Review of DEM stream derivation and spectral hydrographic feature detection for cartographic processing of the USGS National Hydrography Dataset

Paulo Raposo
paulo.raposo@psu.edu

for Seminar in Cartography: Multiscale Hydrography
GEOG 512, taught by Dr. Cynthia A. Brewer, cbrewer@psu.edu
Department of Geography, Pennsylvania State University
December 16, 2009

Introduction
The National Map maintained and published by the USGS has already become available in 7.5-minute quadrangle format as the "Digital Map - Beta" and the "US Topo" series through online download by the public. These map products represent the ongoing realization of a new map vision recently adopted by the USGS (Usery et al., 2005) wherein five layers of information are to be represented: digital orthoimagery, elevation, land cover, hydrography and transportation. Currently, most of the quadrangles published bear the first and last mentioned information layers along with geographic names, and research is underway on how best to incorporate hydrography and contours by automated generalization of USGS cartographic data.

NHD data does not always vertically integrate with candidate orthoimages for production of USGS US Topo-series topographic maps. The inspiration for this research project came from the sense that use of imagery in cartography, as is being done in the US Topo series, makes vertical integration (also termed conflation) with overlying vector layers of great importance. Vertical integration with imagery is crucial for several design considerations. It is believed here that "the principal requirements for cartographic data, ...are communicative adequacy and graphic or visual consistency," (Buckley et al., 2005) and that poorly integrated layers present inconsistent messages to map readers that lead to confusion and, ultimately, reader incredulity. Further, non-integrated layers diminish the aesthetic quality of a map. Vertical integration, in the words of Usery et al. (2005) "means the datasets match geometrically, topologically, and have a correspondence of attributes." Following this definition, vertically integrated datasets are both analytically useful and, by virtue of geometric correspondence, visually appreciable and non-alarming.

At the conception of this project I speculated on methods of stream detection related to orthoimagery in order to create guide data for geometrically transforming NHD such that it vertically integrates with orthoimagery. I conceived of two hypothetical methods, one using DEM-derived streams made from DEMs known to be vertically integrated with the orthoimagery to be used, and another using spectral data to extract streams from multi- or hyper-spectral orthoimages which could be used as map production orthoimages in the US Topo series. To explore the possibilities of these methods, I sought to read widely on the subjects of DEM stream derivation, spectral analysis for hydrographic feature recognition, and image conflation techniques. I review this literature in the following section, while in the section thereafter I elaborate on my ideas regarding possible processing involving either of the methods mentioned above.

Literature review
Derivation of streams from elevation data has been a research topic since the mid 1980s. O’Callaghan and Mark (1984) presented their method for stream derivation, which would later be
known as the D8 method, after the eight directions in which a given pixel could flow into its topographically lowest neighbouring pixel. Each pixel was given the value 1, and accumulation in each was calculated by summing the values of all the pixels upstream from it. Selecting pixels from the accumulation raster at a chosen threshold permits the extraction of pixels representing a stream. O’Callaghan and Mark noted in their article that their method tended to produce streams that flowed in cardinal directions, an artefact of the coarse raster elevation data from which they were calculated.

Hydrologic modeling continued to be a key concern in DEM creation and processing in years to follow. Development of the GTOPO 30 elevation dataset — one of the first global elevation datasets — was undertaken with continental-level hydrographic feature derivation as a key consideration (Verdin and Jenson, 1996). Methods to improve stream derivation from DEMs have been explored by researchers. A method commonly used is stream burning, presented by Maidment (1996). Stream burning is a process using vector stream data to force a stream channel where vector streams overlie the DEM; elevation values at such pixels are dropped, usually in subtraction by a constant value, so that an chasm is created in the DEM, forcing stream accumulation to gather there. Stream burning is used when the analyst is confident the vector stream positions are accurate, and is of interest in deriving streams from a DEM such that they align with these correct positions. Research has suggested that this method provides more accurate results for stream placements, and it has been used to generate diverse stream data sets (e.g., Renssen and Knoop, 2000). Callow et al. (2006) have cautioned, however, that use of stream burned DEMs in processing steps that depend on correct elevation data can be problematic. Alternatives to stream burning that do not affect the elevation values in the DEM have been explored, namely in using vector stream data to influence accumulation rasters (Kenny and Matthews, 2005; Turcotte et al., 2001).

