

# A comparison of geographical information systems–based algorithms for computing the TOPMODEL topographic index

Feifei Pan

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Christa D. Peters-Lidard

Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Michael J. Sale and Anthony W. King

Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Received 2 February 2004; revised 1 March 2004; accepted 20 April 2004; published 24 June 2004.

[1] The performance of six geographical information systems (GIS)-based topographic index algorithms is evaluated by computing root-mean-square errors of the computed and the theoretical topographic indices of three idealized hillslopes: planar, convergent, and divergent. In addition to these three idealized cases, two divergent hillslopes with varying slopes, i.e., concave (slopes decrease from top to bottom) and convex (slopes increase from top to bottom) are also tested. The six GIS-based topographic index algorithms are combinations of flow direction and slope algorithms: i.e., single flow direction (SFD), biflow direction (BFD), and multiple flow direction (MFD) plus methods that determine slope values in flat areas, e.g., W-M method [Wolock and McCabe, 1995] and tracking flow direction (TFD) method. Two combinations of horizontal resolution and vertical resolution of the idealized terrain data are used to evaluate those methods. Among those algorithms the MFD algorithm is the most accurate followed by the BFD algorithm and the SFD algorithm. As the vertical resolution increases, the errors in the computed topographic index for all algorithms decrease. We found that the orientation of the contour lines of planar hillslopes significantly influences the SFD's computed topographic index. If the contour lines are not parallel to one of eight possible flow directions, the errors in the SFD's computed topographic index are significant. If mean slope is small, TFD becomes more accurate because slope values in flat areas are better estimated. *INDEX TERMS*: 1899 Hydrology: General or miscellaneous; 1824 Hydrology: Geomorphology (1625); 1832 Hydrology: Groundwater transport; *KEYWORDS*: GIS, TOPMODEL, topographic index, single flow direction algorithm, biflow direction algorithm, multiple flow direction algorithm

**Citation:** Pan, F., C. D. Peters-Lidard, M. J. Sale, and A. W. King (2004), A comparison of geographical information systems–based algorithms for computing the TOPMODEL topographic index, *Water Resour. Res.*, 40, W06303, doi:10.1029/2004WR003069.

## 1. Introduction

[2] TOPMODEL is a topography-based concept for watershed hydrology modeling. Since the TOPMODEL was first proposed in 1979 [Beven and Kirkby, 1979], it has been widely used to study the effects of topography on hydraulic processes including flood frequency, streamflow generation, flow paths, geomorphic characteristics, and water quality [Wolock and McCabe, 1995]. In addition to the success of the TOPMODEL concept in traditional hydrologic modeling, this concept has also been successfully incorporated into several ecosystem-atmosphere models including the Regional Hydro-Ecological Simulation System (RHESys) [e.g., Ford et al., 1994]; the Land Ecosystem-Atmosphere Feedback model (LEAF-2) [Walko et al., 2000]; the TOPMODEL-based Land Atmosphere Transfer Scheme

(TOPLATS) [Famiglietti and Wood, 1994; Peters-Lidard et al., 1997]; the Catchment Model [Koster et al., 2000]; and the Common Land Model (CLM) [Dai et al., 2003].

[3] To apply the TOPMODEL, a modeled catchment is partitioned with a regular grid or lattice. The so-called 'topographic index' is then calculated for each cell in the catchment. The topographic index,  $\ln(a/\tan \beta)$ , is the natural logarithm of the ratio of the specific flow accumulation area  $a$  to the ground surface slope  $\tan \beta$ . The surface slope can be evaluated from digital elevation model (DEM) data. The specific flow accumulation area is the total flow accumulation area (or upslope area)  $A$  through a unit contour length  $L$ . To compute the total flow accumulation area  $A$ , flow directions are tracked upslope, starting from the cell of interest to the upstream divide of the watershed, and then tracked downslope accumulating cells contributing to the drainage area of the cell of interest. Here we note an uncertainty associated with the definition of the flow accumulation area  $A$ . From the presentations of Beven and

Wood [1983, Figure 2] and Kirkby [1997, Figure 1], one can reasonably conclude that the flow accumulation area is defined along the ground surface. However, using DEMs in geographical information systems (GIS), the computed flow accumulation area is generally the area projected to x-y plane, and this calculation of  $A$  has become standard practice. The difference between these two areas is negligible if the slope is less than 0.5 (m/m), and most of the slopes in the watersheds to which the TOPMODEL is applied are less than 0.5 (m/m) [e.g., Montgomery and Dietrich, 1992]. Therefore, for consistency with standard practice, we adopt the convention of calculating the flow accumulation area or upslope area as the area projected to x-y plane.

[4] The slope term ( $\tan \beta$ ) in the topographic index arises from the assumption that the surface of the water table is parallel to the ground surface. Thus the local hydraulic gradient is assumed to be equal to the slope of the ground. Because flow direction depends on hydraulic gradient, or ground surface slope, flow direction and the calculation of the upslope accumulation area should be consistent with the local slope value that is used to compute  $\ln(a/\tan \beta)$ . Thus the computed topographic index is dependent upon the calculation of both slope and flow direction.

[5] Although many researchers have investigated algorithms for calculation of slope and flow direction, those studies tend to focus on either slope [e.g., Jones, 1998; Zhang et al., 1999] or flow accumulation area [e.g., Tarboton, 1997; Rieger, 1998]. Few studies have examined the combined effects of slope and flow direction algorithms on the topographic index. The works of Quinn et al. [1991], Wolock and McCabe [1995], and Mendicino and Sole [1997] are exceptions. However, these studies only investigated the difference in the statistical moments or distributions of the computed topographic index; no comparisons of errors between the “true,” i.e., analytically solved, and the numerically computed topographic indices were carried out. Comparing against “truth” is difficult because an analytical expression for the real terrain does not normally exist. Therefore one cannot usually determine which topographic index algorithm is more accurate. This paper seeks to resolve that issue and provide objective evaluation of the appropriate numerical algorithm for calculating the topographic index.

## 2. Topographic Index Algorithms

[6] A topographic index algorithm is actually a combination of two algorithms, one to calculate flow direction and another to calculate slope. Here we describe three commonly used flow direction algorithms that allow flow in one, two, and more than two directions. We identify each topographic index algorithm by the name of the associated flow direction algorithm (e.g., single flow direction, biflow direction, and multiple flow direction), because each flow direction algorithm uses its own algorithm to determine slope.

### 2.1. Single Flow Direction (SFD) Algorithm

[7] The single flow direction (SFD) algorithm [O’Callaghan and Mark, 1984] calculates flow direction as the steepest slope direction, which is determined by the Maximum Downward Gradient (MDG). This SFD algorithm, also known as the D8 algorithm, is widely used in DEM

data analysis [e.g., O’Callaghan and Mark, 1984; Band, 1986; Greenlee, 1987; Mark, 1988; Jenson and Domingue, 1988] and GIS software (e.g., the “FLOWDIRECTION” function in ARC/INFO GRID).

[8] MDG computes the downhill elevation gradients of a  $3 \times 3$  cell window along eight directions (i.e., elevation of the center cell minus elevation of each of its eight neighbors divided by the distance between those two cells). The slope of the central cell is calculated as the largest of the eight directions. Steepest slope direction is the direction from the central cell to the neighbor generating the largest downhill elevation gradient.

[9] If the central cell has a lower elevation than one of its neighbors, the downhill gradient along this direction is negative. If the calculated slope (i.e., the largest one among eight directions) is less than zero, this cell is called a sink or a pit. At a sink, there is only inflow, and no outflow. However, it is often computationally required to force watersheds to have outlets. Therefore sinks in DEM data are usually filled before watersheds can be delineated and other hydrologic parameters estimated. One common filling approach for sinks generated by MDG is to raise the elevation of a sink to the lowest elevation among its neighbors [e.g., Jenson and Domingue, 1988; Tarboton et al., 1991]. This algorithm is used in the ARC/INFO GRID (i.e., the “FILL” function). For example, the recalculated slope value of the “filled” cell is now zero; the cell and its neighbors form a flat area. To determine flow direction in a flat area, the method suggested by Jenson and Domingue [1988] (e.g., the ARC/INFO FLOWDIRECTION function) is used.

