River Basin Restoration Prioritization Project: Task 3 Report, Database and database standards

**Technical Memo**

For DENR Division of Mitigation Services (Nancy Daly)

June 15, 2015

By Larry Band, Lydia Olander, Dean Urban, Jon Duncan, John Fay, John Lovette,

Christy Ihlo, Solomon Vimal

# Executive Summary

This document describes the datasets used for the Division of Mitigation Services (DMS) River Basin Restoration Priorities (RBRP) Project . It includes brief descriptions of the data, primary sources, steps taken to procure and process the data, and how the data are stored and accessed for subsequent analysis. A general screening protocol for data selection was maintained that emphasized primary information sources that are developed, distributed and updated by federal and state agencies, with sufficient metadata and QA/QC information to support analysis and updating of information as it becomes available. We note that in an effort to provide as up-to-date information as available, we make use of recently developed sources, but substantial new information is in the process of being developed for public release. As such, methods for updating or adding new data sets need to be maintained, and this will be supported in the software development for the pilot projects on the Tar-Pamlico and Catawba River Basins.

Metrics to evaluate four ecosystem functions are supported by the data sets selected and described here. In all cases, ecosystem functional models are used that can respond to changes in primary environmental variables to allow evaluation of ecosystem uplift potential through management activities. This approach has the desired outcome of reducing the number and redundancy of input variables, using a formal model of ecosystem states and sensitivity. Three of the ecosystem functions: hydrologic flow regime, water quality emphasizing major nutrient loads, and stream channel and floodplain geomorphology and hydraulics, are developed to focus on a parsimonious dataset necessary for description of current state and potential ecosystem uplift. Nutrient loading and hydrologic flow regime are estimated with statistical models developed by the United States Geological Survey, while stream channel and floodplain geomorphology and hydraulics have been extensively measured and modeled by the North Carolina Floodplain Mapping Program, and are distributed as part of the NC Flood Risk Information System or FRIS (www.ncfloodmaps.com). The fourth function, habitat, requires development of ecoregion specific statistical models to predict habitat quality based on the likelihood of key species presence and the statistically significant variables required for the models. A larger number of environmental variables are used as input to these models, from which different subsets are identified as statistically significant at the ecoregion level. The selection of significant variables will continue through the pilot development phase (Task 4), to optimize different modeling approaches.

Much of the information we use is derived from the National Hydrography Dataset Plus (NHDPlus), developed with the USGS and EPA. The NHDPlus stream reach and catchment has been chosen as the smallest geographic scale for aggregating land cover and stream data, along with the scale of the NC FRIS stream reaches and catchments. While finer scale information is available and is increasingly being distributed over the web, our ecosystem functional models were selected to operate at or near the NHDPlus resolution, with results aggregated to the HUC12 level. A large number of environmental datasets, analyses, and models are now being distributed and operated at these levels, providing important leverage for the new RBRP approach.

In this report we describe current storage and access to datasets in contractor (UNC and Duke) computer resources. However, due to security restrictions, additional steps will be developed to access information by web services or by data federation using secure information management systems such as iRODS (irods.org). We note in the report where information is currently in restricted storage space. The final storage resources and location for the datasets will need to be determined in Task 4 with DMS.

Contents

Executive Summary 2

I. Overview 4

II. The National Hydrographic Dataset (v2) Data (NHDPlus) 4

III. The 2011 National Land Cover Dataset (NLCD) 7

IV. ESRI Landscape Layers 8

V. US Army Corps of Engineers Stream Temperature Regimes 9

VI. North Carolina Department of Transportation Roads Layer 9

VII. National Inventory of Dams (NID) Reservoir locations 9

VIII. North Carolina Aquatic Species Occurrence Data 10

IX. NOAA Atlas 14, Volume 2 Rainfall Data 10

X. SPARROW Nitrogen and Phosphorus Estimates 10

XI. Regional Regressions for Flood Return Interval Discharge 11

XII. Stream and Floodplain Geomorphology and Hydraulics 12

Tables 15

Figures 18

References 21

Glossary of acronyms 22

# I. Overview

Base data for the new RBRP methodology emphasizes information produced by national and state agencies, which provide institutional continuity to update and make available open source information. Data sets are well described, quality assured, and distributed by various web access methods, obviating the need to repeatedly download, document and store future releases. Secondary information computed using this information is also described in this report, and initial datasets are available statewide for a set of the base data and computed metrics, all georeferenced to common datums and projections. The envisioned RBRP will access information for analysis from individual HUC 8 or Targeted Resource Areas (TRA), and methods of rapidly accessing all required information for a RBRP project will be developed (with the online mapping contractor and DMS) during the pilot Tar-Pamlico and Catawba phases in Task 4.

Unlike a set of current environmental prioritization methods that are recently released (e.g. EPA Recovery Potential Screening tool: http://water.epa.gov/lawsregs/lawsguidance/cwa/tmdl/recovery/index.cfm) which make use of a large number of potential geographic overlays that require extensive weighting, typically at the resolution of the HUC 12, we adapt a data screening protocol to promote a parsimonious set of significant variables that are organized by specific model and statistical analytical needs for four primary ecosystem functions. This allows us to reduce redundancy, ensure significance and sensitivity of included data, and provide methods to evaluate the potential restoration impacts of different environmental management policies and projects. The statistical methods and models we use are designed to operate at relatively fine scales (primarily at the NHDPlus catchment and reach scale), while supporting prioritization and management recommendations at the HUC 12 to HUC 8 levels. Some of the data sets are developed for different watershed representations, including the NC Floodplain Mapping Program data which is developed from much finer LiDAR terrain data and presented at the scale of FEMA stream reaches, cross sections and drainage areas, but is somewhat less extensive in terms of network completeness compared to the NHDPlus stream and catchment network scale. Additional data sets that are used as “overlays” are defined at the HUC 12 level, and are generally incorporated in the absence of models or statistical analytical methods available at the NHDPlus (or similar) level, as described in the RBRP Task 2 report, and repeated in the Appendix of this report.

