Generalization for *The National Map* with emphasis on the National Hydrography Dataset

Lawrence Stanislawski, Science Applications International Corporation (SAIC)
Michael Starbuck, U.S. Geological Survey, Mid-Continent Mapping Center
Michael Finn, U.S. Geological Survey, Mid-Continent Mapping Center
E. Lynn Usery, U.S. Geological Survey, Mid-Continent Mapping Center

ABSTRACT

The National Map (TNM), being developed by the United States Geological Survey (USGS), is created from many data sources and Web Map Services (WMS) owned by many organizations. *TNM* data sources are electronically stored at various scales and resolutions. Consequently, when data are extracted from several WMS and horizontally integrated in geospatial data applications, the differences between data sources become apparent. The ability to render one or more scales of *TNM* data into an appropriate, or functionally equivalent, representation at a user-specified scale could substantially enhance analysis capabilities of these data. This paper describes ongoing research by the USGS into automated generalization approaches that address this problem, focusing on the National Hydrography Dataset (NHD) layer of *TNM*. The NHD provides vector representations of surface water features depicted at medium, high, and local resolutions, with only medium resolution currently available nationwide.

INTRODUCTION

The National Hydrography Dataset (NHD) is a vector data layer of *The National Map* representing the surface waters of the United States. The NHD includes a set of surface water reaches delineated on the vector data. Each reach consists of a significant segment of surface water having similar hydrologic characteristics, such as a stretch of river between two confluences, a lake, or a pond (USGS, 2000). A unique address, called a reach code, is assigned to each reach, which enables linking of ancillary data to specific features and locations on the NHD.

The NHD is divided and distributed at watershed basin and subbasin boundaries to simplify environmental analyses that are affected by surface drainage. It is stored in an ArcGIS geographic database (geodatabase) model at three levels of detail: medium (1:100,000-scale source), high (1:24,000-scale source), and local (1:12,000 or larger source) resolutions. Currently, only the medium resolution layer is complete for the conterminous U.S, with the high resolution about three quarters complete. Although the available resolutions are suitable for many applications, various uses of the NHD may require different levels of detail than what is available in the geodatabase. While only additional data gathering can furnish more detailed data, generalization of existing data can provide less detailed NHD data that is suitable for large-area regional or national studies.

The purpose of this project is to develop a generalization strategy that can be implemented on subsets of the NHD. The strategy should produce a dataset in the NHD model format that maintains: feature definitions, reach delineations, feature relationships,

and flow connections between remaining generalized features--so that the extracted data will continue to function with pre-existing NHD applications but with less detail and subsequently with faster processing speed. Eventually, we hope to develop software tools that generalize NHD data to a user-specified level of detail, or to a level that is defined by the scale of a user's display. Furthermore, the generalization process should occur in real-time rather than storing several generalized layers at pre-defined scales. Subsequent to this development effort, a data user could download equally functional, generalized NHD data—a significant enhancement to the analysis capabilities of the NHD. This paper summarizes our progress on the development of a generalization strategy for the NHD.

BACKGROUND

Generalization of cartographic data has been the subject of research since early in the twentieth century. Nearly all generalization processes include the initial step of selecting the objects and attributes from an initial database that are to be represented in the generalized dataset (McMaster and Shea, 1992). Subsequent to the selection process, one or more generalization operations—such as line simplification, aggregation, amalgamation, or merging—may be implemented on the remaining features (ibid.).

In this paper we refer to the object selection process as feature pruning. As of the date of this writing, the basic strategy for our NHD generalization approach entails feature pruning, line simplification, and feature simplification. Feature pruning concerns the removal of features that are too small for the desired generalized scale, which is also referred to as spatial refinement by McMaster and Shea (1992). Line simplification is the removal of vertices on the lines that comprise post-pruned features. And, feature simplification consists of converting features to a more generalized basic feature type, which entails collapsing area features to either lines or points or merging features.

Thus far, the majority of our development efforts have focused on the process of pruning or refining drainage network features of the NHD. A drainage network consists of flow-oriented digital lines that represent the connectivity and flow between water features on the ground. Network features types are stream/river, canal/ditch, pipeline, connector, and artificial path. A connector represents a path where surface flow is known to exist but was not included in the source material, and an artificial path represents a flow path through an areal water feature that is connected to the drainage network. Areal feature types that may include artificial paths are: area of complex channel, canal/ditch, estuary, ice mass, lake/pond, playa, reservoir, stream/river, swamp/marsh, and wash.

The first step in identifying a pruning strategy for a drainage network is to determine a goal for data pruning. To do this, some quantities that describe geometric characteristics of a stream drainage network must be defined and estimated. These quantities may be compared before and after pruning to assess the adequacy of the pruning process.

Geometric characteristics of an NHD flowline network that are being evaluated include the following: catchment area, upstream arbolate sum, drainage proportion, upstream drainage area, upstream drainage density, basin drainage density, number of confluence-to-confluence segments, stream order, bifurcation ratio, and basin bifurcation

ratio. These terms are defined as follows:

Catchment area – is an area associated with a segment of a drainage network. Surface runoff in this area flows into the associated network segment.

Upstream arbolate sum – can be computed for at any point on a drainage network. It is the sum of the lengths of all upstream network features that flow into a point of a network.

Drainage proportion – can be computed at any point on a drainage network. It is the proportion of the arbolate sum at that point to the total length of the drainage network.

Upstream drainage area – can be computed for any point on a drainage network. It is the sum of all upstream catchment areas up to a point in the network.

Upstream drainage density – can be computed for any point on a drainage network. It is the ratio of the arbolate sum to the upstream drainage area at a point in the network (Pidwirny 2004).

Basin drainage density – the ratio of length of all network features to the sum of all catchment areas in the network. The sum of all catchment areas in the network may also be referred to as the basin area.

Number of confluence-to-confluence segments – the number of segments between network feature confluences (intersections) that exist in the network.

Stream order – A method of numbering network segments within a drainage system, in which the smallest unbranched mapped tributary is called first order. Segments where two or more first order tributaries intersect becomes a second order segment and so on (Horton 1945).

Bifurcation ratio – the ratio of the number of segments of any order to the number of segments of the next highest order (Pidwirny 2004).

Basin bifurcation ratio – the average of the bifurcation ratios of each stream order in the drainage network (Pidwirny 2004).

METHODS

Computation of NHD Flowline Network Morphology

The subbasin is the smallest unit of NHD data that can be downloaded from the geodatabase. Several Arc Macro Language (AML) and Python programs were developed to process and compute subbasin network characteristics from available subregion NHD geodatabase data. Geometric characteristics for a number of subbasins at all available resolutions have been computed and summarized. Small clusters of subbasins that are located in the northeast, midwest, and southwest United States have been summarized. A rapid method that uses Thiessen polygons to the estimate catchment area for each network drainage feature was developed for these computations.

Flowline Network Pruning

In parallel, a pruning strategy that extracts the most prominent network features has been under development. Identifying an adequate pruning strategy for NHD generalization has been somewhat elusive. The most logical method for pruning is to maintain the features that hold the most surface water or transport the most surface water flow. However, such information is not currently associated with the NHD. Therefore,

the option presented in this research consists of eliminating network features that have an upstream drainage area that is less than a threshold value. This method ensures that network features that drain larger areas are more likely to be maintained in the generalized data than features that drain smaller areas. This approach follows the same logic as the Pfafstetter system for topologically coding river basins and networks (Verdin 1997). Furthermore, using upstream drainage area for pruning can minimize the effect of disparate data collection techniques, which are apparent in the NHD.

Appropriate threshold values that achieve a desired output drainage density can be iteratively determined during the pruning process. However, such a pruning strategy requires that the user understand and select an output drainage density, which may be an acceptable method for many NHD users. But what if a person does not have an output drainage density in mind? Instead, they would like a drainage density that is appropriate for a particular scale.

The difficulty with pruning is determining a drainage density that is appropriate for a desired scale. If we assume that the NHD data is consistent in its representation of drainage, then we could use it to estimate appropriate drainage densities for the scales of available NHD data, and subsequently interpolate density values for other scales. However, a visual inspection of the NHD data shows patches of drainage density differences between adjacent quadrangles even within the smallest watershed divisions. This anomaly is caused by the range of surface hydrographic conditions (wet and dry years) and the various map compilation standards used for topographic mapping by the USGS over the years (USGS, 1955). Figure 1 illustrates an example of this condition.

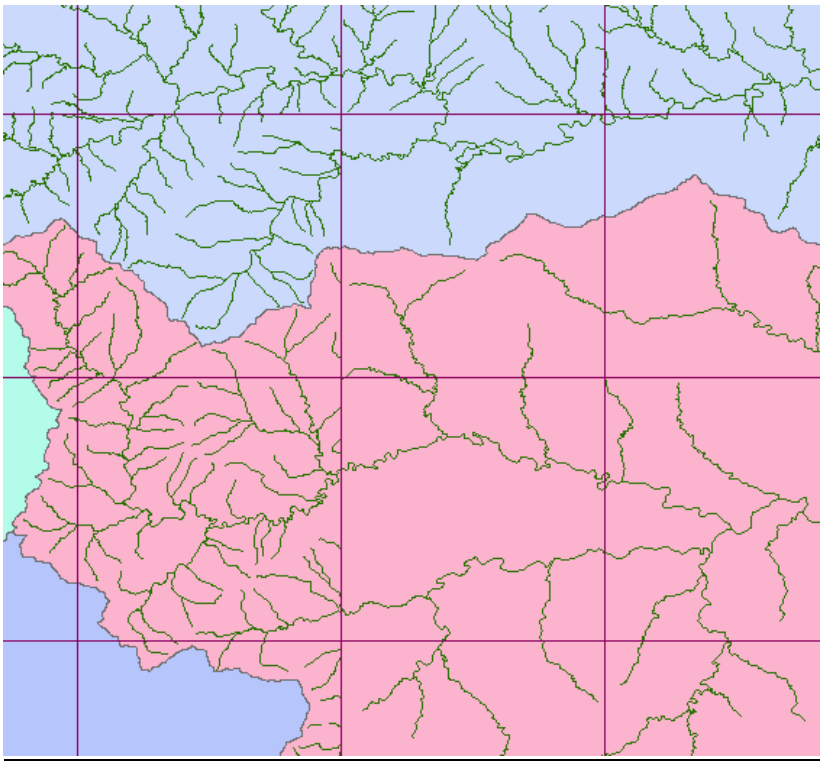

Figure 1. Example of variation in density of mapped drainage features along 7.5-minute quadrangle boundaries for medium resolution NHD. Background colors delineate subbasins.

