Proceedings 16th ICA Generalisation Workshop, Dresden Germany 23-24 August 2013
Topics of Interest: Characterising and Reasoning About Geographic Phenomena

Alternate Methods for Automatic Selection of Primary Paths Through Braided Hydrographic Networks

Barbara P. Buttenfield *, Lawrence V. Stanislawski **, Christopher Anderson-Tarver *, and Michael J. Gleason *

* Department of Geography, University of Colorado, Boulder Colorado USA
** Center of Excellence for Geospatial Information Science, United States Geological Survey, Rolla Missouri USA

Abstract
Hydrography provides an important and commonly included data layer for all scales of topographic mapping. This layer poses special challenges in generalization, including but not limited to preserving topologic connectivity, maximizing channel length at reduced scales, and preserving a connected network that is visually logical and hydrologically valid. Primary paths form an important component for mapping hydrographic networks. At larger scales, primary paths establish continuity among flowlines and water polygons. At smaller scales, a generalized primary path may substitute for the entire network. Manual delineation of primary paths can be labor-intensive and prone to errors, which may cause gaps in the network. This paper presents alternative strategies for automatic delineation of primary paths through braided hydrographic networks. This draft examines three strategies, which delineate a single primary path through the braid, or identify the outermost channels which bound the braid and connect to inflow and outflows, or use a weighted criteria model to prioritize and eliminate channels incrementally.

Problem Context
Hydrography provides an important and commonly included data layer for all scales of topographic mapping. Due to scale sensitivity, hydrographic data requires frequent generalization across mapping scales for proper integration with terrain and other data layers. Vector hydrography poses special challenges for generalization, including but not limited to preserving topologic connectivity among linear and polygonal features, maximizing channel lengths at reduced scales, protecting precise coordinate positions for specific types of hydrographic features (e.g., dams and bridges, stream gauges, and tributary confluences), and accurately reflecting channel hierarchy at every scale. In particular, cartographers want to highlight a primary path (also called a centerline, main stem, or thalweg, depending on scientific discipline) which demarcates the main channel flowing through a network. Primary paths anchor a stream network, lending visual coherence at large mapping scales. At smaller scales, a simplified primary path can act as surrogate for the entire network. Customarily, primary paths are derived from stream order (Horton 1945; Strahler 1957), channel depth or geometry (Benke 2005), or flow rates (Crowder and Diplas, 2000; Merwade et al. 2005). When these attributes are explicitly attributed in a data base, automatic delineation of primary paths is straightforward. When these attributes are not available, primary paths must be inferred by characterization and reasoning about the phenomenological nature of the stream network.

Previous work (Anderson-Tarver et al. 2011; 2012) demonstrates an algorithm that permits progressively inclusive primary path delineations which may be targeted to specific map scales. The algorithm is supported by database enrichment of catchment areas and upstream drainage estimates for each channel (Stanislawski et al. 2006), and delineates a primary path by means of a shared node trace from pour point to headwaters. The authors acknowledge an important limitation of the basic algorithm, whose delineation includes all channels through braided regions of the stream network. The all-inclusive solution occurs because braided channels often carry practically equivalent upstream drainage values, or lack unique names. For topographic mapping, automatic delineation should offer alternatives, either to demarcate a single primary channel through the braid, or alternatively to demarcate two channels which bound the spatial extent of the braid. A third option uses a weighted criteria model (Eastman et al. 1995) to prioritize every channel within a braid, based upon a set of geographic and geometric factors. This paper explores each alternative, demonstrating results for automatic delineation of primary paths through braided hydrographic networks encountered in dry landscapes.

