r.accumulate: Efficient computation of hydrologic parameters in GRASS

February 2, 2020
FOSDEM 2021

Huidae Cho, Ph.D., GISP, PE (MD), M.ASCE, CFM

Institute for Environmental and Spatial Analysis
University of North Georgia

GRASS GIS Core Developer

Me

- Huidae Cho /hidɛ t͡ɕo/
- Teaching geospatial science and computing at the University of North Georgia
- GRASS GIS core developer (20 years)
- ArcGIS developer (12 years)
- Water resources engineer (10 years)
- https://github.com/HuidaeCho
- grass4u@gmail.com

Source code

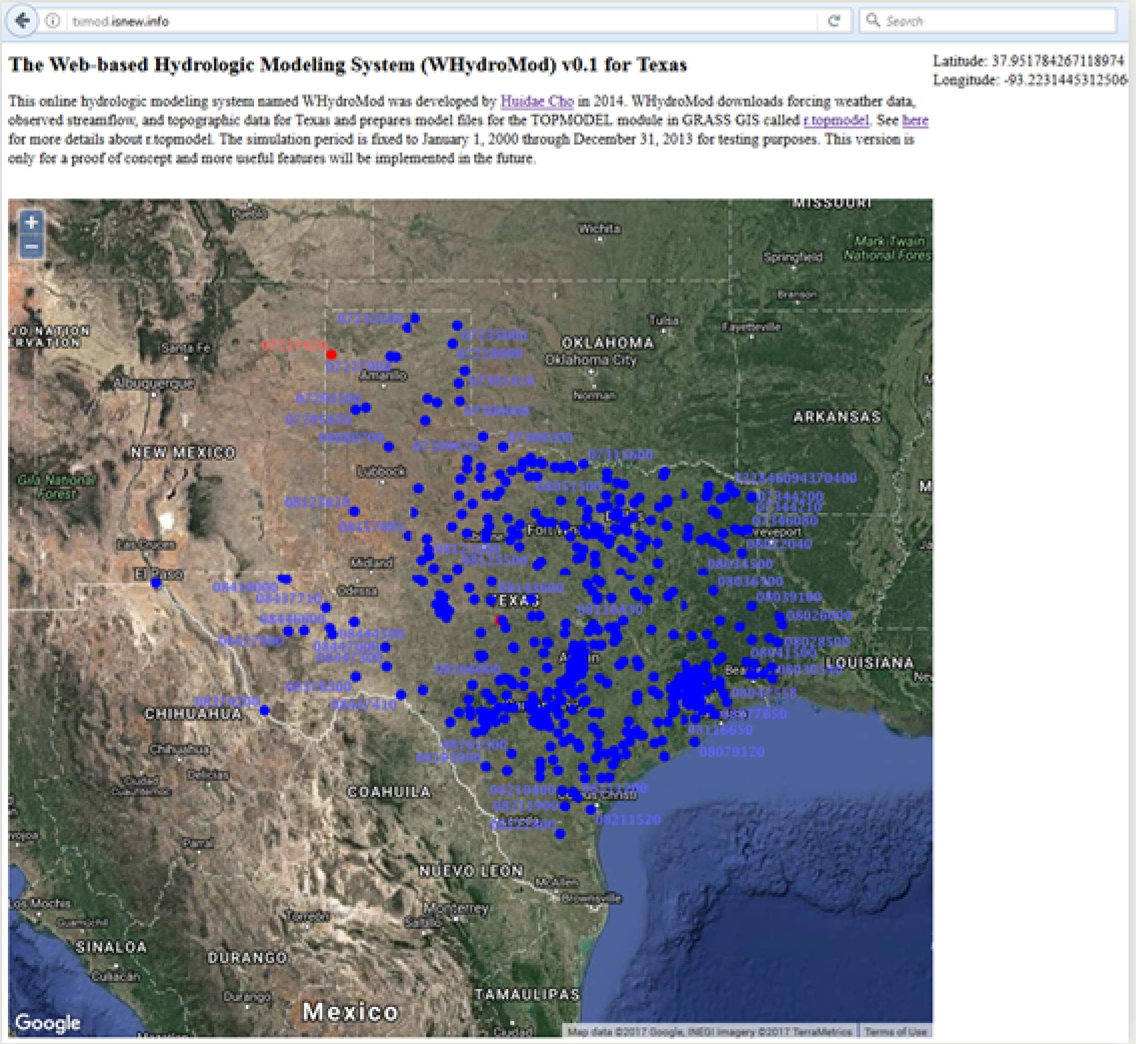
- https://github.com/OSGeo/grass-addons/tree/master/grass7/raster/r.accumulate

Web-based hydrologic modeling system (WHydroMod)

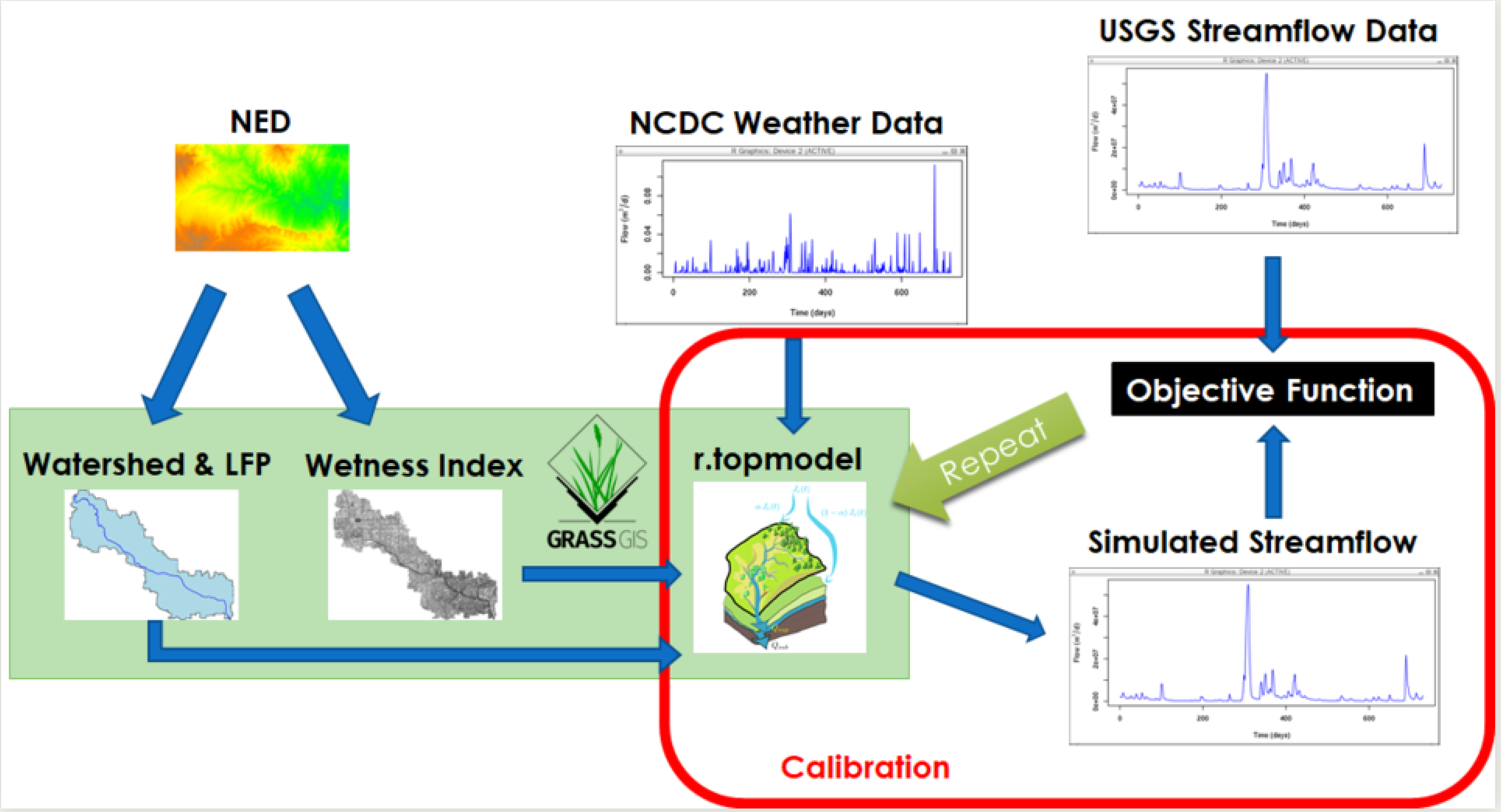
- r.topmodel: A GRASS module for the Topography Model (TOPMODEL)
- USGS online data
- NCDC web API