While the datasets described here are assembled and available on contractor (Duke/UNC) computer systems, access and methods of distributing information for DMS and potentially public use need to be developed as part of the pilot projects (Task 4) in conjunction with the online mapping contractor and DMS personnel. Storage and access described here are interim solutions in preparation for DMS determination of final server and access protocols. At this stage, information downloaded and processed by both UNC and Duke are housed on local resources, and will need to be made available to DMS and other stakeholders by web access or data federation (working with the online mapping contractor).

# II. The National Hydrographic Dataset (v2) Data (NHDPlus)

Overview

The NHDPlus is a suite of geospatial products derived from static snapshots of the National Hydrography Dataset (NHD) stream network (1:100,000-scale), Watershed Boundary Dataset (WBD) hydrologic units (12-digit), and National Elevation Dataset (NED) topography (30m) through a sophisticated data integration process. Additional attribute information was drawn from the National Land Cover Dataset (NLCD), NOAA climate data, and other sources, and aggregated at the NHDPlus reach or catchment levels. Where information was not already available in aggregated form at the reach or catchment level, raster information was processed to this level within this task (see below). Within the NHDPlus, the smallest catchment area is the area draining into each reach. Each reach is an unbroken length of stream between a source and junction or between two junctions, as derived from 1:100,000 hydrographic data. There is a one-to-one mapping between reaches and catchments. NHDPlus catchments typically range up to a few square kilometers in area, while the corresponding stream reach may reach up to a few kilometers in length. The NHDPlus, WBD and NED are all maintained through USGS National Geospatial Program (NGP) led stewardship programs involving states and federal agencies. The derived components of NHDPlus are updated using the latest versions of the National Hydrography Dataset (NHD), the Watershed Boundary Dataset (WBD) and the National Elevation Dataset (NED) on a periodic basis, as determined by programmatic priorities and resources. For example, the production of NHDPlus Version 2 was largely driven by USGS Water Program requirements for improved regional water quality modeling capabilities. USGS Division of Water Resources and the EPA Office of Water worked together to update the NHD in preparation for NHDPlus V2 and also shared the cost for producing it.   
(Source: [http://www.horizon-systems.com/NHDPlus/NHDPlusV2\_home.php](http://www.horizon-systems.com/nhdplus/NHDPlusV2_home.php))

The *DMS River Basin Restoration Prioritization Project* largely relies on the NHDPlus V2 to provide both the spatial units of prioritization (NHDPlus catchments, eventually up-scaled to HUC 12 watersheds) as well as numerous in-stream and up-slope attributes associated with these catchments. The project also uses a number of raster datasets included with NHDPlus to derive additional instream, riparian, upslope, and catchment wide attributes. Table 1 lists all the datasets obtained from NHDPlus used in the project.

The following section outlines the procedures used to obtain and prepare NHDPlus data for use in developing the models used to determine uplift potential.

1. Obtaining the data

Horizon Systems (<http://www.horizon-systems.com/NHDPlus/index.php>) hosts all NHDPlus data and related documentation on its servers and allows data to be downloaded free of charge. Most data are bulk-downloaded for regional sections corresponding to major drainage areas; see <http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php>. North Carolina touches three of these regions: The South Atlantic North (03N), the Tennessee (06), and a small region of the Ohio (05).

1. Uncompressing the files and the NHDPlus directory format

We downloaded all the raster, vector, and tabular datasets listed in Table 1 for each of the three regions intersecting North Carolina. Some datasets for regions 03N and 05 were additionally parsed into sub-regional divisions (e.g. 03a and 03b). All subdivisions for each region were downloaded.

The datasets, which are obtained as compressed zip files, were decompressed into the native NHDPlus directory format (Figure 1). At the end of this step, all required vector, raster, and tabular NHDPlus data can be accessed on a local desktop.

1. Importing data to the Nicholas School GIS Server

To facilitate sharing data across multiple machines (and avoid duplication among team workers), we uploaded the NHDPlus datasets to the Nicholas School’s GIS Server (“NS-GIS.WIN.DUKE.EDU”) - a Windows 2008 Server running MS SQL Server 2008 (R2) integrated with ESRI’s ArcGIS Server (v 10.2). To do this, we created an ArcGIS Enterprise GeoDatabase on the server (“NHDPlusV2”) using Geodatabase administrative tools in ArcGIS desktop. We expect the ArcSDE geodatabases will be stored in major watershed units (e.g. HUC 4 – HUC 6) to limit size to well less than 10 GB.

All but a few (exceptions listed below) of the NHDPlus datasets for a single region (region 03N) were imported directly into the server geodatabase using ArcMap’s *Feature Class to Geodatabase*, *Raster to Geodatabase*, and *Table to Geodatabase* tools. Once region 03N datasets were imported into the server geodatabase, corresponding datasets from the regions 05 and 06 were merged directly into the existing server datasets using the ArcGIS *Mosaic* and *Append* tools for raster and vector/tabular datasets, respectively.

**Projections**

All spatial datasets maintained their original coordinate systems. Raster datasets are referenced to the Albers Equal Area (NAD 83) projection. Vector datasets are referenced to the NAD 1983 geographic coordinate system.

**Metadata**

Federal Geospatial Data Consortium (FGDC) Metadata from the original datasets downloaded from the Horizon Systems servers were copied over to the server datasets using ESRI’s Import Metadata tool.

Special case #1: NHDFlowlines

The NHDFlowlines feature class was handled slightly differently to allow a network dataset to be constructed from it. As ArcMap network datasets can only be created within feature datasets, we first manually created a feature dataset (“NHDFlowlines”) in the server geodatabase, using the coordinate system and XY tolerance information of the NHDFlowlines feature class as a template. We then imported the NHD Flowline dataset for NHD region 03N into this feature dataset and then appended the Flowline datasets from regions 05 and 06 to it (using ArcCatalog’s Load Dataset function)

Special case #2: Mean annual and **monthly** precipitation, temperature, and runoff tables

The NHDPlus provides the tabular data for precipitation, temperature, and runoff as both mean annual averages and as individual monthly averages. Prior to uploading these datasets into the server geodatabase, we merged the annual and monthly means into a single table for each parameter. We did this in two steps, with a Python script written to execute each one.

