National Stream Internet Protocol and User Guide

David Nagel[1], Erin Peterson[2], Daniel Isaak[1], Jay Ver Hoef[3], and Dona Horan[1]

U.S. Forest Service, Rocky Mountain Research Station
Air, Water, and Aquatic Environments Program
322 E. Front St., Boise, ID

United States Department of Agriculture - U.S. Forest Service
Rocky Mountain Research Station
Air, Water, & Aquatic Environments Program

Author affiliations: [1] US Forest Service, Rocky Mountain Research Station, AWAE Program, 322 E. Front St., Suite 401, Boise, ID 83702. [2] Queensland University of Technology, Brisbane, Queensland, Australia. [3] National Oceanic and Atmospheric Administration, Fairbanks, AK.

Version 3-22-2017

Abstract

The rate at which new information about stream resources is being created has accelerated with the recent development of spatial stream-network models (SSNMs), the growing availability of stream databases, and ongoing advances in geospatial science and computational efficiency. To further enhance information development, the National Stream Internet (NSI) project was developed as a means of providing a consistent, flexible analytical infrastructure that can be applied with many types of stream data anywhere in the country. A key part of that infrastructure is the NSI network, a digital GIS layer which has a specific topological structure that was designed to work effectively with SSNMs. The NSI network was derived from the National Hydrography Dataset Plus, Version 2 (NHDPlusV2) following technical procedures that ensure compatibility with SSNMs. This report describes those procedures and additional steps that are required to prepare datasets for use with SSNMs.

1.0 Introduction

1.1 Overview

The USGS National Hydrography Dataset Plus, Version 2 (NHDPlusV2) (McKay et al., 2012) is an attribute rich, GIS stream network developed at 1:100,000 scale by the Environmental Protection Agency (EPA) and U.S. Geological Survey (USGS). It is a nationally consistent, freely available data layer that has proven useful for representing stream characteristics at regional scales. The NHDPlusV2 data have been used extensively for hydrography and watershed applications since its initial development.

To characterize stream resources, physical, chemical, and biological observations are frequently sampled at discreet locations within a stream or river network. However, extrapolating these data to yield semi-continuous predictions throughout the network has proved challenging using standard, non-spatial statistical techniques developed for terrestrial systems (Peterson and Ver Hoef, 2010). Two software products have been developed to address these issues and have proven useful for modeling stream temperature (Isaak et al., 2010) and other network attributes at landscape scales: 1) The Spatial Tools for the Analysis of River Systems (STARS) ArcGIS custom toolset (Peterson and Ver Hoef, 2014) and 2) the Spatial Stream Network (SSN) package (Ver Hoef et al., 2014) for R statistical software (R Core Team, 2013). These tools overcome previous statistical deficiencies by incorporating a covariance structure that accounts for spatial autocorrelation along the drainage network. However, these software packages require a stream network with a degree of topological integrity that was previously unavailable at the national scale. This document describes a modified version of the NHDPlusV2 dataset, the National Stream Internet (NSI) network, which has been modified to work directly with the STARS and SSN software packages. The NSI network is inextricably linked to the STARS and SSN software, as the data were developed specifically for use with these tools. This document is intended to assist users in understanding the unique features of the NSI network and discovering how the data may be used with the STARS custom toolset and the SSN package.

1.2 Purpose and Conventions

The purpose of this document is to describe the NSI dataset and provide an overview of its modifications from the NHDPlus data. This guide also seeks to assist users in utilizing SSNMs by providing instructions and references for using the STARS and SSN software products, including a list of software requirements. Although this document will not provide all of the information necessary to use these products, it is intended that this guidance will lead users to the resources necessary to do so.

Regarding conventions used in this document, italic font is used for emphasis of important terms. Field names for ArcGIS attribute tables are generally printed in bold text. Bold text is also used for section headings. Important instructions are noted with bold-italics. User interface headings from the STARS tools are denoted with quotes.

Section headings are indicated by common numerical standards such as 1.0 and 1.1. When a series of actions is outlined within a section, the actions are referred to as *steps* and are indicated by the style convention such as 1), 2), etc.

2.0 Software Requirements

The following software is required to utilize the NSI network with the STARS custom toolset and the SSN package.

ArcGIS

ArcGIS (Esri 2009) version 10.1 or higher with an Advanced level license and the Spatial Analyst extension is required. It is assumed that the user has a moderate level of GIS experience.

STARS ArcToolbox

STARS version 2.0.3 or higher is required. STARS is a custom ArcGIS toolset that must be loaded with ArcGIS. The software, along with sample datasets and a tutorial may be downloaded here: http://www.fs.fed.us/rm/boise/AWAE/projects/SSN_STARS/latest_releases.html

R statistical software

The SSN package was developed for use with R statistical software (R Core Team, 2013). Information about downloading and installing R can be found here: http://www.r-project.org/

SSN software package

SSN is an extension package for R software. The latest software version, reference manual, and vignette may be downloaded here: http://cran.r-project.org/web/packages/SSN/index.html

Python

Python software (Python Software Foundation 2014) version 2.7 is installed with ArcGIS version 10.2 and higher and has been tested for compatibility with the STARS custom toolset. Python version 3.x has not been tested with STARS.

PythonWin

PythonWin is a Python add-on that is required for some STARS components. PythonWin must be downloaded and installed separately from Python. PythonWin build 218, version 2.7 (pywin32-218.win32-py2.7.exe) has been tested for compatibility with STARS. PythonWin may be downloaded from here: http://sourceforge.net/projects/pywin32/files/pywin32/

3.0 Data and Tools Overview

3.1 Nomenclature

The STARS custom toolset and the SSN package enable users to fit statistically valid models of aquatic phenomenon on stream networks and make subsequent predictions of aquatic attributes at unsampled locations. Field-collected data are needed to fit a SSNM and these data are referred to as *observed data*, which are collected at *observation sites* (Figure 1). When a SSNM is fit to an observed variable, we refer to it as the *response variable* (i.e., dependent variable). In most cases, it is also desirable to include *covariates* in the model (i.e., predictors or independent variables), which help explain variability in the response. For example, if the response variable is stream temperature, then canopy cover might be used as a covariate in the SSNM. A fitted SSNM can also be used to generate predictions at user-defined points along the stream network and we refer to these as *prediction points*.

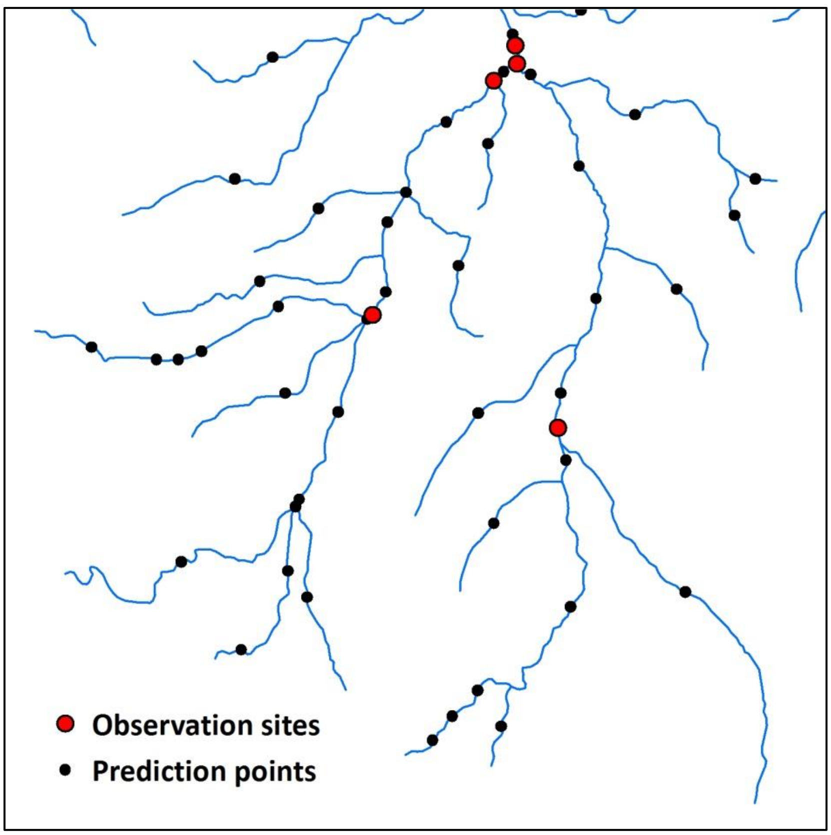


Figure 1. Example observation sites and prediction points. As part of the NSI dataset, prediction points are assigned at the midpoint of each NHDPlusV2 stream segment that contains a unique ComID value.

3.2 Data

3.2.1 NSI Dataset

The *NSI dataset* consists of *stream lines* and *prediction points*. The NHD digital stream lines, or *NHDFlowline* data, were edited to conform to topological constraints required by the STARS and SSN tools, as described in section 4.0 below. The lines alone are referred to as the NSI network. In addition, prediction points have been generated at the midpoint of each NSI stream line that contains a unique ComID value. The data are available for the conterminous U.S. The data may be freely downloaded from here: http://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet.html

3.2.2 Observation Data

Spatially referenced observed data are required to generate the SSNM. Observed data generally include field-collected measurements such as temperature, chemistry, fish presence/absence or counts, or any other variable that can be estimated using one or more covariates. The observation sites are represented as Esri shapefile point locations and must be spatially referenced to the NSI network.

3.2.3 Covariate Data

Covariate data are used to explain variability in the response variable during the modeling process. For example, if stream temperature is the response variable, suitable covariates may include elevation, canopy cover, and drainage area. Covariate data must be available for each observation measurement. If the SSNM is used to make predictions, then covariates must also be available at each prediction point on the network and stored in the attribute table.

The covariates are typically derived from GIS layers that are available throughout the entire stream network. The NHDPlusV2 dataset contains many such variables that may be joined to the NSI dataset.

3.2.4 Prediction Points

Prediction points are user defined locations along the stream network where the response variable can be estimated (Figure 1). If an SSNM is used to make predictions, the model results can be stored at the prediction points. These predictions may then be linked back to the original NSI or NHDPlusV2 stream line data through a common ID field.

Prediction points are distributed with the NSI dataset, but users may also generate their own prediction point features and covariates. Prediction points must be spatially referenced to the NSI network. Prediction points are not required for using the SSNM. Some users may only wish to explore the covariance structure of the response variable, in which case only observation data are necessary.

3.3 Tools

3.3.1 STARS

The purpose of the STARS toolset is to generate and format the data needed to fit SSNMs in R (R Core Team, 2013). The STARS toolset makes use of the Landscape Network (LSN; Peterson and Ver Hoef, 2014), a data structure used to efficiently navigate and calculate indices throughout a stream network. The LSN is an ArcGIS format Personal Geodatabase (Esri 2011) and is discussed further in section 5, step 5 below. Within the STARS toolset, specific programs have been included to 1) pre-process the LSN; 2) calculate the hydrologic distances (with flow-direction preserved), the spatial additive function used to weight converging stream segments, and the covariates for all observed and prediction locations in the stream network; and 3) export the topological, spatial, and attribute information in a format that can be efficiently stored, accessed, and analyzed in R. Note that pre-processing has already been completed for the NSI dataset (function 1 above, this paragraph). Users of the NSI dataset are required to calculate hydrologic distances and export the data for R using STARS (functions 2 and 3 above).

3.3.2 SSN

The SSN package was developed for the R statistical environment (R Core Team, 2013), which is open source statistical computing and graphics software. Once the streams data have been properly formatted using the STARS toolset, the SSN package allows users to: 1) import and store their spatial data in R, 2) calculate pair-wise distances and spatial weights based on the network topology, 3) fit SSNMs to data where autocorrelation is based on three spatial relationships (Euclidean, flow-connected, and flow-unconnected), 4) estimate relationships between stream variables (spatial regression), 5) make predictions at unobserved locations (prediction points), 6) export spatial data for use in other software programs, and 7) visualize the spatial data.

More information about the STARS toolset and SSN package is available at the following website: http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml

4.0 NSI Network

4.1 Reconditioning

“Reconditioning” refers to a set of editing procedures that have been applied to the original NHDFlowline data and used to generate the NSI dataset. These procedures were necessary because certain features of the NHDFlowline data are not compatible with the data structure used in the SSN package and must be modified before the data can be used to fit SSNMs. This section documents the reconditioning procedures that were completed to generate the NSI network and create the prediction points.

1) Preprocessing

First, two files were downloaded from the NHDPlusV2 website:
1) NHDPlusVnn_<dd>_<VPUid>_NHDSnapshot.7z
2) NHDPlusVnn_<dd>_<VPUid>_NHDPlusAttributes.7z

The compressed zip data contains numerous files, but only two files were used: 1) the NHDSnapshot file containing the stream line data in the shapefile NHDFlowline.shp and 2) the NHDPlusAttributes file containing the value added attribute (VAA) table called PlusFlowlineVAA.dbf.

