An Open Source Solution to High Performance Processing for Extracting Surface Water Drainage Networks from Diverse Terrain Conditions

Lawrence V. Stanislawski, Kornelijus Survila, Jeffrey Wendel, Yan Liu, and Barbara P. Buttenfield

ABSTRACT: This paper describes a workflow for automating the extraction of elevation-derived channels using freely available open source tools with parallel computing support, and demonstrates the effectiveness of procedures in various terrain conditions within the conterminous United States. Drainage networks are extracted from the U. S. Geological Survey (USGS) 1/3 arc-second 3D Elevation Program (3DEP) elevation data having a nominal cell size of 10 meters (m). This research demonstrates the utility of open source tools with parallel computing support for filling depressions in 30 HUC8 subbasins distributed across humid, dry and transitional climate regions in terrain conditions exhibiting a range of differing slopes. Special attention is given to low slope terrain, where network connectivity is preserved by generating synthetic stream channels through lake and waterbody polygons. Conflation analysis compares the extracted streams with a 24K NHD flowline network, and shows that similarities are highest for 2nd and higher order tributaries.

KEYWORDS: elevation-derived streams, synthetic streams, 3DEP, high performance computing, National Hydrography Dataset

Introduction

Hydrographic data is an important component of topographic mapping that supports many applications including hydrologic modeling and forecasting, flood inundation mapping, landslide mapping, habitat mapping, river restoration, and geomorphologic change analysis. For more details, Poppenga et al. (2013) review several of these applications. Maidment (2015) describes a framework for the National Flood Interoperability Experiment that integrates topographic and hydrographic data to support hydrologic forecasting systems for the United States. Hydrographic and other data that are used in applications must be both accurate and current to obtain reliable results. Geoprocessing methods to validate and improve hydrographic data using current, high resolution elevation data, including light detection and ranging (lidar) point cloud data and derived products, are being developed and tested (Poppenga et al., 2013; Stanislawski et al., 2015a). These techniques require extensive data processing to apply over large areas. For example, data processing for the United States requires over 200 terabytes of storage space for lidar data alone. Although recent technology and cyberinfrastructure advances are furnishing the capacity to acquire, store and process large volumes of geospatial data, workflows that automate geoprocessing methods to improve hydrographic flow network data have not been fully realized.

Automated methods to extract surface water drainage networks from elevation data can assist initial capture and update of hydrographic data (O'Callaghan and Mark 1984; Jenson and Dominigue, 1988; Tarboton, Bras, and Rodriguez-Iturbe, 1991; Montgomery and Foufoula-Georgiou, 1993; Maidment, 2002; Anderson, 2012; Passalacqua, Belmont, and Foufoula-Georgiou, 2012, Poppenga et al., 2013). Software tools such as ArcGIS® Spatial Analyst Tools, Geographic Resources Analysis Support System (GRASS), LandSerf, System for Automated Geoscientific Analysis (SAGA), and Terrain Analysis Using Digital Elevation Models (TauDEM) can help automate the extraction of surface water drainage networks from elevation data, along with furnishing other capabilities. Using these tools within a Windows® operating system on a desktop computer to process large, high resolution digital elevation models (DEMs) can be tedious and time consuming, and may be impractical depending on DEM size and resolution.

Gong and Xie (2009) decomposed DEM data into watersheds and simultaneously extracted drainage networks for multiple watersheds using a collection of computers running Windows® and ArcGIS®. In this case, simultaneous distributed computing enhanced computational efficiency compared to serial processing methods (Gong and Xie, 2009). Stanislawski et al. (2015b) tested simultaneous extraction of drainage networks for multiple watersheds using ArcGIS® Server on a Linux cluster. However, the Windows® emulation employed for ArcGIS® could not handle multiple processing threads, and this strategy did not enhance performance or data throughput (Stanislawski et al., 2015b). Consequently, the drainage network extraction workflow employed by Stanislawski et al. (2015b) was translated to open source and TauDEM geoprocessing functions and implemented on a Linux high-performance computing cluster to substantially enhance performance (Stanislawski et al., 2016).

Typical methods for extracting drainage networks from elevation data involve filling pits in the elevation model, deriving flow direction and flow accumulation surfaces, defining a minimum contributing area threshold that forms a linear drainage feature, and extracting the drainage network (O'Callaghan and Mark, 1984; Jenson and Dominigue, 1988; Tarboton, Bras, and Rodriguez-Iturbe, 1991; Montgomery and Foufoula-Georgiou, 1993; Maidment, 2002; Gong and Xie, 2009; Passalacqua, Belmont, and Foufoula-Georgiou, 2012). Filling pits or depressions is a challenging problem for hydrologic applications (Band, 1999; Hutchinson and Gallant, 1999; Wang and Liu, 2006; Poppenga et al., 2010; Zhu et al., 2013). Improperly filling pits can extract networks that are disconnected because some or all features are not extracted in one or more filled areas, or networks may include overly straight and parallel features within the filled depressions.

This paper discusses an automated geoprocessing workflow that uses a high performance computing environment to improve the United States National Hydrography Dataset (NHD). Specialized techniques to identify, classify and fill depressions in terrain data are used to extract connected networks in various conditions within the conterminous United States. As in Stanislawski et al. (2016), the automated workflow is implemented through a high-performance Linux cluster to extract drainage networks from the U. S. Geological Survey (USGS) 1/3 arc-second 3D Elevation Program (3DEP) elevation data which has a nominal cell size of 10 meters (m). Extracting networks from this publicly available product yields faster results of sufficient quality to support cartographic production and additional research refinements.

