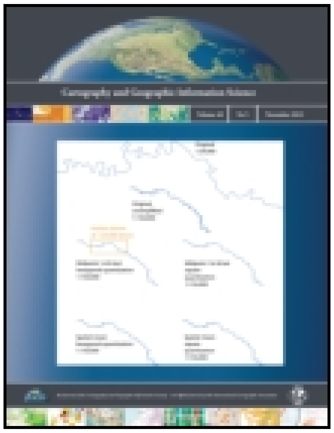

Cartography and Geographic Information Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcag20

A method to generalize stream flowlines in small-scale maps by a variable flow-based pruning threshold

Michael Tinker [a], Peter Anthamatten [b], Jeff Simley [a] & Michael P. Finn [a]

[a] United States Geological Survey, National Geospatial Technical Operations Center, Box 25046, MS 505, Denver, CO, 80225-0046, USA

[b] Department of Geography and Environmental Sciences, University of Colorado Denver, Denver, CO, USA

Published online: 28 May 2013.

Cartography and Geographic Information Science, 2013
Vol. 40, No. 5, 444–457, http://dx.doi.org/10.1080/15230406.2013.801701

A method to generalize stream flowlines in small-scale maps by a variable flow-based pruning threshold

Michael Tinker[a]*, Peter Anthamatten[b], Jeff Simley[a] and Michael P. Finn[a]

[a]*United States Geological Survey, National Geospatial Technical Operations Center, Box 25046, MS 505, Denver, CO 80225–0046, USA;* [b]*Department of Geography and Environmental Sciences, University of Colorado Denver, Denver, CO, USA*

(*Received 15 August 2012; accepted 23 April 2013*)

The aim of this paper is to explore and describe a method of automated generalization designed to produce a map which strikes a balance between cartographic and hydrologic representations. Following a discussion of scholarly literature on generalization, we describe a novel method for automated generalization of hydrographic stream data, using the National Hydrography Data Set (NHDPlus) as an example.

Traditional hydrography shows a fairly uniform density of stream flowlines over space. While this is pleasing to the eye, traditional methods tend to under-represent rivers in humid areas and over-represent them in arid areas. We address this problem through a method in automated generalization to produce a high-quality presentation of hydrographic data, suitable for display as a wall map or in an atlas. Streams are pruned based on a variable flow threshold, derived from the local mean annual precipitation by a regression equation.

After running the model using different parameters, we produce a more satisfactory portrayal of stream networks in the United States that communicates the flow of water through rivers and reflects the regional climate. Specific advantages in generalizing with variable flow threshold include (1) the method allows for fine gradations in output scale; (2) the output maps tend to minimize density variations in the raw data; (3) the subjective criteria are easily derived; and (4) the method can be performed rapidly on large data sets, as long as the stream data has been enriched with reliable flow rates.

Keywords: generalization; hydrography; National Hydrography Data Set; NHDPlus

Introduction

Hydrography is a conspicuous and complex natural theme on topographic maps due to its large number of features (Savino et al. 2011). It must be drawn with strict tolerances in order to appear well integrated with the terrain and is very sensitive to changing scales (Buttenfield, Stanislawski, and Brewer 2011). A significant amount of previous scholarly research has explored systematic methods for generalizing hydrography suitable for automation.

Hydrographic generalization methods are often based on stream ordering schemes. Recent methods emphasize on generalizing by stream drainage area, several of which we briefly review in the following section. In this paper, we employ the National Hydrography Data Set (NHDPlus) to explore a novel generalization method that uses variable flow thresholds. NHDPlus is a publicly available, comprehensive vector spatial data set containing hydrography for the United States at 1:100,000 scale, provided by the US Environmental Protection Agency (EPA). Originally conceptualized as a tool for surface-water modeling, the data are used for a variety of applications, including cartography (EPA, USGS, and Horizon Systems Corporations 2010a, 2010b). NHDPlus data are ideal for exploring generalization methods because they include many additional attributes, such as Strahler stream order and mean annual estimates of natural flow and velocity for all streams.

The intent of this generalization method is to improve upon traditional hydrographic presentation to produce a river map that does not under-represent rivers and streams in high precipitation areas or over-represent them in arid areas. This method represents an advancement over previous work by using mean annual precipitation to derive the flow rate pruning thresholds, serving to emphasize on a more accurate rendering of precipitation and runoff regime. While we use NHDPlus to develop and discuss this work, these methods may be applied to other hydrography data sets, as long as they have been enriched with flow rates (see, for example, ESRI 2012 or Global Runoff Data Center 2013).

Related work

Any method of generalization must address two key issues: w*hich* features to retain (or delete) and *how many* features to retain (or delete), which can be represented by the *selection* and the *target*. The selection process can address either retention or deletion, but the target is generally a number of features to retain. Töpfer's radical law (or principle of selection) has been

*Corresponding author. Email: mdtinker@usgs.gov

adapted for use in hydrography. The original formula is: $n_f = n_a \times \sqrt{M_a/M_f}$, where n_f is the number of features shown on the derived map; n_a is the number of features on the source map; M_a is the scale denominator of the source map; and M_f is the scale denominator of the derived map (Töpfer 1966). Some researchers have used or modified the radical law to establish targets (e.g., Ai, Liu, and Chen 2006; Gardiner 1982; Zhang 2007; Stanislawski 2009; Wilmer and Brewer 2010; Mazur and Castner 1990). While this principle helps to answer the question of *how many* features to retain when generalizing to smaller scales, it provides no guidance about *which* features to retain.

Generalizing by stream order

Most hydrographic generalization methods are based on the stream ordering schemes of Horton (1945), Strahler (1957), or Shreve (1966). In Horton's scheme, a first-order stream is a headwater with no tributaries; a second-order stream is represented by the confluence of two first-order streams, and so on. At confluences, the shorter stream or the stream joining the parent stream at a greater angle is assigned a lower order. Stream order increments only when two streams of the same order flow together. Once all confluence-to-confluence segments of the stream network have been ordered, the main stem of each river, from headwater to outlet, is assigned the same order as its outlet. The cartographer must decide which segments comprise the main stem of the river (Mazur and Castner 1990), though Horton's original criteria were to select the "longest and straightest path" (which implies curvilinear continuity rather than the actual straightness of line). Strahler modified Horton's method by abandoning this subjective retracing stage. Shreve then further modified Horton's scheme to account for the number of upstream tributaries. In Shreve's scheme, stream order increments whenever two streams join, regardless of the tributary length or angle of confluence.

Horton's method may be considered a hierarchy of rivers, while Strahler and Shreve's methods are hierarchies of segments. Generalization based upon the ordering scheme is well represented in recent literature (Catlow and Du 1984; Mazur and Castner 1990; Richardson 1994; Touya 2007; Zhang 2007; Thomson and Brooks 2000; Savino et al. 2011) and other scholars have since refined or expanded upon these methods.

