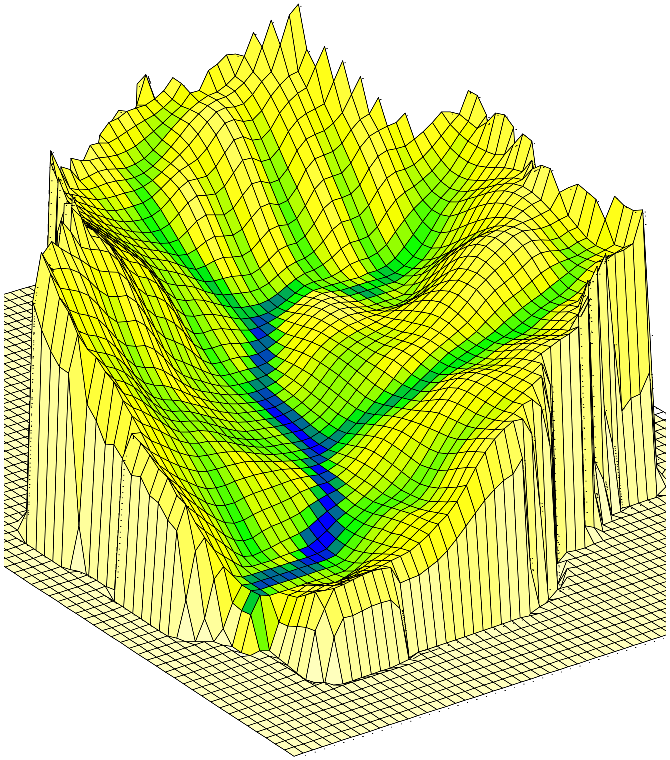
Terrain Analysis Using Digital Elevation Models (TauDEM)



David Tarboton[1], Dan Watson[2], Rob Wallace,[3] Kim Schreuders[1], Jeremy Neff[1]

[1]Utah Water Research Laboratory, Utah State University, Logan, Utah

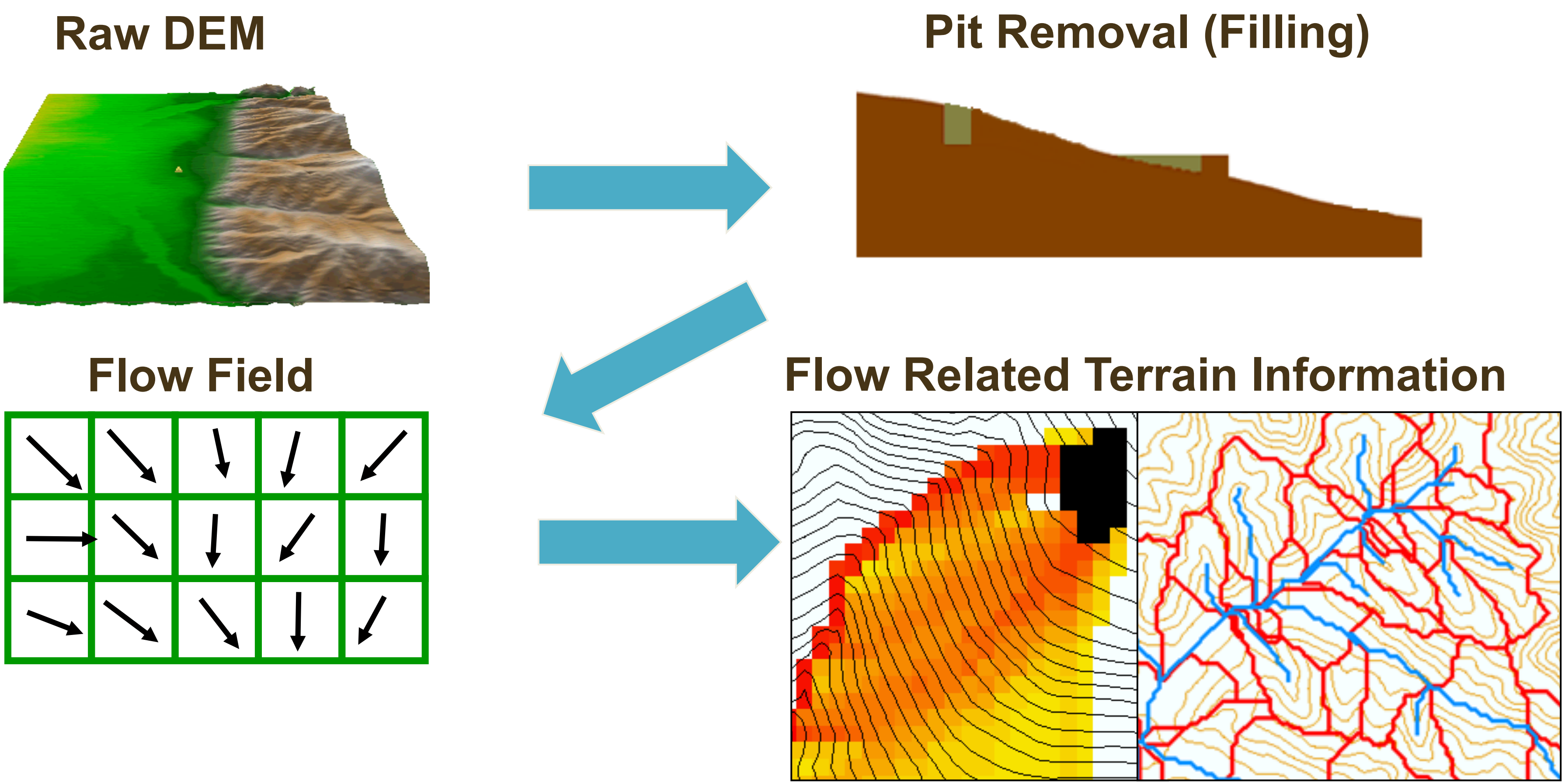
[2]Computer Science, Utah State University, Logan, Utah

[3]US Army Engineer Research and Development Center, Information Technology Lab, Vicksburg, Mississippi

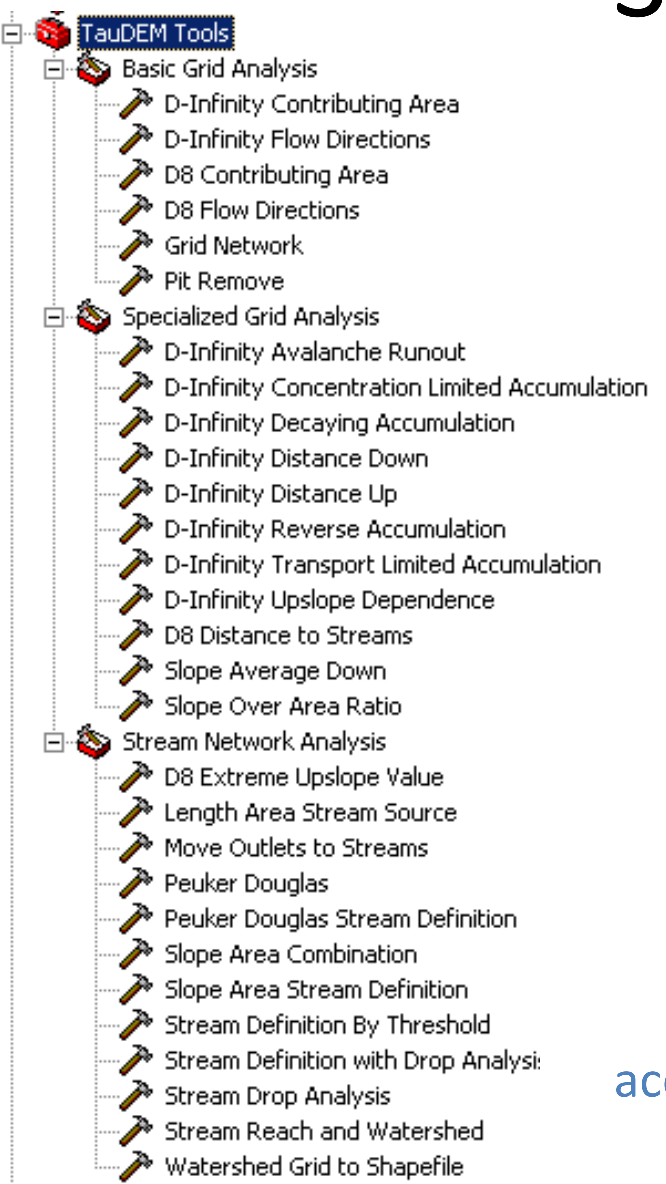
System-Wide Water SWWRP Resources Program

This research was funded by the US Army Research and Development Center under contract number W9124Z-08-P-0420

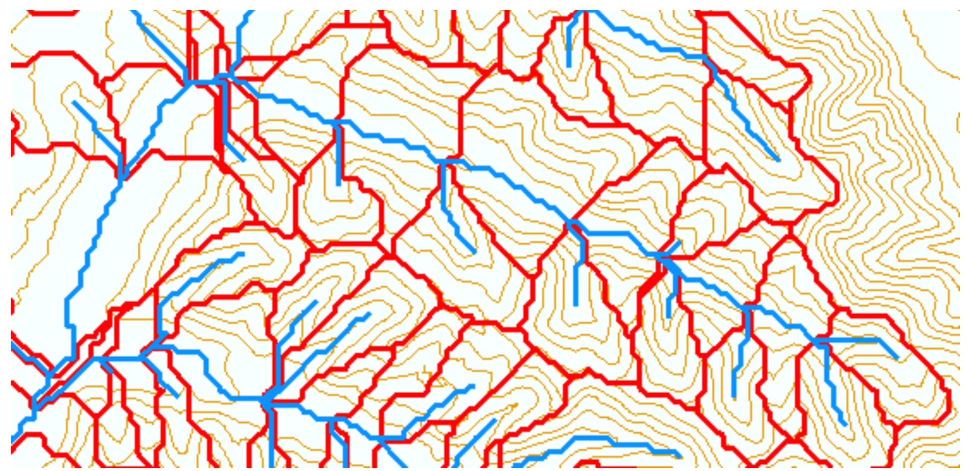
Deriving hydrologically useful information from Digital Elevation Models



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

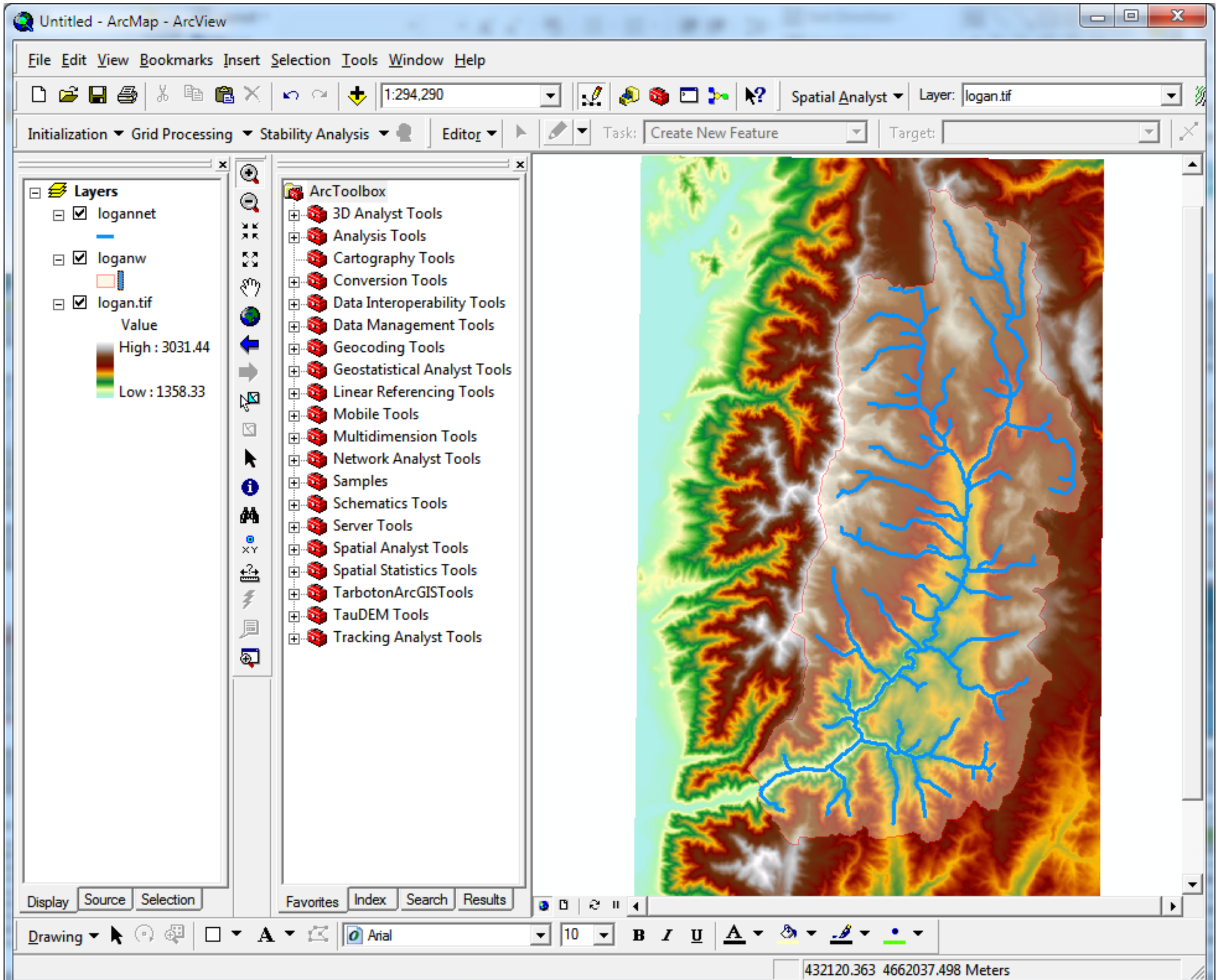
10^4 cells/km^2

2010's LIDAR ~1 m

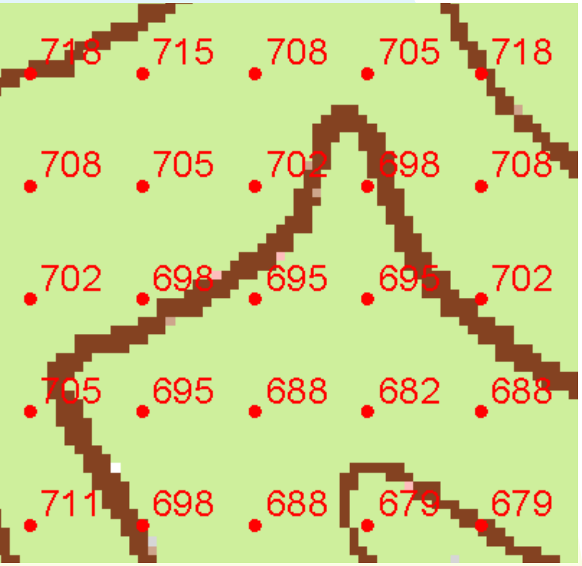
10^6 cells/km^2

Website and Demo

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



Grid Data Format Assumptions



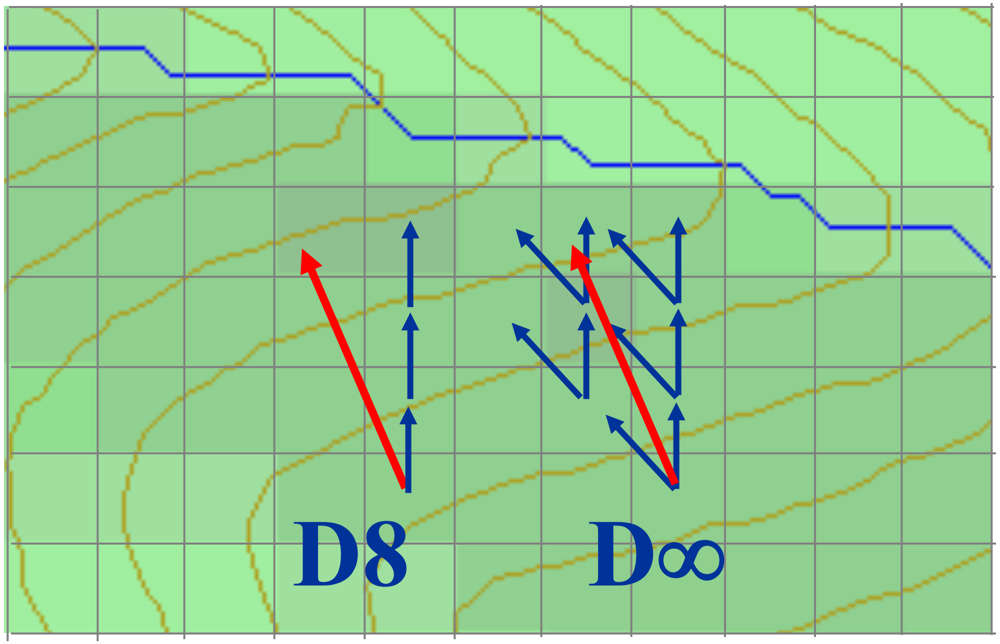
- Input and output grids are uncompressed GeoTIFF
- Maximum size 4 GB
- GDAL Nodata tag preferred (if not present, a missing value is assumed)
- Grids are square ($\Delta x = \Delta y$)
- Grids have identical in extent, cell size and spatial reference
- Spatial reference information is not used (no projection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

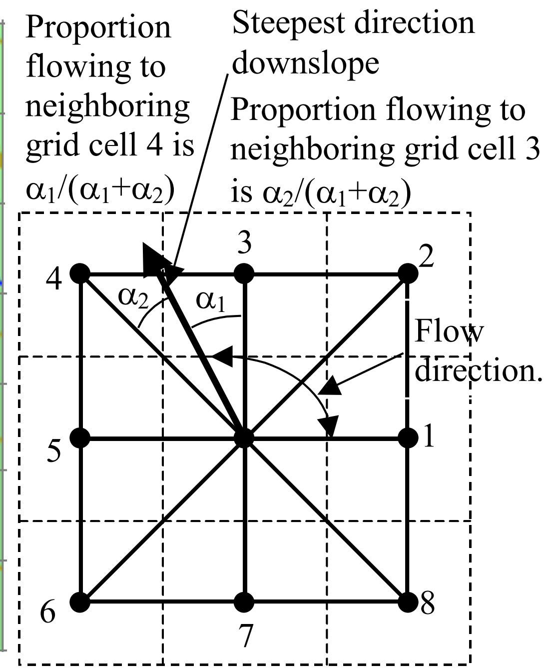
Representation of Flow Field

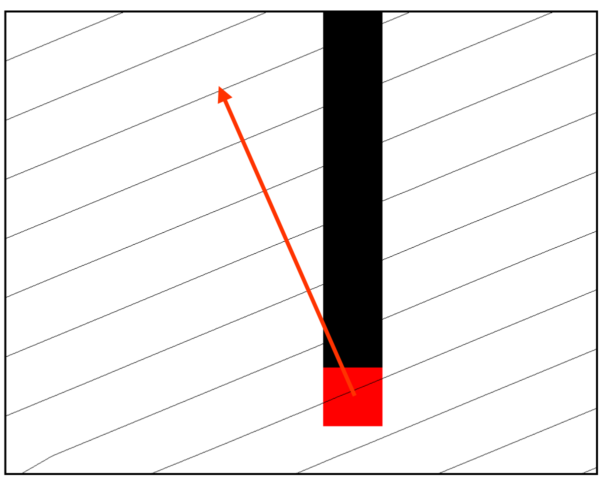
Steepest single direction

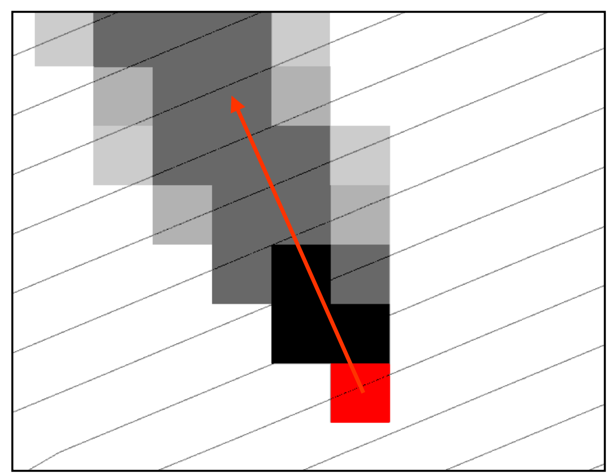
48	52
56	67

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





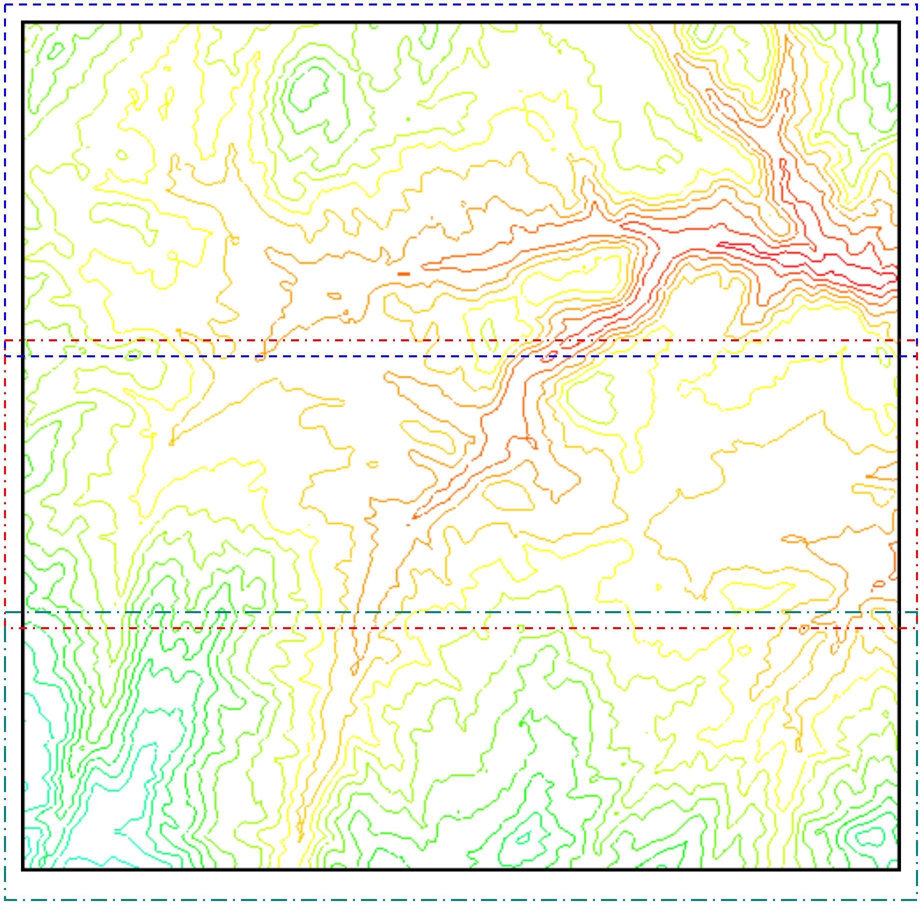




Tarboton, D. G., (1997), "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models," Water Resources Research, 33(2): 309-319.)