Methods

High-resolution (HR) NHD within the conterminous United States is a multi-scale dataset comprised of hydrographic data from the best available data contributed from source scales of 1:24,000 (24K) and larger (Stanislawski et al., 2015b; Stanislawski et al., 2016). Consequently, the HR NHD must be generalized for 24K or smaller scale displays. The NHD is subdivided and distributed in watershed units. There are 18 two-digit region watersheds, and 202 four-digit subregion, and 2,119 eight-digit subbasin watersheds in the conterminous United States (Figure 1). To automate generalization of the HR NHD, elevation-derived 24K natural drainage density patterns are extracted for eight-digit Hydrologic Unit Code (HUC8) subbasin watersheds. Target densities for pruning the HR NHD to 24K and smaller scales are ultimately derived from the extracted 24K drainage patterns (Stanislawski, 2009).

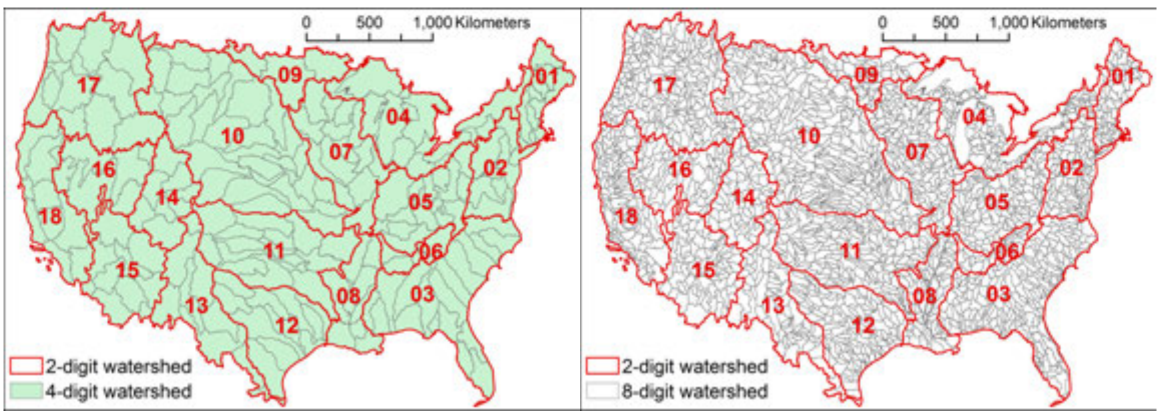

Figure 1. At left, two-digit region (labeled) and four-digit subregion watersheds of the National Hydrography Dataset within the conterminous United States. At right, two-digit region (labeled) and eight-digit subbasin watersheds are shown.

Natural drainage density patterns at 24K are extracted from 1/3 arc-second 3DEP DEM data through weighted flow accumulation (WFA) modelling. Weights for the model are based on runoff data (McCabe and Markstrom, 2007; McCabe and Wolock, 2008) adjusted for terrain slope, soil permeability, soil depth, ground water, and vegetation cover. Geomorphology conditions are estimated at resolutions of 5 kilometers (km) for runoff, nominally 10 m for slope, and 1 km for soil depth, soil permeability, ground water, and vegetative cover. For more details see Stanislawski et al. (2012).

The total amount of vector drainage lines that are extracted from each HUC8 watershed through the WFA model is controlled by parameters derived from the initial version of 24K hydrographic flow line data. In many areas, the initial 24K version of hydrographic data has compilation inconsistencies that do not reflect natural drainage density patterns caused by geomorphologic conditions, and therefore the initial 24K version is not a suitable reference pattern for 24K content. However, the 24K version is the best available synoptic estimate for the *amount* of hydrographic content that should be displayed at 24K within each watershed. Consequently, four parameters are derived from the 24K version for each HUC8 subbasin: (1) total km and (2) total number of natural flow line network features, (3) minimum length of first-order, primarily ephemeral stream tributaries, and (4) polygonal areas that are devoid of 24K flow line features. These parameters along with the associated WFA model control the concentration and distribution of linear drainage features that are extracted from the elevation data for any subbasin.

The first step of the extraction algorithm fills spurious depressions that may obstruct continuous downhill flow over the surface. Spurious depressions are relatively small anomalous depressions in the elevation data. Pit removal algorithms will generally fill these areas adequately. Of concern are other large relatively flat depressions caused by waterbodies such as, ponds, lakes, dry lakes, streams, or possibly flood plains spanning valley bottoms. Filling these features can cause the process to omit drainage line extraction within these features creating improper disconnected networks (Figure 2a). Also a problem is the generation of parallel drainage lines in these large flat areas where the filling process forms one or more adjacent tilted planes (Figure 2b). An accurate set of waterbodies could be overlaid to skip the filling process in the waterbodies, but drainage lines would still be needed in the waterbodies and other flat areas in order to complete and fully connect the network.

