

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.

For clarity, the authors' responses are inserted as blue text.

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 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.

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 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., Schaeffli, 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 Rhei—Everything Flows”: Change in

hydrology and society—The 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).

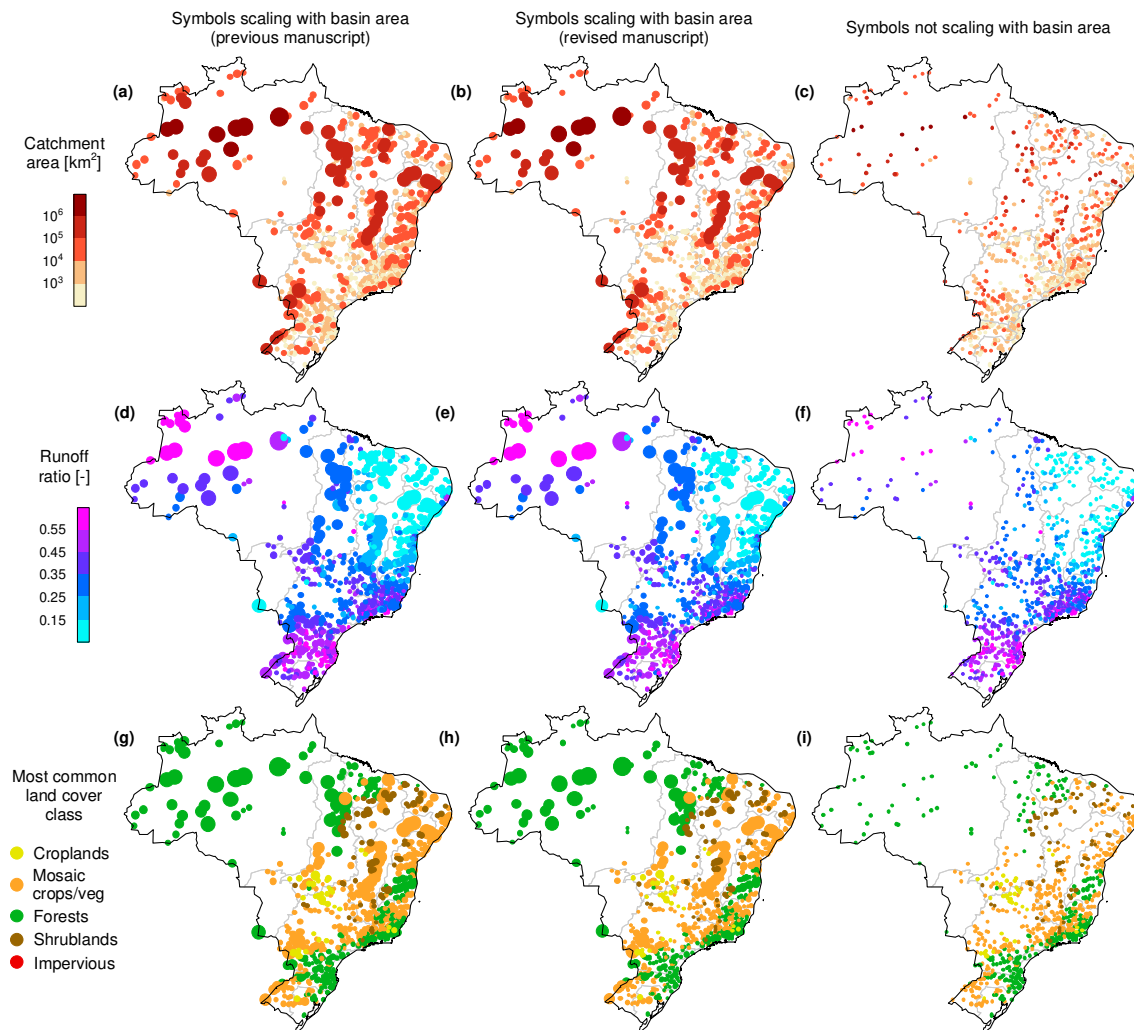
Comment #19: Figures: it is a bit difficult to make out any details in the figure in the south-east region, where gauge density is high. It might be worthwhile to not scale the data points according to catchment size (although keeping this scaling in Figure 1 is quite informative) in the data plots.

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.

Figure R1. Example of attributes with symbols scaling with catchment size in the previous manuscript (a, d, g), in the revised manuscript (b, e, h), and with symbols not scaling with catchment size (c, f, i).



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 $5 \times 10^3 \text{ km}^2$ 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.

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.

For clarity, the authors' responses are inserted as blue text.

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.

For clarity, the authors' responses are inserted as blue text.

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).

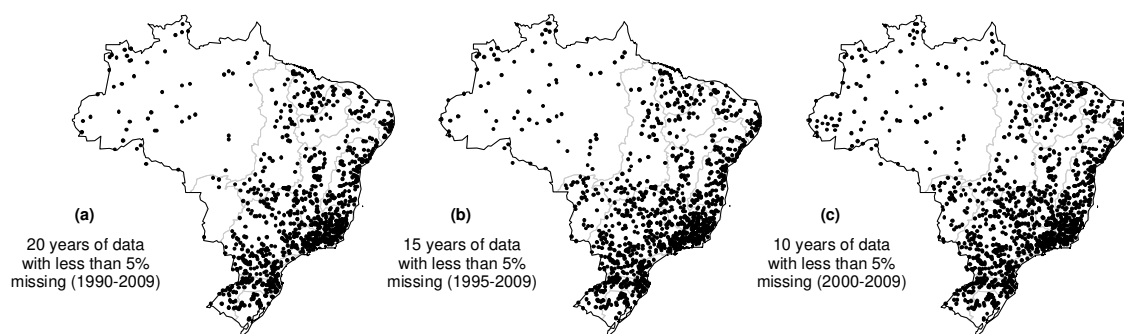
Reply: We added the number of catchments in the captions of Figures 1, 4, 5, 6, 7, 8, 9, and 10 of the revised manuscript.

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 (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.

Figure R2: Gauges selected using three different criteria.



CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil

Vinícius B. P. Chagas¹, Pedro L. B. Chaffe², Nans Addor³, Fernando M. Fan⁴, Ayan S. Fleischmann⁴, Rodrigo C. D. Paiva⁴, Vinícius A. Siqueira⁴

5 ¹Department of Sanitary and Environmental Engineering, Graduate Program of Environmental Engineering, Federal University of Santa Catarina–UFSC, Florianópolis, Brazil

²Department of Sanitary and Environmental Engineering, Federal University of Santa Catarina–UFSC, Florianópolis, Brazil

³~~Department of Geography, Geography, College of Life and Environmental Sciences,~~ University of Exeter, Exeter, United Kingdom

10 ⁴Hydraulic Research Institute, Federal University of Rio Grande do Sul-UFRGS, Porto Alegre, Brazil

Correspondence to: Pedro L. B. Chaffe (pedro.chaffe@ufsc.br)

Abstract We introduce a new catchment dataset for large-sample hydrological studies in Brazil. This dataset encompasses daily time series of observed streamflow from ~~3713–3679~~ gauges, as well as meteorological forcing (precipitation, evapotranspiration, and temperature) for 897 selected catchments. It also includes ~~63–65~~ attributes covering a range of

15 topographic, climatic, hydrologic, land cover, geologic, soil and human intervention variables, as well as data quality indicators. This paper describes how the hydrometeorological time series and attributes were produced, their primary limitations and their main spatial features. To facilitate comparisons with catchments from other countries, the data follow the same standards as the previous CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) datasets for the United States, Chile, and Great Britain. CAMELS-BR complements the other CAMELS datasets by providing data for

20 hundreds of catchments in the tropics and in the Amazon rainforest. Importantly, precipitation and evapotranspiration uncertainties are assessed using several gridded products and quantitative estimates of water consumption are provided to characterize human impacts on water resources. By extracting and combining data from these different data products and making CAMELS-BR publicly available, we aim to create new opportunities for hydrological research in Brazil and to facilitate the inclusion of Brazilian basins in continental to global large-sample studies. We envision that this dataset will

25 enable the community to gain new insights into the drivers of hydrological behavior, better characterize extreme hydroclimatic events, and explore the impacts of climate change and human activities on water resources in Brazil. The CAMELS-BR dataset is freely available at <https://doi.org/10.5281/zenodo.3709338> <https://doi.org/10.5281/zenodo.3709337> (Chagas et al., 2020).

1 Introduction

Large-scale hydrological research relies on data from large samples of catchments to formulate general conclusions on

30 hydrological processes and models (Gupta et al., 2014; Addor et al., 2019). Hydrometeorological datasets with large spatial and temporal coverage are the basis to improve hydrological understanding with appropriate statistical robustness. For

example, multiple studies used large-sample datasets to investigate the drivers of hydrological change (e.g., Slater et al., 2015; Blöschl et al., 2019a; Gudmundsson et al., 2019), the impacts of anthropic activities on the water cycle (e.g., Milliman et al., 2008; Hoekstra and Mekonnen, 2012; Montanari et al., 2013), hydrological similarity and classification (e.g., Berghuijs et al., 2014; Sawicz et al., 2014; Knoben et al., 2018), predictions in ungauged basins (e.g., Yadav et al., 2007; Ehret et al., 2014; Singh et al., 2014), areas where extreme events are a concern (e.g., Van Lanen et al., 2013; Villarini, 2016; Woldemeskel and Sharma, 2016), and to predict future hydrological change (e.g., Luke et al., 2017; Zscheischler et al., 2018). Moreover, large-sample hydrology is needed for ~~validation~~-evaluation of continental to global hydrological models, to identify limitations in model structure, parameterization and forcing according to geographic and climatic regions (Haddeland et al., 2011; Gudmundsson et al., 2012; Beck et al., 2017a; Zhao et al., 2017; Siqueira et al., 2018; Veldkamp et al., 2018), to estimate uncertainty in model estimates (e.g., Müller Schmied et al., 2014; Beck et al., 2016; Hirpa et al., 2018) and to make use of data assimilation techniques (e.g., Wongchuig et al., 2019). Better predictions in such models allow for the quantification of water resources availability over large scales and are fundamental for nationwide water resources planning and management (Schewe et al., 2014; Bierkens, 2015; Doll et al., 2016; Alfieri et al., 2020).

To uncover the hydrological functioning of a catchment, it is key to understand how it is controlled by climate, human interferences (Wohl et al., 2012; Di Baldassarre et al., 2018), and landscape attributes, such as vegetation, topography, soil, and lithology (Fan et al., 2019). For this, researchers must work with a multiple-hypotheses framework (Merz et al., 2012; Pfister and Kirchner, 2017), which frequently leads to processing massive amounts of data and often tedious, repetitive tasks. To deal with this problem, multiple datasets have been created, such as the Global Runoff Data Centre (GRDC, 2019), the Global Streamflow Indices and Metadata Archive (GSIM; Do et al., 2018; Gudmundsson et al., 2018); the HydroATLAS (Linke et al., 2019); and the Global Runoff Reconstruction (GRUN; Ghiggi et al., 2019). A noteworthy dataset, the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS; Newman et al., 2015; Addor et al., 2017), produced a synthesis from multiple catchment attributes. It initially included only catchments in the United States, but later expanded to include Chile with the CAMELS-CL dataset (Alvarez-Garretón et al., 2018) and Great Britain with CAMELS-GB (Coxon et al. (submitted)). The CAMELS datasets facilitated hydrological research by addressing some of the major problems with large datasets, such as a lack of common standards across databases, absence of uncertainty estimation, and open accessibility of hydrological observations (Addor et al., 2019).

Even though there is a growing number of large-sample datasets worldwide (e.g., Addor et al., 2017; Do et al., 2018; Gudmundsson et al., 2018; Ghiggi et al., 2019; Linke et al., 2019), the access to open and readily available data in some regions like South America is still difficult and requires additional quality checks (Crochemore et al., 2019). Particularly in Brazil, large-sample hydrological studies lack a comprehensive dataset to rely on. Brazilian hydrometeorological information is currently collected, maintained and distributed by institutions such as the Brazilian National Water Agency (ANA – *Agência Nacional de Águas*; <http://www.snirh.gov.br/hidroweb/>) and the National Institute of Meteorology (INMET; <http://www.inmet.gov.br/portal/>). The creation of ANA in 2000 led to the release of open hydrological information, which prompted the growth of hydrological studies and fostered water resources management. However, the use of data provided for

Brazilian catchments is challenging because (i) it requires either manual data acquisition of one station at a time through the institutions' local repositories (e.g., ANA, 2019a) or web scraping techniques to access these data in an automated fashion, (ii) there is little consistency in data format across regions and stations, and (iii) current datasets do not systematically provide catchment attributes characterizing the hydroclimate, landscape, and anthropogenic influences. Further, the difficulty of
70 accessing national meteorological daily time series has led users to compute them from other gridded global databases (e.g., Xavier et al., 2016; Beck et al., 2017c; Sun et al., 2018). All these difficulties hinder large-sample hydrological studies in Brazil where, unsurprisingly, nationwide studies (e.g., Siqueira et al., 2018; Bartiko et al., 2019) are less common than in North America or Europe. Consequently, studies in Brazil generally include only a reduced number of stream gauges and catchment attributes, and are restricted to specific regions, such as the Amazon (e.g., Tomasella et al., 2011; Paiva et al., 2013; Latrubesse
75 et al., 2017; Levy et al., 2018), or the La Plata basin (e.g., Collischonn et al., 2001; Pasquini and Depetris, 2007; Melo et al., 2016; Lima et al., 2017; Chagas and Chaffe, 2018).

