A Comparison of Star and Ladder Generalization Strategies for Intermediate Scale Processing of USGS National Hydrography Dataset

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] 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

Abstract

In this study, we will work with the USGS National Hydrography Dataset (NHD) which represents the surface water features of the United States and Territories. NHD is distributed in a high-resolution version, intended for mapping at 1:24,000 and at a medium resolution intended for mapping at 1:100,000. The objective is to compare two strategies for generalization, namely ladder (or incremental) generalization with star processing, in which all generalized versions are generated directly from a single large scale source. These tests are being conducted in support of USGS delivery of hydrographic data at multiple scales. The purpose of testing is to ascertain if all reduced scale versions can be generalized from a single large scale version, and what differences in feature shape, length, and density might be generated by ladder versus star generalizations. Comparisons will involve geometric measurements (e.g. proportions of retained line length and retained coordinates), statistical measurements (Coefficients of Line Correspondence and Area Correspondence) and shape comparisons (areal feature displacement). Comparison of processing steps will also be discussed, for example at what scales stratified pruning or differential simplification is required. The idea is to determine how many source scales of hydrography are needed to provide generalized data usable across a continuous range of mapping scales, and what differences in geometry, statistics and shape emerge when using the two types of generalization.

Introduction: Ladder and Star Strategies for Production of Base Mapping Data

National mapping agencies (NMAs) in many countries are confronting the challenge of fully automated generalization for mapping at multiple scales. European agencies differ in following a star or ladder generalization strategy (Stoter 2005a). The star strategy generates versions at all reduced scales from a single large scale database, generalizing source data in progressively larger scale jumps. The ladder strategy generalizes smaller scale versions incrementally, effectively generalizing the generalized data. Stoter (2005b) reports that NMAs such as Belgium and Germany exclusively applied the ladder strategy. France, Denmark, Catalonia and Switzerland adopted a mixed strategy, where " ... large to middle-scale datasets are derived from the base dataset while smaller scales are derived from one middle-scale dataset." (Stoter 2005b: 3)

In the United States, the U.S. Geological Survey (USGS) national mapping agency has historically utilized a strategy that is neither star nor ladder, exclusively, compiling some multi-scale topographic base map data layers individually, as in the case of the National Elevation Dataset (NED) and the National Landcover

Dataset (NLCD). In contrast, data for the National Hydrography Dataset (NHD) was initially produced at two mapping scales, intermixing a ladder strategy with independent updates and densification. The 1:100 000 (100K) medium resolution (MR) version was compiled from photo-reduced mosaics of the 1:24 000 (24K) high resolution (HR) vector data. Since compilation, the 24K data has been updated, and additionally densified in some areas to mapping scales as large as 1:2 400; and thus portions of the 24K data are now independent of the 100K data. At present, no maintenance is performed to keep the two versions integrated. A current strategy of the USGS National Geospatial Technical Operations Center (NGTOC) for minimizing data maintenance is to compile a multi-scale layer of the best available data for each data theme, and apply automated generalization operations to generate coarser levels of detail. When a national coverage dataset is comprised of multiple scale versions, as is the NHD, smaller scale versions may be compiled from several scales and sources of USGS digital data.

This research addresses the question of whether the choice of strategy in generalization of hydrography results in different processing outcomes. Does the decision to apply a star or a ladder strategy matter? Does it produce consequential differences in feature length and area, or in feature densities? Like terrain, the geometry of water features is known to be sensitive to scale change, which challenges the utilization of NHD data for cartographic base mapping and hydrographic analysis at scales below 100K. Intermediate scale level-of-detail (LoD) versions have been explored (Buttenfield et al. 2011; Stanislawski et al. 2010; Brewer et al. 2009) and these are felt to be appropriate for use down to scales of roughly 1:200 000 (200K). This paper reports an experiment to extend the usable scale ranges of NHD below 200K, applying both ladder and star strategies and comparing generalization results graphically and metrically.

A set of four adjacent NHD subbasins will be generalized by star and ladder methods from 24K source to 50K, 200K and 500K. Generalization processing will include enrichment, pruning, simplification and shape modification. Results will be compared visually, in a cartographic context and conflated metrically, using Coefficients of Line Correspondence and Coefficients of Area Correspondence (Stanislawski 2009; Buttenfield et al. 2010). Benchmark datasets which are currently available include the MR 100K NHD, the 500K River Reach dataset and the 1: 2 000 000 (2M) National Atlas data. Bootstrapping analysis (Stanislawski et al. 2010) will determine if differences in hydrographic detail are significant among the versions, and if these differences vary with landscape differences as evidenced by local stream channel density.

Study Area and Data Set

The four subbasins are situated in central Iowa and include the drainage system for the Raccoon River and the Des Moines River, which covers roughly 20 000 km^2. As shown in Figure 1, the hydrographic pattern reflects characteristics of the landscape history, where a glacial lake bed borders a dissected till plain. The stream network in the lake bed is less dense. Better drainage in the till plain is evident in the denser network to the south, which was developed during an earlier glacial period. It is important to preserve these local densities when generalizing the data, since they reflect significant glacial processes evident in landscapes throughout this part of the country. Blue lines in Figure 1 indicate stream centerlines and primary drainage channels: the Des Moines River flows down the east side of the study area on its way to the Mississippi River, while the Raccoon River flows down from the northwest corner.

