Flow Modeling on Massive Terrains

Laura Toma

Duke University

Flow Modeling

★ Flow direction

The direction water flows at a point in the terrain.

★ Flow accumulation value

Total amount of water which flows through a point in the terrain.

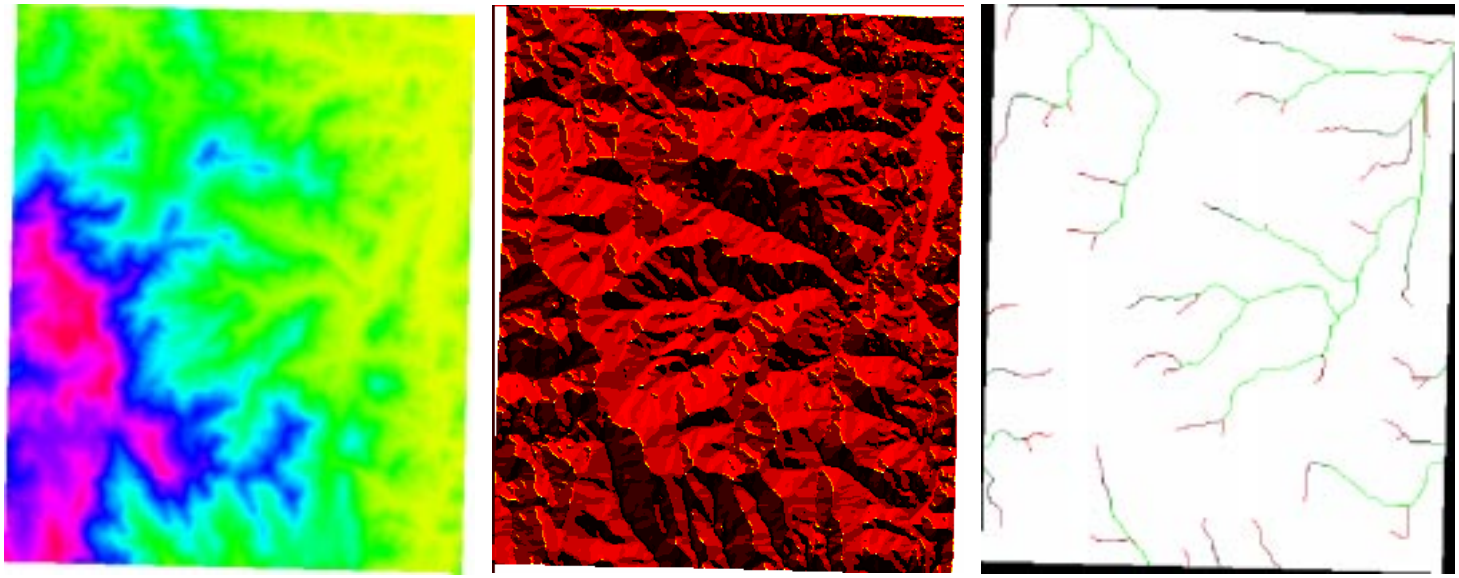
Objective

★ Flow routing: Compute flow directions for **all points in the terrain**.

★ Flow accumulation: Compute flow accumulation values for **all points in the terrain**.

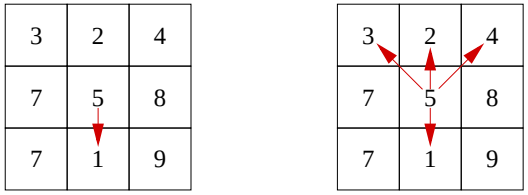
Applications

- ☆ Watersheds, drainage network
- ☆ Erosion, infiltration, drainage, solar radiation distribution, sediment transport, vegetation structure, species diversity



