Density-Stratified Thinning of Road Networks to Support Automated Generalization for *The National Map*

L. V. Stanislawski[1], M. Briat[2], E. Punt[2], M. Howard[1], C. A. Brewer[3], B. P. Buttenfield[4]

[1]Center of Excellence for Geospatial Information Science, U.S. Geological Survey, 1400 Independence Road Rolla MO 65401, USA
Email: lstan@usgs.gov

[2]Esri, Inc. 380 New York St, Redlands CA 92373, USA
Email: [mbriat, epunt]@esri.com

[3]Department of Geography, Pennsylvania State University, University Park PA 16802, USA
Email: cbrewer@psu.edu

[4]Department of Geography, University of Colorado-Boulder, Guggenheim 110, 260 UCB Boulder CO, 80309, USA
Email: babs@colorado.edu

1. Introduction

One of the primary goals of cartographic generalization is the appropriate reduction of content from a larger to a smaller scale representation. When working with vector geospatial data, scale reduction typically involves removal of less important, or less prominent, features. Various strategies and techniques exist for thinning features from road or hydrographic networks to produce reduced scale versions of the data (Buttenfield et al. 2011, Savino et al. 2010, Stanislawski 2009, Thomson and Brooks 2007, Touya 2007). Commercial software packages, such as Environmental Systems Research Institute, Inc. (Esri) ArcGIS® (version 10 and later), include tools that automate the thinning of road networks (Punt and Watkins 2010, Briat et al. 2011); but the quality and capabilities of these tools have not been thoroughly assessed as yet.

The United States Geological Survey (USGS) Center of Excellence for Geospatial Information Science (CEGIS) is researching and developing methods and tools for multi-scale display and delivery of geospatial data in *The National Map*, which the USGS develops and maintains. *The National Map* includes a hydrographic network in the National Hydrography Dataset (NHD) and a transportation network in the Best Practices (BP) Data Model (USGS 2006). Relying on the geoprocessing functionality available through Esri's ArcGIS® framework, CEGIS developed generalization tools that rely on density-stratification to thin the NHD flow network (Stanislawski and Buttenfield 2011a, Stanislawski and Savino 2011). Although a bit more complicated than unstratified approaches, density-stratified thinning maintains natural variations in local stream density that reflect the wide range of geographic conditions that exist within the United States.

Following concepts similar to those applied for hydrography, CEGIS, in collaboration with Esri, is investigating the use of Esri's Thin Road Network tool for automated generalization of the BP transportation data of *The National Map*. The BP database currently (2012) includes three sources of transportation data. Data compiled in 2011 by TomTom® North America, Inc. and by the U.S. Forest Service for U.S. Department of Agriculture forest areas are displayed on US Topo 1:24,000-scale (24K) topographic maps. Data compiled in 2008 by the U.S. Census Bureau are distributed to users to support their mapping or analysis needs. Although these data sources are compiled to slightly varying scales, these transportation datasets typically are used by the USGS for display and analysis at 24K. For purposes of this paper, we will refer to these data as 24K road data.

Preliminary tests indicate that using the Thin Road Network tool with a uniform length tolerance tends to homogenize road network density at smaller scales, thereby reducing density variations reflective of geographic conditions. In an attempt to better control density variations while using the Thin Road Network tool, this paper evaluates generalization results produced by the tool in comparison to results expected from the Radical Law (Töpfer and Pillewizer 1966). Several thinning operations are performed on 24K road data to find a version adequate for 1:100,000-scale (100K) mapping. Resulting line densities are compared for urban and rural density classes for two study areas. Results provide a quantitative assessment of Esri's road thinning tool and demonstrate that density-stratified thinning can improve the road thinning process, making it a more robust solution that can support a greater variety of map uses and geospatial applications.

2. Methods

2.1 Test Data

Thinning tests were performed on the BP 24K road data for two study areas. The first covers four NHD subbasins in central Iowa, a region that is about 24,000 square kilometers (km^2) in a largely rural agricultural section of the Midwestern United States, which includes the Des Moines Metropolitan Statistical Area (MSA), which has a 2011-estimated population of 580,000 (U.S. Census Bureau 2011). This same four-subbasin study area was evaluated in an earlier NHD generalization project by CEGIS (Stanislawski et al. 2009). The 24K road data includes 109,265 features totalling nearly 36,000 kilometers (km) of roads.