[10] Once the single flow direction is determined for each cell, flow accumulation area (or upslope area) ( $A$ ) is calculated using a recursive procedure [e.g., Tarboton, 1997]. The specific flow accumulation area ( $a$ ) is ( $A$ ) divided by a contour length, which is equal to the grid size or horizontal resolution of the DEM. The slope is set to be the maximum downward elevation gradient.

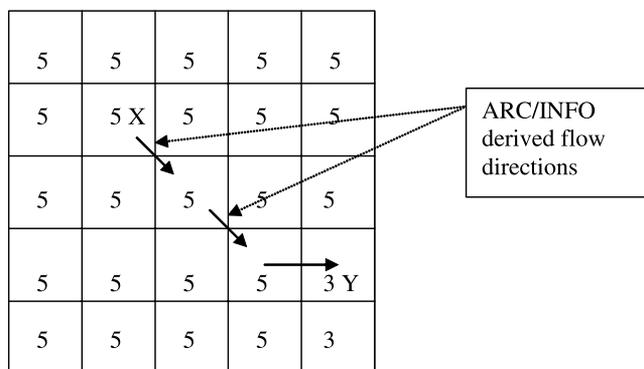
### 2.2. Biflow Direction (BFD) Algorithm

[11] The biflow direction algorithm (BFD), also known as  $D_{\infty}$ , was proposed by Tarboton [1997]. In this algorithm the  $3 \times 3$  cell window is divided into 8 triangular facets. The slope direction and magnitude of each facet are compared. The steepest downward direction is chosen and divided into two directions along the edges forming that facet. The proportion of flow along each edge is inversely proportional to the angle between the steepest downward direction and the edge. Therefore at most two flow directions can be assigned to each cell. The contour length is defined as the grid cell size (DEM resolution), and the slope is set to be the largest slope of 8 facets [Tarboton, 1997].

[12] Like the SFD algorithm, the BFD algorithm has trouble defining flow direction in flat areas. A GIS-based algorithm for returning flow directions [Jenson and Domingue, 1988], such as the ARC/INFO “FLOWDIRECTION” function, can be used to determine the flow direction in flat areas, such that only one flow direction is assigned to each cell in a flat area.

### 2.3. Multiple Flow Direction (MFD) Algorithm

[13] Quinn et al. [1991] first suggested the multiple flow direction (MFD) algorithm to improve representation of the convergence or divergence of flow. Wolock and McCabe



**Figure 1.** An example showing the TFD method for recalculating slope values in flat areas. Numbers in cells are elevation values. Following the ARC/INFO derived flow directions, we identify the closest cell to cell X with a lower elevation is cell Y. The slope value for cell X is computed as the ratio of the elevation difference to the length of flow path between X and Y, which is  $2/(2\sqrt{2}HR + HR)$ , where  $HR$  is the cell size.

[1995] showed how to implement this algorithm using ARC/INFO GRID functions. Unlike the SFD and BFD algorithms, the MFD algorithm allows flow in more than one or two directions. In the MFD algorithm, each flow direction is weighted by the downward elevation gradient (i.e., from the central cell to each of its 8 neighbors) multiplied by a “contour length”. There are two ways to set the contour length: i.e.,  $(HR/2)$  and  $(\sqrt{2}HR/4)$  [Quinn *et al.*, 1991], or  $0.6HR$  and  $0.4HR$  [Wolock and McCabe, 1995], for cardinal and diagonal flow directions, respectively, where  $HR$  is the horizontal resolution,  $0^\circ$  grid cell size of the DEM. Here, we assume the  $HR$  in easting and northing directions are the same (which is true for most DEM data). To differentiate those two algorithms for assigning the contour lengths, we call Quinn’s method the MFD algorithm, and Wolock and McCabe’s algorithm the MFD\* algorithm. The slope is the weighted summation of all positive downward gradients [Quinn *et al.*, 1991]. As for the MFD algorithm, a GIS-based algorithm for returning flow directions [Jenson and Domingue, 1988], such as ARC/INFO flow direction function can be used to determine the flow directions in flat areas.

#### 2.4. Recalculation of Slopes in Flat Areas

[14] Because the topographic index cannot be calculated directly in flat areas, estimating slopes in flat areas is a critical issue. Wolock and McCabe [1995] defined the minimum slope value to be equal to  $(0.5 \times \text{vertical resolution})/(\text{horizontal resolution})$ . We refer to this as the W-M method.

[15] To maintain consistency between flow direction and slope in flat areas, we propose a simple algorithm called the tracking flow direction (TFD) method. This algorithm starts from cell  $X$  and following the flow direction determined by ARC/INFO “FLOWDIRECTION” function that searches for the nearest cell  $Y$  with a lower elevation than cell  $X$ . The slope at cell  $X$  is then set to be the ratio of elevation difference to length of flow path between cells  $X$  and  $Y$  (see Figure 1). If a nonzero slope value cannot be found for the cell over the whole domain, the value of  $-1$  is assigned

to this cell, indicating that slope is undefined. The minimum of all defined slopes is then assigned to each undefined cell.

### 3. Idealized Hillslopes

[16] To compare the computed and theoretical “true” topographic indices, we apply six GIS-based topographic index algorithms to three idealized hillslopes, i.e., planar, divergent and convergent, for which we can calculate the theoretical topographic index analytically. Although those hillslopes are ideal simplifications, they represent three types of real terrain: terrace, divergent and convergent.

[17] For the planar and convergent hillslopes, we use one slope value for each case, i.e., slope is spatially uniform. For the divergent hillslopes, in addition to the constant slope cases, we also test two divergent hillslopes with varying slopes, i.e., concave (slope decreases from top to bottom), and convex (slope increases from top to bottom). Although slopes are not spatially uniform for those concave and convex divergent hillslopes, to simplify the problem, we set slope constant along each contour line. Under such conditions, we can theoretically compute the specific flow accumulation area of each cell along a contour line as the total upslope area divided by the length of the contour line. This calculation is valid only when the contour line is closed for the divergent or convergent hillslopes, or the contour line is infinitely long for the planar hillslopes.

[18] Prior to introducing each idealized case, we first define three domains. The computation domain is defined as the region where we use a recursive algorithm to compute the flow accumulation area based on flow directions determined by the SFD, or BFD or MFD algorithms. We can think of the computational domain as a part of a watershed. However, this part is not just an arbitrary domain. It starts from a divide or a local peak and covers a part of the downstream area, since we will set the top boundary parallel to the contour lines. Therefore, when we compute the flow accumulation area, we do not miss any contribution from upslope crossing the top boundary.

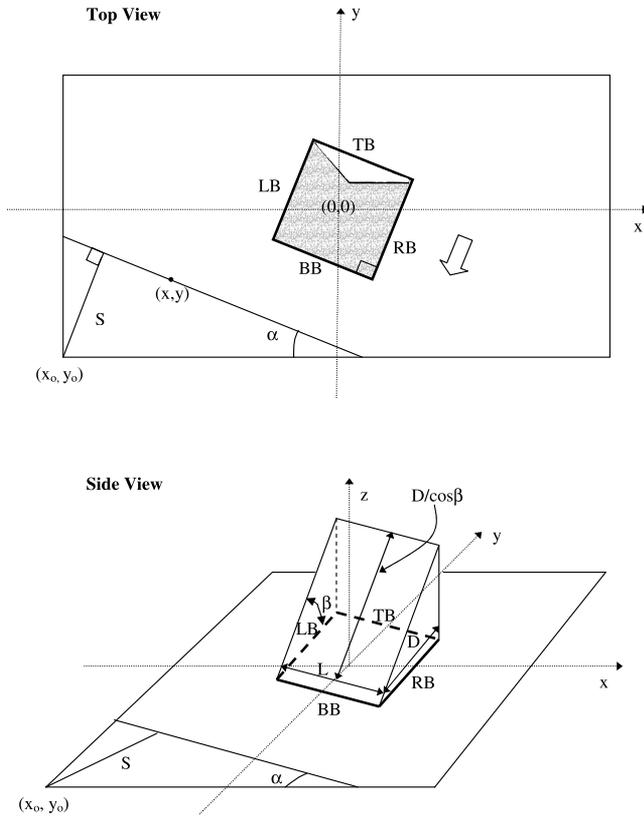
[19] The influence domain is defined as the area inside the computational domain that receives the flow contribution from the lateral boundaries. The comparison domain is inside the computational domain but with the influence domain excluded. Our theoretical calculation of the topographic index is valid only if there is no influence from the lateral boundaries. Therefore, to compare the computed and the theoretical topographic indices properly, we need to carry out such comparison inside the comparison domain, because only inside the comparison domain are the computed flow accumulation area and thus the computed topographic index not influenced by the lateral boundaries.

