Multiscale Representations of Water: Tailoring Generalization Sequences to Specific Physiographic Regimes

Barbara P. Buttenfield[1], Lawrence V. Stanislawski[2], Cynthia A. Brewer[3]

[1] Department of Geography, University of Colorado, Boulder, Colorado USA
Email: babs@colorado.edu

[2] ATA Services, Center for Excellence in Geospatial Information Science (CEGIS),
United States Geological Survey (USGS) , Rolla, Missouri USA
Email: lstan@usgs.gov

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

1. Introduction

This paper reports on generalization and data modeling to create reduced scale versions of hydrographic data for *The National Map* (http://nationalmap.gov) of the U.S. Geological Survey (USGS). The work draws upon previously published methods for estimating upstream drainage area (Stanislawski et al 2007), for automated stream pruning (Stanislawski 2009), for quantitative assessment of generalization results by statistical bootstrap (Stanislawski et al 2010), and for visual evaluation of mapped hydrography (Brewer et al 2009). This paper demonstrates that generalization processing must be varied to preserve specific differences in hydrographic characteristics which reflect differences in landscape type. Specifically tailored processing sequences generalize data compiled for the USGS National Hydrography Dataset (NHD) at 1:24,000 (24K) mapping scale. Results are evaluated metrically against a benchmark compiled for 1:100,000 (100K).

The United States comprises diverse physiographic regions (Figure 1). Early results established that a single processing sequence using uniform pruning and generalization parameters will generate simplified versions of hydrography that are both analytically and cartographically inadequate. Landscape differences made manifest by local geographic and geologic conditions require differing generalization algorithms, parameters and processing sequences for effective multiscale representation. Touya (2008) has proposed similar arguments but does not demonstrate empirical results, as are shown here.

Subbasin Name and Location	NHD subbasin	Regime
A Upper Suwannee River, Florida-Georgia	03110201	Flat Humid
B Lower Beaver River, Utah	16030008	Flat Dry
C Pomme De Terre River, Missouri	10290107	Hilly Humid
D Lower Prairie Dog Town Fork Red River, Texas	11120105	Hilly Dry
E South Branch Potomac, West Virginia	02070001	Mountainous Humid
F Piceance-Yellow Creeks, Colorado	14050006	Mountainous Dry
G Lower Cimarron River, Oklahoma	11050003	Flat Humid

Figure.1. Physiographic divisions of the coterminous United States (Fenneman and Johnson, 1946), with sample NHD subbasins which cover roughly 2,000 - 5,000 square kilometers.

To establish unique generalization sequences sufficient to cover the range of physiographic regimes spanning the nation, we selected a sample of NHD subbasins, characterized by three terrain regimes (flat, hilly or mountainous), and by two precipitation regimes (dry or humid). The sample of subbasins was selected to span all six combinations of terrain and precipitation.

The test subbasin (**C**) forms the watershed for the Pomme de Terre River, Missouri. The subbasin sits in the Ozark Plateau of the Interior Highlands, and covers ~2,190 sq km. The geography of this landscape is a humid climate with hilly but not mountainous terrain. Subbasin **G**, with characteristics similar to subbasin **C** is used for validation.

2. Processing Methods and Approach

Generalization of hydrography is computationally intensive. We store processing output as intermediate scale hydrographic datasets, called Level of Detail (LoD) databases (Cecconi et al. 2002) in full NHD schema. The first set of LoDs is complete, intended for mapping scales ranging from 1:50,000 (50K) to about 1:200,000 (200K) and referred to as 50K LoDs. Data modeling involves four stages of processing.

2.1 Enrichment

Hydrographic attribute tables are enriched with catchment area, estimated upstream drainage areas (Stanislawski 2009) and stream channel densities. Added data fields support subsequent modeling in several ways, e.g., to estimate local density values for each stream reach (confluence-to-confluence) and to guide pruning. Upstream drainage estimates permit relative prominence ranking of stream reaches, which assists automatic centerline delineation (especially in braided flows) as well as tapered stream symbolization for cartographic display.

