Reach Attributes

Modeling transport and fate of nutrients in water streams and reservoirs requires information about physical characteristics of the channel. Previous SPARROW models have modeled transport in streams and reservoirs as first-order contaminant loss, relating loss rates in streams to reach residence time (computed as quotient of reach length and reach mean annual stream velocity); and relating loss rates in reservoirs to areal hydraulic load (computed as quotient of mean annual reservoir outflow and surface area). The estimates used in the SAGT nutrient SPARROW model for reach length (length_m), mean annual velocity (meanv), mean annual streamflow (meanq), and water body surface area (surfarea_km) are taken from ERF1_2 (Enhanced River Reach File 2.0, Nolan and others, 2002). Assignment of estimates to the reaches unique to SAGT ERF1 that were added to accommodate monitoring site locations (described in the section "Hydrologic Network of Reaches and Associated Catchments") required additional steps. For these reaches, the attributes length and time of travel were recalculated using the length of the split reach, and other channel attributes were assigned values from the next downstream segment. The reach-level estimates of channel characteristics for all reaches in SAGT ERF1 are included in the file *SAGT_ERF1_input.xls.zip* (2.1 MB).

Mean Annual Nitrogen and Phosphorus Load at Stream Monitoring Sites

Measurements of nutrient water quality at stream monitoring sites from a combination of monitoring programs were used to develop observations of the response variable—mean annual nitrogen or phosphorus load—in the SPARROW regression equation. Mean annual load is estimated as the product of daily streamflow and estimated daily concentration, which is modeled from nutrient water-quality data and streamflow data.

Selection of Monitoring Sites

The nutrient water-quality data used for instream-load estimation were collected by Federal, State, and local agencies during 1975–2004. Data from the ambient monitoring programs of agencies other than the USGS (table 2) were obtained from either the STOrage and RETrieval (STORET) database of the USEPA or from individual State agency databases. Data from USGS monitoring programs were obtained from the National Water Information System (NWIS) database of the USGS.

Nutrient load was estimated for monitoring sites on streams and rivers—reservoir sites were excluded—from which samples were collected at least quarterly, with a minimum of 10 samples collected since 1995, for at least a 2-year period during which daily streamflow data also were collected (or could be estimated from a nearby gage). Although more than 3,000 sites in the region met the criteria for sampling frequency, only 782 sites had sufficient data for load estimation. Of the 782 sites, 202 were collocated with a USGS streamflow gaging station; the additional 580 sites were located close to and on the same stream as a USGS streamflow gaging station with at least 2 years of concurrent streamflow data. The criterion for close proximity between the paired water-quality and streamflow monitoring sites is based on the ratio between the drainage areas: the streamflow monitoring site was considered sufficiently close if the ratio was between 0.75 and 1.33. Station information for each of the 782 water-quality monitoring sites selected for load estimation, and for each corresponding streamflow gaging station, are included in the file *SAGT_monitoredload.xls* (700 KB).

Review and Revision of Nutrient Concentration Results

Data retrieved from all sources were reviewed and revised to a standard format. Revisions were of two types. First, cases of obviously erroneous concentration results (for example, an ammonia concentration value of 800 milligrams per liter [mg/L]) were identified and revised to missing values. Second, differences among data sources in the format or convention for recording results were resolved to a standard format. For example, an analytical result "less than 0.02" recorded as –0.02, or as 0.02 with remark code "K," would be converted to the qualifying-code convention used in NWIS, which is 0.02 with remark code "<." Two computer programs, *Reformat_ModSTORET_WQdata.sas* and *Convert_remarkcoding_and_otherproblematic.sas*, were used to revise the results to a standard format. The programs are coded in Statistical Analysis Systems (SAS) programming language (Statistical Analysis Systems Institute, 2000); text-file versions of these programs also are provided (*Reformat_ModSTORET_WQdata.txt* and *Convert_remarkcoding_and_otherproblematic.txt*).

Concentrations of total nitrogen (TN) were computed using analytical results for dissolved or total nitrite plus nitrate and total Kjeldahl nitrogen (TKN). Where TKN is missing, values were computed from analytical results for the separate constituents organic nitrogen and ammonia, or results for dissolved Kjeldahl nitrogen and suspended Kjeldahl nitrogen. Whenever two analytical results were combined to produce a value for a calculated parameter and either or both result was censored, rules were applied to produce the value and qualifying code for the calculated parameter. The procedures and rules for combining analytical results to produce a result for TN are described in the SAS computer program *Combine_nutrient_constituents.sas*.

Table 2. Sources of water-quality monitoring data used to estimate mean annual nutrient load.

[Program identifier corresponds to value of attribute 'Program' for the station in the SAGT_monitoredload dataset; agency abbreviation corresponds to the first 8 digits in the station identification (attribute 'station_id') in the SAGT_monitoredload dataset; NA, not applicable (agency abbreviation is not included in the station identification); NWIS, U.S. Geological Survey's National Water Information System; STORET, U.S. Environmental Protection Agency's STOrage and RETrieval system; Leg, Legacy; Mod, Modernized]

Monitoring agency	Program identifer	Agency abbreviation	Database
U.S. Geological Survey	USGS	NA	NWIS
U.S. Geological Survey, National Water-Quality Assessment Program	USGS-NAWQA	NA	NWIS
Tennessee Valley Authority	Tennessee Valley Authority	131TVAC	STORET - Leg for data through 1999, and file provided by Tyler Baker, TVA[1]
Alabama Department of Environmental Management	ALA DEPT ENVIRON MGMT	21AWIC	STORET - Leg
Alabama Department of Environmental Management (collaborating with Auburn University)	AU-ADEMResLd and AUM-ADEMResL	21AWIC	File provided by Lynn Sisk, ADEM[2]
Alabama Department of Environmental Management (collaborating with University of Alabama)	UA-ADEMResLd	21AWIC	File provided by Lynn Sisk, ADEM[2]
Florida Department of Environmental Protection	FLORIDA DEPT ENV PROTECTN	21FLA	STORET - Leg and Mod
Florida Department of Environmental Protection	FL DEPT OF ENVIRON REG	21FLBFA	STORET - Leg and Mod
Florida Department of Environmental Protection	FL Dept. of Environmental Protection	21FLGW	STORET - Leg and Mod
Hillsborough County Environmental Protection Commission	Hillsborough County Environmental (Florida)	21FLHILL	STORET - Leg and Mod
Hillsborough County Environmental Protection Commission	HILLS COUNTY ENV	21FLHILL	STORET - Leg and Mod
IMC Agrico Company	IMC Agrico (Florida)	21FLIMCA	STORET - Leg and Mod
Lake County Water Resource Management	Lake County Water Resource Management (Florida)	21FLLCPC	STORET - Leg and Mod
Manatee County Department of Environmental Management	Manatee County Environmental Management Dept (Florida)	21FLMANA	STORET - Leg and Mod
Manatee County Department of Environmental Management	ENVIRONMENTAL	21FLMANA	STORET - Leg and Mod
Orange County Environmental Protection Division	Orange County Environmental Protection (Florida)	21FLORAN	STORET - Leg and Mod
Orange County Environmental Protection Division	ORANGE COUNCY ENV	21FLORAN	STORET - Leg and Mod
Peace River Manasota Regional Water Supply Authority	Peace River Manasota Regional Water Supply Authority	FLPRMRWS	STORET - Leg and Mod
St. Johns Water Management District	St. Johns Water Management District	21FLSJWM	STORET - Leg and Mod
St. Johns Water Management District	ST. JOHN'S RIVER WATER	21FLSJWM	STORET - Leg and Mod
Suwanee River Water Management District	SUWANNEE R WTR MGNT DIST	21FLSUW	STORET - Leg and Mod
Southwest Florida Water Management District	Southwest Florida Water Management District	21FLSWFD	STORET - Leg and Mod
Volusia County Environmental Health Laboratory	VOLUSIA ENV HEALTH LAB	21FLVEMD	STORET - Leg and Mod
Volusia County Environmental Health Laboratory	Volusia County Environmental Health Lab (Florida)	21FLVEMD	STORET - Leg and Mod
Georgia Department of Natural Resources, Environmental Protection Division	GA DEPT OF NAT RESOURCES	21GAEPD	STORET - Leg and Mod
Kentucky Natural Resources and Environmental Protection Cabinet	KY DEPT NAT RES & ENV PRO	21KY	STORET - Leg and Mod
Mississippi Department of Environmental Quality	MISSISSIPPI DEPT NAT RES	21MSWQ	STORET - Leg for data through 1999, and file provided by Jeff Thomas, MSDEQ[3]
Mississippi Department of Environmental Quality	AMBN	21MSWQ	STORET - Leg for data through 1999, and file provided by Jeff Thomas, MSDEQ[3]
North Carolina Department of Environment and Natural Resources	NCDENR-DWQ (2nd)	21NC02WQ	STORET - Mod
South Carolina Department of Health and Environmental Control	SC DEPT HEALTH & ENV CON	21SC60WQ	STORET - Leg and Mod
South Carolina Department of Health and Environmental Control	SC PUBLIC SERVICE AUTHRTY	21SCSANT	STORET - Leg and Mod
Tennessee Department of Environment and Conservation	Tennessee Department of Environment and Conservation	TDECWPC	STORET - Leg and Mod (changes in station identification provided by Linda Cartwright, TDEC[4])
Virginia Department of Environmental Quality	VA DEPT OF ENVIRONMENTAL	21VASWCB	STORET - Leg for data through 1999, and database retrieval provided by Roger Stewart, VADEQ[5]

The authors gratefully acknowledge the following individuals for providing monitoring data to supplement data available in STORET:

[1] Tyler Baker, Tennessee Valley Authority.

[2] Lynn Sisk, Alabama Department of Environment and Management.

[3] Jeff Thomas, Mississippi Department of Environmental Quality.

[4] Linda Cartwright, Tennessee Department of Environment and Conservation.

[5] Roger Stewart, Virginia Department of Environmental Quality, Water Quality Monitoring Data Retrieval Application, *http://gisweb.deq.virginia.gov/monapp/mon_data_retrieval_app.html*, accessed December 2005.

Estimation of Nutrient Load Representing Long-Term Mean for 1975–2004, Normalized to 2002

Instream loads of nitrogen and phosphorus were estimated using bias-corrected, log-linear regression models, within the program Fluxmaster (Schwarz and others, 2006, based on methods described by Cohn and others, 1989 and 1992, and Gilroy and others, 1990). A special feature is available in Fluxmaster to compute temporally-detrended estimates of long-term mean load. Without detrending, the estimate of mean annual load for each station would represent mean conditions centered on the year at the midpoint of the station's concurrent concentration and streamflow record. The variability in the midpoint year (ranging from 1990 to 2003) for the set of stations in this analysis would introduce temporal bias in the estimates of mean load that could hamper spatial comparison of load. To compensate for this, the estimates of mean load are detrended, or normalized, to a common or base year; 2002 was selected in this analysis because it matches the period of nutrient-input estimates.

Daily mean concentration was modeled by regressing the available instantaneous measurements of nutrient concentration against the variables streamflow, season, and time:

$$\ln[C] = \beta 0 + \beta 1(\ln[Q]) + \beta 2(\ln[T]) + \beta 3(\text{sine}[2\pi T]) + \beta 4(\text{cosine}[2\pi T]) + e$$

where

ln[]	is	natural logarithm function;
C	is	instantaneous daily concentration;
Q	is	daily streamflow;
T	is	decimal time;
π	is	3.14169;
$\beta 0$–$\beta 4$	are	coefficients to be estimated in the regression analysis; and
e	is	model error.

Daily mean load is estimated as the product of estimated daily mean concentration and measured daily streamflow. The series of estimated daily values of mean load is then summed to produce a series of annual values of mean load. Computation of the detrended estimate of load normalized to 2002, however, requires an estimate of temporal trend in streamflow; temporal trend was modeled by regressing the streamflow record (daily mean values) against the variables season and time, incorporating an autoregressive process to specify the serial correlation structure and thus correct for serial correlation in errors inherent to the time-series data (Schwarz and others, 2006):

$$\ln[Q] = \beta 0 + \beta 1(\ln[T]) + \beta 2(\text{sine}[2\pi T]) + \beta 3(\text{cosine}[2\pi T]) + AR + e$$

where

ln[]	is	natural logarithm function;
Q	is	daily streamflow;
T	is	decimal time;
π	is	3.14169;
AR	is	an autoregressive model estimated with a specified number of lags, L (for this application, L = 30);
$\beta 0$–$\beta 4$	are	coefficients to be estimated in the regression analysis; and
e	is	model error.