The data to be processed for this research is the NHD, which represents surface water features for the United States. A high-resolution (HR) version is largely compiled from 24K source data for the United States. A medium-resolution (MR) version compiled for 100K contains fewer and less detailed features, albeit with richer attribution. USGS distributes the HR and MR data versions through *The National Map* (http://nationalmap.gov) web portal. State-level efforts continue to densify regions of the high-resolution version to scales as large as 1:2 400. At present, densified coverage is only available in small pockets across the nation. In addition to HR and MR versions, USGS distributes small scale versions of United States hydrography, at 1:2 000 000 (2M) and 1:10 000 000 (10M), through *The National Atlas* (http://www.nationalatlas.gov).

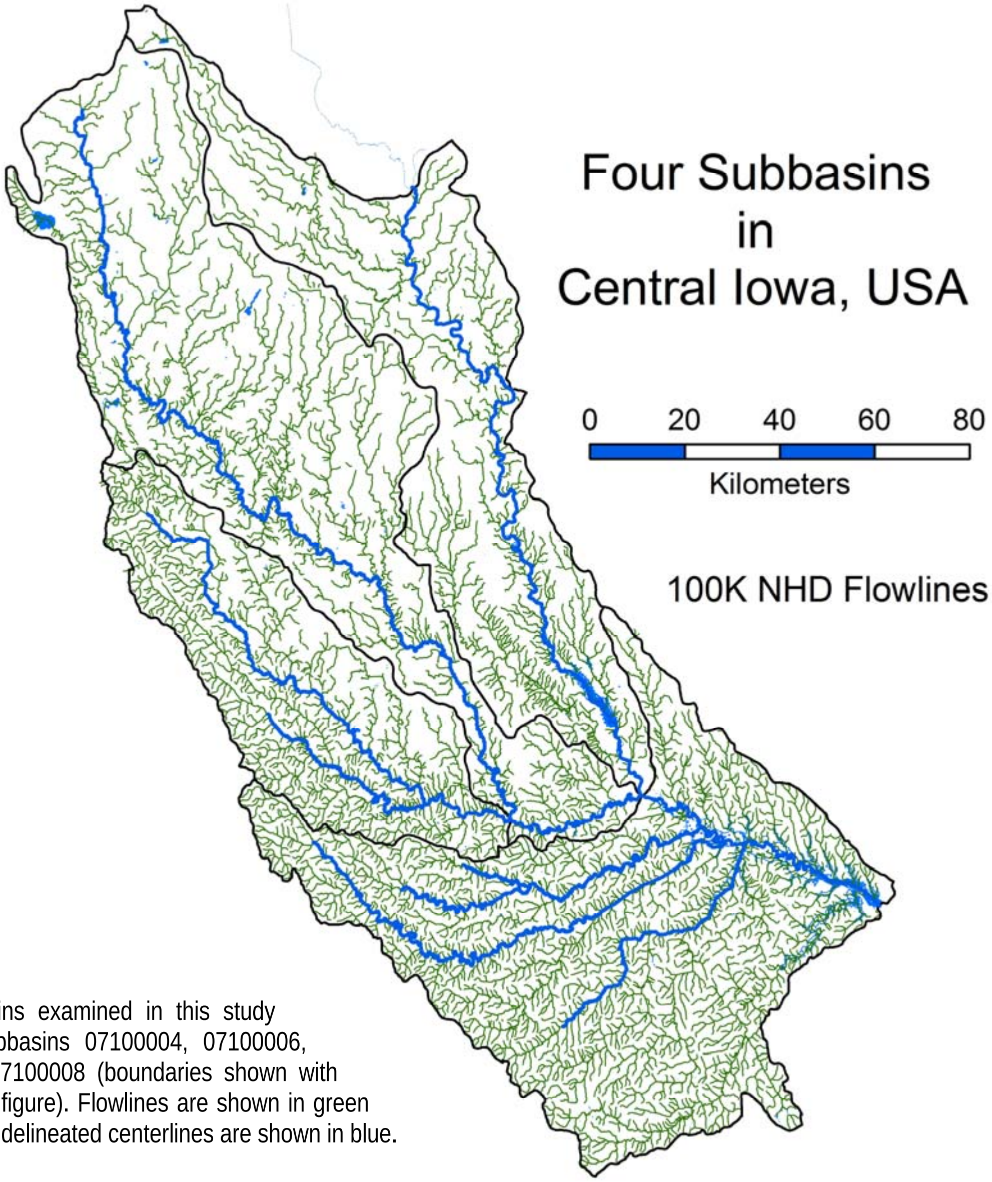

Figure 1.
The four subbasins examined in this study include NHD subbasins 07100004, 07100006, 07100007, and 07100008 (boundaries shown with black lines in the figure). Flowlines are shown in green and automatically delineated centerlines are shown in blue.

Methods

Generalization will be applied to a source data set to generate LoDs at intermediate mapping scales. The processing will include database enrichment, stratified pruning, and simplification. Each LoD will be validated against a benchmark dataset in terms of geometry and conflation. Within the scope of this paper, results will be described for only one source scale. The 24K source scale data will be generalized using a star strategy to produce three product datasets, at 50K (a 2x scale jump), 200K (8x) and 500K (20X). The 50K LoD will be ladder-processed to 200K; and that dataset will be ladder-processed to 500K.

Figure 2 illustrates the complete testing method. The vertical axis is logarithmic, showing successive 4x scale jumps. The magnitude of these scale jumps is not arbitrary, rather they are designed with an eye to available benchmark data sets (100K NHD, 500K River Reach data and 2M National Atlas data).

Dashed lines indicate a scale where a benchmark data set is available for metric and statistical comparison of the LoDs. Gray boxes in the Figure indicate expectations of the range of scales appropriate for mapping using each LoD. At the outset, our working premise is that the 50K LoD should support mapping from 50K to 200K, the 200K LoD should support mapping down to 1 : 1 million, and the 500K LoD may be usable down to 1 : 3 million.