#### 3.1. Planar Hillslopes

[20] A three-dimensional, idealized planar hillslope is illustrated in Figure 2. In a  $NROW \times NCOL$  domain ( $NROW$  and  $NCOL$  are numbers of rows and columns), the coordinates of each cell are given by:

$$\begin{aligned} x &= [j - 0.5(NCOL - 1)] * HR, j = 0, \dots, NCOL - 1 \\ y &= [0.5 * (NROW - 1) - i] * HR, i = 0, \dots, NROW - 1 \end{aligned} \quad (1)$$

where  $(i, j)$  are row and column indices of each cell,  $HR$  is grid cell size or horizontal resolution of DEM data. In this study, we set  $NROW = 2001$ , and  $NCOL = 4001$ .



**Figure 2.** Top and side views of a planar hillslope. The computational domain is formed by four boundaries (LB, RB, TB, and BB). Inside the computational domain, the unshaded area is the comparison domain, and the shaded area is affected by the lateral boundaries (LB and RB). The block arrow is the normal direction of the contour lines and along the downward elevation gradient direction. The angle between each contour line and x axis is  $\alpha$ ,  $S$  is the distance from point  $(x_0, y_0)$  to a contour line,  $D$  is the distance between a cell and the top boundary TB, and  $\beta$  is slope angle.

[21] The theoretical elevation  $Z_t$  of each cell is given by:

$$Z_t = S * slope = S * \tan \beta \quad (2)$$

where  $\tan \beta$  is slope,  $\beta$  is slope angle ( $0^\circ < \beta < 90^\circ$ ), and  $S$  is the distance from  $(x_0, y_0)$  to the contour line passing cell  $(x, y)$  given by

$$S = \frac{|(x - x_0) \tan \alpha + (y - y_0)|}{\sqrt{(\tan \alpha)^2 + 1}}, \quad (3)$$

$$x_0 = -0.5(NCOL - 1)HR, y_0 = -0.5(NROW - 1)HR,$$

where  $\alpha$  is the angle between each contour line and x axis (from the contour line to x axis in the counterclockwise direction) and we call it contour angle (Figure 2).

[22] To investigate the effects of the vertical resolution (VR) of elevation data on the errors in the computed topographic index, we set elevation  $z$  (which is used for determining flow directions and slopes) as follows:

$$z = \begin{cases} VR * \text{int}(Z_t), & \text{if fractional part of } Z_t < 0.5 \\ VR * [1 + \text{int}(Z_t)], & \text{if fractional part of } Z_t \geq 0.5 \end{cases} \quad (4)$$

where  $\text{int}()$  is a function to convert a floating point value to an integer by cutting off all fractional part. Using expressions shown in equation (4), we can mimic the vertical resolution of real DEM data. For example, a theoretical elevation is 125.253m. According to equation (2), the elevation is 125 m at 1-m VR, or 125.3 m at 0.1-m VR, or 125.25 m at 0.01-m VR.

[23] The computational domain for the planar hillslopes is confined by four boundaries (Figure 2), i.e., TB, BB, LB and RB. TB is called the top boundary and BB is the bottom boundary. They are parallel to the contour lines and associated with the highest and lowest elevations in the computational domain, respectively. LB and RB are two lateral boundaries. Facing the direction of downward gradient of elevation (shown as a block arrow in Figure 2), LB is on the left side, and RB is on the right side. We set the computational domain for the planar hillslopes as follows:

$$\begin{aligned} d_1 &\leq 0.1(NCOL - 1)HR, d_2 \leq 0.1(NCOL - 1)HR, \\ d_3 &\leq 0.1(NCOL - 1)HR, d_4 \leq 0.1(NCOL - 1)HR \end{aligned} \quad (5)$$

where  $d_1, d_2, d_3,$  and  $d_4$  are distances from origin  $(0, 0)$  to four boundaries: LB, RB, TB and BB.

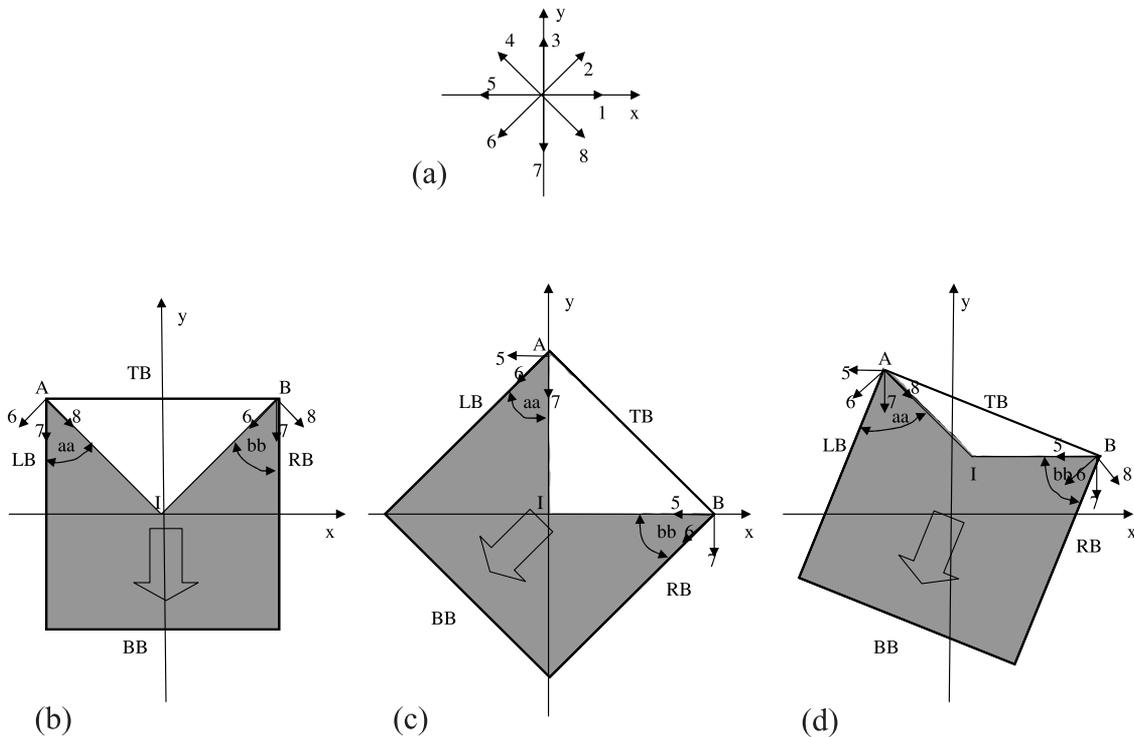
[24] In the computational domain, the theoretical topographic index is given by:

$$\begin{aligned} \ln(a / \tan \beta) &= \ln \left[ \frac{\text{upslope area} / \text{contour length}}{\text{slope}} \right] \\ &= \ln \left[ \frac{D * L}{L \tan \beta} \right] = \ln \left[ \frac{D}{\tan \beta} \right] \end{aligned} \quad (6)$$

where  $D$  is the distance from cell  $(x, y)$  to TB, and  $L$  is the contour length. As we mentioned in section 1, the flow accumulation area of the upslope area is the area projected to x-y plane, i.e.,  $DL$ . Although  $L$  does not appear in the final expression, equation (6) is valid only as  $L$  is infinite, or there is no influence due to the lateral boundaries. However,  $L$  cannot be infinite. Therefore a proper comparison between the computed and theoretical topographic index can only be carried out in a comparison domain where the lateral boundaries have no influence.

