Assessment of a Rapid Approach for Estimating Catchment Areas for Surface Drainage Lines

U.S. Geological Survey
Center of Excellence for Geospatial Information Science
1400 Independence Road, Rolla Missouri, 65401

Lawrence V. Stanislawski
Science Applications International Corporation
Phone: (573) 308-3914
Fax: (573) 308-3652
E-mail: lstan@usgs.gov

Michael P. Finn
United States Geological Survey
Phone: (573) 308-3931
Fax: (573) 308-3652
E-mail: mfinn@usgs.gov

Mark Barnes
United States Geological Survey
Phone: (573) 308-3575
Fax: (573) 308-3652
E-mail: mbarnes@usgs.gov

E. Lynn Usery
United States Geological Survey
Phone: (573) 308-38
Fax: (573) 308-3652
E-mail: usery@usgs.gov

Abstract

The catchment, or catchment area, is the area associated with a segment of a drainage network. Surface runoff in a catchment flows into the associated network segment. Catchment area is commonly used for estimating watershed parameters, such as upstream drainage area, flood frequency, or surface-water flow. Hydrologic analyses require relatively accurate catchment area estimates to furnish substantive conclusions. Conversely, alternative uses for catchment area may not require such precise estimates. Estimation of catchment areas using Thiessen polygons around evenly spaced points along network features is a rapid and computationally simple approach when compared to the alternative of deriving catchments from an elevation model; however, Thiessen-estimated catchments do not precisely follow the ridgelines of surface models and only provide approximate subdivisions of the watershed. This paper assesses the accuracy of Thiessen-derived catchment area estimates for several drainage networks having various drainage densities and landscape types. Thiessen-derived catchment area estimates are compared to those derived from surface-elevation models.

Keywords: Drainage network, catchment, catchment area, hydrography, Thiessen polygons, Voronoi polygons

1. Introduction

The National Hydrography Dataset (NHD) is a vector data layer of *The National Map* representing the surface waters of the United States (USGS, 2000). It is stored in an ArcGIS geographic database (geodatabase) model at two levels of detail: medium (1:100,000-scale source) and high (1:24,000-scale or larger source) resolution. Currently (2007), full coverage for the conterminous U.S exists for the medium resolution layer, and the high resolution layer is nearing completion. Although available resolutions of the NHD are suitable for many applications, various uses may require levels of detail that differ from those available. More detailed data can be acquired through additional collection, but generalization of existing data is required to furnish less detailed NHD data, which are suitable for regional or national studies.

Optimization of geospatial database design and maintenance is a common goal of generalization research (Mackaness, 2006; Chaudry and Mackaness, 2006). A robust automated generalization process could make it feasible to store and maintain only the most accurate, high resolution layer of the NHD, eliminating storage and maintenance of the less accurate, low resolution layer. An objective of this research is to develop an automated generalization process capable of producing a subset of NHD features that maintains the NHD model format but at a user-defined level of detail.

To date (2007), the NHD generalization process consists of feature pruning and simplification. Nearly all generalization processes include an initial step of selecting objects and attributes from the source database that are to be represented in the generalized dataset (McMaster and Shea, 1992). We refer to the object selection process as feature pruning. After pruning the network and associated area features, remaining features must be simplified. Feature simplification is accomplished through two processes: rule-based feature type modifications (such as amalgamation and linearization of polygonal features), and removal of vertices.

Our network pruning strategy extracts the most prominent network features based on the relative extent of the watershed surface that flows into the network features. To accomplish this task, catchment area estimates must be acquired for each network segment. Subsequently, we apply an augmented directed graph approach to assign upstream drainage area (UDA) estimates to the segments (Stanislawski and others, 2006), which are then used to prune less significant network features. The approach of pruning by UDA follows the same logic as the Pfafstetter system for topologically coding river basins and networks (Verdin, 1997). In addition, UDA is the most significant factor for estimating stream flow volumes in the National Flood Frequency Program (USGS, 2002).