WHydroMod for Texas

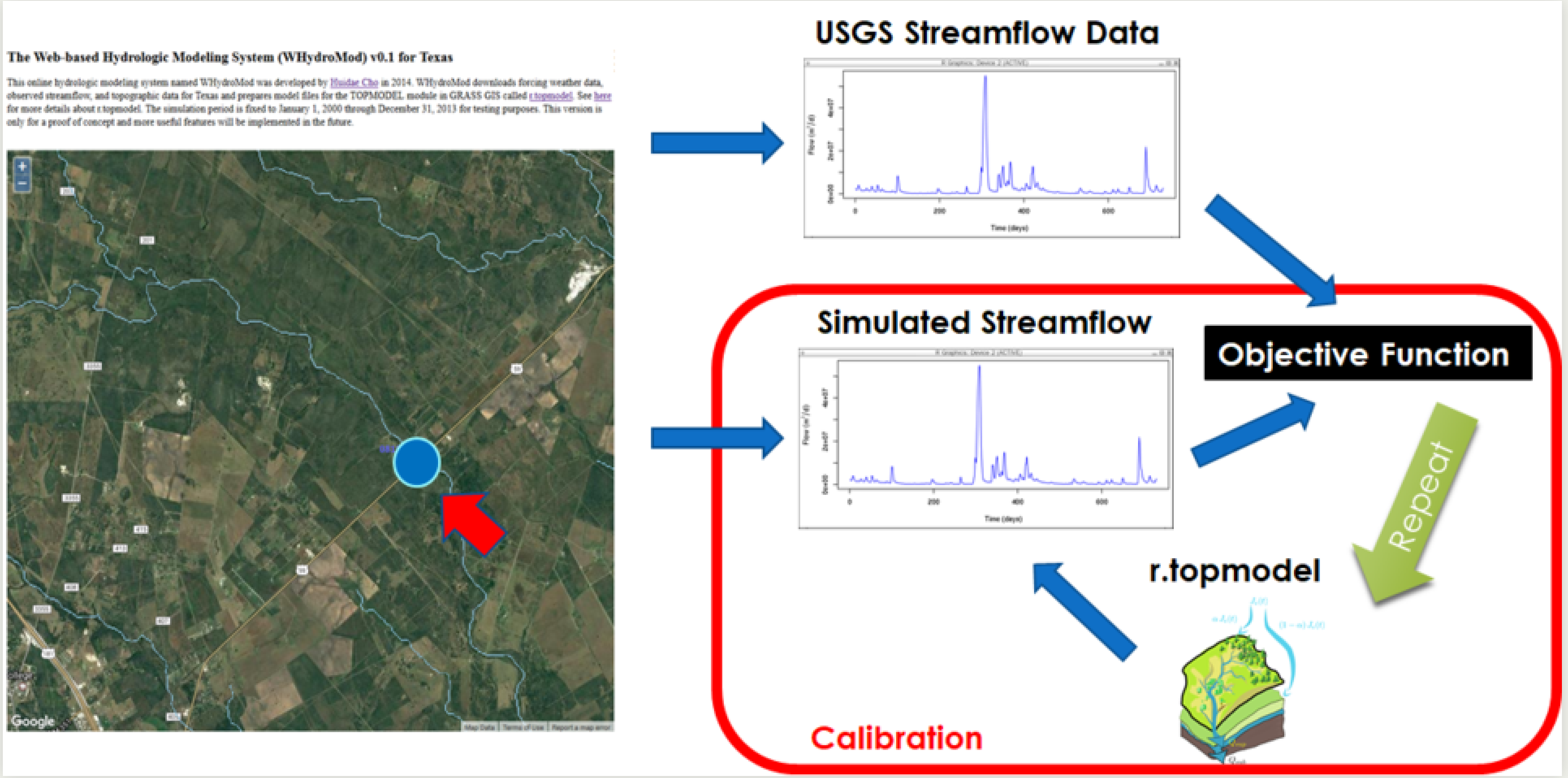
- http://txmod.isnew.info
- Proof-of-concept implementation for Texas
- GNU Affero General Public License Version 3
- Transparent system to the user
- Open source stack
  - GRASS GIS
  - PostgreSQL
  - PyWPS
  - MapServer
  - OpenLayers
  - Apache
- Bing Maps API for base maps



