USGS
science for a changing world

nationalatlas.gov
Where We Are

Transforming NHD to Global Map/National Atlas Hydro—Generalization From 1:100,000 to 1:1,000,000

by Zachary D. Wilson and Robin H. Gary, U.S. Geological Survey, Texas Water Science Center

Abstract

The Global Mapping Project is an international effort to develop and integrate 1:1,000,000-scale (1:1M) geospatial data that will facilitate environmental research at spatial scales ranging from continental to global. Hydrographic drainage datasets that include streams, waterbodies, and canals facilitate environmental research by establishing hydrologic connectivity between regions. This poster explores the various methods used in the production of the 1:1M hydrographic dataset for the United States. Methods tested include network hierarchy analysis based on stream name, classification of 1:1M scanned map data, calculation of zonal statistics within buffered streams, generation of hydrologic derivatives for 30- and 90-meter resolution digital elevation models (DEMs), and thinning of the network on the basis of the National Hydrography Dataset (NHD) Plus Thinner Code. Distinct inadequacies in each approach indicate that combining aspects of several methods is most effective. The final method involves an intelligent downstream trace algorithm and ancillary datasets to select stream segments for the 1:1M hydrographic dataset.

Objective

As part of an effort to fulfill the aims of the Global Mapping project and to recompile the National Atlas at 1:1M-scale, the U.S. Geological Survey (USGS) Texas Water Science Center, in collaboration with the National Atlas program, is producing a 1:1M hydrographic dataset. Our objective was to develop a methodology for producing a 1:1M-scale hydrographic dataset for the entire United States, including Alaska, Hawaii, Puerto Rico, and the U.S. Virgin Islands, as part of the Global Mapping Project and the National Atlas 1:1M-scale recompilation effort. We have attempted to create an unambiguous methodology with a high degree of repeatability. Therefore, we have used only publicly available datasets that cover the entire United States.

Background

The accuracy, detail, and connectivity available in the 1:100,000-scale (1:100K) NHD provides an appropriate and reliable base for the creation of a 1:1M hydrographic dataset. The 1:100K NHD data model can be generalized by referencing ancillary small-scale data that cover the United States and its territories.

Several readily available, small-scale, nationwide hydrographic datasets exist. The National Atlas offers a 1:2,000,000-scale (1:2M) Streams and Waterbodies of the United States dataset. The National Imagery and Mapping Agency (NIMA), now called the National Geospatial-Intelligence Agency (NGA), published the VMAP0 hydrographic dataset at 1:1M-scale using data collected from 1972 through 1992. The USGS offers the Elevation Derivatives for National Applications, which includes a synthetic stream network, produced from 30-meter DEMs. The International Map of the World (IMW) series, produced by numerous organizations, provides coverage of North America with 1:1M-scale maps compiled from the 1920s to the 1970s. Although none of these sources individually provides the detail and accuracy needed to fulfill the requirements of the project, they combine to form an appropriate indicator of stream network density for compilation of the 1:1M-scale hydrographic dataset.

Links

National Atlas of the United States – http://nationalatlas.gov

International Steering Committee for Global Mapping – http://www.iscgm.org

U.S. Geological Survey – http://www.usgs.gov

GeoGratis (Natural Resources Canada) – http://geogratis.cgdi.gc.ca

INEGI (Mexico National Geography Institute) – http://www.inegi.gob.mx

USGS Texas Water Science Center GIS Group – http://tx.usgs.gov/GIS

Methods Testing: Trial and Error

The final method is the result of a process of trial and error where several methods for data model generalization were tested and ultimately synthesized.

1. Utility Network Analyst

The ArcGIS Utility Network Analyst allows the user to trace downstream on any dataset which contains a geometric network. The geometric network stores the directionality of each line feature within the feature class. By placing a flag on a headwater reach, one can trace flow downstream to the outlet of the network. The results of the trace can be converted to a selection of features, the attributes of which can be updated. In our test case, we selected streams for the 1:1M dataset on the basis of this trace downstream procedure. Unfortunately, in areas with low relief and areas with braided streams, all stream reaches were selected instead of only the main flow path.