Parallel Approach

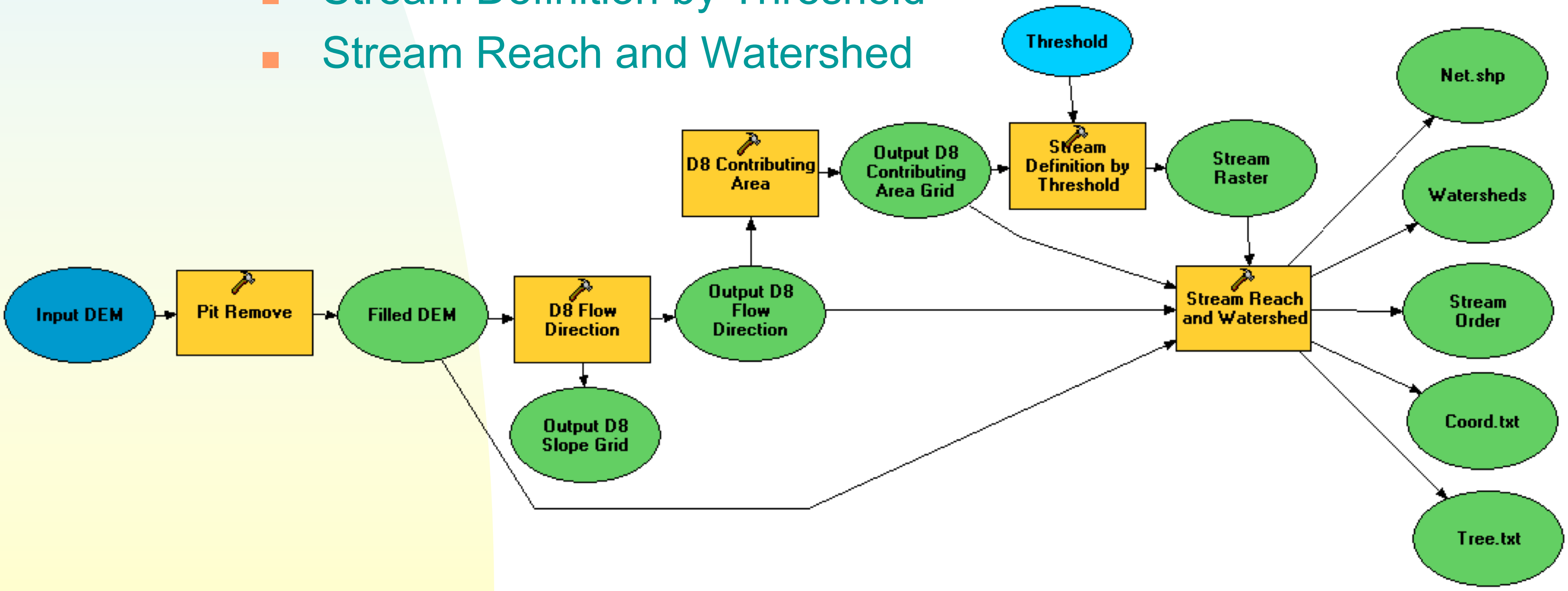
- MPI, distributed memory paradigm
- Row oriented slices
- Each process includes one buffer row on either side
- Each process does not change buffer row



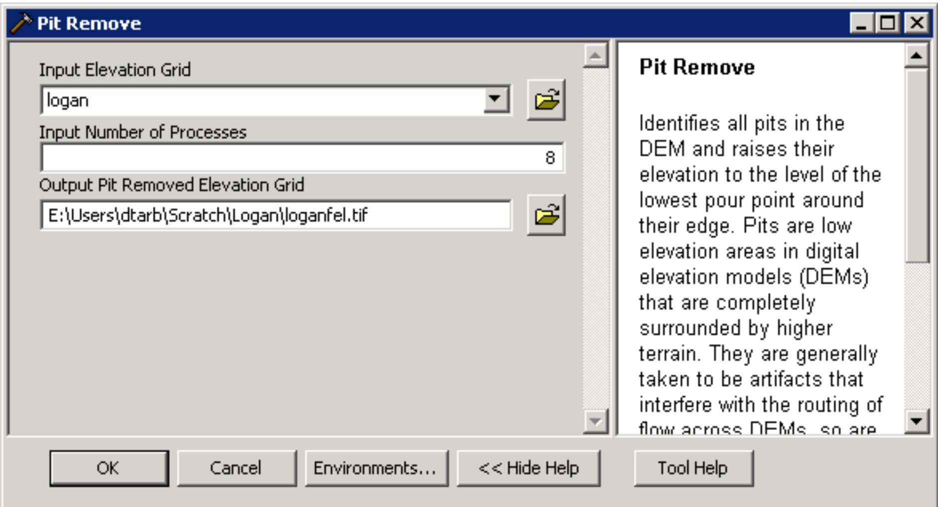
Illustrative Use Case: Delineation of channels and watersheds using a constant support area threshold

Steps

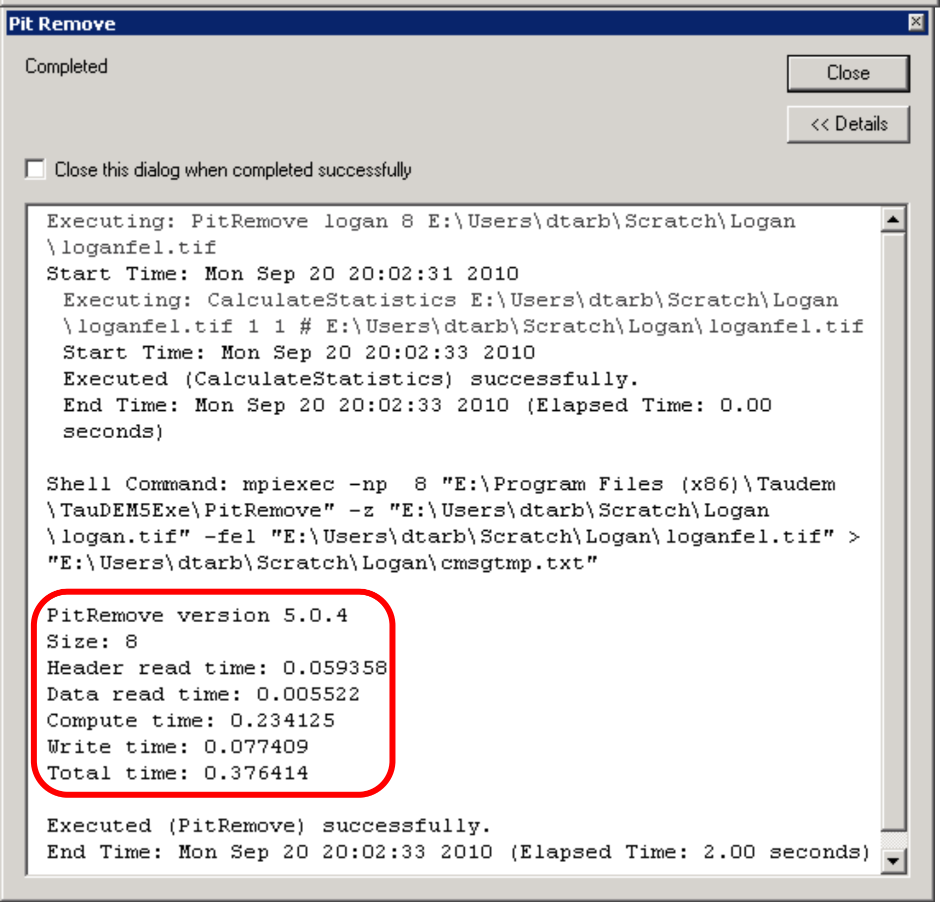
- Pit Remove
- D8 Flow Directions
- D8 Contributing Area
- Stream Definition by Threshold
- Stream Reach and Watershed



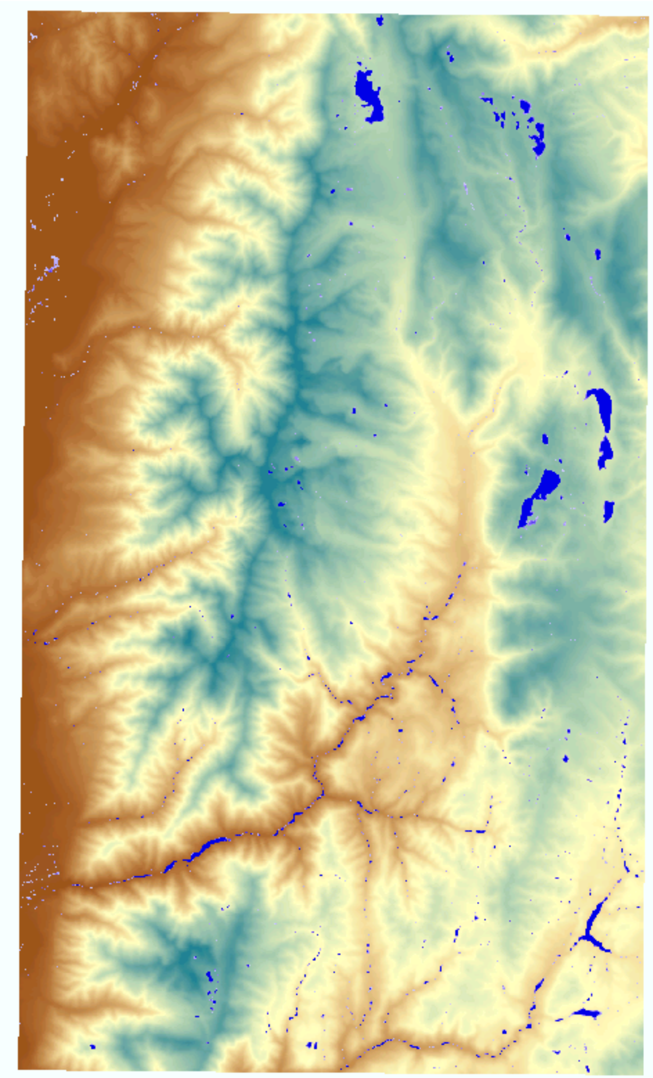
Pit Remove



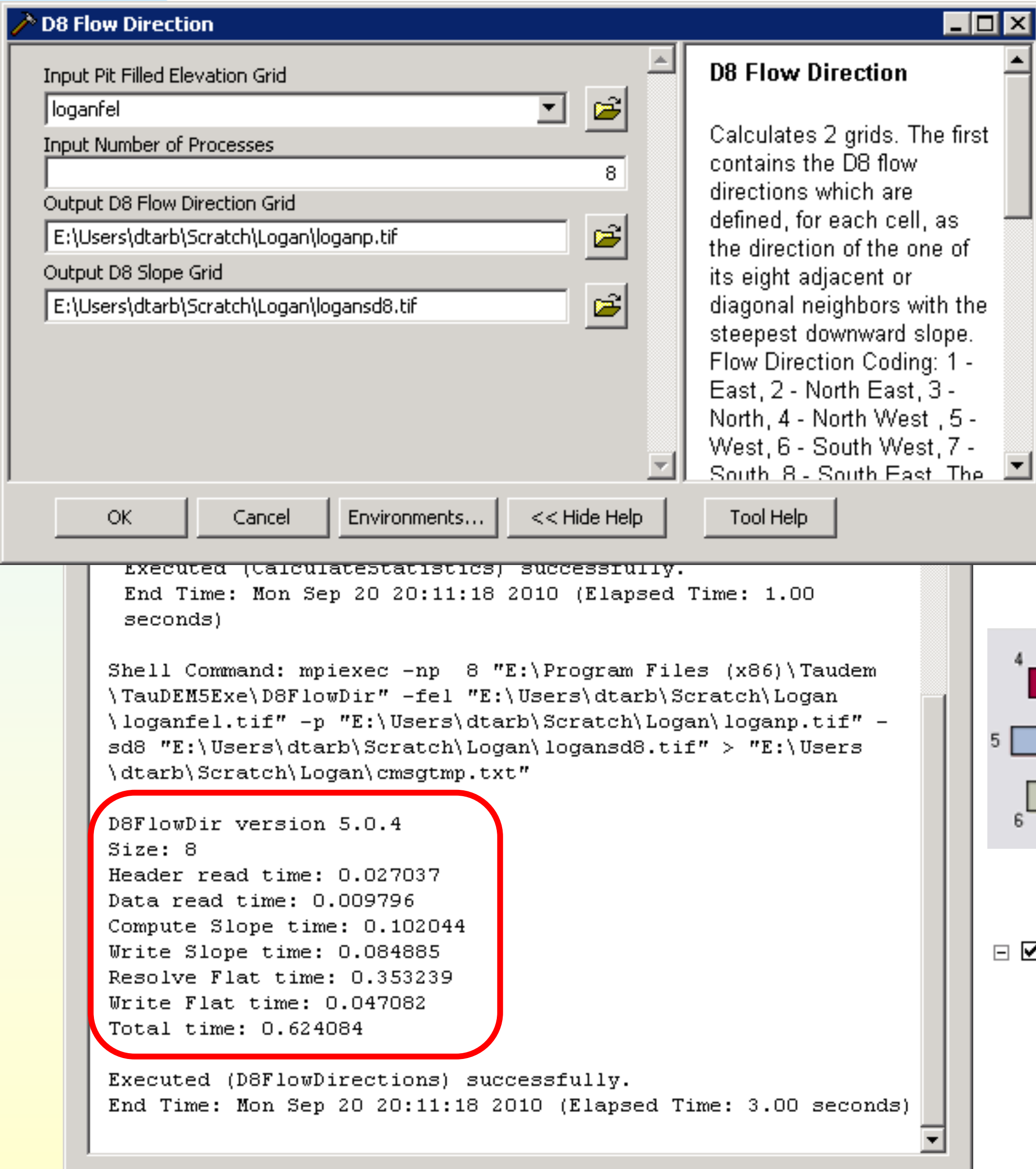
Pit Remove
Input Elevation Grid
logan
Input Number of Processes
8
Output Pit Removed Elevation Grid
E:\Users\dtarb\Scratch\Logan\loganfel.tif
Pit Remove
Identifies all pits in the DEM and raises their elevation to the level of the lowest pour point around their edge. Pits are low elevation areas in digital 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
OK
Cancel
Environments...
<< Hide Help
Tool Help

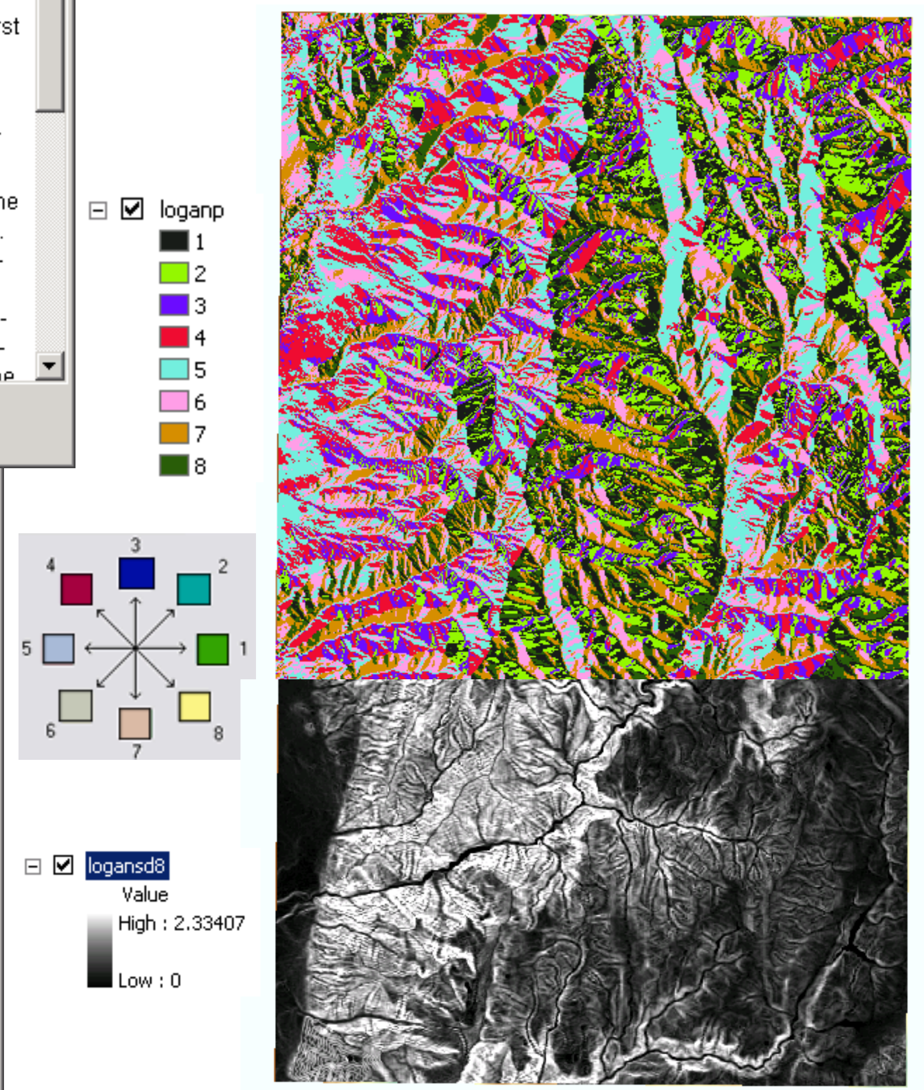


Pit Remove
Completed
Close
<< Details
Close this dialog when completed successfully
Executing: PitRemove logan 8 E:\Users\dtarb\Scratch\Logan\loganfel.tif
Start Time: Mon Sep 20 20:02:31 2010
Executing: CalculateStatistics E:\Users\dtarb\Scratch\Logan\loganfel.tif 1 1 # E:\Users\dtarb\Scratch\Logan\loganfel.tif
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 8 "E:\Program Files (x86)\Taudem\TauDEM5Exe\PitRemove" -z "E:\Users\dtarb\Scratch\Logan\logan.tif" -fel "E:\Users\dtarb\Scratch\Logan\loganfel.tif" > "E:\Users\dtarb\Scratch\Logan\cmsgtmp.txt"
PitRemove version 5.0.4
Size: 8
Header read time: 0.059358
Data read time: 0.005522
Compute time: 0.234125
Write time: 0.077409
Total time: 0.376414
Executed (PitRemove) successfully.
End Time: Mon Sep 20 20:02:33 2010 (Elapsed Time: 2.00 seconds)