Other research into improving stream derivation from elevation data has focussed on spatial resolution. McMaster (2002) has suggested that DEM spatial resolution should be no greater than the average hillslope length, and offers 180 m as a general value for maps at the 1:25,000 scale. Wu et al. (2008) found that higher resolution DEMs performed more accurately in stream derivation, with their best results occurring at 210 m resolution or finer. Some of the rasters in their sample were created by downsampling from 30 m DEMs; this method can be used to derive higher-resolution DEMs when no spatially finer data is available.

Still further work on DEM stream derivation has focussed on variations in landscape, since many common derivation methods assume landscape homogeneity and thus fail to model real variables such as surface permeability. One way of integrating considerations of landscape has been to divide the DEM into separate regions according to landscape characteristics, and treat each of these in a way that reflects the regional characteristics (Colombo et al., 2006).

Hydrological features are often analysed using remotely sensed spectral imaging. These technologies are widely used today by scientists to observe wetland ecological systems (Ozesmi and Bauer, 2002; see Narumalani et al., 1997 and Peixoto et al., 2009 for examples) because water features of various composition and in various states exhibit detectable spectral signatures, particularly in infrared wavelengths. Armenakis et al. (2003) have used techniques such as these to detect lakes in northern Canada, comparing results to lake features detected in scanned Canadian topographic maps and creating composite data layers in an effort to improve mapped lake accuracy. Their research used image fusion methods, wherein the 15 m resolution Landsat ETM-8 panchromatic band is used to interpolate 15 m resolution data from the instrument’s 30 m resolution multispectral bands; this method is commonly used in multispectral remote sensing. It is expected that the utility of hyperspectral imaging will increase with technological development (Goetz, 2009).

Along with DEM stream derivation, issues regarding vertical image integration of geographic datasets have been the focus of research since the mid 1980s. Early attention to data conflation was driven by the joint efforts of the USGS and the U.S. Bureau of Census in integrating their datasets (Saalfeld, 1988).
"Piecewise Linear Homeomorphic transformation" (PLH), or rubber sheeting, was introduced by Saalfeld (1985), as well as White and Griffin (1985). Rubber sheeting processes generally divide the layer of interest into triangles for the transformation process, usually using the Delaunay triangulation method, and then perform affine transformations on each triangle, preserving topological relationships between the vertices of the input layer. Also, attributes of the layer are retained. This method was proposed as a means of increasing the positional accuracy of a dataset with desired attributes by using another dataset known to have better positional accuracy.

Gillman (1985) suggested that the problems in image integration be seen in two parts: first, the identification of corresponding points between data sets, and second the methods of transformation employed. Research since has tended to focus on the first part identified by Gillman, and approached the problem as a matter of point-matching. The method most often chosen in the literature to address Gillman's second part has been rubber sheeting.

Several publications from researchers affiliated with the USGS have cited the need to develop automated methods for feature extraction and vector-to-image conflation (Finn et al., 2004; Usery et al., 2005). Most existing implementations of this procedure are interactive, involving user input in the process of control point identification, but national mapping organizations (NMOs) such as the USGS seek greater levels of automation to address their very large datasets.

The overwhelming majority of published literature on image conflation with vector data has focussed on the problem set posed by vector road data and orthoimagery. Rellier et al. (2002) have used Markov random fields and Bayesian statistical analysis to re-shape road networks into correspondence with detected roads in SPOT imagery. Knoblock et al. (2006) and Chen et al. (2008) have achieved favourable results with automated feature extraction algorithms that recognize geometric properties such as angles of roadways in imagery, and also with neighbourhood (as opposed to global) search methods for finding corresponding points between datasets. Their methods take advantage of the fact that the layers meant for vertical integration (e.g., a vector road layer and an orthoimage) begin in approximate, rough correspondence, requiring relatively small, and usually affine, transformations to achieve conflation. Stanislawski et al. (2007) have developed a means of detecting "blunders," gross errors in automated point matching results, and removing these from the set of control points used in automated rubber sheeting of vector layers to orthoimagery.