Preprocessing a data layer to prepare it for automated generalization is fairly common practice (Yan and others, 2006; Stoter, 2005). Part of the preparation process for our automated generalization strategy includes assignment of catchment area and UDA values to network segments. The area associated with a segment of a drainage network is referred to as the segment's catchment area, or just catchment. Surface runoff in the catchment flows into the associated network segment. Common approaches for estimating catchments use rigorous algorithms involving a digital elevation model (DEM) and can require lengthy processing times.

Efforts to derive catchments and other value-added attributes for the NHD are ongoing but focused on the medium resolution layer (USEPA and USGS, 2006). To test and demonstrate our automated generalization approach on the high resolution NHD, we developed a rapid automated approach that applies Thiessen polygons (Thiessen and Alter, 1911) to estimate catchments for each segment of an NHD network. This paper assesses the quality of Thiessen-polygon-derived catchment estimates, which are a required preprocessing quantity for our network pruning strategy.

2. Methods

Surface drainage networks are used regularly for hydrologic or other environmental studies. A digital surface drainage network depicts channels on the terrain surface through which water will or does flow, depending on various terrain and environmental conditions. A drainage network may be derived from a DEM using tools provided in software packages such as the Topographic Parameterization Software (Garbrecht and Martz, 2005) or Arc Hydro (Maidment, 2002). Exclusively considering effects due to topography, a DEM-derived network ensures that a network channel segment exists for each minimum drainage area within a watershed, thus rendering a fairly homogeneous density of channels throughout a study area.

Alternatively, a drainage network may be compiled from vector hydrographic data, such as that available in the NHD. Hydrographic network data typically include only those channels having water flow that is sufficient for intermittent or perennial classification; therefore, networks compiled from vector hydrographic data may exhibit various drainage densities over an area of interest. Consequently, elevation-derived and hydrographic-feature-compiled drainage networks for the same area can be substantially different in content and location of network segments. For this reason, this study only evaluates catchments estimated for elevation-derived networks.

In this study, we compare two different methods for generating catchments: elevation-derived (ED) and Thiessen-polygon-derived (TPD). We assume ED catchments are more accurate than TPD catchments, and we quantify how well the TPD catchments overlay, or correspond, with the ED catchments.

2.1. Elevation-derived (ED) catchments

Elevation-derived catchments were computed using the terrain processing tools in ArcHydro (version 1.1, Maidment, 2002). A brief synopsis of the processing steps is: 1) fill DEM sinks; 2) build flow-direction grid; 3) build flow-accumulation grid; 4) build a stream grid for the user-specified threshold of cells for stream formation, where the threshold is the minimum number of cells in the flow accumulation grid used to define a stream; 5) build a stream-segment grid with associated start and end locations; 6) delineate a catchment grid that associates a catchment to each stream segment; 7) convert catchment grid to a vector polygon file maintaining the stream segment identifiers; 8) convert the stream-segment grid to vector stream file maintaining stream-segment identifiers; and 9) build adjoint catchments, which aggregates all upstream catchments into one catchment at each confluence, and associates a catchment ID to each stream segment.

The last step (step 9) was required only for the transfer of the catchment ID to the associated stream segments. It did contribute significantly to processing times, and it failed to process some of the smaller stream-formation thresholds.

2.2 Thiessen-polygon-derived (TPD) catchments

Thiessen-polygon-derived (TPD) catchments for a network are computed as follows: 1) each segment is assigned a unique ID; 2) a set of points are systematically located along each segment (15-meter spacing was used for this study); 3) Thiessen polygons are derived for all points on the network; and 4) Thiessen polygons are aggregated by segment ID to provide a catchment for each segment. This process is automated through an Arc Macro Language (AML) program.

2.3 Catchment comparisons

Six NHD subbasins that fall in one of six regimes based on climate and topography were evaluated (table 1). Subbasin boundaries were extracted from the NHD (USGS, 2006b). A 30-meter resolution DEM was extracted from the National Elevation Dataset (NED) (USGS, 2006a) for each subbasin, including a 600-m buffer. Elevation summary statistics were computed for each subbasin (table 1).

