Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR037212

The Drivers of Hydrologic Behavior in Brazil: Insights From a Catchment Classification

André Almagro[1], Antônio Alves Meira Neto[2], Noemi Vergopolan[3], Tirthankar Roy[4], Peter A. Troch[5], and Paulo Tarso S. Oliveira[1]

[1]Faculty of Engineering and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, [2]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA, [3]Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, USA, [4]Department of Civil and Environmental Engineering, University of Nebraska-Lincoln, Omaha, NE, USA, [5]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Key Points:

- We analyzed catchment-scale relationships between streamflow and its drivers by clustering the catchments based on hydrological similarity
- Six catchment groups emerged following mainly an aridity gradient, from the wettest to the driest, and seasonality
- Climate controls the hydrological behavior of water-limited catchments, while landscape characteristics control energy-limited catchments

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. T. S. Oliveira,
paulotarsoms@gmail.com

Citation:

Almagro, A., Meira Neto, A. A., Vergopolan, N., Roy, T., Troch, P. A., & Oliveira, P. T. S. (2024). The drivers of hydrologic behavior in Brazil: Insights from a catchment classification. *Water Resources Research*, *60*, e2024WR037212. https://doi.org/10.1029/2024WR037212

Received 25 JAN 2024
Accepted 8 AUG 2024

Abstract Despite hosting ~16% of the global freshwater and almost 50% of water resources in South America, Brazilian catchment-scale relationships between drivers and streamflow are still poorly understood. Here, we used streamflow signatures and attributes of 735 catchments from the Catchment Attributes for Brazil data set to investigate the dominant hydrological processes for the catchments. We also assess how catchments group based on hydrologic behavior similarities and analyze which climatic/landscape attributes control the streamflow variability. To classify and group the catchments, we used the *k*-means method optimized by the Elbow approach, along with a Principal Component Analysis. Uncertainty on catchment grouping was checked by *k*-fold cross-validation. Then, we used a recursive feature elimination using the random forest technique to assess the most influential catchment attributes to the hydrological signatures. Our results revealed six similarity groups, which followed mainly an aridity gradient ranging from the wettest to the driest, but also seasonality. The climate is the primary driver of hydrological behavior for the water-limited groups, highlighting the influence and importance of the atmospheric demand in several Brazilian catchments. High soil storage capacity in energy-limited catchments associated with high precipitation led to high discharge all year due to the subsurface fluxes' contribution. Our findings may be useful to improve streamflow predictability and hydrological behavior identification by further understanding hydrological similarities and their signatures due to catchment landscape characteristics. Further, by employing an easily reproducible methodology and clear metrics to weigh uncertainty, our study provides a significant step toward establishing a catchment-scale common classification system.

1. Introduction

Catchments are considered poorly defined complex systems (Dooge, 1986) characterized by a significant degree of spatial heterogeneity of the landscape characteristics and spatial and temporal variability of its hydroclimate processes (Beven, 2000; McDonnell & Woods, 2004). This variability might suggest unique landscapes (such as topography, soils, geology, vegetation, and anthropogenic modification) and climate characteristics make it challenging to generalize catchments' hydrological behavior—highlighting its uniqueness (Beven, 2000; Vergopolan et al., 2022). However, given how geomorphology, soils, and vegetation are adaptive to (and a result of) climate, geology, and time, we can consider that catchments present at least some level of self-organization (Blöschl & Sivapalan, 1995; Dooge, 1986; Sivapalan, 2005; Troch et al., 2013).

Identifying catchments' hydrological behavior and controlling characteristics is the first step toward a better understanding of the many complex levels involved in catchment hydrological processes (McDonnell & Woods, 2004). One of the main goals of classifying their behaviors and drivers is to provide insights into catchment functioning and structure, similarities (or dissimilarities), connections, and transfer information between them (Wagener et al., 2007). Deciphering meaningful observation patterns inevitably relies on a catchment classification system capable of predicting the dominant controls on the water fluxes (Sivapalan, 2005). Fundamentally, classifying catchments by their salient and emblematic characteristics enabled the development of suitable methodologies for catchment-scale purposes for hydrologic prediction, hazard assessments, and new theories development (Sivakumar et al., 2013). Furthermore, the development of a catchment classification standard focusing on hydrologic functioning provides an avenue toward a universal understanding within the hydrology sciences (Wagener et al., 2008).

Although there is no standard catchment classification or grouping system within the hydrologic sciences (McDonnell & Woods, 2004; Wagener et al., 2007), some directions and starting points have been proposed (Wagener et al., 2007). A classification system must (a) map the catchment geophysical conditions across spatiotemporal scales, and the catchment functions related to the partition, storage, and release of the water entering the catchment must be considered; (b) the hydrologic catchment functions should initially be based on the streamflow, to be broadly applicable; (c) it must weigh the uncertainty of the metrics adopted for classification. Based on this, it is possible to achieve a reliable catchment classification system capable of limiting all the internal complexity and spatial variability within determined classes (McDonnell & Woods, 2004; K. Sawicz et al., 2011). Previous research has shown five main (and critical) components of the streamflow regime that describe and regulate the hydrological processes in catchments: the magnitude, frequency, duration, seasonality, and rate of change of the flow (H. McMillan et al., 2017; Poff & Ward, 1989; Richter et al., 1996). A comprehensive representation of these components contributes to a more accurate representation of the hydrological behavior within similarity groups.

Data collection is a primary component in developing a meaningful classification system (Beven, 2000). An integrated assessment of the recently emerged large-sample data sets is essential to understanding catchment functioning and its controllers (Addor et al., 2017; Almagro et al., 2021; Lyon & Troch, 2010; Wagener et al., 2007). Despite the lack of a standard classification system, statistical techniques, such as clustering, are often used to find patterns and distinguish similar catchments among these data sets (Sivapalan, 2005; Wagener et al., 2008). Several studies adopted statistical clustering methods to identify similar catchments, from the fuzzy partitioning method (Knoben et al., 2018; K. Sawicz et al., 2011; K. A. Sawicz et al., 2014; Wu et al., 2021) to the hierarchical clustering (Chaney et al., 2021; Jehn et al., 2020), and manually by trial and error (Berghuijs et al., 2014). A common characteristic among all the cited studies is that they use hydrological signatures that cover the main components of entire range of flows (especially the baseflow index, flow duration curve, and mean daily streamflow).

The global hydroclimate can be viewed as a continuous spectrum, with each catchment situated along this spectrum, suggesting that spatial connectivity may play a role (Jehn et al., 2020; Knoben et al., 2018). The research focused on classifying catchments based on hydrological behavior tends to uncover generalized patterns of mechanisms and controls, revealing key similarities despite significant heterogeneity. For example, Wu et al. (2021) expanded the Dunne diagram (Dunne, 1978) by incorporating additional process details obtained through in-depth catchment classification analysis of the runoff generation process. Similarly, Tarasova et al. (2018) identified robust relationships between event characteristics and wetness states, enabling direct inferences about the role of storage in catchment functioning.

Although Brazil possesses approximately 16% of the world's freshwater and accounts for nearly half of South America's water resources, there remains a limited understanding of the interconnections between drivers and streamflow at the catchment scale in the country (Ballarin et al., 2023; Schwamback et al., 2022). Almagro et al. (2021) have paved the way for the first catchment classification efforts in Brazil by compiling more than 100 catchment attributes for 735 catchments available through the Catchment Attributes for Brazil database (CABra). However, to the best of our knowledge, there is no study classifying and grouping catchments in Brazil. Furthermore, few studies realized an in-depth investigation into the single groups of similar catchments (Tarasova et al., 2018; Wu et al., 2021), which can be useful for identifying the drivers of their internal variability. The lack of in-depth analysis opens the possibility (and opportunity) to investigate and better understand dominant controls on hydrological processes, and catchment functioning in environments not yet explored.

In this study, we intend, apart from considering differences among the groups, to focus also on the in-depth investigation of the dominant controls of streamflow variability within the similar groups taking advantage of prior findings that provide valuable insights for directing future analyses, offering a roadmap for targeted exploration. To do so, we leverage statistics and machine learning advances to present the first hydrological characterization and classification of Brazilian catchments. We aim to fill this knowledge gap by (a) proposing an approach to classify catchments by their similarities in hydrological behavior and (b) by determining the catchments' characteristics that control the streamflow variability. The 735 catchments from the CABra data set were clustered using 14 streamflow signatures. We use a random forest algorithm to identify and quantify the main drivers of hydrological processes for the similar catchment groups and determine the controls on streamflow variability. By (a) and (b), we provide insights into a better understanding of the dominant hydrological processes

10.1029/2024WR037212

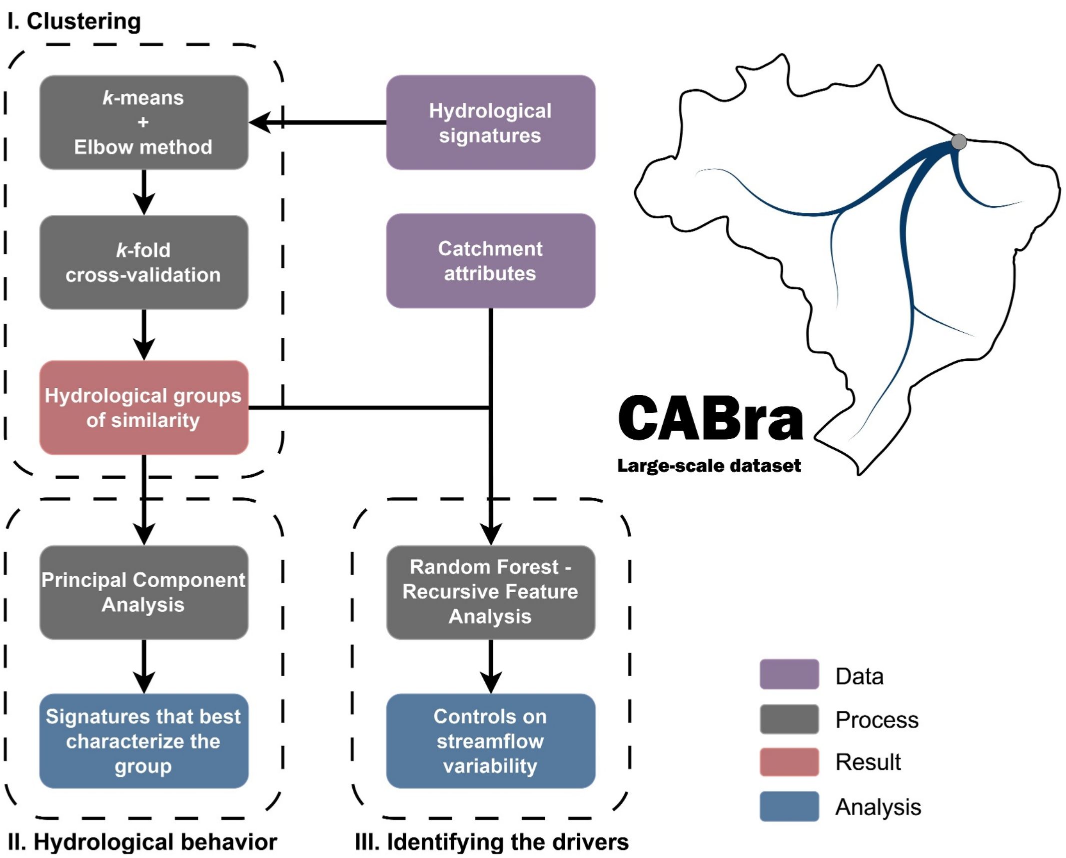

Figure 1. An overview of the methodology used in this study.

and contribute to a common framework formation (Sivakumar, 2008). The integrated analysis of catchments' location, hydrological signatures, and landscape attributes enables us to identify and understand the main features of climate and landscape and their relationship driving hydrological behavior. The catchment classification reveals potential sources of hydrologic predictability, improving our understanding of Brazilian catchments' functioning and behavior and providing insights into the drivers of freshwater availability and its change, given the vital importance of Brazilian water resources for the global water cycle.

2. Materials and Methods

The methodology outlined in this paper involves three main steps using hydrological and landscape data from Brazilian catchments, from the Catchment Attributes for Brazil—CABra data set, as illustrated in Figure 1. The approach presented here attempted to reduce the subjectivity by employing statistical methods with minimal user inferences. First, we utilized the k-means method optimized by the Elbow approach to classify and group the Brazilian catchments based on a set of hydrological signatures describing their hydrological behavior, resulting in hydrological groups of similarity. Second, a Principal Component Analysis (PCA) was applied to the catchment groups to identify the hydrological signatures that best characterize each group. Finally, a Random Forest algorithm was employed to determine the main catchment attributes influencing each group's hydrological behavior and streamflow variability, utilizing Recursive Feature Analysis. This approach extends its applicability beyond Brazilian hydrology, providing a versatile framework for classifying catchments using diverse catchment data sets. Detailed descriptions of the data and processes are provided in Sections 2.1–2.5.

