National Stream Internet Protocol and User Guide

David Nagel[1], Erin Peterson[2], Daniel Isaak[1], Michael Dumelle[3], Jay Ver Hoef[4], and Dona Horan[1]

USDA Forest Service, Rocky Mountain Research Station
Water and Watersheds Program
322 E. Front St., Boise, ID

UNITED STATES DEPARTMENT OF AGRICULTURE - U.S. FOREST SERVICE
Rocky Mountain Research Station
Water and Watersheds Program

Author affiliations: [1] USDA Forest Service, Rocky Mountain Research Station, Water and Watersheds Program, 322 E. Front St., Suite 401, Boise, ID 83702. [2] EP Consulting, Brisbane, Australia. [3]USEPA Office of Research and Development, Corvalis, OR [4] National Oceanic and Atmospheric Administration, Fairbanks, AK.

Version 9-24-2024

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) by editing spatial features in the dataset that ensure compatibility with SSNMs. This document 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 and US Environmental Protection Agency, 2024.) is an attribute rich, digital stream network developed at 1:100,000 scale by the Environmental Protection Agency (EPA) and U.S. Geological Survey (USGS). Link: https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data.
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 proven challenging using standard, non-spatial statistical techniques developed for terrestrial systems (Peterson and Ver Hoef, 2010). Two software packages for R Statistical Software (R Core Team, 2024) 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) SSNbler (Peterson et al. 2024) and 2) SSN2 (Dumelle et al., 2023). 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 SSNbler and SSN2 packages. The NSI network is inextricably linked to the SSNbler and SSN2 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 SSNbler and the SSN2 packages. More information about using the SSNbler and SSN2 packages can be found here: https://research.fs.usda.gov/rmrs/projects/stars-ssn.

1.2 Purpose

The NSI dataset contains two types of data: streamlines and prediction points. The purpose of this document is to describe the NSI dataset, as well as provide an overview of the modifications made to the NHDPlusV2 streamlines and the procedures used to generate prediction points. This guide also seeks to assist users in utilizing Spatial Stream Network Models (SSNMs) by providing a brief overview and references for using the NSI with the SSNbler and SSN2 packages, including a list of software requirements and a summary of terms and procedures for completing the modeling process. A SSNM is a digital spatial model developed for an aquatic theme, such as stream temperature, using the SSN2 package. For more information on SSNMs, see *An Introduction to Spatial Stream Network Modeling in R*

Using SSN2: https://usepa.github.io/SSN2/articles/introduction.html. Also see: https://usepa.github.io/SSN2/.

Although this document will not provide all the information necessary to use these products, links provided herein will lead users to the resources necessary to do so.

2.0 NSI Network

2.1 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 topological features of the NHDFlowline data are not compatible with the SSN object structure (previously Spatial Stream Network) used in the SSN2 package. Thus, the topology of the streams data must be modified before the data can be used to fit Spatial Stream Network Models (SSNMs). This section documents the reconditioning procedures that were completed to generate the topologically corrected NSI network and create the NSI 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 streamline 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. It is necessary to use data a rectangular coordinate system in spatial statistical models because distances measured using decimal degrees cannot be calculated properly by the package. Therefore, 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
WKID:	5070
Authority:	EPSG
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 direction of flow, they are not compatible with the LSN data model. Consequently, these features were eliminated from the NHDFlowline data when the NSI streamlines were created (Figure 1). However, most of the features in the NHDFlowlines shapefile are attributed as "With Digitized" and are included in the NSI streamlines network.