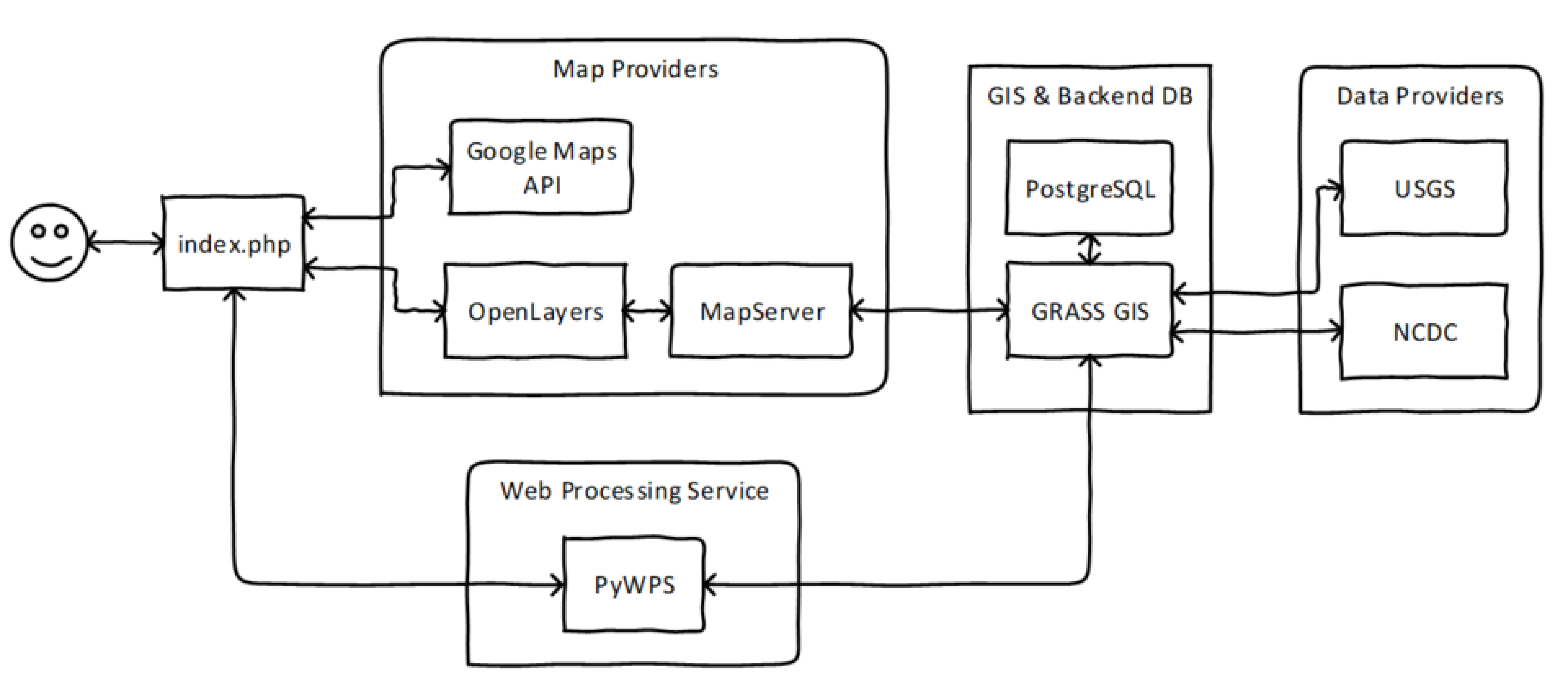
Typical TOPMODELing



The WHydroMod way



Data flow


Map Providers
Google Maps API
index.php
OpenLayers
MapServer
GIS & Backend DB
PostgreSQL
GRASS GIS
Data Providers
USGS
NCDC
Web Processing Service
PyWPS

Challenges

Web users are impatient

But it takes minutes to download online data & process DEM

- Blue & red USGS gages
- Automate heavy data processing over night
- Allow the user to initiate new processing and share results
- **Improve the performance of geospatial computation**

Important hydrologic parameters

- Flow direction
- Flow accumulation
- Longest flow path

Flow direction

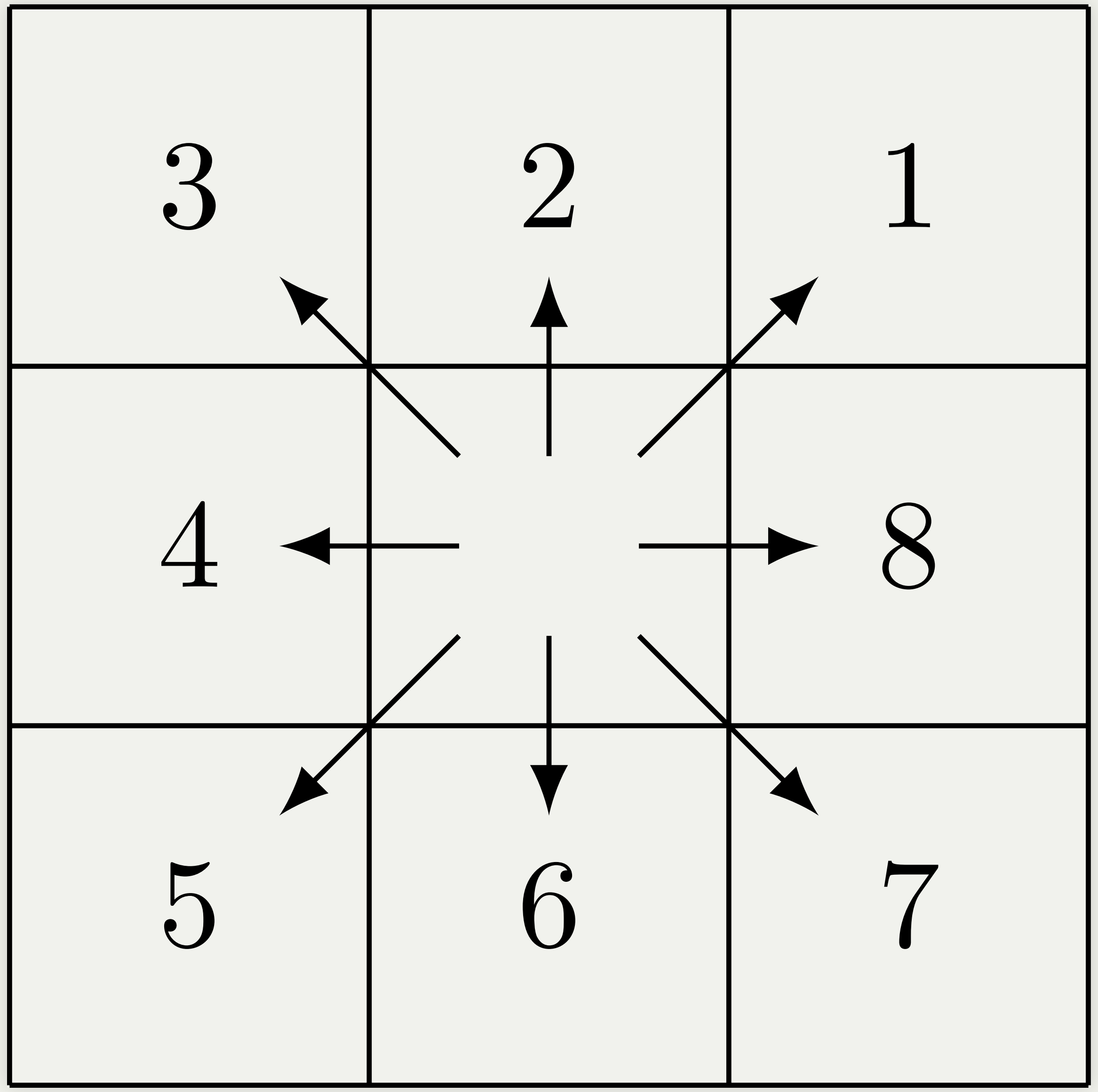
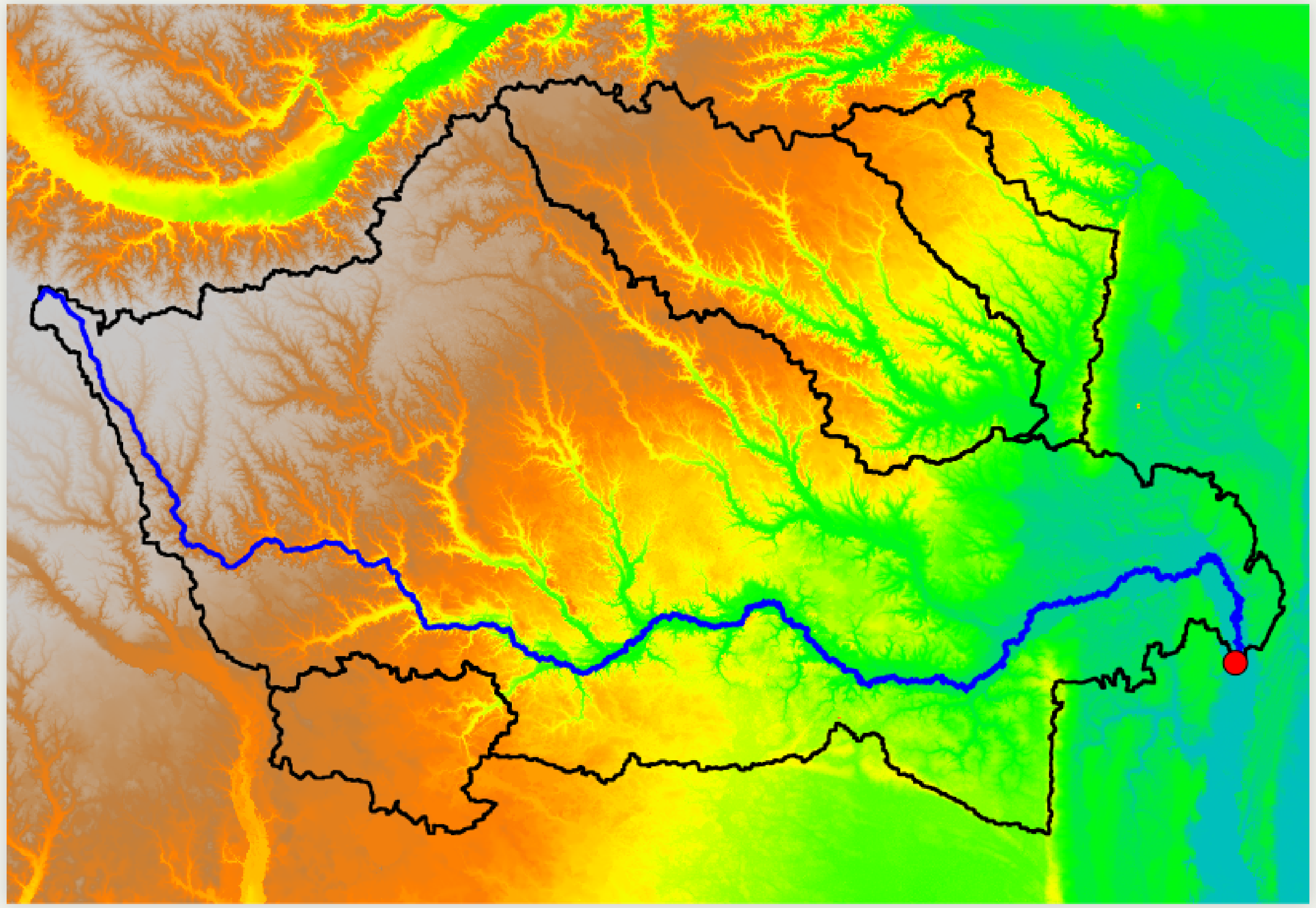