[25] Our first step is to remove the influence domain from the computational domain. In order to determine the influence domain, we need to follow flow directions flowing into the computational domain from two corner cells on the top boundary (TB). The MFD algorithm has a larger influence region than the SFD algorithm or the BFD algorithm, because it allows more flow dispersion. Therefore we choose the MFD's comparison domain as the common comparison domain, because it is in both SFD's and BFD's comparison domains.

[26] Figure 3a shows 8 possible flow directions at a cell. The angles between these flow directions and x axis are  $0^\circ$  (direction 1),  $45^\circ$  (direction 2), ..., and  $315^\circ$  (direction 8). Figures 3b, 3c, and 3d illustrate the computational, influence and comparison domains for three cases corresponding to three different contour angles, i.e.,  $0^\circ, 45^\circ$  and  $22.5^\circ$ . The area between LB and AI is the influence domain due to the lateral boundary LB. The angle formed by LB and AI (i.e.,  $aa$ ) is the maximum angle among the angles formed by LB and each flow direction at point A which points to the computational domain. Similarly, we can determine the



**Figure 3.** (a) The 8 flow directions, i.e., 1, 2, . . . , 8. (b) Computational, influence, and comparison domains as the angle between each contour line and x axis is 0°. The computational domain is formed by the top boundary TB, bottom boundary BB, and two lateral boundaries LB and RB. The influence domain is the shaded area formed by LB, BB, RB, BI, and AI. The comparison domain is the unshaded area inside the computational domain with the influence domain excluded. (c) Same as Figure 3b, except the contour angle is 45°. (d) Same as Figure 3b, except the contour angle is 22.5°.

influence domain due to the lateral boundary RB, which is formed by RB and BI. The angle between RB and BI is *bb*. In Figure 3b, both points A and B have three flow directions, i.e., 6, 7, and 8. However, only direction 8 at point A and direction 6 at B point into the computational domain. Therefore angles *aa* and *bb* are equal to 45°. Similarly, in Figure 3c, both A and B have three flow directions, i.e., 5, 6, and 7. Among those directions, only direction 7 at point A and direction 5 at point B flow into the computational domain. Therefore angles *aa* and *bb* are equal to 45°. In Figure 3d, both A and B have four flow directions, i.e., 5, 6, 7 and 8. Among those directions, only directions 7 and 8 at A, and directions 5 and 6 at B flow into the computational domain. Comparing the angles, we can find that angle *aa* is formed by LB and direction 8, and angle *bb* is formed by RB and direction 5. Therefore *aa* and *bb* are equal to 67.5°. Once *aa* and *bb* are determined, we can determine the influence domain and thus the comparison domain.

### 3.2. Divergent Hillslopes

#### 3.2.1. Divergent Hillslopes With Constant Slopes

[27] A top and a side view of a divergent hillslope are illustrated in Figure 4. In a  $NCOL \times NROW$  domain, the coordinates of each cell are the same as the planar hillslopes (i.e., equation (1)). The theoretical elevation  $Z_t$  is given by:

$$Z_t = f_1 * HR - r * slope = f_1 * HR - r * \tan \beta \quad (7)$$

where  $\tan \beta$  is slope,  $\beta$  is slope angle ( $0^\circ < \beta < 90^\circ$ ),  $r = \sqrt{x^2 + y^2}$  is radius of the contour line passing  $(x, y)$ , and  $f_1 * HR$  is the elevation at the center of the domain. In this study, we set  $NCOL = 4001$ ,  $NROW = 2001$ ,  $f_1 = 3000$ , and the computational domain inside the circle (exclude the origin  $(0, 0)$ ) with a radius equal to  $R$  (Figure 4), which satisfies the following conditions:

$$0 < r \leq R \quad (8)$$

where we set  $R = 500HR$  in this paper.

[28] In the computational domain, the theoretical topographic index is given by:

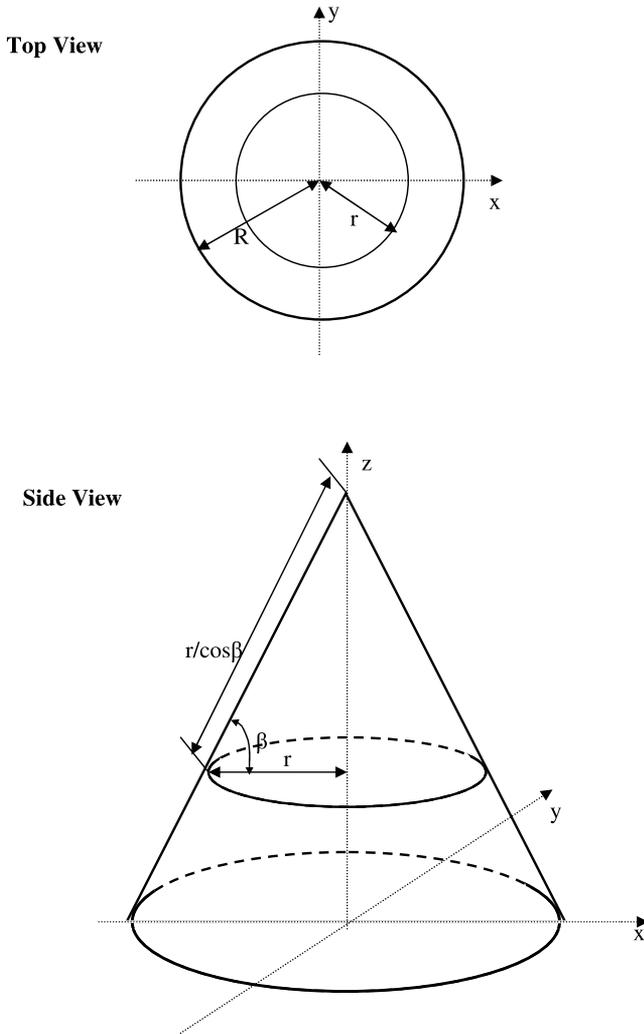
$$\begin{aligned} \ln(a/\tan \beta) &= \ln \left[ \frac{\text{upslope area} / \text{contour length}}{\text{slope}} \right] \\ &= \ln \left[ \frac{\pi r^2}{2\pi r \tan \beta} \right] = \ln \left[ \frac{r}{2 \tan \beta} \right] \end{aligned} \quad (9)$$

The comparison domain is the same as the computational domain, since the contour lines are closed.

#### 3.2.2. Divergent Hillslopes With Varying Slopes

[29] Figure 5 illustrates two divergent hillslopes with varying slopes: slope decreases from top to bottom (concave, Figure 5a) and slope increases from top to bottom (convex, Figure 5b). The theoretical elevation  $Z_t$  is given by

$$\begin{aligned} \text{Concave} & \quad Z_t = c_1(r - c_2)^2 \\ \text{Convex} & \quad Z_t = c_3 - (r/c_4)^2 \end{aligned} \quad (10)$$



**Figure 4.** Top and side views of a divergent hillslope with a constant slope. Both the computational domain and the comparison domain are inside the circle (exclude the origin (0, 0)) with a radius equal to R, and  $\beta$  is slope angle.

where  $c_1, c_2, c_3$  and  $c_4$  are coefficients describing the shapes of the hillslopes. In this study, we set  $c_1 = 10^{-5}$ ,  $c_2 = 2249.49HR$ ,  $c_3 = 3000HR$ ,  $c_4 = 500$ . We set the computational domain to be the same as the divergent cases with constant slopes. In the computational domain, the theoretical topographic index is given by:

$$\begin{aligned} \ln(a/\tan\beta) &= \ln\left[\frac{\text{upslope area} / \text{contour length}}{\text{slope}}\right] \\ &= \ln\left[\frac{\pi r^2}{2\pi r \text{slope}}\right] = \ln\left[\frac{r}{2 \text{slope}}\right] \end{aligned} \quad (11)$$

Slope is computed as:

$$\begin{aligned} \text{Concave} \quad \text{slope} &= \left|\frac{\partial Z_t}{\partial r}\right| = 2c_1(c_2 - r) \\ \text{Convex} \quad \text{slope} &= \left|\frac{\partial Z_t}{\partial r}\right| = 2r/c_4^2 \end{aligned} \quad (12)$$

As with the divergent cases with constant slopes, the comparison domain is the same as the computational domain, since the contour lines are closed.