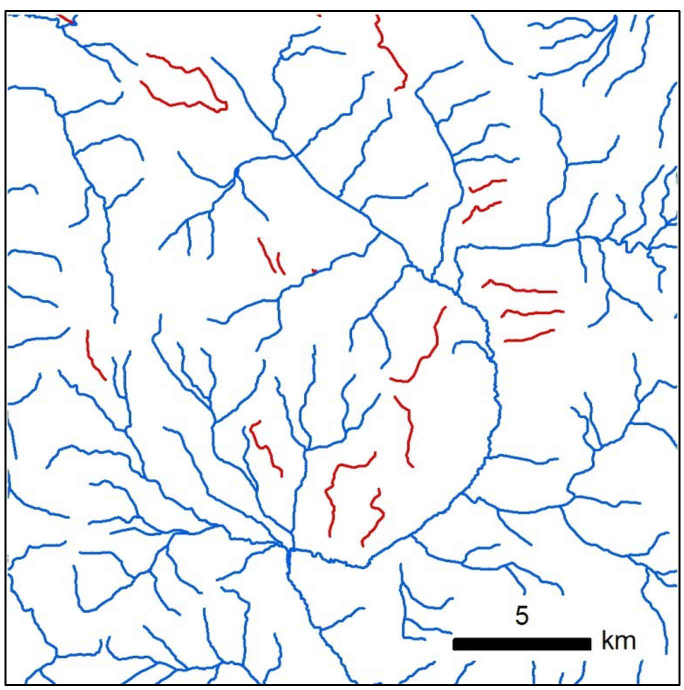

Figure 1. Features attributed as "Uninitialized" (red lines) were deleted from the NHDFlowline streams.

3) Isolated Networks

A note about isolated networks: Isolated networks are comprised of streamlines 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.

4) Removing Braids and Diverging Flow

Although braided, or diverging stream segments are not true topological errors, they are not permitted in the SSN object used in the SSN2 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 2).

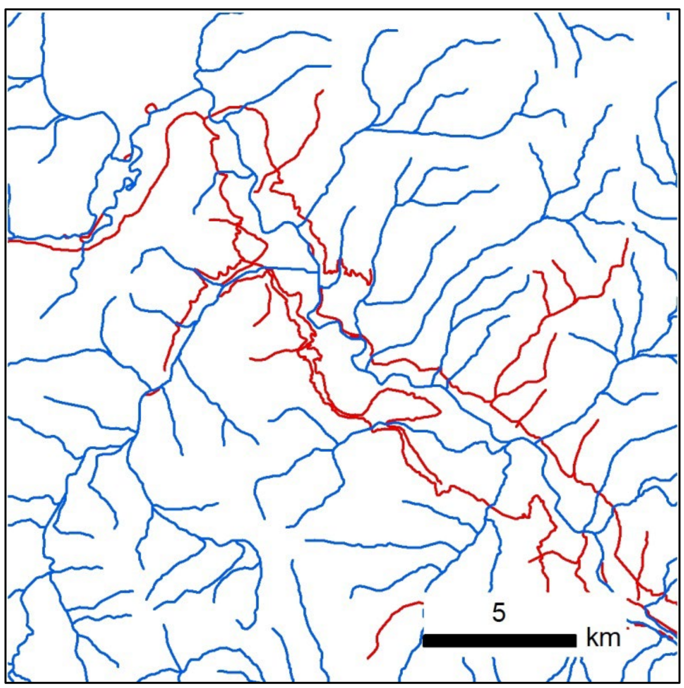

Figure 2. Braids and diversions (red lines) were deleted from the NHDFlowline streams layer.

5) Building the Landscape Network

The NHDFlowline data were imported into a landscape network (LSN) data model with a precursor program to SSNbler, called STARS, using the 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 was used in the STARS toolset, see Peterson and Ver Hoef (2014).

6) Converging Streams and Complex Confluences

Two additional topological restrictions are often encountered in the NHDPlusV2 data that were not addressed by the previous procedures outlined above: converging streams and complex confluences. These features are not considered true topological errors in a GIS but must be addressed before building an SSN object.

Converging streams

Converging streams occur when two or more edges converge, but do not flow into another downstream edge (Figures 3, a, b). Converging streams sometimes occur in the center of waterbodies that do not have an outlet stream (Figure 3, a) or due to a topological error (Figure 3, 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 3b) the junctions were evaluated individually and the location of the end node(s) modified as necessary. Converging nodes were identified using the STARS Check Network Topology tool.

