.dmr_flow_093005;  
116 month=month(datepart(date));  
117 year=year(datepart(date));  
118 quarter=qtr(datepart(date));  
119 if state = 'NC';  
120 keep npdes outfall year quarter month q1flow c2p600 c2p665;  
121 run;  
122  
123 proc sort ; by npdes outfall year quarter month; run;  
124  
125 proc univariate noprint;  
126 var c2p600 c2p665 q1flow;  
127 by npdes outfall year quarter month ;  
128 output out=ncp1 median=med600 med665 medflow;  
129 run;  
130  
131 data test.temp1;  
132 set ncp1;  
133 if month = 12 or month = 1 or month = 2 then season = 'winter';  
134 if month = 5 or month = 3 or month = 4 then season = 'spring';  
135 if month = 6 or month = 7 or month = 8 then season = 'summer';  
136 if month = 11 or month = 9 or month = 10 then season = 'fall';  
137 c2p600=med600;  
138 c2p665=med665;  
139 q1flow=medflow;  
140 pcs_data=1;  
141 drop med600 med665 medflow;  
142 run;  
143  
144 proc sort; by npdes outfall year month; run;  
145 proc sort data=test.ncdmr_101305; by npdes outfall year month;; run;  
146  
147  
148 data test.nc_pcs2;  
149 merge test.ncdmr_101305 test.temp1;  
150 by npdes outfall year month;  
151 run;  
152  
153  
154 proc sort data=test.nc_pcs2; by npdes; run;  
155  
156 data site;  
157 set site.pcs_site_101405;  
158 if state='NC';  
159 run;  
160  
161 data site2;  
162 set site;  
163 keep facility nc_facility npdes sic_code sic_desc epa_noaa permit_flo;  
164 run;  
165  
166 proc sort; by npdes; run;  
167
```

42 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
168 data test.temp78;
169 format npdes $9.;
170 set ncsite.nc_site_101105;
171 npdes=permit;
172 run;
173
174 proc sort data=test.temp78; by npdes; run;
175
176 data test.nc_pcs3;
177 merge test.nc_pcs2 site2 test.temp78;
178 by npdes;
179 run;
180
181 data test.nc_pcs4;
182 set test.nc_pcs3;
183 drop med600mgl med665mgl medflow_n medflow_p date_nc nobs600 nobs665 nobsnflow nobspflow;
184 run;
185
186 proc print data=test.nc_pcs4(obs=500); var npdes outfall month year season nc_data p600_mgl
187 flow_mgd pcs_data c2p600 q1flow sic_code; run;
188
189 /* define single flow variable and subset data for 2000-2002 and epa_noaa=1 and flowtype ne '4'
190 this assumes that flow_mgd is the appropriate flow variable from the state data
191 and q1flow is the appropriate flow data from the PCS data. This needs to be determined on a
192 state-by-state basis. Also assumption is made to use the state data in preference to the PCS data,
193 where both available. Data edits done after 2/23/06 checking; further editing (5/5/06) done
194 pursuant to error noted by Mike templeton for NC6033. Additional editing done based on results
195 of flowcheck_050506.sas program, which identified several discharge records that needed
196 to be deleted or edited. */
197 /*
198 data temp99;
199 set test.nc_pcs4;
200 if npdes eq 'NC0059251' and year=2002 and month = 4 ; flow_mgd = .017;
201 run;
202 proc sort; by outfall year month; run;
203
204 proc print; var npdes outfall year month q1flow flow_mgd name sic_code; run;
205
206
207 data test.nc_pcs5;
208 format flowtype $2. flowclass $15. ;
209 set test.nc_pcs4;
210 if nc_facility eq '' then nc_facility = facility;
211 if npdes = 'NC0064149' and year=2002 and month = 8 then flow_mgd=flow_mgd/1000;
212 if npdes eq 'NC0006033' and year = 2002 and (month = 11 or month = 12 ) then
213     p600_mgl=p600_mgl/10;
214 if not ( npdes eq 'NC0003255' and year = 2002 and month = 11 and outfall = '005' );
215 if npdes eq 'NC0029131' and year = 2000 and (month=6 or month = 7) then flow_mgd = .015;
216 if not ( npdes eq 'NC0030180' and year = 2002);
217 if npdes eq 'NC0052469' and year=2000 and (month=1 or month=2 or month=3 or month=10 or
218     month=11 or month=12) then q1flow=1.254 ;
219 if npdes eq 'NC0059251' and year=2002 and month = 4 then flow_mgd = .017;
220 if flow_mgd gt 0 and (q1flow eq '.' or q1flow eq '') then flowtype = '1';
221 if ( flow_mgd eq '.' or flow_mgd eq '') and q1flow gt 0 then flowtype = '2';
222 if flow_mgd gt 0 and q1flow gt 0 then flowtype = '3';
223 if ( flow_mgd eq 0 or flow_mgd eq '.' or flow_mgd eq '')
```

```

224      and (q1flow eq '.' or q1flow eq 0 or q1flow eq '') then flowtype = '4';
225      if flowtype = '1' then flow=flow_mgd;
226      if flowtype = '2' then flow = q1flow;
227      if flowtype = '3' then flow=flow_mgd;
228      if flowtype = '4' then flow='.';
229      if flow gt 0 and flow le 0.05 then flowclass = '1';
230      if flow gt .05 and flow le .2 then flowclass = '2';
231      if flow gt .2 and flow le 1 then flowclass = '3';
232      if flow gt 1 and flow le 5 then flowclass = '4';
233      if flow gt 5 then flowclass = '5';
234      if year ge 2000 and year le 2002;
235      if epa_noaa = 1;
236      run;
237
238 proc sort; by flowtype; run;
239
240 proc univariate noprint;
241 var flow flow_mgd q1flow;
242 by flowtype;
243 output out=flowstat median=medflo medflomgd medq1flo n=nflo nflomgd nq1flo;
244 run;
245
246 proc print data=flowstat; run;
247
248
249 data test.nc_pcs5; set test.nc_pcs5; if flowtype ne '4'; run;
250
251
252
253 /* analyze p600 and p665 data from state of NC and from PCS and create single, unified values
254 for 600 and 665. This assumes that p600_mgl and p665_mgl are the appropriate concentration
255 variables from the NC state data and c2p600 and c2p665 are the appropriate concentration data
256 from the PCS data. This needs to be determined on a state-by-state basis. Also assumption
257 is made to use the NC data in preference to the PCS data, where both available. Data edits
258 made after 2/23/06 data checking */
259
260 /*
261 data test.nc_pcs6;
262 format p600type $2. p665type $2.;
263 set test.nc_pcs5;
264 if npdes='NC0000272' and year=2002 and month = 12 then p600_mgl=p600_mgl/10;
265 if p600_mgl gt 0 and (c2p600 eq '.' or c2p600 eq 0) then p600type = '1';
266 if p665_mgl gt 0 and (c2p665 eq '.' or c2p665 eq 0) then p665type = '1';
267 if ( p600_mgl eq '.' or p600_mgl eq '') and c2p600 gt 0 then p600type = '2';
268 if ( p665_mgl eq '.' or p665_mgl eq '') and c2p665 gt 0 then p665type = '2';
269 if p600_mgl gt 0 and c2p600 gt 0 then p600type = '3';
270 if p665_mgl gt 0 and c2p665 gt 0 then p665type = '3';
271 if (p600_mgl eq 0 or p600_mgl eq '.') and (c2p600 eq '.' or c2p600 eq 0) then p600type = '4';
272 if (p665_mgl eq 0 or p665_mgl eq '.') and (c2p665 eq '.' or c2p665 eq 0) then p665type = '4';
273 run;
274
275
276 data test.nc_pcs7;
277 format type600 $10. type665 $10.;
278 set test.nc_pcs6;
279 if p600type = '1' or p600type = '3' then do;

```

44 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
280 p600=p600_mgl; orig600=1; type600='orig600'; end;
281 if p665type = '1' or p665type = '3' then do;
282 p665 = p665_mgl; orig665=1; type665='orig665'; end;
283 if p665type = '2' then do;
284 p665=c2p665; orig665=1; type665='orig665'; end;
285 run;
286
287 proc univariate noprint;
288 var p600 p665;
289 output out=test.nc600665 p99=p600_99 p665_99 p90=p600_90 p665_90 q3=p600_75 p665_75
290 median=p600_50 p665_50 q1=p600_25 p665_25 p10=p600_10 p665_10 p1=p600_01 p665_01
291 mean=p600_mean p665_mean;
292 run;
293
294 data test.nc600665;
295 set test.nc600665;
296 state='NC';
297 run;
298
299 proc print data=test.nc600665;
300 title 'Univariate stats for p600 and p665 and flow based on NC or PCS original data';
301 run;
302
303
304
305 /* summary stats for p600/665 */
306 /*
307 proc sort data=test.nc_pcs7; by flowclass sic_code; run;
308
309 proc univariate noprint;
310 var p600 p665;
311 by flowclass sic_code;
312 output out=p600665 median=med600 med665 n=n_600 n_665;
313 run;
314
315 data p600665;
316 format type600 $10. type665 $10.;
317 set p600665;
318 if med600 ne '.' or med665 ne '.';
319 if med600 gt 0 then type600 = 'p600';
320 if med665 gt 0 then type665 = 'p665';
321 run;
322
323 proc sort; by sic_code flowclass type600 type665; run;
324
325 proc print data=p600665;
326 title 'Summary stats for p600/665 observations using p600/665 ';
327 run;
328
329
330 /* Create seasonal median p600 p665 values for each permit/outfall (across 2000-2002 data)
331 and merge back into data set. Median seasonal values will be used in a couple of instances,
332 including where there is no value for p600/p665 from original NC and PCS data and where
333 individual monthly values are VERY large (e.g. gt 10 times the median seasonal value).
334 The goal is ultimatley to have a p600/p665 value for each record where there is a flow value
335 (all of the records have SIC that are associated with TN and TP discharge, per EPA study)
```