**3.3. Convergent Hillslopes**

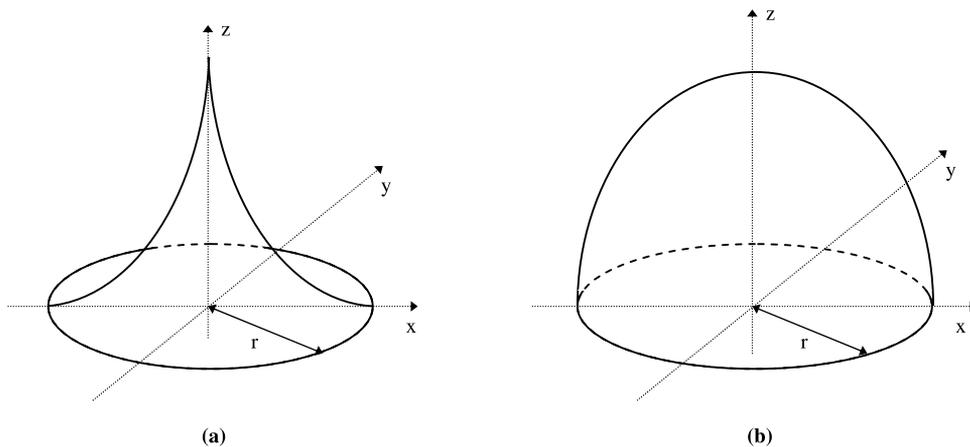
[30] A top view and a side view of a convergent hillslope are illustrated in Figure 6. In a  $NCOL \times NROW$  domain, the coordinates of each cell are given by:

$$\begin{aligned} x &= [j - 0.5(NCOL - 1)]*HR, \quad j = 0, \dots, NCOL - 1 \\ y &= -i*HR, \quad i = 0, \dots, NROW - 1 \end{aligned} \quad (13)$$

In this paper, we set  $NCOL = 4001$ , and  $NROW = 2001$ . Unlike the divergent hillslopes, in the x-y plane we set the origin (0, 0) at the upper boundary, rather than the center, of the domain. Otherwise, the convergent hillslope would be filled and becomes a flat area if we set the origin at the center of the domain, and apply ARC/INFO FILL function. The theoretical elevation of each cell is given by:

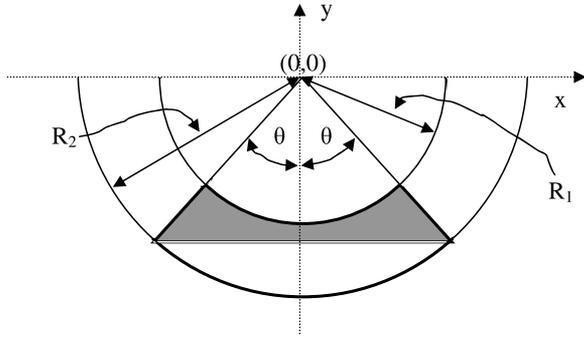
$$Z_t = f_2 + rslope = f_2 + r \tan\beta \quad (14)$$

where  $\tan\beta$  is slope and  $\beta$  is slope angle ( $0^\circ < \beta < 90^\circ$ ), and  $f_2$  is the elevation of the origin (0, 0). In this paper, we set

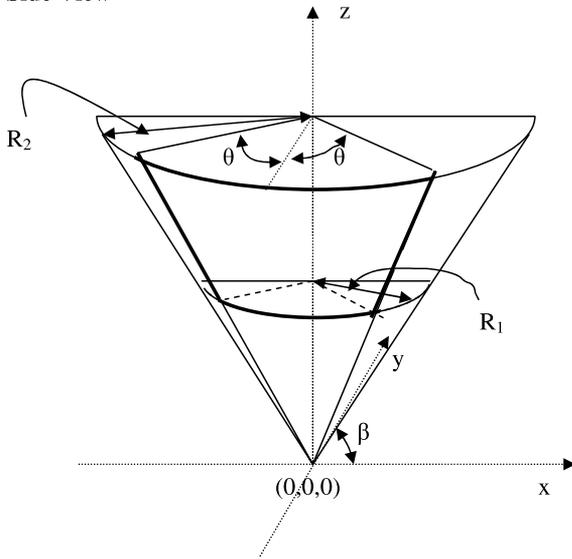


**Figure 5.** Side view of two divergent hillslopes with varying slopes: (a) slope decreases from top to bottom (concave) and (b) slope increases from top to bottom (convex).

Top View



Side View



**Figure 6.** Top and side views of a convergent hillslope. The computational domain is bounded by two open contour lines and two lateral boundaries. The shaded area is the influence domain due to the lateral boundaries, and the unshaded area inside the computational domain is the comparison domain.

$f_2 = 200$ , and the computational domain as the region confined by two thick curves and two lateral boundaries shown in Figure 6, which satisfies the following conditions:

$$R_1 \leq r < R_2, \quad \theta \leq \Phi \quad (15)$$

where  $r = \sqrt{x^2 + y^2}$ ,  $\theta = |\text{atan}(x/y)|$ . In this study, we set  $R_1 = 1400\text{HR}$ ,  $R_2 = 1900\text{HR}$  and  $\Phi = 45^\circ$ . The theoretical topographic index is given by:

$$\begin{aligned} \ln(a/\tan\beta) &= \ln \left[ \frac{\text{upslope area / contour length}}{\text{slope}} \right] \\ &= \ln \left[ \frac{\pi R_2^2 - \pi r^2}{2\pi r \tan\beta} \right] = \ln \left[ \frac{R_2^2 - r^2}{2r \tan\beta} \right] \end{aligned} \quad (16)$$

The above expression is valid only if the contour lines are closed. However, the contour lines in the computational domain are open (Figure 6). Therefore the comparison domain is not the same as the computational domain. As we did for the planar hillslopes, we first identify the influence region where the flow accumulation area is influenced by

the lateral boundaries based on flow directions given by the MFD algorithm.

[31] The procedure to determine the influence and comparison domains for the convergent hillslope is shown in Figure 7. The computational domain is formed by two lateral boundaries: AD and BC, and two arcs: AEB and CFD. Point A has three flow directions, i.e., 3, 4, and 5. Among those directions, only direction 5 flows into the computational domain. Therefore the influence region due to the lateral boundary AD is the region between AD and AG, where the angle formed by those two lines is  $45^\circ$ . Similarly, the influence domain due to the lateral boundary BC is the region between BC and BG, and the angle formed by those two lines is also  $45^\circ$ . Excluding the influence domain (shaded area) from the computational domain, we obtain the comparison domain (unshaded area inside the computational domain).

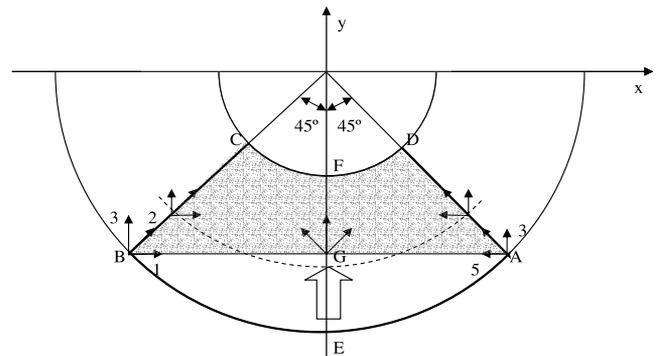
**4. Evaluation Method**

[32] To compute the topographic index of each idealized hillslope we first develop an idealized DEM data set based on equations in section 3. The ARC/INFO FILL and FLOW-DIRECTION functions are applied to determine flow direction at each cell over the whole domain. The ARC/INFO-derived flow directions are used only in flat areas where flow directions cannot be resolved by the SFD, BFD, or MFD algorithms. Those flow directions are also utilized by the TFD algorithm for recalculating slope values in flat areas.