To overcome these limitations, we produced and made publicly available a new dataset for large-sample hydrological studies in Brazil, CAMELS-BR. It includes daily streamflow time series from ~~3,713~~3,679 stream gauges ~~and, for a selected group of 897 catchments,~~ daily meteorological time series ~~for 3,097 catchments,~~ and ~~63-65~~ catchment attributes from properties such
80 as topography, climate, land cover, geology, soil, and human intervention. All catchment attributes and time series are in an easily readable file format and on a quickly accessible database. We follow standards defined by the previous CAMELS and CAMELS-CL datasets, thus allowing direct comparisons with them. Most attributes rely on data products that cover the whole of South America, so they are spatially consistent across Brazil. To reduce the risk of data misinterpretation, we describe the major limitations of the data sources and indices computed. By synthesizing hydrological information from thousands of
85 catchments in Brazil into a single dataset, we allow researchers to skip the arduous task of collecting and preprocessing large quantities of disparate data.

2 Hydrometeorological daily time series

2.1 Streamflow data

We provide daily streamflow time series for two sets of gauges (Table 1). The first set comprises ~~3713~~3679 streamflow gauges
90 and is provided by the Brazilian National Water Agency (ANA, 2019a). We refer to these as “raw streamflow” time series, as they are readily available from ANA (2019a). 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).~~
~~we provide these time series in a different file format to ease their processing.~~
ANA estimates daily streamflow either by (i) taking two daily stream stage measurements, one in the morning (at 7 am) and
95 another in the afternoon (at 5 pm), which are averaged and transformed into discharge using a stage-discharge relationship (rating curve); or (ii) resorting to regionalization methods when no stream stage measurements are available (no further details on the methods are provided by ANA). The raw streamflow time series cover different periods, ranging from a few days to

more than a century. Additionally, although ANA performs data quality checks, these time series include inconsistencies such as typographical errors and days with missing data. The ~~3713-3679~~ gauges are irregularly distributed throughout the country (Fig. 1a). Overall, their spatial distribution is denser and their time series longer in the Southern Atlantic, Southeastern Atlantic, and Paraná hydrographic regions (Fig. 1a and 1b).

The second set of streamflow time series includes 897 gauges, and here we simply refer to them as “streamflow” time series (Table 1). This is the set of gauges used to compute the catchment attributes. It is a subset from the previous ~~3713-3679~~ gauges, which resulted from two selection criteria. Firstly, we selected only gauges that have less than 5 % of missing streamflow data between the water years ~~(starting 1st September)~~ 1990 ~~(starting on September 1, 1989)~~ and 2009 ~~(ending on August 31, 2009)~~. We chose the water years from 1990 to 2009 because (i) it is the period with the largest number of stream gauges with available data (Fig. 2), and (ii) it coincides with the period of analysis from other CAMELS datasets (Addor et al., 2017; Alvarez-Garretón et al., 2018), allowing for direct comparisons with them. Secondly, we only considered catchments for which boundaries have been delimited by Do et al. (2018) and for which there is a good match with the area estimated by the data provider (see Sect. 3). Although the hydrological signatures introduced below were computed using data from 1990 to 2009, the time series for the 897 stream gauges include data from 1980 to 2018 when available, to enable complementary analyses by other users.

We individually screened the ~~897 selected~~ streamflow time series between 1990 and 2009 for ~~the following errors: such as~~ zeroes or repeated values instead of missing values, abrupt changes resulting from changes in measurement instruments or rating curves, annual streamflow larger than annual precipitation, and unrealistic daily streamflow values (i.e., larger than 1,000 mm day⁻¹). Gauges affected by such errors were not included in the set of 897 catchments. In addition, we summarized the streamflow metadata provided by ANA as follows. For each daily streamflow measurement, we provide two pieces of information (Table 1). The first metadata variable, “qual_control_by_ana”, was set to 1 if the data was quality checked by ANA and to 0 otherwise. The second metadata variable, “qual_flag”, indicates the reliability of streamflow estimates. It is also provided by ANA and consists of the following quality flags: 0, when there is no description; 1, streamflow resulted from stream stage measurements and the rating curve; 2, streamflow estimated by ANA without stream stage measurements; 3, streamflow values marked as doubtful; and 4, when the stream water level falls outside the range of the stream stage, e.g., when the river ran dry. To summarize the ~~ANA~~ metadata (i.e., q_qual_control_ ~~frae-perc~~ and q_stream_stage_ ~~fraeperc~~; Table 2), 80 % of the 897 gauges had at least 90 % of their data over 1990-2009 checked for inconsistencies (Fig. 1c). The Amazon, São Francisco, and Paraná regions have the lowest frequency of quality controls in Brazil. Furthermore, the streamflow estimates from 64 % of the 897 catchments were derived from stream stage measurements for 90 % of the days over 1990-2009 (Fig. 1d).

2.2 Meteorological data

Meteorological daily time series data are provided for 897 catchments (Table 1). These include (i) precipitation from CHIRPS v2.0 (Funk et al., 2015), CPC (NOAA, 2019a), and MSWEP v2.2 (Beck et al., 2019); (ii) potential evapotranspiration from

GLEAM v3.3a (Miralles et al., 2011; Martens et al., 2017); (iii) actual evapotranspiration from GLEAM v3.3a and MGB South America (Siqueira et al., 2018); and (iv) minimum, maximum, and average temperature from CPC (NOAA, 2019b). The datasets were selected because of their high spatial resolution, their full coverage of South America allowing for consistency through all catchments, and because they are commonly used which enables comparisons with other studies. 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. (no weight was applied if a cell is only partially covered by the catchment). 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. A significant limitation of the meteorological data is that, because the cell grids of the adopted products have resolutions range from 0.05° (ca. 5.5 km² at the equator) to 0.5° (ca. 55 km² at the equator), some catchments are smaller than a single cell. 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. This limitation has to be kept in mind particularly when using the CPC precipitation data (resolution of 0.5°; NOAA, 2019), as precipitation is the meteorological variable with the highest spatial heterogeneity amongst those used in CAMELS-BR.

In addition to GLEAM v3.3a, estimates of actual evapotranspiration (ET) were obtained from the MGB model version for South America (Siqueira et al., 2018). The MGB is a conceptual, semi-distributed hydrologic-hydrodynamic model that discretizes the basin (or a set of basins) into irregular unit-catchments and further into hydrological response units by combinations of land use and soil types, where both water and energy balance are computed. The model calculates ET using the Penman-Monteith equation based on CRU meteorological data (i.e., temperature, pressure, radiation, and wind speed) and MSWEP v1.1 precipitation data (Beck et al., 2017b). Surface resistance is adjusted according to the availability of water in the soil that is updated during the water budget. The MGB also computes the evaporation of flooded areas and intercepted water from the canopy with the Penman equation. Regular ET cells of 0.5° resolution were generated by aggregating unit-catchments using their areas as weights.

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.

3 Topographic indices

Even though ANA (2019a) provides estimates of the areas of most gauged Brazilian catchments, the catchment boundaries are not publicly available. Hence, in this study, we used the catchment boundaries provided by Do et al. (2018), who used the

HydroSHEDS 15 arc-sec resolution digital elevation model (DEM) and delineated the catchments with a procedure similar to Lehner (2012) for more than 3,000 gauges in Brazil. For each streamflow gauge, Do et al. (2018) positioned the outlet at the center of all the DEM grid cells within a radius of 5 km from the gauge coordinates indicated by the metadata. They then selected the grid cell (and associated catchment boundaries) leading to the catchment area most similar to the one indicated by ANA (2019a). The main limitation of the procedure of Do et al. (2018) is that catchment boundaries were not manually inspected.

Using those catchment boundaries, we computed four topographic attributes (Table 2), namely gauge elevation, catchment mean elevation, mean slope, and area. The area of the catchments ranged from 10.8 km² (i.e., in the upper São Francisco hydrographic region) to 4.7 million km² (i.e., the Amazon basin at Óbidos). Approximately 30 % of the analyzed catchments are smaller than a thousand km², 43 % are between 1 and 10 thousand km², and 27 % are larger than 10 thousand km². The largest basins are in the Amazon and in the Tocantins-Araguaia hydrographic regions (Fig. 3Fig. 4a). Combined with the Paraguay basin, those regions are usually characterized by low elevations (Fig. 3Fig. 4b), flat slopes (Fig. 3Fig. 4c), and large proportions of wetlands (see Sect. 6.2). The smaller catchments are located along the mountain belts on the eastern coast of Brazil, particularly in the southern and southeastern parts of the country. Those are also the catchments with the steepest slopes. Additionally, many catchments with intermediate elevation ranges (i.e., between 500 and 900 m) are in the central part of the country, which comprises the Brazilian highlands. Note that, since we computed the average attribute value (unless otherwise noted) of each catchment, the attributes become less representative as the area of the catchment increases.

4 Climatic indices

4.1 Data and methods

We computed ~~the same eleven~~ thirteen climatic indices (Table 3) over the same period (1990 to 2009, except for the asynchronicity index) as the ones in CAMELS (Addor et al., 2017) and CAMELS-Chile (Alvarez-Garreton et al., 2018). The first water year starts on 1st September 1989 and the last one finishes on 31st August 2009. This is to facilitate inter-dataset comparability. We used precipitation data from CHIRPS v2.0 (Funk et al., 2015) to compute the indices, since it has the highest spatial resolution among the three adopted precipitation products (i.e., CHIRPS v2.0, CPC, and MSWEP v2.2) and relies on both remote-sensing and gauge-based data.

The mean precipitation, mean potential evapotranspiration, and the aridity index are considered to capture long-term climatic conditions. The aridity index is the ratio of mean potential evapotranspiration to mean precipitation, which stands as a first-order control on the partitioning of precipitation into streamflow (Budyko et al., 1974; Blöschl et al., 2013). 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

195 ~~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). computed by fitting the seasonal precipitation and temperature using sine curves representing the seasonal amplitude and time of the year when most of the precipitation occurs (Woods, 2009).~~ The indices of extreme climatic conditions include the frequency, duration, and the most common season of high precipitation events and dry days. Dry days are defined as days with precipitation less than 1 mm, so that the index is not compromised by underdetected
200 precipitation events (Haylock and Nicholls, 2000).

4.2 Spatial variability in climatic indices

The mean daily precipitation in Brazil is highest in the Amazon and in Southern Brazil, where it ~~usually on average~~ exceeds 5 mm day⁻¹ (1825 mm year⁻¹) (~~Fig. 4~~Fig. 5a). The lowest mean precipitation occurs in Northeastern Brazil, which is also where mean potential evapotranspiration exceeds the mean precipitation (aridity index > 1, ~~Fig. 4~~Fig. 5b). ~~Northeastern Brazil (in particular, the states of Maranhão, Piauí, Ceará) also has the highest values of asynchronicity index in the country (not shown), which corresponds to Mediterranean climates.~~ The precipitation regime is highly seasonal in most of the country, particularly in the central-west and southeastern Brazil (~~Fig. 4~~Fig. 5c). This seasonality is regulated by the South American Monsoon System (Raia and Cavalcanti, 2008; Carvalho et al., 2011), with peaks in the austral summer (~~Fig. 4~~Fig. 5f) and several dry months during the austral winter (~~Fig. 4~~Fig. 5i). Southern Brazil has a distinct regime, with ~~a~~ uniform precipitation throughout
210 the year caused by a combination of large-scale phenomena and a diversity of sources of atmospheric moisture (Seager et al., 2010; Martinez and Dominguez, 2014). The Amazon basin, which extends into both hemispheres, has contrasting precipitation regimes between the north (with a peak in austral winter) and the south (with a peak in austral summer) related to alternating warming of each hemisphere (Marengo and Espinoza, 2016). This seasonality is substantially diminished downstream in the Amazon.

215 The number of high precipitation and dry days is highest along the catchments on the coast (~~Fig. 4~~Fig. 5d and ~~4g~~5g), which is also where the smallest catchments are located. Both indices are significantly correlated with catchment area (p-value < 0.001), so a regional analysis of both indices should be carried out with caution since large catchments are located in the Amazon and Tocantins-Araguaia basins. On the other hand, the duration of high precipitation (~~Fig. 4~~Fig. 5e) and dry day events (~~Fig. 4~~Fig. 5h) do not correlate with catchment area. Their spatial distribution is remarkably similar to the aridity index, except for the
220 Tocantins-Araguaia basin, which has long dry periods but not necessarily long high precipitation events. Summer is the most common season of extreme precipitations in the majority of Brazil, with two main exceptions (~~Fig. 4~~Fig. 5f): (i) part of the coast of Northern Brazil; and (ii) Southern Brazil. This is possibly linked to mesoscale convective systems over Southeastern South America (Salio et al., 2007), to sea surface temperature anomalies in the Atlantic ocean (Liebman et al., 2010), and ~~to~~
225 2013).