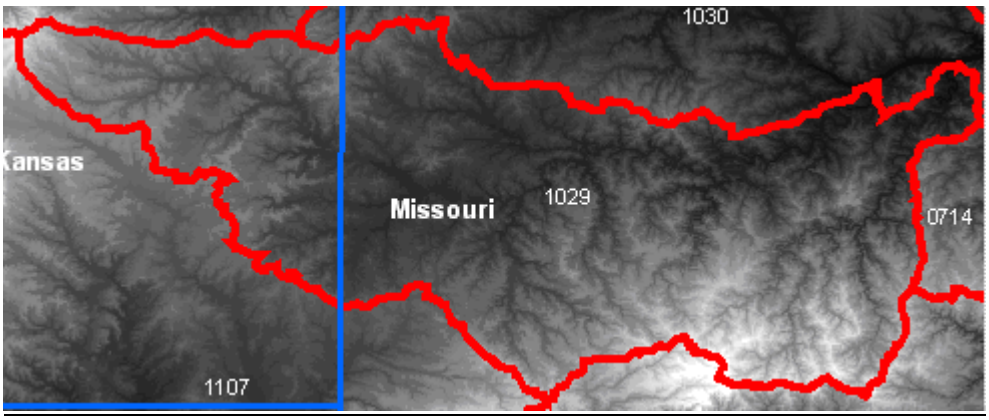

Figure 2. Gasconade-Osage NHD Subregion (1029) covers parts of Missouri and Kansas.

Likewise, disparate drainage density disparity apparent in 1:250,000-scale USGS topographic maps spurred NOAA into supplementing the 1:250,000-scale hydrographic features when compiling NOAA charts. Subsequently, the U.S. Environmental Protection Agency (EPA) developed river Reach File version RF1 from photo-reduced 1:500,000-scale NOAA charts (Horn et al. 1994). River Reach Files are the precursors to the NHD.

To overcome the drainage density problem, approximate drainage densities from four scales of hydrographic vector data were determined for NHD Gasconade-Osage subregion (1029, Figure 2), which falls in the Interior Plains and Interior Highlands physiographic divisions (USGS 2004), and the NHD Colorado Headwaters and Gunnison subregions (1401 and 1402, Figure 3), which fall in Intermontane Plateaus and Rocky Mountain System physiographic divisions (USGS 2004). For these subregions, the scales and respective data sources for which drainage densities were estimated are: 1:24,000, NHD; 1:100,000, NHD; 1:500,000, RF1; and 1:2,000,000, DLG data (USGS 2003). Only the NHDFlowline feature class, which is comprised of the drainage network features, was used in the computation of drainage density for the NHD datasets.

In addition, a model for the relationship between elevation-derived drainage density and minimum drainage area for stream formation was derived for the two study areas: one model for the Gasconade-Osage subregion, and one model for the Colorado Headwaters and Gunnison subregions. Elevation data was extracted from the National Elevation Dataset (NED, USGS 2005), and resampled to a 30 meter cell size. Approximate stream channels were derived from the elevation data through commonly used methods that include the following steps: pre-processing of elevation model to fill sinks and create an approximate depressionless surface, compute flow direction from pre-processed elevation data, compute flow accumulation from flow direction, and then vectorize flow direction (Sanchez 2002, Maidment 1996, Smith and Maidment 1995). The minimum drainage area for stream formation on the flow accumulation grid was varied between 150 and 100,000 cells. Resulting drainage densities were computed, and power function models predicting drainage density from minimum drainage area for stream formation were estimated through regression in Microsoft Excel.

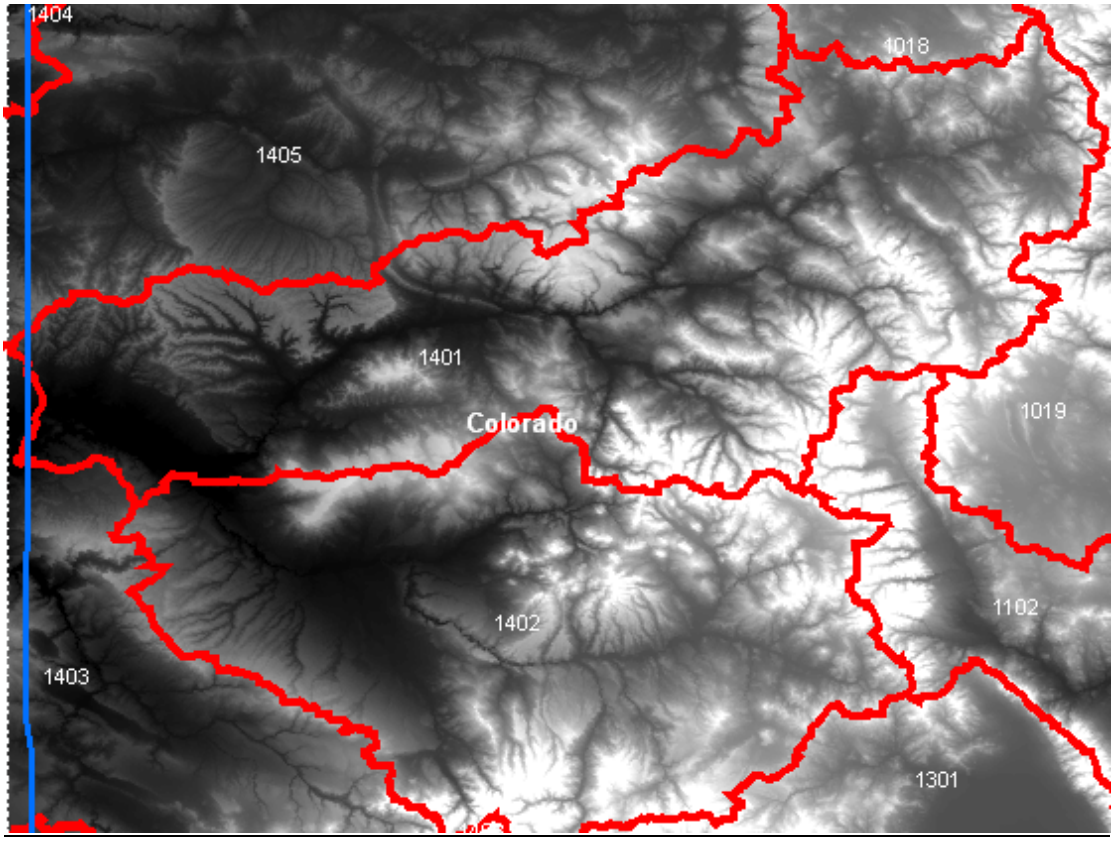

Figure 3. Colorado Headwaters and Gunnison subregions (1401 and 1402) fall almost entirely in Colorado.

Upon evaluating the relationship between elevation-derived drainage density and minimum drainage density for stream formation, it was deduced that an approximate linear relationship should exist between the square root of map scale and drainage density derived from mapped drainage network features. Therefore, a linear model predicting drainage density of mapped network features from the square root of map scale was estimated for each study area. The linear models were estimated through regression in Excel using the four drainage density values resulting from the previously described scales of hydrographic vector data.

.

RESULTS

NHD Flowline Network Morphology

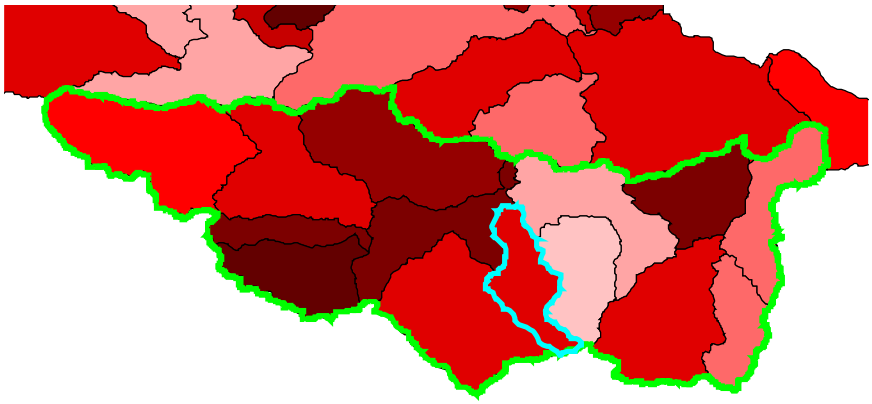

Figure 4. Relative magnitude of drainage density differences between the high resolution and medium resolution NHD around the Gasconade-Osage subregion. Darker red indicates a higher difference in density. Gasconade-Osage subregion is outlined in green, and the Pomme De Terre subbasin is outlined in cyan.

Sample Dataset

Pomme De Terre (PDT) subbasin (NHD# 10290107) was selected as a sample dataset for network generalization and to exemplify quantities determined to describe flowline network morphology. Pomme De Terre subbasin falls within the Gasconade-Osage subregion. Figure 4 is color coded to show the relative magnitude of drainage density differences between the high resolution (1:24,000-scale) layer and the medium resolution (1:100,000-scale) layer of the region, with darker red indicating a higher difference in density. As illustrated, PDT subbasin has a medium level of network drainage density difference between the two resolutions. The difference between the high and medium resolution drainage densities for the PDT is about 0.6421 km/sq km.

High resolution

A summary of the network morphology quantities for the high resolution PDT subbasin follows:

Reach Statistics

Number of network reaches: 3367

Average length of network reaches: 1.02 km

Catchment areas range from 0.001 sq km to 7.903 sq km.

Arbolate sum values range from 0.018 km to 3424 km.

Proportion of total drainage flowing through main outflow reach: 0.99

Upstream drainage areas range from 0.035 sq km to 2187.3 sq km
Upstream drainage densities range from 0.0938 km/sq km to 7.213 km/sq km.
Upstream drainage density for main outflow reach: 1.565 km/sq km
Basin Statistics
Number of confluence-to-confluence segments: 3986
Length of network features: 3428.5 km
Area of subbasin: 2190.3 sq km
Basin drainage density: 1.566 km/sq km
Stream order on confluence-to-confluence segments range from 1 to 15.
Bifurcation ratios range from 1 to 4.18.
Basin bifurcation ratio is 2.25.