The original NHDFlowline shapefile resides in the geographic, decimal degrees coordinate system. The data were reprojected to a rectangular coordinate system because distances measured using decimal degrees can change at different locations on the globe. Using a rectangular coordinate system satisfies the assumptions of the SSN package. The shapefile was reprojected to a national Albers coordinate system that matches the NHDPlusV2 raster data. The following parameters were used:

Projected Coordinate System:	NAD_1983_Albers
Projection:	Albers
False_Easting:	0.00000000
False_Northing:	0.00000000
Central_Meridian:	-96.00000000
Standard_Parallel_1:	29.50000000
Standard_Parallel_2:	45.50000000
Latitude_Of_Origin:	23.00000000
Linear Unit:	Meter

Geographic Coordinate System:	GCS_North_American_1983
Datum:	D_North_American_1983
Prime Meridian:	Greenwich
Angular Unit:	Degree

2) Removing Uninitialized Flow

NHDFlowline features can be assigned to one of two flow categories, either "With Digitized" or "Uninitialized." The uninitialized features generally represent three feature types: 1) isolated stream segments, 2) canal or ditches, and 3) some channels inside braided networks. Because these features lack attributes for determining their role in the flow network, they are not compatible with the STARS data model. Consequently, these features were eliminated from the NHDFlowline data to generate the

NSI stream lines (Figure 2). However, the majority of features in the NHDFlowlines shapefile are attributed as “With Digitized” and are included in the NSI stream lines network.

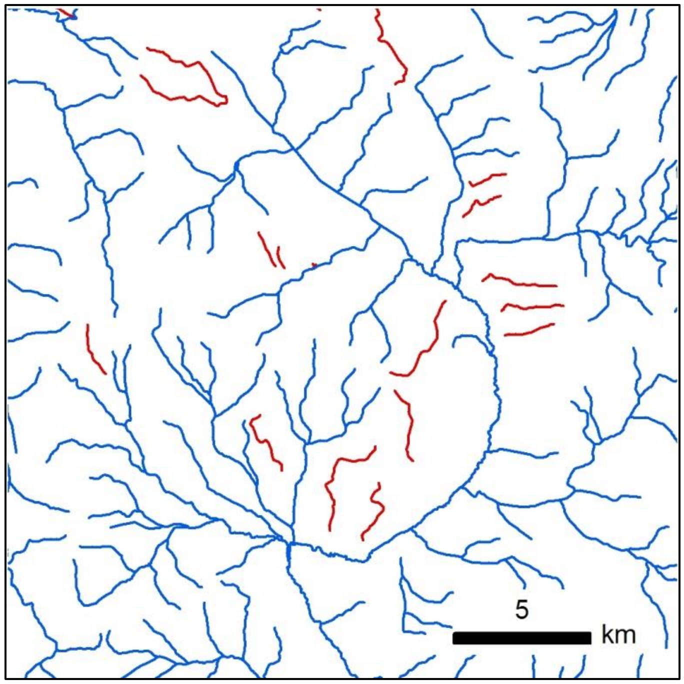


Figure 2. Features attributed as “Uninitialized” (red lines) were deleted from the NHDFlowline streams.

3) Isolated Networks

Isolated networks are disconnected from the main drainage structure; however these networks were retained in the NSI dataset. The pour point of any isolated network is considered an outlet in the LSN and network-based information (e.g., distance from outlet) was computed independently for these features.

4) Removing Braids and Diverging Flow

Braided and diverging stream segments are not permitted in the spatial stream network data model used by the SSN package. These segments were identified and removed from the NHDFlowline shapefile by deleting a selected feature set from the PlusFlowlineVAA table join. Where “StreamOrde” = “StreamCalc”, features were retained. All other features were deleted from the NHDFlowline shapefile (Figure 3).

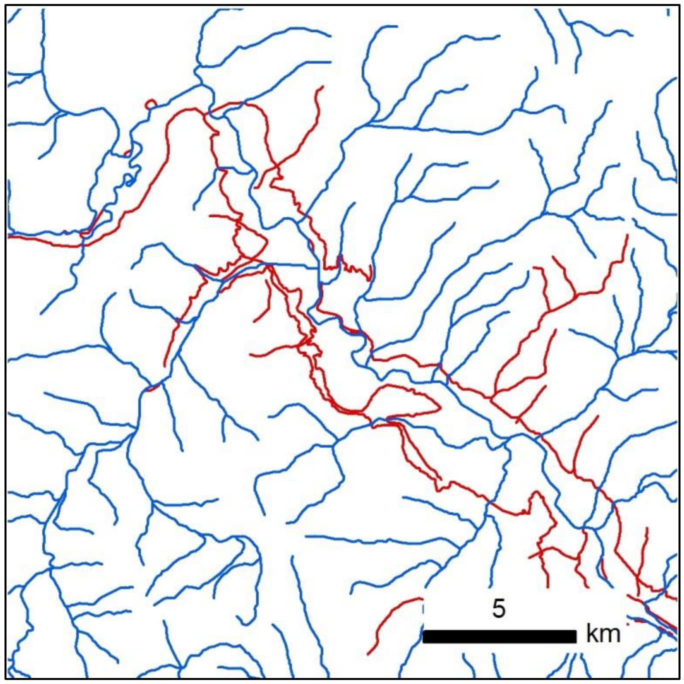


Figure 3. Braids and diversions (red lines) were deleted from the NHDFlowline streams layer.

5) Building the Landscape Network

The NHDFlowline data were imported into a LSN data model using the STARS Polyline to Landscape Network tool. This step was necessary to investigate topological conditions which are discussed in the following paragraphs. For general information about the LSN data model, please see Theobald et al. (2005) or Theobald et al. (2006). For specific information about how the LSN is used in the STARS toolset, see Peterson 2014 or Peterson and Ver Hoef (2014).

6) Identifying and Editing Topological Restrictions

Two topological restrictions are often encountered in the NHDPlusV2 data that were not addressed by the previous procedures outlined above. These include converging streams and complex confluences. These features are not considered true topological errors in a GIS or within an LSN, but must be addressed to satisfy certain SSN requirements.

Converging streams

Converging streams occur when two or more edges converge, but do not flow into another downstream edge (Figures 4, a, b). Converging streams sometimes occur in the center of waterbodies that do not have an outlet stream (Figure 4, a) or due to a topological error (Figure 4, b). In cases of convergence at waterbodies, the nodes from each converging stream segment were moved apart so that the downstream node of each segment was treated as a stream outlet. In instances with topological errors

(Figure 4b) the junctions were evaluated individually and modified as necessary. Converging nodes were identified using the STARS Check Network Topology tool.

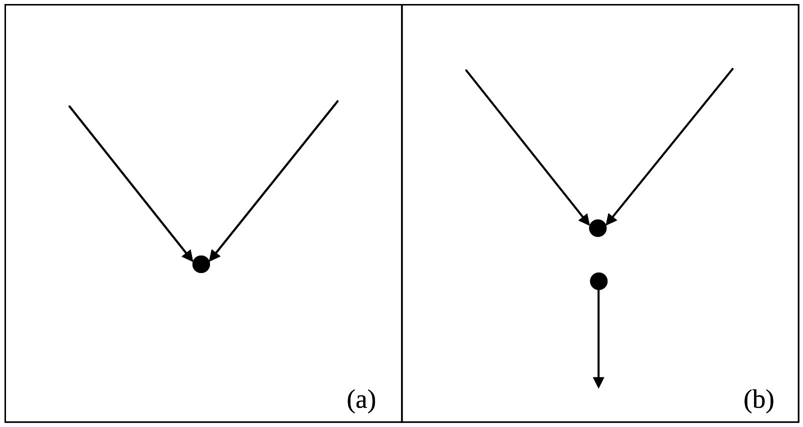


Figure 4. Converging stream nodes occur at the downstream node of two edges that converge (a, b), but do not flow into another downstream edge. This commonly occurs at the edge of the streams shapefile (a) or may be the result of topological errors within the stream network (b).

Complex Confluences

The second restriction is that only two edges may converge and flow into a single downstream edge at a confluence. Stream junctions where three or more tributaries converge upstream of a node are referred to as complex confluences. The STARS Identify Complex Confluences tool was used to identify LSN nodes that violate this condition (Figure 5a). To correct this problem at each location, the main stem stream segment was split and the tributary node with the smallest contributing area was moved downstream approximately 25 m, to the split location (Figure 5b).

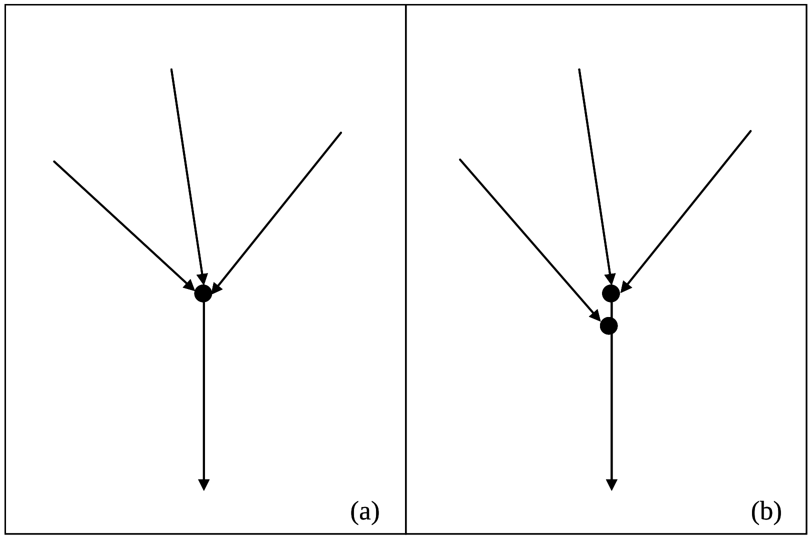


Figure 5. The stream network may contain confluences where three or more edges converge and flow into a single downstream edge (a). If this occurs, the nodes (black circles) must be identified using the Identify Complex Confluences tool and the error manually corrected (b).

Because the downstream main stem segment was split, some attributes must be recalculated to satisfy certain STARS procedures. Three attribute fields were added to accommodate the additional line features added to the network following the slitting procedure:

- *DUP_COMID* - All features in the dataset formally had a unique ComID value linked to the NHDPlusV2 dataset. Following the splitting procedure, a duplicate ComID is created. The DUP_COMID field was added to identify the duplicate ComIDs and their associated line segments. The new short segment created by the splitting procedure was assigned a value of 1, whereas all other segments received a value of 0.

- *DUP_ArSqKM* – In the original NHDPlusV2 dataset, the ArSqKM field represents the catchment contributing area in square kilometers for each feature with a unique ComID. After splitting, the field ArSqKM is assigned value 0 at the shorter split segment, because this new segment does not have a true catchment area. However, to retain the original value, DUP_ArSqKM is calculated with the original value.

- *DUP_Length* – LengthKM represents the length of each feature in the original NHDPlusV2 dataset. DUP_Length is the recalculation of all segment feature lengths following the splitting procedure.

7) Identifying Downstream Divergences

Downstream divergences occur where downstream flow splits into two stream channels. This phenomenon may represent a natural braid in the network, an artificial drainage diversion point, or a true topological error caused by a stream feature that is digitized in the wrong direction. Regardless of the cause, downstream divergences are not permitted in the LSN and were manually edited as necessary to correct the issue. The STARS Check Network Topology tool was used to identify downstream divergence.

8) Outlets and Sinks

The Check Network Topology tool was also used to identify outlets. Outlets are features without a connecting downstream feature. An outlet may be the furthest downstream segment in a basin (pour point); a true sink, where a network terminates within the basin; or a location where a stream terminates at a coastline (Figure 6). These features were identified and checked for integrity.

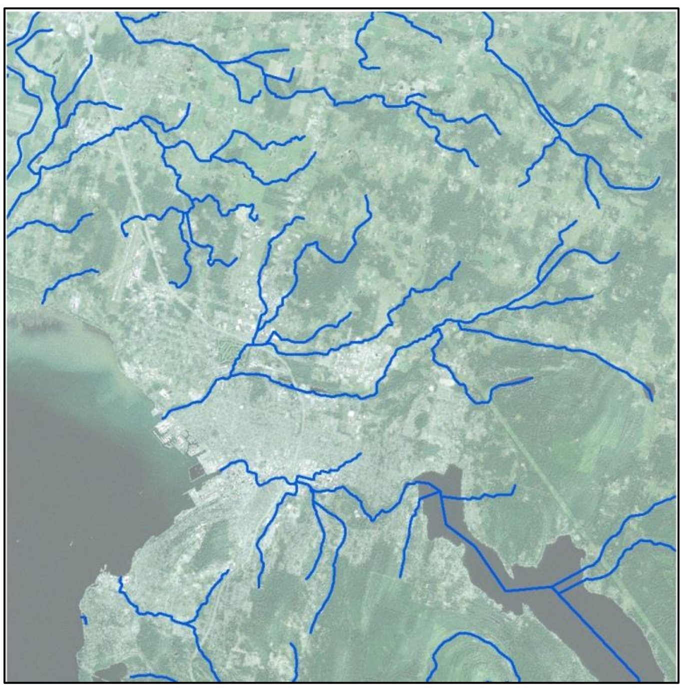


Figure 6. Legitimate outlets terminating at the coast.