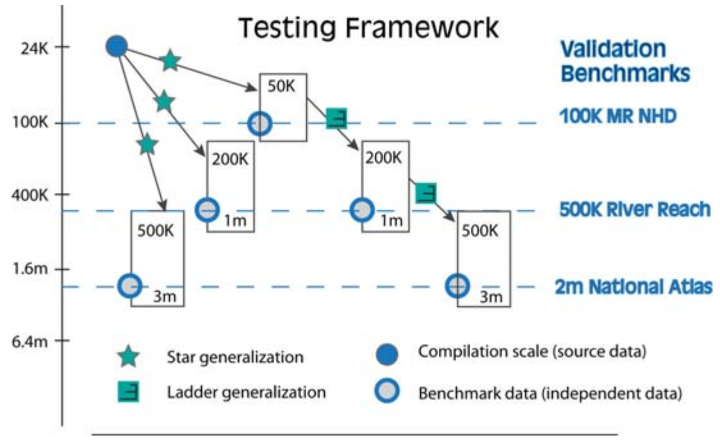

Figure 2. Workplan for testing star and ladder generalization strategies. Five LoDs will be created, and compared against three benchmark data sets.

Analysis

Database enrichment. The 24K source dataset was enriched with estimates of catchment area based on a Thiessen polygon tessellation (Stanislawski et al. 2007). A catchment is defined as the ground area which drains into a stream channel. Density is calculated by dividing summed stream length for all channels by summed catchment areas for the entire subbasin. Density values guide the stratified pruning and simplification, discussed below. A second form of database enrichment calculates upstream drainage area values for each NHD flowline feature, by cumulative summing of catchment areas (Stanislawski 2009). Upstream drainage area values are additionally used for selective elimination following simplification, and to symbolize tapered stream widths in mapping at reduced scales.

A third form of database enrichment automatically delineates stream centerlines and attributes them in the flowlines database. For this research, centerline delineation is accomplished by parsing all stream names to collect features whose name contains “River” to get the primary channels, and then selectively eliminating small primary channels whose upstream drainage area is less than 100 km^2. A more robust method for automated centerline delineation is proposed by Anderson-Tarver et al. (2011) which traces upstream through node tables from the channel having the highest upstream drainage area. Our research shows that in humid areas with very dense drainage networks, such as in Iowa, the two delineation methods produce similar but not identical results.

Pruning. Pruning is the initial step which modifies the level of detail. Pruning extracts a stream network from the source data whose drainage density is appropriate for the target map scale. Two factors constrain pruning: the first preserves topology with other channels and with hydrographic polygon features, by eliminating all upstream channels when any channel is pruned. The second constraint preserves local density variations within the subbasins, thus preserving the glacial variations observed above. Partitioning has been used previously in generalization to differentially process areas with different data densities (Bobzien et al. 2008; Chaudhry and Mackaness 2008; Stanislawski 2009) and to monitor isolated (point) objects over time and space (Downs 2010). It is mandatory when pruning across a sequence of scale jumps, as in this experiment, that features pruned out at larger scales do not suddenly reappear in smaller scale pruning outputs. Our approach to pruning addresses that issue by working from headwaters downstream towards the pour point of a stream network.

Two density classes were established for study area channels (Figure 3), based on a raster-based partitioning analysis of NHD flowline features (Stanislawski and Buttenfield, 2011), which resulted in low- and high-density partitions having densities for the 24k data of 0.401 km / km^2 and 1.211 km / km^2, respectively. The calculation of an appropriate density at reduced scale is determined by means of a variation of Töpfer and Pillewizer's Radical Law. We calculate an expected stream density to be retained as stream length is pruned. Radical Law computations are specific to each LoD (Table 1), depending on source and target scales, and on stream densities at the source scale. To protect stream topology, pruning eliminates entire confluence- to-confluence stream channels along with all upstream channels. Water features vary in length depending on terrain (hilly or flat) and climate (dry or humid) and predictions about what length of hydrographywill be pruned are difficult. Therefore, it is not possible to achieve the Radical Law density estimates precisely.

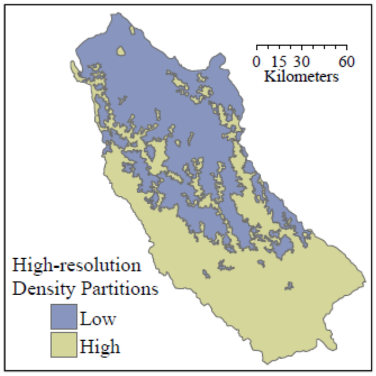

Figure 3
Density partitions for 24K flowlines in the study area.

As can be seen comparing length values in the Table, the amount of pruning increases for larger scale jumps. The three star generalizations demonstrate this trend, pruning 30% at 50K from 2 237 km down to 1565 km), 60% at 200K, and 73% at 500K for density partition1. Pruning reduces lengths roughly the same amount for density partition 2, with reductions in stream length of 31% at 50K, 62% at 200K, and 75% respectively. Comparing star with ladder pruning is easier to understand by comparing Radical Law estimated densities. The 200K star pruning reduces channel density to 0.139 km/km^2 for partition 1 and to 0.420 km/km^2 for partition 2. The 200K ladder pruning is nearly the same (0.138) for partition 1, but only 0.406 (4% lower) for partition 2. The effect is more pronounced for the 500K pruning, in which pruned density differs between star and ladder LoDs by 4% for partition 1 and by 12% for partition 2.

