

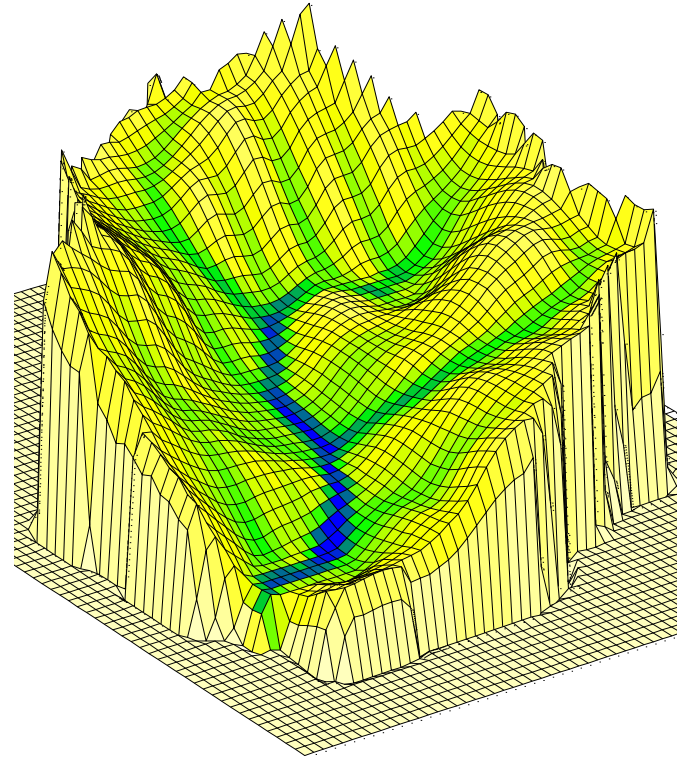
# Terrain Analysis Using Digital Elevation Models (TauDEM)

David Tarboton<sup>1</sup>, Dan Watson<sup>2</sup>,  
Rob Wallace,<sup>3</sup> Kim Schreuders<sup>1</sup>,  
Jeremy Neff<sup>1</sup>

<sup>1</sup>Utah Water Research Laboratory, Utah State  
University, Logan, Utah

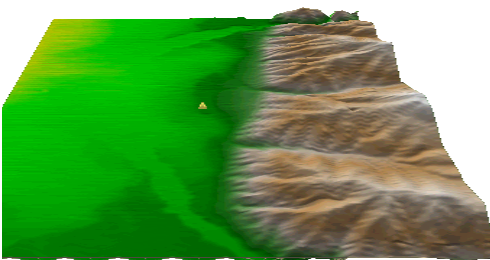
<sup>2</sup>Computer Science, Utah State University,  
Logan, Utah

<sup>3</sup>US Army Engineer Research and Development  
Center, Information Technology Lab, Vicksburg,  
Mississippi



# Deriving hydrologically useful information from Digital Elevation Models

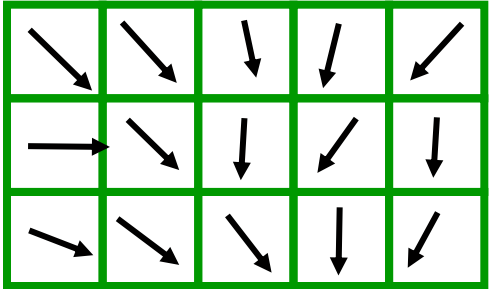
Raw DEM



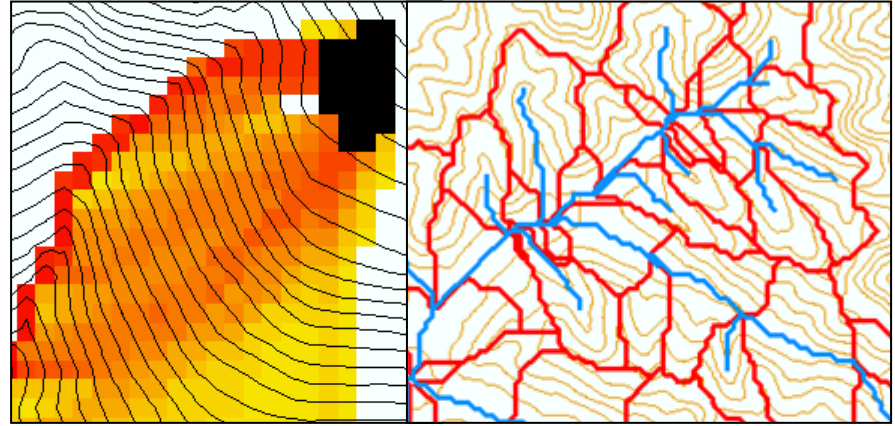
Pit Removal (Filling)



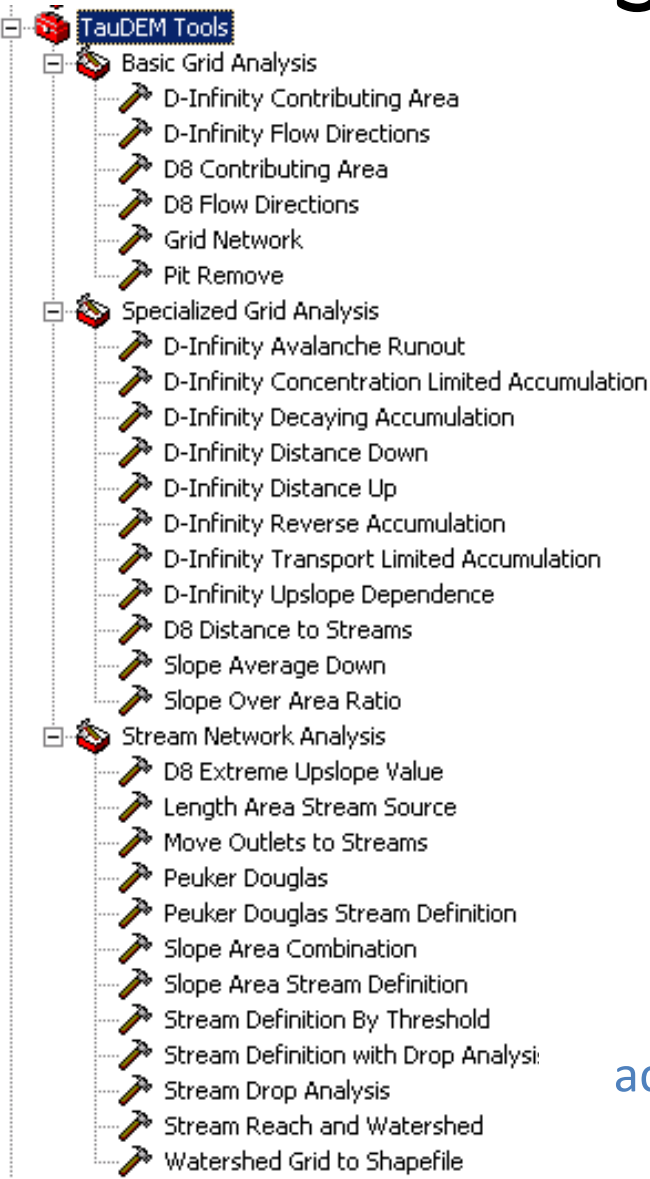
Flow Field



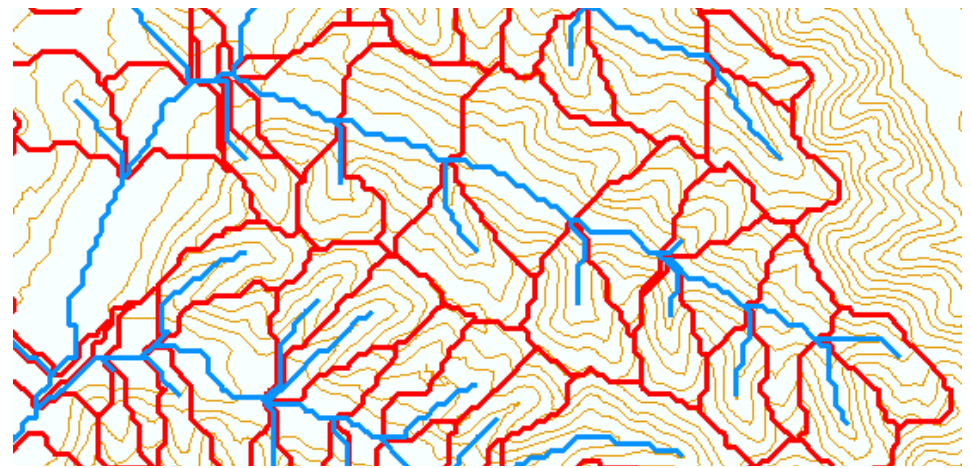
Flow Related Terrain Information



# 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<sup>2</sup>

1990's USGS DEM 30 m

$10^3$  cells/km<sup>2</sup>

2000's NED 10-30 m

$10^4$  cells/km<sup>2</sup>

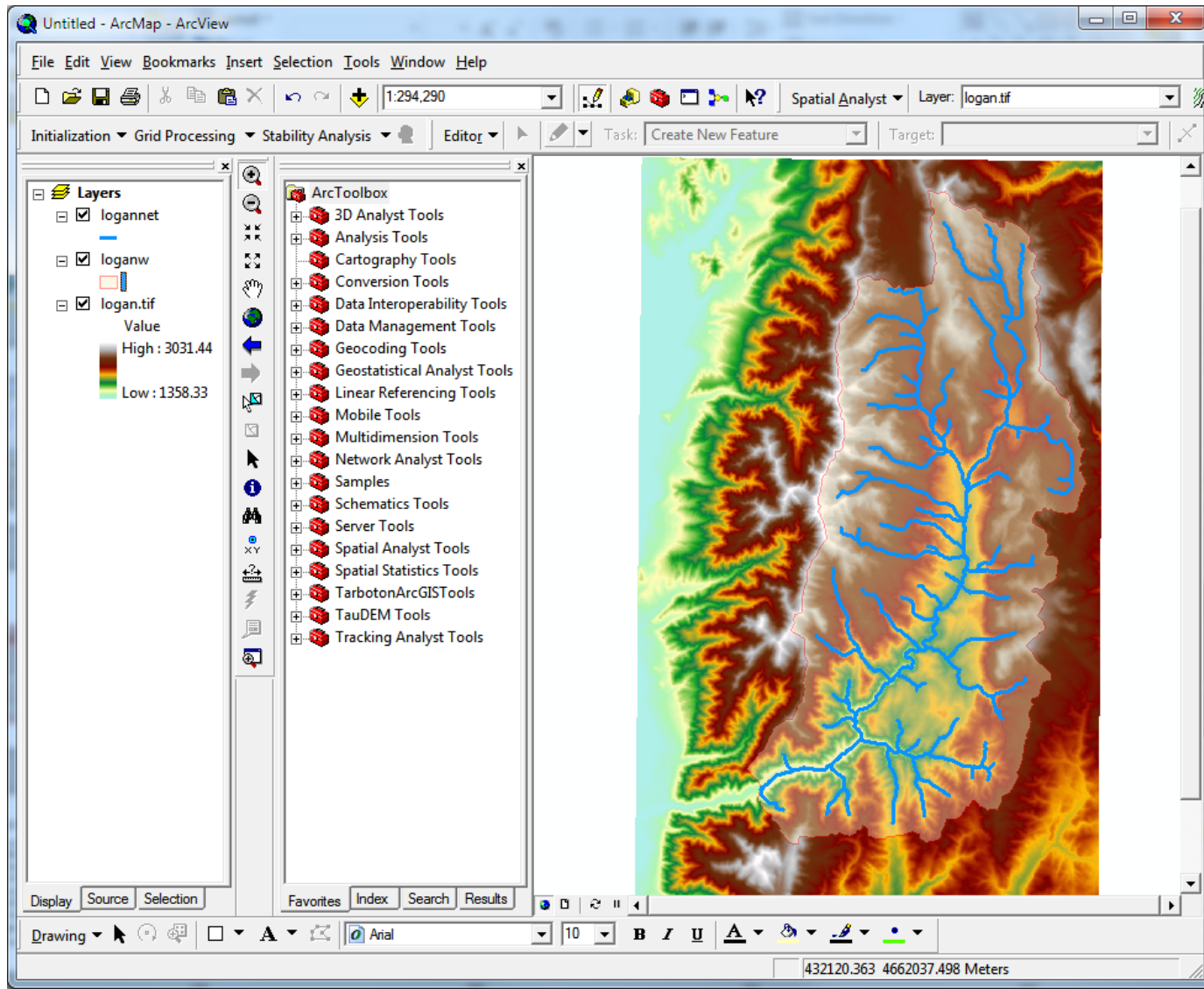
2010's LIDAR ~1 m

$10^6$  cells/km<sup>2</sup>



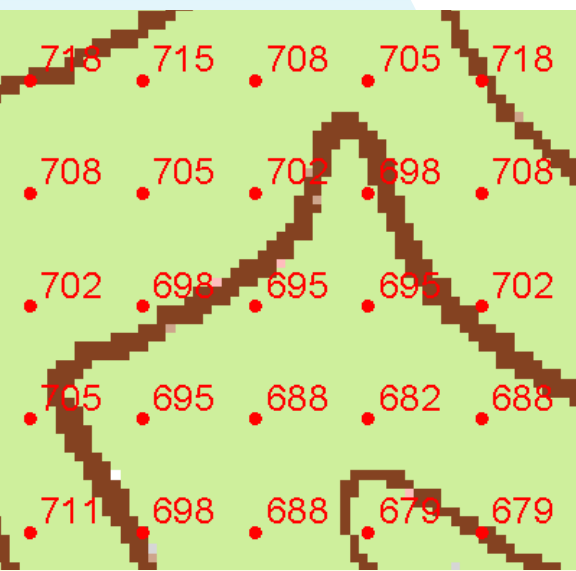
# Website and Demo

- <http://hydrology.usu.edu/taudem>





# Grid Data Format Assumptions



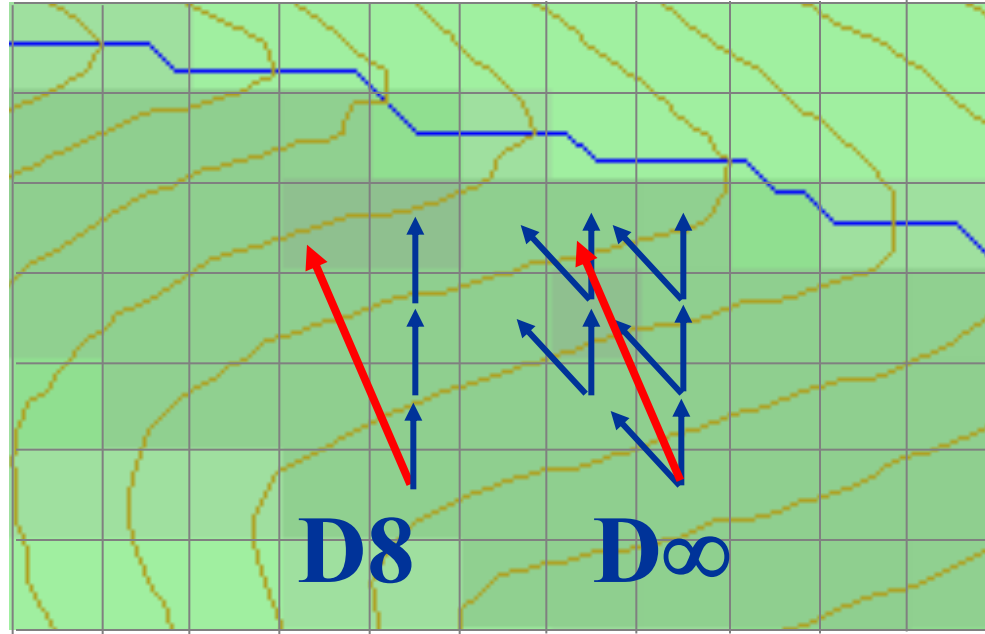
- Input and output grids are uncompressed GeoTIFF
- Maximum size 4 GB
- GDAL Nodata tag preferred (if not present, a missing value is assumed)
- Grids are square ( $\Delta x = \Delta y$ )
- Grids have identical in extent, cell size and spatial reference
- Spatial reference information is not used (no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

# Representation of Flow Field

Steepest  
single  
direction

48	52
56	67

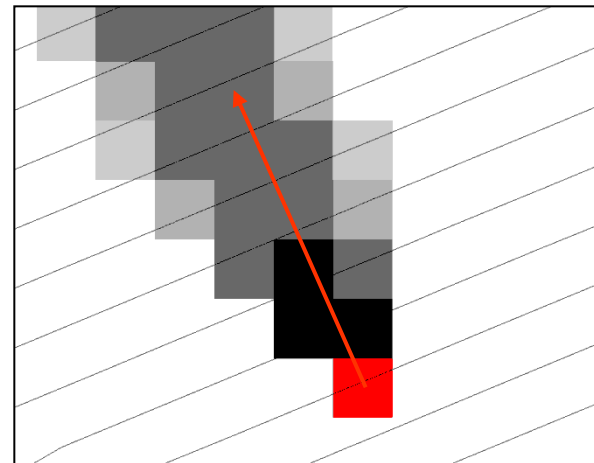
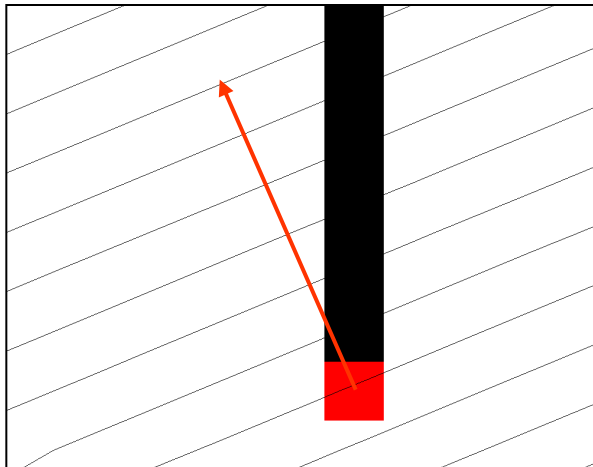
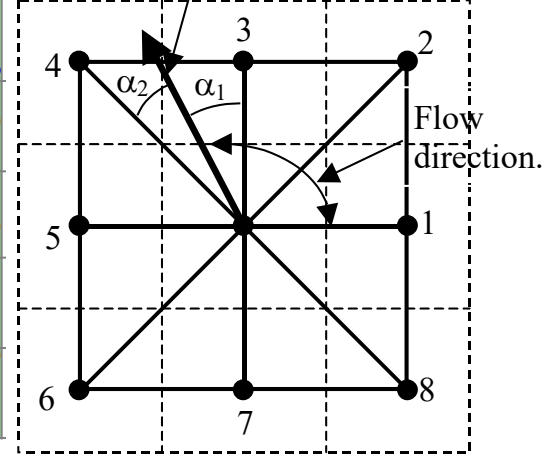
$$\frac{67 - 52}{30} = 0.50$$



Proportion  
flowing to  
neighboring  
grid cell 4 is  
 $\alpha_1/(\alpha_1+\alpha_2)$

