Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-67-AC1, 2020

Open Access 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.

Vinícius B. P. Chagas et al.

vbchagas@gmail.com

Received and published: 16 July 2020

Reply to comments by Reviewer #1, Wouter Knoben.

We appreciate the helpful comments of Reviewer #1. The recommendations improved the clarity and the reproducibility of our work. Please, find below our reply to all the comments.

Comment #1: L37: Suggest to change "validation" to "evaluation". See e.g. Oreskes et al. (1994).

Reply: We have changed from "validation" to "evaluation" (line 38 in the track change

Printer-friendly version

Discussion paper

revised manuscript).

Comment #2: L61: Is a word missing here? “. . . by institutions such as the . . .”

Reply: We have changed from “by institutions as the” to “by institutions such as the” (line 62 in the track change revised manuscript).

Comment #3: L77-78: How do these numbers relate to the 897 catchments in the title?

Reply: The 3,097 catchments mentioned in the manuscript were incorrect. We modified the sentence to refer to the 897 catchments in the title: “It includes daily streamflow time series from 3,679 stream gauges and, for a selected group of 897 catchments, daily meteorological time series and 65 catchment attributes from properties such as topography, climate, land cover, geology, soil, and human intervention.” (lines 78-80 in the track change revised manuscript).

We removed from the dataset 34 catchments with available streamflow data, reducing the total number of gauges from 3,713 to 3,679. None of the 897 selected catchments were removed. Those 34 removed catchments are located outside Brazil and are not monitored by a Brazilian agency (such as the Brazilian National Water Agency), thus are outside the scope of CAMELS-BR. The total number of catchments were updated throughout the manuscript (lines 13, 89, 99, 103, and Table 1 of the track change revised manuscript).

Comment #4: L90: What are the native and new file formats?

Reply: We modified the sentence to specify the file formats: “Their values are unchanged but, to ease their processing, we converted the native files (i.e., Excel files with daily streamflows not disposed in chronological order) to a new file format (i.e., text files with daily streamflow in chronological order).” (lines 91-93 of the track change revised manuscript).

Comment #5: 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?

Reply: CAMELS covers the water years from 1990 to 2009, which corresponds to 1 October 1989 to 30 September 2009. CAMELS-BR also covers the water years from 1990 to 2009, which corresponds to 1 September 1989 to 31 August 2009. So, they cover essentially the same period. This information was not clear in the manuscript, so we modified the sentence to "Firstly, we selected only gauges that have less than 5 % of missing streamflow data between the water years 1990 (starting on September 1, 1989) and 2009 (ending on August 31, 2009)." (lines 104-105 of the track change revised manuscript).

Comment #6: 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).

Reply: The current list of quality control checks is already complete. To make it clearer, we substituted "for errors such as" by "for the following errors" (line 113 of the track change revised manuscript).

Comment #7: 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?

Reply: Yes, that summarizes only the quality control by ANA. To clarify, we modified the sentence to "To summarize the ANA metadata . . ." (line 123 of the track change revised manuscript).

All 897 selected gauges have passed our additional quality control. To clarify, we modified the first sentence of the paragraph to "We individually screened the 897 selected streamflow time series for the following errors: . . ." (line 113 of the track change revised manuscript).

Comment #8: 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.

Reply: We added the climatic attribute et_mean (mean actual evapotranspiration) to the database and included its description in Table 3. Since another attribute was also added to the database (see Comment #12), we updated the total number of attributes from 63 to 65 throughout the manuscript; changed from 11 to 13 climatic attributes in Table 1; and modified the sentence "We computed thirteen climatic indices" (line 182 of the track change revised manuscript).

Comment #9: 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?

Reply: We modified the sentence to clarify it: "The daily values represent the average of all cells with their centroids intersected by the catchment, of which all cells contribute to the average equally, whether the catchment fully covers them or not. However, some catchments do not intersect the centroid of any cell. For those, we computed the daily values as the average of all cells partially covered by the catchment." (lines 134-138 of the track change revised manuscript). We chose this method because it is the most used in most algorithms (Tem alguma referência disso?).