Table 1. Study Area Subbasin Characteristics

[NHD, National Hydrography Dataset; Precip., precipitation; in/yr, inches per year; elev., elevation; m, meters; st. dev., standard deviation; deg, degree]

Subbasin name	State	NHD subbasin number	Regime	Physiographic division[1]	Mean annual precip.[2] (in/yr)	Mean elev. (m)	St. dev. elev. (m)	Mean slope (deg)	St. dev. slope (deg)
Upper Suwannee	FL, GA	03110201	Flat Humid	Atlantic Plain of Coastal Plain	55	41.2	7.6	0.3	0.6
Lower Beaver	UT	16030008	Flat Dry	Intermontane Plateaus of Basin and Range	12.5	1,536.6	166.9	3.7	6.2
Pomme De Terre	MO	10290107	Hilly Humid	Interior Highlands of Ozark Plateaus	42.5	319.7	46.8	3.4	2.8
Lower Prairie Dog Town Fork Red	TX	11120105	Hilly Dry	Interior Plains of Great Plains and Central Lowland	22.5	629.9	88.0	2.5	2.5
South Branch Potomac	WV	02070001	Mountainous Humid	Appalachian Highlands of Valley and Ridge	40	667.6	276.7	14.0	8.6
Piceance-Yellow	CO	14050006	Mountainous Dry	Intermontane Plateaus of Colorado Plateaus	15	2,199.8	200.1	13.0	8.1

[1] Physiographic division (USGS, 2004).

[2] Mean annual precipitation (National Atlas of U.S., 2006).

Streams and catchments were derived from each subbasin elevation model for stream formation thresholds of 50, 100, 200, 300, 400, 800, and 1200 cells, which is equivalent to 0.045, 0.09, 0.18, 0.27, 0.36, 0.72, and 1.08 square kilometers, respectively. These threshold values provide a range of stream and catchment densities for each subbasin; however, because of the previously mentioned Arc Hydro processing limitation, the 50-cell threshold streams and catchments were not generated for the FL-GA, TX, and WV subbasins. Likewise, the 100-cell threshold streams and catchments could not be computed for the FL-GA subbasin. Only those ED catchments having an associated network segment falling completely within the subbasin were included in the analysis. Furthermore, all catchments having a TPD edge defining the edge of the subbasin study area were excluded from comparisons. This was done because TPD catchments on the edge of the subbasin invariably included parts that do not overlap any ED catchment, which was not included in the commission error computations.

TPD catchments were generated for all ED network segments of each subbasin, and the TPD catchments used the ED segment IDs; therefore, a one-to-one association exists between the TBD and ED catchments. For each subbasin and stream-formation threshold, ED catchments and TPD catchments were compared through a spatial union. Percent correct, percent omission, and percent commission that each TPD catchment covered of its associated ED catchment area was computed from the area values in the spatial union. The area of a commission error for a TPD catchment is the area of the TPD catchment that does not overlay any of the associated ED catchment, and is, therefore, committed to some other ED catchment or catchments (figure 1).

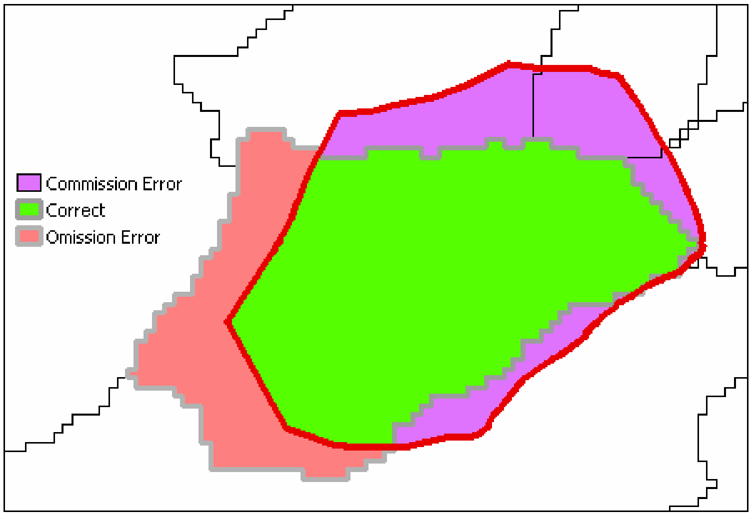

Figure 1. Thiessen-derived catchment (red outline) overlaying associated elevation-derived catchment (gray outline) with correct area in green, and areas of commission error in purple and omission error in pink.