4.2 Prediction Points

Prediction points (Figure 1) were generated at the center of each NHDPlusV2 stream segment containing a unique ComID. The prediction points contain the same set of attributes as the NSI network. These points can be joined to the NHDPlus Flowlines using the ComID field in order to assign the estimated response variable values to each stream segment. However, it is not necessary to use the prediction points provided with the NSI dataset. Users may generate their own points at an interval spacing which suites their needs.

4.3 Distribution Files

Two shapefile types are distributed with the NSI dataset: 1) the reconditioned NHDFlowlines and 2) prediction points.

Reconditioned NHDFlowline features are distributed with the following naming convention:
Flowline<ID><VPU>_NSI.shp
The associated prediction points are named:
PredictionPoints<ID><VPU>_NSI.shp
Where <ID> is the NHDPlusV2 Drainage Area ID and <VPU> is the Vector Processing unit number.

Files are organized and distributed by NHDPlusV2 vector processing unit (VPU) (Figure 7).

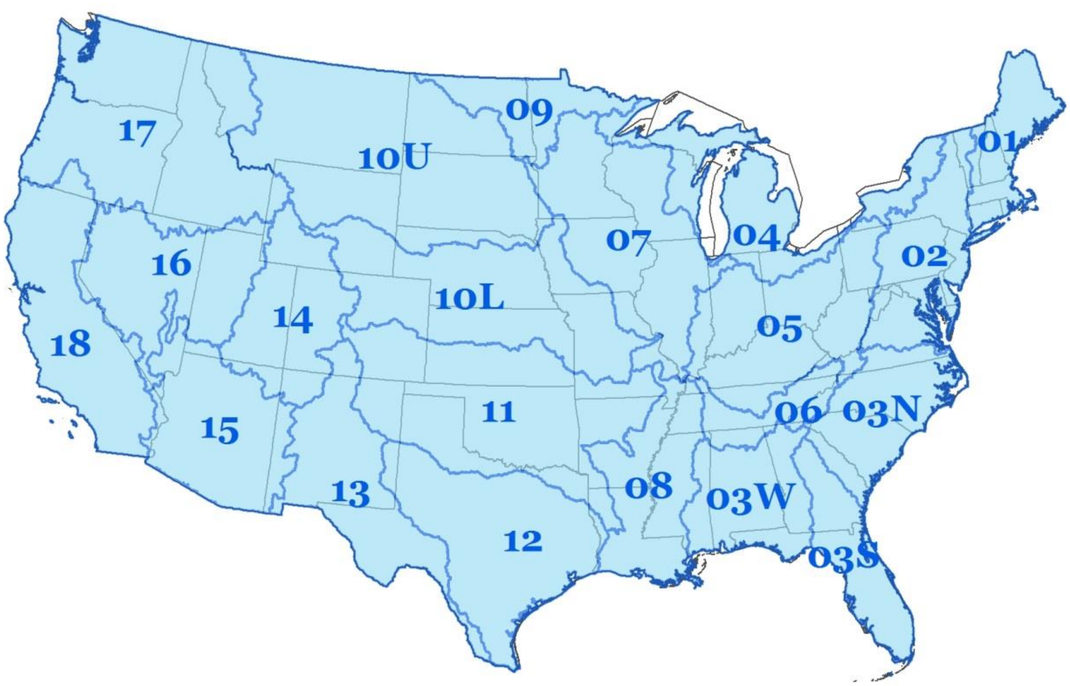


Figure 7. Vector processing units.

4.4 Attributes Description

Most attributes from the NHDFlowlines shapefile were retained for the NSI dataset, however two were removed, **Shape_Leng** and **Enabled**. **Shape_Leng** was computed in decimal degrees and was no longer valid for the reprojected NSI dataset. **Enabled** was deemed unnecessary since the value is always "True." Two attributes were derived from the PlusFlowlineVAA table, **AreaSqKM** and **TotDASqKM** (McKay et al., 2012). Three attributes were generated specifically for the NSI network, **DUP_COMID**, **DUP_ArSqKM**, and **DUP_Length**.

Note that the attribute field names may sometimes be spelled with inconsistent case when comparing this document with the NHDPlusV2 data. This discrepancy occurs because the case sometimes changes among VPUs in the NHDPlusV2 dataset.

NHDPlus V2 attributes

FID – ArcGIS internal ID number

Shape – Describes the geometry type of the features in the shapefile

ComID – Unique ID number assigned by NHDPlusV2 for each NHD stream segment

Fdate – Stream feature currency date

RESOLUTION – NHD database resolution (high, medium, or low)

GNIS_ID – Geographic Names Information System ID for the value in GNIS_Name

GNIS_Name – Feature Name from the Geographic Names Information System

LengthKM – Feature length in kilometers

ReachCode – Reach Code assigned to feature

FLOWDIR – Flow direction is "With Digitized" or "Uninitialized". Only flowlines with FlowDir = "With Digitized" are used to define the surface water network used in NHDPlus. All other flowlines are ignored by NHDPlus.

FTYPE – NHD Feature Type

Fcode – Numeric codes for various feature attributes in the NHDFCode lookup table

AreaSqKM – Reach catchment area in square kilometers, except where DUP_COMID = 1. In these cases, AreaSqKM = 0.

TotDASqKM – Total Upstream Cumulative Drainage Area, in square kilometers at the downstream end of the NHDFlowline feature

NSI attributes

DUP_COMID – Identifies segments created to eliminate 3-segment confluences. DUP_COMID = 1 designates these additional segments that are created following the splitting process.

DUP_ArSqKM – Where DUP_COMID = 1, contains the original value of the AreaSqKM field prior to receiving value 0. 0 is calculated for AreaSqKM following the 3-segment confluence splitting process.

DUP_Length – Length of all features recomputed after 3-segment confluences have been edited.

5.0 Processing the NSI Dataset using the STARS tools

Section 5 provides a set of workflow instructions to assist users in implementing the STARS toolset with the NSI dataset. These instructions provide general guidance for implementing the tools but do not reference a specific tutorial dataset. For specific examples and additional technical details, see Peterson 2014 and other documentation at SSN/STARS website: http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml. Instructions for using the SSN package are available from another source, which is referenced in step 11 below.

1) Download the NSI stream lines from the NSI website

The NSI stream lines can be downloaded from here:
http://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet.html

The zip format file must be extracted using standard compression software tools. The stream lines are stored in Esri shapefile format. The projection information is listed in section 4.0 above.

The NSI data are organized by NHDPlus Vector Processing Units (Figure 7). VPUs correspond to major drainage areas of the United States. Because VPUs are relatively large by desktop processing standards, users will not likely run statistical analyses for an entire VPU. For this reason, the data should be clipped to the project area boundary. It's advisable to use basin boundaries from a source such as the Watershed Boundary Dataset (WBD) (USGS, 2013) for clipping. The WBD was developed with the NHDPlusV2 dataset. Using a watershed boundary rather than a rectangular one will minimize the number of outlets present in the dataset and ensure that most stream lines are network connected.

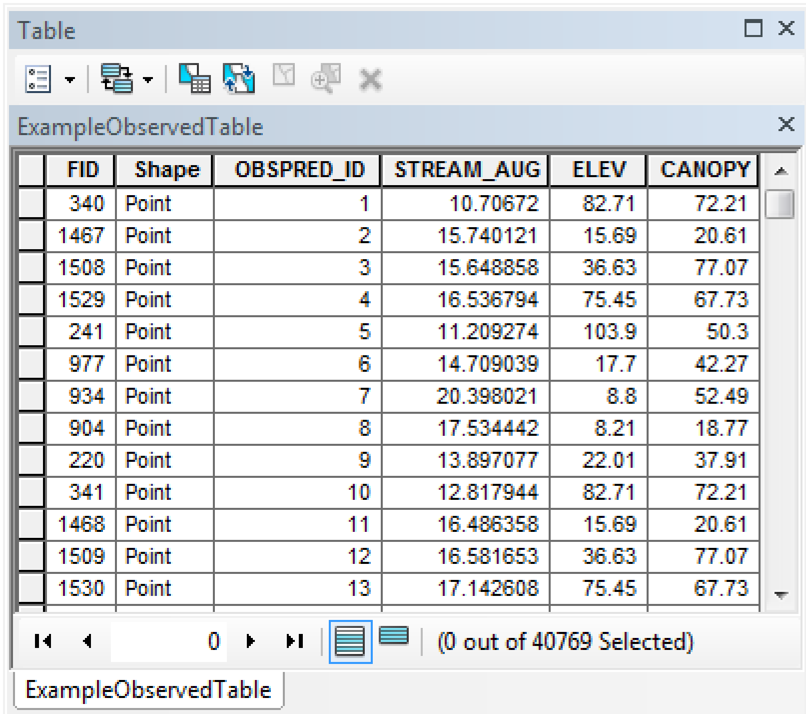
After clipping, the stream line shapefile may be reprojected to another map projection. The projection must be a rectangular one, such as UTM or Albers, to satisfy the assumptions of the SSN package. A geographic coordinate system with units of degrees is not permissible.

If the NSI prediction points will be used they should also be downloaded and extracted. The points must be clipped to match the spatial extent of the stream lines.

2) Obtain and process observation data

A number of considerations must be addressed when collecting and processing the observation data. The distribution and number of observation sites is an important consideration, but beyond the scope of this document. See Peterson et al., 2006 for more guidance on this topic.

It is incumbent on the user to record the location of the observation sites in a form that allows the points to be imported to Esri shapefile format and overlain with the NSI stream lines. After the point shapefile is generated, the field-collected values must be stored in the attribute table of the shapefile for each site. Following standard database protocols, a unique ID value is necessary to identify each site (Figure 8).



FID	Shape	OBSPRED_ID	STREAM_AUG	ELEV	CANOPY
340	Point	1	10.70672	82.71	72.21
1467	Point	2	15.740121	15.69	20.61
1508	Point	3	15.648858	36.63	77.07
1529	Point	4	16.536794	75.45	67.73
241	Point	5	11.209274	103.9	50.3
977	Point	6	14.709039	17.7	42.27
934	Point	7	20.398021	8.8	52.49
904	Point	8	17.534442	8.21	18.77
220	Point	9	13.897077	22.01	37.91
341	Point	10	12.817944	82.71	72.21
1468	Point	11	16.486358	15.69	20.61
1509	Point	12	16.581653	36.63	77.07
1530	Point	13	17.142608	75.45	67.73

Figure 8. An example attribute table for the observation site shapefile. Each record contains a unique ID number (OBSPRED_ID), the observation value or response variable (STREAM_AUG, representing mean August stream temperature), and two predictors, elevation (ELEV), and canopy percent (CANOPY).

Once the observation site shapefile has been generated, the points must be “snapped” to the NSI stream line data. Snapping ensures that the points are exactly coincident with the location of the stream lines. There are many methods for accomplishing this task and it is best for the user to determine a procedure that is most suitable to their dataset and GIS skill level. Users may utilize the ArcGIS editing tools to manually move points, being sure to employ the snapping tolerance function. In addition, the ArcToolbox tool called *Snap* (ArcToolbox > Editing Tools > Snap) may be used. Finally, the STARS toolbox contains a tool for snapping points as they are imported into the Landscape Network (ArcToolbox > STARS_v2.0.2 > Preprocessing > Snap Points to Landscape Network). However, it is recommended that the points should be well aligned with the stream lines before this program is used, because points may be inadvertently snapped to an incorrect network edge. Regardless of the method used, it’s very important to examine the points to be sure that they have been properly snapped before pursuing the statistical analysis.

3) Generate or obtain prediction points

Prediction points represent locations throughout the stream network where the dependent variable will be estimated using the SSN package. The prediction points that are distributed with the NSI dataset are located at the midpoint of each stream segment (Figure 1). Rather than use the NSI points, these locations may be defined by the user to suit the requirements of the project. Others may choose a regular distribution of points, such as 1 km spacing, or a strategy based on the Generalized Random Tessellation Stratified (GRTS) design (Stevens and Olsen, 2004).

In most cases, a GIS procedure will be used to generate the prediction points directly from the NSI stream lines. In this case, the points will already be aligned to the stream lines, so snapping will not be required. However, if the prediction points were not generated from the stream lines, it will be necessary to snap the points to the lines using procedures described for the observation points above.

As with the observation data, it is critical to assign a unique ID value to each prediction point. (Figure 9). Following the model fitting procedure, the output predictions will be joined back to the prediction points using this ID value.

Table

ExamplePredictedTable

FID	Shape	OBSPRED_ID	STREAM_AUG	ELEV	CANOPY
8671	Point	104492	-999	102.47	48
8672	Point	104493	-999	89.38	54.79
8673	Point	104494	-999	81.5	90.5
8674	Point	104841	-999	92.77	58.43
8675	Point	104842	-999	82.25	52.84
8676	Point	105014	-999	143.03	92.4
8677	Point	105015	-999	118.48	91.98
8678	Point	105016	-999	113.92	67.05
8679	Point	105017	-999	103.93	7
8680	Point	104939	-999	167.72	91.05
8681	Point	104940	-999	156.55	91.4
8682	Point	105068	-999	80.45	84.16
8683	Point	105069	-999	69.17	53.77

0 (0 out of 40769 Selected)

ExamplePredictedTable ExampleObservedTable

Figure 9. An example attribute table for the prediction point shapefile. Each record contains a unique ID number (OBSPRED_ID), the observation value (STREAM_AUG, which is flagged for NULL with -999), and two predictors, elevation (ELEV), and canopy percent (CANOPY).