Mazur and Castner (1990) proposed a method to arbitrarily select the Horton orders and compare the results to other published small-scale maps. Richardson (1994) calculated the number of streams retained after generalizing by both ordering schemes, setting the target number of streams by selecting a percentage of the number of original streams. Thomson and Brooks (2000) described a method of selecting by stroke "a curvilinear segment that can be drawn in one smooth movement and without a dramatic change in style." In their method, once strokes are identified, attributes such as their length or representative class can be derived and sorted. Selection then proceeds by arbitrarily deleting strokes on the basis of their attributes. The authors found that selection by strokes generally produces the "longest and straightest path", which is essentially Horton's original criterion for defining the main river stem. Touya (2007), following Thomson and Brooks (2000), used a different set of criteria to determine the stroke, including river name, priority of permanent (perennial) over intermittent streams, priority of artificial path through irrigation zones, river length, and angle of confluence. Once strokes are determined, the Horton orders are assigned. Selection retains higher order strokes and lower order strokes are retained only if they meet a minimum length threshold.

Additional generalization methods have been devised by drawing directly from Strahler's work. Catlow and Du (1984) argued that simple deletion of all low-order streams does not result in acceptable cartography because all headwaters are indiscriminately culled and single-river systems are eliminated. They used length, therefore, as a secondary criterion. First-order single-river systems are retained if they are longer than an arbitrary length threshold. Savino et al. (2011) determined the main course of the river by starting from streams with the largest Strahler order and then moving upstream. They identified the main stem by evaluating attributes such as total distance to the furthest upstream source, total number of branches uphill, and width. River courses that are shorter than an arbitrary length threshold are pruned, and density of the remaining rivers is also considered. Arbitrary buffers are built around each river course, and selection proceeds recursively by pruning river courses below a threshold percentage of overlap.

Zhang (2007) described a method in which stretches of river are numbered, as in the Shreve scheme, but then the main stem of each river is assigned the same order as its outlet, as in the Horton scheme. While both Zhang's and Horton's methods depend on an arbitrary identification of the main stem, Zhang's method also applies arbitrary length thresholds to each river group. In Zhang's method, the target is set with a modified application of the radical law, based on the number of tributaries.

Limitation of Strahler order method to generalize the NHDPlus

The NHDPlus can be generalized by selection of Strahler stream order. The greatest stream order value in the conterminous US is 10, found on the main stem of the lower Mississippi River. In the Pacific Northwest, the greatest value is nine, found on the lower Columbia River, below its confluence with the Snake River (Pierson et al. 2008).

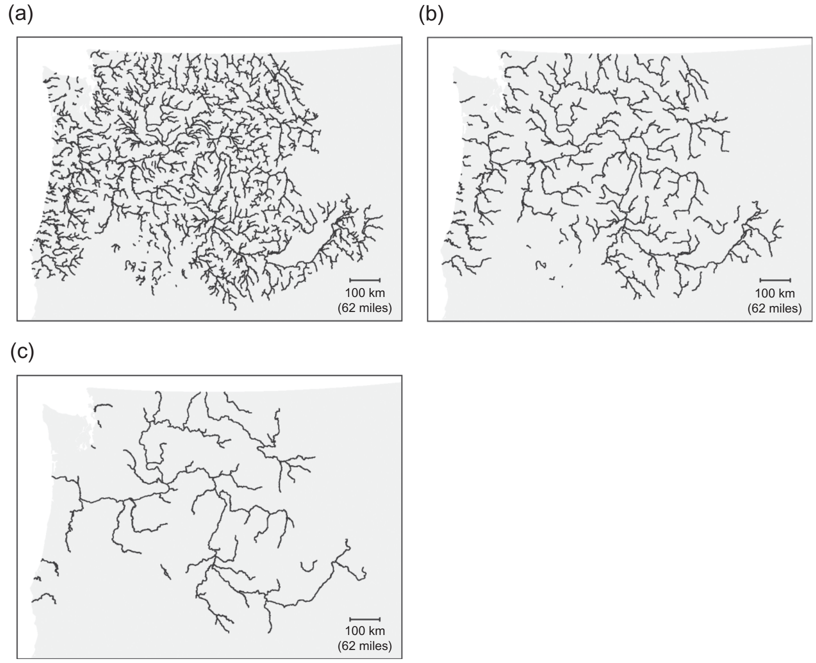

Figure 1. (a, b, and c) US Pacific Northwest (NHDPlus region 17), successively generalized by Strahler orders 4, 5, and 6 or greater.

In Figure 1, the US Pacific Northwest (NHDPlus region 17) is shown to be successively generalized by Strahler orders 4, 5, and 6. Figure 1a shows only streams of order 4 or greater, which could work for display at 1:500,000. Figure 1b shows only streams of order 5 or greater; while this could work for display at 1:1,000,000, it appears to be missing important coastal streams. Figure 1c, which shows streams of order 6 or greater, has lost so many rivers that it would not be useful for a small-scale map. This serves to demonstrate that the NHDPlus can be generalized by Strahler order, but the resulting maps show a very wide range in scale. Strahler order is not fine enough to generalize the 1:100,000 NHDPlus to more than one or two output scales.

Figure 1b and c also show how generalizing by order may indiscriminately cull important rivers. The coastal region of the US Pacific Northwest is one of the wettest areas in the United States, yet few coastal rivers appear in the generalized maps because they were of lower order. In reality, they are likely to be important rivers with high flow and should remain on the map.

Generalizing by drainage area

Other researchers have devised generalization methods based on drainage area, rather than stream order. Ai, Liu, and Chen (2006) recursively culled flowlines whose associated watershed area was less than a specified threshold, working from the premise that watersheds that drain larger areas gather more precipitation and should be retained in the final product.

Stanislawski (2009) used a similar technique by recursively eliminating reaches whose associated cumulative upstream drainage areas were less than a specified threshold. Stanislawski's process begins by assigning an upstream drainage area to all reaches (confluence-to-confluence sections). The upstream drainage areas are estimated by a technique based on the summing of Thiessen polygons generated along evenly-spaced points on the flowlines (Stanislawski et al. 2007). The flowlines in the original data often show wide variation in density, due to natural variation in local physiography, climate, or data collection issues, which are accounted for by partitioning the original data into a range of density classes. Selection proceeds by recursively culling reaches with upstream drainage area that are less than the threshold, which is derived from the density class. The target density is determined by application of the radical law or by comparing the output to a benchmark data set (Buttenfield, Stanislawski, and Brewer 2011). Stanislawski (2009) modified the radical law to compute a relationship based on stream length. The modified equation is

$$\text{length}_{\text{target}} = \text{length}_{\text{source}} \times \sqrt{RF_{\text{source}}/RF_{\text{target}}}\ ,$$

where RF is the denominator of the representative fraction.