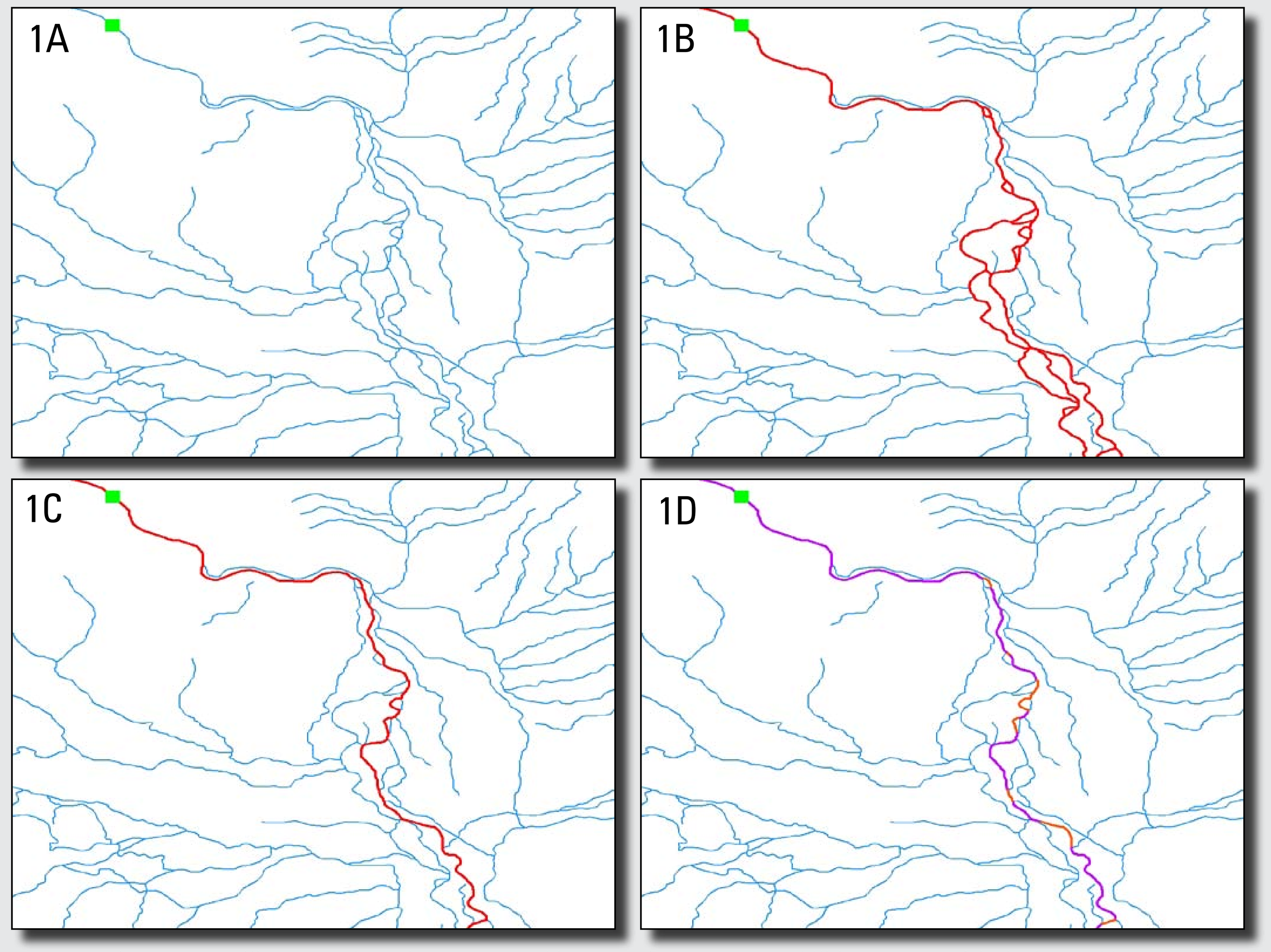

Figure 1. The ArcGIS Utility Network Analyst (UNA) tool has trace downstream capabilities. In parts of the hydrographic network with braided streams (1A), the UNA trace downstream function selects all possible reaches (1B) resulting in a stream network too dense for the target 1:1M-scale. Our feature selection algorithm finds a single path through braided streams (1C) based on several criteria. During processing, the reach attributes are updated to reflect the specific criterion used to include each reach in the 1:1M dataset (1D).