The second study area is in the MSA that includes Atlanta, Georgia with an estimated population of about 5.4 million (U.S. Census Bureau 2011). The study area covers more than 10,000 km^2. The 24K road data for this area includes 393,920 features totalling nearly 49,000 km of roads.

Results for only the second study area, Atlanta, are discussed in this paper. Analysis of the first study area is complete, and results will be presented at the workshop.

2.2 Pre-Processing

The ArcGIS® (version 10.1) Thin Road Network tool requires input road data with feature geometries that are topologically adequate for a road network system to ensure efficient data processing and proper results. Pre-processing of feature geometry to prepare data for the Thin Road Network tool included projecting the data from geographic coordinates to the North American Albers Equal Area Conic projection, removing coincident features, converting multi-part to single-part features, and splitting features at all intersections.

Complete attribution was not available on all features within the road network database. That is, the road classification field was not populated on all features in the original test data. Therefore, aside from aforementioned adjustment of feature geometries, pre-processing of feature attribution included transferring road name attributes to retained features when removing coincident features, populating road classification where needed based on feature names, and calculating an index of road importance based on road classification values. Programs were developed through Arc

Macro Language (AML) and Python with ArcGIS® geoprocessing functions to automate the pre-processing steps.

2.3 Road Thinning Tests

The ArcGIS® Thin Road Network tool identifies relatively insignificant road segments, which, when removed from the display, result in a simplified network suitable for smaller-scales. The connectivity and general morphology of the road pattern are maintained, along with the integrity of navigable routes. The resulting collection reflects the relative importance, significance, and density of input features. Importance is determined through the hierarchical road classification (e.g., Interstate, State Route, local road, etc.) present on the input data as attributes. Feature significance is determined through topological connectivity, and by a feature's participation in long routes, or *itineraries* across the extent of the data. Features that are part of long itineraries are deemed more significant than those required only for local travel. Density is computed similar to any density metric, as a ratio of the length of a street segment to its associated area.

The tool requires three parameters. One parameter identifies the road classification field in the attribute table, which establishes importance. A second parameter is the attribute field that will store the binary (1 or 0) results of the tool, where 1 indicates the feature should not be displayed (i.e., invisible) and 0 means the feature should remain visible. The tool does not actually delete any road segments; rather it populates this "invisibility" field to mark insignificant road features that should not be displayed in the thinned network. The degree of thinning that takes place—and thus the relative density of the resulting network—is driven by the third user-defined, numeric parameter named Minimum Length, which is a value (in map units or page units) that is equal to the shortest segment that is visually sensible to show at final scale.

The relation between minimum length and desired network density is not always intuitive, and depends on road sinuosity, concentration of local roads, and the overall connectivity pattern across the network. These factors obviously will vary from one dataset to another, similar to variations in geography and landscape type, which require tailoring parameters for other types of generalization (Buttenfield et al., 2011; Touya et al, 2010). Road network thinning must be considered as a modeling process rather than a deterministic tool for which a single set of input parameters will be adequate for any geographic region.

After pre-processing the data for the Thin Road Network tool, road data for each study area were subdivided into line-density strata using a raster partitioning process that subdivides line data by means of user-defined density class breaks (Stanislawski and Buttenfield 2011b). In addition, the partitioning process ensures that each density polygon is larger than a user-defined minimum. The Atlanta data were subdivided into three line-density strata, using density class breaks of less than 2.5, between 2.5 and 4.5, and greater than 4.5 km/km^2, with the minimum area tolerance of 45 km^2. The choice of density strata is scale dependent. These selected target density class breaks are approximate values for rural, suburban, and urban areas based on an evaluation of the BP 24K road data for the coterminous US. The stratification process also enriches the feature class by assigning a strata code to each road feature.

Given the fixed area of each study site, a target average road density estimate was determined for 100K mapping for each density stratum in the study areas. The Radical Law (Töpfer and Pillewizer 1966) was applied to the average density values for each stratum to determine the target 100K density estimates. For the purpose of this calculation, the BP road data are considered as 24K data because the USGS uses these data for the 24K US Topo map series. Subsequently, each dataset was thinned multiple times using different Minimum Length values in an effort to find the value that produces a thinned density for each strata that most closely matches the associated Radical Law estimate. The Atlanta data were thinned using minimum lengths of 0.5, 1, 1.5, 2, and 2.5 km.