Simplification. Following pruning, a sequence of generalization operations was applied to the pruned data, to simplify flowline features by the Bend Simplify algorithm (Wang and Muller 1998). In every case an identical simplification tolerance was applied to both density partitions. Flowlines for the 50K LoD were simplified to 75m. The star and ladder 200K LoD flowlines were simplified to 300m; star and ladder 500K flowlines were simplified to 600m. As one might expect, the ladder LoDs contain relatively less detail (as evidenced by shorter channel lengths) in comparison with star LoDs at any scale. Density discrepancies from Radical Law estimates increase at smaller target scales, with largest discrepancies occurring with equal frequency in both density partitions. This is an interesting finding, as one might expect that the ladder generalization densities would drift consistently from Radical Law estimates with each subsequent ladder step (i.e., 50K, 200K, 500k). With each pruning step initiating with less detail than its star equivalent, one might predict that the level of detail in a ladder LoD at any scale would be harder to control than for a star LoD. For this data set at least, that is not the case. We note however that for both 500K LoDs, where the most extreme discrepancies occur, simplification to 300m and to 600m tolerance produced a nearly identical density for partition 2 (0.191 km/km^2 for the ladder, and 0.249 km/km^2 for the star). At this target scale, most of the original channel detail had been pruned away, leaving almost no further detail to remove without corrupting the cartographic validity of the LoD.

Table 1. Length and Density Computations for Ladder and Star Generalizations

Stream lengths catchment areas and densities for each LOD produced to date. The 24K data is not compiled with density stratifications, thus only one source length is available. Data for each LoD is partitioned individually prior to pruning, resulting in differing density stratifications for each one. Length values are given in km and density values are given in km / km^2 .

LoD and source scale	Density Partition	Source Length	Catchment Area	Catchment Density	Radical Law Estimated Length	Radical Law Estimated Density	Pruned Length	Simplified Length	Density after generalization (% discrepancy)
50K star (24K)	1	2 237	5 580	0.401	1 550	0.278	1 565	1 538	0.276 (0.8%)
	2	17 678	14 592	1.211	12 246	0.839	12 481	11 848	0.812 (3.2%)
200K star (24K)	1	2 237	5 580	0.401	775	0.139	890	851	0.153 (-9.9%)
	2	17 678	14 592	1.211	6 123	0.420	6 715	5 840	0.400 (4.6%)
200 ladder (50K)	1	1 538	5 580	0.276	769	0.138	835	813	0.146 (-5.7%)
	2	11 848	14 592	0.812	5 924	0.406	5 805	5 410	0.371 (8.7%)
500k star (24K)	1	2 237	5 580	0.401	490	0.088	605	563	0.101 (-15.0%)
	2	17 678	14 592	1.211	3 872	0.265	4 422	3 637	0.249 (6.1%)
500k ladder (200K)	1	813	5 580	0.146	514	0.092	492	484	0.087 (6.0%)
	2	5 410	14 592	0.371	3 422	0.235	2 787	2 781	0.191 (18.7%)

A second phase of generalization processed details for polygonal features. First the polygon boundaries were simplified, to 60m (NHD waterbodies) and 40m (NHD areas) for the 50K LoD, to 200m (NHD waterbodies and NHD areas) for both star and ladder 200K LoDs, and to 400m for the 500K LoDs. Following simplification, polygon features were selected on size in accordance with minimum size criteria approximated from NHD standards (USEPA and USDOI 1999). The sequencing of simplification and selection is important because the Bend Simplify algorithm can either increase or decrease polygon areas depending on their shape convexity. Size criteria vary by feature type and scale, so for example the minimum size for inclusion of swamp/marsh features at 200K is 0.203 km^2 and the minimum size for inundation areas is 0.072 km^2. Our analysis focused on comparisons of linear features, due more to space limitations for this paper than anything else. Analysis of polygon generalization is underway, but we do not have the detailed comparisons yet that are presented here for linear features.

Figure 4 illustrates simplification outcomes for all LoDs, along with source data (24K). Each panel shows a colored box coded by outline hue to the footprint of the next largest scale. Obvious differences between the star and ladder 200K and 500K LoDs include differences in the number of stream channels and this is especially evident in the channel headwaters. In the ladder approach, simplification from a previous generalization reduces starting total stream length from that of corresponding star values, and consequently target pruned lengths are likewise smaller than star values. Differences in shape are less obvious on visual inspection but will be highlighted in the validation section below.