2.2 Pruning

Pruning eliminates entire reaches without damaging correct topology of the stream network, terminating when remaining reaches approach a channel density limit established by a modification of the Radical Law (Töpfer and Pillewizer 1966). The modification computes remaining stream channel length. Pruning tends to homogenize channel density throughout the subbasin, thus if substantial local differences exist (e.g., in subbasins **C, D** and **E)**, it becomes necessary to stratify density levels and prune them separately (Figure 2).

In subbasin **C**, pruning reduced total channel length from 1,923 km to 1,303 km in the higher density partition, and from 1,507 km to 1,055 km in the lower density partition (total channel length 2,358 km, a 31.3% reduction).

2.3 Additional Generalization

Following pruning, additional generalization either modifies or removes details from individual features. This is the stage at which physiographic differences impose the greatest impact on processing sequences. If they exist, swamps are aggregated; flood zone boundaries are smoothed; ponds and lakes are selected on a minimum size threshold (0.02 sq. km.); centerlines are substituted for polygonal river channels; selected coordinates along flowlines are eliminated. Gaffuri (2007) argues for preservation of network outflow, but that is not performed in our approach. Six separate processing sequences have been fast-prototyped for inclusion in the toolbox. Figure 3 shows results of processing subbasin **C**.

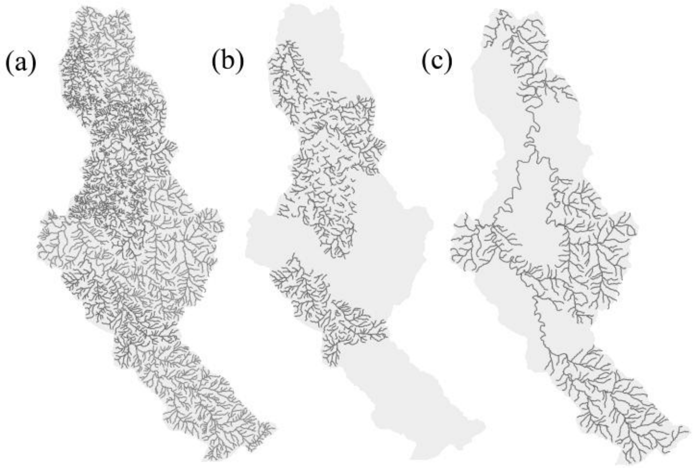

Figure 2. Density stratification and pruning of Subbasin C: (a) original flowlines; (b) higher density channels after pruning; (c) lower density channels after pruning.

We refer to the collective pruning and additional generalization processing as “differential generalization”. Pruning is differential when local density differences are stratified, as for example in regions which are partially glaciated or which cross several types of bedrock. Additional generalization models feature types (streams, canals, ponds, reservoirs, dams, etc.) differentially. In both processing stages, the sequence of operations and/or the parameters are specific to regional physiographic and hydrographic characteristics. Burghardt & Neun (2006) propose a constraint-based approach in which decisions are made automatically about which type of pruning or other generalization methods; our approach cannot accomplish this at present.

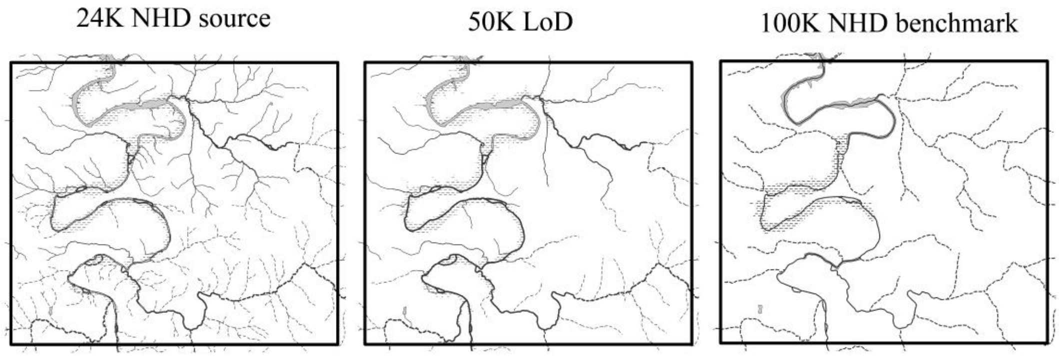

Figure 3. Subbasin C after pruning and additional generalization, compared to original scale (compiled at 24K) and to benchmark (compiled at 100K). Panels displayed at 100K.