For each subbasin, the mean percent correct, mean percent omission, and mean percent commission were computed for each stream-formation threshold. The mean percent correct for any stream-formation threshold is the sum of percent correct values of all TPD catchments in the subbasin divided by the number of TPD catchments in the subbasin. Mean percent omission and mean percent commission are computed the same way, but using percent omission and percent commission values, respectively. Subsequently, the means were averaged for all thresholds for each subbasin. In addition, total percent correct was computed for each threshold as the sum of all correct areas in the TPD catchments, divided by the total area of all ED catchments in the subbasin, and expressed as a percent. Total percent correct also was averaged for all thresholds for each subbasin.

Another statistic evaluated is the coefficient of areal correspondence (CAC), which is computed for any two associated areas as the area of intersection, divided by the area of union (Taylor, 1977). In figure 1, the CAC for this catchment would be computed as the green area divided by the sum of all colored areas. CAC was computed for all catchments of each subbasin and stream-formation threshold, and subsequently summarized in the same manner as percent correct values.

3. Results

The number of catchments computed for any one subbasin ranges from 957 to 24,603. Obviously, catchment density increases with decreasing stream-formation threshold. Using a Pentium(R) 4 CPU, 3.0

GHz machine with 1 GB RAM, computation times for TPD catchments of a subbasin ranged from 2 to 14 minutes. On the other hand, computation times for ED catchments ranged from about 10 minutes to 2 hours, not including the adjoint catchment process, which could take from 3 minutes to more than 5 hours. In some of the small threshold cases, the adjoint catchment process failed because of software and/or hardware limitations. Thus, the TPD catchment process is about 5 to 10 times faster than the ED catchment process. Aside from being a faster and more reliable process, the TPD catchment process can be applied directly to any vector network--such as that furnished by the NHD—without integrating the vector network with a DEM, which is an additional requirement for the ED catchment process when working with non-ED networks.

3.1 Between subbasin comparisons

Comparisons between subbasins are summarized in figure 2. Averages of the mean percent correct values range from about 50 to 65, with averages better than 60 percent on hilly and mountainous subbasins. Average total percent correct values range from about 58 to 75. Furthermore, average total percent correct is greater than average mean percent correct for all subbasins, suggesting that the percent correct values have a skewed distribution with the mode being larger than the mean, and/or a greater portion of the larger catchments have a higher-than-average percent correct value. The former is exemplified in figure 3, where the mode of the distribution is 71, and the latter is exemplified in figure 4. Both figures show percent correct values for all catchments in the 100-cell stream-formation threshold of the mountainous humid subbasin (WV).

