Estimating surface flow over digital elevation models using a new improved form-based algorithm

A. Hasan, P. Pilesjö & A. Persson
Department of Physical Geography and Ecosystems Analysis,
Lund University, Sweden

Abstract

This paper discusses new improvements of a form-based algorithm, which is used to estimate flow distribution over a continuous surface. In the new form-based algorithm (IFBFD), cells in a DEM are classified into five different classes. The classes are Peaks, Complicated, Sinks, Flats and Undisturbed cells. The method of how to estimate the flow distribution from each cell depends on its class. Estimating the flow distribution over flat area cells and sinks is done in an innovative way. The flow over a flat area can be either flow-out or flow-in. Flow-out occurs when one or more cells on the flat area border has an elevation lower than the flat area cells. The flat area is classified as 'flow-in' when all cells on the border of flat area have elevations higher than the flat area cells. The result is that the flow will be converged in the center of the flat area, and that cells will have no outflow (sink). Additionally, a culvert function is added to the new algorithm to enable the user to deal with man-made flow barriers like roads and railway lines. The new culvert function breaches the barrier and connects the flow between two defined points on both sides of it.

The new algorithm is tested using the number of mathematical surfaces, as well as on a real DEM derived from LIDAR data. The results of comparing our new algorithm with some well-known algorithm used in most GIS programs shows that the IFBFD algorithm produces more realistic results than other algorithms. Tests show the capability of the new IFBFD algorithm to deal with different terrain types, flat areas and sinks, making it suitable for simulating the real flow distribution over any DEM without the need to e.g. fill sinks. Moreover, the IFBFD algorithm produces a convincing result when deriving the drainage network.

WIT Transactions on Ecology and the Environment, Vol 146, © 2011 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/RM110181

Keywords: digital terrain modelling, digital terrain analysis, form-based algorithm, hydrological modeling, surface flow estimation, flow routing algorithm, flat areas.

1 Introduction

Catchment topography is critical for models of distributed hydrological processes. Slope controls flow pathways for surface as well as near surface flow, and influences the sub surface flow pattern substantially. The key parameter in catchment topography is flow distribution, which tells us how overland flow is distributed over the catchment area.

Stating flow distribution over a land surface is a crucial measurement in hydrological modeling, the use of Digital Elevation Models (DEM) has made it possible to estimate flow on each location over a surface. Based on the flow distribution estimation on each location represented by a DEM, the drainage pattern over an area, as well as various hydrological parameters, such as catchment area and up-stream flow accumulation, can be modeled.

Generally, surface flow and flow accumulation are estimated by the use of two different types of raster algorithms; either approximating to a single, or multiple, flow directions. If working with raster data, multiple flow algorithms assume transport to more than one adjacent cell, while a single flow algorithm only distributes water to one neighbor cell at a time in the raster. In many cases, single directional flow algorithms produce satisfying results over concave surfaces, while it is often more appropriate to divide the flow into two or more directions if the surface is flat or has a convex form. Combinations of the two types are often to prefer when modeling e.g. water flow over natural surfaces [10].

However, the mentioned raster algorithms are often not optimal, and needs extensive calibration. The latter is mainly due to the problem how to weight the influence of slope when splitting flow between neighbouring cells (see e.g. [7]).

The single flow algorithm was described by [6]. It assumes that flow follows only the steepest downhill slope. Using a raster DEM, implementation of this method resulted in that hydrological flow at a point only follows one of the eight possible directions corresponding to the eight neighbouring grid cells [1, 2, 5, 6]. However, for the quantitative measurement of the flow distribution, this over-simplified assumption must be considered as illogical and would obviously create significant artifacts in the results, as stated by e.g. [3, 4, 9, 12].