4) Obtain covariate data and attribute the observed sites and prediction points

The covariate data, or predictors, are necessary for estimating the aquatic phenomena at the prediction points. One or more predictors are required. The NHDPlus dataset contains a large set of potential predictors. Users may find predictors from other local, state, and national GIS datasets as well.

Both the observation site shapefile and prediction point shapefile must be attributed with the covariate values. The observation sites will already contain the field-collected response variable values, however the covariate values must also be present (Figure 8). For the prediction points, covariate values must be added to the attribute table as well. The observation sites and prediction points shapefiles must have the same attribute field names for the covariate data. Because the prediction point file will not contain any observation values, the observation attribute is assigned a flag value of -999 (Figure 9).

Note that covariate data are not necessary if the user does not wish to estimate response values at prediction points. The SSN package may be used to explore spatial model results without making predictions.

5) Build a Landscape Network in STARS

With the NSI stream lines available and the observation sites and prediction points attributed, the data are ready to be imported into the LSN. The LSN is built during the import process and the stream lines are imported first. The points are imported into the LSN at a later time.

The LSN is an ArcGIS format Personal Geodatabase with a specific spatial topology and associated table structure used to represent the stream network. The line shapefile is converted to line and point feature classes in the Geodatabase. The lines are referred to as "edges" and the locations at the line end-points and junctions are converted to features called "nodes." An identifier is assigned to each edge (rid) and node (pointID) in the landscape network. The geographic coincidence of nodes and edges is recorded in the *noderelationships* table, while the spatial relationship of edges to one another is stored in the *relationships* table. These tables preserve information concerning flow direction in the network. For a complete discussion of the LSN data model, please see Peterson 2014, or Peterson and Ver Hoef (2014).

Before running this tool, be sure that there is a folder C:\Temp, and it is empty.
Create a new folder to hold the LSN. There should be no other files in this folder.

To generate the Landscape Network in ArcGIS, choose ArcToolbox > STARS_v2.0.1 > Pre-processing > Polyline to Landscape Network (Figure 10).

1) Enter the NSI streams shapefile for the "Streams Shapefile" input. ***Be sure that the .shp extension is included in the input name.***

2) Specify an "Output Landscape Network" name. ***Be sure to include the .mdb extension in the output file name.***

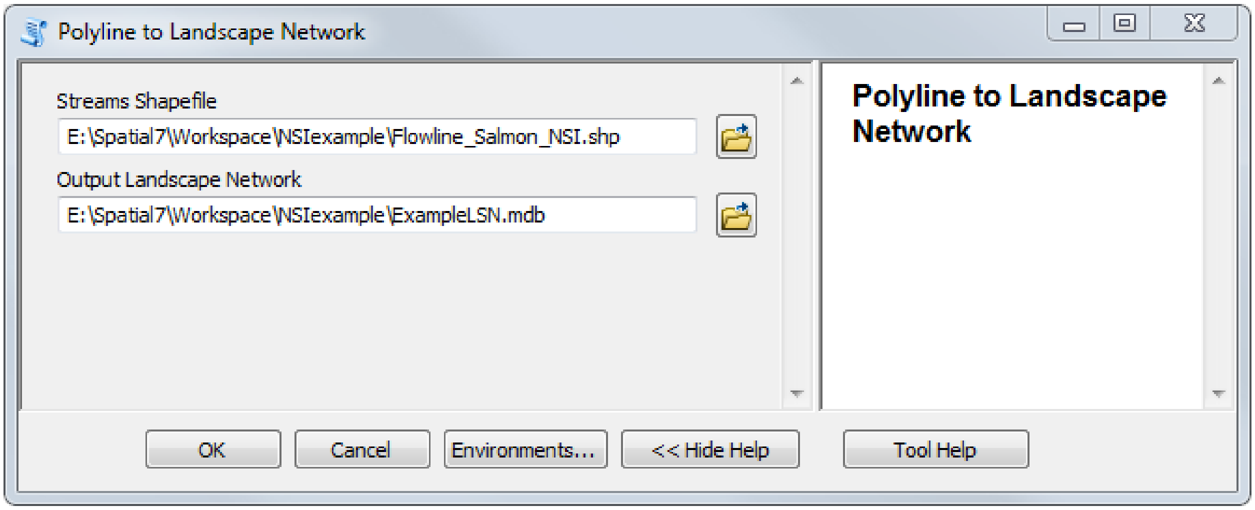


Figure 10. Polyline to Landscape Network tool.

If the tool finishes properly, the message FINISHED Polyline to Landscape Network Script will be shown in green.

The tool produces a personal geodatabase with five components:

- Nodes (point feature class)
- Edges (polyline feature class)
- Relationships (table)
- Noderelationships (table)
- Nodexy (table)

6) Check Network Topology

Although the topology of the NSI data has been reconditioned to conform to the STARS and SSN software requirements, it is import to check the network topology to be sure that clipping the NSI data or merging of adjacent NSI vector processing units has not caused any new errors.

Before running the tool, check the C:\temp folder to be sure it is empty. If not, delete the contents.

Important: ArcCatalog must be closed before running this script. Also close any instances of ArcMap, except for the instance being used to run the STARS tools.

To run the Check Network Topology tool in ArcGIS, choose ArcToolbox > STARS_v2.0.1 > Pre-processing > Check Network Topology (Figure 11).

1) Enter the nodes features class from the LSN generated in the previous step.

2) Enter a field name that will be created to hold attributes describing each node type.

3) Enter the edges features class from the LSN generated in the previous step.

4) Enter a search tolerance for finding gaps between nodes in the stream network. 10 is a good value for use with the NSI dataset.

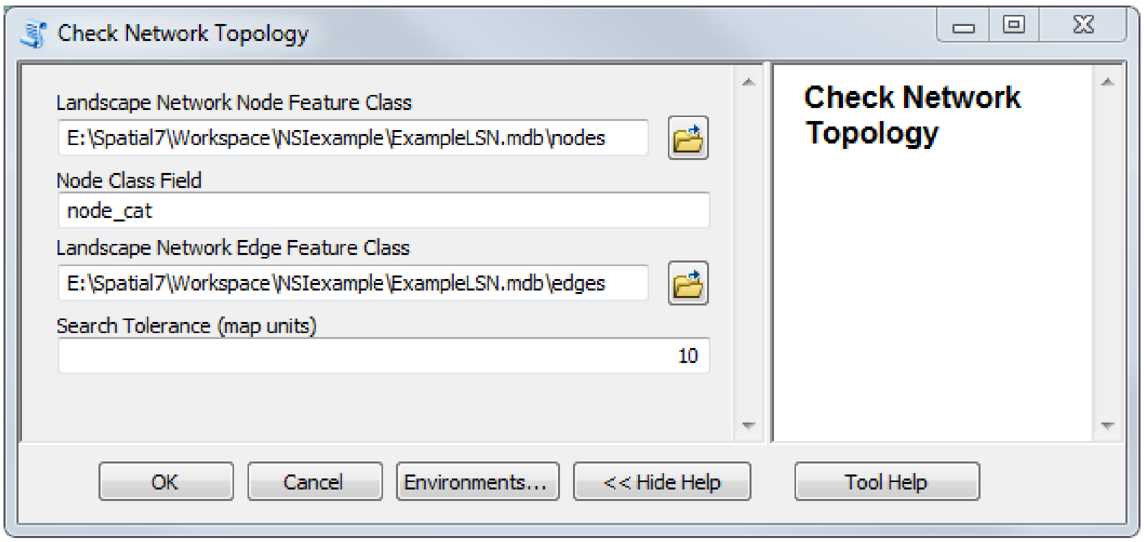


Figure 11. The Check Network Topology tool.

This script adds a new field (node_cat) to the nodes attribute table in the LSN, with a category assigned to each node (Confluence, Converging stream, Downstream Divergence, Outlet, Pseudo Node,Source). It will also create a shapefile of points, node_errors.shp, in the same directory as the LSN, which can be used to identify potential topological errors.

First display the shapefile node_errors.shp, which was created in the same folder with the LSN. Any points in the shapefile indicate the location of potential errors. These should be examined to look for breaks in the network topology. Any breaks that represent errors in the network will need to be corrected. Also, using ArcMap, display the **nodes** feature class from the LSN. Open the attribute table and note the node_cat field. Sort the records in ascending order in this field and look for Downstream Divergence and Outlets. Downstream divergences are not allowed and will need to be corrected. Outlets are allowed, but may also signal a break in network topology if the feature is not a true outlet. These locations should be examined individually to determine the best course of action.

If any errors exist, it is important to fix those in the original shapefile that was imported into the LSN. The errors should be corrected using the steps outlined in the STARS Tutorial, **STARS: Spatial tools for**

the analysis of river systems, Version 2.0.0 – A Tutorial (Peterson 2014). After the errors are corrected, the LSN should be generated again using the Polyline to Landscape Network tool. Check Network Topology should be rerun to be sure that all errors were addressed properly.

7) Import the observed sites and prediction points into the LSN

The Snap Points to Landscape Network Edges tool imports the observed sites and prediction points into the LSN and also snaps the points to the stream network. The tool links each point to a stream line in the LSN and computes a ratio that locates the position of the points on the lines. The points are incorporated into the LSN using dynamic segmentation (Theobald et al., 2006). Dynamic segmentation involves matching each point with the closest edge segment, physically moving the point to that location, and calculating the distance ratio from the end of an edge to the point location. Once this process is complete, distances between any two point locations in the LSN can be calculated using two pieces of information: the reach identifier (*rid* field) and the distance ratio (*ratio* field). Although snapping can be completed with this tool, to avoid potential errors of points snapping to the wrong stream, it is advised that the points should be snapped and verified before reaching this step. See section 5, step 2, above. The Search Radius is set to "1" if the points have already been snapped before they were incorporated into the LSN.

The tool is run twice, first for the observation sites and next for the prediction points (Figure 12). In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Pre-processing > Snap Points to Landscape Network.

1) Enter the shapefile containing the observation points in the "Sample Points Feature Class" input box. These locations are referred to as *sites* throughout this document.

2) Enter the "LSN Edges Feature Class" created in step 5.

3) Choose an "Output Snapped Points" name for the feature class that will reside within the Geodatabase (.mdb) file.

4) Choose a Search Radius of 1 if the points were previously snapped to the stream lines. Otherwise refer to Peterson 2014.

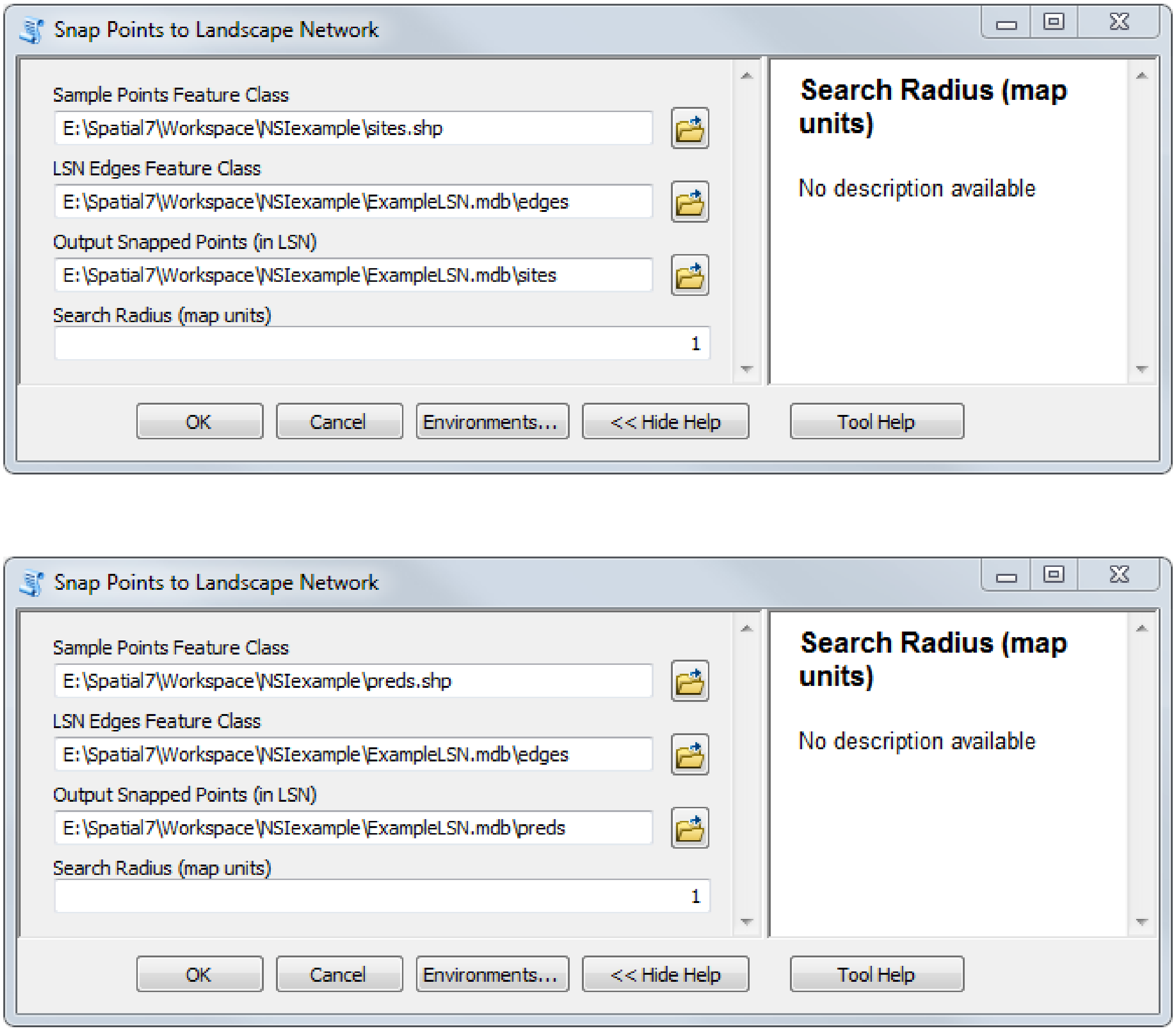


Figure 12. The Snap Points to Landscape Network tool is run twice, once for observation sites (sites.shp) and another time for prediction points (preds.shp).

