Author's Response (Revision Round 2)

2024-10-09 - Daniel Kovacek & Steven Weijs

Authors' thanks: We are grateful to the editors and anonymous reviewers for their time dedicated to reviewing and providing feedback on our manuscript. The points raised have meaningfully improved the quality of the final revision.

A brief response to the editor's (additional) comments regarding figures is presented first below. No comments were provided by Reviewer #2, so we include final responses to comments from Reviewer #3 only. The information is organized in the following order: i) reviewer comment, ii) author response, and iii) line numbers and/or sections identifying related manuscript changes. Please note that page and line numbers referring to manuscript edits correspond to the **revised manuscript**.

Responses to Editor's comments

Editor's Comment: *I find that the quality of the figures in the paper is still not satisfactory for publication in ESSD. For instance, Figures 4 and 5 are particularly difficult to follow..*

Author's Response: We have revised the maps (Figures 1, 3, 4, 5, and 9) to be more consistent in appearance and have added geographic reference features to maps where appropriate. Where gridlines are presented, a North arrow is left out since this information is indicated by the latitude and longitude grid line labels. Thank you for being persistent in the feedback regarding figure quality, we hope these revisions are in line with the standard expected from the journal. In addition, care has been taken to ensure that axis and legend labels in quantitative figures are larger and more consistent in their appearance. Quantitative figures (6, 7, 8, 11, and 12) have been updated to increase font sizes overall and use a consistent font where possible. Note that Figure 8 had to be reprocessed and it drew a different random sample of sub-basins so the distribution has changed slightly but the conclusion about the effect of DEM resolution on the computed slope does not change.

Responses to RC3 comments

RC3 Comment: *"From Table 1, it is not clear if these are the only provided fields because Section 2.1.3 addes some new fields. … It will also be useful to provide a new table (similar to what's provided in README) summarizing the final released data, fields, units, etc."*

Author's Response: Table 1 represents a summary of ***data sources*** used to derive all attributes. Table 2 (referenced at lines 170 and 200) contains all the attributes, descriptions, units, etc as you suggest.   Unfortunately due to the size of the table owing to the number of attributes / fields, it is relegated to the end of the paper following the references automatically by LaTeX. A reference to Table 2 has been added nearer to Table 1 for improved clarity.

Corresponding Manuscript Edits:

- The introduction of section 2.1 (near line 105) has been revised as follows: *"Table 1 provides a summary of the geospatial data sources—including digital elevation, land cover, soil, and climate datasets—that attributes of ungauged basins were extracted from. These datasets were processed through the data preparation pipeline outlined in Figure 2, and the resulting attributes are listed and described in Table 2."*

RC3 Comment: *…many recently published LSH work in ESSD already provides more fields than what's provided here for British Columbia. Can the authors provide more fields, or justifications on why they are not provided here, by comparing it with existing work?*

Author's Response: Since the BCUB dataset does not include streamflow, it doesn't strictly belong with LSH datasets. The dataset was initially created to provide a basis for comparing streamflow monitoring networks to the much larger ungauged areas using similar attributes, which we believe offers an important comparison and a novel one. The primary goal, in our case, is optimizing streamflow monitoring networks—a task, to our knowledge, not previously approached by characterizing the ungauged space to a level of detail approaching that which monitored catchments are described in the LSH literature.

Basin delineation is a source of uncertainty in the estimation of catchment attributes, and this is one reason other datasets, like the Caravan dataset (Kratzert et al. 2023), initially excluded catchments smaller than 100 km² though we not only in very recent months this constraint has been revised in part. However, as noted in lines 78-81 of the manuscript regarding the HYSETS dataset (Arsenault et al. 2020), which is a major component of the Caravan dataset, catchments under 50 km² account for nearly one-third of monitored catchments in British Columbia. We believe this exclusion introduces a significant bias but also presents a meaningful opportunity for further exploration which our dataset aims to support.

A major component of this work was the delineation of catchment boundaries for a very large set of ungauged catchments. LSH datasets typically use catchment polygons provided by official governing bodies, in HYSETS (which is a component of Caravan) the catchment polygons are from the Water Survey of Canada (WSC), the U.S. Geological Survey (USGS), and the (Mexican) Comisión Nacional del Agua (Conagua). The importance of deriving catchment polygons from consistent sources is a component of the design criteria of the dataset (line 50-54), and some consequences of not doing so are described in the Technical validation sections, 2.2.2 and 2.2.3, namely inconsistency of certain terrain attributes when using elevation datasets at different resolutions. The emphasis placed on catchment delineation is a more direct treatment of one source of uncertainty in catchment attributes that is not often addressed in the LSH literature.