Test Data Set

In the United States, the National Hydrography Dataset (NHD) is compiled by the United States Geological Survey (USGS) from either the high-resolution or 1:100,000-scale (100K) layer. The high-resolution (HR) NHD layer is mostly compiled from 24K source data within the 48 coterminous States, and is being densified in specific areas to 1:12,000 or larger scales, depending on the needs of local or state agencies. The NHD represents surface water including natural and human-made hydrographic features. Enrichment of the HR NHD for generalization includes an estimate of upstream drainage area (UDA) for each flowline feature, which gives a relative prominence estimate for each feature. In contrast, Ai et al. (2006) simplify a river network using watershed areas estimated through Delaunay triangulation. Others have used stream order or total upstream channel length (Thompson and Brooks 2000, Savino et al. 2011) to generalize stream channels, but these values are sensitive to inconsistent channel compilation, which is encountered within the HR NHD data. UDA prominence estimates are normalized by area and are better suited for flow networks with compilation inconsistencies. UDA estimates for the HR NHD are derived from Thiessen partitioning of catchments for each flowline feature (Stanislawski 2006). The subbasin processed for this paper is the Lower Prairie Dog Town Fork of the Red River, Texas (HUC 11120105).

Basic Primary Path Delineation Algorithm

The basic algorithm follows a three-stage process, building upon earlier work (Anderson-Tarver et al 2012). First, primary path 'stems' are established from outflow(s) or pour point(s) (i.e., locations of the farthest downstream channel in the subbasin) by selecting on the UDA attribute. To maintain comparability when processing multiple subbasins, the UDA threshold is stipulated as a percentage of the subbasin area drained by a given stream channel. In Figure 1, the reader will see that the choice of UDA threshold affects how many channels are selected as stems. A UDA value of 20 percent is used in this paper: all channels that drain 20 percent or more of the subbasin area are selected as primary stem paths.

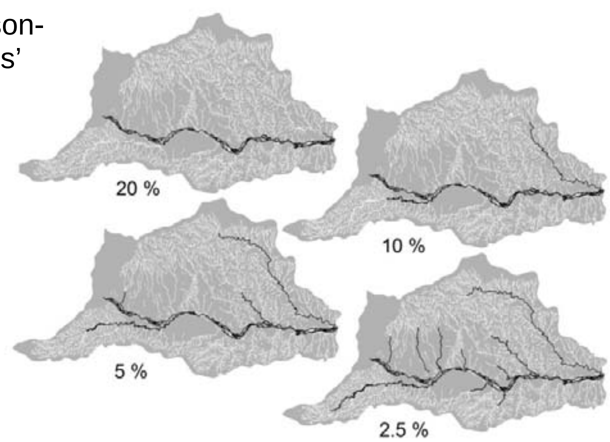


Figure 1. Impact of UDA value on selection of primary path stems. Percentages indicate how much of the subbasin is drained by the selected channels. Lower Prairie Dog Town Fork Red River, Texas (HUC 11120105).

The second step uses a shared node trace moving upstream from the top of each stem to the headwaters, making a decision at each tributary confluence and at this stage the processing expands upon the algorithm reported in Tarver et al 2012. If one tributary shares the Reachcode (a channel identifier) of the downstream flowline, it is selected (Figure 2a). Otherwise, if the name of the downstream channel is shared, it is selected (Figure 2b). If neither or both tributaries share the name of the downstream channel, the tributary with the higher UDA value is selected (Figure 2c). The basic delineation algorithm does not add or delete features, but rather enriches a new attribute characterizing particular channels as primary paths.

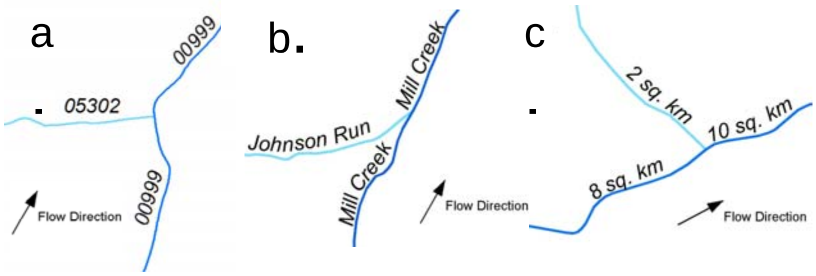


Figure 2. Sequentially tested criteria determine the path of the node trace: reachcode, channel name and UDA value. (a) The reachcode causes the southern tributary to be selected. (b) The channel name causes selection of the Mill Creek tributary. (c) The southern upstream channel carries the higher UDA value and is therefore selected. In all panels, the darker blue line shows the resulting primary path.

