**Stream Classification Framework for the SARP Region**

**A Summary Report to Complete GRANT AGREEMENT # 070111-01**

**Between the Southeast Aquatic Resources Partnership and The Nature Conservancy**

**Arlene Olivero Sheldon and Mark Anderson 3/29/2013**

**Objective**:

The objective of this project was to develop some basic stream classification attributes for the entire Southeast Aquatic Resources Partnership (SARP) region and to provide more detailed attributes in the eastern section of the SARP geography (9 states: AL, FL, GA, KY, NC, SC, TN, WV, VA) where additional data and modeling capacity was available. The final product is a mapped dataset of information linked to the NHDPlus medium resolution hydrography that can be used to classify stream reaches. The results of this work contribute to SARP’s overall objective to develop a river classification framework database consisting of a hierarchical set of hydrologic, morphologic, and biotic parameters for NHDPlus river segments which can be used to identify ecologically similar types of rivers within the region according to the needs of the user.

**Approach:**

The stream classification variables, thresholds, and spatial analysis to develop the regional river classification attributes underwent review by a small committee of regional and topical issue experts via a series of seven webinars during 2011-2013. The objective of this review was to ensure that the approach and methods were scientifically credible and accurate. Detailed webinar notes and the full presentations from the webinars are available at the project Wiki website at http://sifn.bse.vt.edu/sifnwiki/index.php/SIFN\_Classification\_Expert\_Review. The final results of our work are available in this summary report and accompanying dataset for further use and review by scientists and natural resource managers across the region.

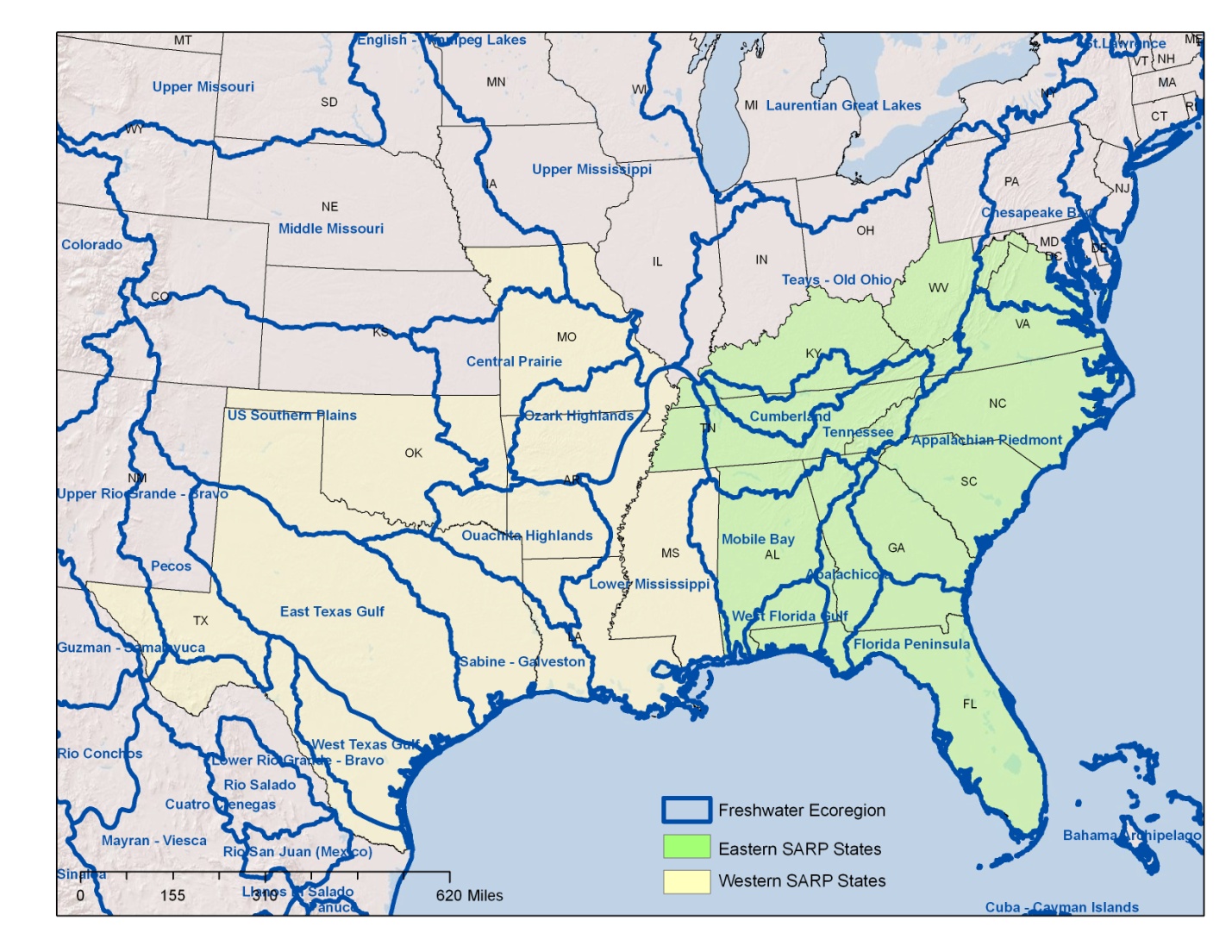
**SALCC River Classification Committee Members:**

During the series of webinars, the follow experts participated in the review:Mary Davis (SARP, Facilitator), John Faustini (Chair, USFWS), Ryan Smith (TNC-TX), Shannon Brewer (USGS), Chris Goudreau (NCWRC), Chris Konrad (USGS), Jim McKenna (USGS), Mark Anderson (TNC-Eastern Division), Don Orth (Va Tech), Paul Blanchard (MO-DNR), Rua Mordecai (SALCC), Ryan McManamay (Oak Ridge NRL), Arlene Olivero Sheldon (TNC-Eastern Division), Emily Watson (SARP/USFWS), Analie Barnett (TNC-Eastern Division).

**Geography:**

The SARP footprint comprised 14 states in the Southeastern US (NHDPlus catchments (n = 899,135). This area was divided into an eastern and western portion for the purposes of this project (Map1). Additional stream classification attributes related to geology, soils, temperature class, and hydrologic variability class were developed for reaches in the eastern SARP footprint as data was available.

Map 1. SARP Region



**Base Hydrology Map:**

The 2006 National Hydrography Dataset (NHDPlus) Version 1, a widely available 1:100,000 GIS dataset, was used as the base hydrology dataset for this project (USGS 2006). The NHDPluslinework is geometrically corrected, augmented with names, and provides line (stream), polygon (lake), and local catchment watersheds for each flowline. The NHDPlusalso comes with a set of important value-added attributes for modeling and navigating upstream/downstream. Many of these pre-calculated attributes were useful in our classification effort. Moreover, USGS has a maintenance infrastructure to improve the NHDPlus dataset and integrate user updates over time.

**Results**

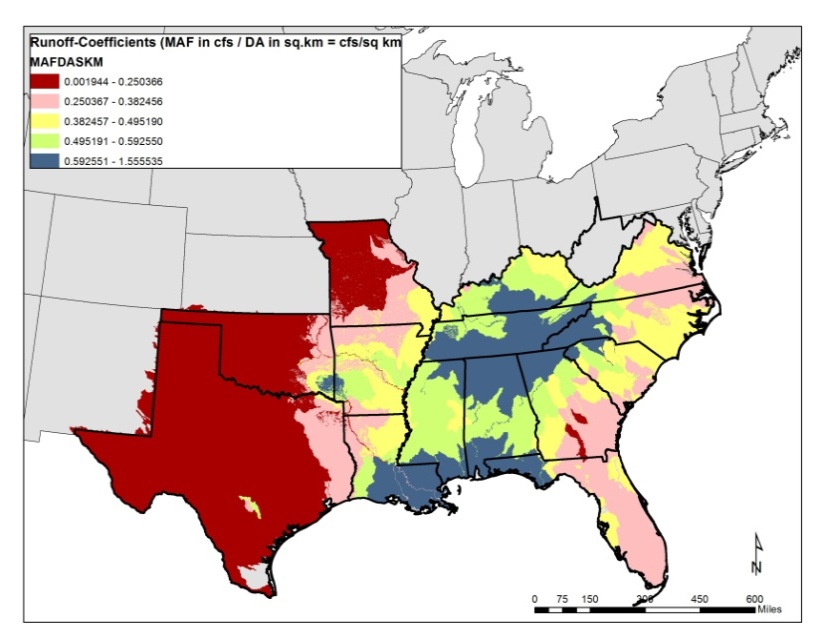
**Part 1. Basic Stream Classification Attributes for Entire SARP Region**

**Stream Size Class**

Stream size has been given the highest classification importance in many reach scale stream classification systems because of its strong effect on determining aquatic biological assemblages at the reach scale (Vannote et al. 1980, Higgins et al. 2005). The well-known "river continuum concept" provides a description of how the physical size of a stream relates to major ecosystem changes from small headwater streams to large river mouths (Vannote et al. 1980). In narrow headwater streams, coarse particulate organic matter (e.g. leaves, twigs etc.) from the riparian zone shades the river and provides the energy source for a consumer community dominated by shredding insects. As a river broadens at mid-order sites, energy inputs change as sunlight reaches the stream to support significant periphyton production and grazing insects. As the river further increases in size, fine particulate organic matter inputs increase and macrophytes become more abundant as reduced channel gradient and finer sediments form suitable conditions for their establishment. In even larger rivers, the main channel becomes unsuitable for macrohphytes or periphyton due to turbidity, fast current, depth and/or lack of stable substrates. Autochthonous production by phytoplankton increases until limited by increasing instream turbidity. Allochthonous organic matter inputs occurring outside the stream channel will then again become the primary energy source as processes such as floodplain scouring increase. These changes in physical habitat and energy source as streams grow in size are correlated with predictable patterns of changes in the aquatic biological communities (Vannote 1980).