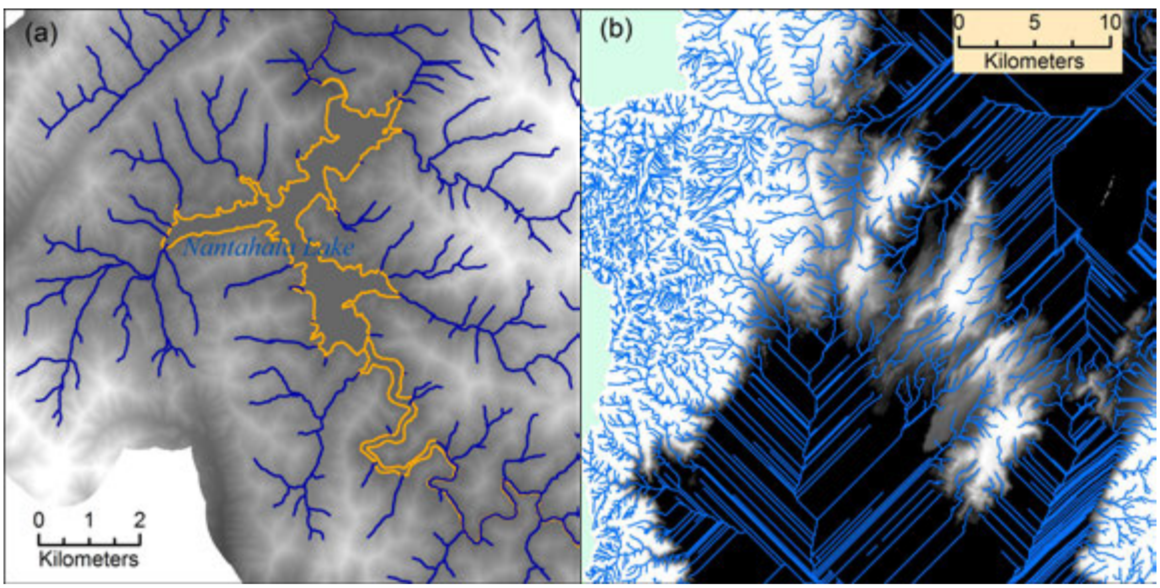

Figure 2. (a) Section of drainage network around Nantahala Lake, North Carolina extracted for Upper Little Tennessee subbasin (NHD HUC8 06010202) after filling depressions. Network features are not extracted in Nantahala Lake (orange boundary) because the filled surface is too flat to generate adequate flow accumulation to form drainage lines. (b) Section of drainage network extracted for Eureka-Saline Valleys subbasin (NHD HUC8 18090201) in California after filling depressions. Filling depressions in these extremely large valleys creates a flat surface causing parallel drainage lines to be extracted. Filled elevation data are represented with gray scale where lighter shades indicate higher elevations.

The key to extracting a fully connected network involves identification and classification of depressions and handling of each class of depression. The method used here classifies the 1/3 arc-second DEM for a watershed based on a terrain roughness estimate. Mean elevation, minimum elevation, and standard deviation of elevation are recorded for a subbasin DEM. In addition, a digital slope surface, recorded as percent rise, is also computed for the subbasin. These data are used to identify large flat areas that are problematic in the drainage line extraction process. Two considerations are built into the process. A rough subbasin has a standard deviation in elevation greater than 77.7 m, otherwise the subbasin is defined as non-rough. A large, relatively flat depression is an area that is greater than 500 acres where slope values are less than 1/30th of the standard deviation of the slope for the subbasin. Threshold values were determined by testing procedures on more than 30 subbasins distributed in various conditions in the conterminous United States and visually verifying the connectivity of the extracted networks.

For a rough subbasin, NHDPlus version 2 (McKay et al., 2012) hydro-DEM data is inserted into the subbasin DEM where the large, relatively flat depressions exist. NHDPlus version 2 hydro-DEM data is hydro-conditioned 1 arc-second (nominally 30-m cell size) elevation data, which has selected 1:100,000-scale (100K) NHD flowlines and waterbody boundaries are burned into the DEM data and HUC12 boundaries incorporated as "walls" during hydro-enforcement of the DEM (McKay et al., 2012). These steps create the hydro-conditioned DEM. At this point, the rough subbasin DEM includes 30-m hydro-conditioned cells that follow 100K hydrography in the flat large depressions and nominally 10-m cells everywhere else. The TauDEM *pitremove* algorithm, which uses the standard flooding approach described by Jenson and Domingue (1988), is applied to remove spurious depressions in the subbasin DEM. Subsequently if the minimum elevation of the filled DEM is not substantially higher (less than or equal to 100 m) than the minimum of the original DEM, then the filled DEM is used for subsequent computations. If the minimum elevation of the filled DEM is substantially higher than the original minimum, then the *pitremove* algorithm is applied to the original DEM, not including the large flat areas, and then the 30-m NHDPlus data is added to the filled DEM in the flat areas.

For non-rough subbasin DEMs (standard deviation of elevation < 77.7 m), the *pitremove* algorithm is applied to the entire subbasin DEM.

After hydro-conditioning of the subbasin DEM is complete, D8 flow directions (O'Callaghan and Mark, 1984) and the WFA model are computed. The total km of drainage lines to extract for the subbasin (parameter estimated from 24K version of flow lines) is expanded to account for pruning of short first order tributaries and simplification that will remove the blockiness after drainage lines are extracted. The expanded length is converted to the number of cells needed from the flow accumulation dataset that will form the required amount of vector drainage lines. The minimum WFA value that extracts the required number of cells is determined through an iterative process that accounts for drainage lines that may be extracted in the devoid polygons, if any exist. Once the minimum WFA value is found, flow accumulation cells greater than the minimum are converted to vector lines and simplified. The conversion process computes stream order values for the extracted vector network. First order features that are shorter than the minimum length, as defined by the input parameters, are pruned.

