Longest Flow Paths, Shortest Compute Times

Huidae Cho hcho@nmsu.edu

Department of Civil Engineering, New Mexico State University

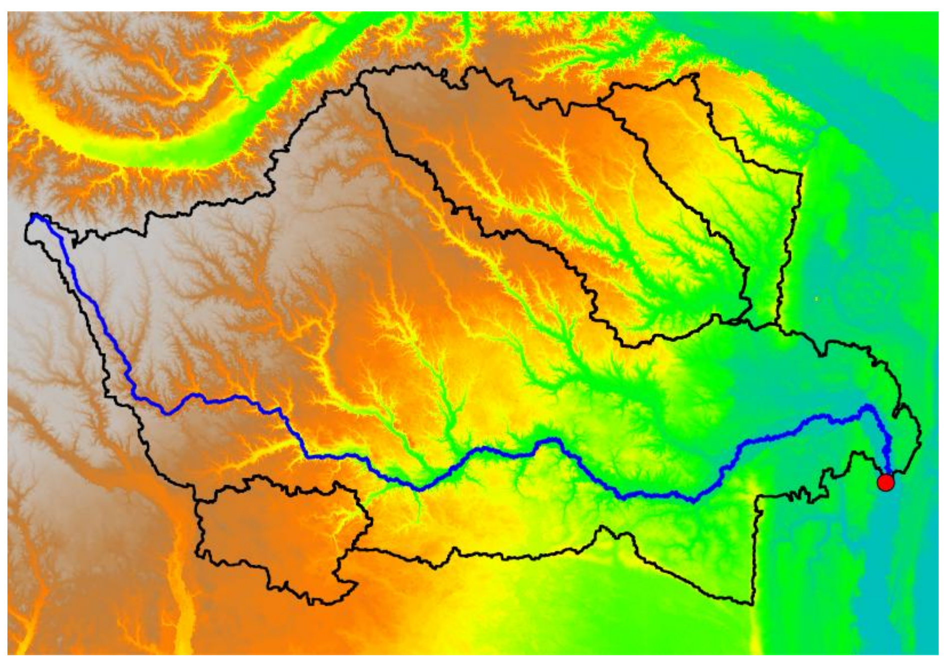
Presented at the 2025 CIROH Developers Conference, Burlington, VT

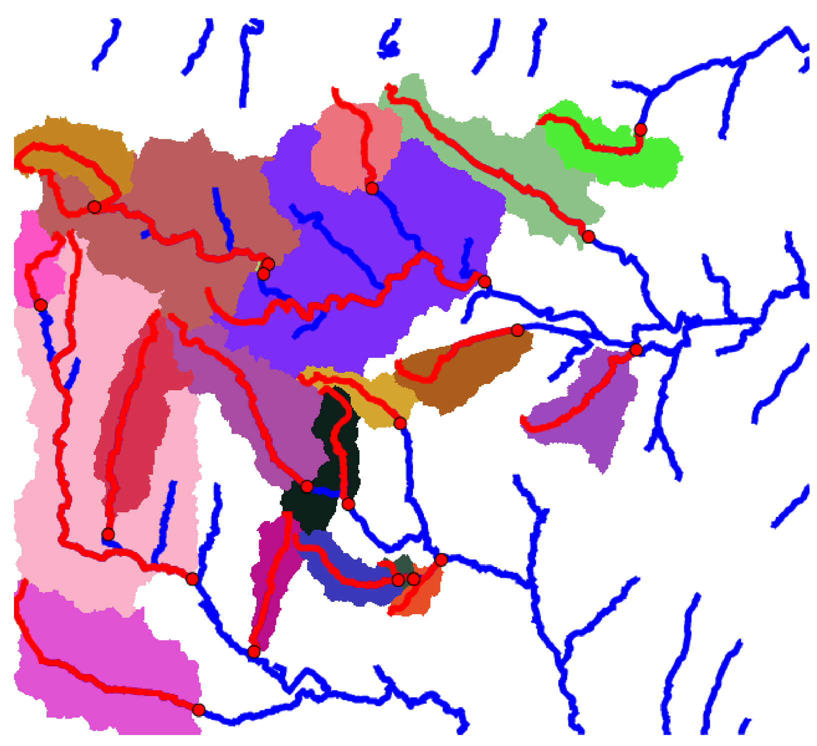
https://github.com/HuidaeCho/melfp

Longest Flow Path

A flow path (FP) is the hydrologic path or watercourse from one point to another in the watershed. The longest flow path (LFP) represents the flow path from a headwater point (typically on the watershed divide) to the outlet that is longer than all other flow paths in the watershed.

