APPLICATION OF THE RADICAL LAW IN GENERALIZATION OF NATIONAL HYDROGRAPHY DATA FOR MULTISCALE MAPPING

J. M. Wilmer, C. A. Brewer

Dept. of Geography, Pennsylvania State University, University Park, Pennsylvania – (juw215, cab38)@psu.edu

KEY WORDS: Radical Law, hydrography, generalization, multiscale mapping

ABSTRACT:

The importance of automating feature generalization is increasing in conjunction with demands for updated data. Determining which features to represent at myriad scales is an important part of this process. The Principles of Selection, as proposed by Töpfer and Pillewizer in 1966, provide cartographers with an empirically based generalization rule. Also known as the Radical Law, the principles provide an equation estimating the number of features depicted at smaller scales based on the relationship of map scale denominators. However, the applicability of the equation at large scales seems limited in its current form. With ongoing initiatives to automate cartographic processing and the digitization of data at large scales, there is utility in determining how many features to display on derived medium and small scale products. We evaluate the USGS National Hydrography Dataset and *National Atlas* hydrography to determine the existing length of features for comparison to expected results based on the Radical Law equation. The rate of feature selection is not the same along the continuum of scale. A new factor is added to the Radical Law equation to account for this variability. Results from a USGS flowline pruning tool by Stanislawski and Buttenfield are used to compare equation results to benchmark USGS hydrography at 5K, 24K, 100K and 2M scales. Increasing knowledge of existing hydrographic features can improve implementation of automated feature generalization and propagation through a wide range of scales.

1. INTRODUCTION

Feature selection is normally the first step to any generalization project, independent of the generalization model being employed. Although acknowledged to be part art form and part scientific process (Li & Openshaw, 1993), automation requires rules to guide generalization. These rules are often based on scale as features become too small to depict. This approach is not likely to remove enough features to respond to the loss of map space. The questions of how many features and which should be retained have to be answered, at least partially, by rules in the automated domain.

The focus of this paper is on the selection of hydrographic flowline features. It extends research sponsored by the USGS to develop automated generalization techniques in support of *The National Map*, an on-line resource for US topographic data. The labor intensive process of updating paper maps means topographic sheets can go decades between updates. Utilizing an automated approach to the data will allow the USGS to more rapidly propagate updates to all scales.

1.1 Research Questions

To address the issue of feature selection as applied to hydrographic flowlines, the following research questions were posed:

- What are the existing display relationships between hydrographic features at 1:4,800 (where available); 1:24,000; 1:100,000; and 1:2,000,000? These datasets will also be referred to by the scale denominators: 5K, 24K, 100K, and 2M.
- Is there a new factor that can be introduced to the equation to address deviations from the Radical Law between large, medium, and small scale representations?
- Can guidelines for density pruning be established to limit under-selecting flowlines at derived scales?

1.2 Background and Related Literature

The Radical Law (Töpfer & Pillewizer, 1966) is a mathematical estimation of how many features should be maintained at smaller scales in the generalization process.

$$n_f = n_a \sqrt{M_a / M_f} \qquad (1)$$

where n_f is the number of objects at the derived scale
n_a is the number of objects on the source material
M_a is the scale denominator of the source map
M_f is the scale denominator of the derived map

Additional factors called the Constant of Symbolic Exaggeration and the Constant of Symbolic Form were included for evaluation at scales smaller than 1:1,000,000. Although not used in this evaluation, the factors indicate an acknowledgement by the authors that different feature types are selected at different rates at different points on the scale continuum. Using the hydrographic data as an example, three scale transitions can be delimited: local to large scale (5K to 24K), large to medium scale (24K to 100K), and medium to small scale (100K to 2M). It is likely that the 2M data does not indicate the beginning of the small-scale data range. Feature selection between these scales may not match what is expected by the Radical Law and variables may be added based on crossing the threshold into the next scale range.