Open Source Tools

The drainage network extraction process is computationally intensive. Therefore, the network extraction step is implemented using portable Python scripts coded with open source tools, including the Geospatial Data Abstraction Library (GDAL, http://gdal.org) and TauDEM (https://github.com/dtarb/TauDEM). GDAL is deployed on a Linux cluster as C/C++/Python programming libraries. TauDEM uses a parallel programming model for data decomposition, runtime communication, and parallel input/output. The software is deployed on a five-node Linux cluster at the U.S. Geological Survey (USGS). Each node has 20 processing cores and 64 gigabytes of shared Random Access Memory (RAM). A parallel Lustre file system on a high-speed Infiniband interconnect provides rapid access to file storage. Scheduling jobs to simultaneously extract drainage networks from multiple subbasins is managed virtually through the Simple Linux Utility for Resource Management (SLURM).

Test Data

A set of 30 test HUC8 subbasins distributed in the conterminous United States within three different climate regions having various terrain conditions is used to test the efficacy of the drainage network extraction methods (Figure 3). Test subbasins range in size from 1300 to 7500 square km. Ten subbasins are situated in dry, humid, and transitional climate regions. The dry region experiences less than 140 millimeters per year (mm/year) of runoff, and the humid region experiences more than 140 mm/year of runoff (McCabe and Markstrom, 2007; McCabe and Wolock, 2008). Test subbasins in the dry and humid regions are distributed within three terrain slope categories: 0.0 to 1.5, 1.5 to 7.0, and greater than 7.0 percent rise. The third set of ten subbasins lie in areas having conditions that transition between more than one slope or climate category.

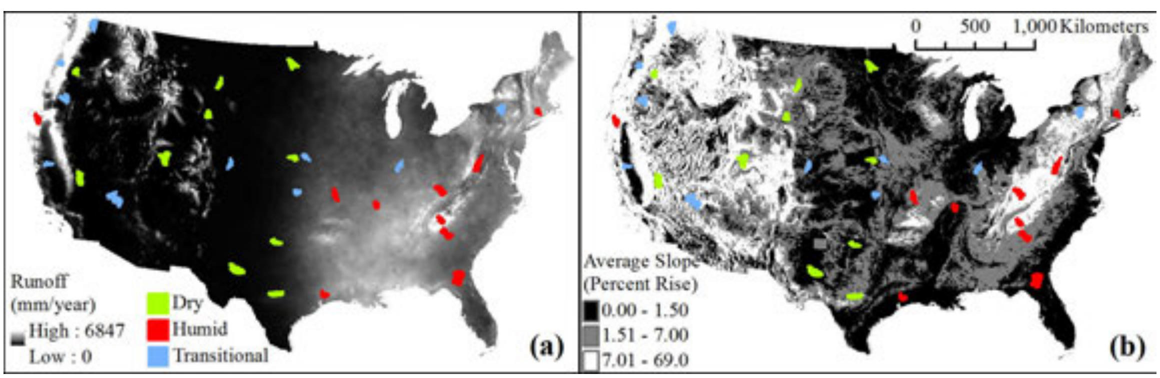

Figure 3. Distribution of 30 test subbasins in the conterminous United States displayed over (a) mean annual runoff and (b) average slope.

Similarity Testing

To validate the quality of the extracted 24K drainage networks, each of the 30 extracted networks is compared to the 24K version NHD flowline network for its associated subbasin. An automated approach estimates the similarity of the two associated networks through the Coefficient of Line Correspondence (CLC) (Stanislawski et al., 2015a). The approach automatically identifies the matching and mismatching linear features in the two networks. The CLC metric is computed as the sum of the length of the matching lines from both datasets divided by the sum of the length of all lines in both datasets. CLC values range from 0.0, indicating no matching features in both datasets, to 1.0, indicating all features are matching in both datasets.

Results

Visual inspection of extracted drainage networks for the 30 subbasins indicate that all networks are well connected in all areas, including the problematic large, flat depression areas illustrated in Figure 2. An example is shown in Figure 4a, illustrating how extracted network features are fully connected through Nantahala Lake, North Carolina. In this case, DEM conditioning includes 30-m hydro-conditioned NHDPlus data in large flat depressions and subsequent depression filling through the *pitremove* algorithm. The 100K channels burned through the NHDPlus data are sufficient to create channels through the flat waterbody areas.