Attempts to solve the problem connected to the single flow algorithms have led to several proposed ‘multiple flow direction’ algorithms [3, 4, 9–11]. These algorithms estimate the flow distribution values proportionally to the slope gradient, or risen slope gradient, in each direction. Holmgren [4] summarizes some of the algorithms as

$$f_i = \frac{(\tan \beta_i)^x}{\sum_{j=1}^{8} (\tan \beta_j)^x}, \text{ for all } \beta > 0 \qquad (1)$$

where i,j = flow directions (1…8), fi = flow proportion (0…1) in direction i, tan ßi = slope gradient between the center cell and the cell in direction i, and x = variable exponent.

By changing the exponent (x) in Equation (1), two extreme approaches in estimating flow distribution can be observed. While x = 1, flow will be distributed to downhill neighboring cells proportionally to the slope gradients, as suggested by [11]. The other extreme is when $x \to \infty$, which will approach towards the 'single flow' drainage distribution mentioned above. Holmgren [4] suggested an x value between 4 and 6. This gives a result between a very homogeneous flow distribution when x = 1, and a distinctive flow which occurs when x becomes greater than 10.

A form-based solution, sometimes referred to as the Pilesjö-Zhou algorithm, was presented by [10]. Given the limitations and problems of the algorithms presented above a 'multiple flow direction' approach based on analysis of the form of individual 3x3-cell surface facets was proposed. It was assumed that flow diverge over convex surfaces, and converge over concave surfaces. There is no absolute way to determine convexity and concavity of the center cell in a 3x3-cell facet. The possible complexity of the surface often implies approximations. One way to approximate, used in the form-based algorithm is to use a trend surface based on the elevation values of all nine cells in the facet. When the topographic form of the center cell in the facet is judged as concave, the flow is distributed fully to the main drainage direction. If the main drainage direction is not equal to the direction to one of the eight neighboring cells, the flow distribution has to be split between two cells. This is done by splitting the drainage vector into two diagonal (i.e. 45° apart) vectors. When the topographic form of the center cell in the facet is judged as convex, the flow is distributed according to equation (1).

The aim of this study is to further develop the form based algorithm, focusing on complex terrain, flat areas, sinks, as well as an optimization of the x value for convex surfaces.

2 Method

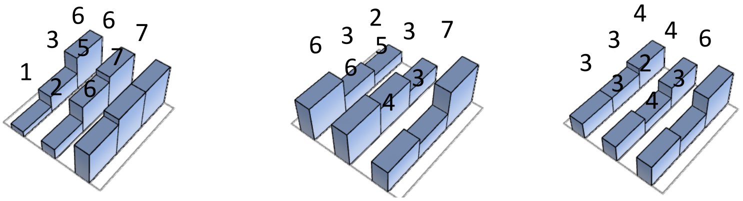
Bellow follows a step-by-step description of the proposed IFBFD algorithm, from the classification of individual cells in a DEM to the estimation of flow paths over the surface.

2.1 Cell classification

The estimation of flow is based on classification of topographic form (of each cell in the DEM), using a 3 by 3cells moving window. In the former form-based algorithm the cells were classified into three classes (Undisturbed, Complicated and Sinks). Here, in the new form-based algorithm (IFBFD), the cells are instead classified into five different categories. The classes are Undisturbed, Complicated, Sinks, Flats and Peaks.

The classification process is based on elevation differences between the center cell in the 3 x 3 window and its eight neighbors.

- ***Undisturbed:*** As shown in figure 1A, a cell is classified as Undisturbed if there is one valley "out" from the center cell in the 3 x 3 cell window. The valley is minimum one and maximum seven cells wide. A valley is defined as cell/cells having lower elevation value/values than the center cell.
- ***Complicated:*** As shown in figure 1B, a cell is classified as Complicated if there is more than one valley in the 3 x 3 window.
- ***Sink:*** As shown in Figure 1C, a cell is classified as a Sink if all 8 surrounding cells are higher than the center cell.
- ***Flats:*** As shown in Figure 1D, a cell is classified as Flat if at least one neighbor cell has the same elevation as the center cell, and the remaining cells are higher.
- ***Peaks:*** As shown in Figure 1E, a cell is classified as a Peak if all surrounding cells are lower than the center cell.



