CHAPTER 31

The Effects of DEM Generalization Methods on Derived Hydrologic Features

D.B. Gesch

INTRODUCTION

Digital elevation model (DEM) data are being used routinely as a source for deriving hydrologic features such as drainage basins and stream networks. Many of the popular geographic information systems (GIS) and image processing software packages include functions for deriving hydrologic data directly from DEMs. The scientific literature includes many examples of both the background of the methods (Mark, 1984; Band, 1986; Jenson and Domingue, 1988; Fairfield and Leymarie, 1991) and application of the techniques (Band, 1989; Jenson, 1991; Moore et al., 1991; Maidment, 1993).

The increasing availability of global elevation data, such as the GTOPO30 global DEM produced by the U.S. Geological Survey (USGS) (Gesch and Larson, 1996), has allowed usage of the techniques for deriving hydrologic features from DEMs to be extended to continental and global scales (Verdin and Jenson, 1996). Development of continental and global scale DEMs usually involves generalization of higher resolution source DEMs. While the effects of terrain data resolution and DEM aggregation on hydrology and geomorphology applications have been well documented (Chang and Tsai, 1991; Wolock and Price, 1994; Vieux, 1995), there is little documentation on the effects of the method of generalization on derivative hydrologic products. Fisher (1996) has described the effects of generalization methods on another DEM derivative, the viewshed.

DEM GENERALIZATION

Generalization of DEM data may be done for several reasons, including "scale" reduction and data volume reduction. Scale reduction may be done strictly for cartographic display purposes, or it may also be performed to adjust the grid spacing of the terrain data to a resolution appropriate for the features being analyzed. The latter is the case when DEMs are used to derive features for use in large area (continental and global) land surface studies. Higher resolution DEMs are generalized to a level that reduces the data volume to be processed but still portrays topographic features at a level of detail suitable for the applications in which they are to be used. A case in point is the USGS GTOPO30 global DEM which has elevations regularly spaced every 30-arc seconds of latitude and longitude (approximately 1 kilometer). A large portion of GTOPO30 was produced by generalizing higher resolution source DEMs having a grid spacing of 3-arc seconds (approximately 100 meters).

Cartographic generalization is an area in which significant research and development have been conducted, although most studies have concentrated on generalization of linear and polygonal features in vector cartographic data. Generalization of raster terrain data, while similar in concept to generalization of other forms of spatial data, poses unique challenges, due in part to the manner in which the information content is represented in gridded data. In vector cartographic data, the features of interest have already been delin-

eated during the compilation process. In gridded terrain elevation data the features of interest; for instance, ridges and stream lines, are included with other features that may not be of interest; for example, broad flat areas. The challenge, then, in generalizing DEMs is to retain the maximum amount of information content (by keeping a proper representation of critical features), but this is made more difficult by not knowing beforehand all the various applications in which the generalized data may eventually be used.

There are several categories of generalization methods for DEM data: statistical, resampling, and morphology-based. Statistical methods use as the aggregated elevation value a representative statistic, such as minimum, maximum, mean, median, mode, or weighted average, calculated from a window of full resolution elevation cells. Resampling methods use a geometric transformation with interpolation to accomplish a change in the grid spacing. Nearest neighbor and bilinear resampling algorithms are often used for raster elevation data. When the generalized cell size is an integer multiple of the full resolution cell size (and there is no change in projection or orientation between the original and generalized grids), nearest neighbor resampling is equivalent to systematic subsampling in which every nth row and column are selected. Statistical and resampling approaches to generalization have the advantage of being widely available in many GIS and image processing software packages. They work well in low relief areas, but smoothing of features can occur in high relief areas and prominent topographic features may not be retained, especially if the degree of generalization is significant.