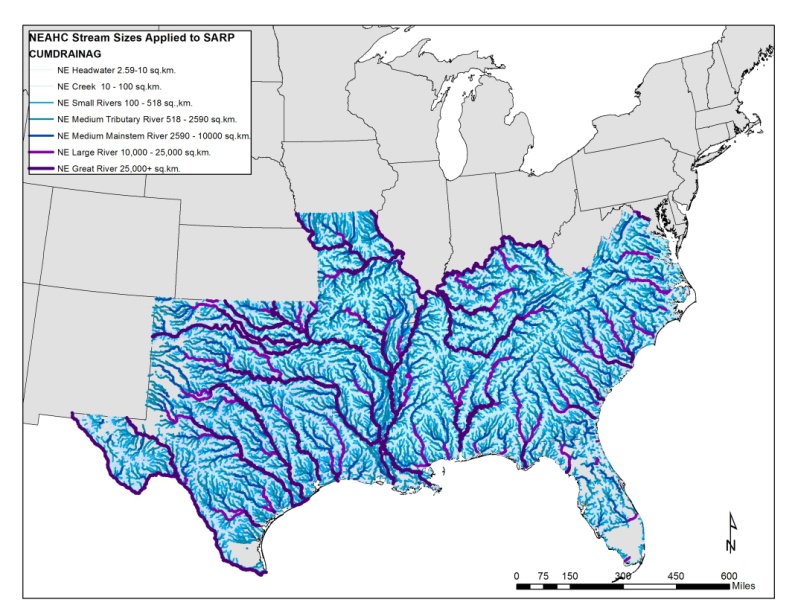
Catchment drainage area, mean annual flow, monthly flow volumes, stream order, number of first order streams above a given segment, and bankfull width can all be used as measures of stream size. We chose **upstream catchment drainage area** as the primary measure of stream size for the SARP region and **mean annual flow** as a secondary measure of stream size. Both variables were available from USGS as part of the NHDPlus value added attributes (USGS 2006). The steering committee discussed and reviewed a variety of measures of size and the specific stream size break thresholds used in the Greater Texas classification (Smith, 2011) , Northeast Aquatic Habitat Classification system (NAHC) (Olivero and Anderson, 2008), and National Fish Habitat Classification (Beard and Whelen, 2006). Committee members felt there was strong support for maintaining both a measure of stream size based on upstream drainage area (basin or watershed size) and also for a measure based on mean annual flow. Mean annual flow provided a consistent comparison of stream size across the regional differences in basin yield. An attribute of runoff coefficient was created to represent the difference in basin yield by taking the NHDPlus mean annual flow (cubic feet/sec) divided by drainage area in sq.km to get unit runoff (cfs/sq km). A map of the runoff coefficient, displayed in 5 quantiles, highlighted distinct variation in basin yield between the western, central, and eastern portions of the SARP geopgraphy for streams with the same upstream drainage area (Map 2). Mean annual flow, however, is a hydrologic metric that some members of the team found confusing if used as a measure of stream size which is more usually considered a geomorphologic attribute. Many of the westernmost streams in the SARP region have highly variable flows and the mean annual flow metric did not represent this variation. The upstream drainage basin size was considered a better metric for channel geomorphic size, particularly in this western region, as it is this upstream drainage area size that drives the important large seasonal floods and channel shaping flows which are geomorphically very important to shaping the channel size. More study is needed to understand the relationship of these two measures of size to instream biota, so both attributes are included in the SARP river classification framework.

Map 2. Runoff Coefficients (cfs/sq. km) for the SARP region were calculated using NHDPlus Mean Annual Flow (MAF) in cfs / NHDPlus Drainage Area (DA) in sq. km.



Size Classes Based on Drainage Area: The drainage area thresholds used to develop seven stream size classes in the northeast (NAHC, Olivero and Anderson, 2008) were adopted for the SARP region (Map 3). These size classes were deemed appropriate and ecologically useful by the review committee. These class breaks were originally determined by study of the size breaks and biological descriptions used in northeast states, testing of fisheries datasets in PA, testing of the distributions of rare freshwater species across size classes, and an attempt to match breaks being used by the National Fish Habitat Framework when possible (Olivero and Anderson, 2008).

Map 3. Size Classes Based on NHDPlus Cumulative Drainage Area



Size Classes Based on Mean Annual Flow: As a secondary measurement of stream size, the team agreed to also maintain the following seven size classes based on mean annual flow breaks (Map 4). These breaks were developed by sampling the mean annual flow values of northeastern reaches within each of the seven Northeast size classes, and then using the class means plus the half-way point between class means to define breaks in mean annual flow. These classes then roughly correspond to the mean annual flow found for headwaters, creeks, small rivers, medium tributary rivers, medium mainstem rivers, large rivers, and great rivers as previously defined by upstream drainage area.

Map 4. Size Classes Based on NHDPlus Mean Annual Flow (Unit Runoff Method)

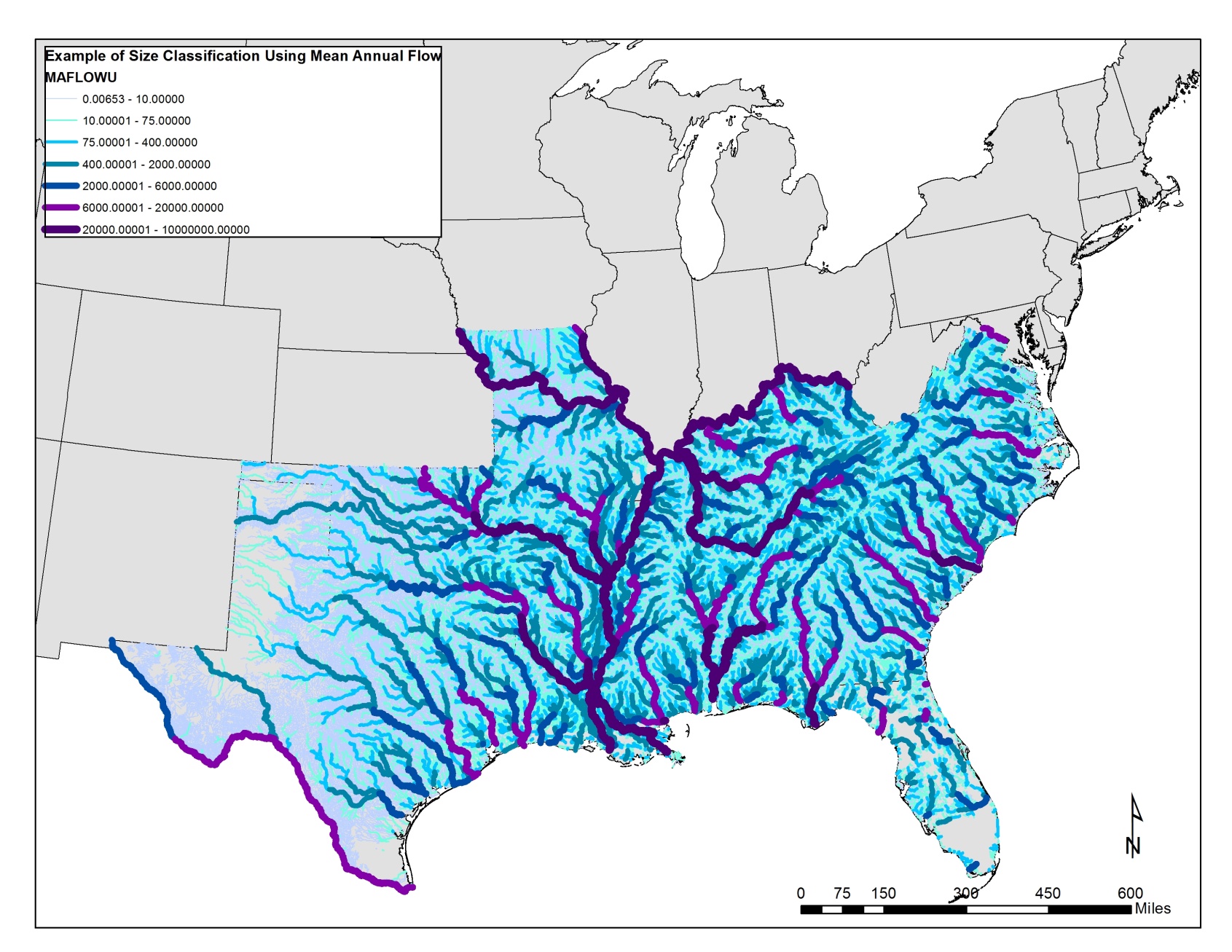


Table 1. Size Class Descriptions and Definitions



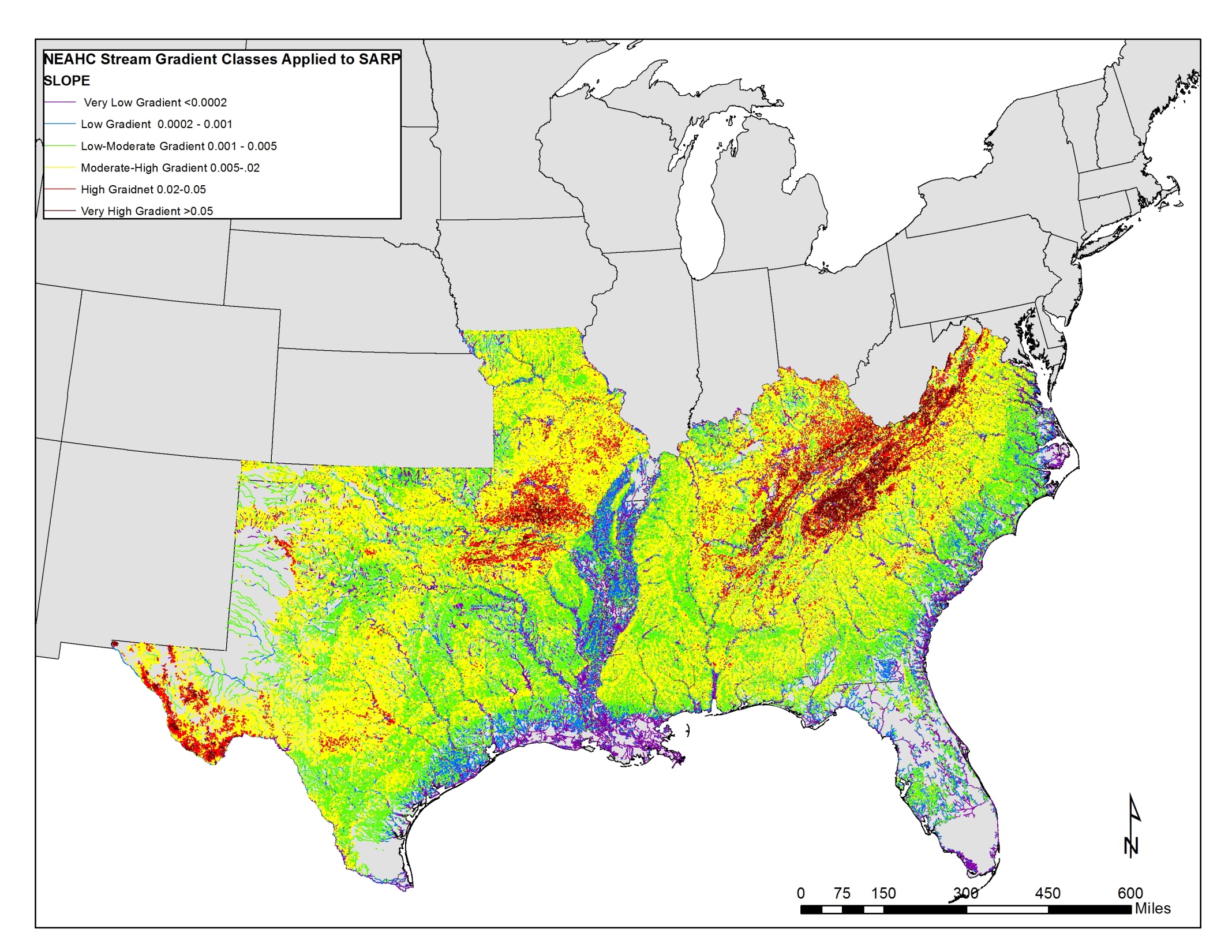
**Stream Gradient Class**

Stream gradient also highly influences aquatic communities at the reach scale due to its influence on stream bed morphology, flow velocity, sediment transport/deposition, substrate and grain size (Rosgen 1994). For example, high gradient streams are dominated by step-pools to plane-bed systems. They have substrates of cobble and boulders, colluvial sediment transport, and are usually highly entrenched, valley confined, and have low sinuosity. Moderate gradient streams are generally plane bed systems with some riffle-pool development. They have substrates of gravel, cobble, and boulders, transport sediment regimes, and are moderately entrenched with narrow valleys with low sinuosity. Low gradient systems are dominated by riffle-pool systems. They have substrates of sand, gravel, and cobble, alluvial storage and depositional sediment regimes, high sinuosity, and are only slightly entrenched with adjacent floodplain ecosystems in their broader valleys. Very low gradient streams are dominated by ripple-dune streams with very high sinuosity. These rivers have sand, gravel and finer sediment substrates, alluvial storage and depositional sediment regime, and slight entrenchment with critical adjacent floodplain systems (Rosgen 1996, Allan 1995).