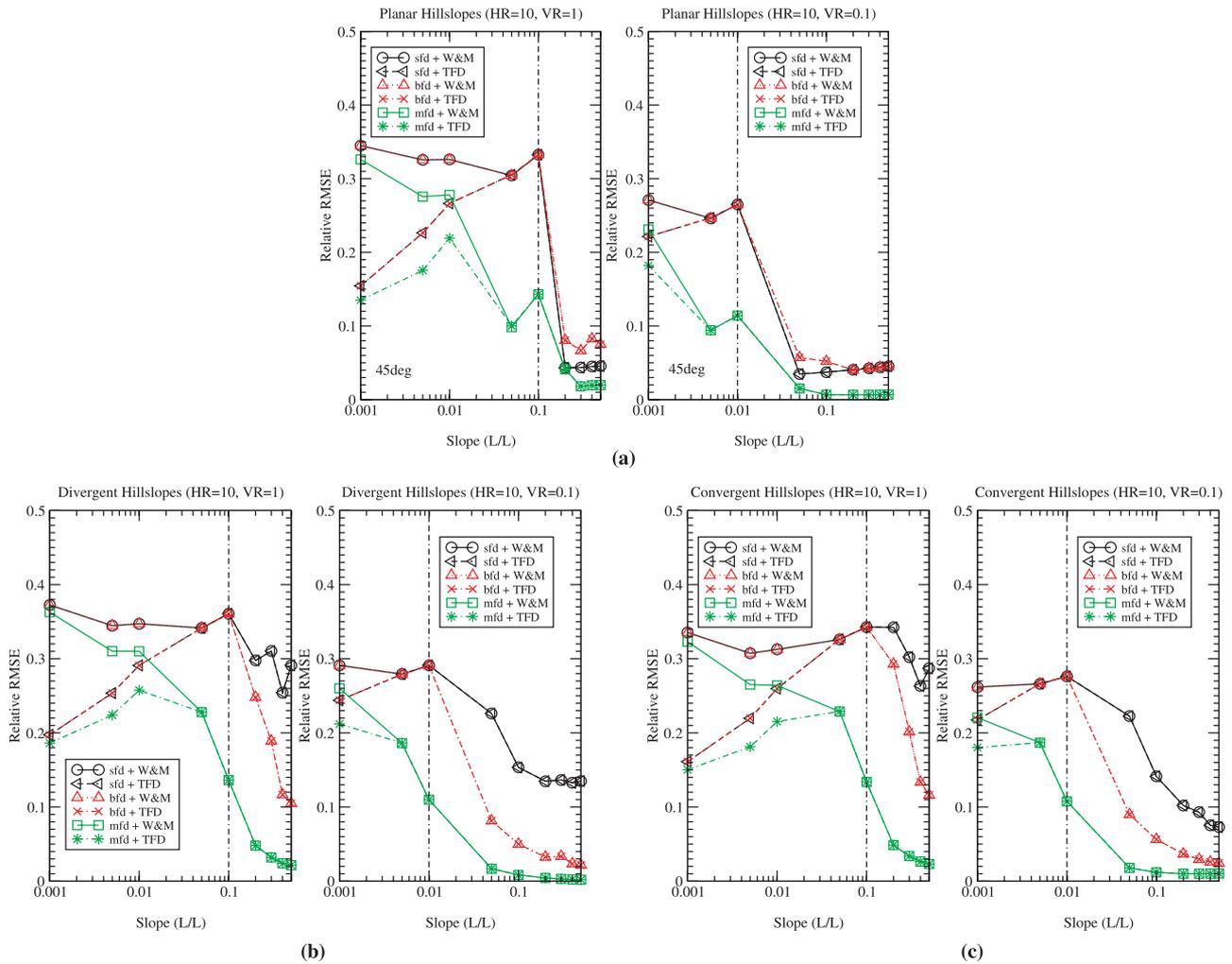
[33] We use a recursive algorithm to compute flow accumulation area inside the computational domain. The slope, contour length, and topographic index at each cell are determined first. Next, the root-mean-square error (RMSE) of the calculated topographic index is computed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{i=n} [\ln(a/\tan\beta)_{Ti} - \ln(a/\tan\beta)_{Ci}]^2}{n}}, \quad (17)$$

where subscripts  $Ti$  and  $Ci$  stand for theoretical and computed topographic indices at cell  $i$  and  $n$  is the number



**Figure 7.** For a convergent hillslope the computational domain is formed by two lateral boundaries AD and BC and the top boundary AEB and the bottom boundary CFD. The influence domain is the shaded area formed by AD, BC, line AGB, and the bottom boundary CFD. The comparison domain is the unshaded area formed by the top boundary AEB and line AGB.



**Figure 8.** Plots of relative RMSE versus slope with two sets of horizontal and vertical resolutions: (1)  $HR = 10$  units,  $VR = 1$  unit (left panel), and (2)  $HR = 10$  units,  $VR = 0.1$  unit (right panel) for (a) planar, (b) divergent, and (c) convergent hillslopes.

of total cells inside the comparison domain. The relative RMSE is:

$$\text{Relative RMSE} = \frac{\text{RMSE}}{\left[ \sum_{i=1}^{i=n} \ln(a / \tan \beta)_{T_i} \right] / n} \quad (18)$$

## 5. Results and Discussions

### 5.1. Comparison of Relative RMSEs

[34] A total of 324 tests were made: three types of idealized hillslopes (i.e., planar, divergent with constant slopes, and convergent) times two sets of data resolutions (i.e., ( $HR = 10$  units,  $VR = 1$  unit), ( $HR = 10$  units,  $VR = 0.1$  unit)) times six combinations of flow direction and slope algorithms (i.e., SFD+W-M, SFD+TFD, BFD+W-M, BFD+TFD, MFD+W-M, MFD+TFD) times nine mean slope values (i.e., 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5). The contour angle is  $45^\circ$  for the planar hillslopes. The relative RMSEs of the computed topographic index are shown in Figures 8a, 8b, and 8c, for planar, divergent, and convergent hillslopes, respectively.

[35] No matter whether the hillslope is planar, divergent, or convergent, we observed five common error characteristics for the computed topographic index (as illustrated in Figure 8): (1) The MFD + TFD algorithm yielded the smallest errors for all cases. (2) As the slope increases, the overall errors decrease. (3) As the vertical resolution of the data increases, the overall errors decrease. (4) If the mean slope is less than the threshold, i.e.,  $VR/HR$  (dotted line on Figure 8), there is no difference between the BFD algorithm and the SFD algorithm. If the mean slope is greater than  $VR/HR$ , the BFD algorithm is better than the SFD algorithm for divergent and convergent hillslopes, and a little worse than the SFD algorithm for planar hillslopes with a  $45^\circ$  angle between the contour line and x axis. The effect of the angle between the contour line and x axis on the computed topographic index of planar hillslopes will be discussed in section 5.2. (5) Error in the computed topographic index due to W-M method and TFD method for recalculating slope values in flat areas appears when the mean slope is less than  $0.5VR/HR$ . The errors associated with TFD are less than those associated with W-M, because W-M only assign  $0.5VR/HR$  as the min-

**Table 1.** Minimum Slope Values Determined by the Tracking Flow Direction (TFD) and *Wolock and McCabe's* [1995] (W-M) Methods

Theoretical Slope Value	HR = 10, VR = 1						HR = 10, VR = 0.1					
	Planar		Divergent		Convergent		Planar		Divergent		Convergent	
	TFD	W-M	TFD	W-M	TFD	W-M	TFD	W-M	TFD	W-M	TFD	W-M
0.001	0.0010	0.05	0.0009	0.05	0.0009	0.05	0.0009	0.005	0.0009	0.005	0.0009	0.005
0.005	0.0048	0.05	0.0050	0.05	0.0045	0.05	0.0041	0.005	0.0041	0.005	0.0033	0.005
0.01	0.0092	0.05	0.0088	0.05	0.0088	0.05	0.0071	0.005	0.0071	0.005	0.0071	0.005
0.05	0.0414	0.05	0.0414	0.05	0.0333	0.05	0.0500	0.005	0.0400	0.005	0.0410	0.005