Morphology-based generalization methods take into account the topographic features present in the DEM. The goal is to retain as much of the significant topographic information (represented by the critical terrain features of ridges, valleys, peaks, and pits) as possible in the generalized DEM. Weibel (1992) presents a very thorough discussion on the types of terrain generalization methods. Morphology-based approaches are similar in concept to the structure line model of generalization described by Weibel. The "breakline emphasis" method of DEM generalization, an implementation of a morphology-based approach, is the subject of the work reported here. The method is named for its emphasis on retaining the topographic breaklines (ridge lines and stream channels) as depicted in the full resolution elevation data.

Breakline Emphasis Generalization Method

The breakline emphasis method of DEM generalization involves two steps: extraction of topographic breaklines from the full resolution DEM, and selection of generalized cell values based on the absence or presence of breaklines. Topographic breaklines are extracted directly from the full resolution data with the widely used methods described by Jenson and Domingue (1988). Figure 31.1 outlines the breakline extraction procedure. Flow direction processing determines the direction of flow from each cell to its steepest downslope neighbor (1 of 8 possible directions). Flow accumulation finds the total number of upslope cells that flow into each cell. Ridges are delineated simply by extracting those cells which have a flow accumulation of zero (there is no flow from any of the neighboring cells). Streams are found by inverting the DEM and then performing the same flow direction and accumulation processing used for ridges. The streams extracted in such a manner may not coincide exactly with the actual stream channels, especially in low relief areas where factors other than just topographic position influence the location of the channel, but use of the inverted DEM allows for complete automation of the procedure. An alternative method of delineating the streams would be to threshold the flow accumulation calculated from the original DEM (where higher accumulation values represent the stream lines where flow accumulates), but such an approach would require intervention by an analyst to interactively select a threshold value.

Figure 31.2 shows how the breaklines are used in selection of a value for each cell in the generalized DEM. If full resolution ridge and/or stream cells are present in the area covered by the generalized cell, a simple majority criteria is applied to classify the generalized cell as either a ridge cell or a stream cell. Ridge cells are assigned the maximum value of the corresponding full resolution elevation values, while stream cells receive the minimum elevation from the area of corresponding full resolution cells. In this manner, the significant topographic structure lines are preserved in the generalized DEM. Generalized cells that do not contain breaklines are simply assigned a representative value from a nearest neighbor resampling of the full resolution DEM.

Breakline emphasis processing was applied to a portion of the 3-arc second source data that were used for the GTOPO30 global DEM. In this case, the generalized 30-arc second data were created for each grid

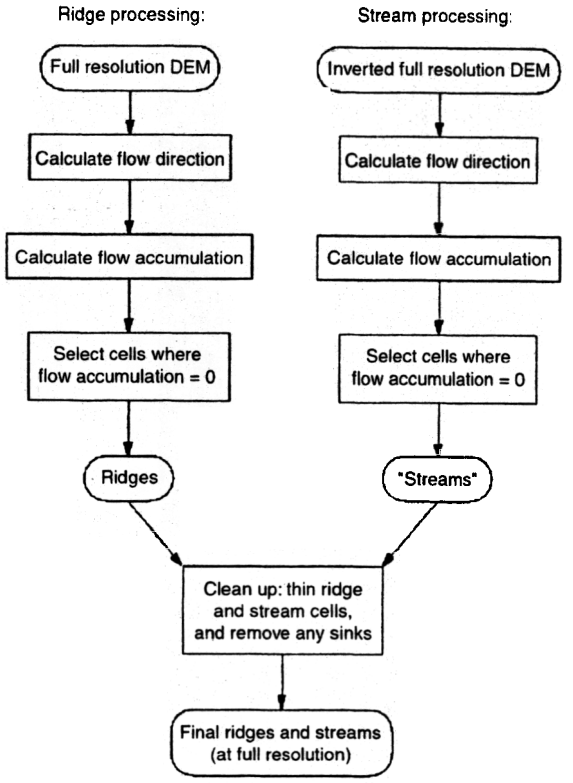

Figure 31.1. Breakline extraction procedure.

cell by selecting one elevation value to represent the area covered by 100 full resolution grid cells (a 10-by-10 matrix).