Delineation of Primary Paths through Braided Regions

The algorithm uses the three criteria (reachcode, name and UDA value) to delineate a primary path as the set of flowlines that run continuously from headwaters to pour point in the subbasin and that drain a substantial percentage of the subbasin area. A limitation of basic demarcation occurs in braided regions, where the three criteria are insufficient in some cases to delineate a single primary path. Figure 3 illustrates the problem for the Texas subbasin, showing that the delineated channels coalesce and, at smaller scales, will overlap even for very thin line weights.

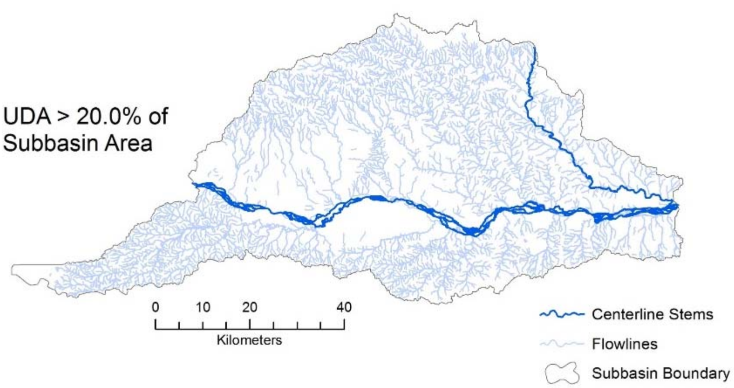


Figure 3. Solution for basic primary path algorithm for the Texas subbasin. Selection of channels within a heavily braided region can produce an overly complex primary path. Dashed box shows the extent of Figure 4.

The solution partitions the braid by establishing polygons in between every channel within the braid, and dissolving the polygons to isolate braided sub-regions (Figure 4). In between the braid polygons, the basic algorithm will operate effectively. Within the braided polygons, three approaches may be applied.

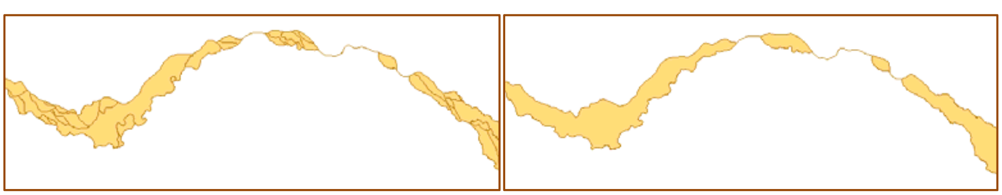


Figure 4 Generating braid polygons by dissolving areal regions in between each braid channel.

Approach #1: Inner Channel Primary Path. After isolating braid polygons, the same basic algorithm can be applied, limiting its operation only to those channels inside a single braid polygon (Figure 5a). Flowlines are already enriched and thus all attributes are available to the subset of channels; UDA values do not need to be recomputed. The Inner Channel solution differs from the basic algorithm in that it traces only from the outflow of each braid polygon in turn, thus reducing the number of channels to be checked. The basic algorithm traces upstream from every primary path "stem"; and the Texas subbasin is unusual in that nearly all of the braid channels have high enough UDA values to be considered as "stems". Use of the Inner Channel algorithm thus avoids delineating every channel within braid polygons. One can also check for braid polygons with multiple contributing inflows (Figure 5b) by testing the edges of the braid polygon for inflows not included in a primary path. Each isolated primary inflow to a braid must be processed by building another braid polygon surrounding associated channels, using the same method as described above, and delineating a primary path through those isolated channels (Figure 5c).

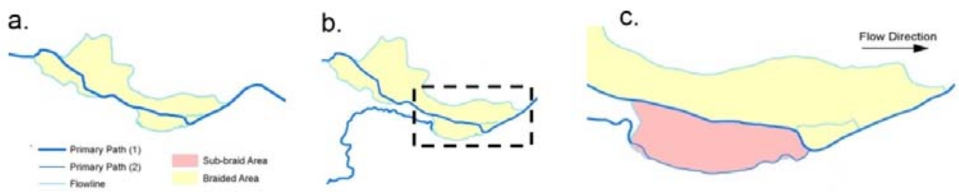


Figure 5. Delineating an Inner Channel Primary Path. (a) Within each braid polygon, the basic algorithm is applied, thus reducing the number of channels which must be traversed and tested. (b) Edges of each braid polygon are tested to detect multiple inflows; where they exist, (c) smaller braid polygons isolate flowlines which can be selected to connect that inflow to the primary path.

