Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-521-RC1, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "CABra: a novel large-sample dataset for Brazilian catchments" *by* André Almagro et al.

Pedro Luiz Borges Chaffe (Referee)

pedro.chaffe@ufsc.br

Received and published: 9 December 2020

This manuscript provides a dataset of catchment attributes for several Brazilian catchments. The dataset (including climate, streamflow, groundwater, and others) was compiled from several data sources and most of the methods and limitations are discussed in the specific sections. The data set is made public through Zenodo. The authors put a lot of effort in delineating the major basins, providing meteorological datasets, calculating potential evaporation with three different methods, and developing a new hydrologic disturbance index. I agree with the authors that ". . .similarities [with CAMELS-BR (Chagas et al., 2020)] highlight nothing but the urgent need for the creation of such a database for Brazilian catchments". Overall the manuscript is well written and it is worthy of publication after minor revisions.

As major comments, I believe the authors could highlight those new products they produced and the novelty of the manuscript by: (1) framing CABra as a complementary dataset that builds on some opportunities left by the CAMELS-BR dataset (e.g., need for drainage density and uncertainty analysis); (2) making use of the CAMELS-BR for a quantitative comparison to show if or how the different approaches and data sources might influence signature values. I would also create a new section after the Introduction called "2 Motivation to extend the CAMELS-BR data set" similarly to Addor et al. (2017). This new section would be a more concise and accurate (see minor comments) replacement of section "3 Comparison with the CAMELS-BR. . ." of the current manuscript.

Please, find bellow some minor comments:

L27-28 – References are highlighted in gray.

L35-39 – Why isn't CAMELS-BR cited here with all the other CAMELS datasets?

L46-47 – "Additionally, there is no repository. . .". How about CAMELS-BR?

L49-51 – This sentence might seem contradictory as CAMELS-BR (and perhaps CABra) had already been developed.

L67-72 – If this paragraph corresponds to Fig.1, should it read "hydrologic disturbance" or "Hydrologic signature"?

L109 – "100 cell accumulating water" is equivalent to what area?

L115 – Can you specify somewhere what are the 132 catchments?

L118-121 – This is not accurate. Do et al. (2018) and Gudmundsson et al. (2018) methodology was based on the areas provided in ANA's dataset (the same you used to check the error). Even though they probably did not visually inspect the boundaries for all the basins they provided, in the CAMELS-BR every boundary from the Do et al. (2018) data set was visually inspected (this procedure may not be explicit in Chagas et

al., 2020). Please, check those references.

L129-130 – What do you mean by “necessitating a better understanding of hydrologic processes”?

L165-172 – I think this ensemble can be another nice product of your data set if better described. Do you think that an ensemble mean product is always “more reliable”? Could you compare and show the differences?

L217 – “estimative” I think it should be “estimation”.

L242-243 – “There are only a few... that have precipitation in the winter”. I think you mean “most of the precipitation”.

L244 – “Amazonian coastal area might not be obvious to the international readership. If this is really relevant, perhaps you could provide an indication in some figure.

L266-267 – What are those outliers? Are they relevant?

L280-282 – Many streamflow gauges have inconsistencies other than those typographical errors cited here. For example, there are abrupt changes resulting from changes in measurement instruments or rating curves, and unrealistic daily streamflow values (i.e., larger than 1000 mm d-1) (Chagas et al, 2020). Have you screened the time series for those inconsistencies? How do you think they would affect the hydrological signatures?

L298-299 – I might have missed in your paper how you defined a hydrologic year. Was it the same for the entire country?

L314-316 – I am not sure if “reaching infinity” is the best expression. I think it means we should not calculate the slope if the value in the denominator is zero.

L351-354 – It is nice that you provided HAND data. The “robust correlations” were found in this work or in Nobre et al.? If the latter, I think it is better to provide the reference again to avoid ambiguity.

L367-370 – You do not need this fractured rocks vs porous rocks discussion here.

L373-347 – You can also delete “This kind of... thereby forming rivers and lakes”.

L415 – “There is a spatial correlation...” Why is it so? Is it an underlying process or a feature of how the data source was produced?

L417 – “we have” should be “there is”.

L417-419 – “These characteristics... present the opposite”. I would delete it or rephrase it.

L478 – You use satellite observations of 2015 but calculate the signatures using 1980-2010 data. It could be argued that this land-cover data does not correspond to the same period. Can you provide the rationale behind your choice?

L480-489 – You should highlight more this NDVI product that you provide as it is a nice addition.

