MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

Morten Revsbæk
University of Aarhus

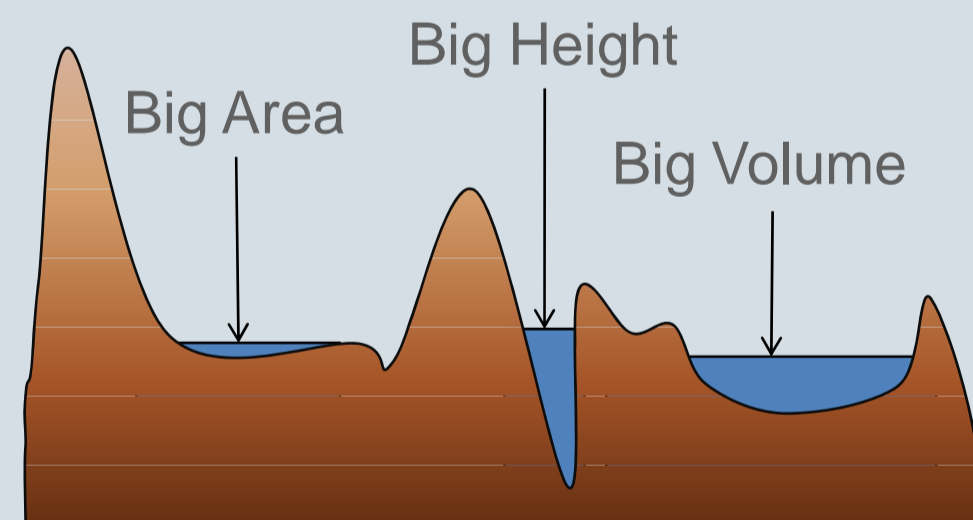# TerraSTREAM: Hydrological Conditioning

Motivation

- **Problem:** Detailed data set → Many small, insignificant sinks.
- **Flow Routing Consequence:** Disconnected river network.
- **Contour Line Consequence:** Many small and insignificant contours.
- **Solution:** Find insignificant sinks and remove them.

Significant Sink
Insignificant Sink
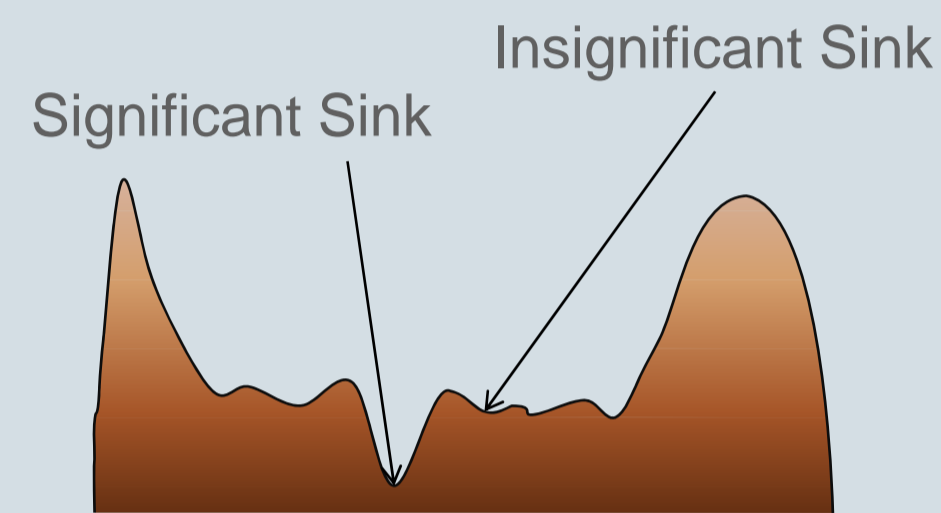
Solution

- We associate a numeric value – **persistence value** - with each sink.
  - Persistence value can be the **height, area or volume** of sink.
- Define a **persistence threshold** and remove all sinks with a persistence value lower than the threshold.


Big Area
Big Height
Big Volume

I/O-Efficient Algorithm

- Hydrological conditioning using height persistence can be done I/O-efficiently using I/O-efficient batched union-find as proposed by Agarwal, Arge and Yi in 2006 [SoCG'06]
- We defined and solved the more general batched union-find with dynamic set properties so that both height, area and volume persistence (or any combination of these) can be calculated I/O-efficiently.

## Hydrological Conditioning and Flow Routing

Calculating Persistence

- The graph to the right shows both the height and the volume of each sink (**red cross**) in the terrain depicted below.
- **The Vertical line** represents the height threshold that removes all but the three highest sinks.
- **The Horizontal line** represents the volume threshold that removes all but the three largest sinks.