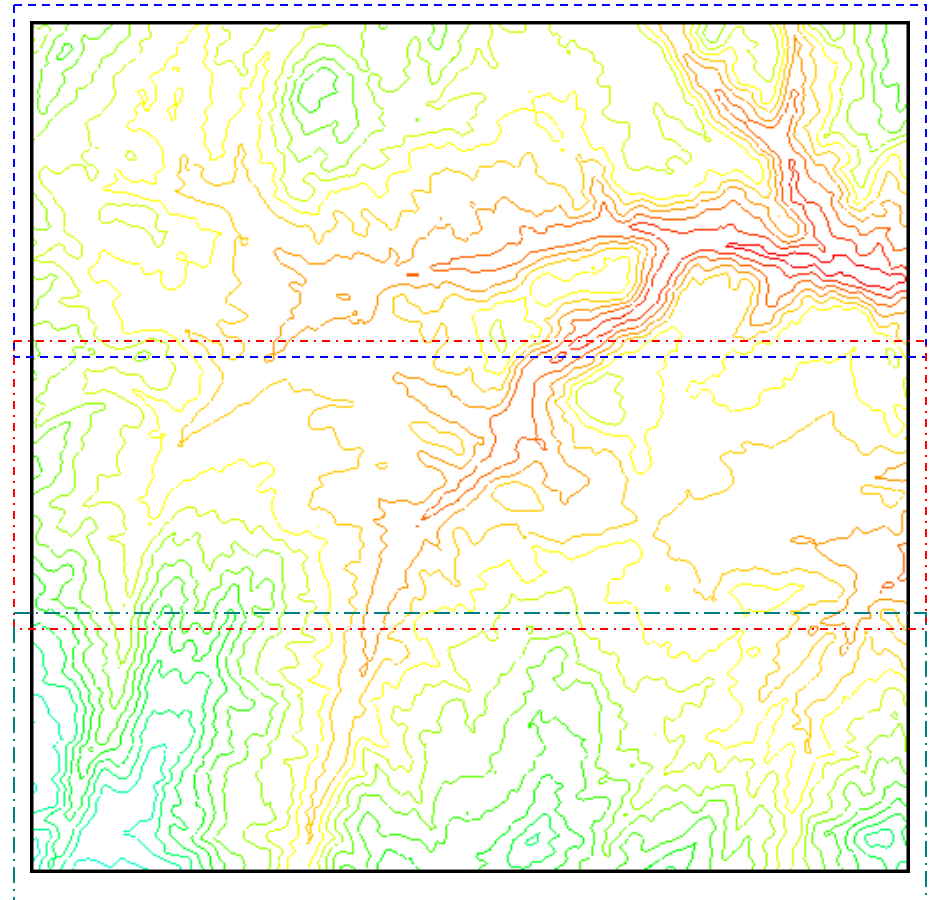
Steepest direction  
downslope

Proportion flowing to  
neighboring grid cell 3  
is  $\alpha_2/(\alpha_1+\alpha_2)$



# 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

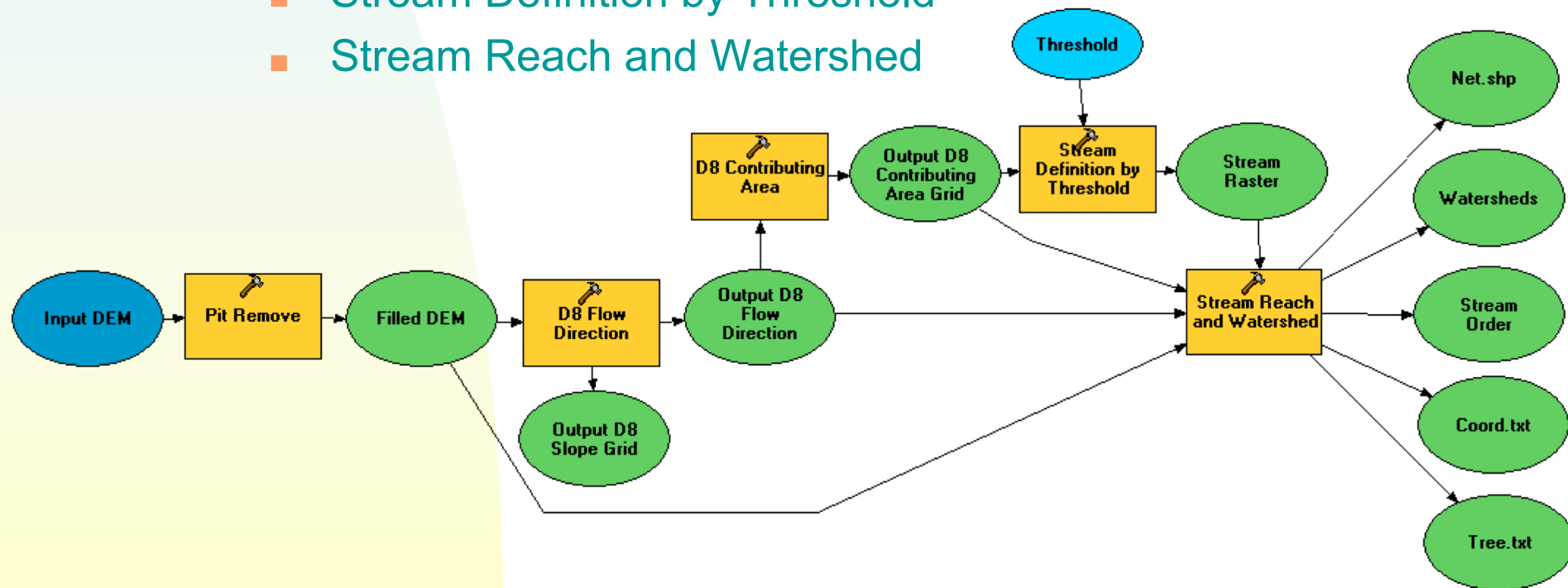


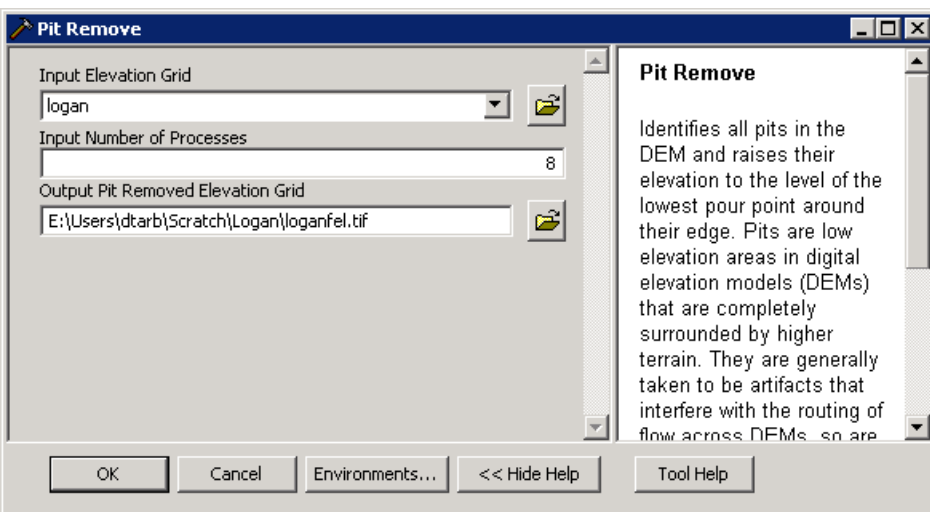


# Illustrative Use Case: Delineation of channels and watersheds using a constant support area threshold

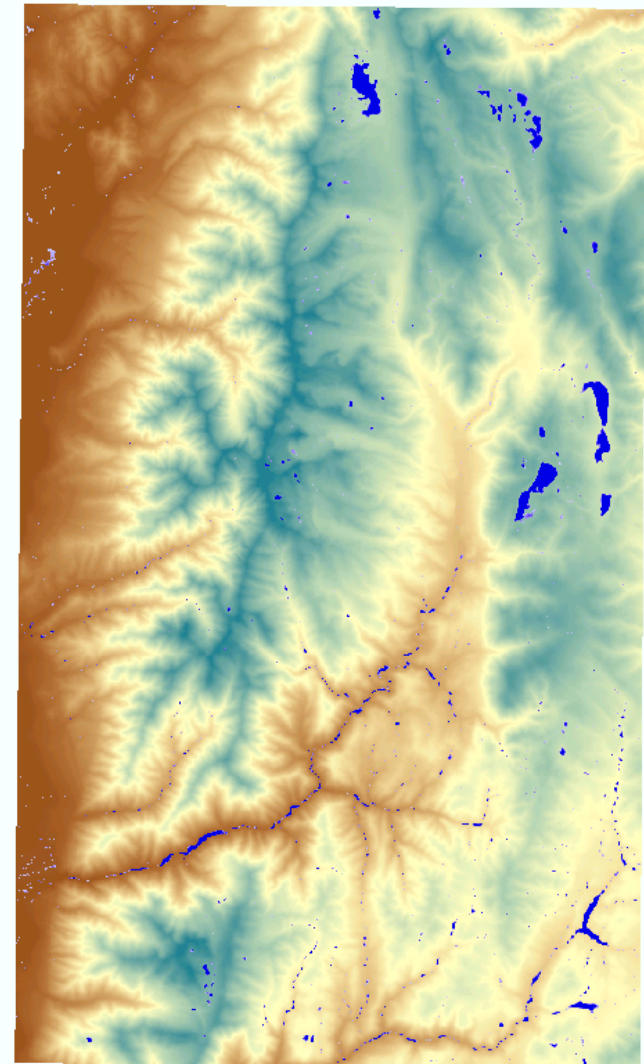
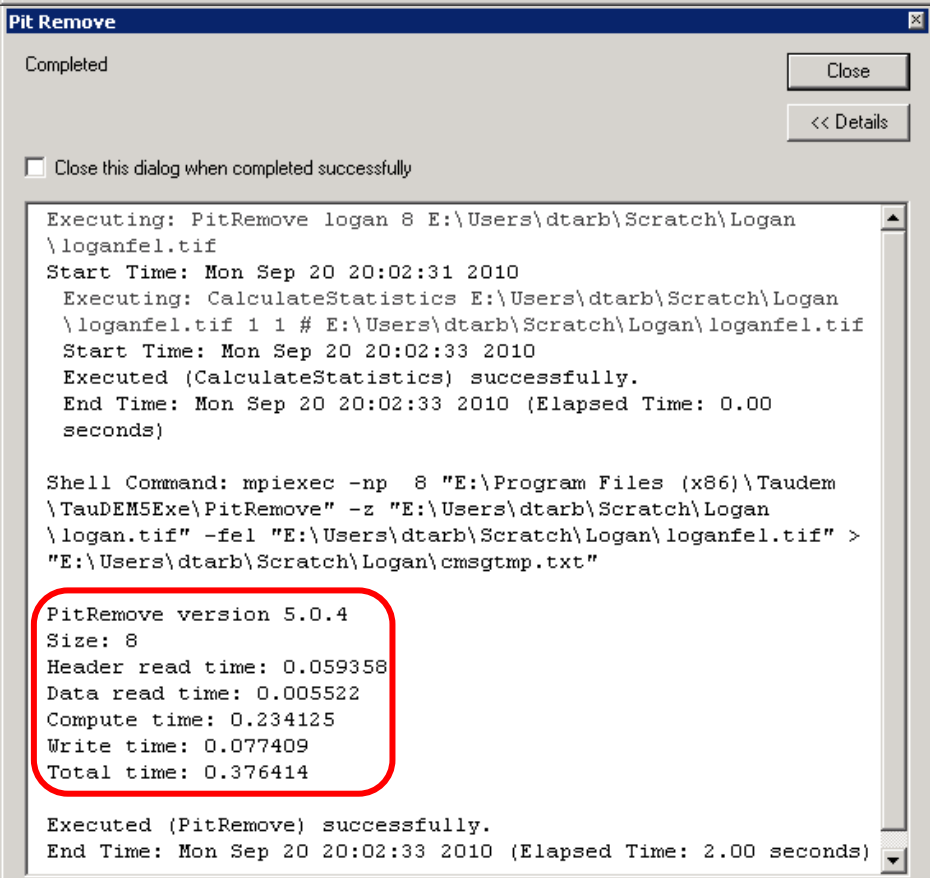
## Steps

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

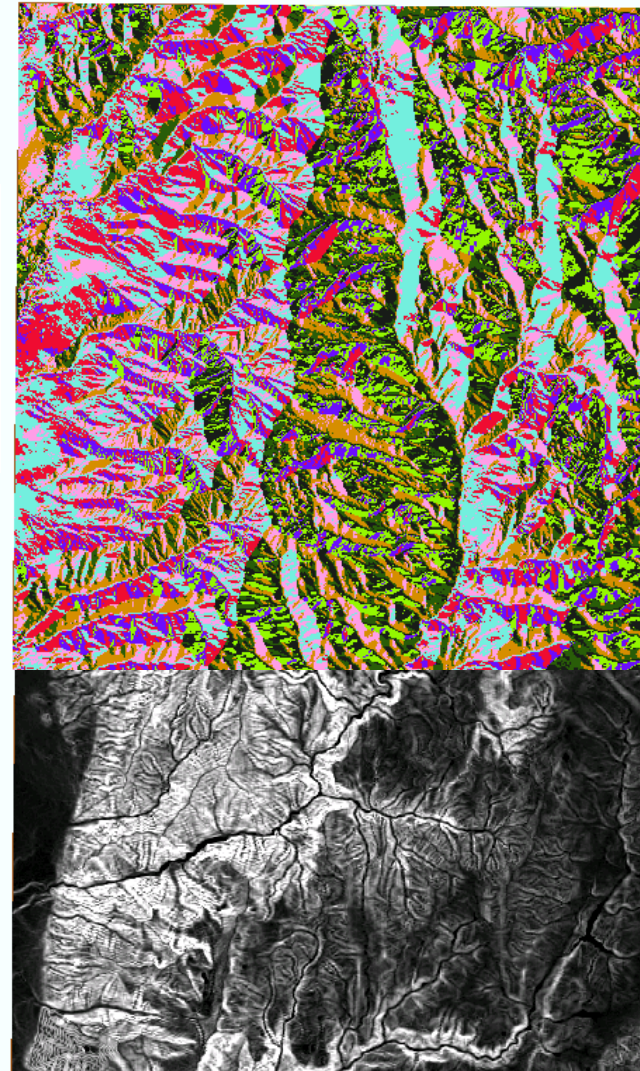
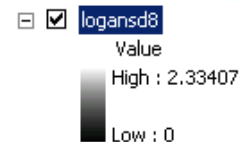
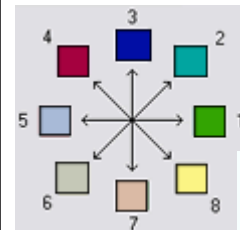
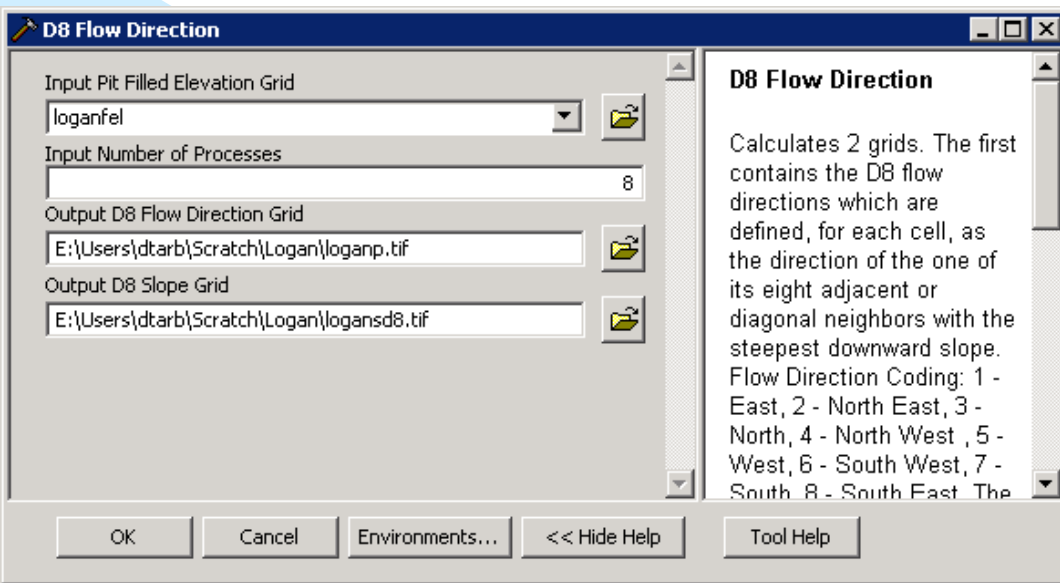




# Pit Remove

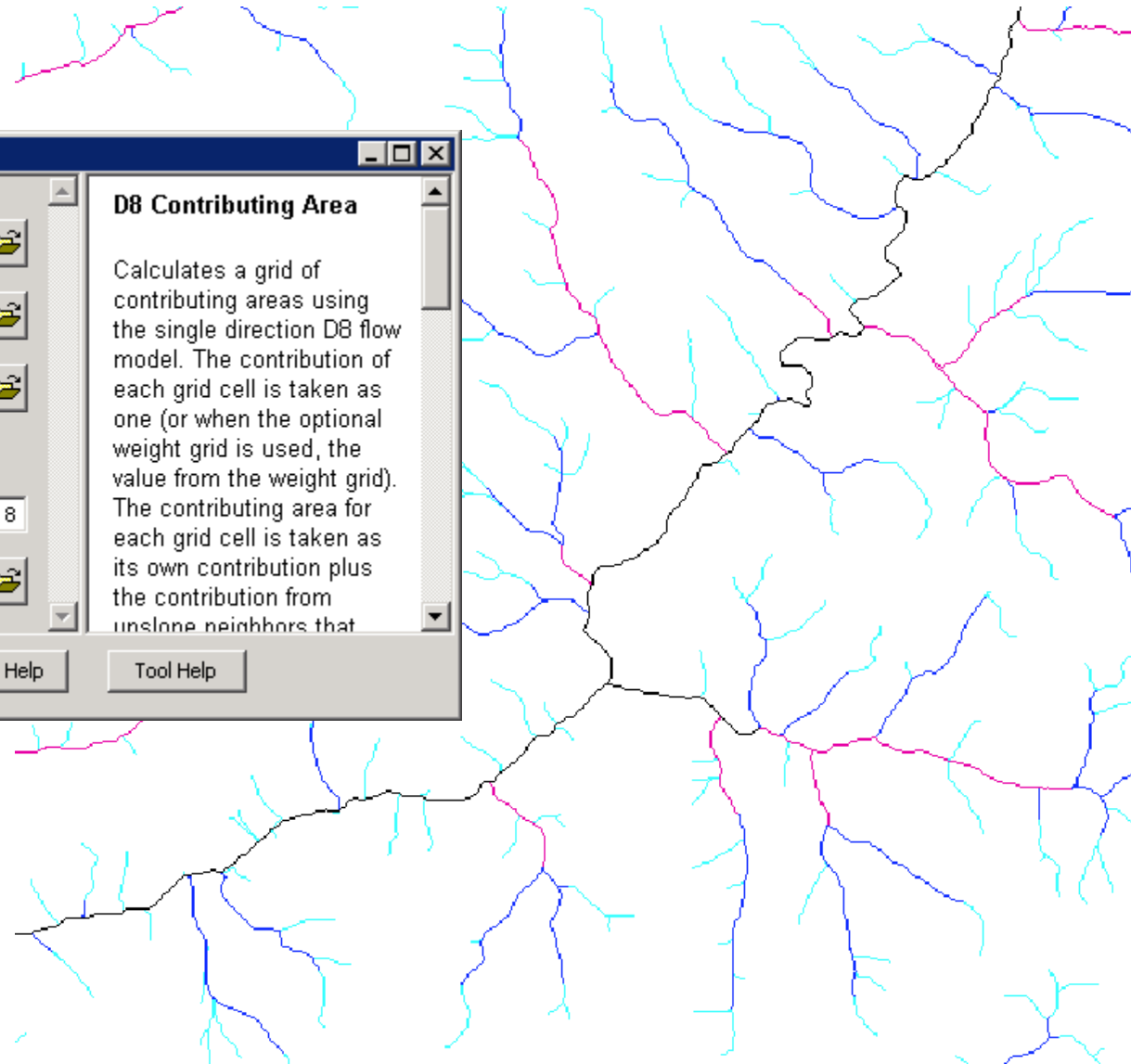
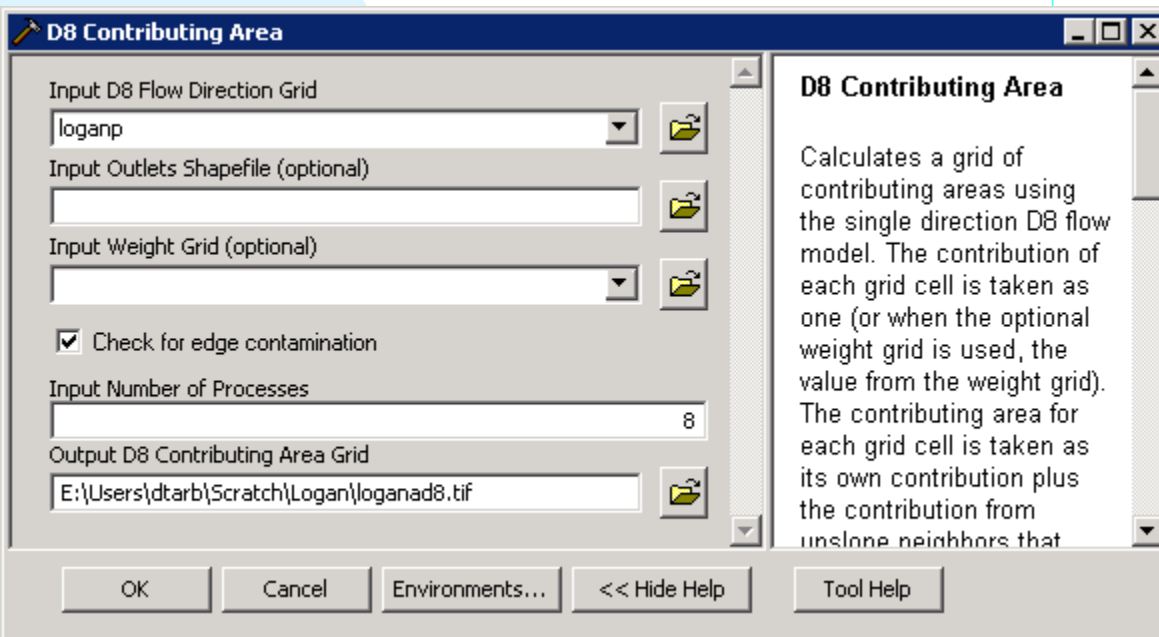