EFFECTS OF GENERALIZATION ON HYDROLOGIC DERIVATIVES

To investigate the effects of DEM generalization methods, tests were conducted for two drainage basins in the United States. DEM data with a 100-meter grid spacing were generalized to 500-meter, 1-kilometer, and 2-kilometer grid spacings using resampling (nearest neighbor), morphology-based (breakline emphasis), and statistical (mean and median) methods. The James River basin in North and South Dakota and the Allegheny River basin in western Pennsylvania were used for the tests. The basins are part of a national data set of hydrologic unit codes (HUCs) (Seaber and Kapinos, 1987), delineated from 1:250,000-scale topographic maps, that are commonly used for watershed level applications. The HUCs serve as the reference against which DEM-derived basins are compared.

Drainage Basins

Drainage basins were derived from the full resolution and generalized DEMs using the hydrologic analysis tools available in the ARC/INFO[1] GIS software package. Figure 31.3 shows graphically the DEM-derived Allegheny River basins (shaded) compared with the reference HUC (outlined) at the various grid spacings. The percentages refer to the percent of the HUC covered by the derived basin. Table 31.1 lists more comparison statistics for two of the generalization methods, nearest neighbor resampling and breakline emphasis. Note that even in the full resolution DEM a significant portion of the HUC is not covered by the DEM-derived basin. Further investigation has indicated that by slightly modifying just a few elevation values along the sub-basin drainage divide, the derived basin would include the previously missing area. The percent difference in total area of the derived basin against the reference HUC can sometimes be misleading. The percent difference in total area should ideally be zero, but even at zero difference the spatial correspondence of the derived and reference basins could be poor if the "undershoot" area (error of omission) is equal to the "overshoot" area (error of commission). Therefore, when comparing the results, it is useful to examine the percentages of undershoot and overshoot (which should ideally both be zero), as well as a graphic portrayal of the basin shapes.

Figure 31.3 and Table 31.1 indicate that for the Allegheny River basin, a basin with high relief and well defined drainage, the breakline emphasis generalization approach performs better than nearest neighbor resampling, in terms of derived basins matching the reference basin. At the 1-kilometer grid spacing the statistical methods perform equally well, but as the generalized grid spacing increases to 2 kilometers, only the basin derived from the breakline emphasis generalized DEM maintains a match of the reference basin consistent with that derived from the full resolution DEM. At 500 meters all the generalization methods perform about the same, presumably because the level of generalization is less coarse, allowing enough of the significant topographic features to be retained in each case.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