2.1. Catchments Data Set

Brazil has significant importance on the global hydrologic cycle, with ∼16% of the globe's freshwaters flowing in its rivers (D. B. B. Rodrigues et al., 2015). To classify and assess the main controls of the streamflow variability in

Brazil, we used data from the Catchment Attributes for Brazil—CABra database (Almagro et al., 2021), comprising 735 catchments. The set of catchments includes all Brazilian biomes and climatic regions, with more than 100 attributes related to topography, climate, streamflow, soils, geology, groundwater, land use, and anthropogenic disturbance. The CABra data set also presents a daily time series of climate variables and river discharge along with the catchment attributes. The main variables used in our study are precipitation (P), actual evapotranspiration (E), potential evapotranspiration (Ep), and streamflow (Q). P was derived from the Brazilian Daily Weather Gridded Data (BR-DWGD) (Xavier et al., 2016) and the globally widely used ERA5 (Hersbach et al., 2020). Ep was calculated by employing the Priestley and Taylor (1972) equation to the BR-DWGD and ERA5 weather data. It is important to note that the reference data set BR-DWGD covers only the Brazilian territory, while the CABra data set has 20 catchments with upstream areas outside Brazil. To overcome this, the ERA5 data was incorporated to the CABra data set to produce a more reliable product for all the CABra catchments, based on an ensemble mean product using both data sets in catchments with areas outside Brazil. Therefore, 715 (97%) out of 735 catchments have the climatic data derived from ground observations and 20 (3%) are using a reanalysis data set. E was derived from the Global Land Evaporation Amsterdam Model version 3 (Martens et al., 2017). Finally, Q was obtained from gauge observations from the Brazilian Water Agency (ANA). It is important to note that all the hydrological signatures were calculated considering the hydrological years (1 October–30 September) from 1980 to 2010. The CABra data set can be accessed at https://thecabradataset.shinyapps.io/CABra (interactive interface) and https://doi.org/10.5281/zenodo.4070146 (data files).

2.2. Streamflow Signatures

The components that characterize the entire range of flows can be described by statistical metrics, which are known as hydrologic signatures. The role of the hydrological signatures is to extract minimal representations of relevant information from the flow regime data (Gupta et al., 2008). Thus, hydrologic signatures represent the hydrological behavior, identify dominant processes, and determine the spatiotemporal variability of the rainfall-runoff response (H. McMillan et al., 2017).

Finding the links between hydrological signatures and hydrological processes is essential to quantify similarity and explain phenomena at the catchment scale. These signatures represent catchments' streamflow and runoff characteristics quantitatively in response to climate forcings and landscape characteristics, limiting redundancy or overlap (H. K. McMillan, 2021). Furthermore, the signatures provide a quantitative measure that can be compared and tested against other catchments, allowing us to determine the hydrological similarity strength (H. McMillan et al., 2017; K. Sawicz et al., 2011; Westerberg & McMillan, 2015). We considered a set of 14 (of 18 available in the CABra data set) streamflow signatures obtained from the CABra data set version 5 (Almagro et al., 2021), which were calculated using observed data from the Brazilian Water Agency (ANA) from 1980 to 2010. The hydrological signatures adopted in this study (Table 1) describe the streamflow's magnitude, frequency, duration, seasonality, and rate of changes (Poff & Ward, 1989). Here, we categorize the hydrological signatures following previous studies (Westerberg & McMillan, 2015) that group them based on the distribution, frequency and duration, and dynamics of streamflow, covering all aspects of the hydrological behavior of the catchments.

To enhance the quality and relevance of the data used, our set of signatures could be comparable to other studies, are capable of capture a broad range of hydrological information and represent the hydrological behavior. At the same time, to ensure that the signatures are also independent among the aspect being represented, signatures strongly correlated with other similar signatures (as determined through a correlation matrix) were removed as they did not add value to our analysis. From the original set of hydrological signatures from the CABra data set representing the distribution (magnitude) of streamflow, the 1st and 99th percentiles of streamflow were removed from the analysis because they are highly correlated with the 5th and 95th percentiles, respectively. We also removed the coefficient of variation for the low- and high-flows because they were also strongly correlated with the coefficient of variation of daily streamflow.

2.2.1. Distribution (Magnitude) Signatures

To compare different water yields between the pool of catchments, we used the mean daily streamflow. In addition, to quantify the flow distribution and its lower and upper limits, we used the 5th and 95th quantiles of daily streamflow (Q_5 and Q_{95}). Along with the coefficient of variation chosen to reflect the interannual streamflow variability, these signatures help identify and resolve differences in streamflow variability between catchments.

Table 1
The Hydrological Streamflow Signatures Used in This Study

Aspect	Attribute	Long name	Unit
Distribution (magnitude)	Mean daily streamflow	Mean daily streamflow	$mm.day^{-1}$
	Q_5	5th percentile of streamflow	$mm.day^{-1}$
	Q_{95}	95th percentile of streamflow	$mm.day^{-1}$
Frequency and duration	Frequency of low-flow	Mean frequency of low-flow events	$day.year^{-1}$
	Duration of low-flow	Mean duration of low-flow events	days
	Frequency of high-flow	Mean frequency of high-flow events	$day.year^{-1}$
	Duration of high-flow	Mean duration of high-flow events	days
	Half-flow day	Mean half-flow date	day of the year
	Zero-flow frequency	Mean frequency of zero-flow events	$day.year^{-1}$
Dynamics (seasonality and rate of change)	Coef. of variation	Coef. of variation of daily streamflow	–
	Elasticity of streamflow	Streamflow elasticity to precipitation	%
	Slope of FDC	The slope of the flow duration curve	–
	Baseflow index	Baseflow index	–
	Annual runoff ratio	Annual runoff ratio of the catchment	–

2.2.2. Frequency and Duration Signatures

The frequency and duration of low- and high-flow events are essential to identify how many events occur in a year and how long they last. Low- and high-flow events were selected based on a threshold based on the daily streamflow, as adopted in the CABra data set (Almagro et al., 2021). For the low-flow, the frequency is defined as the number of separated streamflow events lower than 0.2 times the daily mean (Olden & Poff, 2003). For each of these events, its duration is considered as a separate signature. Similarly, high-flow events are defined as events 9 times higher than the mean daily streamflow (Clausen & Biggs, 2000). These metrics are useful for understanding the flow regime better and the occurrence of hydrological droughts and floods, which may adversely impact ecosystems and society. We also adopted the zero-flow signature, which refers to the average number of days with no flow conditions within a year, which is a metric critical to detect and differentiate non-perennial streams.

The half-flow date is the date (of the hydrological year) in which half of the streamflow is reached. This signature was proposed by Court (1962) and is based on the concept of the mass median of the streamflow for a given time, in which the median is the date that half of the year's flow has passed (and half is yet to pass). It uses the entire streamflow series, avoiding discarding information about streamflow seasonality by adopting the date of its momentary maximum. It shows the interdependence of total streamflow and half-flow date, appearing as a good indicator of streamflow seasonality (and changes in seasonality) (Court, 1962). Low values of the half-flow date indicate that higher streamflow amounts occur in the first portion of the hydrological year (early summer, matching with the precipitation peak in Brazil). In contrast, higher signature values indicate that low flows dominate the first portion of the hydrological year.

2.2.3. Dynamics (Seasonality and Rate of Change) Signatures

The elasticity of streamflow characterizes the sensitivity of a catchment's streamflow response to changes in precipitation at the annual time scale. Although streamflow is sensitive to other factors, the precipitation primarily determines changes in the water availability (Berghuijs et al., 2017). It can be calculated by taking the inter-annual difference between annual streamflow divided by the inter-annual difference between annual precipitation, which is then normalized by the long-term runoff ratio (Dooge, 1992; Sankarasubramanian et al., 2001; Schaake, 1990). Streamflow elasticity is the proportional change in streamflow divided by the proportional change in a climate variable (here, the precipitation). The elasticity value defines the streamflow changes due to the 1% change in the precipitation. Catchments with a value greater than one are considered elastic (sensitive to a shift in the precipitation). In comparison, those with a value lower than one are considered inelastic (insensitive to a change in the precipitation).

10.1029/2024WR037212

The slope of the flow duration curve (FDC) is a signature calculated between the 33rd and 66th percentiles of streamflow since, at a semi-log scale, this represents a relatively linear part of the FDC (Yadav et al., 2007). Catchments with damped and less variable responses would present a low slope value, while those with a high slope value would have a more variable streamflow regime.

The baseflow index (BFI) is the ratio between long-term baseflow and the total streamflow. It is the most popular signature to quantify the importance of baseflow processes (H. McMillan, 2020; K. Sawicz et al., 2011), typically associated with the discharge of water from groundwater storage—or the slow response of the catchment. A high BFI defines a catchment with high baseflow contribution, that is, more water using slow flow paths through the catchment. Among the various algorithms that have been proposed in the literature, the baseflow separation was performed by using the Lyne and Hollick (1979) digital filter, as described in Almagro et al. (2021).

The annual runoff ratio is the ratio of long-term annual streamflow to long-term annual precipitation. It represents the catchment's fast response to the rainfall, being mainly generated by infiltration/saturation excess. As the annual runoff ratio is influenced by water losses to evapotranspiration and deep groundwater, it can be used as a signature of overall water loss to these water balance components (H. McMillan et al., 2014). A high annual runoff ratio describes a streamflow-dominated catchment from which a large portion of water exits as streamflow. In contrast, a low runoff ratio characterizes that a small portion of precipitation is converted into streamflow.

2.3. Clustering Catchments by Their Streamflow Signatures

To classify and group the CABra catchments, we used a cluster analysis methodology, which is widely used in hydrological studies (Ali et al., 2012; Rao & Srinivas, 2006; K. Sawicz et al., 2011; Sivakumar et al., 2013), along with a Principal Component Analysis (PCA), which enables identifying the hydrological signatures that best describe each group, for example, the most outstanding hydrological signature. The clustering analysis groups units according to their similarity in previously defined measures. In this study, the catchments are the units, and the similarity measures are the hydrological signatures.

Among several clustering methods described in the literature, the most popular ones are hierarchical agglomerative, Fuzzy partitioning, *k*-means, PCA, and affinity propagation (García-Escudero et al., 2010; Milligan & Cooper, 1987). Here, we used the *k*-means method optimized by the Elbow approach (Nainggolan et al., 2019). *k*-means is an unsupervised partitional clustering algorithm to classify multivariate observations based on the Euclidean distance method (Equation 1) between data samples, as proposed by MacQueen (1967) and widely used for clustering large data sets (Jehn et al., 2020). This algorithm divides data into *k* sections (user choice), defines an initial centroid value for it, and randomly selects and assigns the samples (e.g., catchments) to one of the *k* clusters. Then, the distance between each sample to the centroid of each cluster is computed, and, after a given number of continuous iterations, when all samples are within one of the *k* groups, it results in an optimal cluster solution, converging to a local minimum of the criteria parameters (Marutho et al., 2018; Nainggolan et al., 2019; Shi et al., 2010; Syakur et al., 2018).

$$d_{(x_i,y_i)} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i\right)^2} \tag{1}$$

where x is the position of data in each cluster, and y is the center of each cluster.