# D8 Flow Direction (and Slope)



```
Executed (CalculateStatistics) successfully.  
End Time: Mon Sep 20 20:11:18 2010 (Elapsed Time: 1.00  
seconds)  
  
Shell Command: mpiexec -np 8 "E:\Program Files (x86)\Taudem  
\TauDEM5Exe\D8FlowDir" -fel "E:\Users\dtarb\Scratch\Logan  
\loganfcl.tif" -p "E:\Users\dtarb\Scratch\Logan\loganp.tif" -  
sd8 "E:\Users\dtarb\Scratch\Logan\logansd8.tif" > "E:\Users  
\dtarb\Scratch\Logan\cmsgtmp.txt"  
  
D8FlowDir version 5.0.4  
Size: 8  
Header read time: 0.027037  
Data read time: 0.009796  
Compute Slope time: 0.102044  
Write Slope time: 0.084885  
Resolve Flat time: 0.353239  
Write Flat time: 0.047082  
Total time: 0.624084  
  
Executed (D8FlowDirections) successfully.  
End Time: Mon Sep 20 20:11:18 2010 (Elapsed Time: 3.00 seconds)
```

# D8 Contributing Area



# 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

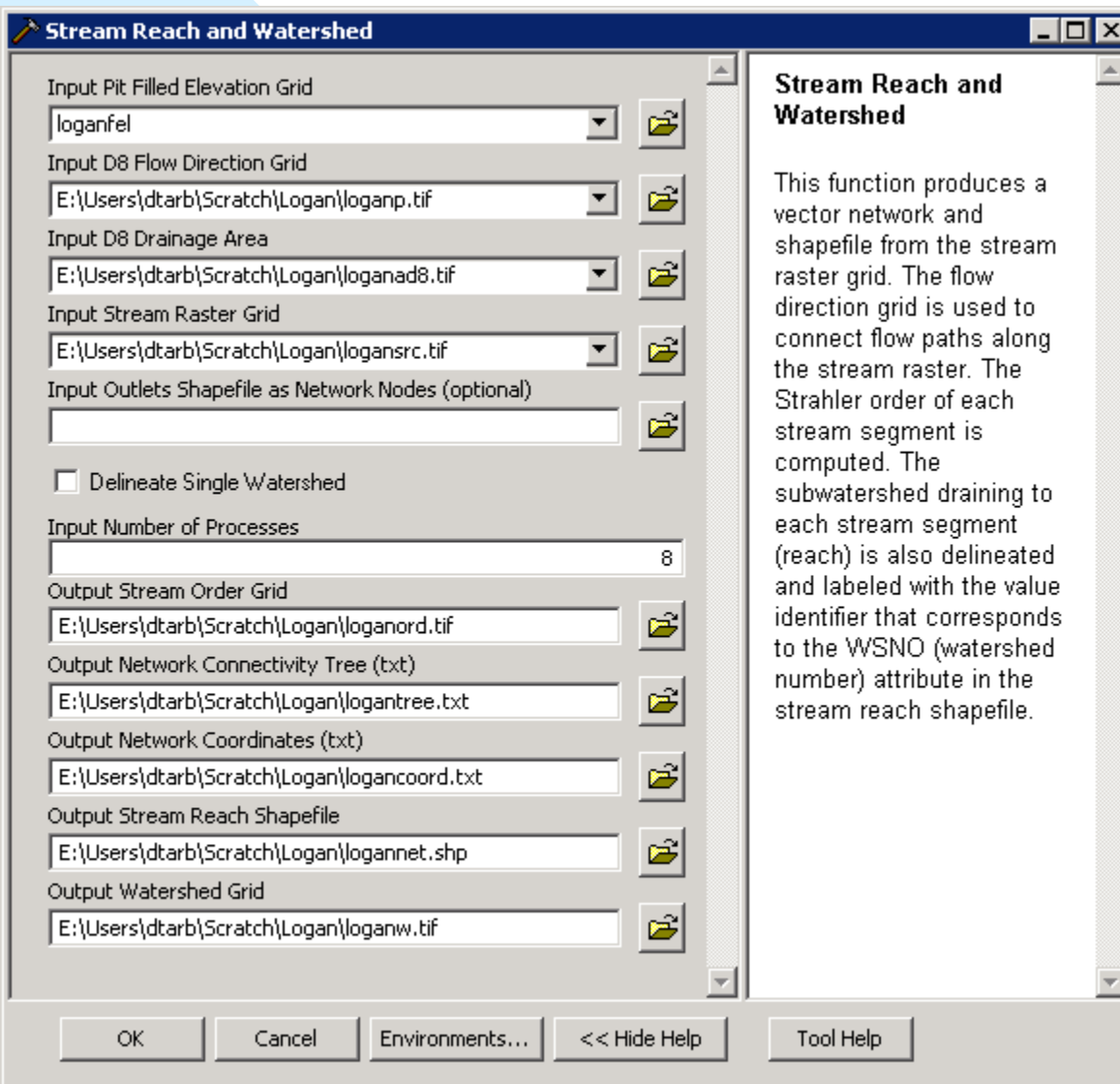
**Stream Definition by Threshold**

Operates on any grid and outputs an indicator (1,0) grid identifying cells with input values  $\geq$  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-

OK Cancel Environments... << Hide Help Tool Help

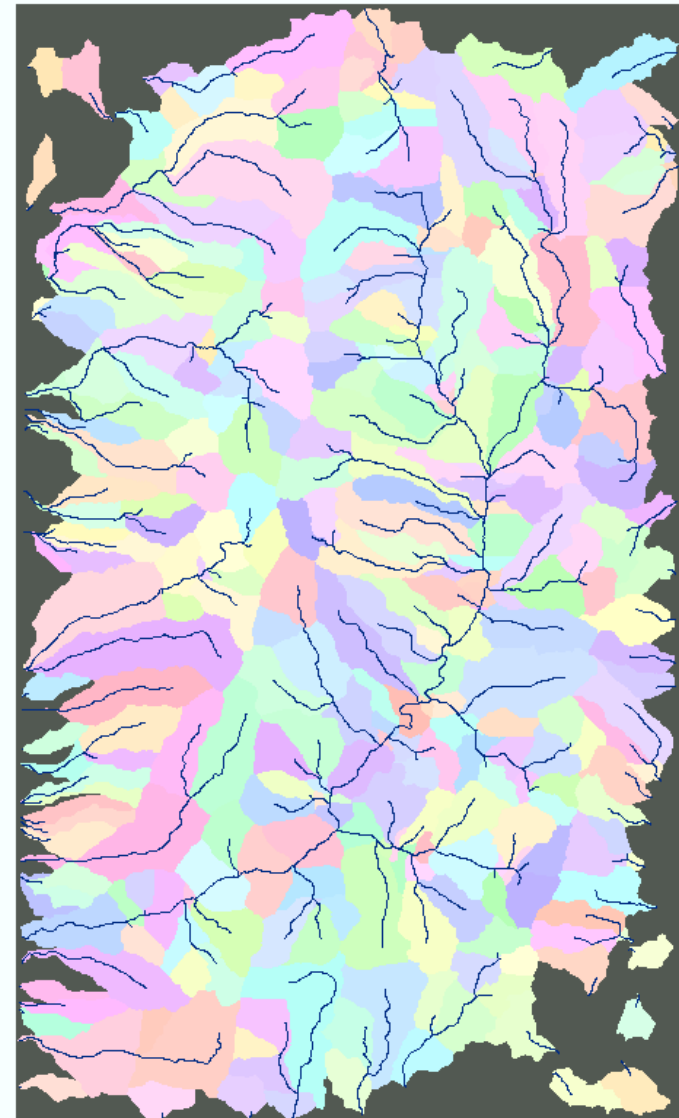


# Stream Reach and Watershed



## Stream Reach and Watershed

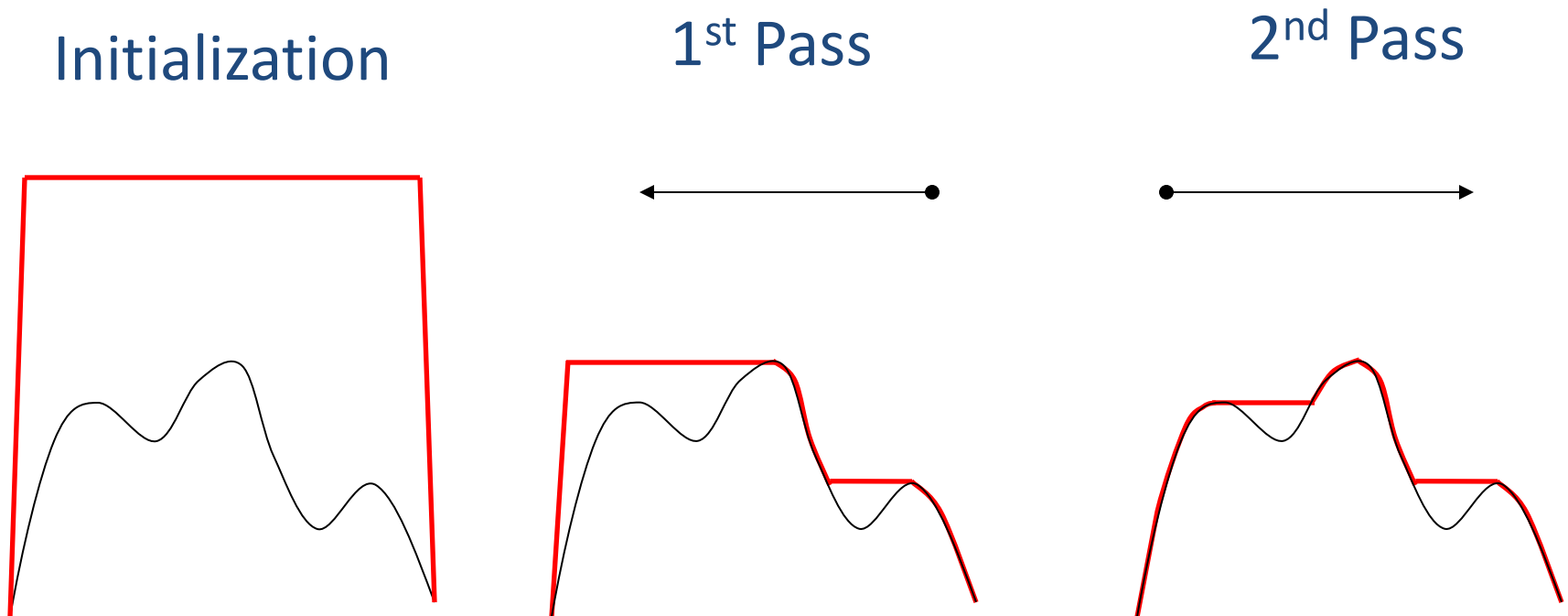
This function produces a vector network and shapefile from the stream raster grid. The flow direction grid is used to connect flow paths along the stream raster. The Strahler order of each stream segment is computed. The subwatershed draining to each stream segment (reach) is also delineated and labeled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





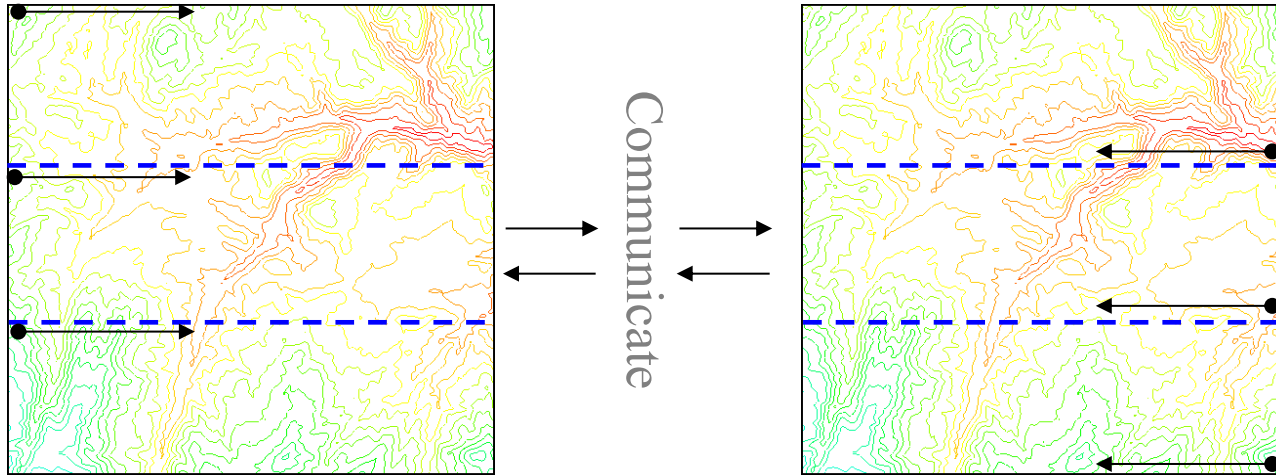
# Some Algorithm Details

## 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  $Z(i) > n$

$F(i) \leftarrow Z(i)$

Else

$F(i) \leftarrow n$

$i$  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

# Parallelization of Contributing Area/Flow Algebra

## 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

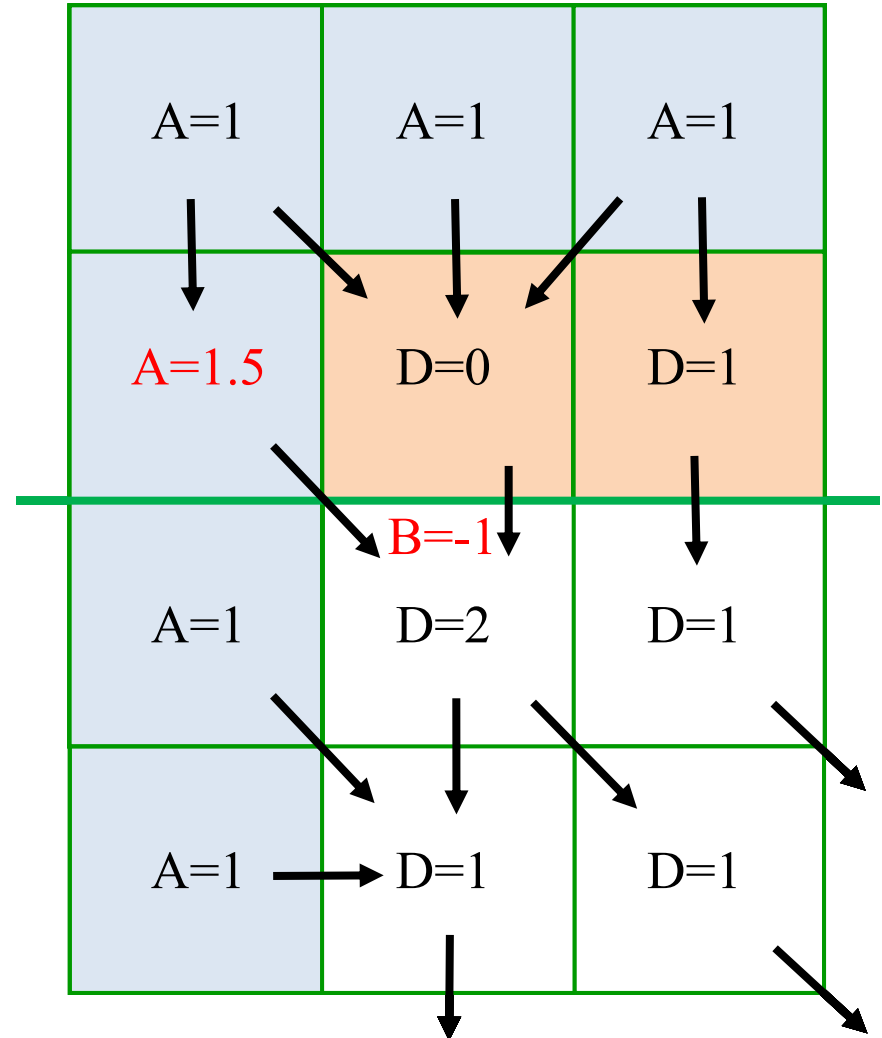
Executed by every process with  $D$  and  $Q$  initialized from FindDependencies.

