A least-cost path approach to stream delineation using lakes as patches and a digital elevation model as the cost surface

Melles S.J.\textsuperscript{a}, Jones, N.E.\textsuperscript{a}, Schmidt, B.\textsuperscript{a} Rayfield, B.\textsuperscript{b,a}\textsuperscript{*}

\textsuperscript{a}Ontario Ministry of Natural Resources, Trent University, 2140 Eastbank Drive, Peterborough, ON K9J 4B8, Canada
\textsuperscript{b}McGill University, 1205 Docteur Penfield, Montreal, QC H3A 1B1, Canada

Abstract

We present two approaches to predict movement of surface waters through heterogeneous terrain: least-cost graphs and circuits. These methods were compared with the commonly applied ‘deterministic eight’ (D8) approach by examining distances from extracted drainage networks to known-flow boundaries. Overlap statistics and classification accuracy estimates indicated that least-cost graphs and the circuit-based approaches hold some promise and avoid many of the pitfalls associated with removing depressions from a DEM. The majority of stream segments were correctly identified by all methods, but errors of omission and commission were fairly high.

Keywords: Drainage network extraction; channel network; graph theory; circuit theory; flow direction

1. Introduction

Predicting surface water movement over variable terrain is one of the principal goals of hydrologic modeling [1]. The most commonly applied method to route overland flow uses the ‘deterministic eight’ algorithm (D8), whereby flow direction for a central cell is determined by examining elevation in the eight neighbouring grid cells of a digital elevation model (DEM). Flow direction is assigned to the steepest down-slope neighbour [2,3,4]. Though several algorithms for extracting drainage networks from DEMs exist, most methods use neighbourhood information like the D8 algorithm. A typical problem encountered with these methods is that predicted flow can get trapped in depressions. Indeed depressions have been described as the ‘nemesis’ of neighbourhood methods [4]. Spurious pits without an outlet or
real lakes must be ‘filled’ before flow can continue, and several techniques to perform this task have been developed [4,1]. In addition, areas with little topographic relief are problematic for neighbourhood methods because flow direction cannot be determined for central cells within a flat area by examining surrounding cells alone. Flat areas, lakes, wetlands, and depressions are common in many regions of the world, and there is a real need to automate drainage network detection for these features.

Two recent approaches developed in the field of landscape ecology to predict connectivity in heterogeneous landscapes hold promise for detecting channel networks: they are surface connectivity models based on graph [5,6] and circuit theory [7]. Neither of these approaches has been applied to the problem of drainage network extraction using DEMs though least-cost paths are intuitively applicable; and similarities between the mathematical laws governing electrical processes and laws governing water movement through a watershed were recognized decades ago [8]. Rather, the majority of research into hydrologic flow algorithms has focused on problems related to routing flow through depressions or pits in a DEM and over flat areas [e.g., 9,10,11]. Other research has focused on problems of partitioning flow between neighbouring cells with equal elevations using multiple flow directions [12,13]. Jones et al. [2002, 13] addressed the problem of drainage extraction by developing a shortest-path approach through depressions using a weighted graph algorithm, the priority-first-search algorithm (PFS). This algorithm finds outlets for each pit iteratively using several criteria. Paths must travel through the lowest elevation cells, paths follow the greatest change in elevation, and paths must be of the shortest total length.

Our approaches are based on a novel treatment of lake and wetland patches as spatially explicit nodes [5]. The channel network is found by spreading outwards from all patches at once using a diffusion algorithm with the DEM as a cost (or resistance) surface [5,7]. Problems identified by [13,14] are avoided because graph-based least-cost paths or conductance circuits can be found without modifying the DEM to ‘fill’ depressions. Additionally, these methods may avoid the problems of parallel flow in flat areas and the issue of routing flow to only one of eight nearest neighbours. The latter issue is circumvented because diffusion occurs in all directions simultaneously and the cumulative costs of all paths between lakes are considered when the LCP is determined [5].

2. Methods

We compared four methods of stream delineation to known-flow based on a hydrologically correct surface water layer. These four methods were the D8 method, two graph-based least-cost path (LCP) analysis methods (using both MPG and CG, see nomenclature), and a method based on circuit theory. We applied a cost threshold to delineate streams only where least-cost flow paths were relatively ‘inexpensive’ (i.e., only links within or below one standard deviation of the mean distribution of flow path ‘costs’ were retained). A current map was used to estimate the location of stream flow based on circuit analysis. These methods were tested in a small area in south-central Ontario, Canada (Black and Hollow River watersheds, 2,289 ha, 10 m resolution) with little topographic relief (311-376 m, Fig. 1).