```
336 Seasonal median values for 600/665/flow also will be used in load calculations when there is not  
337 12 months of flow data for a given year  
338  
339 Also note that there are observations where flow = 0 or flow = blank and where p600/665 (either  
340 original PCS data for medsea values) is gt 0. In these cases it is assumed that flow exists for  
341 that NPDES/outfall. A median seasonal flow value across the 2000-2002 time period is developed  
342 and substituted where flow = 0 or blank and there is a positive value for p600/665 for that discharger. */  
343  
344 /*  
345 proc sort data=test.nc_pcs7; by npdes outfall season; run;  
346  
347 proc univariate noprint;  
348 var p600 p665 flow;  
349 by npdes outfall season;  
350 output out=median1 median=medsea600 medsea665 medseaflo;  
351 run;  
352  
353 proc sort data=median1; by npdes outfall season; run;  
354 proc sort data=test.nc_pcs7; by npdes outfall season; run;  
355  
356 data test.nc_pcs8;  
357 merge test.nc_pcs7 median1;  
358 by npdes outfall season;  
359 run;  
360  
361 data test.nc_pcs8;  
362 format type600 $10. type665 $10.;  
363 set test.nc_pcs8;  
364 if p600 = '.' and medsea600 gt 0 then do;  
365 type600 = 'medsea600'; sea600=1; p600=medsea600; end;  
366 if p665 = '.' and medsea665 gt 0 then do;  
367 type665='medsea665'; sea665=1; p665=medsea665; end;  
368 if flowclass eq ' ' and ( p600 gt 0 or p665 gt 0 ) then flowsub = 1;  
369 if flowsub = 1 then flow = medseaflo;  
370 run;  
371  
372 /* analyze which records have no values for p600/665 */  
373  
374 data temp55;  
375 set test.nc_pcs8;  
376 if flow gt 0 and p600 eq '.';  
377 run;  
378  
379 proc univariate noprint;  
380 var p600 ;  
381 output out=tempn nmiss=n_no600;  
382 run;  
383  
384 data temp56;  
385 set test.nc_pcs8;  
386 if flow gt 0 and p665 eq '.';  
387 run;  
388  
389 proc univariate noprint;  
390 var p665 ;  
391 output out=tempp nmiss=n_no665;
```

46 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
392 run;
393
394 data temp57;
395 set tempn tempp;
396 run;
397
398 proc print; title 'nobs of records with no p600/p665 after using medsea'; run;
399
400
401
402 /* analyze p600 p665 and flow data by flowclass, SIC and season (2000-2002 data) based
403 on combined MRB2 state and PCS data for records with p600/665 concentration data,
404 to create median values of 600 and 665 to use when there is no data from original data
405 (p600/665) or median data (by permit outfall season; medsea600/665. These TPC values
406 are created with Z:\Nutrients\SPARROW\point_source\load_1005\summary\flosicsea_060906.sas.
407 Median values are based on at least 5 observations */
408
409 /*
410 proc sort data=test.nc_pcs8; by flowclass sic_code season; run;
411 proc sort data=summary.flosicsea; by flowclass sic_code season; run;
412
413 data test.nc_pcs9;
414 merge test.nc_pcs8 summary.flosicsea;
415 by flowclass sic_code season;
416 run;
417
418 data test.nc_pcs9;
419 format type600 $10. type665 $10.;
420 set test.nc_pcs9;
421 if p600='.' and flosicsea600 gt 0 then do;
422 type600='fss600'; fss_600=1; p600=flosicsea600; end;
423 if p665='.' and flosicsea665 gt 0 then do;
424 type665='fss665'; fss_665=1; p665=flosicsea665; end;
425 run;
426
427 /* analyze which records have no values for p600/665 after flosicsea */
428
429 data temp55;
430 set test.nc_pcs9;
431 if flow gt 0 and p600 eq '.';
432 run;
433
434 proc univariate noprint;
435 var p600 ;
436 output out=tempn nmiss=n_no600;
437 run;
438
439 data temp56;
440 set test.nc_pcs9;
441 if flow gt 0 and p665 eq '.';
442 run;
443
444 proc univariate noprint;
445 var p665 ;
446 output out=tempp nmiss=n_no665;
447 run;
```

```
448  
449 data temp57;  
450 set tempn tempp;  
451 run;  
452  
453 proc print; title 'nobs of records with no p600/p665 after flosicsea'; run;  
454  
455 /* For situations where there is not an original value for p600/665 and there is not a value from  
456 flosicsea600/665, ./load_1005/summary/flowsicsea_060906.sas analyzes original p600/665 and flow  
457 data by flowclass and SIC (2000-2002 data) to create median seasonal values of 600 and 665 by  
458 flow class/SIC-code to use when there is no p600/665 data . A minimum of 5 observations must be  
459 present to use a median value for a flowclass/sic_code. */  
460  
461 /*  
462 proc sort data=test.nc_pcs9; by flowclass sic_code ; run;  
463 proc sort data=summary.flosic; by flowclass sic_code ; run;  
464  
465 data test.nc_pcs9a;  
466 merge test.nc_pcs9 summary.flosic;  
467 by flowclass sic_code ;  
468 run;  
469  
470 data test.nc_pcs9b;  
471 format type600 $10. type665 $10.;  
472 set test.nc_pcs9a;  
473 if p600='.' and flosic600 gt 0 then do;  
474 type600='fs600'; fs_600=1; p600=flosic600; end;  
475 if p665='.' and flosic665 gt 0 then do;  
476 type665='fs665'; fs_665=1; p665=flosic665; end;  
477 run;  
478  
479 /* analyze which records have no values for p600/665 */  
480  
481 data temp55;  
482 set test.nc_pcs9b;  
483 if flow gt 0 and p600 eq '.';  
484 run;  
485  
486 proc univariate noint;  
487 var p600 ;  
488 output out=tempn nmiss=n_no600;  
489 run;  
490  
491 data temp56;  
492 set test.nc_pcs9b;  
493 if flow gt 0 and p665 eq '.';  
494 run;  
495  
496 proc univariate noint;  
497 var p665 ;  
498 output out=tempp nmiss=n_no665;  
499 run;  
500  
501 data temp57;  
502 set tempn tempp;  
503 run;
```

48 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
504  
505 proc print; title 'nobs of records with no p600/p665 after flosic'; run;  
506  
507  
508 /* Use ./load_1005/summary/flowsicsea_060906.sas to create "typical  
509 facility concentration" (TPC) for TN and TP, using  
510 several data sources, to calculate loads when flow data exists and no  
511 other source of concentration data is available. Primary sources of TPC  
512 data is from a Tetra Tech analysis done for EPA, with data provided by  
513 EPA's Steve Rubin. Also SIC-based TPC are developed from composite MRB2  
514 data for this study. The lower of the nutrient TPC from the EPA/Tetra Tech  
515 and the NC/SC studies are used as the TPC. */  
516  
517  
518 proc sort data=test.nc_pcs9b; by sic_code ; run;  
519  
520 proc sort data=summary.mrb2_tpc_022306; by sic_code ; run;  
521  
522 data test.nc_pcs9c;  
523 merge test.nc_pcs9b summary.mrb2_tpc_022306;  
524 by sic_code ;  
525 run;  
526  
527 data test.nc_pcs9d;  
528 format type600 $10. type665 $10.;  
529 set test.nc_pcs9c;  
530 if p600='.' and tpc600 gt 0 then do;  
531 type600='tpc600'; tpc_600=1; p600=tpc600; end;  
532 if p665='.' and tpc665 gt 0 then do;  
533 type665='tpc665'; tpc_665=1; p665=tpc665; end;  
534 run;  
535  
536 /* analyze which records have no values for p600/665 after tpc */  
537  
538 data temp55;  
539 set test.nc_pcs9d;  
540 if flow gt 0 and p600 eq '.';  
541 run;  
542  
543 proc univariate noprint;  
544 var p600 ;  
545 output out=tempn nmiss=n_no600;  
546 run;  
547  
548 data temp56;  
549 set test.nc_pcs9d;  
550 if flow gt 0 and p665 eq '.';  
551 run;  
552  
553 proc univariate noprint;  
554 var p665 ;  
555 output out=tempp nmiss=n_no665;  
556 run;  
557  
558 data temp57;  
559 set tempn tempp;
```