**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
  
```

Queue is processed until empty.

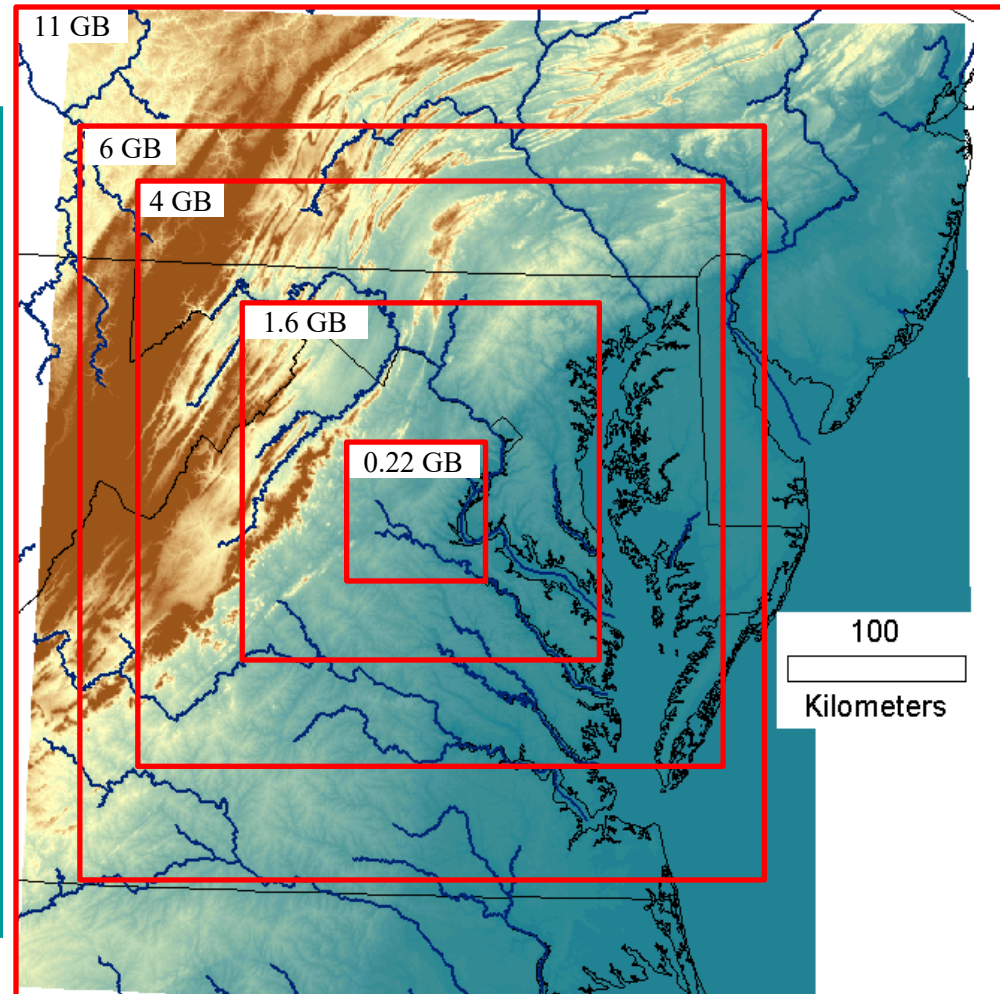


# Capabilities Summary

## Capability to run larger problems

		Processors used	Grid size	
			Theoretical limit	Largest run
2008	TauDEM 4	1	0.22 GB	0.22 GB
Sept 2009	Partial implementation	8	4 GB	1.6 GB
June 2010	TauDEM 5	8	4 GB	4 GB
Sept 2010	Multifile on 48 GB RAM PC	4	Hardware limits	6 GB
Sept 2010	Multifile on cluster with 128 GB RAM	128	Hardware limits	11 GB

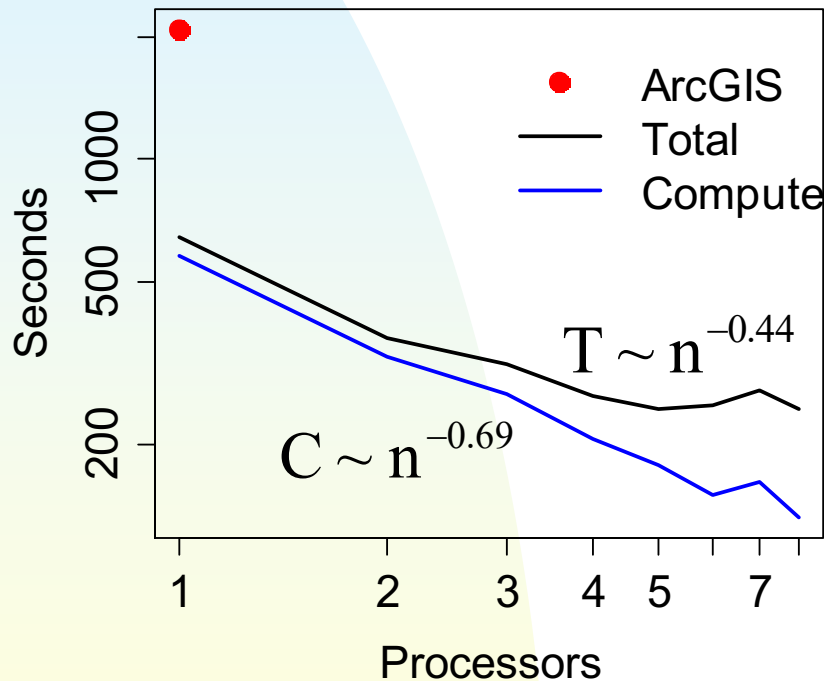
Single file size limit 4GB



At 10 m grid cell size

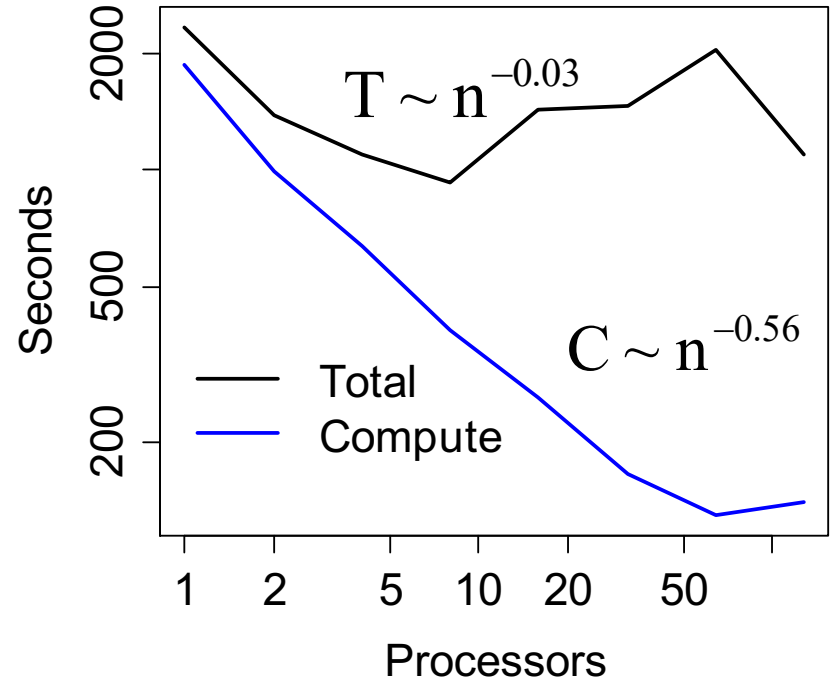
# Improved runtime efficiency

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

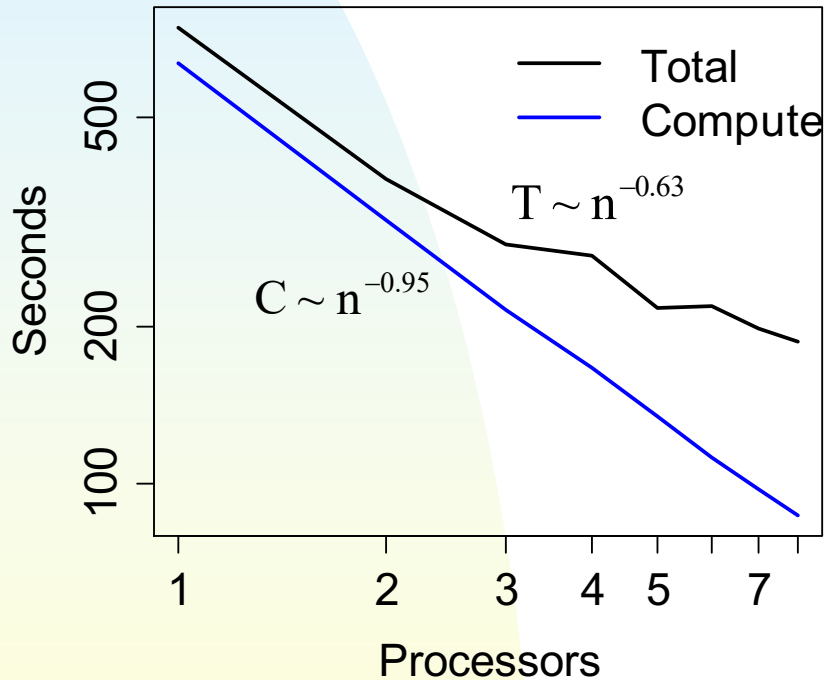


128 processor cluster

16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM

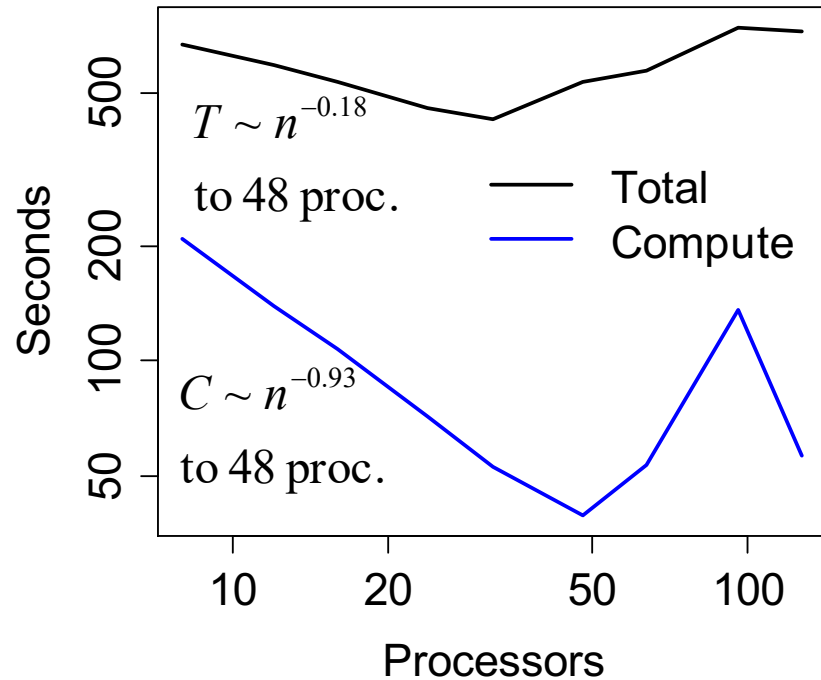
# Improved runtime efficiency

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

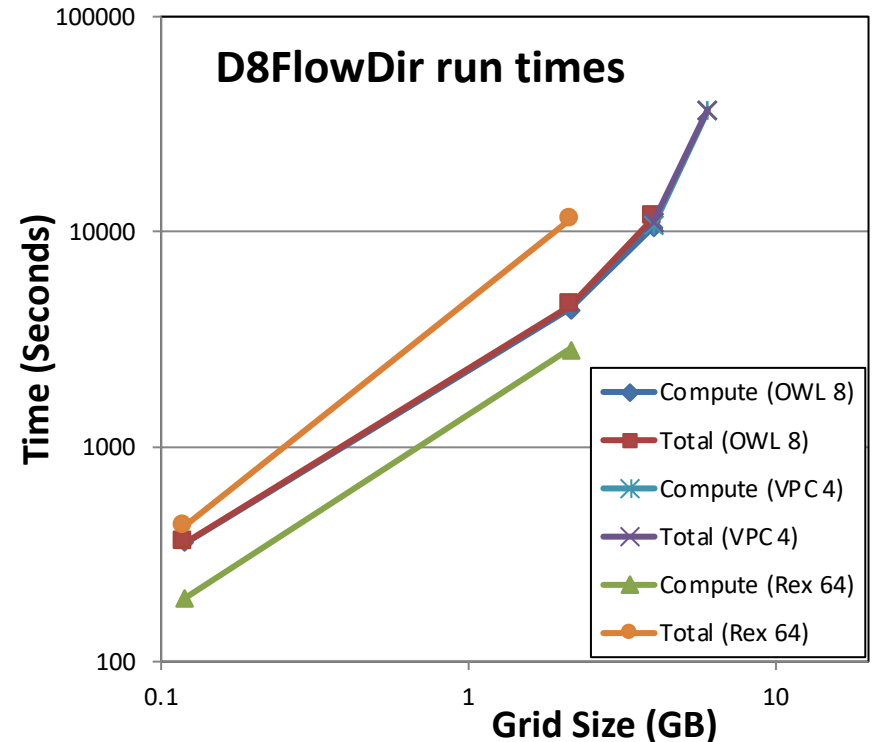
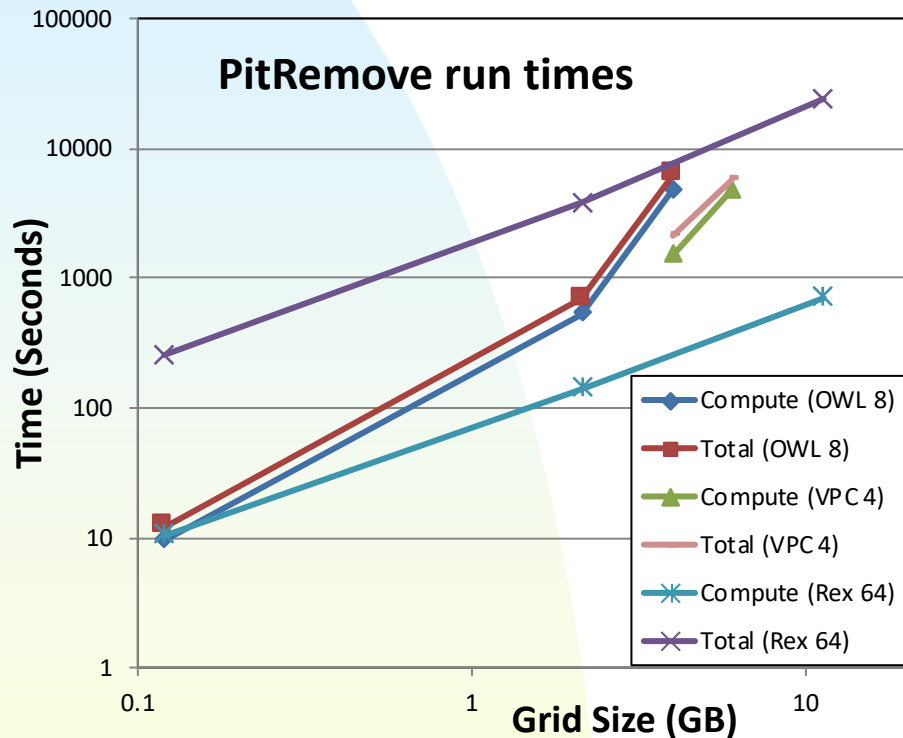


# Scaling of run times to large grids

Dataset	Size (GB)	Hardware	Number of Processors	PitRemove (run time seconds)		D8FlowDir (run time seconds)	
				Compute	Total	Compute	Total
GSL100	0.12	Owl (PC)	8	10	12	356	358
GSL100	0.12	Rex (Cluster)	8	28	360	1075	1323
GSL100	0.12	Rex (Cluster)	64	10	256	198	430
GSL100	0.12	Mac	8	20	20	803	806
YellowStone	2.14	Owl (PC)	8	529	681	4363	4571
YellowStone	2.14	Rex (Cluster)	64	140	3759	2855	11385
Boise River	4	Owl (PC)	8	4818	6225	10558	11599
Boise River	4	Virtual (PC)	4	1502	2120	10658	11191
Bear/Jordan/Weber	6	Virtual (PC)	4	4780	5695	36569	37098
Chesapeake	11.3	Rex (Cluster)	64	702	24045		

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

# Scaling of run times to large grids



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

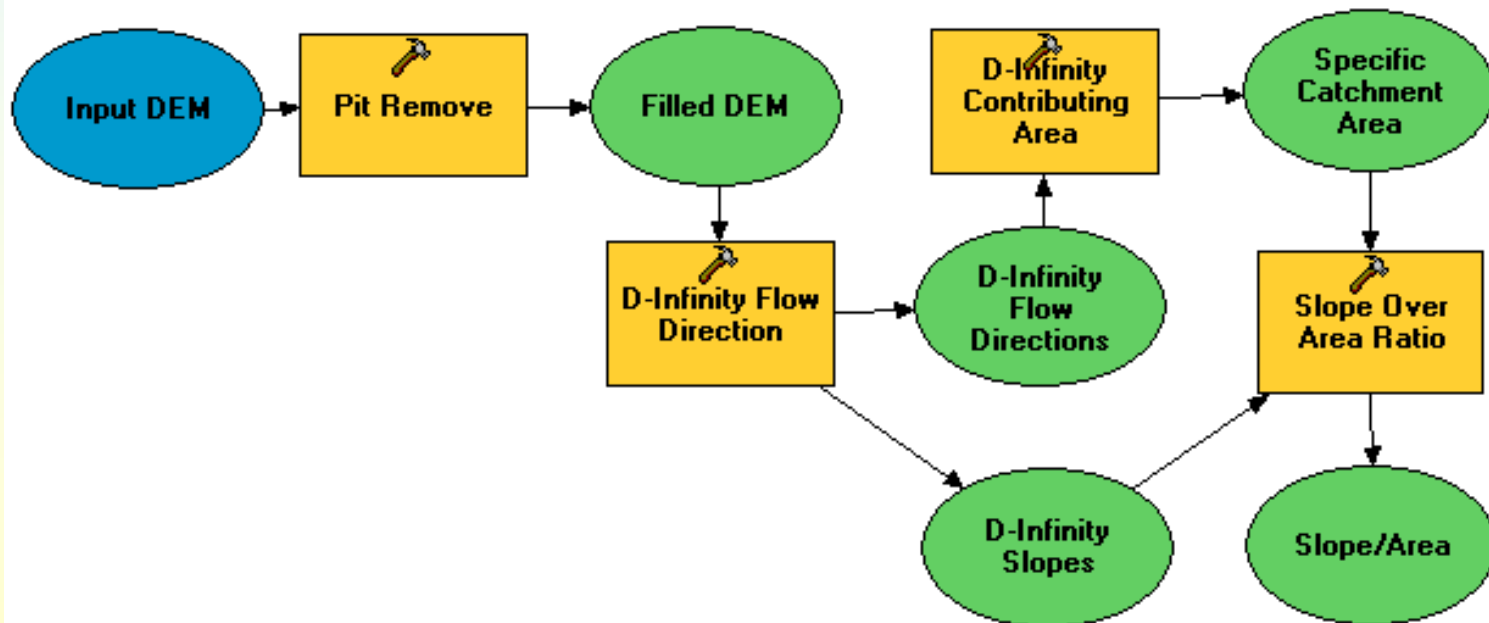
---

- Start with a DEM and end up with a topographic wetness index from the Dinfinit 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 runoff)

# Topographic wetness index from the D-infinity method

## Steps

- Pit Remove
- D-Infinity Flow Directions (and Slopes)
- D-Infinity Contributing Area
- Slope Over Area Ratio



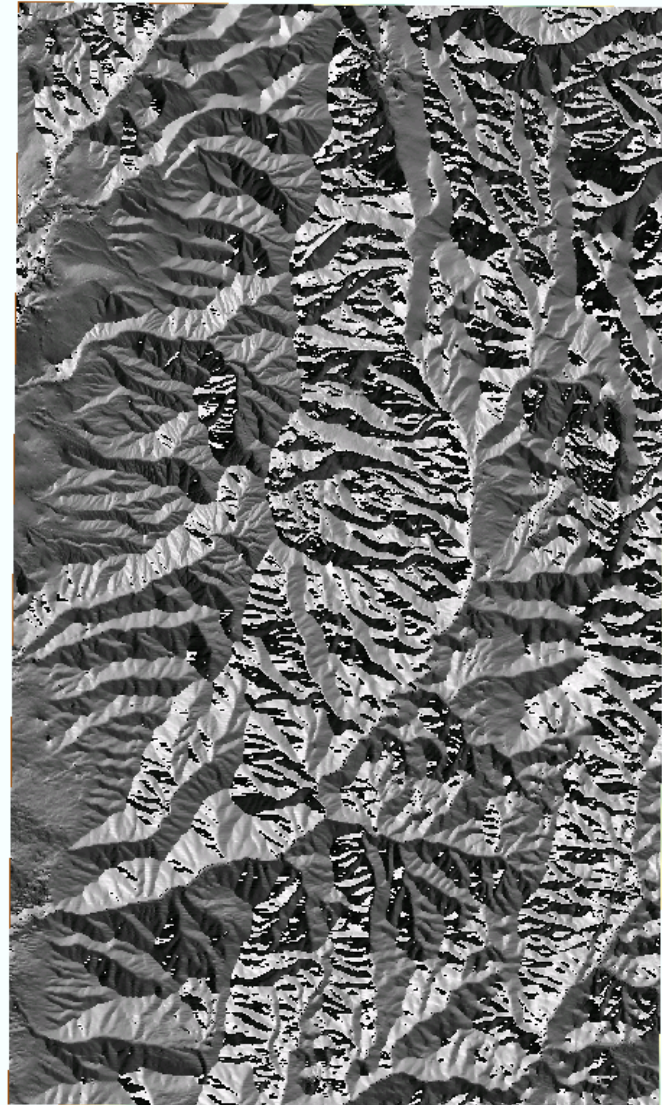
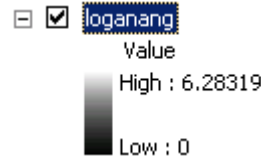
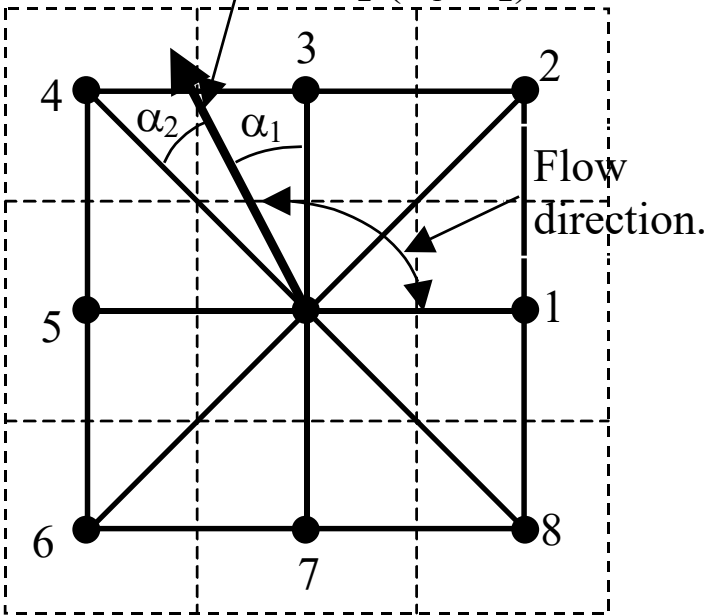


# D-Infinity Flow Direction (and slope)

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  $\alpha_2/(\alpha_1+\alpha_2)$



**D-Infinity Flow Direction**

Input Pit Filled Elevation Grid  