If the tool runs successfully, the message Finished Snap Points to Landscape Network Edges will appear.

A new sites feature class will be written to the LSN geodatabase if the survey locations are successfully incorporated. The new sites attribute table will contain some new fields, two of which are the *ratio* and *rid* fields. The rid field indicates which edge the site has been snapped to. The ratio for each site, r_i, provides the exact location along the edge:

$$r_i = \frac{d(l_j, s_i)}{L_j}$$

where $d(l_j, s_i)$ is the distance travelled along edge j (i.e. hydrologic distance) between the most-downstream location on an edge, l_j, and the site location, s_i, and L_j is the total length of the j^{th} edge. Together, the rid and ratio values are used to identify a site's location within the LSN. This is extremely useful because it allows attributes to be estimated for site-specific locations along the edges. The site ratios range between 0 and 1 (0 ≤ ratio ≤ 1) because they are proportions; an error has occurred if values outside this range are present. Also, be sure to compare the total number of sites in the feature class to the total number of sites in the shapefile to ensure that all of the sites have been incorporated into the LSN.

When the tool is finished, the points in the LSN are lacking projection information. In ArcGIS, use ArcToolbox > Data Management Tools > Projections and Transformations > Define Projection to reset sites and preds with the correct projection, otherwise subsequent tools may fail.

8) Calculate upstream distance for edges and sites in STARS

There are two tools provided in the STARS toolset to calculate the upstream distance between the stream outlet (i.e., the most downstream location in the stream network) and each of the edges and sites: Upstream Distance – Edges and Upstream Distance – Sites. The Calculate>Upstream Distance - Edges tool calculates the total distance from the uppermost location on each line segment (upstream node) to the stream outlet (i.e., the most downstream location in the stream network) and records it in the edges attribute table in the field called **upDist**. The Calculate>Upstream Distance - Sites tool calculates the total distance from each site location to the stream outlet and records it in the sites attribute table. The new attributes in both the edges and sites attribute tables have the same units as the edges Shape_Length attribute. These attribute values provide part of the information needed to calculate flow-connected and flow-unconnected hydrologic distance measures in R.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Calculate > Upstream Distance – Edges (Figure 13).

1) Input the "Edges Feature Class" from the LSN created in step 5.

2) Choose the Length Field **Shape_Length**.

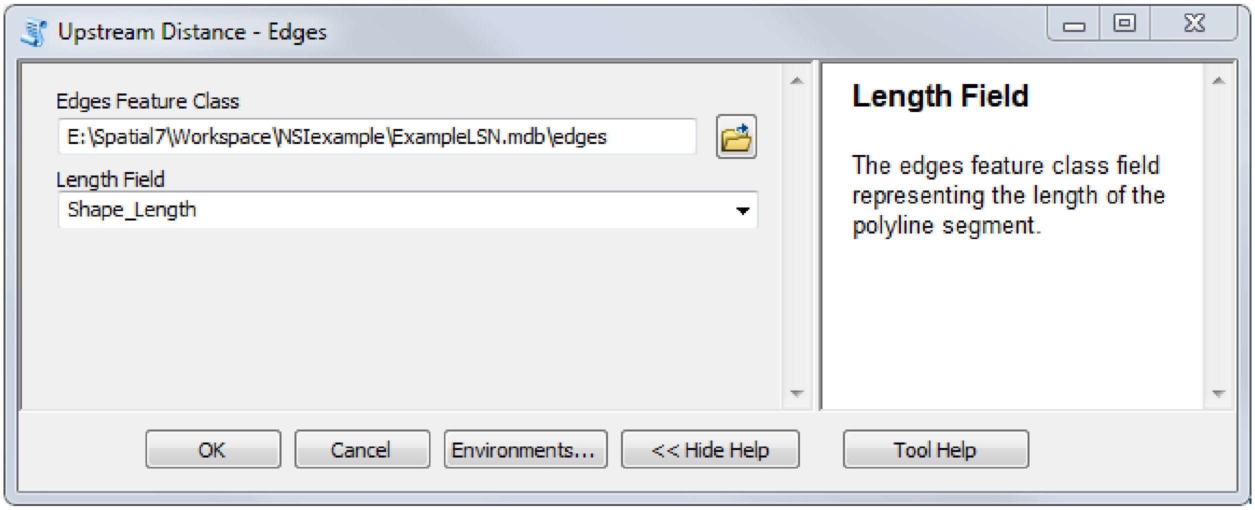


Figure 13. Upstream Distance – Edges tool.

You should see a green message Program finished successfully.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Calculate > Upstream Distance – Sites (Figure 14).

1) Input the "Edges Feature Class" from the LSN created in step 5.

2) Choose the "Length Field" **Shape_Length**.

3) Choose the *sites* and *preds* feature classes that were imported into the LSN in step 6 above.

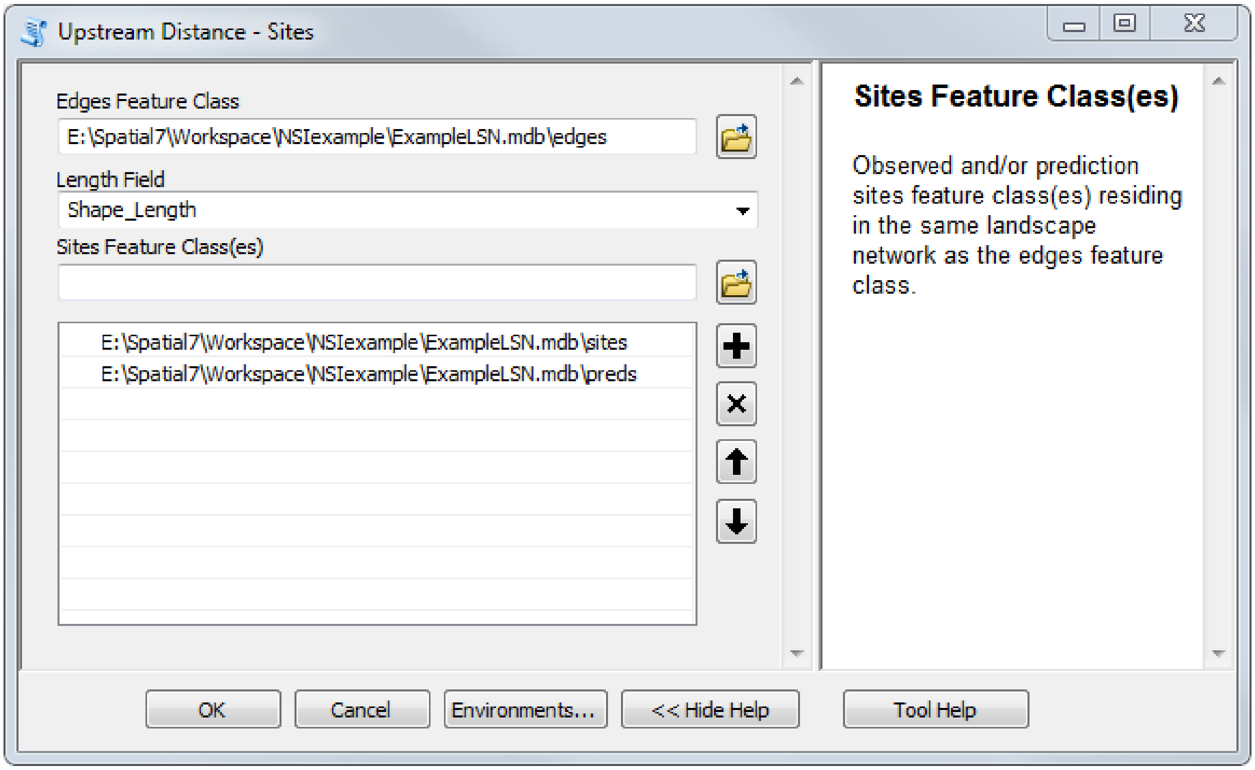


Figure 14. Upstream Distance – Sites tool.

If no errors occurred, a green message, Program finished successfully, will appear.

9) Calculate the proportional influence and additive functions in STARS.

Calculating the spatial weights needed to fit a spatial statistical model to streams data is a three step process: 1) calculating the segment proportional influence (PI), 2) calculating the additive function values, and 3) calculating the spatial weights (Peterson and Ver Hoef 2010). Steps 1 and 2 are performed using the STARS toolset, while step 3 is undertaken in R using the SSN package. The segment PI is defined as the relative influence that a stream segment has on the segment directly downstream (Figure 15).

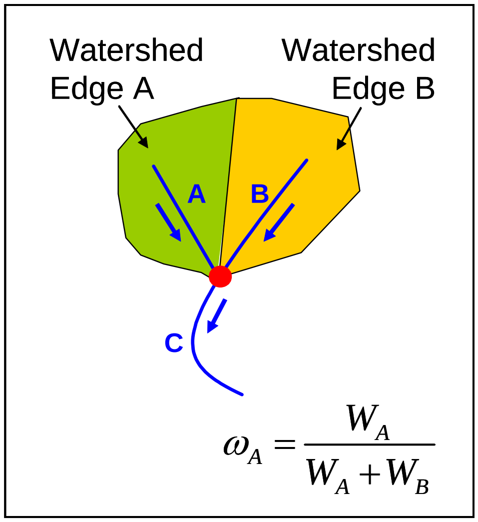


Figure 15. Calculating the segment proportional influence.

When using the NSI dataset with this instruction set the segment PI is based on watershed area. However, other measures (i.e. slope or Shreve's stream order) could also be used if the user computed these measures.

To begin, watershed area is calculated for the downstream node of each edge segment in the network, W_j (refer to section 12 of Peterson 2014). The cumulative watershed area at each confluence, pseudo, or outlet node $\sum_{k=1}^{n} W_k$, is calculated by summing the watershed area for the $n \leq 2$ edges that flow into it. The PI for each edge that flows into the node, ω_j, is then

$$\omega_j = \frac{W_j}{\sum_{k=1}^{n} W_k}$$

The segment PIs directly upstream from a confluence always sum to 1 because they are proportions.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Calculate > Segment PI (Figure 16).

1) Input the "Landscape Network Edges featureclass" called *edges* generated in step 5.

2) Use **TotDASqKM** as the "Edge field to calculate PI for." **TotDASqKM** is the accumulated drainage area calculated in NHDPlusV2 and is the proper field to use for computing the proportional influence if weighting by stream size.

3) Choose an "Output PI Field" name. *areaPI* is the convention used in this document.

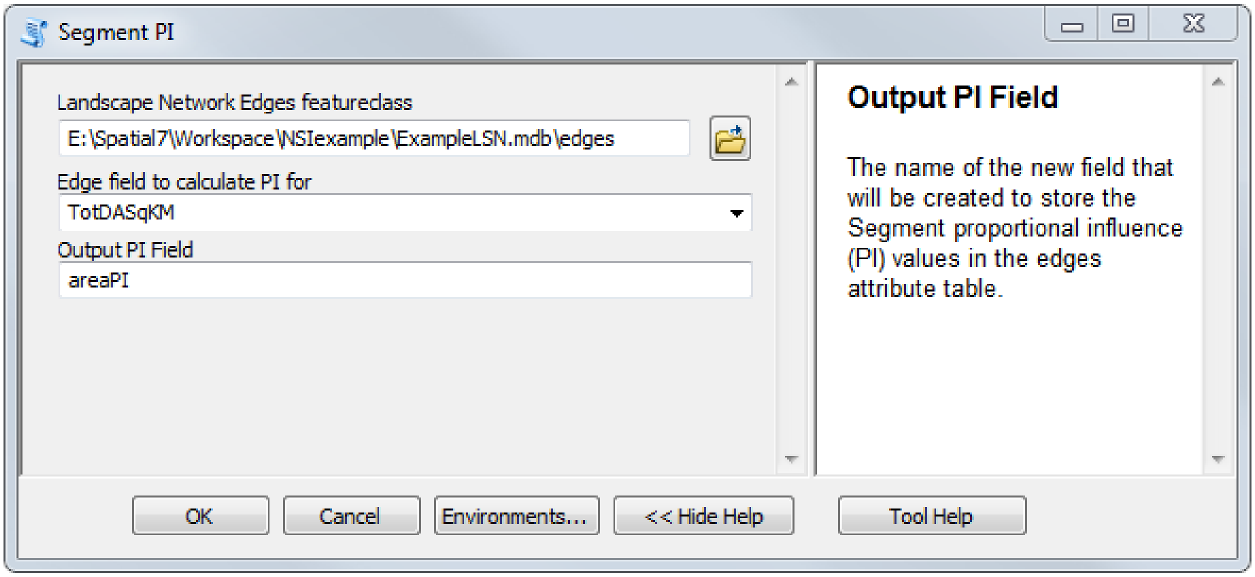


Figure 16. Segment PI tool.