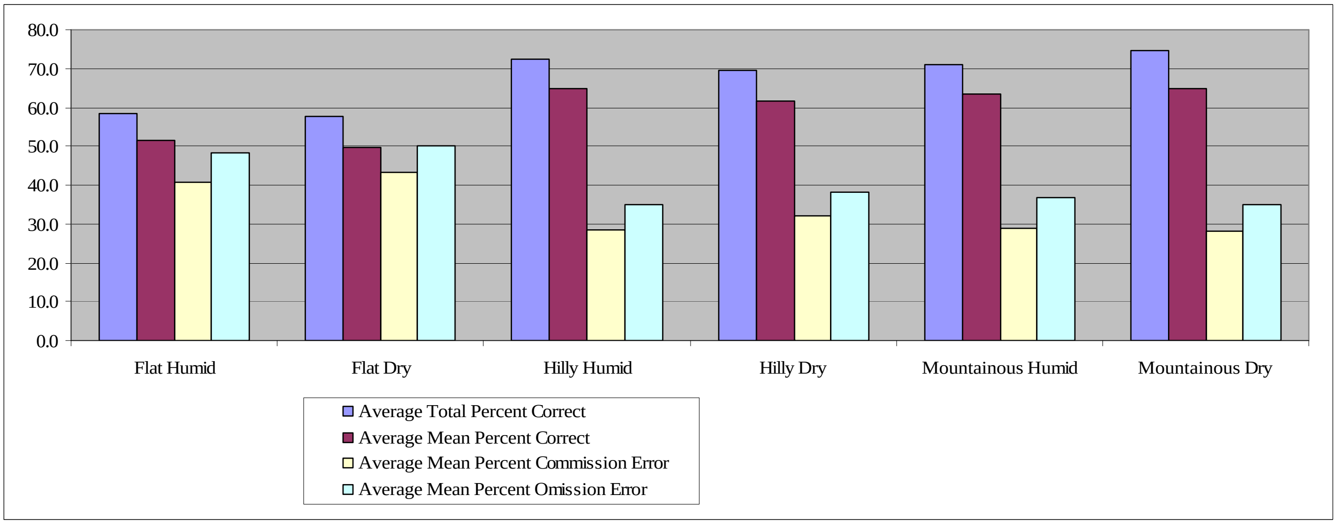

Figure 2. Summary of mean percent correct and error areas of TPD catchments compared to ED catchments for each subbasin. Averages of the mean for each stream-formation threshold are shown for each subbasin.

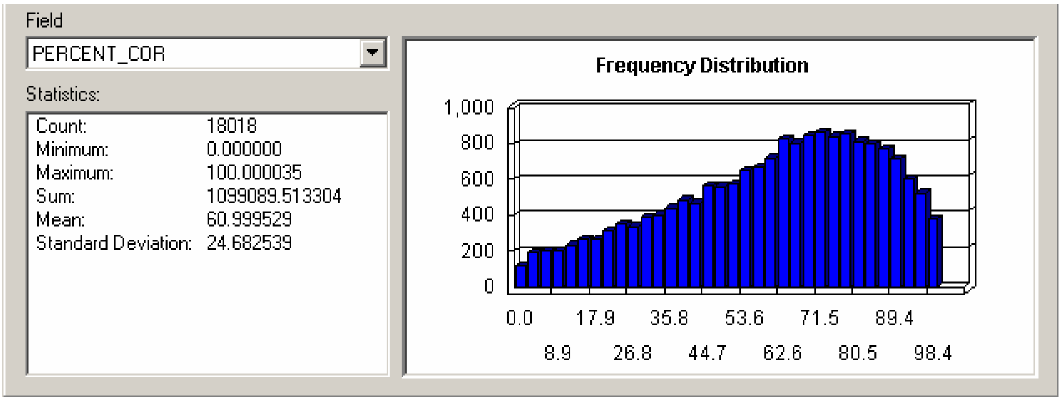
Field
PERCENT_COR
Statistics:
Count: 18018
Minimum: 0.000000
Maximum: 100.000035
Sum: 1099089.513304
Mean: 60.999529
Standard Deviation: 24.682539

Figure 3. Distribution of percent correct values for all catchments from the 100-cell stream-formation threshold for the hilly mountainous subbasin (WV). Mode of distribution is 71.

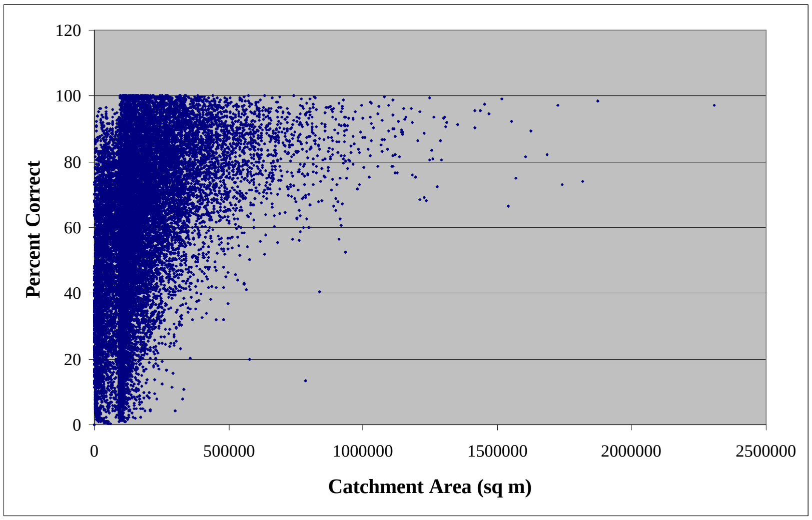

Figure 4. Distribution of percent correct values compared to catchment size for the 100-cell stream-formation threshold in the hilly mountainous subbasin (WV).