Comment #10: 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?

Reply: Meteorological products were computed for all catchments, regardless of their sizes. We clarified the limitation of computing meteorological variables for catchments

Printer-friendly version

Discussion paper

smaller than a single cell by adding the sentence “This leads to the assumption that such a meteorological variable is homogeneous in catchments smaller than a single cell, even though this might not always be the case.” (lines 140-141 of the track change revised manuscript).

Comment #11: 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.

Reply: Please refer to the reply to the two previous comments (Comment #10 and #11).

Comment #12: 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).

We thank the reviewer for pointing out the paper by Feng et al. We extracted the asynchronicity index proposed by Feng et al. (2019) for each catchment and added this new index to CAMELS-BR.

We also added to the manuscript the following information:

(i) A description of the asynchronicity index on Table 3: “Asynchronicity between the annual precipitation and PET cycles, where high values represent high relative magnitude and phase differences”.

(ii) “Those indices are complemented by the precipitation seasonality index (p_seasonality, Table 3), which relies on sine curves to approximate the monthly climatology of temperature and precipitation. While, for Brazil, the annual precipitation cycle is captured quite well, a sine curve provides a relatively rough approximation of the temperature cycle, particularly in the center of the country (around the state of

Goiás; Berghuijs and Woods, 2016). Hence, in addition to p_seasonality, we extracted the asynchronicity index proposed by Feng et al. (2019), which relies on information theory and has the advantage of being non-parametric (in particular, it does not assume sinusoidality).” (lines 190-196 of the track change revised manuscript).

(iii) “Northeastern Brazil (in particular, the states of Maranhão, Piauí, Ceará) has the highest values of asynchronicity index in the country (not shown), which corresponds to Mediterranean climates.” (lines 204-206 of the track change revised manuscript).

References:

Berghuijs, W. R. and Woods, R. A.: A simple framework to quantitatively describe monthly precipitation and temperature climatology, Int. J. Climatol., 36(9), 3161–3174, doi:10.1002/joc.4544, 2016.

Feng, X., Thompson, S. E., Woods, R. and Porporato, A.: Quantifying Asynchronicity of Precipitation and Potential Evapotranspiration in Mediterranean Climates, Geophys. Res. Lett., 46(24), 14692–14701, doi:10.1029/2019GL085653, 2019.

Comment #13: L207: is the Ladson digital filter the same approach as used in CAMELS and CAMELS-CL?

Reply: Yes. All hydrological indices were computed using the same approach as in CAMELS and CAMELS-CL, including the usage of the Ladson digital filter. To clarify this, we added the sentence earlier in the same paragraph: “The hydrological signatures were computed in the same approach as in CAMELS, CAMEL-CL, and CAMELS-GB datasets.” (lines 229-230 of the track change revised manuscript).

Comment #14: 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?

Reply: We added the definition of consumptive water use earlier in the section, when it is mentioned for the first time: “Consumptive water use refers to water withdrawals that do not return to the catchment, for example, by evaporating, transpiring, or being incorporated into manufactured products.” (lines 381-383 of the track change revised manuscript).

The streamflow indices are certainly affected by consumptive water use since it is an essential component of the water cycle (Milly et al., 2008; Hoekstra and Mekonnen, 2012). However, we do not know in which ways those indices are affected, as we do not know in exactly which ways the other ∼50 attributes affect the streamflow indices. This is an extensive and interesting topic of research (Montanari et al., 2012) and we hope that the CAMELS-BR dataset allows further investigations of what drives the hydrological behavior of catchments and in which way calculated streamflow indices are affected by water use.

