Terrain Analysis Using Digital Elevation Models (TauDEM)

David Tarboton¹, Dan Watson², Rob Wallace,³ Kim Schreuders¹, Jeremy Neff¹

¹Utah Water Research Laboratory, Utah State University, Logan, Utah
²Computer Science, Utah State University, Logan, Utah
³US Army Engineer Research and Development Center, Information Technology Lab, Vicksburg, Mississippi

This research was funded by the US Army Research and Development Center under contract number W9124Z-08-P-0420
Deriving hydrologically useful information from Digital Elevation Models

Raw DEM

Pit Removal (Filling)

Flow Field

Flow Related Terrain Information
A parallel version of the TauDEM Software Tools

- Improved runtime efficiency
- Capability to run larger problems
- Platform independence of core functionality

Deployed as an ArcGIS Toolbox with tools that drive accompanying command line executables, available from
http://hydrology.usu.edu/taudem/
The challenge of increasing Digital Elevation Model (DEM) resolution

1980’s DMA 90 m
$10^2$ cells/km$^2$

1990’s USGS DEM 30 m
$10^3$ cells/km$^2$

2000’s NED 10-30 m
$10^4$ cells/km$^2$

2010’s LIDAR ~1 m
$10^6$ cells/km$^2$
Website and Demo

- [http://hydrology.usu.edu/taudem](http://hydrology.usu.edu/taudem)
Grid Data Format Assumptions

- Input and output grids are uncompressed GeoTIFF
- Maximum size 4 GB
- GDAL Nodata tag preferred (if not present, a missing value is assumed)
- Grids are square (Δx = Δy)
- Grids have identical in extent, cell size and spatial reference
- Spatial reference information is not used (no projection on the fly)
Representation of Flow Field

Steepest single direction

\[
\begin{array}{cc}
48 & 52 \\
56 & 67 \\
\end{array}
\]

Proportion flowing to neighboring grid cell 4 is \( a_1/(a_1+a_2) \)

Proportion flowing to neighboring grid cell 3 is \( a_2/(a_1+a_2) \)

Steepest direction downslope

Parallel Approach

- MPI, distributed memory paradigm
- Row oriented slices
- Each process includes one buffer row on either side
- Each process does not change buffer row
Illustrative Use Case: Delineation of channels and watersheds using a constant support area threshold

Steps

- Pit Remove
- D8 Flow Directions
- D8 Contributing Area
- Stream Definition by Threshold
- Stream Reach and Watershed
Pit Remove

Identifies all pits in the DEM and raises their elevation to the level of the lowest pour point around their edge. Pits are low elevation areas in digital elevation models (DEMs) that are completely surrounded by higher terrain. They are generally taken to be artifacts that interfere with the routing of flow across DEMs, so are

```
Executing: PitRemove logan \ Users\dtab\Scratch\Logan\loganfel.tif
Start Time: Mon Sep 20 20:02:33 2010
Executing: CalculateStatistics E:\Users\dtab\Scratch\Logan \loganfel.tif 1 1 # E:\Users\dtab\Scratch\Logan\loganfel.tif
Start Time: Mon Sep 20 20:02:33 2010
Executed (CalculateStatistics) successfully.
End Time: Mon Sep 20 20:02:33 2010 (Elapsed Time: 0.00 seconds)

Shell Command: mpiexec -np 6 "E:\Program Files (x86)\Taudem\TaudEM5\Exe\PitRemove" -z "E:\Users\dtab\Scratch\Logan\logan.tif" -fel "E:\Users\dtab\Scratch\Logan\loganfel.tif" > "E:\Users\dtab\Scratch\Logan\msgtmp.txt"
```

PitRemove version 5.0.4
Size: 8
Header read time: 0.059356
Data read time: 0.005522
Compute time: 0.234125
Write time: 0.077409
Total time: 0.386414

Executed (PitRemove) successfully.
End Time: Mon Sep 20 20:02:33 2010 (Elapsed Time: 2.00 seconds)
D8 Flow Direction (and Slope)

D8 Flow Direction

Calculates 2 grids. The first contains the D8 flow directions which are defined, for each cell, as the direction of the one of its eight adjacent or diagonal neighbors with the steepest downward slope. Flow Direction Coding: 1 - East, 2 - North East, 3 - North, 4 - North West, 5 - West, 6 - South West, 7 - South, 8 - South East. The

