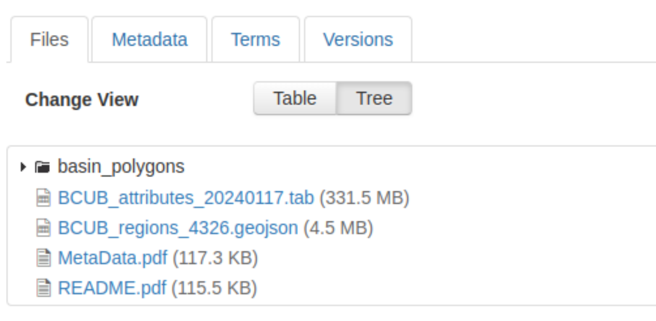
Author's Thanks (2024-05-21): We are grateful for the time and care taken by the reviewer to consider our work and provide feedback. Some of the points raised made us think about the process and the data in new ways and highlight where important clarifications should be made. Please see below for our responses to specific feedback.

> ***RC2 Comment:*** *It would be nice if authors can convince us the necessity to have 1.2 million basins (sub-basins or sub-catchment), it's very difficult to find the one you are interested and very difficult to use all of them in a regional scale.*

Author's Response: The smallest monitored sub-basin operated by the BC Hydrometric Service (and likewise in the HYSETS dataset) is 1 km^2. In developing the BCUB dataset, we aimed to cover the range of basin sizes described by the set of monitored watersheds in BC since these are widely used in research and practice. The reason for this is that we want to comprehensively characterize the ungauged space, i.e. the set of ungauged basins above the size threshold. This will serve further research in what we might be missing with the current monitoring network. We will aim to highlight this goal somewhat more in the introduction.

We agree, the large number of sub-basins does create challenges for working with the dataset as a whole. We provide an example of working with (parts of) the data in a web-based Jupyter Notebook: (https://dankovacek.github.io/bcub_demo/notebooks/7_Dataset_plot_example.html) In an effort to support different use cases, we provide the data in two formats: i) a smaller and widely used (csv) format containing all sub-basin attributes with x,y coordinates of the sub-basin pour point and polygon centroids, and ii) the much larger basin polygon files in (geospatial) Parquet format (saved under basin_polygons in the data repository).



The Parquet format is supported by GDAL as of version 3.5, so QGIS must be compiled with GDAL >= 3.5 which is not default in some environments.
Please see the following for information about versions and compatibility:
https://gis.stackexchange.com/questions/430973/importing-geoparquet-file-in-qgis

Reading/writing Parquet in R:
https://arrow.apache.org/docs/r/reference/read_parquet.html

Reading/writing Parquet in Python:
https://arrow.apache.org/docs/python/parquet.html

Parquet is also implemented in Julia, MATLAB, Rust, Go, Java, C++, and others:
https://arrow.apache.org/docs/

> ***RC2 Comment****: If possible, authors can introduce some specific implementations of these large-sample basins.*

Figures 9 and 10 in the manuscript give examples of specific questions that can be asked of this dataset. The ability to derive customized samples of basins by a wide range of characteristics may support experimental design, for example in temperature monitoring for quantifying the effect of land cover change on stream temperature. We are currently using the BCUB dataset as an input for a streamflow monitoring network optimization study. We will revise the manuscript to more fully describe these examples in the Usage Notes (section 3).

> ***RC2 Comment****: It is also very difficult to find real observation of river discharge to support further analysis.*

Author response: The HYSETS dataset (Arsenault, 2019) provides streamflow observation at a large set of monitored locations along with their catchment polygons and attributes, whereas the BCUB dataset defines and describes attributes for sub-basins at all river network confluences and does not contain streamflow since the vast majority of confluences are ungauged. These two datasets can be combined to extrapolate information from monitored to unmonitored catchments. It should be noted that some work is required to map locations between HYSETS and BCUB datasets, and many historical monitoring locations no longer exist and their location coordinates were recorded with varying degrees of precision.

> ***RC2 Comment****: Please check the unit of precipitation, e.g., 2028mm/day for gauge 1269663 must be wrong?*

Author response: Thank you for catching the typo, the precipitation index represents a mean annual value, which is derived from daily total precipitation from the DAYMET dataset. We have updated Table 2 to read "Mean Annual Precipitation" with the corresponding units [mm/year].

> ***RC2 Comment:*** *What is the difference in the number of detected sub-basins when using the two spatial resolution?*

Author response: The number of sub-basins is a function of the minimum area threshold assumed in the stream network extraction process, which we set at 1 km^2 to match the smallest sub-basin in the streamflow monitoring network. Our hypothesis is that the number of basins is not expected to change significantly as a result of a change in the input DEM resolution. The key factors are:

1. The number of raster cells (pixels) representing the smallest sub-basin (1 km^2) increases by roughly an order of magnitude between the EarthENV (90m at the equator, ~10^2 pixels) and the USGS 3DEP (30m at the equator, ~10^3 pixels), and
2. The raster pixel dimension changes as a function of latitude, meaning the precision of one (integer) number of pixels increases with increasing latitude.

The smallest number of upstream accumulation cells in the BCUB dataset is 1507 which is a result of the projected grid dimension decreasing with increasing latitude (30m raster pixel would yield a minimum threshold of 1111 pixels). In the coarser resolution, this would represent 170 pixels owing to the factor of 9 area difference between the 30 and 90m grid dimensions. This number of pixels represents the worst case scenario where rounding to an integer number of pixels represents 0.6% rounding error. The BCUB dataset contains pour points at confluences, but not headwaters, unless they coincide, so only pour points (confluences) within 0.6% difference from the area threshold would be affected (included/excluded). Approximately 0.25% of the basins in the BCUB are less than 1.01km^2 representing a 1% deviation from the minimum basin area.

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.