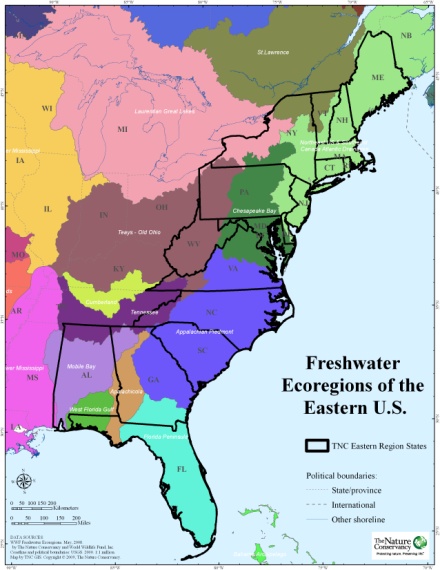
Gradient was measured as the slope of the flow line, calculated as rise height over run length (NHD-Plus, USGS 2006) and notated as a percentage of run length (Table 2, Map 5). The SARP committee discussion for river gradient focused on the number of class categories more than on the break thresholds. Since categories can be collapsed using a decision process relevant to a particular need, it was decided to keep all six gradient categories used in the NAHC northeast classification (Olivero and Anderson, 2008). The original northeast gradient classes were developed by studying breaks used in the existing state classifications, examining the relationship of gradient classes to known places in the region, studying rare species distributions across gradient classes, and review of Rosgen’s gradient classes.

Table 2. Gradient Class Descriptions and Definitions



Map 5. Stream Gradient Classes

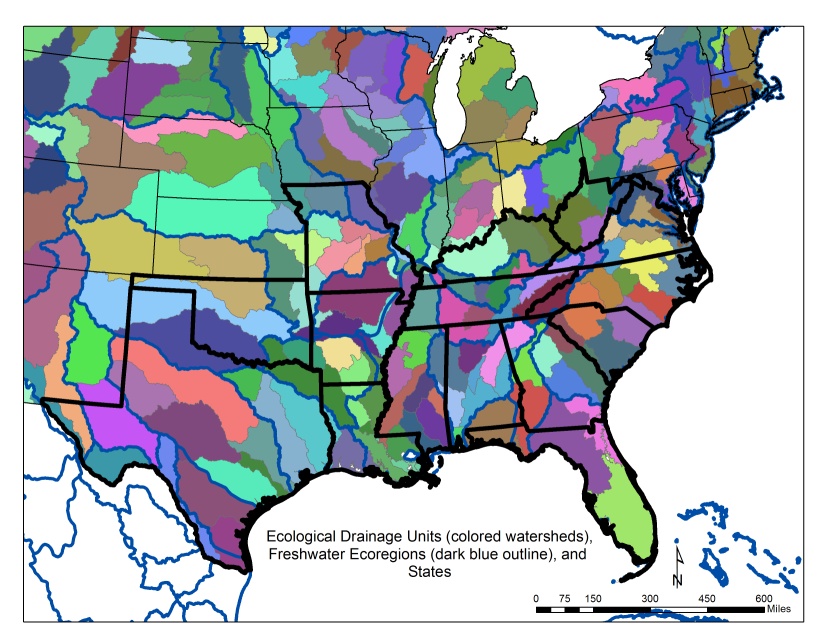
**Freshwater Ecoregion:**  Map 6. Freshwater Ecoregions

The team recommended the use of two geographic stratification units to nest any reach scale stream classification within, Freshwater Ecoregions and Ecological Drainage Units. Freshwater ecoregions (Map 6) were defined and mapped by the World Wildlife Fund (Abell et al 2008). Freshwater ecoregions provide a global biogeographic regionalization of the Earth's freshwater biodiversity. These units are distinguished by patterns of native fish distribution that are a result of large-scale geoclimatic processes and evolutionary history. The freshwater ecoregion boundaries generally, though not always, correspond with those of watersheds. Within individual ecoregions there will be turnover of species, such as when moving up or down a river system, but taken as a whole an ecoregion will typically have a distinct evolutionary history and/or ecological processes (Abell et al. 2008).

**Freshwater Ecological Drainage Unit:**

Within Freshwater Ecoregions, Ecological Drainage Units (EDUs) group watersheds that share a common zoogeographic history, physiographic and climatic characteristics, and therefore likely have a distinct set of freshwater assemblages and habitats. EDUs are hypothesized to account for the variability within freshwater ecoregions due to finer-scale drainage basin boundaries and physiography. EDUs are delineated as groups of 8-digit US Geological Survey Hydrologic Unit watersheds. EDUs were qualitatively defined by the TNC Freshwater Initiative for most of the U.S. using primarily USFS Fish Zoogeographc Subregions, USFS Ecoregions and Subsections, and major drainage divisions. EDUs were developed for a small portion of the country by the Missouri Resource Assessment Partnership (Map 7).

Map 7. Ecological Drainage Units



**Datasets for Part 1: Size, Gradient, Ecological Drainage Unit, Freshwater Ecoregion, Northeast Temperature Class, TNC freshwater portfolio**

Datasets representing the reach attributes compiled for Part 1, along with the Northeast Temperature Model (see Part 3 of this report) , and TNC freshwater portfolio can be found in /Part 1folder which represents the data that were distributed to SARP 3/2012. In addition to this previously available data, ArcGIS .lyr files for the key classification attributes are also now included in a new ArcGIS\_Lyr folder.

* **Reach Tables For Central and Western SARP:**  TNC calculated attributes for size classes, gradient classes, Ecological Drainage Unit, Freshwater Ecoregion, and Unit Runoff Coefficient along with related source NHDPlus raw data attribute fields (13 attributes) with short metadata. Data was provided according to NHDPlusdrainage regions and separated into two groups due to file sizes.
  + /regions7\_8\_10/distribute\_reg7810\_3\_2012.dbf, distribute\_reg7810\_3\_2012\_fielddefinitions.xls
  + /regions11\_12\_13/distribute\_reg111213\_3\_2012.dbf, distribute\_reg111213\_3\_2012\_fielddefinitions.xls
* **Reach Tables for Eastern SARP:** TNC calculated attributes for size classes, gradient classes, Ecological Drainage Unit, Freshwater Ecoregion, and Unit Runoff Coefficient along with related source NHDPlus raw data attribute fields (13 attributes). In addition, for the eastern SARP region we provide additional attributes on baseflow index, local and cumulative air temperature, local and cumulative precipitation, and estimated temperature class based on the Northeast Temperature Model along with short metadata.
  + /regions2\_3\_5\_6/distribute\_reg2356\_3\_2012.dbf, distribute\_reg2356\_3\_2012\_fielddefinitions.xls
* **Layer Files: /ArcGIS\_lyrs:** ArcGIS .lyrs for size, gradient, Ecological Drainage Unit, and Freshwater Ecoregion
* **TNC Eastern Division freshwater portfolio shapefiles:** with short metadata
  + Standardized FW Portfolio View 2\_8\_2012.zip: includes TNC portfolio NHDPlus flowlines, NHDPlus lakes, and NRCS HUC12s/watershed areas.

**Part 2: Geology, Soils, Baseflow Index, Landforms, and NLCD 2006 Attributes for the Eastern SARP Region**

We attributed the local catchment of each stream reach in the eastern portions of USGS drainage regions 2, 3, 5, and 6 with its available bedrock geology, soils (texture, available water capacity, organic carbon), USGS Baseflow Index, TNC modeled landforms, and NLCD 2006 landcover. These attributes were compiled to aid in the SALCC hydrologic modeling effort (Part 3 of this report), and they will be useful to future stream classification efforts in the eastern SARP region. These attributes are briefly described below:

Bedrock Geology

Bedrock geology classes included the following 7 ecologically relevant classes:

1. Acidic Granitic

2. Acidic Sedimentary/Metasedimentary

3. Acidic Shale

4. Mafic/Intermediate Granitic

5. Ultramafic

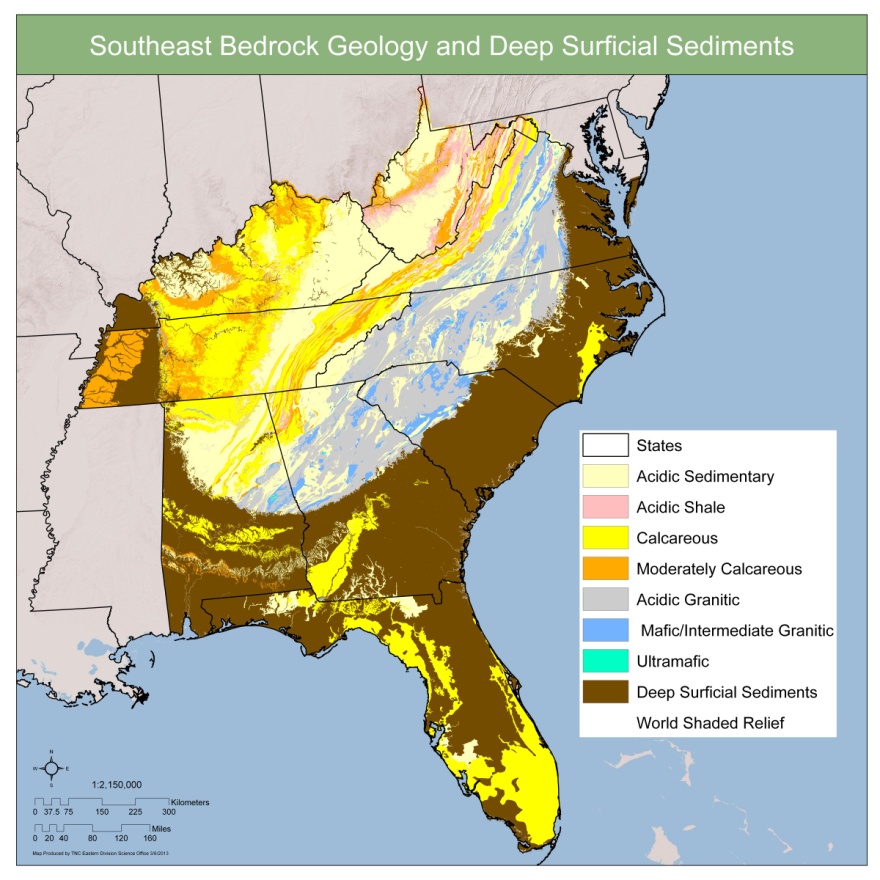
6. Calcareous Sedimentary/Metasedimentary

7. Moderately Calcareous Sedimentary/Metasedimentary

The percentage of each of the above geology types was calculated for the local reach catchments where bedrock geology data was available. Current USGS Bedrock geology maps for the northeastern states were compiled in digital form at a scale of 1:125,000- 1:250,000 (Appendix 1, Map 8). The data was reclassified into seven major bedrock classes according to the rocks' texture, resistance, and chemistry properties (Anderson et al. 1999, Appendix 1). The process of crosswalking the hundreds of state geologic formations into the seven regional major ecological bedrock types took much time, effort, and review throughout 2012-2013. In addition to the bedrock categories, a deep surficial sediments category where bedrock was very deeply buried within the coastal plain was added to highlight areas where bedrock was too deeply buried to be mapped or have high ecological influence on surface natural communities. In these areas, the soils information would likely have more ecological influence on stream community types, water chemistry, and hydrology.