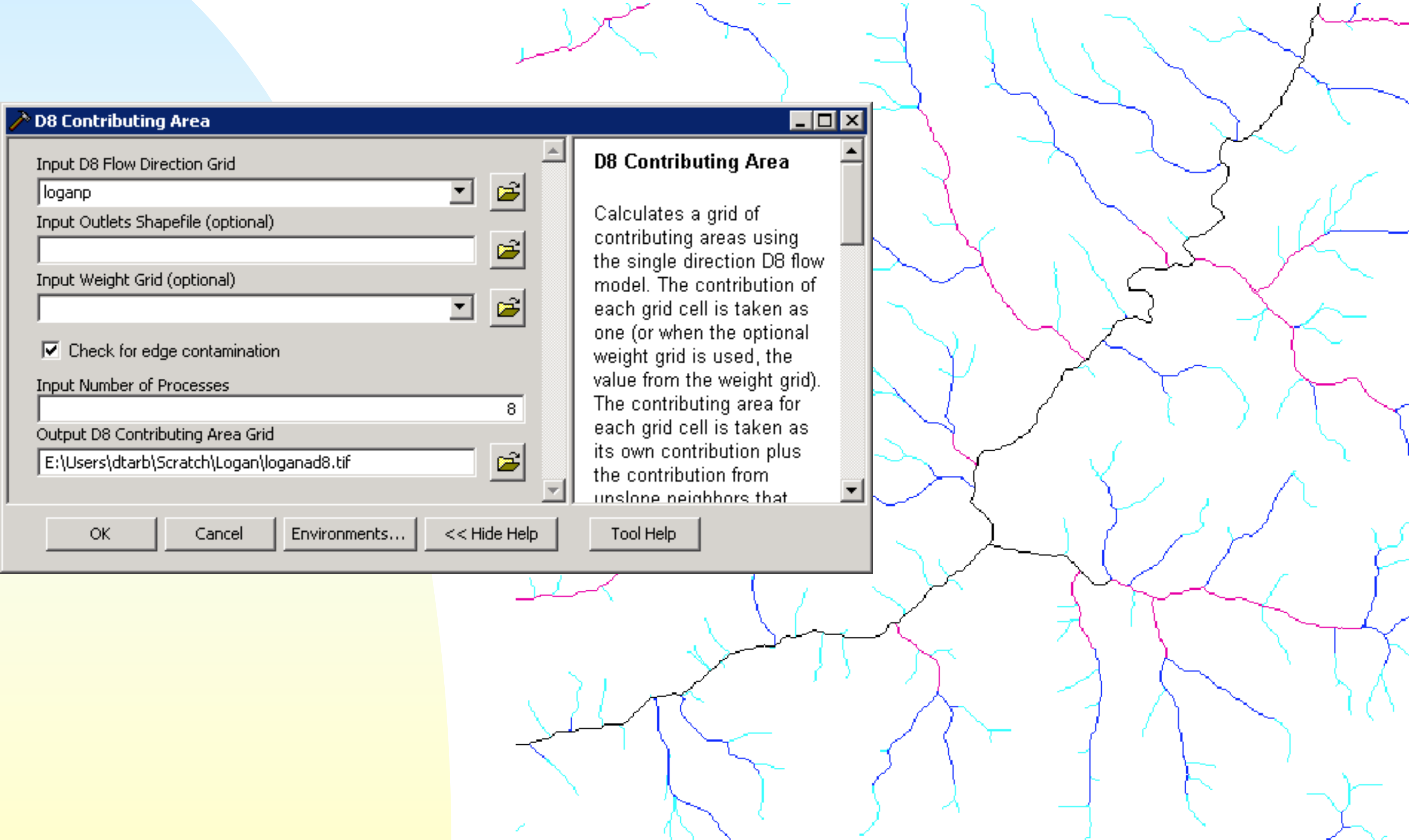


D8 Flow Direction (and Slope)

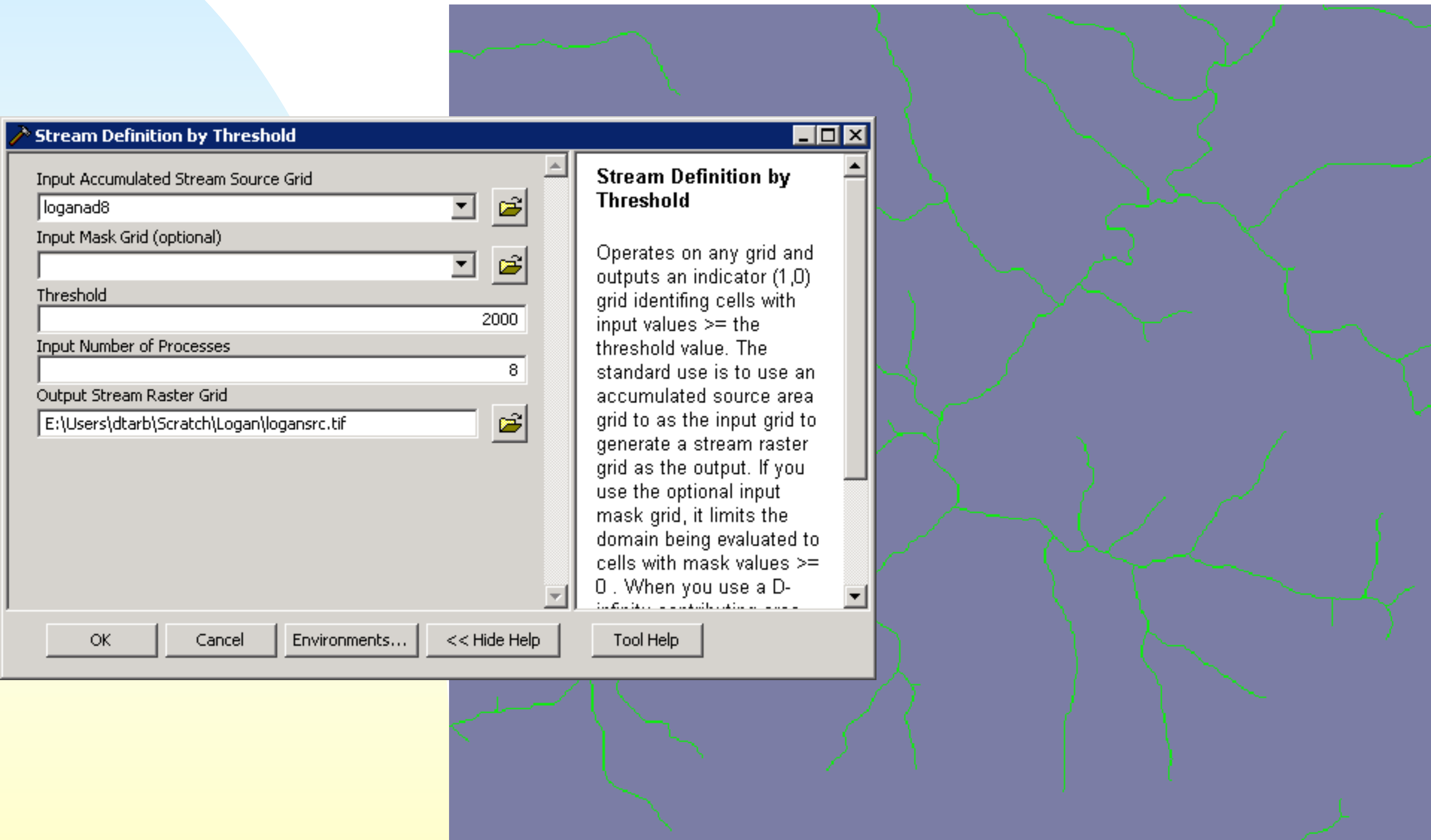




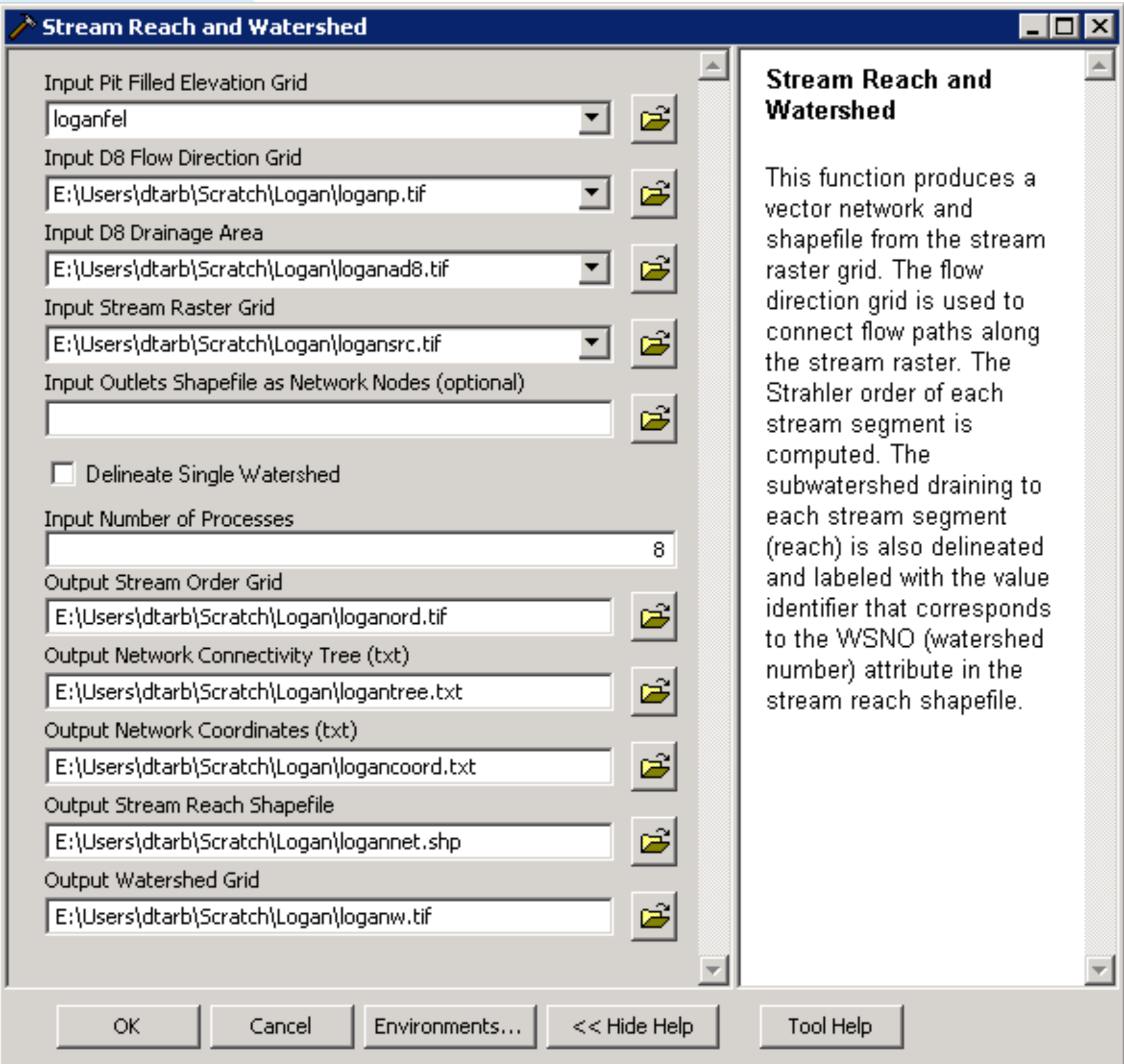
D8 Contributing Area

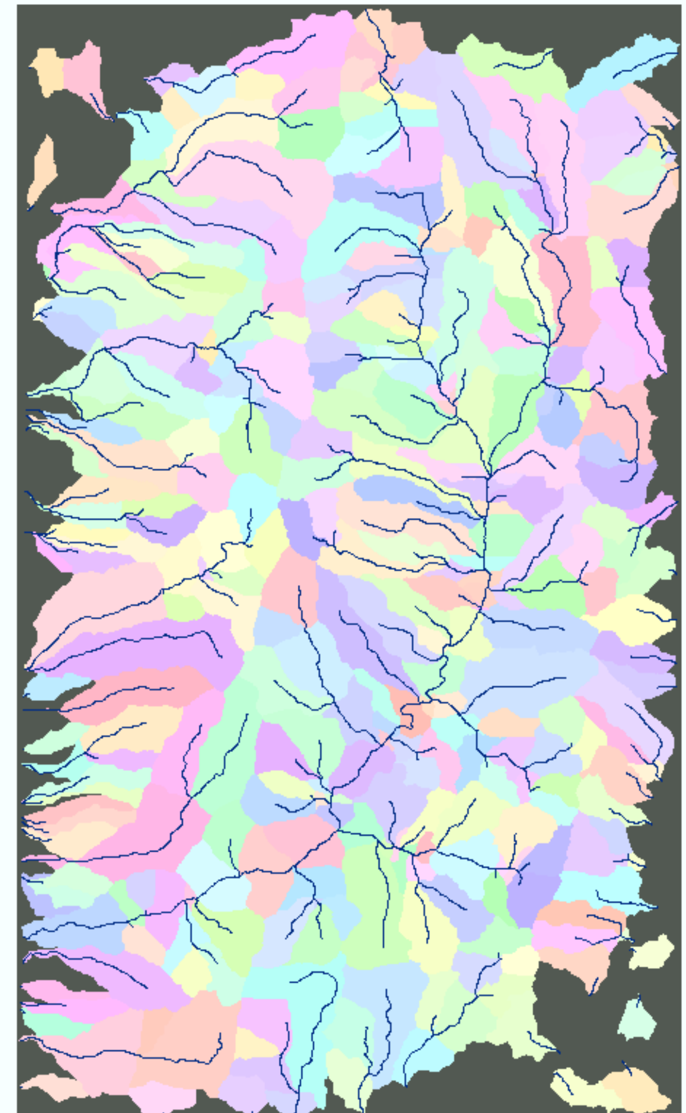


Stream Definition by Threshold



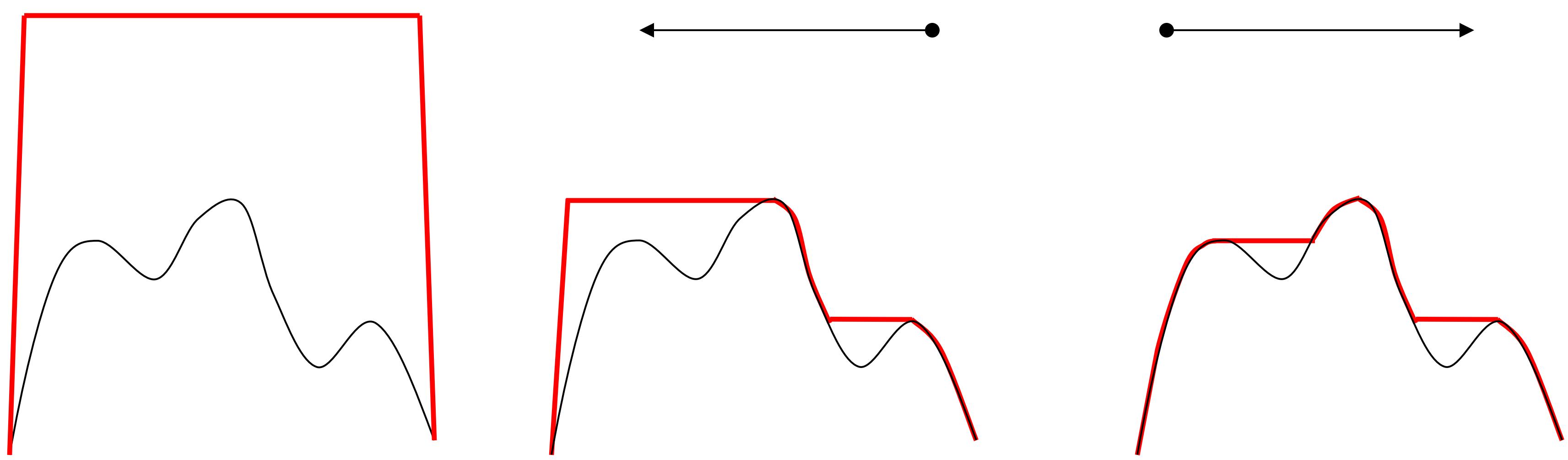
Stream Reach and Watershed





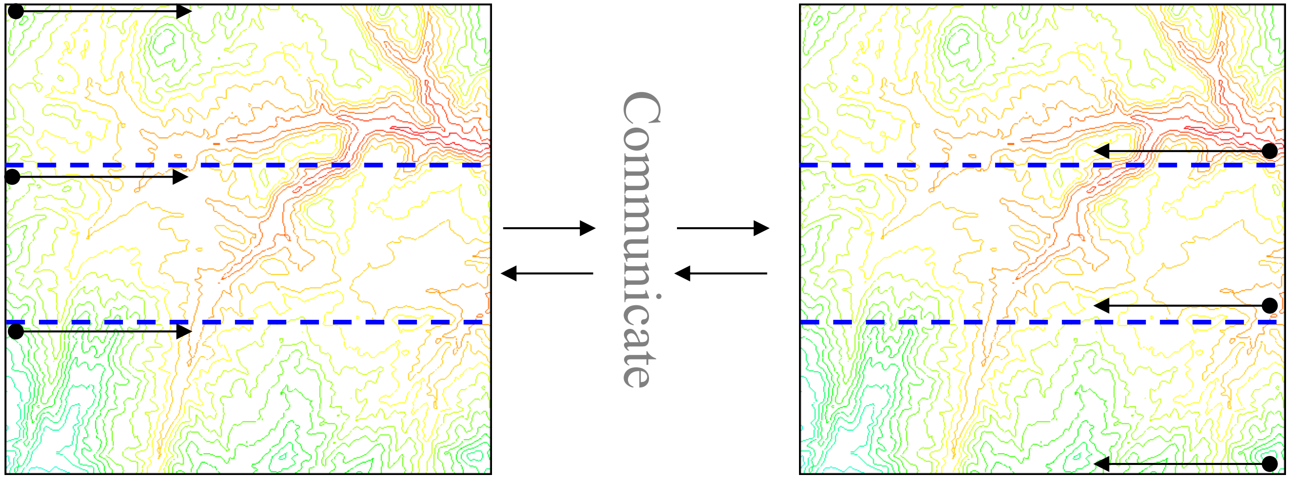
Some Algorithm Details
Pit Removal: Planchon Fill Algorithm

Initialization | 1st Pass | 2nd Pass



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

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)

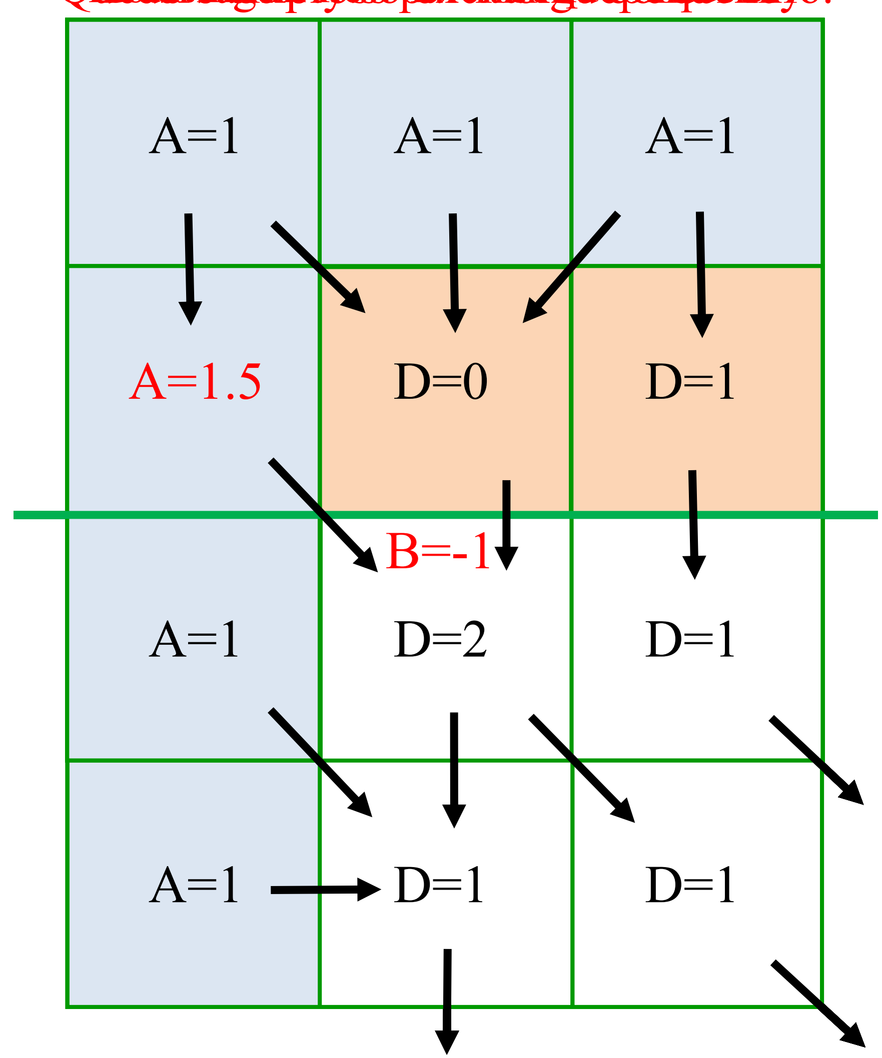
for all i
    for all k neighbors of i
        if $P_{ki}>0$ D(i)=D(i)+1
    if D(i)=0 add i to Q
next