Song et al. (2009) present compelling work wherein they achieved 95.5% accuracy in vertically integrating vector roads with orthoimagery. Their method involves removal of vegetation pixels using an NDVI threshold, and subsequent spectral recognition of potential road pixels. Intersections are found by calculating histograms for each pixel, measuring distance to the furthest pixel of similar spectral signature on the y axis, and binning the data into 360 bins, each corresponding to one angular degree, being cartographic directions in which measurements described for the y value were taken. Pixels at four-way intersections are likely to have

![Figure 1 - Methodology of Song et al. (2009) (reproduced)](image-url)
four peaks in their histograms, pixels at three-way intersections are likely to have three peaks in their histograms, and so on. An iterative "relaxation labelling algorithm" is then used to match vector intersections to road intersections identified in the imagery. A snake algorithm moves intermediate road vertices onto a road-identified image pixel, and finally a refinement algorithm moves all points into the center of road pixel areas to achieve good cartographic vertical integration. Their method is outlined in Figure 1.

**Proposed approach for NHD processing**

Based on literature reviewed and awareness of pre-existing research efforts at the USGS concerning road data vector transformation in order to achieve vertical integration with orthoimagery, it is suggested here that a similar approach can be taken with NHD data. While the overwhelming majority of research in rubber sheeting has taken place using road data in order to opportunistically use the easily identifiable control points offered by intersections and road termini, it is suggested here that a similar operation is possible on hydrographic data. Branch points in stream systems where tributaries merge or streams otherwise diverge can be used in a similar fashion as road intersections have been, as control points for rubber sheeting operations. Further, as research and technology permits, other vertices along hydrographic vectors, such as the apexes of highly salient curves, may be automatically detected and used as control points.

As reviewed in the literature (Knoblock *et al.*, 2006; Song *et al.*, 2009; Usery *et al.*, 2005), feature extraction from imagery has commonly taken advantage of the particular geometries common to orthogonal road networks. Such an *a priori* assumption about the geometries of target features for hydrography images isn't appropriate, as river systems exhibit a variety to geomorphologic forms that translate to a variety of geometric planimetric forms. This dimension of complexity makes detection of hydrographic features in imagery using geometric assumptions more difficult than for roads.

In response to this particular difficulty, I have explored two alternative methods of stream detection, one using elevation data related directly to the orthoimagery to be used in the map, and the other using spectral methods to detect hydrographic features by spectral signatures from the orthoimagery itself, provided the imagery is sufficiently multi- or hyperspectral, and of high enough spatial resolution to do this. Figure 2 outlines the overall processes involved using either method.

The first method involves the provision that the DEM used is known to be vertically-integrated with the orthoimagery to be used in the final map product. DEMs and orthoimagery are not usually created and maintained with mutual
vertical integration in mind, and are not normally compared to assess vertical integration. DEMs, however, are routinely used to orthorectify imagery. It follows then that a DEM used to rectify an image will have exerted considerable influence on the vertical placement of each resampled pixel in the orthoimage produced. It is here suggested, then, that a DEM used to rectify an image may serve an added purpose in providing an alternate plotting of the land forms visible in that image; that is, valleys and ridges in the DEM correspond to valleys and ridges in that image. If this is so, stream derivation from a DEM used to orthorectify an image shall correspond, in theory, with streams visible in the orthoimage.

It is also observed that stream derivation from elevation data will return more streams that are present in NHD. The new streams will generally be headwater streams, which can be added to NHD to improve NHD coverage and analytical utility. It is also noted, however, that some streams among those derived and potentially added to NHD will be dry valley floors or episodic streams at most, and these streams will need to be removed from a set of new streams under consideration for addition to NHD. Towards an automated means of conducting this editing, derived streams may be compared to spectrally-detected streams from spectral imagery.

The second method requires data obtained by multi- or hyper-spectral sensors. These data would need to exist in high (i.e., at least sub-meter) spatial resolution in order to detect narrow streams of the calibre visually detected by USGS cartographers in legacy topographic map compilation. Further, this method would involve careful threshold settings; water is known to be spectrally identifiable by very low infrared values, but different sedimentation and biomass loads in various water bodies make for a considerably wide variation among hydrographic feature spectral signatures. Threshold brackets across sensor bands used to detect hydrographic features by this method would have to be sufficiently wide to catch the variety of hydrographic features present in the area(s) in question, while also effectively excluding non-hydrographic pixels in the image.