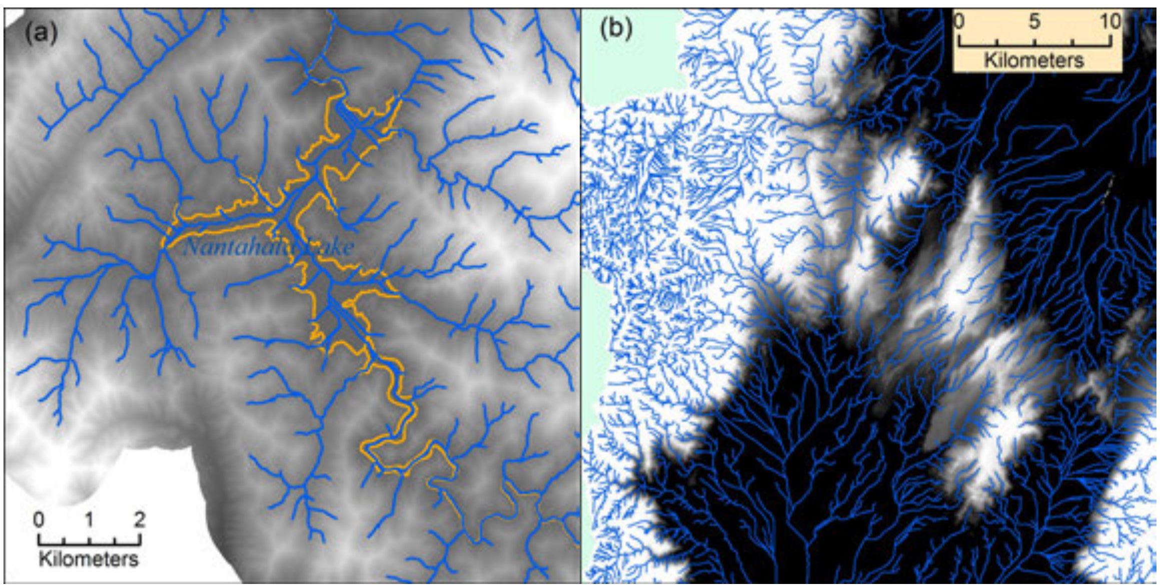

Figure 4. (a) Section of drainage network around Nantahala Lake, North Carolina extracted for Upper Little Tennessee subbasin (NHD HUC8 06010202) after including NHDPlus hydro-conditioned elevation data in large flat depressions and subsequently filling spurious depressions. Network features are extracted in Nantahala Lake (orange boundary). (b) Section of drainage network extracted for Eureka-Saline Valleys subbasin (NHD HUC8 18090201) in California after filling depressions except in large flat depression areas and then including NHDPlus hydro-conditioned elevation data in large flat depression areas. Comparison with Figure 2 shows that this conditioning process does not create unrepresentative parallel flow vectors through the large flat depression areas. Filled elevation data are represented with gray scale where lighter shades indicate higher elevations.

Table 1 summarizes the differences in the statistics of the subbasin DEM datasets before and after conditioning with NHDPlus data in large depressions and subsequent global filling. As highlighted with the yellow-shaded rows in Table 1, about half of the tested subbasins have rough terrain requiring a search for large flat depression areas. For the Eureka-Saline Valleys subbasin (HUC8 18090201), the difference of the minimum elevations before and after conditioning is greater than 100 m; therefore, the NHDPlus data inserted in the large flat depression areas is not reconditioned with the *pitremove* algorithm. As seen in Figure 3b, this step leaves enough relief in the large flat depressions to form adequate flow accumulation to extract more sinuous drainage lines in the depressions than the parallel lines that were extracted from the globally filled DEM as previously shown Figure 2a.

CLC values measuring the similarity of all, first order, and higher order features in the 24K extracted networks and 24K version NHD flowline networks for the 30 subbasins are shown in Figure 5. For all features the average, minimum, and maximum CLC values are 0.73, 0.50, and 0.85, respectively. So, on average about 73 percent of the extracted channels match the NHD flowlines, and in the worst case subbasin, only about 50 percent of the features match between the two datasets. Lower matching (about 0.50 to 0.60) is found in the subbasins having average slope less than about 1.5 to 2 degrees. Low slope

Table 1: Summary statistics of 1/3 arc-second DEM datasets for 30 subbasin watersheds before and after adjustments to include NHDPlus data and remove depressions. Yellow rows indicate standard deviation of elevation is greater than 77.7 meters. Red cell indicates the difference of the minimums of the adjusted and original datasets for the subbasin is greater than 100 meters.