Research gaps

Gardiner (1982) proposed three objective criteria through which hydrography can be generalized: (1) discharge characteristics such as mean flow, peak flow, or recurrence intervals; (2) biological aspects such as water quality; and (3) morphological characteristics such as area of catchment basin, depth, or stream order. The methods discussed above largely generalize from morphology; however, both Ai, Liu, and Chen's (2006) and Stanislawski's (2009) methods use a stream's associated drainage area as a proxy for flow. Ai estimated the area of the watershed for each river course, while Stanislawski estimated the upstream drainage for all reaches. The basic implication is that streams with less flow can be pruned during generalization and that streams with greater flow may be retained. In this work, we address shortcomings of previous generalization efforts by directly incorporating stream flow into the automated generalization process.

Materials and methods

The method described in this paper does not smooth or simplify lines, nor does it generalize water bodies. Methods for smoothing and simplifying linear networks have been studied extensively (e.g., McMaster and Shea 1992; Regnauld and McMaster 2007) and automated systems are available in many geographical information systems (GIS) software packages. Generalization of water bodies can be partially accomplished by scale-dependent minimum area thresholds, which is a relatively straightforward task (Catlow and Du 1984; João 1998; Stanislawski 2009; Buttenfield, Stanislawski, and Brewer 2011). Complete generalization of waterbodies that performs partial pruning of polygon features and boundary simplification, while maintaining network integrity, is a complex task not fully addressed by the methods reviewed here.

We describe a method to generalize flowlines by pruning streams based on a flow threshold. This pruning threshold is tied to the local precipitation by a regression equation. The slope of the regression line can be arbitrarily adjusted by the cartographer to change the output scale for generalization. Cartographers often use an arbitrary metric to generalize hydrography (Horton 1945; Catlow and Du 1984; Mazur and Castner 1990; Thomson and Brooks 2000; Touya 2007; Zhang 2007; Savino et al. 2011).

North American Atlas 1:10M-scale hydrography

The North American Atlas is a joint project of the Atlas of Canada, Mexico's National Institute of Statistics and Geography and the National Atlas of the United States. It is available for free download from the National Atlas (USDOI 2006). North American Atlas maps are intended for display at 1:10,000,000 scale (Figure 2). The hydrography layer was revised and re-released in 2006. For this project, the North American Atlas hydrography is used as the benchmark.

Figure 2. North American Atlas hydrography, intended for display at 1:10,000,000 (USDOI 2006).

While this map shows cartographic balance, the hydrography is misleading. For example, arid areas in the West display a similar density of flowlines as wetter areas in the East. Yet, many western streams have low flow rates and some are intermittent. For those not familiar with climate and hydrology, the map can be misleading.

PRISM precipitation: annual climatology (1971–2000)

The Parameter-elevation Regressions on Independent Slopes Model (PRISM) is a precipitation model, produced at Oregon State University, which shows the mean annual precipitation from 1971–2000 (Figure 3). PRISM is widely considered to be the definitive data set for annual temperature and precipitation in the United States (Oregon State University PRISM Group 2001). It offers an opportunity to improve the representation of where one is more likely to find, or not find, water. We use these PRISM precipitation models to inform the generalization process discussed in this paper.

NHDPlus

We use NHDPlus data to demonstrate our methods. NHDPlus is a product of the USEPA, freely available as a download from Horizon Systems (EPA, USGS, and Horizon Systems Corporations 2010b), conceived as a platform for surface-water modeling in pollution control analysis. It incorporates features of the 1:100,000-scale National Hydrography Data Set (NHD; Simley and Carswell 2009), the National Elevation Data Set (NED; Gesch et al. 2009), and the national Watershed Boundary Data Set (WBD 2012). The NHDPlus includes many additional attributes such as Strahler stream order, cumulative drainage areas, land cover, temperature, precipitation distributions as well as mean annual estimates of natural flow and velocity for all streams. While the 1:100,000-scale US Geological Survey (USGS) NHD data have already been generalized from the 1:24,000 scale by traditional manual compilation methods, the inclusion of flow data into the NHDPlus makes generalizing by flow possible to scales less than 1:100,000. The NHDPlus team suggests using mean annual flow as a way to thin the flowline network, but do not illustrate a method (EPA, USGS, and Horizon Systems Corporations 2010a).

Features in the flowline network include streams/rivers, canals/ditches, pipelines, connectors, and artificial paths. A connector represents a path where flow is known to exist but was not collected on the original data, or its precise location has not been determined. An artificial path connects the network where a polygonal waterbody feature intersects a flowline. Feature types that may include artificial paths include lakes, ponds, streams, rivers, areas of complex or braided channels, washes, canals, ditches, estuaries, ice masses, playas, reservoirs, swamps, bayous, and marshes.

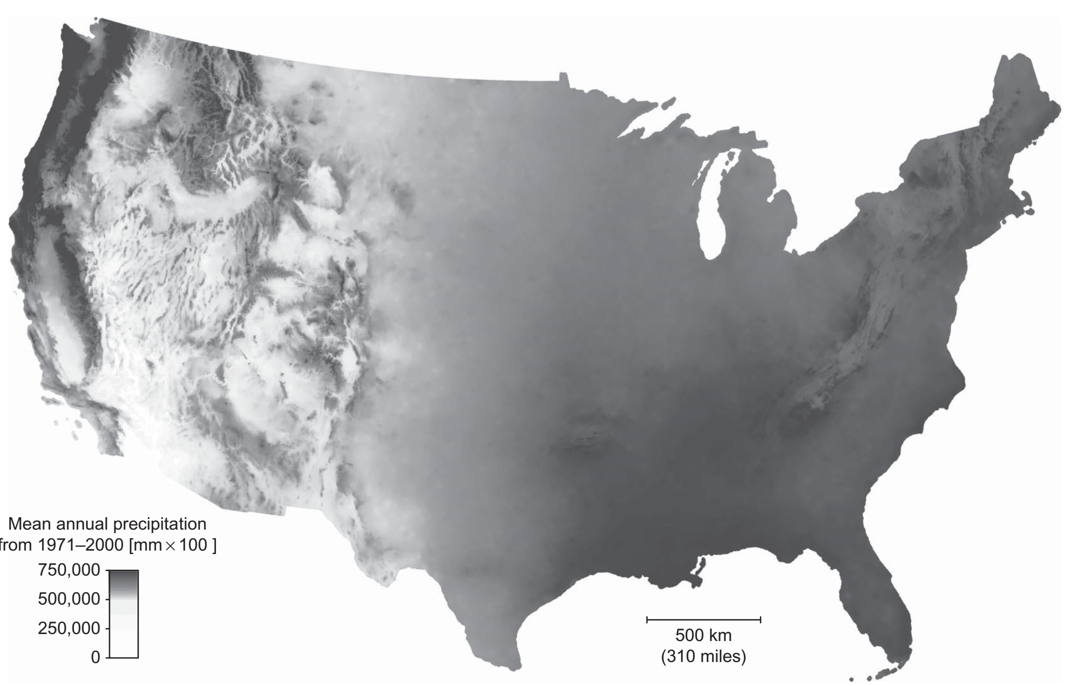

Figure 3. PRISM model, mean annual precipitation from 1971–2000 [mm × 100] (Oregon State University PRISM group 2001).