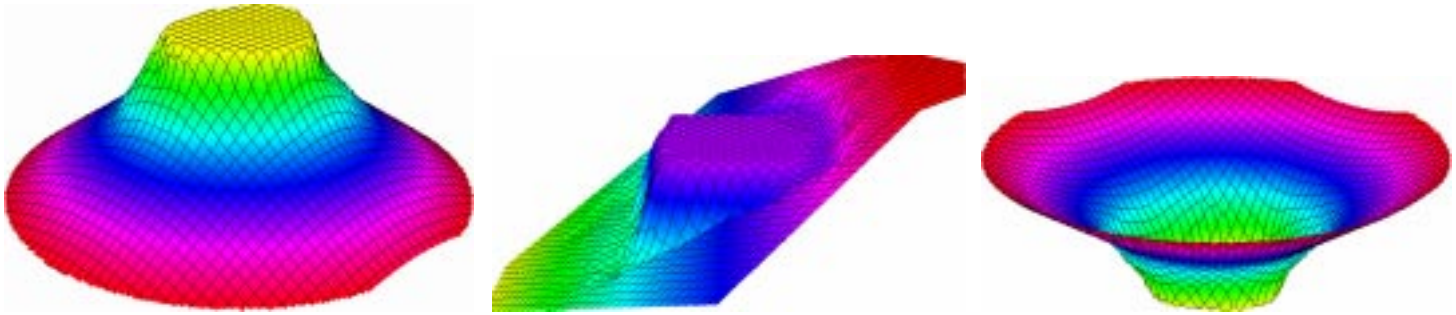
Flow Routing

☆ Water flows downhill.



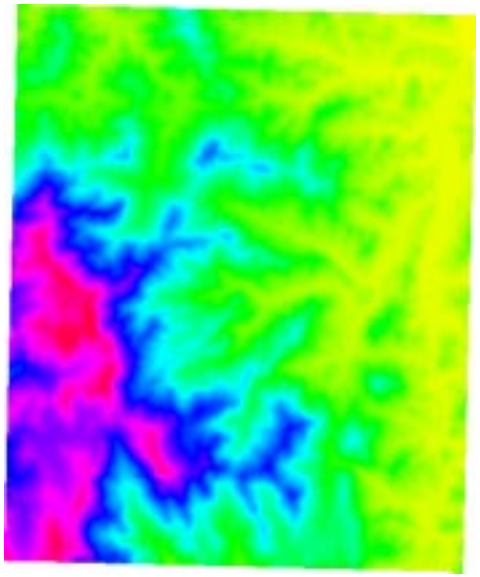
☆ Compute flow directions by inspecting 8 neighbor cells.

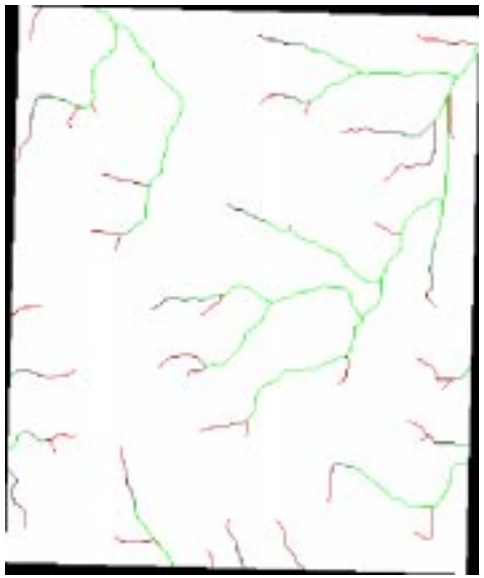
☆ Flat areas: plateaus and sinks.



Flow Accumulation

- ☆ Compute the total amount of flow through each grid point
  - Initially one unit of water on each grid point
  - Every point distributes water to the neighbors pointed to by its flow direction(s)





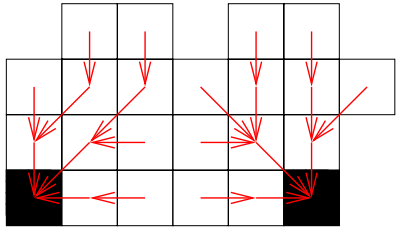
Scalability to Massive Data

- ☆ Massive remote sensing data available
  - USGS: entire US at 10m resolution; 3m and 1m resolution available
  - NASA's Shuttle Radar Topography Mission: collect data for 80% of earth's land mass (10 terabytes)
  - LIDAR
- ☆ Existing software
  - ArcInfo: cannot process files $> 2GB$
  - GRASS, TARDEM: run for weeks..

I/O-Efficient Flow Routing

☆ Flow routing

- Every cell has flow direction
- Flow directions do not induce cycles
- Every cell has a flow path to the edge of the terrain

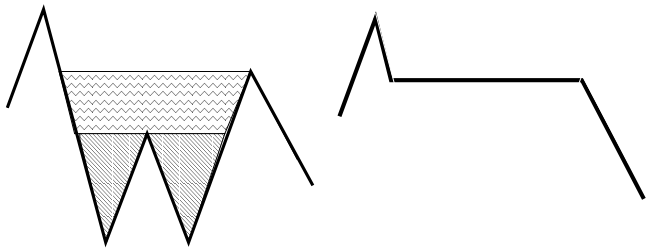


☆ Plateaus

☆ Sinks

- Flooding: Fill the terrain up to the steady state level reached when an infinite amount of water is poured onto the terrain and the outside is viewed as a giant ocean.