The spatial distribution of arbolate sum values is displayed in Figure 5. The high resolution subbasin has a few reaches, totaling about 4.5 km in length that are not connected to the main network. Therefore, the proportion of drainage flowing through the main outflow reach is not quite unity.

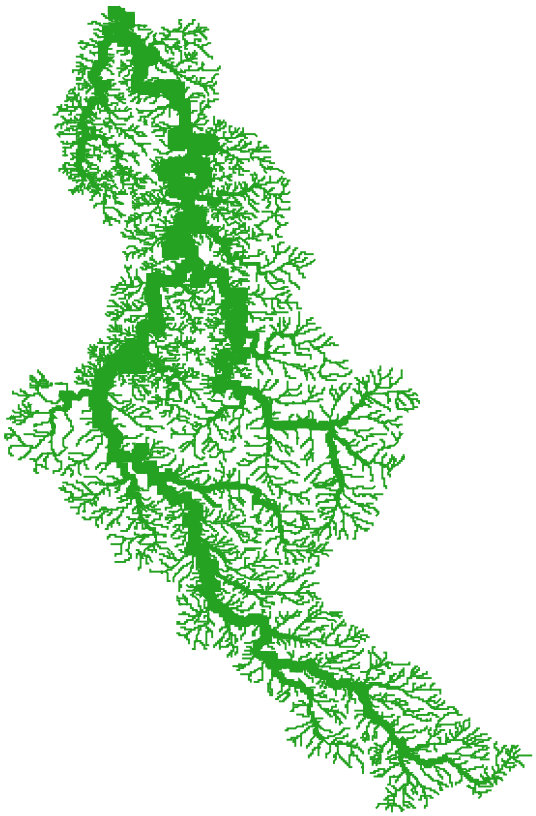

Figure 5. Graduated symbols of arbolate sum values by network reach for Pomme De Terre subbasin of the high resolution NHD.

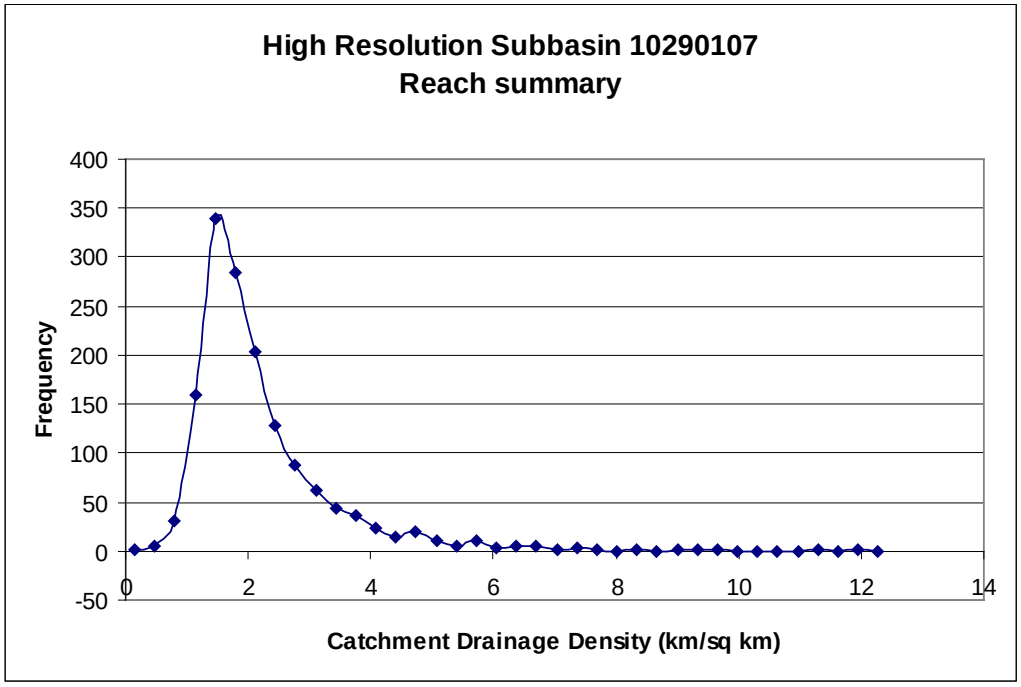

Figure 6. Drainage density distribution of reach catchments for Pomme De Terre subbasin of high resolution NHD.

The distribution of catchment drainage densities for high resolution reaches in the PDT subbasin are displayed in Figure 6. The maximum frequency occurs at about 1.46 km/sq km. Figure 7 displays the relationships of upstream drainage density and catchment area with arbolate sum. As portrayed in Figure 7, the majority of the reach catchment areas are less than the upstream drainage densities.

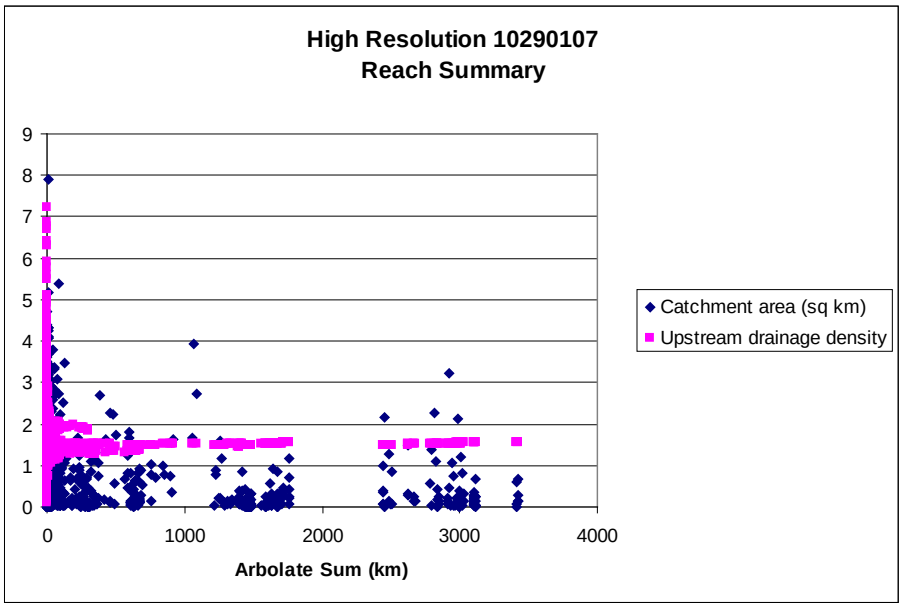

Figure 7. Catchment area and upstream drainage density relationships with arbolate sum for Pomme De Terre subbasin of the high resolution NHD.

Basin drainage density is about 1.57 km/sq km. Typical values range from about 1.5 to 6, with very small drainage densities suggesting a highly permeable landscape (Pidwirny 2004, Northcott 2002). However, describing relationships between the drainage density of mapped features and landscape permeability requires that the features are collected in a specific and consistent manner.

A number of divergent stream segments occur along main channels that converge further downstream, causing confluence-to-confluence stream order values to increase to 15. The stream ordering program was conditioned to increment stream order at the convergent confluence of divergent-to-convergent channel segments where one or both of the converging segments, which have the same order, receive additional flow from other tributaries. If no additional flow is received by either converging segment, then order at the convergence remains the same as the inflowing segments.

Theoretically, bifurcation ratios should range from 2 to about 6 (Northcott 2002). However, values of one are reported here because several divergent-to-convergent channels that increment stream order occur in series along the main channel, which generate the same number channels in consecutive stream orders. Furthermore, this situation generates several low bifurcation ratios among the higher stream order values, and consequently, the basin bifurcation ratio is relatively low (2.25) when compared to the medium resolution basin ratio (4.28).

Medium resolution

A summary of the network morphology quantities for the medium resolution PDT subbasin follows:

Reach Statistics

Number of network reaches: 1079
Average length of network reaches: 1.91 km
Catchment areas range from 0.031 sq km to 12.846 sq km.
Arbolate sum values range from 0.077 km to 2064 km.
Proportion of total drainage flowing through main outflow reach: 1.00
Upstream drainage areas range from 0.254 sq km to 2190.3 sq km
Upstream drainage densities range from 0.2164 km/sq km to 2.400 km/sq km.
Upstream drainage density for main outflow reach: 0.9421 km/sq km

Basin Statistics

Number of confluence-to-confluence segments: 1014
Length of network features: 2063.6 km
Area of subbasin: 2190.3 sq km
Basin drainage density: 0.9421 km/sq km
Stream order on confluence-to-confluence segments range from 1 to 6.
Bifurcation ratios range from 1 to 8.67.
Basin bifurcation ratio is 4.28.

The spatial distribution of arbolate sum values for the medium resolution PDT subbasin is displayed in Figure 8. All reaches in the medium resolution subbasin are connected to the main network. Therefore, the proportion of drainage flowing through the main outflow reach is unity.

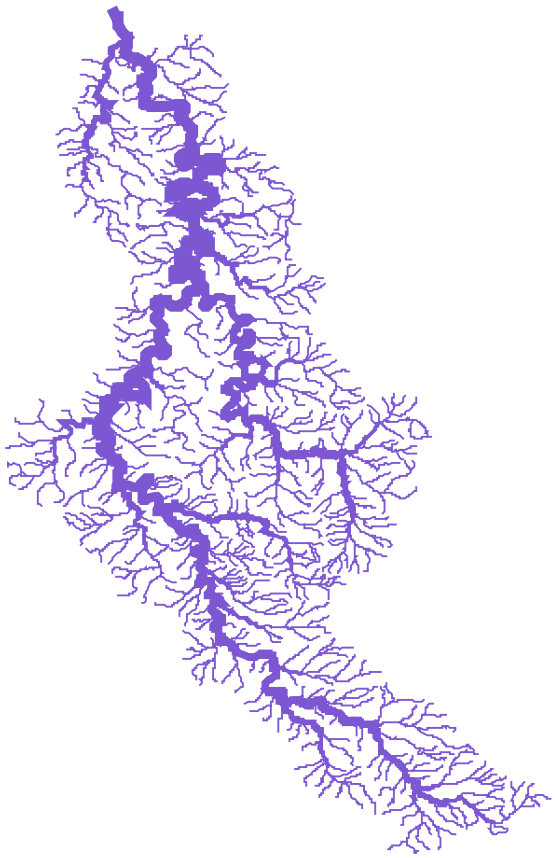

Figure 8. Graduated symbols of arbolate sum values by network reach for Pomme De Terre subbasin of the medium resolution NHD.