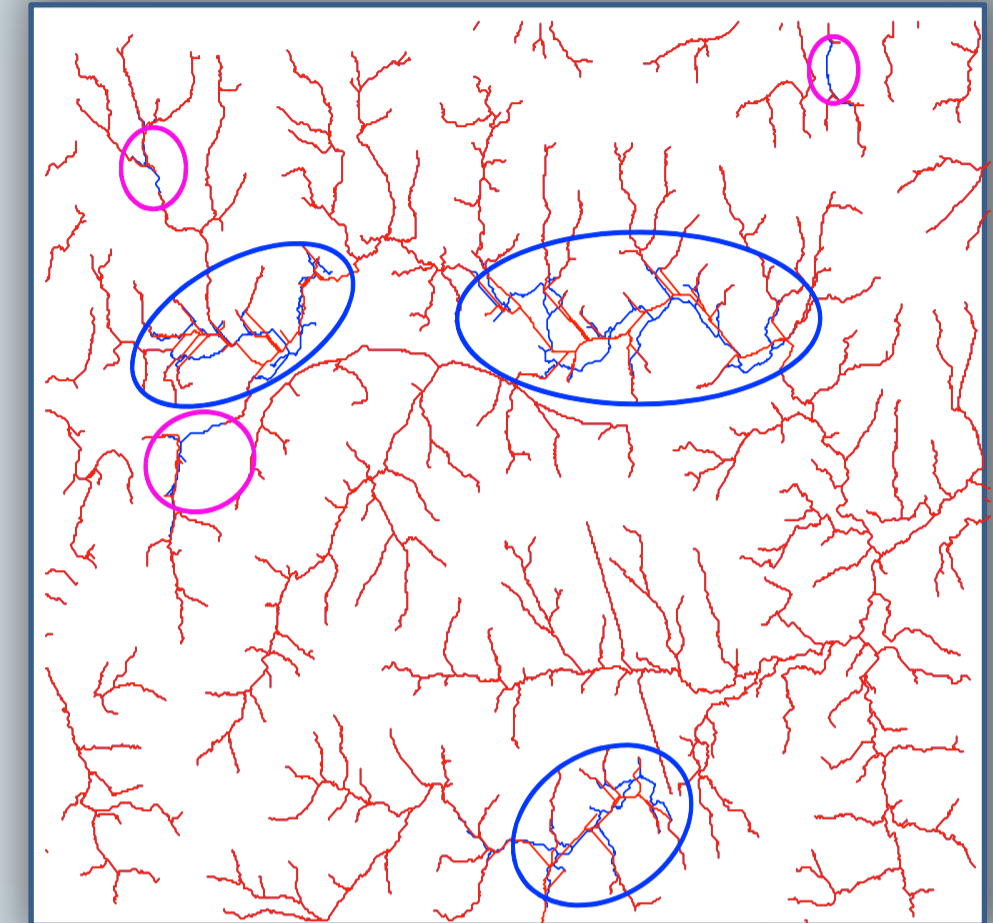Height persistence vs. Volume persistence
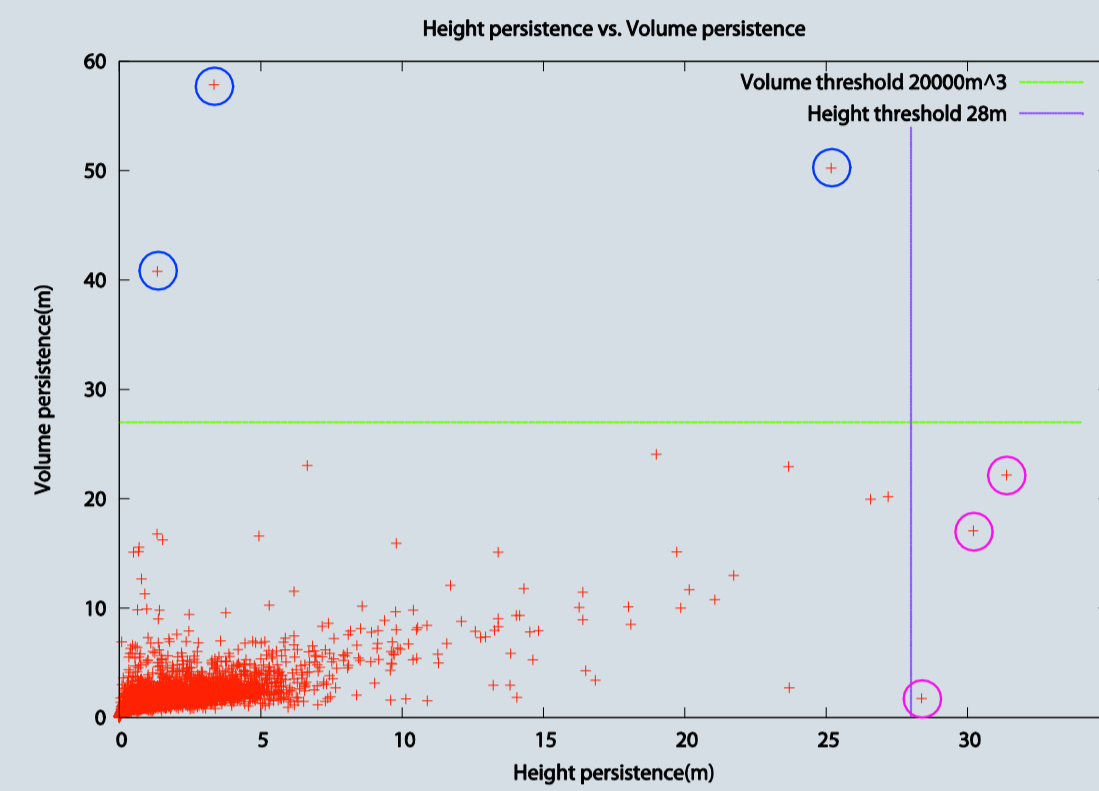
Removing Sinks