Approach #2: Outer Channel Primary Path. For mapping at very small scales (e.g., 1:1 million or smaller), knowledge of braid extent may be as important as identifying the precise primary path. A cartographic convention applied to generalize complex railway sidings is to retain the outermost tracks in the siding, eliminating inner tracks, to preserve overall shape and spatial extent (McMaster and Shea, 1992: 60). A similar principle is applied in the second approach, selecting the outermost channels in each braid polygon, with an added size constraint on the area of each braid polygon. Minimum threshold ground size is established by USGS standards for retention of polygons at scale. For braid polygons larger than the threshold, outer (bounding) channels are incorporated into the primary path. Polygons whose area falls below the threshold are processed with the inner channel approach (Figure 6). In some cases, the braid polygon becomes so narrow that an outer channel is not appropriate. The research team is currently exploring methods to detect narrow "pinchpoints" automatically, and at these locations, the narrow portions of braid polygons should be pruned and the inner channel solution should be applied. Automatic detection based on polygon width forms an area of ongoing research.

Figure 6. Outer chanel delineation for braid polygons above the size threshold.

Processing Notes and Preliminary Results

The primary path algorithm is implemented in Python 2.7 using the Arcpy module in ArcGIS 10.1. Because the primary path algorithm must visit each channel upstream of each UDA selection, it can run slowly if using the Python cursor object. Performance is improved substantially by using native Arcpy and Python objects. Each channel is extracted to a tuple storing a geometry object and associated attributes, including reachcode, UDA value and GNIS name. (The Geographic Names Information System or GNIS is a placename gazetteer containing accepted names and point locations for natural and cultural landmarks in the United States.) A list of tuples comprising the primary path is generated during the UDA selection and subsequent shared node trace.

Results of the first two approaches for delineating a primary path through a braid are compared to the outcome using the basic algorithm, for the Texas subbasin (Figure 7). A larger scale inset with full set of flowlines is shown in Figure 8. At present, these approaches have been implemented and tested on several subbasins across the United States, in humid and dry landscapes, hilly and flat terrain.

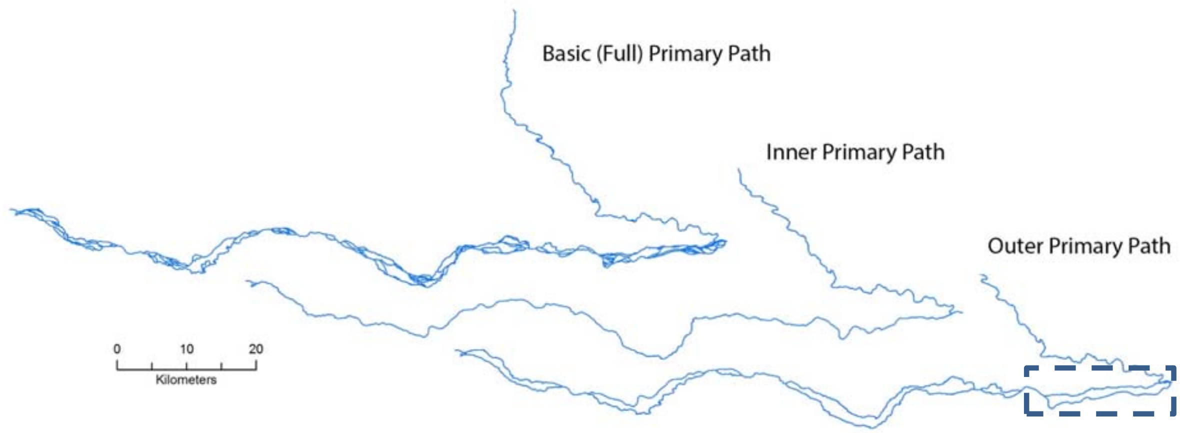
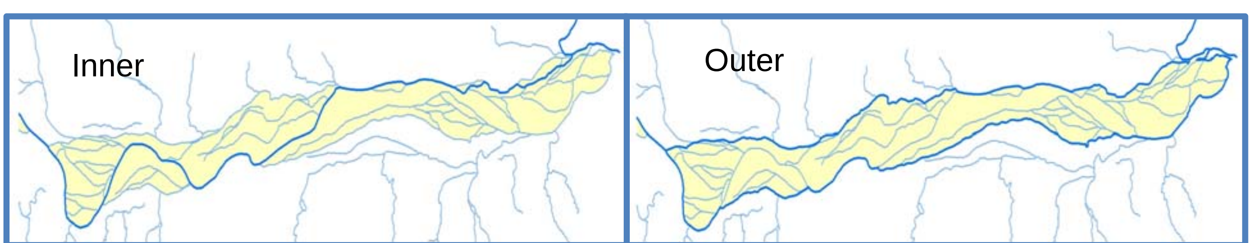