The suggested spectral method can obviate the need to ensure a priori vertical integration of the input datasets that is apparent in the first method using elevation data. If the imagery used is both sufficiently multi- or hyper-spectral to contain hydrographic-diagnostic pixels, and has bands in visible red, green and blue bandwidths (e.g. as the Landsat Thematic Mapper sensor does, despite its coarse spatial resolution for this suggested application), the RGB channels can be used to compose a false-colour image that appears convincingly like a photograph for use in the final map product. Spectrally-recognized hydrographic pixels in such an image, as well as vectorized hydrographic features created thereof, would already be in seamless vertical integration with the rest of the image.

Regardless of which method is used to detect streams, either in the image (as per the spectral method), or related directly to the image (as per the DEM-derived method), vectorization of detected streams constitutes the next step. Upon vectorization, control point pairs can be identified between the NHD data and the vectorized streams. Vectorized streams, particularly when derived from elevation data, may be unrealistically angular; it may prove useful to generalize and/or smooth derived streams upon vectorization in order to produce more cartographically-credible lines. If this is done, both the generalized/smoothed lines, as well as the original (and perhaps angular) lines can be used in subsequent steps, and either usage can be evaluated and compared upon process completion for best results.

The next step would be to identify pairs of related points between the vectorized streams and NHD streams for use as control points in rubber sheeting. While this is generally easily done by a human cartographer, automation is demanded in applications such as USGS topographic production because of high labour costs across vast national data. It is suggested here that identification of control points for hydrographic stream courses could follow a similar method as has been done with roads by training automated systems to detect intersections at points of stream convergence or branching. Further, algorithms for detection of apex points of salient curves can also be developed in order to add such
points to the usable sample set for use in rubber sheeting. Methods such as those presented by Stanislawski et al. (2007) may be used to filter such points in order to remove any that are clearly erroneous.

It is possible to rubber sheet either the orthoimage to match NHD, or NHD to match the orthoimage. Following on the strategies of several authors who transformed their vector layers to match imagery, I propose transforming the NHD vector layer to correspond to the orthoimage. The optimal transformation would be by way of local piecewise linear homeomorphic transformation (i.e., rubber sheeting), so that positional incongruence between NHD and the streams of the image is addressed at multiple, local areas; this addresses the fact that error incongruence between NHD and the orthoimage is most likely to be non-systemic (Song et al., 2009). The rationale of transforming NHD to match the orthoimage is based on the fact that orthoimagery, by definition, has already been processed to remove distortions and thus presents a credible, accurate placement of features visible in the image. By contrast, NHD vectors have been generally hand-digitized from older orthoimages, and it is presumed that greater positional error would exist in the NHD than in the recent orthoimagery selected for contemporary topographic map production.

The goal being the achievement of vertically integrated streams for map production, it is preferable to use geometrically transformed NHD data rather than newly detected streams because NHD is attribute-rich. A benefit of the rubber sheeting transformation method on NHD is that the transformed NHD will retain attributes while becoming more cartographically appropriate. In this way, transformed NHD layers retain their analytical usefulness. Further, transformed NHD layers can be stored in a separate digital cartographic model (DCM) as distinct from original NHD data, with references to the particular orthoimagery they correspond to (i.e., with which they vertically integrate). They can also be used in potential future multi-representation databases (MRDBs) as seed layers upon which generalization may be performed to produce data layers suited to smaller-scale display.

To these effects, I have run preliminary experiments rubber sheeting NHD to match orthoimagery. These are presented in the following two sections.
Stream derivation and rubber sheeting tests

i. Preliminary trial

In order to experiment with my proposed methods, I first assembled data that I knew would yield a non-integrated situation. I used a small sample of the National Elevation Dataset (NED) data at the 1/3 arc-second resolution (approx 10 m) and NHD Medium resolution hydrography. I conducted the processes interactively, using the tools in ESRI ArcGIS 9.3.1 Spatial Analyst extension software to derive streams from the NED data, and then performed “Spatial Adjustment” in ArcMap to rubber sheet the NHD data to these derived streams. The following screen captures illustrate the process undergone.