The first Python script (“[NHD\_MergeVPUAttributes.py](https://github.com/Duke-NSOE/EEP_DataDevelopment/blob/master/Scripts/NHD_MergeVPUAttributes.py)”) merged the individual parameter tables for the different NHDPlus regions into a single table covering all regions. For example, the mean monthly runoff records for the month of January (ROMM01001.txt) for catchments in regions 05 and 06 were each appended to those records in region 03N. The appended tables were stored in a local file geodatabase as inputs for the second step. A list of the tables resulting from this step is shown in Table 2.

The second Python script (“[NHD\_JoinVPUAttributes.py](https://github.com/Duke-NSOE/EEP_DataDevelopment/blob/master/Scripts/NHD_JoinVPUAttributes.py)”) joined the 12 monthly mean values to the annual mean tables for cumulative total precipitation, cumulative total temperature, incremental precipitation, incremental temperature, and runoff, respectively. Records were joined using the *FeatureID* attribute, and the value fields were renamed to include the month of record (e.g. “PRECIPVC\_01”). The result here is a single table for each parameter (precipitation, temperature, and runoff) that includes both annual and monthly mean values.

We wrote an additional Python script to calculate the minimum and the monthly mean values for each catchment. This script is titled “<https://github.com/Duke-NSOE/EEP_DataDevelopment/blob/master/Scripts/NHD_CalculateVPUSummaries.py>”.

Special case #3: Incremental and Cumulative NLCD tables

The NHDPlus also provides incremental and upstream cumulative area of NLCD land cover classes for each catchment via its VPU Attribute Extension (<http://www.horizon-systems.com/NHDPlus/V2NLCD2011.php>). As in the case with the precipitation, temperature, and runoff tables, we merged NLCD records for each region into a single table of all regions. We used the same Python script as above (“[NHD\_MergeVPUAttributes.py](https://github.com/Duke-NSOE/EEP_DataDevelopment/blob/master/Scripts/NHD_MergeVPUAttributes.py)”) to automate this and then uploaded the resulting tables into the ArcGIS Server geodatabase.

Additionally, we also combined several land cover attributes to produce more generalized land cover classes for analysis. Table 3 lists the regroupings of the original NLCD land cover classes. The Python script “NHD\_CalculateNLCDSummaries.py” created these merged NLCD class tables for incremental and cumulative land cover data, respectively.

1. Creating the NHD Flowlines network dataset

To enable network analyses on stream courses (e.g. finding upstream or downstream distances to dam locations), we constructed a network dataset from the NHDPlus flowlines. We constructed the network dataset (“NHDFlowlines\_ND”) in the ArcGIS 10.1 format, using the Z-coordinate values from the NHDFlowlines feature class to define elevations and maintaining the default length cost attribute, defined in meters. We also added two other network attributes: one that can restrict flow in the downstream direction and another that can restrict flow in the upstream direction. These are called FlowDirection and Upstream, respectively.

1. Setting permissions/access to the NHD datasets

We set up an administrative “database owner” (DBO) account to upload datasets to the ArcGIS geodatabase. We created a separate read-only account (“EEPreader”) for consuming the data. The datasets are accessed by creating an ArcGIS Database Connection to NS-GIS2.WIN.DUKE.EDU, supplying EEPreader and the password, and setting the database to NHD. *For security purposes, however, the NS-GIS2 server can only be accessed within the Duke University firewall.* The databases developed within the project will be transferred to NCDENR DMS servers by the end of this project.

The complete list of NHDPlus data imported into the ArcGIS geodatabase is listed in Figure 1.

# III. The 2011 National Land Cover Dataset (NLCD)

Overview

The Multi-Resolution Land Characteristic Consortium (MRLC) created and hosts the 2011 National Land Cover Database (NLCD 2011) on its servers. The NLCD comprises wall-to-wall land use/land cover, tree canopy cover, and percent developed impervious surface at a 30 m cell resolution for the United States. Additional information on the NLCD 2011 is found on the MLRC website: <http://www.mrlc.gov/nlcd2011.php>.

We use the NLCD to derive a number of environment layers for habitat evaluation: flowline land cover, percent stream shading (from riparian canopy cover), riparian zone composition, catchment percent canopy cover, and catchment percent impervious. The NLCD percent developed impervious layer is also used to calculate a metric for total accumulated upslope impervious area for use in the hydrology evaluation. Additional information provided by or derived from the NLCD, NHDPlus and other federal and state datasets are described in figure 2 and Appendix I.

1. Obtaining the data

We obtained raster datasets for land cover, impervious surface, and tree canopy cover from the MLRC download site: <http://www.mrlc.gov/nlcd11_data.php>. The respective datasets include the entire contiguous US and arrive as zipped Imagine IGE format files. The editions of the NLCD data used are listed in Table 4.

1. Uncompressing and uploading the datasets to the Nicholas School GIS Server

After downloading to a local drive, we unzipped each of the three NLCD data files and imported them to an ArcGIS enterprise geodatabase created on the Nicholas School server using the *Raster to Geodatabase (multiple)* tool.

Projections

The three NLCD datasets maintained their original coordinate system: the Albers Conical Equal Area (NAD 83) projection.

Metadata

FGDC Metadata from the original datasets downloaded from the MRLC servers were copied over to the Nicholas School geodatabase server datasets using ESRI’s *Import Metadata* tool.

1. Setting permissions/access to the NLCD datasets

We set up an administrative “database owner” (DBO) account to upload datasets to the ArcGIS geodatabase. We created a separate read-only account (“EEPreader”) for consuming the data. The datasets are accessed by creating an ArcGIS Database Connection to NS-GIS2.WIN.DUKE.EDU, supplying EEPreader and the password, and setting the database to NLCD. *For security purposes, however, the NS-GIS2 server can only be accessed within the Duke University firewall.*

The complete list of NLCD data imported into the ArcGIS geodatabase is listed in Table 7.

# IV. ESRI Landscape Layers

Overview

Environmental Systems Research Institute (ESRI) provides access to a growing number of nationwide raster and vector datasets pertaining to the physical structure of the land – what ESRI terms Landscape Layers (Source: <http://blogs.esri.com/esri/arcgis/2013/09/11/welcome-to-esri-landscape-layers/>). We calculate catchment summary statistics on a number of these layers to evaluate their relative importance in predicting habitat.