Figure 1: GRASS GIS drainage encoding

Flow accumulation

A raster (think of a matrix with cell values) that downtraces rain drops following flow directions.

Longest flow path



FP is the watercourse from one point to another.

$$\overrightarrow{\mathbf{LFP}} \in \left\{ \overrightarrow{\mathbf{FP}}_i \;\middle|\; \left\| \overrightarrow{\mathbf{FP}}_i \right\| \geq \left\| \overrightarrow{\mathbf{FP}}_j \right\| \; \forall j \neq i \right\}$$

Yes! LFP is the “longest” flow path.

There can be more than one in some case.

Motivation

The current LFP algorithm

- Peter Smith (1995) $\rightarrow$ the resolution of DEM was limited
- grid-based $\rightarrow$ can take a long time
- grid output $\rightarrow$ potentially **invalid** vector output
- calculation of two flow length grids for each outlet
- not for **a large number of watersheds**

Smith, P. N. H., 1995. Hydrologic data development system. Transportation Research Record: Journal of the Transportation Research Board 1599, 118–127.

Peter Smith’s method

Downstream flow length

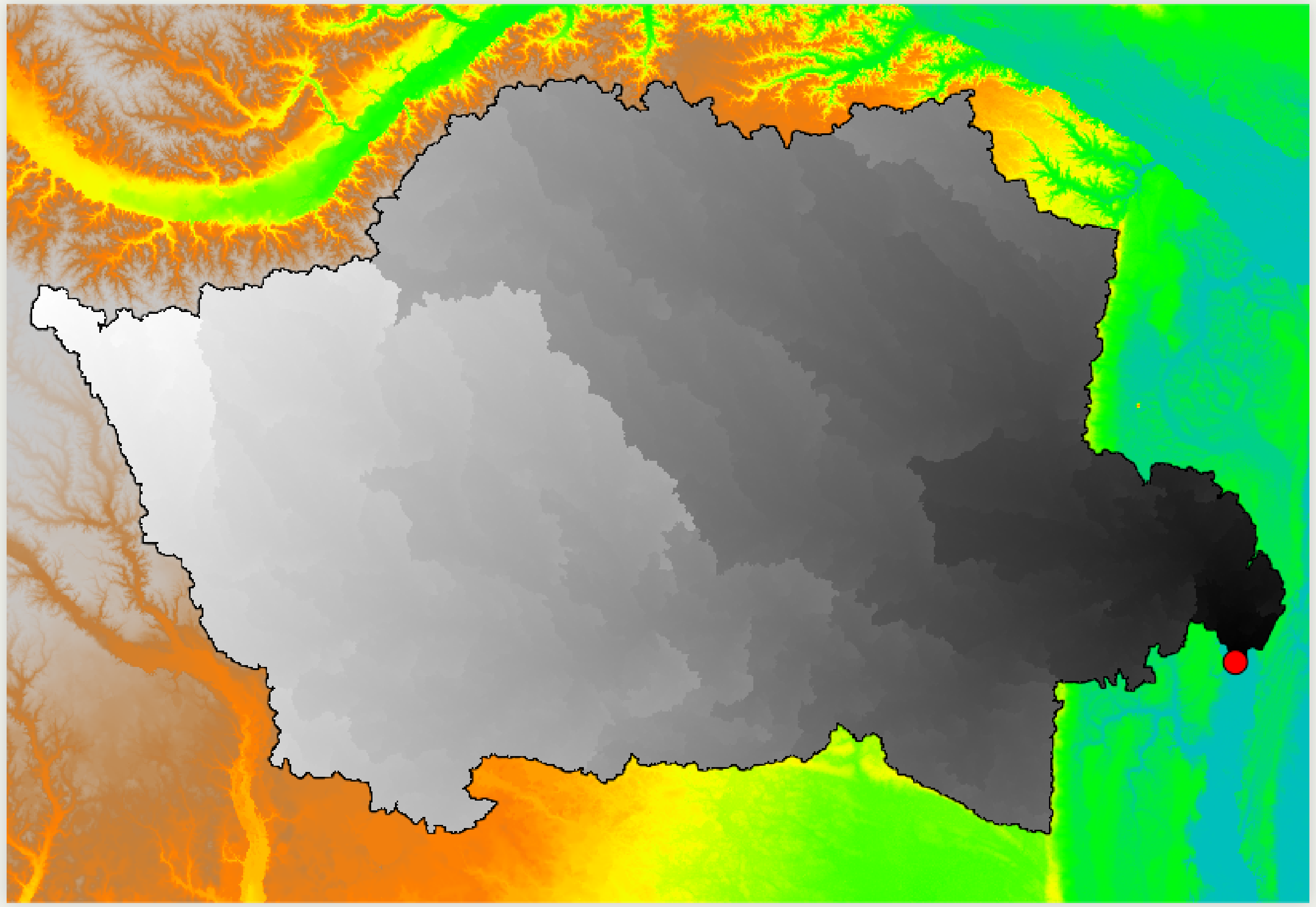
Flow length starting from the outlet

$\text{FL}_{i,j}$ is the flow length from cell i to cell j.

$$\text{DFL}_i = \begin{cases} 0 & i = 0 \text{ at the outlet} \\ \text{FL}_{i-1,i} + \text{DFL}_{i-1} & i \geq 1 \end{cases}$$