```
560 run;  
561  
562 proc print; title 'nobs of records with no p600/p665 after TPC'; run;  
563  
564  
565 /* count number of months and quarters with data, by permit/outfall, for 2000-2002 */  
566  
567 proc sort data=test.nc_pcs9d; by npdes outfall year ; run;  
568  
569 proc univariate noint;  
570 var flow;  
571 by npdes outfall year ;  
572 output out=mon_flow n=mon_flow;  
573 run;  
574  
575 proc sort data=test.nc_pcs9d; by npdes outfall year quarter ; run;  
576  
577 proc univariate noint;  
578 var flow;  
579 by npdes outfall year quarter ;  
580 output out=qtr1 n=qtr_flow;  
581 run;  
582  
583 data qtr2;  
584 set qtr1;  
585 if qtr_flow>0 then qtr_flow=1;  
586 run;  
587  
588 proc sort; by npdes outfall year; run;  
589  
590 proc univariate noint;  
591 var qtr_flow;  
592 by npdes outfall year;  
593 output out=qtrstat sum=qtrflow;  
594 run;  
595  
596 proc sort data=qtrstat; by npdes outfall year; run;  
597 proc sort data=mon_flow; by npdes outfall year; run;  
598 proc sort data=test.nc_pcs9; by npdes outfall year ; run;  
599  
600 data test.nc_pcs10;  
601 merge qtrstat mon_flow test.nc_pcs9d;  
602 by npdes outfall year ;  
603 run;  
604 proc contents data=test.nc_pcs10; run;  
605  
606  
607 /* create final data set, replacing very large values of p600/665 with median values  
608 (by npdes/outfall) First, look at distribution of p600/665. To be replaced, values of p600/665  
609 will have to be greater than 10 times the size of the median concentration value for all values  
610 for that discharger/outfall during 2000-2002 and greater than the 95th percentile value of all  
611 p600/665 values in the entire data set. */  
612  
613 proc sort data=test.nc_pcs10; by npdes outfall; run;  
614  
615 proc univariate data=test.nc_pcs10;
```

50 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
616 var p600 p665;
617 run;
618
619 /* TN stats */
620
621 proc sort data=test.nc_pcs10; by npdes outfall; run;
622
623 data tn;
624 set test.nc_pcs10;
625 if p600 gt 0;
626 run;
627
628 proc univariate noint;
629 var p600;
630 by npdes outfall;
631 output out=temptn median=median_p600;
632 run;
633
634 proc sort data=temptn; by npdes outfall; run;
635
636 /* TP stats */
637
638 proc sort data=test.nc_pcs10; by npdes outfall; run;
639
640 data tp;
641 set test.nc_pcs10;
642 if p665 gt 0;
643 run;
644
645 proc univariate noint;
646 var p665;
647 by npdes outfall;
648 output out=temptp median=median_p665;
649 run;
650
651 proc sort data=temptp; by npdes outfall; run;
652
653 data test.temp88;
654 merge test.nc_pcs10 temptn temptp;
655 by npdes outfall;
656 run;
657
658
659 data test.nc_pcs11;
660 format type600 $10. type665 $10. high_tn $10. high_tp $10.;
661 set test.temp88;
662 state ='NC';
663 if p600 gt (10*median_p600) then high_tn='10*med_tn';
664 if p665 gt (10*median_p665) then high_tp='10*med_tp';
665 if high_tn='10*med_tn' and p600 gt 37.1 then do;
666 p600=median_p600; hi600=1; type600='hi600'; end;
667 if high_tp='10*med_tp' and p665 gt 6.6 then do;
668 p665=median_p665; hi665=1; type665='hi665'; end;
669 run;
670
671 quit;
```

1 Attachment 5: Typical Pollutant-Concentration Program

```
2
3 /* ****
4
5 program : flowsicsea_060906.sas
6 date: 6/9/2006
7 from: flowsicsea_012706.sas (4/21/06)
8
9 This program is used to develop estimates of p600/665 based on composite data sets.
10 flowsicsea600/665 variables are median p600/665 values by flow class, SIC code,
11 and season, calculated using NC, VA, GA, MS, AL, TN, and SC data.
12 flowsic600/665 are median p600/665 values by flow class and SIC code. The program was revised
13 (6/9/06)
14 to incorporate code from rubin_tpc_022206.sas, allowing calculation of all SIC TPC values
15 in this program.
16
17 **** */
18
19 libname test 'Z:\Nutrients\SPARROW\point_source\load_1005\summary';
20 libname nc 'z:\Nutrients\sparrow\point_source\nc\nitrogen';
21 libname ncp 'z:\Nutrients\sparrow\point_source\nc\phosphorus';
22 libname sc 'z:\Nutrients\sparrow\point_source\sc\Flow';
23 libname pcs 'z:\nutrients\sparrow\point_source\PCS_donahoo_0805';
24 libname site 'Z:\Nutrients\SPARROW\point_source\site_1005';
25 libname ncsite 'Z:\Nutrients\SPARROW\point_source\nc\site';
26 libname scsite 'Z:\Nutrients\SPARROW\point_source\sc\site_1005';
27 libname nCDATA 'Z:\Nutrients\SPARROW\point_source\load_1005\NC';
28 libname fldata 'Z:\Nutrients\SPARROW\point_source\load_1005\FL';
29 libname scdata 'Z:\Nutrients\SPARROW\point_source\load_1005\SC';
30 libname msdata 'Z:\Nutrients\SPARROW\point_source\load_1005\MS';
31 libname aldata 'Z:\Nutrients\SPARROW\point_source\load_1005\al';
32 libname gadata 'Z:\Nutrients\SPARROW\point_source\load_1005\ga';
33 libname tndata 'Z:\Nutrients\SPARROW\point_source\load_1005\TN';
34 libname vadata 'Z:\Nutrients\SPARROW\point_source\load_1005\va_ky';
35 libname rubin 'Z:\Nutrients\SPARROW\point_source\load_1005\Rubin_0206';
36
37 /* analyze p600 p665 and flow data by flow class, SIC and season (2000-2002 data)
38 from SC, NC, VA, GA, AL, MS, and TN to create median seasonal values of 600 and 665 by flow class
39 to use when there is no year/month-specific original data (p600/665) or median data
40 (by permit outfall season; medsea600/665. A minimum of 5 observations must be
41 present to estimate a seasonal value for a flow class/sic_code. Use year-month-specific
42 data from states not data summarized by season (i.e. nc_pcs7 data sets).
43 */
44
45 data sctemp;
46 set scdata.nc_pcs7a;
47 keep npdes outfall year season sic_code p600 p665 flow flow class;
48 run;
49
50 data nctemp;
51 set nCDATA.nc_pcs7;
52 keep npdes outfall year season sic_code p600 p665 flow flow class;
53 run;
54
55 data vatemp;
```

52 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
56      set vadata.nc_pcs7;
57      keep npdes outfall year season sic_code p600 p665 flow flow class;
58  run;
59
60  data tntemp;
61  set tndata.nc_pcs7;
62  keep npdes outfall year season sic_code p600 p665 flow flow class;
63  run;
64
65  data mstemp;
66  set msdata.nc_pcs7;
67  keep npdes outfall year season sic_code p600 p665 flow flow class;
68  run;
69
70  data gatemp;
71  set gadata.nc_pcs7;
72  keep npdes outfall year season sic_code p600 p665 flow flow class;
73  run;
74
75  data altemp;
76  set alldata.nc_pcs7;
77  keep npdes outfall year season sic_code p600 p665 flow flow class;
78  run;
79
80  data test.ncscfl1;
81  set sctemp nctemp vatemp gatemp mstemp altemp tntemp;
82  run;
83
84 /* create flowsicsea variables. Examine values of flosicsea600/665 gt 75th pctl */
85
86  proc sort data=test.ncscfl1 ; by flow class sic_code season; run;
87
88  proc univariate noprint;
89  var p600 p665;
90  by flow class sic_code season;
91  output out=flosicseal median=flosicsea600 flosicsea665
92  n=n600flosicsea n665flosicsea ;
93  run;
94
95  data flosicsea;
96  set flosicseal;
97  if flosicsea600 gt 0 or flosicsea665 gt 0;
98  run;
99
100 data flosicseal;
101 set flosicsea;
102 if n600flosicsea ge 5 ;
103 keep flow class sic_code season flosicsea600 n600flosicsea;
104 run;
105
106 proc sort; by flow class sic_code season; run;
107
108 data flosicsea2;
109 set flosicsea;
110 if n665flosicsea ge 5 ;
111 keep flow class sic_code season flosicsea665 n665flosicsea;
```

