Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR030619

Key Points:

- Baseflow and aridity help classify suitable model structures for different catchments in Brazil
- A single bucket model performs relatively well in wet catchments with low baseflow, mainly in southern Brazil
- Independently of model structure, performance is generally lower in drier than in wetter catchments

Supporting Information:

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

Correspondence to:

P. C. David,
paulacunhadavid@gmail.com

Citation:

David, P. C., Chaffe, P. L. B., Chagas, V. B. P., Dal Molin, M., Oliveira, D. Y., Klein, A. H. F., & Fenicia, F. (2022). Correspondence between model structures and hydrological signatures: A large-sample case study using 508 Brazilian catchments. *Water Resources Research*, *58*, e2021WR030619. https://doi.org/10.1029/2021WR030619

Received 18 JUN 2021
Accepted 5 FEB 2022

Author Contributions:

Conceptualization: Paula C. David, Pedro L. B. Chaffe, Vinicius B. P. Chagas, Marco Dal Molin, Debora Y. Oliveira, Antonio H. F. Klein, Fabrizio Fenicia
Data curation: Paula C. David, Vinicius B. P. Chagas
Formal analysis: Paula C. David, Marco Dal Molin, Fabrizio Fenicia
Funding acquisition: Antonio H. F. Klein
Investigation: Paula C. David, Marco Dal Molin, Fabrizio Fenicia
Methodology: Paula C. David, Pedro L. B. Chaffe, Vinicius B. P. Chagas, Marco Dal Molin, Debora Y. Oliveira, Antonio H. F. Klein, Fabrizio Fenicia
Resources: Pedro L. B. Chaffe, Antonio H. F. Klein

Correspondence Between Model Structures and Hydrological Signatures: A Large-Sample Case Study Using 508 Brazilian Catchments

Paula C. David[1], Pedro L. B. Chaffe[2], Vinicius B. P. Chagas[1], Marco Dal Molin[3], Debora Y. Oliveira[1,4], Antonio H. F. Klein[5], and Fabrizio Fenicia[3]

[1]Graduate Program in Environmental Engineering, Federal University of Santa Catarina, Florianópolis, Brazil, [2]Department of Sanitary and Environmental Engineering, Federal University of Santa Catarina, Florianópolis, Brazil, [3]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, [4]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA, [5]Coastal Oceanography Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil

Abstract Daily streamflow dynamics can be accurately simulated by conceptual models as simple as a single bucket in some catchments, while they require more complex configurations in other catchments. However, without resorting to calibration, anticipating where and why a given model structure may be appropriate remains difficult. In this work, we explored the feasibility of relating suitable model structures to the climate and streamflow characteristics of 508 catchments in Brazil. Specifically, we tested four model structures using up to three reservoirs, where each reservoir is intended to represent a catchment function: the rainfall-runoff threshold, the fast, and the slow hydrograph response. We hypothesized a relationship between suitable model structures and hydrological signatures of aridity (I_A) and baseflow index (I_B). Our results show that different classes of signatures resulted in distinct patterns of model performance. Wet catchments ($I_A < 0.9$) with low baseflow ($I_B < 0.4$) were the easiest to model, with a single-reservoir model presenting a relatively good performance. In the case of low baseflow, adding a rainfall-runoff threshold reservoir resulted in better performance than adding a slow response reservoir, whereas in the case of high baseflow ($I_B < 0.6$) the opposite occurred. In the case of low baseflow, the inclusion of a slow response reservoir helped the simulation of dry catchments ($I_A < 1.1$), but not of wet ones, which we attributed to the impact of permeability in dry catchments. These results indicate a path toward model structure identification from streamflow signatures and potentially from landscape features.

1. Introduction

Conceptual hydrological models are useful tools to study and simulate catchment-scale processes and behavior. Determining a suitable conceptual model structure for a given catchment, however, remains difficult given the variability of processes, data, and intended objectives that characterize catchment model applications (Beven, 2000). Large model intercomparison studies showed that no single model is able to outperform the others in a broad range of catchments, but rather that relative model performance is catchment specific (Duan et al., 2006; Knoben et al., 2020). Even when model applications are limited to streamflow simulations at daily or subdaily time resolution, the range of appropriate model structures is diverse. Following the principle of parsimony, as exemplified in the "top-down" approach (Sivapalan et al., 2003), models as simple as a single reservoir were found to be appropriate in some catchments, whereas more complex models were necessary in others (Bai et al., 2009; Kavetski & Fenicia, 2011; Kirchner, 2009).