To validate thinning results, a line-density raster dataset with a 300-meter (m) cell resolution was generated separately for the thinned road networks of each study site. In a similar manner, 300-m resolution line-density raster datasets were compiled from 100K road data available in USGS digital line graph (DLG) format. The 100K DLG data were compiled from maps produced between 1979 and 1981 for the Atlanta site (these are markedly older sources than the 24K road data). Evaluation of the spatial distribution of differences between the 24K road data thinned to 100K and the 100K DLG data was done by subtracting the 100K DLG density raster from the density raster for the 24K road data that was thinned to 100K.

3. Results of Thinning Atlanta Roads to 1:100,000-scale

The density strata generated for the 24K road data in the Atlanta study area are shown in Figure 1(b). The raster partitioning process using the defined class breaks formed density strata having average densities of 1.65, 3.35, and 7.20 km/km^2 for the rural, suburban, and urban classes, respectively (Figure 1).

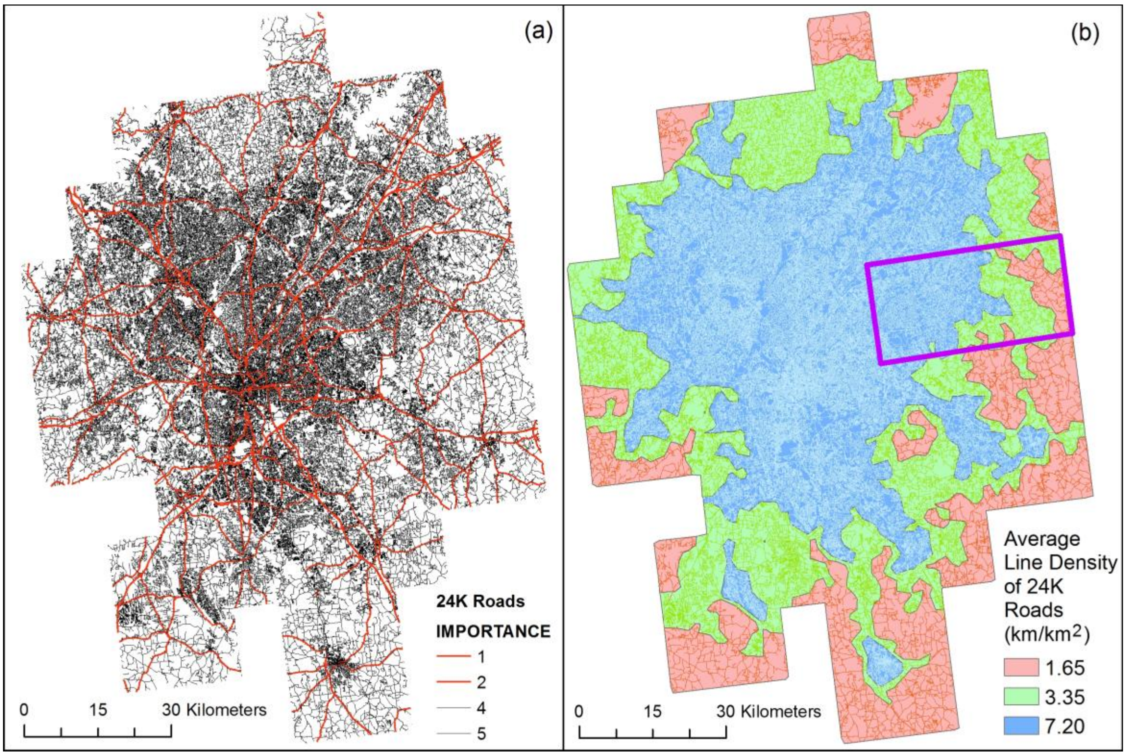

Figure 1. 2011 24K road data for metropolitan Atlanta, Georgia (a), and density strata generated from the raster partitioning process using class breaks of less than 2.5, 2.5 to 4.5, and greater than 4.5 kilometer per square kilometer (km/km^2), and a 45 square kilometer minimum polygon tolerance (b). Values shown in the panel b legend are average densities that result from partitioning. Purple inset box in (b) is the extent of area displayed in Figure 3.