The distribution of catchment densities for medium resolution reaches in the PDT subbasin are displayed in Figure 9. The maximum frequency occurs at about 0.998 km/sq km. Figure 10 displays the relationships of upstream drainage density and catchment area with arbolate sum. The majority of the reach catchment areas are greater than the upstream drainage density values, which is reverse of what occurs for the high resolution data. Basin drainage density is about 0.942 km/sq km.

A number of divergent stream segments occur along main channels that converge further downstream. However, with few to no tributaries entering the divergent segments, the convergence of these segments does not cause an inflated maximum stream order value, as seen in the high resolution dataset. Thus, the basin bifurcation ratio of 4.28 is about twice as large as the basin ratio for the high resolution dataset.

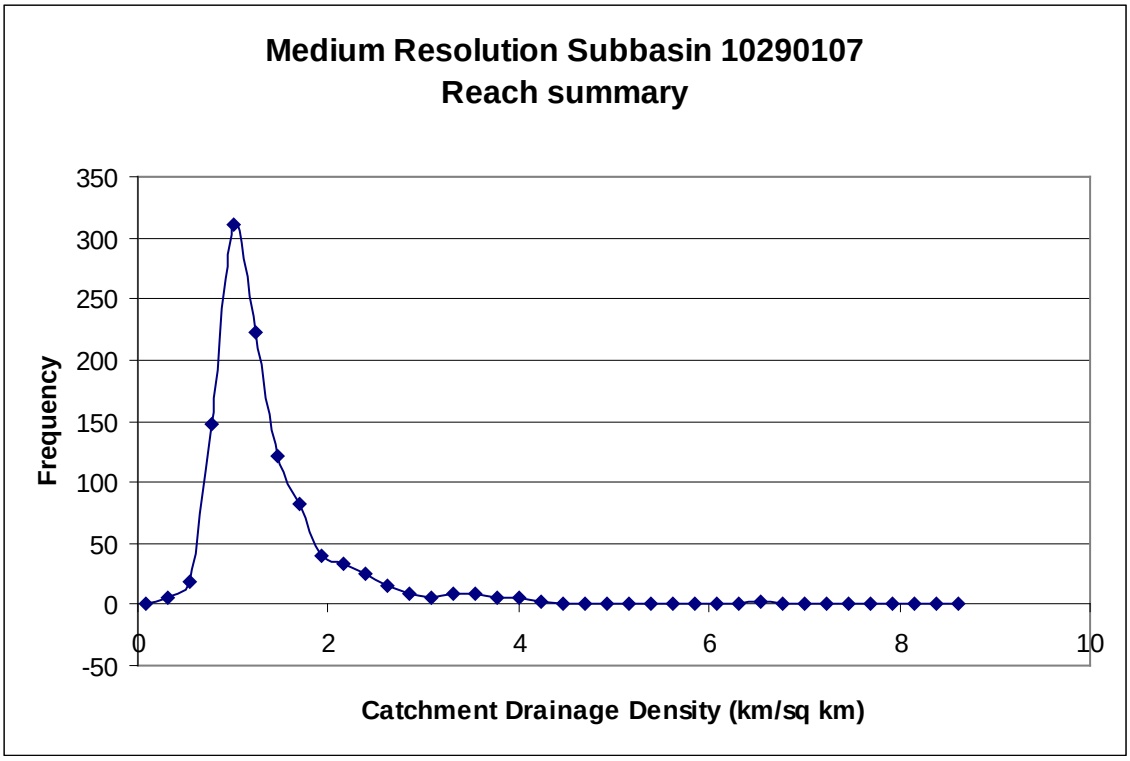

Figure 9. Drainage density distribution of reach catchments for Pomme De Terre subbasin of medium resolution NHD.

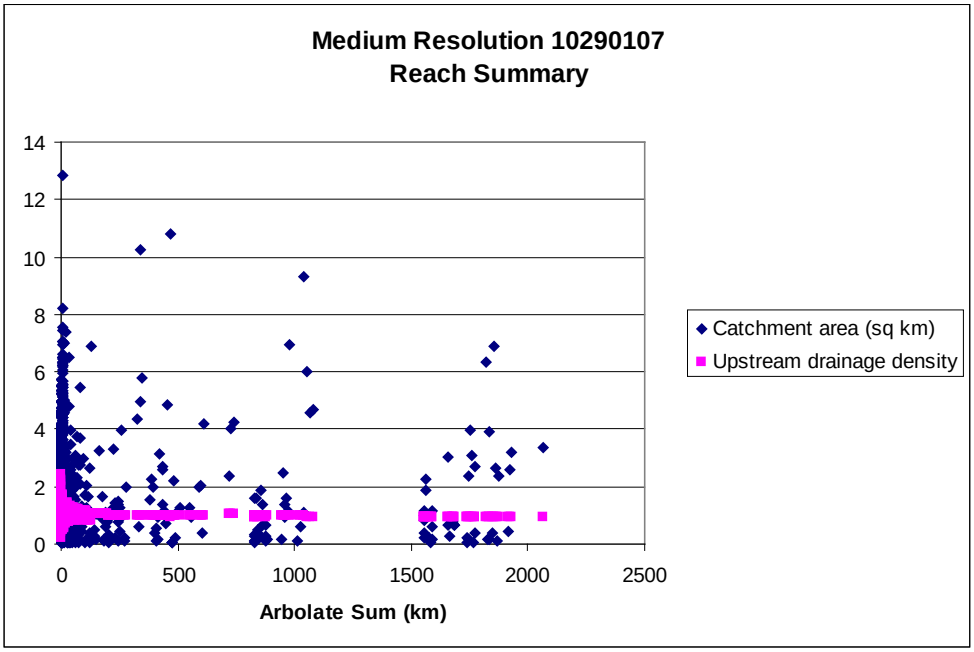

Figure 10. Catchment area and upstream drainage density relationships with arbolate sum for Pomme De Terre subbasin of the medium resolution NHD.

Flowline Network Pruning

Gasconade-Osage subregion (1029).

Elevation data. Drainage density values of synthetic streams that were derived by varying the minimum drainage area for 30-meter resolution NED data are displayed Figure 10. The minimum drainage areas for synthetic stream generation ranged from 0.135 sq km (150 cells) to 90 sq km (100,000 cells). The regression equation predicting drainage density from minimum drainage area has a high coefficient of determination (0.9988), suggesting that the model is reliable. The relationship suggests that drainage density is approximately inversely related to the square root of minimum drainage area, which also appears reliable (Figure 11, R^2 0.996).

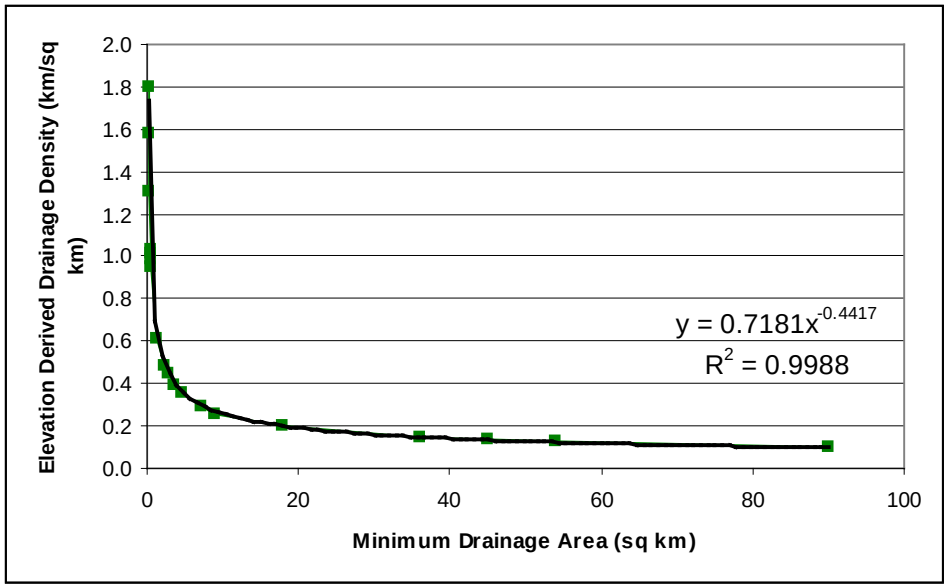

Figure 10. Relationship between minimum drainage area and elevation derived drainage density from 30-resolution elevation data of the Gasconade-Osage subregion.

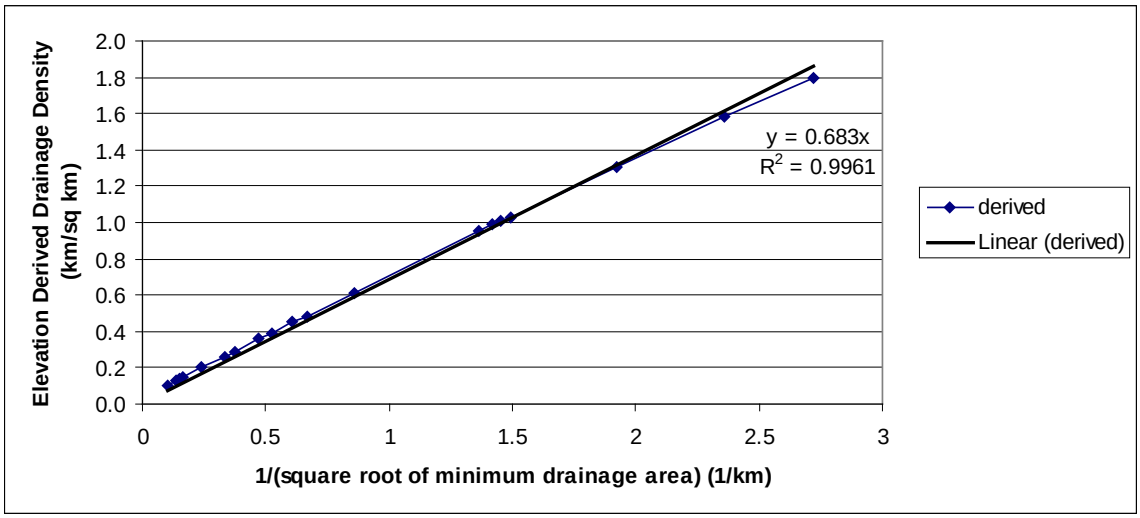

Figure 11. Linear relationship predicting elevation derived drainage density from the

inverse of the square root of minimum drainage area for Gasconade-Osage subregion.