(A) *Undisturbed* ***(B)*** *Complicated* ***(C)*** *Sink*

(D) *Flat* ***(E)*** *Peak*

Figure 1: (A-D). Illustrations of the different topographic forms classified by the use of a 3 x 3 cell window.

2.2 Filling sinks

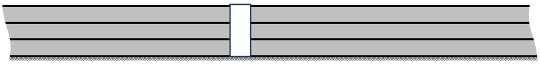
Sinks are natural features that frequently occur in the terrain. When carrying out hydrological modeling sink cells should normally not be filled unless we are sure that this sink is an artifact. In the presented IFBFD algorithm there is no need to fill the sink in order to run the flow model. Also including sinks the software estimate flow directions from all cells and connects them into flow distribution and flow accumulation. However, any unfilled sink will result in stopping flow at that cell. Sometimes this is not desirable.

In order to remove sinks caused by errors in the input data or errors in the interpolation algorithm (or if the user wants to remove natural inks) a function was created in MATLAB [14] to fill/remove sinks in a DEM. The user can choose between filling all the sinks or only fill sinks with a specified maximum depth. The latter alternative gives the user the option to, based on her/his knowledge 'filter2 the DEM to get rid of artificial sinks. Filling all sinks will of course result in continuous flow towards the DEM edges.

2.3 Breaching culverts

A function to breach culverts has been added in order to enable users to deal with e.g. man-made flow barriers like roads and railway lines. This function breaches the barrier by connecting two defined points/cells on opposite sides of the obstruction. This is done in a semi-automatic way, where the program is suggesting points with the possibility for the user to change their positions. The function is illustrated in Figure 2 below.

A: Culvert connecting the two sides of a barrier, the path is Invisible with top view



B: Breached culvert to enable water path between the two ends of the barrier

Figure 2: The breaching culvert function do not change the actual elevations in the DEM (A), but only let flow pass the barrier (B).

2.4 Flow distribution

The method of how to estimate the flow distribution from each cell depends on its classification. Below we present how flow is estimated to be distributed to surrounding neighbors depending on terrain type classification.

Sinks: As mentioned above there is an option to remove some are all cells in the DEM. If so, all cells in the filled sinks have got the same elevation as the closest neighboring cell able to transport water further in the DEM. This means that the sink has been transformed to a flat area, described below. Unfilled sinks have no outgoing flow, resulting in a flow from these cells equaling zero.

Complicated and Peak cells: Flow distribution from Complicated and Peak cells is estimated using Equation (1) above. The flow will be distributed to all cells lower than the center cell. The reason behind having two different classes for cells while the flow distribution is done using the same equation is that the flow accumulation calculation is done starting from peak cell.

Undisturbed cells: The flow from undisturbed cells depends on the topographic form of the surface surrounding the undisturbed cell. The surface can be either concave or convex. Since there is no absolute way to determine convexity and concavity of the center cell in a 3 x 3cell facet we have to use approximations. One way to approximate is to use a trend surface based on elevation values of all cells in the facet. If the surface is convex flow is distributed to all lower cells around the center cell. If the surface is concave then the direction of the flow resultant is calculated, and the flow is split between the two cells closest to this ‘main drainage direction’ (if the direction is not directly pointing at one neighboring cell, i.e. 0, 90, 180 or 270 degrees). A detailed description of flow distribution from undisturbed cells can be found in [10].

Flat cells: Flat cells can exist in any DEM naturally, or be a result of errors in input data, an inappropriate interpolation algorithm, or the ‘new surfaces’ created by filling sinks (see above). Estimating flow distribution over flat area cells was introduced in IFBFD since water naturally always flow from higher elevations to lower, ever if the terrain is ‘terraced’. Flow over a flat area was defined to be either ‘flow-out’ or ‘flow-in’. Flow-out occurs when one or more cells on the flat area border have an elevation lower than the flat area cells. This implies that the flow from the flat area cells is directed out of the flat area. The flat area cells are defined as flow-in when all cells on the border of the flat area have elevations higher than the flat area cells, this will result in a flow pattern converging in the center of the flat area. The cell in the center of the flat area will have no flow out and assigned as sink. Two functions were created in order to define flow over flat surfaces.