Criticisms of the equation's limited ability to identify which features to retain are abundant (Shea 1988; Buttenfield & McMaster 1991; João, 1998; Li & Choi, 2002; Jiang & Harrie, 2003). Some method of prioritizing the features must be implemented. Rank ordering features assists in selecting which ones to eliminate. In mapping cities, Kadmon (1972) suggested that the remoteness variable be weighted most heavily, acknowledging that larger cities may be eliminated. Calculating upstream drainage area for flowlines is similar in that isolated segments will be attributed with higher areas, increasing their relative importance to the overall network.

An evaluation of road networks indicated that feature selection is not consistent at all scales as the Radical Law suggests (Li & Choi, 2002). The authors counted features at scales between 1:1K and 1:200K with the number of roads selected appearing to indicate a threshold at approximately 1:20K. Ninety-seven percent of the roads were maintained in a scale change from 1:1K to 1:20K. By comparison, 65 percent of the roads were maintained between 1:20K and 1:50K which is in line with Radical Law expectations. Variations in selection rates were also found in Austrian maps for rivers (Leitner & Buttenfield, 1995). The percentage of rivers remaining between 1:200K and 1:500K was almost twice the percentage between 1:50K and 1:200K. The findings of Li & Choi (2002) and Leitner & Buttenfield (1995) point toward thresholds at particular scales that lead to variations in feature selection.

Stanislawski (2009) used the Radical Law for comparisons of hydrographic network segments, pruning 1:24K data to 1:100K. His findings showed approximately one-fourth of the 1:24K segments remained in the derived 1:100K data. This is in line with what the Radical Law says when employing the Constant of Symbolic Form for linear symbols which is the square root of the source data scale denominator (24K) divided by the target or derived scale denominator (100K). Applying this factor establishes an expectation of 24 percent of features remaining, or about one-fourth.

Stanislawski et al. (2009) proposed using the equation on flowline channel length rather than on the number of flowlines. The term channel is used here to refer to all flowline types. Comparisons between 24K and 100K data indicate a close similarity between Radical Law expectations and the percent of channel length displayed at the smaller scale. However, they found that pruning the 100K data to 500K based on Radical Law expectations provided a derived river network that was too sparse. This may be partly explained by the order in which the generalization occurs. Selecting the expected channel length and then simplifying that data will result in a dataset with less length than expected. Addressing the impact of line simplification algorithms with other factors may provide an acceptable final result in line with Radical Law expectations. The authors proposed line expansion factors for each scale to address the simplification differences. Based on the factors of 1.08 for the 100K data and 1.16 for the 500K data, there is approximately a 7.4 percent difference in line lengths between the two scales. This does not fully account for the 19 percent increase in data suggested, but it does provide for some of the difference.

Automated feature selection can be aided by partitioning the data. This allows for quicker access and selectively updating portions of a large dataset (Goodchild, 1989). Among others ways, partitioning map space has been accomplished based on homogenous geographic regions (Hardy & Lee, 2005; Chaudry & Mackaness, 2008) and density variations (Bobzien et al., 2008; Chaudhry & Mackaness, 2008; Stanislawski et al., 2009). While Goodchild (1989) expected rectangular partitions would be the norm based, understanding has grown that features are geographically dependent, and partitions should not be geometric in nature (Chaudhry & Mackaness).

1.3 About the Data

As part of *The National Map*, the National Hydrography Dataset (NHD) "is a comprehensive set of digital spatial data that represents the surface water of the United States" (Simley & Carswell, 2009). The USGS provides hydrographic data at three scales: 1:4,800 (5K); 1:24,000 (24K); and 1:100,000 (100K). The 24K data are collectively referred to as high resolution while the 100K data are medium resolution. High resolution data were compiled by digitizing 24K topographic sheets (USGS, 2010a). The 5K data are local resolution, but this is currently only available through the NHD for Vermont. However, other state organizations have created larger scale hydrography layers. For example, New Jersey's Department of Environmental Protection created 1:2,400 hydrographic data in 2002 (Thornton, 2008) and Massachusetts' DEP created 1:12,000 wetlands data in 2006 (MassGIS, 2007). The USGS is working to propagate local resolution data for subbasins intersecting Vermont into the high resolution dataset (USGS, 2010b) potentially altering comparisons between high and medium resolution data.

