Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-67-RC1, 2020

Earth System Science Data Discussions

Interactive comment on "CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil" *by* Vinícius B. P. Chagas et al.

Wouter Knoben (Referee)

wouter.knoben@usask.ca

Received and published: 28 April 2020

Summary

The authors present a unified dataset for large-sample hydrology in Brazil. They provide raw streamflow data for 3713 gauges across Brazil, and catchment attributes for a subset of 897 catchments for which the authors consider the streamflow data to be of good quality. The catchment attributes are compiled from multiple sources and broadly cover topographic, climatic, hydrologic, land cover, geologic, soil and human intervention variables. The data are made publicly available through Zenodo.

First off, I would like to congratulate the authors on what must have been a massive amount of work. This looks like a tremendous dataset and it is clear that the authors must have spent considerable effort compiling it. From the success of the original CAMELS data and the more recent CAMELS-CL, it is clear that datasets such as this are very highly appreciated by the community. I think CAMELS-BR forms a strong addition to these two existing datasets. I have no major comments but compiled a list of smaller points that hopefully help the authors in clarifying a few things that are not entirely clear to me.

Review

L37: Suggest to change "validation" to "evaluation". See e.g. Oreskes et al. (1994)

L61: Is a word missing here? ". . . by institutions such as the . . ."

L77-78: How do these numbers relate to the 897 catchments in the title?

L90: What are the native and new file formats?

L101: As far as I can tell CAMELS and CAMELS-CL cover the period 1989 to 2009 (at least). Why was the year 1989 not included in CAMELS-BR?

L108: Is this a complete list of quality control checks that were performed? The text ". . . errors such as . . ." seems to imply that more checks were done but not listed here. A complete list of all quality control steps taken would be good (or rewriting the sentence without the words "such as" if the current list is already complete).

L117: To clarify, lines 117 to 121 only summarize the quality control by ANA? All 897 selected gauges have passed the authors' additional quality control described in lines 108-111?

L125: It might be good to expand the current meteorological indices with ae_mean (mean actual evaporation). This goes beyond what CAMELS and CAMELS-CL provide, but it might be a good way to remind the reader that actual evaporation data is

also provided.

L129: “no weight was applied if a cell is only partially covered by the catchment”. Does this mean that partially covered cells are not used for calculating the catchment average or that all cells contribute to the average equally, whether the catchment fully covers them or not? Why was this particular choice made and can this be justified in some way?

L130-134: I don’t fully understand the description of this limitation (maybe because of the previous comment). Does this mean that for some catchments and meteo products no data could be calculated?

L130-134: I think this limitation section can be stronger if the authors describe how they deal with this limitation during preparation of the data set.

L173-175: Are sine curves an appropriate representation of the temperature and precipitation regimes in Brazil? Was the accuracy of the calibrated sine curves comparable to the results in Berghuijs and Woods (2016)? Given how large the study area is, and that seasonality metrics tend to be somewhat specialized towards certain climate types, it might be useful to compute a few additional seasonality metrics (see e.g. Feng et al., 2019).

L207: is the Ladson digital filter the same approach as used in CAMELS and CAMELS-CL?

L378-388: It might be worthwhile to briefly discuss here what happens with consumptive water after it has been used. Does the predominantly evaporate/transpire or is it released back into the stream? In which way are the calculated streamflow indices affected by water use?

L408; “Lehner et al. (2011, Technical Documental)” Should this be “Technical Document”?

L441: “a new dataset comprising more than 3000 catchments in Brazil”. It would be

helpful to add a line here to clarify that there is a subset of 897 basins, and which kind of data and attributes are available for the 3000+ and the 897 set.

Table 1: there is some inconsistency between time periods for various forcing variables. For consistency with CAMELS and CAMELS-CL, it would be nice if all variables are provided for 1979-2009.

Table 5: is “bare_frac” the same variable as “barren_frac” in CAMELS-CL? If so, it would be good to stick with consistent naming.

Figures: it is a bit difficult to make out any details in the figure in the south-east region, where gauge density is high. It might be worthwhile to not scale the data points according to catchment size (although keeping this scaling in Figure 1 is quite informative) in the data plots.

Figures: a follow-up suggestion to the previous comment is to add histograms to each data plot that summarize the information on the map (as was done in the original CAMELS paper). This makes it easier for the reader to see how the catchment attributes vary across their respective ranges.

Figures: I’m not sure whether a diverging colour scheme is very appropriate for continuous variables that have no clear breakpoint in the middle of the range. For example in Fig. 3a, I don’t fully understand why catchments smaller than 5*10ˆ3 kmˆ2 are green and larger ones red. This implies some critical change between the smaller and larger catchments that I don’t think is there. A continuous color scheme (e.g. Fig. 4d) would be more appropriate. This applies to multiple figures. Note, in cases such as Fig. 4c I think a diverging colour scheme is justified, because this makes it easier to distinguish positive and negative values.

Figure 4b: Do no aridity index values exceed 1.2?

Figure 4c: If I remember this metric correctly, values of -0.5 and +0.5 should be equivalent. Why do these values exceed -0.8 And +0.8?

References

Berghuijs, W. R., & Woods, R. A. (2016). A simple framework to quantitatively describe monthly precipitation and temperature climatology. International Journal of Climatology, 36, 3161–3174. https://doi.org/10.1002/joc.4544

Feng, X., Thompson, S. E., Woods, R., & Porporato, A. (2019). Quantifying asynchronicity of precipitation and potential evapotranspiration in Mediterranean climates. Geophysical Research Letters.https://doi.org/10.1029/2019GL085653

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, 263(5147), 641–646. https://doi.org/10.1126/science.263.5147.641

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-67, 2020.