Vector hydrographic data. Given the aforementioned relationship, we surmised that mapped hydrographic features should follow a similar relationship, with drainage density being directly related to the square root of map scale ratio. Drainage density values for the four resolutions of vector hydrographic data in the Gasconade-Osage subregion are the following: 1.656 km/sq km for 1:24,000; 0.925 km/sq km for 1:100,000; 0.17 km/sq km for 1: 500,000; and 0.119 km/sq km for 1:2,000,000. The relationship predicting vector data drainage density from the square root of the scale ratio for this subregion is displayed in Figure 12. The relationship should be acceptable to estimate a drainage density that can be used to prune high resolution network features for generalization.

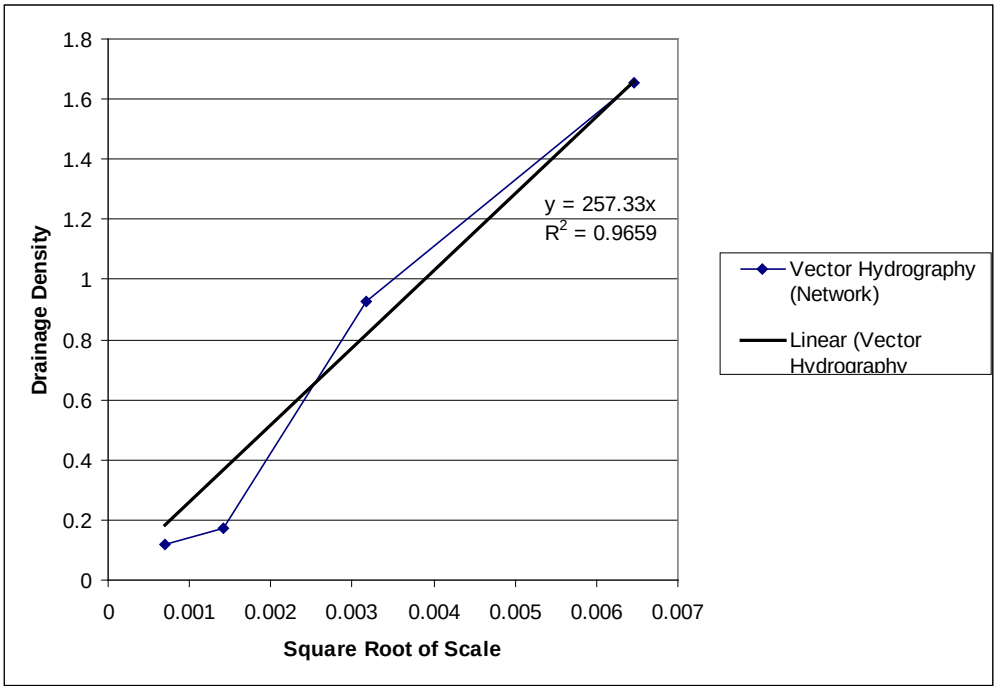

Figure 12. Relationship between drainage density and the square root of map scale ratio of source hydrographic features in the Gasconade-Osage subregion.

High resolution. Additional details about the high resolution NHD flowline network features are presented in this section. Fourteen subbasins exist in the Gasconade-Osage subregion. A total of 121,419 catchment areas were estimated for the high resolution subregion. Catchment areas for individual drains in the region ranged from 0.000045 sq km to 9.550 sq km. Estimation of catchment areas using Thiessen polygons around evenly spaced event points along the network features is a rapid and computational 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 (Figure 13), and are only used as approximate subdivisions of the drainage area for generalization purposes. Upstream drainage densities computed for each drain feature in this subregion range from 0.00336 km/sq km

to 33.137 km/sq km, when computed separately for each subbasin.

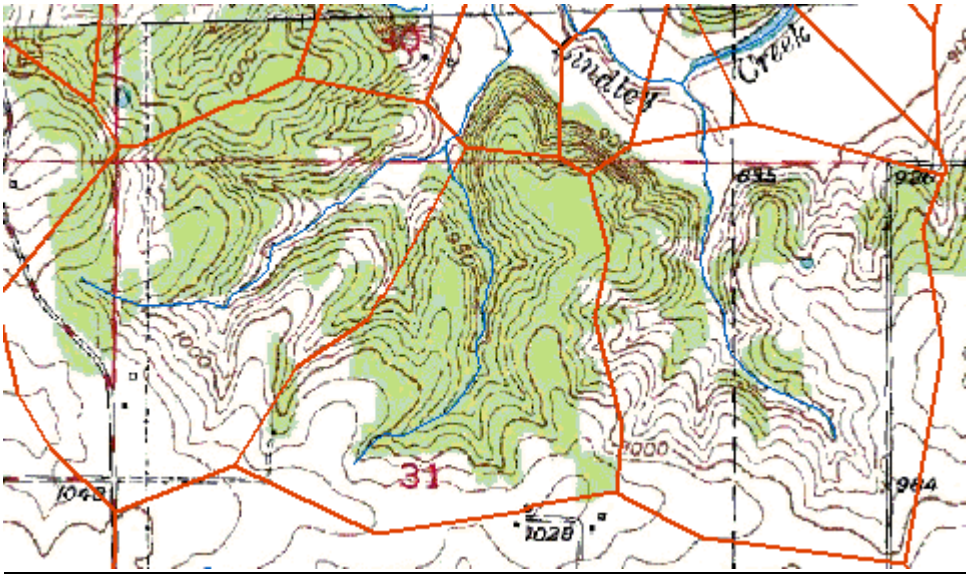

Figure 13. Catchment area estimates for 1:24,000-scale NHD drain features generated from Thiessen polygons around evenly distributed points on drainage features.

Medium resolution. A total of 27,624 catchment areas were estimated for the medium resolution Gasconade-Osage subregion. Catchment areas for individual drains in the regions ranged from 0.000470 sq km to 57.386 sq km. Upstream drainage densities computed for drain features in this region range from 0.0290 km/sq km to 27.113 km/sq km, when computed separately for each subbasin.

Colorado Headwaters and Gunnison subregions (1401 and 1402).

Elevation data. Drainage density values of synthetic streams that were derived by varying the minimum drainage area for 30-meter resolution NED data are displayed Figure 14. The minimum drainage areas for synthetic stream generation in these subregions ranged from 0.135 sq km (150 cells) to 90 sq km (100,000 cells). The regression equation predicting drainage density from minimum drainage area has a high coefficient of determination (0.9994), suggesting that the model is reliable. A highly reliable (R^2 = 0.996) linear equation predicting drainage density from the inverse of the square root of minimum drainage area is also found (Figure 15). The slope of this relationship is slightly larger than the slope for the same relationship in the Gasconade-Osage subregion, suggesting that a smaller density of elevation-derived streams is available in the Gasconade-Osage area with decreasing minimum drainage area than in the mountainous Colorado Headwaters and Gunnison subregions. Thus, differing terrain characteristics other than minimum drainage area between the two Gasconade and Colorado subregions may be influencing these relationships. Perhaps a greater amount of flatter (plateau) regions in the Gasconade-Osage region is affecting the start of flow accumulation. This inference is supported by a cursory understanding of hydrologic engineering whereby areas having greater slope more readily form stream channels than flatter areas.

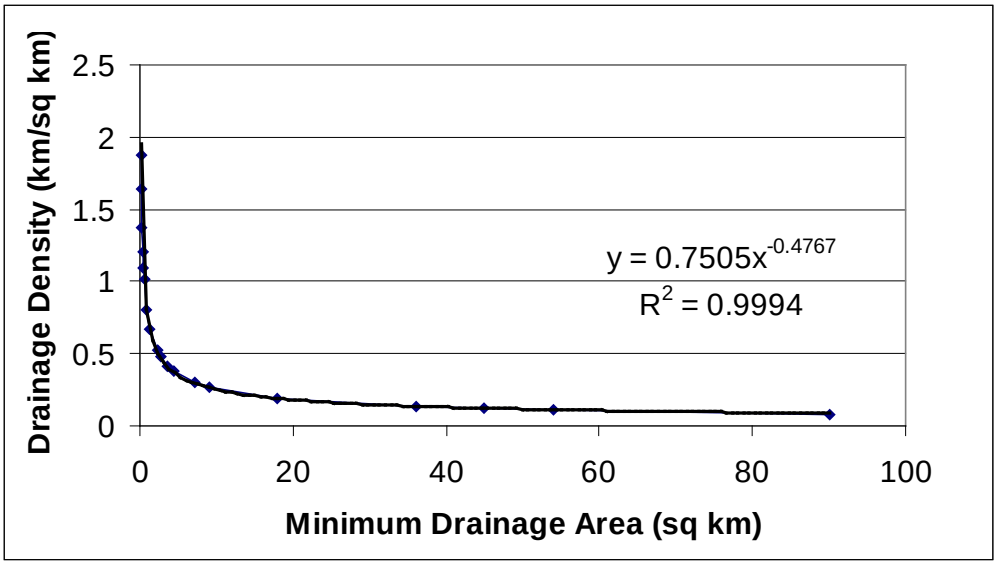

Figure 14. Relationship between minimum drainage area and elevation derived drainage density from 30-resolution elevation data of Colorado Headwaters and Gunnison subregions.