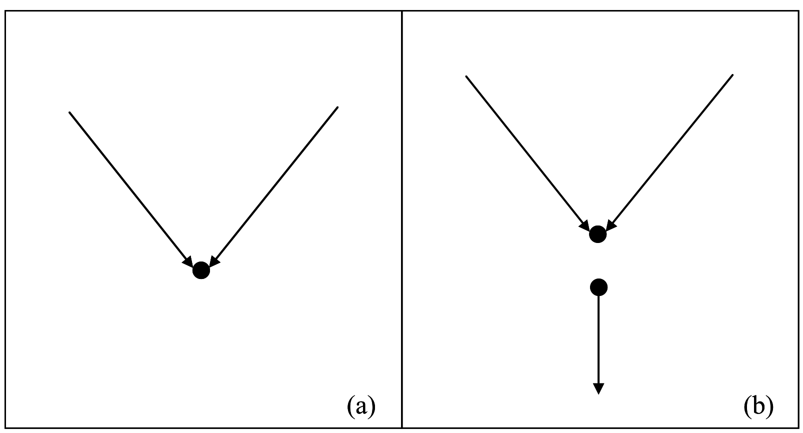

Figure 3. Converging stream nodes occur at the downstream node of two flowlines that converge (a, b), but do not flow into another downstream line feature. This commonly occurs at the boundaries of the streams shapefile (a) or may be the result of topological errors within the stream network (b).

Complex Confluences

The second restriction is that only two flowlines may converge and flow into a single downstream flowline 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 nodes that violate this condition (Figure 4a). To correct this problem at each location, the flowline representing the main stem was split approximately 25m downstream and the tributary node with the smallest contributing area was moved to the split location (Figure 4b).

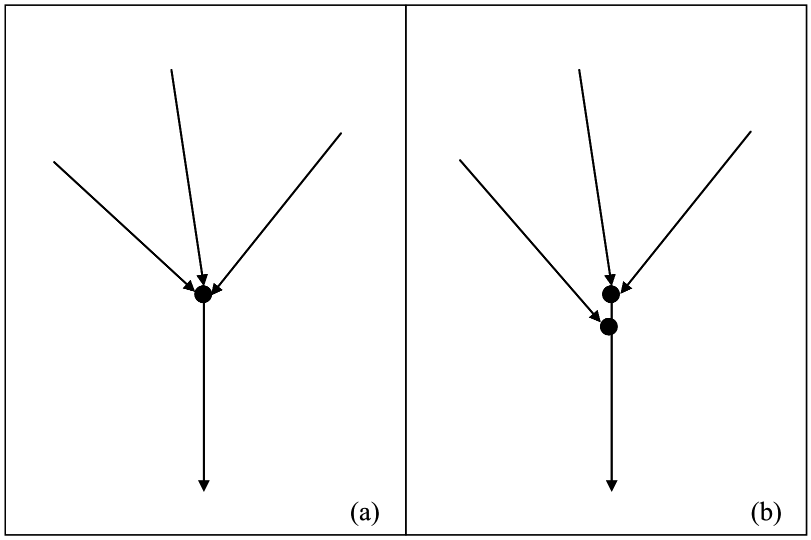

Figure 4. The stream network may contain confluences where three or more flowlines converge and flow into a single downstream flowline (a). If this situation occurs, the nodes (black circles) must be identified using the Identify Complex Confluences tool and the error manually corrected (b).

Because the main stem segment was split downstream, some attributes must be recalculated to avoid duplicate identifiers and incorrect areas and lengths. 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 5). These features were identified and checked for integrity.

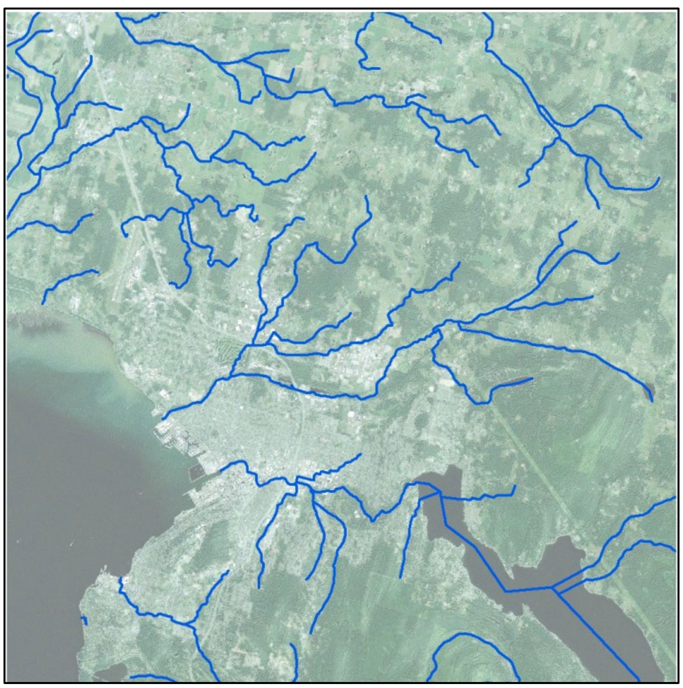

Figure 5. Legitimate outlets terminating at the coast.