1. Accessing the ESRI Landscape Layers

An ESRI ArcGIS Online Organization Account is required to access these datasets, and unlike the NHDPlus and NLCD datasets, these Landscape Layers cannot be downloaded in bulk; instead, the data are accessed directly through links from ArcGIS Desktop to ESRI’s ArcGIS Online server. Once this connection is made, however, spatial subsets of the data can be used in geoprocessing or downloaded locally. These subsets are limited to 24,000 x 24,000 pixels for raster data and up to 1000 features for vector data. The raster access limit is usually not an issue at the HUC 6 level, envisioned as the level of partitioning the full state into datasets within the capacity of DENR DMS servers.

The landscape layers reside on five different ArcGIS servers. We have created connection files to these servers, each stored within the *Data/ESRILayers* folder in the DMS workspace. Our analysis of these layers did not require creating local copies of data subsections. Instead, we performed analyses (zonal statistics) directly on geospatial layers created in Python scripts that access the copies of the data hosted on ESRI cloud. The Python script “[EEP\_ExtractESRILandscapeData.py](https://github.com/Duke-NSOE/EEP/blob/master/Scripts/EEP_ExtractESRILandscapeData.py)” includes code on how the ArcMap *MakeImageServerLayer* process can connect to an ESRI Landscape Layer (via the ArcGIS Server connections mentioned above) and how subsequent geoprocessing tools can work on these layers.

Projection

ESRI Landscape layers are referenced to the North America Albers Equal Area Conic projection.

# V. US Army Corps of Engineers Stream Temperature Regimes

DMS personnel provided us with a dataset on in-stream thermal habitat for all North Carolina streams. We uploaded the original file, in ESRI shapefile format, to a server geodatabase (“NC”) on the Nicholas School server as “ThermalStreams” and granted read-only access to the “EEPreader” account. The dataset can be accessed via the NC.sde link found in the Data folder of the DMS workspace. (Access is available only to machines within the Nicholas School firewall.)

Projection

The Stream Temperature Regimes layer is referenced to the North Carolina State Plane (feet) coordinate system.

# VI. North Carolina Department of Transportation Roads Layer

The NC Department of Transportation (NC DOT) provides a statewide dataset of primary and secondary road arcs as part of its Linear Referencing System. We downloaded this dataset (in Shapefile format) from the NC DOT GIS resources page (<https://connect.ncdot.gov/resources/gis/pages/gis-data-layers.aspx>) to a local drive. The dataset was then imported to the “NC” server geodatabase on Nicholas School GIS server as NCDOT\_Roads. The dataset can be accessed via the NC.sde link found in the Data folder of the EEP workspace. (Access is available only to machines within the Nicholas School firewall.)

Projection

The NC DOT Roads Layer is referenced to the North Carolina State Plane (feet) coordinate system.

# VII. National Inventory of Dams (NID) Reservoir locations

We downloaded federal, state, local, utility, private, and unlisted reservoir data from the US Army Corps of Engineers’ (USACE) National Inventory of DAMs (<http://nid.usace.army.mil/cm_apex/f?p=838:12>). From the national datasets, we extracted features from NC, VA, TN, and SC. We cleaned the data by removing features with duplicate NID ID (often levees around the dams or saddle dikes). We also removed features missing latitude/longitude coordinates, as these tend to be ash ponds, fossil fuel complex, and other features that are not part of a hydrologically connected network. Finally, we clipped the dataset to our study area (all HUC 6s intersecting North Carolina). Table 5 lists the number of reservoirs at the various stages of cleaning.

The cleaned reservoirs feature class was uploaded to the “NC” server geodatabase on the Nicholas School GIS server as “NDI\_Reservoirs”. The dataset can be accessed via the NC.sde link found in the Data folder of the EEP workspace. (Access is available only to machines within the Nicholas School firewall.)

Projection

The NID reservoir layer is referenced to the NAD 1983 Geographic Coordinate System.

# VIII. North Carolina Aquatic Species Occurrence Data

Mark Endries of the US Fish and Wildlife Service compiled a comprehensive data of aquatic species observation records in North Carolina (<http://www.fws.gov/asheville/htmls/Maxent/Maxent.html>). He provided us with these locations as multi-point features in a shapefile. We uploaded this shapefile to the “NC” server geodatabase on Nicholas School GIS server as “AquaticSpeciesOccurrences”. The dataset can be accessed via the NC.sde link found in the Data folder of the EEP workspace. (Access is available only to machines within the Nicholas School firewall.)

Projection

The Aquatic Species Occurrence Layer is referenced to the North Carolina State Plane (meter) coordinate system.

# IX. NOAA Atlas 14, Volume 2 Rainfall Data

The NOAA Atlas provides a database of precipitation frequency estimates for the Ohio River Basin and surrounding states, including North Carolina, which is utilized in the USGS Regional Peak Flow Regression equations for Hydroregion 4, the Coastal Plain. We downloaded the 50-year, 24-hour rainfall data (in ASCII format) from the NOAA webpage (<http://hdsc.nws.noaa.gov/hdsc/pfds/pfds_gis.html>). We then converted to an ESRI grid format and extracted rainfall depth for North Carolina. The grid cell precipitation is provided in units of inches multiplied by 1000 (thousandths of an inch) in the original dataset and has been converted to inches.

Projection

The NOAA Atlas precipitation frequency estimates are referenced to the WGS 1972 Geographic Coordinate System.

Metadata

FGDC compliant metadata for the original dataset can be accessed from the NOAA Atlas webpage (<http://hdsc.nws.noaa.gov/hdsc/pfds/meta/na14_vol2_orb_grid_metadata.xml>).