Note

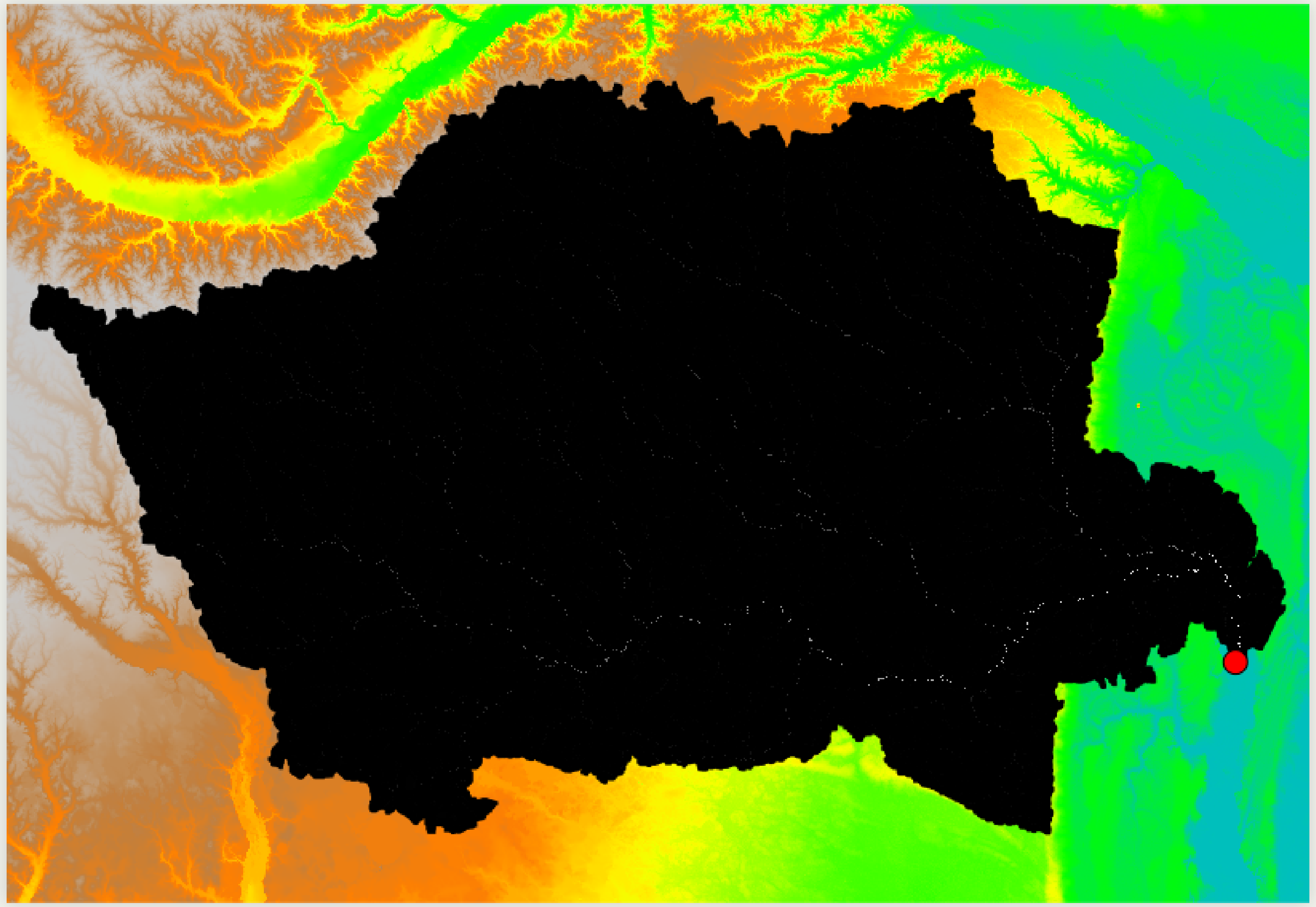
- 0 at the outlet
- high at a headwater



Upstream flow length

Flow length starting from a headwater

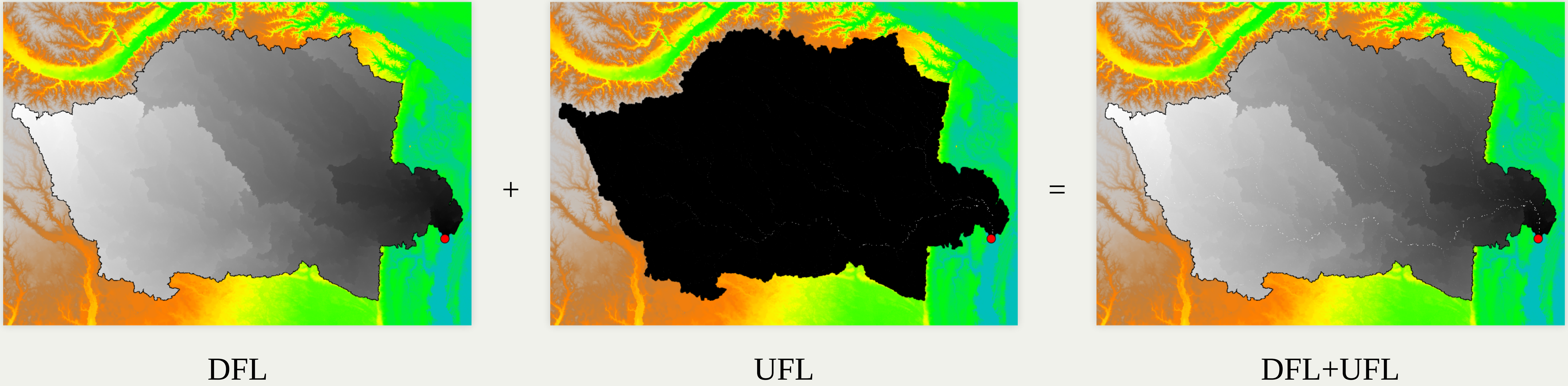
$$\text{UFL}_i = \begin{cases} 0 & i = 0 \text{ at a headwater} \\ \text{FL}_{i-1,i} + \max(\text{UFL}_{i-1}) & i \geq 1 \end{cases}$$



Note

- 0 at a headwater
- high at the outlet
- taking the maximum at a confluence

DFL+UFL



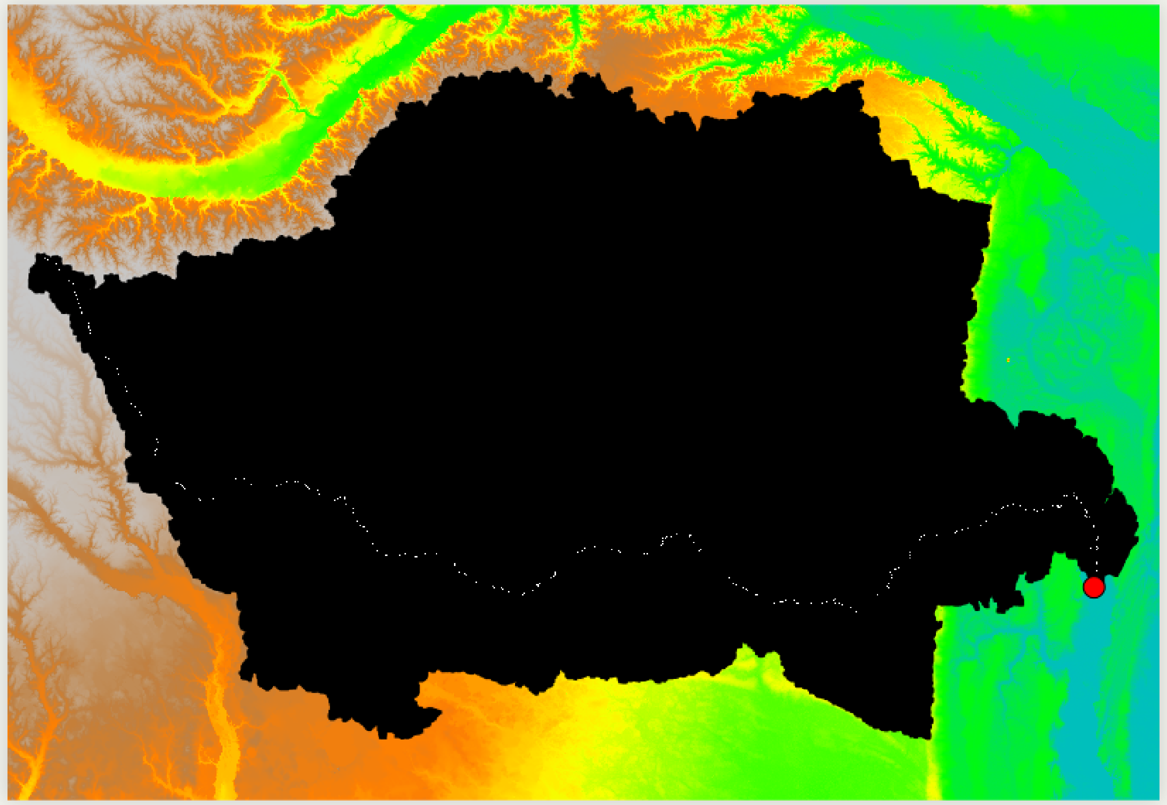
Note

- maximum DFL at the headwater on the LFP
- maximum UFL at the outlet
- both maximums are the same $\rightarrow$ LFL (longest flow length)
- any DFL+UFL cells on the LFP have this maximum value