In recognition of model specificity, alternative development strategies have been pursued for modeling streamflow time series. One strategy has been to develop models that represent the best compromise over a wide range of applications. The pioneering study by Jakeman and Hornberger (1993) explored model identification in seven catchments covering a wide range of scales and climatic conditions and determined that a three-element model structure with two reservoirs in parallel driven by a rainfall excess element was the most commonly identified configuration. A more systematic approach to model identification is represented by the data-based mechanistic approach, whose application to multiple catchments confirmed such three-element structures as the most commonly identified model configuration (Young, 2003). Conceptual models such as the GR4J (Perrin et al., 2003) and HBV (Lindstrom et al., 1997) can be regarded as emblematic examples of the search for a general

Software: Paula C. David, Pedro L. B. Chaffe, Marco Dal Molin, Debora Y. Oliveira
Supervision: Pedro L. B. Chaffe, Fabrizio Fenicia
Validation: Paula C. David, Pedro L. B. Chaffe
Visualization: Paula C. David, Vinicius B. P. Chagas, Marco Dal Molin, Fabrizio Fenicia
Writing – original draft: Paula C. David, Pedro L. B. Chaffe, Vinicius B. P. Chagas, Marco Dal Molin, Debora Y. Oliveira, Antonio H. F. Klein, Fabrizio Fenicia
Writing – review & editing: Paula C. David, Pedro L. B. Chaffe, Vinicius B. P. Chagas, Marco Dal Molin, Debora Y. Oliveira, Antonio H. F. Klein, Fabrizio Fenicia

purpose conceptual model as they have been subject to a process of continuous refinement over time. Interestingly, these models are also generally consistent with such three-element model structures.

"Best compromise" models have important practical advantages since they can, for example, facilitate operational use (Le Moine et al., 2007). However, even if such models have better performance than others on average, they may not be the best choice in each individual catchment. For example, such fixed models (e.g., HBV) may have advanced to reflect a certain type of hydrology (e.g., Sweden), but not global variability in hydrologic systems. Therefore, an alternative strategy has been to develop approaches that enable model identification in each specific catchment. The most basic approach to pursue this objective has been to develop a set of a priori defined model structures, which are subsequently evaluated on a set of catchments (e.g., Bai et al., 2009; Coxon et al., 2014; Jakeman & Hornberger, 1993; van Esse et al., 2013). More advanced approaches include model selection in an optimization setup, thus enabling direct model structure identification (e.g., Seiller et al., 2017; Spieler et al., 2020; Young, 2003). Such model identification approaches have generally confirmed that the streamflow response of different catchments may be best described by models that differ in terms of structure and parameters.

If streamflow dynamics in different catchments may be best represented by distinct models, the question is whether it is possible to relate suitable model structures to a given catchment directly based on data, therefore without resorting to calibration. Such an ability would contribute to the representation of intercatchment differences and advance catchment classification, hence furthering capacity to sort and group the variability of catchment systems (McDonnell & Woods, 2004; Wagener et al., 2007). In particular, a major objective of catchment classification is to predict the catchment "functions" of "partition" of incoming flows, such as those relating to precipitation in different flow paths, "storage" of water in different catchment compartments, and "release" as evaporation or outflow (Wagener et al., 2007). The representation of these functions through model components designed for that purpose is a common objective of process-based hydrological models. Hence, building a mapping between catchment data and suitable models is a way to characterize hydrological diversity. This process does not necessarily imply a unique mapping between catchments and models, but rather that certain catchment classes may be associated with distinct model classes (Coxon et al., 2014).