	Original Elevation Data (meters)				Elevation Data After Inserting NHDPlus in Low Relief Areas (for some subbasins) and Removing Depressions (meters)				Adjusted – Original (meters)	
Subbasin	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum	Minimum Difference	Maximum Difference
1100002	171.5	65.3	-1.8	399.4	172.0	64.9	0.0	399.4	1.8	0.0
2050102	439.8	72.2	249.5	653.6	439.8	72.2	249.8	653.6	0.3	0.0
2070001	650.6	267.6	160.2	1482.5	650.6	267.6	160.2	1482.5	0.0	0.0
3060104	211.6	54.7	98.6	556.3	211.6	54.8	100.4	556.3	1.8	0.0
3110201	41.1	7.8	8.4	96.6	41.2	7.7	8.4	96.6	0.1	0.0
5100201	393.2	97.1	189.3	1002.7	393.3	97.1	189.9	1002.7	0.6	0.0
5120203	224.1	39.7	149.3	317.3	224.3	39.6	149.3	317.3	0.0	0.0
6010202	906.7	244.2	410.3	1700.8	905.6	248.5	-643.4	5380.9	-1053.7	3680.1
8020201	94.4	5.5	81.7	182.9	94.6	5.4	83.1	182.9	1.5	0.0
9020109	370.0	59.8	251.9	482.0	370.1	59.8	251.9	482.0	0.0	0.0
10090204	1602.6	102.8	1337.2	1923.9	1602.6	102.8	1337.2	1923.9	0.0	0.0
10090210	961.8	78.9	755.9	1263.5	961.8	78.9	755.9	1263.5	0.0	0.0
10190013	1472.2	103.8	1264.3	1825.4	1472.2	103.8	1264.3	1825.4	0.0	0.0
10240006	350.8	34.6	268.2	445.6	350.8	34.6	269.0	445.6	0.8	0.0
10270204	466.9	31.7	384.7	542.4	466.9	31.7	384.7	542.4	0.0	0.0
10290107	318.2	46.6	214.9	468.6	318.5	46.4	215.3	468.6	0.4	0.0
11030017	424.5	20.7	361.0	498.2	424.5	20.7	361.2	498.2	0.2	0.0
11130102	343.3	29.7	265.5	439.7	343.3	29.7	265.5	439.7	0.0	0.0
12020007	20.1	10.0	0.6	68.7	20.1	10.0	0.6	68.7	0.0	0.0
12080005	899.1	64.1	766.2	1047.6	899.3	64.2	766.2	1047.6	0.0	0.0
12090206	487.8	109.9	196.1	697.1	487.8	109.9	196.3	697.1	0.2	0.0
14060005	1995.2	387.1	1261.8	3102.4	1995.3	387.9	-754.0	6556.7	-2015.9	3454.4
15010005	959.1	457.1	348.2	2466.3	951.1	480.1	-1038.1	6448.2	-1386.3	3981.9
17020008	1419.5	503.2	238.5	2720.2	1419.5	503.1	238.5	2720.2	0.0	0.0
17070307	995.1	241.5	391.0	1808.6	995.1	241.5	391.0	1808.6	0.0	0.0
17090009	369.9	346.1	18.2	1528.7	370.7	349.3	18.6	5057.1	0.3	3528.4
18010202	1599.9	216.8	1269.2	2549.7	1599.9	216.8	1269.2	5311.0	0.0	2761.2
18010212	1079.9	299.6	135.8	2387.0	1079.9	299.6	135.8	2387.0	0.0	0.0
18040002	196.6	242.9	4.8	1160.4	213.9	337.5	-1163.7	5060.3	-1168.4	3899.9
18090201	1565.4	656.2	322.5	3525.3	1799.0	422.4	1489.9	6547.5	1167.3	3022.2

subbasins may have predominantly swampy, karst, or coastal conditions where man-made drainage features may be prevalent, and consequently the actual hydrographic features may not match very well with extracted natural flow patterns inherent in the topography.

CLC values comparing first order features in associated subbasins range from 0.31 to 0.73, with an average of 0.59, and CLC values comparing second and higher order features range from 0.66 to 0.99, with an average of 0.89. Thus, the majority of mismatching is found in the first order features, which largely can be attributed to cartographic constraints that limited the initial collection of headwater features (first order tributaries) to a minimum length or minimum distance from nearest ridge (U.S. EPA and U.S. DOI, 1999). Results corroborate findings by other researchers, who suggest that NHD flowlines do not adequately represent headwater features for engineering and hydrologic purposes (Colson et al., 2008; Fritz et al., 2013; Caruso, 2014). Consequently extracted features may help guide or prioritize updating the NHD with additional headwater content.

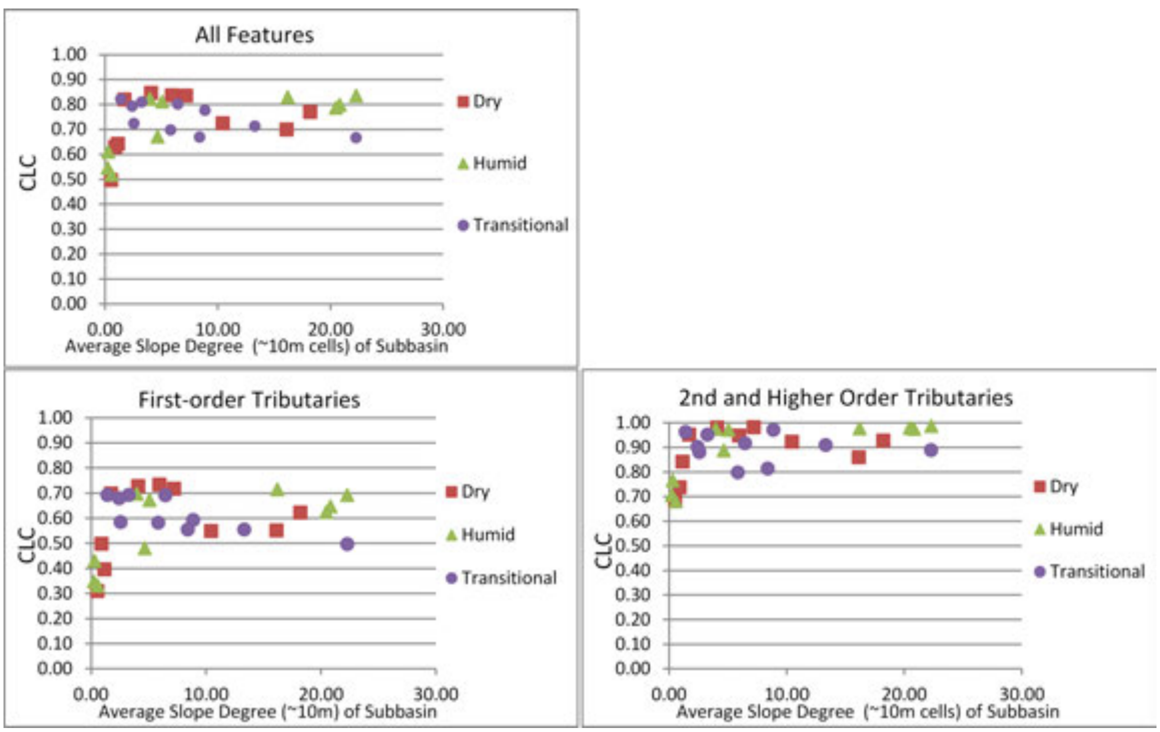

Figure 5. Coefficient of Line Correspondence (CLC) values estimating the similarity of the 1:24,000-scale (24K) elevation-derived network and the 24K version NHD flowline network. CLC values can range from 0.00 for complete mismatch of all features to 1.00 for perfectly matching line sets. Subbasin CLC values are plotted against average elevation slope values. Top chart shows CLC values for all features, bottom left chart shows CLC values comparing only first order features in both networks, and bottom right chart shows CLC values comparing 2nd and higher order features in both networks.