Long-term mean nitrogen and phosphorus loads, normalized to 2002, were estimated for 637 (for nitrogen) and 747 (for phosphorus) of the 782 sites. The fewer number of sites with nitrogen load estimates reflects sparser concentration data for a chemical constituent, organic nitrogen, required to estimate total nitrogen concentration and load. The load estimates are included in the file *SAGT_monitoredload.xls* (700 KB). Careful consideration should be given to the fact that these estimates represent a hypothetical condition—the load that would have occurred at each station in 2002 if streamflow, and the relation between water quality and streamflow and season, corresponded to conditions detrended to 2002 from the available record during the period 1974–2005. This hypothetical load is useful for regional-scale assessments of water-quality conditions but should be used with caution for local-scale interpretations. For local-scale interpretations use of loads estimated for actual time periods and employing a more detailed regression analysis, such as stepwise linear regression and consideration of additional explanatory variables, is suggested.

Error Associated with Estimating Mean Nitrogen and Phosphorus Load from Monitoring Data

The standard error of the mean annual nitrogen load estimates, expressed as a percentage of the estimated value, for the 637 sites where nitrogen load was estimated was typically (for 80 percent of the sites) less than 15 percent. The standard error of the mean annual phosphorus load estimates for the 747 sites with load estimates was typically (for 80 percent of the sites) less than 25 percent. Large values of standard error (up to 90 percent of the estimate for nitrogen and more than 150 percent for phosphorus) at some sites reflect uncertainty in the calibration of the daily concentration model, or in the detrended estimate of mean streamflow and load (Schwarz and others, 2006, p. 27).

The 4-parameter log-linear regression approach used to model daily concentration for this analysis may be inadequate for estimating annual load accurately at some sites, for example where the concentration-streamflow relation is influenced by hysteresis or antecedent conditions. The decision to employ the 4-parameter model uniformly for all stations was considered appropriate for this regional-scale assessment of water quality and in view of the available resources.

Load estimates for water-quality monitoring sites that are not collocated with the associated streamflow monitoring sites are less certain due to uncertainty in the estimates of daily streamflow because the streamflow record for these sites had to be estimated based on an area-weighted adjustment of streamflow record from the nearby gage. Load estimates for water-quality sites for which the corresponding streamflow gage record is relatively short also are less certain due to uncertainty in estimating long-term mean streamflow. Loads estimated based on streamflow records shorter than 5 years may be biased due to short-term variation in streamflow, for example below-normal streamflow for 3 consecutive years. Estimates based on streamflow record of 5 years or less were screened for this bias by comparing the value of mean annual streamflow computed from the streamflow record with the value of runoff computed from the Unit Runoff Method (Bondelid and others, 1999); the sites for which these values differed by more than 40 percent were excluded from further analysis. These sites (17) are listed in the file *SAGT_monitored-load.xls* (700 KB) along with information about station location and record; load estimates are not shown for these sites.

Characteristics of Monitored Mean Annual Nitrogen and Phosphorus Load and Streamflow

Estimates of observed mean annual nitrogen and phosphorus load and mean annual streamflow, normalized to the base year 2002, are summarized in table 3. To facilitate spatial comparisons of instream loads at sites draining watersheds of differing size and streamflow characteristics, the load estimates at each site were scaled in two ways. First, the load estimate was divided by the total upstream area for the monitoring site, producing an estimate of yield in kilograms per hectare per year. Estimates of yield are useful for comparisons among sites of mass output, and comparison with inputs in a mass balance analysis. Second, the load estimate was divided by the mean annual streamflow at the monitoring site, producing the equivalent of the flow-weighted mean of the model-estimated daily concentrations in milligrams per liter. Estimates of flow-weighted mean concentration are useful for evaluating average water-quality conditions at the site and for comparisons with national datasets. For the purpose of spatial comparisons of mean annual streamflow among sites, the streamflow estimate at each site was scaled by dividing by the total upstream area, producing an estimate of runoff rate over the upstream area in inches per year.

Nitrogen

The mean value of the nitrogen yield estimates, normalized to 2002, for the 637 stations in the SAGT area for which nitrogen load could be estimated is 4.7 kilograms per hectare (kg/ha), median value is 3.8 kg/ha, and 10- and 90-percentile values are 1.9 and 7.7 kg/ha, respectively (table 3). This distribution is placed in context with the national distribution of stream nitrogen yield, by comparing with estimates of mean annual nitrogen yield for 477 sites monitored by the NAWQA Program during 1992–2001 (Mueller and Spahr, 2005). The median value of nitrogen yield estimates for the SAGT area (3.8 kg/ha) is similar to the median value for the national set (4.1 kg/ha); however the 90-percentile value for the national distribution is much larger (22 kg/ha compared to 7.7 kg/ha), as is the mean value for the national distribution (8.1 kg/ha compared to 4.7 kg/ha) (table 3).

The spatial pattern of mean annual nitrogen yield for 2002 is shown in figure 6, along with the boundaries of the hydrologic subregions. The highest 10 percent of observations of nitrogen yield (>7.9 kg/ha) occur at sites throughout the SAGT area; however, clusters of high-yield observations occur in the Peace-Tampa Bay subregion (0310), near metropolitan areas in central Georgia, Alabama, and North Carolina, and in the northeastern part of the Mobile-Tombigbee subregion (0316). The lowest 10 percent of observations of nitrogen yield (<1.9 kg/ha) occur throughout the eastern half of the SAGT area, and especially in the Chowan-Roanoke (0301) and Peace-Tampa Bay (0310) subregions.

The spatial pattern in monitored nitrogen yield was evaluated with respect to the hydrologic subregion boundaries using Tukey's multiple comparison test. Although nitrogen yield distribution overlapped among many subregions, distributions between some subregions were sufficiently different (at alpha = 0.05) to enable the division of the subregions into three groups: (1) the Lower Tennessee subregion (0603 and 0604) with consistently high observations of yield (mean value 12 kg/ha); (2) a grouping of 8 subregions with consistently low observations of yield (mean value ≤3.9 kg/ha), including the Chowan-Roanoke subregion (0301), the drainages to the Atlantic in South Carolina and Georgia, and three drainages to the Gulf (Suwanee, Ochlockonee, and Choctawhatchee-Escambia subregions; 0311, 0312, and 0314, respectively); and (3) a grouping of 11 subregions with yield observations ranging too widely within each subregion to permit characterization as consistently high or low. The hydrologic subregion framework is clearly not appropriate for delineating regions of relatively homogeneous nitrogen yield in the SAGT area; however, the subregion boundaries are useful for describing certain local-scale patterns.

The mean value of nitrogen flow-weighted mean concentration for the SAGT station set is 1.2 mg/L, median value is 0.95 mg/L, and 10- and 90-percentile values are 0.47 and 2.0 mg/L, respectively (table 3). As with estimates of yield, values of flow-weighted mean concentration for the SAGT area for mean and 90-percentile values are substantially lower than the national set. The spatial pattern of mean annual nitrogen flow-weighted concentration for 2002 is shown in figure 6B. Nitrogen yield and nitrogen flow-weighted mean concentration for monitoring sites in the SAGT are strongly correlated (r^2 = 0.68), suggesting that, in general, concentration variability explains variability of yield (fig. 7). In many cases, however, sites with the highest observations of nitrogen flow-weighted concentration do not have the highest observations

Table 3. Variation across the SAGT SPARROW model area in monitored mean annual load and concentration of nitrogen and phosphorus, and in monitored mean annual streamflow, normalized to 2002.

[km², square kilometer; kg/ha/yr, kilogram per hectare per year; mg/L, milligram per liter; Tukey class, interpreted to indicate difference in means such that means with the same Tukey class letter are not significantly different at 0.05; color indicator of Tukey rank: white/black (L/H) - subregions with values that are significantly different from and lower (L) / higher (H) than (respectively) each other; light gray/dark gray (O) - subregions for which distribution of values overlap distribution in other subregions]

Area	Subregion number	Subregion area, km²	Median of basin area for sites, km²	Number of sites	Nitrogen load, in kg/ha/yr							Nitrogen flow-weighted mean concentration, in mg/L						
					Mean value for subregion	Standard deviation	10-percentile	Median	90-percentile	Tukey class	Color indicator of Tukey class	Mean value for subregion	Standard deviation	10-percentile	Median	90-percentile	Tukey class	Color indicator of Tukey class
SAGT (all subregions combined)				637	**4.7**	3.9	1.9	3.8	7.7			**1.2**	0.8	0.47	0.95	2.0		
(Compared to nationwide set of NAWQA Cycle 1 sites)				481	**8.06**	11	0.54	4.1	22			**2.7**	4.7	0.33	1.3	6.8		
Subregion name																		
Chowan–Roanoke	0301	47,300	1,722	59	**2.7**	0.7	1.8	2.7	3.6	D	L	**0.73**	0.19	0.49	0.70	1.0	EF	O
Neuse–Pamlico	0302	33,900	437	39	**5.3**	4.2	2.3	4.0	9.3	BCD	O	**1.4**	1.1	0.80	1.1	2.3	ABCDE	O
Cape Fear	0303	25,100	838	26	**8.5**	7.8	3.0	5.1	21	AB	O	**1.9**	1.2	0.84	1.4	3.6	AB	H
Pee Dee	0304	47,900	3,357	34	**3.4**	1.5	2.1	3.2	4.6	D	L	**0.91**	0.37	0.52	0.90	1.2	DEF	O
Edisto–Santee	0305	61,100	884	74	**3.9**	4.5	1.5	2.5	5.6	D	L	**0.85**	0.81	0.36	0.62	1.2	DEF	O
Ogeechee–Savannah	0306	42,200	530	22	**2.3**	1.4	1.3	2.1	4.3	D	L	**0.50**	0.33	0.19	0.47	0.70	F	L
Altamaha–St. Marys	0307	53,000	2,846	22	**2.8**	1.4	1.7	2.0	4.8	D	L	**0.80**	0.32	0.53	0.66	1.3	DEF	O
St. Johns	0308	30,000	5,798	74	**4.8**	1.5	2.9	5.1	6.4	BCD	O	**1.5**	0.44	1.0	1.5	2.0	ABCD	O
Peace–Tampa Bay	0310	26,000	365	88	**5.3**	3.3	1.4	5.1	9.4	BCD	O	**1.7**	0.83	0.91	1.5	2.8	ABC	O
Suwanee	0311	35,700	4,838	28	**3.5**	3.5	2.2	2.8	4.6	D	L	**1.1**	0.81	0.82	1.0	1.3	BCDEF	O
Ochlockonee	0312	9,500	2,042	6	**3.1**	1.2	1.7	2.7	5.0	D	L	**0.76**	0.26	0.36	0.78	1.1	DEF	O
Apalachicola	0313	53,100	6,866	33	**5.7**	3.7	1.9	4.1	10	BCD	O	**1.0**	0.45	0.48	0.89	1.7	CDEF	O
Choctawhatchee–Escambia	0314	38,800	1,358	17	**3.4**	0.9	2.6	3.2	4.2	D	L	**0.5**	0.14	0.31	0.51	0.65	EF	O
Alabama (Coosa–Tallapoosa)	0315	58,800	1,051	27	**5.3**	1.7	3.9	4.8	6.9	BCD	O	**0.9**	0.38	0.62	0.79	1.2	DEF	O
Mobile–Tombigbee	0316	56,700	675	16	**7.9**	6.0	2.7	6.0	15	ABC	O	**1.3**	1.1	0.51	1.0	2.3	ABCDE	O
Pascagoula	0317	31,300	886	17	**4.5**	1.7	2.7	4.1	6.0	CD	O	**0.7**	0.37	0.41	0.64	1.1	DEF	O
Pearl	0318	22,600	4,715	8	**5.3**	1.2	3.6	5.3	7.9	BCD	O	**1.1**	0.26	0.73	1.0	1.6	CDEF	O
Upper Tennessee[a]	0601 and 0602	57,900	989	35	**4.7**	2.3	2.2	4.2	7.3	BCD	O	**0.8**	0.63	0.28	0.71	1.5	DEF	O
Lower Tennessee[b]	0603 and 0604	47,400	1,128	12	**12**	10	5.2	9.4	17	A	H	**2.1**	1.8	0.84	1.7	3.5	A	H
Sites with nitrogen load estimates placed on the SAGT ERF1_2 network				333	**4.8**	4.4	2.0	3.6	7.9			**1.1**	0.84	0.45	0.87	1.9		