# X. SPARROW Nitrogen and Phosphorus Estimates

Overview

The Spatially Referenced Regressions of Watersheds (SPARROW) model (Smith et al, 1997; Schwarz et al, 2006) used for estimation of Total Nitrogen (TN) and Total Phosphorus (TP) is based on nonlinear regression equations describing TN and TP sources, land to water delivery, and instream and riparian decay rates. This information was developed by a SPARROW model for TN and TP at the NHDPlus scale (Hoos et al., 2013) for southeast Atlantic and Gulf drainages, as well as the Tennessee River basin. The analysis is based on a time-centered estimate of loads derived from USGS stream gauging stations, as the data based method requires a sufficient number of stream flow volume and nutrient concentration observations, which requires a multi-year dataset. The estimates are calibrated to the spatial variation of nutrient loads estimated from the USGS gauge network in the region. Therefore, they provide general nutrient loading conditions over this period, and should not be interpreted as providing current (e.g. 2015) loads. We use these estimates as they interpolate estimates of nutrient loading for every reach in the NHDPlus river network, without requiring detailed, time consuming and more data demanding mechanistic models, which would be difficult to apply and update at the scale of the full state. SPARROW loading estimation methods are well described, are going through a process of updating and improvement over time as new information becomes available, and are posted as web services by the USGS (e.g. <http://cida.usgs.gov/sparrow/>). The statistical model identifies a set of input landscape and management variables that enter the regressions and provide statistically significant explanation (association) with the measured loads. Therefore, interpretation of model coefficients for individual (and significant) variables should be considered in a more abstract and generalized fashion. That is, interpretation of management effects on nutrient loading should be at land to water delivery and in-stream removal rather than on specific types of wetlands or species within riparian buffers. DMS approved management handles (and other methods of management) will be grouped with respect to which aspect of SPARROW components they address (e.g. nutrient loading, land-to-water retention, aquatic retention), rather than individual practices.

1. Obtaining the data

NHDPlus v1 flow line reach estimates of TN and TP were made available by Anne Hoos and Ana Garcia of the USGS, respectively. While both SPARROW datasets were created using the NHDPlus v1 dataset, the spatial output data were joined to the NHDPlus v2 network dataset with > 98% compliance except within the New River Basin, and limited reaches which are false centerlines in water bodies or in reaches classified as ditches or other artificial features (largely in the Outer Coastal Plain). As these models were designed for the Southeast and the New River drains to the Ohio River Basin, the New River Basin was not included in estimates from the N and P models. Note, however, that drainage areas flowing to the Tennessee River in western North Carolina are included in the database. We extracted the total and incremental yield fields from the SPARROW output dataset, representing total upslope TN or TP output and local TN or TP output, respectively. As this data is tabular, it inherits the projection of NHDPlus v2 flowline or catchment dataset to which it is joined.

Projection

Same as NHDPlus v2.

Metadata

See Hoos et al (2013) for details on all input data sources to the model. Here we only use model results.

# XI. Regional Regressions for Flood Return Interval Discharge

Overview and methods

The USGS has developed regional regression equations to estimate peak flood flow for North Carolina. These equations have been developed for the four hydrologic regions that cover the state (Piedmont-Ridge and Valley, Blue Ridge, Sand Hills, and Coastal Plain) for a variety of flood frequency periods. We extracted the equations from the USGS SIR 2014-5030, “Methods for Estimating the Magnitude and Frequency of Floods for urban and Small, Rural Streams in Georgia, South Carolina, and North Carolina, 2011” (<http://pubs.usgs.gov/sir/2014/5030/>), for all hydrologic regions but the Blue Ridge. The equations for the Blue Ridge hydroregion were extracted from USGS Fact Sheet 007-00, “The National Flood-Frequency Program—Methods for Estimating Flood Magnitude and Frequency in Rural and Urban Areas in North Carolina, 2001” (<http://pubs.usgs.gov/fs/fs-007-00/>), as they were not provided in the more recent USGS publication. We used these equations to estimate instantaneous peak flood flows at 2-, 10-, 50-, and 100-year recurrence intervals with input datasets coming from NHDPlus v2, NLCD, and the NOAA Atlas. The equations can be found in Table 6. We note that this equation set does not match what is currently available through the USGS Streamstats online site, as the national Streamstats program has not yet updated the previous equation set due to resource limitation. Therefore, the equation set we are using is more up-to-date and can be modified if and when this version is updated.

The hydrologic regions used in the regional regression equations were adapted from the USEPA ecoregions and developed by the USGS for use in these equations. The GIS data for these hydrologic regions is associated with the USGS report USGS SIR 2014-5030 and was downloaded from (<http://pubs.usgs.gov/sir/2009/5043/downloads/Hydrologic_Regions_GIS_Coverage.zip>) for use in our analysis.

1. Calculation of peak flows

Using the regional regression equations, along with calculated layers representing upslope area percent impervious and developed cover and the rainfall data from NOAA, we calculated the 2-, 10-, 50-, and 100-year flood peak flow levels using the ArcGIS Calculate Field function. This information is derived for each NHDPlus flow reach within North Carolina, and then joined to the tables of the NHDPlus v2 catchments and flowline attributes described above.

Projection

Same as NHDPlus v2

# XII. Stream and Floodplain Geomorphology and Hydraulics

In response to catastrophic flooding from Hurricane Floyd in 1999 and other major flood, surge and inundation events, the State of North Carolina embarked on an ambitious program of modeling and remapping flood frequency and hazard zones for the entire state. The current NC Flood Risk Information System (FRIS) is widely considered the most advanced state flood hazard system in the country. The North Carolina Flood Mapping Program (NCFMP) has developed detailed hydraulic models [Hydrologic Engineering Center-River Analysis System (HEC-RAS)] to characterize the flood extents for different flow return periods (i.e., 25, 50, 100, 200, 500 Years). These models are developed with highly detailed river geometry information with several variables as inputs estimated and measured using a variety of methods. Notably, the cross-section geometries are surveyed in the field and station-elevation pairs are manually entered, and in some regions cross-sections are supplemented by creating them with high resolution LiDAR data that is available for the entire state. The river cross-section information includes surveyed measurements of man-made structures in places where flows are artificially altered. This data and model output provides the advanced ability to map geomorphic and hydraulic information pertinent to restoration prioritization.

We note that the FRIS stream network, while providing higher resolution information on stream channel geomorphology than is available through NHDPlus, does not include all reaches of streams that are considered to be non-hazardous, and therefore is somewhat less spatially extensive than the NHDPlus. However, the resolution of the stream and floodplain network, and the state LiDAR dataset provide sufficient information to extract find grained current conditions at scales that can be aggregated to the HUC 12. In addition, limited HEC RAS modeling to explore sensitivity of flood extent to changes in flow levels due to impervious conditions (and by extension other causes of return period discharge levels) provides process based estimates of potential flood exacerbation or abatement by a set of management activities.