References:

Hoekstra, A. Y. and Mekonnen, M. M.: The water footprint of humanity, Proceedings of the National Academy of Sciences, 109(9), 3232–3237, doi:10.1073/pnas.1109936109, 2012.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319(5863), 573–574, doi:10.1126/science.1151915, 2008.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z. and Belyaev, V.: “Panta RheiâĂŤEverything Flows”: Change in hydrology and societyâĂŤThe IAHS Scientific Decade 2013–2022, Hydrological Sciences Journal, 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

Comment #15: L408; "Lehner et al. (2011, Technical Documental)" Should this be "Technical Document"?

Reply: We changed from "Technical Documental" to "Technical Document" (line 438 in the track change revised manuscript).

Comment #16: 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.

Reply: We clarified by modifying the sentence to "Here, we introduced the CAMELS-BR, a new dataset comprising streamflow time series for 3,679 catchments in Brazil and, for a selected quality-controlled set of 897 catchments, meteorological time series and 65 catchment attributes." (lines 470-473 of the track change revised manuscript).

Comment #17: 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.

Reply: Consistency between CAMELS datasets refers to the water years from 1990 to 2009. All attributes in CAMELS-BR cover this period and every time series includes at least those 20 years of data.

We have modified Table 1 to clarify the coverage period of each data source. Each data source covers distinct and non-coincidental periods. If we restricted all meteorological variables to include only coincident time periods, it would reduce the coverage periods from 1980-2018 to 1981-2014. We believe that it is more beneficial for the users of the dataset to include the entire cover period that each data source provides.

Comment #18: 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.

Reply: Yes, both refer to the same variable. We renamed the variable from "bare_frac" to "barren_perc" because it refers to percentages instead of fractions (Table 5 of the revised manuscript), as recommended by Comment #26 (Reviewer #2).

Comment #19: 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.

Reply: We have tried your suggestion and it seems that not scaling the symbols with catchment area did not solve this problem. It seems that it also hinders the visualization in low gauge density regions (see Fig R1c, f, i). Although the southeastern region has a high gauge density, those are usually the smallest catchments in the country (Fig R1a). We believe that scaling the symbols with catchment size facilitates the data visualization, particularly in the southeast.

We tried to improve the visualization by decreasing the size of all symbols but keeping them scaling with catchment size. In this way, it enhanced the visualization of high gauge density regions without hindering the visualization of low gauge density regions (Fig R1a-b, d-e, g-h). The following Figures were modified: Fig 1, 4, 5, 6, 7, 8, 9, and 10.

Comment #20: 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.

Reply: We added a histogram to each map that shows a catchment attribute. The following figures were modified: Fig 1, 4, 5, 6, 7, 8, 9 and 10 in the revised manuscript.

Comment #21: 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*10ËĘ3 kmËĘ2 are green and larger ones red. This implies some critical change be-

Printer-friendly version

Discussion paper

tween 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.

Reply: We changed to a sequential color scheme all Figures that had diverging color schemes and no clear breakpoint. The following Figures were modified: Fig 1c-d; Fig 4a and 4c; Fig 6c-f; Fig 7f; Fig 8c-d and f-g; Fig 9a-d; Fig 10a-b (of the revised manuscript).

Additionally, to improve visualization, we changed the classes of the color schemes of Fig 1a and 1c; Fig 4a and 4c; Fig 6f; Fig 8d; and Fig 10a (of the revised manuscript) and included additional gauges that were missing in Fig 1a.

Comment #22: Figure 4b: Do no aridity index values exceed 1.2?

Reply: The catchments in the most arid parts of Brazil frequently exceed aridity index values of 1.2. The color class with the largest values refers to aridity greater than 1.0 but not limited to 1.2. We believe it is clearer with the addition of histograms to represent the range of values in the figure (Fig 5 of the revised manuscript).

Comment #23: 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?

Reply: We imagine that the reviewer refers to another metric, because this metric, computed using Eq. 14 in Woods (2009), typically takes values between -1 (precipitation out of phase with temperature) and 1 (precipitation in phase with temperature, i.e. simultaneous peaks), while values close to 0 indicate uniform precipitation throughout the year. Hence, values smaller than -0.8 or greater than 0.8 are possible, and values of -0.5 and +0.5 are not equivalent. We have clarified the description of this metric in Table 3, added the references to this equation, and added the equations used to compute other indices in Table 4.