Table 3. Variation across the SAGT SPARROW model area in monitored mean annual load and concentration of nitrogen and phosphorus, and in monitored mean annual streamflow, normalized to 2002.—Continued

[km^2, square kilometer; kg/ha/yr, kilogram per hectare per year; mg/L, milligram per liter; Tukey class, interpreted to indicate difference in means such that means with the same Tukey class letter are not significantly different at 0.05; color indicator of Tukey rank: white/black (L/H) - subregions with values that are significantly different from and lower (L) / higher (H) than (respectively) each other; light gray/dark gray (O) - subregions for which distribution of values overlap distribution in other subregions]

Area	Sub-region number	Sub-region area, km²	Median of basin area for sites, km²	Number of sites	Phosphorus load, in kg/ha/yr							Phosphorus flow-weighted mean concentration, in mg/L						
					Mean value for subregion	Standard deviation	10-percentile	Median	90-percentile	Tukey class	Color indicator of Tukey class	Mean value for subregion	Standard deviation	10-percentile	Median	90-percentile	Tukey class	Color indicator of Tukey class
SAGT (all subregions combined)				747	**0.66**	1.1	0.13	0.33	1.3			**0.17**	0.32	0.03	0.08	0.32		
(Compared to nationwide set of NAWQA Cycle 1 sites)				454	**0.71**	1.3	0.05	0.32	1.6			**0.28**	1.0	0.02	0.12	0.54		
Subregion name																		
Chowan–Roanoke	0301	47,300	1,722	59	**0.26**	0.10	0.16	0.24	0.41	C	L	**0.07**	0.03	0.04	0.06	0.11	B	L
Neuse–Pamlico	0302	33,900	437	39	**0.54**	0.43	0.21	0.44	0.95	C	L	**0.14**	0.10	0.07	0.12	0.23	B	L
Cape Fear	0303	25,100	838	26	**0.96**	0.88	0.35	0.61	3.0	ABC	O	**0.21**	0.13	0.09	0.17	0.49	B	L
Pee Dee	0304	47,900	3,214	34	**0.29**	0.24	0.11	0.27	0.49	C	L	**0.08**	0.07	0.03	0.07	0.13	B	L
Edisto–Santee	0305	61,100	1,043	84	**0.58**	1.04	0.14	0.28	0.95	C	L	**0.13**	0.20	0.03	0.07	0.21	B	L
Ogeechee–Savannah	0306	42,200	1,656	37	**0.28**	0.16	0.10	0.27	0.45	C	L	**0.07**	0.04	0.02	0.06	0.11	B	L
Altamaha–St. Marys	0307	53,000	2,031	41	**0.31**	0.34	0.09	0.23	0.64	C	L	**0.08**	0.06	0.03	0.06	0.14	B	L
St. Johns	0308	30,000	4,776	80	**0.32**	0.17	0.11	0.32	0.50	C	L	**0.09**	0.04	0.05	0.09	0.14	B	L
Peace–Tampa Bay	0310	26,000	361	88	**1.9**	2.31	0.09	1.3	4.1	A	H	**0.62**	0.73	0.05	0.37	1.5	A	H
Suwanee	0311	35,700	4,307	31	**0.50**	1.00	0.17	0.32	0.51	C	L	**0.15**	0.24	0.06	0.11	0.18	B	L
Ochlockonee	0312	9,500	2,403	7	**0.30**	0.11	0.13	0.33	0.44	C	L	**0.09**	0.06	0.02	0.08	0.16	B	L
Apalachicola	0313	53,100	3,281	70	**0.48**	0.80	0.12	0.27	0.70	C	L	**0.09**	0.14	0.03	0.05	0.13	B	L
Choctawhatchee–Escambia	0314	38,800	1,507	14	**0.21**	0.06	0.11	0.22	0.28	C	L	**0.03**	0.01	0.01	0.03	0.05	B	L
Alabama (Coosa–Tallapoosa)	0315	58,800	1,148	44	**0.65**	0.38	0.23	0.59	1.2	C	L	**0.11**	0.07	0.04	0.09	0.18	B	L
Mobile–Tombigbee	0316	56,700	461	14	**1.2**	1.4	0.14	0.55	3.8	ABC	O	**0.20**	0.24	0.02	0.10	0.58	B	L
Pascagoula	0317	31,300	886	17	**0.47**	0.31	0.07	0.45	0.95	C	L	**0.08**	0.06	0.01	0.08	0.18	B	L
Pearl	0318	22,600	4,715	8	**0.84**	0.40	0.41	0.74	1.7	BC	O	**0.17**	0.08	0.08	0.14	0.34	B	L
Upper Tennessee[a]	0601 and 0602	57,900	1,109	42	**0.42**	0.35	0.15	0.29	1.0	C	L	**0.07**	0.07	0.02	0.04	0.17	B	L
Lower Tennessee[b]	0603 and 0604	47,400	1,128	12	**1.8**	1.67	0.42	1.1	4.3	AB	O	**0.31**	0.27	0.06	0.22	0.73	B	L
Sites with phopshorus load estimates placed on the SAGT ERF1_2 network				378	**0.67**	1.2	0.14	0.35	1.4			**0.16**	0.33	0.03	0.08	0.32		

Table 3. Variation across the SAGT SPARROW model area in monitored mean annual load and concentration of nitrogen and phosphorus, and in monitored mean annual streamflow, normalized to 2002.—Continued

[km², square kilometer; kg/ha/yr, kilogram per hectare per year; mg/L, milligram per liter; Tukey class, interpreted to indicate difference in means such that means with the same Tukey class letter are not significantly different at 0.05; color indicator of Tukey rank: white/black (L/H) - subregions with values that are significantly different from and lower (L) / higher (H) than (respectively) each other; light gray/dark gray (O) - subregions for which distribution of values overlap distribution in other subregions]

Area	Subregion number	Subregion area, km²	Median of basin area for sites, km²	Number of sites[c]	Streamflow, in inches per year: Mean value for subregion	Standard deviation	10-percentile	Median	90-percentile	Tukey class	Color indicator of Tukey class
SAGT (all subregions combined)				759	**17**	7.2	11	16	26		
(Compared to nationwide set of NAWQA Cycle 1 sites)				490	**17**	12	2.4	14	25		
Subregion name											
Chowan–Roanoke	0301	47,300	1,722	59	**15**	1.9	13	14	18	EFGH	O
Neuse–Pamlico	0302	33,900	437	39	**14**	1.7	12	14	17	FGH	O
Cape Fear	0303	25,100	838	26	**16**	3.4	13	14	21	EFGH	O
Pee Dee	0304	47,900	3,245	35	**15**	2.8	12	14	16	EFGH	O
Edisto–Santee	0305	61,100	1,043	86	**17**	4.2	13	16	23	DEFGH	O
Ogeechee–Savannah	0306	42,200	1,572	38	**19**	11	11	16	43	CDEFG	O
Altamaha–St. Marys	0307	53,000	2,031	41	**14**	3.5	11	13	18	FGH	O
St. Johns	0308	30,000	4,776	80	**14**	4.9	9.0	15	17	GH	O
Peace–Tampa Bay	0310	26,000	365	89	**13**	8.6	5.7	12	21	H	L
Suwanee	0311	35,700	4,307	31	**12**	2.3	9.1	12	13	H	L
Ochlockonee	0312	9,500	2,403	7	**16**	6.8	9.4	13	26	EFGH	O
Apalachicola	0313	53,100	3,281	70	**21**	8.3	14	20	27	BCDE	O
Choctawhatchee–Escambia	0314	38,800	1,358	17	**28**	9.8	22	23	36	A	H
Alabama (Coosa–Tallapoosa)	0315	58,800	1,148	44	**23**	3.2	20	23	27	ABCD	O
Mobile–Tombigbee	0316	56,700	523	18	**23**	3.1	19	23	28	ABC	O
Pascagoula	0317	31,300	886	17	**25**	4.6	20	24	30	ABC	O
Pearl	0318	22,600	4,715	8	**20**	1.4	18	20	23	BCDE	O
Upper Tennessee[a]	0601 and 0602	57,900	1,109	42	**26**	7.7	17	24	34	AB	O
Lower Tennessee[b]	0603 and 0604	47,400	1,128	12	**22**	3.1	19	24	25	ABCD	O

[a] The Upper Tennessee subregion is defined for this summary as the combined subregions 0601 and 0602.

[b] The Lower Tennessee subregion is defined for this summary as the combined subregions 0603 and 0604. Monitored basins that include areas upstream from this subregion (that is, in the Upper Tennessee subregion) are excluded from the summary by subregion.

[c] Mean annual streamflow, and nitrogen or phosphorus or both nitrogen and phosphorus load, was estimated for a total of 782 sites in the SAGT model area. Streamflow and load are not reported or included in this summary for 21 of these sites, however, due to uncertainty about the boundary of the watershed contributing to the site or due to concern that the streamflow and load estimates may be biased by short streamflow record.

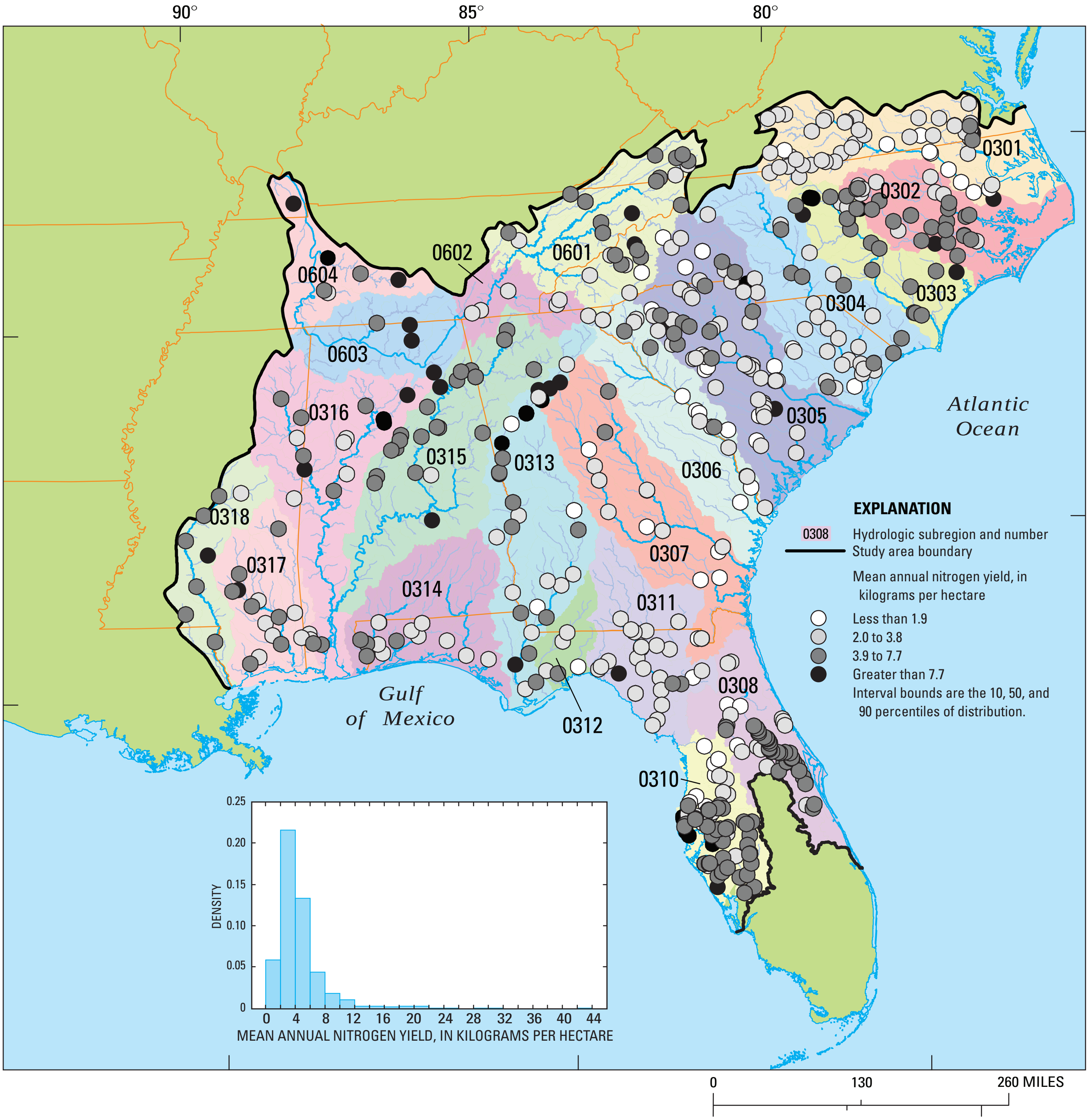

Figure 6A. Mean annual nitrogen yield estimated from stream monitoring data from 637 sites in the SAGT river basins, normalized to 2002.