5 Hydrological signatures

5.1 Data and methods

We computed thirteen hydrological signatures (Table 4) that represent a wide range of hydrological information for the water years from 1990 to 2009. The hydrological signatures were computed in the same approach as in CAMELS, CAMEL-CL, and CAMELS-GB datasets. Intermediate streamflow conditions were evaluated with the mean daily flow and its ratio to mean daily precipitation. These were complemented by baseflow information, a fundamental component that sustains streamflow during dry periods (Smakhtin, 2001). The baseflow index is the ratio of long-term baseflow to long-term total streamflow. We used the digital filter from Ladson et al. (2013) to separate the baseflow component from the hydrograph. The variability of streamflow was evaluated with the slope of the flow duration curve and the streamflow elasticity indices. The slope of the flow duration curve is defined as the slope between the log-transformed 33rd and 66th long-term percentiles of daily streamflow (Yadav et al., 2007; Sawicz et al., 2011). High values of that index suggest highly variable streamflow, caused either by a high seasonality of streamflow or by a flashy response to precipitation events (Yokoo and Sivapalan, 2011; McMillan et al., 2017). Streamflow elasticity is an indicator of the sensitivity of mean annual flow to changes in mean annual precipitation (Sankarasubramanian et al., 2001). For example, a streamflow elasticity value of 2 indicates that a 1 % change in mean annual precipitation generates a 2 % change in mean annual flow. Extreme streamflow conditions were analyzed using signatures based on the magnitude, frequency, and duration of high and low flow events. High and low flow events were defined through long-term thresholds, based on the median and mean flow, respectively (Olden and Poff, 2003). The magnitude of high and low flow events was characterized using the 5th and the 95th percentiles. There are two primary limitations to the hydrological signatures used in this study. First, several signatures might scale with catchment area. Since catchment area varies substantially among hydrographic regions, spatial analyses should be carefully conducted. Second, we did not check for temporal dependencies of consecutive high or low flow events, for example when two flood peaks occur within a couple of days from each other and both may be related to a single extreme precipitation event. Many criteria exist to identify independent high flow events (Hall et al., 2014; Archfield et al., 2016) and low flow events (Fleig et al., 2006; Van Loon, 2015), which might lead to differences in the analyzed signatures.

5.2 Spatial variability in hydrological signatures

The spatial distribution of mean daily flows (Fig. 5Fig. 6a) and runoff ratio (Fig. 5Fig. 6b) closely resembles that of mean daily precipitation. These are notably high in Southern Brazil and in parts of the Amazon, and low in Northeastern Brazil. The mean half-flow date (i.e., when the cumulative discharge since 1st September reaches half of the annual discharge) follows a gradient ranging from February and March in the Eastern Atlantic region to May in the Amazon and on the northern coast (Fig. 5Fig. 6c). Steep slopes of the flow duration curve occur especially in the tributaries of Southern Amazon, the Tocantins-Araguaia basin, the Eastern Atlantic hydrographic region and in parts of Southern Brazil (Fig. 5Fig. 6d). Some catchments have undefined values, meaning that they have zero flow for more than 33 % of the time. Since the slopes of the flow duration

curve indicate the overall streamflow variability, they are spatially similar to several other hydrological signatures. They are, most noticeably: (i) negatively correlated with the baseflow index (~~Fig. 5~~Fig. 6e), hence catchments with high baseflow may be highly resilient to dry periods (Fan, 2015); (ii) positively correlated with streamflow precipitation elasticity (~~Fig. 5~~Fig. 6f), which indicates variability at the interannual timescale; (iii) negatively correlated with the 5th percentile of streamflow (i.e., low flows; ~~Fig. 5~~Fig. 6l); and (iv) positively correlated with the frequency and duration of low flow events (~~Fig. 5~~Fig. 6j and ~~5k~~6k). However, note that some regions do not follow those patterns. In particular, catchments in Southern Amazon and in the Tocantins-Araguaia basin have high baseflow indices despite steep slopes of the flow duration curve. It possibly implies that the variability in those catchments is related to a high seasonality, rather than to a flashy response to precipitation events. High flow days are more frequent and their events are longer in Southern Brazil, in the Eastern Atlantic region, and on the coast of northeastern Brazil (~~Fig. 5~~Fig. 6g and ~~5h~~6h). Those regions also have the most frequent and longest low flow events. This suggests that, in addition to the catchments in Southwestern Amazon and in the Tocantins-Araguaia basin, the extremes of both high and low flows might be related. Catchments seldom have high values in all three high flow signatures, except for Southern Brazil, revealing that this might be the region with the most problematic flood episodes. On the other hand, long and frequent low flows are found in nearly all hydrographic regions. The majority of catchments in the Eastern Atlantic and the East Northeastern Atlantic regions are characterized by long and frequent low flows, where nearly half of those have at least 100 days of low flows in the year.

6 Land cover characteristics

6.1 Data and methods

Each catchment was described using ten land cover classes (Table 5) based on GlobCover2009 (Arino et al., 2012). GlobCover2009 uses imagery from Envisat's Medium Resolution Imaging Spectrometer Instrument Fine Resolution (MERIS FR). The classification has a spatial resolution of 300 m, a global coverage every three days, and is based on images from January until December 2009. GlobCover2009 classification includes 22 land cover classes but, to simplify the dataset, we combined these into similar classes. In particular, the class "forests" is a combination of broad-leaved and needle-leaved forests, either evergreen or deciduous. Note that GlobCover 2009 does not differentiate between natural or planted forests. There are three main limitations of the GlobCover2009 dataset. High confusion rates between croplands and grasslands show that the separation of crops, pastures, and meadows can be problematic, especially in Brazil where those land covers occur extensively. Identification of wetlands is also an issue (Arino et al., 2012) and flooded forests might be underrepresented in the classification. Lastly, GlobCover2009 used the Shuttle Radar Topography Mission (SRTM) Water Body Data, which is based on data from 2000 and does not coincide with the 2009 MERIS data.

6.2 Spatial variability in land cover characteristics

Croplands are widespread in Brazil, especially in the highlands, in Southern Brazil, and on the eastern coast of Northeastern Brazil (~~Fig. 6~~[Fig. 7a](#) and ~~6d~~[7d](#)). Out of the 897 CAMELS-BR catchments, 52.4 % have croplands or mosaics of croplands and natural vegetation as the dominating land cover (~~Fig. 6~~[Fig. 7c](#)). Croplands are most noticeable particularly in the Uruguay and Paraná hydrographic regions. Even though GlobCover2009 does not cover the same period as the hydrological signatures (i.e., 1990-2009), croplands were already extensive in almost all states in Brazil in the 1980s and pastures in the 1960s (Leite et al., 2012; Dias et al., 2016). This is true except for Southern Amazon, where agricultural expansion has led to one of the highest deforestation rates in the world since the 1980s (Song et al., 2018).

Aside from the Amazon, catchments dominated by forests are located in mountain belts, i.e., in steep slope regions in Southern and Southeastern Brazil (~~Fig. 6~~[Fig. 7b](#)). Shrublands occur mainly in the driest regions of the country (~~Fig. 6~~[Fig. 7e](#)), but they are not the predominant land cover in these regions. Natural wetlands or water bodies are largely present in the Amazon, Tocantins-Araguaia, and Paraguay hydrographic regions (not shown). Some catchments in the Paraná, Uruguay, and São Francisco basins are also substantially covered by water bodies. However, those are mainly artificial reservoirs (see Sect. 9.3).

The CAMELS-BR catchments typically have a low fraction of their area considered to be “impervious areas”, such as urban land covers; only 0.2 % of the catchments have more than 5 % of impervious areas (not shown). ~~In addition~~[Besides](#), grasslands, bare soil areas, and permanent snow are rare in the CAMELS-BR catchments (not shown).

7 Geologic characteristics

7.1 Data and methods

The geology of the catchments was described using seven geologic attributes (Table 6). The first and second most common geologic class, their fractions, and the percentage of the catchment covered by carbonate rocks were extracted from the Global Lithological Map (GLiM; Hartmann and Moosdorf, 2012). GLiM was created by assembling information from 92 regional lithological maps. In the Brazilian territory, it relies on data from the Brazilian Geological Survey at the 1:1 million scale (Schobbenhaus et al., 2004). We considered only the first level of the GLiM geologic classes, which classifies lithology into 16 groups. The additional second and third levels provide more specific geologic information but were not included in this study. We note that two geologic classes cover a particularly broad variety of rocks. First, the “unconsolidated sediments” class is quite unspecific with regards to the sediment types and grain sizes (it includes sediments originated by areas as alluvial, swamp, and dune deposits). Second, catchments dominated by the “metamorphic rocks” class can have a wide range of lithologies, from shales to gneiss and quartzite.

We extracted the subsurface permeability and porosity indices from the GLoBal HYdrogeology MaPS 2.0 (GLHYMPS; Gleeson et al., 2014; Huscroft et al., 2018), which is modeled based on information from the GLiM and the Global Unconsolidated Sediments Map (GUM; Börker et al., 2018). Subsurface permeability indicates how easily water can flow

through the subsurface. GLHYMPS modeled it only for saturated conditions (Huscroft et al., 2018), so it is not adequate to characterize regions dominated by unsaturated processes, e.g., deeply weathered soils. The subsurface porosity indicates the fraction of void spaces in a material and controls the water storage capacity in the subsurface. A major caveat of GLHYMPS data is that it is only adequate for analyses at the regional scale, i.e., over spatial units greater than 5 km (Gleeson et al., 2014).

7.2 Spatial variability in geological characteristics

The catchments on the eastern coast have lithologies dominated by either metamorphic or acid plutonic rocks (Fig. 7a and 7b), related to high elevation and steep slopes in this region. These catchments also have low subsurface permeability (Fig. 7g) and the lowest subsurface porosity rates in the country (Fig. 7f). In Southern Brazil, basic volcanic lithology is widespread, which encompasses basaltic rocks. Southern Brazil has the most homogeneous lithological types of the country (Fig. 7c and 7d), where more than 80 % of the catchment areas are usually characterized by a single lithological type. However, subsuperficial porosity and permeability are highly heterogeneous, extending from middle-range to high porosity values and from middle-range to low permeability values.

Sedimentary rocks occur on a large scale at São Francisco, Parnaíba, Western Northeast Atlantic, and part of the Amazon hydrographic regions. The Northern Amazon is characterized mostly by metamorphic or plutonic rocks, while the Western Amazon has either siliciclastic or mixed sedimentary lithologies. On the other hand, unconsolidated sedimentary lithologies occur particularly downstream in the Amazon, the Tocantins, and the Paraguay basins. These basins also have flat slopes and large proportions of wetlands, which allows for alluvial particles to settle down. Most of the catchments with a high proportion of sedimentary rocks have high subsurface porosity, although their permeability varies according to the grain sizes of these rocks. Carbonate sedimentary rocks, such as karst or limestone, are more common in the São Francisco basin and in western Amazon (Fig. 7e). Those rocks are also present in some isolated and smaller catchments in Southern Brazil, in the Paraguay, and in the Tocantins-Araguaia basins.

8 Soil characteristics

8.1 Data and methods

We provide six soil characteristics (Table 7). Five of those were extracted from SoilGrids250m (Hengl et al., 2017; Shangguan et al., 2017), a collection of soil maps for the world at the 250 m resolution. SoilGrids250m maps are the result of a model using approximately 150,000 soil profiles, with predictions based on machine learning methods and 158 remote-sensing covariates including climate, vegetation, geomorphology, and lithology (Hengl et al., 2017). Although SoilGrids250m generated predictions for several soil depths, in this work we only computed soil characteristics over a depth of 30 cm. SoilGrids250m is based on a machine-learning model that explains large proportions of the variance of most observed variables, including 69 % of the variance of organic carbon content and more than 70 % of the soil textures (i.e., clay, silt, and sand content).

The soil characteristics might be highly correlated with other attributes from CAMELS-BR since they are modeled based on climatic and landscape covariates. Organic carbon content and clay content have modeled depth to bedrock as a predominant variable (Hengl et al., 2017). Other variables are also important, such as temperature and geomorphological characteristics (e.g., surface slope). The predictions of sand content are based primarily on depth to bedrock and precipitation, both at similar weights. Out of the five variables considered from the SoilGrids250m, predictions of depth to bedrock is the most problematic, with 59 % of its variance explained by the model (Shangguan et al., 2017). It has precipitation as the predominant covariate, which accounts for the control of weathering rates and soil production. Other decisive covariates are vegetation dynamics and geomorphological characteristics, which accounts for factors such as soil erosion (Shangguan et al., 2017).

The sixth soil characteristic is the water table depth, based on a 1 km resolution global model by Fan et al. (2013). Combined with depth to bedrock, water table depth can be an indicator of water storage potential in the catchment, which is related to baseflow and the supply of water for the vegetation during dry periods (Fan et al., 2013, 2019). The most important variables in the predictions of the water table depth of that model are, in decreasing order of importance, surface slope, elevation, precipitation, and temperature (Fan et al., 2013). Note that groundwater abstractions are not represented in the model, so water table depth data must be used with caution when analyzing catchments with intense anthropogenic intervention.

8.2 Spatial variability in soil characteristics

Soil texture in CAMELS-BR is characterized by (i) a predominance of clay content in Southern Brazil, in parts of Southeastern Brazil, particularly in higher elevations, and in northeastern Amazon (~~Fig. 8~~[Fig. 9c](#)); (ii) similar values of clay, sand, and silt content in the southern tributaries of Western Amazon (~~Fig. 8~~[Fig. 9a to 8c](#)); and (iii) a wide predominance of sand content in the rest of the country (~~Fig. 8~~[Fig. 9a](#)). As expected, the aridity index is closely related ~~with~~[to](#) the spatial distribution of the soil texture, since climatic attributes are important covariates in SoilGrids250m predictions (Hengl et al., 2017). The predominance of clay in Southern Brazil and in part of Southeastern Brazil might be linked to their lithological classes, i.e., with basic volcanic rocks in the former and acid plutonic rocks in the latter, since they have coincidental spatial distributions.

Organic carbon content is most pronounced in parts of the Amazon and in regions with high clay content (~~Fig. 8~~[Fig. 9d](#)). The depth to bedrock is higher in Central and Northeastern Brazil, frequently above 30 m (~~Fig. 8~~[Fig. 9e](#)). On the other hand, only 14 % of the catchments have depths to bedrock lower than 10 m, all of them located in Southern Brazil. Regarding water table depths, there is a clear gradient with higher depths on the eastern coast of Brazil (i.e., exceeding 30 m deep) to lower depths towards the Amazon (i.e., less than 10 m deep). Amongst all six soil characteristics indices, water table depth has the lowest correspondence to climate, i.e., to mean precipitation and aridity index. It is mostly correlated with catchment slopes, as previously indicated by Fan et al. (2013).