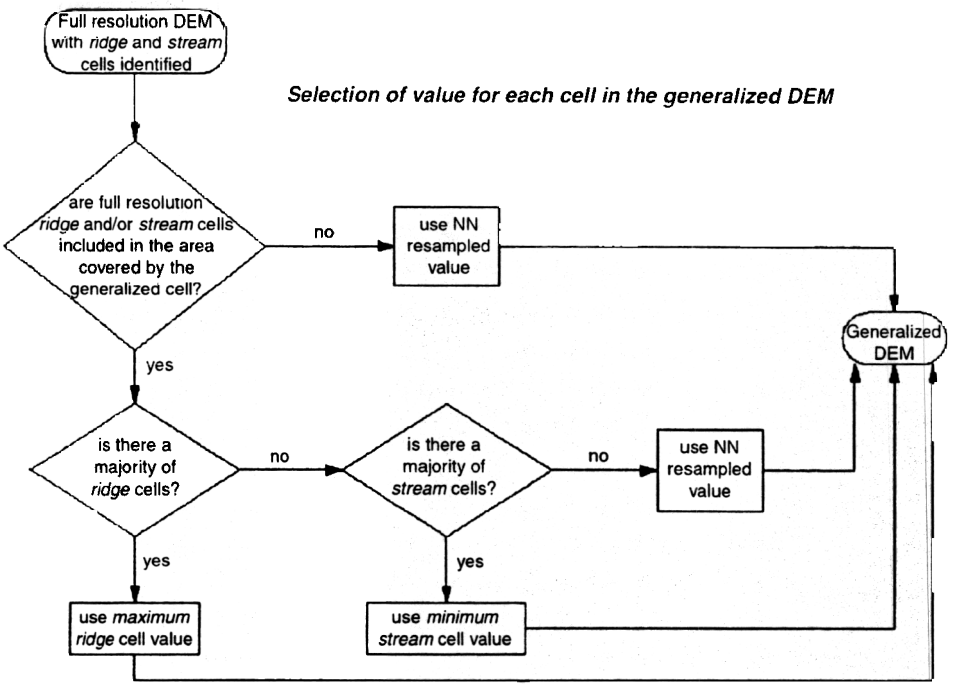

Figure 31.2. Use of breaklines to guide selection of generalized DEM values.

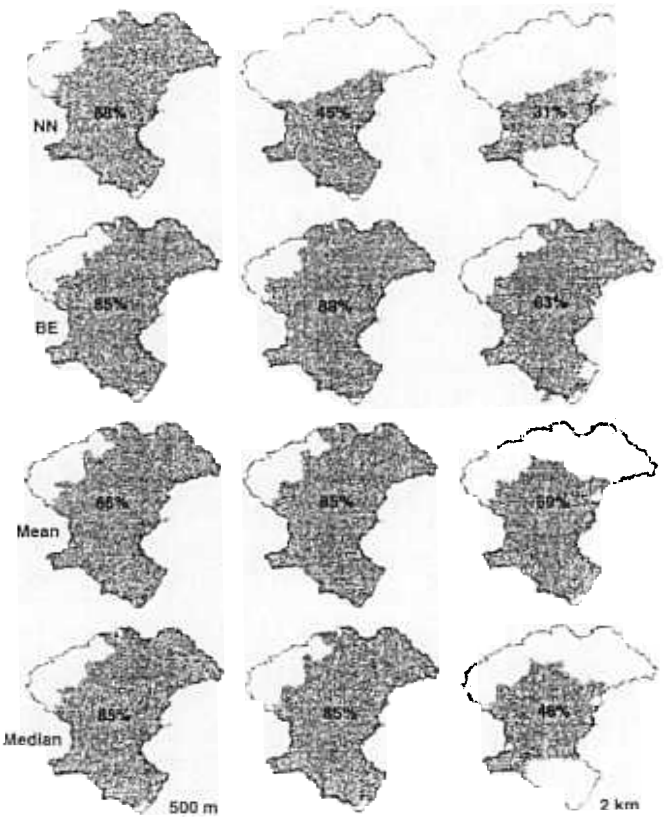

Figure 31.3. DEM-derived Allegheny River drainage basins (shaded) compared with reference basin outline.

Table 31.1. Comparison of Drainage Basins Derived from DEMs Generalized with Breakline Emphasis (BE) and Nearest Neighbor Resampling (NN).

Generalization Method	Grid Spacing	Percent Difference in Total Area: Derived Basin vs. HUC	Percent of HUC Covered by Derived Basin	Percent of HUC Not Covered by Derived Basin (Undershoot)	Percent of HUC in "Extra" Cells (Overshoot)
None		−13.95%	85.81%	14.82%	0.87%
NN		−13.54%	85.57%	14.43%	0.89%
BE		−13.41%	85.40%	14.60%	1.19%
NN	km	−53.93%	45.40%	54.60%	0.67%
BE		−10.49%	88.36%	11.64%	1.15%
NN	2 km	−67.47%	30.57%	69.43%	1.96%
BE		−15.07%	83.12%	16.88%	1.81%