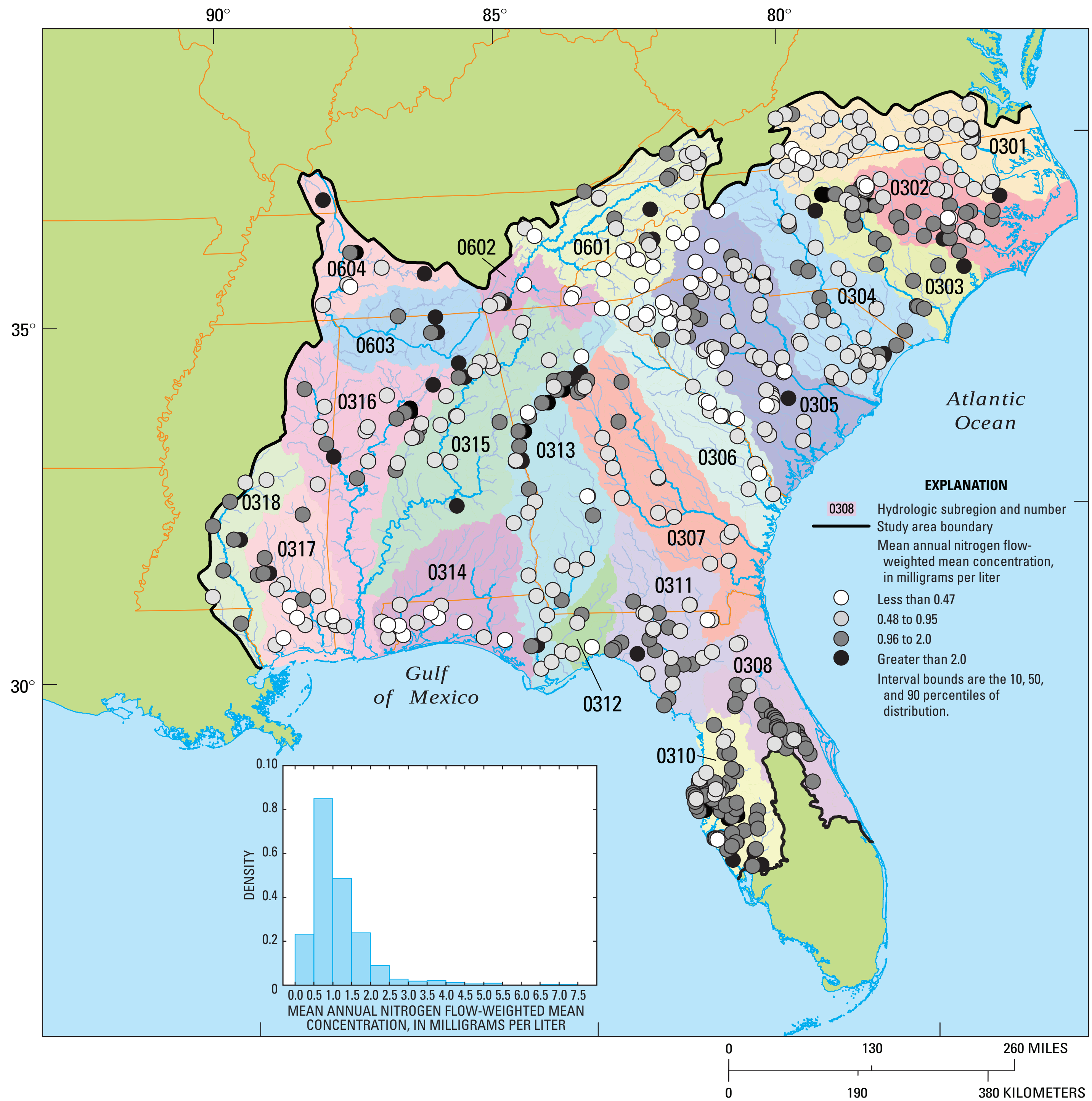

Figure 6B. Mean annual nitrogen flow-weighted mean concentration estimated from stream monitoring data from 637 sites in the SAGT river basins, normalized to 2002.

of nitrogen yield. For many of the sites in the St. Johns and Peace-Tampa Bay subregions with flow-weighted mean concentration among the highest (>2.0 mg/L) in the SAGT area, observations of nitrogen yield are among the lowest in the SAGT area (<1.0 kg/ha) (figs. 6 and 7). Conversely, many sites with relatively high yield values have relatively low values of flow-weighted concentration, such as sites in the Coosa-Tallapoosa subregion (0315) (fig. 6).

The noted divergence from a directly proportional relation is due to the fact that flow-weighted mean concentration varies as a function not only of mass yield, but also of streamflow yield, or runoff. The discrepancies between the spatial distribution of high and low values for nitrogen yield compared with the spatial distribution of nitrogen flow-weighted mean concentration (figures 6A and 6B) are, therefore, a function of differences in streamflow yield. Streamflow yield is relatively low for many sites in the St. Johns (0308) and Peace-Tampa Bay (0310) subregions, and relatively high for many sites in the Coosa-Tallapoosa (0315) and Upper Tennessee (0601 and 0602) subregions (figure 8A). The general pattern of variation in streamflow yield for the SAGT area is evident from the contoured surface prepared by Gebert and others (1987) and shown in figure 8B: streamflow yield is generally higher (>20 inches) in drainages to the Gulf extending eastward to the Ochlockonee subregion (0312) and in the Tennessee River basin, and generally lower (<20 inches) in drainages to the Atlantic and in the Peace-Tampa Bay (0310) and Suwanee (0311) subregions.

The influence of streamflow yield on the relation between yield and concentration is illustrated in figure 7 by the three lines showing the expected value of yield for a specified value of flow-weighted mean concentration assuming a specific streamflow yield of 11, 16, or 26 inches, which corresponds with the 10, 50, and 90 percentile of the distribution of streamflow yield for the load estimation sites in the SAGT area. For sites with low values of streamflow yield (plotting position to the left of the 11-inch line), nitrogen yield may be relatively low (<3 kg/ha) and flow-weighted concentration relatively high (>2 mg/L). Conversely, for some sites with high values of streamflow yield (plotting position to the right of the 26-inch line), nitrogen yield is relatively high (>20 kg/ha) and flow-weighted mean concentration relatively low (1.4 mg/L). Many of the sites with streamflow yield <11 inches or greater than 26 inches are located in drainage basins influenced by large springs (Miller, 1990) or losing reaches (Rumenik, 1988).

Variation of streamflow yield in the SAGT area is caused by variation in the volume of water from precipitation that is available for direct runoff, termed excess precipitation (fig. 8B). Calculated as the difference between precipitation and potential evaporation, estimates of excess precipitation (Wolock and others, 2003) correspond closely to contoured values of streamflow yield for most of the SAGT area. In many areas in Florida, however, contoured streamflow yield does not compare closely with estimates of excess precipitation because streamflow yield is affected by factors other than direct runoff from the

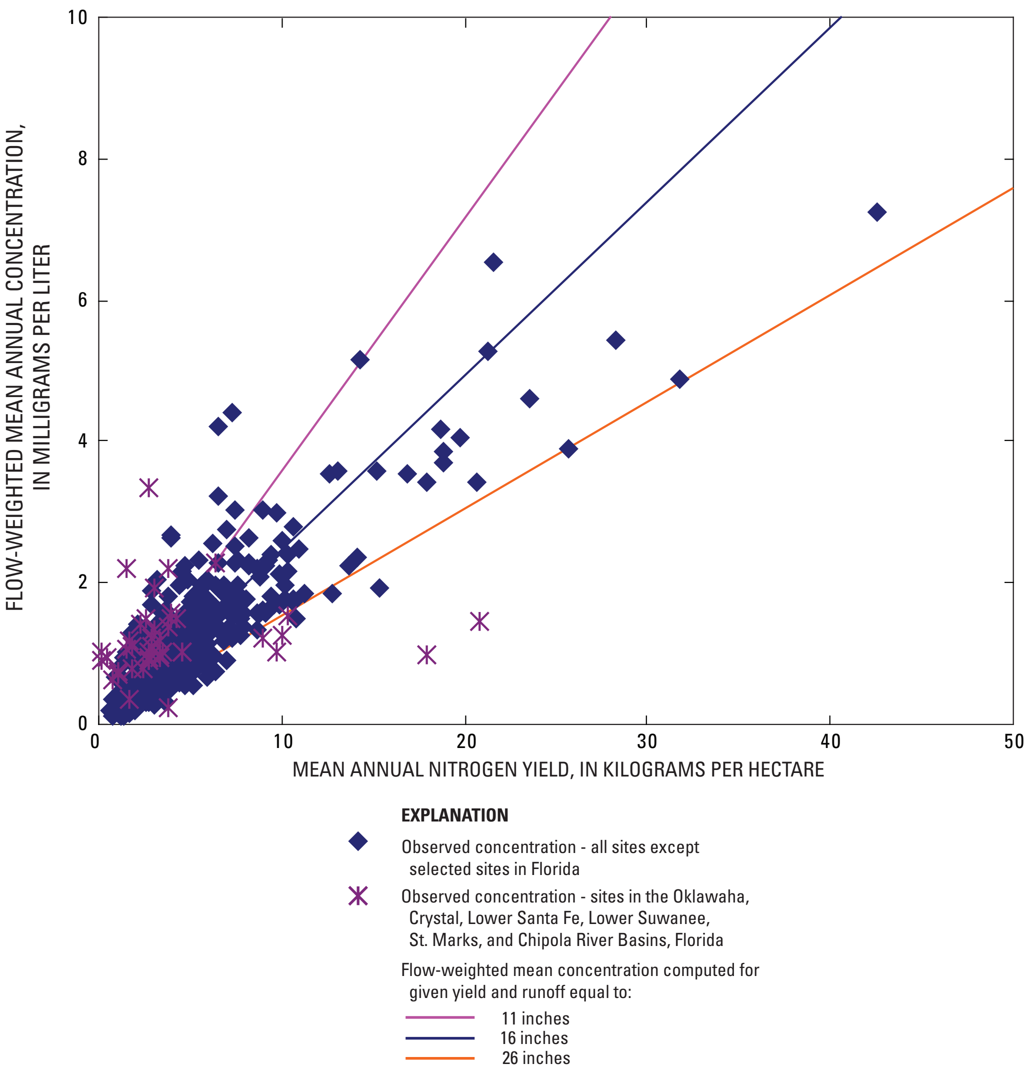

Figure 7. Relation of nitrogen flow-weighted mean concentration to nitrogen yield and runoff rate for the SAGT river basins.