NHDPlus data for the conterminous United States are available in 18 hydrologic *regions*. The boundaries of the regions in the NHDPlus are derived from the Watershed Boundary Dataset (WBD), a companion data set to the NHD. Watershed boundaries define the extent of surface-water drainage areas. The selection and delineation of hydrologic boundaries are determined by hydrologic principles, not favoring any administrative or special projects nor particular program or agency (WBD 2012).

The WBD broadly terms drainages as hydrologic units and each is assigned a unique Hydrologic Unit Code (HUC). Hydrologic units are nested in a multi-level hierarchical drainage system. The largest unit is called a *region* and is assigned a two-digit hydrologic unit code (HUC2). Each region is composed of hydrologic units called *subregions* which are assigned a four-digit hydrologic unit code (HUC4) and are then further divided into six-digit hydrologic units called basins (HUC6). This process of nesting hydrologic units continues through subbasins (HUC8), *watersheds* (HUC10), *and subwatersheds* (HUC12). USGS data stewards are currently conducting work to define smaller nested hydrologic units at the HUC16 level.

In the NHD model, a *reach* "is a continuous, unbroken stretch or expanse of surface water" or is "a significant segment of surface water that has similar hydrologic characteristics, such as a stretch of stream/river between two confluences or a lake/pond" (USGS 2000). Reaches are assigned *reachcodes* (EPA, USGS, and Horizon Systems Corporations 2010a), which enumerate the hydrologic unit in which the reach resides, to the level of the subbasin (HUC8). Reaches are further composed of smaller continuous features which are assigned a unique and permanent 10-digit *common identifier* (COMID).

Reachcodes are used primarily as data addressing system to link scientific information to the NHD, while COMIDs are used to manage the geometry in editing and flow networking. For example, if an inflowing tributary is added to a reach of river, the associated reachcode would remain the same, but the associated COMID would be retired and two new COMIDs would identify each of the two segments on either side of the confluence. A single reachcode can therefore be comprised of a single or multiple COMIDs. In this study, the process is performed on the geometry represented by the COMIDs, however, it may be useful to preserve the full reachcodes for additional work.

The 1:100,000-scale NHDPlus contains several attribute tables beyond what are available from the original NHD. One such table is the flow attribute table, which contains flow rates in cubic feet per second (cfs) for all COMIDs. The NHDPlus uses two methods to estimate mean annual flow: the Vogel method (Vogel, Wilson, and Daly 1999) and the Unit Runoff Method (UROM), which was developed for the National Water Pollution Control Assessment Model, derived from real annual flow rates from the Hydro Climatic Data Network (HCDN) of stream gages. The HCDN gages are minimally affected by flows from human activities such as dam releases or irrigation withdrawals (Slack, Lumb, and Landwehr 2006) and are representative of natural flow conditions (EPA, USGS, and Horizon Systems Corporations 2010a).

The NHDPlus team performed a general assessment of the UROM model by comparing the modeled flow values of 9990 points to 10-year averages for a set of USGS stream gages (EPA, USGS, and Horizon Systems Corporations 2010b), as reported by the National Water Information System (NWIS; USGS 2012). For our own characterization of the UROM model, we removed four outliers from their data and then calculated trend lines for all 18 scatter plots. The R^2 values range from 0.6259 to 0.9996, with an average R^2 value of 0.8915.

This project relies on the UROM flow rates to determine pruning reaches of streams with flow rates less than a threshold flow. Figure 4 shows all perennial flowlines in the NHDPlus data set with flow volume greater than 10 cfs. Although this representation is instructive, the cartographic balance of the map is poor. The density of the flowlines renders the map practically unreadable in the eastern parts of the United States, while the more arid areas in the western United States are comparatively sparse.

Figure 5 shows all streams with a flow threshold value 500 cfs. The higher threshold value improves the cartography in the East, but rivers are under-represented in arid areas in the West; large areas in Arizona and Nevada are completely devoid of hydrographic features and many tributaries of the upper Missouri River are missing. A single pruning threshold for the entire NHDPlus is insufficient to produce a cartographic product suitable for a wall map or atlas. Customizing the pruning thresholds to specific climatic conditions is useful because the morphology and flow of streams in different regions vary greatly, depending on many factors, such as the local precipitation or physiography.

Manually flow-generalized NHDPlus hydrography, 1:10M scale

We manually produced a flow-generalized small-scale wall map of the conterminous United States from the 1:100,000 scale NHDPlus, giving each of the 18 hydrological regions a unique base pruning threshold. These flow-volume thresholds are enumerated in a threshold index; a section of this table is shown in Table 1. In many cases, hydrologic regions have been divided into subregions, each with its own localized base-flow volume. The pruning threshold was chosen to emulate the North American Atlas 1:10M-scale

Figure 4. NHDPlus streams with flow greater than 10 cfs.

Figure 5. NHDPlus streams with flow greater than 500 cfs.

hydrography (Figure 6). The flow-generalized map represents a balance between cartographic and hydrologic representations.

Although the manual flow-generalized map emulates the 1:10,000,000 scale North American Atlas hydrography, the production process is time consuming and tedious. The primary goal of this project was to make the flow-generalization process repeatable, automatable, and applicable to other hydrographic data sets that contain flow data. In the next phase of the project, we correlated this initial arbitrary pruning threshold to local precipitation.

Table 1. Examples of threshold tables.

Region 11 (base cfs = 500)	
Subregion	Threshold (cfs)
1	50
2	100
3	100
4	75
5	75
6	50
7	500
8	50
9	50
10	50
11	500
12	50
13	50
Region 12 (base cfs = 200)	
Subregion	Threshold (cfs)
1	200
2	200
3	200
4	200
5	25
6	200
7	200
8	25
9	200
10	200
11	200
Region 13 (base cfs = 25)	
Subregion	Threshold (cfs)
1	50
2	25
3	25
4	25
5	25
6	25
7	25
8	25
9	25

Generation of an pruning regression line

The PRISM precipitation model (Oregon State University PRISM group 2001) provided mean annual precipitation data for the conterminous United States for the period of 1971–2000. In order to correlate the pruning threshold to local precipitation, the mean precipitation of the HUC 4, 6, and 8 polygons were determined with the *ArcGIS Zonal Statistics* tool (Figure 7). Separate tables for HUC 4, 6, and 8 polygons were generated and merged into one HUC precipitation table.