- The **top right terrain** has been conditioned using height persistence.
- The **bottom right terrain** has been conditioned using volume persistence.
- The thresholds used are the ones given in the graph above, so that in each of the terrains only three sinks are kept after conditioning.
- **Conclusion:** The sinks kept when using volume conditioning seem much more significant and intuitively they correspond to the sinks that water will realistically flow towards.
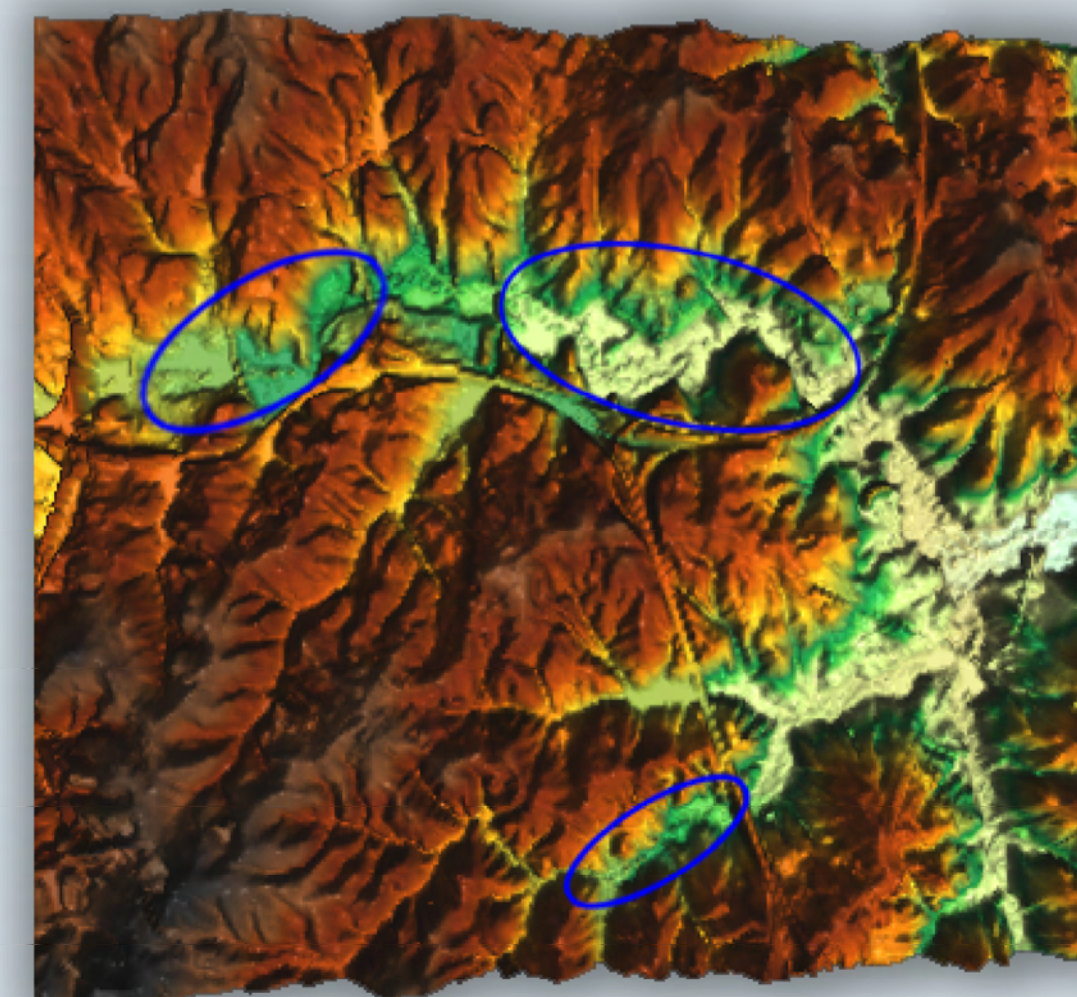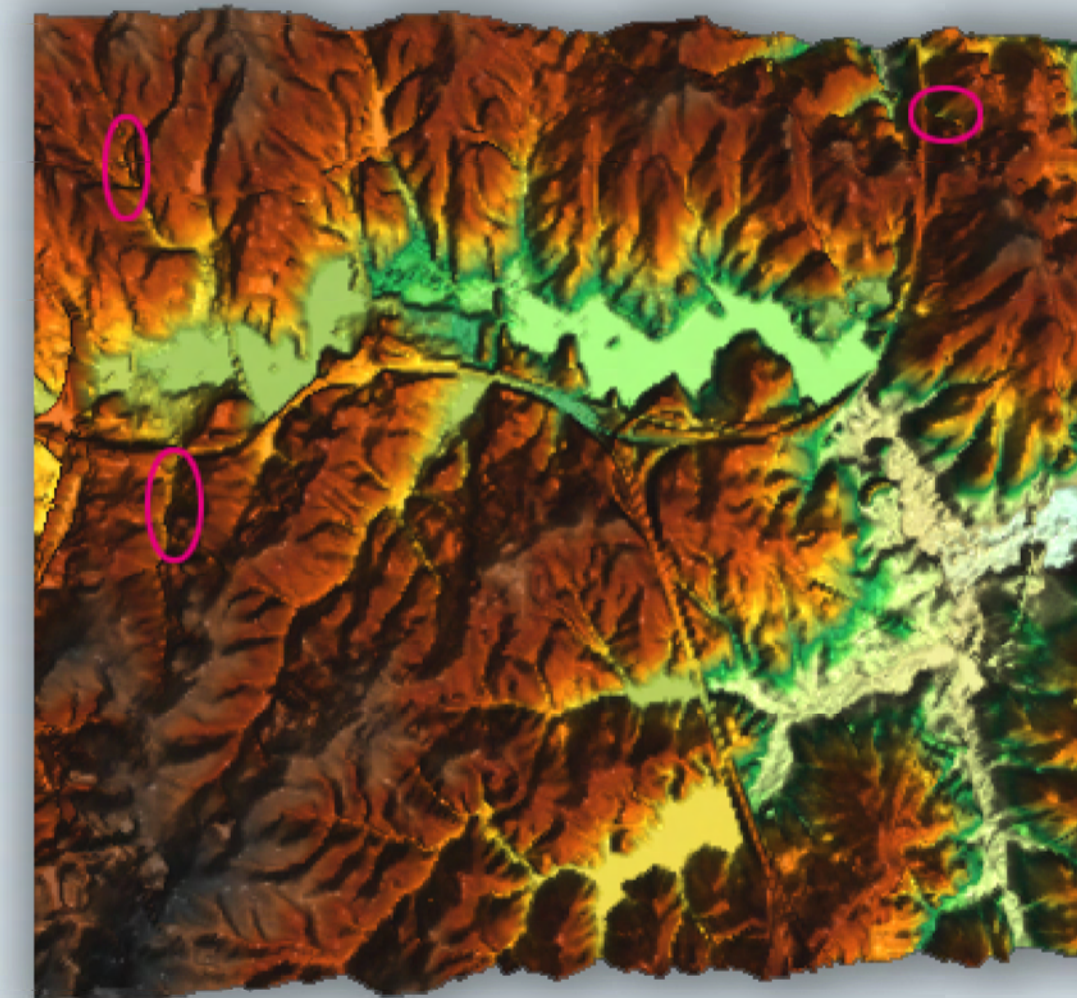
Flow Routing

- The **red river network** is generated from the height conditioned terrain.
- The **blue river network** is generated from the volume conditioned terrain.
- **Conclusion:** Blue network is connected in the purple circles and the flow is more "natural" in the blue circles.

Conditioning and Contour Line Generation

Contour Lines

- We can enhance the output of contour line generation, by removing sinks that result in insignificant contours.
- **Black contour lines** were generated from a terrain conditioned using volume persistence.
- **Red contour lines** were generated from a terrain conditioned using height persistence.
- In the **top figure** the black lines have been drawn on top of the red lines and vice versa in the **bottom figure**.
- **Conclusion:** Volume persistence removes many insignificant contours (top) that are kept when using height persistence, whereas height persistence removes contours that seem significant (bottom).