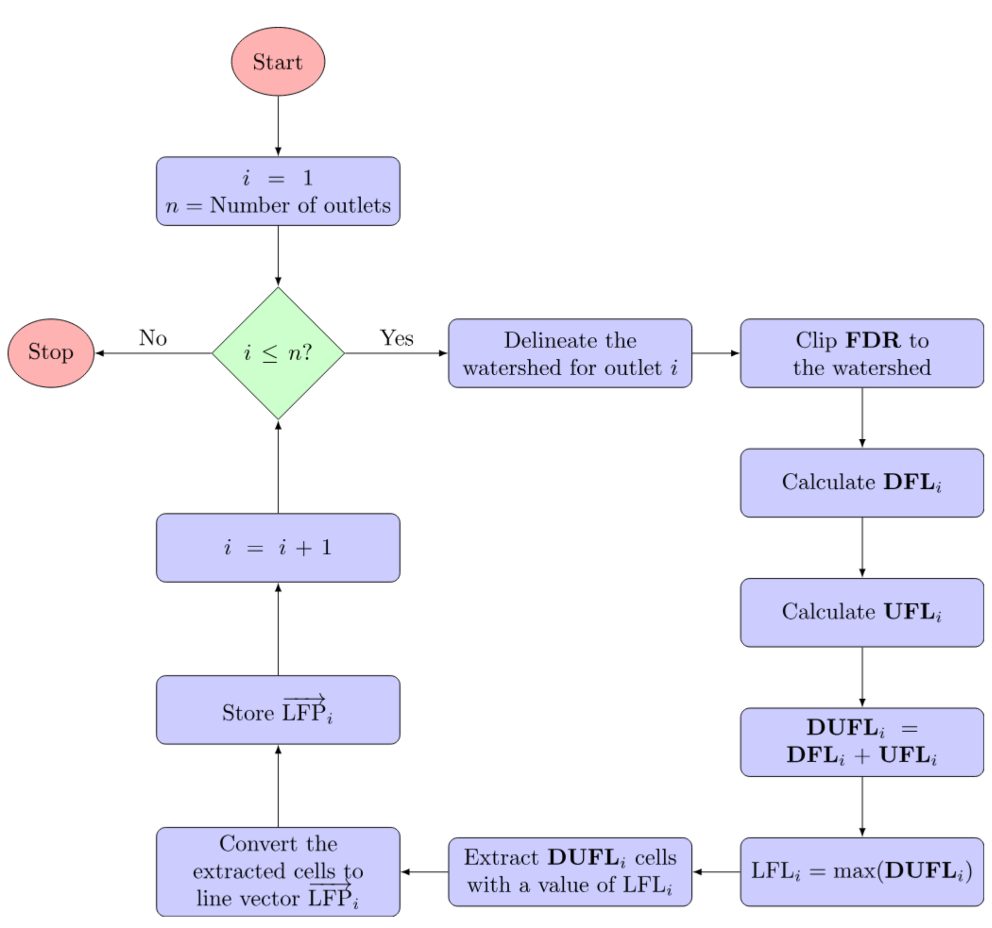
$$\overrightarrow{\mathrm{LFP}}_i \in \begin{cases} \left\{\overrightarrow{\mathrm{LFP}}_j + \vec{P}_{ji} \,\middle|\, \left|\overrightarrow{\mathrm{LFP}}_j + \vec{P}_{ji}\right| \geq \left|\overrightarrow{\mathrm{LFP}}_k + \vec{P}_{ki}\right| \forall j,k \in \mathbf{UP}_i, j \neq k\right\} & \text{if } \mathbf{UP}_i \neq \emptyset \\ \{\vec{0}\} & \text{otherwise} \end{cases}$$