The final joined table contained 223 unique HUC polygons, each tied to its respective pruning threshold and mean annual precipitation. A scatter plot and its regression line were subsequently generated from this work (Figure 8). The resulting graph is a scatter plot of two ratios: mean precipitation per HUC (range 0–225,686 mm × 100) versus the arbitrary pruning threshold per HUC (range 0–500 cfs). The HUCs essentially drop out, resulting in a graph of pruning threshold per precipitation. Because the pruning threshold was arbitrary chosen, the scattered points are arranged in tiers. The range of pruning threshold values is narrow (0–500) compared to the wide range of precipitation values (0–225,686), so any regression line for this scatter plot will necessarily have a shallow slope:

Threshold (cfs) = 0.0035 × precipitation (mm × 100) – 60.021 (cfs)

This regression line equation enables the determination of a pruning threshold in cfs for any hydrologic unit, given its precipitation. Because the regression line has a *y*-intercept of –60.021 cfs, the pruning threshold drops to less than zero when the mean annual precipitation is less than 171.48 mm. A cumulative distribution of the PRISM precipitation data shows that about 5.2% of the data is less than 171.48 mm.

To achieve a representative hydrography, arid areas of the map should contain fewer flowlines than humid areas. However, in arid areas where the mean annual precipitation is less than 171.48 mm, the pruning threshold would be less than zero cfs and no streams would be pruned, leaving unpruned patches of flowlines in arid areas of the map. An example of this is found in a section of hydrologic region 16, shown in Figure 9. This problem can be addressed by dropping the intercept term from the regression equation. The modified regression equation is

Pruning threshold (cfs) = 0.0035 × precpitation (mm × 100)

Pruning that is driven by this equation achieves the desired results – namely, patches of unpruned flowlines do not appear in the arid areas. The slope of the regression line can be manipulated to achieve different cartographic results. Indeed, if the *y*-intercept can be dropped to improve the final hydrography, as shown in Figure 9, it is reasonable to assume that other adjustments to the slope of the regression line could also be useful.

Automated generalization

An automated generalizer tool was built for this study with the *Arc Model Builder* in *ArcGIS*, version 10 (ESRI, Redlands, CA, USA) that allows the user to input any desired slope in the regression equation and generalize features for the entire data set within minutes. This is a useful tool for testing the effects of different regression lines on the final hydrographic product. The majority of the tool's operations consists of enrichments to the raw data – a common practice in generalization.

The tool operates on individual flowlines by common identifier (COMID). It initially selects only streams and artificial paths from the flowline layer, ignoring features such as canals/ditches, pipelines, and other connectors. After selection of the flowlines and artificial paths, the tool adds a HUC8 field (populated by parsing the first

Figure 6. Manually flow-generalized NHDPlus, approximately 1:10M scale.

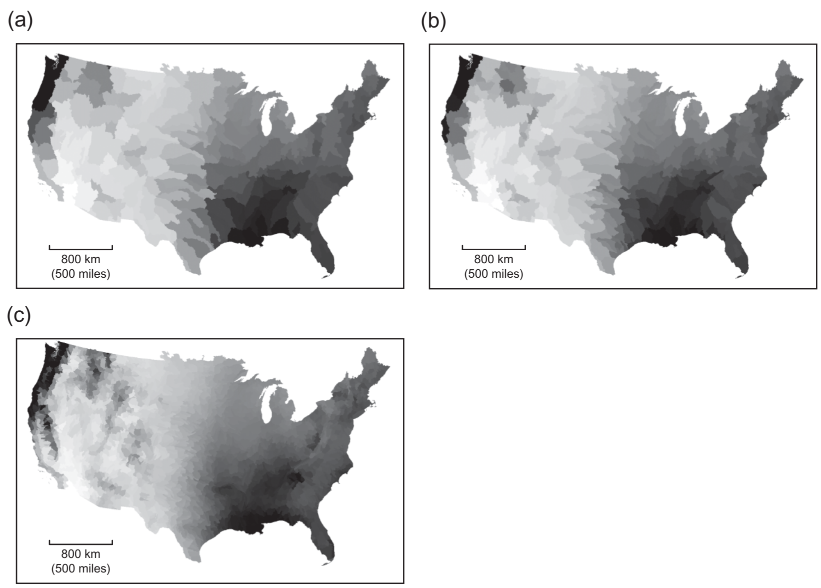

Figure 7. PRISM mean annual precipitation, per (a) HUC4, (b) HUC6, and (c) HUC8.

eight digits from the reachcode) to the flowline table. The tool then joins PRISM mean annual precipitation per HUC8 via the HUC8 field. A third field, the UROM mean annual flow, is joined via the COMIDs. A threshold field is then added to the flowline table and the pruning threshold is calculated by the regression equation. The slope of the regression line is entered by the cartographer. The tool enriches each unique COMID in the flowline table with four new fields – HUC8, mean annual precipitation per HUC 8, mean annual flow, and the pruning threshold – then selects individual features by COMID where the flow is greater than the threshold and outputs the results to a separate geodatabase feature class.

We ran the tool with five different regression lines, with slopes of 0.00001, 0.0001, 0.001, 0.01, and 0.1. We then performed a quality assessment of the results by

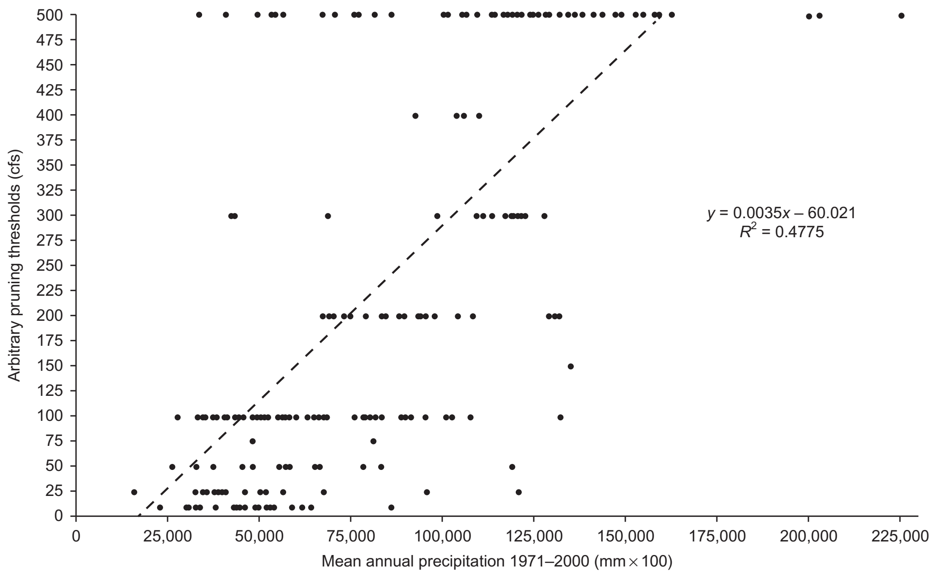

Figure 8. Arbitrary pruning threshold per mean annual precipitation.