If hydrological models represent the landscape properties that control hydrological processes (Freeze, 1974), one can expect that suitable model structures can be directly associated with such properties. Although this connection appeared feasible for small headwater catchments (Fenicia et al., 2014; Kavetski & Fenicia, 2011) or distributed models (Fenicia et al., 2016; Samaniego et al., 2010; Singh et al., 2012), it has been difficult to establish for lumped conceptual models and large catchments. For example, while investigating the differences in performance of 36 lumped conceptual models in 559 catchments across the United States, Knoben et al. (2020) found "no clear relationships between the catchments where any model performs well and descriptors of those catchments' geology, topography, soil, and vegetation characteristics." Similarly, in evaluating the performance of 12 model structures in 99 catchments in Germany, Ley et al. (2016) did not find a strong connection between relative model performance and catchment characteristics. Massmann (2020), on a model comparison study in 574 US catchments, found it difficult to link relative model performance to landscape attributes (topography or catchment area). These results indicate that relationships between catchment characteristics and model structures or parameters are easier to establish on a smaller rather than on a larger scale (Singh et al., 2012).

Alternatively, by taking into account that hydrological models are typically intended to transform climate input into streamflow output, one may seek a connection between suitable model structure and the characteristics or signatures of such input–output variables. The possible relationship between suitable model structure and hydrograph characteristics was already noted by Jakeman and Hornberger (1993). While searching for a best compromise model structure, Jakeman and Hornberger (1993) observed that the optimal model was catchment dependent and that, for example, a slow reservoir was not needed if baseflow was negligible. A correspondence between model complexity and baseflow index was found in other studies, suggesting more complex models for catchments with higher baseflow (Coxon et al., 2014; Massmann, 2020). In determining the minimum level of model complexity required to predict runoff in New Zealand catchments, Atkinson et al. (2002) established a hypothetical relationship between model complexity, timescale, and climate characteristics, suggesting that more complex models are needed for finer time resolutions and drier catchments. The fact that wetter catchments are easier to model than drier catchments is generally consistent with model intercomparison studies (e.g., Coxon et al., 2014; Massmann, 2020; Parajka et al., 2013; Poncelet et al., 2017). To a certain extent, the ease with

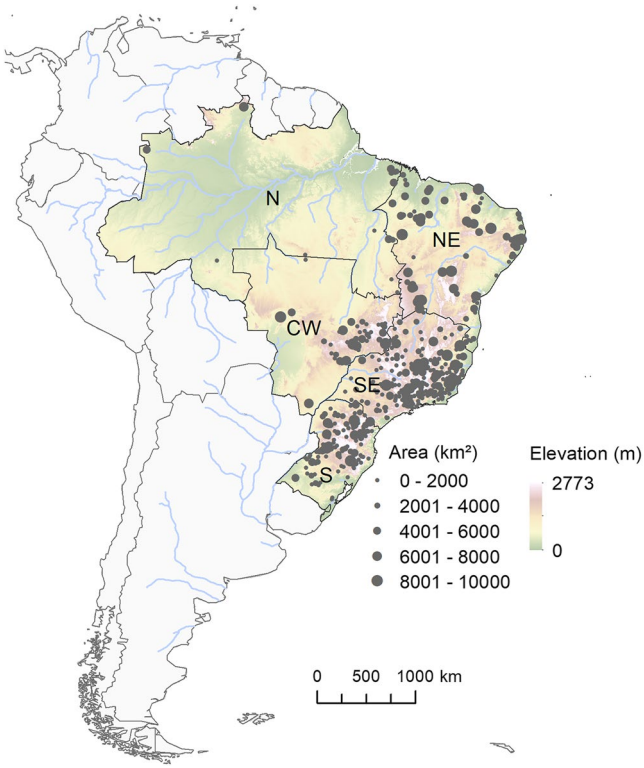

Figure 1. Brazil and the 508 stream gauges included in this study. Gray lines indicate countries. Black lines indicate the regions of Brazil: north (N), northeast (NE), central-west (CW), southeast (SE), and south (S). Blue lines indicate the river drainage.

which wetter and therefore more dynamic catchments can be modeled also reflects the properties of widely used objective functions like Nash–Sutcliffe efficiency (Knoben et al., 2019).