As mentioned above, the number of clusters needs to be defined as a priori. In some cases, when the user has expertise in the field, it is acceptable to specify a fixed number of clusters. In our study, however, we automatically determined the number of groups using the Elbow approach, which minimizes human subjectivity in the clustering analysis. In this context, the Elbow method is one of the most popular approaches to determining the optimal number of clusters (k) for a given data set. It is based on the notion that adding the value k of clusters beyond an optimal k will not significantly contribute to the analysis. Starting from 1, the value of k is added one by one, looking for the percentage of variance explained as a function of k. During the process, the Sum of Squared Error (SSE—Equation 2)—the sum of the average Euclidean distance of each sample to the centroid—is recorded for each step. When the SEE drops significantly, forming a small angle, the value of k is achieved. SEE is calculated as follows:

Table 2
Catchment Attributes Chosen for Dominant Attributes Analysis

Class	Attribute	Long name	Unit	Source
Topography	Area	Area of the catchment	km^2	Almagro et al. (2021)
	Elevation	Mean elevation of the catchment	m	Almagro et al. (2021)
	Slope	Mean slope of the catchment	%	Almagro et al. (2021)
Climate	Aridity	Aridity index	–	Almagro et al. (2021)
	Precip. seasonality	Seasonality of the precipitation	–	Almagro et al. (2021)
Soil	Soil sand	The sand fraction in the soil	%	Hengl et al. (2017)
	Soil carbon	The carbon fraction in the soil	%	Hengl et al. (2017)
	Soil depth	Depth of the soil to the bedrock	m	Hengl et al. (2017)
Geology	HAND	Height above the nearest drainage	–	Nobre et al. (2011)
	Porosity	Subsurface soil porosity	–	Gleeson et al. (2014)
	Permeability	Subsurface soil permeability	–	Gleeson et al. (2014)
Land-use	Forest	Forest fraction of cover	%	Buchhorn et al. (2019)
	Bare soil	Bare soil fraction of cover	%	Buchhorn et al. (2019)
	Grass	Grass fraction of cover	%	Buchhorn et al. (2019)
	Summer NDVI	NDVI in DJF (rainy season)	–	Buchhorn et al. (2019)
Anthropogenic	Disturbance	Hydrological disturbance index	–	Almagro et al. (2021)

$$\mathrm{SSE}_i = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_k \right\|^2 \tag{2}$$

where x is the data in each cluster; c_k is the kth cluster.

To minimize the uncertainties and produce reliable groups, we preprocessed the data by normalizing the hydrologic signatures, due to their different units, by a min-max scaling. The parameters for the initialization of the cluster analysis were also set so that the catchments of a given group need to explain at least 80% of the hydrologic signatures' internal variance through 10 individual repeats by the employment of the Elbow method. To ensure the stability of the clusters generated, we performed a cross-validation test, which used different resampled portions of the data set (with different sizes: 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, and 10%) to train and evaluate the clustering model.

2.4. Characterizing Groups by Their Streamflow Signatures

We employed a PCA on the streamflow signatures to identify and interpret the hydrological signatures that best characterize each hydrological group. Similar approaches have been used for identifying the main characteristics and governing processes in the hydrologic sciences (Haag & Westrich, 2002; Jehn et al., 2020; Singh et al., 2009; Vergopolan et al., 2022). The PCA is based on the correlation matrix between the input variable (here the streamflow signatures) and in the loading matrix by the principal components, which results in a set of uncorrelated variables (the principal components) that can be used to explain most of the variation in the original data. The first principal component is the linear combination of the streamflow signatures that captures the highest level of variation in the data. The second principal component, independent (and uncorrelated) of the first one, captures the highest level of residual variance. We only evaluate the two principal components, which described at least 60% of the internal variance together.

2.5. Identifying the Drivers of Streamflow Variability

The most influential attributes of the hydrologic signatures were also assessed in our analysis. Among the >100 catchment attributes of the CABra data set, a subset of 16 attributes were selected (Table 2) to perform a recursive feature elimination using the random forest technique, as described in Vergopolan et al. (2021). The attributes were selected based on their representativeness of the attribute class by the employment of a recursive feature elimination and, as reported in the literature (Addor et al., 2018), some of the attributes presented in the CABra data set were excluded to avoid redundant information through the analyses, respecting the following criteria: (a) at least one attribute from each attribute class; (b) it must be a numeric attribute; and (c) most of processes and features of the attribute class represented. The most influential catchment attributes among each class of attributes, represented by the results of a preprocessing (recursive feature elimination) were selected (see Table S1 in Supporting Information S1 for the complete list of catchment attributes in CABra data set).

For topographical features, we specifically chose catchment area, mean elevation, and slope, as they were the only numerical attributes influencing the model's estimation of hydrological signatures (with feature importance >0). When considering climate attributes, the aridity and seasonality of precipitation emerged as the most influential. This is unsurprising given that aridity and seasonality are widely recognized as primary controls on long-term water balance (Budyko, 1974; Woods, 2009). In terms of land cover, we focused on the three most prevalent types in Brazil: forest, grass, and crops. Additionally, we included the Normalized Difference Vegetation Index (NDVI) in the summer, which correlates with increased biomass (and consequently leaf area index) due to Brazil's typical rainy season during this time. Soil attributes were represented by soil texture, with sand and carbon fractions selected as the most impactful. Soil depths were also included, reflecting soil water storage capacity, which significantly influences water partitioning and catchment functioning. From the geological perspective, we opted for Height Above the Nearest Drainage (HAND) as it offers easy access and applicability, requiring only the application of the HAND equation to any elevation map. Moreover, HAND is highly correlated with water table depth, as presented in the CABra data set. Static and dynamic well levels were excluded due to reported inconsistencies in records (Almagro et al., 2021). Finally, to account for human impacts on catchments, we used the Hydrological Disturbance Index (HDI), developed by Almagro et al. (2021), which integrates all relevant attributes in this category.

2.5.1. Random Forest Analysis

Random forests are a machine learning algorithm that relies on numerous regression trees to generate an ensemble of predictions (Breiman, 2001; Wang et al., 2015). Random forest regressor algorithms, for instance, relate predictors to a response variable. In this paper, we consider the catchment attributes as the predictors and the hydrological signatures as the response variables. Once the regression tree is grown, each predictor (catchment attribute) is randomly shuffled through the regression, and the prediction is performed. Then, prediction accuracy is assessed by removing one-by-one the predictors, indicating how much a given catchment attribute is essential to the hydrological signature prediction, in a methodology called Recursive Feature Elimination (RFE) which indicates the most relevant and redundant predictors (Vergopolan et al., 2021). For each predictor removed, the prediction's mean squared error (MSE) is assessed. The more the error (expressed in MSE) when an attribute is removed, the more important and influential it is to that hydrological signature. This way, we can assess the most important features (or attributes) on the hydrological signatures of each catchment group, making it possible to identify the main drivers of streamflow variability.

The advantages of using random forest algorithms applied to large-sample data sets can be summarized in four main aspects (Addor et al., 2018; Tyralis et al., 2019): (a) random forests are adapted to capture non-linear relationships between the hydrological signatures and catchment attributes; (b) physical principles do not limit random forests, making it possible to identify unknown relationships which would be impossible by traditional hydrological modeling; (c) as random forests use an ensemble of regression trees, there is a lowered chance of data overfitting when using big training data; and (d) and perhaps more critical, random forests are interpretable, based on the feature importance of each predictor. Regarding the final aspect, it is important to state that the interpretability of random forests is completely based on insights provided by the influence of each variable in the model's prediction, and unveiling various process complexity and learn the physic of hydrological processes still need some kind of coupled framework (De la Fuente et al., 2023; Feng et al., 2022).

Here, we used random forests to assess the most influential catchment attributes to the hydrological signatures. We used 100 trees to ensure convergence, avoid overfitting, and produce the best transferability and generalization. As our main goal here is not to predict or simulate the hydrological behavior but rather to understand the system, we used the entire data set in the random forest regression analysis training to ensure that all the hydrological information would be considered for the identification of the main drivers. The evaluation of the random forest algorithm will be illustrated by the r-squared and mean-squared error metrics.

10.1029/2024WR037212

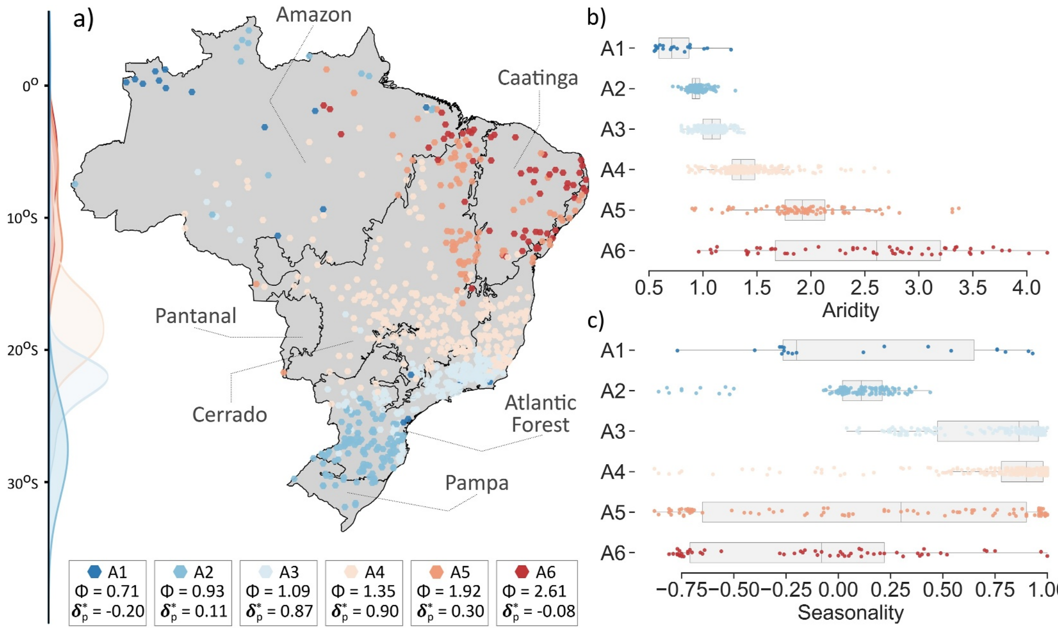

Figure 2. Clusters of catchments with similar hydrologic behavior. Panel (a) shows their spatial distribution over Brazilian biomes (Amazon, Atlantic Forest, Cerrado, Caatinga, Pampa, and Pantanal) along with each catchment group statistic, and a *y*-axis showing the adjusted curve for the latitudinal histograms for each group. Panels (b and c) show their within-clustered variability over aridity index and seasonality, respectively. Φ is the long-term median aridity index; δ_p^* is the long-term median seasonality of precipitation.

We also calculated each attribute class's importance on each catchment group's hydrological behavior, to provide an overall view. To do so, we summed all the importance resulted from the recursive feature elimination from attributes of the same class and then divided the resulting value by the number of attributes of the class. The relative importance of each attribute class to the sum of all importance of the hydrological signatures was calculated by Equation 3.

$$Imp_{class} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} imp_i^j}{k} \tag{3}$$

where Imp_{class} is the weighted importance for a given attribute class, k is the number of attributes in each attribute class, and n is the number of hydrological signatures.

3. Results and Discussions

3.1. Catchments With Similar Hydrologic Behavior

Our classification of Brazilian catchments based on the *k*-means method optimized by the Elbow approach resulted in six main groups, as shown in Figure 1a (and Figure S1 in Supporting Information S1). Through cross-validation using various sample sizes of catchments, we demonstrated that the proposed classification system is stable. We found that catchments exhibit not only a similarity pattern in streamflow data but also a spatial grouping pattern, regardless of the amount of available data (Figures S2 and S3 in Supporting Information S1).

Figures 2b and 2c show the distribution of the aridity index ($\varnothing = \overline{Ep}/\overline{P}$) and precipitation seasonality ($\delta_p^* = \delta_p \cdot sgn(\Delta_T) \cdot \cos(2\pi(S_P - S_T)/\tau)$) within each group, which are known to be the main controls of the long-term water balance (Budyko, 1974; Woods, 2009). The aridity index is the ratio between mean-annual potential evapotranspiration ($\overline{Ep}$) and precipitation ($\overline{P}$), indicating how much of the water input can be lost to the atmosphere. At the same time, precipitation's seasonality gives insights into precipitation seasonality, if in-phase with energy demand (indicating less storage) or out-of-phase with energy demand (indicating greater

10.1029/2024WR037212

storage). This index combines important details about the amplitude (δ_p and Δ_T) and phase (S_P and S_T) of temperature and precipitation during the annual cycle (τ) into a singular variable. The processes represented by the two indices regulate the water partition, storage, and release in the catchment scale.

We found that the catchment grouping matches an aridity gradient (from the wettest to the driest), which showed to be the most valuable basis for explaining variations between catchment hydrological behaviors (Addor et al., 2018; De la Fuente et al., 2023; K. Sawicz et al., 2011). Even though one can expect this behavior, it is essential to note that no information about the aridity was provided to the clustering algorithm, which used only the streamflow signatures. Thus, we first labeled the groups based on their aridity level (e.g., aridity level 1 = A1, aridity level 2 = A2, and so on), resulting in groups A1, A2, A3, A4, A5, and A6. Even if the resulting groups did not follow any pattern or gradient of the precipitation seasonality, it is possible to infer some interesting results of groups A2, A3, and A4. For all the streamflow signatures used, the clustering method was effective, and the resulting clusters accurately reflect the overall characteristics of their respective catchment. This is demonstrated in Figure S4 of the Supporting Information S1, where all the streamflow signatures are plotted against the mean daily flow—which is also a function of the aridity (Meira Neto et al., 2020)—indicating low dispersion of the catchments to the center point of the group.