The relationship of the mapped bedrock geology types in the eastern U.S. to the acid neutralizing capacity of streams and rivers has been investigated by a number of studies and used as a key stream classification attribute. The Northeast Aquatic Habitat Classification (Olivero and Anderson, 2008) investigated the relationship of underlying geology to known stream pH locations, examined the relationships between rare aquatic species and geology, and used geologic buffering capacity as a stream classification attribute. In the report and accompanying state atlas "Geologic Control of Sensitivity of Aquatic Ecosystems in the United States to Acidic Deposition" (Norton 1980), the sensitivity of aquatic ecosystems to acidic precipitation was also shown to depend largely on the capacity of the drainage basin bedrock to assimilate acid during chemical weathering. Even small amounts of limestone in a drainage basin can exert an overwhelming influence on terrains that otherwise would be very vulnerable to acidification (Norton 1980). Because aquatic organisms need water pH to be within a certain range for optimal growth, reproduction, and survival, the geologic setting’s influence on stream buffering capacity can play a large role in structuring aquatic communities at the reach scale. Although this project compiled bedrock geology and some initial stream pH point data (Herily, per com.), there was not sufficient time to develop a full stream buffering capacity model or pH stream classification attribute for the region. Further research in the southeast should be done to confirm and investigate the relationship of bedrock geology to pH and instream natural communities.

Map 8. Bedrock and Deep Surficial Sediments



Soils

Soil characteristics influence stream habitat conditions in terms of instream substrates, water chemistry, and hydrologic regime. Detailed fine scale maps of soil characteristics, however, have not previously been available across broad regions. As part of NRCS’s efforts to make the county soil survey information (SSURGO) more easily available across a broader geographic area, TNC obtained selected soils attributes from the SSURGO dataset from the NRCS December 30, 2009 snapshot (Bliss, 2013, per comm). The soil attributes were then linked to the appropriate “map unit” and these millions of mapunit polygons were transformed into 30m grid surfaces (Maps 9, 10, and 11). Where SSURGO soil attributes were not available for a particular county, the coarser scale NRCS STATSGO data was used to fill the holes, per the recommendation from NRCS (Waltman per. comm). Thirteen soils attributes related to soil texture (sand, silt, clay), available water capacity, and organic carbon attributes were sampled in each available NHDPlus reach catchment to provide a mean value for each attribute for each reach catchment (Table 3). Currently some county lines can still be seen in the soils data given the differences in methods and time periods used between counties and there was no way to improve this issue. NRCS is aware of this county to county variation issue, and it is the goal of their “harmonization project” over the next three years to address and resolve these issues through detailed study and reassignment of soil types across county lines. Despite this issue in the current soils data, these data are still a great improvement over previous soils maps and their utility was shown as these variables were highly used by the Random Forest Hydrologic Class model (see Part 3 of this report).

Table 3: Soils Attributes

*Available Water Capacity*

awc\_tp: Available Water Capacity, Total Profile, in cm \* 100

awc05 = Available Water Capacity, 0-5cm layer zone, in cm \* 100

awc520 = Available Water Capacity, 5-20 cm layer zone, in cm \* 100

awc2050 = Available Water Capacity, 20-50 cm layer zone, in cm \* 100

awc50100 = Available Water Capacity, 50-100 cm layer zone, in cm \* 100

awc\_deep = Available Water Capacity, Thickness this variable was sampled to in cm

*Soil Organic Carbon*

soc\_tp = Soil Organic Carbon, total profile g/sq.m

soc05 = Soil Organic Carbon, 0-5 cm g/sq.m

soc520 = Soil Organic Carbon, 5-20 cm g/sq.m

soc\_deep = Soil Organic Carbon, thickness cm

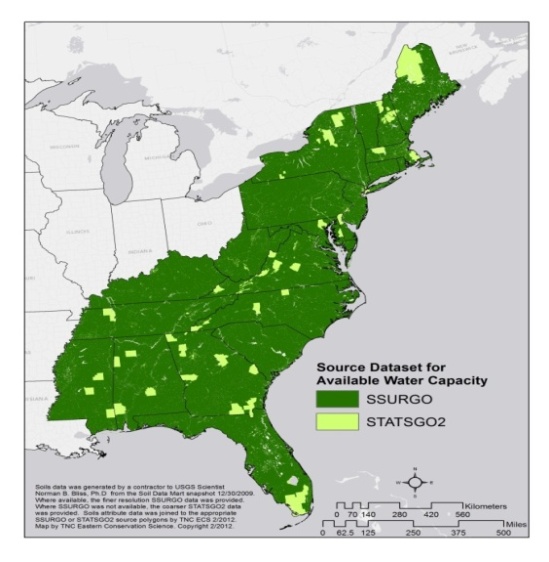
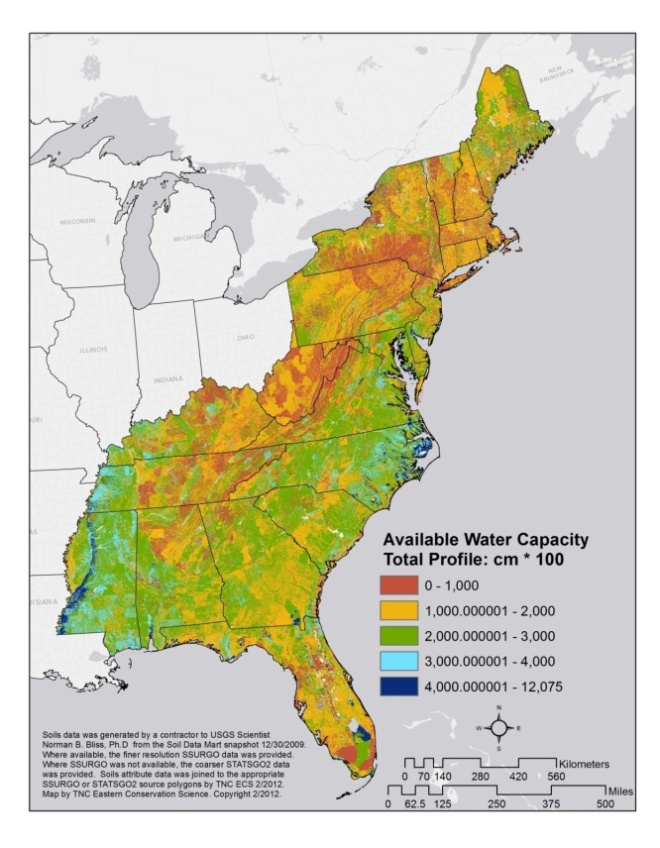
*Soil Texture*

sand0\_20 = % sand in the 0-20cm thickness zone

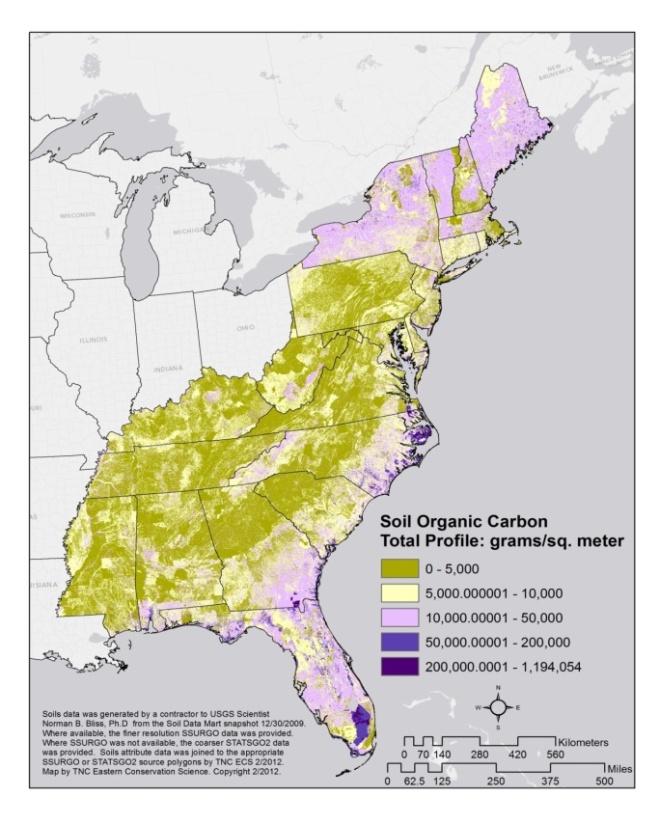
silt0\_20 = % silt in the 0-20cm thickness zone

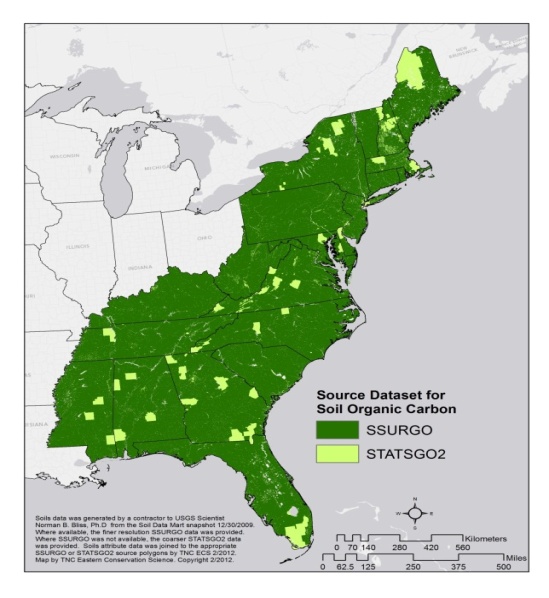
clay0\_20 = %caly in the 0-20cm thickness zone

Maps 9: 9A.) Source Datasets for Available Water Capacity and 9B.) Available Water Capacity Total Profile