Stream Networks

Stream lines (as defined by cells of concentrated flow accumulation) also were derived from the full resolution and generalized DEMs using the ARC/INFO hydrologic analysis tools. The reference data set for the James River basin was the U.S. Environmental Protection Agency Reach File (Dewald et al., 1996), a national hydrology data set derived from 1:500,000-scale map sources. Figure 31.4 shows the comparison of stream lines derived from DEMs generalized with nearest neighbor resampling and breakline emphasis approaches. The derived stream lines (shown as shaded cells) are overlaid on the reference stream lines for a portion of the drainage network of the James River basin. A low relief area such as the James River basin is the type of area where the effectiveness of automatic stream extraction procedures is controlled by the quality and resolution of the DEM. The gross mismatches between the derived and reference streams at the 500-meter grid spacing likely occur in nearly flat areas where the resolution of the original source DEM was too coarse to capture the subtle topographic gradients. Overall, the streams derived from DEMs generalized with breakline emphasis correspond much better spatially to the mapped hydrography than do those derived from nearest neighbor resampled DEM's. The drainage pattern is also much more consistent across the three grid spacings of the breakline emphasis DEMs than it is for the nearest neighbor DEMs.

Surface Curvature

Surface curvature (second derivative of the surface) is a topographic parameter that can be used in modeling erosion, deposition, and sediment transport (Zeverbergen and Thorne, 1987). Curvature was calculated for the Allegheny River basin from the full resolution DEM and from 1-kilometer generalized DEMs. Table 31.2 shows a comparison of the areas classified as flat, concave, and convex in each DEM. Because of the significant difference in cell sizes it is difficult to compare quantitatively the spatial correspondence of the curvature classes from the generalized DEMs with those from the full resolution DEM. This is reflected in the much larger percentage of area classified as flat in the full resolution DEM. The 3 by 3 cell window used to calculate curvature covers a much smaller ground area in the full resolution DEM, thus increasing the likelihood of calculations based on nearly identical elevations. However, the relative proportion of concave and convex areas is a convenient way to compare the performance of the generalization methods. As seen in Table 31.2, breakline emphasis was the only generalization method of the four tested which resulted in a DEM that maintained the correct ratio of concave to convex area (as found in the full resolution DEM).

CONCLUSIONS

Limited tests indicate that breakline emphasis, an implementation of a morphology-based DEM generalization approach, is effective in preserving topographic structure in generalized DEMs, resulting in better derived hydrologic features. Drainage basins and stream networks derived from DEMs generalized by breakline emphasis correspond better to reference data sets than do those derived from DEMs generalized with

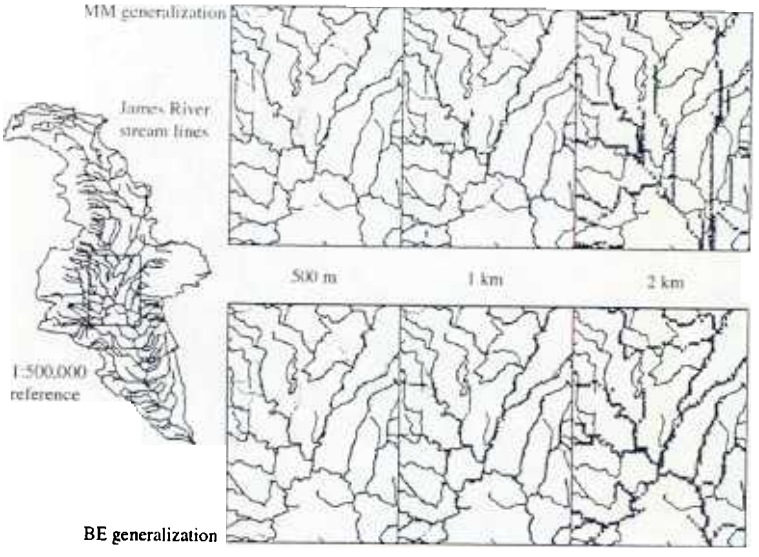

Figure 31.4. Comparison of DEM-derived stream lines.