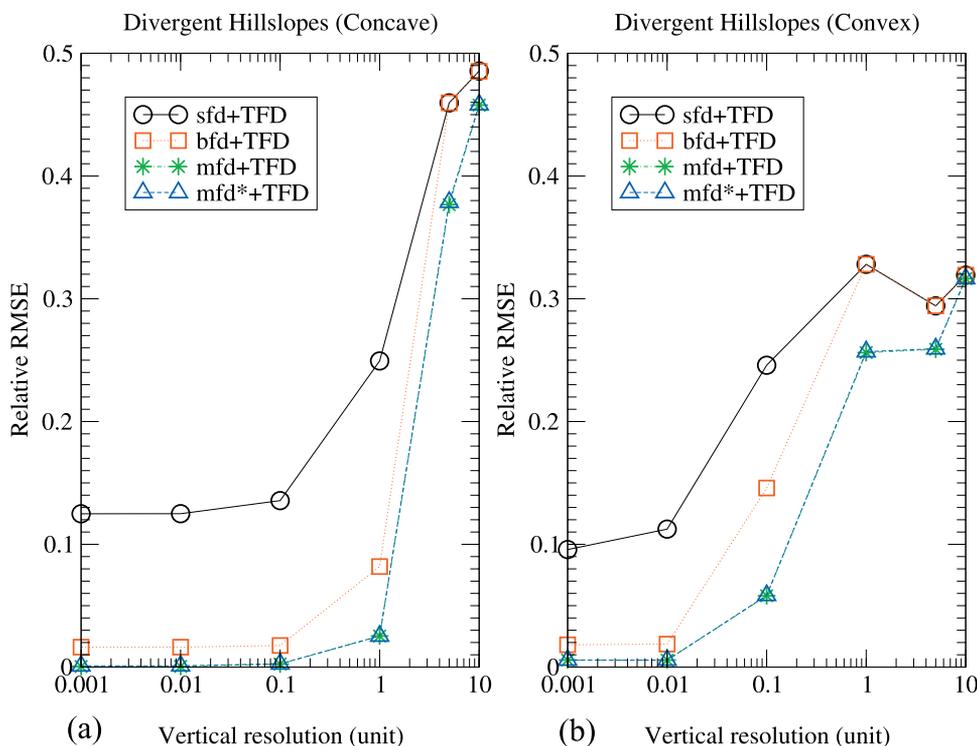
imum slope value whereas TFD assigns a smaller value than 0.5VR/HR for replacing zero-slope values in flat areas. On the other hand, if mean slope is greater than 0.5VR/HR, TFD can also replace zero slopes with a larger value (see Table 1).

[36] Another set of 40 tests were also carried out to investigate the effects of the definition of the contour length, the variation of slopes and the vertical resolution on the errors of the computed topographic index. Those 40 tests are combinations of two divergent hillslopes (i.e., concave and convex) times four algorithms (i.e., SFD+TFD, BFD+TFD, MFD+TFD, and MFD\*+TFD) times five vertical resolutions (i.e., 0.01, 0.1, 1, 5, and 10 units). The results are shown in Figure 9. These results indicate that the MFD+TFD algorithm is the best for all cases. The difference due to the different algorithms to assign the contour lengths (i.e., MFD and MFD\*, as discussed in section 2.3) is negligible. The errors decrease and finally become stable for all cases as the vertical resolution becomes finer and finer. However, a fine vertical resolution cannot eliminate the difference

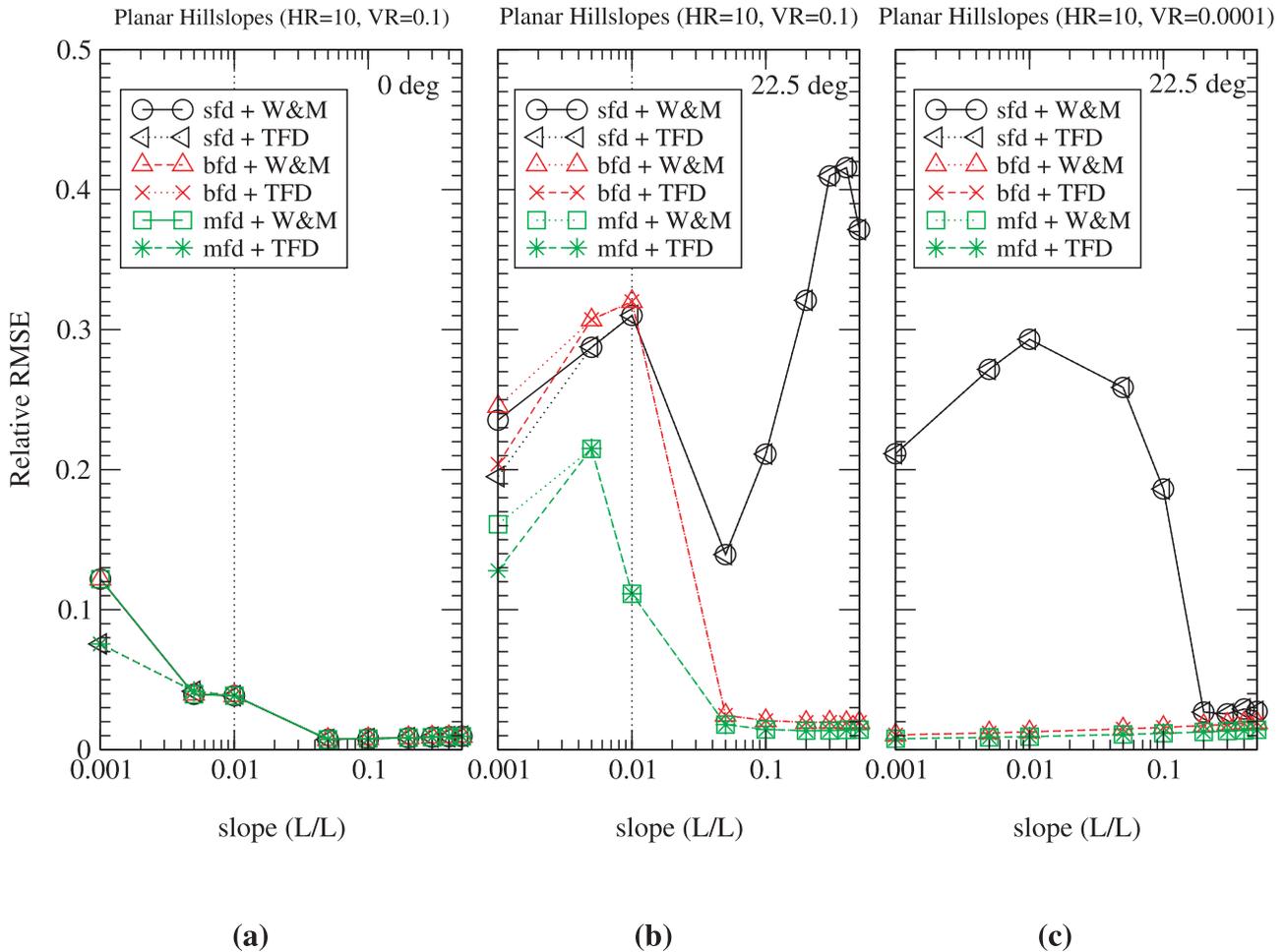
among the errors due to different algorithms, i.e., MFD is better than BFD, and BFD is better than SFD.