 loganfcl

Input Number of Processes  
 8

Output D-Infinity Flow Directions Grid  
 E:\Users\dtarb\Scratch\Logan\loganang.tif

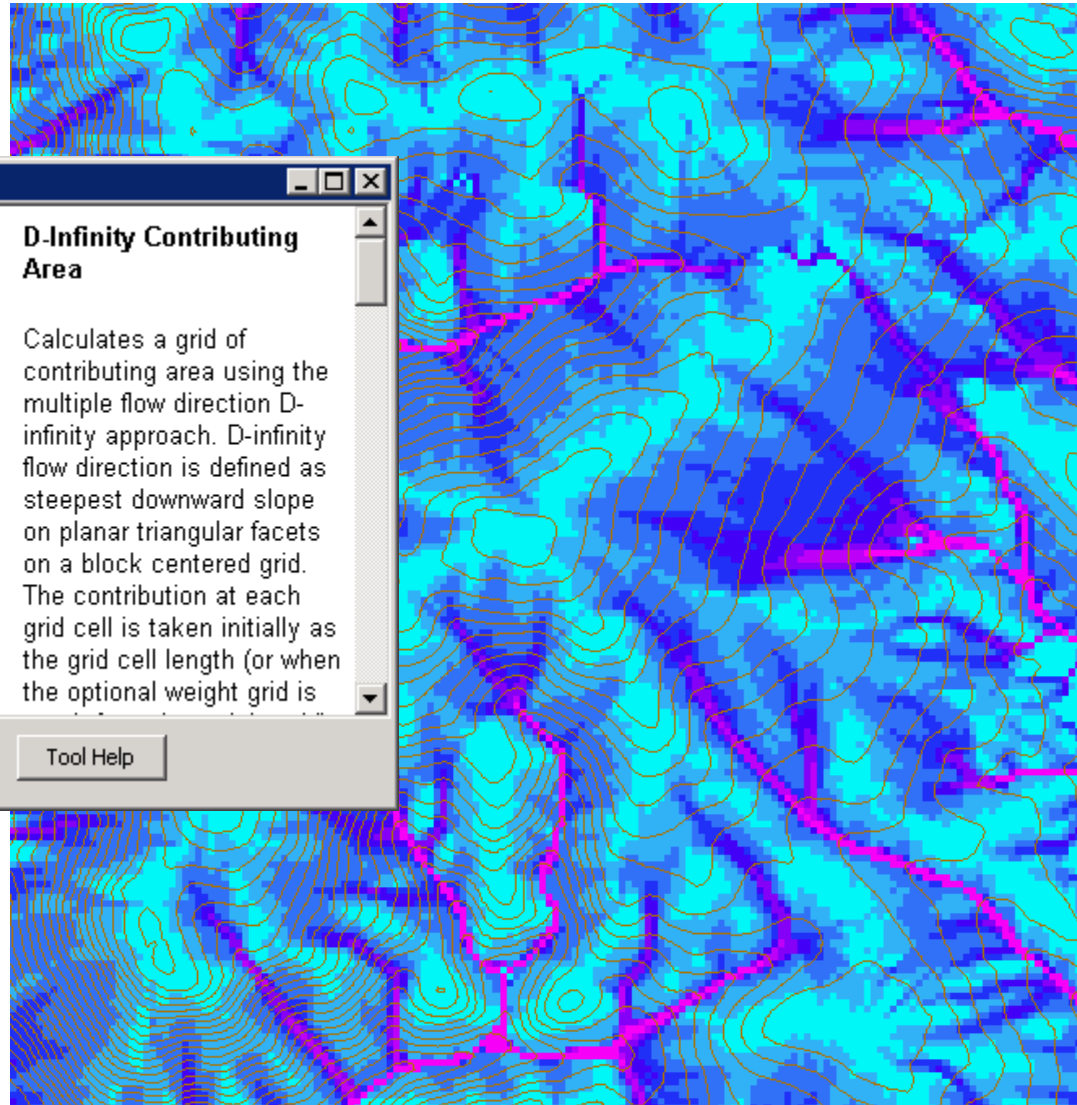
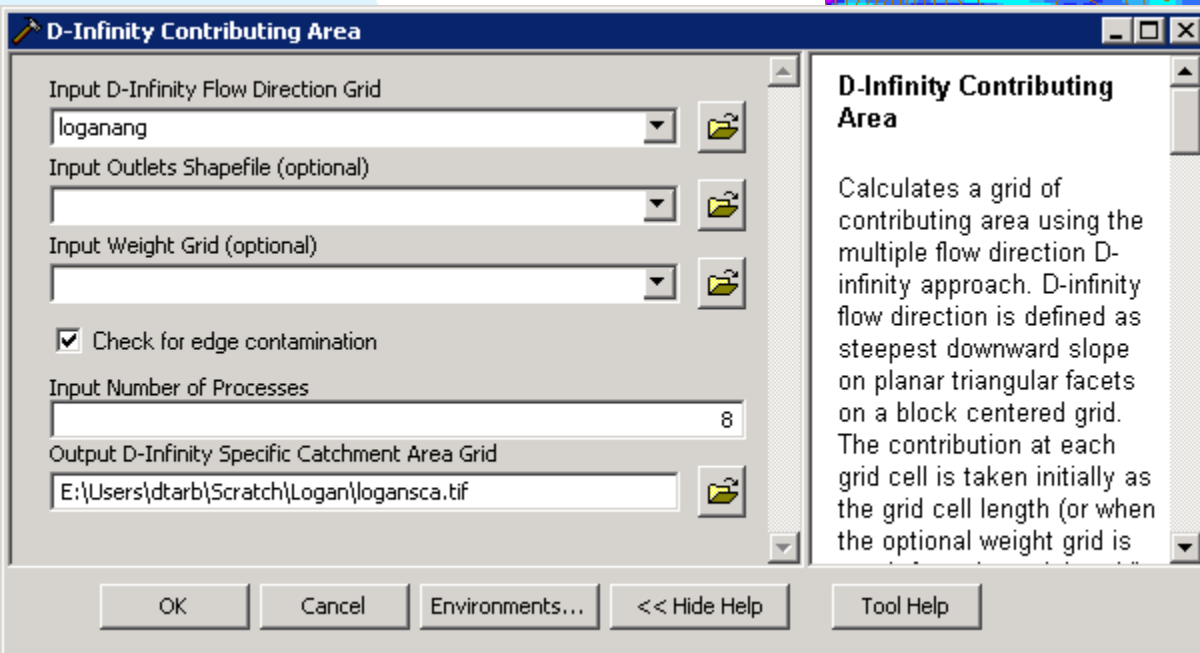
Output D-Infinity Slope Grid  
 E:\Users\dtarb\Scratch\Logan\loganslp.tif

**D-Infinity Flow Direction**

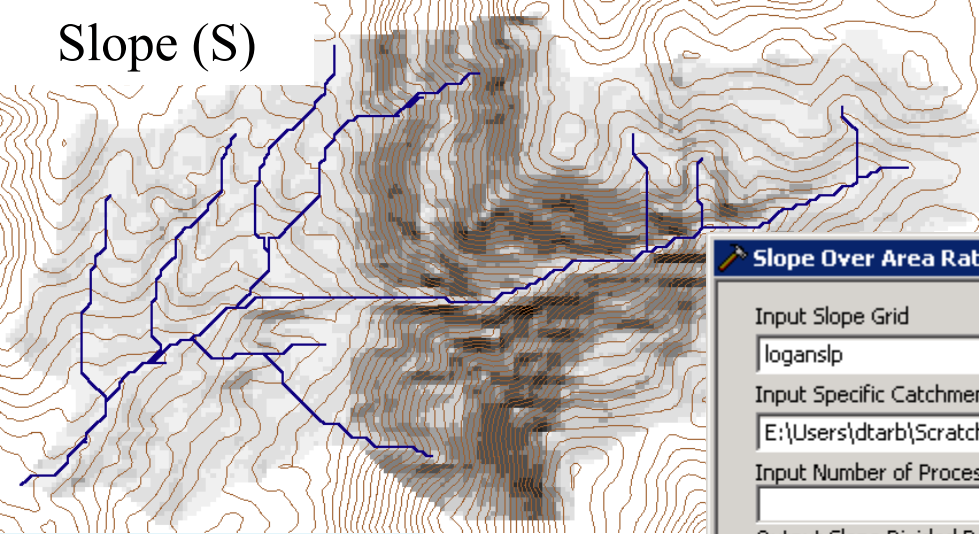
Assigns a flow direction based on the steepest slope of a triangular facet (Tarboton, 1997). Flow direction is defined as steepest downward slope on planar triangular facets on a block centered grid. Flow direction is encoded

OK Cancel Environments... << Hide Help Tool Help

# D-Infinity Contributing Area



Slope (S)



# Wetness Index

**Slope Over Area Ratio**

Input Slope Grid  
loganslp

Input Specific Catchment Area Grid  
E:\Users\dtarb\Scratch\Logan\logansca.tif

Input Number of Processes  
8

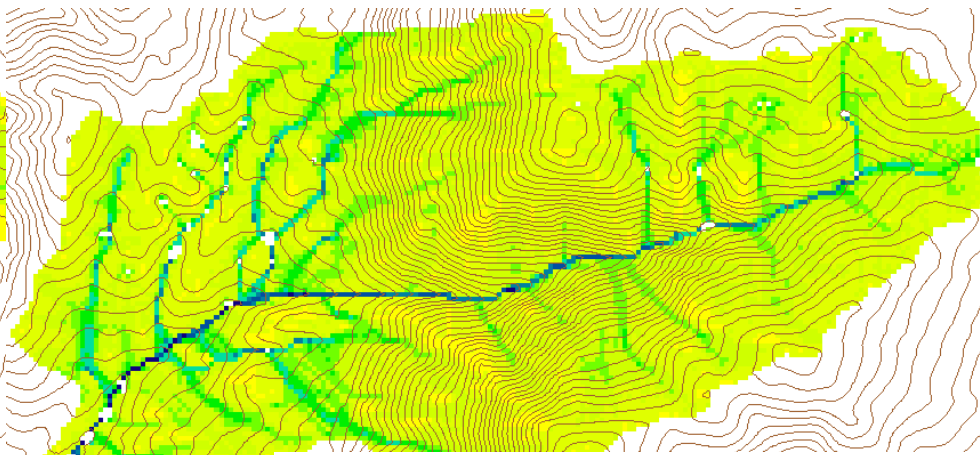
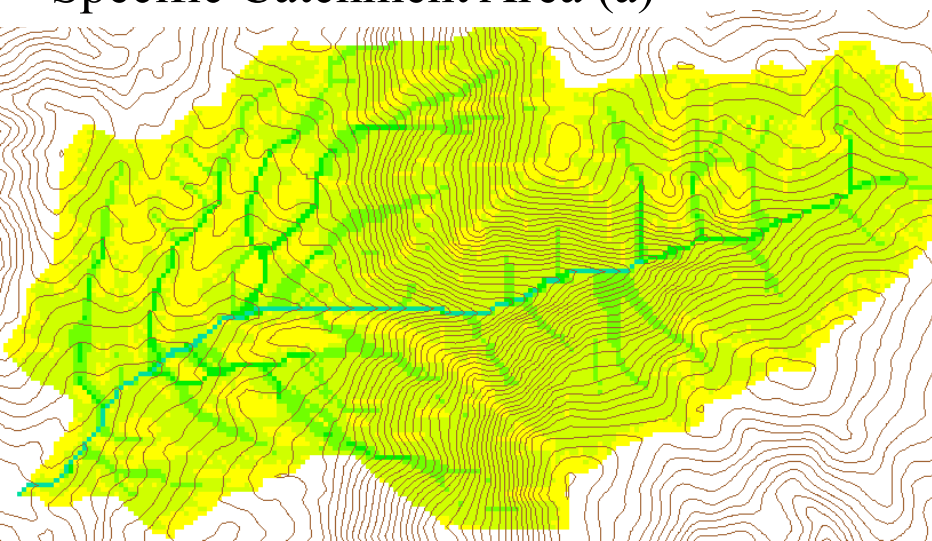
Output Slope Divided By Area Ratio Grid  
E:\Users\dtarb\Scratch\Logan\logansar.tif

**Slope Over Area Ratio**

Calculates the ratio of the slope to the specific catchment area (contributing area). This is algebraically related to the more common  $\ln(a/\tan \beta)$  wetness index, but contributing area is in the denominator to avoid divide by 0 errors when slope is 0.

OK Cancel Environments... << Hide Help Tool Help

Specific Catchment Area (a)



Wetness Index  $\ln(a/S)$

# Channels and watersheds with spatially variable drainage density using Peuker Douglas threshold objectively chosen by drop analysis

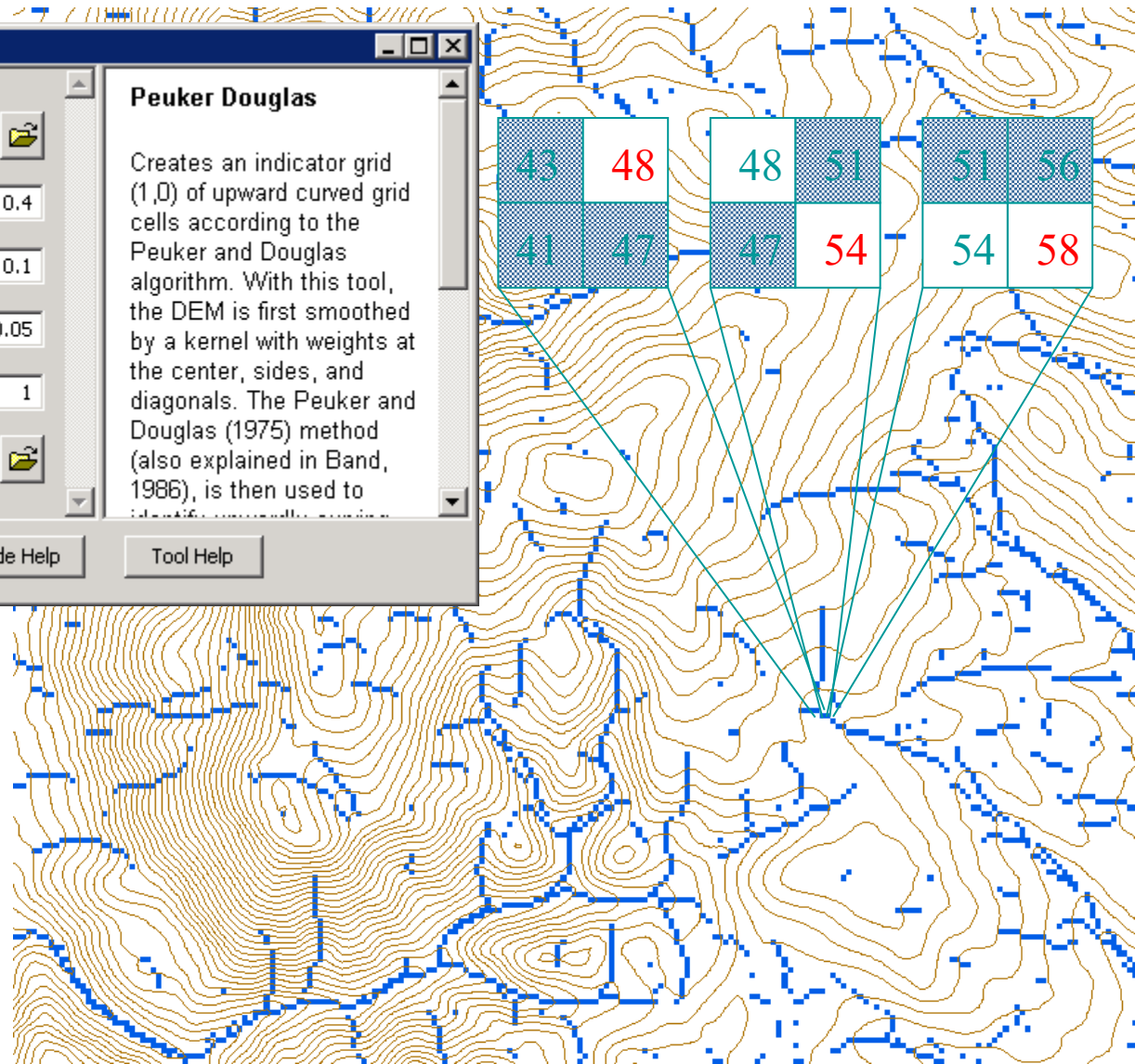
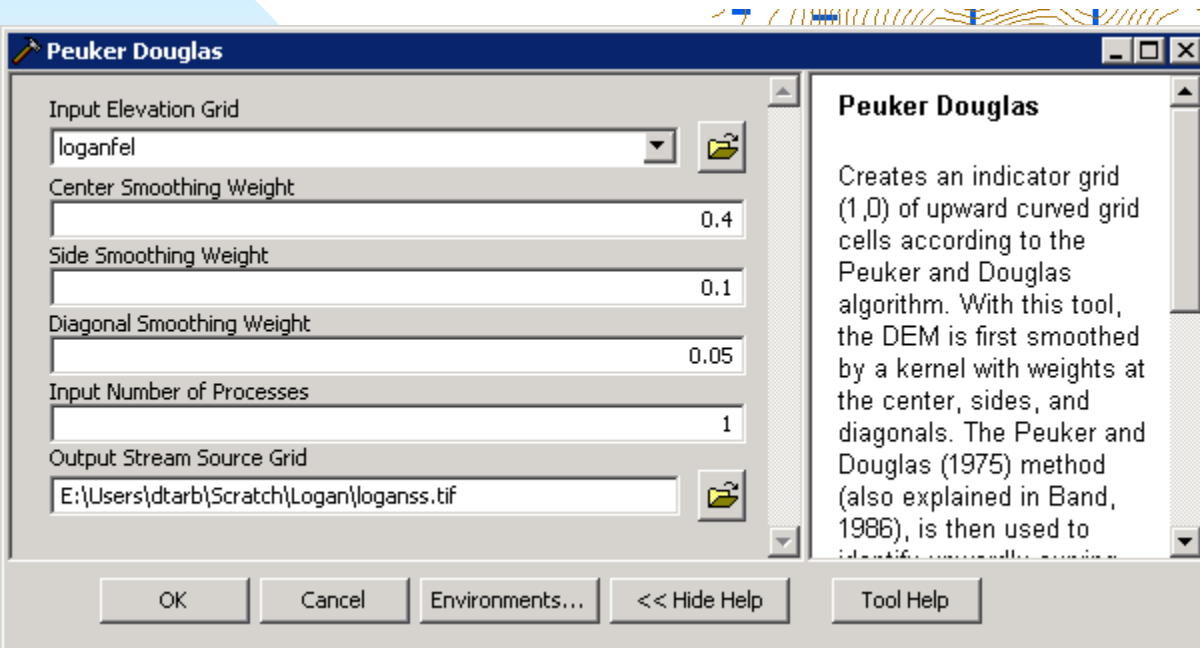
---

## 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

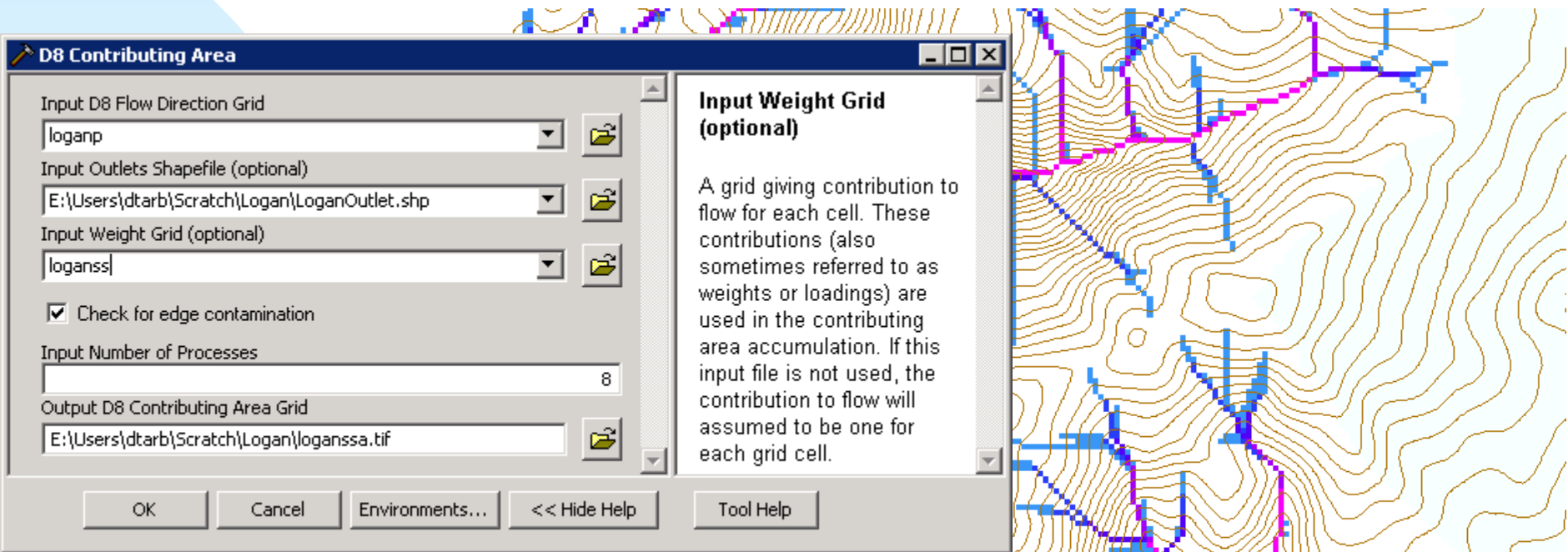


# Peuker Douglas

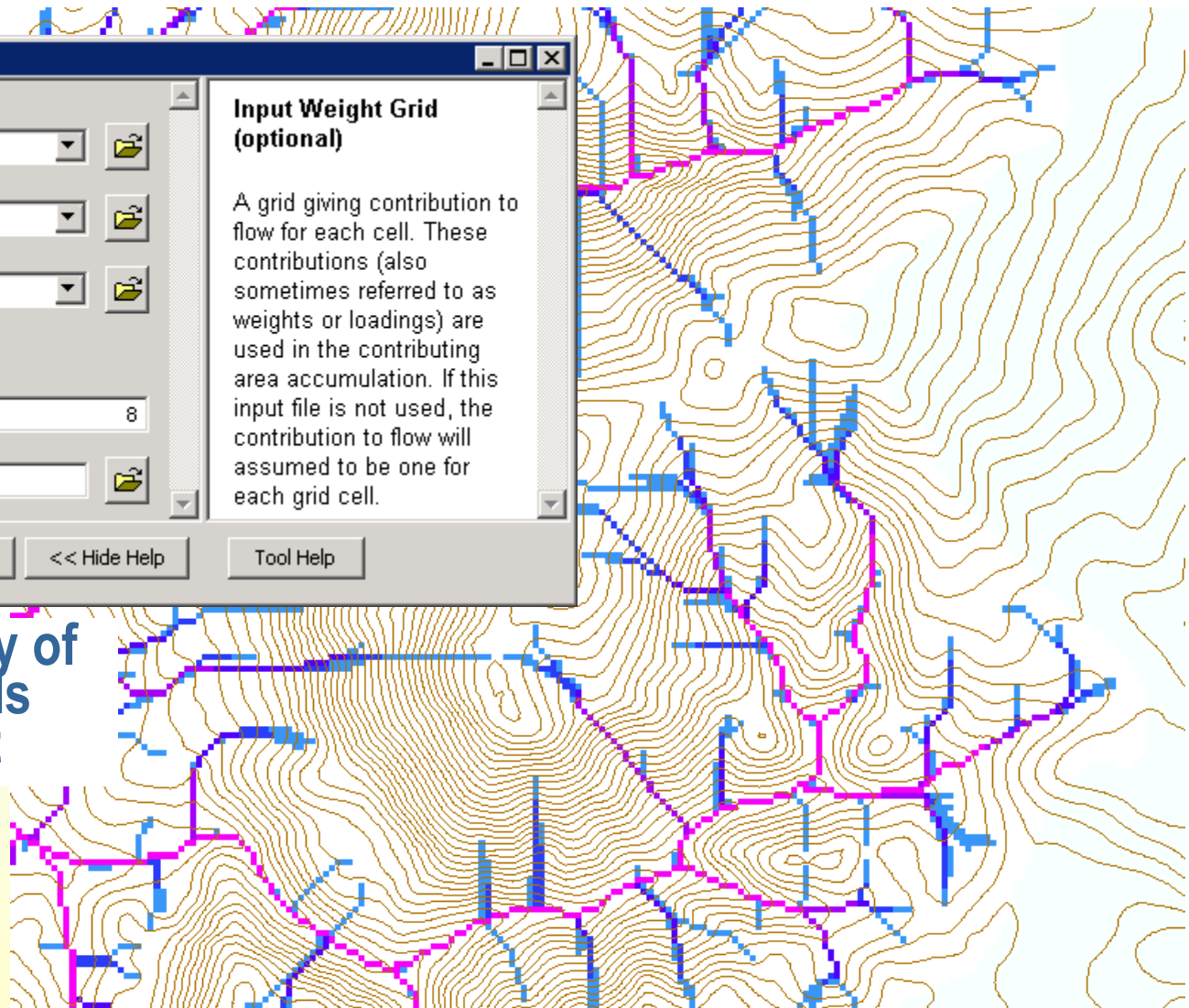
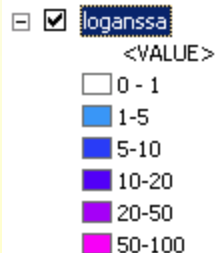


**Local Valley Form  
Computation**  
(Peuker and Douglas, 1975,  
Comput. Graphics Image Proc.  
4:375)

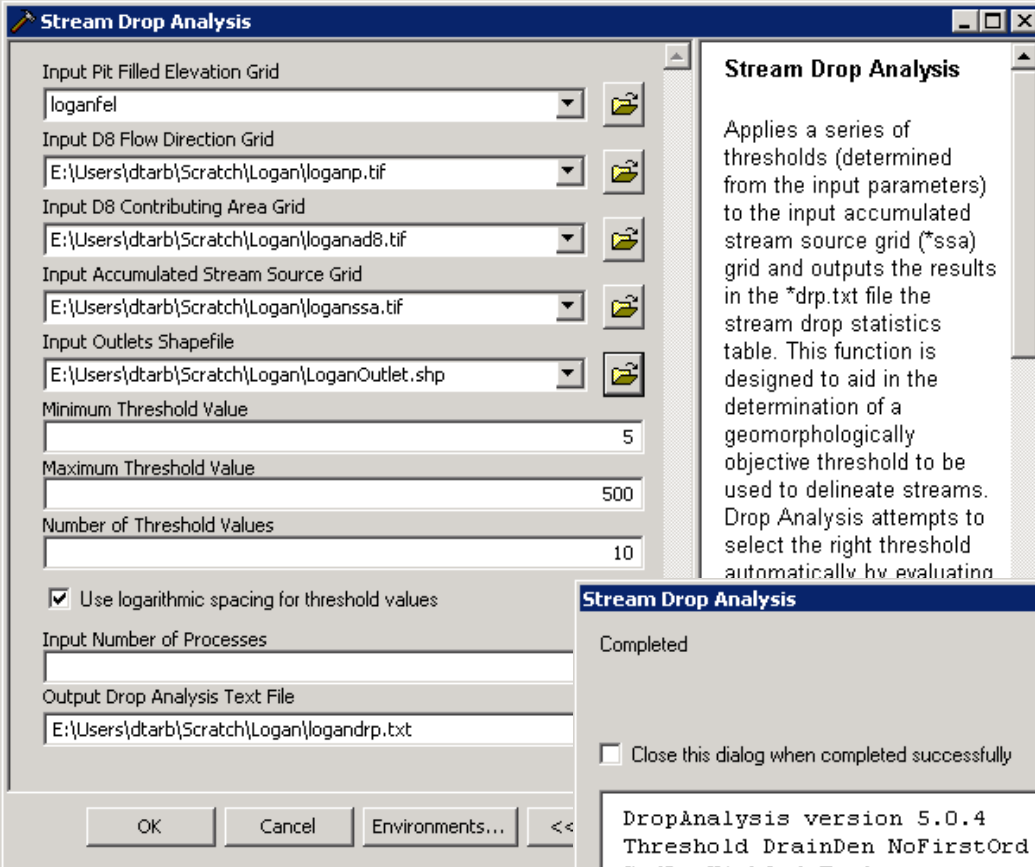
# Weighted D8 Contributing Area



Contributing area only of valley form grid cells upstream of outlet



# Stream Drop Analysis



## Stream Drop Analysis

Applies a series of thresholds (determined from the input parameters) to the input accumulated stream source grid (\*ssa) grid 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

## Stream Drop Analysis

Completed

Close

<< Details

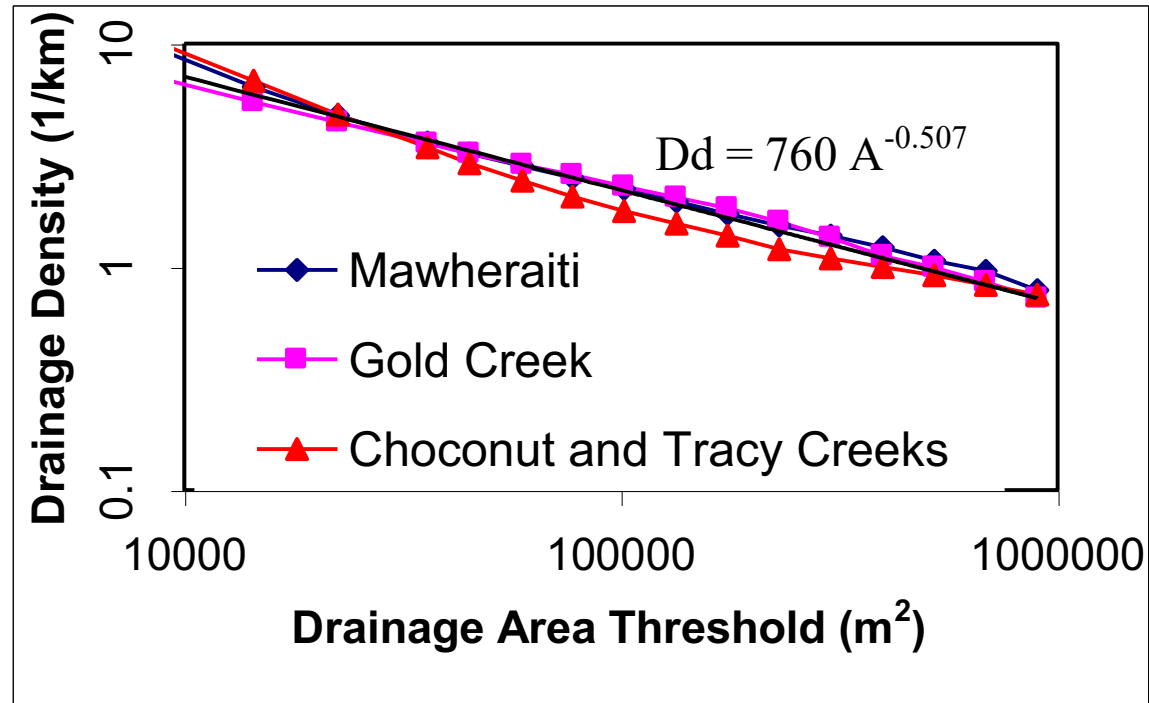
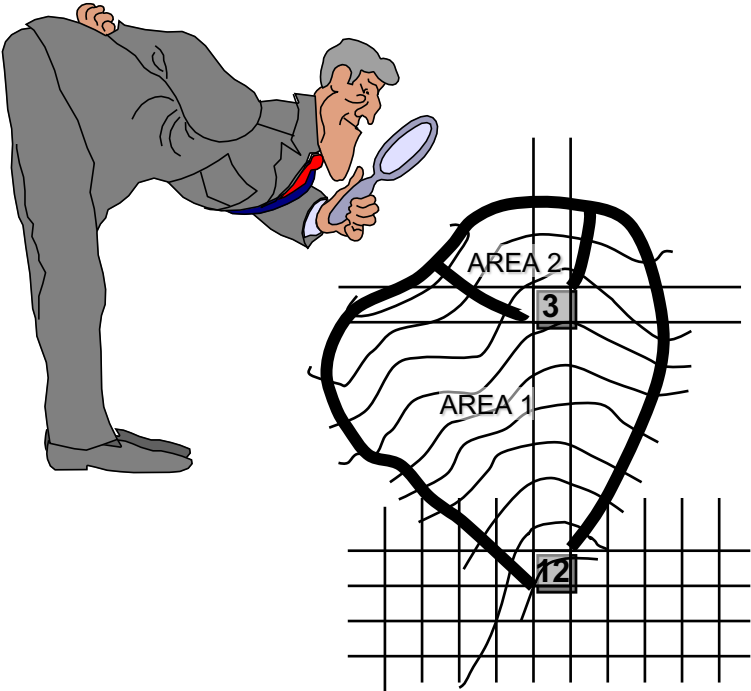
Close this dialog when completed successfully

