WATER RESOURCES RESEARCH, VOL. 34, NO. 4, PAGES 897–901, APRIL 1998

Extraction and representation of nested catchment areas from digital elevation models in lake-dominated topography

D. Scott Mackay

Department of Forest Ecology and Management, and Institute for Environmental Studies
University of Wisconsin–Madison

Lawrence E. Band[1]

Department of Geography, University of Toronto, Toronto, Ontario, Canada

Abstract. This paper presents a new method for extracting flow directions, contributing (upslope) areas, and nested catchments from digital elevation models in lake-dominated areas. Existing tools for acquiring descriptive variables of the topography, such as surface flow directions and contributing areas, were developed for moderate to steep topography. These tools are typically difficult to apply in gentle topography owing to limitations in explicitly handling lakes and other flat areas. This paper addresses the problem of accurately representing general topographic features by first identifying distinguishing features, such as lakes, in gentle topography areas and then using these features to guide the search for topographic flow directions and catchment marking. Lakes are explicitly represented in the topology of a watershed for use in water routing. Nonlake flat features help guide the search for topographic flow directions in areas of low signal to noise. This combined feature-based and grid-based search for topographic features yields improved contributing areas and watershed boundaries where there are lakes and other flat areas. Lakes are easily classified from remotely sensed imagery, which makes automated representation of lakes as subsystems within a watershed system tractable with widely available data sets.

1. Introduction

1.1. Problem

This paper presents a new method to extract and represent watersheds with lakes and other flat areas from digital elevation models (DEMs). A large proportion of high-latitude areas are for the most part of flat to moderate relief, modulated by the effects of glacial erosion and deposition. Identification of flow directions, contributing areas, and nested catchments is needed to understand topographic controls on water, carbon, nutrient, and sediment flows within and over full watersheds. Existing automated methods for determining flow directions and catchment areas from DEMs have been designed and operated in moderate to steep topography using "pitless" DEMs [e.g., *O'Callaghan and Mark*, 1984; *Band*, 1986a, 1989; *Jenson and Domingue*, 1988; *Quinn et al.*, 1991; *Martz and Garbrecht*, 1992; *Costas-Cabral and Burges*, 1994; *Tarboton*, 1997].

In order to understand the role of topography on processes over large, heterogeneous areas, it is important that automated catchment algorithms be applicable to nonmountainous areas. Current algorithms determine flow directions between adjacent grid cells. These methods are susceptible to errors on the DEM, particularly in relatively flat areas where true topographic relief and DEM vertical error are comparable in magnitude. In addition, they typically involve inferring local flow directions in some areas, such as treating lakes as wide streams rather than explicitly defining these areas as hydrologically distinct features. A different approach, which is suggested here, is to route flow for lakes using different algorithms from the ones used for topographic flow routing.

This paper presents a general method for extracting and representing watersheds in both mountainous and nonmountainous areas. Features such as lakes are explicitly represented and included as contiguous surface elements during flow path analysis. These surface elements comprise one or more connected DEM grid cells. They are used to determine flow connectivity in areas where there is little or no topographic variability, such as the surface of a lake, or through other flat areas. The method should be applicable to any of the grid-based methods used to extract flow connectivity and catchment areas on DEMs.

The explicit accounting of flat and low signal to noise areas in the identification of surface drainage features from DEMs is the main contribution of this paper. Water bodies and low signal to noise areas are defined from remotely sensed imagery and DEM analysis prior to drainage identification. During subsequent drainage identification, these special areas are individually processed with optimized parameters, which results in a more accurate drainage pattern in and around these problematic areas.

1.2. Background

This paper addresses grid-based DEMs, which are widely available and used in flow path and contributing area calculations. The first of these flow path algorithms were the steepest descent, or D8, algorithms of *O'Callaghan and Mark* [1984] and

---

[1] Now at Department of Geography, University of North Carolina, Chapel Hill.

Paper number 98WR00094.
0043-1397/98/98WR-00094$09.00

Figure 1. Flat objects used to guide search for flow paths. Dark objects are classified as water bodies.