FLAT_Flow_out function is a function directing the flow from flat cells where a way out of the flat area can be found. An example is illustrated in Figure 3A. Estimating the flow directions in this case is done by starting by giving neighboring cells to the ‘out glow’ cell flow directions ‘pointing’ to that cell, and then stepwise move further and further away from the out flow cell assigning flat cells flow directions to neighbors with a defined direction.

The ***FLAT_Flow_in*** function is used when there is no way out from the flat area. An example is presented in Figure 3B. All cells on the border of the flat area have elevations higher than the flat area cells, and this result in a converging flow in the center. This center cell will have no defined flow and will be treated as a sink (see above). The flow directions of the surrounding cells will be estimated (by vector addition) starting from the flat cells that have the maximum number of known distributed flow directions cells (i.e. the border cells of the flat area).

2.5 Flow accumulation/drainage area

In order to estimate flow accumulation (sometimes referred to as drainage area) the drainage paths over the surface have to be traced, and the number of cells (area) transporting water to cells down-slope has to be calculated. This is done by starting a 'search function' from peak cells, transporting water to surrounding cells but not receiving anything. The drainage area for these peak cells will be one cell, and the drainage areas for the cells they are transporting water is updated proportionally to who much of the total flow (1 cell) that is distributed to each of them. Then these peak cells are flagged as visited and they are neglected in further processing. After this we again search for cells with only out flow. Since the original peaks are excluded this will be cells neighboring the peaks. They are treated like the peaks, we 'jump down' one further level, and go on like this until all cells have been examined. A detailed description of the estimation of flow accumulation can be found in [10].

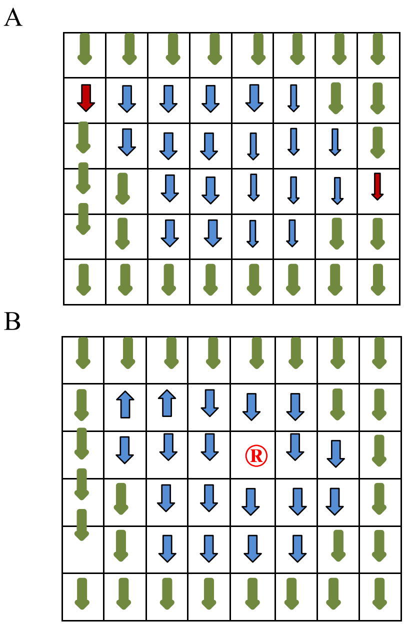


Figure 3: Illustrations of how the flow distribution is assigned to cells in two different types of flat areas. In A the flat area has an outflow, resulting in a continuous flow pattern, and in B the flat area has no outflow, resulting in a sink in the middle of the area.

3 Results

In order to calibrate and test the improved IFBFD algorithm three different comparisons were made:

1. The estimated flow accumulation using the IFBFD algorithm was compared to mathematical solutions of flow accumulated calculated from pre-defined mathematically created surfaces (plane and ellipse) in order to find the best x value (see Equation (1)) for convex surfaces.
2. The estimated flow accumulation using the IFBFD algorithm was compared to other well known, frequently used, algorithm referring to the mathematical solutions mentioned above.
3. The estimated flow accumulation using the IFBFD algorithm was applied on a natural surface and visually compared to the single flow algorithm described by [6].

Table 1 shows the result of the different x values tested for the convex surfaces. The root mean square error (RMSE) between the IFBFD result and the mathematical solution is presented for a number of different x values. For a detailed description of the test and the creation of the mathematical results and generated surfaces (see [13]). Based on the test, an x value of 1.4 was selected, judged to be most appropriate for convex surfaces.