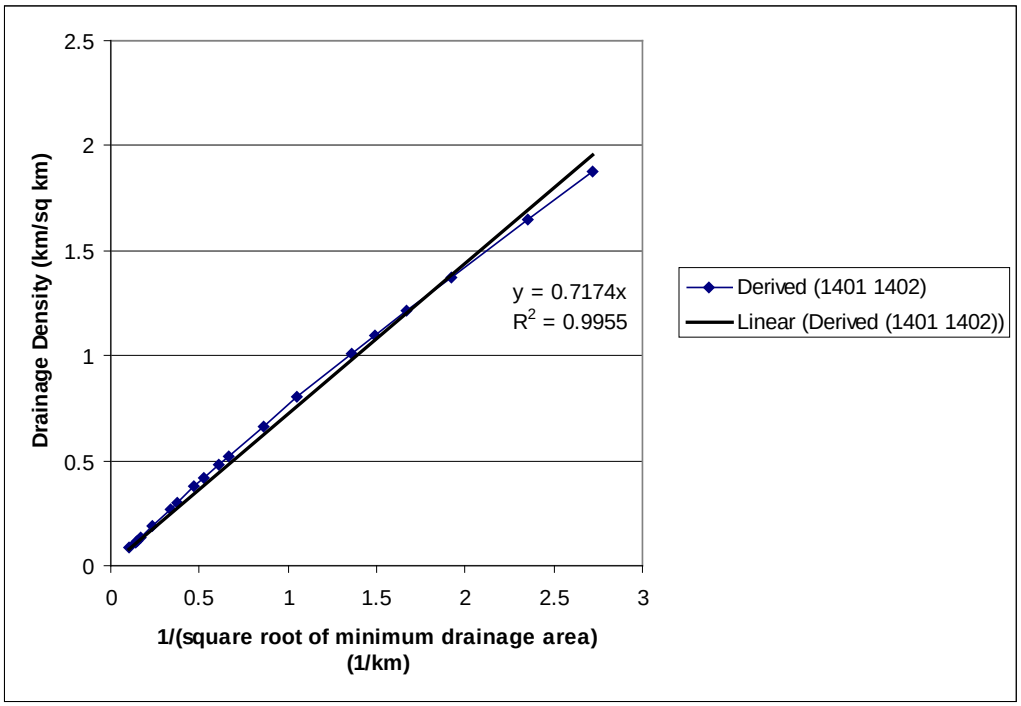

Figure 15. Linear relationship predicting elevation derived drainage density to the inverse of the square root of minimum drainage area for Colorado Headwaters and Gunnison subregions.

Vector hydrographic data. Drainage density values for the four resolutions of vector hydrographic data in the Colorado subregions are the following: 1.598 km/sq km for 1:24,000; 0.695 km/sq km for 1:100,000; 0.155 km/sq km for 1: 500,000; and 0.112

km/sq km for 1:2,000,000. The relationship predicting drainage density from the square root of the scale ratio for subregions 1401 and 1402 is displayed in Figure 16. As with the Gasconade subregion, the modeled relationship should be adequate to estimate a drainage density for pruning higher resolution network features during generalization.

The slope of the relationship predicting the density of mapped vector hydrographic features from map scale is greater for the Gasconade-Osage subregion model (Figure 12, y = 257.33x) than the Colorado subregions model (Figure 16, y = 236.5x), which is the reverse of what the elevation-derived drainage density relationships suggest (Figure 11; y = 0.683x and Figure15; y = 0.7174x). Thus, factors not related to terrain variations—such as, surface permeability, surface water conditions at time of data collection, or inconsistent data collection standards—may be influencing these vector relationships.

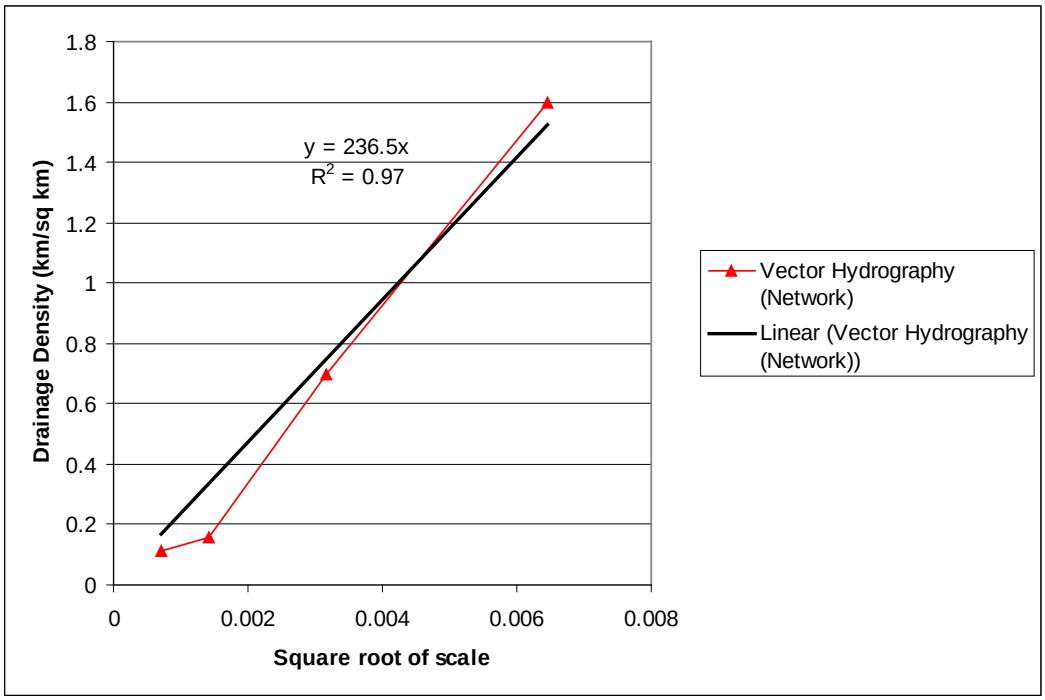

Figure 16. Relationship between drainage density and the square root of map scale ratio for source hydrographic network features in the Colorado Headwaters and Gunnison subregions.

High resolution. Subregions 1401 and 1402 include 12 subbasins. Catchment areas for 105,366 flowline features were determined for these two subregions. Catchment areas range from 0.000078 sq km to 19.087 sq km. Upstream drainage densities computed for each drain feature in the two subregions range from 0.000394 km/sq km to 40.214 km/sq km, when computed separately for each subbasin.

Medium resolution. A total of 13,037 catchment areas were estimated for the medium resolution Colorado Headwaters and Gunnison subregions. Catchment areas for individual drains in the regions ranged from 0.001690 sq km to 66.817 sq km. Upstream drainage densities computed for each drain feature in the two subregions range from

0.014105 km/sq km to 25.143 km/sq km, when computed separately for each subbasin.

Generalization results

A selection process to remove the less significant network features was tested on the PDT high resolution subbasin. The goal was to prune the high resolution dataset to resemble the medium resolution drainage network. An AML program was written that prunes the high resolution network to a desired basin drainage density. In this case, a basin drainage density of 0.942 km/sq km was the selected target. Since NHD addresses are defined by reach code, entire reaches are pruned from the network. Therefore, AML summarizes arbolate sum and upstream drainage area by reach to perform the pruning process.

Initially, a minimum arbolate sum value was used for pruning smaller tributaries. However, this method can eliminate significant streams in areas where drainage density is relatively low due to disparate data collection standards. Therefore, it was opted to prune using a minimum upstream drainage area that is estimated through the Thiessen polygons around drainage segments. With this method, pruning is performed by iteratively increasing a minimum upstream drainage area until the desired drainage density is achieved. However, upstream drainage area quickly rises to overwhelming sizes. Therefore, the square root of area values is used to reduce the magnitude of these values for automation purposes. Alternative numeric scaling techniques will need to be developed to implement this pruning strategy.

Pruning of the PDT high resolution to a minimum upstream drainage area of 2.158 sq km yielded a basin drainage density of 0.939 km/sq km, which is a bit lower than the requested value. Further work is necessary to develop more precision in the iterative pruning process.

A summary of the network morphology quantities for the generalized PDT subbasin follows:

Reach Statistics

Number of network reaches: 1648
Average length of network reaches: 1.25 km
Catchment areas range from 0.00135 sq km to 8.263 sq km.
Arbolate sum values range from 0.0169 km to 2056 km.
Proportion of total drainage flowing through main outflow reach: 1.00
Upstream drainage areas range from 0.0943 sq km to 2189.97 sq km
Upstream drainage densities range from 0.0261 km/sq km to 2.722 km/sq km.
Upstream drainage density for main outflow reach: 0.9387 km/sq km

Basin Statistics

Number of confluence-to-confluence segments: 1332
Length of network features: 2056.7 km
Area of subbasin: 2190.3 sq km
Basin drainage density: 0.9390 km/sq km
Stream order on confluence-to-confluence segments range from 1 to 14.
Bifurcation ratios range from 1 to 4.31.
Basin bifurcation ratio is 2.13.

Figure 17. Graduated symbols of arbolate sum values by network reach for Pomme De Terre subbasin of the high-resolution NHD network generalized to 0.942 km/sq km basin drainage density.

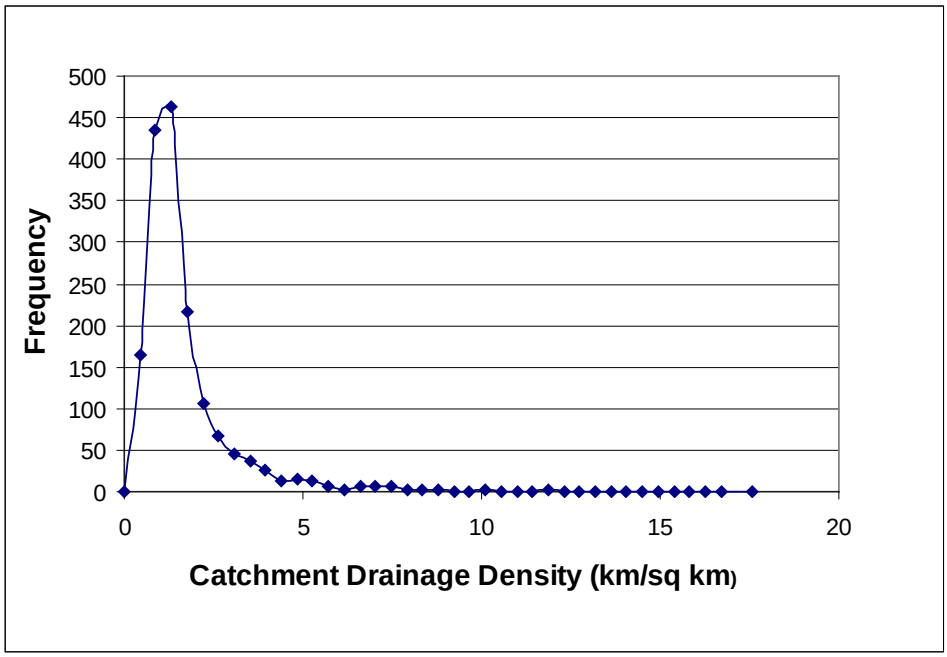

Figure 18. Drainage density distribution of reach catchments for Pomme De Terre subbasin of high-resolution NHD network generalized to 0.942 km/sq km basin drainage density.