<table>
<thead>
<tr>
<th>Nomenclature &amp; definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM</td>
</tr>
<tr>
<td>D8</td>
</tr>
<tr>
<td>LCP</td>
</tr>
<tr>
<td>CG</td>
</tr>
<tr>
<td>MPG</td>
</tr>
</tbody>
</table>
Fig. 1. Comparison of known stream and predicted networks. Stream delineation was carried out using the D8 method and the proposed least-cost path approaches. The DEM used in these analyses was hydrologically corrected.

2.1. DEM & D8

An Ontario Provincial DEM (v2.0.0, 10 m resolution) was used in our analysis. This DEM was created using ANUDEM (v5.1.0) interpolation of base mapping contour data in conjunction with Spot heights (meters above sea level). ANUDEM interpolation allows inclusion of a flow-directed hydrology network as a boundary condition, which ensures that sinks do not intersect the known hydrology network [1]. However, all sinks are not removed by this method. Arc GIS – Arc Hydro was used to detect and fill spurious sinks in the DEM and to detect streams using the D8 flow routing algorithm [4].

2.2. Using graph-theory to model hydrologic flow

A hydrologic network can be modeled as a spatial graph consisting of a set of “nodes” (or patches) that represent lakes and “links” that represent the potential stream paths that connect them. In spatial graphs, nodes are 2-dimensional shapes with a fixed location, and links connect the nearest edges of nodes rather than their centroids [5]. The spatial location of each graph link can be predicted using least-cost path algorithms that minimize the cumulative cost of a link between pairs of nodes, based on the suitability of the intervening matrix. The DEM was translated into a cost surface by subtracting the minimum elevation from all values and rounding floating point values to nearest integers. Hydrologic graphs were produced using the spatial graph models implemented in SELES v3.3 [5,15].

2.3. Using circuit theory to model hydrologic flow

Circuit theory can be used to quantify the overall resistance in the watershed to hydrologic flow among lakes, creating a cumulative current map rather than a specific predicted stream path [7]. The current at a given grid cell corresponds to the density of random walkers that pass through that cell when moving among lakes across the watershed. Current maps were produced with the software Circuitscape v3.5 [16].

2.4. Comparison and overlay statistics

We examined the distance between stream cells delineated using the D8, two LCP algorithms, and the known-flow layer. This was done by determining the distance to known-flow for all cells in the study area, and then extracting values beneath D8 and least-cost delineated streams within a one cell buffer in
ArcGIS. Frequency distributions of distances to known-flow were created for both predicted stream paths and all off-path cells within the study area (excluding lakes and wetland patches). We expected that predicted paths would be closer to known-flow paths than to all other cells in the watershed. Predicted path distributions should have a median value close to zero (i.e. distance to known-flow is zero). Nonparametric Mann-Whitney tests were performed using R v2.12 [17] to test alternate hypotheses that a true shift in the median location of these distributions had occurred. For conductance maps, values below the known-flow layer (+/- one cell) were compared with conductance in off-known-flow path areas.

We also computed a direct overlap statistic, \(O_s\) [18] as a measure of number of candidate stream segments at the same location. A stream segment is defined as a flow path that is unbroken by tributaries, connections, or lakes and wetlands. Stream delineation accuracy was also estimated for both D8 and MPG LCP segments by calculating number of missing segments (omission errors) and number of spurious segments (commission errors). Only these two algorithms were compared in this way because the CG approach produced too many spurious segments. CG analysis will necessarily require thinning, but we did not apply a threshold-based approach to the CG results.

3. Results and discussion

Problem areas where streams fragments did not connect up with lakes (Fig. 1a-c, e.g. at character a), or where spurious streams were delineated (Fig. 1a-c, e.g. at character b), were evident in results for both D8 and least-cost path approaches (MPG & CG). Issues with parallel flow lines in flat areas were also apparent in both D8 and LCP approaches (Fig. 1a-c, e,g. at character c). The latter two problems (spurious streams and parallel flow paths) were amplified in the CG analysis (Fig. 1d). Both LCP analyses had visible edge effects (Fig 1cd, e.g. at character d), which produced stream segments in areas that would otherwise be disconnected given a larger study area.