9 Human intervention indices

9.1 Data and methods for consumptive water use

380 We computed four indices of human intervention in the catchments (Table 8). Two are the total consumptive water use in the
catchment in 2017, one normalized by catchment areas and another normalized by mean annual streamflow. 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. The water uses are based on the Manual of Consumptive Water Use in Brazil (ANA,
2019c), which estimated the monthly water use of each municipality in Brazil. These estimates are the sum of water demands
385 from six categories:

(i) Irrigation: demand based on water balance models that estimate the quantity of water needed by irrigated crops but not
supplied by precipitation or soil moisture ANA (2019c). The spatial extent of irrigated croplands was characterized using the
national censuses of agriculture (e.g. IBGE, 2007) and remote-sensing images from Landsat, Sentinel-2, and Moderate-
390 Resolution Imaging Spectroradiometer (MODIS) (ANA, 2019b).

(ii) Livestock: demand estimated by multiplying the number of livestock units with their corresponding daily drinking
requirements. The number and type of livestock of each municipality were mapped on the national censuses of agriculture.

(iii) Households: demand estimated by multiplying the number of people in a municipality by their per capita domestic water
use.

395 (iv) Industry: demand estimated by multiplying the number of employees in several industrial categories from each
municipality by its per capita water use.

(v) Mining: demand estimated by combining the water use coefficient with the annual production of several types of mineral
extraction.

(vi) Thermoelectricity: demand estimated by applying a water use coefficient to the annual electricity production of each
400 thermoelectric plant in the country.

These water demand estimates do not differentiate surface water from groundwater. Even though groundwater abstraction is
extensive in the eastern part of Brazil (Fan et al., 2013), it is estimated that most of the water use in South America comes
from surface water (Wada et al., 2014). To estimate the consumptive water use of each catchment, we divided the values of
405 each municipality by its area. We assumed the water use to be spatially homogeneous throughout the municipality territory
and transferred the data for each municipality onto a 500 m spatial resolution raster.

There are three major limitations of using ANA (2019c) estimated consumptive water use. First, evaporation from artificial
reservoirs was not included in the computation. Thus, water use might be underestimated, particularly in the northeastern part
of Brazil, i.e., in the driest part of the country. Second, the dataset comes on an irregular grid, since municipalities areas vary
410 significantly. The smallest municipalities are usually within 500 km of the coast and their areas are mostly a few hundred km².

In contrast, the western part of Brazil and the Amazon usually have municipalities larger than a thousand km². Hence, consumptive water use of small catchments in the western part of Brazil should be interpreted with caution because they are smaller than the input data. The third limitation is that the consumptive water use of South America outside Brazil was not estimated by ANA (2019c) and was not considered in this study. This affects particularly the basins in the Amazon since they cover large parts out of Brazil. That said, anthropic intervention in these basins is low: only three basins with international borders in the Amazon are more than 10 % covered by croplands or croplands and natural vegetation mosaic, and none has more than 0.05 % of impervious land covers such as urban areas.

9.2 Data and methods for reservoirs

The other two indices for human intervention are related to flow regulation (Table 8), i.e., the sum of the total storage capacity of all reservoirs in the catchment and its ratio to the total annual flow of the catchment (i.e., the degree of regulation). We worked with estimated storage capacities from 1406 reservoirs in South America. The reservoirs were mapped by combining three data sources: (i) the Global Reservoirs and Dam database v1.3 (GRanD; Lehner et al., 2011); (ii) the hydroelectric power plants database of the National Electrical System Operator (ONS, 2019); and (iii) the 2017 National Dam Safety Report (ANA, 2018) database. The GRanD database includes reservoirs throughout South America, while the other two provided data only for Brazil. The procedure for combining the three databases was:

(i) We included all reservoirs from GRanD v1.3 in South America.

(ii) For each GRanD reservoir, we visually compared the inundated area with the one indicated by the polygons from the water bodies maps from Pekel et al. (2016). When the inundated areas differed substantially, we substituted the former with the latter and updated the size of the inundated area.

(iii) Out of more than 24,000 reservoirs from ONS (2019) and ANA (2018) databases, we included only those that have their inundated areas (Pekel et al., 2016) visible at the 1:500,000 scale. Although our goal was to only include reservoirs larger than approximately 0.5 km², some smaller reservoirs were also included. We computed the size of the inundated areas of those reservoirs according to the polygons from Pekel et al. (2016).

(iv) To check for duplicates in the databases, we manually inspected all dam points and their inundated areas.

(v) Finally, the storage capacities of reservoirs updated in step (ii) or included in step (iii) were recalculated using their inundated areas and, when available, information on dam height. We applied two equations determined by Lehner et al. (2011, Technical Documenta) with a statistical regression using data from 5824 reservoirs worldwide. When information on dam height was available, we applied Eq. (1):

$$V = 0.678 (A h)^{0.9229}, \quad (1)$$

where V is the reservoir storage capacity in 10⁶ m³; A is the size of the inundated area in km²; and h is dam height in m. When information on dam height was not available, we used Eq. (2):

$$V = 30.684 A^{0.9578}$$

(2)

9.3 Spatial variability in human intervention indices

445 The spatial distribution of human interference indices reveals that, unlike the catchments in the original CAMELS for the United States, catchments in CAMELS-BR can be significantly impacted by human activities. There are 17.8 % of catchments with annual consumptive water uses greater than 5 % of the mean annual flow. Those are principally in the driest parts of the country, i.e., in the São Francisco, Eastern Atlantic, Eastern Northeast Atlantic, and upper Paraná hydrographic regions (Fig. [109b](#)). Nevertheless, water uses greater than 20 % of the mean annual flow are rare, occurring in only 3.9 % of the catchments.

450 The similarity encountered between arid climates and high consumptive water uses may be attributed to two main causes. First, in the most arid catchments, the mean annual flow is typically a third of that of the rest of the country, which, unsurprisingly, leads to higher water uses proportional to the annual flow. Second, crops in drier climates require frequent irrigation and considerable rates of water withdrawal. On the other hand, we observe that the central and southeastern regions of Brazil have the greatest values of water uses normalized by catchment area (Fig. [109a](#)). Catchments in those regions are

455 commonly occupied by either irrigated croplands or populous metropolitan areas, which are respectively the first and second categories with the highest water demands in Brazil (ANA, 2019c).

The degree of regulation is related to catchment area (Fig. [109c](#)), meaning that the most regulated basins are downstream in the river basins. The main rivers with high regulations are the Paraná, Uruguay, São Francisco, Tocantins-Araguaia, Parnaíba, and Paraíba do Sul rivers. In those regions, 19.2 % and 7.2 % of the catchments have a degree of regulation greater than 10 %

460 and 50 %, respectively. These values nearly double in the driest regions of the country (i.e., the Eastern Atlantic, São Francisco, Eastern Northeast Atlantic, and Parnaíba hydrographic regions): 37.6 % and 22.1 % of the catchments have a degree of regulation greater than 10 % and 50 %, respectively. Therefore, the driest catchments of CAMELS-BR dataset have the highest human intervention rates, both in terms of consumptive water use and reservoir regulation

10 Data availability

465 The CAMELS-BR dataset is freely available at <https://doi.org/10.5281/zenodo.3709338> <https://doi.org/10.5281/zenodo.3709337> (Chagas et al., 2020). The files provided are (i) the ~~63-65~~ attributes in a zip file, (ii) the daily time series in zip files, (iii) the catchment boundaries used to compute the attributes and extract the time series, computed by Do et al. (2018) and Gudmundsson et al. (2018), and (iv) a readme file.

11 Conclusions

470 So far, large-sample hydrological studies in Brazil lacked a comprehensive and easily accessible dataset. Here, we introduced the CAMELS-BR, a new dataset comprising ~~streamflow time series for more than 3,000~~[3,679](#) catchments in Brazil ~~and, for a~~

~~selected quality-controlled set of 897 catchments. The data provided include daily streamflow and~~ meteorological time series and 63–65 catchment attributes. The attributes cover a wide range of fundamental properties for large-sample hydrological research, such as topography, land cover, geology, soil, and human intervention characteristics. We strived to make CAMELS-BR as comparable as possible to the other CAMELS datasets (Addor et al., 2017; Alvarez-Garreton et al., 2018) by using common naming conventions, scripts, and datasets. We also discuss the major limitations of the data to limit the risk of misinterpretation and misuse.

Even though CAMELS-BR is a step forward for hydrological research in Brazil, there are several opportunities for expanding the dataset in the future. For example, future versions of CAMELS-BR could include additional catchment attributes critical to understand hydrological processes, such as drainage density and basin morphometry (Shen et al., 2017). Further, an updated version should better characterize heterogeneities within each catchment, both for the time series and attributes. Additionally, since data uncertainties are omnipresent (Montanari, 2007; Blöschl et al., 2019b; Addor et al., 2019), they should be further explored by including additional data sources.

By simplifying the access to hydrological data, we aim to encourage further large-sample hydrological studies in Brazil, to facilitate the inclusion of Brazilian catchments in global large-sample studies, and to increase the transparency and reproducibility of these studies. We believe the data introduced here will, in particular, prove useful to explore the drivers of catchment behavior, to anticipate hydrological changes, and to study the impacts of human activities on the water cycle. We see CAMELS-BR as a resource designed to serve the broad water science community and to help with water resources management at regional, national, and continental scales.

Author contribution

PC and VC initiated the investigation. VC, PC, NA, FF, AF, RP, and VS designed the study. VC processed the data and created the figures. VC and NA computed the catchment attributes. VC prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

VC would like to thank CNPq (Brazilian National Council for Scientific and Technological Development) for the scholarship. NA acknowledges support from the Swiss National Sciences Foundation (P400P2_180791).

References

- 500 Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences*, 21, 5293–5313, 2017.
- Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K. and Mendoza, P. A.: Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges, *Hydrological Sciences Journal*, 1–14, doi:10.1080/02626667.2019.1683182, 2019.
- 505 Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X*, 6, 100049, 2020.
- Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, C., Cortes, G., Garreaud, R., McPhee, J. and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, *Hydrology and Earth System Sciences*, 22, 5817–5846, 2018.
- ANA - Brazilian National Water Agency: Relatorio de Seguranca de Barragens 2017, 2018.
- 510 ANA - Brazilian National Water Agency, HIDROWEB: <http://www.snirh.gov.br/hidroweb>, last access 15 June 2019, 2019a.
- ANA - Brazilian National Water Agency: Levantamento Da Agricultura Irrigada Por Pivôs Centrais No Brasil (1985-2017), 2a edição, 2019b.
- ANA - Brazilian National Water Agency: Manual De Usos Consuntivos Da Água No Brasil, 2019c.
- 515 Archfield, S. A., Hirsch, R. M., Viglione, A. and Blöschl, G.: Fragmented patterns of flood change across the United States, *Geophys. Res. Lett.*, 43(19), 10,232–10,239, doi:10.1002/2016GL070590, 2016.
- Arino, O., Perez, J. J. R., Kalogirou, V., Bontemps, S., Defourny, P. and Bogaert, E. V.: Global Land Cover Map for 2009 (GlobCover 2009), PANGAEA., 2012.
- Bartiko, D., Oliveira, D. Y., Bonumá, N. B. and Chaffe, P. L. B.: Spatial and seasonal patterns of flood change across Brazil, *Hydrological Sciences Journal*, 64(9), 1071–1079, doi:10.1080/02626667.2019.1619081, 2019.
- 520 Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, 52(5), 3599–3622, 2016.
- Beck, H. E., van Dijk, A. I., de Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. Global evaluation of runoff from ten state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, 21(6), 2881–2903, 2017a.
- 525 Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, *Hydrol. Earth Syst. Sci.*, 21, 589–615, <https://doi.org/10.5194/hess-21-589-2017>, 2017b.
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J. and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, *Hydrol. Earth Syst. Sci.*, 21(12), 6201–6217, doi:10.5194/hess-21-6201-2017, 2017c.

- 530 Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., McVicar, T. R. and Adler, R. F.: MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment, *Bull. Amer. Meteor. Soc.*, 100(3), 473–500, doi:10.1175/BAMS-D-17-0138.1, 2019.
- Bierkens, M. F. Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51(7), 4923–4947, 2015.
- 535 Berghuijs, W. R., Sivapalan, M., Woods, R. A. and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, *Water Resour. Res.*, 50(7), 5638–5661, doi:10.1002/2014WR015692, 2014.
- 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.
- 540 Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T. and Viglione, A. (Eds.): *Runoff prediction in ungauged basins: synthesis across processes, places and scales*, Cambridge University Press, 2013.
- 545 Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K. and Živković, N.: Changing climate both increases and decreases European river floods, *Nature*, doi:10.1038/s41586-019-1495-6, 2019a.
- 550 Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chiffard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., 555 Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbuloglu, E., Jarihani, B., et al.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, *Hydrological Sciences Journal*, 64(10), 1141–1158, 560 doi:10.1080/02626667.2019.1620507, 2019b.
- Börker, J., Hartmann, J., Amann, T. and Romero-Mujalli, G.: *Terrestrial Sediments of the Earth: Development of a Global Unconsolidated Sediments Map Database (GUM)*, *Geochem. Geophys. Geosyst.*, 19(4), 997–1024, doi:10.1002/2017GC007273, 2018.
- Budyko, M. I., Miller, D. H. and Miller, D. H.: *Climate and life*, Academic press New York, 1974.
- 565 Carvalho, L. M. V., Jones, C., Silva, A. E., Liebmann, B. and Silva Dias, P. L.: The South American Monsoon System and the 1970s climate transition, *Int. J. Climatol.*, 31(8), 1248–1256, doi:10.1002/joc.2147, 2011.
- Chagas, V. B. P. and Chaffe, P. L. B.: The Role of Land Cover in the Propagation of Rainfall Into Streamflow Trends, *Water Resour. Res.*, 54(9), 5986–6004, doi:10.1029/2018WR022947, 2018.