```
DropAnalysis version 5.0.4
Threshold DrainDen NoFirstOrd NoHighOrd MeanDFirstOrd MeanDHighOrd StdDevFirstOrd
StdDevHighOrd Tval
5.000000 0.002461 2256 688 66.508453 125.044464 76.240814 131.867966 -14.564702
8.340503 0.001854 1165 351 85.638748 145.378479 97.830666 142.423080 -8.938837
13.912798 0.001537 774 239 96.581406 159.873108 103.330826 151.388107 -7.345115
23.207947 0.001226 452 141 115.005356 182.002914 109.692078 158.783463 -5.642088
38.713192 0.000999 294 96 116.624161 211.537094 107.394669 166.852936 -6.479936
64.577499 0.000790 188 70 116.728371 209.407593 123.760880 156.084854 -4.967545
107.721756 0.000635 109 38 153.991043 239.083878 144.088898 162.634705 -3.030640
179.690720 0.000524 75 19 187.208069 269.439911 158.242188 156.966827 -2.026490
299.742218 0.000412 50 14 197.519684 255.433441 137.707306 168.146484 -1.324365
508.000183 0.000304 30 4 214.549347 289.485138 153.106644 135.973572 -0.928733
299.742218 Value for optimum that drop analysis selected - see output file for details.
Processes: 8
Read time: 0.142451
Compute time: 1.028273
Total time: 1.170725