If no errors occurred, a green message, Program finished successfully, will appear.

The segment PI values are proportions (0 ≤ segment PI ≤ 1). However, PI values equal to zero will occur if the segment attribute used to calculate the PI is equal to 0. If any PI values are greater than 1, then an error has occurred and the segment PIs should be recalculated.

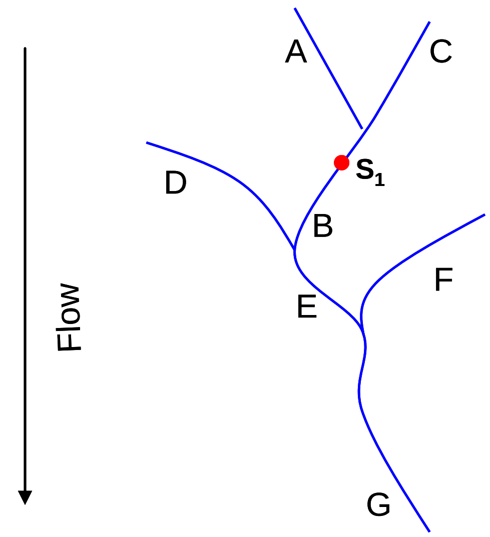
10) Calculate the additive function using STARS

Two tools have been provided in the STARS toolset which are used to calculate the additive function value (AFV) for every edge and site in the LSN (Figure 17): Additive Function – Edges and Additive Function – Sites. The AFV for a given edge, *j*, is equal to the product of the segment PIs found in the path downstream to the stream outlet.

$$AFV_j = \prod_{m=1}^{n} \omega_{D_{j,m}}$$

Note that the set of edges in the downstream path to the stream outlet, $D_{[j]}$, also includes the *j*th edge.

The AFV for the outlet segment (Figure17, edge G) is equal to 1. The AFV for a site, AFV_i, is simply equal to the AFV_j of the edge it lies on. For additional details, please see (Peterson and Ver Hoef 2010, Appendix A).



$$D_{[B]} = [B, E, G] = [D_{B,1}, D_{B,2}, D_{B,3}]$$

$$AFV_B = \prod_{m=1}^{n} \omega_{D_{B,m}} = \omega_B \times \omega_E \times \omega_G$$

$$AFV_1 = AFV_B$$

Figure 17. Calculating the additive function values (AFV) for every site and edge in the landscape network. Note that the ith site, in the example above, can be located anywhere on segment B.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Calculate > Additive Function – Edges (Figure 18).

1) Input the "Edges Feature Class" called *edges* generated in step 5.

2) Create an "Output Field" name. *afvArea* is the convention used in this document.

3) Choose the "Segment PI" field generated in step 8 above.

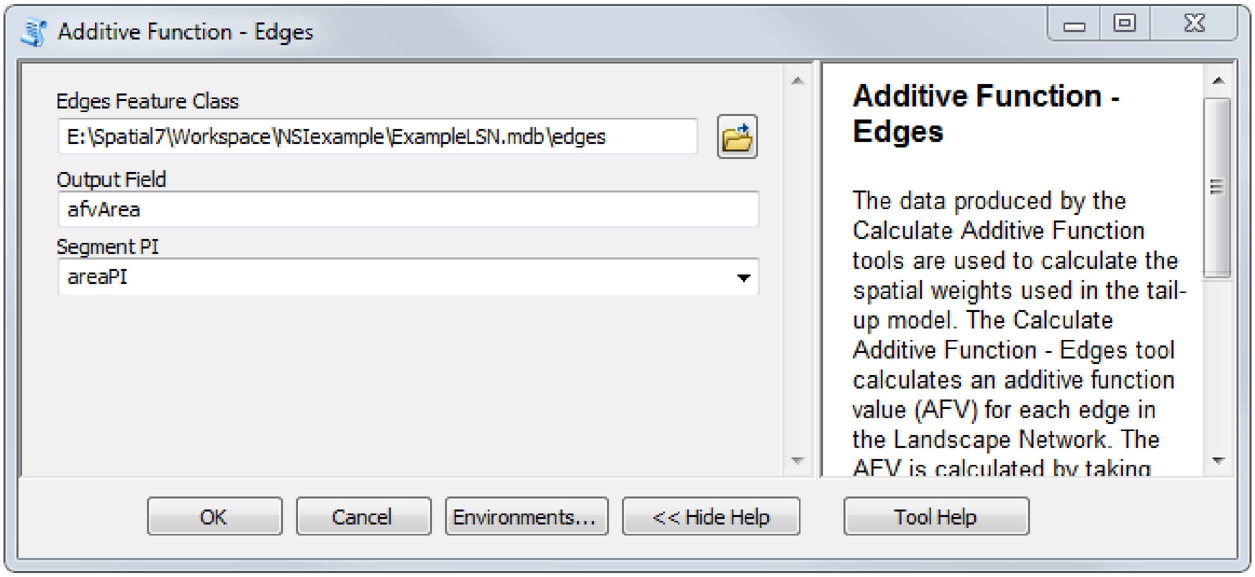


Figure 18. Additive Function – Edges tool.

If the script runs without errors a green message will appear: Finished Get Additive Function - Edges.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Calculate > Additive Function – Sites (Figure 19).

1) Input the "Edge Feature Class" called *edges* generated in step 5 above.

2) Choose the "Edges AFV Field" *afvArea* generated in Figure 17.

3) Choose the *sites* and *preds* feature classes that were imported into the LSN in step 6 above.

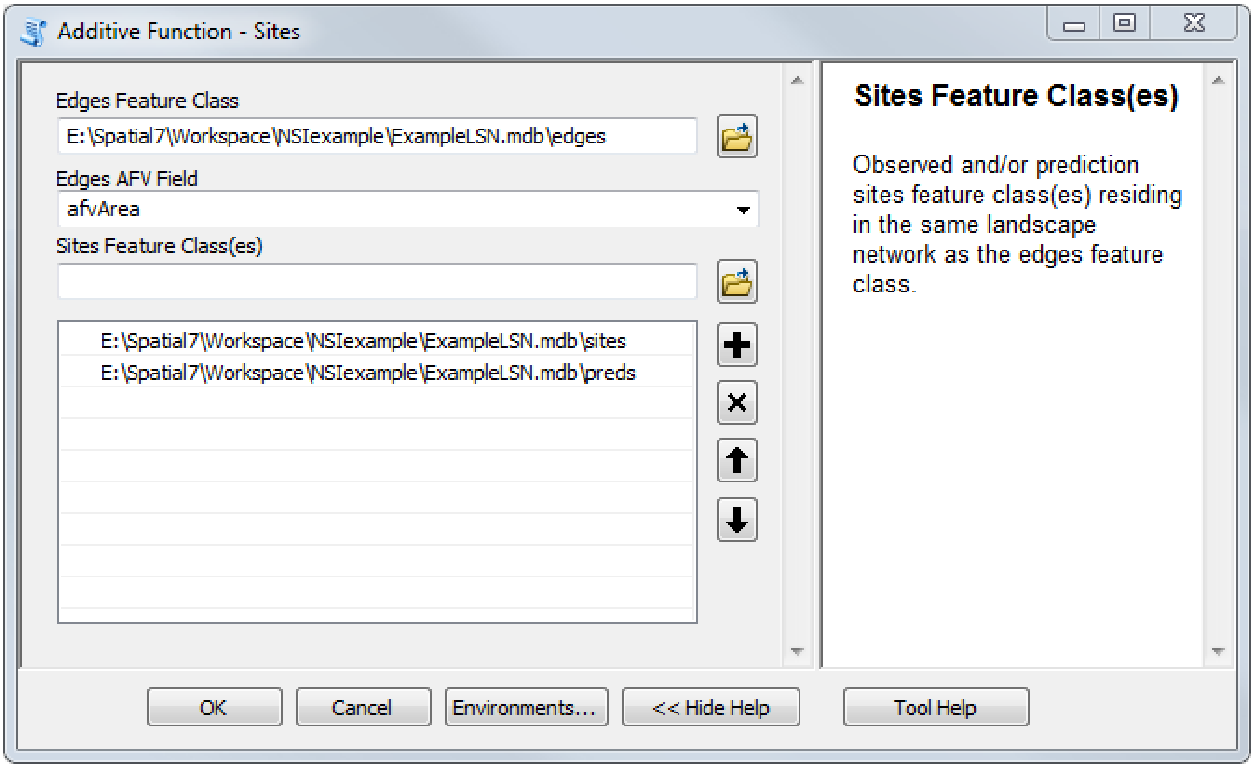


Figure 19. Additive Function – Sites tool.

When the tool finishes running, a green message should appear: Finished Additive Function Script.

The Additive Function tools create new fields in both the sites and edges attribute table representing the AFV values. The AFV is a product of proportions (segment PI values) and so the AFV should always range between 0 and 1. Check to ensure that this is the case.

11) Create the SSN object using STARS.

The purpose of the Create SSN Object tool is to reformat the LSN as a Spatial Stream Network (.ssn) object. The .ssn object represents the spatial data and the topology of the network in a format that can be easily accessed and efficiently stored and analysed in R statistical software using the SSN package. In a LSN, the relationships, nodexy, and noderelationships tables provide an efficient way to store and quickly analyse topological relationships between edges, sites, and nodes in the network. This information is powerful because it allows the spatial relationship between any two locations to be analysed, accounting for connectivity, flow direction, and distance (see Theobald et al. 2006 or Peterson and Ver Hoef 2014, for a detailed description of the relationships tables). The LSN is stored as a personal geodatabase in ArcGIS, which is a Microsoft Access file (Figure 20). Unfortunately, it is difficult to access the feature geometry of feature classes in R when they are stored in this format. As an alternative, we have chosen to create the .ssn object, which is used to store the feature geometry, attribute data, and topological relationships of each spatial dataset contained in the LSN in a format that can be efficiently stored, accessed, and analysed in R.

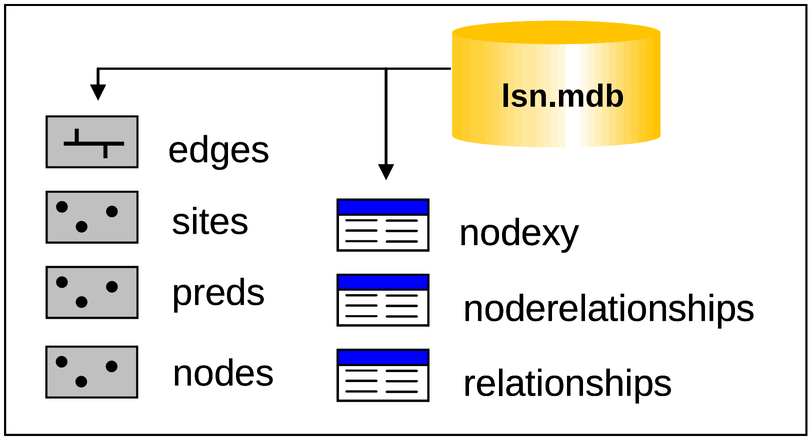


Figure 20. A landscape network (LSN) must contain six datasets before it can be used to calculate the data needed to fit the spatial statistical models: three feature classes: edges, nodes, and sites, as well as, three Access tables: nodexy, noderelationships, and relationships. Additional feature classes representing prediction locations may also be included.

The spatial datasets are stored as shapefiles in the .ssn object, since shapefiles can be easily imported into R, with all of the associated shape geometry and attributes. However, shapefiles of the edges and sites in their present form cannot be used to represent the topological relationships in the LSN. Our solution was to use network and binary identifiers (IDs). Before the binary IDs are calculated, a new folder is created to hold the files that make up the .ssn object. The folder is located in the same directory as the lsn.mdb and the naming convention is *lsn name*.ssn (e.g. lsn.ssn).

The process of assigning binary IDs is relatively straightforward. The outlet edge is identified for a network in the LSN and assigned a binary ID = 1 (Figure 21). The upstream node of one outlet edge represents the downstream node of the two edges directly upstream (Figure 21, black circles). Binary IDs are assigned to the two upstream edges by appending a 0 or a 1 to the downstream binary ID (i.e., 1 ← 10 and 11, Figure 20). This process of moving upstream and assigning binary IDs continues until every edge in the stream network has been assigned a binary ID.