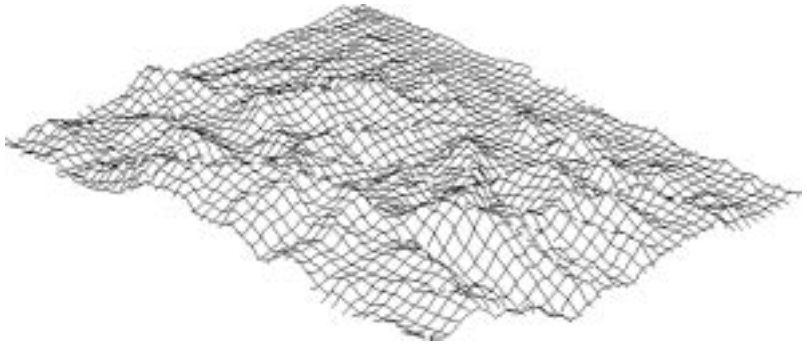
Flooding

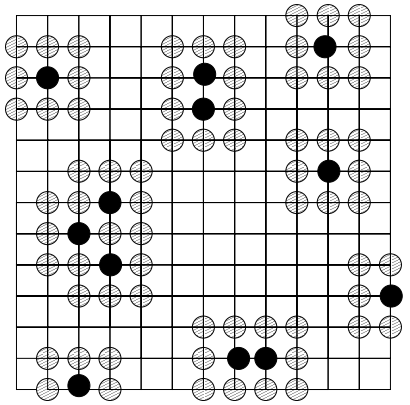


- ☆ Previous work: Jenson & Domingue '88
  - Watershed: part of the terrain that flows into the sink.
  - Partition the terrain into watersheds $\longrightarrow$ watershed graph
  - Identify and collapse cycles in the watershed graph
  - W= number of watersheds: $O(W^2)$ time, $O(W^2)$ I/Os
- ☆ I/O-efficient flooding: $O(W \cdot \alpha(W, N))$ time and I/Os

Flow Accumulation–Internal memory algorithm

☆ Process (sweep) points in decreasing order of heights, distributing flow to neighbors
one sweep enough $\Longrightarrow$ $O(N \log N)$ time





☆ Problem: algorithm uses $O(N)$ I/Os if directions and flow stored as grids (not fitting in memory)

- Points with same height are distributed over the terrain
  $\Longrightarrow$ scattered accesses

I/O-Efficient Flow Accumulation

- ☆ Eliminate scattered accesses to flow grid
  - Idea: neighbor only needs the distributed flow when the sweep plane reaches its elevation
  - Use a $O(\frac{1}{B} \log_{M/B} \frac{N}{B})$ priority queue [A95, BK98]
    - Distribute flow by inserting it in priority queue with priority equal to neighbor's height (and grid position as secondary key)
    - Augment each height with heights of neighbors (trade space for I/Os)
  - $O(N)$ priority queue operations $\Rightarrow$ $O(\frac{N}{B} \log_{M/B} \frac{N}{B})$ I/Os