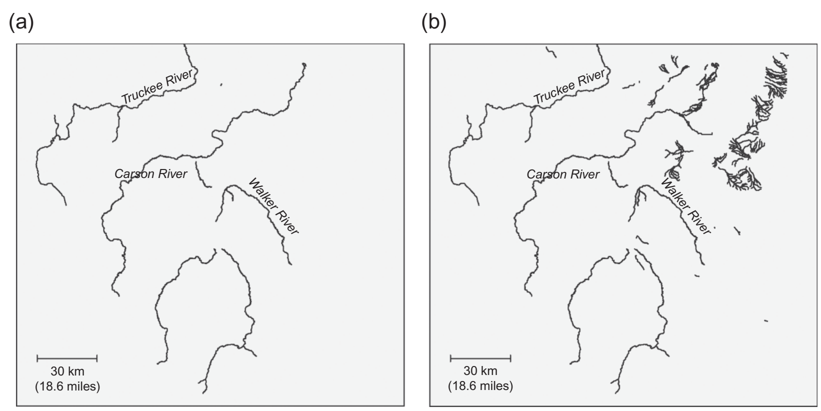

Figure 9. (a) Areas of region 16 showing full flowline density, with a pruning threshold (cfs) = 0.0035 × precipitation (mm × 100) − 60.021 (cfs) and (b) Areas of region 16, *y*-intercept = 0. Pruning threshold (cfs) = 0.0035 × precipitation (mm × 100).

comparing the results to the National Atlas 1:10,000,000-scale hydrography, with the goal of matching its general line density and overall aesthetic quality. We also judged our results by the degree to which the flowlines correspond to the macro-scale precipitation trends, as represented by the PRISM data.

Results

The results approximate 1:10,000,000-scale North American Atlas hydrography when the slope of the regression line is 0.001. However, there is also an improved balance of cartographic and hydrologic representations. Arid areas of the conterminous United States appear with more negative space on the map than high precipitation areas. In most cases, streams do not extend to their headwaters.

Results are presented for the Pacific Northwest with regression lines of slopes 0.0001 and 0.001, with a *y*-intercept of zero (Figure 10). The slope of the regression equation essentially serves as a scaling factor. Steeper slopes produce maps with greater density of flowlines; shallower slopes produce maps with lesser density of flowlines. Output based on variable flow thresholds tends to minimize or smooth out variations in the density of the original data. Where the original flowline data are very dense, the flow volumes of many flowlines may be similar. A pruning threshold based on flow indiscriminately culls

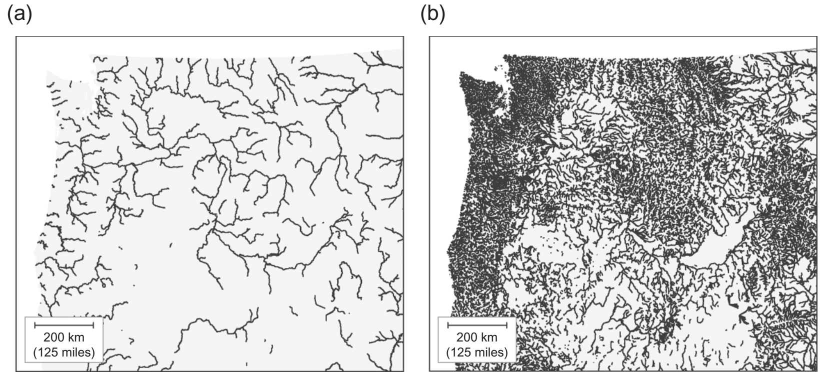

Figure 10. (a) Pacific Northwest. Pruning threshold (cfs) = 0.0001 × precipitation (mm × 100) and (b) Pacific Northwest. Pruning threshold (cfs) = 0.001 × precipitation (mm × 100).

the many flowlines with similar flow, minimizing density disparities of the original data.

Discussion

An excerpt from this work, taken from the Platte River basin, illustrates a shortcoming of the generalizer tool (Figure 11). The network has been disrupted along the South Platte River, due to irrigation canals that created zero flow gaps in the NHDPlus. Braided streams, which commonly introduce complications to generalization algorithms (Stanislawski et al. 2006; Thomson and Brooks 2007; Savino et al. 2011; Touya 2007), similarly cause flow rate disruptions, though in NHPlus, flow is assigned to the main channel. Disruptions in the flowline network occur in all hydrologic regions, though most commonly in shorter sections. These disruptions are not usually visible at small scales.

Figure 12 shows results for the entire conterminous United States with a regression line of slope of 0.001 and a *y*-intercept of zero. We believe that this represents an optimum representation of the river network of the United States. The selection and density of rivers communicate the flow of water through rivers and give an indication of the climate of the nation. In this sense, the map satisfies a hydrologic portrayal of surface water; yet, it also portrays a distribution of features necessary for mapping the landscape.

The EPA and USGS made version 2 of NHDPlus (NHDPlusV2) available to the public in 2012. The key improvements over the NHDPlus are higher quality data, addition of attributes, inclusion of Canadian and Mexican data, and improvements to a variety of the hydrological data models, such as UROM flow rates and network integrity (EPA, USGS, and Horizon Systems Corporations 2012). Application of this generalization method to NHDPlusV2 and other data is the next logical step for future research on automated stream pruning. Work on the NHDPlusV2 may yield improved balance between cartographic and hydrologic representations. For example, the pruning threshold could be correlated to a raster data set that accounts for multiple environmental factors, such as McCabe and Wolock's (2011) runoff estimates, which include both precipitation and

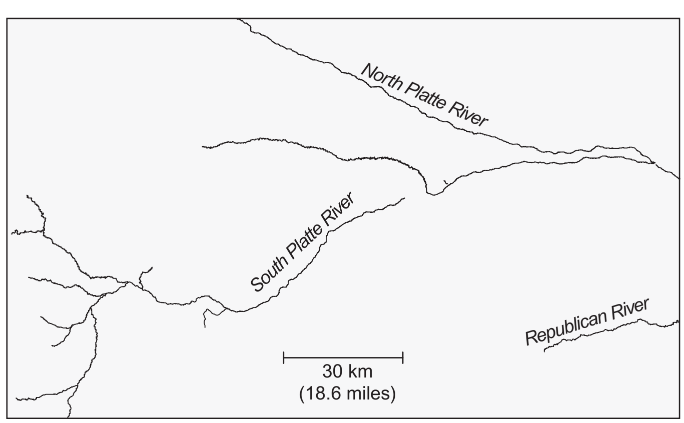

Figure 11. Platte River basin, showing disrupted stream network. Pruning threshold (cfs) = 0.001 × precipitation (mm × 100).