2. Flow algebra function

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

FlowAlgebra(P,Q,D,θ,γ)

while Q isn't empty
    get i from Q
    $\underline{\theta}_i = FA(\underline{\gamma}_i, \underline{P}_{ki}, \underline{\theta}_k, \underline{\gamma}_k)$
    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

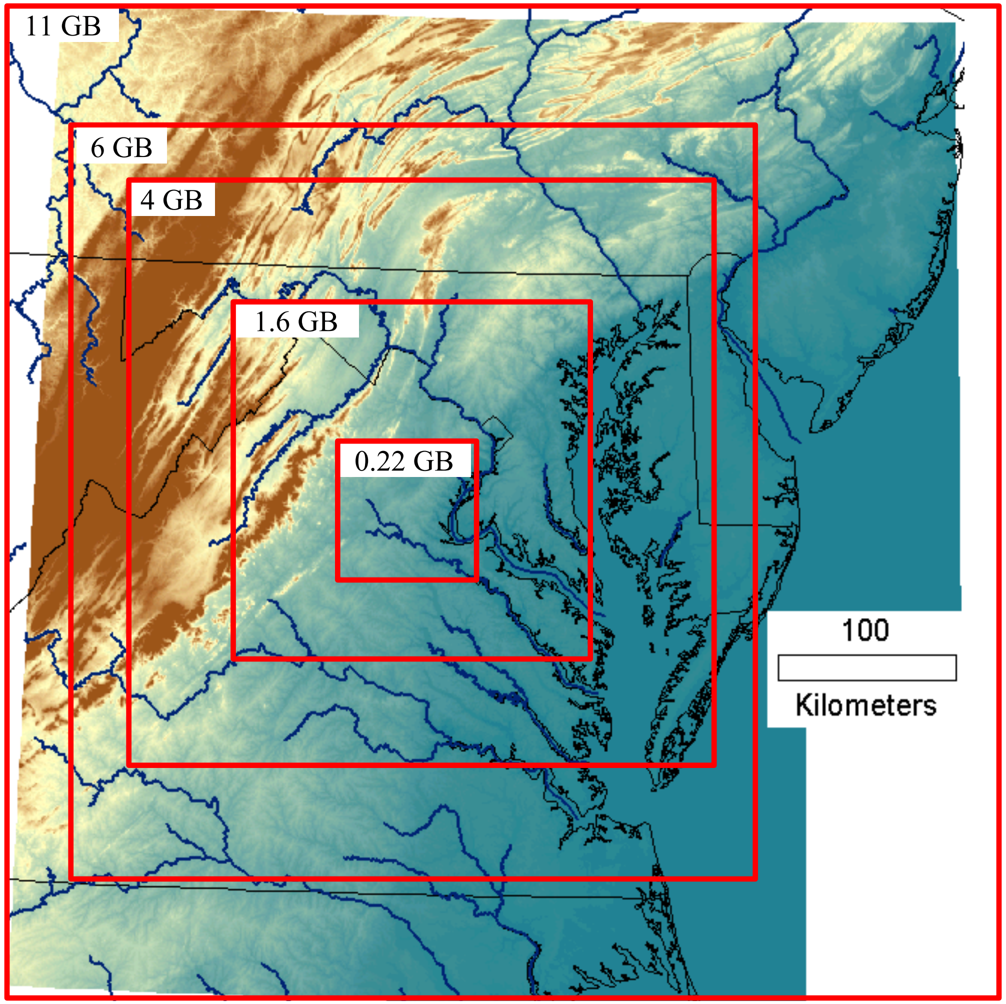


Capabilities Summary

Capability to run larger problems

		Processors used	Grid size	
			Theoretcal limit	Largest run
2008	TauDEM 4	1	0.22 GB	0.22 GB
Sept 2009	Partial implement-ation	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

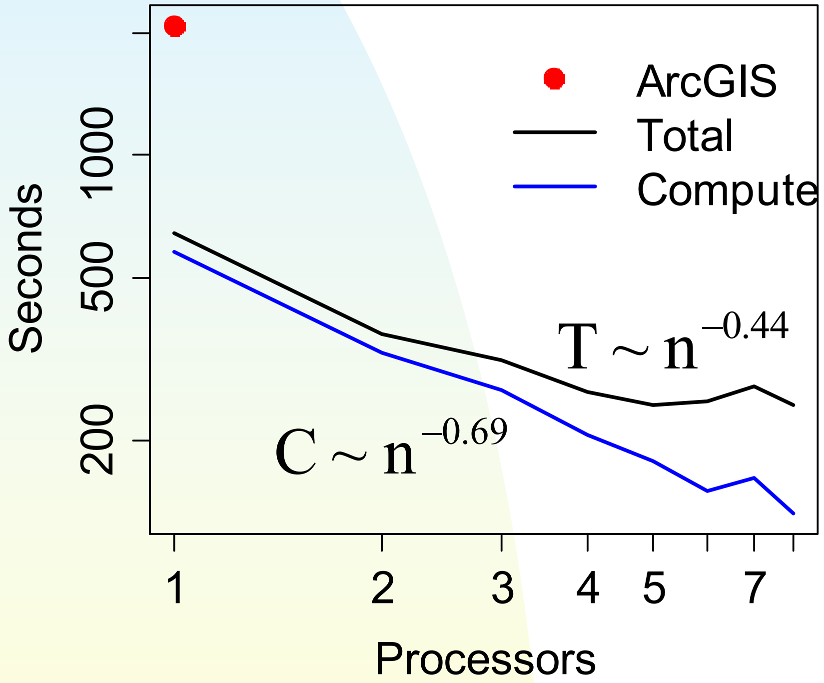
Single file size limit 4GB

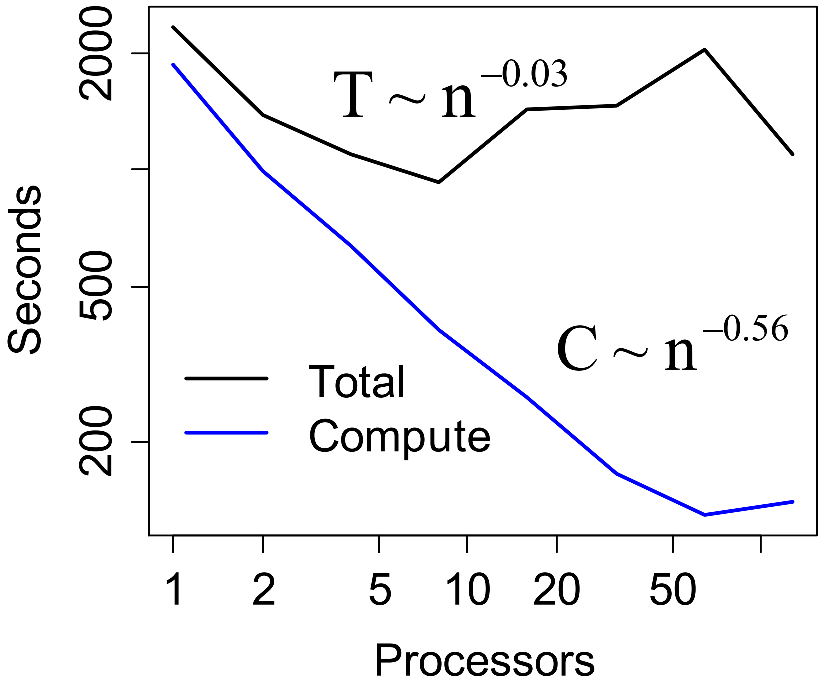


At 10 m grid cell size

Improved runtime efficiency

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





8 processor PC

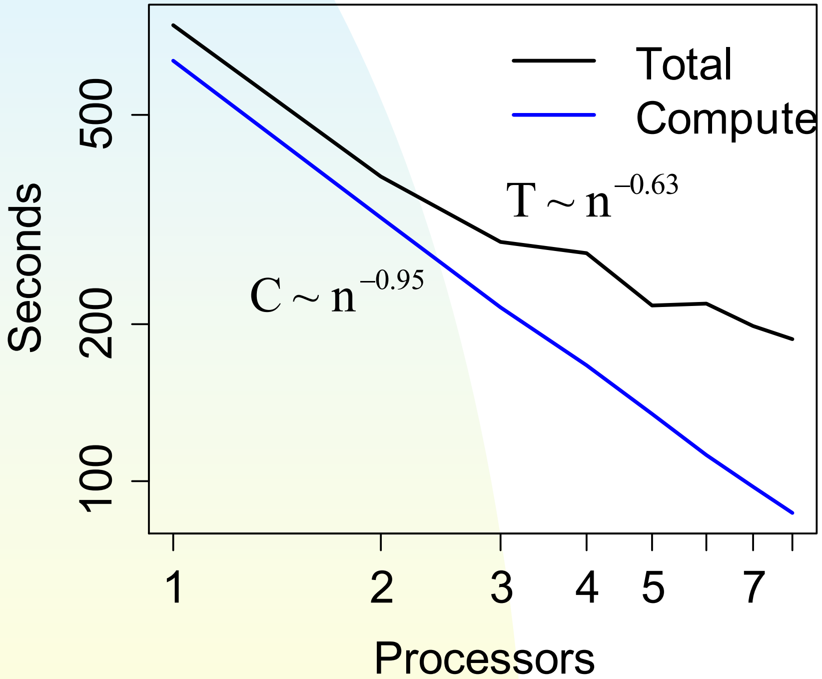
Dual quad-core Xeon E5405 2.0GHz PC with 16GB RAM

128 processor cluster

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

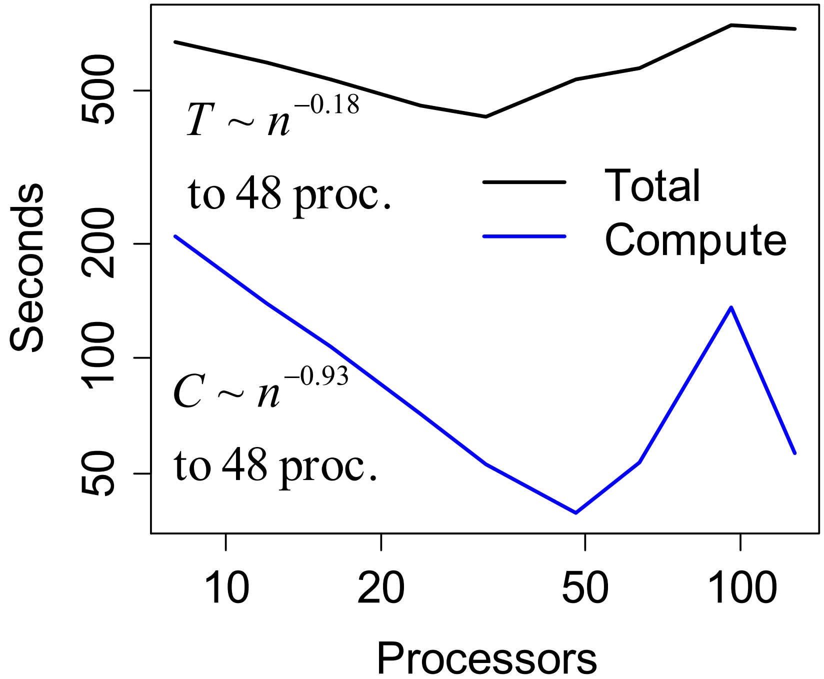
Improved runtime efficiency

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

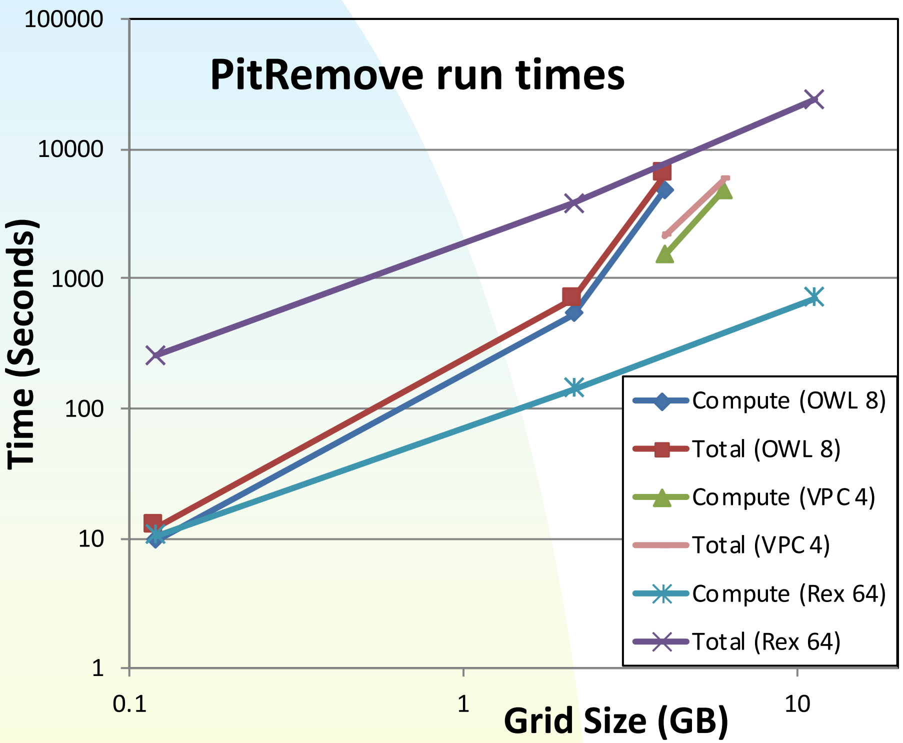
16 diskless Dell SC1435 compute nodes, each with 2.0GHz dual quad-core AMD Opteron 2350 processors with 8GB RAM

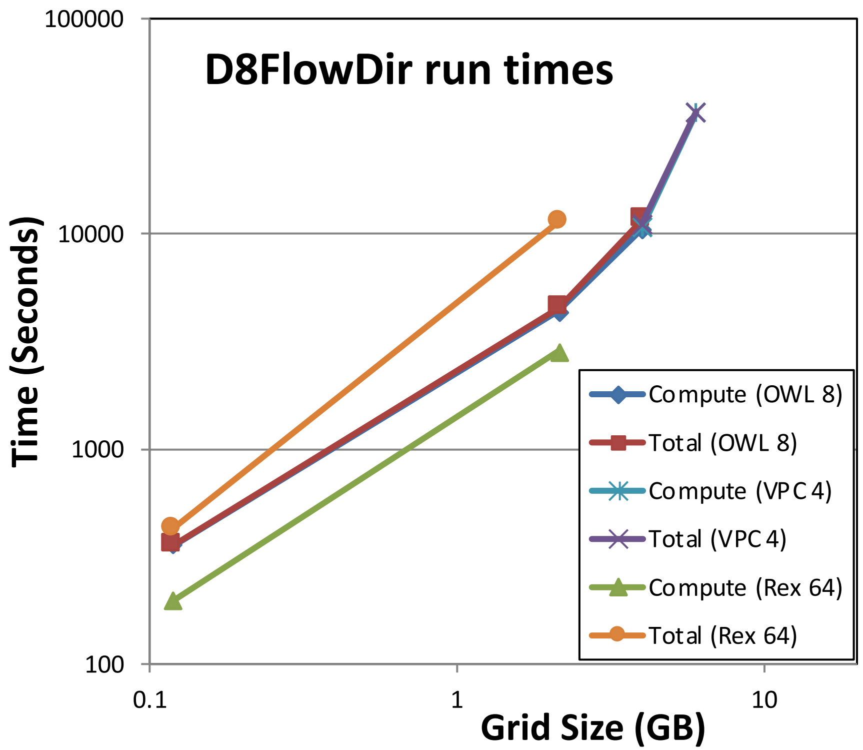
Scaling of run times to large grids

Dataset	Size	Hardware	Number of Processors	PitRemove (run time seconds)		D8FlowDir (run time seconds)	
	(GB)			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

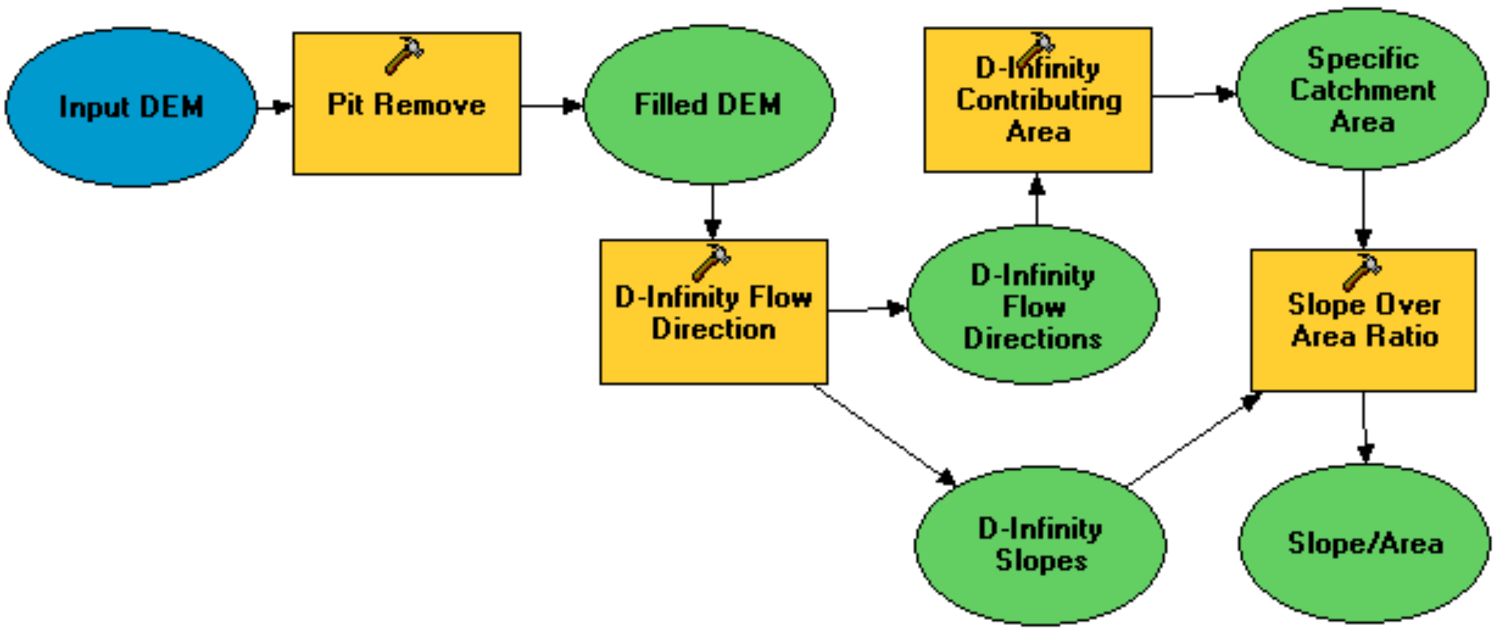
Additional Illustrative Use Cases

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

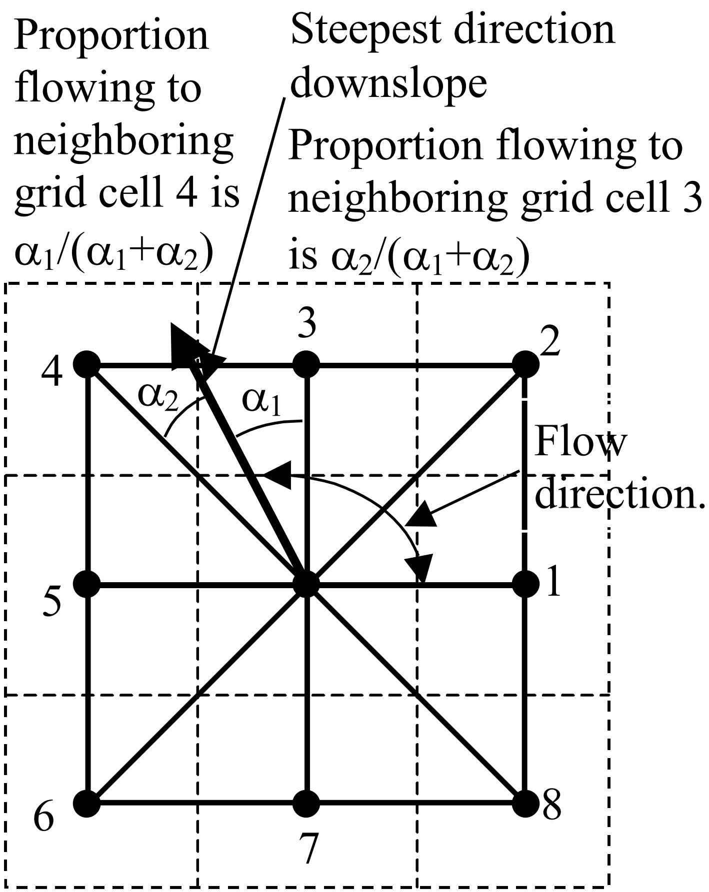
Topographic wetness index from the Dinfinity method