TerraFlow

http://www.cs.duke.edu/geo*/terraflow/

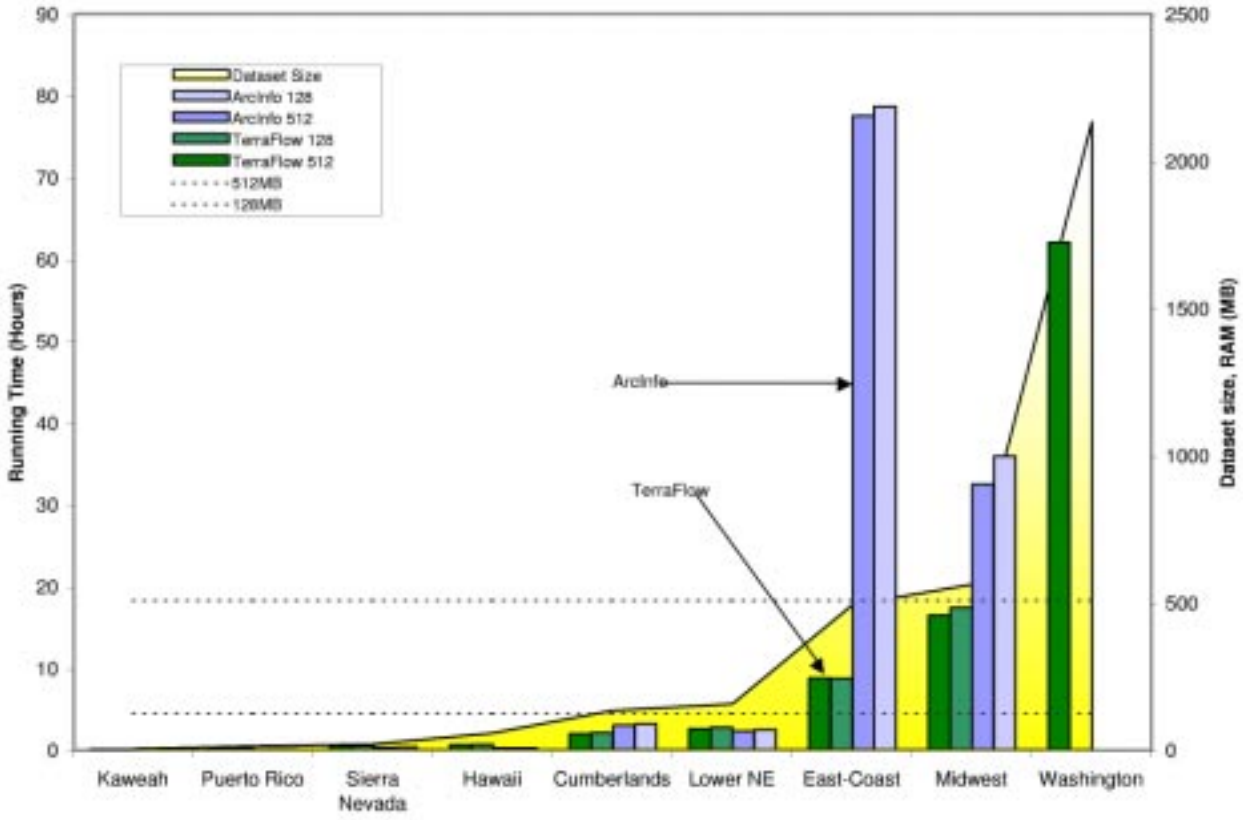
Collection of programs for flow routing and flow accumulation on massive grids.

- ⋆ Efficient
  - 2-1000 times faster on massive grids than existing software
- ⋆ Scalable
  - 1 billion elements! (> 2GB)
- ⋆ Flexible
  - different flow models

Experimental Results: Datasets

Dataset	Resolution	Dimensions	Grid Size
Kaweah	30m	1163 x 1424	3.2MB
Puerto Rico	100m	4452 x 1378	12MB
Sierra Nevada	30m	3750 x 2672	19MB
Hawaii	100m	6784 x 4369	56MB
Cumberlands	80m	8704 x 7673	133MB
Lower New England	80m	9148 x 8509	156MB
Central Appalachians	30m	12042 x10136	232MB
East-Coast USA	100m	13500 x 18200	491MB
Midwest USA	100m	11000 x 25500	561MB
Washington State	10m	33454 x 31866	2GB

TerraFlow Performance



TerraFlow Performance

- ☆ Significant speedup over ArcInfo for large grids
  - East-Coast dataset
    - TerraFlow: 8.7 hours
    - ArcInfo: 78 hours
  - Washington state dataset
    - TerraFlow: 63 hours
    - ArcInfo: cannot process it!
- ☆ Other software
  - GRASS: killed after 17 days on Hawaii
  - TARDEM: Can handle Hawaii. Killed after 20 days on Cumberlands (CPU utilization 5%, 3GB swap file)

Future Directions

- ☆ Flow modeling on TINs
  - Flow along edges. Compute flow accumulation of nodes.
  - Extend grid approach: assign flow at triangle level. Flow across edges and along channel edges. Compute flow accumulation of triangles and channel nodes.
  - Compute contributing area directly: trace steepest downslope paths across triangles.
- ☆ Grid/TIN conversion
  - Maintain global features