D8FlowDir version 5.0.4
Size: 8
Header read time: 0.027037
Data read time: 0.000796
Compute Slope time: 0.102044
Write Slope time: 0.054885
Resolve Flat time: 0.353239
Write Flat time: 0.047082
Total time: 0.624034

Executed [D8FlowDirections] successfully.
End Time: Mon Sep 20 20:11:16 2010 (Elapsed Time: 3.00 seconds)
D8 Contributing Area

Calculates a grid of contributing areas using the single direction D8 flow model. The contribution of each grid cell is taken as one (or when the optional weight grid is used, the value from the weight grid). The contributing area for each grid cell is taken as its own contribution plus the contribution from upslope neighbors that...
Stream Definition by Threshold

Stream Definition by Threshold

- Input Accumulated Stream Source Grid: loganad8
- Input Mask Grid (optional)
- Threshold: 2000
- Input Number of Processes: 8
- Output Stream Raster Grid: E:\Users\dtarb\Scratch\Logan\logansrc.tif

Stream Definition by Threshold

Operates on any grid and outputs an indicator (1,0) grid identifying cells with input values >= the threshold value. The standard use is to use an accumulated source area grid to as the input grid to generate a stream raster grid as the output. If you use the optional input mask grid, it limits the domain being evaluated to cells with mask values >= 0. When you use a D-infinity contribution area grid, another option is to set the mask to 0 to identify a drainage divide.
Stream Reach and Watershed

Stream Reach and Watershed

This function produces a vector network and shapefile from the stream raster grid. The flow direction grid is used to connect flow paths along the stream raster. The Strahler order of each stream segment is computed. The subwatershed draining to each stream segment (reach) is also delineated and labeled with the value identifier that corresponds to the WSNO (watershed number) attribute in the stream reach shapefile.
Some Algorithm Details

Pit Removal: Planchon Fill Algorithm

Planchon, O., and F. Darboux (2001), A fast, simple and versatile algorithm to fill the depressions of digital elevation models, *Catena*(46), 159-176.
Parallel Scheme

Initialize( Z,F)

Do
  for all grid cells i
    if Z(i) > n
      F(i) ← Z(i)
    Else
      F(i) ← n
      i on stack for next pass
  endfor
Send( topRow, rank-1 )
Send( bottomRow, rank+1 )
Recv( rowBelow, rank+1 )
Recv( rowAbove, rank-1 )

Until F is not modified

Z denotes the original elevation.
F denotes the pit filled elevation.
n denotes lowest neighboring elevation
i denotes the cell being evaluated

Iterate only over stack of changeable cells
1. Dependency grid

Executed by every process with grid flow field P, grid dependencies D initialized to 0 and an empty queue Q.

**FindDependencies**\( (P,Q,D) \)

for all \( i \)

- for all k neighbors of \( i \)
  - if \( P_{ki} > 0 \) \( D(i) = D(i) + 1 \)
  - if \( D(i) = 0 \) add \( i \) to Q

next

2. Flow algebra function

Executed by every process with D and Q initialized from FindDependencies.

**FlowAlgebra**\( (P,Q,D,\theta,\gamma) \)

while Q isn’t empty

- get \( i \) from Q
- \( \theta_i = FA(\gamma_i, P_{ki}, \theta_k, \gamma_k) \)
- for each downslope neighbor \( n \) of \( i \)
  - if \( P_{in} > 0 \)
    - \( D(n) = D(n) - 1 \)
  - if \( D(n) = 0 \)
    - add \( n \) to Q

next \( n \)
end while

swap process buffers and repeat
## Capabilities Summary

### Capability to run larger problems

<table>
<thead>
<tr>
<th>Year</th>
<th>Version</th>
<th>Processors used</th>
<th>Grid size</th>
<th>Theoretical limit</th>
<th>Largest run</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>TauDEM 4</td>
<td>1</td>
<td>0.22 GB</td>
<td>0.22 GB</td>
<td>0.22 GB</td>
</tr>
<tr>
<td>Sept 2009</td>
<td>Partial implementation</td>
<td>8</td>
<td>4 GB</td>
<td>1.6 GB</td>
<td>1.6 GB</td>
</tr>
<tr>
<td>June 2010</td>
<td>TauDEM 5</td>
<td>8</td>
<td>4 GB</td>
<td>4 GB</td>
<td>4 GB</td>
</tr>
<tr>
<td>Sept 2010</td>
<td>Multifile on 48 GB RAM PC</td>
<td>4</td>
<td>Hardware limits</td>
<td>6 GB</td>
<td>6 GB</td>
</tr>
<tr>
<td>Sept 2010</td>
<td>Multifile on cluster with 128 GB RAM</td>
<td>128</td>
<td>Hardware limits</td>
<td>11 GB</td>
<td>11 GB</td>
</tr>
</tbody>
</table>

Single file size limit 4GB

At 10 m grid cell size
Improved runtime efficiency

Parallel Pit Remove timing for NEDB test dataset (14849 x 27174 cells \(\approx 1.6\) GB).