Group A1 comprises 19 catchments in the Amazon and Atlantic Forest, presenting the lowest median aridity index among all groups ($\varphi = 0.71$). Group A2 also presents an energy-limited condition ($\varphi = 0.93$) but differs from other groups in the seasonality of the precipitation ($\delta_p^* = 0.11$), not presenting a rainy/dry season (and also, it is not in-phase or off-phase with temperature). This group comprises 127 catchments in southern Brazil, in the Pampa and the Atlantic Forest biomes, with a few in the Amazon. Group A3 consists of 174 catchments in the Atlantic Forest, the Cerrado, and the Amazon biomes. In median terms, the catchments in this group are in-phase ($\delta_p^* = 0.87$) and present water-energy equilibrium ($\varphi = 1.09$), with the aridity index ranging from 0.8 to 1.5. Group A4 consists of 269 in-phase catchments in a water-limited condition ($\delta_p^* = 0.90$ and $\varphi = 1.35$) mainly located in the Brazilian Cerrado but also includes catchments in Atlantic Forest and Amazon. Group A5 comprises 89 catchments primarily located in the northeastern portion of Brazil. Nevertheless, A5 catchments are distributed across five out of the six Brazilian biomes, exhibiting a notable concentration along the interface between the Cerrado and the semi-arid region. Aridity is high in A5, with median $\varphi = 1.92$. Finally, the A6 group presents 57 catchments and the highest median aridity index among all the clustered groups ($\varphi = 2.61$). Groups A1, A5, and A6 did not show a clear seasonality pattern, even though the distribution shows us that individual catchments spam over a wide range of seasonality. We can also note that this transitional behavior is also observed throughout the Brazilian territory, where catchment grouping varies through the latitudinal gradient, from Group A3 to Group A5, with Group A4 in the middle. As reported by K. Sawicz et al. (2011), we also noted the critical role of catchment proximity for the catchment similarity through Brazilian catchments (Figure 2), even if the same is not valid in other regions (Jehn et al., 2020).

Each group's climatological and hydrological behavior is presented in the context of the Budyko space (Budyko, 1948), considering the relationship between the long-term water balance components (Figure 3). The majority of catchments in group A1 are in an energy-limited condition, below the Budyko curve, with up to 0.4 of the evaporative index (E/P), indicating that a more portion of the precipitation is available to generate streamflow, likely to exceed evaporation rates. While the aridity index values for groups A2 and A3 are quite similar, the primary difference lies in the mean evaporative index. Specifically, A3 exhibits higher values of E/P, indicating that more water is evaporated than is converted into streamflow. Lower amounts of precipitation in A3 can be inferred as the cause of this behavior. Groups A4, A5, and A6 exhibit higher and more scattered mean values of the aridity index, indicating a more arid climate. However, they also display an increasing gradient in the evaporative index, similar to that of A1, A2, and A3. Catchments in A4 fall below the Budyko curve, whereas those in A5 and A6 do not follow this pattern. Although there is a large variability and a wide range of aridity values among the catchments in group A6, they are located on the upper portion of the Budyko space. High values of E/P are found in A6 (of at least 0.8), indicating that most of the precipitated water is evaporated in long-term means, being highly water limited and vulnerable to water scarcity.

According to the PCA (Figure 4), the two main components explain 63% of the streamflow signatures in Brazilian catchments. The first principal component accounts for 41% of the signature's variance. It strongly correlates with the frequency and duration of high-, low-, and zero-flow events, elasticity, and coefficient of variation. This component can be viewed as a measure of the variability of the streamflow in Brazil. The second principal

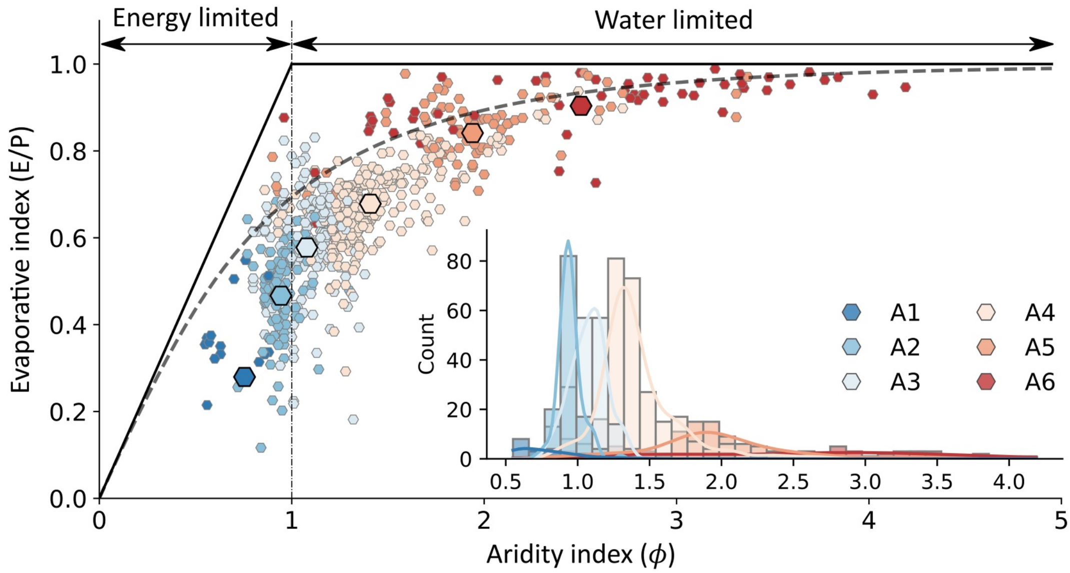

Figure 3. The distribution of the Brazilian catchments (color-coded into the six groups—A1, A2, A3, A4, A5, and A6) in the context of the Budyko relationship. The solid lines are representing the energy and water limits to the evaporative index, while the dashed line represents the theoretical Budyko curve. The inset shows the histogram of the aridity index. Regular hexagons are the catchments among each group, while large hexagons are the median values of each group.

component accounts for 22% of the variance in hydrological signatures. It increases with signatures such as streamflow's 5th and 95th percentiles, half-flow day, mean streamflow, and annual runoff ratio. This principal component can be seen as a measure of the overall streamflow behavior (mean, low-flow, and high-flow) and if the mean values are maintained by baseflow or runoff.

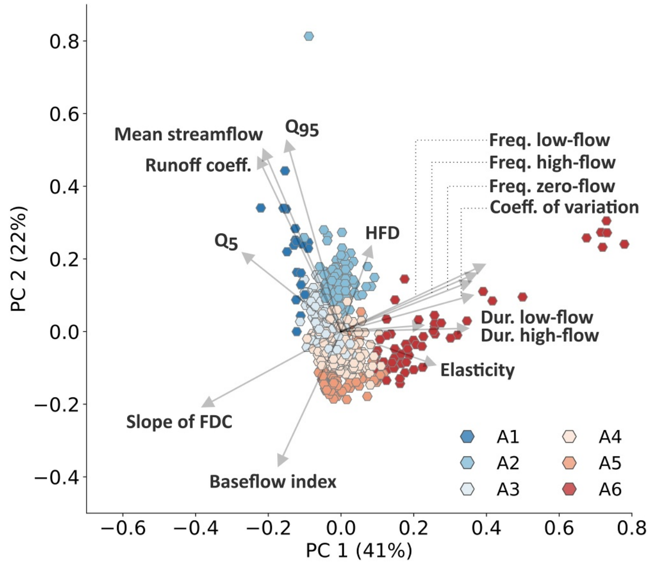

Figure 4. A biplot of the two principal components of the hydrological signatures of Brazilian catchments, grouped by the six hydrological clusters (A1, A2, A3, A4, A5, and A6). The biplot shows the relationship between the principal components and the hydrological signatures, aside from showing how the hydrological signatures correlate between themselves. The dots are the principal component scores of each catchment, and the vectors are the principal component loadings of each catchment. The angle between the vectors indicates how correlated they are. The closer the angle, the more correlated signatures are. The larger the angle, the less correlated signatures are. An angle of 90° indicates the signatures are not correlated.

Groups A1, A2, and A3 hydrologic signatures are mainly subjected to distribution characteristics of streamflow, presenting the high values of mean streamflow, 5th and 95th quantiles, and annual runoff ratio. Group A6, in turn, shows high positive values for the variables strongly correlated with the second principal component, which indicates enhanced streamflow variability in this group. There is also a subset of A6 with exceptionally high-frequency values of high-, low-, and zero-flow. Groups A4 and A5 are in the middle of the biplot and opposites to the loading vectors, between extreme values of the streamflow signatures, indicating that the catchments within these groups present low signature values. These groups resemble, differentiating on the magnitude of mean streamflow and baseflow index. Groups A4 and A5 are transitional between the wettest and driest catchment groups.

3.2. The Hydrological Behavior of Brazilian Catchments

Regime curves for precipitation, actual evapotranspiration, streamflow, and storage change for the six groups are shown in Figure 5, and their hydrological signatures are presented in Figure 6. As mentioned previously, aridity appears to be the main physio-climatic control of catchment classification based on hydrological signatures. In this way, our discussion will follow the distinct characteristics of each group along the aridity gradient.

A1, A2, and A3 represent the wet groups. Their catchments are energy-limited (A1 and A2) and at the threshold between energy and water-limited conditions (A3) (see Figure 3), having the largest values of mean daily streamflow, annual runoff ratio, and 95th percentile of streamflow (Q_{95}),

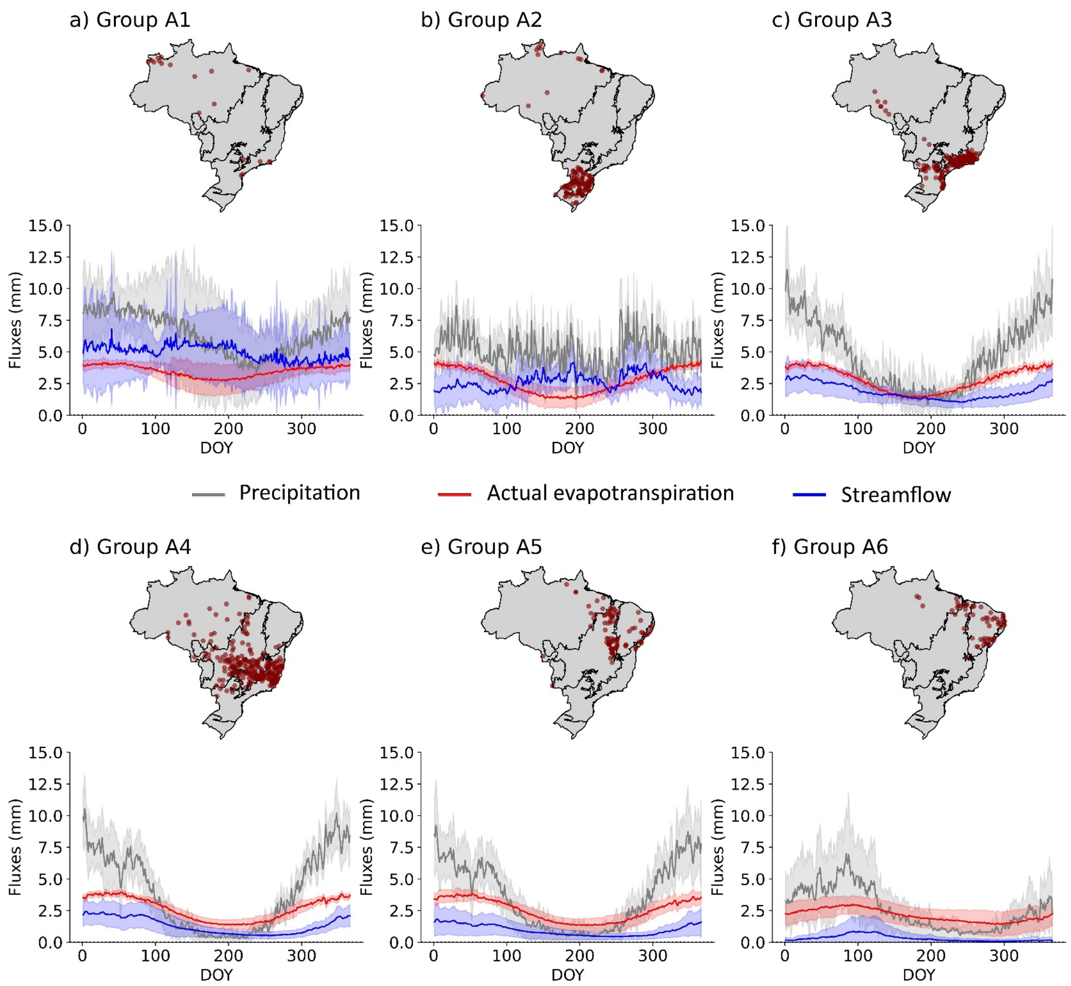

Figure 5. Spatial distribution (upper panels) and daily regime curves (lower panels) from water balance components (precipitation, actual evapotranspiration, and streamflow) for each of the hydrological groups. (a) Is the A1 group; (b) is the A2 group; (c) is the A3 group; (d) is the A4 group; (e) is the A5 group; and (f) is the A6 group. All the units are in mm. To generate the regime curves for each group, we have used precipitation (P), actual evaporation (E), and streamflow (Q).