2.2 Prediction Points

Prediction points represent unsampled locations on the stream network, where predictions will be made using a SSNM in the SSN2 package. Prediction points (Figure 7, page 14) were generated at the midpoint of each NHDPlusV2 stream segment containing a unique COMID. The NSI prediction points contain the same set of attributes as the NSI stream network. These points can be joined to the NHDPlusV2 flowlines or additional attribute tables for NHDPlus Catchments (e.g. Wieczorek and LaMotte 2010) using the COMID field to assign additional attributes to prediction points or the estimated response 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.

2.3 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 6).

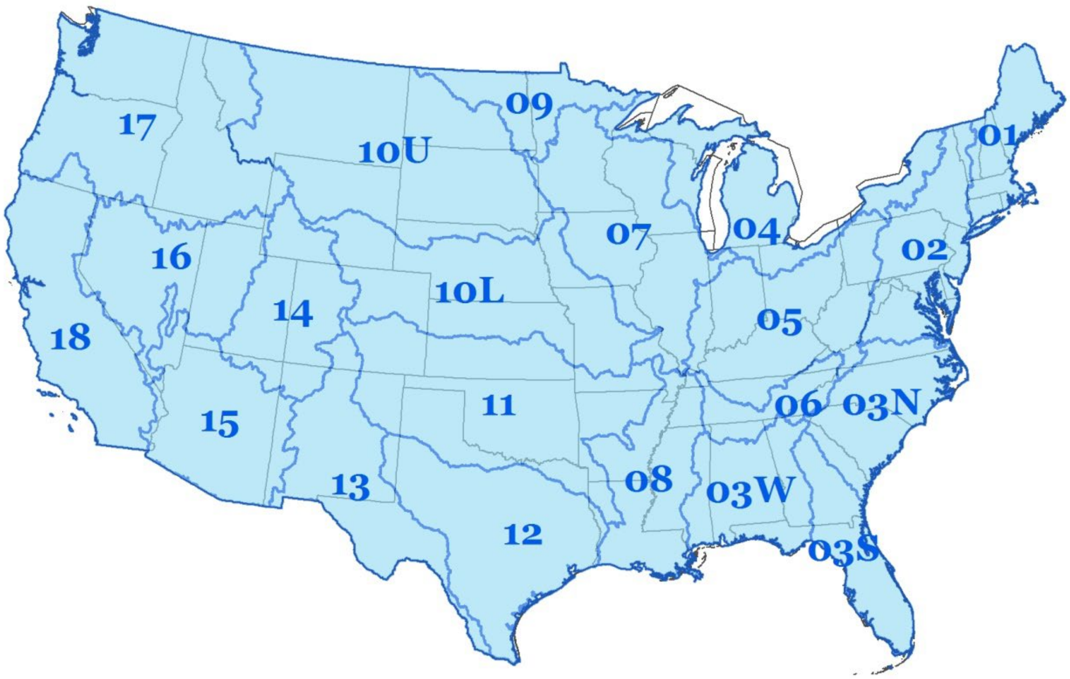

Figure 6. Vector processing units.

2.4 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

<u>**NSI attributes**</u>

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.

3.0 Software Requirements

The following software is required or recommended to utilize the NSI network with the SSNbler and the SSN2 R packages.