NED raster was filled to eliminate pits. A flow accumulation and flow direction raster was calculated in ArcMap using the D8 method.

Streams were queried out of the accumulation raster by selecting those cells at or above an accumulation threshold.

Raster streams were vectorized.
Vector streams were smoothed using the PAEK algorithm.

A subset of these were extracted to run the rubber sheeting test. (ESRI orthoimagery included for visual reference.)

NHD Medium data was introduced, and also sampled for a test subset (in blue).
Spatial adjustment was carried out using the rubber sheeting method, with the establishment of "displacement links" (common points between the feature to be spatially adjusted and the feature to adjust it towards; arrows) and "identity links" (points at which the feature to be spatially adjusted is already in correspondence with the feature to adjust it towards; red cross hairs).

The NHD data geometry was adjusted to more closely follow the derived-stream guide.

**Discussion**

It was concluded from these first attempts that rubber sheeting methods for stream conflation with orthoimagery was a promising research direction. For further experimentation, I turned to examples of data that might actually be used by the USGS in topographic map production at the 1:24,000 scale.

**ii. Experimental study on a Louisiana watershed**

The data assembled and used in the following illustrated experiment were downloaded from online USGS data portals. They cover a small test area in central Louisiana, arbitrarily chosen for public internet access to downloadable data sets for all three data themes involved. All data were constructed on the 1983 North American Datum and projected Plate Carrée. The data sets were as follows:

- NHD High Resolution (1:24,000)
- 1/9 arc-second NED
- National Agriculture Imagery Program (NAIP) 1 m color orthoimagery

Processing was conducted interactively using ArcMap 9.3.1, Spatial Analyst extension and "Spatial Adjustment" capabilities built into the software. The following series of screen shots illustrates the processing steps involved.
The NED 1/9 arc-second DEM at 1:24,000 in the ArcMap display.

Blue lines are NHD High Resolution streams for this area. It is observed that they do not correspond with visible valley depths in the DEM. The red polygon identifies an arbitrarily chosen area in which experimentation was conducted.

The image has been panned to the left to center on the study area. Blue NHD High Resolution flowlines are overlaid on NAIP 1m orthoimagery, and the viewing scale is set to 1:24,000. It is observed that the NHD streams are not successfully vertically integrating with the streams in the orthoimage at 1:24,000.
Streams were derived using the D8 method built into ArcMap 9.3.1 Spatial Analyst Extension - detail shown.

Derived streams were smoothed using the PAEK method to eliminate angularity - detail shown.

Vectorized (i.e., derived) streams are shown in orange. It is observed that there is positional disagreement between the NHD and derived streams.
Rubber sheeting, first performed on derived streams. Red cross-hairs indicate points of existing overlap where process is to preserve the position. Arrows indicate pairs of control points, where arrow ends are at points along derived streams to be moved, and arrow points are at points along NHD to be arrived at.

Result.

Result seen over orthoimage at 1:24,000.
Rubber sheeting, now performed on NHD data, moving toward derived streams.

Result seen over orthoimage at 1:24,000.

Detail of original NHD stream (thinner blue line) and transformed NHD stream (thicker) over orthoimage.
Discussion

It was observed that rubber sheeting NHD to match derived streams produced a better vertically-integrated result; however, the result was not perfect. One speculation is that using a greater number of control points in the rubber sheeting process could generate better results, as more local transformations would occur, bringing more points along the stream courses into alignment.

One immediately apparent observation made was that derived streams number many more than do NHD streams. This is due to the fact the DEM derivation produces streams in any depression, provided accumulation values in pixels in a given depression are at or above the accumulation threshold defined by the analyst. It is speculated that DEM stream derivation may prove an effective way to find headwater streams that NHD data lack. It is also understood, however, that DEM derivation will return valley channels that do not normally host water streams, or host them episodically at best, and these derived streams need to be edited out of the dataset before they are committed to map data. It is speculated that stream detections made with spectral data may assist in this process: a derived stream layer can be compared to a spectral layer wherein headwater streams are detected spectrally, and streams in the derived layer that are not represented in the spectral layer can be deleted.