albeit in descending order following their aridity values (Figure 4). A1 has the largest mean daily precipitation, being characterized by streamflow consistently larger than evapotranspiration all year and is referred to further as "perennial wet." A2 also presents high streamflow values, not as high as A1, but greater than A3, referred to as the "transitional wet" group. Precipitation seasonality is the climate variable that distinguishes A2, with median seasonality close to zero and slight variation, indicating no clear relationship between the precipitation and temperature annual cycles, with no well-defined rainy and dry seasons. This establishes a direct link between A2's response and actual evaporation, such that a decrease in evapotranspiration leads to a corresponding increase in streamflow. On the other hand, Group A3 presents the most distinctive in-phase (high seasonality) regimes of precipitation and evapotranspiration of the wet catchments, referred to as the "seasonal wet" group. A3 catchments exhibit one of the shortest half-flow dates, the lowest streamflow elasticity, and noteworthy low-quantiles (Q_5) of streamflow, indicating that for most portion of the year, their catchments preserve water in the soil, keeping the baseflow and, consequently, perennial streamflow (Figure 5). All three groups are also mainly covered by forests (Atlantic and Amazon Forest biomes). Such catchments consist of perennial rivers with similar streamflow elasticity values. The main observed streamflow feature is that Group A2 has an overall more considerable streamflow variability characterized by larger median values of low and high-flow frequencies, steeper FDC slopes, larger coefficient of variation, and lower BFI, indicating less groundwater contribution. The latter can also be inferred by A2 having lower Q_5 values than groups A1 and A3.

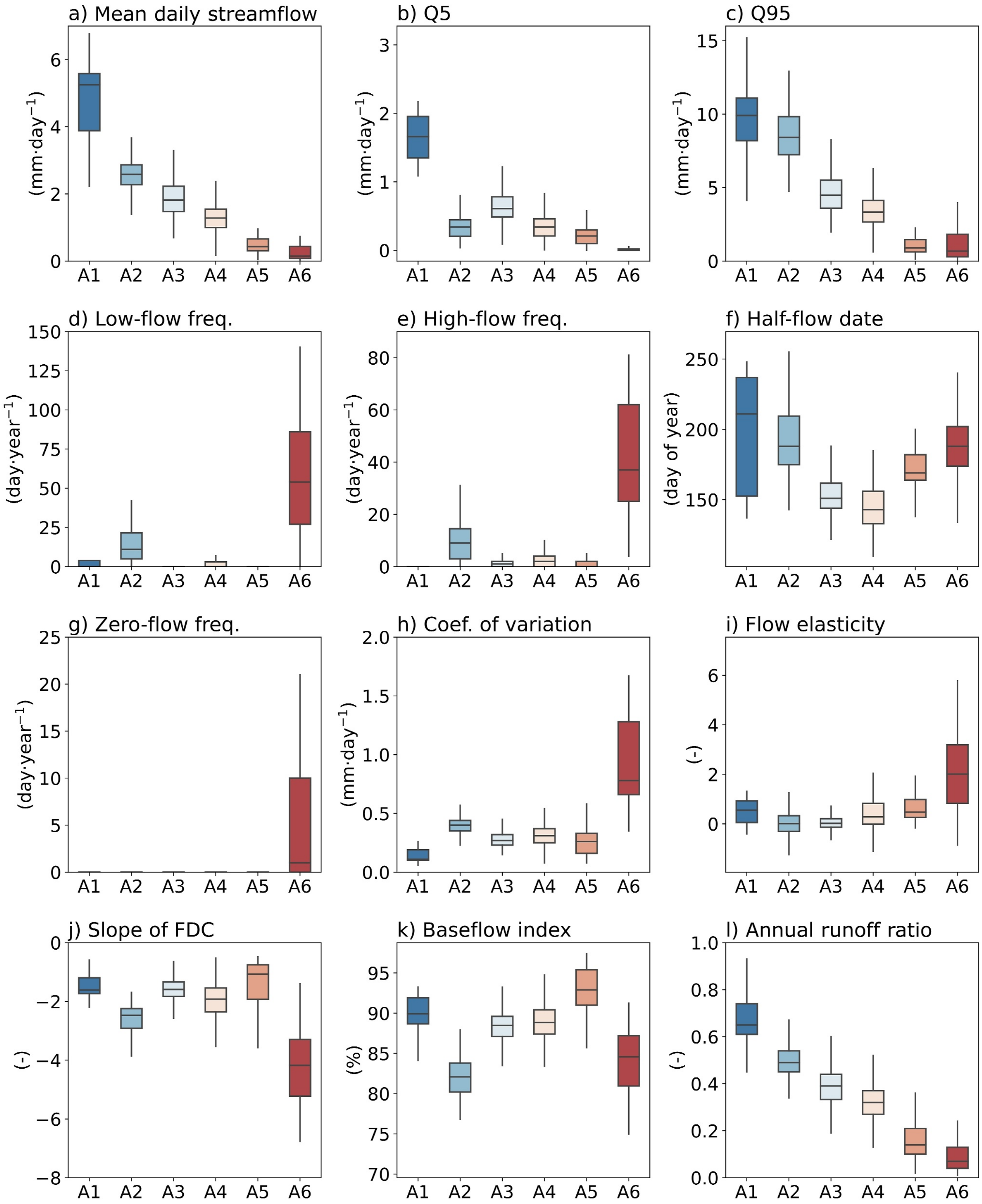
a) Mean daily streamflow
(mm·day^{-1})
b) Q5
(mm·day^{-1})
c) Q95
(mm·day^{-1})
d) Low-flow freq.
(day·year^{-1})
e) High-flow freq.
(day·year^{-1})
f) Half-flow date
(day of year)
g) Zero-flow freq.
(day·year^{-1})
h) Coef. of variation
(mm·day^{-1})
i) Flow elasticity
(-)
j) Slope of FDC
(-)
k) Baseflow index
(%)
l) Annual runoff ratio
(-)
A1 A2 A3 A4 A5 A6

Figure 6.

A4, A5, and A6 represent the dry catchments of Brazil. During the dry season, there is more bottom-up water movement through evapotranspiration than precipitation in these groups (Figures 5d–5f). This occurs because most of the storage is held in the unsaturated zone as capillary water, which evaporates in the interstorm period (Wagener et al., 2007), leading to low streamflow values. They show a similar trend of increasing aridity leading to decreasing values of mean streamflow, annual runoff ratio, $Q_{5,}$ and Q_{95}. Similar to the pattern seen in the first three humid groups, one of the groups stands out, in this case, group A6, with higher variability as seen in the coefficient of variation, the slope of the FDC, low and high-flow frequencies, as well as in the zero flow frequencies, resulting in the classification of these groups as the "intermittent dry" catchments. It is worth mentioning that although A6 is the only group that displays intermittent streamflow, not all its catchments are of this kind. In this context, a lower BFI seems to be correlated with increased variability, with A6 exhibiting the lowest BFI among the various groups. Group A6 is also characterized by the largest elasticity of all the six groups, which is somewhat expected from catchments in arid climates, being a result from the large interannual variability of precipitation. Groups A4 and A5 are very similar, with wetter conditions for the A4 (runoff ratio, mean daily streamflow, and daily streamflow percentiles). A4 also presents a pronounced precipitation seasonality, referred to further as the "seasonal dry" group. Despite exhibiting a hydrological behavior similar to Group A4, where streamflow is predominantly influenced by low precipitation and high evapotranspiration, Group A5 is generally drier, functioning as an intermediate group between Groups A4 and A6, referred through the text as "transitional dry." The dry groups are located in a spatial gradient that starts in the Atlantic Forest, passes through the Cerrado, and ends in the Caatinga biome. One of the main features observed is that A5 is located on the border of the Cerrado and Caatinga biome, acting as a transitional group between A4 and A6, not only in the aridity gradient but also in the biome features (such as landscape and biodiversity, as seen in Figure S5 of the Supporting Information S1), indicating some degree of coevolution between climate and vegetation.

3.3. The Dominant Attributes of Streamflow Variability in Brazil

This section presents the results and discusses the main drivers of streamflow variability in Brazilian catchments. First, the results from the random forest analysis followed by the recursive feature elimination, showing the attributes' importance grouped by attribute classes, are shown in Figure 7. Those results are intended to provide an overall view of the controls on catchment behavior in our data set. To aid the discussion, we show in more detail the influence of each catchment attribute on the hydrologic signatures of the groups through heatmaps describing both the importance and correlation between the catchment attributes and the hydrological signatures of the A1–A6 (Figure 8 and Tables S3–S9 in Supporting Information S1). The random forest algorithm performed well in representing each of the streamflow signatures using the catchment attributes as predictors, as can be seen in Table S2 of the Supporting Information S1, with a mean $R^2 = 0.76$ for A1, $R^2 = 0.84$ for A2, $R^2 = 0.76$ for A3, $R^2 = 0.80$ for A4, $R^2 = 0.83$ for A5, and $R^2 = 0.77$ for A6. Overall, we found $R^2 > 0.70$ for 82% of the signatures analyzed. It was also noted that the attributes used here are not good predictors of low-flows and zero-flow through the groups.

By analyzing the interactions between hydrological processes and their impact on catchment functioning (L'vovich, 1979), we found that distinct attribute classes exhibit varying degrees of influence on hydrological behavior within different groups. Figure 7 shows two distinct patterns controlling the water fluxes in Brazilian catchments, one for the wet groups and another for the dry groups. Groups A4, A5, and A6—the dry groups—have climate as the main driver of the hydrological signatures' variability. It is not surprising that climate is the most important attribute controlling the water fluxes since the catchments within these groups are mainly water-limited. This means that the water fluxes in these catchments are primarily subjected to the availability of water inputs. Once the water is available, could contribute significantly to runoff generation. On the other hand, groups A1, A2, and A3—the wet groups—are mainly energy-limited, making it possible for other attributes to control the water fluxes on their catchments. However, the climate still presents a great influence among the attributes. The soil and topography also exert important control over these groups, highlighting the importance of the soil water

Figure 6. Boxplots of the streamflow signatures within each hydrological group. (a) Is the mean daily streamflow, in mm.day^{-1}; (b) is the 5th percentile of daily streamflow, in mm.day^{-1}; (c) is the 95th percentile of daily streamflow, in mm.day^{-1}; (d) is the frequency of low-flows, in days.year^{-1}; (e) is the frequency of high-flows, in days.year^{-1}; (f) is the half-flow day, in the day of the year; (g) is the frequency of zero-flows, in days.year^{-1}; (h) is the coefficient of variation of daily streamflow, in mm.day^{-1}; (i) is the streamflow elasticity to the precipitation, dimensionless; (j) is the slope of the flow duration curve (FDC), dimensionless; (k) is the baseflow index, dimensionless; and (l) is the annual runoff ratio (annual runoff divided by annual rainfall) of the catchment, dimensionless.