Table 1: RMSE for different x values (see Equation (1)) based on a comparison between the IFBFD algorithm and mathematical solutions for a plane and an ellipse.

x	1	1.1	1.15	1.2	1.25	1.3	1.35	**1.4**	1.5	2
Plane RMSE	35.97	28.62	25.24	22.13	19.38	17.04	15.35	**14.37**	14.95	37.23
Ellipse RMSE	11.4	10.33	9.95	9.65	9.43	9.28	9.19	**9.14**	9.19	11.29

The comparison between the IFBFD algorithm and other commonly used algorithms, based on RMSE values for a plane and an ellipse indicated satisfactory performance of the IFBFD algorithm. A test against the single flow algorithm, used e.g. in ArcGIS gave RMSE values (plane and ellipse respectively) of 154.53 and 64.66 for the single flow algorithm, considerably higher than the solution presented in this paper. For the commonly used multiple flow algorithm ([4] and Equation (1)) the RMSE was the same, only differing because of the x value. However, since the x value is often set to a default value using other software, and we set it to an optimized value of 1.4 it was concluded that the IFBFD algorithm was considered equal or better.

The result of the comparison between the single flow algorithm (ArcGIS) and the IFBFD algorithm applied on a natural surface is presented in Figure 4 below. It is obvious, by a visual interpretation, that the proposed IFBFD algorithm better reflect the flow accumulation over a natural landscape then the relatively simplified single flow algorithm.