The hydrographic data in the NHD is composed of nine feature types with the flowline, waterbody, and area features generally being of greatest interest (USGS, 2010a). Flowlines contain stream/river, pipeline, underground conduit, coastline, canal/ditch, connector, and artificial path types. Connectors "establish a known, but non-specific connection between two non-adjacent segments that have flow" and artificial paths maintain network connectivity through 2-dimensional features like lakes or area rivers (USGS, 2010a).

The National Atlas hydrography data are provided for the entire United States and its territories at a scale of 1:2,000,000 (2M). Streams and waterbodies are represented together in one data layer. Artificial paths are not included in these data, eliminating flow network connectivity. Waterbodies are also line features, not polygons as in the NHD, defined by their shoreline boundaries. Waterbody and area river features were removed from the 2M data and supplanted by artificial paths from the 24K dataset for the research reported.

There are quality issues with the NHD in both the high and medium resolutions. One concern is the variability in compilation of the data when it was digitized. Compilation rules were established to create effective paper maps and occasionally the results do not match what is on the ground (Buttenfield & Hultgren, 2005). Also, these rules may also be unevenly applied. There are distinct boundaries defining the physical products that show differences in individual technique or understanding of the rules as shown in Figure 1.

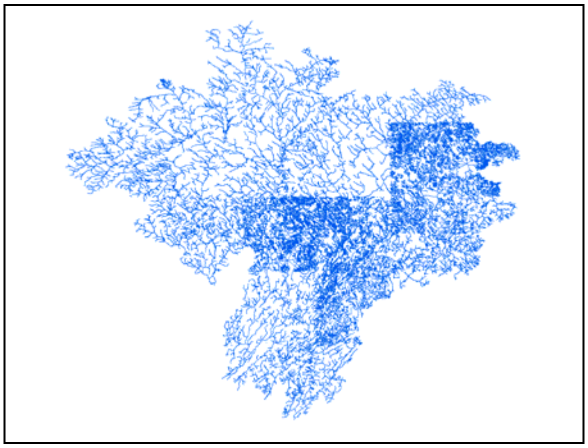

Figure 1. 24K data compilation issues. Lower Kennebec subbasin, Maine, at 1:2,000,000**.**

1.4 Subbasins

Most of the subbasins selected for this study are part of on-going research (Stanislawski, et al., 2005; Brewer et al. 2009) and are distributed across the continental US from each of six landscapes: humid mountainous, humid hilly, humid flat, dry mountainous, dry hilly, and dry flat. Three subbasins were selected to represent urban landscapes to evaluate if there are feature selection differences in areas with higher proportions of man-made features. One urban subbasin, New Haven, is also in a coastal area providing more diversity. Five more subbasins were chosen from Vermont to explore the relationships between local resolution data and the other data scales. The subbasins, locations, and their landscape types are in Table 2. The total length of flowlines in the subbasins ranges between 1500 and 6000 km.

Subbasin name	*State*	*NHD subbasin*	*Regime*
Upper Suwannee	FL, GA	03110201	Flat Humid
Lower Beaver	UT	16030008	Flat Dry
Pomme de Terre	MO	10290107	Hilly Humid
Lower Prairie Dog Town Fork Red	TX	11120105	Hilly Dry
South Branch Potomac	WV	02070001	Mountainous Humid
Piceance-Yellow	CO	14050006	Mountainous Dry
Upper Chattahoochee	GA	03130001	Urban (Atlanta)
Cahokia-Joachim	IL	07140101	Urban (St Louis)
Quinnipiac	CT	01100004	Urban (New Haven)
Passumpsic	VT	01080102	Hilly Humid
White	VT	01080105	Hilly Humid
Otter	VT	02010002	Hilly Humid
Winooski	VT	02010003	Hilly Humid
Lamoille	VT	02010005	Hilly Humid

Table 2. Study subbasins and landscape types.