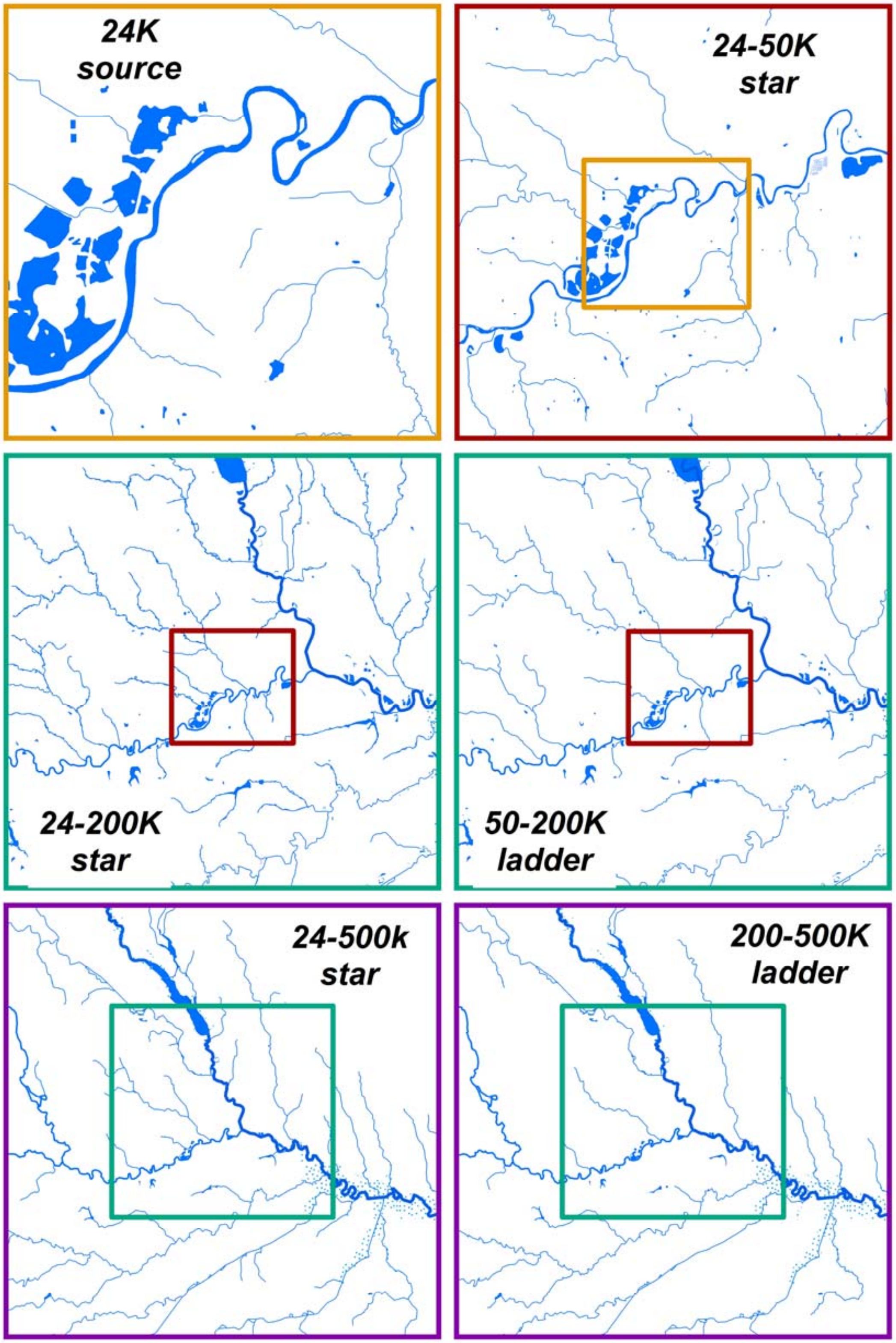

Figure 4. Sample images of data following enrichment, pruning, simplification and selection for 24K source and all five LoDs. Boxes are color coded against panel frames to indicate spatial footprint of the next scale.

Validation

The 50K LoD can be assessed in comparison to the medium resolution 100K NHD shown in Figure 1. The star and ladder LoDs can be assessed in comparison to a 500K benchmark of independently compiled data, the River Reach data set. This was originally compiled for use by the Environmental Protection Agency, "… to establish hydrologic ordering, to perform hydrologic navigation for modeling applications, and to provide a unique identifier for each surface water feature". (Horn et al. 1994:1) As a data base intended for analysis rather than for cartographic base mapping, the River Reach data contains more detail than may be necessary for topographic base mapping, particularly at smaller scales (Figure 5). The initial version was completed in 1975, at 500K. This dataset contains over 3 million hydrographic line features for the continental United States and Hawaii.