Table 1 shows the densities that result from thinning the Atlanta roads using various minimum length parameter values with the Thin Road Network tool. The target 100K densities estimated through the Radical Law for the rural, suburban, and urban strata are 0.81, 1.64, and 3.53 km/km^2, respectively. Given the minimum length values tested, thinning the 24K road data using 2.5-, 1.5-, and 1-km minimum lengths achieves densities that match the target 100K densities most closely in the rural, suburban, and urban strata, respectively.

Table 1. Summary of thinned road network densities produced by the Thin Road Network tool from 1:24,000-scale road data within rural, suburban, and urban density strata (classes) within metropolitan Atlanta area in Georgia. Percentage values highlighted in light pink indicate the minimum length that produced the density which best matches the target 1:100,000-scale density estimated through the Radical Law. [diff., difference; km, kilometers; km/km^2, kilometers per square kilometer]

Density Class	Rural		Suburban		Urban	
	Density (km/km^2)	Percent diff. from Radical Law	Density (km/km^2)	Percent diff. from Radical Law	Density (km/km^2)	Percent diff. from Radical Law
Density Class Break	less than 2.50		2.50 to 4.50		more than 4.50	
Average Density of class at 1:24,000	1.65		3.35		7.20	
Radical Law Density Estimate for 1:100,000	0.81		1.64		3.53	
Minimum Length 0.5 km	1.53	88.83	2.54	54.88	5.03	42.59
Minimum Length 1 km	1.32	63.04	1.94	18.06	3.74	5.89
Minimum Length 1.5 km	1.20	47.59	1.67	1.37	3.18	-9.71
Minimum Length 2 km	1.13	39.59	1.54	-6.26	2.88	-18.34
Minimum Length 2.5 km	1.09	34.49	1.47	-10.38	2.69	-23.85

A stratified thinning selection of the 24K road features to most closely meet the Radical Law 100K target densities is shown in Figure 2. The density difference map between the density of the 100K DLG roads and the thinned 24K roads is shown in Figure 2b. The average density difference determined from the mean of the raster difference dataset is 0.14 km/km^2, indicating the 100K DLG roads are slightly more dense than the thinned roads overall.

Comparison data sets are lacking for accurate benchmarking of thinned data results, so Figure 2b results show as much about urban change as mapping practices. The difference map shows DLG data to be substantially denser in urban areas than the thinned roads dataset. This suggests that either the archived 100K DLG roads are not properly thinned in urban areas, or the Radical Law target estimate results in over-thinning the urban area roads, or some combination of these two. In selecting the 24K features to retain on 100K USGS topographic maps, compilation standards for roads indicate that entire street patterns, except alleys, were retained in urban areas; only unimproved roads, trails, and special service roads were eliminated by minimum length criteria (USGS 1985). Thus, a vast majority of 24K features were retained for 100K mapping, particularly in urban areas, which is consistent with displayed raster difference values. Low difference values around the central urban area likely indicate urban sprawl that has occurred in the 30 years since the DLG source maps were made. In fact, the Atlanta MSA population has grown by more than 2 million since the 1990 census.