8 processor PC
Dual quad-core Xeon E5405 2.0GHz PC with 16GB RAM

128 processor cluster
16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
Improved runtime efficiency

Parallel D-Infinity Contributing Area Timing for Boise River dataset (24856 x 24000 cells ~ 2.4 GB)

8 processor PC
Dual quad-core Xeon E5405 2.0GHz PC with 16GB RAM

128 processor cluster
16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
## Scaling of run times to large grids

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size (GB)</th>
<th>Hardware</th>
<th>Number of Processors</th>
<th>PitRemove (run time seconds)</th>
<th>D8FlowDir (run time seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Compute</td>
<td>Total</td>
</tr>
<tr>
<td>GSL100</td>
<td>0.12</td>
<td>Owl (PC)</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>GSL100</td>
<td>0.12</td>
<td>Rex (Cluster)</td>
<td>8</td>
<td>28</td>
<td>360</td>
</tr>
<tr>
<td>GSL100</td>
<td>0.12</td>
<td>Rex (Cluster)</td>
<td>64</td>
<td>10</td>
<td>256</td>
</tr>
<tr>
<td>GSL100</td>
<td>0.12</td>
<td>Mac</td>
<td>8</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Yellowstone</td>
<td>2.14</td>
<td>Owl (PC)</td>
<td>8</td>
<td>529</td>
<td>681</td>
</tr>
<tr>
<td>Yellowstone</td>
<td>2.14</td>
<td>Rex (Cluster)</td>
<td>64</td>
<td>140</td>
<td>3759</td>
</tr>
<tr>
<td>Boise River</td>
<td>4</td>
<td>Owl (PC)</td>
<td>8</td>
<td>4818</td>
<td>6225</td>
</tr>
<tr>
<td>Boise River</td>
<td>4</td>
<td>Virtual (PC)</td>
<td>4</td>
<td>1502</td>
<td>2120</td>
</tr>
<tr>
<td>Bear/Jordan/Weber</td>
<td>6</td>
<td>Virtual (PC)</td>
<td>4</td>
<td>4780</td>
<td>5695</td>
</tr>
<tr>
<td>Chesapeake</td>
<td>11.3</td>
<td>Rex (Cluster)</td>
<td>64</td>
<td>702</td>
<td>24045</td>
</tr>
</tbody>
</table>

1. Owl is an 8 core PC (Dual quad-core Xeon E5405 2.0GHz) with 16GB RAM
2. Rex is a 128 core cluster of 16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
3. Virtual is a virtual PC resourced with 48 GB RAM and 4 Intel Xeon E5450 3 GHz processors
4. Mac is an 8 core (Dual quad-core Intel Xeon E5620 2.26 GHz) with 16GB RAM
1. Owl is an 8 core PC (Dual quad-core Xeon E5405 2.0GHz) with 16GB RAM
2. Rex is a 128 core cluster of 16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM
3. Virtual is a virtual PC resourced with 48 GB RAM and 4 Intel Xeon E5450 3 GHz processors
Summary and Conclusions

- Parallelization speeds up processing and partitioned processing reduces size limitations
- Parallel logic developed for general recursive flow accumulation methodology (flow algebra)
- Documented ArcGIS Toolbox Graphical User Interface
- 32 and 64 bit versions (but 32 bit version limited by inherent 32 bit operating system memory limitations)
- PC, Mac and Linux/Unix capability
- Capability to process large grids efficiently increased from 0.22 GB upper limit pre-project to where < 4GB grids can be processed in the ArcGIS Toolbox version on a PC within a day and up to 11 GB has been processed on a distributed cluster (a 50 fold size increase)
Limitations and Dependencies

- Uses MPICH2 library from Argonne National Laboratory
  http://www.mcs.anl.gov/research/projects/mpich2/

- TIFF (GeoTIFF) 4 GB file size (for single file version)

- [Prototype with capability to use multiple files to cover domain not yet on web site]