The average mean CAC ranges from 0.34 to 0.51, with better correspondence in the hilly and mountainous subbasins (figure 5). Thus, in the flat subbasins, a TPD catchment generally should overlay about one-half of the area that it should define, that is, the “correct” catchment area. The incorrect part of the TPD catchment, which is about one-half, should be about 7 percent smaller than the area omitted from the “correct” area. In hilly or mountainous subbasins, a TPD catchment generally should overlay nearly two-thirds of the “correct” area, and the incorrect part (less than one-third of the TPD catchment) should be about 7 percent smaller than the area omitted from the “correct” area.

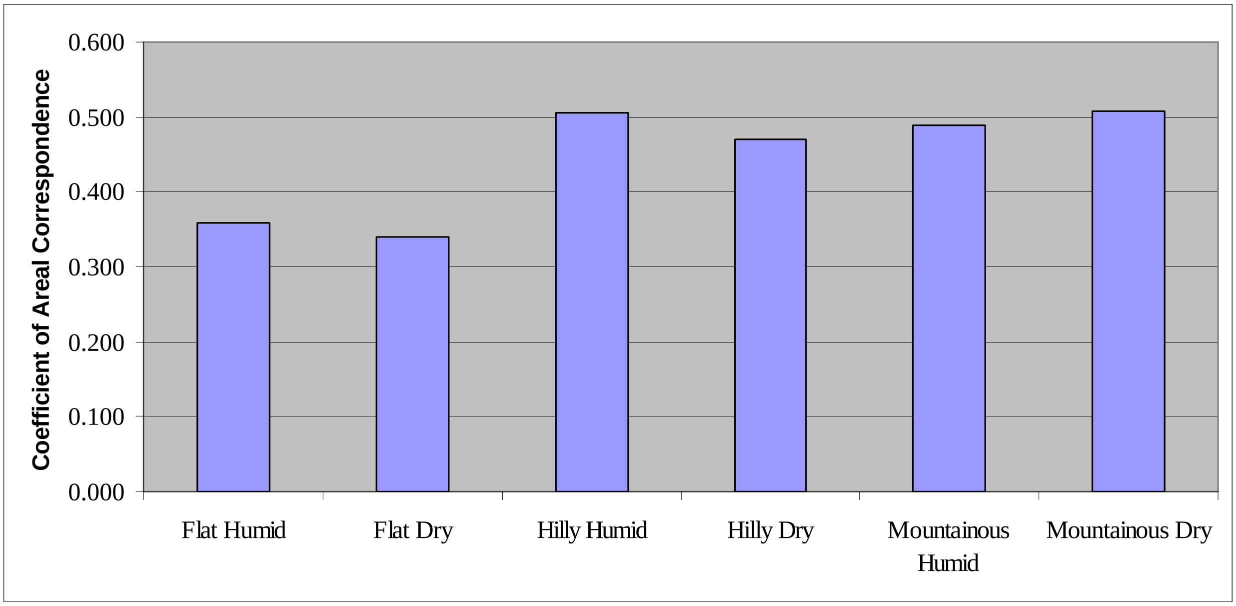

Figure 5. Average of the mean coefficient of areal correspondence (CAC) for each formation threshold is shown for each subbasin.

3.2 Within-subbasin comparisons

An evaluation of the relations between drainage density and CAC shows that CAC has a minor relation with drainage density that affects CAC by about 1/100th for each kilometer per square kilometer. This relation is negative for all subbasins except the flat dry subbasin in Utah, where the relation is positive. However, the Utah subbasin is classified as flat and dry because the majority of the subbasin is composed of dry lake, but the remainder consists of surrounding high plateaus and associated steep slopes that drain into the lake. Restricting the catchment analysis for the Utah subbasin to the fairly homogeneous dry lake region could reverse the drainage density to CAC relation for that subbasin.