Although such previous studies provided useful indications on how signatures can potentially affect identifiable model complexity, they did not establish predictive relationships with regard to such interrelationships. In this study, we framed a set of hypotheses on the relationship between signatures and identifiable model complexity, which are verified using a large-sample data set. Specifically, we investigated whether signatures of baseflow and aridity help identify suitable model structures in different catchments, leveraging the recently developed CAMELS-BR data set (Chagas et al., 2020), containing a large sample of catchments in Brazil. Our exploration of model alternatives included four model structures which complement the three-element model identified by Jakeman and Hornberger (1993) with simpler model alternatives using one or two elements. Our hypothesis was that the relative performance of such models can be related to catchment signatures and therefore help identify different dominant processes in different catchments.

Our specific objectives are to do the following:

- Assess when different hydrological processes should be added to the model structure.
- Understand the correspondence between model structures and hydrological signatures.

This paper is structured as follows. Section 2 describes the study area and data. Section 3 describes the methodology, including model structure and hydrological signatures selection, model evaluation, and the key hypothesis. Section 4 presents the main results, including catchment classification and analysis of streamflow simulations for each model structure and catchment class. Section 5 discusses the results according to the key hypothesis presented in Section 3.4. Section 6 summarizes the main conclusions of the study.

2. Study Area and Data

We used 508 catchments from the CAMELS-BR data set (Figure 1), which contains hydrometeorological time series and catchment attributes for 897 catchments in Brazil (Chagas et al., 2020). We excluded catchments larger than 10,000 km^2 to avoid strong climate variability within a single catchment. In addition, we considered only catchments in which less than 20% of the area is covered by urban use to limit the effects of anthropogenic changes.

The hydrometeorological time series includes 20 years of observations ranging from 1 January 1985 to 31 December 2004 at daily resolution. The observed variables include streamflow, precipitation, and potential evaporation (averaged for each catchment). Precipitation gridded data are obtained from Xavier et al. (2016) and potential evaporation is obtained from the Global Land Evaporation Amsterdam Model (GLEAM), as described in Chagas et al. (2020).

The study area encompasses a wide range of climate, hydrology, and landscape properties. It can be divided into three major regions: the semiarid, the humid tropics, and the humid subtropics. None of such regions have dominant snow-related processes (Chagas et al., 2020). The semiarid region includes most of the northeastern area of Brazil. Mean annual precipitation is 750 mm and mean annual potential evaporation is 1,200 mm. Streams in the northern portion of the semiarid region are ephemeral and highly responsive to rainfall events. They have low subsurface porosity, permeability, and baseflow indices (on average 0.35), which suggests a predominance of quick flow pathways. On the other hand, streams in the southwestern semiarid region are perennial and have low

responsiveness to rainfall events. Therein, they have high subsurface porosity, permeability, and baseflow indices (on average 0.74), suggesting a predominance of slow flow pathways.

The humid tropics and subtropics have contrasting hydrological regimes. The humid tropical catchments are located mostly in central-western and southeastern Brazil. Mean annual precipitation and potential evaporation are 1,500 and 1,025 mm, respectively. Precipitation is highly seasonal due to the South American Monsoon System (Marengo et al., 2012), which leads to wet summers and dry winters. Streamflow responds slowly to rainfall events as compared to the humid subtropics and is sustained by high baseflow rates during the dry winters. The soils are highly permeable and commonly more than 25 m deep (Figure S1). Such characteristics indicate the potential for a high water storage capacity, which gets replenished during the wet season and released slowly in the dry season. In the humid subtropics, by contrast, precipitation and streamflow regimes are uniform throughout the year. The subtropical catchments are located mostly in southern Brazil, with mean annual precipitation and potential evaporation of 1,700 and 890 mm, respectively. Streamflow is highly responsive to rainfall, with considerable fluctuations between low and high flow events. Compared with the humid tropics, baseflow rates are low, soils are not as deep (less than 10 m deep) and have high clay content (Figure S1), indicating lower water storage capacity.

3. Methodology

In the process of investigating a mapping between hydrological signatures and conceptual model structures, the following key steps were envisaged.

1. Selection of conceptual model structures.
2. Selection of hydrological signatures.
3. Model evaluation.
4. Development of key hypotheses.

These steps are detailed in the following sections.