Datasets were downloaded from the USGS NHD Geodatabase viewer and *The National Map* website. An NHD Generalization Toolbox being developed by the USGS National Geospatial Technical Operations Center (NGTOC) and the Center for Excellence in Geospatial Information Science (CEGIS) (Stanislwaski, 2010) was used in this research.

Before examining the data, some "cleaning" was required to make effective comparisons. The New Haven subbasin 24K data included coastline features and artificial path segments that were in the Atlantic Ocean. These features were deleted from the attribute table as they are not part of the flow network. Some of the subbasins had artificial paths that extended beyond their boundaries. Segments were selected that had their centroid within the subbasin. This mitigated the need to recalculate segment lengths in the attribute table as would have been required by clipping. Since the 2M flowline data do not have artificial paths, they were added using simplified 24K data. Shoreline and bank feature segments that define waterbodies and area features respectively were removed and then 24K artificial paths were added.

There are noticeable differences between the 24K and 100K data in some of the subbasins. Utah has a connector flowline in the 24K data not present in the 100K data that is given significant drainage area. The 24K data in Texas appears to be missing flowlines in one corner of the subbasin that are present in the 100K.

2. METHODS

Flowline pruning was accomplished through the Simple Network Pruning tool in the NHD Generalization toobox. This tool prunes stream segments based on upstream drainage area (UDA) until the desired stream density is reached. Density is measured in kilometers of channel length per square kilometer of area. UDA is estimated using Thiessen polygons for each flowline segment. The tool was run several times for various density measures for generalizing 24K to 100K and 2M as well as 5K to the 24K, 100K, and 2M scales. All densities are multiples of 0.05. Shorter tributaries were then removed using the Prune Short Dangling Tributaries tool. Flowlines having from nodes that are "dangling" (no upstream feature) and are too short to be depicted at the derived scale are removed in this step. This additional round of feature selection after density-based pruning drove repetitive processing to identify densities that would provide the best results. Although an expected density could be calculated based on the Radical Law, taking into account line simplification, it was not possible to estimate the reduction in channel length from removing these dangling tributaries.

Line simplification using the Bend Simplify algorithm in ArcGIS 9.3.1 was then applied to the pruned coverages using Tobler's (1987) rule for minimum detection. The minimum size an object, in meters, that can be detected is calculated by dividing the scale denominator by 1,000. I extended this to calculate the minimum size of a line's curve that could be detected. Thus, pruned 5K data were simplified using 24m, 100m, and 2,000m tolerances for 24K, 100K, and 2M comparisons respectively. 24K data used 100m for deriving 100K results and 2,000m for 2M. Comparisons of the generalized data to Radical Law expectations resulted in a higher percentage of the original features being retained which is supported by Stanislawski et al. (2009). They identified the need for more features and length than expected to generalize the 24K data to 100K and that the resultant 100K data generalized to 500K was too sparse. Densities with the simplified flowline length that best matched Radical Law expectations were identified for validation measures.

I did not use density partitions for two reasons. The first is that partitioning can improperly create small sub-networks resulting in topological inconsistencies in the pruning results (Stanislawski et al., 2009). The second reason is that compilation inconsistencies in both the 24K and 100K data would likely be maintained if partitioning was used. Although density partitioning will produce a product that more accurately represents natural variations, it will also keep artificats of over-compilation. I feel that subbasins create a sufficient partioning system based on natural differences.