```
112 run;  
113  
114 proc sort; by flow class sic_code season; run ;  
115  
116 data test.flosicsea;  
117 merge flosicsea1 flosicsea2;  
118 by flow class sic_code season;  
119 run;  
120  
121 proc print; 'MRB2 regional p600/665 values — FLO/SIC/SEA'; run;  
122  
123  
124 /* create flowsic variables */  
125  
126 proc sort data=test.ncscfl1 ; by flow class sic_code ; run;  
127  
128 proc univariate noprint;  
129 var p600 p665;  
130 by flow class sic_code ;  
131 output out=flosic1 median=flosic600 flosic665  
132 n=n600flosic n665flosic ;  
133 run;  
134  
135 data flosic;  
136 set flosic1;  
137 if flosic600 gt 0 or flosic665 gt 0;  
138 run;  
139  
140 data flosic1;  
141 set flosic;  
142 if n600flosic ge 5 ;  
143 keep flow class sic_code flosic600 n600flosic;  
144 run;  
145  
146 proc sort; by flow class sic_code ; run;  
147  
148 data flosic2;  
149 set flosic;  
150 if n665flosic ge 5 ;  
151 keep flow class sic_code flosic665 n665flosic;  
152 run;  
153  
154 proc sort; by flow class sic_code ; run ;  
155  
156 data test.flosic;  
157 merge flosic1 flosic2;  
158 by flow class sic_code ;  
159 run;  
160  
161 proc print; title 'MRB2 regional p600/665 values — FLO/SIC'; run;  
162  
163 /* create sic values */  
164  
165 proc sort data=test.ncscfl1 ; by sic_code ; run;  
166  
167 proc univariate noprint;
```

54 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
168 var p600 p665;
169 by sic_code ;
170 output out=sic1 median=sic600 sic665
171 n=n600sic n665sic ;
172 run;
173
174 data sic;
175 set sic1;
176 if sic600 gt 0 or sic665 gt 0;
177 run;
178
179 data sic1;
180 set sic;
181 if n600sic ge 5 ;
182 keep sic_code sic600 n600sic;
183 run;
184
185 proc sort; by sic_code ; run;
186
187 data sic2;
188 set sic;
189 if n665sic ge 5 ;
190 keep sic_code sic665 n665sic;
191 run;
192
193 proc sort; by sic_code ; run ;
194
195 data test.sic042106;
196 merge sic1 sic2;
197 by sic_code ;
198 run;
199
200 proc print; title 'MRB2 regional p600/665 values — SIC'; run;
201
202
203 /* This is the code incorporated from rubin_tpc_022206.sas.
204
205 import EPA/Tetra Tech TPC data from Steve Rubin 2/22/06 */
206
207
208 PROC IMPORT OUT= WORK.epa_tpc
209 DATAFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\Rubin
210 _0206\NEWTPC4.XLS"
211 DBMS=EXCEL REPLACE;
212 SHEET="MRB2";
213 GETNAMES=YES;
214 MIXED=NO;
215 SCANTEXT=YES;
216 USEDATE=YES;
217 SCANTIME=YES;
218 RUN;
219
220 data epa_tpc;
221 format sic_code $4.;
222 set epa_tpc;
223 sic_code=sic;
```

```
224 drop sic2 sic sicdesc_epa tn_epa tp_epa;  
225 run;  
226  
227 proc sort nodupkey ; by sic_code; run;  
228 proc contents; run;  
229 proc print; run;  
230  
231 /* import sic code/epa_noaa information and merge with pcs site information and merge */  
232  
233  
234 PROC IMPORT OUT= work.siccode  
235 DATAFILE= "Z:\GIS\PointSources\PCS_site_0205\sic_codes.xls"  
236 DBMS=EXCEL2000 REPLACE;  
237 GETNAMES=YES;  
238 RUN;  
239  
240 data sic_code;  
241 format sic_code $4.;  
242 set siccode;  
243 sic_code=sic3;  
244 drop sic_i sic3 sic;  
245 run;  
246  
247 proc sort; by sic_code; run;  
248  
249 data test.epa_tpc;  
250 merge epa_tpc sic_code;  
251 by sic_code;  
252 run;  
253  
254 data test.epa_tpc;  
255 set test.epa_tpc;  
256 if epa_noaa=1;  
257 run;  
258  
259 Proc sort; by sic_code; run;  
260  
261 proc contents; run;  
262 proc print; title 'Rubin-based SIC TPC'; run;  
263  
264 /* import MRB2 states-based sic TPC and merge */  
265  
266 proc sort data=test.sic042106; by sic_code; run;  
267  
268 data temp;  
269 merge test.epa_tpc test.sic042106;  
270 by sic_code;  
271 run;  
272  
273 data test.mrb2_tpc_022306;  
274 set temp;  
275 tpc600=epa_tn;  
276 tpc665=epa_tp;  
277 if epa_tn = '.' then tpc600=sic600;  
278 if epa_tp='.' then tpc665=sic665;  
279 if sic600 gt 0 and epa_tn gt sic600 then tpc600=sic600;
```

56 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
280 if sic665 gt 0 and epa_tp gt sic665 then tpc665=sic665;
281 drop epa_noaa count n600sic n665sic;
282 run;
283
284 proc print; title 'TN and TP concentrations, combined EPA national database and MRB2 PCS database';
285 run;
286
287
288
289 /* EXPORT FLOWSICSEA AND FLOSIC TABLES TO EXCEL */
290
291 PROC EXPORT DATA= TEST.FLOSIC
292 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\flosicsea_060906t.xls"
293 DBMS=EXCEL REPLACE;
294 SHEET="flosic";
295 RUN;
296
297 PROC EXPORT DATA= TEST.flosicsea
298 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\flosicsea_060906t.xls"
299 DBMS=EXCEL REPLACE;
300 SHEET="flosicsea";
301 RUN;
302
303 PROC EXPORT DATA= test.mrb2_tpc_022306
304 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\flosicsea_060906t.xls"
305 DBMS=EXCEL REPLACE;
306 SHEET="sic_tpc";
307 RUN;
308
```

1 Attachment 6: Load Calculation Program

```

2
3
4 /* ****
5
6 program: nc_loadcalc_021606.sas
7 date: 10/13-19/05 11/3/05 11/8-9/05 11-17-05 1-13-06 1-20-06 2/6-23/06
8   3/21/06 6/9/06 6/19/06
9
10 **** */
11
12 libname test 'Z:\Nutrients\SPARROW\point_source\load_1005\NC';
13 libname ncn 'z:\Nutrients\sparrow\point_source\nc\nitrogen';
14 libname ncp 'z:\Nutrients\sparrow\point_source\nc\phosphorus';
15 libname summary 'Z:\Nutrients\SPARROW\point_source\load_1005\summary';
16 libname pcs 'z:\nutrients\sparrow\point_source\PCS_donahoo_0805';
17 libname site 'Z:\Nutrients\SPARROW\point_source\site_1005';
18 libname ncsite 'Z:\Nutrients\SPARROW\point_source\nc\site';
19
20 options ps=54 ls=80;
21
22 /* Calculate loads for permits/outfall with 12 months of flow */
23
24 data test.temp1;
25 set test.nc_pcs11;
26 if mon_flow=12,
27 if month = 1 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
28 if month = 2 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*28;
29 if month = 3 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
30 if month = 4 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
31 if month = 5 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
32 if month = 6 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
33 if month = 7 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
34 if month = 8 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
35 if month = 9 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
36 if month = 10 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
37 if month = 11 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
38 if month = 12 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
39 if month = 1 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
40 if month = 2 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*28;
41 if month = 3 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
42 if month = 4 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
43 if month = 5 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
44 if month = 6 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
45 if month = 7 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
46 if month = 8 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
47 if month = 9 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
48 if month = 10 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
49 if month = 11 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
50 if month = 12 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
51 keep type600      type665    npdes    outfall    nc_facility      sic_code sic_desc
52 year      Month     qtrflow   mon_flow    flowclass      quarter   season   p600_mgl
53 flow_mgd p665_mgl    c2p600  c2p665  q1flow   flow      p600    orig600  p665
54 orig665 medsea600    medsea665   medseaflo      flowsub flosicsea600
55 flosicsea665

```

58 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
56 flosic600  flosic665 tpc600 tpc665 high_tn high_tp      kgnmg112      kgpmgl12;
57 run;
58
59 proc sort; by year npdes sic_code; run;
60
61 proc univariate noint;
62   var kgnmg112 kgpmgl12;
63   by year npdes sic_code;
64   output out=temp1 sum=kgn_12 kgp_12;
65 run;
66
67 data temp1;
68   set temp1;
69   calc12=1;
70 run;
71
72 proc sort data=temp1; by npdes year; run;
73
74
75
76
77 /* calculate loads for permits with 3 or 4 quarters of flow data (mon_flow lt 12)
78 Use seasonal values for p600/665. Load calculated for
79 a permit/outfall by season and summed for each discharger in each year.
80 proc univariate gets a single value for each year/permit/outfall/season for flow
81 (which varies across this combination of variables) and the various p600/665
82 variables (which are constant across this combination of variables and can be used
83 as by variables and carried over into next data set.*/
84
85 data test.temp2;
86   set test.nc_pcs11;
87   if mon_flow lt 12 and qtrflow ge 3;
88 run;
89
90 proc sort ; by npdes outfall sic_code year season ; run;
91
92 proc univariate noint;
93   var p600 p665 flow;
94   by npdes outfall sic_code year season ;
95   output out=temp3 median=med600 med665 medseaflo;
96 run;
97
98 data temp4;
99   set temp3;
100  p600=med600;
101  p665=med665;
102  drop med600 med665;
103 run;
104
105
106 data test.temp5;
107   set temp4;
108   if season='winter' then kgn_34qtr=((p600)*(medseaflo*3785000)*(0.000001))*90;
109   if season='spring' then kgn_34qtr=((p600)*(medseaflo*3785000)*(0.000001))*92;
110   if season='summer' then kgn_34qtr=((p600)*(medseaflo*3785000)*(0.000001))*92;
111   if season='fall' then kgn_34qtr=((p600)*(medseaflo*3785000)*(0.000001))*91;
```

```

112 if season='winter' then kgp_34qtr=((p665)*(medseaflo*3785000)*(0.000001))*90;
113 if season='spring' then kgp_34qtr=((p665)*(medseaflo*3785000)*(0.000001))*92;
114 if season='summer' then kgp_34qtr=((p665)*(medseaflo*3785000)*(0.000001))*92;
115 if season='fall' then kgp_34qtr=((p665)*(medseaflo*3785000)*(0.000001))*91;
116 run;
117
118 proc sort; by year npdes sic_code; run;
119
120 proc univariate noprint;
121 var kgn_34qtr kgp_34qtr;
122 by year npdes sic_code;
123 output out=temp6 sum=kgn_34qtr kgp_34qtr;
124 run;
125
126 data temp6;
127 set temp6;
128 calc34qtr=1;
129 run;
130
131 proc sort data=temp6; by npdes year; run;
132
133 /* calculate load for permit/outfall with less than 3 quarters of data */
134
135 data test.temp60;
136 set test.nc_pcs11;
137 if qtrflow lt 3;
138 if month = 1 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
139 if month = 2 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*28;
140 if month = 3 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
141 if month = 4 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
142 if month = 5 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
143 if month = 6 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
144 if month = 7 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
145 if month = 8 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
146 if month = 9 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
147 if month = 10 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
148 if month = 11 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
149 if month = 12 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
150 if month = 1 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
151 if month = 2 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*28;
152 if month = 3 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
153 if month = 4 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
154 if month = 5 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
155 if month = 6 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
156 if month = 7 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
157 if month = 8 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
158 if month = 9 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
159 if month = 10 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
160 if month = 11 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
161 if month = 12 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
162 keep type600      type665 npdes    outfall   nc_facility      sic_code sic_desc
163 year      Month   qtrflow   mon_flow      flowclass      quarter   season   p600_mgl
164 flow_mgd  p665_mgl   c2p600  c2p665   q1flow   flow      p600   orig600  p665
165 orig665  medsea600  medsea665  medseaflo  flowsub  flosicsea600
166 flosicsea665
167 flosic600  flosic665 tpc600  tpc665  high_tn  high_tp      kgnmgl12      kgpmgl12;