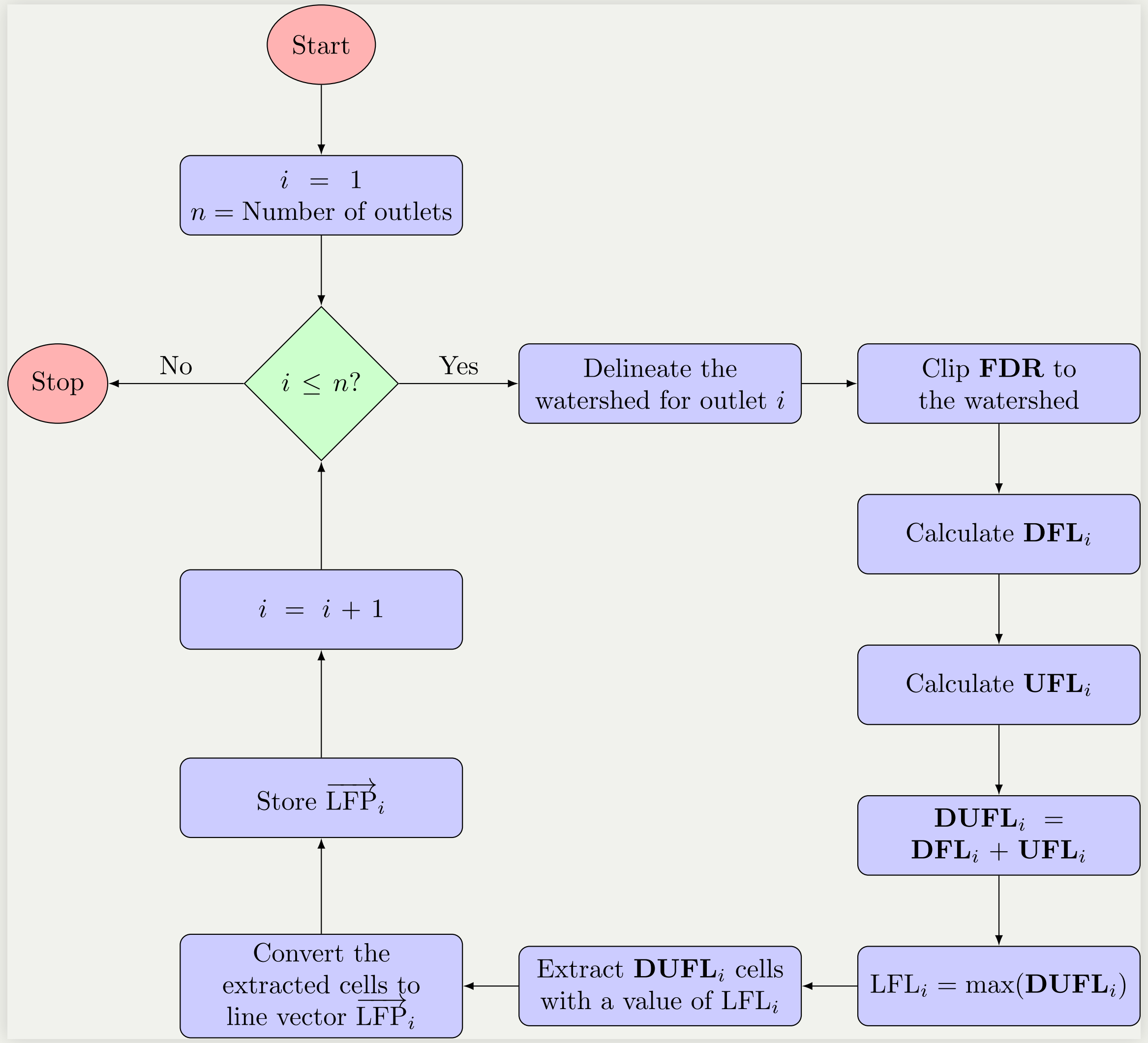
LFP

$\overrightarrow{\text{LFP}}$ defined by all the cells with a value of LFL

$$\text{LFL} = \max(\mathbf{DFL} + \mathbf{UFL})$$

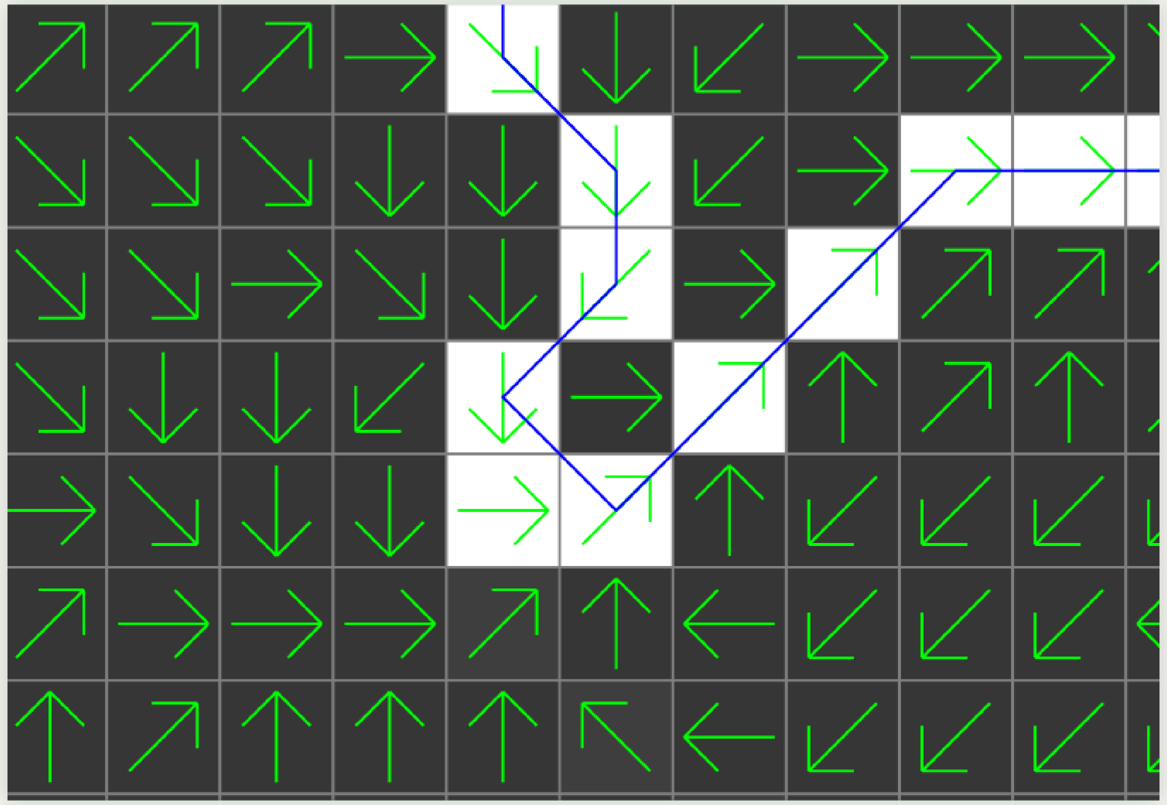


Typical procedure for multiple outlets



Critical problem

Green arrows: flow directions, blue line: LFP



Hydrologically **invalid!**

Other problems

Slow!