2. GNIS Name Hierarchy

This method counts the number of stream segments with the same name, as indicated by the Geographic Names Information System (GNIS) attribute, to determine whether to include them in the 1:1M dataset. It aims to establish a hierarchy such that small streams with only one or two named segments are not selected. A relative threshold value for the stream segment GNIS count is estimated by comparing density to ancillary maps compiled at 1:1M; however, this limits repeatability because the threshold value is subject to interpretation. Results contain breaks in network connectivity and inconsistent stream density. This could be because that this method relies solely on GNIS data, which often is more dense in areas with larger population and might favor complicated stream network areas because they have smaller named segments.

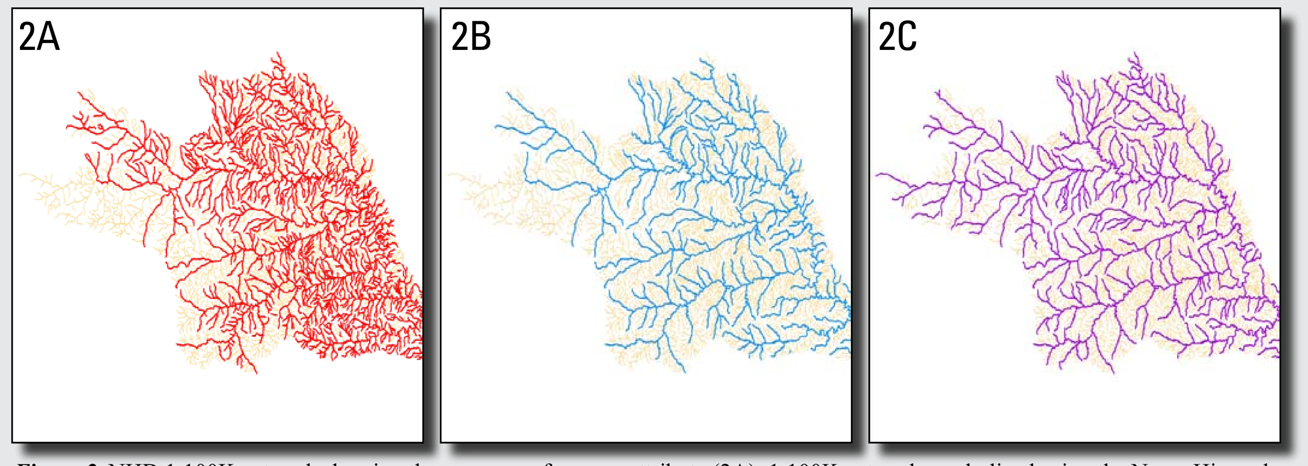

Figure 2. NHD 1:100K network showing the presence of a name attribute (2A); 1:100K network symbolized using the Name Hierarchy method (2B); and 1:100K network showing the final method results (2C).

3. Hydrologic Derivatives

Zonal statistics can be calculated using a polygon feature class and raster data. We created a buffer for each reach in the 1:100K NHD streams feature class and calculated statistics from USGS Elevation Derivatives for National Applications (EDNA) flow accumulation data derived from 30-meter DEMs. This procedure used zonal statistics to indicate which reaches were most important hydrologically. This approach successfully selected reaches that correlated with the 1:1M ancillary datasets in the middle parts of watersheds but had limited success finding the important headwater reaches and reaches in areas with low relief.