```

60 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
168 run;  
169  
170  
171  
172 proc sort; by year npdes sic_code; run;  
173  
174 proc univariate noprint;  
175 var kgnmgl12 kgpmgl12;  
176 by year npdes sic_code;  
177 output out=temp61 sum=kgnqtrlt3 kgpqtrlt3;  
178 run;  
179  
180 data temp61;  
181 set temp61;  
182 calcqtrlt3=1;  
183 run;  
184  
185 proc sort data=temp61; by npdes year; run;  
186  
187 data temp7;  
188 merge temp1 temp6 temp61;  
189 by npdes year;  
190 run;  
191  
192 data temp8;  
193 set temp7;  
194 if kgnqtrlt3 eq ‘.’ then kgnqtrlt3=0;  
195 if kgpqtrlt3 eq ‘.’ then kgpqtrlt3=0;  
196 if kgn_12 eq ‘.’ then kgn_12=0;  
197 if kgn_34qtr eq ‘.’ then kgn_34qtr=0;  
198 if kgp_34qtr eq ‘.’ then kgp_34qtr=0;  
199 if kgp_12 eq ‘.’ then kgp_12=0;  
200 kg_n=kgn_12+kgn_34qtr+kgnqtrlt3 ;  
201 kg_p=kgp_34qtr + kgp_12+kgpqtrlt3;  
202 run;  
203  
204 proc print; run;  
205  
206 /* create data set with indiv TN and TP load estimate variables by year */  
207  
208 data load00;  
209 set temp8;  
210 if year=2000;  
211 kgn_00=kg_n;  
212 kgp_00=kg_p;  
213 calc1200=calc12;  
214 calc34qtr00=calc34qtr;  
215 calcqtrlt300=calcqtrlt3;  
216 keep npdes sic_code kgn_00 kgp_00 calc1200 calc34qtr00 calcqtrlt300;  
217 run;  
218  
219 proc sort; by npdes sic_code; run;  
220  
221 data load01;  
222 set temp8;  
223 if year=2001;
```

```
224 kgn_01=kg_n;
225 kgp_01=kg_p;
226 calc1201=calc12;
227 calc34qtr01=calc34qtr;
228 calcqtrlt301=calcqtrlt3;
229 keep npdes sic_code kgn_01 kgp_01 calc1201 calc34qtr01 calcqtrlt301;
230 run;
231
232 proc sort; by npdes sic_code; run;
233
234 data load02;
235 set temp8;
236 if year=2002;
237 kgn_02=kg_n;
238 kgp_02=kg_p;
239 calc1202=calc12;
240 calc34qtr02=calc34qtr;
241 calcqtrlt302=calcqtrlt3;
242 keep npdes sic_code kgn_02 kgp_02 calc1202 calc34qtr02 calcqtrlt302;
243 run;
244
245 proc sort; by npdes sic_code; run;
246
247 data nc_load;
248 merge load00 load01 load02;
249 by npdes sic_code;
250 run;
251
252 proc sort; by npdes; run;
253
254 /* import NC site information and create NC load + site information */
255
256 data temp78;
257 format npdes $9.;
258 set ncsite.nc_site_101105;
259 npdes=permit;
260 run;
261 proc sort out=tempnc nodupkey ; by npdes ; run;
262
263 proc sort data=tempnc; by npdes; run;
264
265 data site;
266 set site.pcs_site_101405;
267 if state='NC';
268 keep npdes facility_0805 name_0805 flow_0805 lat_0805 long_0805
269 lat_0405 long_0405 sic_desc sic_code;
270 run;
271
272 proc sort; by npdes; run;
273
274
275 data temp99;
276 merge tempnc nc_load site;
277 by npdes;
278 run;
279
```

62 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
280 proc contents; run;
281
282 data tempload;
283 set temp99;
284 run;
285
286
287 /* flag load values where the percent difference from year to year exceeds
288 the 95th pctl or less than the 5 th pctl of all differences for those pairings of years.
289 Also flag load values greater than the 95th percentile of all load values */
290
291 data test.nc_load_site_110905;
292 format lat0405 $10. long0405 $10.;
293 set tempload;
294 flag=0;
295 pctn0001=(kgn_01-kgn_00)/kgn_00;
296 if pctn0001 ge 1 or pctn0001 lt -.44 then flag = 1;
297 if pctn0001 = ‘ ’ or pctn0001 = ‘.’ then flag =0;
298 pctn0102=(kgn_02-kgn_01)/kgn_01;
299 if pctn0102 ge 1 or pctn0102 lt -.53 then flag = 1;
300 if pctn0102 = ‘ ’ or pctn0102 = ‘.’ then flag =0;
301 pctp0001=(kgp_01-kgp_00)/kgp_00;
302 if pctp0001 ge 1 or pctp0001 lt -.47 then flag = 1;
303 if pctp0001 = ‘ ’ or pctp0001 = ‘.’ then flag =0;
304 pctp0102=(kgp_02-kgp_01)/kgp_01;
305 if pctp0102 ge 1 or pctp0102 lt -.47 then flag = 1;
306 if pctp0102 = ‘ ’ or pctp0102 = ‘.’ then flag =0;
307 if kgn_00 ne ‘ ’ or kgp_00 ne ‘ ’ or kgn_01 ne ‘ ’ or kgp_01 ne ‘ ’ or
308 kgn_02 ne ‘ ’ or kgp_02 ne ‘ ’;
309 if ( kgn_00 gt 75400 or kgn_01 gt 73900 or kgn_02 gt 76500 )or
310 ( kgp_00 gt 14300 or kgp_01 gt 13000 or kgp_02 gt 14700 ) then flag=1;
311 state=‘NC’;
312 lat0405=lat_0405; long0405=long_0405;
313 drop lat_0405 long_0405;
314 run;
315
316 proc contents; run;
317
318
319 proc univariate;
320 title ‘Summary stats for North Carolina TN and TP annual loads and for inter-annual changes’;
321 var kgn_00 kgn_01 kgn_02 kgp_00 kgp_01 kgp_02 pctn0001 pctn0102 pctp0001 pctp0102;
322 run;
323
324 proc contents data=TEST.temp1; run;
325 proc contents data=TEST.temp60; run;
326
327 /* export data to spreadsheet */
328
329
330
331 PROC EXPORT DATA= TEST.nc_load_site_110905
332 OUTFILE= “Z:\Nutrients\SPARROW\point_source\site_1005\nc\nc_loa
333 d_site_061206t.xls”
334 DBMS=EXCEL REPLACE;
335 SHEET=“load summary by discharger”;
```

```
336 RUN;
337
338
339 PROC EXPORT DATA= TEST.temp1
340 OUTFILE= "Z:\Nutrients\SPARROW\point_source\site_1005\nc\nc_loa
341 d_site_061206t.xls"
342 DBMS=EXCEL REPLACE;
343 SHEET="12 months";
344 RUN;
345
346 PROC EXPORT DATA= TEST.temp5
347 OUTFILE= "Z:\Nutrients\SPARROW\point_source\site_1005\nc\nc_loa
348 d_site_061206t.xls"
349 DBMS=EXCEL REPLACE;
350 SHEET="gt 3-4 qtrs";
351 RUN;
352
353 PROC EXPORT DATA= TEST.temp60
354 OUTFILE= "Z:\Nutrients\SPARROW\point_source\site_1005\nc\nc_loa
355 d_site_061206t.xls"
356 DBMS=EXCEL REPLACE;
357 SHEET="all the rest";
358 RUN;
359
360 /* export high flow values created in program flowcheck_050506.sas */
361
362 PROC EXPORT DATA= summary.flowcheck061306
363 OUTFILE= "Z:\Nutrients\SPARROW\point_source\site_1005\nc\nc_loa
364 d_site_061206t.xls"
365 DBMS=EXCEL REPLACE;
366 SHEET="flowcheck";
367 RUN;
368
369
370 quit;
```