Figure 7. Three variants on primary path delineation for the Texas subbasin; for clarity, only the primary path through the braided region is shown. (Top) Basic primary path delineation including all channels within braided regions. (Middle) Selection of a single ("inner") channel through braid polygons. (Bottom) Selection of ("outer") channels bounding the braid polygons associated with 100K scale. Dashed box shows location of inset in Figure 8.



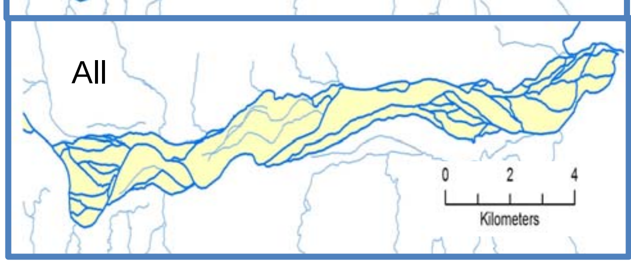


Figure 8. A comparison of the basic algorithm with Inner and Outer Channel solutions for the braid polygon highlighted in Figure 7, illustrating three alternative primary paths against the full set of flowlines.

One might surmise that the level of detail in three solutions would lend themselves to large, medium and small scale mapping situations. The question now arises: "is there an automated method to generate intermediate scale displays containing progressively less content?", essentially generating a set of versions of the braid which illustrate incremental transitions from the All Channel (basic) solution to the Inner Channel solution. This is the purpose of the final alternative, a heuristically weighted braid delineation, as described below.

Approach #3: Weighted Channel Solution
The third approach prioritizes every channel within each braid polygon, enabling progressive elimination of channels as mapping scale decreases, effectively providing a continuous and incremental primary path delineation. A widely accepted methodology, called a weighted criteria model (also called a multi-criteria model) is a reasoning tool and decision-support method which evaluates alternatives relative to individual criteria which have been assigned a specific importance ranking. Criteria can be weighted independently to avoid bias or over-dependence, and to test relative impacts and sensitivity. These tests can lend insight to better understand and evaluate the outcome. The method is often applied to decide among alternative strategies in planning, forecasting, and operations research (Taylor 2008). Weighted criteria modeling is useful in environmental modeling and site suitability analysis (Eastman et al. 1995), or any situation in which a set of optional strategies appear to be equally feasible. So long as the ranking of each criterion is accomplished on a unified scale, ranked criteria may be combined as a weighted sum or a weighted product (Triantaphyllou 2002).

Weights are established on the basis of existing or enriched attributes, specifically on GNIS name, feature type (e.g., perennial or intermittent stream), presence of an underlying water polygon, UDA value, and whether the channel is already a part of the basic, inner or outer channel primary path. The goal is to progressively eliminate channels until all that is left is the Inner Channel primary path. Weights for this example are arbitrary, prioritizing GNIS name and membership in the primary path (Figure 9).