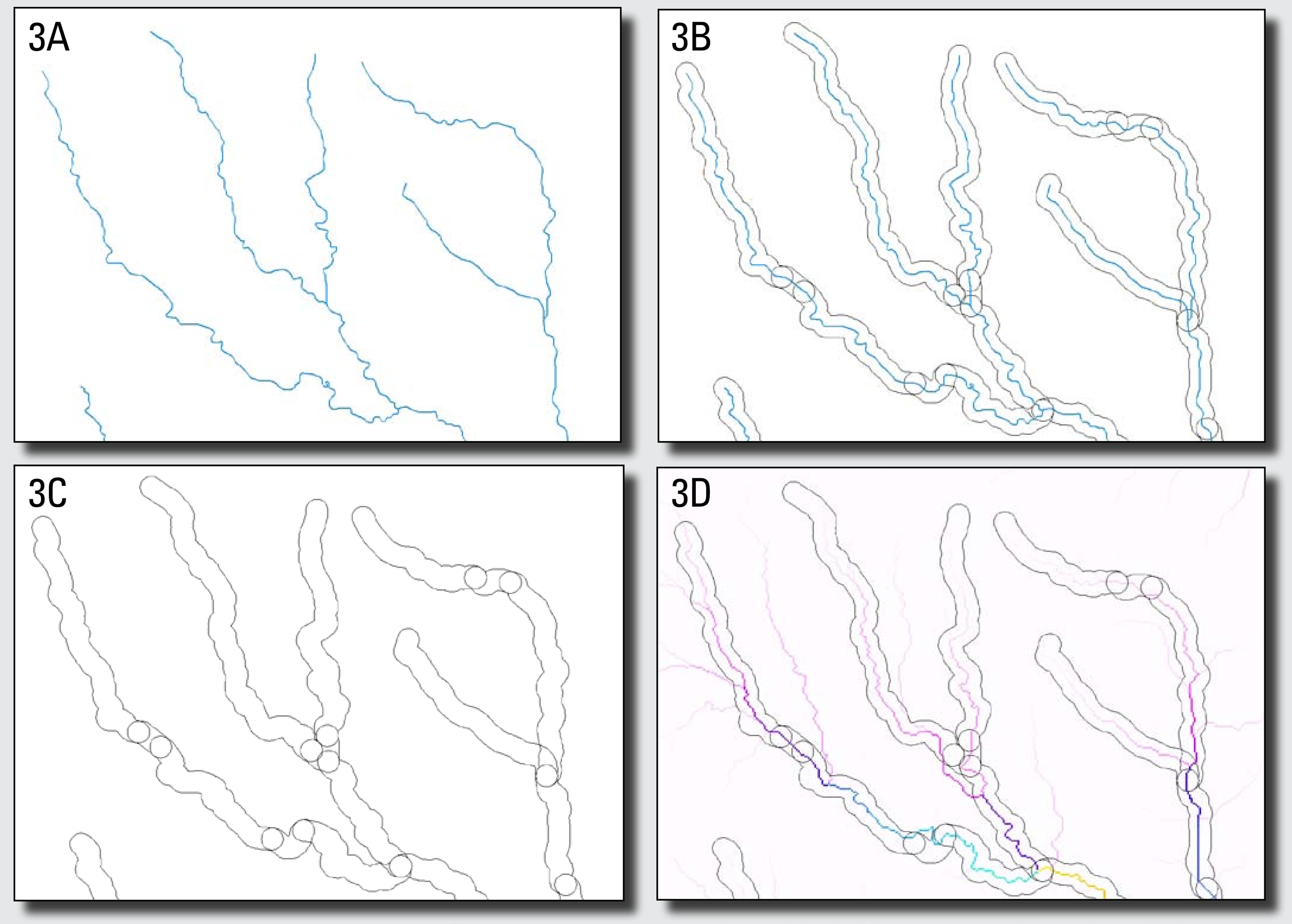

Figure 3. NHD 1:100K streams (3A) were buffered by 500 meters (3B). The resulting buffers (3C) were then used to calculate zonal statistics based on EDNA flow accumulation data (3D).

4. NHD Plus Thinner Code

The U.S. Environmental Protection Agency, Horizon Systems Corporation, and USGS are producing NHD Plus, a networked hydrographic dataset with value-added attributes (VAAs) based on 1:100K NHD. Among the VAAs is a Thinner Code attribute which is designed to allow the user to progressively thin the representation of the hydrographic network by selecting among seven values (0–6). If values 0–6 are represented in the network, all reaches will be present. Starting with the Thinner Code value of 6, the network can be thinned by eliminating more values from the cartographic representation. In our project, the goal was to select reaches in this way to represent a 1:1M-scale network, however the NHD Plus Thinner Code values were not consistent enough across the entire United States to allow us to use them as the sole source for the feature selection phase of the stream network generalization.