Esri’s Flow Length tool is limited.

Divide-and-conquor approach

Purely vector-based approach

r.accumulate GRASS GIS addon

Divide

Divide the problem into smaller pieces in a recursive way.

$$\overrightarrow{\text{LFP}}_i \in \left\{ \overrightarrow{\text{LFP}}_j + \overrightarrow{\text{FP}}_{ji} \ \forall j \in \mathbf{UP} \right\}$$

where

- $\mathbf{UP} = \{\text{Upstream neighbors of cell } i\}$
- $0 \le \|\mathbf{UP}\| \le 8$

Define a function of j that returns the length of the longest flow path at cell i.

$$f(j) = \begin{cases} \left\|\overrightarrow{\text{LFP}}_j\right\| + \left\|\overrightarrow{\text{FP}}_{ji}\right\| = \left\|\overrightarrow{\text{LFP}}_i\right\| & \text{if } \mathbf{UP} \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

Now, the problem becomes finding all $\arg\max_{j \in \mathbf{UP}} f(j)$ by traversing upstream cells starting from the outlet. The search stops when $\mathbf{UP} = \emptyset$.

Conquor

Intuitively, the longest longest flow length

$$\mathrm{LFL}_{\max} = s \cdot \mathrm{FAC}\sqrt{2}$$

where s and FAC are the cell size and flow accumulation, respectively

Based on Hack's law, the shorted longest flow length

$$\mathrm{LFL}_{\min} = s\sqrt{\mathrm{FAC}}$$

Between upstream neighbor cells i and j, cell i is on the $\overrightarrow{\mathrm{LFP}}$ if $\mathrm{LFL}_{\min,i} > \mathrm{LFL}_{\max,j}$.

Stack overflow and solution

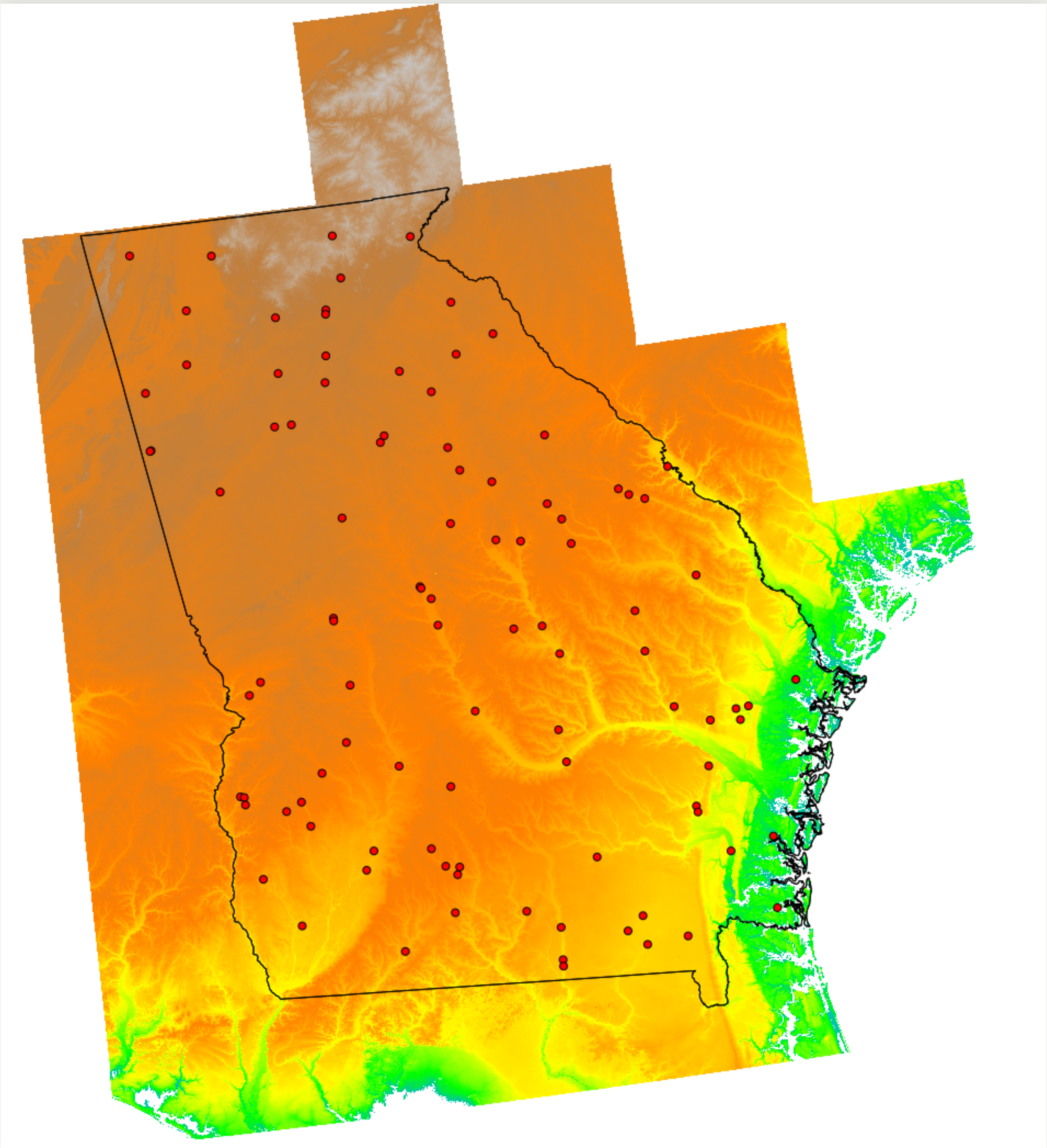
- Depending on the watershed size, recursion can consume all stack memory and cause a stack overflow
- Convert recursion to iteration using a heap-based stack
- A single loop until the stack with upstream cells depletes

Benchmark experiments: r.accumulate vs. Arc Hydro

Is paid software always better than open and free software?

Data

- National Elevation Dataset (NED) 1 arc-second (approximately 30 meters)
- 27 NED maps
- National Atlas of the United State state boundaries for masking
- Generated 100 outlets randomly