Executed (StreamDropAnalysis) successfully.
End Time: Mon Sep 20 21:26:59 2010 (Elapsed Time: 2.00 seconds)
```



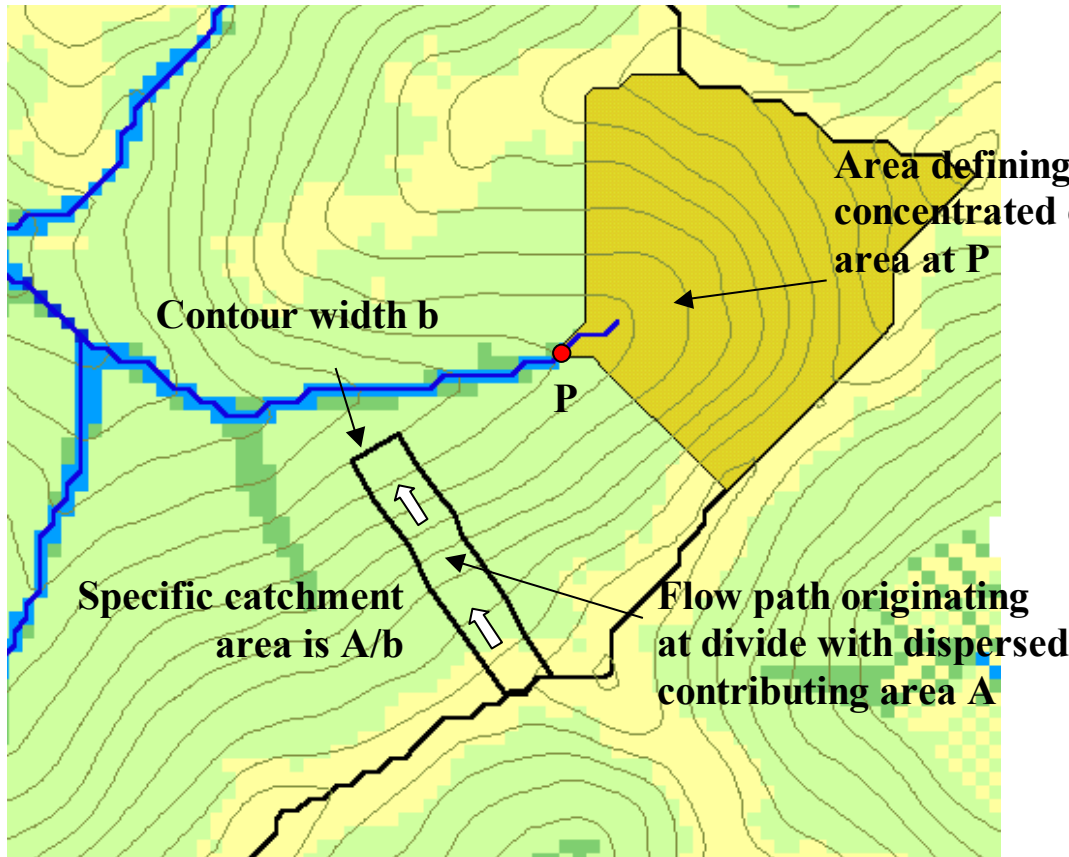
# How to decide on stream delineation threshold ?



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.

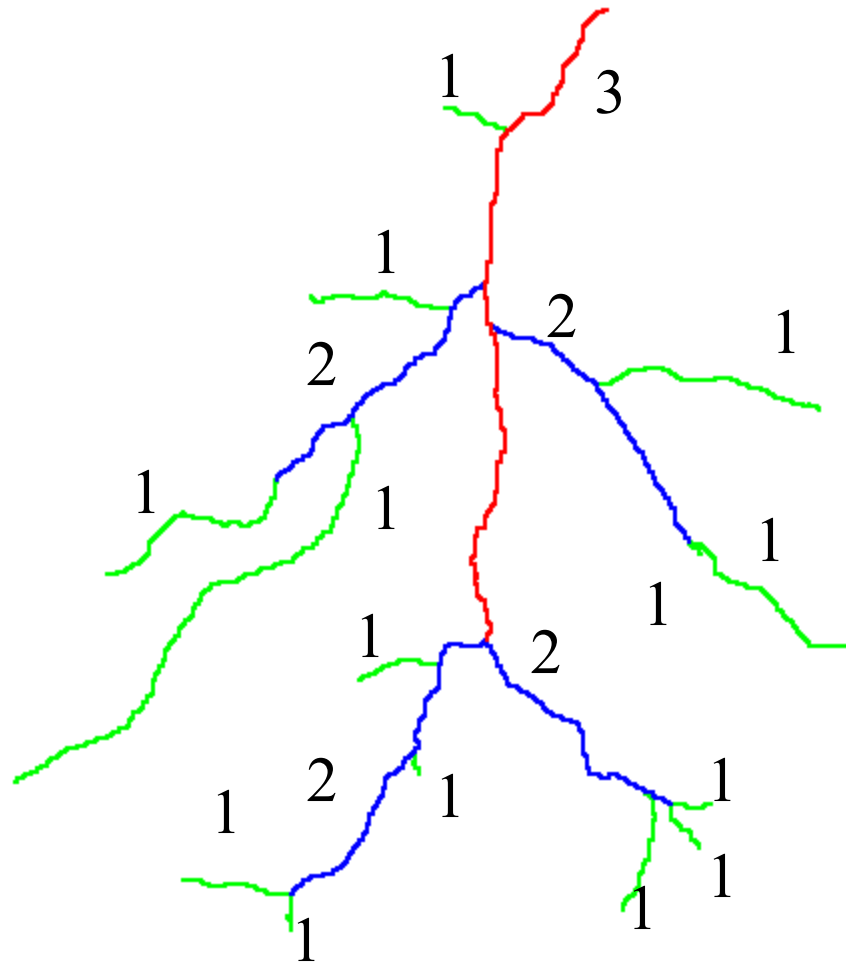
## Why is it important?

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

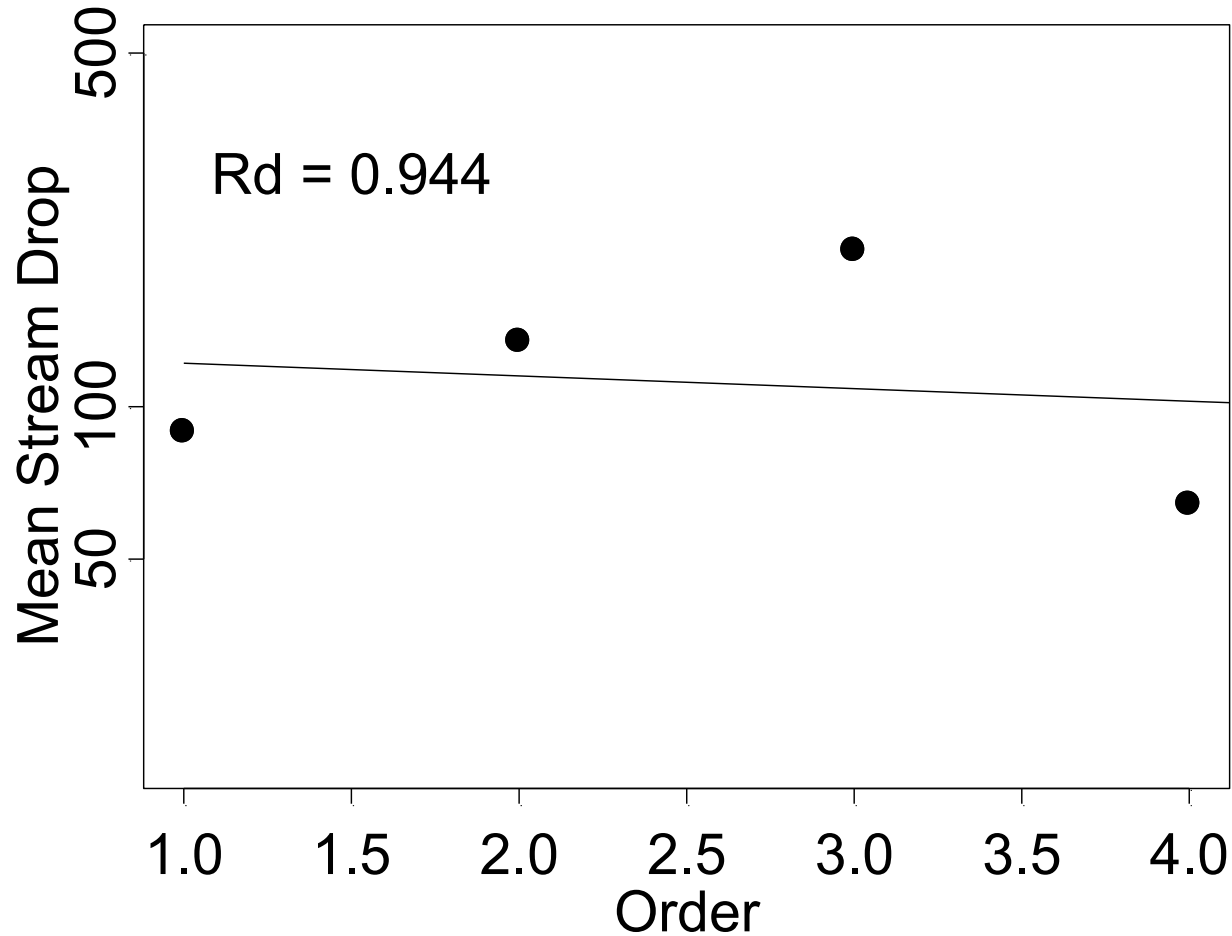


**Drainage area can be concentrated or dispersed (specific catchment area) representing concentrated or dispersed flow.**

# Hortons Laws: Strahler system for stream ordering

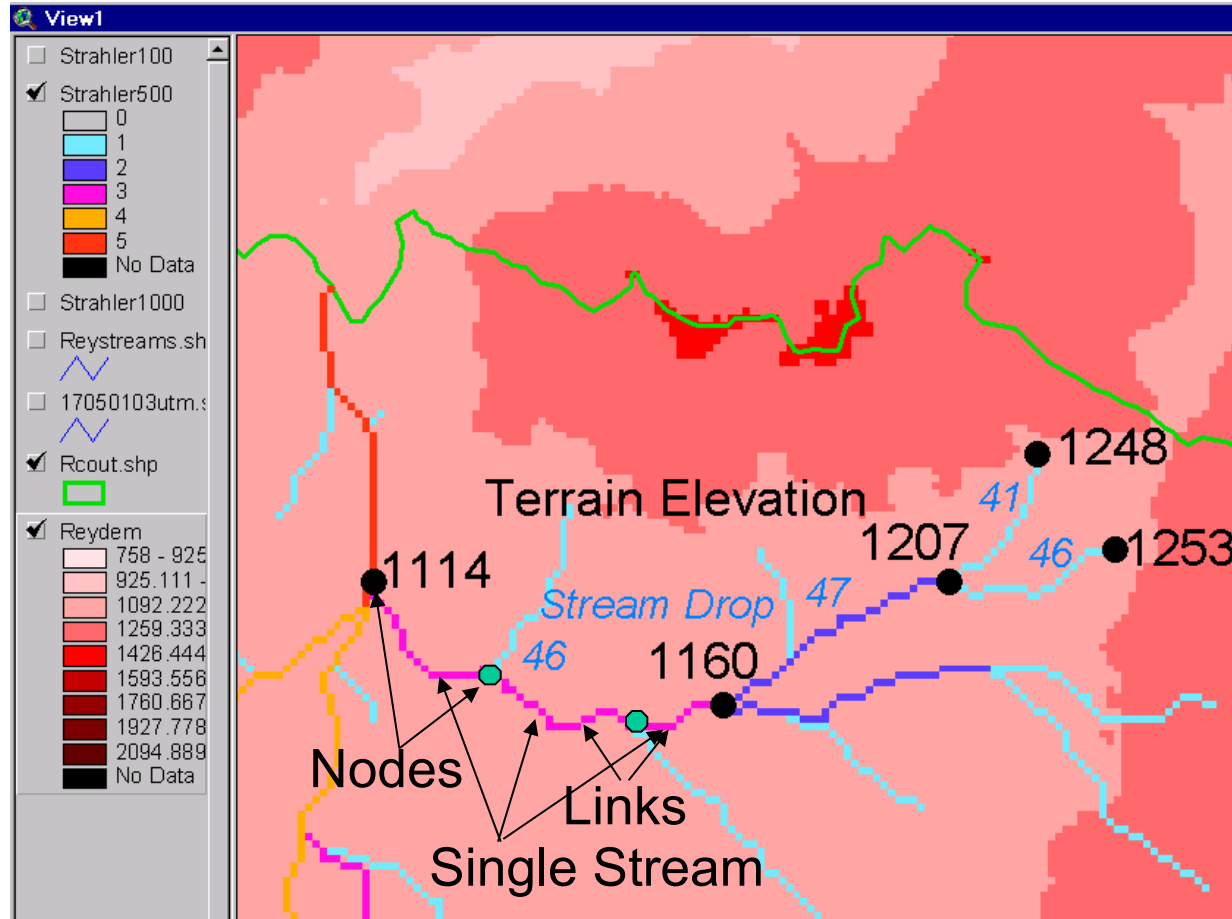


# Constant Stream Drops Law



# Stream Drop

Elevation difference between ends of stream

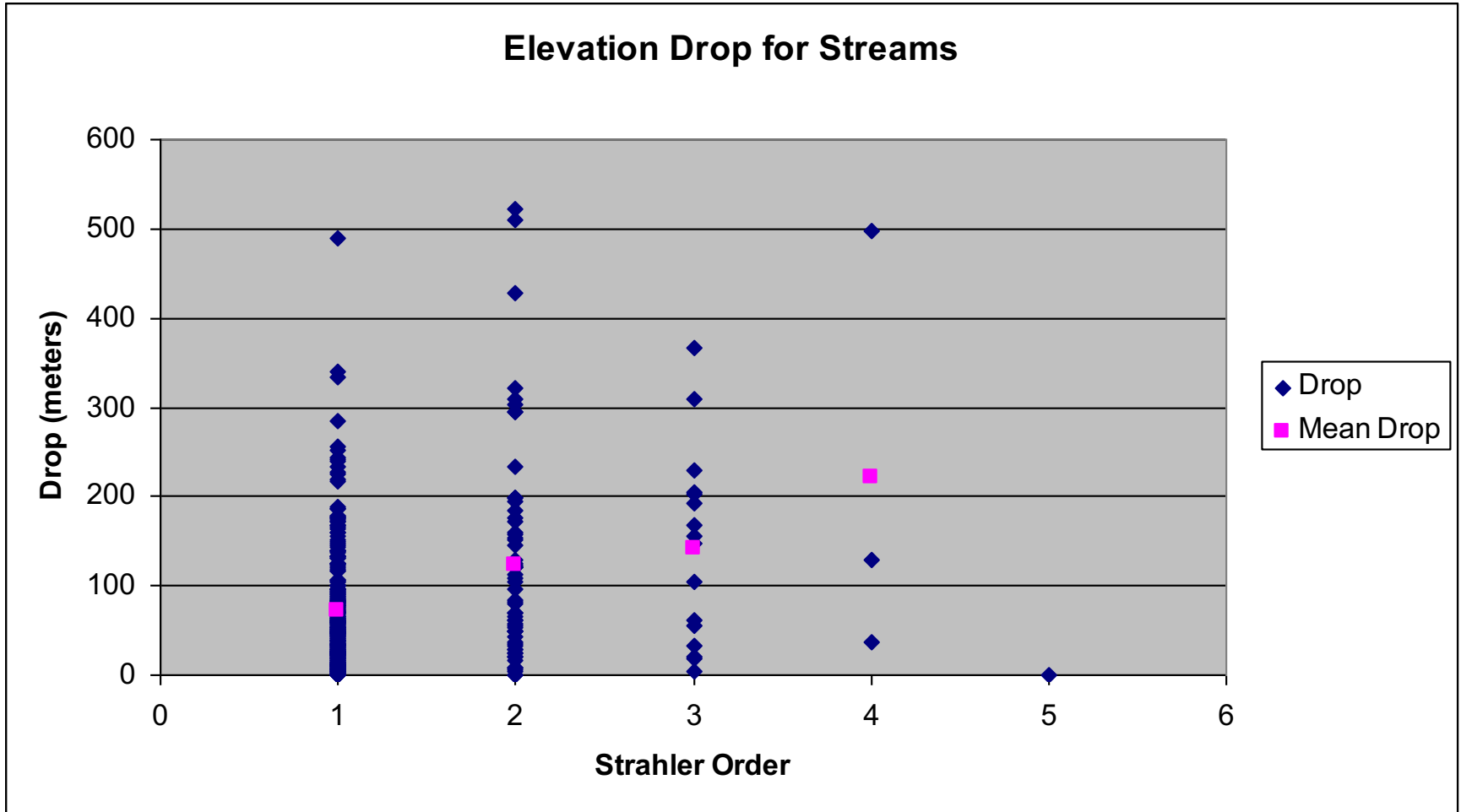


Note that a “Strahler stream” comprises a sequence of links (reaches or segments) of the same order

**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

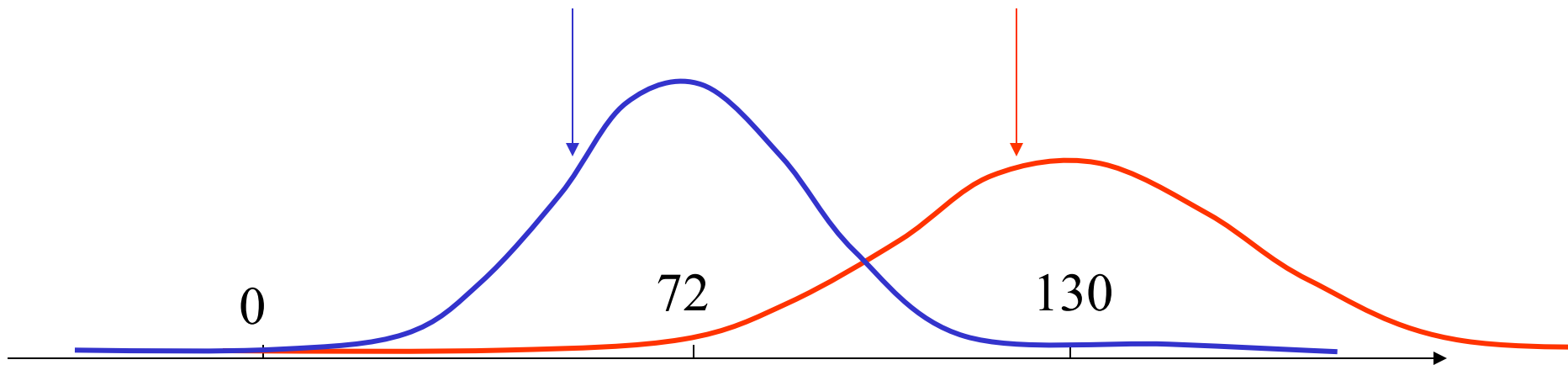
# Statistical Analysis of Stream Drops





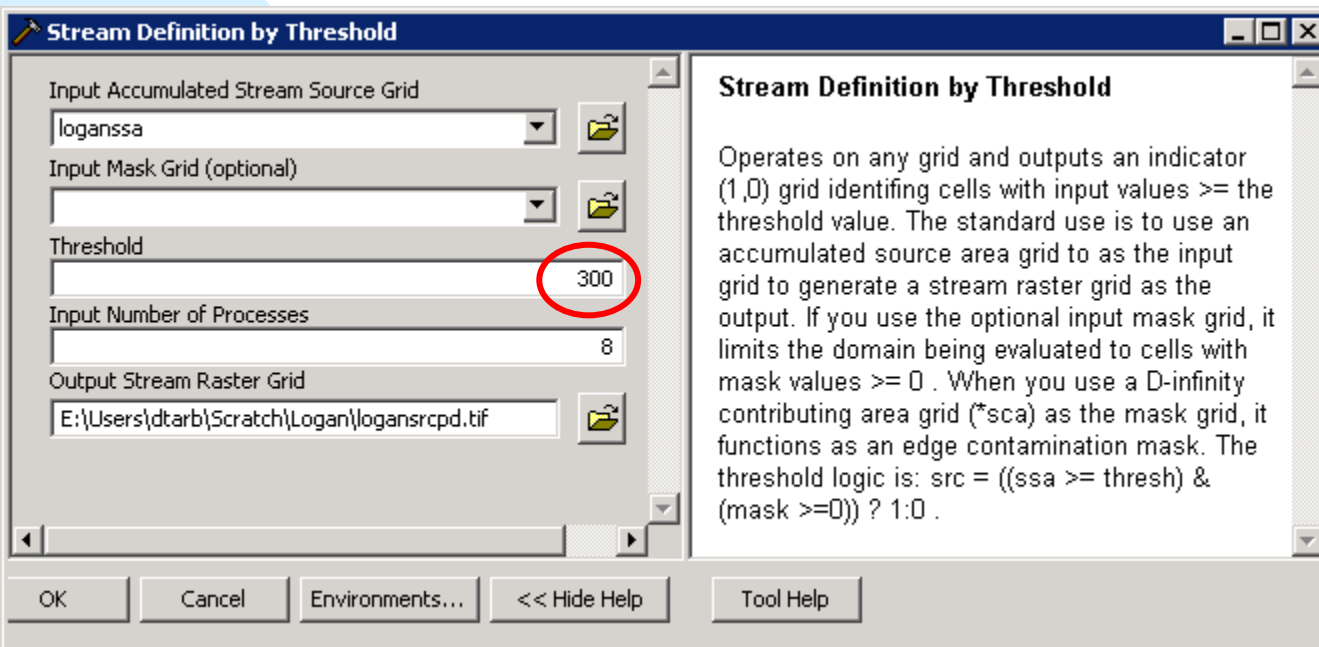
# T-Test for Difference in Mean Values

	Order 1		Order 2-4
Mean X	72.2	Mean Y	130.3
Std X	68.8	Std Y	120.8
Var X	4740.0	Var Y	14594.5
Nx	268	Ny	81



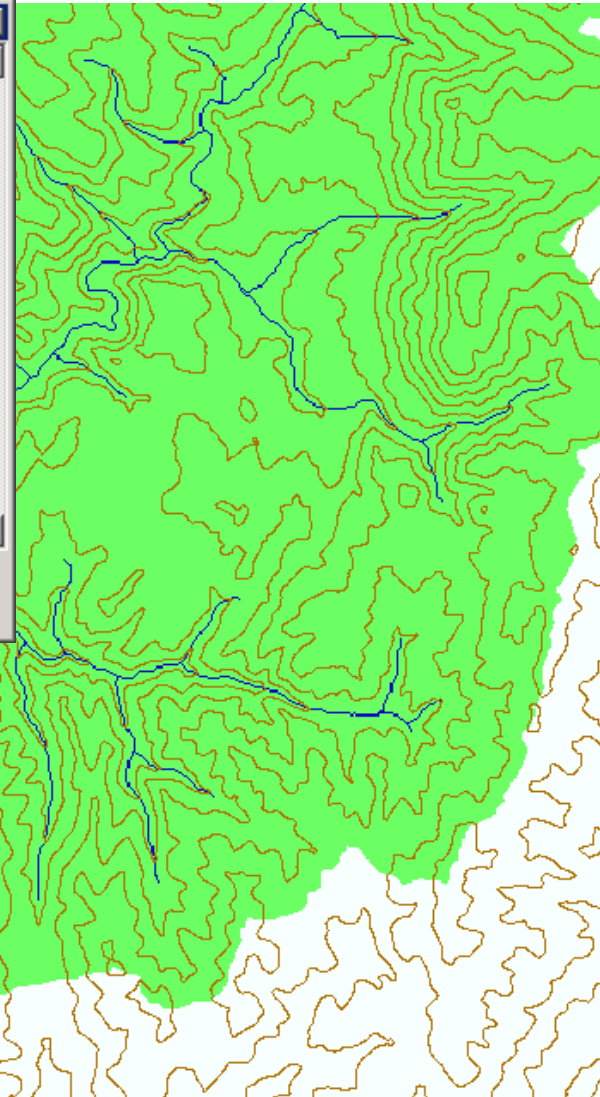
T-test checks whether difference in means is large ( $> 2$ ) when compared to the spread of the data around the mean values

# 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 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:  $\text{src} = ((\text{ssa} \geq \text{thresh}) \& (\text{mask} \geq 0)) ? 1:0$ .



# Stream Reach and Watershed

The screenshot displays the 'Stream Reach and Watershed' tool interface. The left pane contains input parameters and output options. The right pane shows the 'Output Stream Reach Shapefile' section with a list of attributes. The background is a map showing a watershed with stream reaches highlighted in blue.

**Stream Reach and Watershed**

Input Pit Filled Elevation Grid  
loganfcl

Input D8 Flow Direction Grid  
E:\Users\dtarb\Scratch\Logan\loganp.tif

Input D8 Drainage Area  
E:\Users\dtarb\Scratch\Logan\loganad8.tif

Input Stream Raster Grid  