Marks et al. [1984]. D8 has been widely used to partition watersheds into subcatchment areas [*Band*, 1986a; *Jenson and Domingue*, 1988; *Tarboton et al.*, 1991], as well as in calculating contributing area [*Band*, 1989; *Ehlschlaeger*, 1989; *Morris and Heerdegen*, 1988; *Lammers and Band*, 1990; *Jenson and Domingue*, 1988; *Martz and Garbrecht*, 1992]. Fractional, or F8, flow algorithms partition flow from a cell to all its eight neighbors by weighting flow according to relative slope [*Freeman*, 1991; *Quinn et al.*, 1991]. Uncertainty associated with F8 weighting schemes prompted development of a flow tube analogy in which flow across a planar surface is resolved for each cell using both aspect and gradient of the plane [*Lea*, 1992; *Costa-Cabral and Burges*, 1994; *Tarboton*, 1997].

Regardless of how flow is routed, all cell-based algorithms attempt to find surface flow directions and upslope areas either during an ascent from concave points on the DEM or during descent from convex points on the DEM. The information gathered by these algorithms is affected by error in the DEM, such as pits or dams. Pits and dams occur as a result of insufficient or missing data during DEM production. In small, steep watersheds, pits and dams are usually negligible because of the high local topographic relief. However, in flatter areas of larger watersheds or in lake-dominated areas, pit depths and dam heights often exceed local true elevation differences. This results in a number of pathological drainage conditions, including gaps [*Chorowicz et al.*, 1992; *Costas-Cabral and Burges*, 1994] and loops [*Band*, 1989; *Smith et al.*, 1990].

A number of simple rules have been devised to overcome gaps and loops, including pit filling [*Marks et al.*, 1984; *Band*, 1986a; *Jenson and Domingue*, 1988; *Martz and Garbrecht*, 1992], dam breaching [*Garbrecht and Martz*, 1996], and slope tolerances [*Band*, 1989]. During pit filling, a surface is formed by filling a pit to some new pour height. This produces a surface through which flow paths can be inferred from the surrounding topography [*Garbrecht and Martz*, 1997]. Slope tolerances permit flow connections as long as the slope gradient of a cell is below some threshold value. They also allows for catchment area spillage over shallow divides and into adjacent catchments, and so simple rules should be used with caution.

By combining lake and other flat features with cell-based flow path algorithms, an overall improvement in representation of lake-dominated watersheds is anticipated. Here the method is demonstrated for both D8 [*Band*, 1989] and F8 [*Freeman*, 1991; *Quinn et al.*, 1991] approaches and is tested in a lake-dominated watershed in the Algoma Highlands of Ontario, Canada.

2. Methodology

2.1. Identification of Features

The approach begins with the identification of all flat features. Two distinct types of features are identified: (1) lakes and (2) relatively flat areas on the DEM in which slope tracking is likely to fail. Identification of the second type of feature can take into consideration the specific tolerances used by a given flow path algorithm. Every such algorithm uses some technique to avoid problems during in relatively flat areas on a DEM. For instance, to prevent the upslope climb from failing in low signal to noise areas, such as in a wide flood plain, *Band* [1989] incorporated a slope tolerance σ, following *Marks et al.* [1984]. Slope β is computed for all grid cells on the DEM using the method reported by *Lammers and Band* [1990]. A flow path is formed between a given cell and any of its neighbors when $\beta \leq \sigma$. *Band* [1989] uses this slope threshold for the whole watershed. As σ is raised, an increasing number of flow paths tend to spill over divides in addition to climbing out of low signal-to-noise areas. It is conceivable that the magnitude of slope tolerance needed for successful climbing may change with position on a DEM as a result of differences in signal-to-noise ratio.

To accommodate this spatially variable information content in the DEM, algorithms were developed to identify flat areas, classify them by cover type (e.g., water or land), and optimally assign the lowest threshold that permits full marking of each of these areas. A cell is considered to be flat if it has a slope gradient below σ. This identifies a priori the grid cells that would be marked during flow path analysis as a result of satisfying the slope tolerance. A limitation of this approach is that an initial slope threshold must be provided. The initial value is then optimally adjusted for each collection of flat cells during slope tracking. If the initial σ is too low, the size of the flat region may be underestimated, and slope tracking may still fail. However, if σ is much higher than the optimally adjusted threshold, then much of the initial flat region will simply be marked as regular topography during slope tracking. Thus an initially high estimate of σ will generally produce a more complete watershed marking.

Contiguous groups of flat areas are formed into labeled regions by using local region growing [*Ballard and Brown*, 1982]. The labeled regions are classified into water bodies or land areas using supervised classification of remotely sensed imagery. Water has a very low near-infrared reflectance compared with land areas (e.g., forested cover), and so simple image classification algorithms produce good separation between water and land. Figure 1 illustrates the features derived from a combination of a DEM using a σ of 2 degrees to identify flat areas, and Landsat Thematic Mapper band 4 (near-infrared) satellite imagery to distinguish water from land areas.