- Processor memory
Additional Illustrative Use Cases

- Start with a DEM and end up with a topographic wetness index from the Dinfinity method.
- Start with a DEM and end up with a delineation of channels and watersheds that are sensitive to spatial variability in topographic texture with spatially variable drainage density using the Peucker Douglas approach and channelization threshold objectively chosen by drop analysis.
- Flow algebra functions (Transport limited accumulation, decaying accumulation, upslope dependence, distances up and down, avalanche runout).
Topographic wetness index from the Dinfinity method

Steps

- Pit Remove
- D-Infinity Flow Directions (and Slopes)
- D-Infinity Contributing Area
- Slope Over Area Ratio
Proportion flowing to neighboring grid cell 4 is \( \alpha_1/(\alpha_1 + \alpha_2) \)

Steepest direction downslope
Proportion flowing to neighboring grid cell 3 is \( \alpha_2/(\alpha_1 + \alpha_2) \)

Flow direction.

D-Infinity Flow Direction (and slope)
D-Infinity Contributing Area

Calculates a grid of contributing area using the multiple flow direction D-infinity approach. D-infinity flow direction is defined as steepest downward slope on planar triangular facets on a block centered grid. The contribution at each grid cell is taken initially as the grid cell length (or when the optional weight grid is used, the cell weight) times the drainage area of the upslope element.
Wetness Index

\[ \ln(a/S) \]
Channels and watersheds with spatially variable drainage density using Peucker Douglas threshold objectively chosen by drop analysis

Steps

- Pit Remove
- D8 Flow Directions
- D8 Contributing Area
- Peuker Douglas
- Weighted D8 Contributing Area
- Stream Drop Analysis
- Stream Definition by Threshold
- Stream Reach and Watershed
Local Valley Form Computation
Weighted D8 Contributing Area

Contributing area only of valley form grid cells upstream of outlet
Stream Drop Analysis

Applies a series of thresholds (determined from the input parameters) to the input accumulated stream source grid (*ssg) and outputs the results in the *drp.txt file the stream drop statistics table. This function is designed to aid in the determination of a geomorphologically objective threshold to be used to delineate streams. Drop Analysis attempts to select the right threshold automatically by evaluating:

- Use logarithmic spacing for threshold values
- Number of threshold values
- Minimum Threshold Value
- Maximum Threshold Value

Example output:

<table>
<thead>
<tr>
<th>Drop Analysis version 5.0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold DrainDen NoFirstOrd NoHighOrd MeanDFirstOrd MeanDHighOrd StdDevFirstOrd StdDevHighOrd Tval</td>
</tr>
<tr>
<td>5.000000 0.002461 22.56 600 66.500453 125.04446 76.240034 101.67966 -14.564702</td>
</tr>
<tr>
<td>5.340503 0.001854 11.65 351 55.638748 145.378479 97.830666 142.423080 -8.938377</td>
</tr>
<tr>
<td>13.912796 0.001537 77.4 239 96.581466 159.673108 103.33026 151.38107 -7.345115</td>
</tr>
<tr>
<td>23.837947 0.001226 45.2 141 115.005356 182.002914 109.692078 150.783463 -5.642063</td>
</tr>
<tr>
<td>38.713192 0.000999 294 96 116.624161 211.537094 107.394669 166.852936 -5.479936</td>
</tr>
<tr>
<td>54.577499 0.000790 186 70 116.728371 209.407593 123.760880 156.084854 -4.967545</td>
</tr>
<tr>
<td>107.721756 0.000635 109 38 153.991043 239.083878 144.086898 162.634705 -3.03640</td>
</tr>
<tr>
<td>179.690720 0.000524 75 19 157.208069 269.439911 158.242188 156.966827 -2.022420</td>
</tr>
<tr>
<td>299.742216 0.000412 50 14 197.516684 255.433441 137.707306 168.146484 -1.324365</td>
</tr>
<tr>
<td>599.591163 0.000304 30 4 214.549347 289.485138 153.106644 135.973572 -0.598755</td>
</tr>
</tbody>
</table>

Value for optimum that drop analysis selected - see output file for details.

Processes: 6
Read time: 0.142451
Compute time: 1.028273
Total time: 1.170725

Executed (StreamDropAnalysis) successfully.

End Time: Mon Sep 20 15:26:59 2010 (Elapsed Time: 2.00 seconds)
How to decide on stream delineation threshold?

Drainage density (total channel length divided by drainage area) as a function of drainage area support threshold used to define channels for the three study watersheds.

Why is it important?
Hydrologic processes are different on hillslopes and in channels. It is important to recognize this and account for this in models.