R statistical software

The SSN package was developed for use with R statistical software (R Core Team, 2021-2024). Information about downloading and installing R can be found here: http://www.r-project.org/

RTools

RTools is necessary for installing the SSNbler package from the GitHub developer platform. This software will not be necessary once SSNbler is available on the CRAN package repository.
https://cran.rstudio.com/bin/windows/Rtools/

SSNbler
Available from GitHub:
https://github.com/pet221/SSNbler

SSN2
Available from the CRAN package repository:
https://cloud.r-project.org/web/packages/SSN2/index.html

GIS Software
The SSNbler and SSN2 packages generate several GIS datasets in GeoPackage format representing streamlines, observation sites, and prediction sites (if predictions will be made). There are several steps in the SSNbler workflow where a GIS may be needed to ensure that the stream network is free of topological errors and that it contains additional attributes needed to fit SSNMs in SSN2. Because GeoPackage is an open-source format (https://www.geopackage.org/), many GIS software packages can read the format. Two packages are recommended here, although others may also be adequate:

ArcGIS Pro (Esri, 2023)
https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview
QGIS (QGIS, 2023)
https://www.qgis.org/en/site/

4.0 Brief Overview of Nomenclature, Data, and Tools for SSN Modeling

This section provides a brief overview of the required inputs for modelling with the SSN2 R package. A more detailed tutorial entitled *Preparing Input Data for Spatial Stream Network Modeling* is available at this website: https://research.fs.usda.gov/rmrs/projects/spatial-stream-network

4.1 Nomenclature

The SSNbler and the SSN2 packages enable users to fit SSNMs to field collected data at sample locations and make subsequent predictions of aquatic attributes at unsampled locations. These field-collected data are needed to fit an SSNM and are referred to as *observed data*, which are collected at *observation sites* (Figure 7). 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*.

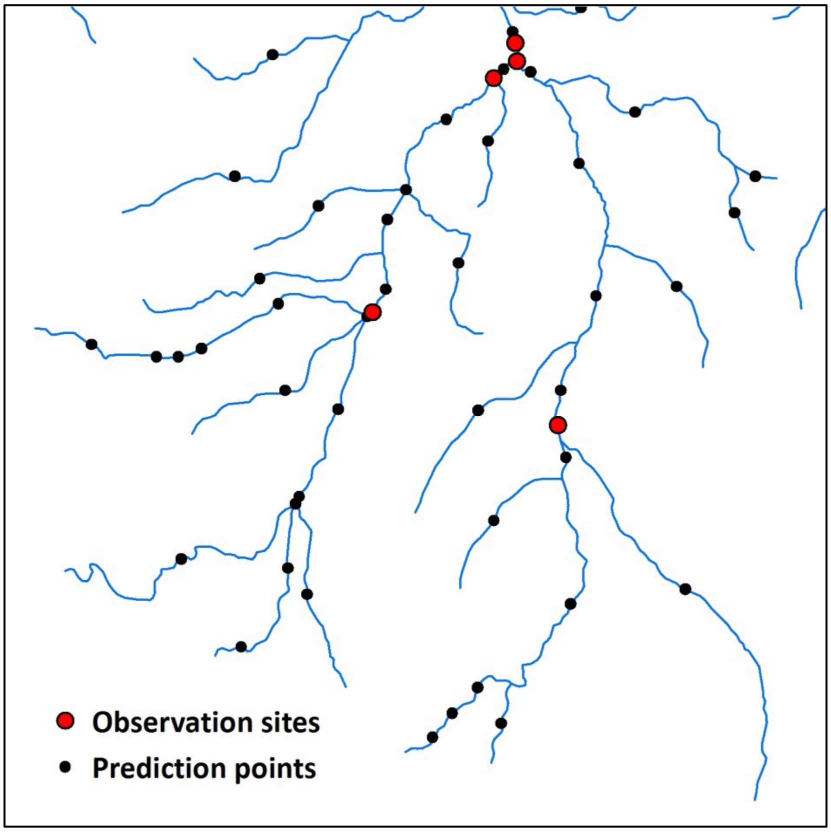

Figure 7. Example observation sites and prediction points. As part of the NSI dataset, prediction points are assigned at the midpoint of each NHDPlusV2 stream segment that contains a unique COMID value.

4.2 Data

4.2.1 NSI Dataset

The *NSI dataset* consists of *streamlines* and *prediction points*. The NHD digital streamlines, or *NHDFlowline* data, were edited to conform to topological constraints required by the SSNbler and SSN2 packages, as described in section 4.0 below. The lines alone are referred to as the *NSI network*. In addition, prediction points (see below) have been generated at the midpoint of each NSI streamline that contain a unique COMID value. The data are available for the contiguous U.S. These data may be freely downloaded from here: https://research.fs.usda.gov/rmrs/projects/national-stream-internet.