3.1. Selection of Conceptual Model Structures

In this study, we aimed to select a minimal set of model structures, covering the range of complexities and corresponding processes that are typically identifiable using daily streamflow data in nonsnow-dominated catchments. As discussed in Section 1, a "best compromise" model structure that adequately models streamflow response in many catchments is constituted by three key elements, represented by a reservoir that partitions precipitation between evaporation and effective rainfall, followed by two reservoirs operating in parallel, representing the fast and the slow hydrograph response. These three elements are often associated with distinct catchment compartments and corresponding processes. The first reservoir in the cascade of processes is typically associated with the unsaturated zone and characterizes the nonlinearity in the partitioning of precipitation between evaporation and effective rainfall, depending on the reservoir storage. The two parallel reservoirs are associated with the (sub)surface and groundwater compartments and are represented as linear (Jakeman & Hornberger, 1993; Young, 2003) or nonlinear (e.g., in HBV) storage–discharge processes. Many widely used conceptual models such as GR4J or HBV can be assimilated to such a structure, which also turns out to be one of the most commonly identified using various model identification strategies in several catchments (e.g., Jakeman & Hornberger, 1993; Young, 2003).

We here hypothesized that this model structure is rather inclusive in terms of processes and that simpler structures can be found suitable, depending on the characteristics of catchment forcings and responses. Thus, we complemented this model using simpler versions of it, using either one or two reservoirs. In principle, the set of models could have been more inclusive, for example, by introducing additional elements such as lag functions or reservoirs (e.g., Fenicia et al., 2014). However, we assumed that these model structures were representative of the processes that are most easily identified using streamflow data in a simple calibration setup. Moreover, large-sample studies are unavoidably confronted with the need to balance "depth" with "breadth," which results in some compromises in the number of feasible model experiments (Gupta et al., 2014).

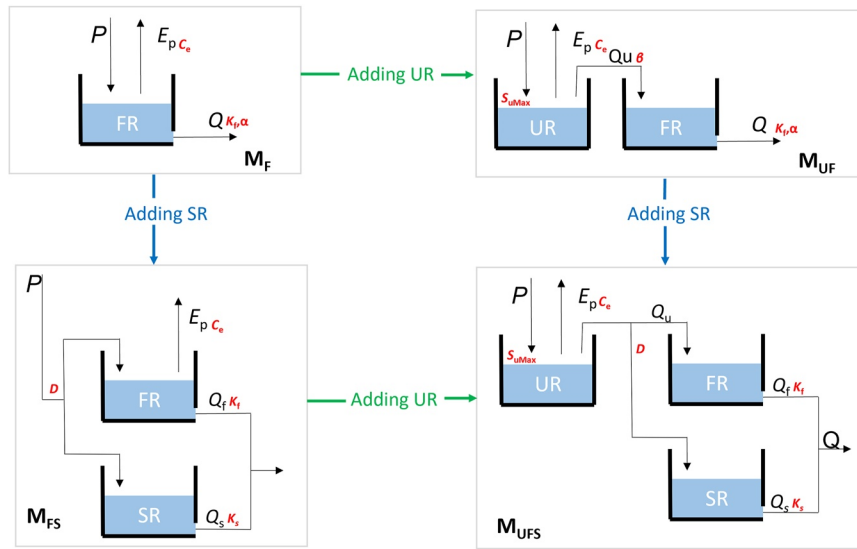

Figure 2. Model structures used. FR indicates fast reservoir, UR indicates unsaturated zone reservoir, and SR indicates slow reservoir. Green pointers indicate the comparisons when adding the UR, blue pointers indicate the comparisons when adding the SR. The parameters are displayed in red. *Ce* is the evaporation parameter, S_{uMax} is the unsaturated reservoir capacity, β is the unsaturated reservoir exponent, D is the proportion of flow directed to the slow reservoir, α is the fast reservoir exponent, K_f is the fast reservoir coefficient, and K_s is the slow reservoir coefficient.