- 570 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 - link to files, Zenodo, <http://doi.org/10.5281/zenodo.3709338>, 2020.
- Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: groupings based on covariance, Journal of Hydrology, 237(3–4), 184–197, doi:10.1016/S0022-1694(00)00306-1, 2000.
- 575 Collischonn, W., Tucci, C. E. M. and Clarke, R. T.: Further evidence of changes in the hydrological regime of the River Paraguay: part of a wider phenomenon of climate change?, Journal of Hydrology, 245(1–4), 218–238, doi:10.1016/S0022-1694(01)00348-1, 2001.
- Coxon, G., N. Addor, J. Bloomfield, J. Freer, M. Fry, J. Hannaford, N. Howden, C. Jackson, E. Robinson, T. Wagener and R. Woods: CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, submitted to Earth System Science Data.
- 580 Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., & Arheimer, B. Lessons learnt from checking the quality of openly accessible river flow data worldwide. Hydrological Sciences Journal, 1-13, 2019.
- Di Baldassarre, G., Wanders, N., AghaKouchak, A., Kuil, L., Rangelcroft, S., Veldkamp, T. I. E., Garcia, M., van Oel, P. R., Breinl, K. and Van Loon, A. F.: Water shortages worsened by reservoir effects, Nat Sustain, 1(11), 617–622, doi:10.1038/s41893-018-0159-0, 2018.
- 585 Dias, L. C. P., Pimenta, F. M., Santos, A. B., Costa, M. H. and Ladle, R. J.: Patterns of land use, extensification, and intensification of Brazilian agriculture, Glob Change Biol, 22(8), 2887–2903, doi:10.1111/gcb.13314, 2016.
- 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(2), 765–785, doi:10.5194/essd-10-765-2018, 2018.
- 590 Döll, P., Douville, H., Güntner, A., Schmied, H. M., & Wada, Y.: Modelling freshwater resources at the global scale: challenges and prospects. Surveys in Geophysics, 37(2), 195–221, 2016.
- Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., Gelfan, A. N., Harman, C., Kleidon, A., Bogaard, T. A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M. F. P., Di Baldassarre, G., Parajka, J., van Beek, L. P. H., van Griensven, A., Westhoff, M. C. and Winsemius, H. C.: Advancing catchment hydrology to deal with predictions under change, Hydrol. Earth Syst. Sci., 18(2), 649–671, doi:10.5194/hess-18-649-2014, 2014.
- 595 Fan, Y.: Groundwater in the Earth’s critical zone: Relevance to large-scale patterns and processes: Groundwater at large scales, Water Resour. Res., 51(5), 3052–3069, doi:10.1002/2015WR017037, 2015.
- Fan, Y., Li, H. and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339(940), 1–5, doi:10.1126/science.1229881, 2013.
- 600 Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., Hazenberg, P., McNamara, J., Pelletier, J., Perket, J., Rouholahnejad-Freund, E., Wagener, T., Zeng, X., Beighley, E., Buzan, J., Huang, M., Livneh, B., Mohanty, B. P., Nijssen, B., Safeeq, M., Shen, C., Verseveld, W., Volk, J. and Yamazaki, D.: Hillslope Hydrology in Global Change Research and Earth System Modeling, Water Resour. Res., 55(2), 1737–1772, doi:10.1029/2018WR023903, 2019.
- 605

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.

Fleig, A. K., Tallaksen, L. M., Hisdal, H. and Demuth, S.: A global evaluation of streamflow drought characteristics, *Hydrology and Earth System Sciences*, 5, 535–552, 2006.

- 610 Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A. and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Sci Data*, 2(1), doi:10.1038/sdata.2015.66, 2015.

Ghiggi, G., Humphrey, V., Seneviratne, S. I. and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, *Earth Syst. Sci. Data*, 11(4), 1655–1674, doi:10.5194/essd-11-1655-2019, 2019.

- 615 Gleeson, T., Moosdorf, N., Hartmann, J. and van Beek, L. P. H.: A glimpse beneath earth’s surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, *Geophys. Res. Lett.*, 41(11), 3891–3898, doi:10.1002/2014GL059856, 2014.

GRDC - Global Runoff Data Centre, GRDC: https://www.bafg.de/GRDC/EN/Home/homepage_node.html, last access: 24 December 2019.

- 620 Grimm, A. M.: Interannual climate variability in South America: impacts on seasonal precipitation, extreme events, and possible effects of climate change, *Stoch Environ Res Risk Assess*, 25(4), 537–554, doi:10.1007/s00477-010-0420-1, 2011.

Gudmundsson, L., Wagener, T., Tallaksen, L. M., & Engeland, K.: Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. *Water Resources Research*, 48(11), 2012.

- 625 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(2), 787–804, doi:10.5194/essd-10-787-2018, 2018.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S. and Seneviratne, S. I.: Observed Trends in Global Indicators of Mean and Extreme Streamflow, *Geophys. Res. Lett.*, 46(2), 756–766, doi:10.1029/2018GL079725, 2019.

- 630 Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M. and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, *Hydrol. Earth Syst. Sci.*, 18(2), 463–477, doi:10.5194/hess-18-463-2014, 2014.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., ... & Gomes, S. Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, 12(5), 869–884, 2011.

- 635 Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriaučiūnienė, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A. and Blöschl, G.: Understanding flood regime changes in Europe: a state-of-the-art assessment, *Hydrol. Earth Syst. Sci.*, 18(7), 2735–2772, doi:10.5194/hess-18-2735-2014, 2014.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, *Geochem. Geophys. Geosyst.*, 13(12), doi:10.1029/2012GC004370, 2012.

- 640 Haylock, M. and Nicholls, N.: Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998, *Int. J. Climatol.*, 9, 2000.

- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, edited by B. Bond-Lamberty, PLoS ONE, 12(2), e0169748, doi:10.1371/journal.pone.0169748, 2017.
- Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J.. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of hydrology*, 566, 595-606, 2018.
- 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.
- 650 Huscroft, J., Gleeson, T., Hartmann, J. and Börker, J.: Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLoBAL HYdrogeology MaPS 2.0 (GLHYMPS 2.0), *Geophys. Res. Lett.*, 45(4), 1897–1904, doi:10.1002/2017GL075860, 2018.
- IBGE - Brazilian Institute of Geography and Statistics: Censo Agropecuário 2006, 2007.
- Knoben, W. J. M., Woods, R. A. and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, *Water Resour. Res.*, 54(7), 5088–5109, doi:10.1029/2018WR022913, 2018.
- 655 Ladson, A., Brown, R., Neal, B. and Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter, *AJWR*, 17(1), doi:10.7158/W12-028.2013.17.1, 2013.
- Latrubesse, E. M., Arima, E. Y., Dunne, T., Park, E., Baker, V. R., d’Horta, F. M., Wight, C., Wittmann, F., Zuanon, J., Baker, P. A., Ribas, C. C., Norgaard, R. B., Filizola, N., Ansar, A., Flyvbjerg, B. and Stevaux, J. C.: Damming the rivers of the Amazon basin, *Nature*, 546(7658), 363–369, doi:10.1038/nature22333, 2017.
- 660 Lehner, B.: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network, Global Runoff Data Centre in the Federal Institute of Hydrology (BFG)., 2012.
- 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, *Frontiers in Ecology and the Environment*, 9(9), 494–502, doi:10.1890/100125, 2011.
- Leite, C. C., Costa, M. H., Soares-Filho, B. S. and de Barros Viana Hissa, L.: Historical land use change and associated carbon emissions in Brazil from 1940 to 1995, *Global Biogeochem. Cycles*, 26(2), n/a-n/a, doi:10.1029/2011GB004133, 2012.
- Levy, M. C., Lopes, A. V., Cohn, A., Larsen, L. G. and Thompson, S. E.: Land Use Change Increases Streamflow Across the Arc of Deforestation in Brazil, *Geophys. Res. Lett.*, 45(8), 3520–3530, doi:10.1002/2017GL076526, 2018.
- 670 Liebmann, B., Kiladis, G. N., Allured, D., Vera, C. S., Jones, C., Carvalho, L. M., ... & Gonzáles, P. L. Mechanisms associated with large daily rainfall events in Northeast Brazil. *Journal of Climate*, 24(2), 376-396, 2011.
- Lima, Carlos HR, Amir AghaKouchak, and Upmanu Lall. Classification of mechanisms, climatic context, areal scaling, and synchronization of floods: the hydroclimatology of floods in the Upper Paraná River basin, Brazil. *Earth System Dynamics* 8.4, 1071, 2017.
- 675

- Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F. and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, *Sci Data*, 6(1), 283, doi:10.1038/s41597-019-0300-6, 2019.
- 680 Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R. and Sanders, B. F.: Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States, *Water Resour. Res.*, 53(7), 5469–5494, doi:10.1002/2016WR019676, 2017.
- Marengo, J. A. and Espinoza, J. C.: Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts, *Int. J. Climatol.*, 36(3), 1033–1050, doi:10.1002/joc.4420, 2016.
- 685 Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, 10(5), 1903–1925, doi:10.5194/gmd-10-1903-2017, 2017.
- Martinez, J. A. and Dominguez, F.: Sources of Atmospheric Moisture for the La Plata River Basin, *J. Climate*, 27(17), 6737–6753, doi:10.1175/JCLI-D-14-00022.1, 2014.
- 690 McMillan, H., Westerberg, I. and Branger, F.: Five guidelines for selecting hydrological signatures, *Hydrological Processes*, 31(26), 4757–4761, doi:10.1002/hyp.11300, 2017.
- Melo, D. de C. D., Scanlon, B. R., Zhang, Z., Wendland, E. and Yin, L.: Reservoir storage and hydrologic responses to droughts in the Paraná River basin, south-eastern Brazil, *Hydrol. Earth Syst. Sci.*, 20(11), 4673–4688, doi:10.5194/hess-20-4673-2016, 2016.
- 695 Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J. and Hundecha, Y.: More efforts and scientific rigour are needed to attribute trends in flood time series, *Hydrol. Earth Syst. Sci.*, 16(5), 1379–1387, doi:10.5194/hess-16-1379-2012, 2012.
- Milliman, J. D., Farnsworth, K. L., Jones, P. D., Xu, K. H. and Smith, L. C.: Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000, *Global and Planetary Change*, 62(3–4), 187–194, doi:10.1016/j.gloplacha.2008.03.001, 2008.
- 700 Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, *Hydrol. Earth Syst. Sci.*, 15(2), 453–469, doi:10.5194/hess-15-453-2011, 2011.
- Montanari, A.: What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology, *Hydrol. Process.*, 21(6), 841–845, doi:10.1002/hyp.6623, 2007.
- 705 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., Schaeffli, 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 Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022, *Hydrological Sciences Journal*, 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.
- 710 Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, 18(9), 3511–3538, 2014.

- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrol. Earth Syst. Sci.*, 19(1), 209–223, doi:10.5194/hess-19-209-2015, 2015.
- NOAA: CPC Global Temperature: <https://www.esrl.noaa.gov/psd/>, last access 15 June 2019, 2019a.
- NOAA: CPC Global Unified Precipitation: <https://www.esrl.noaa.gov/psd/>, last access: 15 June 2019, 2019b.
- Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, *River Res. Applic.*, 19(2), 101–121, doi:10.1002/rra.700, 2003.
- ONS - National Electrical System Operator: SIGEL - Sistema Geográfico de Informações do Sistema Elétrico: <https://sigel.aneel.gov.br/>, last access: 10 December 2019.
- Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S. and Bulhões Mendes, C. A.: Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin, *Water Resour. Res.*, 49(3), 1226–1243, doi:10.1002/wrcr.20067, 2013.
- Pasquini, A. I. and Depetris, P. J.: Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview, *Journal of Hydrology*, 333(2–4), 385–399, doi:10.1016/j.jhydrol.2006.09.005, 2007.
- Pfister, L. and Kirchner, J. W.: Debates-Hypothesis testing in hydrology: Theory and practice, *Water Resour. Res.*, 53(3), 1792–1798, doi:10.1002/2016WR020116, 2017.
- Raia, A. and Cavalcanti, I. F. A.: The Life Cycle of the South American Monsoon System, *J. Climate*, 21(23), 6227–6246, doi:10.1175/2008JCLI2249.1, 2008.
- Salio, P., Nicolini, M., Zipser, E.J. Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet. *Mon Weather Rev.*, 135(4): 1290-1309, 2007.
- Sankarasubramanian, A., Vogel, R. M. and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, *Water Resour. Res.*, 37(6), 1771–1781, doi:10.1029/2000WR900330, 2001.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A. and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15(9), 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.
- Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M. and Carrillo, G.: Characterizing hydrologic change through catchment classification, *Hydrol. Earth Syst. Sci.*, 18(1), 273–285, doi:10.5194/hess-18-273-2014, 2014.
- Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., ... & Gosling, S. N. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, 111(9), 3245-3250, 2014.
- Schobbenhaus, C., Gonçalves, J., Santos, J., Abram, M., Leão Neto, R., Matos, G., Vidotti, R., Ramos, M. and Jesus, J. de: *Carta geológica do Brasil ao milionésimo: Sistema de Informações Geográficas, CPRM - Serviço Geológico do Brasil*, Brasília, 2004.