logansrcpd

Input Outlets Shapefile as Network Nodes (optional)  
LoganOutlet

Delineate Single Watershed

Input Number of Processes  
8

Output Stream Order Grid  
E:\Users\dtarb\Scratch\Logan\loganordpd.tif

Output Network Connectivity Tree (txt)  
E:\Users\dtarb\Scratch\Logan\logantreepd.txt

Output Network Coordinates (txt)  
E:\Users\dtarb\Scratch\Logan\logancoordpd.txt

Output Stream Reach Shapefile  
E:\Users\dtarb\Scratch\Logan\logannetpd.shp

Output Watershed Grid  
E:\Users\dtarb\Scratch\Logan\loganwpd.tif

**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 - Link Number of the downstream link. -1 indicates that this does not exist
- USLINKNO1 - Link Number of first upstream link. (0 indicates no link upstream, i.e. for a source link)
- USLINKNO2 - 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.)
- DSNODEID - Node identifier for node at

# Illustration of some other functions

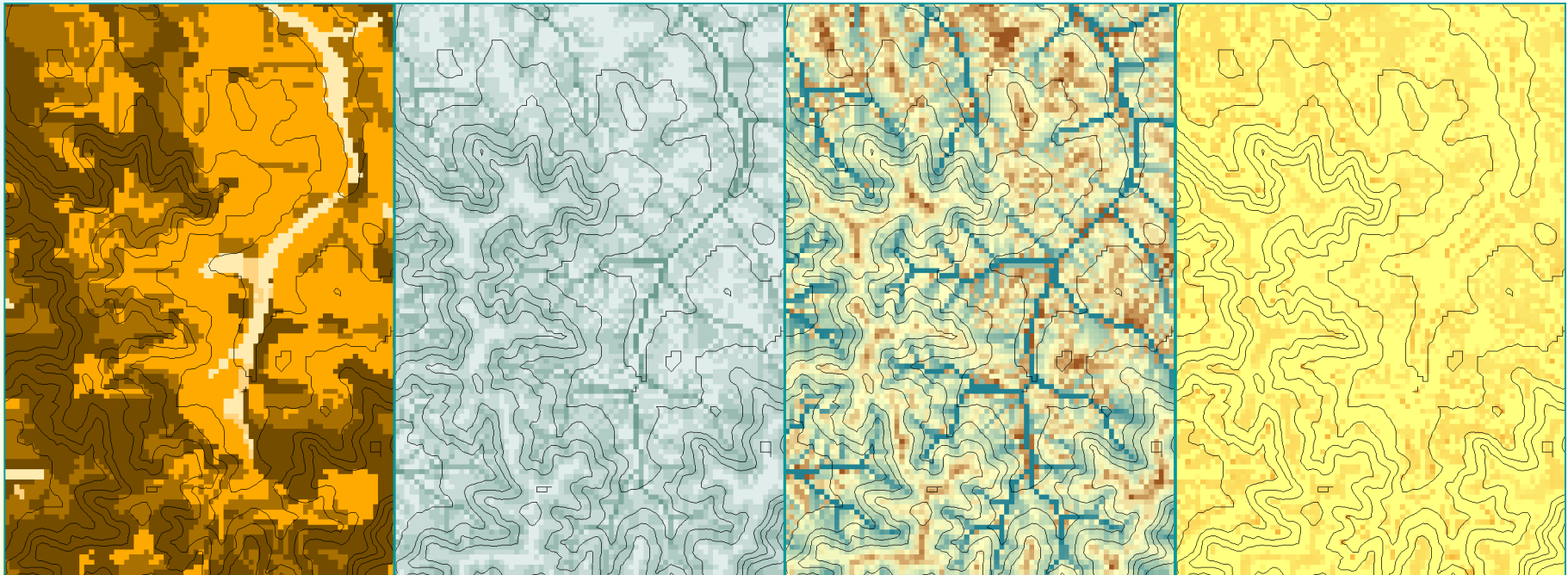
## Transport limited accumulation

Supply

Capacity

Transport

Deposition



$S$

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

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

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

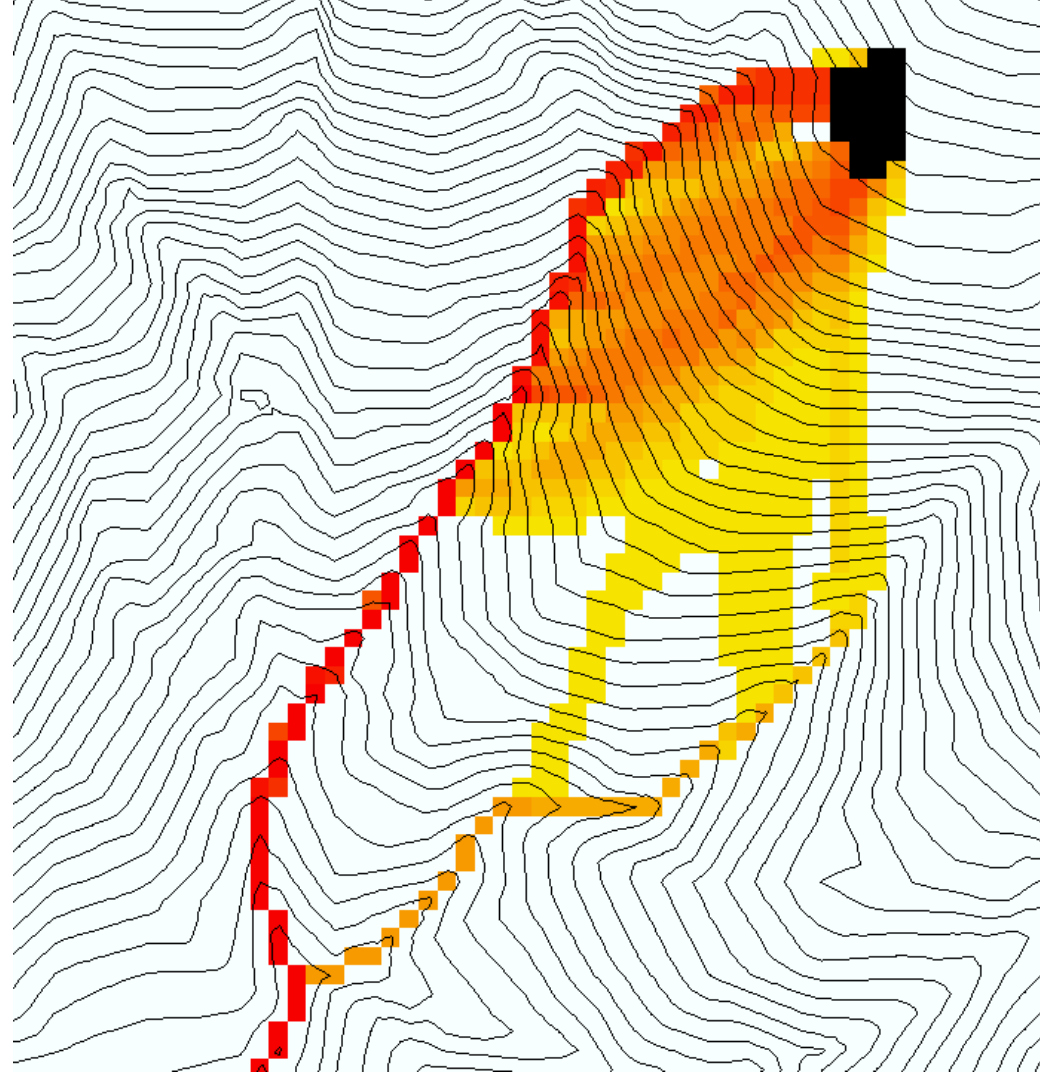
Useful for modeling erosion and sediment delivery, the spatial dependence of sediment delivery ratio and contaminant that adheres to sediment

# 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

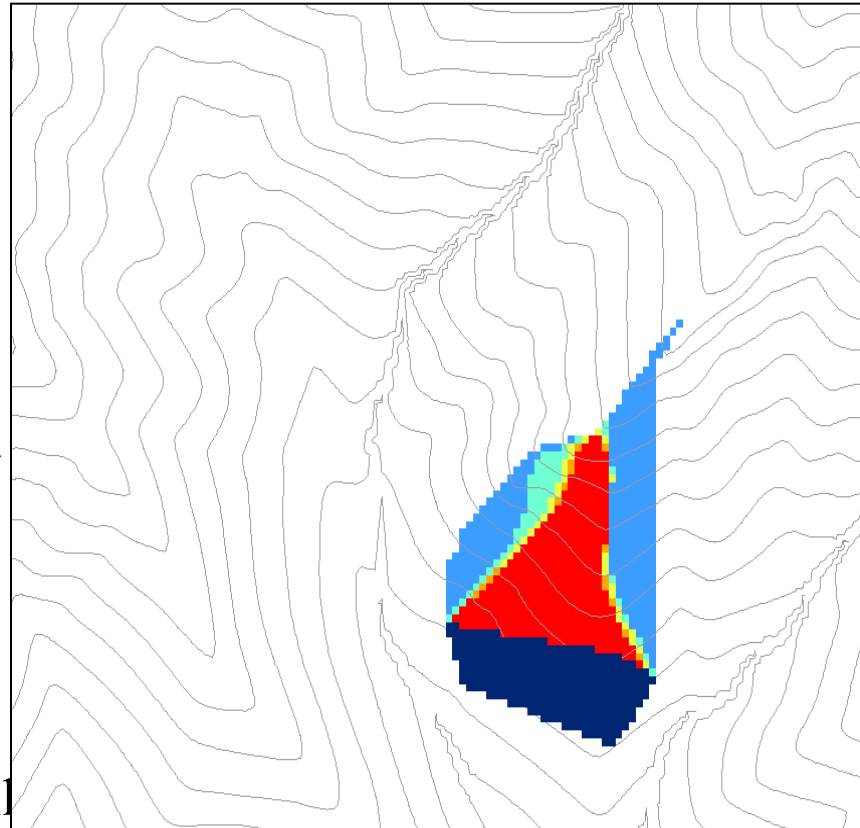
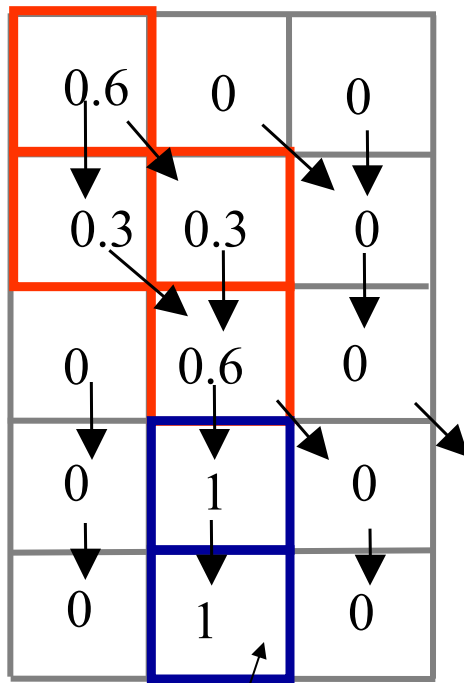
$$DA[m(x)] = DA(i, j) = m(i, j)\Delta^2 + \sum_{k \text{ contributing neighbors}} p_k d(i_k, j_k) DA(i_k, j_k)$$

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  $y$ .



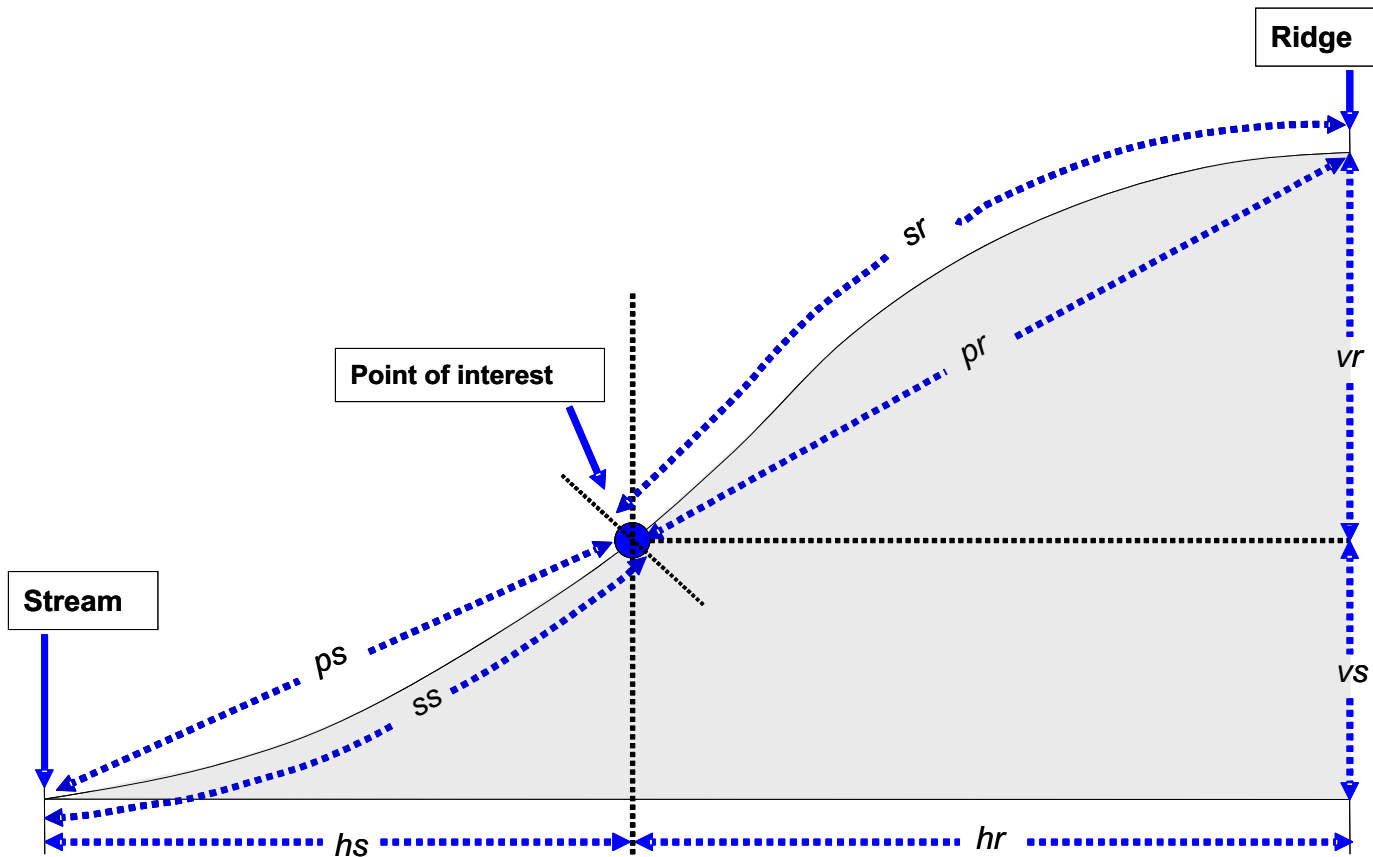
Grid cells  $y$

Dependence function of grid cell

Useful for example to track where a contaminant may come from



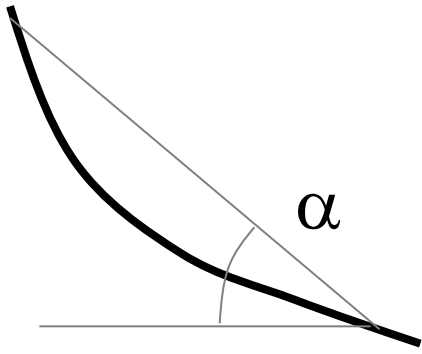
# Distance Down and Distance Up



Types of distance measurements possible in distance down and distance up functions.



# Avalanche Runout



Upslope recursion to determine elevation and distance to point in trigger zone that has the highest alpha angle