2.2. Cell-Based Analysis

Cell-based algorithms compute grid-to-grid and grid-to-feature flow directions. Features provide a priori terrain knowledge that allows the cell-based algorithms to adapt to the flatter, problematic areas on the DEM. Different cell-based algorithms can be associated with each class of flat feature. For lakes the goal is to deliver the total upslope area contributed to the lake and the lake area itself to the lake outlet. Since actual flow of water through a lake requires additional information, such as lake basin and outlet channel geometry, a lake boundary-following procedure is used (Figure 2). This algorithm preserves surface topology by connecting the lake feature to its incoming flow paths and outgoing flow path, while avoiding the need for working its way through the flat lake area. Each grid cell along the boundary acts as a depression point for catchments that drain into the lake. Contributing area assigned to the lake outlet cell is the total area marked in the climb from this cell plus the area of the unmarked cells in the interior of the lake, shown as the dark grey area in Figure 2. The intermediate flow paths around the lake boundary are later dissolved when the watershed is partitioned into subcatchments, streams, and lakes. This retains the topology of the land-lake features, but eliminates cell-to-cell flow within the lake itself. Flow within the lake could be added, given data on lake basin and outlet channel geometry, but is beyond the scope of this paper.

Streams and subcatchment areas are marked using the methods of *Band* [1989], in which stream segments are topologically ordered with their respective subcatchment areas labeled as left and right polygons. Grid cells on the lake boundary are masked out by the lake feature, and so they are excluded from the subcatchment partition data layer. However, lakes are topologically linked in a watershed database such that incoming and outgoing streams and adjacent subcatchments are stored for each lake feature.

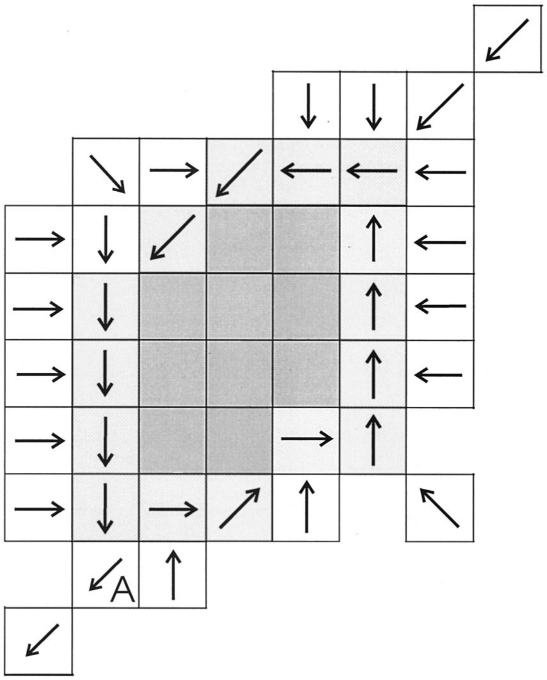

Figure 2. Diagram showing how flow paths are directed along the boundary (light grey) of a lake in order to topologically link the lake to its incoming flow paths and outgoing flow path. The total area accumulated at the outlet cell A is given by the upslope area traversed plus the interior area (dark grey) of the lake.