It is common for LSNs to contain multiple stream networks, with unique stream outlets. For instance, there are 16 individual stream networks in the edges feature class provided in the example. Two edges may have the same binary ID if they reside on different stream networks (Figure 21). Therefore, a network identifier (netID) is also assigned to the edges, sites, and prediction sites attribute tables to differentiate between two edges with the same binary ID. In addition, the observed and prediction sites are assigned a location ID (locID) and a point ID (pid). The locID for a record will only be unique if the dataset does not contain repeated measurements at a single location (e.g. measurements taken over time at a single location). However, the pid for each record is always unique.

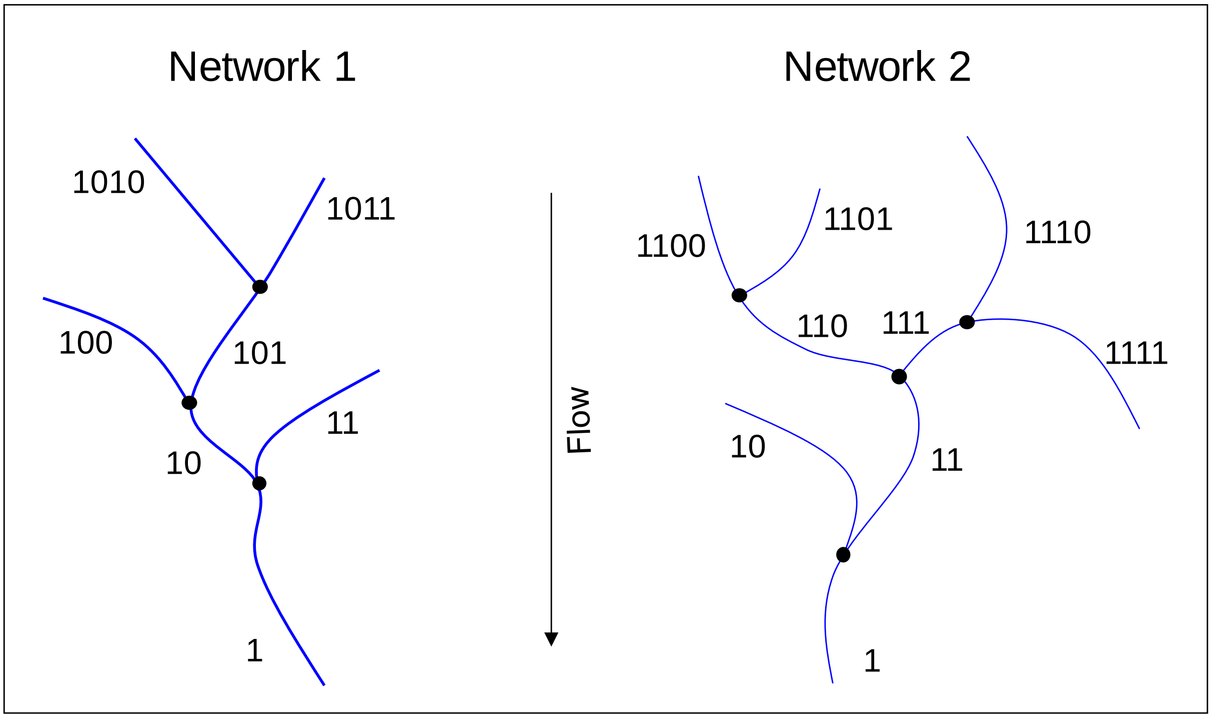


Figure 21. Binary IDs are assigned to each edge in the LSN. Edges are represented by blue lines and nodes (black circles).

The rid and binary ID for each edge are stored in a comma delimited text file (Figure 22), with a separate binary ID file for each network. The naming convention for these files corresponds to the network ID (e.g., net1.dat, net2.dat, etc.). All files are stored in the lsn.ssn folder.

```
"rid", "binaryID"
418,  1
409,  10
410,  11
407,  100
408,  101
176,  110
405,  111
347,  1010
```

Figure 22. Binary ID text file format.

Once the binary, network, location, and point IDs have been calculated and assigned, the edges, sites, and prediction sites feature classes are converted to shapefiles, which are stored in the lsn.ssn folder. When the Create SSN Object tool is complete, the .ssn object contains the spatial, attribute, and topological information of the LSN (Figure 23).

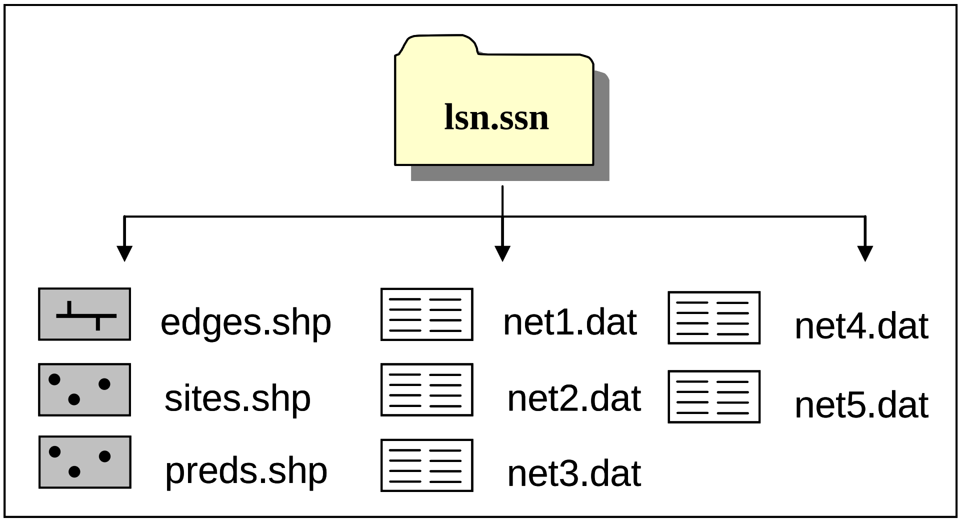


Figure 23. The .ssn object contains the spatial, attribute, and topological information of the LSN. It always contains at least two shapefiles edges and sites, as well as multiple text files containing the edge binary IDs.

In ArcGIS, go to ArcToolbox > STARS_v2.0.1 > Export > Create SSN Object (Figure 24).

1) Input the “Edges Feature Class” called *edges* generated in step 5 above.

2) Choose the *sites* feature class that was imported into the LSN in step 6 above.

3) Choose the” Site ID Field” associated with the *sites* feature class. In the examples used in this document, **OBSPRED_ID** was the name of the ID field (Figure 8).

4) Choose the preds feature class that was imported into the LSN in step 6 above.

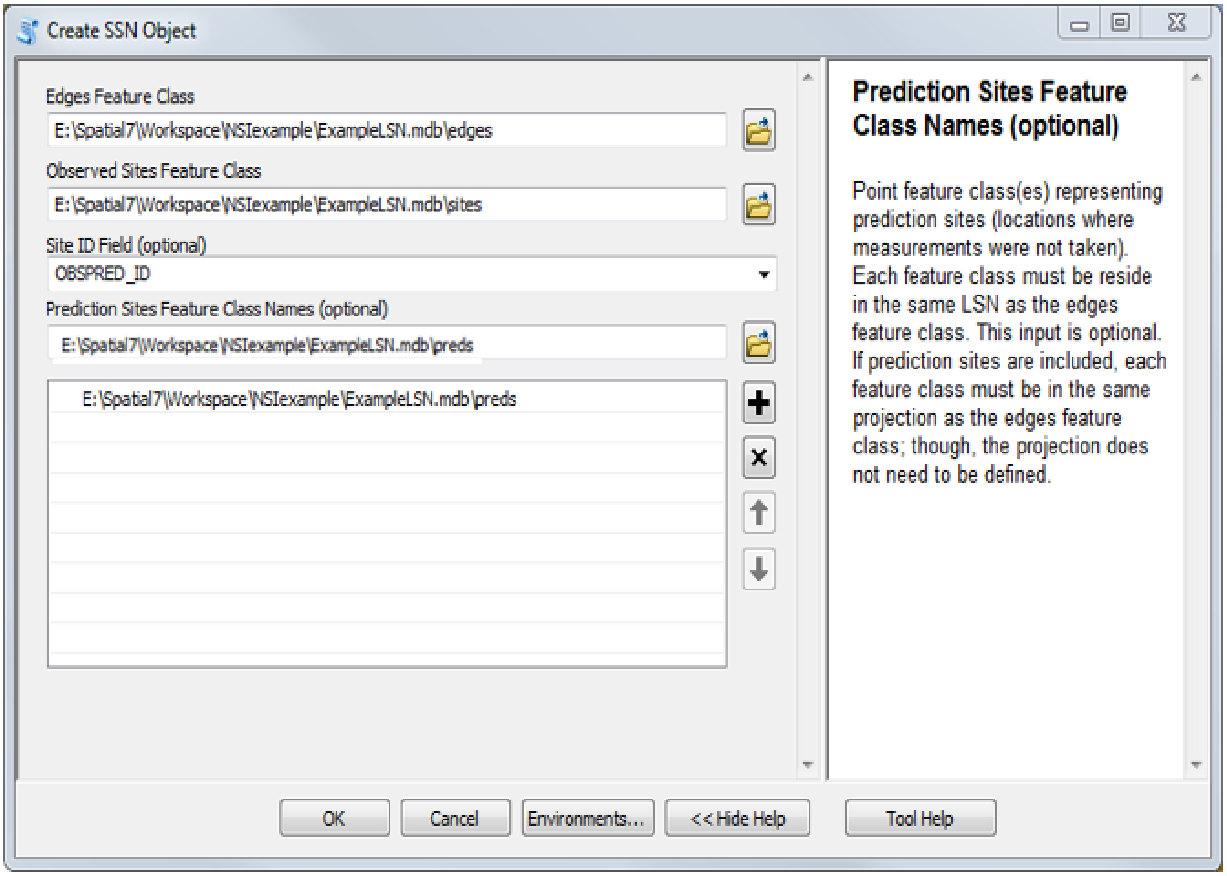


Figure 24. Create SSN Object tool.

The Site ID Field is optional, but must be specified if the "Observed Sites Feature Class" contains repeated measurements (multiple measurements at a single location). The Site ID Field should represent a *site identifier* (ID) and ***must be present and populated*** in the observed sites feature class. *Site identifiers* must be unique for each unique location.

The Create SSN Object tool may take a while to run depending on the number of edges and sites in the LSN. When the tool has finished successfully, a green message, Successfully Finished Create SSN Object Script, will appear. An .ssn object will also be created in the same directory as the LSN used to create it. It will have the same file structure as shown in Figure 19.

12) Fit spatial statistical models using SSN.

The next step involves fitting the statistical models using the SSN package in R. A guide for completing this procedure is available here (Ver Hoef et al., 2014):
http://www.fs.fed.us/rm/boise/AWAE/projects/SSN_STARS/downloads/SSN/SSNvignette2014.pdf

13) Predictions from the SSN package are joined back to prediction point shapefile.

Following the model fitting process in R, the SSN package will produce a text file with the prediction results (if prediction points were used). Each record in the results table will contain prediction values along with the *site ID* number from the prediction point shapefile. The *site ID* number in the table is used to join the SSN prediction records back to the prediction point shapefile attribute table.

14) Predictions may be joined to the NSI stream lines, or back to the original NHDFlowline data.

After the prediction values are joined back to the point shapefile, the ComID value in the shapefile may be used to join the data back to the NSI stream lines or the NHDFlowline features.

6.0 Acknowledgements

Funding was provided by the U.S. Fish and Wildlife Service Landscape Conservation Cooperative. Sharon Parkes and Sherry Wollrab contributed to the protocol development and geoprocessing of the NHDPlus data. Dona Horan reviewed this document.

7.0 References Cited

Esri 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Isaak D., Luce, C., Rieman B., Nagel D., and Peterson E., Horan D., Parkes S., and G. Chandler (2010) **Effects of recent climate and fire on thermal habitats within a mountain stream network**, Ecological Applications, 20(5):1350-1371.

McKay L., Bondelid T., and Dewald T. (2012). **NHDPlus Version 2: User Guide**. U.S. EPA and USGS.

Peterson E.E. (2014) **STARS: Spatial tools for the analysis of river systems, Version 2.0.0 – A Tutorial**. CSIRO.

Peterson E.E., Merton A.A., Theobald D.M., and Uruhart N.S. (2006) Patterns of spatial autocorrelation in stream water chemistry. Environmental Monitoring and Assessment, 121: 571-596.

Peterson E.E. and Ver Hoef J.M. (2010) **A mixed-model moving-average approach to geostatistical modeling in stream networks**. Ecology, 91(3): 644-651. **Appendix A** | **Appendix B**

Peterson E.E. and Ver Hoef J.M. (2014) **STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data**. Journal of Statistical Software, 56(2).

Python Software Foundation (2014) Python Language Reference, version 2.7. Available at http://www.python.org

R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Stevens D.L., Jr., Olsen A.R. (2004) Spatially-balanced sampling of natural resources. Journal of American Statistical Association 99: 262-278

Theobald D.M., Norman J., Peterson E.E., and Ferraz S. (2005) Functional Linkage of Watersheds and Streams (FLoWS): network-based ArcGIS tools to analyze freshwater ecosystems. Proceedings of the ESRI User Conference 2005, San Diego, CA, July 26.