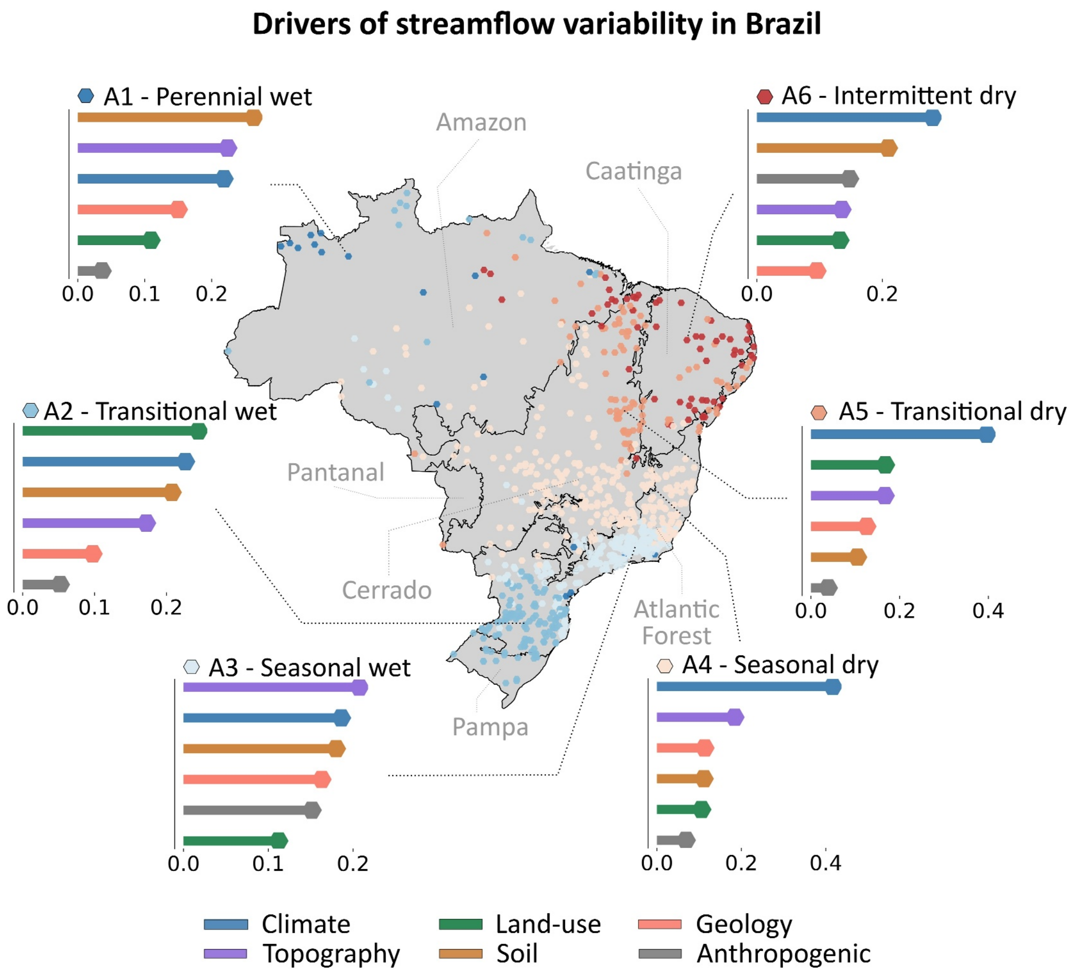

Figure 7. Attributes' importance on controlling the hydrological behavior in the Brazilian catchments. The *x*-axis in the plot is the percentage of streamflow variability explained by each attribute class.

movement on the streamflow variability. It is worth noting the relatively strong influence of the anthropogenic attributes on the hydrologic behavior of the A3 and A6 groups, against very low (almost none) influence in the other groups. This feature's importance in group A3 can be explained by the high population density within their catchments, which are in the country's most populated and developed area, and require large amounts of water for human consumption, industries, and irrigation. In turn, A6 has its catchments within the Brazilian semi-arid region, which required the construction of thousands of reservoirs to store the water, which is responsible for high water regulation in those catchments. One can expect the geology to be more relevant to the water fluxes among the groups of catchments, even though this is not observed through the analysis performed here. This may be due to the geology information included in the large-sample data set. The attributes contained here (subsurface porosity and subsurface permeability) are based on coarse global estimations (Gleeson et al., 2014) that use insufficient validation points worldwide, especially in Brazil and South America.

In addition to presenting an overall hydrological behavior, some specific drivers contribute to each streamflow signature within each hydrological group. Below, we explore the correlation and influence of catchment attributes on streamflow signatures, from the wettest to the driest group, to assess the internal controls of hydrological behavior.

3.3.1. Group A1—Perennial Wet Catchments

Although we consider the A1 group as a wetter condition of A2 and A3, there are differences in the main drivers of the streamflow variability in this group, as seen in Figure 8a. As the water is available in abundance and there is

Figure 8. Diagram of the main controls (catchment attributes on the *x*-axis) of hydrological behavior (streamflow signatures on the *y*-axis) of catchment groups A1, A2, A3, A4, A5, and A6. The size of each hexagon indicates the importance of a given catchment attribute to a given hydrological signature (the larger the circle, the greater its importance). The color of each hexagon indicates the product-moment Pearson correlation coefficient between a given catchment attribute to a given streamflow signature of the catchments within a group.

not enough energy throughout the year, the climate does not play the primary role in the streamflow variability across these catchments. In turn, landscape features take place in controlling the hydrological variability of the catchments in this group. Soil depth is the primary control of the mean streamflow, Q_{95}, frequency of high-flows, and runoff ratio. It is inversely correlated to these streamflow signatures, implying that deeper soils attenuate streamflow events by storing the water in the subsurface. In turn, topography attributes are good indicators of the coefficient of variation, baseflow index, and slope of FDC. The size of the catchment area is inversely correlated with the streamflow coefficient of variation, indicating that larger catchment areas respond more slowly to precipitation inputs, influencing in the overall streamflow variability. Conversely, the catchment area is positively correlated with the baseflow index and the slope of the FDC, indicating variability of specific aspects of streamflow. These correlations suggest that the slow response of larger catchment areas allows water to infiltrate and flow in the subsurface, contributing to baseflow and shaping the slope of the FDC. Still, regarding the frequency and duration of high- and low-flow events, there is a great influence of the height above the nearest drainage (HAND), indicating that water table depth acts as a dampener. Our results reinforce the idea that landscape features (such as topography, geology, and soils) are of central importance in controlling the hydrological behavior of energy-limited catchments (A1–A3) due to their capabilities of controlling the seasonality of the climatological input.

3.3.2. Group A2—Transitional Wet Catchments

The diagram relating the catchment attributes' importance to the streamflow signatures of the A2 group is presented in Figure 8b. The hydrological behavior of the group is primarily influenced by land use, as evidenced by strong correlations between land-cover attributes and various hydrological variables. Specifically, the fraction of bare soil shows a strong direct relationship with the coefficient of variation, runoff ratio, Q_{95}, and mean daily streamflow. Conversely, a strong inverse relationship is observed between the percentage of grass and Q_5, the baseflow index, and the slope of the FDC, while a positive correlation is found for the frequency of high-flow events. These connections suggest that the conversion of natural land cover to non-native vegetation (grass and bare soil) promotes runoff, increases streamflow variability, flood generation, and, in the long-term, mean streamflow. Additionally, reduced infiltration results in less water available for subsurface flow, further altering the shape of the FDC and the baseflow index. The climate is the second most important attribute class affecting streamflow, relying on the strong importance of the aridity index and the precipitation seasonality to the long-term mean streamflow and the half-flow date, respectively. This is due to the uniformly distributed precipitation throughout the year, which means that changes in aridity (caused by evapotranspiration) directly impact the water partitioned available to streamflow. Also, catchments with seasonal precipitation (in-phase or out-of-phase behavior), among the non-seasonal overall behavior of the group, can significantly impact the temporal evolution of annual streamflow accumulation.

3.3.3. Group A3—Seasonal Wet Catchments

The A3 group is on the transition between water- and energy-limited. Topography and soil attributes are important to water partitioning in this group (Figure 8c). Landscape features, such as topography and soil properties, are important factors driving hydrological signatures such as the mean streamflow, percentiles (Q_5 and Q_{95}), frequency and duration of low-flows, streamflow elasticity, and coefficient of variation. Catchments within this group have the shallowest soil depth compared to all other groups. This limitation on water storage increases streamflow elasticity, and consequently, the coefficient of variation, reflecting the streamflow's responsiveness to precipitation changes. A substantial fraction of carbon in the soils also favors the mean streamflow, $Q_{5,}$ and Q_{95}. It helps to store water in the soil (Jha et al., 2022; Minasny & McBratney, 2018; Yost & Hartemink, 2019), increasing streamflow during drier periods and improving overall water availability. Those characteristics reinforce the importance of the surface and subsurface layers in controlling and attenuating the quick variations in water input (precipitation) by storing the water through the soil layer. Group A3 exhibits the strongest influence of the attributes related to the anthropogenic intervention compared to all groups, represented by the hydrological disturbance index. Those attributes mainly affect the mean and percentiles of streamflow, with a negative correlation, especially by the regulation capacity of reservoirs and high demand of water, which may impact the water availability to streamflow. The influence presented here by the climate is less relevant for the streamflow than in other groups (like A4 and A5), mainly because the available energy for evaporation is lower than the precipitation all over the hydrological year. However, the influence of the aridity index and precipitation seasonality on mean streamflow and half-flow date is significant, indicating that in-phase dry season reduces the streamflow. Consequently, the streamflow is a direct response to precipitation, following the same pattern of the annual cycle.

3.3.4. Group A4—Seasonal Dry Catchments

The hydrological behavior of A4 is primarily influenced by climate (Figure 8d). This is demonstrated by the significant impact of precipitation seasonality and aridity index on all streamflow signatures, with a potential influence of up to 80% on mean daily streamflow, as demonstrated by the feature importance analysis. In essence, the hydrological behavior of this group is entirely dependent on the portion of precipitation that is not evaporated, which is typically a small fraction of total precipitation. Topographic slope also plays a significant role in shaping hydrological signatures, with an indirect correlation observed between slope and Q_5, half-flow day, and baseflow index, and a direct correlation between slope and coefficient of variation. Steeper slopes increase runoff, decrease BFI (and low-flows), and contribute to flooding events, leading to more streamflow variability across catchments.

3.3.5. Group A5—Transitional Dry Catchments

The mean streamflow from the A5 group is dominated by the climate, as seen through the high importance shown by the aridity index (with negative correlation), as indicated in Figure 8e. The high evaporative demand of their catchments stresses out the streamflow signatures, particularly during the dry season. Aside from the mean streamflow, the aridity index is the primary control of most of the hydrological signatures from A5, showing the importance of climate to this kind of catchment. Although our analysis shows that forest cover delivers more river streamflow (positive correlation between forest cover and mean daily streamflow), an interesting correlation exists between the forest cover fraction and some signatures. Q_5 inversely responds to the presence of forest, while a direct response is verified in the low-flow frequency and duration. This may be due to the high values of actual evapotranspiration in these catchments, generated by tall trees (typical from forests), which present more canopy interception that increases the evaporation in drier periods (A. F. Rodrigues et al., 2021) and developed an efficient root system to achieve deep soil moisture for their survival (Harper et al., 2014). The water achieved in deep soil layers evaporates by transpiration, lowering the soil water available to recharge and streamflow (Anache et al., 2019; Oliveira et al., 2017). Alternatively, it may be due to specific catchments already being in wetter conditions and presenting a more significant forest cover fraction as an adaptation to the local climate. Attributes related to the topography of the catchments are mainly important for the streamflow variability, the slope being one of the main drivers of low-flow frequency, half-flow date, elasticity, and BFI. Catchments with flat topography, like A5, tend to hold water in the most superficial layers of soil, which is quickly evaporated during interstorm and doesn't reach deeper layers, contributing less to the BFI. As a result, the reduced water availability in the subsurface increases the frequency of low-flow events. Although it presents the highest mean hydrological disturbance index across all groups, it is primarily associated with non-natural vegetation cover. It is, therefore, less critical to streamflow variability compared to other groups.

3.3.6. Group A6—Intermittent Dry Catchments

Group A6 presents more complexity than other groups (Figure 8f). In this group of catchments, most of the time, the precipitated water remains on the most superficial layers of the soil, in the unsaturated zone, being evaporated due to the high amounts of energy available, similar to A5. Through the competition of hydrological processes mediated by the climate and landscape, the climate limits the hydrological behavior of this group during a significant part of the year: there is no water available to generate runoff (which is rationed by the aridity index), resulting in a high frequency of zero-flow—intermittent regime—that characterizes this group. Other characteristics mediate the hydrological behavior when water becomes available. Soil properties, such as soil depth and carbon fraction considerably influence the Q_5 and Q_{95}. Among the groups, A6 holds one of the deepest soils, contributing to sustaining water flow (if water inputs are available) during a portion of the year due to its higher water storage capacity. The impact of the forest coverage fraction on streamflow signatures is noteworthy. Catchments with the highest forest fraction among the A6 group also presented the less dry conditions. This could be attributed to vegetation characteristics, such as studies suggesting that plant stomata close during arid air conditions, preventing transpiration (Werth & Avissar, 2004). Moreover, there is a considerable influence (inversely correlated) of the topography with the streamflow signatures, with the increase in the altitude and slope lowering the runoff ratio, baseflow index, slope of the FDC, and mean streamflow, making this group of catchments to probably be an exporter of water (i.e., leaky catchments). This is in accordance with previous research on leaky catchments in Brazil (Schwamback et al., 2022), highlighting the importance of considering an open water balance on inferences about streamflow (Gordon et al., 2022). Finally, this group is one of the most impacted by human disturbance. The significant number of reservoirs (and their regulation capacity) modify the natural hydrological behavior of the catchments, artificially decreasing the runoff ratio, percentiles, and mean streamflow.