- Seager, R., Naik, N., Baethgen, W., Robertson, A., Kushnir, Y., Nakamura, J. and Jurburg, S.: Tropical Oceanic Causes of Interannual to Multidecadal Precipitation Variability in Southeast South America over the Past Century, *J. Climate*, 23(20), 5517–5539, doi:10.1175/2010JCLI3578.1, 2010.
- Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H. and Dai, Y.: Mapping the global depth to bedrock for land surface modeling, *J. Adv. Model. Earth Syst.*, 9(1), 65–88, doi:10.1002/2016MS000686, 2017.
- Shen, X., Anagnostou, E. N., Mei, Y. and Hong, Y.: A global distributed basin morphometric dataset, *Sci Data*, 4(1), 160124, doi:10.1038/sdata.2016.124, 2017.
- Singh, R., Archfield, S. A. and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – A comparative hydrology approach, *Journal of Hydrology*, 517, 985–996, doi:10.1016/j.jhydrol.2014.06.030, 2014.
- 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.
- Slater, L. J., Singer, M. B. and Kirchner, J. W.: Hydrologic versus geomorphic drivers of trends in flood hazard, *Geophys. Res. Lett.*, 42(2), 370–376, doi:10.1002/2014GL062482, 2015.
- Smakhtin, V. U.: Low flow hydrology: a review, *Journal of Hydrology*, 240(3–4), 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001.
- Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F. and Townshend, J. R.: Global land change from 1982 to 2016, *Nature*, 560(7720), 639–643, doi:10.1038/s41586-018-0411-9, 2018.
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S. and Hsu, K.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, *Rev. Geophys.*, 56(1), 79–107, doi:10.1002/2017RG000574, 2018.
- Tedeschi, R. G., Cavalcanti, I. F. A. and Grimm, A. M.: Influences of two types of ENSO on South American precipitation, *Int. J. Climatol.*, 33(6), 1382–1400, doi:10.1002/joc.3519, 2013.
- Tomasella, J., Borma, L. S., Marengo, J. A., Rodriguez, D. A., Cuartas, L. A., A. Nobre, C. and Prado, M. C. R.: The droughts of 1996–1997 and 2004–2005 in Amazonia: hydrological response in the river main-stem, *Hydrol. Process.*, 25(8), 1228–1242, doi:10.1002/hyp.7889, 2011.
- Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M. and Van Loon, A. F.: Hydrological drought across the world: impact of climate and physical catchment structure, *Hydrol. Earth Syst. Sci.*, 17(5), 1715–1732, doi:10.5194/hess-17-1715-2013, 2013.
- Van Loon, A. F.: Hydrological drought explained, *WIREs Water*, 2(4), 359–392, doi:10.1002/wat2.1085, 2015.
- Veldkamp, T. I. E., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C., Schmied, H. M., ... & Satoh, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. *Environmental Research Letters*, 13(5), 055008, 2018.
- Villarini, G.: On the seasonality of flooding across the continental United States, *Advances in Water Resources*, 87, 80–91, doi:10.1016/j.advwatres.2015.11.009, 2016.

- 780 Wada, Y., Wisser, D. and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, *Earth Syst. Dynam.*, 5(1), 15–40, doi:10.5194/esd-5-15-2014, 2014.
- Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, *Hydrol. Earth Syst. Sci.*, 19(9), 3951–3968, doi:10.5194/hess-19-3951-2015, 2015.
- 785 Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., Goldsmith, S., Harmon, R., Hendrickx, J. M. H., Juvik, J., McDonnell, J. and Ogden, F.: The hydrology of the humid tropics, *Nature Clim Change*, 2(9), 655–662, doi:10.1038/nclimate1556, 2012.
- Woldemeskel, F. and Sharma, A.: Should flood regimes change in a warming climate? The role of antecedent moisture conditions, *Geophys. Res. Lett.*, 43(14), 7556–7563, doi:10.1002/2016GL069448, 2016.
- 790 Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, *Advances in Water Resources*, 32(10), 1465–1481, doi:10.1016/j.advwatres.2009.06.011, 2009.
- Wongchuig, S. C., de Paiva, R. C. D., Siqueira, V., & Collischonn, W. Hydrological reanalysis across the 20th century: A case study of the Amazon Basin. *Journal of hydrology*, 570, 755–773, 2019.
- Xavier, A. C., King, C. W. and Scanlon, B. R.: Daily gridded meteorological variables in Brazil (1980–2013), *Int. J. Climatol.*, 36(6), 2644–2659, doi:10.1002/joc.4518, 2016.
- 795 Yadav, M., Wagener, T. and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Advances in Water Resources*, 30(8), 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007.
- Yokoo, Y. and Sivapalan, M.: Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis, *Hydrol. Earth Syst. Sci.*, 15(9), 2805–2819, doi:10.5194/hess-15-2805-2011, 2011.
- 800 Zhao, F., Veldkamp, T. I., Frieler, K., Schewe, J., Ostberg, S., Willner, S., ... & Leng, G. The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, 12(7), 075003, 2017.
- Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T. and Zhang, X.: Future climate risk from compound events, *Nature Clim Change*, 8(6), 469–477, doi:10.1038/s41558-018-0156-3, 2018.

805 **Table 1. Summary of the data provided by CAMELS-BR.**

	Variable	Description	N. of catchments
Daily time series	Raw streamflow	As read from the Brazilian National Water Agency ^a , in m ³ s ⁻¹ , with varying coverage periods, with data quality flags (qual_control_by_ana and qual_flag) for each time step	3713 3679
	Streamflow	Transformed to mm day ⁻¹ , covering from 1980-2018, with data quality flags (qual_control_by_ana and qual_flag) for each time step	897
	Precipitation	Catchment averages (mm day ⁻¹), with varying coverage periods between 1979-2018 , using three products: CHIRPS v2.0 (0.05° res., covering from 1981-2018) ^b ; CPC Global Unified (0.5° res., 1980-2018) ^c ; and MSWEP v2.2 (0.1° res., 1980-2016) ^d .	897
	Potential evapotranspiration	Catchment averages (mm day ⁻¹), from 1980-2018 , using GLEAM v3.3a (0.25° res., 1980-2018) ^e	897
	Actual evapotranspiration	Catchment averages (mm day ⁻¹), with varying coverage periods between 1980-2018 , using two products: GLEAM v3.3a (0.25° res., 1980-2018) ^e ; and MGB (0.5° res., 1980-2014) ^f	897
	Minimum, maximum, and average temperature	Catchment averages (degrees Celsius day ⁻¹), from 1979-2018 , using CPC Global Unified (0.5° res., 1980-2018) ^g	897
Catchment attributes	Location	5 attributes (Table 2)	3713 3679
	Topography	4 attributes (Table 2)	897
	Data quality checks	2 attributes (Table 2)	897
	Climatic indices	44-13 attributes (Table 3)	897
	Hydrological signatures	13 attributes (Table 4)	897
	Land cover characteristics	11 attributes (Table 5)	897
	Geologic characteristics	7 attributes (Table 6)	897
	Soil characteristics	6 attributes (Table 7)	897
	Human intervention indices	4 attributes (Table 8)	897

^a ANA (2019a).

^b Climate Hazards Group InfraRed Precipitation with Station data v2.0 (Funk et al., 2015).

^c Climate Prediction Center Global Unified Gauge-Based Analysis of Daily Precipitation (NOAA, 2019a).

^d Multi-Source Weighted-Ensemble Precipitation (Beck et al., 2019).

810 ^e Global Land Evaporation Amsterdam Model v3.3a (Miralles et al., 2011; Martens et al., 2017).

^f Large-Scale Hydrological Model (Siqueira et al., 2018).

^g Climate Prediction Center Global Daily Temperature (NOAA, 2019b).

Table 2. Location, topographic characteristics, and data quality checks.

Attribute	Description	Units	Data source
gauge_id	Catchment identifier provided by ANA	-	ANA (2019a)
gauge_name	Gauge name provided by ANA	-	ANA (2019a)
gauge_region	Hydrographic region	-	ANA (2019a)
gauge_lat	Gauge latitude	°N	ANA (2019a)
gauge_lon	Gauge longitude	°E	ANA (2019a)
elev_gauge	Gauge elevation	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
elev_mean	Catchment mean elevation	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
slope_mean	Catchment mean slope	m km ⁻¹	HydroSHEDS 15 arc-sec DEM
area	Catchment area	km ²	Do et al. (2018)
q_quality_control_fraeq_qual ity_control_perc	Fraction-Percentage of streamflow data (1990-2009) with quality control checks by ANA	%	ANA (2019a)
q_stream_stage_fraeq_strea m_stage_perc	Fraction-Percentage of streamflow data (1990-2009) derived from stream stage measurements	%	ANA (2019a)

815

Table 3. Climatic indices.

Attribute	Description	Units	Data source
p_mean	Mean daily precipitation	mm day ⁻¹	CHIRPS v2.0
pet_mean	Mean daily potential evapotranspiration (PET)	mm day ⁻¹	GLEAM v3.3a
<u>et_mean</u>	<u>Mean daily actual evapotranspiration</u>	<u>mm day⁻¹</u>	<u>GLEAM v3.3a</u>
aridity	Aridity, computed as the ratio of mean PET to mean precipitation	-	GLEAM v3.3a and CHIRPS v2.0
p_seasonality	Seasonality and timing of precipitation; (estimated using sine curves to represent the annual temperature and precipitation cycles; <u>values are typically between -1 (precipitation out of phase with temperature) and 1 (precipitation in phase with temperature, i.e., simultaneous peaks), positive (negative) values indicate that precipitation peaks in summer (winter); values close to 0 indicate uniform precipitation throughout the year). See equation 14 in Woods (2009)</u>	-	CHIRPS v2.0
<u>asynchronicity</u>	<u>Asynchronicity between the annual precipitation and PET cycles, where high values represent high relative magnitude and phase differences</u>	-	<u>Feng et al. (2009)</u>
frac_snow	Fraction of precipitation falling as snow (i.e., on days colder than 0 °C)	-	CHIRPS v2.0 and CPC
high_prec_freq	Frequency of high precipitation days (≥ 5 times the mean daily precipitation)	days yr ⁻¹	CHIRPS v2.0
high_prec_dur	Average duration of high precipitation events (number of consecutive days ≥ 5 times the mean daily precipitation)	days	CHIRPS v2.0
high_prec_timing	Season during which most high precipitation days (≥ 5 times the mean daily precipitation) occur	season	CHIRPS v2.0
low_prec_freq	Frequency of dry days (< 1 mm day ⁻¹)	days yr ⁻¹	CHIRPS v2.0
low_prec_dur	Average duration of dry periods (number of consecutive days < 1 mm day ⁻¹)	days	CHIRPS v2.0
low_prec_timing	Season during which most dry days (< 1 mm day ⁻¹) occur	season	CHIRPS v2.0

820

Table 4. Hydrological signatures. Thresholds for high and low flow frequency and duration were obtained from Clausen and Biggs (2000) and Westerberg and McMillan (2015).

Attribute	Description	Units	Data source
q_mean	Mean daily discharge	mm day ⁻¹	ANA (2019a)
runoff_ratio	Runoff ratio, computed as the ratio of mean daily discharge to mean daily precipitation	-	ANA (2019a)
stream_elas	Streamflow precipitation elasticity (i.e., the sensitivity of streamflow to changes in precipitation at the annual timescale, using the mean daily discharge as reference). <u>See equation 7 in Sankarasubramanian et al. (2001), with the last element being \bar{P}/\bar{Q} not \bar{Q}/\bar{P}</u>	-	ANA (2019a)
slope_fdc	Slope of the flow duration curve between the log-transformed 33rd and 66th streamflow percentiles	-	ANA (2019a)
baseflow_index	Baseflow index, computed as the ratio of mean daily baseflow to mean daily discharge, with the hydrograph separation performed using the Ladson et al. (2013) digital filter	-	ANA (2019a)
hfd_mean	Mean half-flow date (i.e., the date on which the cumulative discharge since 1 st September reaches half of the annual discharge)	day of the year	ANA (2019a)
Q5	5% flow quantile (low flow)	mm day ⁻¹	ANA (2019a)
Q95	95% flow quantile (high flow)	mm day ⁻¹	ANA (2019a)
high_q_freq	Frequency of high-flow days (> 9 times the median daily flow)	days yr ⁻¹	ANA (2019a)
high_q_dur	Average duration of high-flow events (number of consecutive days > 9 times the median daily flow)	days	ANA (2019a)
low_q_freq	Frequency of low-flow days (< 0.2 times the mean daily flow)	days yr ⁻¹	ANA (2019a)
low_q_dur	Average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	days	ANA (2019a)
zero_q_freq	Percentage of days with zero discharge	%	ANA (2019a)

Table 5. Land cover characteristics.

Attribute	Description	Units	Data source
crop_ f raeperc	Percentage covered by croplands	%	ESA GlobCover2009
crop_mosaic_ f raeperc	Percentage covered by a mosaic of croplands and natural vegetation	%	ESA GlobCover2009
forest_ f raeperc	Percentage covered by broadleaved or needleleaved forests, either evergreen or deciduous	%	ESA GlobCover2009
shrub_ f raeperc	Percentage covered by shrublands	%	ESA GlobCover2009
grass_ f raeperc	Percentage covered by grasslands or areas with sparse (<15%) vegetation	%	ESA GlobCover2009
bare_ f raebarren_perc	Percentage covered by barren areas	%	ESA GlobCover2009
imperv_ f raeperc	Percentage covered by artificial surfaces or urban areas	%	ESA GlobCover2009
wet_ f raeperc	Percentage covered by water bodies or wetlands	%	ESA GlobCover2009
snow_ f raeperc	Percentage covered by permanent snow or ice	%	ESA GlobCover2009
dom_land_cover	Dominant land cover	-	ESA GlobCover2009
dom_land_cover_ f raeperc	Percentage covered by the dominant land cover	%	ESA GlobCover2009
c			

Table 6. Geologic characteristics.