References:

Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, Adv. Water Resour., 32, 1465–1481, https://doi.org/10.1016/j.advwatres.2009.06.011, 2009.

Reply to comments by Reviewer #2.

We appreciate the helpful comments of Reviewer #2. We have agreed to all recommendations. Please, find below our reply to all the comments.

Comment #24: To what extent do the ET estimates match P-Q when several years of data are available. This might be good to know, to get a first-order idea if the estimates seem somewhat reasonable.

Reply: To analyze to what extent ET estimates matches P-Q, we added a new scatterplot with the long-term water balance (Fig. 3a-b in the revised manuscript).

The following paragraph was added to describe the conclusions from the figure: "The long-term water balance is accurate for most catchments, using either the estimated evapotranspiration from GLEAM (Fig. 3a) or MGB (Fig. 3b). Both evapotranspiration data sources indicate that the highest data uncertainties occur in the Amazon and smaller catchments in the Paraná and the Southeastern Atlantic regions, since those catchments are further away from the 1:1 line in Fig. 3a-b. The same conclusions are derived from the runoff coefficient as a function of the humidity index (Fig. 3c). In addition, there are remarkable differences between GLEAM and MGB estimates, where evapotranspiration from GLEAM is substantially higher in the Amazon basin and substantially lower in the Eastern and the Western NE Atlantic regions." (lines 153-159 of the track change revised manuscript).

Comment #25: "The mean daily precipitation in Brazil is highest in the Amazon and in Southern Brazil, where it usually exceeds 5 mm day-1" I would replace "usually" to "on average" since the first is more often associated with a median than a mean.

Reply: We changed from "usually" to "on average" (line 202 in the track change revised manuscript).

Comment #26: Figures often refer to "fractions" (which suggest 0-1) when instead "percentages" are displayed. Either is fine, but it would be nice if the use was consistent.

Reply: We consistently modified all attribute names and values to refer to percentages instead of fractions. We changed the names (from "_frac" to "_perc") and the descriptions of attributes in the following: Tables 2, 5, 6, and 7; Figures 1, 7, 8, and 9; and line 123 of the track change revised manuscript.

Additionally, to maintain consistency across CAMELS datasets, we have changed the names of the following attributes: bedrock_depth, reservoirs_vol, regulation_degree, consumptive_use, and consumptive_use_perc (Tables 7 and 8).

Reply to comments by Reviewer #3, Thibault Mathevet.

We appreciate the helpful comments of Reviewer #3. The recommendations improved the clarity and the reproducibility of our work. Please, find below our reply to all the comments.

Comment #27: L227 : please clarify "The mean half-flow date".

Reply: We modified the sentence to "The mean half-flow date (i.e., when the cumulative discharge since 1st September reaches half of the annual discharge) . . ." (line 253 of the track change revised manuscript).

Comment #28: To illustrate §4, 5 & 9, I encourage the authors to add a figure with Turc-Mezentsev water balance representation, with the runoff coefficient (Q/P) as a function of the humidity index (P/PET) (897 watersheds, 1990-2009 period). This figure would give a good representation of the water balance variability of the datasets, and the impact of some major human influences or uncertainties in the climatic/streamflow observations.

Reply: To analyze the variability of the water balance, we added a new scatterplot with the runoff coefficient as a function of the humidity index (Fig. 3c in the revised manuscript).