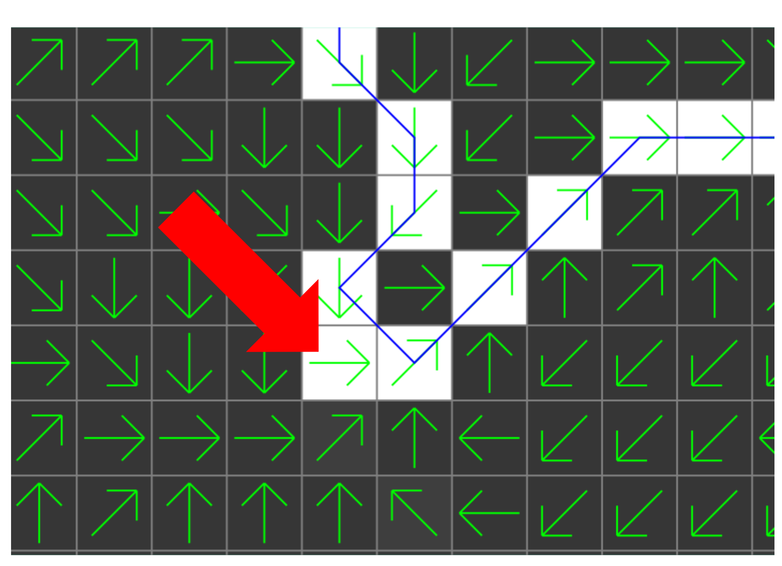


Smith's Method

Smith's (1995) (original? LFP) method uses a raster-based approach. LFL=max(DFL+UFL) and the typical workflow for multiple outlets involves a lot of raster operations!

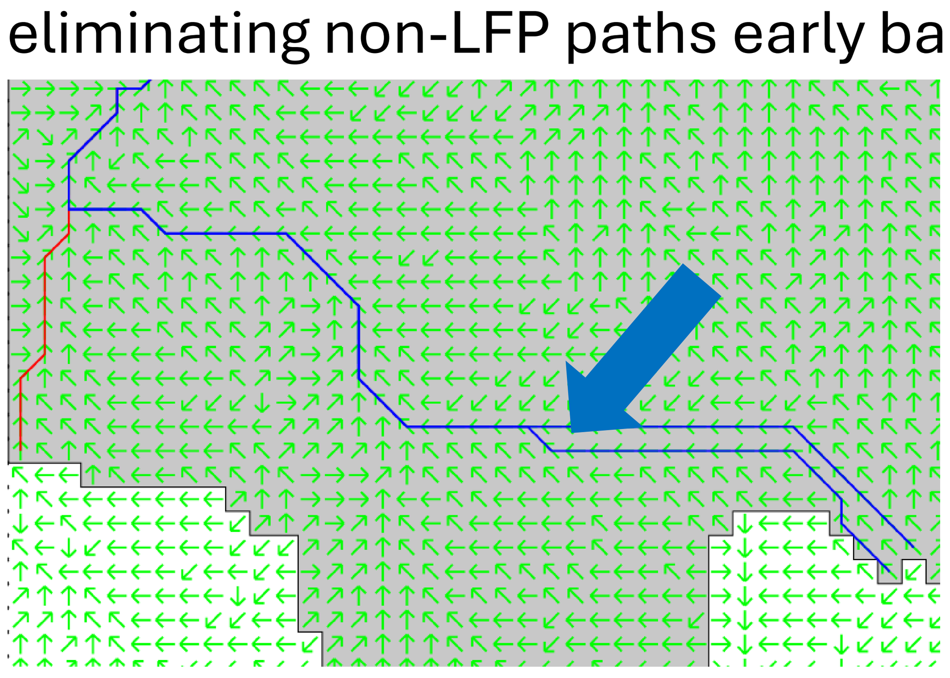


Vectorized results are not always guaranteed to be hydrologically correct!