We recognize the set of attributes we provide in the BCUB is not as comprehensive or numerous as other datasets. Given the rapid development of attributes in the LSH literature in recent years, we did not aim to provide the most complete and current set of attributes as this is a moving target. In addition, the utility and uncertainty in data sources underlying certain attributes (i.e. soil, geology) have been noted in the literature (Beck et al. 2015; Addor et al. 2018). Given that our dataset is roughly two orders of magnitude larger than the largest LSH dataset (Caravan), processing attributes is a large undertaking, but our paper provides a data processing pipeline template to support continuous development of customizable attribute sets.

As far as the number of catchment attributes, our goal was to provide an initial set that is representative of the dominant groups of attributes appearing in the literature, namely terrain, land cover, climate, and soil, and to provide complete code and tutorial-like instructions to offer a highly extensible and transparent data product. We believe this approach supports dataset extension and customization for specific needs across disciplines in light of the accelerating development of remote sensing data products. By providing the full code in addition to accompanying tutorials, we believe this dataset sets a precedent as far as being explicit about what attributes represent and how they are derived.

Corresponding Manuscript Edits:

- Lines 34-35 in the introduction have been revised to clarify the point about ungauged dataset as a basis of comparison for monitoring networks.
- A clearer gap statement between LSH datasets and hydrographic datasets has been added to the end of Section 1.1 to describe the need for the BCUB.
- Section 1.2 has been subdivided into hydrographic datasets (1.2.1) and LSH datasets (1.2.2). The new section 1.2.2 includes a concise description of key points in the evolution of LSH.
- A paragraph was added at the end of section 1.3 as follows: *"Our goal with the BCUB dataset was to provide a representative set of catchment attributes that cover key groups commonly found in the literature–terrain, land cover, climate, and soil. While our attribute set is not as extensive as those found in the LSH literature, we prioritized creating a transparent, extensible data product with complete code and tutorial-like information. Given the rapid development of attributes in LSH research, we focused on providing a solid framework rather than the most exhaustive or up-to-date set of attributes."*

RC3 Comment: *Why are streamflow not served here?*

Author's Response: The British Columbia Ungauged Basin dataset focuses on the much larger set of catchments that are unmonitored. The novelty of this dataset is in providing catchment attributes similar to those found in the LSH literature but for a much larger set of catchments where streamflow measurements have not been collected.

Corresponding Manuscript Edits:

- Splitting section 1.2 into two sub-sections (as described above) should hopefully make it more clear how this dataset is positioned in the gap statement, namely the closing statement of section 1.1 (~ line 60).

RC3 Comment*: The current introduction is lacking a comprehensive background on the state-of-the-art knowledge on other existing global-scale or continental-scale LSH datasets. For example, some of the cited literature in Line 12 is only the geospatial datasets but not the LSH datasets made available. I suggest the authors to do a more comprehensive literature summary, and place the BC work into better context of the community development.?*

Author's Response:
We agree that a clearer link between the availability of geospatial datasets and the emergence of LSH datasets should be made in the introduction. We focused the description of state of the art datasets on comparable hydrographic datasets for brevity, however since this dataset is intended to represent a bridge between the two, it is important to incorporate the LSH literature as the reviewer points out.

Corresponding Manuscript Edits:

- The first paragraph of the introduction has been changed to explicitly describe the link between geospatial (hydrographic) and LSH datasets.

References:

1. Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., Gonzalez, M. C., Ameli, A., and Poulin, A.: "A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds", Scientific Data, 7, 1-12, 2020.
2. Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., et al.: Caravan-A global community dataset for large-sample hydrology, Scientific Data, 10, 61, 2023.
3. Beck, H. E., De Roo, A., and van Dijk, A. I.: Global maps of streamflow characteristics based on observations from several thousand catchments, Journal of Hydrometeorology, 16, 1478–1501, 2015.
4. Addor, N., Nearing, G., Prieto, C., Newman, A., Le Vine, N., and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resources Research, 54, 8792–8812, 2018.395
5. Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2019). Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. Hydrological Sciences Journal, 65(5), 712–725. https://doi.org/10.1080/02626667.2019.1683182