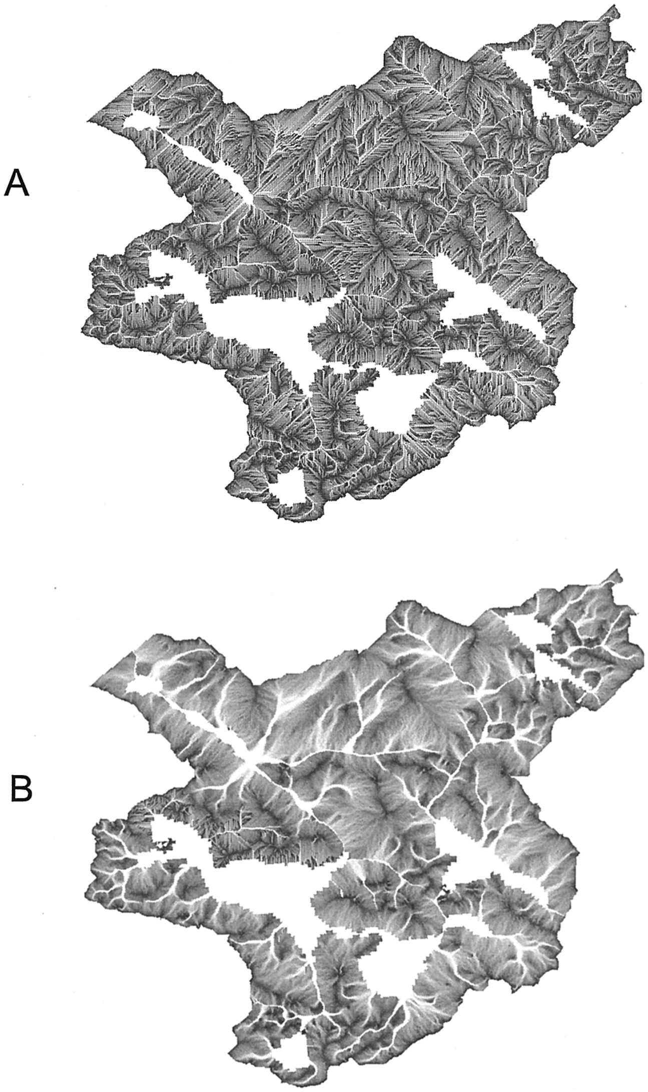

Figure 3. (a) The full D8 upslope contributing area for Turkey Lakes Watershed. (b) The full F8 upslope contributing area.

Another important modification is assigning to each nonlake labeled region k a unique slope threshold σ_k that is optimally determined [*Liang and Mackay*, 1997]. In brief, optimization is performed on each region in order to (1) make σ_k large enough to mark k and (2) minimize spillage over divides. Since some flat regions straddle divides, the algorithm must optimize for constraint 1 in such a way that these flat areas are marked but that their grid cells are included in the appropriate catchments. Since spillage over divides produces excessively high drainage areas organized in large clumps, a reasonable approach is to compute σ_k values that minimize differences in the spatial arrangement of contributing areas between labeled regions and the respective surrounding nonflat, high signal to noise areas. *Liang and Mackay* [1997] found an optimal solution with an algorithm that iteratively lowers or raises σ_k until a difference of means (Student's T) test of mean contributing area within k to mean contributing area outside k fails. The algorithm stops iterating when it accepts the null hypothesis

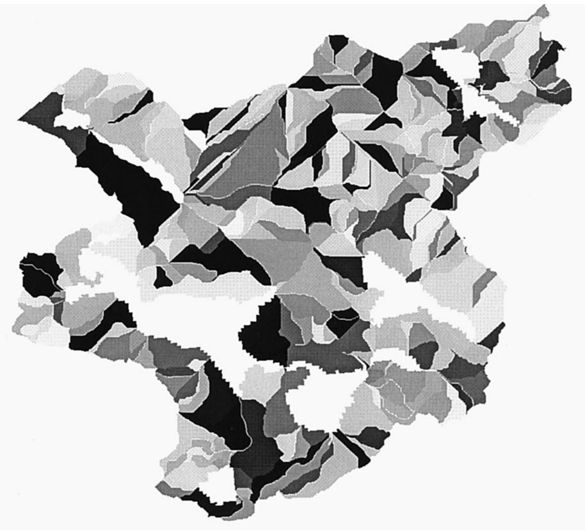

Figure 4. Full watershed partitioning for Turkey Lakes Watershed, showing streams, hillslopes, and lakes. Many of the hillslopes are directly connected to lakes.

that mean contributing areas are not significantly different. This yields slope thresholds high enough to push through flat features but low enough to split the features that straddle divides into the appropriate catchments.

It should be noted that the efficiency of the difference of means test can be quite low in valley bottom areas owing to the high contributing areas. However, the test is needed only along divides where accumulated areas are low. The algorithm also does not require that *k* regions not be completely flat, but only that they have sufficiently low gradients to be of low signal to noise. Features that fail to completely mark during the optimization process after a fixed number of iterations are marked using the lake boundary following algorithm. An alternative solution could use a *Garbrecht and Martz* [1997] or a similar algorithm in these areas, since they are identified prior to slope tracking.