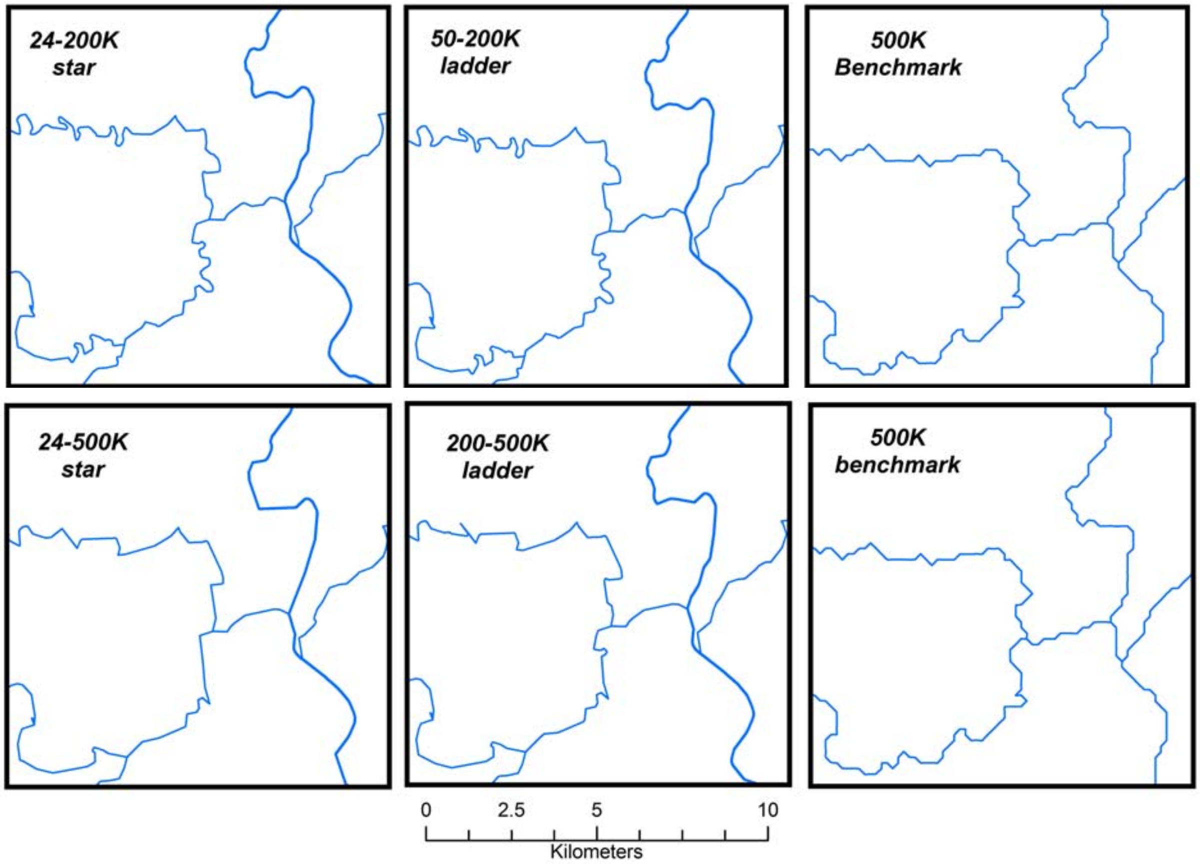

Figure 5. Visual comparison of the 200K and 500K star and ladder LoDs, with flowline features selected for display at 500K. As reflected in length and density values given in Table 1, the two 200K LoDs appear quite similar. The 500K star LoD appears to hold less detail than the 500K ladder, whose flowlines begin to collapse upon each other (center left, just below the panel label). The benchmark data set used for metric validation is shown on the far right.

To compensate for scale-based differences in content between the 200K LoDs and the 500K benchmark data set, it is important to adjust the content of the 200K LoD to the smaller scale. The adjustment was made by selecting on Upstream Drainage Area (UDA) values, discussed above in the section on enrichment. A UDA threshold value of 35 km^2 was chosen to compare the 200K and 500K LoDs with the River Reach benchmark data.

To compare the simplified versions to the benchmark, we utilize the Coefficient of Line Correspondence (CLC), a measure of conflation adapted by Stanislawski et al. (2010) from a measure for areal feature coincidence discussed by Taylor (1977). CLC measures the proportion of flowline features which match in the LoD and the benchmark with respect to the summed length of matches and errors. CLC values can be thought of as a percentage, with a perfect match (total correspondence) indicated by a CLC of 1.0; and a lack of any matches between the two data sets indicated by a CLC of 0.0. The aggregate CLC values over the four subbasins for the star and ladder 200K LoDs are 0.757 and 0.720, respectively. At 200k, the star LoD conflates better with the benchmark. At 500K, the ladder LoD shows better correspondence. CLC values for the star and ladder 500K LoDs are 0.749 and 0.767.

To explore the spatial pattern of line feature correspondence, we overlaid a grid of 200 cells over the study area and sampled line features in each cell, computing CLC values for each cell. Grid cell CLC values are weighted by the amount of area in each cell which lies within the four subbasin study area. Weighted CLC values are illustrated in Figure 6. Extremely low CLC values are evident around the edges of the study area, and in the southern half of the subbasins, where channel densities are highest at larger scales. Recall that channel densities in the 500K benchmark are relatively uniform, while density differences are still apparent at 200K; and this contributes to the lower CLC values.

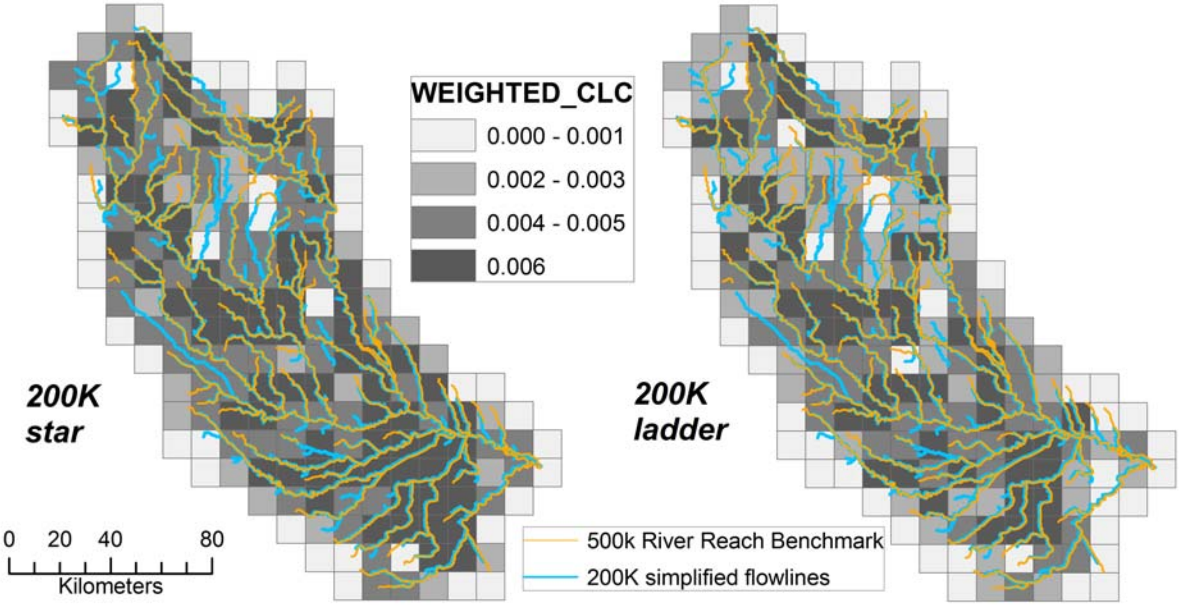

Figure 6. Weighted CLC values sampled for each of 200 grid cells for the 200K Star and Ladder LoDs. Blue lines indicate simplified features not contained in the benchmark; bright orange indicate benchmark features pruned out of the simplified features; green features indicate matching channels. Weighted CLC values for both grids indicate highest correspondence among line features in the higher density partitions, with a mixed degree of conflation in the northern less dense part of the stream network.