The spatial distribution of arbolate sum values for the generalized data set is displayed in Figure 17. All reaches are connected to the main network making the proportion of drainage flowing through the main outflow reach to be one. A reach by reach comparison of the drainage proportion values between the generalized and ungeneralized high resolution PDT datasets may provide some insight on the effect of pruning by upstream drainage area.

The most obvious difference between the generalized dataset and the medium resolution dataset is that the average length of reaches is about 35 percent reduced in the generalized dataset, and a greater number of reaches exist in the high resolution data. Consequently, larger ranges in the reach drainage area and upstream density are evident in the high resolution data. Much of this effect can be attributed to the production process of high resolution NHD. During production of the high resolution NHD, medium resolution reaches are conflated, where possible, to the higher resolution vector data. Often, the more detailed data do not conform to the same drainage pattern represented in the medium resolution data, requiring that some reaches be redelineated to a greater number of smaller confluence-to-confluence reaches.

The distribution of catchment drainage densities for generalized high resolution reaches in the PDT subbasin are displayed in Figure 18. The maximum frequency occurs at about 1.32 km/sq km, which is about 0.3 km/sq km greater than and 0.14 km/sq km less than the maximum frequency for the medium resolution and high resolution datasets, respectively. More open space in the generalized data provides smaller catchment drainage densities than the ungeneralized data. Figure 19 displays the relationships of upstream drainage density and catchment area with arbolate sum. The majority of the reach catchment areas are greater than the upstream drainage densities particularly at the lower end of the distribution, which is similar to this distribution for the medium resolution data. Given that basin drainage density is a little less than 1.0 km/sq km, this result is expected.

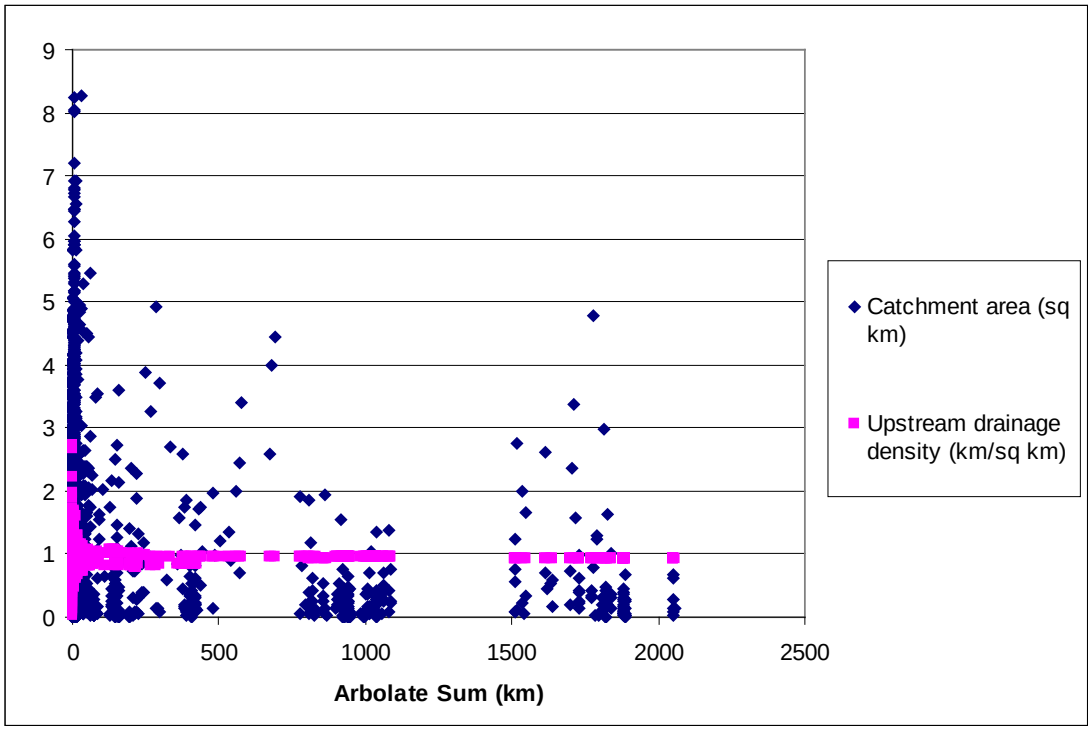

Figure 19. Catchment area and upstream drainage density relationships with arbolate sum for high-resolution Pomme De Terre subbasin generalized to 0.942 km/sq km basin drainage density.

Stream order computations are handled differently for divergent and reconvergent stream channels. A number of divergences and convergences along main stream segments exist in the ungeneralized high resolution dataset that are maintained in the generalized dataset. As seen in the high resolution data, the main stream segments in the generalized dataset that diverge and then reconverge downstream generate additional higher stream order segments through the applied algorithm. The basin bifurcation ratio of 2.13 on the generalized dataset is similar to the ratio for the ungeneralized high resolution dataset, which suggests that similar branching patterns are maintained among stream orders between the generalized and ungeneralized datasets..

Figure 20 shows a summary of stream length by stream order for the three resolutions of data reported for the Pomme De Terre subbasin. Generalization of the high resolution data via upstream drainage area produces nearly the same sum of stream lengths for stream orders 1 through 5 when compared to the medium resolution data. A more thorough analysis, shows that the sum of the lengths for corresponding stream orders differ by about 2 to 7 percent for stream orders 1 through 5, when comparing the medium resolution and the generalized datasets. These facts suggest that the data structure of the generalized dataset mimics that of the medium resolution data fairly well. The main difference in stream orders between the generalized high resolution and medium resolution datasets is caused by the different number of divergent and reconvergent main stream segments between the two resolutions, with the medium resolution having very few of these segments and subsequently no higher order streams are generated by the algorithm.

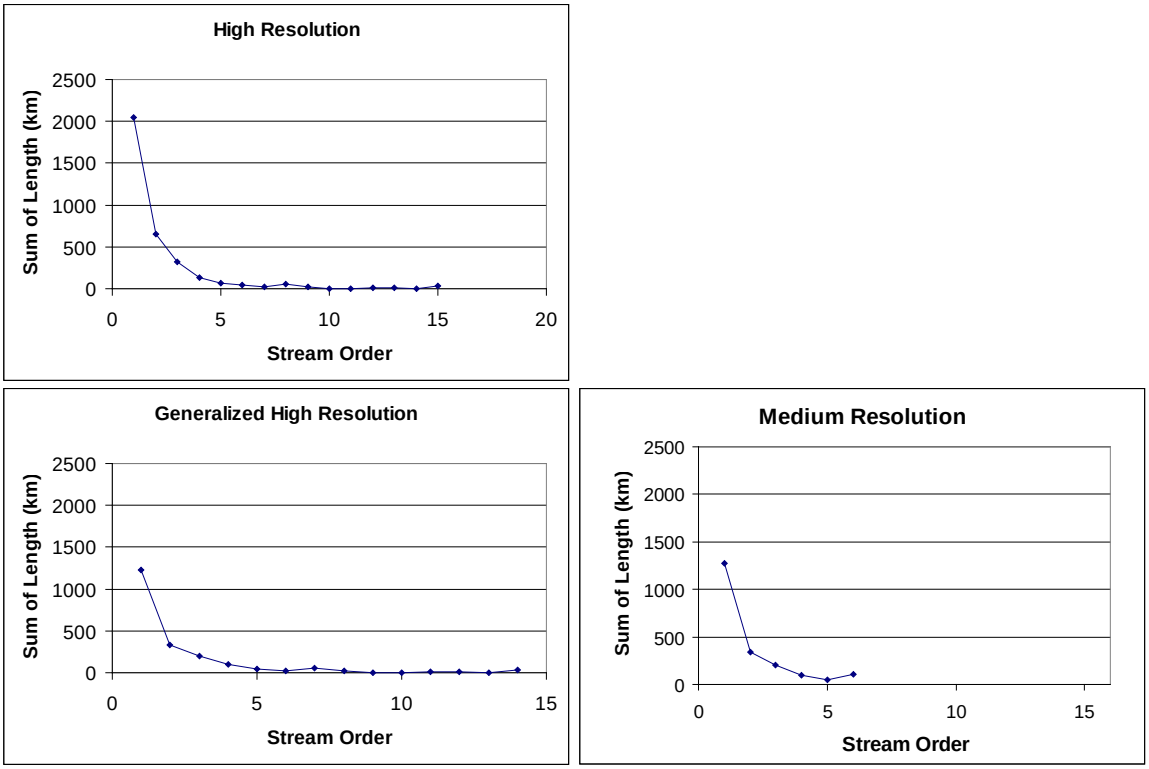

Figure 20. Summary of stream length by stream order for high resolution, medium resolution, and generalized high resolution Pomme De Terre subbasin.

In addition, Figure 21 displays the frequency of stream order segments in the three resolutions of data for the PDT subbasin. Again we see that the generalized data mimics the frequency pattern displayed by the medium resolution, but the difference in frequencies among corresponding orders ranges from about 19 to 500 percent, for stream orders 1 through 5, with the generalized data having larger stream order frequencies than the medium resolution. Frequency differences increase with stream order up to order 5. An explanation is that a larger number of confluence-to-confluence segments exist in the generalized data (1332) versus the medium resolution data (1014). Basically, a larger number of shorter confluence-to-confluence segments exist in the generalized network than in the medium resolution data, but the distribution of network length by stream order, up to order 5, is similar between these two datasets.