2.3. Testing and Results

The new algorithms were tested on a DEM for the Turkey Lakes Watershed (TLW), a 10.5-km^2 watershed in the Algoma Highlands near Sault Ste. Marie, Ontario, Canada. TLW contains a nested chain of lakes. The lake and other flat features are shown in Figure 1. Figure 3 shows the contributing areas using the D8 and the approaches, marked from an outlet selected from several rows and columns inside the DEM boundary. The algorithm will also mark all watersheds starting from the edge of the DEM but is used here to select only TLW. Figure 4 shows a full watershed partitioning into lakes, streams, and subcatchments, as an extension to the watershed database described by *Band* [1989] and *Lammers and Band* [1990].

Computed contributing areas were compared with field-measured [*Jeffries et al*., 1988] areas for each lake basin and the TLW outlet (Table 1). There is a good correspondence between the observed and computed areas. Deviations in accumulated area are not proportional to the area extracted but are generally underestimated. Subtle variations in topography along divides may account for the larger field-surveyed areas. This is consistent with previously studied discrepancies between field-surveyed stream networks and stream networks derived from contour maps [*Mark*, 1984]. The important point to be made about the nested set of areas is that the algorithms seem to perform reasonably well even as a greater number of lakes and flat areas are added to the accumulated upslope area. Results using a single slope threshold and no lake handling result in either an underestimation of the TLW area by about 50% (σ = 2 degrees) or an overestimation of area by 20% (σ = 4 degrees).

Table 1. Comparison of Lake Catchment Areas Generated Using the Automated Method Reported in This Paper to Areas Measured in the Field

Catchment	Observed*	Computed	Difference, %†
South Batchawana Lake	85.6	87.6	+2.3
Wishart Lake	337.0	333.4	−1.1
Lower Turkey Lake	491.0	464.6	−5.0
Turkey Lake	803.0	786.1	−2.0
Turkey Lakes Watershed	1050.0	1052.9	+0.3

*Source: *Jeffries et al*. [1988].

†Calculated as (computed-observed)/observed.

3. Summary

This paper has addressed some shortcomings of existing grid-based watershed extraction and representation algorithms for low-relief areas by identifying distinguishing flat areas as features prior to cell-based flow path analysis. Previous image processing approaches to geomorphological feature extraction from DEMs demonstrated the potential for feature-based approaches [e.g., *Toriwaki and Fukumura*, 1978; *Band*, 1986b]. However, feature-based approaches have not been widely adopted for flow path and contributing area calculation algorithms. Here it has been demonstrated that a feature-based approach provides a way of automatically selecting suitable algorithms to handle different parts of watersheds dominated by lakes and other flat areas. When a feature is encountered, the effective neighborhood or search space is expanded to the full area of the feature, rather than being limited to the immediate neighbors of a grid cell. The algorithms adapt to the DEM by identifying lakes as salient features of the topography. Since lakes are distinguished from streams, all linkages between the lakes and the surrounding terrain can be explicitly accounted for in constructing watershed databases.

By also identifying nonlake flat areas on the DEM, it was possible to optimize a set of slope thresholds that allow for a full watershed marking with a minimal amount of error being introduced by the algorithms. An important contribution of the spatially variable slope thresholds is that a locally high slope threshold has no effect on other areas. This enables the watershed extraction and representation algorithms to adapt to the variable information content of the DEM.

This paper has focused on lakes and other flat areas that pose problems for flow path and contributing area algorithms. However, the method can be generalized to other watershed features, as it assumes only that the features be defined prior to cell-based analysis. Since the approach presented in this paper uses specific knowledge of known features within an area, it is potentially more adaptable to both application requirements and limitations in data quality.

Acknowledgments. This research was partially supported by a Tri-Council Greenplan EcoResearch Doctoral Fellowship to D.S.M. while he attended the University of Toronto. Further funding was provided by the UW-Madison Graduate School, College of Agriculture and Life Sciences, and Institute for Environmental Studies. Funding to L.E.B. was provided by NASA, NSERC, and the Ontario Ministry of Natural Resources. The Turkey Lakes Watershed project has been supported by Environment Canada and Forestry Canada. A contour map of the Turkey Lakes Watershed, from which the DEM used in this paper was generated, was generously provided by NWRI, Environment Canada.

References

Ballard, D. H., and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, N. J., 1982.

Band, L. E., Analysis and representation of drainage basin structure with digital elevation data, in *Proceedings of the Second International Conference on Spatial Data Handling*, pp. 437–450, Int. Geogr. Union, Williamsville, N. Y., 1986a.

Band, L. E., Topographic partition of watersheds with digital elevation models, *Water Resour. Res.*, *23*(1), 15–24, 1986b.

Band, L. E., A terrain-based watershed information system, *Hydrol. Processes*, *3*, 151–162, 1989.