Upon reviewing the spatial distribution of catchment percentages, it appears that larger errors occur in flood plain sections of subbasins; therefore, catchments were separated into headwater and non-headwater catchments based on whether or not they contained a dangling node of a stream line. Mean percentages were recomputed for headwater and non-headwater catchments. Using this approach on each stream-formation threshold, about one-half of the catchments were classified as headwater catchments, and about one-half were classified as non-headwater catchments. Results of this analysis indicate that TPD headwater catchments exhibit slightly better correspondence to ED headwater catchments than the correspondence between non-headwater catchments in 5 of 6 catchments (figure 6). Only the flat dry Utah subbasin does not show a difference in correspondence between headwater and non-headwater catchments. The improvement of average correspondence of headwater catchments over non-headwater catchments ranges from 0.2 percent to 17 percent, with greatest improvement in the flat humid subbasin.

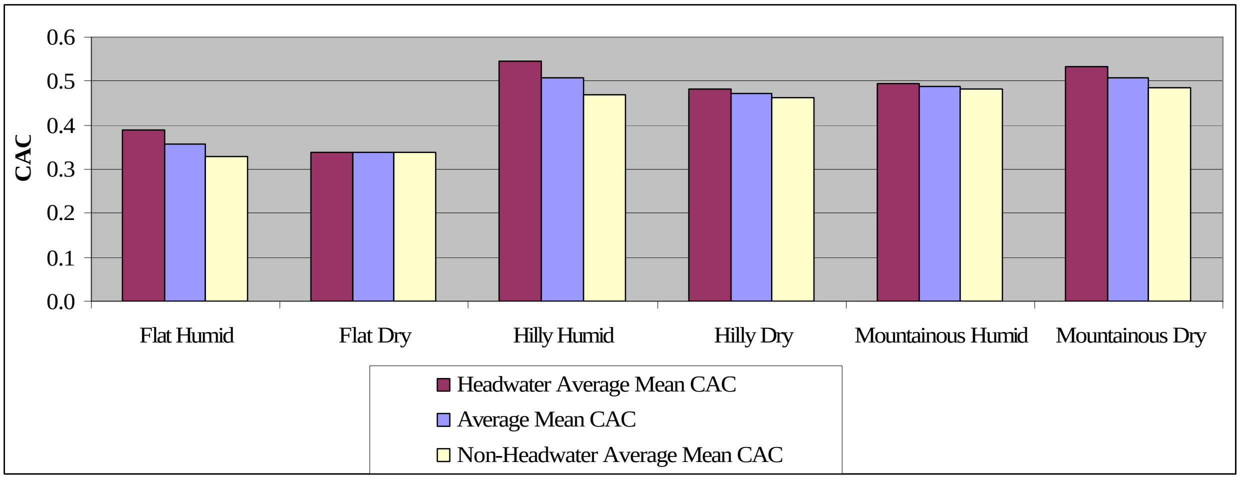

Figure 6. Headwater average, average, and non-headwater average of the mean coefficient of areal correspondence (CAC) for each formation threshold is shown for each subbasin.

4. Conclusions

This paper describes and assesses an automated processing approach that uses Thiessen polygons for estimating catchments for each segment of a drainage network. Thiessen-polygon-derived catchments were compared to catchments derived from a more accurate approach that uses a DEM. Comparisons were completed for six NHD subbasins that fall in six different climatic and topographic regimes.

Results indicate that the TPD catchment process is less likely to fail because of hardware or software limitations and about 5 to 10 times faster than the ED catchment process. Aside from these advantages, the TPD catchment process can be applied directly to any vector network--such as that furnished by the NHD—without integrating the vector network with a DEM, which is an added logistical burden of the ED catchment process.