2.4 Metric Assessment

Our benchmark for assessment is the medium resolution (100K) NHD. Metric assessment incorporates two measures of feature conflation, identifying features which correspond in the LoD and in the benchmark. The Coefficient of Line Correspondence (CLC) (Stanislwski et al. 2010) computes conflation among stream channels on the basis of length. Length preservation forms one of the most important measures of the amount of preserved detail in a generalized line (Cromley and Campbell 1990).

$$\text{CLC} = \frac{\sum \text{conflation}}{\sum \text{conflation} + \sum (\text{omissions} + \text{commissions})} \quad \textbf{(1)}$$

where:

conflation → length of channels common to LoD and benchmark;
omissions → length of channels in 100K benchmark but not in LoD;
comissions → length of channels in LoD but not in 100K benchmark.

CLC values range from 1.0 (perfect correspondence) to 0.0 (total mismatch). Features are buffered to correctly pair generalized features with benchmark features. Buffer size combines horizontal positional accuracy estimates for each network, spanning twice the tolerance for well-defined points from the U.S. National Map Accuracy Standards (NMAS) at the two scales. The tolerance at 24K and 100K is 0.02 inch, or 0.5 mm (U.S. Bureau of the Budget 1947) at each scale, which extends a total of 128 ground meters. The coefficient of area correspondence (CAC) is analogous to the CLC and compares polygonal features and is computed on the basis of area. While the CLC measures conflation of full stream reaches, the CAC includes full and partial conflations for polygonal features.

To get a clear sense of how conflation varies across a subbasin, we create a grid and compute CLC and CAC values for each grid cell (weighted by the amount of subbasin coverage in each cell, to avoid edge bias) (Figure 4). CAC values also range from 1.0 (perfect match) to 0.0 (no match).

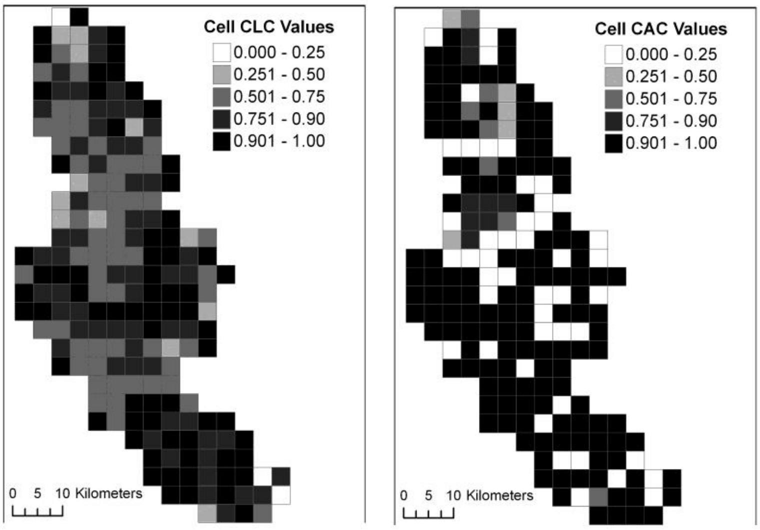

Figure 4. Gridded CLC (left) and CAC (right) metrics for subbasin C comparing 50K~200K LoD with 100K NHD benchmark. Better length correspondence is evident in the less dense portions of the stream network, where pruning has a weaker impact on overall channel length. Overall, we find a 79.2% correspondence in channel length and a 71.9% correspondence among water polygons.

3. Validation

The identical processing sequence was also applied to NHD features for a nearby subbasin (**G**). The subbasin lies within 300 km of subbasin **C** and covers 3,570 sq km. Terrain for subbasin **G** is hilly, similar to subbasin **C** but at a lower elevation. Runoff estimates for subbasin **G** are about half those of subbasin **C,** and channel density is uniformly high. Figure 5 shows CLC and CAC distributions for subbasin **G.** Comparison of CLC and CAC values (Table 1) indicates that applying the processing sequence designed for subbasin **C** to subbasin **G** results in a very good line correspondence, but a lesser quality area correspondence. A bootstrap analysis will generate confidence intervals to infer if differences between the two pairs of metrics are significant, and is described in Stanislawski et al (2010).