Chorowicz, J., C. Ichoku, S. Riazanoff, Y.-K. Kim, and B. Cervelle, A combined algorithm for automated drainage network extraction, *Water Resour. Res.*, *28*(5), 1293–1302, 1992.

Costa-Cabral, M., and S. J. Burges, Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas, *Water Resour. Res.*, *30*, 1681–1692, 1994.

Ehlschlaeger, C., Using the A* search algorithm to develop hydrologic models from digital elevation data, paper presented at Proceedings of International Geographic Information Systems (IGISD) Symposium '89, Baltimore, MD, March 18–19, 1989.

Freeman, T. G., Calculating catchment area with divergent flow based on a regular grid, *Comput. Geosci.*, *17*, 413–422, 1991.

Garbrecht, J., and L. W. Martz, Digital landscape parameterization for hydrological applications, in *HydroGIS 96: Applications of Geographic Information Systems in Hydrology and Water Resources Management*, *IAHS Publ.*, *235*, 1996.

Garbrecht, J., and L. W. Martz, The assignment of drainage direction over flat surfaces in raster digital elevation models, *J. Hydrol.*, *193*, 204–213, 1997.

Jeffries, D. S., J. R. M. Kelso, and I. K. Morrison, Physical, chemical, and biological characteristics of the Turkey Lakes Watershed, central Ontario, Canada, *Can. J. Fish. Aquat. Sci.*, *45*, Suppl. 1, 3–13, 1988.

Jenson, S. K., and J. O. Domingue, Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogramm. Eng. Remote Sens.*, *54*, 1593–1600, 1988.

Lammers, R. B., and L. E. Band, Automating object representation of drainage basins, *Comput. Geosci.*, *16*(6), 787–810, 1990.

Lea, N. L., An aspect driven kinematic routing algorithm, in *Overland Flow: Hydraulics and Erosion Mechanics*, edited by A. J. Parsons and A. D. Abrahams, Chapman and Hall, New York, 1992.

Liang, C., and D. S. Mackay, Feature-based optimization of flow directions and upslope areas on grid-based digital elevation models, in *Proceedings of GIS/LIS'97*, pp. 45–52, Am. Soc. for Photogramm. and Remote Sens., Bethesda, Md., 1997.

Mark, D. M., Relations between field surveyed channel networks and map-based geomorphometric measures, *Ann. Am. Assoc. Geogr.*, *73*, 358–372, 1984.

Marks, D., J. Dozier, and J. Frew, Automated basin delineation from digital elevation data, *GeoProcessing*, *2*, 299–311, 1984.

Martz, L. W., and J. Garbrecht, Numerical definition of drainage network and subcatchment areas from digital elevation models, *Comput. Geosci.*, *18*(6), 747–761, 1992.

Moore, I. D., E. M. O'Loughlan, and G. J. Burch, A contour-based topographic model for hydrological and ecological applications, *Earth Surf. Processes Landforms*, *13*, 305–320, 1988.

Morris, D. G., and R. G. Heerdegen, Automatically derived catchment boundaries and channel networks and their hydrological applications, *Geomorphology*, *1*, 131–141, 1988.

O'Callaghan, J. F., and D. M. Mark, The extraction of drainage networks from digital elevation data, *Comput. Vision Graphics Image Process.*, *28*, 323–344, 1984.

Quinn, P., K. Beven, P. Chevalier, and O. Planchon, The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrol. Processes*, *5*, 59–79, 1991.

Smith, T. R., C. Zhan, and P. Gao, A knowledge-based, two-step procedure for extracting channel networks from noisy DEM data, *Comput. Geosci.*, *16*(6), 777–786, 1990.

Tarboton, D. G., A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resour. Res.*, *33*(2), 309–319, 1997.

Tarboton, D. G., R. J. Bras, and I. Rodriguez-Iturbe, On the extraction of channel networks from digital elevation data, *Hydrol. Processes*, *5*(1), 81–100, 1991.

Toriwaki, J., and T. Fukumura, Extraction of structural information from grey pictures, *Comput. Graphics Image Process.*, *7*, 30–51, 1978.

L. E. Band, Department of Geography, University of North Carolina, Chapel Hill, NC 27599.

D. S. Mackay, Department of Forest Ecology and Management, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706. (e-mail: dsmackay@facstaff.wisc.edu)

(Received August 26, 1997; revised December 30, 1997;
accepted January 7, 1998.)