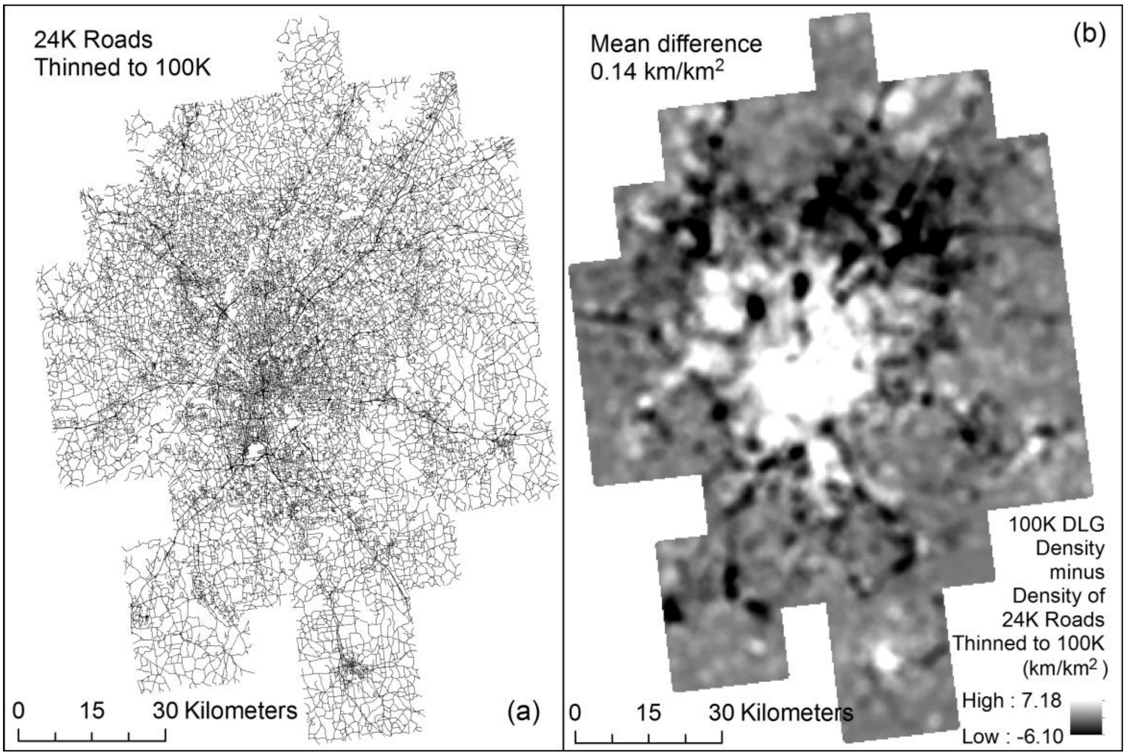

Figure 2. Atlanta area BP 24K road data thinned from 1:24,000 to 1:100,000 through stratified thinning (a). Difference grid comparing road densities for 1:100,000-scale DLG data and the 1:100,000-scale thinned BP road data (b).

Three versions of the roads—unthinned, thinned using a 2-km minimum length, and the stratified thinning results—are shown in Figure 3. The dark grey background in the figure shows the urban density stratum (thinned using 1-km minimum length), the medium grey is the suburban stratum (thinned using 1.5-km minimum length), and the light grey is the rural stratum (thinned using 2.5-km minimum length). These figures are reduced to 1:250,000-scale for visual comparison, though the thinning levels are intended for use at 100K. The stratified thinning result appears to better maintain density transitions between strata than the non-stratified thinning, which tends to homogenize density between strata. A quantitative comparison of this is underway, along with an evaluation of the number isolated features that are generated around partition boundaries from the stratified thinning process.