Attribute	Description	Units	Data source
geol_class_1st	Most common geologic class in the catchment	-	GLiM
geol_class_1st_ fraeperc	Fraction-Percentage of the catchment covered by the most common geologic class	-%	GLiM
geol_class_2nd	Second most common geologic class in the catchment	-	GLiM
geol_class_2nd_ fraeperc	Fraction-Percentage of the catchment covered by the second most common geologic class	-%	GLiM
carb_rocks_ fraeperc	Fraction-Percentage of the catchment covered by carbonate sedimentary rocks	-%	GLiM
geol_porosity	Subsurface porosity of the catchment	-	GLHYMPS v2.0
geol_permeability	Subsurface permeability (log10 scale) of the catchment, extract for each catchment using the geometric mean	m ²	GLHYMPS v2.0

830 **Table 7. Soil characteristics.**

Attribute	Description	Units	Data source
sand_ fraeperc	Fraction-Percentage of sand content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
silt_ fraeperc	Fraction-Percentage of silt content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
clay_ fraeperc	Fraction-Percentage of clay content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
org_carbon_content	Soil organic carbon content at a soil depth of 30 cm	g kg ⁻¹	SoilGrids250m
depth_bedrock bedrock_depth	Depth to bedrock	cm	SoilGrids250m
water_table_depth	Median water table depth	cm	Fan et al. (2013)

Table 8. Human intervention indices.

Attribute	Description	Units	Data source
consumptive_use_mm consumptive_use	Total consumptive water use in 2017, normalized by catchment area	mm yr ⁻¹	ANA (2019c)
consumptive_use_perc	Total consumptive water use in 2017, normalized by mean annual streamflow	%	ANA (2019c)
vol_reservoirs reservoirs_vol	Total maximum storage capacity of the reservoirs in the catchment	10 ⁶ m ³	GRanD v1.3, ONS, and ANA (2018)
degree_of_regulation regulation_degree	Ratio of total reservoir storage capacity of the catchment to its total annual flow	%	GRanD v1.3, ONS, and ANA (2018)

840

Figure 1. (a) South America and the total river discharge data availability from of the 3679 stream gauges included in this study. The black line surrounded by a white line indicates rivers. The dashed line is Brazil's borders. (b) Hydrographic regions of Brazil according to ANA (2019a). (c) Fraction-Percentage of streamflow data with quality control checks by ANA of the 897 selected catchments. (d) Fraction-Percentage of streamflow data derived from stream stage measurements of the 897 selected catchments. The circles are located at the outlet of the catchments and their sizes are proportional to the sizes of the catchments. The grey line in (c) and (d) indicates the limits of hydrographic regions.

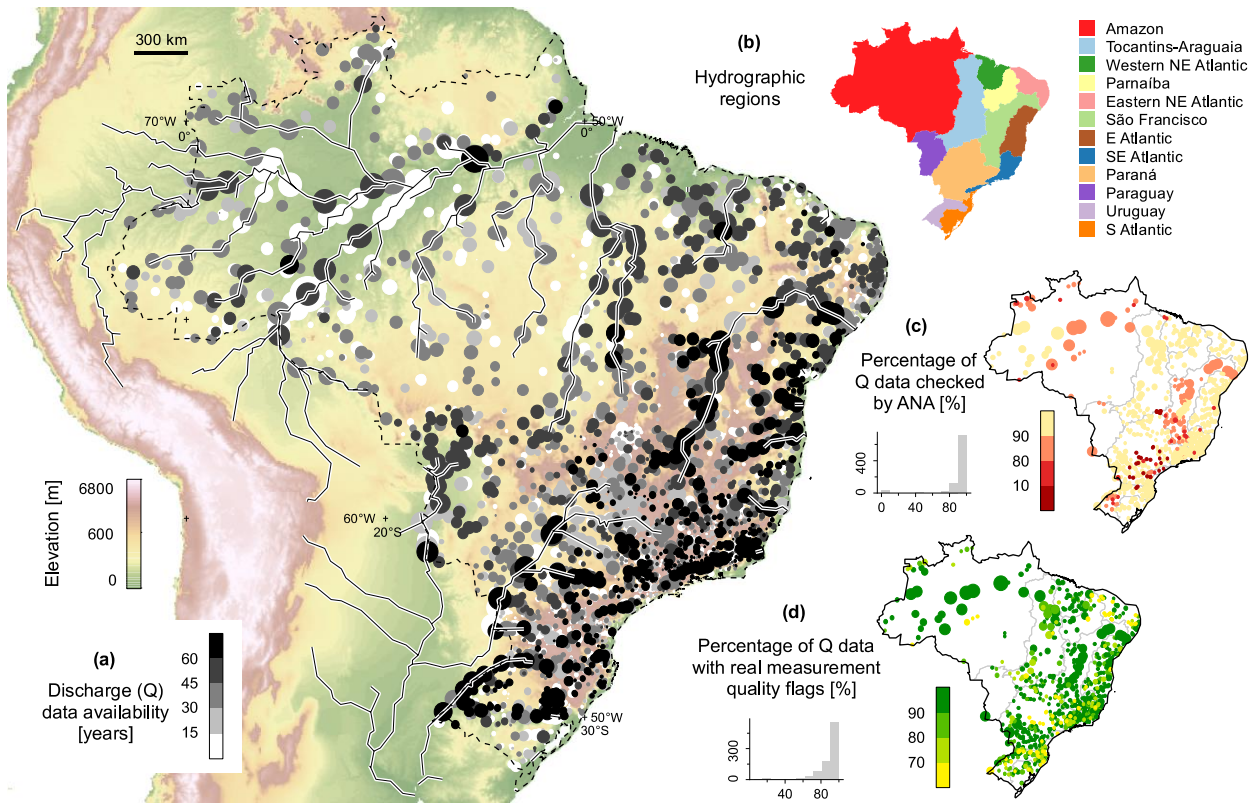
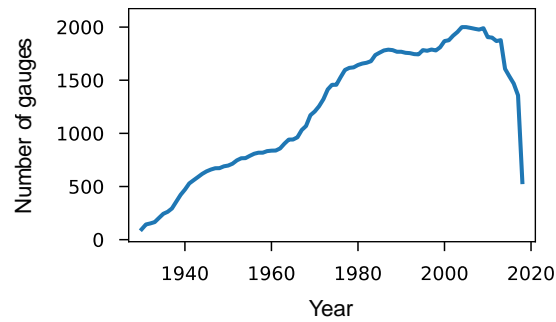
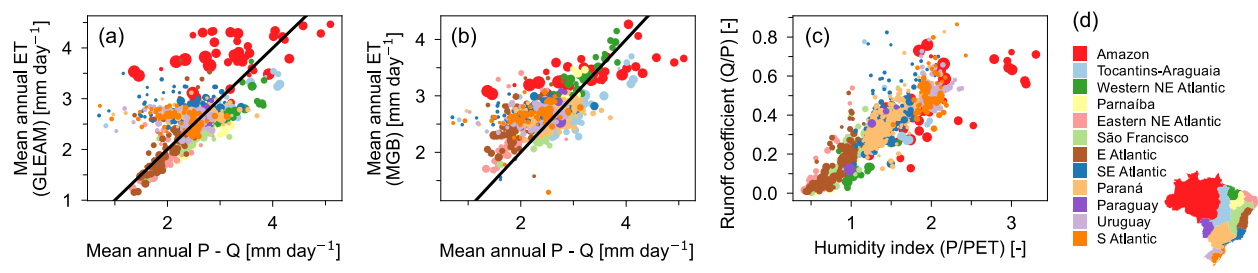


Figure 2. Time series with the number of streamflow gauges with at least one measurement for a given year in Brazil.



845
|

Figure 3. Long-term water balance of the 897 selected catchments for the water years 1990-2009. Mean annual evapotranspiration from (a) GLEAM or (b) MGB as a function of the difference between mean annual precipitation (P) from CHIRPS and streamflow (Q). (c) Runoff coefficient as a function of the humidity index. Line 1:1 is shown in black. The symbol size is proportional to the catchment area. The symbol color indicates the hydrographic region of the catchment (from panel d).



855 **Figure 34.** Topographic characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

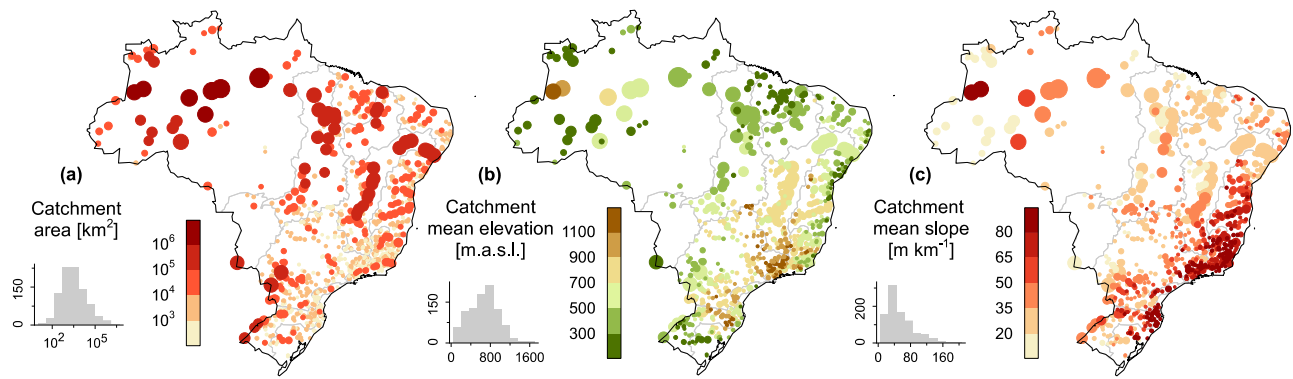


Figure 45. Climatic indices of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