Attachment 7: Load Summary Program

```

1  /*
2
3  program: summary_072106.sas
4  date: 1/25/06 2/7/06 2/15/06 2/21/06 3/21/06 4/10/06 6-9-2006
5  7/17-25/06
6  notes: This programs summarizes data developed in the loadprep and loadcalc
7  programs for individual states. It summarizes the original p600/665 data available
8  from state and PCS databases. It summarizes various attributes of the calculated loads.
9
10 ****
11 ****
12 ****
13
14 libname test 'Z:\Nutrients\SPARROW\point_source\load_1005\summary';
15 libname ncn 'z:\Nutrients\sparrow\point_source\nc\nitrogen';
16 libname ncp 'z:\Nutrients\sparrow\point_source\nc\phosphorus';
17 libname sc 'z:\Nutrients\sparrow\point_source\sc\Flow';
18 libname pcs 'z:\nutrients\sparrow\point_source\PCS_donahoo_0805';
19 libname site 'Z:\Nutrients\SPARROW\point_source\site_1005';
20 libname ncsite 'Z:\Nutrients\SPARROW\point_source\nc\site';
21 libname scsite 'Z:\Nutrients\SPARROW\point_source\sc\site_1005';
22 libname ncload 'Z:\Nutrients\SPARROW\point_source\load_1005\NC';
23 libname scload 'Z:\Nutrients\SPARROW\point_source\load_1005\SC';
24 libname flload 'Z:\Nutrients\SPARROW\point_source\load_1005\FL';
25 libname msload 'Z:\Nutrients\SPARROW\point_source\load_1005\MS';
26 libname alload 'Z:\Nutrients\SPARROW\point_source\load_1005\al';
27 libname gaload 'Z:\Nutrients\SPARROW\point_source\load_1005\ga';
28 libname tnload 'Z:\Nutrients\SPARROW\point_source\load_1005\TN';
29 libname vaload 'Z:\Nutrients\SPARROW\point_source\load_1005\VA_KY';
30 libname sparrow 'D:\sagtsparrow_rf1\data';
31
32 options ps=54 ls=80;
33
34 /*
35
36 /* summarize p600/665 data for states that have data — original files prepared in
37 loadprep programs for various states, in section creating single p600/665 variable */
38 /*
39 proc sort data=flload.fl600665; by state; run;
40 proc sort data=ncload.nc600665; by state; run;
41 proc sort data=scload.sc600665; by state; run;
42 proc sort data=msload.ms600665; by state; run;
43 proc sort data=alload.al600665; by state; run;
44 proc sort data=gaload.ga600665; by state; run;
45 proc sort data=tnload.tn600665; by state; run;
46
47 data test.p600665sum_012506;
48 merge flload.fl600665 ncload.nc600665 scload.sc600665 msload.ms600665
49 alload.al600665 gaload.ga600665 tnload.tn600665;
50 by state;
51 run;
52
53 proc print; run;
54
55 PROC EXPORT DATA= test.p600665sum_012506

```

66 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
56      OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\  
57 npdes_sum_071706t.xls"  
58 DBMS=EXCEL REPLACE;  
59 SHEET="p600665_stats";  
60 RUN;  
61  
62 /* ***** */  
63  
64 /* summarize types of p600/665 data used and number of flow obs */  
65 /*  
66 data test.summary;  
67 set aloload.nc_pcs11 ncload.nc_pcs11 flload.nc_pcs11 scload.nc_pcs11 msload.nc_pcs11  
68 gaload.nc_pcs11 tnload.nc_pcs11 vaload.nc_pcs11;;  
69 run;  
70  
71 proc contents; run;  
72  
73 data test.tempflo;  
74 set test.summary;  
75 if flow gt 0;  
76 run;  
77 proc sort ; by state ; run;  
78  
79 proc univariate noprint;  
80 var flow;  
81 by state ;  
82 output out=sumflo n=n_flow;  
83 run;  
84  
85  
86 PROC EXPORT DATA= work.sumflo  
87 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
88 DBMS=EXCEL REPLACE;  
89 SHEET="observations with flow";  
90 RUN;  
91  
92 proc sort data=test.summary ; by state ; run;  
93  
94 proc univariate noprint;  
95 var orig600 orig665 sea600 sea665 fss_600 fss_665 fs_600 fs_665  
96 tpc_600 tpc_665 hi600 hi665 flowsub;  
97 by state ;  
98 output out=sum sum=sumorig600 sumorig665 sumsea600 sumsea665 sumfss600 sumfss665  
99 sumfs600 sumfs665 sumtpc600 sumtpc665 sumhi600 sumhi665 sumflosub;  
100 run;  
101  
102  
103 PROC EXPORT DATA= work.sum  
104 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
105 DBMS=EXCEL REPLACE;  
106 SHEET="observations data source";  
107 RUN;  
108  
109 proc print data=sum; title 'data sources, by state';run;  
110  
111
```

```
112 data test.wwtp;
113 set test.summary;
114 if sic_code=4952;
115 run;
116
117 proc sort; by flowclass type600 ; run;
118
119 proc univariate noprint;
120 var p600 ;
121 by flowclass type600;
122 output out=flowstat median=med600 n=nobs;
123 run;
124
125 proc print data=flowstat; title 'WWTP(4952) median p600 and number of obs, by flowclass'; run;
126
127 PROC EXPORT DATA= WORK.FLOWSTAT
128 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
129 DBMS=EXCEL REPLACE;
130 SHEET="median_wwtp_type600";
131 RUN;
132
133 proc sort data=test.wwtp ; by flowclass type665 ; run;
134
135 proc univariate noprint;
136 var p665;
137 by flowclass type665;
138 output out=flowstat median= med665;
139 run;
140
141 proc print data=flowstat; title 'WWTP(4952) median p665 and number of obs, by flowclass'; run;
142
143 PROC EXPORT DATA= WORK.FLOWSTAT
144 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
145 DBMS=EXCEL REPLACE;
146 SHEET="median_wwtp_type665";
147 RUN;
148
149
150
151
152
153 proc sort data=test.summary ; by state type600 ; run;
154
155 proc univariate noprint;
156 var p600;
157 by state type600;
158 output out=type600 median=med600 n=n_600 ;
159 run;
160
161 PROC EXPORT DATA= work.type600
162 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
163 DBMS=EXCEL REPLACE;
164 SHEET="type600";
165 RUN;
166
167 proc print data=type600; title 'median p600 and number of obs, by state'; run;
```

68 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
168  
169  
170  
171  
172  
173 proc sort data=test.summary ; by state type665 ; run;  
174  
175 proc univariate noprint;  
176 var p665;  
177 by state type665;  
178 output out=type665 median=med665 n=n_665 ;  
179 run;  
180  
181 PROC EXPORT DATA= work.type665  
182 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
183 DBMS=EXCEL REPLACE;  
184 SHEET="type665";  
185 RUN;  
186  
187 proc print data=type665; title 'median p665 and number of obs, by state'; run;  
188  
189  
190  
191  
192  
193 proc sort data=test.summary ; by state flowclass ; run;  
194  
195 proc univariate noprint;  
196 var flow;  
197 by state flowclass;  
198 output out=flow median=medflow n=n_flow ;  
199 run;  
200  
201 PROC EXPORT DATA= work.flow  
202 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
203 DBMS=EXCEL REPLACE;  
204 SHEET="flow_state_flowclass";  
205 RUN;  
206  
207 proc print data=flow; title 'median flow and number of obs, by state and flowclass'; run;  
208  
209  
210  
211  
212 /* merge load data sets from individual states...final data has about 3000 obs */  
213  
214  
215 proc sort data=alload.al_load_011206;; by state; run;  
216 proc sort data=ncload.nc_load_site_110905; by state; run;  
217 proc sort data=scload.sc_load_121905; by state; run;  
218 proc sort data=flload.fl_load_site_112205; by state; run;  
219 proc sort data=msload.ms_load_010906; by state; run;  
220 proc sort data=gaload.ga_load_011206; by state; run;  
221 proc sort data=tnload.tn_load_032006; by state; run;  
222 proc sort data=vaload.va_load_072006; by state; run;  
223
```

```
224  
225  
226 data test.merge011306;  
227 merge alload.al_load_011206 ncload.nc_load_site_110905 scload.sc_load_121905  
228 flload.fl_load_site_112205 msload.ms_load_010906 gaload.ga_load_011206  
229 tnload.tn_load_032006 vaload.va_load_072006;  
230 by state;  
231 run;  
232  
233  
234 data test.temp1;  
235 set test.summary;  
236 if flow gt 0 ;  
237 run;  
238  
239 proc sort; by descending flow; run;  
240  
241 proc univariate;  
242 var flow;  
243 run;  
244  
245 proc sort; by year npdes outfall; run;  
246  
247 proc univariate noprint;  
248 var flow;  
249 by year npdes outfall;  
250 output out=temp99 median=med_flow_yr;  
251 run;  
252  
253 data y2000a;  
254 set temp99;  
255 if year=2000;  
256 run;  
257  
258 proc sort; by npdes descending med_flow_yr ; run;  
259  
260 proc sort out=temp2000 nodupkey; by npdes; run;  
261  
262 data y2000b;  
263 set temp2000;  
264 medflo2000=med_flow_yr;  
265 major2000=0;  
266 if medflo2000 gt 0 and medflo2000 le 0.05 then flowclass2000 = '1';  
267 if medflo2000 gt .05 and medflo2000 le .2 then flowclass2000 = '2';  
268 if medflo2000 gt .2 and medflo2000 le 1 then flowclass2000 = '3';  
269 if medflo2000 gt 1 and medflo2000 le 5 then flowclass2000 = '4';  
270 if medflo2000 gt 5 then flowclass2000 = '5';  
271 if medflo2000 ge 1 then major2000=1;  
272 keep npdes medflo2000 flowclass2000 major2000;  
273 run;  
274 proc sort; by npdes; run;  
275  
276 data y2001a;  
277 set temp99;  
278 if year=2001;  
279 run;
```