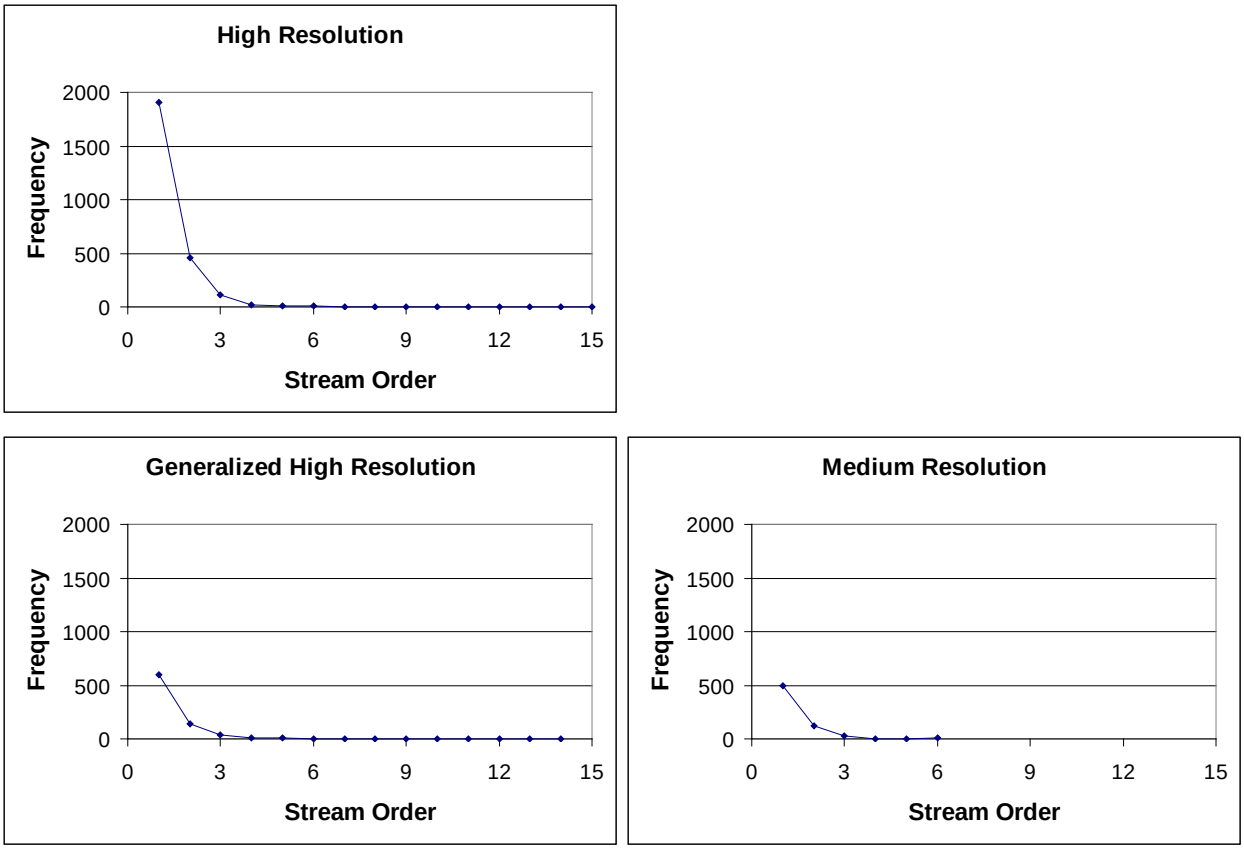

Figure 21. Summary of stream frequency by stream order for high resolution, medium resolution, and generalized high resolution Pomme De Terre subbasin.

SUMMARY

Thus far, quantities that characterize the network morphology of an NHD flowline network have been identified, and automated methods have been developed to compute network morphology statistics. The reach and basin based statistics, along with visual inspections, have assisted the analysis and comparisons of the network drainage for selected watersheds.

NED elevation data resampled to 30-m resolution for three subregions--one in Missouri and two in Colorado—were used to develop a linear relationship to estimate the watershed drainage density of elevation-derived synthetic streams from the minimum upstream drainage area that was used for synthetic stream formation. Reliable linear relationships were determined for the two study areas. Different slope parameters for the two models suggest that terrain differences between the two study areas influence the derived network drainage. This fact reinforces the idea that a network pruning strategy for generalization can be enhanced by applying a strategy that considers regional terrain variations—such as designing different treatments for several terrain strata. Other factors--such surface permeability or climate variables--can also be incorporated into the stratification, which may be similar to the physiographic divisions for the U.S (USGS 2004).

A network pruning strategy that iteratively prunes features based on minimum upstream drainage area until a desired drainage density is achieved was developed and tested on the high-resolution NHD for the Pomme De Terre subbasin. Using upstream drainage area rather than values determined from existing mapped features--such as arbolate sum--can eliminate drainage density variations due to disparate data collection standards. A drainage density of 0.942 km/sq km was selected for the generalization parameter to produce a dataset that mimics the medium resolution NHD for this subbasin. Reach and basin drainage statistics were computed for the generalized dataset and compared to the same quantities derived from the high and medium resolution NHD subbasins for the Pomme De Terre. From a basin perspective, the pruning process furnished a network that is structurally similar to the target medium resolution dataset. However, when compared to the medium resolution dataset, a larger number of shorter, high resolution, confluence-to-confluence segments and reaches were maintained in the generalized network causing dissimilarities in the frequency of features by stream order. This outcome is expected due to greater number of features in the high resolution data that are more detailed and smaller than those in the associated medium resolution dataset.

Linear relationships that approximate the watershed drainage density that should be depicted at a particular map scale were developed for the Gasconade Missouri and Colorado subregions. This relationship, in conjunction with the network pruning strategy, can be used to assist generalization of NHD data when a drainage density requirement is not known for a study area.

Additional work is needed to fine-tune the pruning strategy to ensure functionality in complex braided areas--such as in swampy coastal regions—and with very large subsets of the NHD. Furthermore, the pruning process must be adapted to work through internet web services that function with the National Map, and it must be computationally optimized.

In conjunction with the network pruning process, a suitable line simplification process must be identified and implemented in the generalization process to further eliminate extraneous detail. To implement generalization in real time, the line simplification algorithm must be fast, it should not generate new intersections or gaps in the pruned data, and it must maintain the connectivity among features. One candidate method is the skeleton retraction algorithm. The skeleton retraction algorithm uses the Medial Axis Transform, which can be computed with Thiessen polygons (Gold and Thibault 2001). We have been reviewing the feasibility using this algorithm, and we are currently trying to collaborate with the University of Glamorgan to acquire software algorithm that may be tailored for this task. Although the skeleton retraction algorithm may appear more appropriate than other approaches because it is less likely to generate intersections during line simplification, computationally, it may be too intense to be practical.

Multiple rules must be developed to handle the many feature types of the NHD data model during the feature simplification processes of generalization. Feature relationships that exist in the NHD model can be used to redefine artificial paths to other linear feature types when areal features must be simplified to lines. However, some area feature types—such as, swamp/marsh or lake/pond—do not have equivalent linear feature types in the NHD model. Therefore, some form of feature conversion or additional feature types may need to be defined.

Future work of transferring elevation derived surface flow values to the high resolution dataset may be simplified through NHD generalization. The pruning strategy based on upstream drainage area normalizes density variations due to inconsistent data collection standards. Elevation derived synthetic streams that are being developed by the National Center for Earth Resources Observation and Science (EROS, EDNA) can be designed to conform to the high resolution data or a slightly smaller scale. Subsequent conflation of the synthetic streams and associated flow characteristics can be used to update the NHD feature types—in particular the distinction between intermittent and perennial features—or to add missing features or remove extraneous ones. Furthermore, this process could link additional information to the NHD that is valuable to hydrologic analyses.

Acknowledgments:
Special thanks to Pat Turley and Josh Wilkerson for their assistance computing network morphology statistics.

REFERENCES

Gold, C. M. and D. Thibault. 2001. Map Generalization by Skeleton Retraction, *In Proceedings, International Cartographic Association*, Beijing,China, pp.2072-2081.

Horn, R.C., L. McKay, and, S.A. Hanson. 1994. " History of the U.S. EPA's River Reach File: A National Hydrographic Database Available for ArcInfo Applications ," in *Proceedings of the Fourteenth Annual ESRI User Conference*, Environmental Systems Research Institute, Redlands, CA.

Horton, R.E. 1945. Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology. Geological Society of America Bulletin, 56, 275-370.

Maidment, D. R. 1996. GIS and hydrologic modeling – an assessment of progress. Third International Conference on GIS and Environmental Modeling. January 22-26, 1996. Sante Fe, New Mexico. Accessed June 16, 2005 at URL: http://www.ce.utexas.edu/prof/maidment/gishydro/meetings/santafe/santafe.htm.

Northcott, W. 2002. Lecture notes to Biosystems Engineering 481: Agricultural and Small Watershed Hydrology. College of Engineering, Michigan State University. Accessed June 17, 2005, at URL: http://www.egr.msu.edu/~northco2/BE481/WshedChar.htm.

Pidwirny, M. 2004. Stream Morphometry. Accessed June 9, 2005, at URL: http://www.physicalgeography.net/fundamentals/10ab.html

Sanchez, R.R. 2002. GIS-Based Upland Erosion Modeling, Geovisualization, and Grid Size Effects on Erosion Simulations with CASC2D-SED. Ph.D. Dissertation, Civil Engineering Dept., Colorado State University. Accessed June 16, 2005 at URL: http://www.engr.colostate.edu/~pierre/ce_old/Projects/CASC2D-Rosalia/RRS_Diss4WWW/.

Smith, P.N.H, and D.R. Maidment. 1995. Hydrologic Data Development System. Report 95-1 Center for Research in Water Resources. University of Texas at Austin. Accessed June 16, 2005 at URL: http://www.ce.utexas.edu/prof/maidment/gishydro/docs/reports/smith/rep95_1.htm.

U.S. Geological Survey. 2005. National Elevation Dataset. Accessed June 7, 2005, at URL: http://ned.usgs.gov/.

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. 2003. Streams and Waterbodies of the United States. Accessed June 14, 2005, at URL: http://nationalatlas.gov/atlasftp.html.

U.S. Geological Survey. 2000. The National Hydrography Dataset: Concepts and Contents. Accessed June 8, 2005, at URL http://nhd.usgs.gov/chapter1/index.html.

U.S. Geological Survey. 1955. Map Publication Scales. Book 1, Part B, Chapter 1 of the Geological Survey Topographic Instructions. March 1955. 13pp.

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

Author Information

Lawrence Stanislawski, GIS Developer
SAIC National Mapping Division
1400 Independence Road, Rolla MO, 65401
573-308-3914 Fax: 573-308-3652
e-mail: lstan@usgs.gov

Michael Starbuck, Cartographer
USGS Mid-Continent Mapping Center
1400 Independence Road, Rolla MO, 65401
573-308-3687 Fax: 573-308-3652

Michael Finn, Programmer/Research Scientist
USGS Mid-Continent Mapping Center
1400 Independence Road, Rolla MO, 65401
573-308-3931 Fax: 573-308-3652

E. Lynn Usery, Research Scientist
USGS Mid-Continent Mapping Center
1400 Independence Road, Rolla MO, 65401
573-308-3837 Fax: 573-308-3652