Table 31.2. Comparison of Land Surface Curvature Classes Derived from Generalized DEMs.

Generalization Method	Grid Spacing	Percent of Basin Classified as:			Ratio of Concave to Convex Areas
		Flat	Concave	Convex	
None	100 m				1.046
NN	1 km				0.975
BE	1 km				1.048
Mean	1 km				0.971
Median	1 km				0.966

resampling and statistical methods, especially at coarser grid spacings. Further tests resulting in quantitative comparisons of derived features are required on basins in a wide range of terrain relief conditions. The epsilon band concept for assessing the accuracy of area measurements as applied to drainage basins (McAlister et al., 1996) would be useful for quantitative comparison of derived basins. Suitability of the breakline emphasis method for generalizing very high resolution elevation data also should be investigated. Comparisons should also be done against terrain data generalization schemes that use fractal interpolation (Bindlish and Barros, 1996).

The breakline emphasis method offers the advantage that it can be completely automated using the hydrologic analysis tools and raster data processing routines available in many popular GIS and image processing packages. The approach is also flexible, in terms of the criteria that can be used for selection of a generalized grid cell value. As currently implemented, the criteria applied for selection of either the maximum or minimum full resolution elevation value is whether there is a majority of ridge or stream cells (Figure 31.2), but this is just one option. Other options would include taking into account the locations of the breaklines within the generalized cell, or perhaps using a threshold at which a certain percentage of a generalized cell must be classified as ridge or stream before it is handled as such. Selection of the generalized cell value could also be done differently than simply using the maximum or minimum value. For example, a weighted average of the values of the full resolution breakline cells could be used instead of just the maximum or minimum. Another modification to the procedure would be to use an alternative method of identifying the significant topographic features to be emphasized. A more sophisticated landform description that depicts concave, convex, and horizontal and

sloping flat areas in addition to ridge and stream lines, such as that described by Blaszczynski (1997), could be used to guide selection of generalized values.

The need for effective DEM generalization techniques will remain strong as better quality and higher resolution data become more available, and as applications of DEM data become more widespread. Newer gridding algorithms, such as that developed by Hutchinson (1989), produce source DEMs that are already optimized for hydrologic applications, so use of morphology-based generalization schemes appropriately retain the increased topographic information inherent to the original DEM. As coverage of higher resolution elevation data increases, DEMs at coarser grid spacings will be routinely produced through generalization. One such example is the data to be generated from the Shuttle Radar Topography Mission (SRTM), scheduled for launch in September 1999. The near global 100-meter terrain data to be produced from SRTM data will be an excellent source for producing generalized DEM's for continental scale hydrologic studies.

ACKNOWLEDGMENT

This work was performed under U.S. Geological Survey contract no. 1434-CR-97-CN-40274.

REFERENCES

Band, L.E. Topographic Partition of Watersheds with Digital Elevation Models, *Water Resour. Res.*, 22, pp. 15–24, 1986.

Band, L.E. A Terrain-Based Watershed Information System, *Hydrol. Proc.*, 3, pp. 151–162, 1989.

Bindlish, R. and A.P. Barros. Aggregation of Digital Terrain Data Using a Modified Fractal Interpolation Scheme, *Comput. Geosci.*, 22, pp. 907–917, 1996.

Blaszczynski, J.S. Landform Characterization with Geographic Information Systems, *Photogrammetric Eng. Remote Sensing*, 63, pp. 183–191, 1997.

Chang, K. and B. Tsai. The Effect of DEM Resolution on Slope and Aspect Mapping, *Cartogr. Geogr. Inf. Syst.*, 18, pp. 69–77, 1991.