The following paragraph was added to describe the conclusions from the figure (as mentioned in Comment #24 from Reviewer #2): “The long-term water balance is accurate for most catchments, using either the estimated evapotranspiration from GLEAM (Fig. 3a) or MGB (Fig. 3b). Both evapotranspiration data sources indicate that the highest data uncertainties occur in the Amazon and smaller catchments in the Paraná and the Southeastern Atlantic regions, since those catchments are further away from the 1:1 line in Fig. 3a-b. The same conclusions are derived from visualizing the runoff coefficient as a function of the humidity index (Fig. 3c). In addition, there are remarkable differences between GLEAM and MGB estimates, where evapotranspiration from GLEAM is substantially higher in the Amazon basin and substantially lower in the Eastern and the Western NE Atlantic regions.” (lines 153-159 of the track-change revised manuscript).

Comment #29: This datasets will probably be very usefull for Rainfall-Runoff model intercomparison studies (recently, Mathevet et al., 2020). In order to give a benchmark of hydrological model performances, I would encourage the authors to calibrate a commonly used conceptual Rainfall -Runoff model (such as GR4J model, freely available, Coron et al. 2017 or any other Rainfall-Runoff model). A very simple modeling framework might gives the expected level of model performances on this datasets and the spatial variability of model performances. Providing such a benchmark could slightly improve the paper.

Reply: We agree that providing a set of hydrological simulations to be used as a benchmark is very valuable. We will provide in the database the streamflow simulation for a set of approximately 500 of the catchments. Those simulations are extracted from Siqueira et al. (2018) that used a fully coupled hydrologic–hydrodynamic model (MGB; Modelo Hidrológico de Grandes Bacias) to the continental domain of South America.

While calibrating and analyzing a different rainfall-runoff model would be very valuable, we believe that we are already covering a lot of ground in this paper, by providing a wide range of time series and catchment attributes for a country in which such a dataset does not exist yet. The data processing to produce CAMELS-BR took more than two years and we are concerned that properly setting up another hydrological model (even a simple one) and analyzing its simulations for hundreds of catchments would add a further delay. It is our intention to keep adding to CAMELS-BR and we anticipate that hydrological simulations for these catchments using different models will be produced in the near future and be shared with the community.

References:

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., Paris, A., Calmant, S. and Collischonn, W.: Toward continental hydrologic–hydrodynamic modeling in South America, Hydrol. Earth Syst. Sci., 22(9), 4815–4842, doi:10.5194/hess-22-4815-2018, 2018.

Comment #30: Add the number of watershed represented in the Figure captions (such as indicated in Table 1).

Reply: We added the number of catchments in the captions of Figures 1, 4, 5, 6, 7, 8, 9, and 10 of the revised manuscript.

Comment #31: Is there a possibility to improve the density of watersheds in the western part of the country ? I understand that the spatial density of observations/stations is lower and that these stations might have been excluded for some reasons ? But, hypotheses of exclusion might be relaxed in regions where station density is lower, in order to have a more homogeneous spatial coverage of the county ?

Reply: Even if we use a more relaxed selection criterion, the gauge density increase would be noticeable only in a small portion of the western part of the country (see Fig. R2). If we consider all basins with 10 years of data and less than 5% missing (from

2000-2009), only the upper Paraguay basin would have increased gauge density (Fig. R2c). Changes in other regions, such as the Amazon, Tocantins-Araguaia, and lower Paraguay, would barely be noticeable.

Changing to a more relaxed selection criterion would remove the consistency among CAMELS-BR and the other CAMELS datasets. Additionally, reducing the coverage period would lead to substantial increases in the uncertainty of climatic and hydrological indices. We are aware that the users of the dataset might want to consider those additional catchments in further research, which is why we have included streamflow data from all 3679 catchments in Brazil in the manuscript.