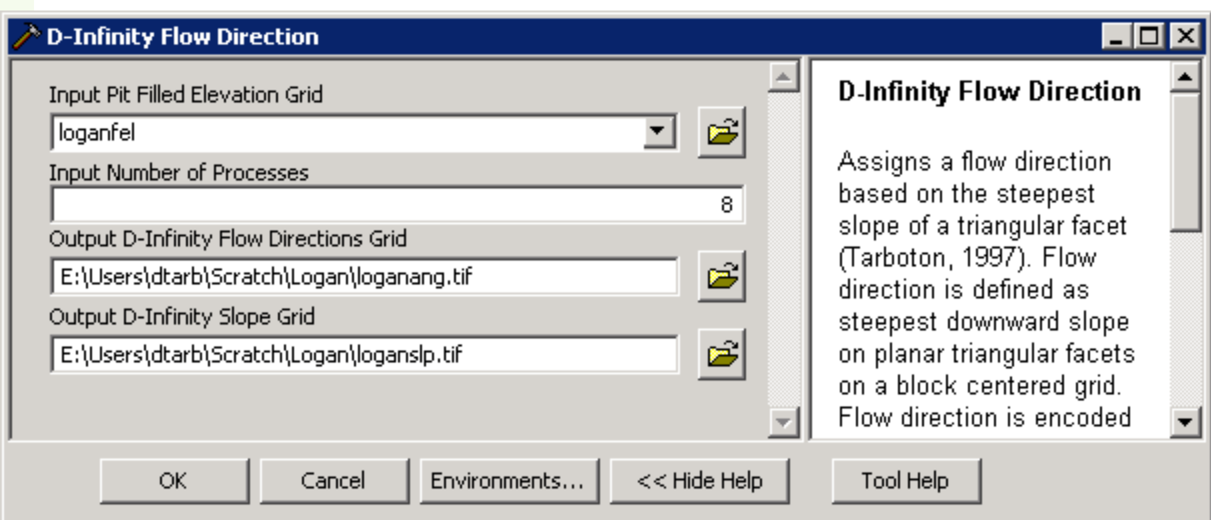
Steps

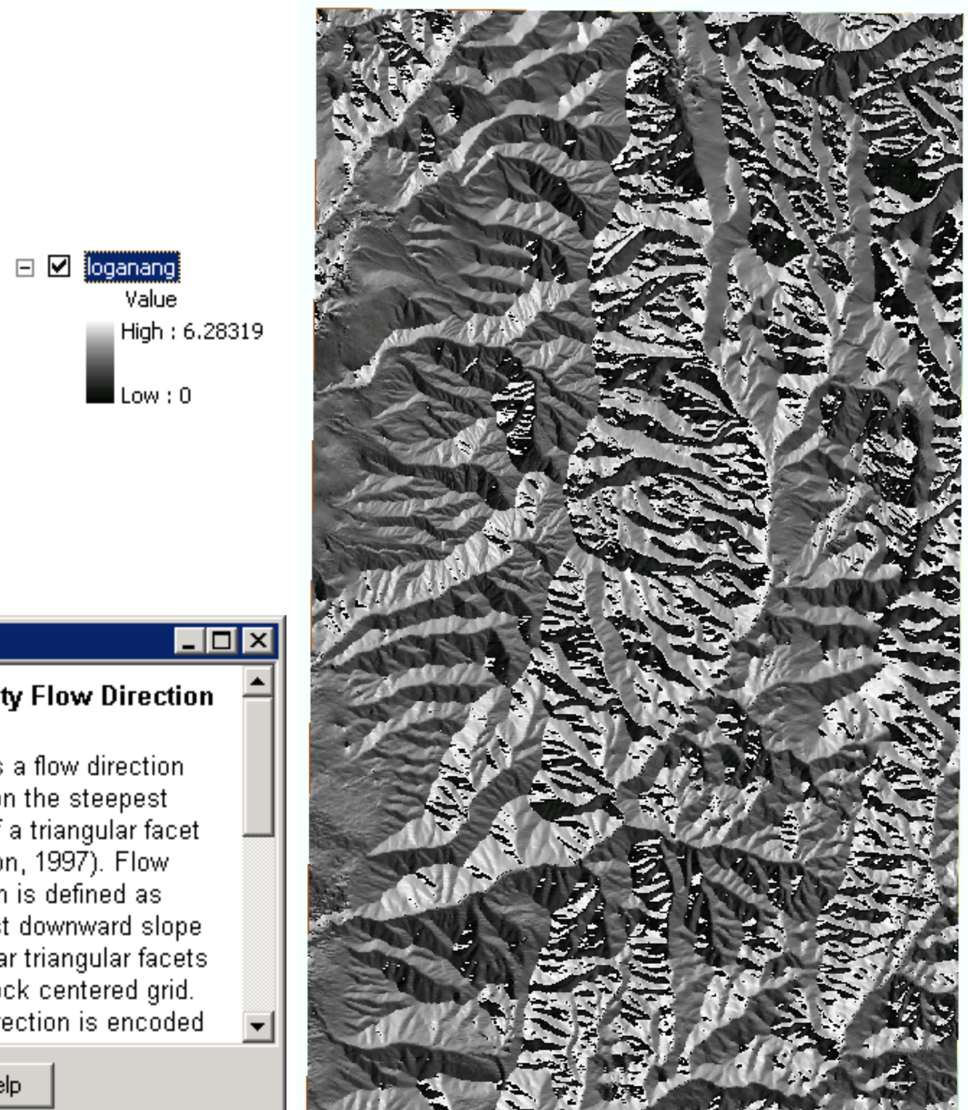
- Pit Remove
- D-Infinity Flow Directions (and Slopes)
- D-Infinity Contributing Area
- Slope Over Area Ratio



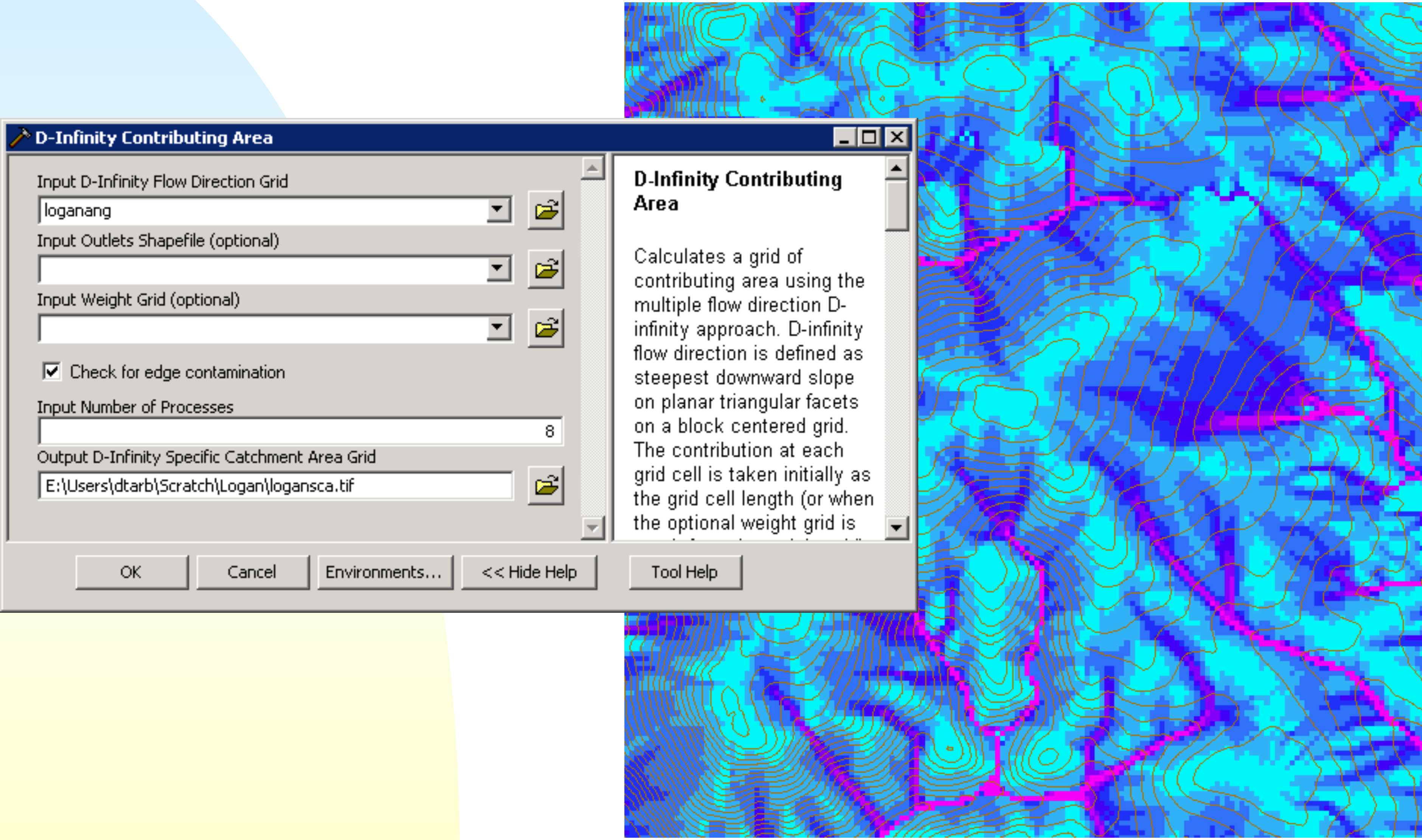
D-Infinity Flow Direction (and slope)





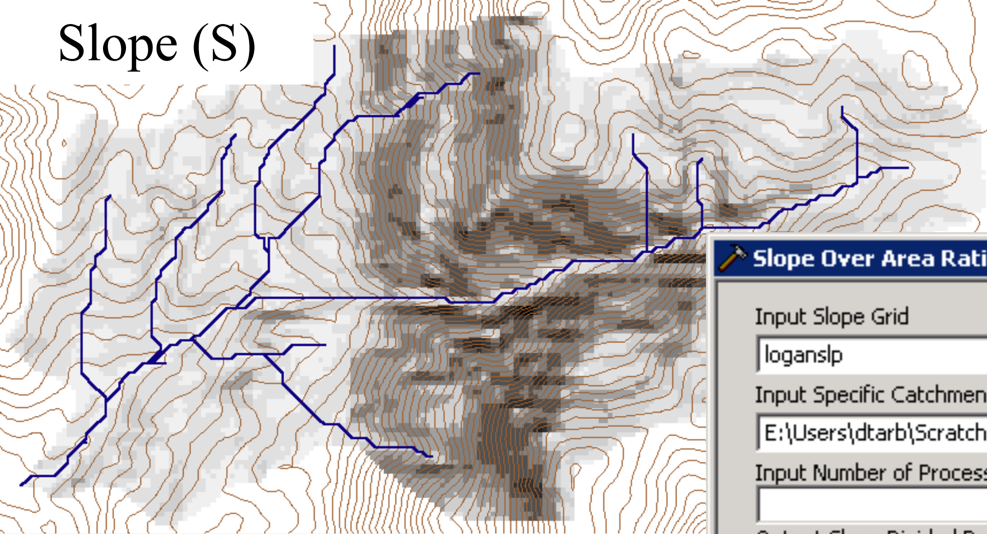


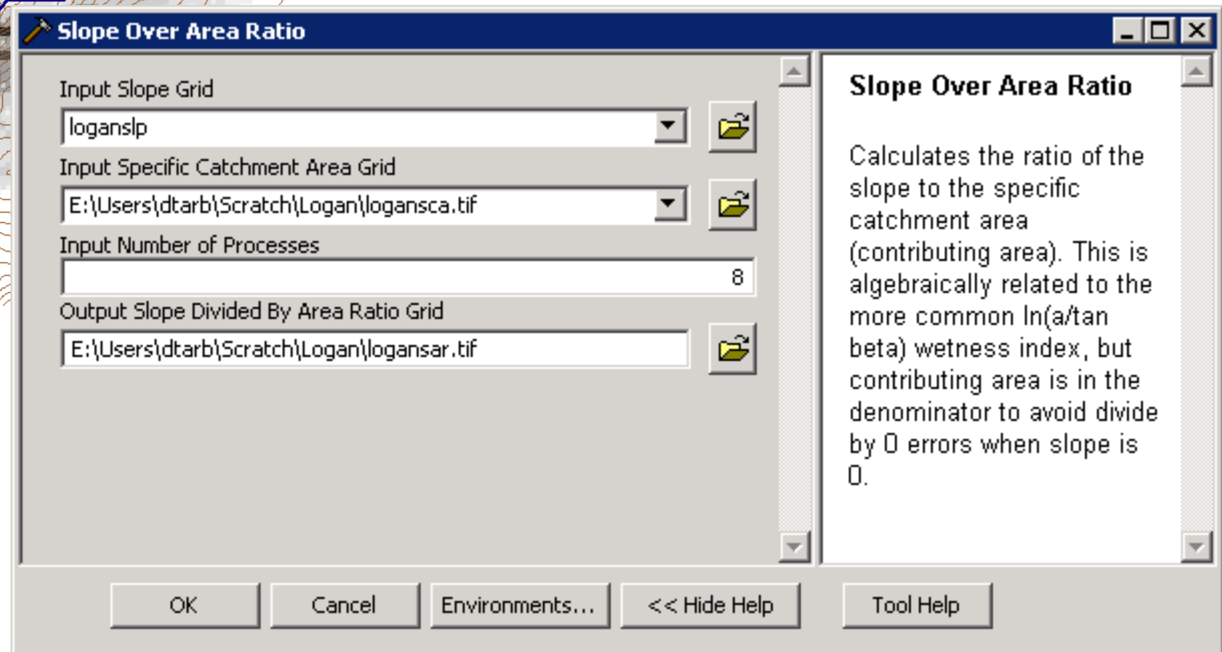
D-Infinity Contributing Area



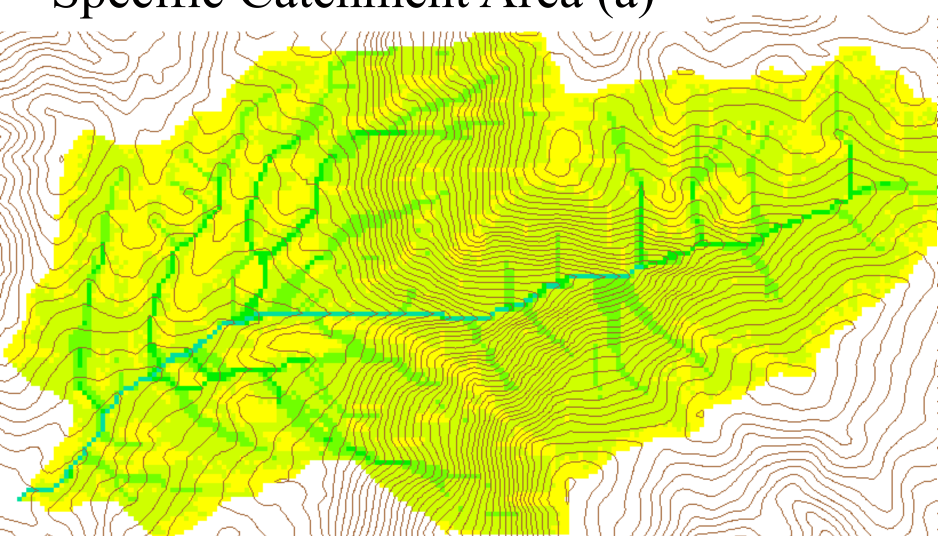
Wetness Index

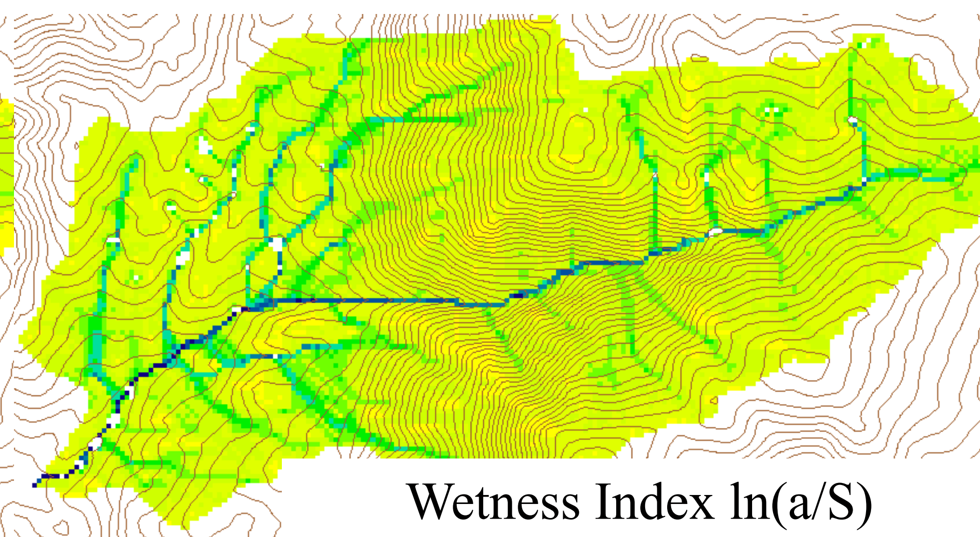
Slope (S)





Specific Catchment Area (a)





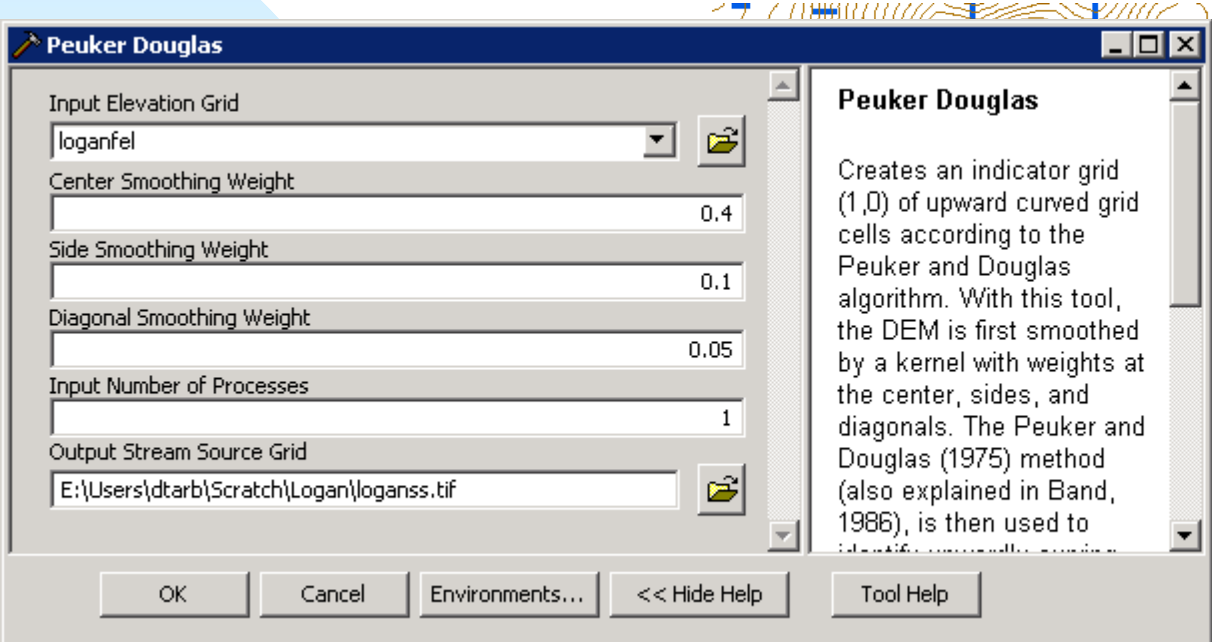
Wetness Index ln(a/S)