Theobald D.M., Norman J.B., Peterson E., Ferraz S., Wade A., and Sherburne M.R. (2006) Functional Linkage of Water basins and Streams (FLoWS) v1 Users Guide: ArcGIS tools for Network-based analysis of freshwater ecosystems. Natural Resource Ecology Lab, Colorado State University, Fort Collins, CO. 43 pg.

Ver Hoef J.M. and Peterson E.E. (2010) **A moving average approach to spatial statistical models of stream networks**. The Journal of the American Statistical Association, 489: 6-18.

Ver Hoef J.M., Peterson E.E., Clifford D., and Shah R. (2014) **SSN: An R package for spatial statistical modeling on stream networks**. Journal of Statistical Software, 56(3).

USGS (2013) **Watershed Boundary Dataset**. Coordinated effort between the United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), the United States Geological Survey (USGS), and the Environmental Protection Agency (EPA). The Watershed Boundary Dataset (WBD) was created from a variety of sources from each state and aggregated into a standard national layer for use in strategic planning and accountability. Available URL: http://nhd.usgs.gov/wbd.html [Accessed: 27/01/2015]

Appendix A: Metadata

National Stream Internet NHD

Tags

Stream / River, Lake / Pond, Canal / Ditch, Reservoir, Spring / Seep, Swamp / Marsh, Artificial Path, Reach, Watershed, Catchment, EARTH SCIENCE, Land Surface, Topography, Cartography, GEODATA, GIS, USGS, EPA, Elevation, National Elevation Dataset, NED, National Hydrography Dataset, NHD, NHDPlus, Stream flow, Stream velocity, Spatially Referenced Regressions on Watershed Attributes, StreamStats, Water-quality, Hydrologic modeling, River Coding Systems, Hydrography, inlandWaters, STARS, Network, National Stream Internet

Summary

This stream layer is a modified version of the USGS NHDPlus Version 2 stream flowline data. These data are intended to be used with two spatial statistical toolsets called STARS and SSN (Peterson E.E. and Ver Hoef J.M. 2014, and Ver Hoef J.M. and Peterson E.E. 2010) . The purpose of the toolsets is to model biological and physical characteristics of stream networks using digital stream line data. To implement the toolsets it is necessary that the streams data are specifically structured to ensure that hydrologic distances and spatial relationships are calculated properly. The STARS toolbox (version 2.0.1) and ArcGIS software were used to identify and modify topological entities of this stream line network so that users may utilize the data directly with the spatial statistical tools. More information about STARS and SSN may be obtained at: http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml

Description

This stream layer is a modified version of the USGS NHDPlus Version 2 (NHDPlusV2) stream flowline data. The original geospatial data included in the NHDPlusV2 are intended to support a variety of water-related applications. [U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS)]. This modified version is intended to be used with the STARS and SSN statistical toolsets (Peterson E.E. and Ver Hoef J.M. 2014, Ver Hoef J.M. and Peterson E.E. 2010). The NHDPlusV2 stream network was simplified by removing braids and diverging channels, as well as modifying confluences with more than two tributaries (complex confluences). Locations where more than 2 segments joined at a stream junction were separated and snapped to a newly created node more than 25 meters downstream. Where possible, the segment with the smallest contributing area (TotDASqKM) value was moved. Where multiple stream segments converged into sinks, the terminal end nodes were moved 25 meters apart. In addition, segments where nodes created downstream divergences or where stream flow converged were separated as deemed most appropriate. This simplification of the stream network supports statistical calculations with the STARS Toolbox while maintaining the majority of network features for modeling purposes.

Peterson E.E. and Ver Hoef J.M. (2014) STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. Journal of Statistical Software, 56(2).

Ver Hoef J.M. and Peterson E.E. (2010). A moving average approach to spatial statistical models of stream networks. The Journal of the American Statistical Association, 489: 6-18

More Information: http://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet.html

Credits:

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS) .

http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php

U.S. Forest Service, Rocky Mountain Research Station, Air, Water, and Aquatic Environments Program, Boise, ID

http://www.fs.fed.us/rm/boise/awae_home.shtml

Credits
U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php
ftp://ftp.horizon-systems.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf

U.S. Forest Service, Rocky Mountain Research Station, Air, Water, and Aquatic Environments Program, Boise, ID

Use limitations
The USDA Forest Service makes no warranty, expressed or implied, including the warranties of merchantability and fitness for a particular purpose, nor assumes any legal liability or responsibility for the accuracy, reliability, completeness or utility of these geospatial data, or for the improper or incorrect use of these geospatial data. These geospatial data and related maps or graphics are not legal documents and are not intended to be used as such. The data and maps may not be used to determine title, ownership, legal descriptions or boundaries, legal jurisdiction, or restrictions that may be in place on either public or private land. Natural hazards may or may not be depicted on the data and maps, and land users should exercise due caution. The data are dynamic and may change over time. The user is responsible to verify the limitations of the geospatial data and to use the data accordingly.

Extent
There is no extent for this item.

Scale Range

Maximum (zoomed in) 1:5,000
Minimum (zoomed out) 1:150,000,000

ArcGIS Metadata ►

Topics and Keywords ►

PLACE KEYWORDS United States

THEME KEYWORDS Rivers, streams

Hide Topics and Keywords ▲

Citation ►

TITLE National Stream Internet NHD
CREATION DATE 2015-01-13 00:00:00

OTHER CITATION DETAILS
http://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet.html

Hide Citation ▲

Citation Contacts ►

RESPONSIBLE PARTY
INDIVIDUAL'S NAME David Nagel
ORGANIZATION'S NAME U.S. Forest Service
CONTACT'S POSITION GIS Analyst
CONTACT'S ROLE publisher

Hide Citation Contacts ▲

Resource Details ►

DATASET CHARACTER SET utf8 - 8 bit UCS Transfer Format

SPATIAL REPRESENTATION TYPE vector

SPATIAL RESOLUTION
DATASET'S SCALE
SCALE DENOMINATOR 1:100,000

PROCESSING ENVIRONMENT ArcGIS

CREDITS
U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php
ftp://ftp.horizon-systems.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf

U.S. Forest Service, Rocky Mountain Research Station, Air, Water, and Aquatic Environments Program, Boise, ID

ARCGIS ITEM PROPERTIES

Hide Resource Details ▲

Extents ►

EXTENT
GEOGRAPHIC EXTENT
BOUNDING RECTANGLE
EXTENT TYPE Extent used for searching

Hide Extents ▲

Resource Maintenance ►

RESOURCE MAINTENANCE
UPDATE FREQUENCY not planned

Hide Resource Maintenance ▲

Resource Constraints ►

CONSTRAINTS
LIMITATIONS OF USE

The USDA Forest Service makes no warranty, expressed or implied, including the warranties of merchantability and fitness for a particular purpose, nor assumes any legal liability or responsibility for the accuracy, reliability, completeness or utility of these geospatial data, or for the improper or incorrect use of these geospatial data. These geospatial data and related maps or graphics are not legal documents and are not intended to be used as such. The data and maps may not be used to determine title, ownership, legal descriptions or boundaries, legal jurisdiction, or restrictions that may be in place on either public or private land. Natural hazards may or may not be depicted on the data and maps, and land users should exercise due caution. The data are dynamic and may change over time. The user is responsible to verify the limitations of the geospatial data and to use the data accordingly.

Hide Resource Constraints ▲

Fields ►

DETAILS FOR OBJECT Flowline_GB11_NSI ►

DEFINITION
NHD Flowline for the entire NHD Version 2 Plus 11 RPU

DEFINITION SOURCE
USDA Forest Service RMRS Boise Aquatic Sciences Lab

FIELD FID
►

FIELD DESCRIPTION
Internal feature number.

DESCRIPTION SOURCE
Esri

DESCRIPTION OF VALUES

Sequential unique whole numbers that are automatically generated.

Hide Field FID ▲

Field Shape
▶

Field description
: Feature geometry.

Description source
: Esri

Description of values
: Coordinates defining the features.

Hide Field Shape ▲

Field COMID
▶

Field description
: Native NHDPlus attribute. A unique ID for each NHD event assigned during the central NHD update process.

Description source
: U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field COMID ▲

Field FDATE
▶

Field description
: NHDPlus attribute. Feature currency date. The date that the stream line location feature was updated in the NHDPlus database.

Description source
: U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field FDATE ▲

Field RESOLUTION
▶

Field description
: NHDPlus attribute. NHD database resolution (i.e. "high", "medium" or "local")

Description source
: U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field RESOLUTION ▲

Field GNIS_ID
▶

Field description

NHDPlus attribute. Geographic Names Information System ID for the value in GNIS_Name

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field GNIS_ID ▲

▶ Field GNIS_NAME

Field description

NHDPlus attribute. Feature Name from the Geographic Names Information System

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field GNIS_NAME ▲

▶ Field LENGTHKM

Field description

NHDPlus attribute. Feature length in kilometers as measured in the NHDPlus database. This length reflects the feature length before the National Stream Internet reconditioning process was completed.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field LENGTHKM ▲

▶ Field REACHCODE

Field description

NHDPlus attribute. Reach Code assigned to feature.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field REACHCODE ▲

▶ Field FLOWDIR

Field description

NHDPlus attribute. Flow direction is “With Digitized” or “Uninitialized” Only flowlines with FlowDir = “With Digitized” are used to define the surface water network for the National Stream Internet. All other flowlines are deleted.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field FLOWDIR ▲

Field FTYPE

▶

Field description

NHDPlus attribute. NHD Feature Type.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field FTYPE ▲

Field FCODE

▶

Field description

NHDPlus attribute. Numeric codes for various feature attributes in the NHDFCode lookup table

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field FCODE ▲

Field AreaSqKM

▶

Field description

NHDPlus attribute. Reach catchment area in square kilometers as computed in the NHDPlus database, Exception: where DUP_COMID = 1, AreaSqKM was calculated to 0 for the National Stream Internet dataset.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Accuracy information

Accuracy A value of 0 (zero) is given to segments that were creating while splitting the original NHD Plus version 2 and creating a new duplicate segment.

Hide Field AreaSqKM ▲

Field TotDASqKM

▶

Field description

NHDPlus attribute. Total Upstream Cumulative Drainage Area, in square kilometers, at the downstream end of the NHDFlowline feature.

Description source

U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS).

Hide Field TotDASqKM ▲

Field DUP_COMID

▶

Field description

National Stream Internet attribute. Indicates a duplicate COMID created by splitting the original segment to correct the complex confluence issue. DUP_COMID = 1 for the short segment that was created by splitting. All other segments have DUP_COMID = 0.

DESCRIPTION SOURCE
USDA Forest Service RMRS Boise Aquatic Sciences Lab

ACCURACY INFORMATION
ACCURACY A value of "1" indicates that the original NHDPlus version 2 segment was split,

Hide Field DUP_COMID ▲

FIELD DUP_ArSqKM
▶
FIELD DESCRIPTION
National Stream Internet attribute. This value is the ArSqKM value for the original stream segment before it was split to modify complex confluences.

DESCRIPTION SOURCE
USDA Forest Service RMRS Boise Aquatic Sciences Lab

Hide Field DUP_ArSqKM ▲

FIELD DUP_Length
▶
FIELD DESCRIPTION
National Stream Internet attribute. This value is the calculated length (in kilometers) of the new segment after the segment was split to modify complex confluences.

DESCRIPTION SOURCE
USDA Forest Service RMRS Boise Aquatic Sciences Lab

Hide Field DUP_Length ▲

Hide Details for object Flowline_GB11_NSI ▲

Hide Fields ▲

Metadata Details ▶

METADATA CHARACTER SET utf8 - 8 bit UCS Transfer Format

SCOPE OF THE DATA DESCRIBED BY THE METADATA dataset

ARCGIS METADATA PROPERTIES
METADATA FORMAT ArcGIS 1.0
STANDARD OR PROFILE USED TO EDIT METADATA FGDC

LAST MODIFIED IN ARCGIS FOR THE ITEM 2015-01-13 15:16:30

Hide Metadata Details ▲

Metadata Contacts ▶

METADATA CONTACT

INDIVIDUAL'S NAME David Nagel
ORGANIZATION'S NAME U.S. Forest Service, Rocky Mountain Research Station, AWAE Program
CONTACT'S POSITION GIS Analyst
CONTACT'S ROLE publisher

CONTACT INFORMATION ►

PHONE
VOICE 208-373-4391

ADDRESS
TYPE both
DELIVERY POINT 322 East Front St; Suite 401
CITY Boise
ADMINISTRATIVE AREA ID
POSTAL CODE 83702

Hide Contact information ▲

METADATA CONTACT
INDIVIDUAL'S NAME Sharon (Parkes) Payne
ORGANIZATION'S NAME USDA Forest Service RMRS Boise ASL
CONTACT'S POSITION GIS Specialist
CONTACT'S ROLE distributor

CONTACT INFORMATION ►

PHONE
VOICE 208-373-4356

ADDRESS
TYPE both
DELIVERY POINT 322 East Front St; Suite 401
CITY Boise
ADMINISTRATIVE AREA ID
POSTAL CODE 83702
E-MAIL ADDRESS sparkes@fs.fed.us

Hide Contact information ▲

Hide Metadata Contacts ▲

Metadata Maintenance ►

MAINTENANCE
UPDATE FREQUENCY not planned

Hide Metadata Maintenance ▲