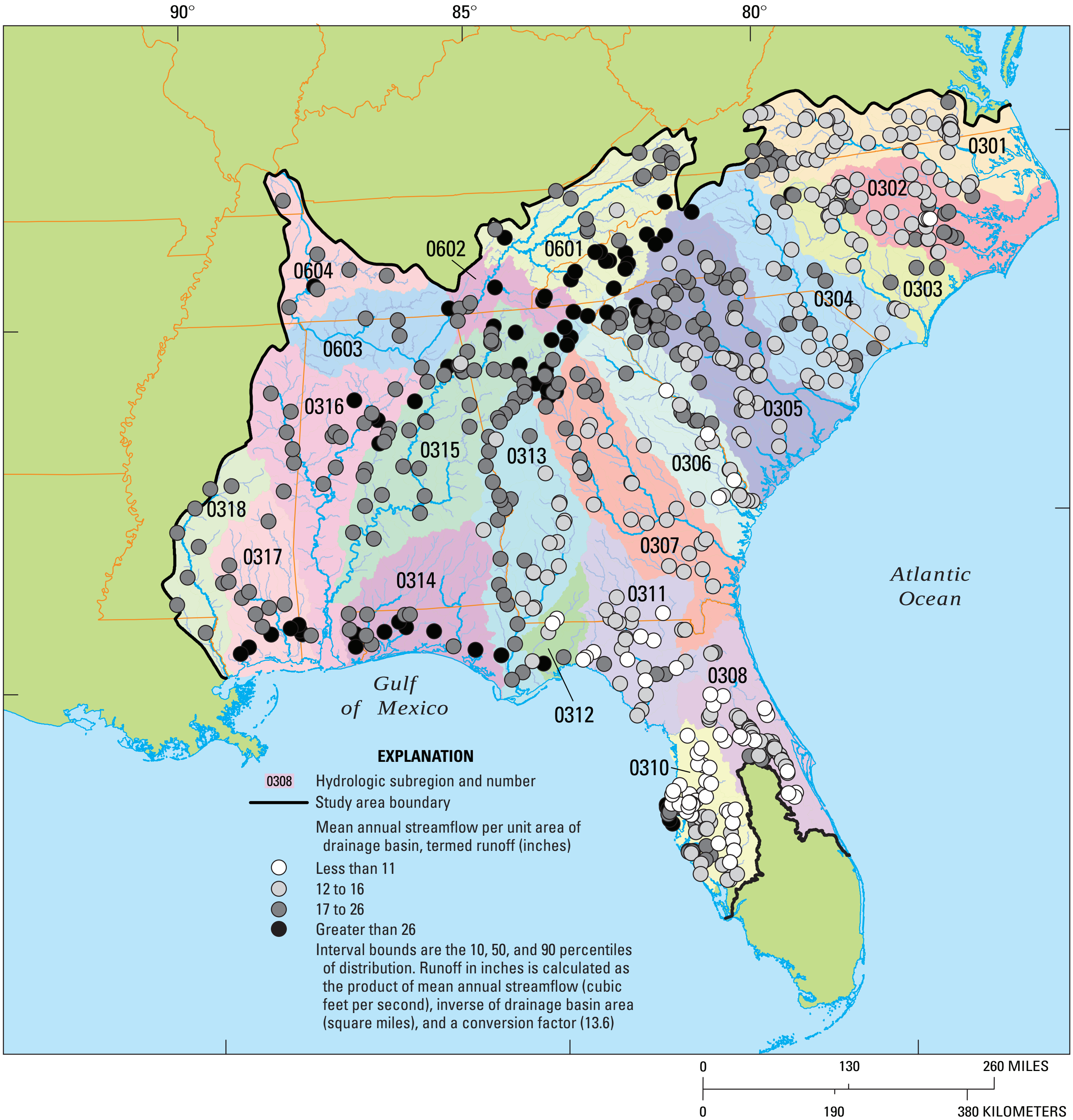

Figure 8A. Mean annual streamflow yield in the SAGT river basins estimated from stream monitoring data from 759 sites, normalized to 2002.

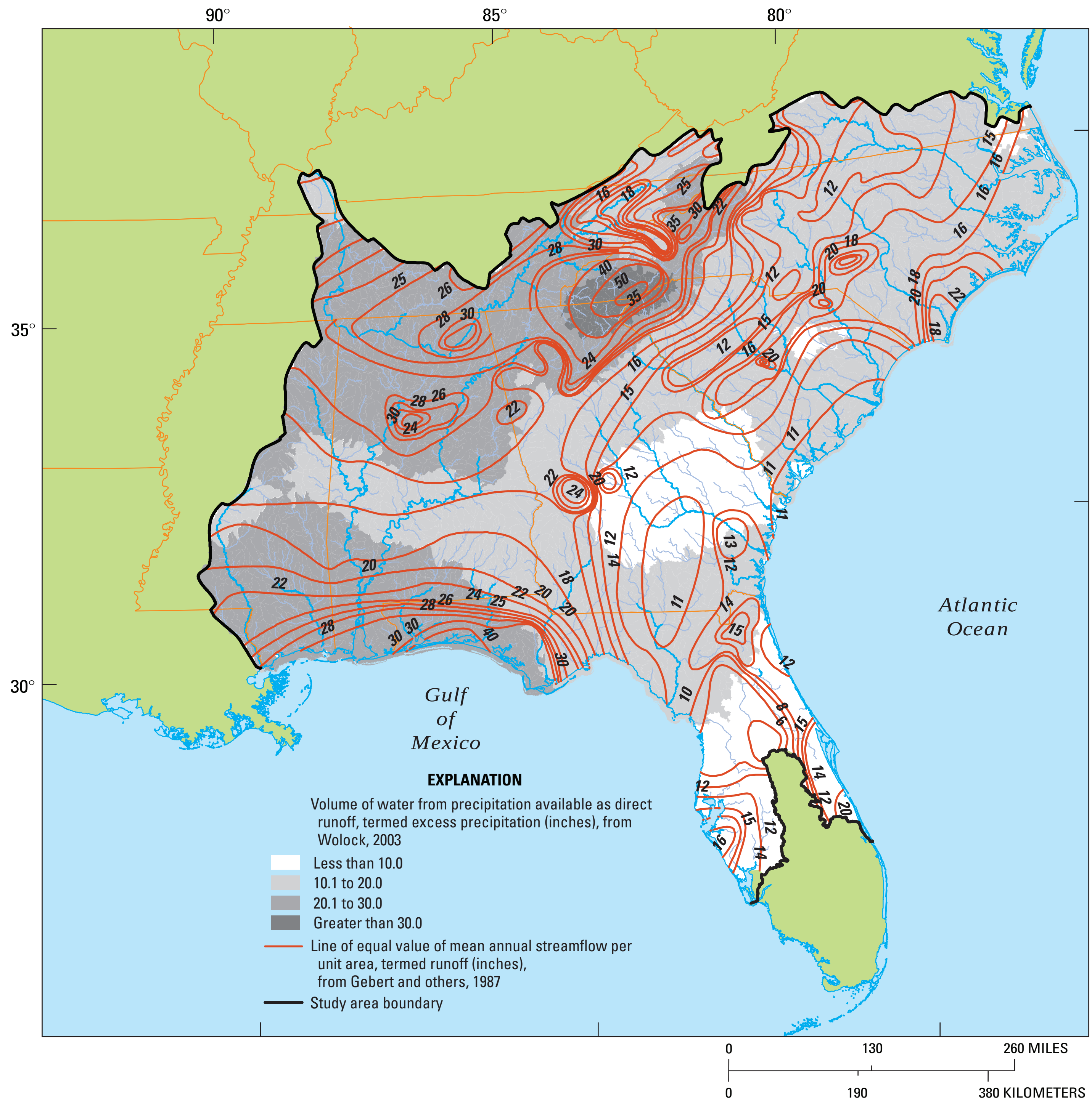

Figure 8B. Mean annual streamflow in the SAGT river basins shown as contour lines and compared with estimates of excess precipitation.

surface-water basin (Rumenik, 1988). Instream nutrient load at stream sites in these areas may not reflect conditions in the associated topographic watershed, and thus the SPARROW approach of explaining instream loads based on watershed attributes may be inappropriate. River basins identified with this concern include south Florida (where surface-water flow paths have been extensively altered) and the Oklawaha, Crystal, Lower Sante Fe, Lower Suwanee, St. Marks, and Chipola River basins in central and northern Florida (where flow exchange with the underlying regional aquifer may represent substantial nitrogen influx to and outflux from the surface-water basins; Rumenik, 1988; Miller, 1990).

Of the 637 stations with estimates of nitrogen load, only 333 can be placed on the SAGT ERF1_2 digital segmented network and used to calibrate a nitrogen SPARROW model based on SAGT ERF1_2 (shown in figure 9 as black triangles). The other 304 sites (shown in figure 9 as white triangles) were excluded for a variety of reasons: sites are located on tributaries too small to be represented in the relatively coarse 1:500,000 ERF1_2 network; sites lack independent information for calibration due to proximity (for example, within 1 kilometer) to another site with a nitrogen load estimate; or the SAGT ERF1_2 network failed to reliably model the flow path upstream from the site (judged to be the case if the ratio of the site drainage area to the upstream drainage area for the ERF1_2 reach associated with the site is outside the range of 0.75–1.33). Summary statistics of estimates of nitrogen yield and flow-weighted mean concentration for this subset of 333 sites are listed in table 3 for comparison with the more complete set of sites. The distribution of yield and concentration estimates for this subset of 333 sites is almost identical to the distribution for the complete set. Concern about flow exchange with the underlying regional aquifer representing substantial nitrogen influx to and outflux from the surface-water basins further reduces the set of stations used to calibrate the nitrogen model from 333 to 321.

Phosphorus

The mean value of the phosphorus yield estimates, normalized to 2002, for the 747 stations in the SAGT area is 0.66 kg/ha, median is 0.33 kg/ha, and 10- and 90-percentile values are 0.13 and 1.3 kg/ha, respectively (table 3). This distribution is almost identical to the national distribution of stream phosphorus yield (Mueller and Spahr, 2005); the mean, median, and 90-percentile values for the SAGT distribution are within 5 percent of the values for the national distribution. The similar values for the 90-percentile indicate that the estimates for some streams in the SAGT area are among the highest in the Nation. This contrasts with results from comparing stream nitrogen yield distribution among the two sets, in which values for the SAGT area are substantially lower than the national set.

The mean value of phosphorus flow-weighted mean concentration for the SAGT station set is 0.17 mg/L, median value is 0.08 mg/L, and 10- and 90-percentile values are 0.03 and 0.32 mg/L, respectively (table 3). These values are lower than the corresponding values for the national set. This result is expected, despite the comparable values for stream phosphorus yield, because the lower values of flow-weighted mean concentration are due to the higher mean annual streamflow yields in the SAGT area.

Estimates of mean annual phosphorus load for the 747 monitored sites are shown in figure 10. The highest 10 percent of observations of phosphorus yield (>1.3 kg/ha) occur at sites throughout the SAGT area; however, high-yield observations, as well as low-yield observations, appear to be clustered in the Peace-Tampa Bay subregion (0310). Clustering of high-yield observations in metropolitan areas is not as pronounced as it is with high-yield observations of nitrogen.

The spatial pattern in monitored phosphorus yield was evaluated with respect to the hydrologic subregion boundaries using Tukey’s multiple comparison test. The analysis divides the observed values into two statistically distinct groupings of subregions: a grouping of 14 subregions with consistently low observations of yield (mean value ≤0.66 kg/ha), and a grouping with consistently high observations of yield (mean value ≥1.8 kg/ha) that includes the Peace-Tampa Bay (0310) and Lower Tennessee (0603 and 0604) subregions. These high-yield subregions include areas of phosphate-rich soil and regolith. Complete characterization of watershed inputs of phosphorus in these subregions requires data on the phosphorus content of natural surficial materials. Regionalization, based on lithologic boundaries, of chemical analyses of soils and streambed sediment could provide estimates of soil phosphorus content for each catchment in the SAGT area (S.E. Terziotti, U.S. Geological Survey, written commun., 2007).

Of the 747 stations with estimates of phosphorus load, only 378 can be placed on the SAGT ERF1_2 digital segmented network and used to calibrate a phosphorus SPARROW model based on SAGT ERF1_2 (shown in figure 11 as black triangles). The other 369 sites (shown in figure 11 as white triangles) were excluded for the same set of reasons described for the nitrogen load station set. Summary statistics of estimates of phosphorus yield and flow-weighted mean concentration for this subset of 378 sites are listed in table 3, for comparison with the more complete set of sites. The distribution of yield and concentration estimates for this subset of 378 sites is almost identical to the distribution for the complete set. Concern about flow exchange with the underlying regional aquifer representing substantial phosphorus influx to and outflux from the surface-water basins further reduces the set of stations used to calibrate the nitrogen model from 378 to 368.