4.2.2 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 GIS point locations and must be spatially referenced to the NSI network.

4.2.3 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 location. 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 through the COMID field. Another source of nation-wide covariate data linked to the NHDPlusV2 is the StreamCat dataset: https://www.epa.gov/national-aquatic-resource-surveys/streamcat-dataset.

4.2.4 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 streamline data through a common ID field or spatial join.

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.

4.2.5 Geographic Projections

All data including streamlines, observation points, and prediction points must be in the same geographic projection. In addition, these spatial datasets must be in a rectangular coordinate system such as Albers or UTM to ensure that distances between features are consistent regardless of their location on the globe. The NSI projection parameters are described below Table 1.

Projected Coordinate System	NAD 1983 Contiguous USA Albers
Projection	Albers
WKID	5070
Authority	EPSG
Linear Unit	Meters (1.0)
False Easting	0.0
False Northing	0.0
Central Meridian	-96.0
Standard Parallel 1	29.5
Standard Parallel 2	45.5
Latitude Of Origin	23.0

Table 1. NSI projection parameters.

4.3 Tools

4.3.1 SSNbler

The *SSNbler package* imports, creates, and assembles data needed to fit spatial-statistical stream-network models using the SSN2 R package. Within the SSNbler package, specific functions have been included to 1) generate information describing the topological relationships between features (e.g. lines and points) , 2) check for topological errors; 3) calculate data used to generate hydrologic distances and weighting metrics; and 4) save 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 to correct for topological errors and restrictions, but additional topological errors may be introduced when streams are clipped to study areas. It is therefore important to follow all of the steps in the SSNbler workflow, regardless of which streams dataset is used (https://pet221.github.io/SSNbler/articles/introduction.html).

4.3.2 SSN2

Once the streams data have been properly formatted using the SSNbler toolset, the SSN2 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 SSN2 package can be found at the GitHub link here: https://usepa.github.io/SSN2/ and here: https://usepa.github.io/SSN2/articles/introduction.html.

5.0 A Brief Example Workflow for Spatial Stream Network Modeling with the NSI Dataset

This section provides a set of workflow instructions to assist users in implementing the SSNbler and SSN2 packages with the NSI dataset. These instructions provide general guidance for implementing the tools. Links are provided below for accessing a specific example dataset with additional technical details. A more detailed tutorial entitled *Preparing Input Data for Spatial Stream Network Modeling* is available on this same website.

1) Download the NSI streamlines from the NSI website

The NSI streamlines can be downloaded from here:
https://research.fs.usda.gov/rmrs/projects/national-stream-internet.

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

The NSI data are organized by NHDPlusV2 Vector Processing Units (VPUs) (Figure 6). 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 user's project area boundary. It's advisable to use basin boundaries from a source such as the Watershed Boundary Dataset (WBD) (USGS, 2024) for clipping. The WBD was developed to closely honor the NHDPlusV2 dataset streamlines. Using a watershed boundary rather than a rectangular one will minimize the number of outlets present in the dataset and ensure that most streamlines are network connected.

After clipping, the streamline 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 spatial statistical modelling. 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 streamlines.

2) Check topology

After downloading the NSI streamlines, they should be checked for topological integrity. Although the NSI data were edited to remove most topological errors, it's possible that some features of the streamlines need to be corrected. There are two tutorials that provide guidance for checking and correcting topology, one using ArcGIS Pro software, and the other QGIS.

The ArcGIS Pro document is called *Correcting Topological Errors Using SSNbler and ArcGIS Pro*. The document and tutorial data are available on the SSN2 website here: https://research.fs.usda.gov/rmrs/projects/stars-ssn.

The QGIS document is called *Correcting Topological Errors Using SSNbler and QGIS*. The document and tutorial data are available on the SSN2 website here: https://research.fs.usda.gov/rmrs/projects/stars-ssn.

2) Obtain and process observation data, prediction points, and covariate data

As mentioned previously, observation data are required to fit a SSNM to data collected on stream networks. Covariates are not technically required but including meaningful covariates in an SSNM usually leads to a stronger model fit. Prediction sites are optional. Please see *Preparing Input Data for Spatial Stream Network Modeling* for helpful tips on how to prepare these datasets for use in the SSNbler and SSN2 packages.