2.1 Validation

Evaluating the resulting generalized data was accomplished by calculating a Coefficient of Line Correspondence (CLC) between the generalized data and the existing data (Buttenfield et al., 2010). For example, 24K data that was generalized to 100K was compared to the existing NHD 100K data. The calculation "is the ratio of the sum of the lengths of matching lines divided by the sum of the lengths of matching lines plus the sum of the lengths of omission and commission errors" (Stanislwaski, 2010). Omission errors are those features in the benchmark data that are not included in the derived data. Commission errors are those additional features in the derived data that are not in the benchmark data. CLC's were calculated for flowline densities previously identified based on Radical Law expectations. The overall goal was to maximize the correlation while keeping differences in omission and commission to a minimum. For those subbasin's whose omission and commission errors were widely divergent, additional CLC calculations were performed for other densities. Results for generalized data to 2M were evaluated using visual methods as existing 2M data were not compatible with the validation tool. Densities were selected based on visually balanced omission and commission rates.

3. RESULTS

Examining flowline lengths in each subbasin (excluding Vermont) at all scales indicated some general patterns exist in the percentage of length maintained. Beginning with 24K data, the Radical Law predicts 49.0 percent will be maintained at 100K and 11.0 percent will be present at 2M. The average flowline length retained in the 100K data was 52 percent. However, there was only 6.6 percent of the 24K length in the 2M data. Comparing the 100K to 2M for these subbasins also fell short of Radical Law expectations as 12.7 percent was maintained while 22.4 percent was expected.

Comparing the flowline lengths in Vermont also indicated some general patterns. While the expected percent of features in the 24K data compared to the 5K data is 44.7, the average percent of features remaining is 71.5. There are greater than expected flowline lengths in the 24K data for all five subbasins, although Otter is noticeably lower than the others at 59.3 percent. The Radical Law estimates 22.4 percent of the 5K data's length would be displayed at 100K, but all five subbasins have at least that much, averaging 28.2 percent, and the White subbasin has almost twice that level. However, there is less length retained in the 100K data than expected when compared to the 24K data. In comparing the 5K to 2M, the expected level is 5 percent. The data show an average of 7 percent of the length remains. In comparing the 24K to the 2M, 11.0 percent of the channel length is expected with the existing relationship at 9.9 percent.

The patterns in the data indicate that there could be an additional factor added to the equation to address the differences between what the Radical Law predicts and what is observed in the data. Three scale transition points exist: between local (5K) and large scale (24K) layers, large to medium (100K), and medium to small (2M). We propose a new factor called the *Constant of Flowlines* (C_f) to explain variations from the basic equation at these transition points. The new equation for flowline comparisons is:

$$n_f = n_a C_f \sqrt{M_a / M_f} \qquad (2)$$

There are three possible values:
C_{f1} = 1 for large scale to medium scale comparisons
C_{f2} = 1.7 for local scale to other scale comparisons
C_{f3} = 0.6 for comparisons to, but not within, small scales

One argument against this factor is that it is not scale-based like the basic equation or other factors proposed by Töpfer & Pillewizer (1966). The scale-based equivalent is approximately the cubed root of M_f/M_a (24K/4.8K) for the 5K to 24K comparison. The cubed root of 100K/5K is 2.7. This relationship would break the mathematic flow between scales relationships. For example, there should be 76 percent of the 5K data remaining when generalized to 24K, and 49 percent of that when further generalized to 100K leaving 37.3 percent. The result should be the same if 5K is generalized directly to 100K, but if a scale-based adjustment is used, the basic equation's expectation of 22.4 percent is increased to 60.4 percent. Therefore, numeric constants are proposed.

Yet, the new factor creates a disparity between the overall expected and observed 5K to 100K comparisons. The White subbasin is a good example of how this new factor could work. There is 73.6 percent of the 5K channel length displayed in the 24K data which is similar to the new expected level of 76.4 percent. There is also 40.6 percent in the 100K data, again similar to the new expected level of 38 percent. Additionally, comparisons between the 24K data and the 100K data are close to the expected results from the basic equation with 55 percent of channel length displayed and 49 percent expected. For the other four subbasins, increases in the 100K flowlines would make the relationships with the new 5K equation and existing 24K equation closer to expected levels.