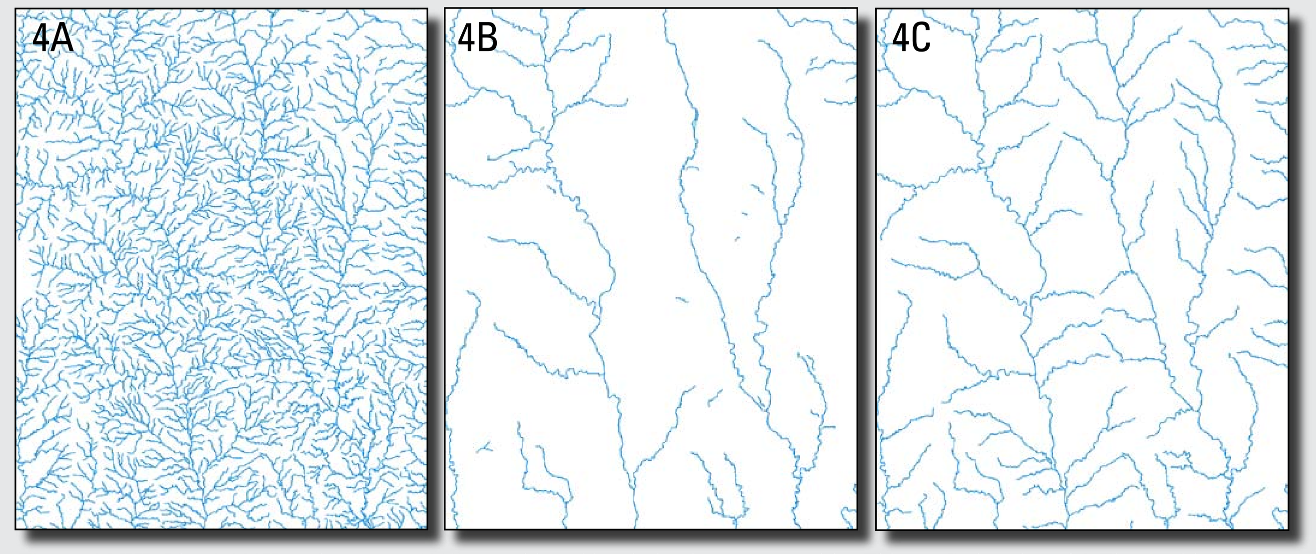

Figure 4. NHD Plus 1:100K-scale streams (4A) were thinned by elminating streams with a Thinner Code value of 6. The resulting stream network density (4B) was insufficiently dense when compared to the target 1:1M network density (4C).

Final Method

The final method can be divided into two main steps: data model generalization (specifically, feature selection) and cartographic generalization. The final data model generalization step is a synthesis of procedures developed by testing several different models (figures 1 - 4), and the cartographic generalization step uses established simplification algorithms.

Data Model Generalization (Feature Selection)

Headwater reaches are selected on the basis of the National Atlas (1:2M), VMAP0 (1:1M), and IMW (1:1M) hydrography datasets, which serve as primary indicators of 1:1M-scale stream density. The USGS EDNA data and Digital Orthophoto Quarter Quads (DOQQs) are used where National Atlas, VMAP0, or IMW are ambiguous about which reaches to select.

The downstream trace tool applies an algorithm to the indicated headwater reaches to determine the preferred downstream route on the NHD geometric network by analyzing stream name, flow direction, and feature type (for example, stream, canal, or artificial path). If there is only one downstream reach, it is included in the 1:1M dataset. If there is more than one downstream reach, a set of criteria is used to decide which of the downstream reaches should be included in the 1:1M dataset. Priority is given to named segments, segments coded as streams, and segments with flow direction.

When all possible reaches have been selected, the network is examined for connectivity. Connectivity is checked visually and by tracing upstream on the proposed 1:1M network to identify reaches that need to be added manually.

Cartographic Generalization

All reaches indicated in the feature selection process are exported as a new feature class, and then simplified using the Bend Simplify algorithm and a tolerance of 500 meters. The feature class is further simplified using the Douglas-Peucker algorithm with a 1-meter tolerance to remove any extraneous vertices. The new, simplified and generalized dataset is then tested for network connectivity. A new regional geometric network is created, and upstream traces are used to search for missing or disconnected reaches.

Acknowledgments

We are indebted to Jay Donnelly of the National Atlas of the United States for his invaluable guidance and constructive criticism. We also acknowledge the contributions of John Hutchinson (USGS, Sioux Falls, S.D.); Anna Regan and Andrew Murray (Natural Resources Canada, Ottawa, Canada); Jose Luis Ornelas (INEGI, Aguacalientes, Mexico); Sue Greenlee and Kris Verdin (USGS, Sioux Falls, S.D.); and Christina Bowden, Christy-Ann Archuleta, Case Watkins, Jean Parcher, Brian Reece, Daniel Pearson, William Asquith, and Peter Bush (USGS, Austin, Tex.).

Contact Information

Zachary Wilson (zdwilson@usgs.gov) and Robin H. Gary (rhgary@usgs.gov)
USGS Texas Water Science Center
8027 Exchange Drive
Austin, Texas 78754