Stratified Database Pruning to Support Local Density Variations in Automated Generalization of the United States National Hydrography Dataset

Lawrence V. Stanislawski, Science Applications International Corporation
U.S. Geological Survey, Center of Excellence for Geospatial Information Science
1400 Independence Road, Rolla, MO 65401, USA
lstan@usgs.gov

Barbara P. Buttenfield, Department of Geography
UCB 260, University of Colorado—Boulder, Boulder, CO 80309, USA
babs@colorado.edu

Michael P. Finn, U.S. Geological Survey
Center of Excellence for Geospatial Information Science
Denver Federal Center, Box 25046, Mail Stop 510, Denver, CO 80225, USA
mfinn@usgs.gov

Keven Roth, GIS Consultant, Former USGS NHD Coordinator
Victor, MT 59875, USA
kevenroth@yahoo.com

Abstract

Introduction: The U.S. Geological Survey's (USGS) Center of Excellence in Geospatial Information Science (CEGIS) is conducting generalization research in cooperation with the University of Colorado—Boulder and Pennsylvania State University to support display and delivery of *The National Map* and other USGS geospatial data at multiple scales. This paper focuses on generalization of the National Hydrography Dataset (NHD).

The process of generalizing detailed geospatial data to render less detailed data and cartographic products can be enhanced substantially with a multiple representation database (MRDB). One method to build a MRDB is to assign a level of importance, or prominence, to each feature in a theme's base layer containing the most detailed feature representations. Subsequently, a hierarchical set of layers can be sub-selected from the base layer and stored in the database. A MRDB built in this manner can enhance cartographic generalization if the level of prominence for each sub-selected layer is appropriately tailored for a specific range of map scales. Ensuing generalization operations and symbology changes associated with base-map generation can be efficiently focused on layers required for each desired scale range.

The NHD is a comprehensive vector data model representing surface-water features of the United States. The NHD includes an address system, known as reach codes, for

linking to other environmental databases. The NHD database includes three levels of detail—medium, high, and local resolutions. Features in the medium-resolution layer are compiled from 1:100,000-scale (100K) source hydrographic data. The local-resolution layer includes 1:4,800-scale data only for the state of Vermont.

Initially, the high-resolution (HR) NHD layer was compiled from 1:63,360-, 1:24,000-, and 1:20,000-scale data for Alaska, the conterminous 48 states, and Puerto Rico and the Virgin Islands, respectively. During recent years (2007-2009) maintenance has focused on correcting features and collecting more detailed features for the HR layer. In some cases, 1:4,800-scale source data are being included in the HR layer. Consequently, the HR NHD layer is a multi-resolution, or multi-scale, layer generally with the most current and accurate available data. Development of the multi-scale HR layer has been conducted, anticipating that database pruning will satisfy the needs to generate uniformly scaled data.

Objectives: Objectives of this research are to develop methods to sub-select, or prune, features from the multi-scale HR NHD layer to automate generating a MRDB and simplify, or further generalize, remaining features for cartographic display. Methods should maintain hydrographic network connectivity and local density variations that typify physiographic or climate variations.

Methodology: *:* A four-subbasin region of HR NHD data in Iowa, having obvious natural network density variations, was pruned to four smaller scales—100K; 1:500,000 (500K); 1:2,000,000; and 1:10,000,000. Methods employ a stratified database pruning approach that partitions HR NHD data based on localized network densities and prunes features to densities appropriate for smaller map scales based on reach codes and upstream drainage area (UDA) estimates. Database enrichment, partitioning, pruning, and subsequent generalization operations are completed through automated processes.

A coefficient of line correspondence (CLC) with the 100K NHD layer as the reference validates the pruning process for the 100K level of detail. The CLC is computed between two linear datasets and estimates the ratio of the length of matching features to the length of matching features, plus the length of omission and commission errors.

In addition, numbers of confluence-to-confluence segments, associated lengths, and stream orders are compared between linear features in pruned and validation data sets at associated scales. Similar comparisons are reported for polygon features.

Results: For the tested region, stratified density pruning produced reasonably good results that maintain local density variations where needed and mimic data conditions existing in other sources of available hydrographic data. The CLC is 0.88 when HR NHD pruned to 100K is compared to 100K NHD. Automated stratified pruning to 500K or smaller scales produces some network discontinuities near partition boundaries, but stratification may not be needed for these smaller scales.

Conclusions: Although implementation on the dynamic, multi-scale HR NHD layer will be a challenge, results indicate stratified pruning in conjunction with tailored generalization and symbolization operations can achieve the goals of this research and meet the cartographic generalization, mapping, and essential hydrographic analysis needs of the user community. It appears that the pruning process can be guided by the Radical Law, but with refinements for large and small scale data conditions. Future research should eliminate the creation of isolated sub-networks during stratification and further compare generalization operations performed after data pruning.