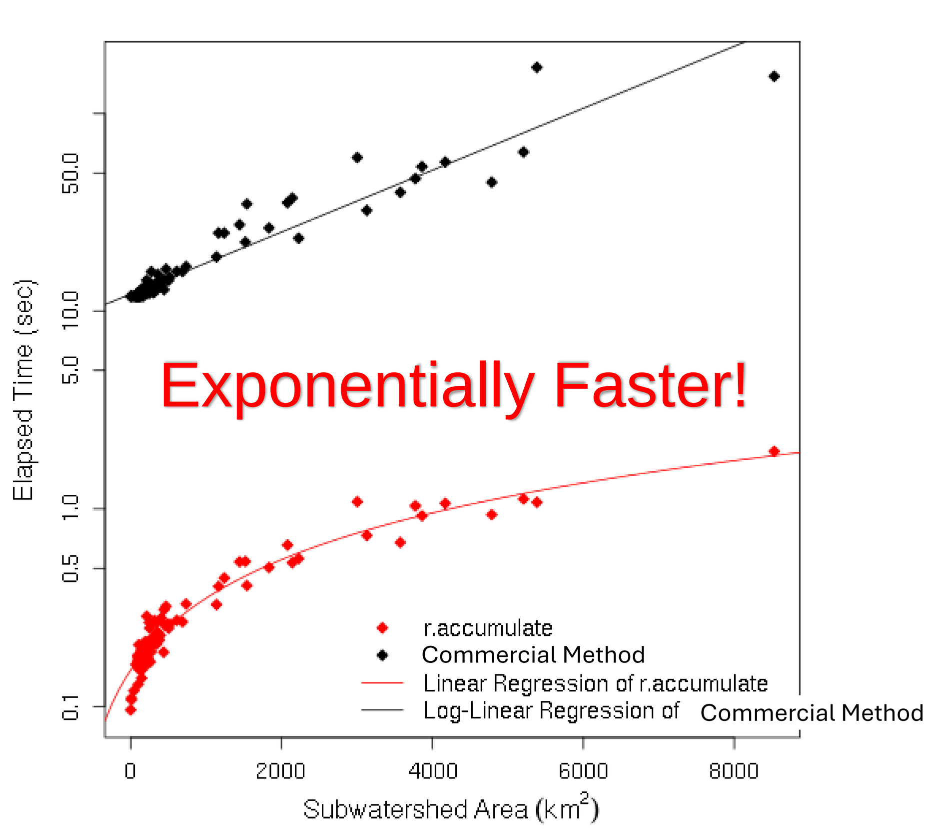
Cho's (2020) Method

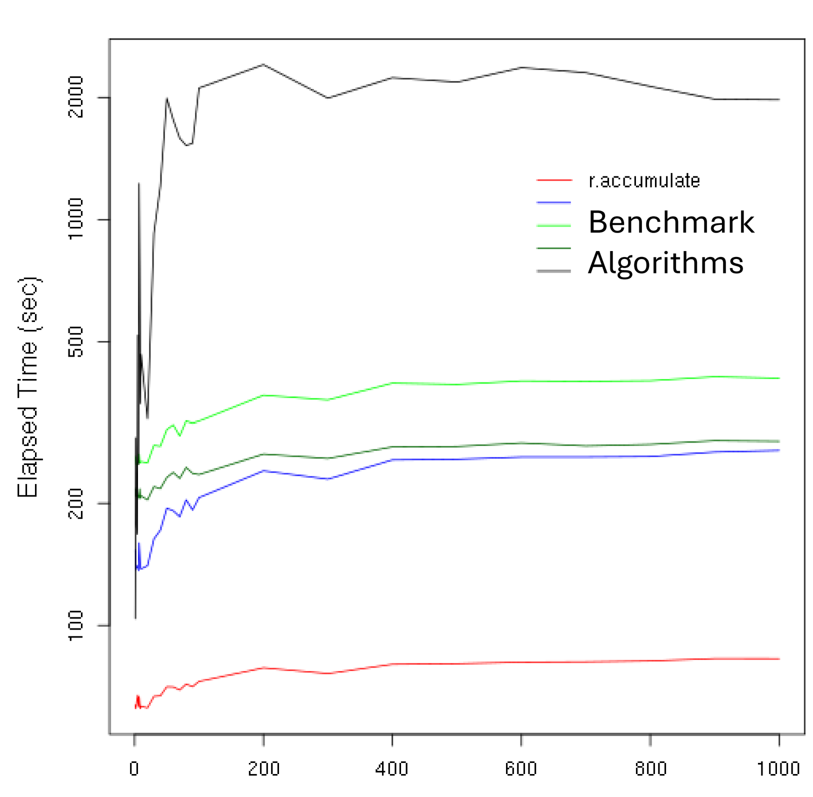
Cho's (2020) method uses a divide-and-conquer approach, eliminating non-LFP paths early based on Hack's law.



Equally LFPs were identified. See commercial software's result.

It's implemented as a GRASS addon, r.accumulate.