Table 1. Subbasin metrics compared for two physiographically similar subbasins, processed with a single processing sequence.

Subbasin	CLC	CAC
Pomme de Terre (**C**)	0.792	0.719
Lower Cimarron (**G**)	0.830	0.623

The CLC and CAC measures provide a method of evaluating the consistency of pruning and generalization across subbasins in comparison to an existing benchmark. Comparison of values in Table 1 indicates 74% average correspondence between the between the 50~200K LoDs and 100K NHD benchmark, which we consider to be a relatively high level of consistency. The CLC and CAC take a first step towards metric assessment of generalization outcomes, and we look forward to other researchers suggesting additional metrics.

4. Summary

Methods described in this paper are designed for processing hydrographic data. To date, we have worked with roughly a dozen hydrographic subbasins situated in rural areas. We are currently testing our approach on two metropolitan areas to identify possible issues caused by urban features, such as differentiating ditches and canals from natural stream channels, working with stream channel discontinuities, etc. Our outcomes could be compared with DEM-derived streams for completeness, and to insure that total displacement does not compromise overall generalization objectives.

Acknowledgements

This research is supported by USGS-CEGIS grant # 04121HS029, “Generalization and Data Modeling for New Generation Topographic Mapping”. Thanks to Chris Anderson-Tarver and Ryan Haney for programming assistance.

References

Brewer, C. A., Buttenfield, B.P., and Usery, E.L. 2009 Evaluating Generalizations of Hydrography in Differing Terrains for The National Map of the United States. *Proceedings, International Cartographic Conference,* Santiago, Chile.

Burghardt and Neun, M 2006 Automated Sequencing of Generalization Services Based on Collaborative Filtering. *Proceedings of the 4th International Conference on Geographic Information Science (GIScience’2006),* 28, 41-46.

Cecconi, A. Weibel, R, and Barrault, M., 2002 Improving automated generalization for on-demand web mapping by multiscale databases, *Proceedings International Symposium on Geospatial Theory, Processing, and Applications*, Ottawa Canada.

Cromley, R.G. and Campbell, G.M. 1990. A Geometrically Efficient Bandwidth Line Simplification Algorithm. *Proceedings of the 4th International Symposium on Spatial Data Handling,* Zurich, Switzerland, 1, 77-84.

Fenneman, N.M. and Johnson, D.W. 1946 Physical Divisions of the United States: Washington, D.C., USGS special map series, scale 1:7,000,000. http://water.usgs.gov/GIS/metadata/usgswrd/XML/physio.xml.

Gaffuri J. 2007 Outflow Preservation of the Hydrographic Network on the Relief in Map Generalization. *Proceedings of 23rd International Cartographic Conference*, Moscow, Russia.

Stanislawski, L.V, Finn, M., Usery, E.L. and Barnes, M. 2007. Assessment of a Rapid Approach for Estimating Catchment Areas for Surface Drainage Lines. *Proceedings ACSM-IPLSA-MSPS*, St. Louis Missouri.

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(5): 325-333.

Stanislawski, L.V., Buttenfield, B.P., and Smaranayake, V.A. 2010, Automated Metric Assessment of Hydrographic Feature Generalization Through Bootstrapping. *12th ICA Workshop on Generalization and Multiple Representations*, September, 2010, Zurich, Switzerland.

Töpfer, F., and Pillewizer, W. 1966 The Principles of Selection, *The Cartographic Journal*. 3: 10-16.

Touya, G. 2008 First Thoughts for the Orchestration of Generalisation Methods on Heterogeneous Landscapes. *Proceedings, ICA Workshop on Generalizations*. Montpellier, France. http://aci.ign.fr/montpellier2008/papers/01_Touya.pdf

U.S. Bureau of the Budget 1947, United States National Map Accuracy Standards.