In total, we considered four model structures, which are schematically represented in Figure 2. For ease of referencing, the four model structures are named M_F (fast reservoir), M_{UF} (unsaturated and fast reservoir), M_{FS} (fast and slow reservoir), and M_{UFS} (unsaturated, fast, and slow reservoir), as shown in Figure 2. The most complex model, M_{UFS}, exemplifies the three-element model recalled above. Precipitation enters the "unsaturated reservoir" (UR), which determines the "effective rainfall," hence the portion of precipitation that eventually becomes streamflow. Effective rainfall is then partitioned between a quick flow and a slow flow component, determined by the "fast reservoir" (FR) and the "slow reservoir" (SR), respectively. The reservoirs UR, FR, and SR are consistent with the "rainfall excess," "quick flow," and "slow flow" elements of the most common configuration identified by Jakeman and Hornberger (1993), and with the "soil water storage," "quick flow pathway," and "slow flow pathway" of Young (2003), as well as with the structure of conceptual models such as GR4J and HBV. The simplest model, M_F (upper left panel of Figure 2), is represented by a single power law reservoir, simulating both catchment streamflow and evaporation. This model is intended to represent catchments as "simple dynamical systems," hence where streamflow is determined by their total water storage (Kirchner, 2009). A single-reservoir model can be regarded as a lower bound in terms of typical complexity in conceptual modeling. But experience has shown that even such a simple model can provide accurate streamflow simulations in some catchments (Kavetski & Fenicia, 2011; Kirchner, 2009).

The models M_{UF} and M_{FS} are two-reservoir models intended to represent two different pathways from M_F to M_{UFS}. Compared to M_{UFS}, M_{UF} contains the reservoirs UR and FR but lacks SR. Hence, effective precipitation from UR is routed through a single reservoir. Whereas in M_{UFS}, the reservoir FR is linear, in M_{UF}, FR is a power law reservoir in order to provide this reservoir with some flexibility to reproduce both fast and slow hydrograph dynamics. M_{FS}, compared to M_{UFS}, contains the reservoirs FR and SR but lacks UR. Hence, M_{FS} has a fast and a slow release component but lacks a component designed to estimate effective rainfall. Similar to M_F, evaporation is taken directly from FR.

The model structures were generated using a MATLAB implementation of the SUPERFLEX framework (Fenicia et al., 2011; Kavetski & Fenicia, 2011). The water balance equations and the constitutive relationships are presented in Tables 1 and 2, respectively. The models were implemented with a second-order accurate explicit method with adaptive time stepping (Schoups et al., 2010), with absolute and relative tolerances fixed at 10^{-3}, as in David et al. (2019).

Table 1
Water Balance Equations of the Models Used in the Experiments (✓ and "–" Indicate Presence or Absence, Respectively)

Water balance equations	M_F	M_{UF}	M_{FS}	M_{UFS}
$\frac{dS_f}{dt} = P_f - Q_f - E$	✓	–	✓	–
$\frac{dS_f}{dt} = P_f - Q_f$	–	✓	–	✓
$\frac{dS_u}{dt} = P - Q_u - E$	–	✓	–	✓
$\frac{dS_s}{dt} = P_s - Q_s$	–	–	✓	✓
$P = P_f$	✓	–	–	–
$Q_u = P_f$	–	✓	–	–
$Q_f = Q$	✓	✓	–	–
$P = P_f + P_s$	–	–	✓	–
$Q_f + Q_s = Q$	–	–	✓	✓
$Q_u = P_f + P_s$	–	–	–	✓

Note. S represents the conceptual storage value, P represents the precipitation, Q represents the discharge, and E represents the evapotranspiration. The subscripts f, u, and s represent fast, unsaturated, and slow, respectively.

Table 2
Constitutive Functions of the Models Used in the Experiments (✓ and "–" Indicate Presence or Absence, Respectively)

Constitutive functions	M_F	M_{UF}	M_{FS}	M_{UFS}
$Q_f = k_f {S_f}^{\alpha}$	✓	✓	–	–
$Q_f = k_f S_f$	–	–	✓	✓
$E = C_e E_p \left(1 - e^{-\frac{S_f}{m}}\right)$	✓	–	✓	–
$\overline{S}_u = S_u / S_{uMax}$	–	✓	✓	✓
$Q_u = P {\overline{S}_u}^{\beta}$	–	✓	✓	✓
$E = C_e E_p \left(\frac{\overline{S}_u(1+m)}{\overline{S}_u + m}\right)$	–	✓	–	✓
$P_s = DP$	–	–	✓	✓
$Q_s = k_s S_s$	–	–	✓	✓