flow_weights_braid_poly2

GNIS_Name	ReachCode	WBAreaC	FType	TOT_CM_SQ_KM	inr_pa	all_pat	outr_	Braid_wt	GNIS_wt	Fcode_wt	inFcode_wt	IP_wt	OP_wt	JDA_wt	ALLP_wt
<Null>	11120105001474	91860479	558	3002273.48103	FLW	CTR1	CTR1	17	0	3	0	0	10	4	0
<Null>	11120105001475	91860479	558	3002299.484245	FLW	CTR1	CTR1	17	0	3	0	0	10	4	0
Prairie Dog Town For	11120105001533	<Null>	558	3002227.905993	CTR1	CTR1	CTR1	52	20	3	0	15	10	4	0
<Null>	11120105001534	<Null>	460	3002243.379563	FLW	CTR1	FLW	10	0	0	1	0	0	4	5
<Null>	11120105001534	<Null>	558	3002243.365024	FLW	CTR1	FLW	10	0	0	1	0	0	4	5
Prairie Dog Town For	11120105001535	<Null>	558	3002252.921182	CTR1	CTR1	FLW	42	20	3	0	15	0	4	0
Prairie Dog Town For	11120105001536	<Null>	558	3002253.17834	CTR1	CTR1	FLW	42	20	3	0	15	0	4	0
Prairie Dog Town For	11120105001536	<Null>	558	3002252.956609	CTR1	CTR1	FLW	42	20	3	0	15	0	4	0
Prairie Dog Town For	11120105001537	<Null>	558	3002254.119431	CTR1	CTR1	FLW	42	20	3	0	15	0	4	0
<Null>	11120105001539	<Null>	460	0.376885	FLW	FLW	FLW	3	0	0	1	0	0	2	0
<Null>	11120105001539	<Null>	460	0.370465	FLW	FLW	FLW	3	0	0	1	0	0	2	0
<Null>	11120105001539	<Null>	460	0.405738	FLW	FLW	FLW	3	0	0	1	0	0	2	0
<Null>	11120105001540	<Null>	558	3002244.278496	FLW	CTR1	FLW	10	0	0	1	0	0	4	5
<Null>	11120105001540	<Null>	460	3002244.275736	FLW	CTR1	FLW	10	0	0	1	0	0	4	5
Prairie Dog Town For	11120105001542	<Null>	558	3002268.803883	CTR1	CTR1	FLW	42	20	3	0	15	0	4	0
<Null>	11120105001543	<Null>	558	1.120726	FLW	FLW	FLW	5	0	3	0	0	0	2	0

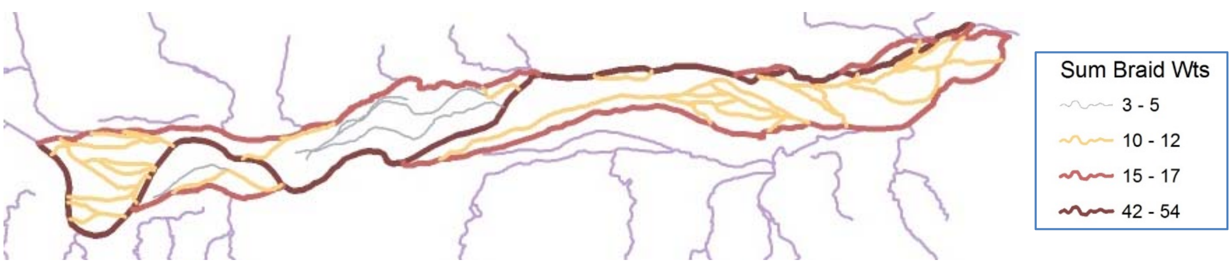


Figure 9. A section of the weighted table and the solution for braid polygon highlighted in Figure 8.

Weights can be adjusted heuristically, to comparatively assess the logic of progressive elimination of channels within a braid. The overarching constraint is that inflow and outflow continuity must be maintained at all scales, as well as cartographically important characteristics, such as feature type, GNIS name, and spatial relationship with polygonal water features. The algorithm traverses every channel within the braid, assessing each criterion and assigning a weight. An additive weights model is computed, and channels with higher scores are retained at smaller scales (Figure 10). Weights adjustment has not been performed as yet, which forms an obvious limitation of the proposed solution. Additionally, while there is an implied scale progression evident in Figure 10, no systematic exploration has been undertaken to understand how to select weights and how to associate specific ranges of mapping scales to a specific weight range. These aspects of braid delineation provide the current focus of research which continues to address the various aspects of automatic braid delineation.