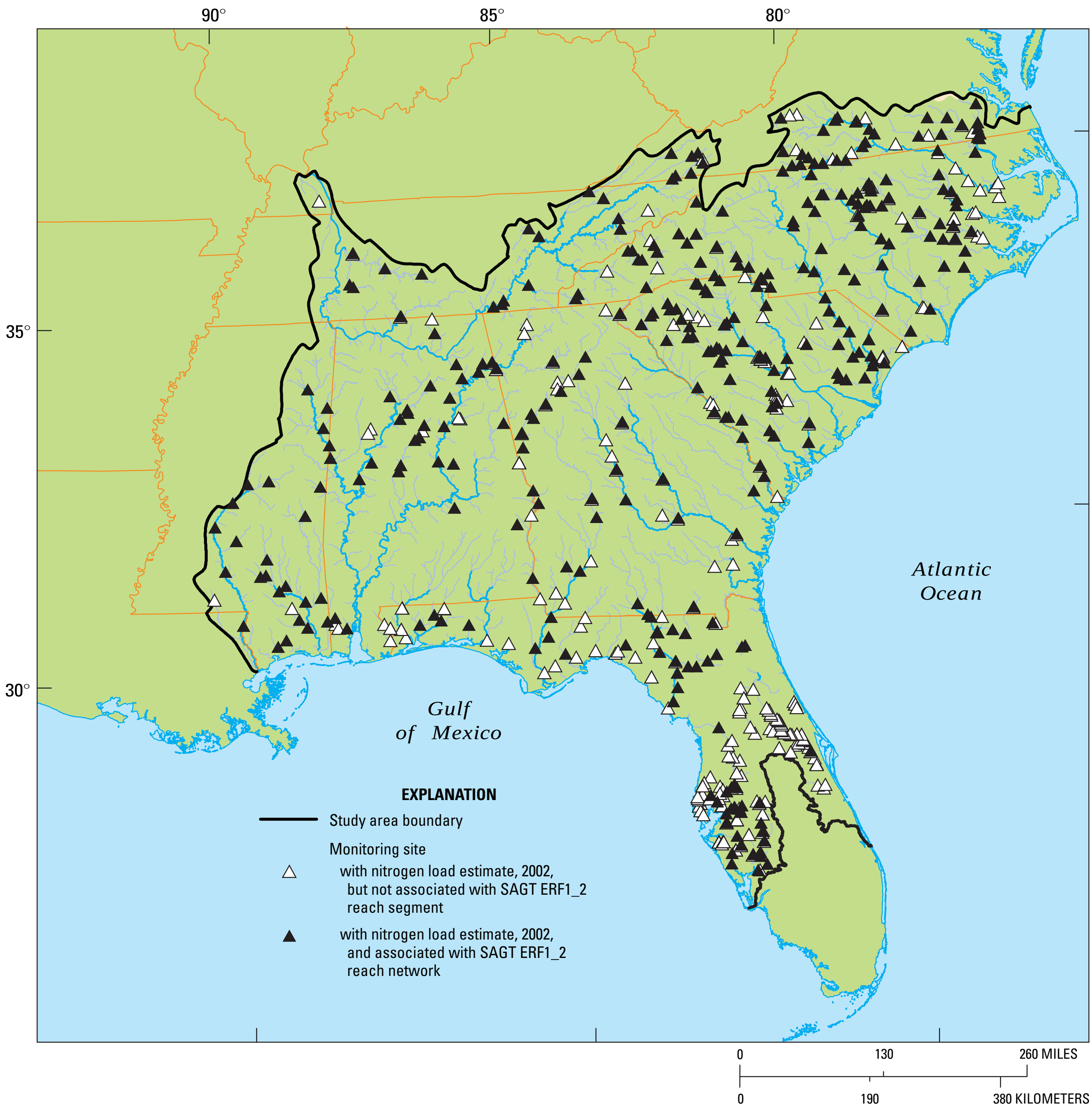

Figure 9. Water-quality monitoring sites for which nitrogen load is estimated for 2002, with the subset of sites associated with a SAGT ERF1_2 reach segment.

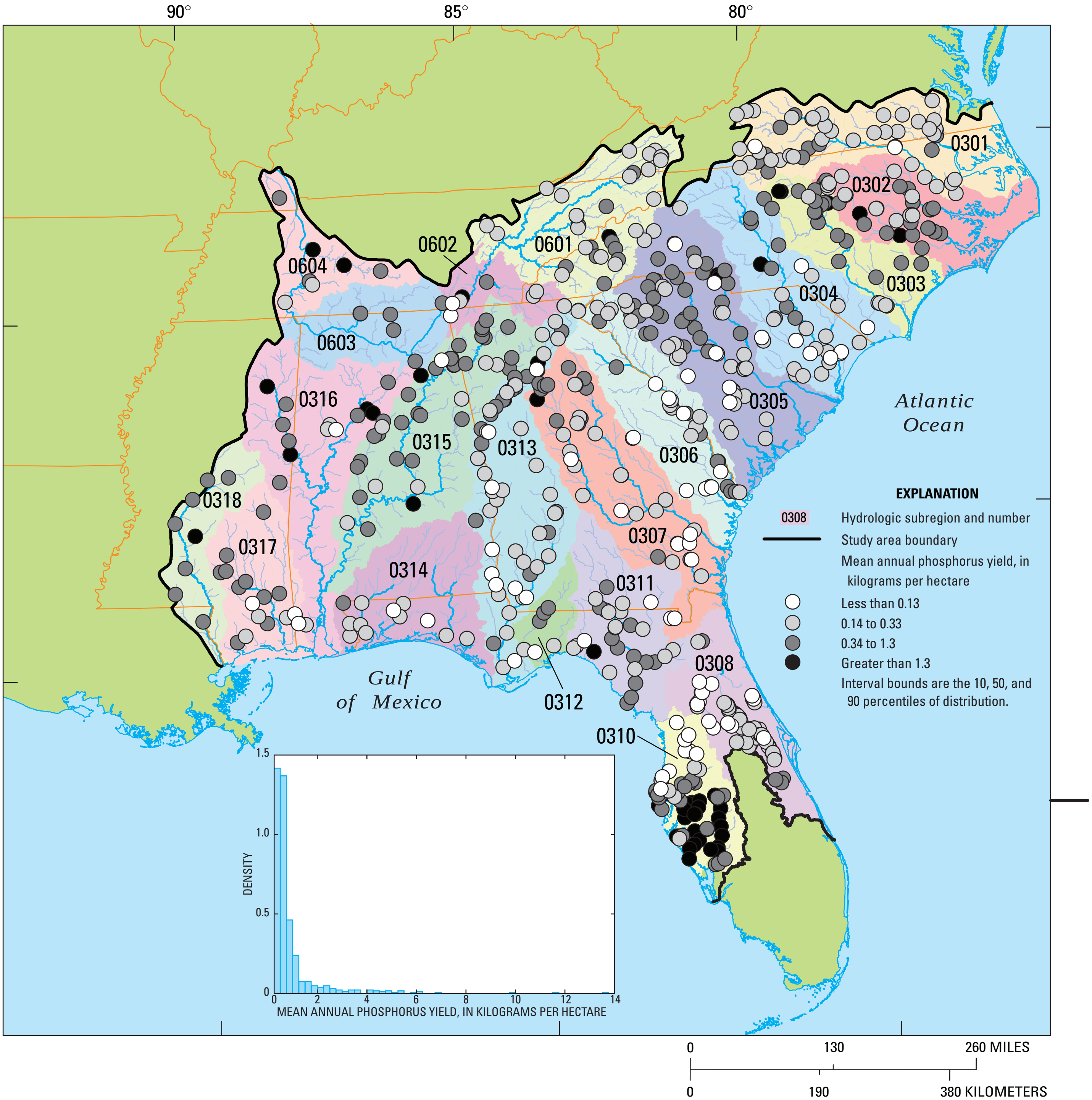

Figure 10A. Mean annual phosphorus yield estimated from stream monitoring data from 747 sites in the SAGT river basins, normalized to 2002.

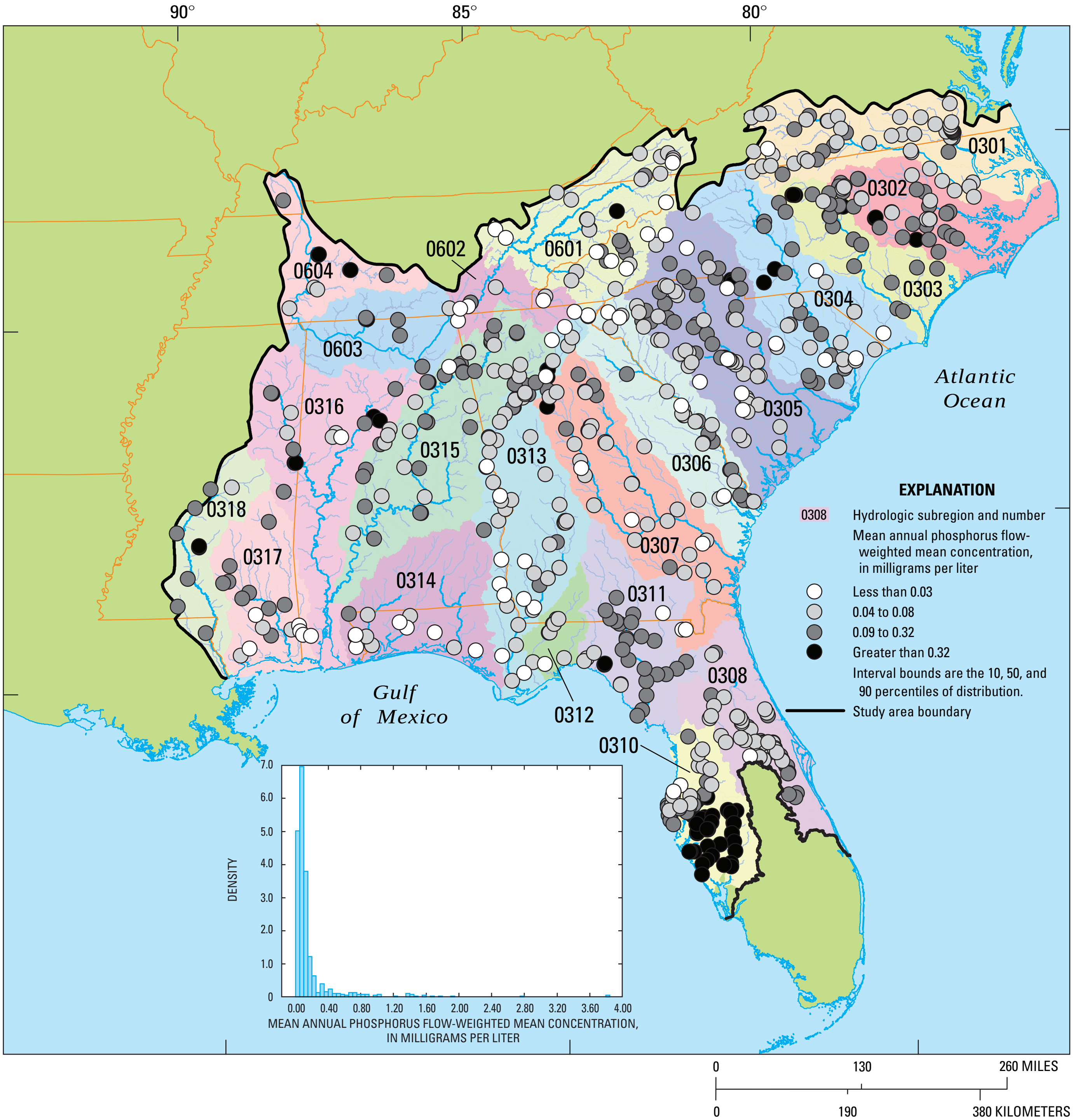

Figure 10B. Mean annual phosphorus flow-weighted mean concentration estimated from stream monitoring data from 747 sites in the SAGT river basins, normalized to 2002.