A bootstrap analysis of these grid-sampled 200K CLC values indicates 95% confidence intervals of 0.6810 - 0.7590 for the star LoD, and 0.7200 -0.7951 for the ladder LoD. Thus we can conclude that at 500K mapping scales, the pattern of coincidence for the two generalization strategies produces essentially the same output ($p < 0.05$). At larger mapping scales (e.g. 200K), dissimilarities between the star and ladder simplifications might become significant.

Figure 7 shows weighted CLC values for the 500K LoDs, comparing to the same 500K River Reach benchmark. The pattern of matching channels is more uniform overall in these LoDs, most likely because at this mapping scale, density differences are not as pronounced as at larger scales. Aggressive pruning and simplification required to achieve appropriate levels of detail at 500K, which homogenizes channel densities throughout the basins. Bootstraps for the 500K CLC values will be presented at the workshop.

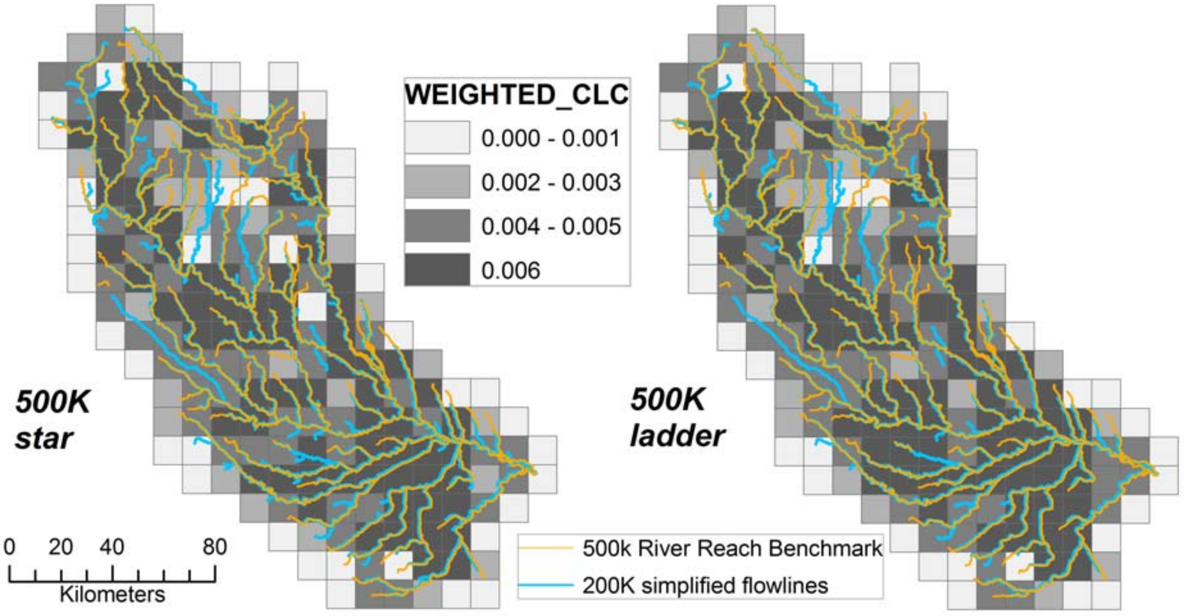

Figure 7. Weighted CLC values sampled for each of 200 grid cells for the 500K Star and Ladder LoDs. Blue lines indicate simplified features not contained in the benchmark; bright orange indicate benchmark features pruned out of the simplified features; green features indicate matching channels. Weighted CLC values for both grids indicate highest correspondence among line features in the higher density partitions, with a mixed degree of conflation in the northern, less dense part of the stream network.

In considering these initial findings, it is important to keep in mind that these LoDs are intended to span a range of mapping scales. The 200K LoDs are intended to span from 200K to 1M. The 500K benchmark falls in the middle of this scale range, where correspondence may not be as high as it would for a 200K benchmark, if one existed at that particular scale. The 500K LoDs span from 500K to 2M, and the 500K benchmark falls at the high end of the scale range, where correspondence should be highest. Results for validation against the 2M scale benchmark will be presented at the workshop.

Discussion

This paper reports on initial experimentation with star and ladder generalization strategies, working with a four subbasin area at the edge of a humid and relatively flat glaciated region in the central United States. One part of the region contains stream channels which are much more dense than the other, permitting comparative exploration of star and ladder processing Findings suggest that for scales below 200K, local channel densities become less prominent, for both star and ladder solutions. The star LoDs retain more length and more channel density relative to the ladder processing, at both 200K and 500K. Conflation analysis by CLC validation for the 200K LoDs shows that at a 500K mapping scale, patterns of correspondence for the star and ladder LoDs do not differ significantly, which means the two types of processing create essentially interchangeable data products at 500K. For the presentation, we will run

CLC analyses to perform direct comparisons between star and ladder versions, to further assess which portions of each strategy are in fact unique.

One should not conclude from this single experiment that the two types of processing are interchangeable at every mapping scale. At larger scales, where smaller details begin to differ more prominently, the star generalization might be preferable for reference or topographic mapping while the ladder strategy would be preferred for creating thematic or web map databases. In the former case, one anticipates overlaying additional data, and in the latter, the objective would be smaller data volume for faster delivery. On the basis of a single set of experiments however, this is mere speculation.