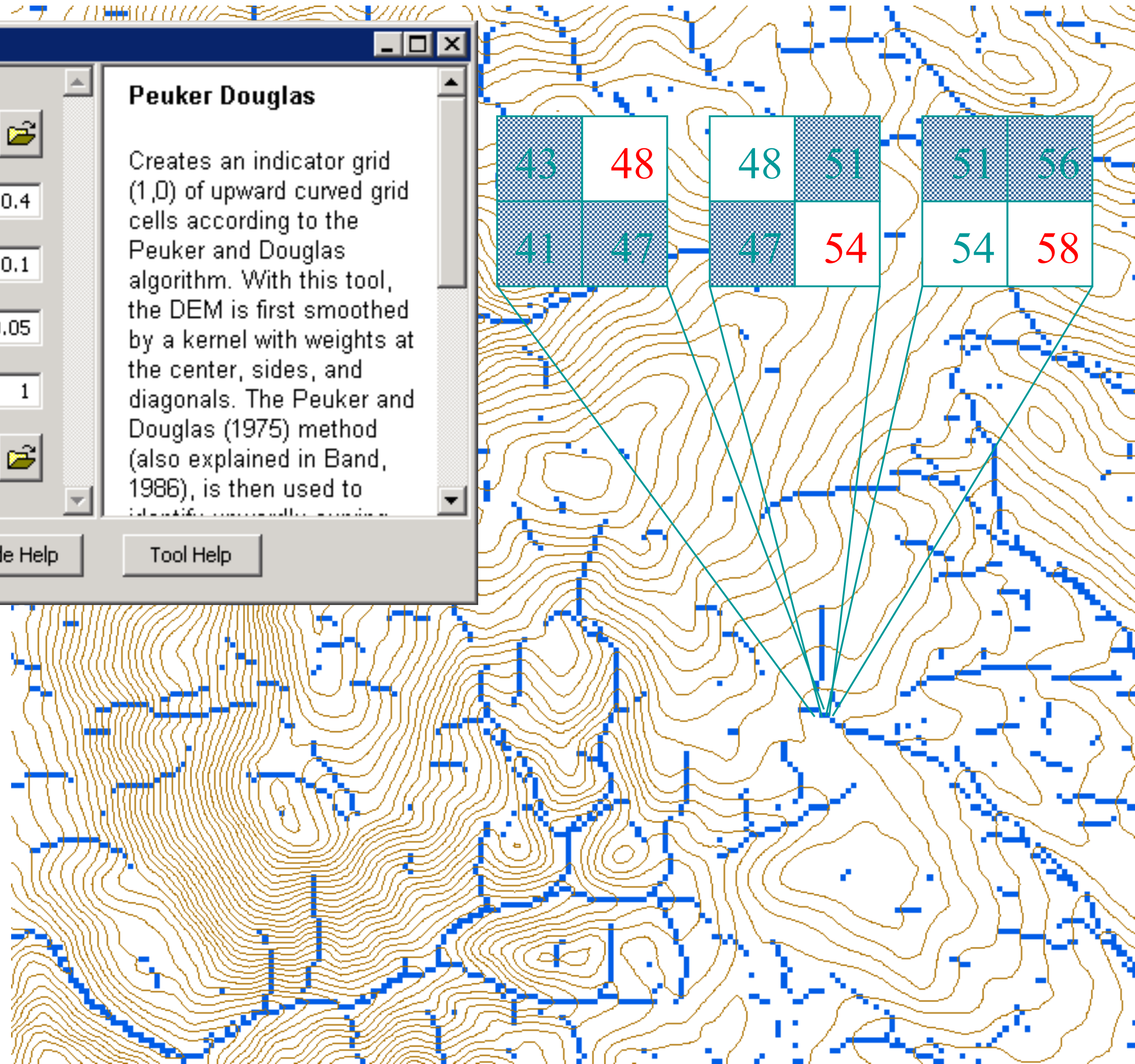
Channels and watersheds with spatially variable drainage density using Peuker Douglas threshold objectively chosen by drop analysis

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

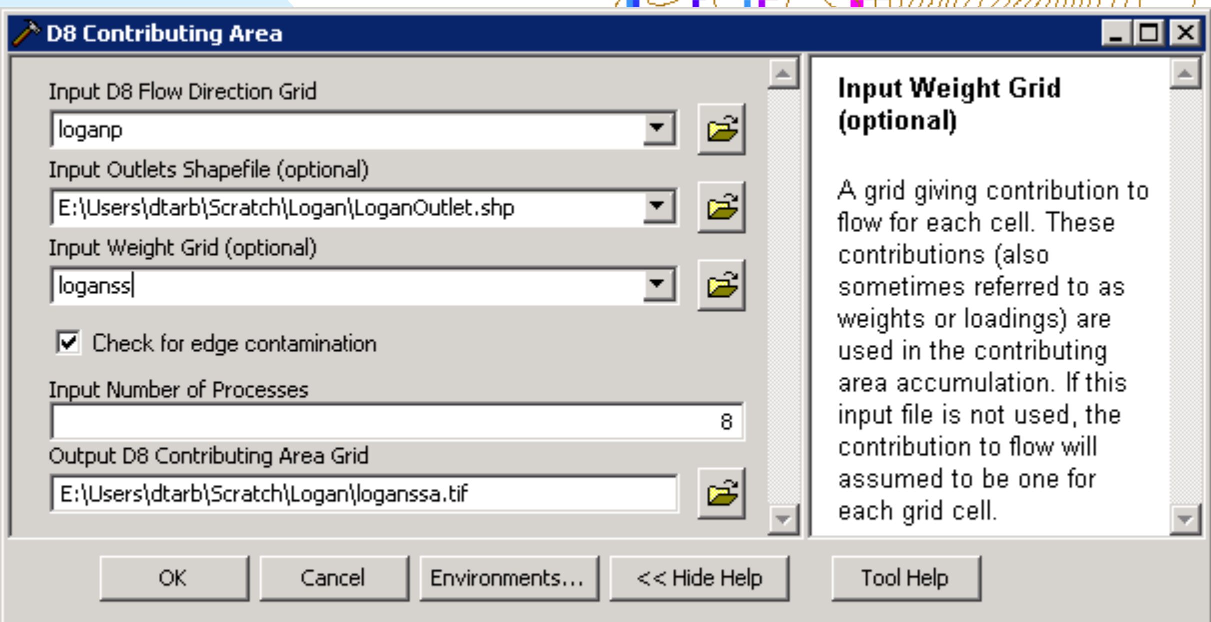
Peuker Douglas

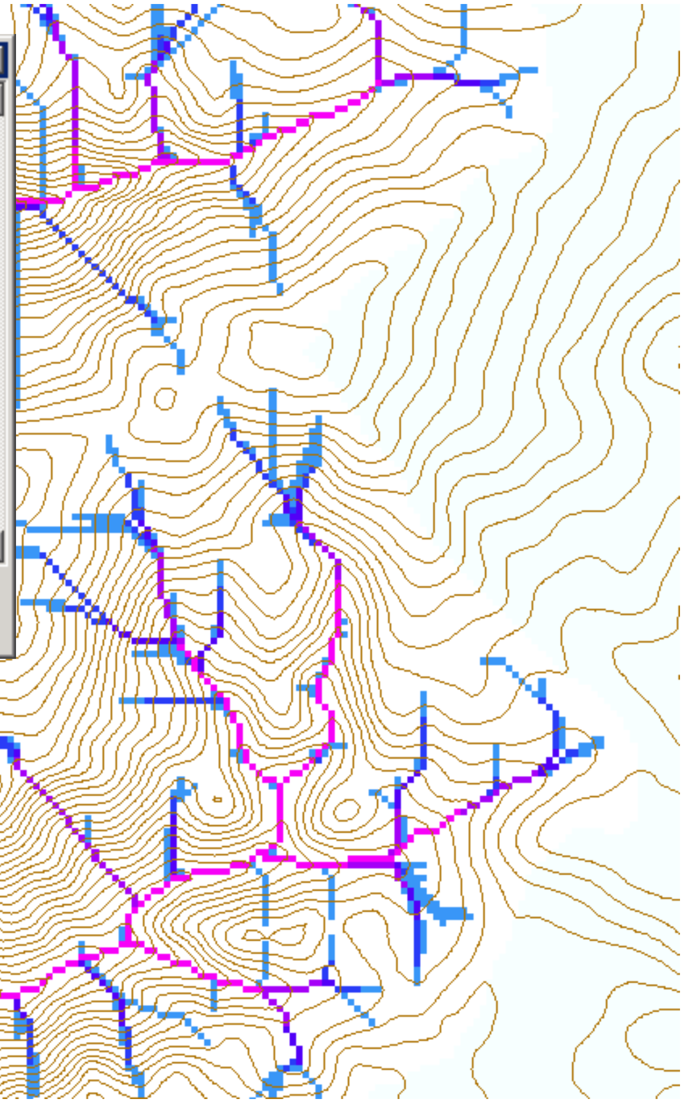




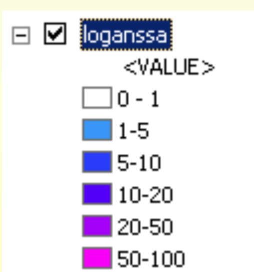
Local Valley Form Computation
(Peuker and Douglas, 1975, Comput. Graphics Image Proc. 4:375)

Weighted D8 Contributing Area

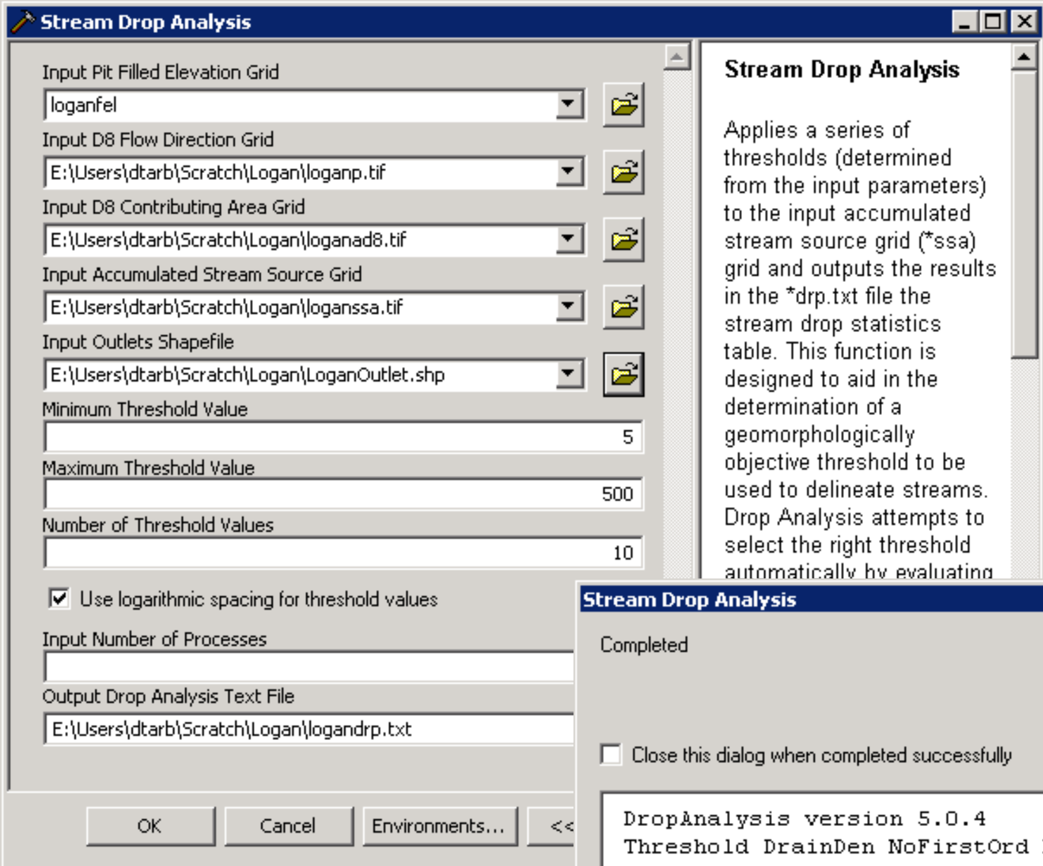


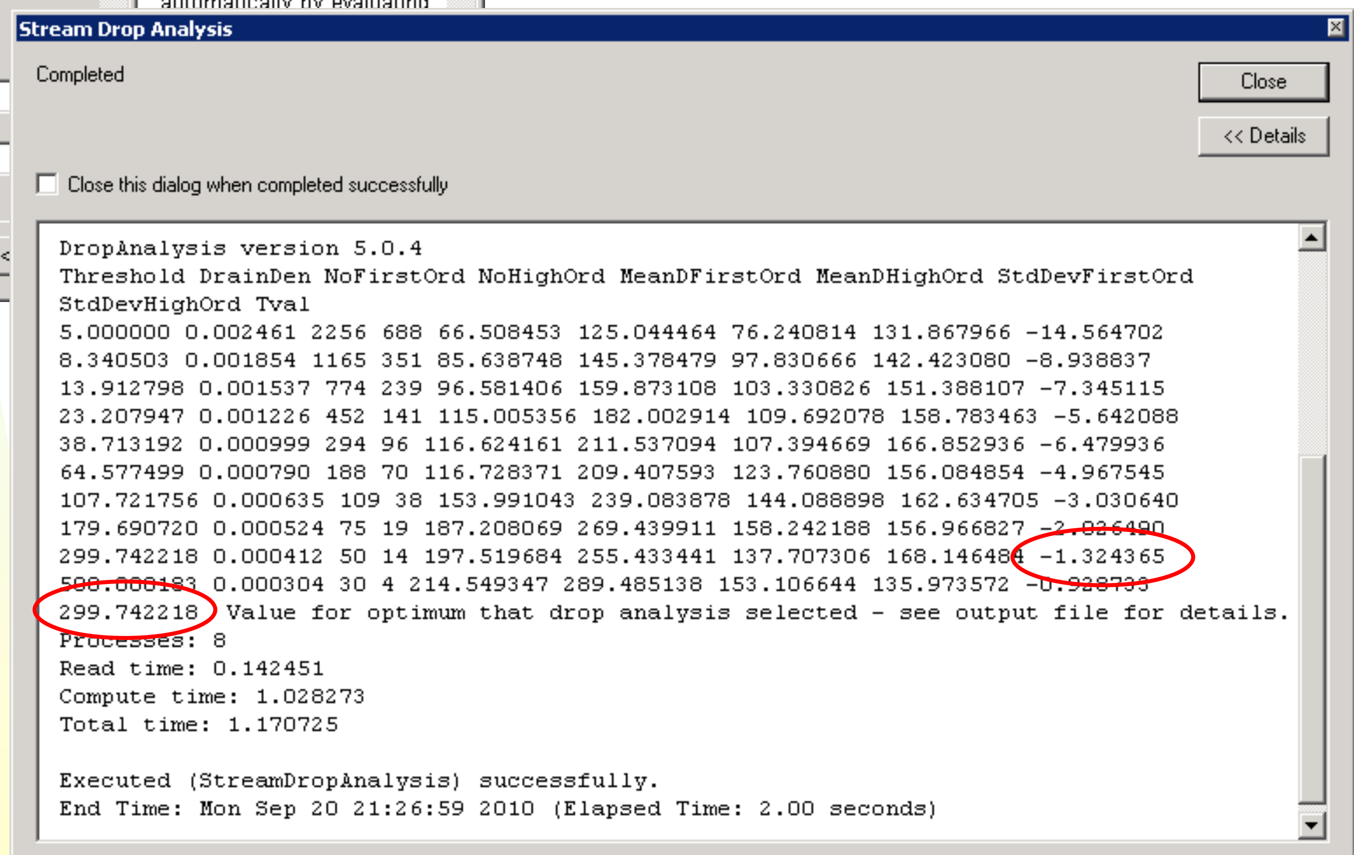


Contributing area only of valley form grid cells upstream of outlet

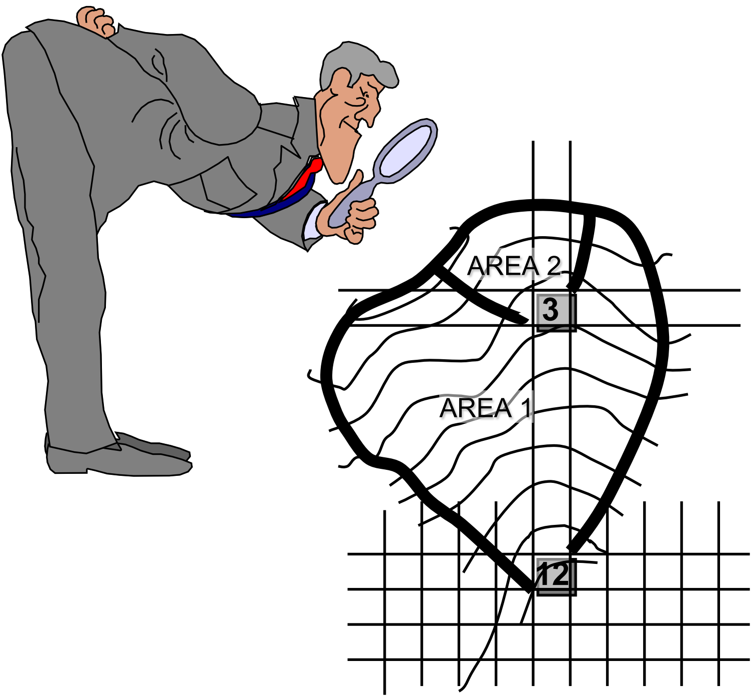


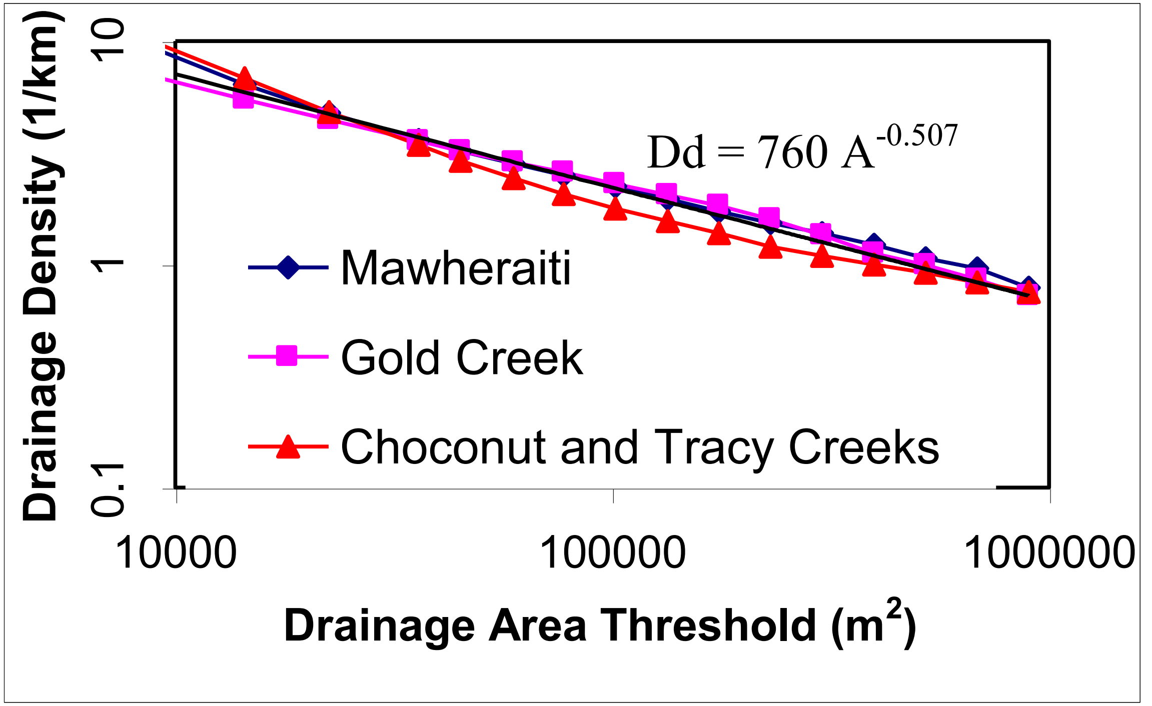
Stream Drop Analysis





How to decide on stream delineation threshold ?