Generally, the fractional part that TPD catchments overlay associated ED catchments is about one half for subbasins in flat terrain and about two thirds for hilly or mountainous subbasins. Furthermore, headwater TPD catchments exhibit better areal correspondence (up to 17 percent) with ED catchments than the areal correspondence between non-headwater TPD catchments and associated ED catchments.

Although this paper quantitatively assesses how well TPD catchments overlay ED catchments, results are depicted through averages. In each subbasin test, a number of TPD catchments have low or no overlap with ED catchments. What impact do these large errors have on applications that may use TPD catchments? Further research is necessary to assess the effects of using TPD catchments for UDA computations and subsequent applications, such as the network pruning process in our generalization approach, or for predicting flow volumes through regression equations in the National Flood Frequency Program (USGS, 2002).

References

Chaudry, O. and Mackaness, W.A., 2006, Modeling geographic phenomena at multiple levels of detail, AutoCarto 2006: June 26-28, 2006, Vancouver, WA.

Garbrecht, J and Martz, L., 2005, TOPAZ Topographic Parameterization Software, Accessed Jan. 26, 2007 at URL: http://www.ars.usda.gov/Research/docs.htm?docid=7834

Mackaness, W.A., 2006, Automated cartography in a bush of ghosts, Cartography and Geographic Information Systems 33(4):245-256.

Maidment, D. R., 2002, Arc Hydro: GIS for water resources: ESRI Press. Redlands, CA.

McMaster, R. and Shea, K., 1992, Generalization in Digital Cartography: Association of American Geographers, Washington, D.C.

National Atlas of the U.S., 2006, United States average annual precipitation 1961-1990, Accessed Jan 30, 2007 as URL: http://www.nationalatlas.gov/mld/prism0p.html.

Stanislawski, L.V., Finn, M., Starbuck, M., Usery, E.L., and Turley, P., 2006, Estimation of accumulated upstream drainage values in braided streams using augmented directed graphs, AutoCarto 2006: June 26-28, 2006. Vancouver, WA.

Stoter, J.E., 2005, Generalization within NMA's in the 21st century, XXII International Cartographic Conference: Coruna, Spain, The International Cartographic Association.

Taylor, P.J., 1977, Quantitative methods in geography: an introduction to spatial analysis, Chapter 5: Areal association, Houghton Mifflin, Boston, 396pp.

Thiessen, A.H. and Alter, J.C., 1911, Precipitation averages for large areas, Monthly weather review 39:1082-1084.

U.S. Environmental Protection Agency and U.S. Geological Survey, 2006, NHDPlus User Guide: June 1, 2006. 97pp. Accessed Jan. 29, 2007 at URL: http://www.horizon-systems.com/NHDPlus/data/NHDPLUS_UserGuide.pdf.

U. S. Geological Survey, 2006a, USGS national elevation dataset, Accessed Jan 30, 2007 at URL: http://ned.usgs.gov/.

_____2006b, USGS NHD geodatabase, Accessed Jan 30, 2007, at URL: http://nhdgeo.usgs.gov/viewer.htm.

U.S. Geological Survey, 2004, Physiographic divisions of conterminous U.S., Accessed June 14, 2005, at URL: http://aa179.cr.usgs.gov/metadata/wrdmeta/physio.htm.

U.S. Geological Survey, 2002, The national flood frequency program, version 3: a computer program for estimating magnitude of flood for ungaged sites, Accessed Jan. 26, 2007 at URL: http://pubs.usgs.gov/wri/wri024168/#pdf.

U.S. Geological Survey, 2000, The national hydrography dataset: concepts and contents, Accessed May 5, 2006, at URL: http://nhd.usgs.gov/chapter1/index.html.

Verdin, K. L., 1997, A system for topologically coding global drainage basins and stream networks, 1997 ESRI international GIS user conference proceedings, URL: http://gis.esri.com/library/userconf/proc97/proc97/to350/pap311/p311.htm,
Accessed June 29, 2005 at URL: http://edcdaac.usgs.gov/gtopo30/hydro/P311.asp

Yan H., Li Z., and Ai T., 2006, System for automatic generalization of topographic maps, Chinese Geographical Science 16(2):165-170.