70 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
280  
281 proc sort; by npdes descending med_flow_yr ; run;  
282  
283 proc sort out=temp2001 nodupkey; by npdes; run;  
284  
285 data y2001b;  
286 set temp2001;  
287 medflo2001=med_flow_yr;  
288 major2001=0;  
289 if medflo2001 gt 0 and medflo2001 le 0.05 then flowclass2001 = '1';  
290 if medflo2001 gt .05 and medflo2001 le .2 then flowclass2001 = '2';  
291 if medflo2001 gt .2 and medflo2001 le 1 then flowclass2001 = '3';  
292 if medflo2001 gt 1 and medflo2001 le 5 then flowclass2001 = '4';  
293 if medflo2001 gt 5 then flowclass2001 = '5';  
294 if medflo2001 ge 1 then major2001=1;  
295 keep npdes medflo2001 flowclass2001 major2001;  
296 run;  
297  
298 proc sort; by npdes; run;  
299  
300 data y2002a;  
301 set temp99;  
302 if year=2002;  
303 run;  
304  
305 proc sort; by npdes descending med_flow_yr ; run;  
306  
307 proc sort out=temp2002 nodupkey; by npdes; run;  
308  
309 data y2002b;  
310 set temp2002;  
311 medflo2002=med_flow_yr;  
312 major2002=0;  
313 if medflo2002 gt 0 and medflo2002 le 0.05 then flowclass2002 = '1';  
314 if medflo2002 gt .05 and medflo2002 le .2 then flowclass2002 = '2';  
315 if medflo2002 gt .2 and medflo2002 le 1 then flowclass2002 = '3';  
316 if medflo2002 gt 1 and medflo2002 le 5 then flowclass2002 = '4';  
317 if medflo2002 gt 5 then flowclass2002 = '5';  
318 if medflo2002 ge 1 then major2002=1;  
319 keep npdes medflo2002 flowclass2002 major2002;  
320 run;  
321  
322 proc sort ; by npdes ; run;  
323  
324 proc sort data=test.merge011306; by npdes ; run;  
325  
326  
327 data test.merge041006;  
328 merge test.merge011306 y2000b y2001b y2002b;  
329 by npdes ;  
330 run;  
331  
332 proc contents; run;  
333 proc contents data=vaload.va_load_072006; run;  
334  
335
```

```

336 data loadcov;
337   set test.merge041006;
338   keep npdes lat_0405 long_0405 lat0405 long0405 lat_0805 long_0805 lat_nc long_nc lat_ms long_ms
339     lat_sc long_sc
340   lat_tn long_tn dd_lat dd_long ;
341 run;
342
343 PROC EXPORT DATA= work.loadcov
344   OUTFILE=
345     "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_latlong_071706t.xls"
346   DBMS=EXCEL REPLACE;
347   SHEET="npdes_lat_long_temp";
348 RUN;
349
350
351
352 proc sort data=test.merge041006; by state ; run;
353
354 proc univariate data=test.merge041006 nopolish;
355   var calc1200 calc1201 calc1202 calc34qtr00 calc34qtr01 calc34qtr02
356   calcqtrlt300 calcqtrlt301 calcqtrlt302;
357   by state ;
358   output out=amtdata sum= scalc1200 scalc1201 scalc1202 scalc34qtr00 scalc34qtr01 scalc34qtr02
359   scalcqtrlt300 scalcqtrlt301 scalcqtrlt302;
360 run;
361
362 PROC EXPORT DATA= work.amtdata
363   OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
364   DBMS=EXCEL REPLACE;
365   SHEET="load_calc_type_st_yr";
366 RUN;
367
368 proc print data=amtdata; title 'summary of ways that loads were calculated'; run;
369
370
371 data temp4952;
372   set test.merge041006;
373   if sic_code = '4952';
374 run;
375
376 proc sort ; by sic_code sic_desc flowclass2002 state; run;
377
378 proc univariate nopolish;
379   var kgn_02 kgp_02 ;
380   by sic_code sic_desc flowclass2002 state;
381   output out=kgsum_sicstfl sum=sicstfl_tn02 sicstfl_tp02;
382 run;
383
384 proc print data=kgsum_sicstfl; title 'TN and TP loads for SIC=4952, 2002, summed by state and
385   flowclass'; run;
386
387 PROC EXPORT DATA= work.kgsum_sicstfl
388   OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
389   DBMS=EXCEL REPLACE;
390   SHEET="kgsum2002_sicstfl";
391 RUN;

```

72 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
392  
393  
394 proc sort data=test.merge041006; by sic_code sic_desc; run;  
395  
396 proc univariate noprint;  
397 var kgn_00 kgn_01 kgn_02 kgp_00 kgp_01 kgp_02;  
398 by sic_code sic_desc;  
399 output out=kg_sum_sic sum=sic_tn00 sic_tn01 sic_tn02 sic_tp00 sic_tp01 sic_tp02  
400 n=nobs_n00 nobs_n01 nobs_n02 nobs_p00 nobs_p01 nobs_p02;  
401 run;  
402  
403 proc print data=kg_sum_sic; title 'TN and TP loads, summed by SIC-code'; run;  
404  
405 PROC EXPORT DATA= work.kg_sum_sic  
406 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
407 DBMS=EXCEL REPLACE;  
408 SHEET="kg_sum_sic";  
409 RUN;  
410  
411  
412  
413 proc sort data=test.merge041006; by state; run;  
414  
415 proc univariate noprint;  
416 var kgn_02 kgp_02 ;  
417 by state;  
418 output out=n_discharger n=n_discharger_tn02 n_discharger_tp02;  
419 run;  
420  
421 proc print data=n_discharger; title 'TN and TP dischargers,2002, summed by state'; run;  
422  
423 PROC EXPORT DATA= work.n_discharger  
424 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
425 DBMS=EXCEL REPLACE;  
426 SHEET="2002_dischargers";  
427 RUN;  
428  
429  
430  
431  
432 proc sort data=test.merge041006; by state flowclass2000; run;  
433  
434 proc univariate noprint;  
435 var kgn_00 kgp_00;  
436 by state flowclass2000;  
437 output out=kgnp00_flow sum=kgn00 kgp00 n=nob_n nobs_p;  
438 run;  
439  
440 PROC EXPORT DATA= work.kgnp00_flow  
441 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
442 DBMS=EXCEL REPLACE;  
443 SHEET="kgnp00_state_flow";  
444 RUN;  
445  
446  
447
```

```
448 proc sort data=test.merge041006; by state flowclass2001; run;
449
450 proc univariate noprint;
451 var kgn_01 kgp_01;
452 by state flowclass2001;
453 output out=kgnp01_flow sum=kgn01 kgp01 n=nob_n nobs_p;
454 run;
455
456 PROC EXPORT DATA= work.kgnp01_flow
457 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
458 DBMS=EXCEL REPLACE;
459 SHEET="kgnp01_state_flow";
460 RUN;
461
462
463
464 proc sort data=test.merge041006; by state flowclass2002; run;
465
466 proc univariate noprint;
467 var kgn_02 kgp_02;
468 by state flowclass2002;
469 output out=kgnp02_flow sum=kgn02 kgp02 n=nob_n nobs_p;
470 run;
471
472 PROC EXPORT DATA= work.kgnp02_flow
473 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
474 DBMS=EXCEL REPLACE;
475 SHEET="kgnp02_st_flow";
476 RUN;
477
478
479 /* import sic code information and merge into summary table */
480
481
482 PROC IMPORT OUT= sic_code
483 DATAFILE= "Z:\GIS\PointSources\PCS_site_0205\sic_codes.xls"
484 DBMS=EXCEL2000 REPLACE;
485 GETNAMES=YES;
486 RUN;
487
488 data sic_code;
489 format sic_code $4. sic_desc $70.;
490 set sic_code;
491 sic_code=sic;
492 keep sic_code sic_desc epa_noaa;
493 run;
494
495 PROC EXPORT DATA= work.sic_code
496 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
497 DBMS=EXCEL REPLACE;
498 SHEET="sic_code";
499 RUN;
500
501 /* compare loads calculated for dischargers with 12 months of 2002 data with original p600 and
502 p665 data with loads for same dischargers using TPC (sic only) concentrations */
503
```

74 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
504 data test.origdata;
505 set test.summary;
506 if year=2002 and mon_flow=12 and (orig600=1 or orig665=1);
507 if not(state='FL');
508 run;
509
510 data orig;
511 set test.origdata;
512 if month = 1 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
513 if month = 2 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*28;
514 if month = 3 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
515 if month = 4 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
516 if month = 5 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
517 if month = 6 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
518 if month = 7 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
519 if month = 8 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
520 if month = 9 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
521 if month = 10 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
522 if month = 11 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*30;
523 if month = 12 then kgnmgl12=((p600)*(flow*3785000)*(0.000001))*31;
524 if month = 1 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
525 if month = 2 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*28;
526 if month = 3 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
527 if month = 4 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
528 if month = 5 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
529 if month = 6 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
530 if month = 7 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
531 if month = 8 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
532 if month = 9 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
533 if month = 10 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
534 if month = 11 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*30;
535 if month = 12 then kgpmgl12=((p665)*(flow*3785000)*(0.000001))*31;
536 run;
537
538 proc sort; by year npdes sic_code flowclass; run;
539
540 proc univariate noprint;
541 var kgnmgl12 kgpmgl12;
542 by year npdes sic_code flowclass;
543 output out=orig2 sum=kgn_orig kgp_orig;
544 run;
545
546 proc sort data=orig2; by year npdes sic_code flowclass; run;
547
548 data tpc;
549 set test.origdata;
550 if month = 1 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
551 if month = 2 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*28;
552 if month = 3 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
553 if month = 4 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*30;
554 if month = 5 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
555 if month = 6 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*30;
556 if month = 7 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
557 if month = 8 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
558 if month = 9 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*30;
559 if month = 10 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
```