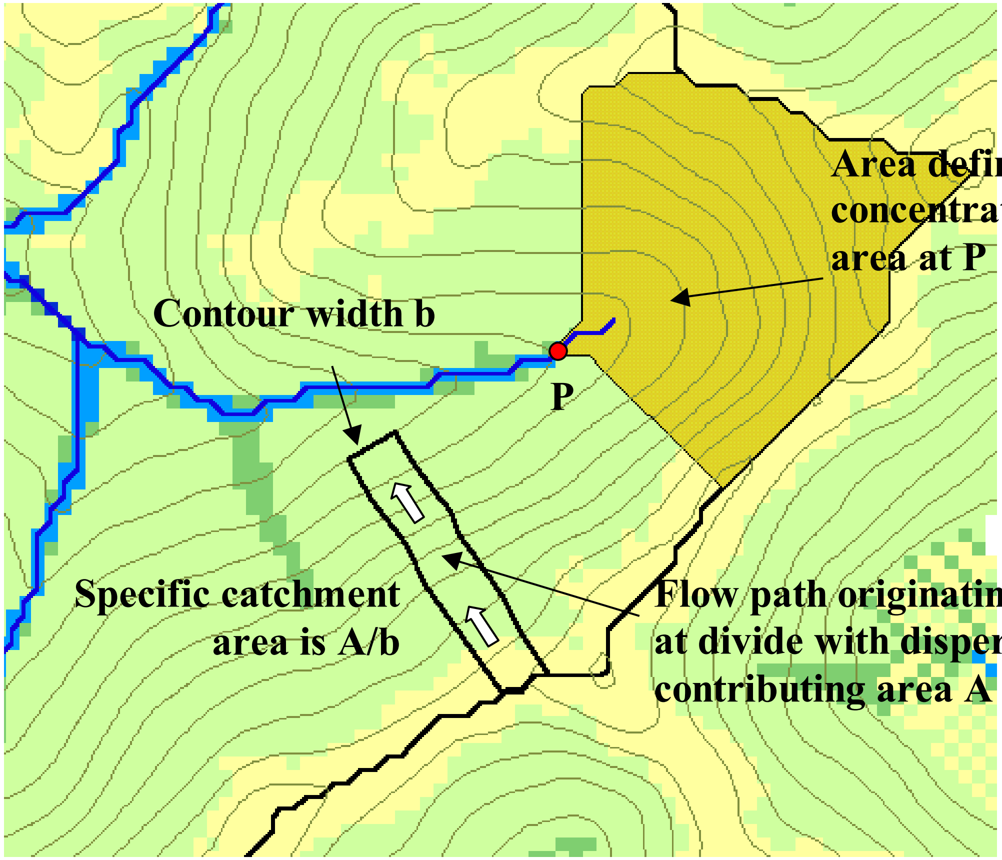




Drainage density (total channel length divided by drainage area) as a function of drainage area support threshold used to define channels for the three study watersheds.

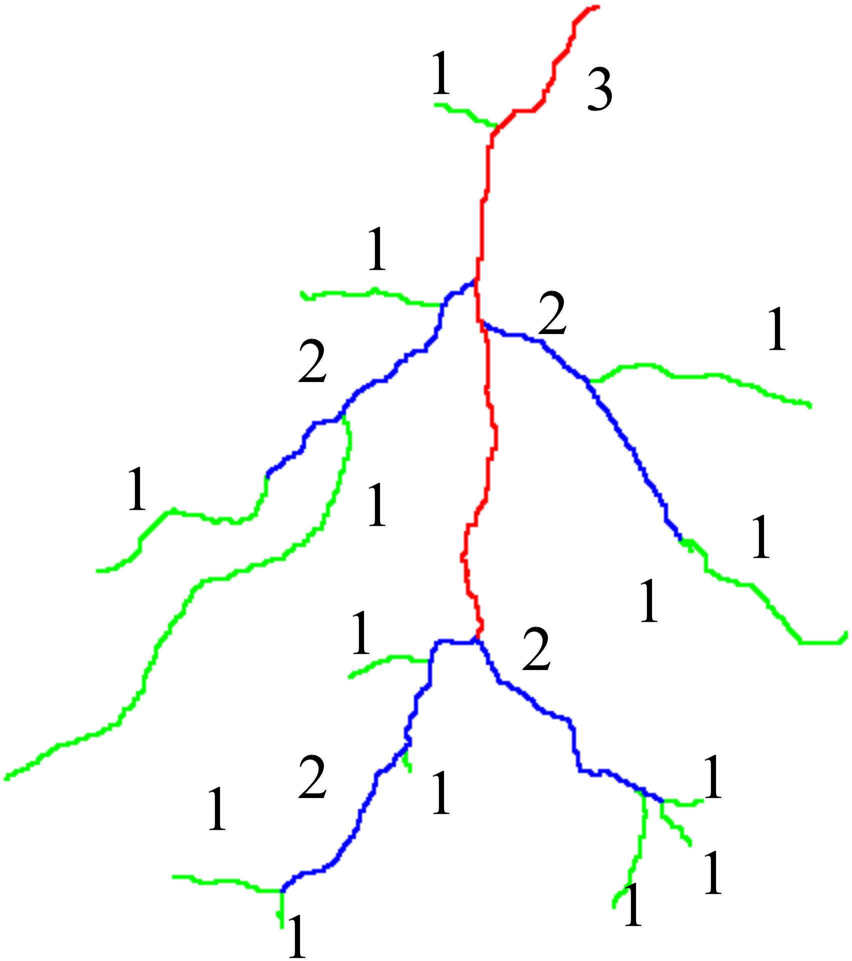
Why is it important?

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

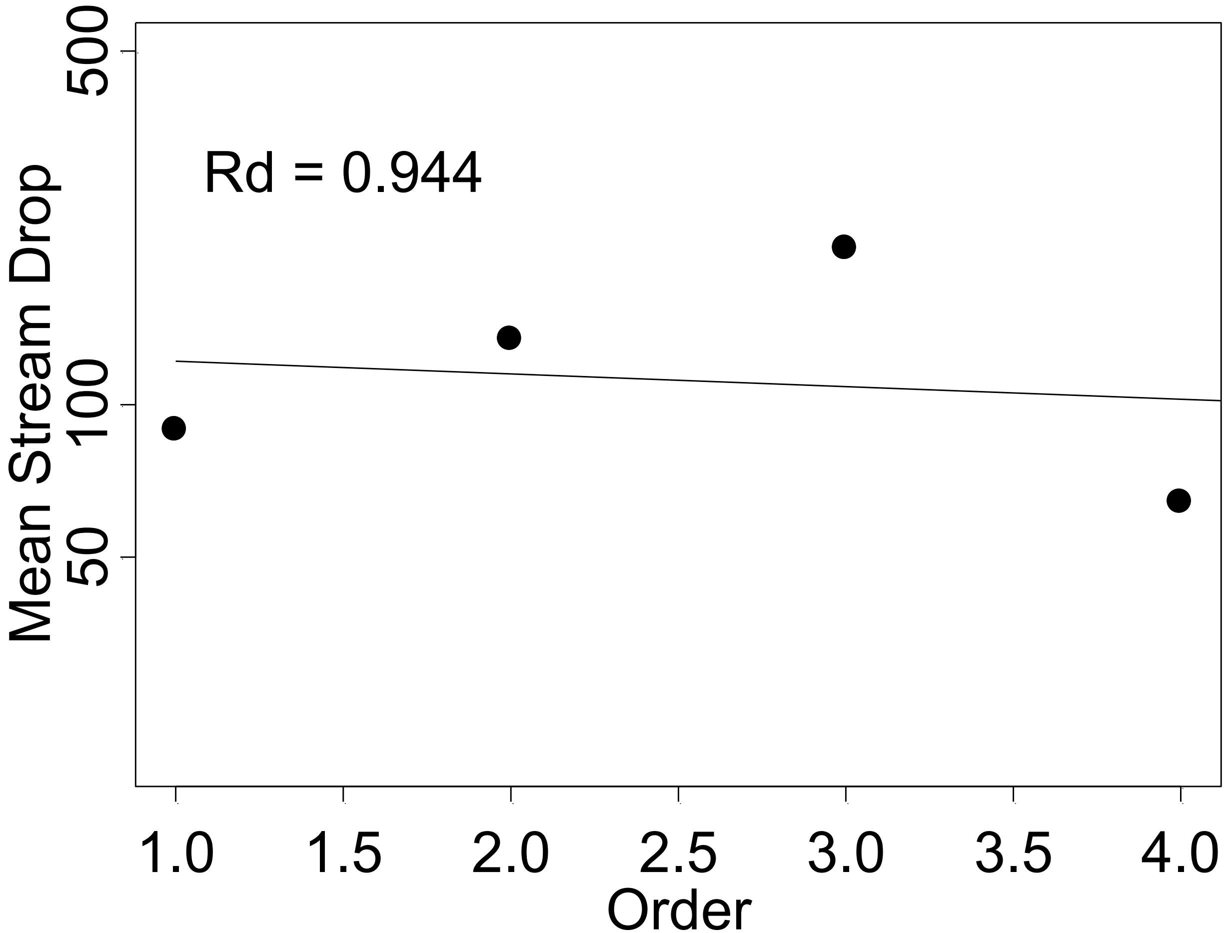


Drainage area can be concentrated or dispersed (specific catchment area) representing concentrated or dispersed flow.

Hortons Laws: Strahler system for stream ordering



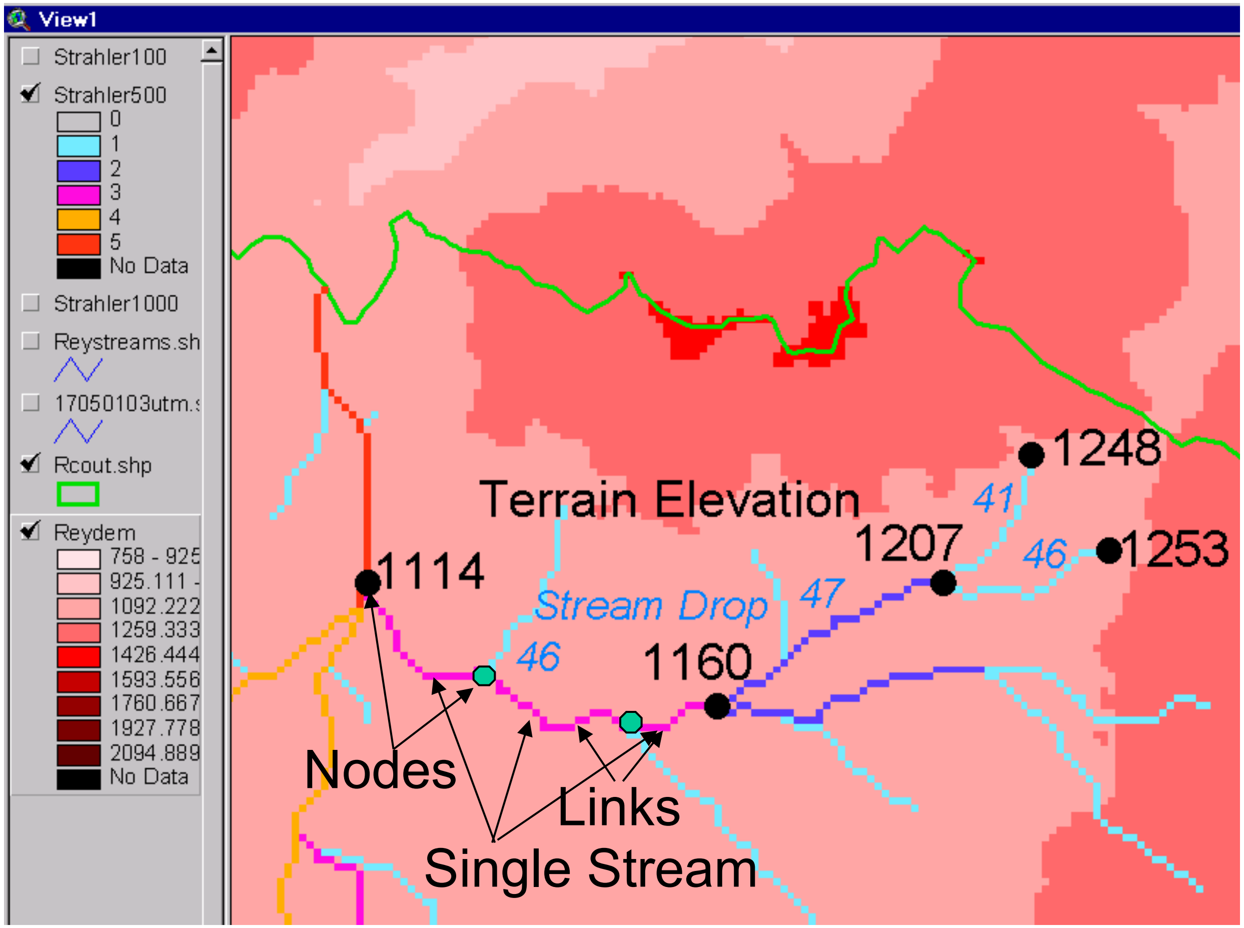
Constant Stream Drops Law



Broscoe, A. J., (1959), "Quantitative analysis of longitudinal stream profiles of small watersheds," Office of Naval Research, Project NR 389-042, Technical Report No. 18, Department of Geology, Columbia University, New York.

Stream Drop

Elevation difference between ends of stream

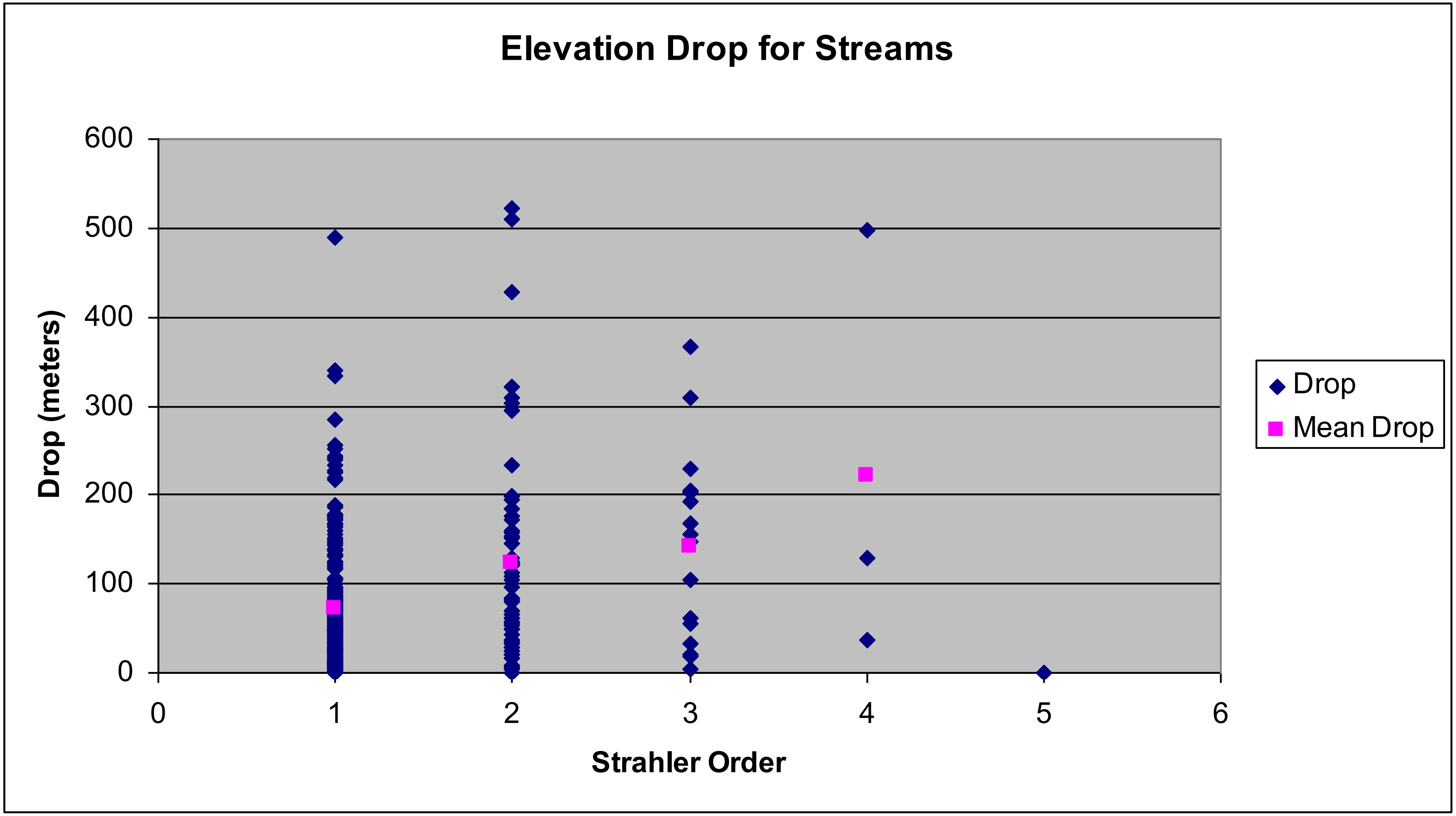


Note that a "Strahler stream" comprises a sequence of links (reaches or segments) of the same order

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

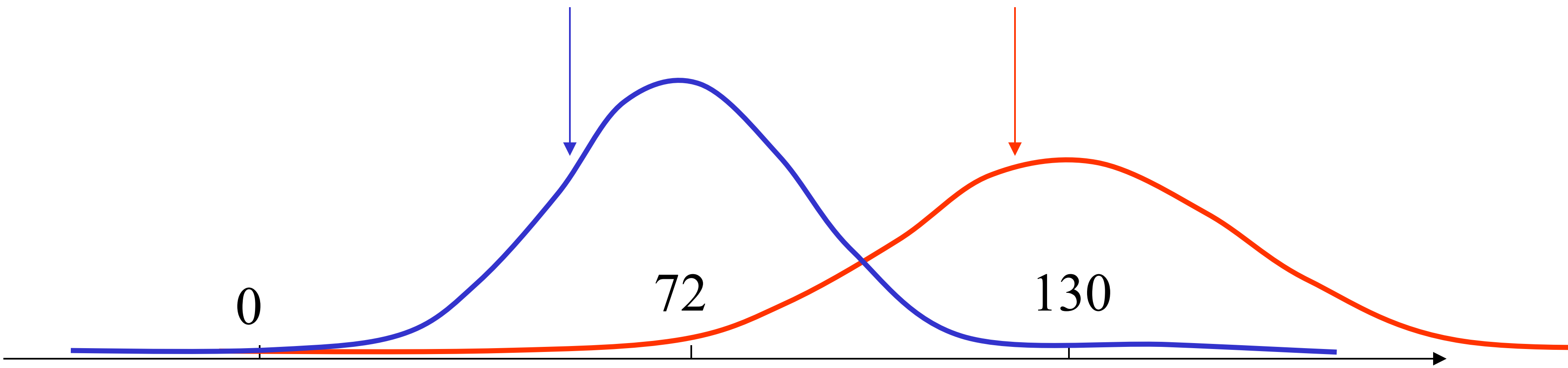
Statistical Analysis of Stream Drops



T-Test for Difference in Mean Values

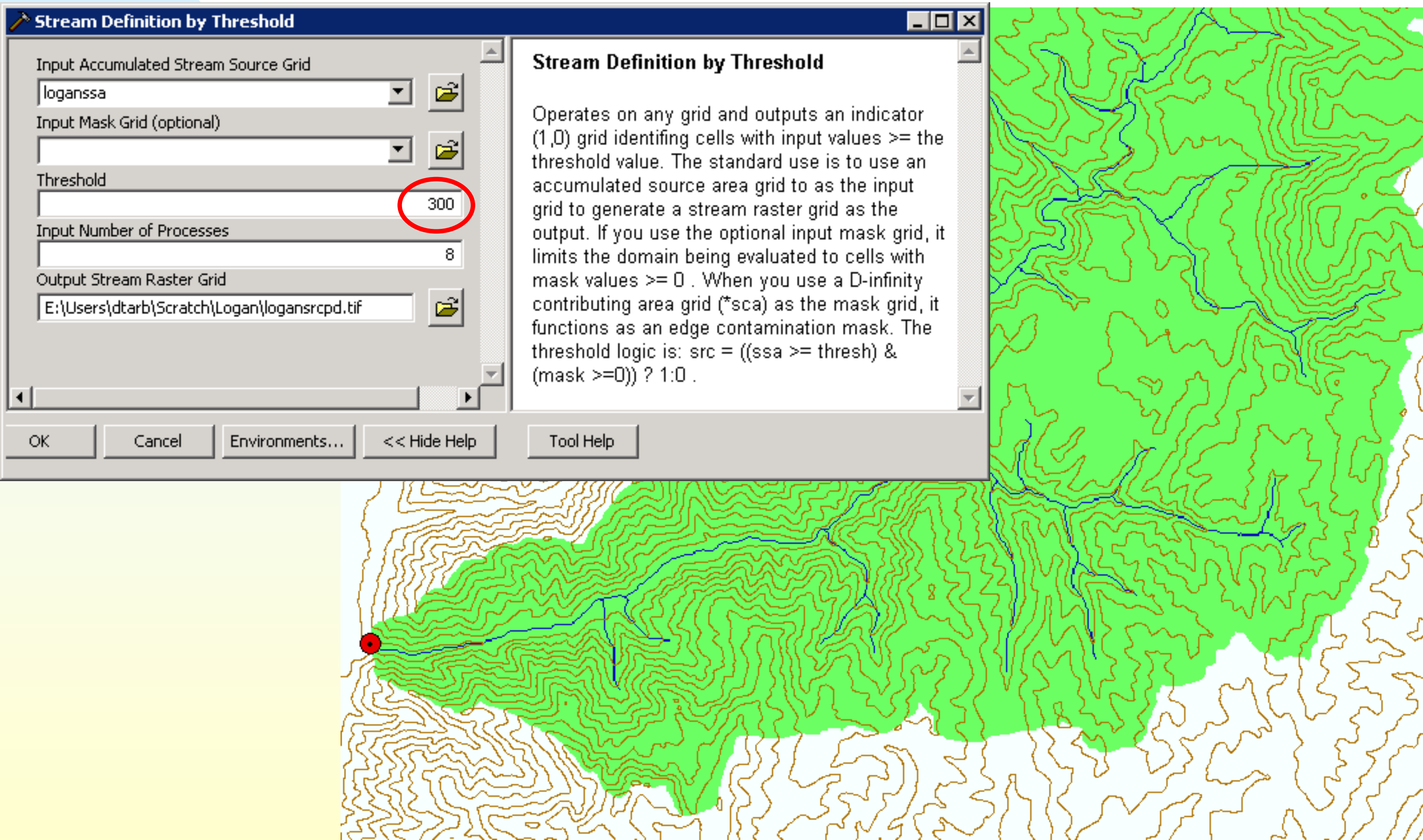
	Order 1
Mean X	72.2
Std X	68.8
Var X	4740.0
Nx	268