The Linux cluster implemented for this research contains 100 processing nodes, which allows simultaneous processing of up to 100 subbasins. Tests on the 30 subbasins and other data demonstrate the capacity to extract networks for about 500 subbasins in about 16 hours. This level of throughput opens new research possibilities, such as a thorough sensitivity analysis and refinement of the weighted flow accumulation and drainage network extraction process, and review of the 24K parameter estimation process. Furthermore, additional work is needed to validate the location and concentration of extracted channels.

Conclusions

Hydrographic data is relevant to many if not most mapping and analytical applications; and yet it is challenging to process because it is highly sensitive to scale-change and because it must be integrated carefully with other data layers such as terrain. These two characteristics mandate advanced processing methods that are labor- and computationally intensive, making a national coverage hydrographic dataset expensive to maintain, especially in an expansive geographic area such as the conterminous United States. A particularly difficult step in data processing involves identifying and filling terrain depressions (sinks) that exist in the data but not in the actual landscape. Advances in open source software, concurrent processing and high performance computing reduce processing costs and time, as well as improve consistency of results.

The research reported here demonstrates the utility of open source tools with parallel computing support for filling depressions in 30 HUC8 subbasins distributed across humid, dry and transitional climate regions in terrain conditions exhibiting a range of differing slopes. Special attention is given to low slope terrain, where network connectivity is preserved by generating synthetic stream channels through lake and waterbody polygons. Conflation analysis compares the extracted streams with a 24K NHD flowline network, and shows that similarities are highest for 2nd and higher order tributaries. This result is expected, given that 1st order tributaries in the vector NHD database are not always complete. One goal of extracting stream channels from terrain data is to improve existing data stores for 1st order tributaries, in addition to improving processing speed and accuracy of recorded data. Future research will continue to refine results, with further examination of processing in subbasins situated in low slope terrain.

References

Anderson, D.L. (2012) Detailed Hydrographic Feature Extraction from High-Resolution LiDAR Data. Doctoral Dissertation, Idaho State University, 115 p.

Band, L.E. (1999) Spatial Hydrography and Landforms, in Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., eds., *Geographical Information Systems: Principles and Technical Issues (2d ed.)*, New York, Wiley, v. 1, pp. 527–542.

Caruso, B.S. (2014) GIS-Based Stream Classification in a Mountain Watershed for Jurisdictional Evaluation. *JAWRA Journal of the American Water Resources Association* 50, 5, pp. 1304–1324. doi:10.1111/jawr.12189.

Colson, T., Gregory, J., Dorney, J., and Russell, P. (2008) Topographic and Soil Maps Do Not Accurately Depict Headwater Stream Networks. *National Wetlands Newsletter* 30, 3, pp. 25–28.

Fritz, K. M., Hagenbuch, E., D'Amico, E., Reif, M., Wigington Jr., P.J., Leibowitz, S.G., Comeleo, R.L., Ebersole, J.L., and Nadeau, T.-L. (2013) Comparing the Extent and Permanence of Headwater Streams from Two Field Surveys to Values from Hydrographic Databases and Maps. *JAWRA Journal of the American Water Resources Association* 49, 4, pp. 867–882. doi:10.1111/jawr.2013.49.issue-4.

Gong, J., and Xie, J. (2009) Extraction of Drainage Networks from Large Terrain Datasets Using High Throughput Computing. *Computers & Geosciences*. 35, 2009, pp. 337-346. doi:10.1016/j.cageo.2008.09.002.

Hutchinson, M.F., and Gallant, J.C. (1999) Representation of Terrain, in Longley, P.A., Goodchild, M.F., Maguire, D.J., and Rhind, D.W., eds., *Geographical Information Systems: Principles and Technical Issues (2d ed.)*, New York, John Wiley, v. 1, pp. 105–124.

Jenson, S. K., and Dominigue, J.O. (1988) Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis. *Photogrammetric Engineering and Remote Sensing*, 54, 11, pp. 1593–1600.

Maidment, D.R. (2015) A Conceptual Framework for the National Flood Interoperability Experiment. Center for Research in Water Resources, University of Texas at Austin, Revised February 9, 2015, 22 p., https://www.cuahsi.org/Files/Pages/documents/13623/nfieconceptualframework_revised_feb_9.pdf Last visited 07/06/2016.

Maidment, D. R., ed. (2002) *Arc Hydro: GIS for Water Resources*. Redlands, CA: ESRI Press.

McCabe, G. J., and Markstrom, S. L. (2007). A Monthly Water-Balance Model Driven by a Graphical User Interface, in *U.S. Geological Survey, Open-File Report 2007-108*. 6 pp., U.S. Geological Survey, Reston, Virginia.