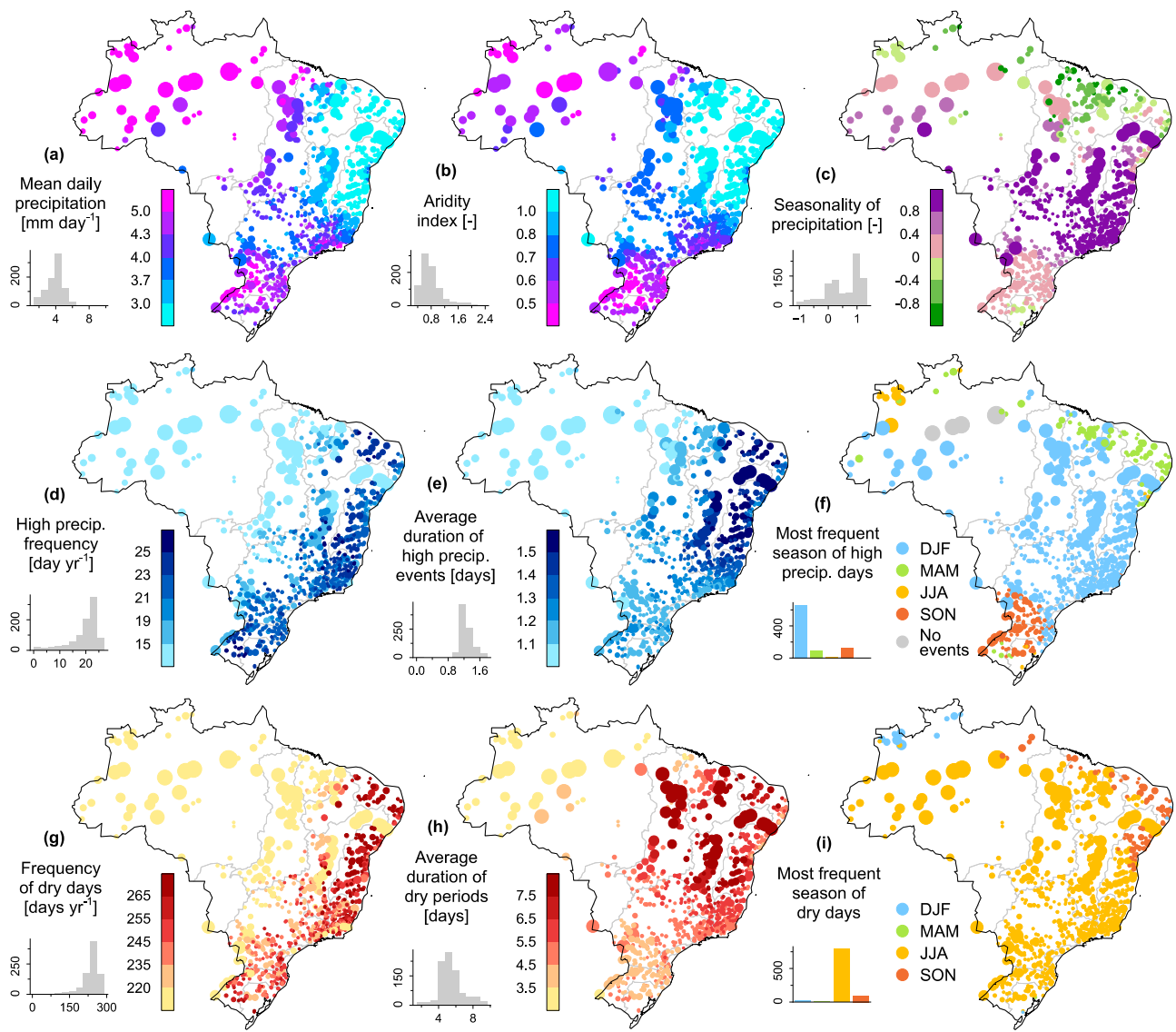
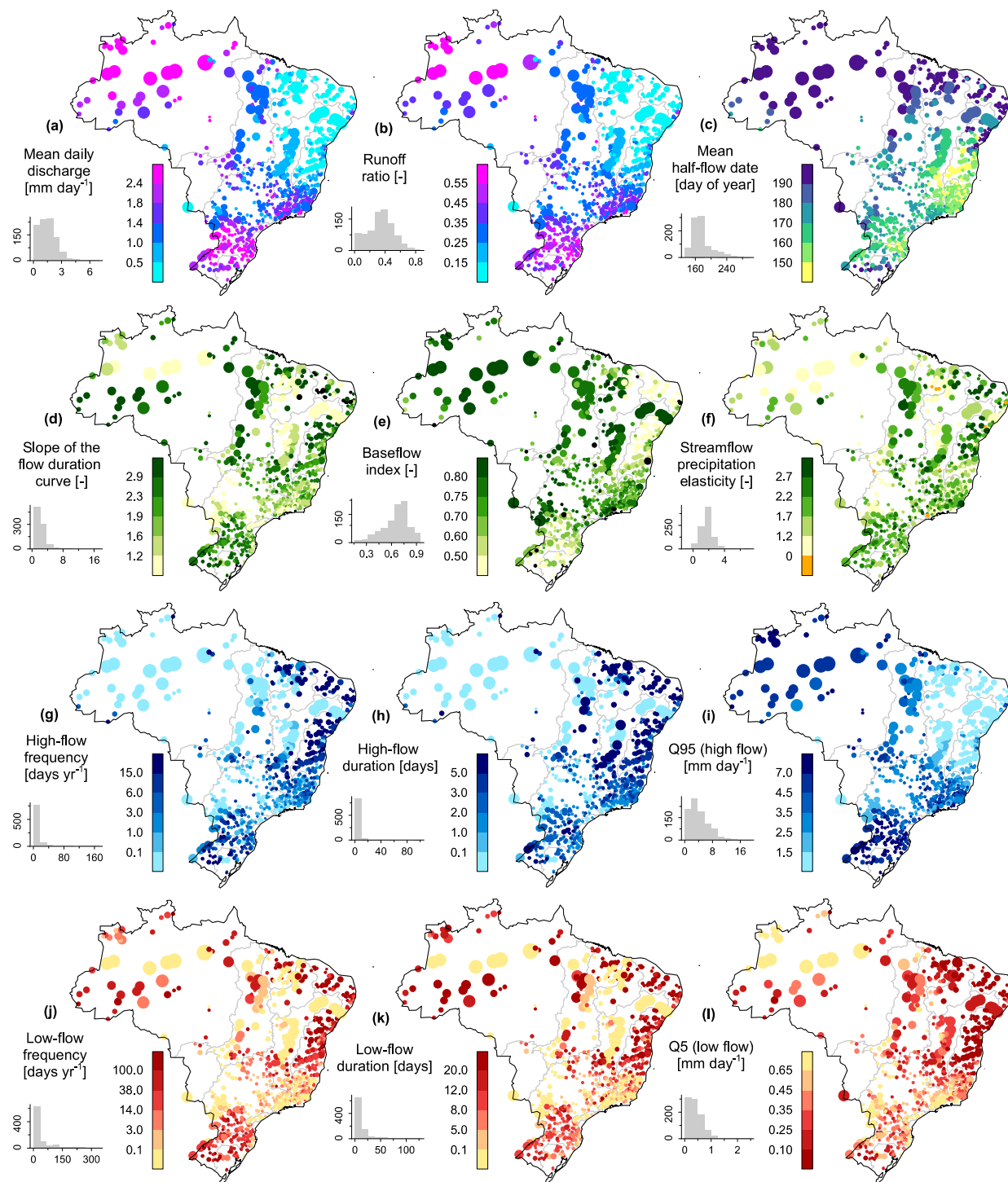
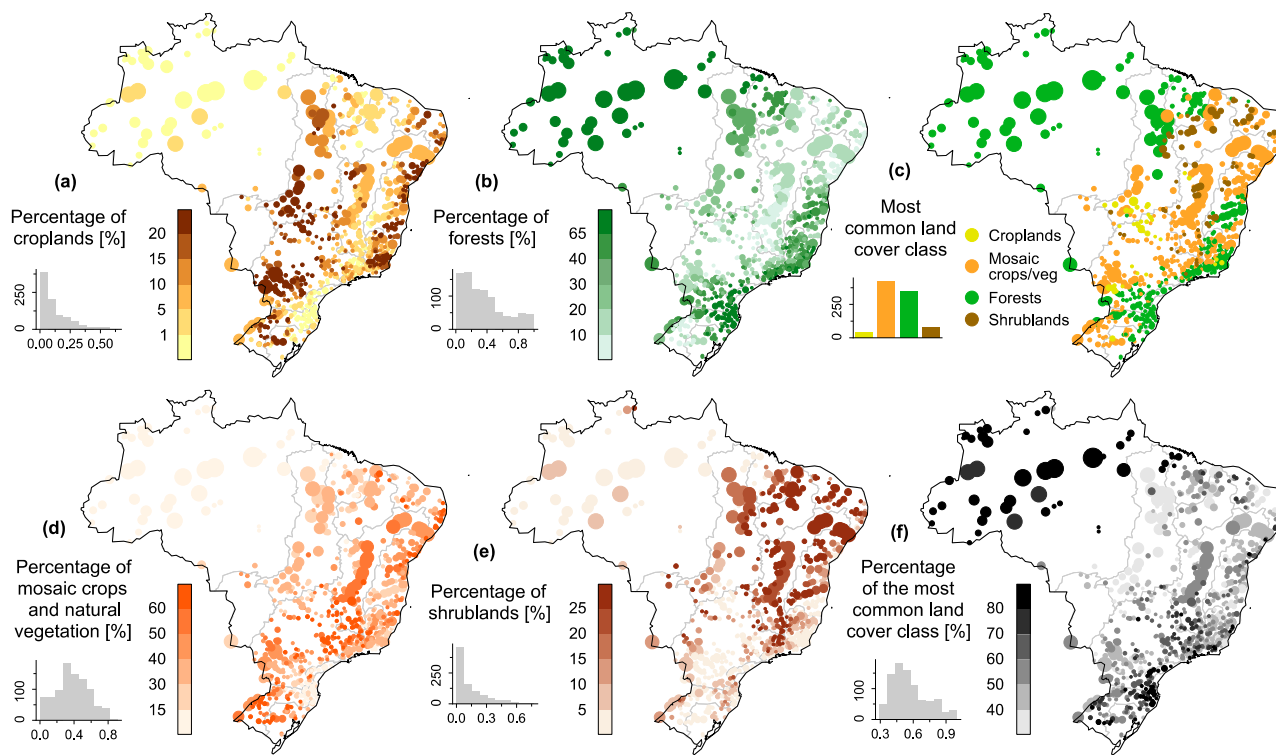


Figure 56. Hydrological signatures of the 897 selected catchments. The black circles are catchments with undefined values. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.



865 **Figure 67.** Land cover characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.



870 **Figure 78.** Geologic characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

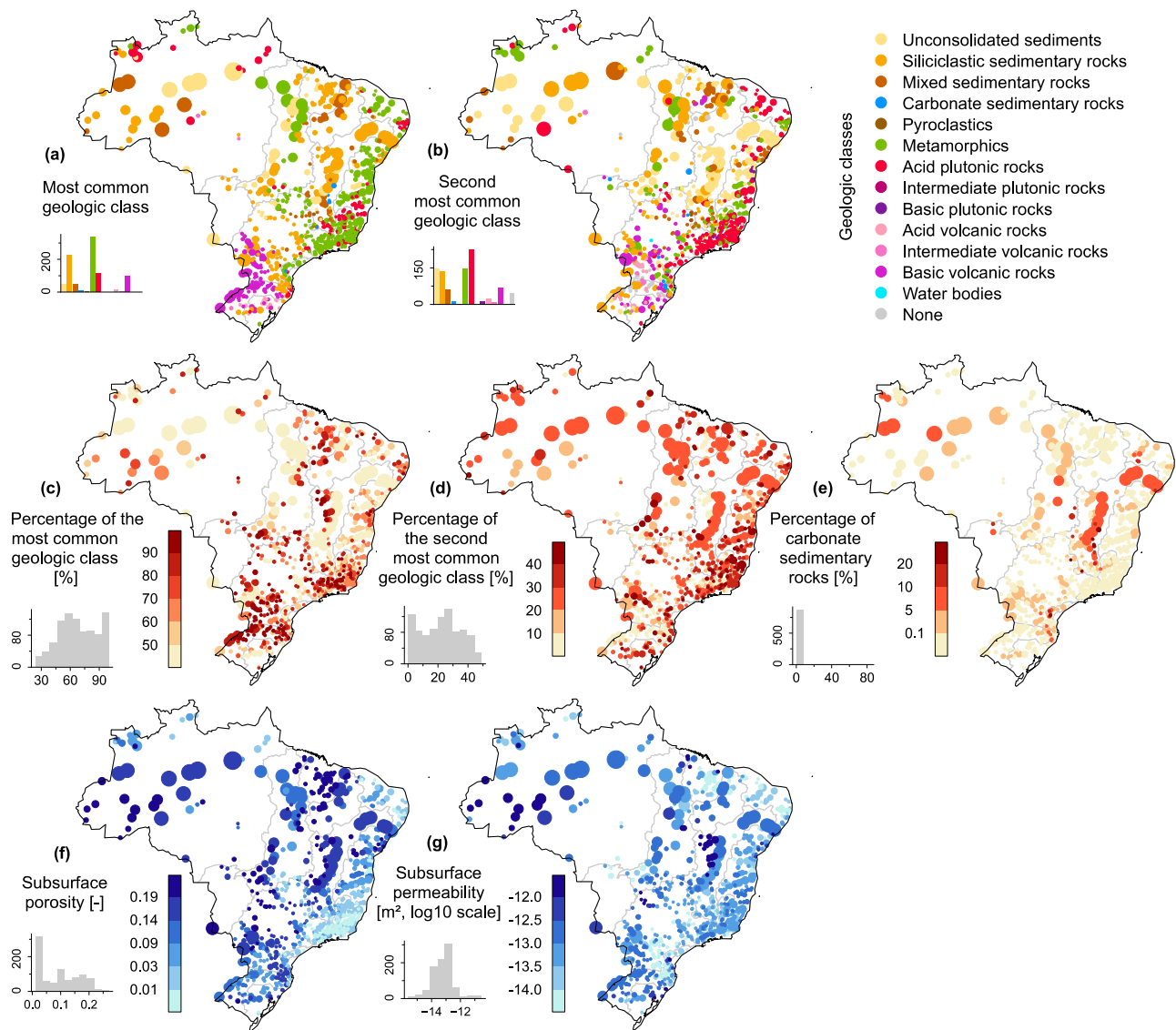


Figure 89. Soil characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

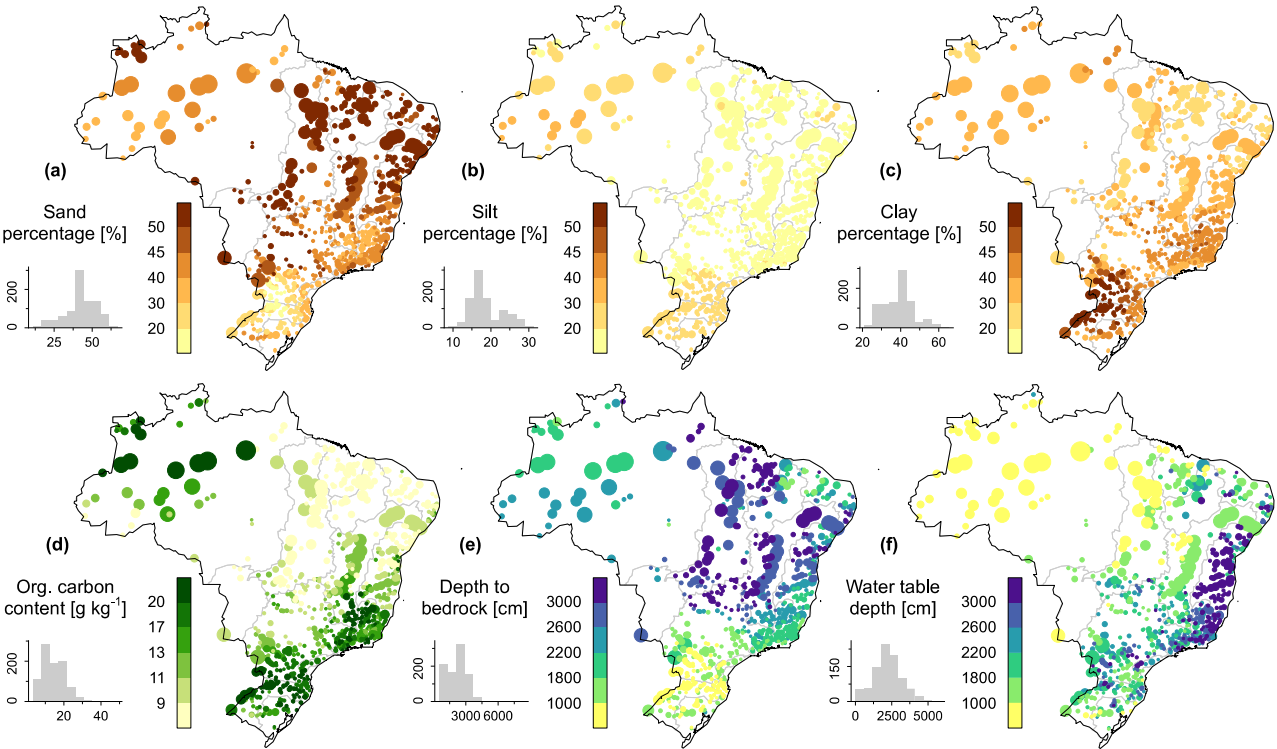


Figure 910. Human intervention indices of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