The state floodplain mapping and modeling proceeded from east to west, and there is some inconsistency between studies and datasets produced along this gradient, and between different contractors. In particular, georeferencing of HEC-RAS models was not directly incorporated in the eastern basins, and we have written additional Python code to map model output to the stream reach and cross sectional information in the geodatabases distributed by the State. This work is proceeding, but we envision that channel and floodplain geomorphic and hydraulic parameters will be extracted from the online state databases at by the python scripts we are building at the HUC 8 or TRA project level. As part of the Task 4 pilot studies of the Tar-Pamlico and Catawba River basins we are experimenting with methods of extracting required information directly from the FRIS web site, or in developing a full state database for delivery to DMS. We note that personnel from the NCFMP have been very cooperative in sharing information and in discussing methods to extract information from their extensive data holdings.

Detailed stream centerline and cross section information are incorporated into HEC-RAS model output, along with a set of useful flow and flood levels and extents for different return periods. This information will produce metrics (to be experimented with and discussed with DMS personnel) on the extent, depth and frequency of flooding, and on stream and floodplain cross-section morphology. The latter are frequently the target of in-stream restoration to alter channel morphology, such as reconnecting the channel to the floodplain (if sensitive structures are not present). We note that these types of activities have the potential to benefit local conditions, but would also benefit downstream areas in terms of reduction of flood peaks.

Python workflows have been developed to georeference HEC-RAS models (note that central and western basins are already mostly georeferenced in the FRIS database), open HEC-RAS models using the HEC RAS Controller (Goodall, 2014) and estimate a set of stream channel and floodplain form and hydraulic metrics.

Example case: Entrenchment Ratio

Entrenchment ratio is the measure of the vertical containment of a river. It suggests whether the river is depth dominated or width dominated. By our definition, entrenchment ratio is the ratio of the top width (width of inundated area) at a given recurrence interval and the channel width at top of bank. Higher values indicate less vertical containment of the river and more flood inundation extent. The entrenchment ratio at one given cross-section however may not be representative of overall reach condition. Therefore, we take all cross-sections on a given reach, weight them by the distance between cross sections (Equation 1). Our python workflow for computing and mapping entrenchment ratios is then merged with the geospatial data for visualization and assessment (Figure 3). Note that this definition differs, and is more precise and specific to a flood return interval, than commonly used methods of assuming a flood height and consequent topwidth at some multiple of the bankfull channel depth.

(1)

This methodology can be extended to a large number of other channel geomorphology and hydraulic variables, including channel width and depth, topwidth (inundated flood extent) and depth, discharge and velocities within channels and on floodplains, and form metrics including entrenchment and incision ratios (bankfull width divided by bankfull depth), for different return intervals. Final choices of metrics will be carried out in Task 4 with DMS personnel. This information will be produced at both the FRIS stream reach resolution, and also aggregated to provide summary information at the HUC 12 or TRA levels.

Projection

The NC FRIS data is referenced to the North Carolina State Plane (feet) coordinate system.

Metadata

See (ncfloodmap.com) for details on HEC-RAS analysis and methodology. Here we use model results and geodatabases developed by NCFMP.

# Tables

**Table 1**. NHDPlus (v2) Datasets obtained from Horizon Systems for NC Regions (03N, 05, & 06).

|  |  |  |
| --- | --- | --- |
| Format | Name | Description |
| Raster | elev\_cm | Elevation, in cm |
| Raster | cat | NHDPlus catchments |
| Raster | fac | Flow accumulation |
| Raster | fdr | Flow direction |
| Raster | fdrnull | Flow direction; streams set to null values |
| Raster | shdrelief | Shaded relief |
| Vector | NHDflowline | NHDPlus flowline features |
| Vector | BoundaryUnit | NHDPlus regional boundary units |
| Vector | CatchmentFeatures | NHDPlus catchment features |
| Vector | WBD\_Subwatershed | National watershed boundary database (HUC12 features) |
| Vector | WBDHU2 | 2-digit hydrologic unit polygons |
| Vector | WBDHU4 | 4-digit hydrologic unit polygons |
| Vector | WBDHU6 | 6-digit hydrologic unit polygons |
| Vector | WBDHU8 | 8-digit hydrologic unit polygons |
| Table (dbf) | CumulativeArea | Cumulative area upstream of an NHDFlowline feature |
| Table (dbf) | ElevSlope | Elevation and slope derived for NHDFlowline features |
| Table (dbf) | PlusFlowlineVAA | NHDPlus “Value Added Attributes” for NHDFlowline features |
| Table (txt) | CumTotNLCD2011 | Area of each NLCD class found upstream of a given catchment |
| Table (txt) | IncrTotNLCD2011 | Area of each NLCD class found within a given catchment |
| Table (txt) | CumTotPrecipMA | Mean annual rainfall across the area upstream of a given catchment |
| Table (txt) | CumTotTempMA | Mean annual temperature across the area upstream of a catchment |
| Table (txt) | IncrPrecipMA | Mean annual rainfall over the of a given NHDPlus catchment |
| Table (txt) | IncrPrecip*mm* (x12) | Mean monthly rainfall over the area of a given NHDPlus catchment |
| Table (txt) | IncrTempMA | Mean annual temperature over the of a given NHDPlus catchment |
| Table (txt) | IncrTemp*mm* (x12) | Mean monthly temperature over the of a given NHDPlus catchment |
| Table (dbf) | EROM\_MA0001 | Extended unit runoff method (EROM) annual flow estimates |
| Table (dbf) | EROM\_*mm*0001 | Extended unit runoff method (EROM) monthly flow estimates |
| Table (txt) | ROMA0001 | Mean annual runoff recorded within a given NHDPlus catchment |
| Table (txt) | ROMA*mm*0001 (x12) | Mean monthly runoff recorded within a given NHDPlus catchment |

**Table 2.** List of appended tables of annual and monthly means for precipitation, temperature, and runoff.

|  |  |
| --- | --- |
| IncrPrecipMA | Mean annual precipitation |
| IncrPrecip*mm* (x12) | Mean monthly precipitation (mm = month) |
| IncrTempMA | Mean annual temperature |
| IncrTemp*mm* (x12) | Mean monthly temperature (mm = month) |
| ROMA0001 | Mean annual runoff |
| ROMA010001 (x12) | Mean monthly runoff (mm = month) |