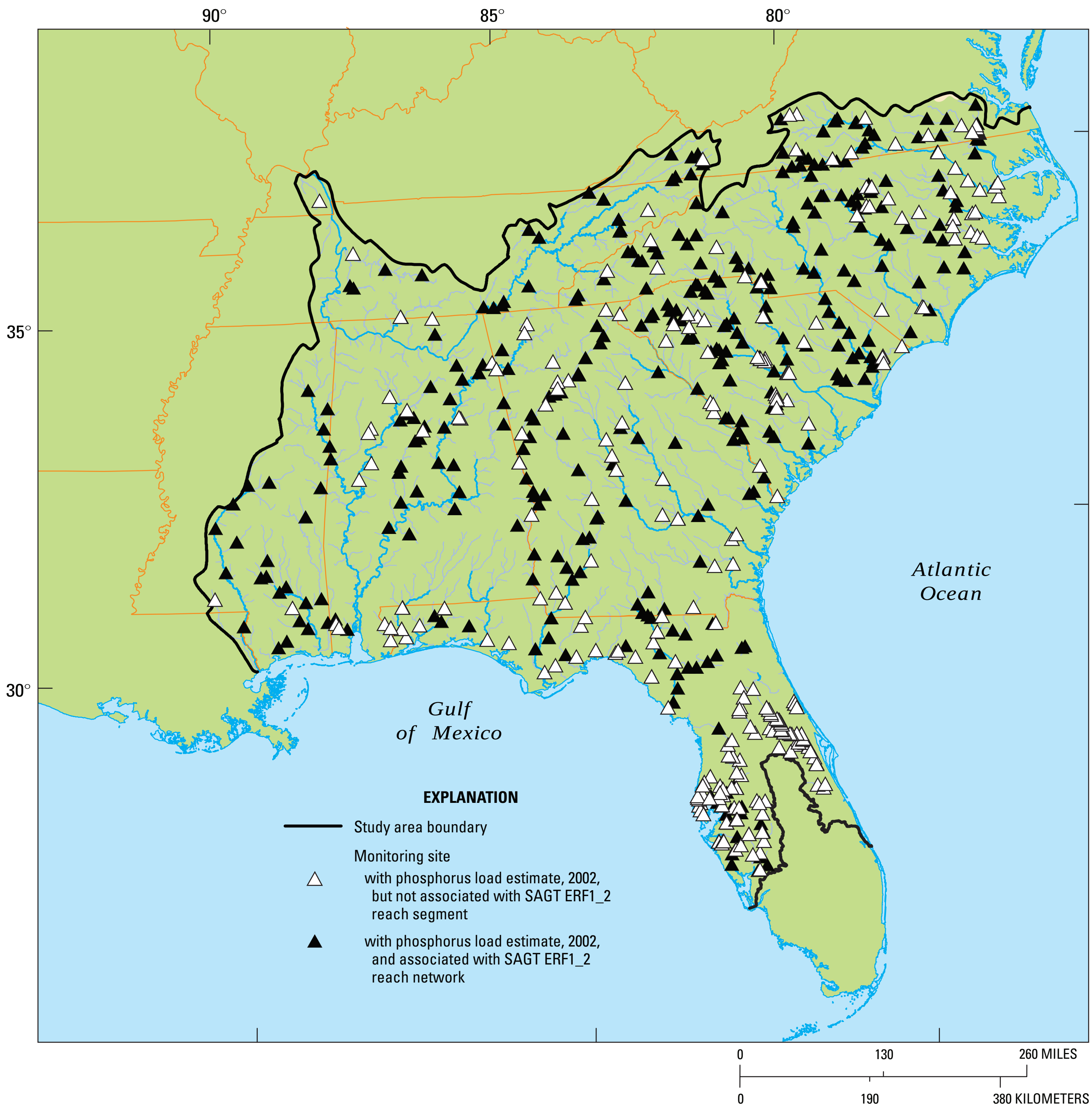

Figure 11. Water-quality monitoring sites for which phosphorus load is estimated for 2002, with the subset of sites associated with a SAGT ERF1_2 reach segment.

Summary

This report describes the digital datasets that characterize nutrient source inputs, environmental characteristics, and instream nutrient loads for the purpose of calibrating and applying a nutrient water-quality model for the southeastern United States for 2002. The water-quality model SPARROW (SPAtially-Referenced Regression On Watershed attributes) uses a regression equation to describe the relation between watershed attributes (predictors) and measured instream load (response). This application of the SPARROW model is based on a 1:500,000-scale description of the stream network and a 1:100,000-scale delineation of the catchments associated with the stream reaches.

Watershed attributes that are considered to describe nutrient input conditions are included as source variables in the regression equation; the nutrient-source variables to be tested in the SAGT SPARROW model include atmospheric deposition, fertilizer application to farmland, manure from livestock production, permitted wastewater discharge, and land cover. Watershed attributes that are considered to affect rates of nutrient transport from land to water are included in the regression equation as land-to-water transport variables; the nutrient-transport variables to be tested in the SAGT SPARROW model include characteristics of soil, landform, and climate. Channel attributes considered as nutrient transport predictors for the SAGT SPARROW model include reach time of travel and reservoir hydraulic loading.

Measurements of nutrient water quality at stream monitoring sites from a combination of monitoring programs were used to develop observations of the response variable—mean annual nitrogen or phosphorus load—in the SPARROW regression equation. Nutrient load was estimated for monitoring sites on streams and rivers (reservoir sites were excluded) for which samples were collected at least quarterly, with a minimum of 10 samples collected since 1995, for at least a 2-year period during which daily streamflow data also were collected (or could be estimated from a nearby gage). Instream loads of nitrogen and phosphorus were estimated from bias-corrected, log-linear regression models using the program Fluxmaster. A special feature available in Fluxmaster to compute detrended estimates of long-term mean load corrects for bias introduced by uneven record length among the stations and thus produces load estimates more suitable for spatial comparisons. The 4-parameter log-linear regression approach used to model daily concentration for this analysis may be inadequate for estimating annual load accurately at some sites, such as where the concentration-streamflow relation is influenced by hysteresis or antecedent conditions. The decision to employ the 4-parameter model uniformly for all stations was considered appropriate for this regional-scale assessment of water quality and in view of the available resources.

The mean value of the nitrogen yield estimates, normalized to 2002, for the 637 stations in the SAGT area is 4.7 kilograms per hectare (kg/ha), median value is 3.8 kg/ha, and 10- and 90-percentile values are 1.9 and 7.7 kg/ha, respectively. The mean value of nitrogen flow-weighted mean concentration for the SAGT station set is 1.2 mg/L, median value is 0.95 mg/L, and 10- and 90-percentile values are 0.47 and 2.0 mg/L, respectively. The highest 10 percent of observations of nitrogen yield (>7.7 kg/ha) occur at sites throughout the SAGT area; however, clusters of high-yield observations are in the Peace River-Tampa Bay basin in Florida, near metropolitan areas in central Georgia, Alabama, and North Carolina, and in the northeastern part of the Tombigbee River basin. The lowest 10 percent of observations of nitrogen yield (<1.9 kg/ha) occur throughout the eastern half of the SAGT area, and especially in the Chowan–Roanoke River basins and the Peace River–Tampa Bay basins.

The mean value of the phosphorus yield estimates, normalized to 2002, for the 747 stations in the SAGT area is 0.66 kg/ha, median is 0.33 kg/ha, and 10- and 90-percentile values are 0.13 and 1.3 kg/ha, respectively. The mean value of phosphorus flow-weighted mean concentration for the SAGT station set is 0.17 mg/L, median value is 0.08 mg/L, and 10- and 90-percentile values are 0.03 and 0.32 mg/L, respectively. The highest 10 percent of observations of phosphorus yield (>1.3 kg/ha) occur at sites throughout the SAGT area; however, high-yield observations, as well as low-yield observations, appear to be clustered in the Peace River–Tampa Bay basins. The areas with high instream yield of phosphorus correspond to areas known to contain phosphate-rich soil and regolith. Complete characterization of watershed inputs of phosphorus in the SAGT area would require data on the phosphorus content of natural surficial materials.

Sites with the highest observations of flow-weighted concentration do not, in many cases, have the highest observations of yield. The noted divergence from a directly proportional relation is due to the fact that flow-weighted mean concentration varies as a function not only of mass yield, but also of streamflow yield. The discrepancies between the spatial distribution of high and low values for mass yield compared with the spatial distribution of flow-weighted mean concentration are, therefore, a function of differences in streamflow yield.

Nutrient conditions measured in streams affected by substantial influx or outflux of water and nutrient mass across surface-water basin divides do not reflect nutrient source and transport conditions in the topographic watershed; inclusion of such streams in the SPARROW modeling approach is considered inappropriate. River basins identified with this concern include south Florida (where surface-water flow paths have been extensively altered) and the Oklawaha, Crystal, Lower Sante Fe, Lower Suwanee, St. Marks, and Chipola River basins in central and northern Florida (where flow exchange with the underlying regional aquifer may represent substantial nitrogen influx to and outflux from the surface-water basins).

References Cited

Alexander, R.B., Brakebill, J.W., Brew, R.E., and Smith, R.A., 1999, ERF1—Enhanced River Reach File 1.2, accessed September 19, 2007, at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/erf1.xml*

Alexander, R.B., Smith, R.A., and Schwarz, G.E., 2000, Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico: Nature, v. 403, p. 758–761.

Bondelid, T., Griffiths, C., and Van Houten, G., 1999, A national water pollution control assessment model—draft technical report prepared for U.S. Environmental Protection Agency, Office of Science and Technology: Durham, North Carolina, Research Triangle Park.

Brabec, Elizabeth, Schulte, Stacey, and Richards, P.L., 2002, Impervious surfaces and water quality: a review of current literature and its implications for watershed planning: Journal of Planning Literature, v. 16, no. 4, p. 499–514.

Brakebill, J.W., Preston, S.D., and Martucci, S.K., 2001, Digital data used to relate nutrient inputs to water quality in the Chesapeake Bay Watershed, version 2.0: Open-File Report 01–251, 17 p.

Bricker, S.B., Clement, C.G., Pirhalla, D.E., Orlando, S.P., and Farrow, D.R.G., 1999, National Estuarine Eutrophication Assessment—Effects of nutrient enrichment in the Nation's estuaries: Silver Spring, Maryland, National Oceanic and Atmospheric Administration, National Ocean Service, Special Projects Office and the National Centers for Coastal Ocean Science, 71 p.

Cohn, T.A., Delong, L.L., Gilroy, E.J., Hirsch, R.M., and Wells, D.K., 1989, Estimating constituent loads: Water Resources Research, v. 25, no. 5, p. 937–942.

Cohn, T.A., Gilroy, E.J., and Baier, W.G., 1992, Estimating fluvial transport of trace constituents using a regression model with data subject to censoring, *in* Proceedings of the Section on Statistics and the Environment: American Statistical Association, Boston, Mass., August 9–13, 1992, p. 142–151.

Daly, Christopher, Gibson, W.P., Taylor, G.H., Johnson, G.L., and Pasteris, P.A., 2002, A knowledge-based approach to the statistical mapping of climate: Climate Research, v. 22, p. 99–113.

Dewald, T.G., Horn, Robert, Greenspun, Robert, Taylor, Phillip, Manning, Lee, and Montalbano, Ann, 1985, STORET Reach Retrieval Documentation: Washington, D.C., U.S. Environmental Protection Agency.

Driver, N.E., and Tasker, G.D., 1990, Techniques for estimation of storm-runoff loads, volumes and selected constituent concentrations in urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2363, 44 p.

Environmental Systems Research Institute, Inc., 2008, ArcGIS Desktop Help 9.2—Spatial Analyst: Environmental Systems Research Institute, Inc., Redlands, California, accessed March 26, 2008, at *http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=An_overview_of_the_Surface_tools*

Falcone, James, 2003, National elevation data, resampled to 100m: U.S. Geological Survey, Reston, VA, accessed October 22, 2007, at *ftp://disftp.er.usgs.gov/pub/NAWQA/ecology/GIS_DATA/elevation*

Gebert, W.A., Graczyk, D.J., and Krug, W.R., 1987, Average annual runoff in the United States, 1951–80: U.S. Geological Survey Hydrologic Atlas 170.

Gilroy, E.J., Hirsch, R.M., and Cohn, T.A., 1990, Mean square error of regression-based constituent transport estimates: Water Resources Research, v. 26, no. 9, p. 2069–2077.

Graham, W.F., and Duce, R.A., 1979, Atmospheric pathways of the phosphorus cycle: Geochimica et Cosmochimica Acta, v. 43, no. 8, p. 1195–1208.

Hellweger, F.L., and Maidment, D.R., 1997, AGREE-DEM surface rconditioning system: Austin, Texas, University of Texas, accessed October 26, 2007, at *http://www.crwr.utexas.edu/gis/gishyd98/quality/agree/agree.htm*

Homer, Collin, Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J., and Wickham, J., 2007, Completion of the 2001 National Land Cover Database for the Conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, no. 4, p. 337–341.

Homer, Collin, Huang, C., Yang, L., Wylie, B., and Coan, M., 2004, Development of a 2001 National Landcover Database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, p. 829–840.