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

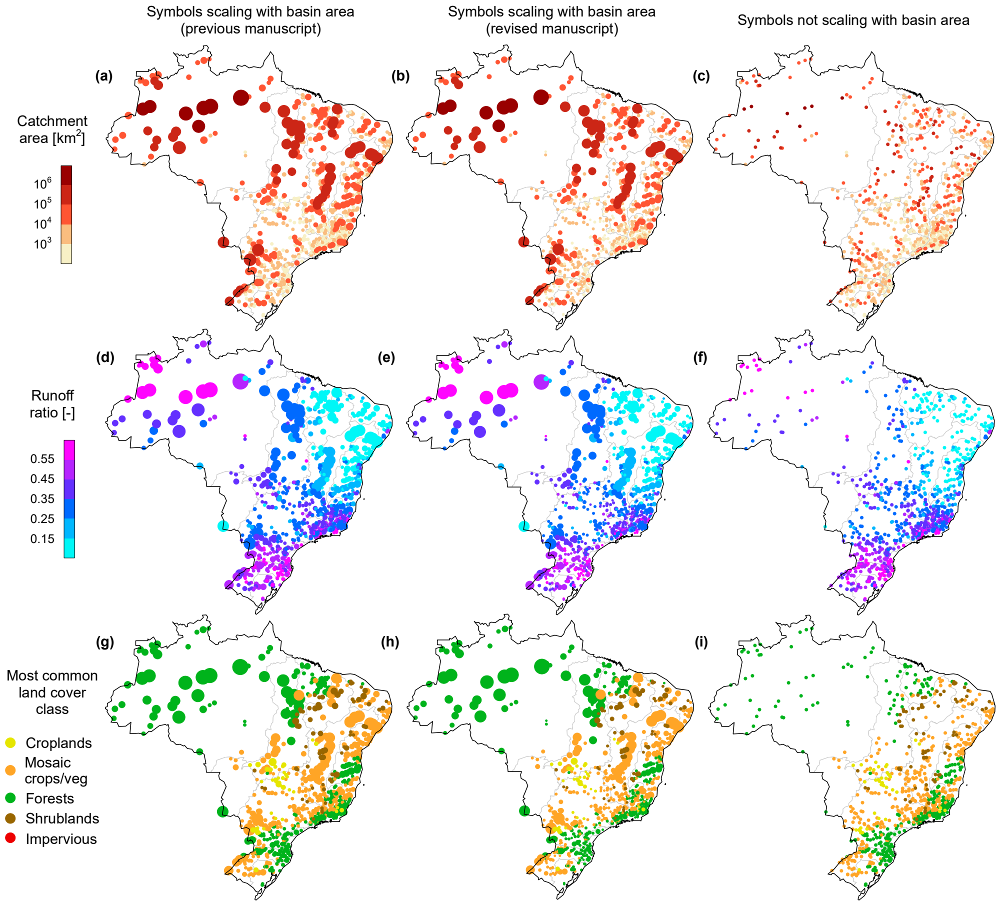

Fig. 1. Figure R1. Example of attributes with symbols scaling with catchment size in the previous manuscript, in the revised manuscript, and with symbols not scaling with catchment size.

Printer-friendly version

Discussion paper

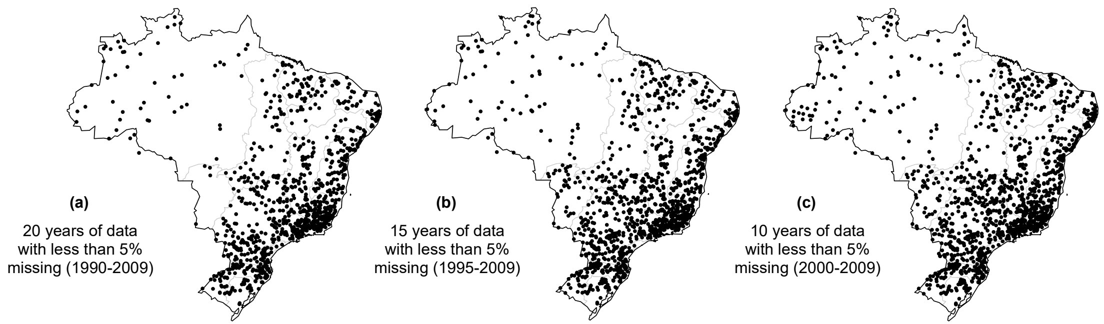

Fig. 2. Figure R2: Gauges selected using three different criteria.