**Table 3.** Reclassification of NLCD classes into more generalized land cover classes

|  |  |
| --- | --- |
| Original class | Generalized class |
| 11 – Open water | 1 – Open Water |
| 12 – Perennial Ice/Snow | *- Omitted -* |
| 21 – Developed, Open Space | 2 – Developed |
| 22 – Developed, Low intensity |  |
| 23 – Developed, Medium Intensity |  |
| 24 – Developed, High intensity |  |
| 31 – Barren | 3 – Barren |
| 41 – Deciduous forest | 4 – Forest |
| 42 – Evergreen forest |  |
| 43 – Mixed forest |  |
| 51 – Dwarf shrub | *- Omitted -* |
| 52 – Scrub shrub | 50 – Shrub |
| 71 – Grassland/herbaceous | 71 – Grassland/herbaceous |
| 72 – Sedge/herbaceous | *- Omitted -* |
| 73 – Lichens | *- Omitted -* |
| 74 – Moss | *- Omitted -* |
| 81 – Pasture/hay | 8 – Pasture/hay |
| 82 – Cultivated crops | 8 – Cultivated crops |
| 90 – Woody wetlands | 9 – Woody wetlands |
| 91 – Emergent herbaceous wetlands | 9 – Emergent herbaceous wetlands |

**Table 4.** Editions of NLCD (2011) layers downloaded and stored

|  |  |
| --- | --- |
| Land Cover | 2014\_10\_10 |
| Percent Developed Imperviousness | 2014\_03\_31 |
| Tree Canopy Analytical | 2014\_03\_31 |

**Table 5.** Number of National Inventory of Dams reservoir locations by state.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Phase** | **NC** | **SC** | **TN** | **VA** | **Total** |
| Federal | Filter by state | 63 | 41 | 96 | 68 | 268 |
|  | Remove Duplicates | 56 | 40 | 54 | 64 | 214 |
|  | Clip to study area | 56 | 40 | 24 | 19 | **139** |
| State | Filter by state | 61 | 78 | 155 | 74 | 368 |
|  | Remove Duplicates | 61 | 77 | 148 | 74 | 360 |
|  | Clip to study area | 61 | 77 | 5 | 18 | **161** |
| Local | Filter by state | 286 | 182 | 177 | 271 | 916 |
|  | Remove Duplicates | 286 | 179 | 174 | 265 | 904 |
|  | Clip to study area | 286 | 179 | 15 | 73 | **553** |
| Utility | Filter by state | 101 | 46 | 9 | 52 | 208 |
|  | Remove Duplicates | 86 | 25 | 9 | 50 | 170 |
|  | Clip to study area | 86 | 24 | 2 | 17 | **129** |
| Private | Filter by state | 2610 | 2089 | 793 | 1184 | 6676 |
|  | Remove Duplicates | 2603 | 2062 | 789 | 1166 | 6621 |
|  | Clip to study area | 2603 | 2058 | 181 | 302 | **5144** |
| Not Listed | Filter by state | 141 | 0 | 0 | 413 | 554 |
|  | Remove Duplicates | 141 | 0 | 0 | 402 | 543 |
|  | Clip to study area | 141 | 0 | 0 | 46 | **187** |

**Table 6.** Regional Regression equations for peak flood flows.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Flood Recurrence Interval | Hydrologic Region | | | |
| 1 - Piedmont/Valley and Ridge | | 2 - Blue Ridge | |
| 0.1 sq.mi. ≤ DA ≤ 3 sq.mi. | 3 sq.mi. ≤ DA ≤ 436 sq.mi. | ImpArea ≤ 10% (rural) | 10% < ImpArea  (urban) |
| 2 year | 163(DA)0.7089 \* 10(0.0133\*ImpArea) | 198(DA)0.5735 \* 10(0.0101\*ImpArea) | 135(DA)0.702 | 33.3(DA)0.739 \* (ImpArea)0.686 |
| 10 year | 381(DA)0.7536 \* 10(0.0076\*ImpArea) | 484(DA)0.5539 \* 10(0.0060\*ImpArea) | 334(DA)0.662 | 122(DA)0.655 \* (ImpArea)0.515 |
| 50 year | 632(DA)0.7903 \* 10(0.0037\*ImpArea) | 794(DA)0.5428 \* 10(0.0037\*ImpArea) | 602(DA)0.635 | 296(DA)0.602 \* (ImpArea)0.396 |
| 100 year | 753(DA)0.8038 \* 10(0.0024\*ImpArea) | 941(DA)0.5386 \* 10(0.0028\*ImpArea) | 745(DA)0.625 | 374(DA)0.593 \* (ImpArea)0.358 |

|  |  |  |
| --- | --- | --- |
| Flood Recurrence Interval | Hydrologic Region | |
| 3 - Sand Hills | 4 - Coastal Plain |
| 0.22 sq.mi. ≤ DA ≤ 459 sq.mi. | 0.1 sq.mi. ≤ DA ≤ 53.5 sq.mi. |
| 2 year | 30(DA)0.6605 \* 10(0.0122\*DevArea) | 26.3(DA)0.5908 \* 10(0.0173\*ImpArea) \* 10(0.0515\*MRF) |
| 10 year | 68.4(DA)0.6507 \* 10(0.0102\*DevArea) | 51.8(DA)0.6004 \* 10(0.0101\*ImpArea) \* 10(0.0666\*MRF) |
| 50 year | 114(DA)0.6451 \* 10(0.0090\*DevArea) | 78.4(DA)0.6111 \* 10(0.0058\*ImpArea) \* 10(0.0738\*MRF) |
| 100 year | 138(DA)0.6430 \* 10(0.0086\*DevArea) | 90.5(DA)0.6154 \* 10(0.0043\*ImpArea) \* 10(0.0762\*MRF) |

DA = Drainage Area (sq.mi.), ImpArea = Percent upslope impervious area1, DevArea = Percent upslope developed area (NLCD classes 21-24), MRF = 50 yr, 24 hr max rainfall (inches)

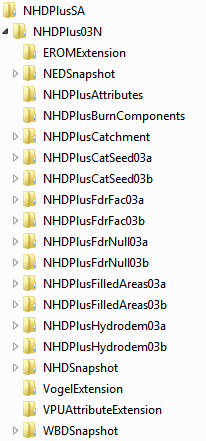
1 The cumulative percent upslope impervious area was derived from the NLCD impervious fraction layer, accumulated along the NHDPlus fdir (flow direction) dataset (described above).