Hoos, A.B., 2005, Evaluation of the SPARROW model for estimating transport of nitrogen and phosphorus in streams of the Interior Low Plateau Ecoregion, Tennessee, Kentucky, and Alabama, during 1992–2002 [abs.], *in* Tennessee Water Resources Symposium, 15th, Burns, Tenn., 2005, Proceedings: Tennessee Section of the American Water Resources Association, p. 2A-47.

Kuntz, K.W., 1980, Atmospheric bulk precipitation in the Great Lakes basin: Scientific Series No. 115, Inland Waters Directorate, Water Quality Branch, Burlington, Ontario, Canada.

McMahon, Gerard, 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.

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., accessed August 6, 2007, at *http://pubs.usgs.gov/of/2007/1040/*

Miller, J.A., 1990, Ground water atlas of the United States; Alabama, Florida, Georgia, and South Carolina: U.S. Geological Survey Hydrologic Atlas 730-G.

Moore, R.B., Johnston, C.M., Robinson, K.W., and Deacon, J.R., 2004, Estimation of total nitrogen and phosphorus in New England streams using spatially referenced regression models: U.S. Geological Survey Scientific Investigations Report 2004–5012, 42 p.

Mueller, D.K., and Spahr, N.E., 2005, Water-quality, streamflow, and ancillary data for nutrients in streams and rivers across the Nation, 1992–2001: U.S. Geological Survey Data Series 152, accessed August 2007 at *http://pubs.usgs.gov/ds/2005/152/index.htm*

Murphy, J.J., 1974, Sources of phosphorus inputs from the atmosphere and their significance to oligotrophic lakes: University of Illinois, Research Report No. 92 UJLC-WRC-74-0092.

National Atmospheric Deposition Program, 2006, The National Atmospheric Deposition Program Data Access, accessed June 2006 at *http://nadp.sws.uiuc.edu/sites/ntnmap.asp*

National Resources Council, 2001, Assessing the TMDL approach to water quality management: Washington, D.C., National Academy Press, 405 p.

Nolan, J.V., Brakebill, J.W., Alexander, R.B., and Schwarz, G.E., 2002, ERF1_2—Enhanced River Reach File 2.0, accessed August 2007 at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/erf1_2.xml*

Oregon State University, PRISM Group, 2007, 800m normals (1971–2000): Corvallis, Oregon, Oregon State University, accessed August 2007 at *http://prism.oregonstate.edu/*

Potter, S.R., Andrews, Susan, Atwood, J.D., Kellogg, R.L., Lemunyon, Jerry, Norfleet, Lee, and Oman, Dean, 2006, Model simulation of soil loss, nutrient loss and change in soil organic carbon associated with crop production: U.S. Department of Agriculture, Natural Resources Conservation Service, Conservation Effects Assessment Project, accessed September 19, 2007, at *http://www.nrcs.usda.gov/technical/nri/ceap/croplandreport/*

Preston, S.D., and Brakebill, J.W., Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay Watershed: U.S. Geological Survey Water-Resources Investigations Report 99–4054, 12 p.

Redfield, G., Efron, S., Goforth, G., James, R.T., Van Horn, S., Adorisio, C., Jones, B., Piccone, T., Petro, K., and Zhang, J., 2007, An integrative perspective on regional water quality and phosphorus: 2007 South Florida Environmental Report, Chapter 1B, South Florida Water Management District, West Palm Beach, Florida, variously paginated.

Ruddy, B.C., Lorenz, D.L., and Mueller, D.K., 2006, County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982–2001: U.S. Geological Survey Scientific Investigations Report 2006–5012, 17 p.

Rumenik, R.P., 1988, Runoff to streams in Florida: Florida Geological Survey Map series 122, Tallahassee, Florida, Florida Geological Survey, 1 sheet.

Saunders, W., 2000, Preparation of DEMs for use in environmental modeling analysis, *in* Maidment, D.R., and Djokic, D., eds., 2000, Hydrologic and Hydraulic Modeling Support: Redlands, CA, ESRI Press, p. 29–51.

Schwarz, G.E., Hoos, A.B., Alexander, R.B., and Smith, R.A., 2006, The SPARROW surface water-quality model—theory, application, and user documentation: U.S. Geological Survey Techniques and Methods, book 6, section B, chap. 3, accessed August 6, 2007, at *http://pubs.usgs.gov/tm/2006/tm6b3/*

Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1987, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p.

Smith, R.A., and Alexander, R.B., 2000, Sources of nutrients in the Nation's watersheds, *in* Managing nutrients and pathogens from animal aagriculture, proceedings from the Natural Resource, Agriculture, and Engineering Service Conference for Nutrient Management Consultants, Extension Educators, and Producer Advisors, March 28–30, 2000, Camp Hill, Pennsylvania.

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.

Statistical Analysis Systems Institute, 2000, SAS language reference—dictionary, version 8: Cary, NC, SAS Institute, Inc., 500 p.

U.S. Department of Agriculture, 1994, State soil geographic (STATSGO) database data use information: U.S. Department of Agriculture-Natural Resources Conservation Service Miscellaneous Publication 1492, 113 p.

U.S. Environmental Protection Agency, 1996, USEPA Reach File Version 1.0 (RF1) for the Conterminous United States (CONUS): Washington, D.C., U.S. Environmental Protection Agency, accessed August 2007 at *http://www.epa.gov/waters/doc/rf1_meta.html*

U.S. Environmental Protection Agency, 2002, National Water Quality Inventory, 2000 Report: U.S. Environmental Protection Agency Office of Water, EPA-841-R-02-001, variously paginated.

U.S. Geological Survey, 2001, National Land Cover Database 2001 (NLCD 2001): U.S. Geological Survey, accessed August 2007 at *http://www.mrlc.gov/mrlc2k_nlcd_map.asp/*

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, digital data release, accessed October 2007 at *http://water.usgs.gov/GIS/metadata/usgswrd/muid.xml*

Wolock, D.M., 2003, Hydrologic landscape regions of the United States: U.S. Geological Survey Open-File Report 03–145, digital data release, accessed October 2007 at *http://water.usgs.gov/GIS/metadata/usgswrd/XML/hlrus.xml*

Wolock, D.M., and McCabe, G.J., 1999, Explaining variability in mean annual runoff in the conterminous United States: Climate Research, v. 11, p. 140–159.

Wolock, D.M., Winter, T.C., and McMahon, Gerard, 2004, Delineation and evaluation of hydrologic-landscape regions in the United States using geographic information system tools and multivariate statistical analyses: Environmental Management, v. 34, p. S71–S88.

Yong, S.T.Y., and Chen, Wenli, 2002, Modeling the relationship between land use and surface water quality: Journal of Environmental Management, v. 66, p. 377–393.

Data Files

Geospatial datasets are available to download as Arc Info shapefiles (zipped using Winzip). Data files of attributes are available to download in Excel (version 2003) format and in tab-delimited text format. Each Excel workbook contains a data sheet and a sheet (named README) with variable name definitions and notes.

Description	Downloadable datafile and metadata	Section of report describing data
SAGT ERF1_2 digital segmented network (geospatial dataset)	*erf1_spar.zip* (5.1 MB), *erf1_spar.html*	Hydrologic network of reaches and associated catchments
SAGT ERF1_2 segmented catchments (geospatial dataset)	*shed_cov.zip* (13 MB), *shed_cov.html*	Hydrologic network of reaches and associated catchments
Catchment-level estimates of watershed and reach attributes evaluated for incremental catchments and reaches	Excel version: *SAGT_ERF1_input.xls.zip* (2.1 MB) (metadata included in the README sheet) Textfile version: *SAGT_ERF1_input.txt* (2.1 MB), *README_SAGT_ERF1_input.txt*	Watershed attributes, reach attributes
Estimates of watershed attributes accumulated for the total upstream watershed contributing to the reach segment	Excel version: *SAGT_accumulatedfortotalwatershed.xls.zip* (2.1 MB) (metadata included in the README sheet) Textfile version: *SAGT_accumulatedfortotalwatershed.txt* (2.3 MB), *README_SAGT_ERF1_input.txt*	Accumulation of catchment-level estimates of watershed attributes to estimates for the total upstream watershed
Monitoring sites, station characteristics, and nutrient load estimates	Excel version: *SAGT_monitoredload.xls* (700 KB) (metadata included in the README sheet) Textfile version: *SAGT_monitoredload.txt* (28 KB), *README_SAGT_monitoredload.txt*	Mean annual nitrogen and phosphorus load at stream monitoring sites

Routines used to modify nutrient-constituent concentration data for load estimation are available to download in SAS (version 9) format and in text format.

Description	Downloadable program file	Section of report describing data
Reformats the water-quality data file from modernized STORET (tilde-delimited) to a SAS datafile in the format used by Fluxmaster (more details provided in paragraphs following this table)	*Reformat_ModSTORET_WQdata.sas* (42 KB) Text version: *Reformat_ModSTORET_WQdata.txt* (42 KB)	Review and revision of nutrient concentration results
Resolves the differences among data sources in the format or convention for recording results (more details provided in paragraphs following this table)	*Convert_remarkcoding_and_otherproblematic.sas* (16 KB) Text version: *Convert_remarkcoding_and_otherproblematic.txt* (16 KB)	Review and revision of nutrient concentration results
Assigns or calculates a value for a total nitrogen (TN) parameter code, P60000, and for a total phosphorus (TP) parameter code, P66500 (more details provided in paragraphs following this table)	*Combine_nutrient_constituents.sas* (8 KB) Text version : *Combine_nutrient_constituents.txt* (8 KB)	Review and revision of nutrient concentration results

The file "Reformat_ModSTORET_WQdata.sas" reformats the water-quality data file from modernized STORET (tilde-delimited) to a tab-delimited file, interpreting information from several variables (characteristic name, sample fraction, and media) into an assignment of parameter code (pcode) following the convention used in Legacy STORET and in NWIS, and populating the associated remark code variable for results below detection. The tab-delimited file is then converted to a SAS datafile in the format (one line per sample) used by the load estimation program Fluxmaster (Schwarz and others, 2006). Multiple stations may be included in the analysis.

The program "Convert_remarkcoding_and_otherproblematic.sas" resolves the differences among data sources in the format or convention for recording results, by revising the data records retrieved from Legacy and Modernized STORET to match the NWIS format or convention. (The load estimation program, Fluxmaster, is programmed to work with data coded using the NWIS convention.) This routine also corrects cases of obviously erroneous concentration results, such as extremely large values.

Summary of changes for legacy STORET data records:

1. Replace the nonsense numbers (positive and negative) with missing values.
2. Replace the zero and negative values that indicate below detection with appropriate detection limit values, and set remark code to ‘<’.
3. Replace all remark codes that mean ‘<’ (K and U) with ‘<’.
4. Replace remark codes that mean ‘>’ (L) to ‘>’.

Summary of changes for modernized STORET data records:

1. For less than result for which detection-limit was not available in the retrieved data in order to populate the value field (P field) during reformatting: set value field equal to a reasonable estimate of detection limit (75 percentile of all detection limits reported in the SAGT project dataset from STORET, which can be obtained from distribution of values in the detection-limit field, or D field).
2. Replace all remark codes that mean ‘<’ (U) with ‘<’.

The program “Combine_nutrient_constituents.sas” assigns or calculates a value for a total nitrogen (TN) parameter code, P60000, and for a total phosphorus (TP) parameter code, P66500. The code P60000 is assigned a value equal to P00600; or if P00600 is missing, it is calculated by combining total Kjeldahl nitrogen results and nitrate results (if available), or by combining ammonia nitrogen results, organic nitrogen results, and nitrate results. The code P66500 is assigned a value equal to P00665, or if P00665 is missing, it is calculated by combining dissolved and suspended phosphorus (if available, although this is rarely the case). The rules for combining results include how to handle the case of one or more of the constituents having censored values, and how to populate the remark code for the calculated parameter.

Prepared by:
USGS Publishing Network
Raleigh Publishing Service Center
3916 Sunset Ridge Road
Raleigh, NC 27607

For additional information regarding this publication, contact:
Anne B. Hoos, Hydrologist
USGS Tennessee Water Science Center
640 Grassmere Park, Suite 100
Nashville, TN 37211
email: abhoos@usgs.gov

Or visit the USGS National Water-Quality Assessment Program website at:
http://water.usgs.gov/nawqa/

USGS Hoos and others—**Data to Support Statistical Modeling of Instream Nutrient Load Based on Watershed Attributes, Southeastern United States**—Open-File Report 2008–1163