LCP approaches correctly identified main stream segments in many areas as evidenced by the direct boundary overlap statistic \(O_s(MPG) = 24\), \(O_s(D8) = 21\), out of 37 possible known-stream segments). Omission and commission errors were fairly high, but this was true for the D8 approach as well. The MPG approach missed 13 known segments whereas the D8 algorithm missed 16 (omission error \((MPG) = 35\%\) and \(D8 = 43\%\)). In contrast, the MPG approach produced more spurious segments than the D8 approach (commission error \((MPG) = 41\%\) and \(D8 = 36\%\)). These measures of overlap and accuracy seem to contradict an apparent better visual agreement between the D8 approach and known-flow (Fig. 1). Boundary overlap statistics measure the intersection between overlapping stream segments on a feature by feature basis, but stream segments in our study area were not all similar in length. Shorter stream segments are weighted as heavily as longer ones. The same is true for our map comparison of boundary classification errors. Classification errors are generally estimated on a pixel-by-pixel basis in thematic map comparisons, but a feature-based map comparison is potentially more relevant in this case [19].

Main stream segments are also readily apparent in the circuit conductance map (Fig 2). Each cell in the conductance map represents current density such that areas with higher densities are more likely to hold flow, and current is negatively correlated with elevation. Conductance is perhaps a more intuitive representation of flow than graph-based LCP outputs. However, users would still have to select a threshold in order to extract a channel network. The D8 and graph-based LCP algorithms also rely on user input to set flow accumulation thresholds for stream extraction or to thin graph networks, and therefore all methods compared rely on at least one subjective decision.

Shifts in the frequency distributions for all approaches are shown in Figure 3. As expected, shifts in these distributions were considerable. LCP methods resulted in predicted streams that were closer to known-flow paths than off-path regions and shifts in these distributions were statistically significant (two-
tailed Mann-Whitney tests, Table 3). D8 delineated streams were about 10 m closer in median values to known-flows than streams delineated using the MPG approach (Table 3, Mann–Whitney test for differences in D8 and MPG distributions). Overall, the concordance between known-flow and the D8 method seemed to be slightly better according to visual estimation, frequency distributions, and Mann-Whitney tests. However, both LCP methods and the circuit-based approach produced very promising results. Median conductance values at known-flow paths were significantly higher than values in off-flow areas (Fig. 2, 3, Table 3). In fact all of the extreme circuit densities (values > 2, Fig. 3gh) were located at

Fig. 2. Current map used to estimate location of stream flow. Warmer colours indicate areas with higher current densities where stream flow is most probable.

Fig. 3. Frequency distributions summarizing distance to known-flow (a-f) and current (g-h) for graph-based and circuit-based approaches to stream delineation in comparison with the deterministic 8 (D8) method. Histograms of distances to known-flows for the areas off-stream paths are also shown.
Table 1. Non-parametric (Mann-Whitney) test results for shifts in the medians of frequency distributions shown in Figure 3.

<table>
<thead>
<tr>
<th>Frequency distribution</th>
<th>Median shift</th>
<th>95% CL around shift</th>
<th>$n_1$</th>
<th>$n_2$</th>
<th>Estimated $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8 path VS off-D8 path</td>
<td>-236.98</td>
<td>-242.07 to -230.83</td>
<td>3,973</td>
<td>19,373</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
</tr>
<tr>
<td>LCP (MPG) VS off-LCP (MPG)</td>
<td>-205.40</td>
<td>-210.42 to -200.00</td>
<td>4,949</td>
<td>19,274</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
</tr>
<tr>
<td>LCP (CG) VS off-LCP (CG)</td>
<td>-164.33</td>
<td>-167.56 to -161.21</td>
<td>17,189</td>
<td>18,049</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
</tr>
<tr>
<td>Conductance, known-flow VS off-known</td>
<td>0.62</td>
<td>0.61 to 0.64</td>
<td>3,974</td>
<td>19,370</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
</tr>
<tr>
<td>D8 path VS LCP (MPG) path</td>
<td>10.00009</td>
<td>9.99998 to 10.00007</td>
<td>3,973</td>
<td>4,949</td>
<td>$&lt; 2.2 \times 10^{-16}$</td>
</tr>
</tbody>
</table>

known-flow paths. Given these results, we believe that the proposed least-cost graph- and circuit-based approaches could have great utility in delineating streams in areas with many lakes and low topographic relief. Future research will test these methods across watershed extents and under varying topographies (flat, lake-dominated, wetland-dominated, etc.) to determine where they have advantages or are limited.

References