Conclusions, and suggestions to the USGS

The cartographic position taken here is that use of orthoimagery in the US Topo-series maps makes vertical integration with the orthoimage of ultimate importance, both with respect to information clarity and for good graphic design. Further, it is here presumed that the locations of features seen in the image will have been correctly georectified, such that these locations are to be taken as correct and accurate above those in other data layers with which the image is to be conflated. In the case of hydrographic features, the USGS seeks to use the National Hydrography Dataset (NHD) as the data with which to cartographically depict the nation’s water features. It is therefore recommended that the locations of water features in the orthoimagery used by the USGS for topographic map production be used to geometrically alter the locations of those same features in NHD data such that the two layers successfully integrate vertically.

Two methods for identifying water features in reference to imagery are suggested:

a. that water feature locations in imagery be derived from digital elevation models (DEMs) which are directly related to the orthoimagery used in that they were used to rectify the imagery, or
b. that multi- or hyper-spectral imagery be used to identify water features by their spectral signatures, and that the same spectral imagery be used to constitute topo map imagery using RGB bands.

Regardless of which method is used to identify water features in the raster data, subsequent processing steps involve the vectorization of water features from raster features, generalization and smoothing as necessary, and the geometric alteration of NHD data by rubber sheeting to correspond with the imagery-derived hydrographic features. Figure 1 outlines the essential processing steps of these methods.

A key benefit to the proposed methodology is the fact that attributes in the NHD vector features are retained, and thus an altered NHD vector layer resultant from the processes here recommended maintains both analytical utility while also becoming cartographically compatible with orthoimagery. Upon generation of such data, these may be retained in a digital cartographic model (DCM) by the USGS. Further, these may be used as seed layers in a multi-representation database (MRDB).

It is suggested that the USGS continue research efforts into the automation of the processes involved until vector-to-image integration can be successfully achieved within acceptable operational costs.
Works Cited

Armenakis, C., F. Leduc, I. Cyr, F. Savopol, and F. Cavayas

Callow, John Nikolaus, Kimberly P. Van Niel, and Guy S. Boggs

Chen, Ching-Chien, Craig A. Knoblock, and Cyrus Shahabi

Colombo, Roberto, Jürgen V. Vogt, Pierre Soille, Maria L. Paracchini and Alfred de Jager
2006 Deriving river networks and catchments at the European scale from medium resolution digital elevation data. *Catena*, 70, pp. 296-305.

Finn, Michael P., E. Lynn Usery, Michael Starbuck, Bryan Weaver, and Gregory M. Jaromack

Gillman, D.

Goetz, Alexander F.H.

Kenny, Frank and Bryce Matthews

Knoblock, Craig A. and Cyrus Shahabi

Knoblock, Craig A., Cyrus Shahabi, Ching-Chien Chen, and E. Lynn Usery

McMaster, Kevin J.

Narumalani, Sunil, Yingchun Zhou, and John R. Jensen
1997 Application of remote sensing and geographic information systems to the delineation and analysis of riparian buffer zones. *Aquatic Botany*, 58, pp. 393-409.

O’Callaghan, John F. and David M. Mark
Ozesmi, Stacy and Marvin E. Bauer

Peixoto, Juliana Maershner Aguiar, Bruce Walker Nelson, and Florian Wittmann

Rellier, Guillaume, Xavier Descombes, and Josiane Zerubia

Renssen, H. and J.M. Knoop

Saalfeld, Alan

Song, Wenbo, James M. Keller, Timothy L. Haithcoat, and Curt H. Davis

Stanislawski, Lawrence V., Michael P. Finn, E. Lynn Usery, and Paul M. Robinette

Turcotte, R., J.-P. Fortin, A.N. Rousseau, S. Massicotte, and J.-P. Villeneuve

Usery, E. Lynn, Michael P. Finn and Michael Starbuck

Verdin, Kristine L. and Susan K. Jenson

White, M.S. and P. Griffin

Wu, Shuo-sheng, E. Lynn Usery, Michael P. Finn, and David D. Bosch