3.1 Density Pruning

The density required for each scale change is shown in Table 3. Channel Length % represents the total length of flowlines after

Subbasin	24K to 100K Density	Flowline Length % (49%)	24K to 2M Density	Flowline Length % (6.6%)
UT	0.50	49.5	0.10	7.8
CO	0.70	48.7	0.15	6.8
TX	0.70	49.8	0.15	9.7
WV	0.80	48.9	0.15	7.6
FLGA	0.30	48.8	0.05	7.0
MO	0.80	49.7	0.10	5.7
New Haven	0.85	49.1	0.15	6.0
Atlanta	0.65	49.6	0.10	6.6
St. Louis	0.70	48.3	0.10	5.9

Table 3. Selected densities for pruning and remaining flowline percentages. Expected percentages are shown in parentheses.

toolbox's inability to prune the subbasin to a density of 0.1. density pruning, pruning short dangling tributaries, and line simplification. Areas with flat terrain, in this case Utah and Florida/Georgia, have lower densities at both the 100K and 2M scales. Texas' higher channel length for 2M is due to the Atlanta and St. Louis have lower densities at 100K than the other humid subbasins while New Haven has a higher density in comparison. Table 4 shows the densities that best meet Radical Law predictions for the Vermont subbasins excluding White, which could not be processed.

Subbasin	5K to 24K Density	Flowline Length % (76%)	5K to 100K Density	Flowline Length % (38%)	5K to 2M Density	Flowline Length % (5.1%)
Lamoille	1.5	75.7	0.75	35.7	0.15	5.5
Otter	1.7	76.5	0.80	36.1	0.20	6.5
Passumpsic	1.2	78.2	0.60	37.2	0.15	6.3
Winooski	1.4	76.3	0.70	36.1	0.15	6.1

Table 4. Selected densities for network pruning and remaining channel length for Vermont subbasins

3.2 Validation

Results were validated using Coefficient of Line Correspondence (CLC) calculations for those data layers that could be compared, which are the 24K to 100K for the nine primary subbasins, and 5K to 24K for the Vermont subbasins. The 5K data were compared to the existing 100K layers in Vermont using CLC although the 100K channel length comparisons to the 5K and 24K data foretold lower CLC values. Visual evaluations were used for comparing pruned results to the 2M data since that data is not properly attributed to allow use of the NHD Generalization Toolbox.

As density partitions were not used in this analysis, a CLC of 0.7 with a balance between omission and commission errors was determined to be sufficient. The goal was not necessarily to replicate the 100K data, but to find a density that best met adjusted Radical Law expectations informed by CLC calculations. Figure 5 shows the calculated densities and CLC values for each subbasin. Subbasins falling below the 0.7 threshold include Utah, Texas, and St Louis. This is not surprising for Texas and St Louis given the differences between their existing densities (0.5 and 0.9 respectively) and Radical Law densities (0.7 for both). Addressing the disparities in the 24K and 100K data for Utah and Texas would improve the CLCs for those areas. St. Louis simply has more data in the 100K than expected (65.6 percent versus 49 percent expected).