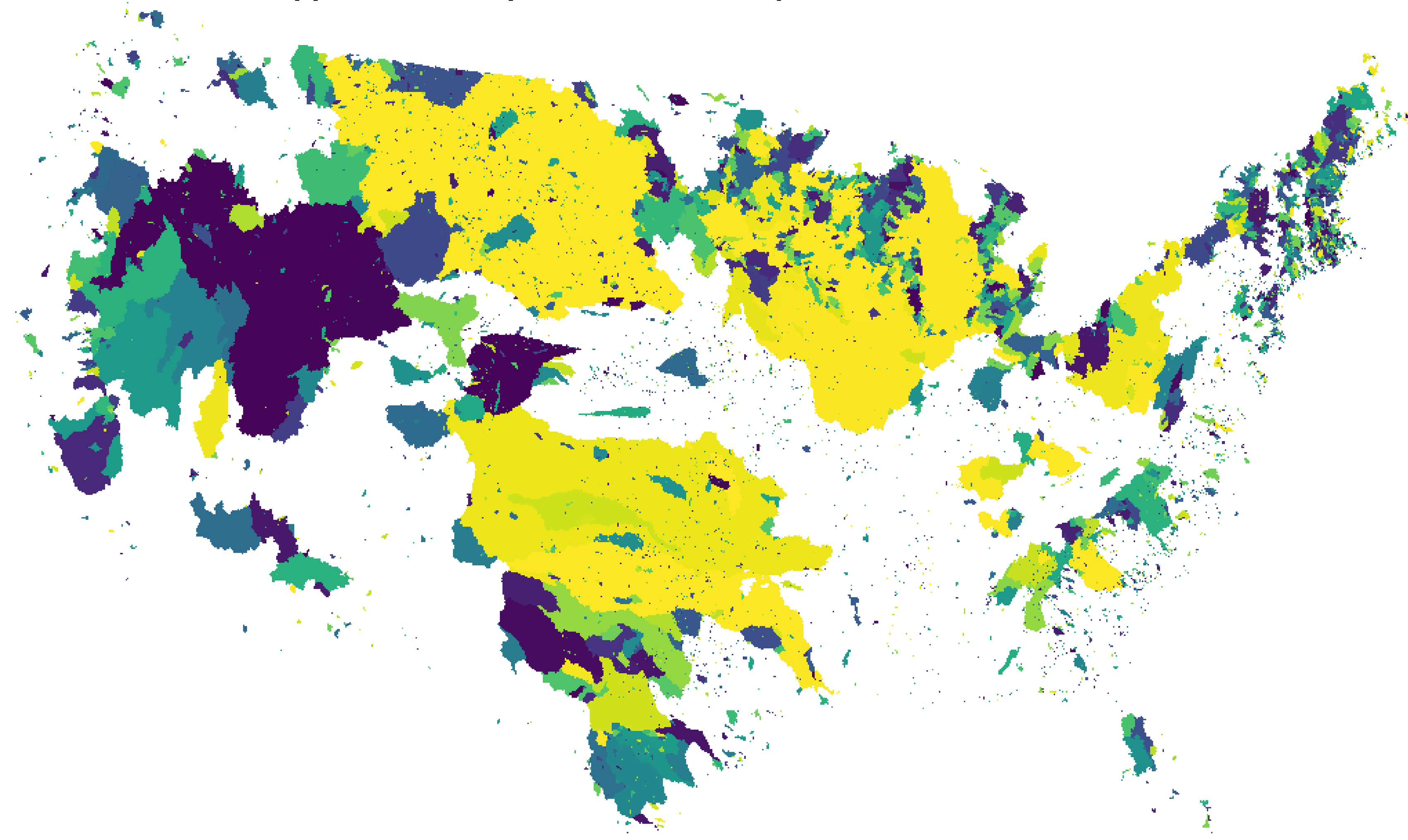


The Computational Lab for Advanced Water Resources Informatics and Modeling

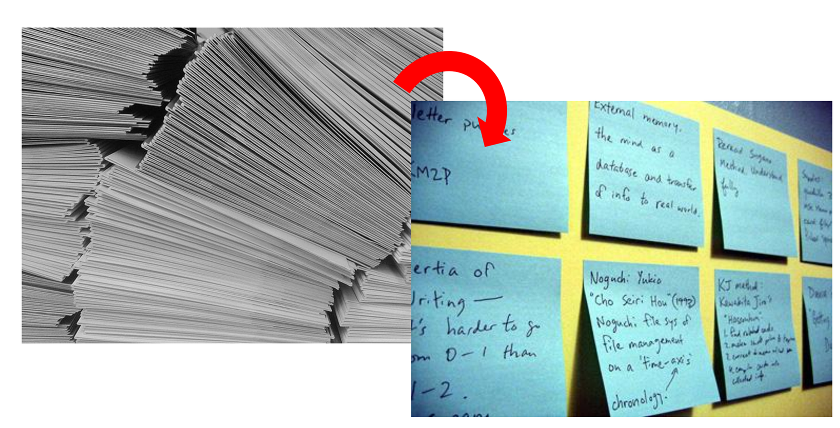
https://clawrim.isnew.info/

Cho's (Under Revision) Method

Cho's (2020) method is faster than other sequential algorithms, but it is still sequential. Cho's (Under Revision) new method called Memory-Efficient Longest Flow Path (MELFP) is an OpenMP parallel algorithm for computing longest flow paths for many outlets. The **motivation** was to calculate longest flow paths for 91,611 dams in the CONUS.

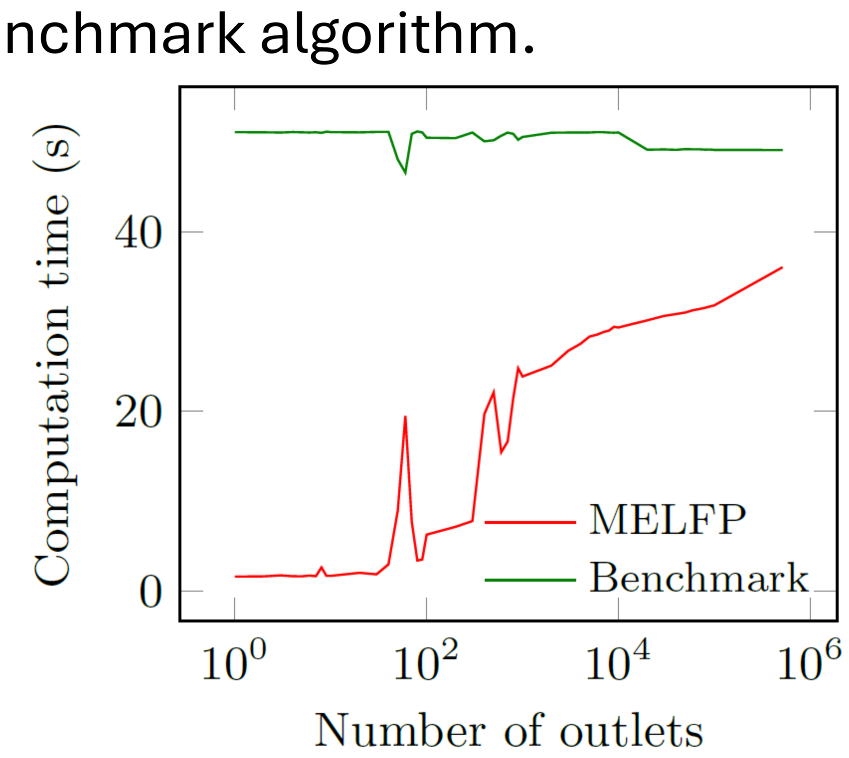


The major **challenge** was data size (15 billion cells, 14GB input data, large 14GB intermediate data required by existing methods). Input + Intermediate + Output = 84GB+ > System 64GB!



Stack-based looping then tasking for improved load balancing, No more large intermediate data!

MELFP achieved a 66% reduction in computation time utilizing 79% less peak memory (20GB vs. 96GB) and 33% higher CPU resources (like it or not?) compared to its **only** parallel OpenMP benchmark algorithm.



Acknowledgments

The author gratefully acknowledges the support from CUAHSI HydroLearn for travel to attend the CIROH Developers Conference. Their support has enabled valuable collaboration and knowledge exchange within the hydrologic modeling and software development community. Image attribution:

- https://www.pickpik.com/stack-white-papers-letters-envelopes-letter-stack-post-60116
- https://www.flickr.com/photos/ayalan/317935661

References

- Cho, H., Under Revision. *Loop then task: Hybridizing OpenMP parallelism to improve load balancing and memory efficiency in continental-scale longest flow path computation.* Environmental Modelling & Software.
- Cho, H., 2020. *A Recursive Algorithm for Calculating the Longest Flow Path and Its Iterative Implementation.* Environmental Modelling & Software 131, 104774.
- Smith, P. N. H., 1995. *Hydrologic Data Development System.* Transportation Research Record: Journal of the Transportation Research Board 1599, 118–127.