The preparation document and accompanying tutorial dataset are available here: https://research.fs.usda.gov/rmrs/projects/spatial-stream-network.

4) Assemble the spatial stream network (SSN) objects in R with SSNbler

The *SSNbler* package is used to import the input data created in the previous step (streamlines, observation points, prediction points, and associated covariate data), which is used to generate information about topological relationships in the network. These topological, spatial, and attribute are assembled into an SSN object, which is used to fit SSNMs using the SSN2 R package

The SSNbler workflow is described in *An Introduction to SSNbler: Assembling Spatial Stream Network (SSN) Objects in R:* https://pet221.github.io/SSNbler/articles/introduction.html.

5) Build spatial stream network models with SSN2

The final step in the process is fitting SSNMs to data stored in the SSN object. Again, the procedures are documented in a tutorial document called *An Introduction to Spatial Stream Network Modeling in R Using SSN2:* https://usepa.github.io/SSN2/articles/introduction.html and https://usepa.github.io/SSN2/.

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 NHDPlusV2 data. Dona Horan reviewed this document.

7.0 References Cited

Dumelle, M., E. Peterson, J. M. Ver Hoef, A. Pearse, and D. Isaak (In Press). SSN2: The next generation of spatial stream network modeling in R. Journal of Statistical Software.

Esri (2023). ArcGIS Pro: Release 3.x. 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., Dewald, T., Johnston, J., Moore, R., and Rea, A., (2012) "NHDPlus Version 2: User Guide". https://www.epa.gov/system/files/documents/2023-04/NHDPlusV2_User_Guide.pdf. [Accessed: 07/02/2024]

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).

Peterson E. E., M. Dumelle, A. R. Pearse, D. Teleki, and J. M. Ver Hoof. (2024) An Introduction to SSNbler: Assembling Spatial Stream Network (SSN) Objects in R. https://pet221.github.io/SSNbler/articles/introduction.html

R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

QGIS (2024). QGIS Geographic Information System, Release 3.x. QGIS Association. GNU General Public License, Boston, MA. http://www.qgis.org

R Core Team (2021-2024) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/. [Accessed: 07/02/2024]

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.

US Environmental Protection Agency (2024) NHDPlusV2 Dataset. https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data. [Accessed: 07/02/2024]

USGS (2024) 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. URL: https://www.usgs.gov/national-hydrography/watershed-boundary-dataset [Accessed: 07/02/2024]

Wieczorek M. and LaMotte A.E. (2010). Attributes for NHDPlus Catchments (Version 1.1): Basin characteristics, 2002-2010. U.S. Geological Survey. https://catalog.data.gov/dataset/attributes-for-nhdplus-catchments-version-1-1-for-the-conterminous-united-states-average-s

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 were originally 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 data may now be used with the SSNbler and SSN2 R packages (Peterson et al., 2024 and Demele et al., 2023). The purpose of the toolsets is to model biological and physical characteristics of stream networks using digital streamline data. To implement the tool sets 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 streamline network so that users may utilize the data directly with the spatial statistical tools. More information about SSNbler and SSN2 may be obtained at:
http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml
Update URL

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 (2024) and the U.S. Geological Survey (USGS)]. This modified version was originally 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 data may now be used with the SSNbler and SSN2 R packages (Peterson et al., 2024 and Demele et al., 2023). 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.

Dumelle, M., E. Peterson, J. M. Ver Hoef, A. Pearse, and D. Isaak (In Press). SSN2: The next generation of spatial stream network modeling in R. Journal of Statistical Software.

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).

Peterson E. E., M. Dumelle, A. R. Pearse, D. Teleki, and J. M. Ver Hoof. (2024) An Introduction to SSNbler: Assembling Spatial Stream Network (SSN) Objects in R. https://pet221.github.io/SSNbler/articles/introduction.html

US Environmental Protection Agency (2024) NHDPlusV2 Dataset. https://www.epa.gov/waterdata/get-nhdplus-national-hydrography-dataset-plus-data.

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. Update URL

Credits:

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

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).

USDA 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 ▲