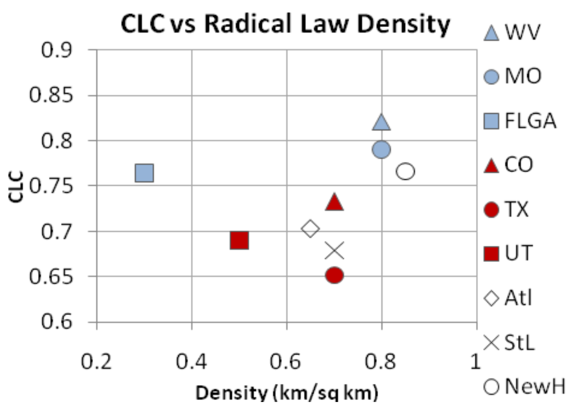

Figure 5. CLC values for Radical Law determined densities

Subbasin	Radical Law Density	Visual Density	Subbasin	Radical Law Density	Visual Density
WV	0.15	0.15	Atl	0.10	0.10
MO	0.10	0.15	StL	0.10	0.15
FLGA	0.05	0.10	New Haven	0.15	0.15
CO	0.15	0.20	Passumpsic	0.15	0.25
TX	0.15	0.20	Lamoille	0.15	0.20
UT	0.10	0.10	Winooski	0.15	0.25
			Otter	0.20	0.20

Table 6. 2M adjusted Radical Law expected densities and visually selected densities.

Each subbasin's density can be calculated and derived from a source dataset, either 5K or 24K for the selected subbasins. Using line expansion factors (Stanislawski et al., 2009) and removing short dangling tributaries leads to estimated changes from initial and final pruning densities of 10 percent. Similarly, 2M densities can be calculated and then adjusted upwards. I recommend adding 0.05 to estimated humid 2M densities and 0.1 to dry estimations. This is greater than the line expansion factor of 1.25

proposed by Stanislawski et al. (2009). These factors tend to create initial pruning densities comparable to the visual recommendations. The Vermont subbasins will be more aggressively pruned due to the greater than expected density in the existing 2M data there. Interestingly, the flat subbasins should both be treated as humid areas. This is based on the adjusted Radical Law estimation and visual comparison results. This may just be particular to the Utah subbasin.

3.3 Test case with additional subbasin

The Lower Penobscot subbasin in Maine, 01020005, was chosen as a test case for these various possibilities. High resolution flowlines exhibit compilation errors in the southwest portion of the subbasin. The 24K density is comparitively low for a humid area at 1.05 km/sq km, but similar to densities found in the Vermont subbasins. The existing 100K density is 0.65 km/sq km with 61.8 percent of the 24K channel length. The 24K data were pruned to 0.57 km/sq km which is the density for the adjusted Radical Law expectation plus 10 percent. There was 56.6 percent of the 24K channel length remaining at the 0.57 density after removing dangling tributaries and simplifying the line. The CLC was calculated to be 0.724. The primary driver for the low CLC was the 0.22 rate of omission because of the higher than expected density in the existing 100K data (commission was 0.05). Simple network pruning removed over compilation from the 24K data, but it also reduced natural density variations.

For the 2M evaluation, the 24K data were pruned to a density of 0.12 for the adjusted Radical Law expectation (6.6 percent of the 24K data) plus 0.05 km/sq km. The existing density was calculated to be 0.12 km/sq km. I believe the 0.12 pruning density provides a visually acceptable product while maintaining 8.6 percent of the 24K channel length. This result is lower than both the existing data and what the basic Radical Law equation would estimate at 11 percent. However, it is higher than, and closer to, the adjusted Radical Law expectation of 6.6 percent.

4. CONCLUSION

New constants are proposed to address differences in observed flowline retention and Radical Law expectations for local to large scale generalization and for generalization into the small scale regime. The Constant of Flowlines for local scale work (C_{f2}) adjusts the Radical Law upwards by a factor of 1.7 for work with local scale data. Similarly, the Constant of Flowlines for small scales (C_{f3}) decreases Radical Law expectations by a factor of 0.6 for work with small-scale flowlines. The threshold for small scale is not established here, but implemented generalizing flowlines to 2M.

These new constants are implemented in flowline generalization through density pruning. Multiple densities were used in each subbasin to explore the impact of removing short dangling tributaries from the pruning results and simplifying the lines. An increase in density of 10 percent is suggested for generalization to 100K for all subbasin types. Generalization to 2M used increases of 0.05 for humid subbasins and 0.1 for dry subbasins. Flat subbasins achieved better visual results using the humid adjustment.