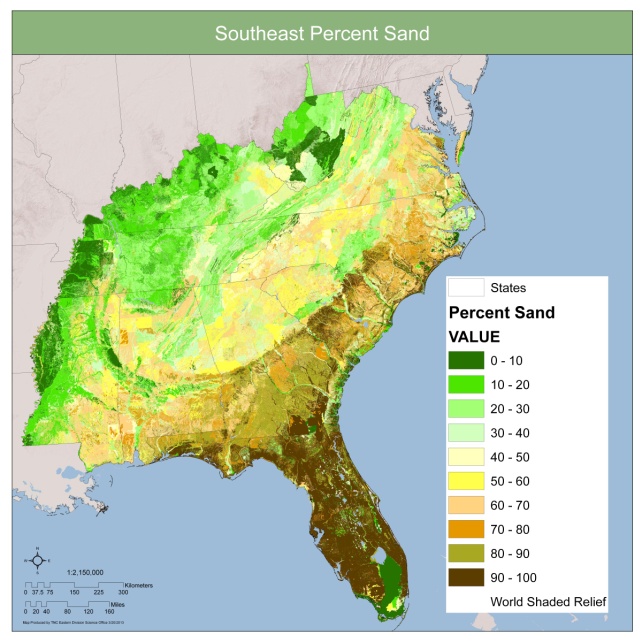
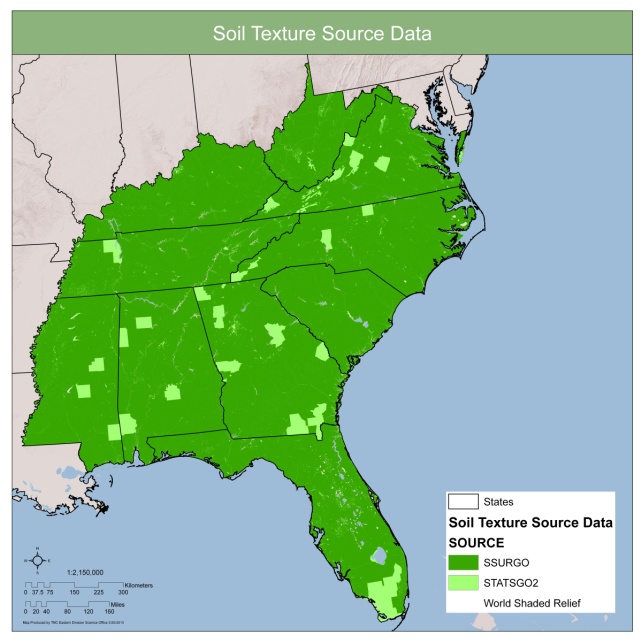


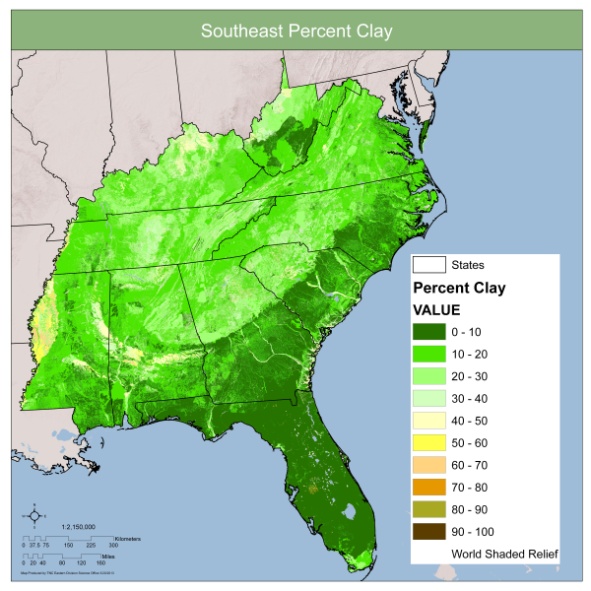
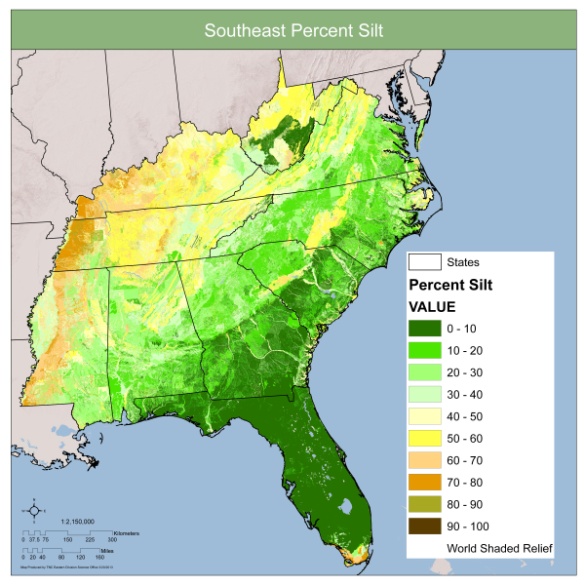
Map 10: 10A.) Source Datasets for Soil Organic Carbon and 10B.) Soil Organic Carbon





Map11: 11A.) Source Datasets for Soil Texture, 11B.) Percent Sand, 11C.)Percent Silt, 11D.) Percent Clay

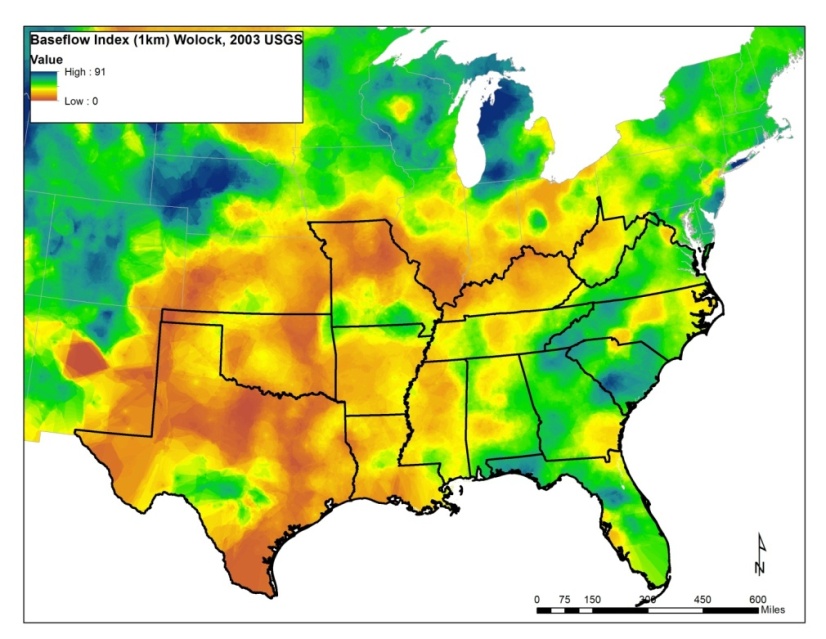




Baseflow Index

The only regionally available dataset for mapping groundwater inflow or baseflow was a Baseflow Index (BFI) USGS Wolock 2003 dataset. BFI is the ratio of base flow to total flow volume for a given year. This 1-kilometer raster (grid) dataset for the conterminous United States was created by interpolating base-flow index (BFI) values estimated at U.S. Geological Survey (USGS) streamgages (Wolock, 2003). The BFI values at the gages were generated using a complex program developed by USGS. The BFI Web page (http://www.usbr.gov/pmts/hydraulics\_lab/twahl/bfi/index.html) states: “The method combines a local minimums approach with a recession slope test. The program estimates the annual base-flow volume of unregulated rivers and streams and computes an annual base-flow index.” To calculate the mean BFI for each local stream catchment, the 1km BFI grid was resampled to a 20m grid cell resolution and the mean BFI for each stream catchment in regions 2, 3, 4, and 5 was generated.

Map 12. Baseflow Index

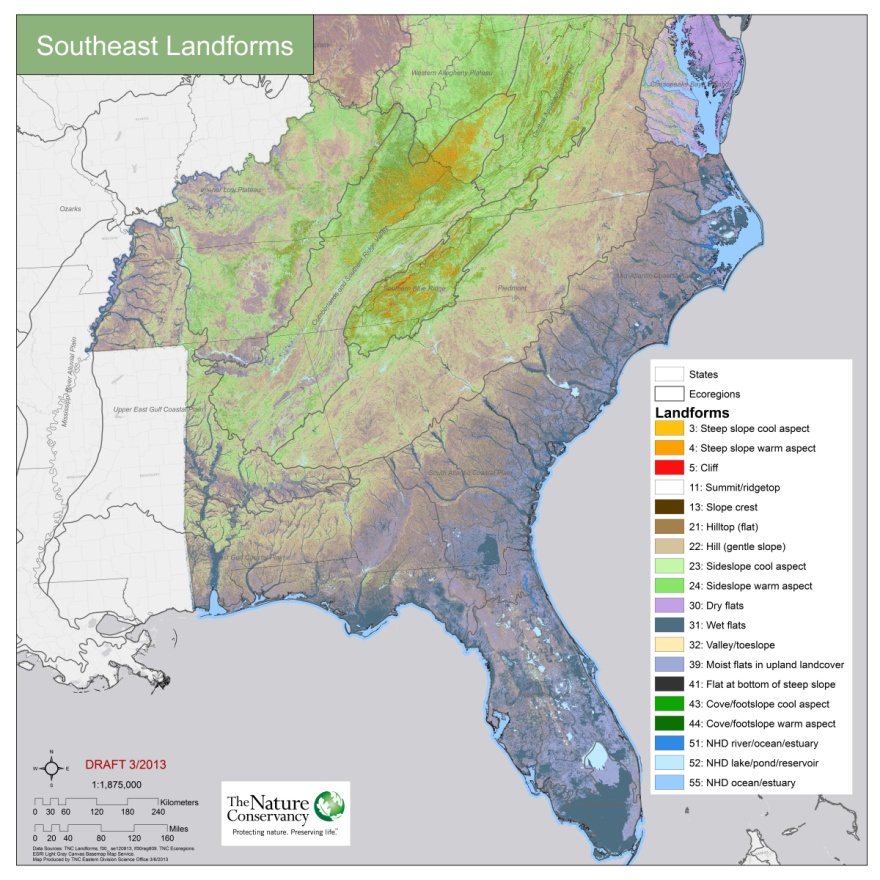


Landforms

The amount of each 30m landform type (n=19) modeled by TNC was calculated for the local catchment of each reach in USGS drainage regions 2, 3, 5, and 6. These tabulations provide a context regarding the confinement of the river and its local catchment topographic setting. Stanley Rowe called landform "the anchor and control of terrestrial ecosystems." Landforms break up broad landscapes into local topographic units, and in doing so provide for meso- and microclimatic expression of broader climatic character. Landform is largely responsible for local variation in solar radiation, soil development, moisture availability, and susceptibility to wind and other disturbance. As one of the five "genetic influences" in the process of soil formation, it is tightly tied to rates of erosion and deposition, and therefore to soil depth, texture, and nutrient availability. These are, with moisture, the primary edaphic controllers of plant productivity and species distributions. If the other four influences on soil formation (climate, time, parent material, and biota) are constant over a given space, it is variation in landform that drives variation in the distribution and composition of natural communities.