Dewald, T.G., S.A. Hanson, L.D. McKay, and W.D. Wheaton. Managing Watershed Data with the USEPA Reach File, in *Proceedings, Watershed '96 Technical Conference and Exposition, Baltimore, Maryland, June 8–12, 1996, Water Environment Federation*, Alexandria, VA, 1996. http://www.epa.gov/OWOW/watershed/Proceed/

Fairfield, J. and P. Leymarie. Drainage Networks from Grid Digital Elevation Models, *Water Resour. Res.*, 27, pp. 709–717, 1991.

Fisher, P. The Effect of Database Generalization on the Accuracy of the Viewshed, in *Spatial Accuracy Assessment in Natural Resources and Environmental Sciences: Second International Symposium*, May 21–23, 1996, Fort Collins, CO, H.T. Mowrer, R.L. Czaplewski, and R.H. Hamre, Eds., General Technical Report RM-GTR-277. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, 1996, p. 272–280.

Gesch, D.B. and K.S. Larson. Techniques for Development of Global 1-km Digital Elevation Models, in *Pecora Thirteen, Human Interactions with the Environment: Perspectives from Space, Sioux Falls South Dakota, August 20–22, 1996, Proceedings*, American Society for Photogrammetry and Remote Sensing, Bethesda, MD (in press). http://edcwww.cr.usgs.gov/pecora/contents.html

Hutchinson, M.F. A New Procedure for Gridding Elevation Data and Stream Line Data with Automatic Removal of Spurious Pits, *J. Hydrol.*, 106, pp. 211–232, 1989.

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

Jenson, S.K. Application of Hydrologic Information Automatically Extracted from Digital Elevation Models, *Hydrol. Proc.*, 5, pp. 31–44, 1991.

Maidment, D.R. GIS and Hydrologic Modeling, in *Environmental Modeling with GIS*, M.F. Goodchild, B.O. Parks, and L.T. Steyaert, Eds., Oxford University Press, New York, 1993, p. 147–167.

Mark, D. Automated Detection of Drainage Networks from Digital Elevation Models, *Cartographica*, 21, pp. 168–178, 1984.

McAlister, E., N. Domburg, T. Edwards, and R. Ferrier. Hydrological Modelling of the River Ythan Using ARC/INFO GRID, in *Proceedings, Sixteenth Annual ESRI User Conference*, May 20–24, 1996. Environmental Systems Research Institute, Inc., Redlands, California, 1996. http://www.esri.com/base/common/userconf/proc96/WELCOME.HTM

Moore, I.D., R.B. Grayson, and A.R. Ladson. Digital Terrain Modeling: A Review of Hydrological, Geomorphological, and Biological Applications, *Hydrol. Proc.*, 5, pp. 3–30, 1991.

Seaber, P.R. and F.P. Kapinos. Hydrologic Unit Maps, *U.S. Geological Survey Water-Supply Paper* 2294, 1987.

Verdin, K.L. and S.K. Jenson. Development of Continental Scale Digital Elevation Models and Extraction of Hydrographic Features, in *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, NM, January 21–26, 1996. National Center for Geographic Information and Analysis, Santa Barbara, CA, 1996. http:/

/www.ncgia.ucsb.edu/conf/SANTA_FE_CD_ROM/main.html

Vieux, B.E. DEM Aggregation and Smoothing Effects on Surface Runoff Modeling, in *Wetland and Environmental Applications of GIS*, J.G. Lyon and J. McCarthy, Eds., Lewis Publishers, Boca Raton, FL, 1995, pp. 205–229.

Weibel, R. Models and Experiments for Adaptive Computer-Assisted Terrain Generalization, *Cartogr. Geogr. Inf. Syst.*, 19, pp. 133–153, 1992.

Wolock, D.M. and C.V. Price. Effects of Digital Elevation Model Map Scale and Data Resolution on a Topography-Based Watershed Model, *Water Resour. Res.*, 30, pp. 3041–3052,1994.

Zeverbergen, L.W. and C.R. Thorne. Quantitative Analysis of Land Surface Topography, *Earth Surface Processes and Landforms*, 12, pp. 47–56, 1987.