L496 – Perhaps it would be better to use “forest and grasslands”.

L508 – Is the MATOPIBA region really significant for discussion?

L509-511 – Is this a feature or a choice for the data set?

L521 – “Higher values were found in timing with...” – please rephrase it.

L539-540 – What is the data source for the reservoirs? The ANA (2017) reference was not provided in the reference list. In the CAMELS-BR the reservoir data from ANA (2018) and GRaND (Lehner et al., 2011) were combined and further checked against Pekel et al. (2016) in order to exclude very small (insignificant) reservoirs.

L542-543 – It is not clear the meaning and how you calculated “distance to the nearest urban area of each catchment”. Distance to the outlet? How do you define urban area? An isolated urban pixel is considered an urban area or spurious data? How can this distance affect the streamflow signatures?

L545-551 – While the Hydrologic Disturbance index is an interesting and novel product, you do not explain how you determined those coefficients in Eq.6. You should investigate how to evaluate the usefulness of the index against the hydrologic signatures that you calculated. What does it capture? Can we use it to somehow classify what you observed?

Section 3 Comparison with the CAMELS-BR and broader implications for hydrological studies – In some parts of this section you provided advantages of CABra over CAMELS-BR, but you did not point out the limitations of CABra. Therefore, I do not consider it to be a "comparison" section. I would suggest you to make this section more concise and to create a new section "2 Motivation to extend the CAMELS-BR data set" similarly to Addor et al. (2017).

L602-603 – Even though you did a better job than CAMELS-BR at delineating basin boundaries, I do not think that a 2% error compared to the ANA values is the best or most correct standard for two reasons: (1) the methodology used by ANA to calculate the areas is not provided; (2) the areas provided by ANA are many times rounded to the nearest hundred or thousand.

L606-609 – "while considering 20 hydrologic years" – should read 20 or more hydrologic years. Even though the attributes were calculated using 1990-2009 (for consistency with other CAMELS data set), data for 1980-2010 was also provided when available. You should also clearly explain that some of the non typographical error types that were checked in CAMELS-BR (e.g., abrupt change, zero in place of missing data etc) might not have been checked in CABra.

L613-620 – Xavier et al. (2016) is a great interpolated product that has been used extensively. However, you should clarify the limitations of choosing to use Xavier et al. (2016) as you cited in the climate section of your paper. Since that data is interpolated inside Brazil, you cannot use this rainfall data sets in basins such as Amazon, Paraguay and Parana. Besides, Xavier et al. (2016) used many rainfall gauges (which is great)

without checking the homogeneity of the data. This is not a criticism at all. I just want to point out that there are advantages and disadvantages for each choice along the way and here lies the opportunity for a comparison of how different are the attributes calculated based on different data sources.

L625-632 – I think this is a great point of your paper (PET calculations) that should be better highlighted. Could other climate indices that were not provided in CABra (timing, frequency, duration, etc) be relevant for hydrological analysis?

L633 – "attributes of from" delete "from"

L633-638 – Could the second geological class or the relative percentages be relevant for hydrological analysis?

L639-643 – Does a higher spatial resolution always mean a smaller classification error?

L644-650 – This is a very nice addition that was not covered in the CAMELS-BR and I think you should highlight it more. You should also try to better describe and test this index. See comments to L539-540 (I do not think ANA provides reservoir sizes) and L545-551.

TABLE 4 – What is the meaning and how did you calculate q_hf (Max streamflow frequency) and q_lf (Min streamflow frequency)?

FIGURE 2 – What are the shades of blue in (b)? "Km" should be in lower case.

FIGURE 3 – (a) it is difficult to differentiate the order of smaller catchments. In this case, I think it is better to use a point at the outlet of each catchment. Some of the opportunities to expand the Chagas et al. (2020) data set were to provide drainage data and you could use that to highlight this important part of your work. It is not clear that the x labels of the inset are the same as the colorbars, perhaps you should increase the tick marks. You chose to keep the Biome delineation in the background, is there a specific reason for that choice? It might not be obvious to the international readership.

References Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

ANA – Brazilian National Water Agency: Relatorio de Seguranca de Barragens 2017, 2018. Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D. and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth Syst. Sci. Data, 12(3), 2075–2096, doi:10.5194/essd-12-2075-2020, 2020.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787–804, https://doi.org/10.5194/essd-10-787-2018, 2018.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, https://doi.org/10.1890/100125, 2011.

Pekel, JF., Cottam, A., Gorelick, N. et al. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422 (2016).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-521, 2020.