Drainage area can be concentrated or dispersed (specific catchment area) representing concentrated or dispersed flow.
Hortons Laws: Strahler system for stream ordering
Constant Stream Drops Law

Rd = 0.944

Stream Drop

Elevation difference between ends of stream

Note that a “Strahler stream” comprises a sequence of links (reaches or segments) of the same order
Suggestion: Map channel networks from the DEM at the finest resolution consistent with observed channel network geomorphology ‘laws’.

- Look for statistically significant break in constant stream drop property as stream delineation threshold is reduced
- Break in slope versus contributing area relationship
- Physical basis in the form instability theory of Smith and Bretherton (1972), see Tarboton et al. 1992
Statistical Analysis of Stream Drops

Elevation Drop for Streams

Drop (meters)

Strahler Order
T-Test for Difference in Mean Values

<table>
<thead>
<tr>
<th></th>
<th>Order 1</th>
<th>Order 2-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean X</td>
<td>72.2</td>
<td>130.3</td>
</tr>
<tr>
<td>Std X</td>
<td>68.8</td>
<td>120.8</td>
</tr>
<tr>
<td>Var X</td>
<td>4740.0</td>
<td>14594.5</td>
</tr>
<tr>
<td>Nx</td>
<td>268</td>
<td>81</td>
</tr>
</tbody>
</table>

T-test checks whether difference in means is large (> 2) when compared to the spread of the data around the mean values.
Stream Definition by Threshold

Stream Definition by Threshold

Operates on any grid and outputs an indicator (1,0) grid identifying cells with input values $\geq$ the threshold value. The standard use is to use an accumulated source area grid to as the input grid to generate a stream raster grid as the output. If you use the optional input mask grid, it limits the domain being evaluated to cells with mask values $\geq 0$. When you use a D-infinity contributing area grid (*sca) as the mask grid, it functions as an edge contamination mask. The threshold logic is: $\text{src} = ((\text{sca} \geq \text{thresh}) \& (\text{mask} \geq 0)) ? 1:0$.
Stream Reach and Watershed

Output Stream Reach Shapefile

This output is a polyline shapefile giving the links in a stream network. The columns in the attribute table are:

- LINKNO: Link Number. A unique number associated with each link (segment of channel between junctions). This is arbitrary and will vary depending on number of processes used.
- DSLINKNO: Downstream Link Number of the downstream link. -1 indicates that this does not exist.
- USLINNO1: Link Number of first upstream link. (0 indicates no link upstream, i.e., for a source link)
- USLINNO2: Link Number of second upstream link. (0 indicates no second link upstream, i.e., for a source link or an internal monitoring point where the reach is logically split but the network does not bifurcate.)
- NODIDENT: Node Identifier for node at
Illustration of some other functions

Transport limited accumulation

Supply | Capacity | Transport | Deposition
--- | --- | --- | ---

\[
S = \chi a^2 \tan(b)^2
\]

\[
T_{cap} = \chi a^2 \tan(b)^2
\]

\[
T_{out} = \min\{S + \sum T_{in}, T_{cap}\}
\]

\[
D = S + \sum T_{in} - T_{out}
\]

Useful for modeling erosion and sediment delivery, the spatial dependence of sediment delivery ratio and contaminant that adheres to sediment
Decaying Accumulation

A decayed accumulation operator DA[.] takes as input a mass loading field m(x) expressed at each grid location as m(i, j) that is assumed to move with the flow field but is subject to first order decay in moving from cell to cell. The output is the accumulated mass at each location DA(x). The accumulation of m at each grid cell can be numerically evaluated

\[
DA[m(x)] = DA(i, j) = m(i, j)\Delta^2 + \sum_{\text{k contributing neighbors}} p_k d(i_k, j_k) DA(i_k, j_k)
\]

Here d(x) = d(i, j) is a decay multiplier giving the fractional (first order) reduction in mass in moving from grid cell x to the next downslope cell. If travel (or residence) times t(x) associated with flow between cells are available d(x) may be evaluated as \(\exp(-\lambda t(x))\) where \(\lambda\) is a first order decay parameter.

Useful for a tracking contaminant or compound subject to decay or attenuation
Dependence function. Quantifies the amount a point $x$ contributes to the point or zone $y$.

Useful for example to track where a contaminant may come from.
Distance Down and Distance Up

Types of distance measurements possible in distance down and distance up functions.
Upslope recursion to determine elevation and distance to point in trigger zone that has the highest alpha angle.