4. Broader Implications of Grouping Brazilian Catchments

Catchment classification based on the streamflow signatures can have several implications beyond generating an improved understanding of the hydrological processes within a catchment class. When analyzing a large set of catchments, we move from individual inferences to global (or regional) generalizations, identifying the most critical controls on water balance and limiting the variability within catchment classes (Addor et al., 2020; McDonnell & Woods, 2004).

Besides reproducing and predicting the hydrological behavior of a catchment system, the groups of similar catchments can be helpful for some applications in hydrology. For instance, identifying catchments (or areas) with similar hydrological behavior and attributes can support the transfer of information to ungauged basins (e.g., Marchezepe et al., 2023). Also, by using the random forest regressor algorithm to predicts runoff rates (Ghiggi et al., 2019), as well as long-term streamflow signatures in the present paper. Even though our study was not designed and aimed for streamflow predictions, we found suitable simulations for long-term streamflow signatures, such as mean daily streamflow, 5th, and 95th quantiles, baseflow index, and runoff ratio (see Table S2 in Supporting Information S1 for the *r*-squared values). Hence, it can also be useful for a prior understanding of the main drivers of water fluxes that may impact climatic extremes, such as droughts and floods.

The patterns and behaviors identified and grouped through the catchment classes allow us to predict better and understand the catchments' responses to climate change (Ballarin et al., 2023; K. A. Sawicz et al., 2014; Sivakumar et al., 2013). Using the catchment's responses to climate change, it is possible to capture the movement between groups and detect potential impacts and hydrological behavior trends early, ensuring water availability throughout the century (K. A. Sawicz et al., 2014). Since the overall behavior of six different groups is known, it would be easier for decision-makers to develop mitigating plans and define priorities (Viviroli et al., 2011). These mitigation plans should rely on investing in measures to build resilience to the impacts of climate change, such as improving infrastructure (Dong et al., 2019; Kolokytha & Malamataris, 2020; Sant'Anna et al., 2022), strengthening natural systems (Fossey & Rousseau, 2016; Hessen & Vandvik, 2022), and diversifying supply chains of food and water (Gomez et al., 2021; Zeff et al., 2014). Looking at the intra-annual variability, enhancement of forecast models may take advantage of the hydrological similarity of groups, adjusting and developing models specifically for each group rather than relying on generalized models that fail to capture the unique characteristics and peculiarities of hydrological behavior.

Despite the advantages of employing such a simple classification framework, there are also some limitations on the method that must be stated. Although we carefully selected the streamflow signatures to represent the hydrological behavior of the catchments, there is a possibility that classification systems are biased, leading to inaccuracies or unfair treatment of specific catchments. However, we demonstrated this was insignificant in our analysis. By using area-averaged attributes, there is also a possibility of oversimplifying the complexity of a catchment's hydrological system, leading to inaccurate predictions or management decisions. As all the methods employed here are data-driven, our results are strictly constrained by the choice of the streamflow signatures, as already discussed in K. Sawicz et al. (2011). Moreover, the constrained sample size within certain groups made applying cross-validation impractical for the random forest analysis. While not the primary focus of this study, it is important to recognize that this limitation could impact the obtained results. As a recommendation for further investigations, more signatures (not only based on streamflow) relying on the land-atmospheric processes should be added to the clustering processes (e.g., soil moisture, evapotranspiration, terrestrial water storage, among others), and categorical attributes may be considered (as long as not compromising the random forest performance). Although it is essential to consider these implications when developing and applying catchment classification systems based on hydrologic signatures, the lack of a generally accepted classification makes our proposed framework—with a clear goal, easily reproducible methodology, and clear metrics to weigh uncertainty and group stability—a significant step towards a standard classification system.

5. Summary and Conclusions

In this study, we propose a new approach for catchment classification, which involves selecting a wide-ranging set of hydrological signatures, covering various aspects of streamflow, including distribution, frequency, duration, and dynamics, allowing for a comprehensive understanding of hydrological behavior. Utilizing an optimized clustering method, we automatically determined the catchment groups, reducing subjectivity. Additionally, we investigated the hydrological signatures as catchment descriptors and the controls on the streamflow variability. Subsequently, we applied this methodology for 735 catchments from the CABra large-sample data set to classify Brazilian catchments for the first time. This enabled an unprecedented and comprehensive analysis of the hydrological behavior of Brazilian catchments, highlighting the primary drivers of streamflow variability and revealing insights into the influence of climate and landscape attributes on hydrological processes. Moreover, through the employment of an easily reproducible methodology and clear metrics for assessing uncertainty and group stability, our study represents a significant step toward establishing a standardized catchment-scale classification system.

Our catchment classification grouped the Brazilian catchments into six groups. Group A1 (perennial wet) is the wettest one, presenting high amounts of precipitation, large areas, and being primarily controlled by topography features. Group A2 (transitional wet) is mainly located in the south, is in an energy-limited condition, and does not present rainy/dry season. Group A3 (seasonal wet) is distributed primarily in mountainous areas of the Atlantic Forest, with high values of seasonal precipitation and streamflow, but at the edge of an energy/water-limited condition. Group A4 (seasonal dry) is mainly located in the Brazilian savannah and shows well-defined rainy and dry seasons—with a high seasonality of precipitation—and the streamflow is controlled primarily by aridity. Group A5 (transitional dry) is mainly located northeast, with high evapotranspiration through the year, sometimes larger than precipitation, leading to a consistent dry season. Finally, Group A6 (intermittent dry) is the extreme condition of Group A5, with high evapotranspiration values and low amounts of precipitation throughout the year, leading to an intermittent condition for its catchments.

In some way, even with the striking dissimilarities between A1 (perennial wet) and A6 (intermittent dry) catchments, we noted that there is a transitional behavior from the first to the last through the other catchment groups (in order: A1–A3, A3–A4, A4–A5, and A5–A6), driven by the aridity index. The sole exception is the A2 group, which presents a distinctive hydrological behavior, primarily attributable to their climatological condition of non-seasonality between water and energy availability. However, this should not be interpreted as a claim that similarities cannot be identified with other groups.

As done in this work, single employment of clustering catchments cannot ensure a general framework for a classification system because of the kind of data and subjectivity inserted. Moreover, in other catchments around the world, different processes and attributes may play a role. Despite this, the approach presented here attempted to reduce the subjectivity by employing statistical methods with minimal user inferences, such as the elbow method to define the number of clusters and the recursive feature elimination to feature selection and importance to the streamflow variability.

A changing climate and increasing climate variability expose large risks to Brazil's security since streamflow behavior at the local scale largely depends on climate. As hydrologic understanding is key to overcoming water resources challenges, our classification provides a step towards a better understanding of catchment hydrology in Brazil. More specifically, our results may be useful for understanding the hydrological processes involved through the catchment groups, making it possible to regionalize further hydrological information, generate new hypotheses, and better predict how catchment responses to climate change are expected to be.

Data Availability Statement

The data used in this study is entirely based on hydrometeorological and catchment attributes from the CABra data set, which is freely available online at https://doi.org/10.5281/zenodo.4070146 (Almagro et al., 2021).

Acknowledgments

This study was supported by grants from the Ministry of Science, Technology, and Innovation—MCTI and the National Council for Scientific and Technological Development—CNPq (Grants 309752/2020-5, 406647/2022-4, 309048/2023-0, and 446044/2023-7) and Fundect (Grant 71/032.795/2022). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001 and CAPES Print.

References

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2020). Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, *65*(5), 712–725. https://doi.org/10.1080/02626667.2019.1683182

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. *Advances in Water Resources*, *40*, 11–22. https://doi.org/10.1016/j.advwatres.2012.01.008

Almagro, A., Oliveira, P. T. S., Meira Neto, A. A., Roy, T., & Troch, P. (2021). CABra: A novel large-sample dataset for Brazilian catchments. *Hydrology and Earth System Sciences*, *25*(6), 3105–3135. https://doi.org/10.5194/hess-25-3105-2021

Anache, J. A. A., Wendland, E., Rosalem, L. M. P., Youlton, C., & Oliveira, P. T. S. (2019). Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. *Hydrology and Earth System Sciences*, *23*(3), 1263–1279. https://doi.org/10.5194/hess-23-1263-2019

Ballarin, A. S., Sousa Mota Uchôa, J. G., dos Santos, M. S., Almagro, A., Miranda, I. P., da Silva, P. G. C., et al. (2023). Brazilian water security threatened by climate change and human behavior. *Water Resources Research*, *59*(7), e2023WR034914. https://doi.org/10.1029/2023WR034914

Berghuijs, W. R., Larsen, J. R., van Emmerik, T. H. M., & Woods, R. A. (2017). A global assessment of runoff sensitivity to changes in precipitation, potential evaporation, and other factors. *Water Resources Research*, *53*(10), 8475–8486. https://doi.org/10.1002/2017WR021593

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. https://doi.org/10.1002/2014WR015692

Beven, K. J. (2000). Uniqueness of place and process representations in hydrological modelling. *Hydrology and Earth System Sciences*, *4*(2), 203–213. https://doi.org/10.5194/hess-4-203-2000

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Breiman, L. (2001). Random forests. *Machine Learning*, *1*(45), 5–32. https://doi.org/10.1023/A:1010933404324

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M., & Fritz, S. (2019). Copernicus global land service: Land cover 100m: Epoch 2015: Globe. https://doi.org/10.5281/ZENODO.3243509

Budyko, M. I. (1948). *Evaporation under natural conditions*. Israel Program for Scientific Translations.

Budyko, M. I. (1974). *Climate and life*. Elsevier.

Chaney, N. W., Torres-Rojas, L., Vergopolan, N., & Fisher, C. K. (2021). HydroBlocks v0.2: Enabling a field-scale two-way coupling between the land surface and river networks in Earth system models. *Geoscientific Model Development*, *14*(11), 6813–6832. https://doi.org/10.5194/gmd-14-6813-2021

Clausen, B., & Biggs, B. J. F. (2000). Flow variables for ecological studies in temperate streams: Groupings based on covariance. *Journal of Hydrology*, *237*(3–4), 184–197. https://doi.org/10.1016/S0022-1694(00)00306-1

Court, A. (1962). Measures of streamflow timing. *Journal of Geophysical Research*, *67*(11), 4335–4339. https://doi.org/10.1029/JZ067i011p04335

De la Fuente, L. A., Gupta, H. V., & Condon, L. E. (2023). Toward a multi-representational approach to prediction and understanding, in support of discovery in hydrology. *Water Resources Research*, *59*(1), e2021WR031548. https://doi.org/10.1029/2021WR031548

Dong, N., Yu, Z., Gu, H., Yang, C., Yang, M., Wei, J., et al. (2019). Climate-induced hydrological impact mitigated by a high-density reservoir network in the Poyang Lake Basin. *Journal of Hydrology*, *579*, 124148. https://doi.org/10.1016/j.jhydrol.2019.124148

Dooge, J. C. I. (1986). Looking for hydrologic laws. *Water Resources Research*, *22*(9S), 46S–58S. https://doi.org/10.1029/WR022i09Sp0046S

Dooge, J. C. I. (1992). Sensitivity of runoff to climate change: A Hortonian approach. *Bulletin of the American Meteorological Society*, *73*(12), 2013–2024. https://doi.org/10.1175/1520-0477(1992)073<2013:sortcc>2.0.co;2

Dunne, T. (1978). Field studies of hillslope flow processes. In M. J. Kirkby (Ed.), *Hillslope hydrology* (pp. 227–293). John Wiley & Sons, Inc. https://doi.org/10.18172/cig.1099

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*, *58*(10), e2022WR032404. https://doi.org/10.1029/2022WR032404

Fossey, M., & Rousseau, A. N. (2016). Can isolated and riparian wetlands mitigate the impact of climate change on watershed hydrology? A case study approach. *Journal of Environmental Management*, *184*, 327–339. https://doi.org/10.1016/j.jenvman.2016.09.043

García-Escudero, L. A., Gordaliza, A., Matrán, C., & Mayo-Iscar, A. (2010). A review of robust clustering methods. *Advances in Data Analysis and Classification*, *4*(2), 89–109. https://doi.org/10.1007/s11634-010-0064-5