Landform is a compound measure, which can be decomposed into the primary terrain attributes of elevation, slope, aspect, surface curvature, and upslope catchment area. The wide availability and improving quality of digital elevation data has made the quantification of primary terrain attributes possible. TNC adopted the Fels and Matson (1997) approach to landform modeling. They described a metric that combines information on slope and landscape position to define topographic units such as ridges, sideslopes, coves, and flats on the landscape (Map 13). Recognizing the ecological importance of separating occurrences of “flats” into primarily dry areas and areas of high moisture availability, we also calculated a simple moisture index that maps variation in moisture accumulation and soil residence time. We used National Wetlands Inventory (NWI) datasets to calibrate the index and set a wet/dry threshold, then applied it to the flats landform to make the split. The parent dataset used to construct the landforms is the 30m National Elevation Dataset (NED) digital elevation model (DEM) developed by the USGS.

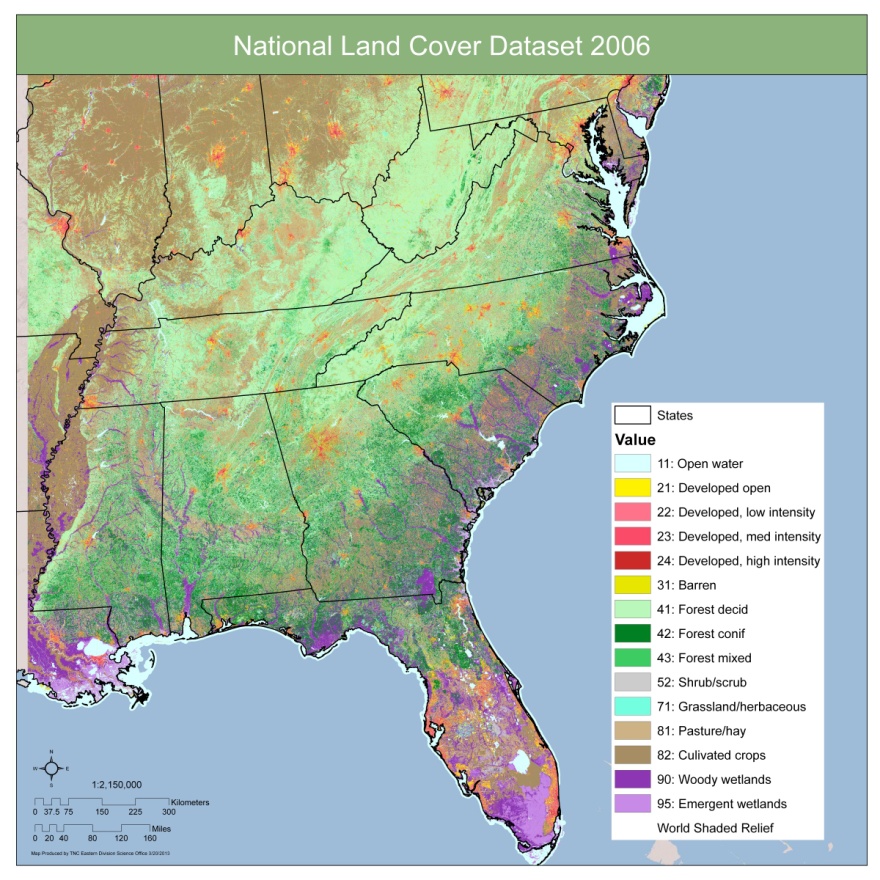
Map 13. TNC Modeled Landforms, 30m resolution



Landcover

The amount of each of the National Land Cover Database 2006 (NLCD2006) classes within the local NHDPlus catchment of each reach in USGS drainage regions 2, 3, 5, and 6 were calculated. NLCD 2006 is a 16-class land cover classification scheme that has been applied consistently across the conterminous United States at a spatial resolution of 30 meters (Map 14). NLCD2006 is based primarily on the unsupervised classification of Landsat Enhanced Thematic Mapper+ (ETM+) circa 2006 satellite data.

Map 14. National Land Cover Dataset 2006



**Datasets for Part 2: Geology, Soils, Baseflow Index, Landform, and NLCD 2006 for the Eastern SARP Region**

Datasets for Part 2 are divided into USGS drainage regions 2, 3, 5, and 6. Tables were developed that have the amount of the various classes for each input dataset within the catchment for all reaches where a particular attribute was available.

Reach Tables

* + /Region2: reg2\_geo.dbf, reg2\_soiltexture.dbf, reg2\_soilsocawc.dbf, reg2\_landform.dbf, region2\_NLCD06.dbf
  + /Region3: reg3\_geo.dbf, reg3\_soiltexture.dbf, reg3\_soilsocawc.dbf, reg3\_landform.dbf, region2\_NLCD06.dbf
  + /Region5: reg5\_geo.dbf, reg5\_soiltexture.dbf, reg5\_soilsocawc.dbf, reg5\_landform.dbf, region2\_NLCD06.dbf
  + /Region6: reg6\_geo.dbf, reg6\_soiltexture.dbf, reg6\_soilsocawc.dbf, reg6\_landform.dbf, region2\_NLCD06.dbf
  + Metadata: part2\_datatables\_field\_definitions.xlsx

(Please note that the baseflow index was an attribute provided in the previous Part 1 reach attribute tables.)

**Part 3: Models for Temperature Class and Hydrologic Class**

Temperature Class:

Stream temperature has been noted as a key stream classification variable as it sets the physiological limits where stream organisms can persist. Seasonal changes in water temperature often cue development or migration, influence growth rates of eggs and juveniles, and can affect the body size, and therefore the fecundity of adults. In addition to limiting effects on biological productivity, temperature extremes may directly preclude certain taxa from inhabiting a water body. Stream temperatures vary on seasonal and daily time scales, and among locations due to climate, elevation, and the relative importance of groundwater inputs (Allan 1995). High elevation areas with low average air temperatures tend to maintain coldwater streams year-round. In low elevation areas, groundwater inflow can also play a role in maintaining cold and cool water streams.

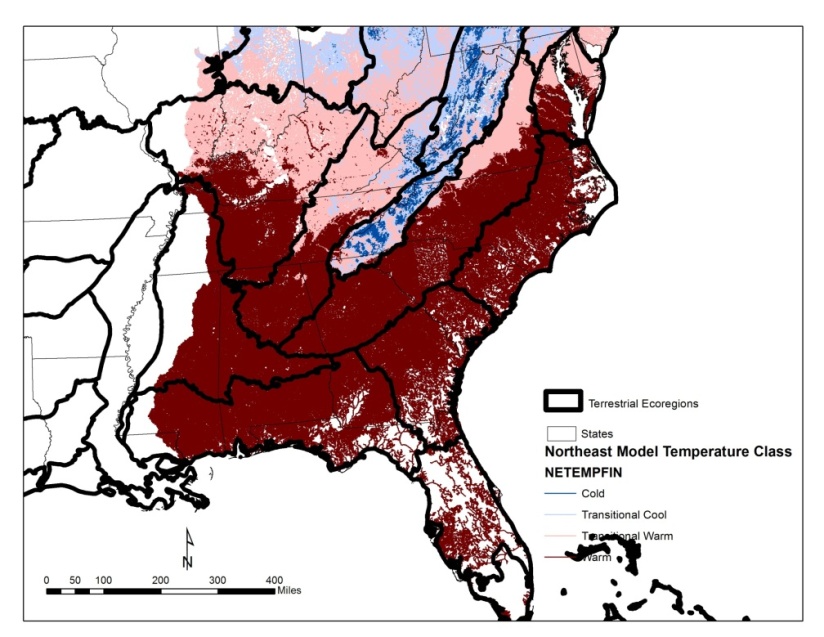
No widely accepted method exists for estimating the expected natural instream water temperature at the stream reach scale in the eastern US. In the northeast, the NAHCS developed a classification and regression tree analysis (CART, Steinberg, and Colla 1997) model relating four major water temperature classes (Table 4) to differences in stream size, air temperature, gradient, and groundwater inputs. For headwaters to small rivers (sizes 1a, 1b, and 2), the most useful classification variables included the cumulative upstream air temperature, stream gradients, and the local baseflow index. For larger rivers (sizes 3a, 3b, 4, and 5), the most predictive variables were cumulative upstream air temperature and stream size class. A detailed set of final decision rules was used to place all NHDPLUSreaches in the northeast into these four temperature classes based on GIS variables measured for each reach (Olivero and Anderson, 2008).

Table 4: Conceptual Guidance for Northeast Water Temperature Classes



There was consensus that the northeast water temperature model was generally applicable to the southeast. The SARP review committee asked TNC to extend the northeast model to the southeast and allow review of the mapped results (Map 15). Although the application of this model to the southeast stream reaches highlighted areas that made sense to the team, there was disagreement over whether the southeast region really needed an additional warm class. Some team members felt an additional tropical class was necessary, while others felt the southern fish fauna was dominated by cyprinids and centrarchids that had a very wide range of tolerances for temperature and it was other factors (biogeographic history/separation, gradient, substrate, river confinement, flow) that determined stream aquatic biological communities within these warm water streams and rivers. Although the team desired to use a fish trait data to help investigate these questions, no fish trait information on upper water temperature preference was available. A cluster analysis of fish sample data provided by SARP also did not prove useful for quick refinement of the temperature map. Future work must be done to justify and refine the temperature classes and map before it can be deemed appropriate to represent the freshwater aquatic communities of the southeast.

Map 15. Northeast Model of Temperature Class Applied to the Southeast



Hydrologic Class

The stream classification workgroup reviewed a number of hydrologic classifications that could potentially be used in the southeast. These included Konrad 2008, McManamay 2011, and Environmental Flow Specialists, Inc (EFS 2012). After an initial test run to predict McManamay classes for ungaged stream reaches in a portion of the SALCC using GIS attributes, SARP recommended implementation of a final hydrologic class model using Environmental Flow Specialists, Inc.’s new national hydrologic hierarchical classification. The EFS classification has been considered by some to be more intuitive and flexible given its hierarchical approach, and the SARP review committee found the new EFS approach to be ecologically relevant and a good alternative to the other hydrologic class multivariate clustering approaches.

The EFS classification is based on the following attributes: perennial/intermittent, size, variability, flood frequency, and seasonality. These attributes can be considered in a hierarchy or “dichotomous key” chart form, and/or can be considered as separate components. After detailed review of the specific attributes of the EFS classification, only the **variability class** was chosen for final model implementation. The review committee felt variability of stream flow is important and the EFS classification worked relatively well in the SALCC region. The other EFS attributes were either found to have problems in definition or in practice were not relevant in the southeast. For example, the team found problems with how EFS defined perennial/intermittent and persistency using a zero flow day value. The committee also did not see the utility in this region for the flood frequency attribute which is the median # days/yr of very high flooding (90th percentile flow). In the east, this flow level roughly corresponds to bank full events, and the committee felt higher flows (e.g., 95th or 99th percentile) were need to represent overbank flood duration events in the east. For the flood timing attribute, there is not much variation in the seasonality of high flows in the SALCC region so this attribute could be dropped. Although stream size is a widely recognized attribute of river ecology, we already had classified stream size by mean annual flow and drainage area and did not wish to model the EFS size classes which were based on median annual flow.