Figure 12. Conterminous United States, approximately 1:10M scale. Pruning threshold (cfs) = 0.001 × precipitation (mm × 100).

temperature, or by multiple physiographic regimes (Brewer, Buttenfield, and Usery 2009; Buttenfield, Stanislawski, and Brewer 2010, 2011; Stanislawski and Buttenfield 2011; Stanislawski, Finn, and Buttenfield 2010). Touya (2008) has also shown that differences in local physiography or climate require different generalization processes.

Additional work is needed to explore methods that avoid disruptions in the flowline network. The current version of our tool prunes individual COMIDs, but it may be preferable to prune by reachcode and to implement an additional set of rules during the pruning process to avoid fragmenting the flowline network. The pruning process should target headwater reaches over reaches in the middle of the connected network.

Traditional hydrography portrays rivers as the headwaters. Future research may determine whether larger rivers should be portrayed all the way to their source or be culled back at the point at which the flow becomes insignificant. This could be resolved by allowing for the adjustment of the length of headwaters to be pruned.

Furthermore, statistical assessment of the flow-generalized results is needed. It would be possible, for example, to relate the line density of the results with the environmental data or to calculate a coefficient of line correspondence (CLC) with a benchmark data set, as demonstrated by Stanislawski (2009). Work is also needed to determine how the slope of the regression line relates to the output scale in order to generate guiding principles for this method. The slope of the regression line may be correlated with the output scale, similar to the relation articulated by the radical law. This project uses a regression line with arbitrary slope to demonstrate the concept because the slope of a line is a simple parameter to adjust. However, other regression equations, such as a power regression or exponential regression, may yield better cartographic results. Preliminary investigations show positive results with a power regression. Finally, this method should be tested on high-resolution hydrographic data, such as the 1:24,000-scale NHD, as it becomes available. This should be possible, as long as the high-resolution NHD data are enriched with stream flow rates.

Conclusion

This study explores a method to generalize the 1:100,000 scale NHDPlus to smaller scales based on estimations of natural stream flow. Streams are pruned based on a variable flow threshold, derived from the local mean annual precipitation, by a regression equation with a user-specified slope. The project aims to improve upon traditional hydrographic technique by adding a hydrologic component – generalizing by variable flow thresholds. The intent is to produce a river map that does not under-represent rivers in high precipitation areas or over-represent rivers in arid areas.

The results show that there are advantages to generalizing with variable flow thresholds: (1) the method allows for fine gradations in output scale; (2) the output maps tend to minimize density variations in the raw data; (3) the method is no less arbitrary than other methods surveyed, but the subjective criteria are more easily derived. For instance, it is less challenging to arbitrarily choose the slope of a regression line than to deliberate over the identification of the main stem of a river basin; and finally, (4) the method can be performed rapidly on large data sets, as long as the stream data have been enriched with reliable flow rates. The cartographer can choose the slope of the regression line and then immediately inspect the output.

References

Ai, T., Y. Liu, and J. Chen. 2006. "The Hierarchical Watershed Partitioning and Data Simplification of River Network." *Progress in Spatial Data Handling* 11: 617–632. doi:10.1007/3-540-35589-8_39.

Berelson, W. L., P. A. Caffrey, and J. D. Hamerlinck. 2004. "Mapping Hydrologic Units for the National Watershed Boundary Dataset." *Journal of the American Water Resources Association* 40 (5): 1231–1246. doi:10.1111/j.1752-1688.2004.tb01582.x.

Brewer, C. A., B. P. Buttenfield, and E. L. Usery. 2009. "Evaluating Generalizations of Hydrography in Differing Terrains for the National Map of the United States." In *24th International Cartographic Conference*. Santiago: International Cartographic Association.

Buttenfield, B. P., L. V. Stanislawski, and C. A. Brewer. 2010. "Multiscale Representations of Water: Tailoring Generalization Sequences to Specific Physiographic Regimes." In *Sixth International Conference on Geographic Information Science*. Zurich: International Cartographic Association.

Buttenfield, B. P., L. V. Stanislawski, and C. A. Brewer. 2011. "Adapting Generalization Tools to Physiographic Diversity for the United States National Hydrographic Dataset." *Cartography and Geographic Information Science* 38 (3): 289–301. doi:10.1559/15230406382289.

Catlow, D. R., and D. Du. 1984. "The Structure and Cartographic Generalization of Digital River Data." In *Proceedings of the 44th Annual ASP-ASCM Convention*, 511–520. Washington DC: American Congress on Surveying and Mapping.

EPA (Environmental Protection Agency), USGS (United States Geological Survey), Horizon Systems Corporations. 2010a. *NHDPlus User Guide*. Accessed April 27, 2012. http://www.horizon-systems.com/nhdplus/.

EPA (Environmental Protection Agency), USGS (United States Geological Survey), Horizon Systems Corporations. 2010b. *NHDPlusV1 Data*. Accessed April 27, 2012. http://www.horizon-systems.com/nhdplus/.

EPA (Environmental Protection Agency), USGS (United States Geological Survey), Horizon Systems Corporations. 2012. *NHDPlus Version 2: User Guide (Data Model 2.1)*. Accessed April 27, 2012. ftp://ftp.horizonsystems.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.

ESRI (Economic and Social Research Institute). 2012. *World Hydro Reference Overlay*. Accessed April 4, 2013. http://www.arcgis.com/home/item.html?id=f7c73101a09c44058f8f029eefd37bd6.

Gardiner, V. 1982. "Stream Networks and Digital Cartography ". *Cartographica* 19 (2): 38–44.

Gesch, D., G. Evans, J. Mauck, J. Hutchinson, W. J. Carswell Jr. 2009. *The National Map – Elevation: U.S. Geological Survey Fact Sheet 2009*. Accessed July 16, 2012. http://pubs.usgs.gov/fs/2009/3053/.

Global Runoff Data Centre. 2013. *Welcome to the Global Runoff Data Centre*. Global Runoff Data Centre. Accessed April 4, 2013. http://www.bafg.de/GRDC/EN/Home/homepage_node.html.

Horton, R. 1945. "Erosional Development of Streams and their Drainage Basins: Hydrophysical Approach to Quantitative Morphology." *The Geological Society of America Bulletin* 56: 275–370.

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

Mazur, E. R., and H. W. Castner. 1990. "Horton's Ordering Scheme and the Generalisation of River Networks." *The Cartographic Journal* 27 (2): 104–112.