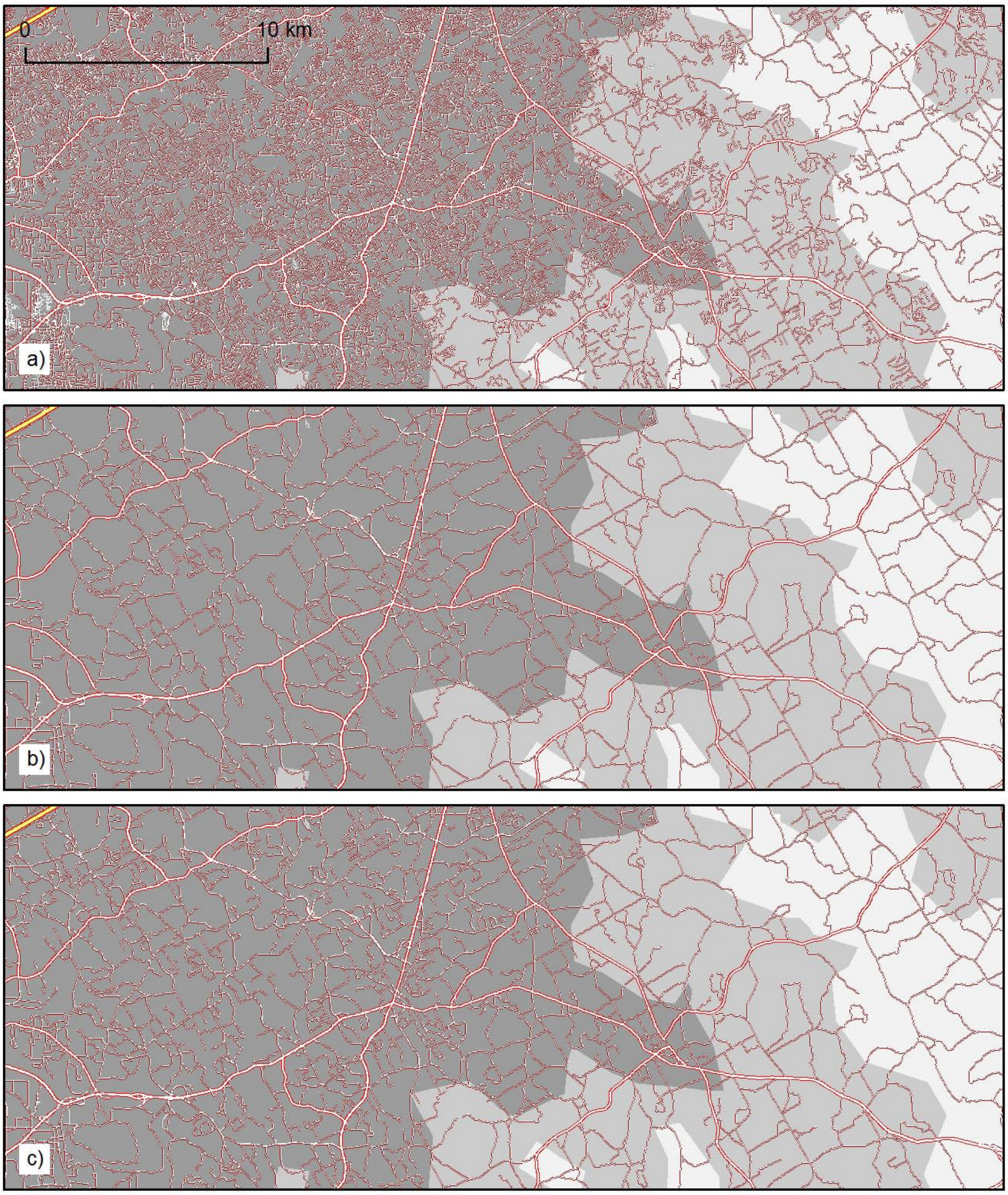

Figure 3. Comparison of a part of the Atlanta study area showing a) 24K road data, b) 24K road data thinned using a uniform (2-km) minimum length, and c) 24K road data thinned to the three densities that best match Radical Law densities.

4. Discussion

The goal of the work reported here is to improve on road thinning through scale by retaining relative differences in road densities between urban and rural areas. The research reported here is driven by the recognition that road networks in urban areas tend to be significantly denser than road networks in suburban or rural areas, and local density differences should be preserved to achieve appropriate generalization results.

This paper presents an evaluation of the latest released version of the Thin Road Network tool, which at the time of this study was in ArcGIS® version 10.1 Pre-Release. The tool uses a simulated annealing technique that retains connectivity and establishes a relative hierarchy of feature importance across a network's spatial extent based on each feature's contribution to longer or shorter linear routes, or itineraries. Road segments that participate in shorter itineraries are considered less significant and more likely to be eliminated. Although the tool strives to retain a representative sense of relative density, character, and connectivity, the tested version of the tool tends to homogenize density somewhat as the amount of scale difference between the compiled source and target scale products increases. This homogenization detrimentally diminishes differentiation between sparsely populated rural areas and more densely populated towns, suburbs, and cities.

A density partitioning method originally developed for hydrographic networks was applied to the road networks, establishing three density classes for the Atlanta region (urban, suburban and rural). Radical Law estimates of network density were computed for each class, to thin roads to a target scale of 100K, and these estimates were used to validate a series of thinning parameters. Results demonstrate that (for both study areas), the urban, suburban and rural density partitions require different minimum length parameter values to meet the Radical Law estimates. To be clear, test results show that no single minimum length can adequately thin urban, suburban and rural road networks equally well. This demonstrates the need for local density stratification for road networks, similar to requirements for hydrographic networks.

Visual inspection of the Atlanta data also indicates that stratified thinning preserves local density differences in thinned data. Additional analysis in preparation for the workshop presentation will quantify this, building density grids for the non-stratified thinning results and differencing them from the 100K DLG benchmark. Comparison of this difference grid with the stratified density difference grid will provide metrics for both study areas to further explore the performance of road network thinning in specific geographic conditions.