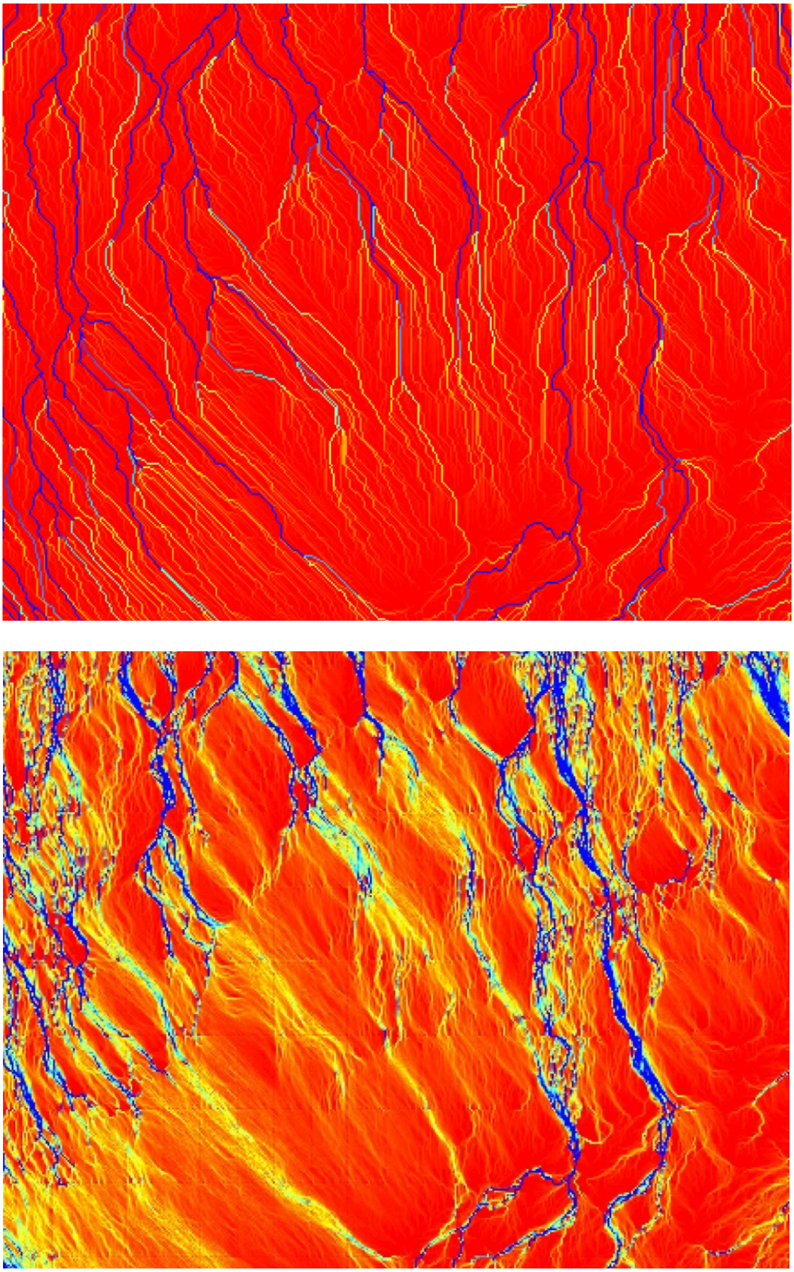


Figure 4: A comparison between the single flow (upper) and IFBFD algorithms (lower), illustrating estimated flow accumulation over a natural surface in Northern Sweden.

4 Discussion and conclusion

The improved form-based flow algorithm presented above seems to perform better than comparable, commonly used, flow algorithms. It gives the user possibilities to fill sinks, natural or artificial, as well as letting flow pass barriers in the landscape.

Flat areas are treated in a realistic way, where water is always ‘forced’ to lower elevations if possible. Flat areas without an outflow are treated as sinks.

The calibration of the x value in Equation (1) has optimized the flow estimation over convex surfaces. For concave surfaces the approach from the original form-based algorithm is kept, distributing flow to one or two neighboring cells down-stream.

Overall, the proposed IFBFD algorithm seems to represent flow over a natural landscape in a satisfactory way. More tests have to be carried out, especially on different mathematical surfaces, comparing with mathematical solutions, but the results presented in this study are promising.

References

[1] Band, L.E., 1986, Topographic partition of watersheds with digital elevation models, *Water Resources Research*, **22**(1), pp 15-24.

[2] ESRI, 1991, *Cell-based Modelling with GRID*, Environmental System Research Institute, Redlands, CA.

[3] Freeman, T.G., 1991, Calculating catchment area with divergent flow based on a regular grid, *Computers & Geosciences*, **17**(3), pp 413-422.

[4] Holmgren, P., 1994, Multiple flow direction algorithms for runoff modelling in grid based elevation models: An empirical evaluation, *Hydrological processes*, **8**, pp 327-334.

[5] Mark, D.M., 1984, Automated detection of drainage networks from digital elevation models, *Cartographica*, **21**(2-3), pp 168-178.

[6] O’Callaghan, J.F. and Mark, D.M., 1984, The extraction of drainage networks from digital elevation data, *Computer Vision, Graphics, and Image Processing*, **28**, pp. 323-344.

[7] Olsson, L. and Pilesjö, P., 2002, Approaches to Spatially Distributed Hydrological Modelling in a GIS Environment. In Skidmore (Ed.) (2002) – *Environmental Modelling with GIS and Remote Sensing*. Taylor & Frances, London.

[8] Pilesjö, P., 1994, Estimation of drainage directions - A new method, in *Proceedings of the 7th Australasian Remote Sensing Conference*, 1-4 March, Melbourne, pp. 328-332.

[9] Pilesjö, P. and Zhou, Q., 1996, A multiple flow direction algorithm and its use for hydrological modelling, in *Geoinformatics’96 Proceedings*, April 26-28, West Palm Beach, FL, pp. 366-376.

[10] Pilesjö, P., Zhou, Q. and Harrie, L., 1998, Estimating flow distribution over Digital Elevation Models using a Form-Based Algorithm. *Geographic Information Science*, **4**, pp. 44-51.

[11] Quinn, P., Beven, K., Chevallier, P. and Planchon, O., 1991, The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrological Processes*, **5**, pp. 9-79.

[12] Wolock, D.M. and McCabe Jr., G.J., 1995, Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL, *Water Resources Research*, **31**(5), pp. 1315-1324.

[13] Zhou, Q. and Liu, X. (2002), Error assessment of grid-based flow routing algorithms used in hydrological models, *International Journal of Geographical Information Science*, 16(8): 819–842.

[14] MathWorks (2008). MATLAB R2008B. Natick, Massachusetts.