	Order 2-4
Mean Y	130.3
Std Y	120.8
Var Y	14594.5
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 Reach and Watershed

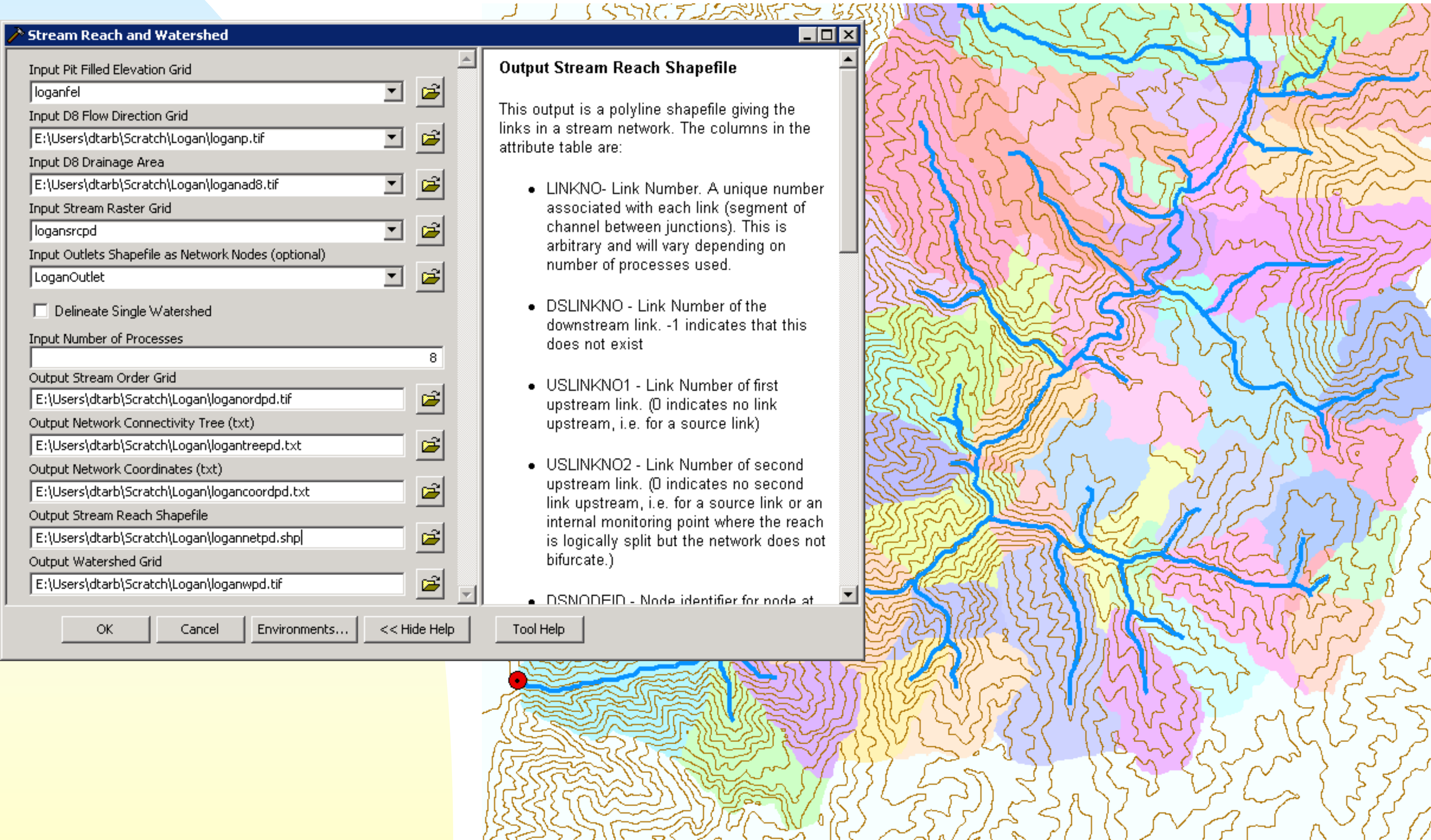
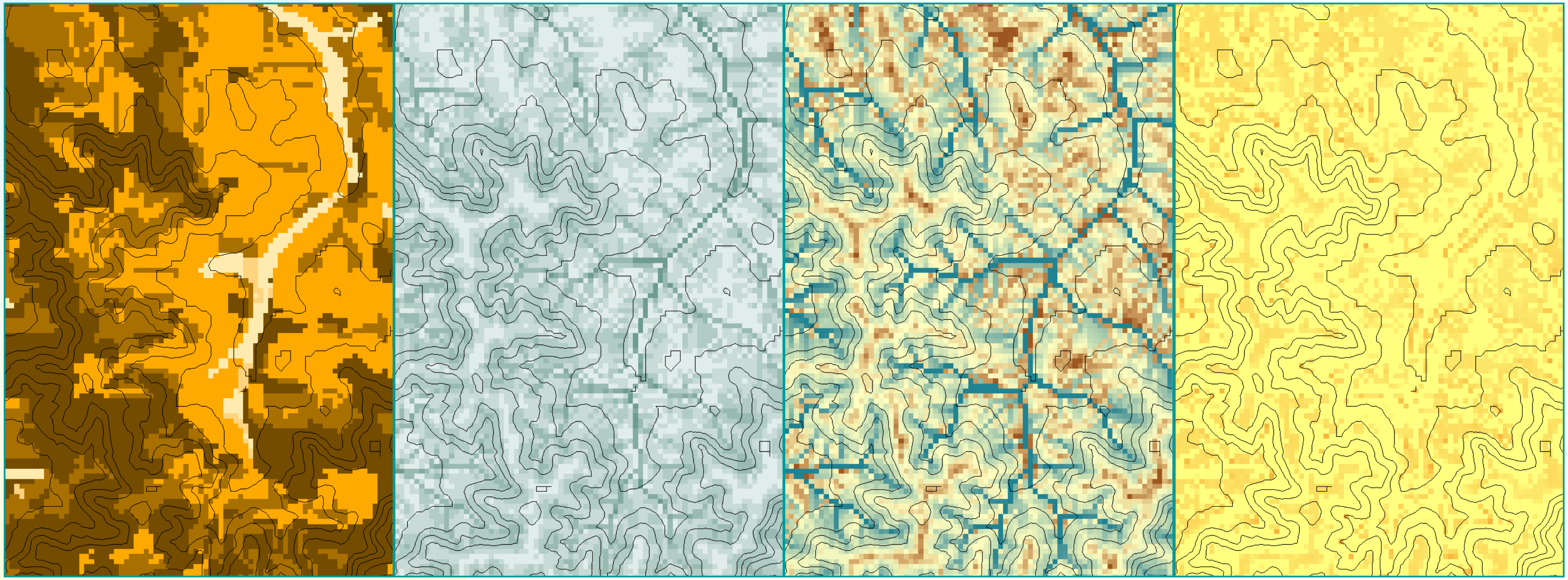


Illustration of some other functions

Transport limited accumulation

Supply Capacity Transport Deposition



$$S \qquad T_{cap} = \chi a^2 \tan(b)^2 \qquad T_{out} = \min\{S + \sum T_{in}, T_{cap}\} \qquad D = S + \sum T_{in} - T_{out}$$

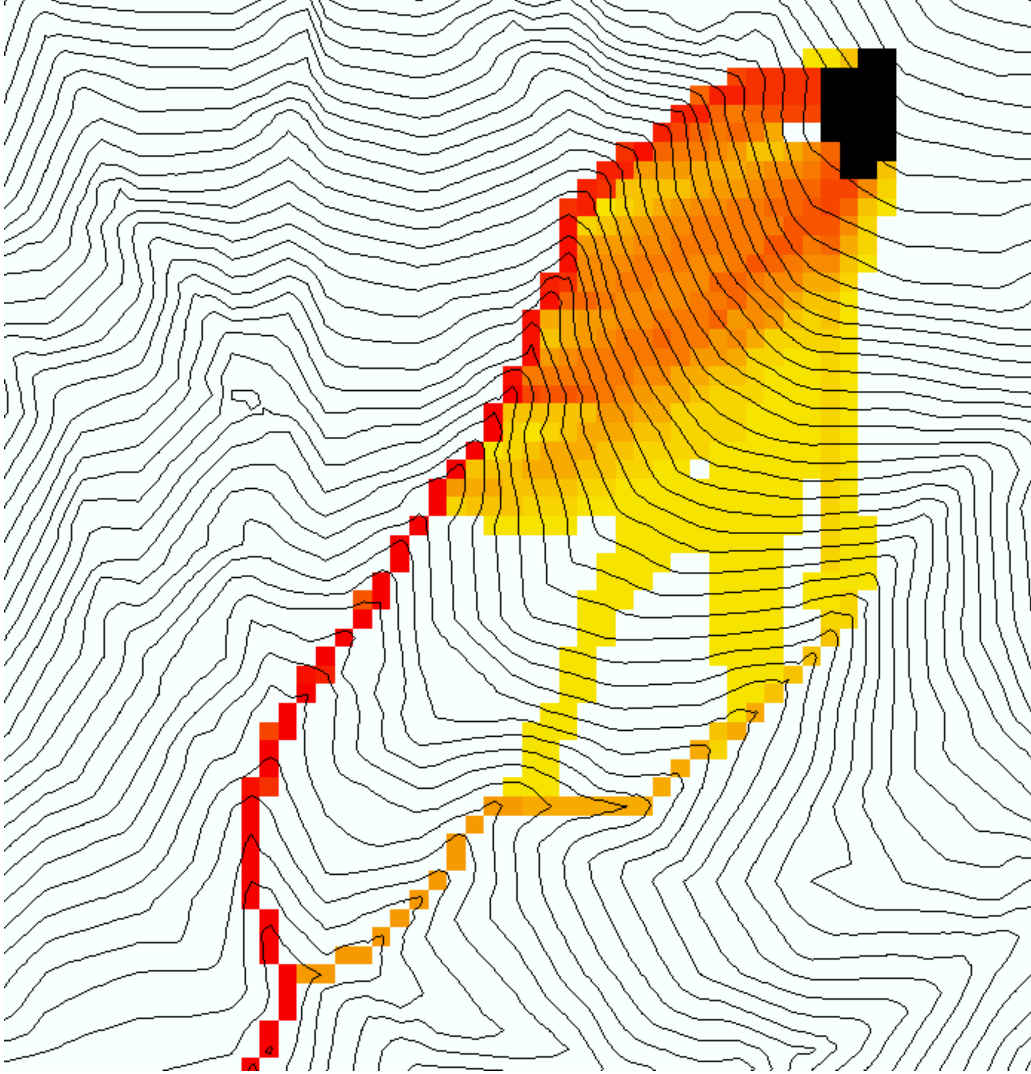
Useful for modeling erosion and sediment delivery, the spatial dependence of sediment delivery ratio and contaminant that adheres to sediment

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

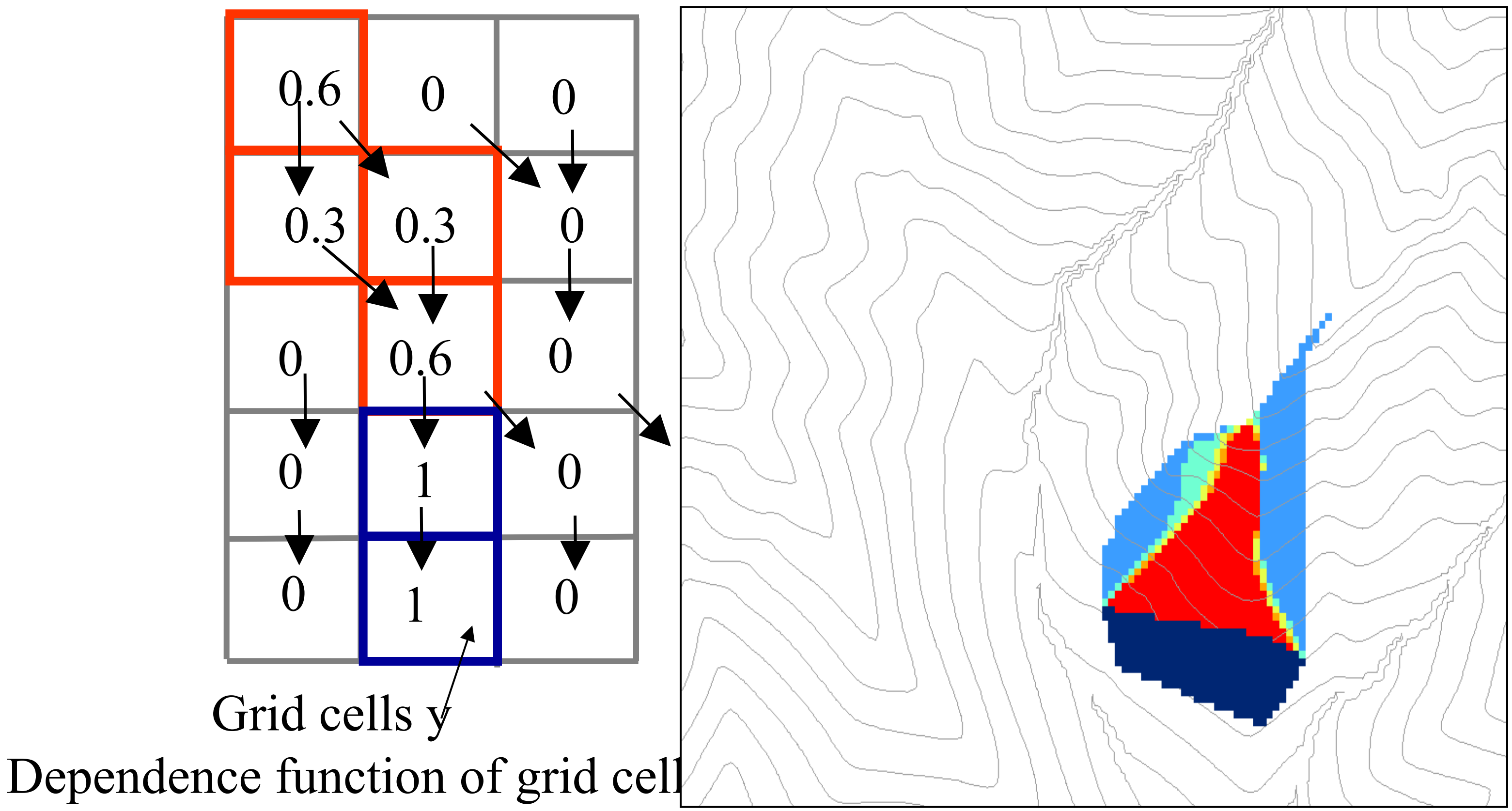
$$DA[m(x)] = DA(i,j) = m(i,j)\Delta^2 + \sum_{k \text{ contributing neighbors}} p_k d(i_k, j_k) DA(i_k, j_k)$$

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 λ is a first order decay parameter.



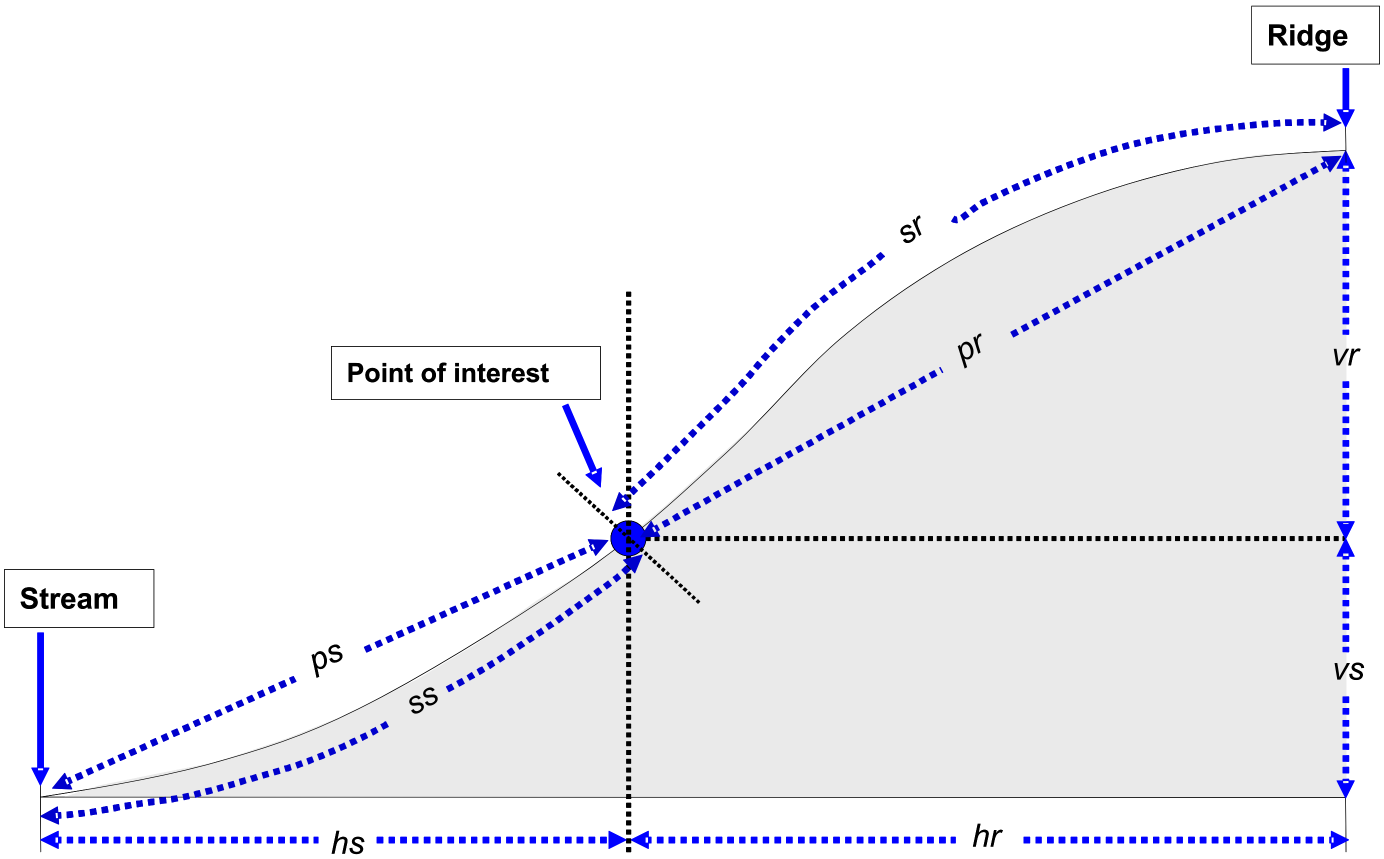
Useful for a tracking contaminant or compound subject to decay or attenuation

Dependence function. Quantifies the amount a point x contributes to the point or zone y.



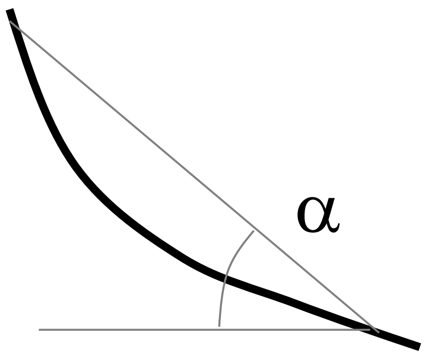
Useful for example to track where a contaminant may come from

Distance Down and Distance Up



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

Avalanche Runout



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