The possibility of a mixed strategy has been considered, that is, to generate star versions at two or three key scales and generate cartographic data products (DCMs, effectively) using ladder processing from these key scales. This is very close to the current strategy used by the USGS. It is workable, but data integration problems are coupled with a requirement for permanent feature IDs to support automatic feature updates. In a country the size of the United States, implementation of permanent feature IDs into existing databases presents an insurmountable obstacle, given present staffing levels and existing data volume.

As stated earlier, our expectation was that the 50K LoD should support mapping from 50K to 200K, the 200K LoD should support a range from 200K to 1M, and the 500K LoD should extend down to 1 : 3 million. That did not happen, in these experiments, and examples will be shown at the presentation. The 50K LoD carries to 200K for sparse channels, but begins to break down in dense channels at about 150K. A similar pattern is seen with the 200K LoD, with denser channels tending to fail on visual inspection at about 800K. Sparser channels hold up well at the smallest scales for both the 200K and 500K versions, especially for the ladder versions. Overall, the 500K LoD seems most appropriate down to 2M rather than to 3M. One reason for the varied success with density partitions may be related to the loss of distinction between local densities at smaller mapping scales. Cartographers have argued that local density becomes unimportant at some as yet unspecified mapping scale. This work indicates that for the Iowa dataset, the scale at which this happens might fall near 200K. Densities might be homogenized by varying simplification tolerance thresholds, or by pruning to homogenize local densities in the first place. For the work reported here, we worked only with stratified pruning and uniform simplification. Here lies another area for future experimentation.

In comparing the two approaches, one must be vigilant to propagate geometrical adjustments based on target scales as well as on the size of the scale jump. The Radical Law in its basic form compensates for the size of the scale jump, however a 4x jump from 5K to 25K generates the same Radical Law coefficient as does a 4x jump from 500K to 2M. While this makes pruning and simplification consistent, it does not account for an important cartographic principle. Clearly, the relative prominence between natural and cultural features shifts: at larger scales, natural features are more important, and at smaller scales, cultural features take precedence. The Radical Law does not account for this in the form used here. Refinements of the Radical Law to compensate for pruning and simplification across very large scale jumps, as well as to acknowledge the inevitable disappearance of density differences at smaller scales form directions for future research.

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 Michael Gleason, CU-Boulder, for assistance with data processing. We also acknowledge Mamata Akella and Kevin Ross for undertaking student projects on star vs. ladder comparisons at Penn State University in 2009.

References

Bobzien, M., Burghardt, D., Petzold, I., Neun, M. and Weibel, R. 2008 ‘Multi-Representation Databases with Explicitly Modeled Horizontal, Vertical, and Update Relations’, *Cartography and Geographic Information Systems,* 35(1): 3-16.

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 (ICC2009)*, Santiago, Chile.

Buttenfield, B.P., Stanislawski, L.V. and Brewer, C.A. 2011 Adapting Generalization Tools to Physiographic Diversity for the USGS National Hydrography Dataset. *Cartography and Geographic Information Science*, special issue for the ICA US National Report, Ed. Robert Edsall (in press).

Buttenfield, B.P., Stanislawski, L.V., and Brewer, C.A. 2010 Multiscale Representations of Water: Tailoring Generalization Sequences to Specific Physiographic Regimes. *Short Abstract Proceedings of the 6th International Conference on Geographic Information Science (GIScience 2010)*.

Chaudhry, O. and Mackaness, W.A. 2008 Partitioning to Make Manageable the Generalization of National Spatial Datasets, *11th ICA Workshop on Generalization and Multiple Representation,* Montpellier, France, June.

Downs, J. 2010 Time-Geographic Density Estimation for Moving Point Objects. *Proceedings of the 6th International Conference on Geographic Information Science (GIScience 2010), Zurich, Switzerland, September.*

Horn, C.R., Hanson, S.A. and McKay, L.D. 1994 *History of the U.S. EPA’s River Reach File: A National Hydrographic Database Available for Arc/Info Applications.* Washington DC: Environmental Protection Agency (EPA) Technical paper. http://www.epa.gov/waters/doc/historyrf.pdf

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. and Buttenfield, B.P. 2011 A Raster Alternative for Partitioning Line Densities to Support Automated Cartographic Generalization. *Proceedings, International Cartographic Congress,* Paris, France July (in press).

Stanislawski, L.V., Buttenfield, B.P. and Samaranayake,V.A. 2010 Automated Metric Assessment of Hydrographic Feature Generalization Through Bootstrapping. *Proceedings, 13th Workshop of the International Cartographic Association Commission on Generalisation and Multiple Representation*, Zurich, Switzerland, September. http://ica.ign.fr/2010_Zurich/genemr2010_submission_11.pdf

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

Stoter, J. E. 2005a Generalization Within NMA's in the 21st century. *Proceedings, International Cartographic Conference (ICC)* A Coruna, Spain.

Stoter, J. E. 2005b Generalisation: The Gap Between Research and Practice. *Proceedings*, *8th ICA WORKSHOP on Generalisation and Multiple Representation.* A Coruña, Spain. http://aci.ign.fr/Acoruna/Papers/Stoter.pdf

Taylor, P.J. 1977. *Quantitative Methods in Geography: An Introduction to Spatial Analysis*, Chapter 5: Areal Association. Houghton Mifflin, Boston, 396pp

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

USEPA and USDOI 1999 Standards for National Hydrography Dataset, United States Environmental Protection Agency and United States Department of the Interior, United States Geological Survey, National Mapping Program Technical Instructions, July. http://rockyweb.cr.usgs.gov/nmpstds/acrodocs/draft/dlg-f/nhd/NHD0799.pdf

Wang, Z. and Muller J.C. 1998 Line Generalization Based on Analysis of Shape Characteristics. *Cartography and Geographic Information Science* 25(1): 3-15.