Hydrologic Variability: *Southeast Classes*

Streamflow varies over many different time scales. Day-to-day, streamflow rises and falls in response to runoff and groundwater inflows. Median Daily Variability % was the gage statistic used by Environmental Flow Specialists, Inc. to separate low variability (<= 189%) from moderate variability (>189%) for streams and rivers in their national classification system. Although the EFS variability classification worked relatively well in the SALCC region, the SARP review team recommended two additional subclasses be modeled if possible. A subclasses of high and very high variability was developed based on the break points for variability of intermittent streams (>189-272% and >272%; Davis, per comm). Low variability subclasses were developed based on natural breaks in the region. Given the 328 gages in the SALCC, the natural break for those with values <= 189 was found to be <=118 and 119 – 189. Using these four classes would keep the model consistent with the national classification by maintaining the 189 split, but make the classes more applicable to regional conditions in the southeast.

Hydrologic Variability*: Model*

The modeling work consisted of four major steps.

1. Compile set of gages, assign hydrologic class, and link them to the appropriate NHDPlusreach
2. Attribute each stream reach and gage with GIS predictor variables
3. Build random forest (RF) classification models using the *randomForest* package in in R
4. Apply the best RF model to each stream reach and map each stream reach according to the “highest probability” class.

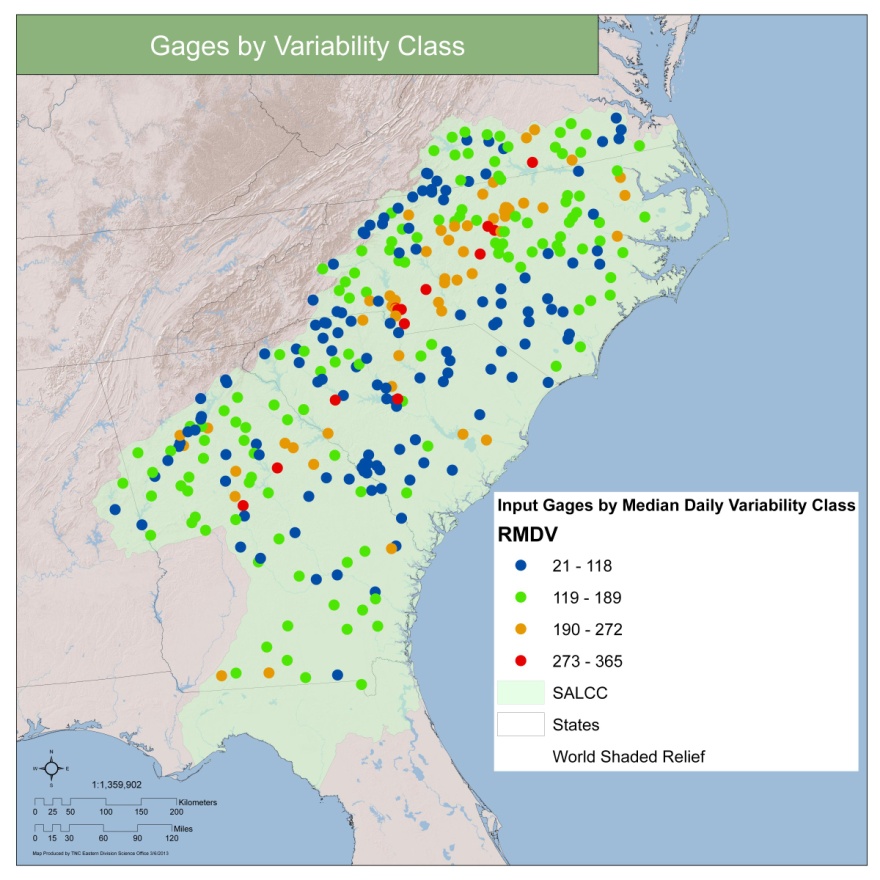
Each of these steps will be described below along with a short discussion of the results relevant to each step.

Step 1. Gage Compilation and Linkage: There were 328 gages in the SALCC area provided by Environmental Flow Specialists, Inc. (Map 16). These included 134 in the very low, 127 in the low, 54 in the moderate, and 13 in the high variability classes (Table 5). The gages were all linked to the appropriate NHDPlus reach based on the streamgage to NHDPlus comid linkage table provided by USGS. Most gages fell on our small river size class, however creeks through medium mainstem rivers also had a large number of gages present (Table 4).

Table 5: Number of Input Gages in the SALCC by Hydrologic Variability Class and Stream Size



Map 16. Input Gages for Hydrologic Variability Class Model



Step 2. Attribute each stream reach with GIS predictor variables

The gages covered parts of USGS drainage regions 2 and 3. Each reach in these drainages had been tagged with over 80 environmental and geographic variables as part of this stream classification project. The region 2 and region 3 attribute tables were merged and 75 potential predictor variables were used as inputs for the Random Forest model (Table 6). This included a variety of variables related to stream size, gradient, elevation, baseflow index, local and cumulative precipitation, local and cumulative air temperature, soils, geology, landforms, and landcover. The selection of variables was guided by the previous model run where 34 predictor variables were used to explore prediction of the McManamay hydrologic classes. After this work, the team recommended eliminating the geographic stratification variables of EDU, freshwater and terrestrial ecoregions. The group also recommended inclusion of additional variables related to soil texture, bedrock geology, landforms, and land cover.

Table 6. GIS Predictor Attributes Available for each Gage and each NHDPlus Stream Reach in SALCC



Step 3. Build Random Forest classification models

To build the hydrologic classification model, we used the Random Forest (RF) algorithm as implemented in the *randomForest* package in R. Random Forest (RF) is a machine learning technique that builds hundreds of decision trees to assess the relationship between a response variable and potential predictor variables (Breiman 2001). Regression trees are built for continuous response variables while classification trees are created for categorical variables. Depending on the type of response variable, the resultant model can then be used to predict class membership or estimate a continuous response for unknown samples.

We experimented with various RF parameters including the number of trees, input source class sample sizes, and number of predictor variables. When attempting to model all four source variability classes, no model was found that could correctly place samples into the high variability class with less than 80% error in that class and less than 30% overall model error. This was likely due to the very small number of input gages (n=13) in this category. This high variability class had to thus be combined with the moderate variability class for further model development. After combining this high variability class with the moderate class, it was possible to obtain a model with an acceptable separate class and overall error rate when trying to model just the three variability classes of very low (Var 1:<= 118%), low (Var 2: 118-189%), and moderate-high (Var 3-4: >189%). After more experimentation with various numbers of trees, input class sample sizes, and numbers of predictor variables, we obtained the best model using equal input source sample sizes (67, 67, 67) to evenly allocate model development space to each of the three input variability categories, 1000 trees built, and 20 variables tried at each node split in the model. The best model had an overall error rate of 23.17% and class error rates ranged from a low of 16% for the very low variability class to 28% for both the low and moderate-high variability classes (Table 7).

Table 7. Random Forest Error Report



Reviewing the variable importance table (Table 8) reveals which of the 75 predictor variables were most helpful in predicting the hydrologic variability class, the variables that were most important to the model overall, and also for predicting specific classes. Overall, the most important variable was the mean baseflow index (BSFLMEANI) variable. This variable was also the most important in predicting the individual three classes. Other variables that were particularly important include measures of stream size (mean annual flow (MAFLOWU) and cumulative drainage area (CUMDRAINAG)) and the run-off coefficient (MAFDASKM). Additionally in the top 10 overall importance variables, we find in decreasing order of importance, the cumulative upstream area weighted annual precipitation (AREAWTMAP) , % silt, % pasture (NLCD81), cumulative upstream area weighted mean annual temperature (AREAWTMAT), % water (NLCD11), and % sand. Within the top 11th-20th overall importance variable list, we find strong influence from % hill gentle slope landforms (LF22), mean soil organic carbon total profile (M\_SOC\_TP), the local catchment precipitation (PRECIP), % clay, reach slope, mean available water capacity 50-100cm (M\_AWC50100), longitude, latitude, mean soil organic carbon in 5-20cm layer (M\_SOC\_520), and % of all hilly landforms (toeslope, hill gentle slope, hilltop).

The individual class importance table columns highlight that to build the model for each specific variability class, certain variables were more or less important. For example, the top 10 variables for VAR\_1 class are identical to the top ten overall, except for the addition of the % hill gentle slope landforms (LF22) and omission of % sand. The top 5 variables overall and for VAR\_1 are also identical and in the same order of importance to the overall model importance variable ranking. However, the models for VAR\_2 and VAR\_3-4 are more different from the overall model. They share seven of their top ten variables in common with the overall model rankings, however these seven are in different orders of importance after baseflow index. For example, we can see by studying the different order of importance that % sand and % silt were more important to the VAR\_3 model, while % sand and % silt were less important to the VAR\_2 model than these variables were to the overall model or to the VAR\_1 model. For VAR\_2, we find three additional variables in its top ten list: longitude, mean available water capacity 50-100cm, and local catchment precipitation. For the VAR\_3-4 model, we three other new important variables in its top ten: % total hilly landforms, % cultivated crops, and mean soil organic carbon in deeper than 20cm.

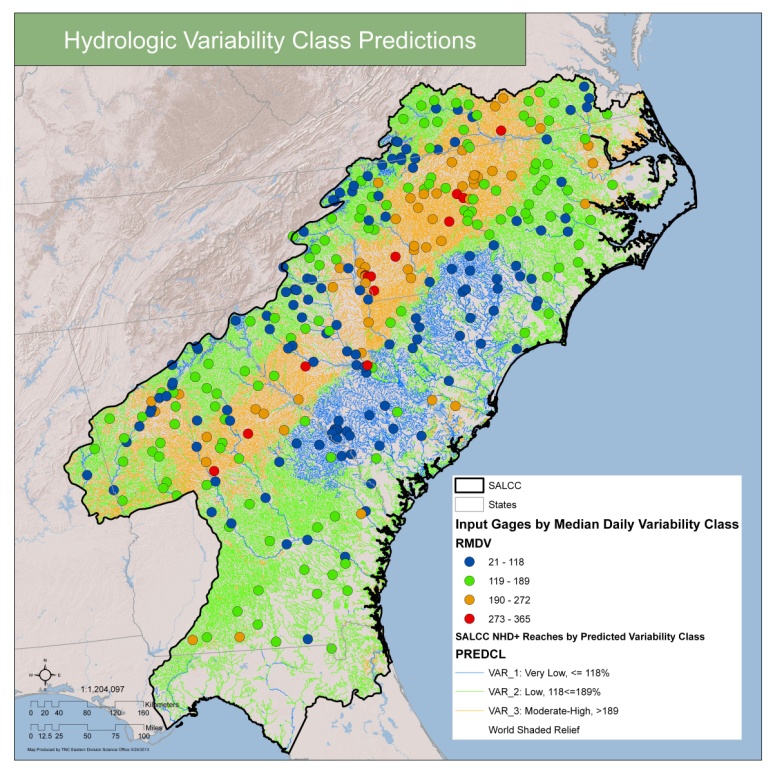
Table 8: Random Forest Variable Importance Table. Within each class, variables are ordered by their importance in the classification model.