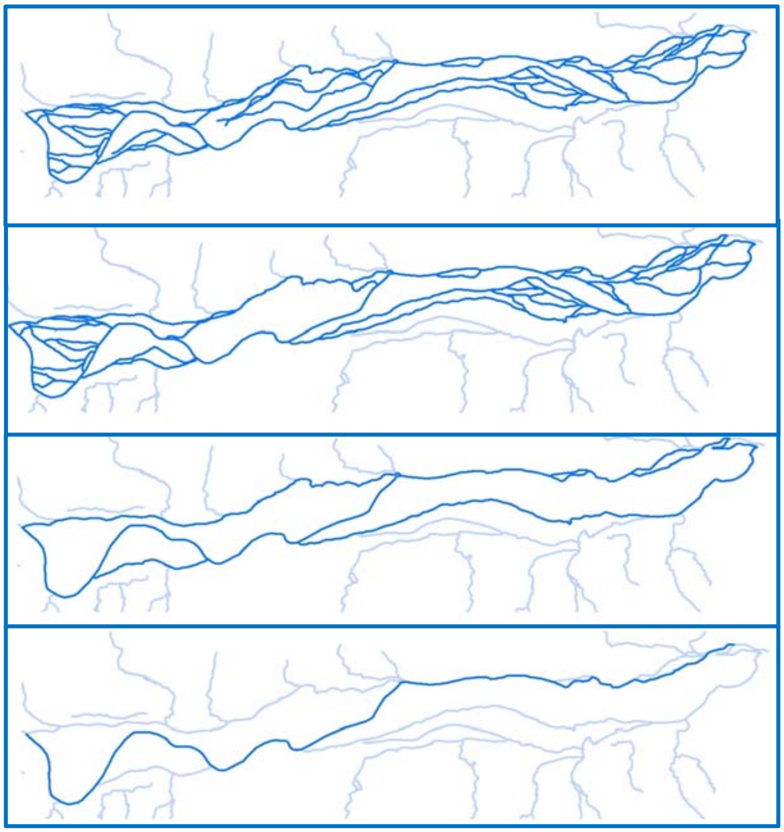


Figure 10. The weighted criteria model provides an automated mechanism to selectively eliminate channels within a braid polygon, transitioning from the Basic solution to the Inner Channel primary path.

Discussion

The topic of delineating hydrographic primary paths through braids has not received much attention in the literature. Traditional network thinning algorithms perform poorly on braided streams, yet braids require generalization because their spatial pattern is often too complex to be represented at smaller scales. This paper presents alternate strategies for delineating primary paths through braided regions of a stream network, demonstrating results for a basic delineation and three alternate strategies. Multi-scale primary path delineation can support mapping applications, for example to highlight display of cartographic centerlines, or to represent a simplified version at smaller mapping scales. Analytical applications for primary paths include demarcation of major flow routes through a hydrographic network.

For large volume databases, or data with an irregular update cycle, primary path delineation becomes particularly challenging for several reasons. First, important attributes may not be consistently available in a database. Even with complete attribution, multiple criteria may conflict. Criteria which are important at larger scales may become irrelevant at smaller scales. However, delineating a primary path by manual methods or by intuition can invoke inconsistent or erroneous results. Systematic reasoning and automated methods which evaluate criteria explicitly are especially useful when a data production task is complex, has many conflicting factors, or is ill-structured. All of these issues can arise with delineation of primary paths through braided regions, making this an excellent case study in which to explore the advantages and limitations of automated reasoning about geographic phenomena.

The alternate approaches to primary path delineation demonstrated in this paper provide several examples of automatic characterization about geographic phenomena. The outer channel strategy provides an interesting example of generalization by aggregation, wherein a set of features of lower dimensionality (lines) are substituted at smaller scales with a set of features (braid polygons) of higher dimensionality. A Hausdorff distance (Rote 1991) could be computed between proximal stream channels to reason automatically at what mapping scales the individual channels would be more appropriately mapped with a polygonal representation. In some cases, a hybrid solution integrating an outer channel with an inner channel could provide a spatial extent and a main stem, concurrently. Model generalization and LoD processing is required in anticipation of such cartographic flexibility.

