Comment for "Catchment Attributes and Meteorology for Large-Sample SPATially distributed analysis (CAMELS-SPAT): Streamflow observations, forcing data and geospatial data for hydrologic studies across North America" by Knoben et al.

General comments:

This study meaningfully contributes to an improvement and expansion of an important testbed for hydrologic modeling community, CAMELS, by enabling a spatial-distributed capacity. This is a significant undertaking which involves huge amount of work, including data selection, pre-processing, quality control, optimization, and validation. I can see the great potential and popularities of the resulting CAMELS-SPAT datasets being widely used in the hydrological modeling community as a testbed. I especially appreciate the following two points about this manuscript

- I appreciate authors combining methods and outcomes, which enhance the readability given the complexity of this manuscript.
- Kudos to authors about their effort in getting the details right, especially the differences in backward-looking and forward-looking timestamps for meteorological forcing datasets, which are usually overlooked.

However, I have noticed several caveats or limitations that I would like the clarifications from the authors.

- First, there is no single basins in Alaska, which is an important area for studying hydrologic change. I completely understand the data sparsity in Alaska, but I am still slightly surprising that there is no single catchment in Alaska that fit into the criteria for selection. Can the authors explain what is the biggest problems with the data in Alaska that they did not make it to the list?
- Since the authors use soil properties from SOILGRIDs, which means that the soil properties only available to the top 2 meters. For hydrologic modeling applications with deep soil columns, can the authors provide suggestions concerning the data source or assumptions of deeper-than-2-meter soil properties?
- I got slightly confused when I started reading Section 3, because I thought the catchment attributes, such as climatology, vegetation, land cover, etc., have already been discussed in Section 2.5. Why do we need another section to discuss it? After reading it through, apparently Section 3 provides a more in-depth analysis of the catchment attributes. However, I would still recommend the authors making some changes to help the readers understand the flow of this manuscript. A possible way to do so is to add a short description of paper structure at the end of Introduction.

Specific comments:
L45: Please cite the original VIC paper

- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research: Atmospheres, 99(D7), 14415-14428.

L62: It is unclear to me how to differentiate 1) sub-basin, 2) basin, and 3) catchment in this context. Is basin equivalent to catchment in this sentence? Do sub-basins mean delineation of entire basin or catchment to different vector-based hydrologic response units (HUCs)?

L70: "Native" was used to describe the data characteristics multiple times in this paragraph. I guess "native" would refer to that the authors keep the "original" spatial resolution of the source met forcing data or other geospatial data sets. I would recommend the authors add one sentence defining "native", deleting the parenthesis in lines 62, 70, 72.

I am not sure about it but the subsections 2.X.1 with title "context" seems a bit redundant. If you do not provide this subtitle, readers could still know this part is the context or background information for this section. The subsections "2.X.2" with title "Methods and outcomes" seems ok.

L102: Delete "for" in this sentence, i.e., "~~for~~ e.g. agricultural and industrial use, …"

L223: Kudos for accounting for daylight savings.

L229-233: Kudos for explaining what the timestamp represents.

L277: "Days" should be plural in the context, i.e., "between the preceding and following day**s**"

L287: What is the area of interest in this study? Do authors refer to all CAMELS-SPAT basins?

L317: What do "whole hours" mean? Please clarify

L389: Do you mean 5^{th} percentile, rather than 0.05^{th} percentile? Maybe a typo?

The title for Section 3 is confusing. I thought Section 2.5 already covers the catchment attributes.

L407-408: "Only" and "rarely" are redundant in this sentence, i.e., "…the existing literature **only rarely** uses catchment descriptors…" Please delete one of them.

L449. Please spell out "m.a.s.l.".

L454: Please use SI and avoid using hectare (ha).

L497: Does the author mean "skewness" values?

L587: Did the authors mean the selection of attributes **for data-driven models** in this context? In most process-based hydrologic models, the attributes that serve as input are usually pre-defined and not much selection should be required.