**5.2. Effect of Contour Line Orientation**

[37] In section 5.1 we only tested the planar hillslopes with a 45° angle between the contour line and x axis. To investigate the effect of the contour line orientation, two contour angles, i.e., 0° and 22.5° were examined. The vertical and horizontal resolutions are 0.1 unit and 10 units respectively. Figure 10a shows that there is no difference among the SFD, the BFD, and the MFD algorithms for 0°, and the SFD algorithm gives the largest error for 22.5° (Figure 10b). To further understand the effect of contour line orientation, we also tested 22.5° at HR = 10 units, and VR = 0.0001 unit, and the results were plotted in Figure 10c. Figure 10c shows that only when the planar hillslope is relatively steep (e.g., slope >0.1) will using a finer vertical resolution DEM reduce the errors for the SFD algorithm. For a relatively flat planar hillslope, even using a finer resolution DEM, errors remain in the SFD algorithm. However, this is



**Figure 9.** Plots of relative RMSE versus vertical resolution for two divergent hillslopes: (a) concave (slope decreases from top to bottom) and (b) convex (slope increases from top to bottom).



**Figure 10.** Plots of relative RMSE versus slope for planar hillslopes: (a)  $HR = 10$ ,  $VR = 0.1$ ,  $0^\circ$  between each contour line and x axis; (b)  $HR = 10$ ,  $VR = 0.1$ ,  $22.5^\circ$  between each contour line and x axis; and (c)  $HR = 10$ ,  $VR = 0.0001$ ,  $22.5^\circ$  between each contour line and x axis.

not the case for either the BFD algorithm or the MFD algorithm.

[38] Figure 10 indicates that the errors in the SFD's computed topographic index of a planar hillslope are not only a function of vertical resolution of DEM data, but also the orientation of the contour lines. When the contour angle is an integer multiple of  $45^\circ$  (e.g.,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ...), the theoretical flow direction (i.e., normal direction of the contour lines) is parallel to one of 8 possible flow directions in the SFD algorithm. Thus the SFD algorithm can resolve the flow directions without any approximation. However, if the normal direction of the contour lines is not parallel to one of 8 possible flow directions (e.g., contour angle =  $22.5^\circ$  and the angle between the normal direction of the contour lines and x axis is  $67.5^\circ$ ), the SFD algorithm must approximate the flow direction (e.g., either  $45^\circ$  or  $90^\circ$ ), and the difference between theoretical flow direction and the SFD's flow direction is  $22.5^\circ$ .

## 6. Conclusions

[39] By comparing the numerically computed and theoretical (analytically calculated) topographic indices for 526 cases, we found that the MFD algorithm is best in

terms of accuracy for all idealized hillslope cases. Although the BFD algorithm is a little worse than the SFD algorithm when the angle between the contour line and x axis is  $45^\circ$  for planar hillslopes, overall the BFD algorithm is better than the SFD algorithm. The difference in the errors of the computed topographic index due to the different algorithms for assigning the contour lengths is negligible (i.e., MFD [Quinn *et al.*, 1991] and MFD\* [Wolock and McCabe, 1995]). We found that the orientation of the contour lines of the planar hillslopes significantly influences the SFD's computed topographic index. If the contour lines are not parallel to one of eight possible flow directions, the errors in the SFD's computed topographic index are significant.

[40] Recalculating nonzero slopes in flat areas increases errors only when the mean slope value is less than  $0.5VR/HR$ . The tracking flow direction (TFD) method is more accurate than the W-M method [Wolock and McCabe, 1995] because TFD can give a minimum slope value less than  $0.5VR/HR$ . Therefore, to achieve the highest accuracy, we recommend using a combination of the MFD algorithm and the TFD method to compute topographic index, especially when the vertical resolution of DEM data is low.

[41] **Acknowledgments.** The authors wish to thank H. I. Jager for her valuable comments and suggestions. Useful comments were provided by the Associate Editor and two anonymous reviewers. Research was supported by an appointment to the ORNL Postdoctoral Research Associates Program (F. Pan), which is sponsored by Oak Ridge National Laboratory and administered jointly by Oak Ridge National Laboratory and by the Oak Ridge Institute for Science and Education under contract numbers DE-AC05-84OR21400 and DE-AC05-76OR00033, respectively. This research was also supported by the U.S. Department of Energy's Office of Science and performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. Additional research support was provided by NASA grant NAG5-8698 (Peters-Lidard) and the United States Environmental Protection Agency through Grants R825210 (Peters-Lidard) and R825211 (Peters-Lidard). Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through Grants R825210 and R825211, it has not been subjected to the Agency's required peer review and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

## References

- Band, L. E. (1986), Topographic partition of watersheds with digital elevation models, *Water Resour. Res.*, 22(1), 15–24.
- Beven, K. J., and M. J. Kirkby (1979), A physical based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24(1), 43–69.
- Beven, K., and E. F. Wood (1983), Catchment geomorphology and the dynamics of runoff contributing areas, *J. Hydrol.*, 65, 139–158.
- Dai, Y., et al. (2003), The Common Land Model (CLM), *Bull. Am. Meteorol. Soc.*, 84, 1013–1023.
- Famiglietti, J. S., and E. F. Wood (1994), Multiscale modeling of spatial variable water and energy balance processes, *J. Geophys. Res.*, 30(D11), 3061–3078.
- Ford, R., S. Running, and R. Nemani (1994), A modular system for scalable ecological modeling, *IEEE Comput. Sci. Eng.*, 1(3), 32–44.
- Greenlee, D. D. (1987), Raster and vector processing for scanned linework, *Photogramm. Eng. Remote Sens.*, 53(10), 1383–1387.
- Jenson, S. K., and J. O. Domingue (1988), Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogramm. Eng. Remote Sens.*, 54, 1593–1600.
- Jones, K. (1998), A comparison of algorithms used to computer hill slope as a property of the DEM, *Comput. Geosci.*, 24, 315–323.
- Kirkby, M. J. (1997), TOPMODEL: A personal view, *Hydrol. Processes*, 11, 1087–1097.
- Koster, R. D., M. J. Suarez, A. Duchame, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure, *J. Geophys. Res.*, 105(D20), 24,809–24,822.
- Mark, D. M. (1988), Network models in geomorphology, in *Modeling Geomorphologic Systems*, edited by M. G. Anderson, pp. 73–79, John Wiley, New York.
- Mendicino, G., and A. Sole (1997), The information content theory for the estimate of topographic index distribution used in TOPMODEL, *Hydrol. Processes*, 11(9), 1099–1114.
- Montgomery, D. R., and W. E. Dietrich (1992), Channel initiation and the problem of landscape scale, *Science*, 255, 826–830.
- O'Callaghan, J. F., and D. M. Mark (1984), The extraction of drainage networks from digital elevation data, *Comput. Vision Graphics Image Processes*, 28, 323–344.
- Peters-Lidard, C. D., M. S. Zion, and E. F. Wood (1997), A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance process, *J. Geophys. Res.*, 102(D4), 4303–4324.
- Quinn, P. F., K. J. Beven, P. Chevallier, and O. Planchon (1991), The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models, *Hydrol. Processes*, 5, 59–80.
- Rieger, W. (1998), Phenomenon-based approach to upslope contributing area and depressions in DEMs, *Hydrol. Processes*, 12(6), 857–872.
- Tarboton, D. G. (1997), A new method for the determination of flow directions and upslope areas in grid digital elevation models, *Water Resour. Res.*, 33(2), 309–319.
- Tarboton, D. G., R. L. Bras, and I. Rodriguez-Iturbe (1991), On the extraction of channel networks from digital elevation data, *Hydrol. Processes*, 5, 81–100.
- Walko, R. L., et al. (2000), Coupled atmosphere-biophysics-hydrology models for environmental modeling, *J. Appl. Meteorol.*, 39(6), 931–944.
- Wolock, D. M., and G. J. McCabe (1995), Comparison of single and multiple flow direction algorithms for computing topographic parameters, *Water Resour. Res.*, 31(5), 1315–1324.
- Zhang, X., N. A. Drake, J. Wainwright, and M. Mulligan (1999), Comparison of slope estimates from low resolution DEMs: Scaling issues and a fractal method for their solution, *Earth Surf. Processes Landforms*, 24, 763–779.

A. W. King, F. Pan, and M. J. Sale, Environmental Sciences Division, MS 6335, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. (panf@ornl.gov)

C. D. Peters-Lidard, Hydrological Sciences Branch, Code 974, NASA Goddard Space Flight Center, Greenbelt, MD 20071, USA.