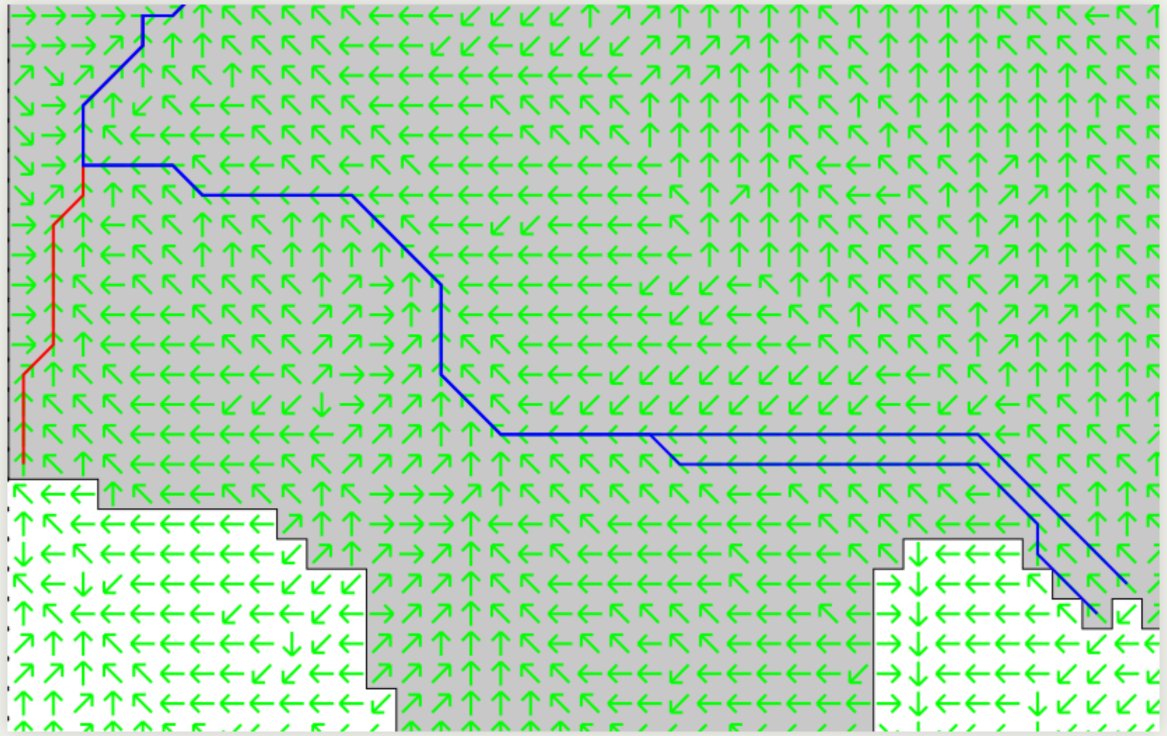
System specifications

- CPU: Intel Xeon E5620 2.40 GHz
- Memory: 48 GB
- OS: Linux kernel version 4.4.14
- GIS: GRASS development version 7.7.svn revision r74124

Results and discussions

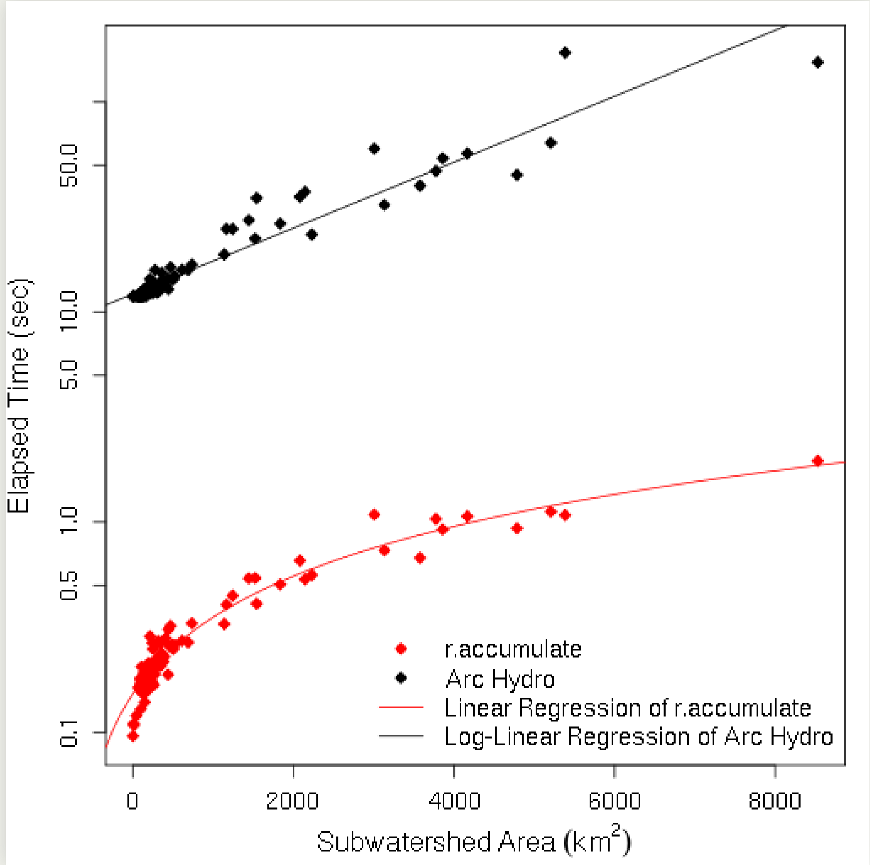
LFP geometries

Both methods produced almost identical LFPs except for some areas

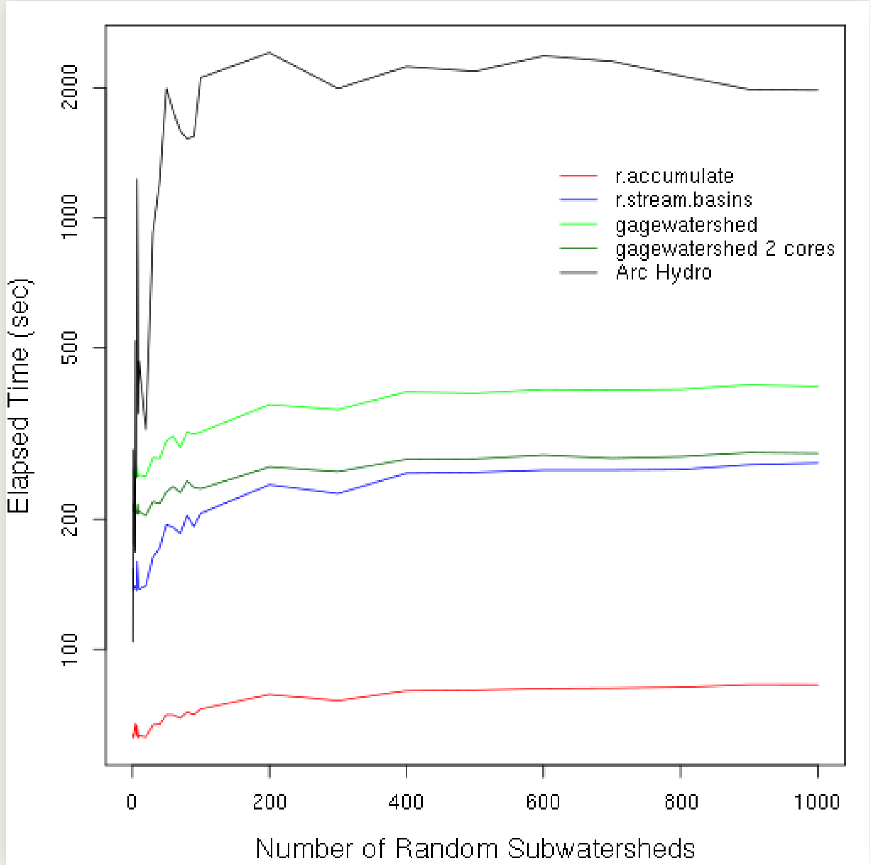


Blue: r.accumulate, Red: Arc Hydro

Performance comparison with Arc Hydro



Efficient delineation of a massive number of subwatersheds



Conclusions

- The fewer raster operations, the faster!
- The performance of r.accumulate is linearly growing with the subwatershed size.
- While that of Arc Hydro exponentially growing.
- The new approach is cost-efficient and can be used for interactive hydrologic modeling (e.g., web).

References

- Cho, H., 2020. *A Recursive Algorithm for Calculating the Longest Flow Path and Its Iterative Implementation.* Environmental Modelling & Software 131, 104774. doi:10.1016/j.envsoft.2020.104774.
- Smith, P. N. H., 1995. *Hydrologic Data Development System.* Transportation Research Record: Journal of the Transportation Research Board 1599, 118–127. doi:10.3141/1599-15.
- Hack, J. T., 1957. *Studies of Longitudinal Stream Profiles in Virginia and Maryland.* Geological Survey Professional Paper 294-B, 45–97. https://pubs.usgs.gov/pp/0294b/report.pdf.