Step 4. Apply best model to stream reaches

Each NHDPlus reach in the SALCC was mapped according to the “highest probability” class membership (Map 17) predicted by the model. Results show an area of very low variability throughout the South Carolina coastal plain, extending a bit north into North Carolina and south into Georgia. Most of the larger rivers throughout the SALCC are also in this class. Outside this very low variability class in the coastal plain, the rest of the coastal plain is in the low variability class. As you move out of the coastal plain into the foothills of the Piedmont there is a large swath of moderate to high variability class streams and small rivers. To the north of this swath, low variability streams are again present.

Map 17. Hydrologic Variability Class Predictions



**Datasets for Part 3: Models**

The temperature model was completed in 3/2012 and this attribute for reaches in regions 2, 3, 5, and 6 was included in the original data for Part 1.

The hydrologic class model was completed 3/29/2013. Within the folder /Part3/R\_Runs\_March\_29\_2013 we have included the SALCC input gages, SALCC input reaches, R code and output model object, and SALCC reaches assigned to their hydrologic variability class in the shapefile: salcc\_nhdflow\_model3\_29\_2013.shp, with .lyr “Reaches by Predicted Variability Class.lyr “

**References**

# Abell R.A. Editor. 2008. Freshwater Ecoregions of the World. The Nature Conservancy and World Wildlife Fund, Inc. <http://www.feow.org>

Allan, J. D. 1995. Stream Ecology: Structure and function of running waters. Kluwer Academic Publishers. Dordrecht, The Netherlands.

Anderson, M.G. 1999. Viability and spatial assessment of ecological communities in the Northern Appalachian Ecoregion. Ph.D. Dissertation. University of New Hampshire.

Beard, D. and Whelan, G. 2006. A framework for assessing the nation’s fish habitat. National Fish Habitat Science and Data Committee. Draft Report. 75p.

Breiman, Leo. Random forests. 2001. *Machine Learning Journal*, 45:5–32.

Environmental Flow Specialists, Inc (EFS). 2012. Draft national hydrologic hierarchical classification, www.eflowsppecialists.com

Konrad, C. 2011. Draft Hydrologic classification of rivers and streams in the southeastern United States.*,* The Nature Conservancy and US Geological Survey.

McManamay, R. Orth, D.J, Dolloff, C.A., and Frimpong, E.A. 2011. A Regional Classification of Unregulated Stream Flows: Spatial Resolution and Hierarchical Frameworks. River Research and Applications 2011: DOI 10.1002/rra.1493

Norton, S.A. 1980. Distribution of Surface Waters Sensitive to Acidic Precipitation: A State-Level Atlas. National Atmospheric Deposition Program. 65p.

NLCD2006 citation: Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J., 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, PE&RS, Vol. 77(9):858-864.

Olivero, A. and Anderson, M. 2008. Northeast Aquatic Habitat Classification. The Nature Conservancy. 88p. http://rcngrants.org/spatialDataRosgen, D.L. 1994. A classification of natural rivers. *Catena* 22: 169-99.

Smith R. 2011. A Classification and Threat Condition Assessment of the Rivers and Streams of Texas. Texas Chapter American Fisheries Society Annual Meeting presentation.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Exported for TNC based on December 30, 2009 Snapshot by Norman Bliss PhD. Contractor to U.S.G.S.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. U.S. General Soil Map (STATSGO2). Exported for TNC based on December 30, 2009 Snapshot by Norman Bliss PhD. Contractor to U.S.G.S.

USGS National Hydrography Dataset Plus (NHD-Plus), 2006. 100,000.

Vannote, RL,G. W. Minshall, K. W. Cummins, J.R. Sedell, and E. Gushing 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37: 130-137.

Wolock, D.M. 2003. Base-Flow Index Grid for the Conterminous United States. 1km grid dataset. USGS Open-File Report: 2003-263

Appendix I: Geology Metadata

Bedrock geology classes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Geology class* | *Lithotypes* | *Meta-equivalents* | *Comments* | *Some characteristic communities* |
| 100: ACIDIC SEDIMENTARY / METASEDIMENTARY: fine- to coarse-grained, acidic sed/metased rock | Mudstone, claystone, siltstone, non-fissile shale, sandstone, conglomerate, breccia, greywacke, arenites | (Low grade:) slates, phyllites, pelites; (Mod grade:) schists, pelitic schists, granofels | Low to moderately resistant rocks typical of valleys and lowlands with subdued topography; pure sandstone and meta-sediments are more resistant and may form low to moderate hills or ridges | Many: low- and mid-elevation matrix forests, floodplains, oak-pine forest, deciduous swamps and marshes |
| 200: ACIDIC SHALE: Fine-grained acidic sedimentary rock with fissile texture | Fissile shales |  | Low resistance; produces unstable slopes of fine talus | Shale cliff and talus, shale barrens |
|  |  |  |  |  |
| 300: CALCAREOUS SEDIMENTARY / META-SEDIMENTARY: basic/alkaline, soft sed/metased rock with high calcium content | Limestone, dolomite, dolostone, other carbonate-rich clastic rocks | Marble | Lowlands and depressions, stream/river channels, ponds/lakes, groundwater discharge areas; soils are thin alkaline clays, high calcium, low potassium; rock is very susceptible to chemical weathering; often underlies prime agricultural areas | Rich fens and wetlands, rich woodlands, rich cove forests, cedar swamps, alkaline cliffs |
|  |  |  |  |  |
| 400: MODERATELY CALCAREOUS SEDIMENTARY / METASED: Neutral to basic, moderately soft sed/metased rock with some calcium but less so than above | Calc shales, calc pelites and siltstones, calc sandstones | Lightly to mod. metamorphosed  calc pelites and quartzites, calc schists and phyllites, calc-silicate granofels | Variable group depending on lithology but generally susceptible to chemical weathering; soft shales often underlie agricultural areas | Rich coves, intermediate fens |
|  |  |  |  |  |
| 500: ACIDIC GRANITIC: Quartz-rich, resistant acidic igneous and high grade meta-sedimentary rock; weathers to thin coarse soils | Granite, granodiorite, rhyolite, felsite, pegmatite | Granitic gneiss, charnockites, migmatites, quartzose gneiss, quartzite, quartz granofels | Resistant, quartz-rich rock, underlies mts and poorly drained depressions; uplands & highlands may have little internal relief and steep slopes along borders; generally sandy nutrient-poor soils | Many: matrix forest, high elevation types, bogs and peatlands |
|  |  |  |  |  |
| 600: MAFIC / INTERMEDIATE GRANITIC: quartz-poor alkaline to slightly acidic rock, weathers to clays | (Ultrabasic:) anorthosite (Basic:) gabbro, diabase, basalt (Intermediate, quartz-poor:) diorite/ andesite, syenite/ trachyte | Greenstone, amphibolites, epidiorite, granulite, bostonite, essexite | Moderately resistant; thin, rocky, clay soils, sl acidic to sl basic, high in magnesium, low in potassium; moderate hills or rolling topography, uplands and lowlands, depending on adjacent lithologies; quartz- poor plutonic rocks weather to thin clay soils with topographic expressions more like granite | Traprock ridges, greenstone glades, alpine areas in Adirondacks |
|  |  |  |  |  |
| 700: ULTRAMAFIC: magnesium-rich alkaline rock | Serpentine, soapstone, pyroxenites, dunites, peridotites, talc schists | | Thin rocky iron-rich soils may be toxic to many species, high magnesium to calcium ratios often contain endemic flora favoring high magnesium, low potassium, alkaline soils; upland hills, knobs or ridges | Serpentine barrens |

**SOUTHEAST GEOLOGY SOURCES:**

ALABAMA Citation\_Information:

Originator: Geological Survey of Alabama

Publication\_Date: 2006

Title: Digital Geologic Map of Alabama Polygons

Edition: first

Geospatial\_Data\_Presentation\_Form: vector digital data

Series\_Information:

Series\_Name: GSA Special Map Series

Issue\_Identification: Special Map 220A

Publication\_Information:

Publication\_Place: Tuscaloosa, Alabama

Publisher: Geological Survey of Alabama

Online\_Linkage: <http://www.gsa.state.al.us>

FLORIDA *Citation:*

*Citation\_Information:*

*Originator:* Florida Department of Environmental Protection

*Publication\_Date:* 2001

*Title:* GEOLOGY (ENVIRONMENTAL)

*Geospatial\_Data\_Presentation\_Form:* Shapefile

*Other\_Citation\_Details:* State of Florida

*Online\_Linkage:* [<http://www.dep.state.fl.us/gis/>](http://www.dep.state.fl.us/gis/)

GEORGIA DIGITAL GEOLOGIC MAP OF GEORGIA (Ver. 2)

georgia department of natural resources

environmental protection division

georgia GEOLOGIC survey

Atlanta

1999

DOCUMENTATION REPORT 99-20

KENTUCKY Geology of Kentucky. Based on Geologic Map of Kentucky,1988, scale 1:500,000. Compiled by Martin C. Noger, Kentucky Geological Survey, from the Geologic Map of Kentucky, scale 1:250,000, 1981 by Robert C. McDowell, George J. Grabowski, and Samuel L. Moore, U.S. Geological Survey. Tectonic and karst interpretations added by Claude S. Dean, 2002.

NORTH CAROLINA Citation\_Information:

Originator: NC DEHNR-Division of Land Resources, NC Geological Survey

Publication\_Date: 19981201

Title: onemap.SDEADMIN.geol

Geospatial\_Data\_Presentation\_Form: vector digital data

Publication\_Information:

Publication\_Place: Raleigh, North Carolina

Publisher: NC DEHNR-Division of Land Resources, NC Geological Survey

Other\_Citation\_Details: NCCGIA distributes this dataset

SOUTH CAROLINA *Citation Information:*

*Originator:* South Carolina Geological Survey

*Publication Date:* 2005

*Title:*

ggms\_poly

*Geospatial Data Presentation Form:* vector digital data

*Series Information:*

*Series Name:* General Geologic Map Series

*Issue Identification:* 1

*Publication Information:*

*Publication Place:* South Carolina

*Publisher:* South Carolina Geological Survey

*Online Linkage:* [\\scdnradmin\data\gisdata\scdata\geology\ggms\_poly.shp](file://scdnradmin/data/gisdata/scdata/geology/ggms_poly.shp)

TENNESSEE *Citation\_Information:*

*Originator:* Greene, D.C., and Wolfe, W.J.

*Publication\_Date:* 2000

*Title:*

Superfund GIS - 1:250,000 Geology of Tennessee.

*Geospatial\_Data\_Presentation\_Form:* vector digital data

*Publication\_Information:*

*Publication\_Place:* Nashville, Tennessee

*Publisher:* U.S. Geological Survey

*Online\_Linkage:* <http://water.usgs.gov/lookup/getspatial?geo250k>