Another example of automatic characterization relates to the weighted traversal strategy. Exhaustive prioritization of all channels within braid polygons may support automatic demarcation of areas of complex channels, deltas and non-hydrographic naturally occurring features such as mountain summits or canyons. In the United States at present, these feature types are recorded in the GNIS database as place names and point locations, but the spatial extent is neither compiled nor stored. Deltas, coastal hazards and areas of complex channels are not stored in every case within NHD, in part because of the problems of consistency and manual delineation mentioned above. Application of the primary path strategies could automate their incorporation into NHD, by identifying feature instances which have been detected at one scale but not others, or which have not been identified at all (as for example in areas experiencing dramatic storm events which modify coastal features and hydrography. In this context, the weighted primary path strategy forms the basis for automating multiscale semantic identification of exemplar feature patterns, and this forms an area for continued research.

Acknowledgements. The work of the Colorado authors is supported by USGS-CEGIS grant # 04121HS029, “Generalization and Data Modeling for New Generation Topographic Mapping”.

References

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

Anderson_Tarver, C. Gleason, M., Buttenfield, B.P., and Stanislawski, L.V. 2012 Automated Centerline Delineation to Enrich the National Hydrography Dataset. *Proceedings GIScience 2012* Columbus, Ohio: Springer LNCS 748: 15-28 Anderson-Tarver, C. and Buttenfield, B.P. Stanislawski, L.V. and Koontz, J. 2011 Automated Delineation of Stream Centerlines for the USGS National Hydrography Dataset. *Proceedings 25th International Cartographic Congress* Paris France, Vol 1: 409-423

Benke, A. C., Cushing, C. E. 2005 *Rivers of North America*. Elsevier Academic Press, Burlington

Crowder, D. W., Diplas, P.. 2000 Using Two-Dimensional Hydrodynamic Models at the Scales of Ecological Importance. *Journal of Hydrology* 230(3-4), 172-191

Eastman, J.R , Jin, W., Kyem, P.A.K., and Toledano, J. 1995 Raster Procedures for Multi-Criteria / Multi-Objective Decisions. *Photogrammetric Engineering and Remote Sensing* 61(5): 539-547

Horton, R. E. 1945 Erosional Development of Streams and Their Drainage Basins: Hydrophysical Approach to Quantitative Morphology. Geological Society of America Bulletin 56(3) 275–370

Merwade, V.M., Maidment, D.R., Hodges, B.R. 2005 Geospatial Representation of River Channels. *Journal of Hydrologic Engineering* 10(3): 243-251

Neun, M., Burghardt, D., Weibel, R. 2008 Web Service Approaches for Providing Enriched Data Structures to Generalisation Operators. *International Journal of Geographical Information Science* 22(2): 133-165

Rote, G. 1991 Computing the Minimum Hausdorff Distance Between Two Point Sets on a Line Under Translation. *Information Processing Letters* 38(3): 123-127

Savino S, Rumor M, Canton F, Langiu G, Reineri M (2011) Model Generalization of the Hydrography Network in the CARGEN Project, In Ruas A (ed), Advances in Cartography and GIScience (1):439-457, Lecture Notes in Geoinformation and Cartography, Springer-Verlag Berlin Heidelberg

Stanislawski, L.V., Finn, M., Starbuck, M., Usery, E.L., Turley, P. 2006 Estimation of Accumulated Upstream Drainage Values in Braided Streams Using Augmented Directed Graphs. *Proceedings AutoCarto 2006,* Vancouver, Washington

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

Taylor, J.W. 2008 Exponentially Weighted Information Criteria for Selecting Among Forecasting Models. *International Journal of Forecasting* 24: 513-524.

Thomson, R.C., Brooks, R. 2000 Efficient Generalization and Abstraction of Network Data Using Perceptual Grouping. In: Proceedings of the 5th International Conference on GeoComputation.

Triantaphyllou, E. 2002 *Multi-Criteria Decision Making Methods: A Comparative Study*. Kluwer Applied Optimization Series Vol. 44. (Kluwer is now Springer-Verlag).