

Overall, this is a technically solid and potentially useful data product, particularly because it provides a globally consistent, 90-m, compute-ready representation of basin topology and upstream attributes. However, the manuscript currently overstates the methodological novelty and computational advantage relative to existing products such as MERIT-Hydro and MERIT-Basins. The main contribution is better framed as reducing user-side preprocessing through DFS indexing and regional partitioning, rather than making these analyses globally possible for the first time. The manuscript would also benefit from clearer distinctions between preprocessing cost, query cost, and full catchment extraction, as well as resolution of the licensing issue for derivative products.

<Major Concerns>

Product name: MERIT-FullBasin

The name “MERIT-FullBasin” may give users the impression that this dataset is an official extension or component of the MERIT product family. As the dataset is independently developed by the authors using MERIT-Hydro as an input, I suggest clarifying this relationship and considering whether the current naming could cause confusion regarding provenance, endorsement, or responsibility for the product.

Positioning of novelty

The manuscript somewhat overstates that these upstream calculations become globally feasible for the first time. Similar analyses are already possible with MERIT-Hydro, or with vector products such as MERIT-Basins combined with Pfafstetter-type hierarchical IDs. In the latter case, upstream sub-basins can be identified efficiently from the hierarchy, and only the most downstream sub-basin containing the target point requires additional pixel-level calculation.

The stronger and more defensible contribution is that the proposed DFS indexing and regional partitioning provide a globally consistent, 90-m, compute-ready dataset that substantially reduces user-side preprocessing.

I therefore suggest reframing the novelty around compute-readiness and ease of use, while acknowledging that comparable calculations are possible from existing MERIT-Hydro/MERIT-Basins products with additional processing.

Data License:

Because MERIT-Hydro is distributed under a more restrictive license (CC BY-NC / applicable derivative-product terms), releasing a representation from which the original D8 direction can be fully reconstructed under the more permissive CC BY 4.0 license appears incompatible

with the licensing conditions of the source dataset. In practical terms, this would allow users to obtain an effectively equivalent version of the MERIT-Hydro flow-direction data under a less restrictive license.

This issue should be resolved before publication.

<Specific Comments>

Abstract:

The Abstract does not clearly explain the core idea that enables the efficient large-scale computation. In particular, it is difficult for a first-time reader to understand why the DFS/tree-depth representation is computationally advantageous, and how it differs from simply storing basin IDs or tracing the original D8 network. I suggest briefly explaining the key mechanism—i.e., that the flow network is converted into an interval-indexed tree structure so that upstream relationships can be queried without repeated recursive tracing. This would make the technical strength and practical usefulness of the product much clearer.

L43: Per-pixel queries reduce to a single GeoTIFF read or a single table lookup.

This statement also appears too broad. Direct lookup applies to precomputed attributes and basin-level records, whereas arbitrary upstream extraction or aggregation over a new variable still requires additional processing (e.g., scanning the relevant domain or constructing a DFS-ordered prefix sum). Please clarify which queries are genuinely direct lookups and which require preprocessing or domain traversal.

L279: (f) Under the UPA-descending sibling order, any maximal run of consecutive `dfs_in` values traces an eldest-child chain:

This explanation and eldest-child concept appears suddenly, and difficult to follow by readers. Please explain more clearly why the UPA-descending child order leads to a hydrologically meaningful “main stem,” and how this definition relates to conventional main-stem definitions such as longest flow path or highest-order channel.

L324: Adjacent 324 targets on the same stream share most of their boundary (Fig. 7b), so each stream

Please state the computational motivation at the beginning of this part (better to start a new paragraph here, as it focusing on efficiency). The key idea is to avoid tracing the catchment boundary independently for every target pixel by reusing the largely shared boundary between adjacent targets along the same stream. Without this motivation, the transition to the incremental-update scheme is rather abrupt.

L371: each region fits in memory under standard GIS or array tools.

The statement that each region is “sized so it fits in memory under standard GIS or array tools” is too vague, because available memory differs widely among users. Please provide the actual size of the largest regional tile, preferably both compressed file size and uncompressed in-memory size for one raster layer. It would also be useful to indicate the approximate memory required for a typical upstream-query workflow involving `dfs_in`, `dfs_out`, and one additional raster.

L510: vector products such as MERIT-Basins (Lin et al., 2021) represent each basin as a polygon and lose per-pixel addressability in the conversion.

The statement that vector products such as MERIT-Basins “lose per-pixel addressability” seems somewhat overstated. Basin membership at the pixel level can in principle be recovered by rasterizing the basin polygons back onto the target grid. The practical distinction is rather that pixel-level lookup is not provided directly and requires an additional rasterization/preprocessing step. I suggest revising this comparison to avoid undervaluing existing vector products. The advantage of the proposed product must be “ready to compute format”.

L518: precision at the fixed per-query cost of the segment-based tools.

The claim of a “fixed per-query cost” needs to be qualified more carefully. For external rasters, the manuscript itself states that the data must first be reordered by `dfs_in` and prefix-summed before upstream sums, means, or variances can be obtained at unit cost. Thus, the constant-time query applies only after a potentially substantial preprocessing step over the relevant region/domain. Please distinguish clearly between preprocessing cost and subsequent query cost.

Similarly, the interval test provides $O(1)$ membership testing for a given pixel, but extracting or enumerating the complete upstream catchment still requires reading a number of pixels that scales with catchment size (or scanning the relevant region). The current wording risks conflating constant-time membership/aggregate queries with full catchment extraction.