Future research with the Thin Road Network tool will focus on methods to streamline required pre-processing steps, minimize creation of isolated features around density class boundaries, and automatically assess navigability of retained features. In addition, a parallel implementation of the tool with density stratification will be developed and tested for use on the road data for the entire United States. However, a primary consideration for this work and for the U.S. National Geospatial Technical Operations Center is to better define standard target densities or feature-type hierarchies for transportation features on the various scales of USGS map products.

Acknowledgements

The collaboration between public sector, private sector, and academic researchers to evaluate proprietary software tools marks an important milestone in the North American generalization community. David Watkins of Esri helped coordinate and assist with this collaborative research effort.

The Penn State and CU-Boulder research was funded by CEGIS of the USGS through the Department of Interior Cooperative Ecosystem Studies Unit (CESU). Paulo Raposo of Penn State made thoughtful contributions to this work. The work of Barbara Buttenfield is supported by USGS-CEGIS grant # 04121HS029, "Generalization and Data Modeling for New Generation Topographic Mapping".

References

Briat M-O, Monnot J-L and Punt E M, 2011, Scalability of contextual generalization processing using partitioning and parallelization. *14th ICA Workshop on Generalization and Multiple Representation*, Paris, June 30-July 1.

Buttenfield B P, Stanislawaski L V and Brewer C A, 2011, Adapting generalization tools to physiographic diversity for the USGS National Hydrography Dataset. *Cartography and GIS* 38(3): 289-301.

Punt E M and Watkins D, 2010, User-directed generalization of roads and buildings for multi-scale cartography. *13th ICA Workshop on Generalization and Multiple Representation*, Zurich, September 12-13.

Savino S, Rumor M, Zanon M and Lissandron I. 2010 Data enrichment for road generalization through analysis of morphology in the CARGEN project. *13th ICA Workshop on Generalization and Multiple Representation*, Zurich, September 12-13.

Stanislawski L V, 2009, Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. *Computers, Environment and Urban Systems*, 33: 325-333.

Stanislawski L V and Buttenfield B P, 2011a, Hydrographic generalization tailored to dry mountainous regions. *Cartography and Geographic Information Science*, 38(2):117-125.

Stanislawski L V and Buttenfield B P, 2011b, A raster alternative for partitioning line densities to support automated cartographic generalization. *25th International Cartography Conference*, Paris, July 3-8.

Stanislawski L V, Buttenfield B P, Finn MP, and Roth, K, 2009, Stratified database pruning to support local density variations in automated generalization of the United States National Hydrography Dataset. *24th International Cartography Conference*, Santiago, November 15-21.

Stanislawski L V, and Savino S, 2011, Pruning hydrographic networks: a comparison of two approaches, *14th ICA Workshop on Generalization and Multiple Representation*, Paris, June 30-July 1.

Thomson R C and Brooks R, 2007, Generalization of geographical networks. In W.A. Mackaness W A, Ruas A and Sarjakoski L T (eds), *Generalization of Geographic Information: Cartographic Modelling and Applications* (pp.). Elsevier for International Cartographic Association, 255–267.

Töpfer F and Pillewizer W, 1966, The principles of selection: a means of cartographic generalization. *The Cartographic Journal*, 3(1):10-16.

Touya G, 2007, A road network selection process based on data enrichment and structure detection. *10th ICA Workshop on Generalization and Multiple Representation*, Moscow, August 2–3.

Touya G, Duchêne C and Ruas A, 2010, Collaborative generalization: Formalisation of generalisation knowledge to orchestrate different cartographic generalization processes. *Proceedings, GIScience 2010*, Zurich. Berlin: Springer LNCS 6292: 264-278.

U.S. Census Bureau, 2011, Metropolitan and Micropolitan Statistical Areas: Tables, http://www.census.gov/popest/data/metro/totals/2011/index.html (accessed May 21, 2012).

USGS, 1985, Standards for 1:100,000-scale quadrangle maps. *National Mapping Program Technical Instructions*, U.S. Geological Survey, Department of the Interior.

USGS, 2006, The Best Practices Data Model – The National Map, March 1 2006, http://services.nationalmap.gov/bestpractices/model/acrodocs/Poster_BPTrans_03_01_2006.pdf.