ACKNOWLEDGEMENTS

Thank you to Larry Stanislawski for his assistance in running the NHD Generalization Toolbox that made this research possible. Thanks also to the Gould Center for providing the hardware and work space.

REFERENCES

Bobzien, M., Burghhardt, D., Petzold, I., Neun, M., & Weibel, R., 2008. Multi-representation Databases with Explicitly Modeled Horizontal, Vertical, and Update Relations. *Catography and Geographic Information Science, 25* (1), 3-16.

Brewer, C. A., Buttenfield, B. P., & Usery, E. L. (2009). Evaluating Generalizations of Hydrography in Differing Terrains for The National Map of the United States. *Proceedings of the 24th Internaional Cartographic Conference.* Santiago, Chile.

Buttenfield, B., & Hultgren, T., 2005. Managing Multiple Representations of “Base Carto” Features: A Data Modeling Approach. *Proceedings, AUTO-CARTO 2005.* Las Vegas, Nevada.

Buttenfield, B. P., & McMaster, R. B. (Eds.). (1991). *Map Generalization: Making Rules for Knowledge Representation.* Essex, England: Longman Scientific & Technical.

Buttenfield, B., Stanislawski, L., & Brewer, C., 2010. Multiscale Representations of Water: Tailoring Generalization Sequences to Specific Physiographic Regimes. *GIScience 2010*, Zurich, Switzerland.

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

Goodchild, M. F., 1989. Tiling Large Geographical Databases. *Symposium on the Design and Implementation of Large Spatial Databases*, 137-146. Santa Barbara, California: Springer: Berlin.

Hardy, P., & Lee, D. (2005). GIS-Based Generalization and Multiple Representation of Spatial Data. *Proceedings, International CODATA Symposium on Generalization of Information.* Berlin, Germany.

Jiang, B., & Harrie, L., 2003. Cartographic Selection Using Self-Organizing Maps. *ICA Commission on Generalisation and Multiple Representation* (pp. 1-7). Paris: ICS.

João, E. M., 1998. *Causes and Consequences of Map Generalisation.* Bristol, PA: Taylor & Francis, Inc.

Kadmon, N. (1972). Automated Selection of Settlements in Map Generalisation. *The Cartographic Journal*, *9* (2), 93-98.

Leitner, M., & Buttenfield, B. P., 1995. Acquisition of Procedural Cartographic Knowledge by Reverse Engineering. *Cartography and Geographic Information Systems*, *22* (3), 232-241.

Li, Z., & Choi, Y. H., 2002. Topographic Map Generalization: Association of Road Elimination with Thematic Attributes. *The Cartographic Journal*, *39* (2), 153-166.

MassGIS., 2007. *DEP Wetlands (1:12,000)*. http://www.mass.gov/mgis/wetdep.htm (accessed 12 July , 2010).

Shea, K., 1988. Cartographic Generalization. *NOAA Technical Report NOS 127 CGS 12*. Reston, Virginia, United States of America: US Department of Commerce.

Stanislwaski, L., 2010. Operation Guide for The National Hydrography Dataset (NHD) Generalization Toolbox. USGS.

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., Finn, M. P., & Roth, K., 2009. Stratified Database Pruning to Support Local Density Variations in Automated Generalization of the United States National Hydrography Dataset. *Proceedings of the 24th International Cartographic Conference (ICC2009).* Santiago, Chile.

Thornton, L. L., 2008. New Technology for Hydrologic Features. *National Monitoring Conference*. Atlantic City, New Jersey.

Töpfer, F., & Pillewizer, W., 1966. The Principles of Selection. *The Cartographic Journal*, *3* (1), 10-16.

USGS, 2010a. *National Hydrography Dataset*. http://nhd.usgs.gov/ (accessed 12 Sep, 2010).

USGS, 2010b. *NHD Stewardship*. http://usgs-mrs.cr.usgs.gov/steward/st2maps.html (accessed 20 Aug 2010).