McCabe, G. J., and D. M. Wolock. 2011. "Independent Effects of Temperature and Precipitation on Modeled Runoff in the Conterminous United States." *Water Resources Research* 47. doi:10.1029/2011WR010630.

McMaster, R. B., and K. S. Shea. 1992. *Generalization in Digital Cartography*. Washington, DC: Association of American Geographers.

Oregon State University PRISM Group. 2001. *Precipitation: Annual Climatology (1971–2000)*. Corvallis, OR: PRISM Climate Group, Oregon State University.

Pierson, S. M., B. J. Rosenbaum, L. D. McKay, and T. G. Dewald. 2008. *Strahler Stream Order and Strahler Calculator Values in NHDPlus*. Accessed April 27, 2012. ftp://ftp.horizonsystems.com/nhdplus/NHDPlusV1/NHDPlusExtensions/SOSC/SOSC_technical_paper.pdf.

Regnauld, N., and R. B. McMaster. 2007. "A Synoptic View of Generalisation Operators." In *Generalization of Geographic Information: Cartographic Modeling and Applications*, edited by W. A. Mackaness, A. Ruas, and L. T. Sarjakoski, 37–66. Oxford: Elsevier.

Richardson, D. E. 1994. "Generalization of Spatial and Thematic Data Using Inheritance and Classification and Aggregation Hierarchies." In *Sixth International Symposium on Spatial Data Handling*, edited by T. C. Waugh and R. G. Healey, 901–920. London: Taylor and Francis.

Savino, S., M. Rumor, F. Canton, G. Langiu, and M. Reineri. 2011. "Model Generalisation of the Hydrography Network in the CARGEN Project." In *25th International Cartographic Conference*. Paris: International Cartographic Conferences.

Shreve, R. L. 1966. "Statistical Law of Stream Numbers." *The Journal of Geology* 74: 178–137.

Simley, J. D., and W. J. Carswell Jr. 2009. *The National Map – Hydrography: U.S. Geological Survey Fact Sheet* 2009–*3054*, 4. Accessed July 16, 2012. http://pubs.usgs.gov/fs/2009/3054/.

Slack, J. R., A. Lumb, and J. M. Landwehr. 2006. *Hydro-Climatic Data Network: A U.S. Geological Survey Streamflow Data Set for the United States for the Study of Climate Variations*, 1874–1988. USGS Open-File Report 92–129. Accessed March 15, 2012. http://water.usgs.gov/GIS/metadata/usgswrd/XML/hcdn.xml.

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. doi:10.1016/j.compenvurbsys.2009.07.004.

Stanislawski, L. V., and B. P. Buttenfield. 2011. "Adapting Generalization Tools to Physiographic Diversity for the United States National Hydrography Dataset." *Cartography and Geographic Information Science* 38 (3): 117–125. doi:10.1559/15230406382289.

Stanislawski, L. V., B. P. Buttenfield, M. P. Finn, and K. Roth. 2009. "Stratified Database Pruning to Support Local Density Variations in Automated Generalization of the United States National Hydrography Dataset." In *Proceedings of 24th International Cartographic Conference*. Santiago: Center of Excellence for Geospatial Information Science (CEGIS), U. S. Geological Survey (USGS).

Stanislawski, L. V., M. P. Finn, M. Barnes, and E. L. Usery. 2007. "Assessment of a Rapid Approach for Estimating Catchment Areas for Surface Drainage Lines." In *American Congress on Surveying and Mapping Annual Conference*. St. Louis, MO: Center of Excellence for Geospatial Information Science (CEGIS), United State Geological Survey (USGS).

Stanislawski, L. V., M. P. Finn, and B. P. Buttenfield. 2010. "Integrating Hydrographic Generalization over Multiple Physiographic Regimes." In USGS Symposium on Generalization and Data Integration, Boulder, CO, June 20–22, 2010.

Stanislawski, L. V., M. Finn, M. Starbuck, E. L. Usery, and P. Turley. 2006. "Estimation of Accumulated Upstream Drainage Values in Braided Streams Using Augmented Directed Graphs." In *AutoCarto 2006*. Vancouver, WA: Center of Excellence for Geospatial Information Science (CEGIS), United State Geological Survey (USGS).

Strahler, A. 1957. "Quantitative Analysis of Watershed Geomorphology." *Transactions of the American Geophysical Union* 8 (6): 913–920.

Thomson, R. C., and R. Brooks. 2000. "Efficient Generalisation and Abstraction of Network Data Using Perceptual Grouping." In *5th International Conference on GeoComputation*. Greenwich: University of Greenwich. Accessed January 13, 2012. http://www.geocomputation.org/2000/GC029/Gc029.htm.

Thomson, R. C., and R. Brooks. 2007. "Generalization of Geographical Networks." In *Generalization of Geographic Information: Cartographic Modeling and Application*, edited by W. A. Mackaness, A. Ruas, and L. T. Sarjakoski, 255–267. Oxford: Elsevier.

Töpfer, F. P., and W. Pillewizer. 1966. "The Principles of Selection." *The Cartographic Journal* 3: 7.

Touya, G. 2007. "River Network Selection Based on Structure and Pattern Recognition." In *23rd International Cartographic Conference*. Moscow, August 4–10.

Touya, G. 2008. "First Thoughts for the Orchestration of Generalisation Methods on Heterogeneous Landscapes." In *ICA Workshop on Generalizations*. Montpellier, June 20–21.

USDOI (United States Department of Interior). 2006. "North American Atlas." *Hydrography*. Accessed November 11, 2011. http://www.nationalatlas.gov/atlasftp-na.html.

USGS (United States Geological Survey). 2000. "The National Hydrography Dataset Concepts and Contents." USGS. Accessed November 11, 2011. http://nhd.usgs.gov/.

USGS (United States Geological Survey). 2012. *National Water Information System*. Accessed March 15, 2012. http://waterdata.usgs.gov/nwis.

Vogel, R. M., I. W. Wilson, and C. Daly. 1999. "Regional Regression Models of Annual Streamflow for the United States." *Journal of Irrigation and Drainage Engineering* 125 (3): 148–157.

WBD (Watershed Boundary Dataset). 2012. "Watershed Boundary Data Set." United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), United States Geological Survey (USGS). Accessed April 27, 2012. ftp://ftp.ftw.nrcs.usda.gov/wbd.

Wilmer, J. M., and C. A. Brewer. 2010. "Application of the Radical Law in Generalization of the National Hydrography Data for Multiscale Mapping." In ASPRS/CaGIS Fall Specialty Conference, Orlando, FL, November 15–19, 2010.

Zhang, Q. 2007. "Drainage Typification Based on Dendritic Decomposition." *The Cartographic Journal* 44: 321–328.