McCabe, G. J., and Wolock, D. M. (2008). Joint Variability of Global Runoff and Global Sea Surface Temperatures. *Journal of Hydrometeorology*. 9, pp. 816–824.

McKay, L., Bondelid, T., Dewald, T., Rea, A., Johnston, C., Moore, R. (2012) *NHDPlus Version 2: User Guide*, United State Environmental Protection Agency, Washington, D.C., 170 p. ftp://ftp.horizon-systems.com/nhdplus/nhdplusv21/documentation/nhdplusv2_user_guide.pdf, Last visited 07/13/2016.

Montgomery, D. R., and Foufoula-Georgiou, E. (1993) Channel Network Source Representation Using Digital Elevation Models. *Water Resources Research.* 29, 12, pp. 3925–3934. doi:10.1029/93WR02463.

O'Callaghan, J. F., and Mark, D.M. (1984) The Extraction of Drainage Networks from Digital Elevation Data. *Computer Vision, Graphics, and Image Processing.* 28, pp. 323–344. doi:10.1016/S0734-189X(84)80011-0.

Passalacqua, P., Belmont, P., and Foufoula-Georgiou, E. (2012) Automatic Geomorphic Feature Extraction from Lidar in Flat and Engineered Landscapes. *Water Resources Research.* 48, pp. 1–18. doi:10.1029/2011WR010958.

Poppenga, S.K., Worstell, B.B, Stoker, J.M, and Greenlee, S.K. (2010) Using Selective Drainage Methods to Extract Continuous Surface Flow From 1-Meter Lidar-Derived Digital Elevation Data. *U.S. Geological Survey Scientific Investigations Report 2010-5059*, 12 pp. http://pubs.usgs.gov/sir/2010/5059/pdf/SIR2010_5059.pdf, Last visited 07/15/2016.

Poppenga, S.K., Gesch, D.B., Worstell, B.B. (2013) Hydrography Change Detection: The Usefulness of Surface Channels Derived from Lidar DEMS for Updating Mapped Hydrography. *Journal of the American Water Resources Association*. 49, 2, pp. 371-389.

Stanislawski, L.V. (2009). Feature Pruning by Upstream Drainage Area to Support Automated Generalization for the United States National Hydrography Dataset. *Computers, Environment and Urban Systems*, 33, pp. 325–333.

Stanislawski, L.V., Doumbouya, A.T., Miller-Corbett, C.D.,Buttenfield, B.P., and Arundel-Murin, S.T. (2012) Scaling Stream Densities for Hydrologic Generalization, in Seventh International Conference on Geographic Information Science, September 18-21, Columbus, Ohio.

Stanislawski, L.V., Buttenfield, B.P., and Doumbouya, A. (2015a) A Rapid Approach for Automated Comparison of Independently Derived Stream Networks. *Cartography and Geographic Information Science*. 42, 5, pp. 435-448, doi: 10.1080/15230406.2015.1060869.

Stanislawski, L.V., Falgout, J., Buttenfield, B.P. (2015b) Automated Extraction of Natural Drainage Density Patterns for the Conterminous United States through High Performance Computing. *The Cartography Journal* 52, 2, pp. 185-192. doi: 10.1080/00087041.2015.1119466.

Stanislawski, L.V., Liu, Y., Buttenfield, B.P., Survila, K., Wendel, J., and Okok, A. (2016) High Performance Computing to Support Multiscale Representation of Hydrography in the Conterminous United States. Proceedings, 19th ICA Workshop on Generalization and Multiple Representation, June 14, 2016, Helsinki, Finland, 10 p.

Tarboton, D. G., Bras, R.L., and Rodriguez-Iturbe, I. (1991) On the Extraction of Channel Networks from Digital Elevation Data. *Hydrological Processes.* 5, pp. 81–100. doi:10.1002/(ISSN)1099-1085.

U.S. EPA, and U.S. DOI. 1999. Standards for National Hydrography Dataset – High Resolution. U.S. Environmental Protection Agency and U.S. Department of the

Interior, National Mapping Program Technical Instructions, July. U.S. Geological Survey.

Wang, L. and Liu, H. (2006) An Efficient Method for Identifying and Filling Surface Depressions in Digital Elevation Models for Hydrologic Analysis and Modelling. *International Journal of Geographical Information Science*, 20, pp. 193–213.

Zhu, D., Ren, Q., Xuan, Y., Chen, Y., and Cluckie, I.D. (2013) An Effective Depression Filling Algorithm for DEM-Based 2-D Surface Flow Modelling. *Hydrology and Earth System Sciences.* 17, pp. 495-505. doi:10.5194/hess-17-495-2013.

Lawrence V. Stanislawski, U.S. Geological Survey, Department of the Interior, 1400 Independence, Rolla, MO 65401.

Kornelijus Survila, CyberInfrastructure and Geospatial Information Laboratory (CIGI), University of Illinois at Urbana-Champaign, Davenport Hall, 607 South Mathews Avenue, Urbana, IL 61801

Jeffrey Wendel, U.S. Geological Survey, Department of the Interior, 1400 Independence, Rolla, MO 65401.

Yan Liu, CyberInfrastructure and Geospatial Information Laboratory (CIGI), University of Illinois at Urbana-Champaign, Davenport Hall, 607 South Mathews Avenue, Urbana, IL 61801.

Barbara P. Buttenfield, Department of Geography, University of Colorado at Boulder. Boulder, CO 80309.