**Table 7**. ESRI Landscape Layers imported into the spatial database

|  |  |
| --- | --- |
| **ESRI Source Layer** | **Description** |
| [USA Soils Flooding Freq.](http://www.arcgis.com/home/item.html?id=2bbf8c1f28ae46d5b807592ccfe970b0) | Sum of flooding frequency values within a catchment |
| [Terrain: Slope Map](http://www.arcgis.com/home/item.html?id=a1ba14d09df14f42ad6ca3c4bcebf3b4) | Mean slope (degrees) |
| [USA Road Density](http://www.arcgis.com/home/item.html?id=64a95b092457466388f09136e331ff09) | Sum of road length w/in 1km cells, averaged across all cells w/in a catchment |
| [Water Table depth](http://www.arcgis.com/home/item.html?id=8739f213277943e390aa2111b95ab72a) | Mean depth to water table (cm) |
| [Soils Erodibility Factor](http://www.arcgis.com/home/item.html?id=93be1788338d492e8d9079abb65d5722) | Mean SSURGO Erodibility (K-factor) value |
| [USA Food Risk](http://www.arcgis.com/home/item.html?id=6b09b1c163c740559dc31cce9144222e) | Proportion of catchment area in flood risk zone |

# Figures

**Figure 1.** Screen shot for directory format created when NHDPlus downloaded zip files are decompressed. Each folder corresponds to different components of the NHDPlus dataset.



**Figure 2.** Screen shotof datasets included in the NHD server geodatabase located on NS-GIS2.WIN.DUKE.EDU. Note that the operational RBRP data set will be developed with online mapping contractors to allow access by DMS personnel and other stakeholders.

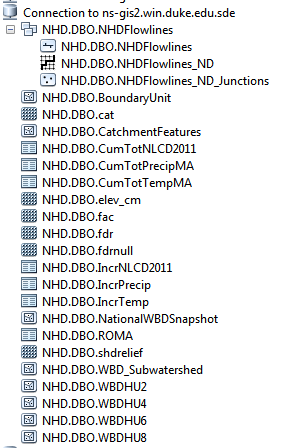
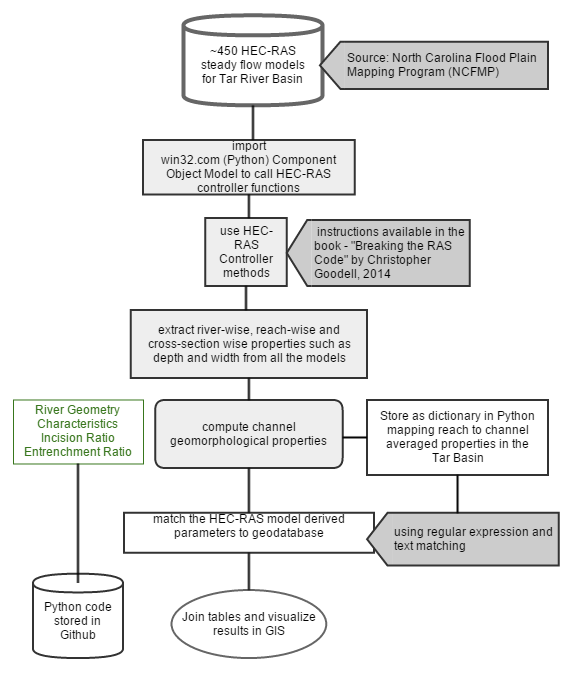


Figure 3. Geomorphology workflow to calculate entrenchment ratio



# References

Feaster, T.D., Gotvald, A.J., and Weaver, J.C., 2014, Methods for estimating the magnitude and frequency of floods for urban and small, rural streams in Georgia, South Carolina, and North Carolina, 2011: Scientific Investigations Report 2014*–* 5030, 104 p., *http://dx.doi.org/10.3133/sir20145030*

Goodall, Christopher, 2014, Breaking the HEC-RAS Code: A User’s Guide to Automating HEC-RAS. h2ls publishing. 278pg.

Hoos, A.B., Moore, R.B., Garcia, A.M., Noe, G.B., Terziotti, S.E., Johnston, C.M., and Dennis, R.L., 2013, Simulating stream transport of nutrients in the eastern United States, 2002, using a spatially-referenced regression model and 1:100,000-scale hydrography: U.S. Geological Survey Scientific Investigations Report 2013-5102, 33 p., http://pubs.usgs.gov/sir/2013/5102/.

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

Smith RA, Schwarz GE, Alexander RB. Regional Interpretation of Water-Quality Monitoring Data. Water Resources Research. 1997;33:2781–2798.

United States Geological Survey, 2002. “The National Flood-Frequency Program—Methods for Estimating Flood Magnitude and Frequency in Rural and Urban Areas in North Carolina, 2001” (<http://pubs.usgs.gov/fs/fs-007-00/>),

# Glossary of acronyms

DMS, Division of Mitigation Services

ESRI, Environmental Systems Resource Institute

FEMA, Federal Emergency Management Agency

FRIS, Flood Resources Information System

HEC-RAS, Hydrologic Engineering Center (US Army Corp Engineers)-River Analysis System

HUC, Hydrologic Unit Code

iRODS, Integrated Rule Oriented Data System

LiDAR, Light Detection and Ranging

NAD, North American Datum

NCDOT, North Carolina Department of Transportation

NCFMP, North Carolina Flood Mapping Program

NED, National Elevation Dataset

NHD, National Hydrography Dataset

NHDPlus, National Hydrography Dataset Plus (additional reach and catchment attributes)

NID, National Inventory of Dams

NLCD, National Land Cover Dataset

RBRP, River Basin Restoration Prioritization

SPARROW, Spatially Referenced Regressions of Watersheds

TN, Total Nitrogen

TP, Total Phosphorus

TRA, Targeted Resource Area

UNC, University of North Carolina (Chapel Hill)

WBD, Watershed Boundary Dataset