```

560 if month = 11 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*30;
561 if month = 12 then kgnmgl12=((tpc600)*(flow*3785000)*(0.000001))*31;
562 if month = 1 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
563 if month = 2 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*28;
564 if month = 3 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
565 if month = 4 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*30;
566 if month = 5 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
567 if month = 6 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*30;
568 if month = 7 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
569 if month = 8 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
570 if month = 9 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*30;
571 if month = 10 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
572 if month = 11 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*30;
573 if month = 12 then kgpmgl12=((tpc665)*(flow*3785000)*(0.000001))*31;
574 run;
575
576 proc sort; by year npdes sic_code flowclass; run;
577
578 proc univariate noprint;
579 var kgnmgl12 kgpmgl12;
580 by year npdes sic_code flowclass;
581 output out=tpc2 sum=kgn_tpc kgp_tpc;
582 run;
583
584 proc sort data=tpc2; by year npdes sic_code flowclass; run;
585
586 data test.compare;
587 merge orig2 tpc2;
588 by year npdes sic_code flowclass;
589 run;
590
591 data test.compare;
592 set test.compare;
593 kgn_dif=((kgn_orig-kgn_tpc)/kgn_orig)*100;
594 kgp_dif=((kgp_orig-kgp_tpc)/kgp_orig)*100;
595 kgn_dif=kgn_orig-kgn_tpc;
596 kgp_dif=kgp_orig-kgp_tpc;
597 run;
598
599 proc univariate;
600 title1 'Wilcoxon signed rank test to see if there is a difference in loads';
601 title2 'calculated with orig vs TPC';
602 var kgn_dif kgp_dif;
603 run;
604
605
606
607 /* **** */
608 /* **** */
609
610 /* input data about reach of dischargers and the stream meanq and watershed area
611 and summarize total 2002 TN and TP by by the type of receiving stream, in terms
612 of the mean annual flow of the reach receiving the discharge and the drainage area,
613 and also by major and minor dischargers. */
614
615 /* **** */

```

76 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
616 /* **** */
617
618 PROC IMPORT OUT= WORK.pts_catch
619 DATAFILE= "Z:\GIS\PointSources\sparrow\pts_catch.xls"
620 DBMS=EXCEL REPLACE;
621 SHEET="pts_catch$";
622 GETNAMES=YES;
623 MIXED=NO;
624 SCANTEXT=YES;
625 USEDATE=YES;
626 SCANTIME=YES;
627 RUN;
628
629 data pts_catch;
630 format npdes $9. wshed $5.;
631 set pts_catch;
632 state=substr(npdes,1,2);
633 npdes=npdes_arc;
634 wshed=e2rf1__;
635 drop npdes_arc e2rf1__;
636 run;
637
638 proc sort; by wshed; run;
639
640 data temp99;
641 set sparrow.sparrow_data1;
642 keep wshed meanq demtarea pskgn_02 pskgp_02;
643 run;
644
645 proc sort; by wshed; run;
646
647 data temp88;
648 merge temp99 pts_catch;;
649 by wshed;
650 run;
651
652 data temp88;
653 format areaclass $7.;
654 set temp88;
655 if npdes ne ' ';
656 if demtarea le 100 then areaclass = '1';
657 if demtarea gt 100 and areaclass le 1000 then areaclass = '2';
658 if demtarea gt 1000 and areaclass le 10000 then areaclass = '3';
659 if demtarea gt 10000 then areaclass = '4';
660 if wshed gt 80000 then areaclass = 'coastal';
661 run;
662
663 proc sort; by npdes; run;
664
665 proc sort data=test.merge041006; by npdes; run;
666
667 data test.merge072406;
668 merge temp88 test.merge041006;
669 by npdes;
670 run;
671
```

```
672  
673 proc sort; by state areaclass; run;  
674  
675 proc univariate noprint;  
676 var kgn_02 kgp_02;  
677 by state areaclass;  
678 output out=temp999 sum=sumkgn sumkgp;  
679 run;  
680  
681 PROC EXPORT DATA= work.temp999  
682 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"  
683 DBMS=EXCEL REPLACE;  
684 SHEET="rch_area_tn_tp_load";  
685 RUN;  
686  
687 data major1;  
688 set test.merge072406;  
689 if major2002=1;  
690 run;  
691  
692 proc sort ; by wshed ; run;  
693  
694 proc univariate noprint;  
695 var kgn_02 kgp_02;  
696 by wshed;  
697 output out=major2 sum=psn02_major psp02_major;  
698 run;  
699  
700 proc sort data=major2; by wshed; run;  
701  
702 data minor1;  
703 set test.merge072406;  
704 if major2002=0;  
705 run;  
706  
707 proc sort ; by wshed ; run;  
708  
709 proc univariate noprint;  
710 var kgn_02 kgp_02;  
711 by wshed;  
712 output out=minor2 sum=psn02_minor psp02_minor;  
713 run;  
714  
715 proc sort data=minor2; by wshed; run;  
716  
717 data test.point02;  
718 merge major2 minor2 temp999;  
719 by wshed;  
720 run;  
721  
722 data test.point02;  
723 set test.point02;  
724 if psn02_major='.' then psn02_major=0;  
725 if psp02_major='.' then psp02_major=0;  
726 if psn02_minor='.' then psn02_minor=0;  
727 if psp02_minor='.' then psp02_minor=0;
```

78 Methods for Estimating Annual Wastewater Nutrient Loads in the Southeastern United States

```
728 psn02_total=psn02_major+psn02_minor;
729 psp02_total=psp02_major+psp02_minor;
730 run;
731
732 data test.point02;
733 set test.point02;
734 if not(psn02_total=0 and psp02_total=0);
735 run;
736
737 proc print ;
738 title 'comparison of ps_02, before and after recnet round of program edits';
739 var wshed psn02_total pskgn_02 psp02_total pskgp_02;
740 run;
741
742
743 PROC EXPORT DATA= test.point02
744 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
745 DBMS=EXCEL REPLACE;
746 SHEET="point_source_02";
747 RUN;
748
749
750
751
752 PROC EXPORT DATA= TEST.merge072406
753 OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summary\npdes_sum_071706t.xls"
754 DBMS=EXCEL REPLACE;
755 SHEET="ncscflmsal";
756 RUN;
757
758
759 quit;
```

Attachment 8: Flow Magnitude Checking Program

```
/* ****
program: flowcheck_050506.sas
date: 05-05-06 6-12-06 6-22-06

*****
libname test 'Z:\Nutrients\SPARROW\point_source\load_1005\summary';

libname nCDATA 'Z:\Nutrients\SPARROW\point_source\load_1005\NC';
libname sCDATA 'Z:\Nutrients\SPARROW\point_source\load_1005\SC';
libname fldata 'Z:\Nutrients\SPARROW\point_source\load_1005\FL';
libname msdata 'Z:\Nutrients\SPARROW\point_source\load_1005\MS';
libname aldata 'Z:\Nutrients\SPARROW\point_source\load_1005\al';
libname gadata 'Z:\Nutrients\SPARROW\point_source\load_1005\ga';
libname tndata 'Z:\Nutrients\SPARROW\point_source\load_1005\TN';
libname vadata 'Z:\Nutrients\SPARROW\point_source\load_1005\VA_KY';
libname sparrow 'D:\sagtsparrow_rfl\data';

options ps=54 ls132;

/*
****

data sctemp;
set sCDATA.nc_pcs7a;
keep npdes outfall year month sic_code p600 p665 flow name;
run;

data nctemp;
format name $30.;
set nCDATA.nc_pcs7;
name=facility;
keep npdes outfall year month sic_code p600 p665 flow name;
run;

data vatemp;
set vadata.nc_pcs7;
keep npdes outfall year month sic_code p600 p665 flow name;
run;

data nttemp;
set tndata.nc_pcs7;
keep npdes outfall year month sic_code p600 p665 flow name;
run;

data mstemp;
set msdata.nc_pcs7;
keep npdes outfall year month sic_code p600 p665 flow name;
run;

data gatemp;
set gadata.nc_pcs7;
```

```

keep npdes outfall year month sic_code p600 p665 flow name;
run;

data altemp;
set alldata.nc_pcs7;
keep npdes outfall year month sic_code p600 p665 flow name;
run;  
  

data ftemp;
set fldata.nc_pcs7;
keep npdes outfall year month sic_code p600 p665 flow name;
run;  
  

data test.temp1;
set sctemp nctemp vatemp gatemp mstemp altemp tntemp ftemp;
run;  
  

proc sort data=test.temp1; by npdes outfall; run;  
  

data test.temp1;
set test.temp1;
if flow gt 0;
run;  
  

proc univariate noint;
var flow;
by npdes outfall;
output out=tempflow median=median_flow;
run;  
  

proc sort data=tempflow; by npdes outfall; run;  
  

data test.mrb2_flow_050506;
merge test.temp1 tempflow;
by npdes outfall;
run;  
  

proc univariate; var flow; run;  
  

data test.mrb2_flow_050506;
format highflow $10.;
set test.mrb2_flow_050506;
if flow gt 100 then highflow='flo>100MGD';
if flow gt (10*median_flow) then highflow='10*medflo';
if flow gt (100*median_flow) then highflow='100*medflo';
run;  
  

/* id flow values ge 1 (major dischargers, the source of most load) where the flow value
is either very high (ge 100 MGD; there was a significant break point in the distribution of flow values
beyond 100 MGD) or is much higher than the median flow value of that discharger */  
  

data test.flowcheck061306;
set test.mrb2_flow_050506;;
if flow ge 1 and (highflow='10*medflo' or highflow='100*medflo' or highflow='flo>100MGD');
run;
```

```
proc sort; by npdes; run;

proc print;
title 'records with flow ge 1 and flow gt 100 MGD or with flow gt 10/100* median flow';
var name npdes outfall year month sic_code flow median_flow highflow;
run;

/* export high flow values created in program flowcheck_050506.sas */

PROC EXPORT DATA= TEST.FLOWCHECK061306
OUTFILE= "Z:\Nutrients\SPARROW\point_source\load_1005\summar
y\flowcheck_062206.xls"
DBMS=EXCEL REPLACE;
SHEET="flowcheck_all";
RUN;
```

Quit;