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). GRUN: An observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, *11*(4), 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath Earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, *41*(11), 3891–3898. https://doi.org/10.1002/2014GL059856

Gomez, M., Mejia, A., Ruddell, B. L., & Rushforth, R. R. (2021). Supply chain diversity buffers cities against food shocks. *Nature*, *595*(7866), 250–254. https://doi.org/10.1038/s41586-021-03621-0

Gordon, B. L., Crow, W. T., Konings, A. G., Dralle, D. N., & Harpold, A. A. (2022). Can we use the water budget to infer upland catchment behavior? The role of data set error estimation and interbasin groundwater flow. *Water Resources Research*, *58*(9), e2021WR030966. https://doi.org/10.1029/2021WR030966

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Haag, I., & Westrich, B. (2002). Processes governing river water quality identified by principal component analysis. *Hydrological Processes*, *16*(16), 3113–3130. https://doi.org/10.1002/hyp.1091

Harper, A., Baker, I. T., Denning, A. S., Randall, D. A., Dazlich, D., & Branson, M. (2014). Impact of evapotranspiration on dry season climate in the Amazon forest. *Journal of Climate*, *27*(2), 574–591. https://doi.org/10.1175/JCLI-D-13-00074.1

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hessen, D. O., & Vandvik, V. (2022). Buffering climate change with nature. *Weather, Climate, and Society*, *14*(2), 439–450. https://doi.org/10.1175/WCAS-D-21-0059.1

Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., & Houska, T. (2020). Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrology and Earth System Sciences*, *24*(3), 1081–1100. https://doi.org/10.5194/hess-24-1081-2020

Jha, P., Lakaria, B. L., Meena, B. P., Biswas, A. K., & Patra, A. K. (2022). Plans and policies for soil carbon storage. In *Plans and policies for soil organic carbon management in agriculture* (pp. 123–140). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-6179-3_5

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resources Research*, *54*(7), 5088–5109. https://doi.org/10.1029/2018WR022913

Kolokytha, E., & Malamataris, D. (2020). Integrated water management approach for adaptation to climate change in highly water stressed basins. *Water Resources Management*, *34*(3), 1173–1197. https://doi.org/10.1007/s11269-020-02492-w

L'vovich, M. I. L. (1979). *World water resources and their future*. AGU.

Lyne, V., & Hollick, M. (1979). Stochastic time-variable rainfall-runoff modeling. In *Hydrology and water resources symposium* (pp. 89–92). Institution of Engineers National Conference Publication.

Lyon, S. W., & Troch, P. A. (2010). Development and application of a catchment similarity index for subsurface flow. *Water Resources Research*, *46*(3), 1–13. https://doi.org/10.1029/2009WR008500

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

Marchezepe, B. K., Almagro, A., Ballarin, A. S., & Oliveira, P. T. S. (2023). Streamflow prediction in ungauged catchments by using the Grunsky method. *International Soil and Water Conservation Research*, *11*(4), 700–712. https://doi.org/10.1016/j.iswcr.2023.06.004

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Marutho, D., Hendra Handaka, S., & Wijaya, E., & Muljono. (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *Proceedings of 2018 international seminar on application for technology of information and communication: Creative technology for human life, ISemantic 2018* (pp. 533–538). https://doi.org/10.1109/ISEMANTIC.2018.8549751

McDonnell, J. J., & Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology*, *299*(1–2), 2–3. https://doi.org/10.1016/j.jhydrol.2004.09.003

McMillan, H. (2020). Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, *34*(6), 1393–1409. https://doi.org/10.1002/hyp.13632

McMillan, H., Gueguen, M., Grimon, E., Woods, R., Clark, M., & Rupp, D. E. (2014). Spatial variability of hydrological processes and model structure diagnostics in a 50 km^2 catchment. *Hydrological Processes*, *28*(18), 4896–4913. https://doi.org/10.1002/hyp.9988

McMillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761. https://doi.org/10.1002/hyp.11300

McMillan, H. K. (2021). A review of hydrologic signatures and their applications. *Wiley Interdisciplinary Reviews: Water*, *8*(1), 1–23. https://doi.org/10.1002/wat2.1499

Meira Neto, A. A., Roy, T., Oliveira, P. T. S., & Troch, P. A. (2020). An aridity index-based formulation of streamflow components. *Water Resources Research*, *56*(9), 1–14. https://doi.org/10.1029/2020WR027123

Milligan, G. W., & Cooper, M. C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement*, *11*(4), 329–354. https://doi.org/10.1177/014662168701100401

Minasny, B., & McBratney, A. B. (2018). Limited effect of organic matter on soil available water capacity. *European Journal of Soil Science*, *69*(1), 39–47. https://doi.org/10.1111/ejss.12475

Nainggolan, R., Perangin-Angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the elbow method. *Journal of Physics: Conference Series*, *1361*(1), 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., et al. (2011). Height above the Nearest Drainage—A hydrologically relevant new terrain model. *Journal of Hydrology*, *404*(1–2), 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. https://doi.org/10.1002/rra.700

Oliveira, P. T. S., Leite, M. B., Mattos, T., Nearing, M. A., Scott, R. L., de Oliveira Xavier, R., et al. (2017). Groundwater recharge decrease with increased vegetation density in the Brazilian cerrado. *Ecohydrology*, *10*(1), e1759. https://doi.org/10.1002/ECO.1759

Poff, N. L., & Ward, J. V. (1989). Implications of streamflow variability and predictability for lotic community structure: A regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, *46*, 1805–1818. https://doi.org/10.1139/f89-228

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Rao, A. R., & Srinivas, V. V. (2006). Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology*, *318*(1–4), 57–79. https://doi.org/10.1016/j.jhydrol.2005.06.004

Richter, B. D., Baumgartner, J. V., Powell, J., & Braun, D. P. (1996). A method for assessing hydrologic alteration within ecosystems. *Conservation Biology*, *10*(4), 1163–1174. https://doi.org/10.1046/j.1523-1739.1996.10041163.x

Rodrigues, A. F., de Mello, C. R., Nehren, U., de Ribeiro, J. P. C., Mantovani, V. A., & de Mello, J. M. (2021). Modeling canopy interception under drought conditions: The relevance of evaporation and extra sources of energy. *Journal of Environmental Management*, *292*, 112710. https://doi.org/10.1016/j.jenvman.2021.112710

Rodrigues, D. B. B., Gupta, H. V., Mediondo, E. M., & Oliveira, P. T. S. (2015). Assessing uncertainties in surface water security: An empirical multimodel approach. *Water Resources Research*, *51*(11), 9013–9028. https://doi.org/10.1002/2014WR016691

Sankarasubramanian, A., Vogel, R. M., & Limbrunner, J. F. (2001). Climate elasticity of streamflow in the United States. *Water Resources Research*, *37*(6), 1771–1781. https://doi.org/10.1029/2000WR900330

Sant'Anna, C., Tilmant, A., & Pulido-Velazquez, M. (2022). A hydrologically-driven approach to climate change adaptation for multipurpose multireservoir systems. *Climate Risk Management*, *36*, 100427. https://doi.org/10.1016/j.crm.2022.100427

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011). Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, *15*(9), 2895–2911. https://doi.org/10.5194/hess-15-2895-2011

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing hydrologic change through catchment classification. *Hydrology and Earth System Sciences*, *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014

Schaake, J. C. (1990). From climate to flow. In P. E. Waggoner (Ed.), *Climate change and US water resources* (pp. 177–206). John Wiley & Sons, Ltd.

Schwamback, D., Gesualdo, G. C., Sone, J. S., Kobayashi, A. N. A., Bertotto, L. E., Garcia, M. V. S., et al. (2022). Are Brazilian catchments gaining or losing water? The effective area of tropical catchments. *Hydrological Processes*, *36*(3), 1–13. https://doi.org/10.1002/hyp.14535

Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *3rd international symposium on intelligent information technology and security informatics, IITSI 2010* (pp. 63–67). https://doi.org/10.1109/IITSI.2010.74

Singh, P. K., Kumar, V., Purohit, R. C., Kothari, M., & Dashora, P. K. (2009). Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. *Water Resources Management*, *23*(2), 325–339. https://doi.org/10.1007/s11269-008-9277-1

Sivakumar, B. (2008). Dominant processes concept, model simplification and classification framework in catchment hydrology. *Stochastic Environmental Research and Risk Assessment*, *22*(6), 737–748. https://doi.org/10.1007/s00477-007-0183-5

Sivakumar, B., Singh, V. P., Berndtsson, R., & Khan, S. K. (2013). Catchment classification framework in hydrology: Challenges and directions. *Journal of Hydrologic Engineering*, *20*(1), A4014002. https://doi.org/10.1061/(asce)he.1943-5584.0000837

Sivapalan, M. (2005). Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale. In M. Anderson (Ed.), *Encyclopedia of hydrological sciences* (pp. 193–219). John Wiley. https://doi.org/10.1002/0470848944.hsa012

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: Materials science and engineering* (Vol. 336, p. 012017). https://doi.org/10.1088/1757-899X/336/1/012017

Tarasova, L., Basso, S., Poncelet, C., & Merz, R. (2018). Exploring controls on rainfall-runoff events: 2. Regional patterns and spatial controls of event characteristics in Germany. *Water Resources Research*, *54*(10), 7688–7710. https://doi.org/10.1029/2018WR022588

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., & Sawicz, K. (2013). Climate-vegetation-soil interactions and long-term hydrologic partitioning: Signatures of catchment co-evolution. *Hydrology and Earth System Sciences*, *17*(6), 2209–2217. https://doi.org/10.5194/hess-17-2209-2013

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, *11*(5), 910. https://doi.org/10.3390/w11050910

Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., et al. (2022). High-resolution soil moisture data reveal complex multi-scale spatial variability across the United States. *Geophysical Research Letters*, *49*(15), 1–13. https://doi.org/10.1029/2022GL098586

Vergopolan, N., Xiong, S., Estes, L., Wanders, N., Chaney, N. W., Wood, E. F., et al. (2021). Field-scale soil moisture bridges the spatial-scale gap between drought monitoring and agricultural yields. *Hydrology and Earth System Sciences*, *25*(4), 1827–1847. https://doi.org/10.5194/hess-25-1827-2021

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., et al. (2011). Climate change and mountain water resources: Overview and recommendations for research, management and policy. *Hydrology and Earth System Sciences*, *15*(2), 471–504. https://doi.org/10.5194/hess-15-471-2011

Wagener, T., Sivapalan, M., & McGlynn, B. (2008). Catchment classification and services-toward a new paradigm for catchment hydrology driven by societal needs. In *Encyclopedia of hydrological sciences* (pp. 1–12). https://doi.org/10.1002/0470848944.hsa320

Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

Werth, D., & Avissar, R. (2004). The regional evapotranspiration of the Amazon. *Journal of Hydrometeorology*, *5*(1), 100–109. https://doi.org/10.1175/1525-7541(2004)005<0100:TREOTA>2.0.CO;2

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Woods, R. A. (2009). Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources*, *32*(10), 1465–1481. https://doi.org/10.1016/j.advwatres.2009.06.011

Wu, S., Zhao, J., Wang, H., & Sivapalan, M. (2021). Regional patterns and physical controls of streamflow generation across the conterminous United States. *Water Resources Research*, *57*(6), 1–31. https://doi.org/10.1029/2020WR028086

Xavier, A. C., King, C. W., & Scanlon, B. R. (2016). Daily gridded meteorological variables in Brazil (1980-2013). *International Journal of Climatology*, *36*(6), 2644–2659. https://doi.org/10.1002/joc.4518

Yadav, M., Wagener, T., & Gupta, H. V. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

Yost, J. L., & Hartemink, A. E. (2019). Effects of carbon on moisture storage in soils of the Wisconsin Central Sands, USA. *European Journal of Soil Science*, *70*(3), 565–577. https://doi.org/10.1111/ejss.12776

Zeff, H. B., Kasprzyk, J. R., Herman, J. D., Reed, P. M., & Characklis, G. W. (2014). Navigating financial and supply reliability tradeoffs in regional drought management portfolios. *Water Resources Research*, *50*(6), 4906–4923. https://doi.org/10.1002/2013WR015126

References From the Supporting Information

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, *339*(6122), 940–943. https://doi.org/10.1126/science.1229881

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12), 1–37. https://doi.org/10.1029/2012GC004370

Melo, D. C. D., Anache, J. A. A., Borges, V. P., Miralles, D. G., Martens, B., Fisher, J. B., et al. (2021). Are remote sensing evapotranspiration models reliable across South American ecoregions? *Water Resources Research*, *57*(11), 1–23. https://doi.org/10.1029/2020WR028752

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., et al. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874