3.2. Selection of Hydrological Signatures

As discussed in Section 1, previous work established that signatures of aridity and baseflow correlate with identifiable model complexity. In particular, arid catchments appear to require more complex models than wet catchments (Atkinson et al., 2002; Massmann, 2020), and catchments with high baseflow typically need more complex models than catchments with low baseflow (Coxon et al., 2014; Massmann, 2020). Such behavior lends itself to process-based interpretation, which is an important prerequisite when attempting to hypothesize a relationship between signatures and model structures. In particular, the effect of aridity on identifiable model complexity can be explained by taking into consideration that the more arid the catchment, the more variable the hydrological regime, potentially uncovering storage–discharge nonlinearities or thresholds that may remain hidden when catchments are permanently in a wet state. The effect of baseflow on model complexity is often related to the presence of deep groundwater, whose contribution to the hydrograph may require an independent model representation. Such considerations underlie the hypotheses about the mapping between model structure and signatures, which are illustrated in Section 3.4. The aridity and baseflow signatures used in this study are defined as follows:

Aridity index. I_A, which is defined as the ratio of long-term average potential evapotranspiration to long-term average precipitation:

$$I_A = \frac{\sum_{t=1}^{N_T} E_{Pot,t}}{\sum_{t=1}^{N_T} P_t} \tag{1}$$

where E_{Pot} indicates the potential evaporation at an individual catchment, P is precipitation, t is the time index, and N_T is the number of observations.

Baseflow index. I_B, which is defined as

$$I_B = \frac{\sum_{t=1}^{N_T} Q_t^{(b)}}{\sum_{t=1}^{N_T} Q_t} \tag{2}$$

where Q stands for streamflow, and $Q^{(b)}$ for baseflow (other variables are defined above). We use a popular filter for calculating baseflow as proposed by Lyne and Hollick (1979):

$$Q_t^{(b)} = \min\left(Q_t, \vartheta_b Q_{t-1}^{(b)} + \frac{1-\vartheta_b}{2}(Q_{t-1} + Q_t)\right) \tag{3}$$

As recommended by Nathan and Mcmahon (1990) for daily streamflow, the filtering parameter ϑ_b was set to 0.925, and three passes (forward, backward, and forward) of the filter were used.

We note that different baseflow separation methods exist (e.g., Carlotto & Chaffe, 2019; Eckhardt, 2008). Although such methods may result in different baseflow index values, they should provide a similar ranking between catchments (i.e., if catchment A has a lower baseflow index than catchment B, then the same ranking will be maintained with different methods). As we

focused on relative catchment differences in this work, we expected that our results are not particularly sensitive to the choice of baseflow separation methods or their parameterization.

3.3. Model Evaluation

3.3.1. Calibration–Validation Approach

We adopted a "temporal split sample" validation strategy (Klemeš, 1986), based on subdividing the time series of the observed streamflow into two periods. In particular, we partitioned the 20-year observation period into two consecutive 10-year periods: period 1, from 1 January 1985 to 31 December 1994, and period 2, from 1 January 1995 to 31 December 2004. Each period was then expanded by prepending five additional years of model inputs, which were used for model warm-up. The calibration–validation was repeated in each period, and the corresponding streamflow time series for calibration or validation were concatenated. In this way, the calibration and validation time series spanned the (same) full period, which facilitated their comparison. All results are presented in validation, unless otherwise indicated.

Model calibration was conducted using the Nash and Sutcliffe efficiency (Nash & Sutcliffe, 1970) of the square root of the streamflow as an objective function:

$$F_{\mathrm{NS}} = 1 - \frac{\sum_{t=1}^{N_{\mathrm{T}}} \left(Q_{\mathrm{sim},t}^{\lambda} - Q_{\mathrm{obs},t}^{\lambda} \right)^2}{\sum_{t=1}^{N_{\mathrm{T}}} \left(Q_{\mathrm{obs},t}^{\lambda} - \mathrm{ave}\left(\mathbf{Q}_{\mathrm{obs},1:N_{\mathrm{T}}}^{\lambda} \right) \right)^2} \tag{4}$$

where sim stands for simulated, obs for observed, ave indicates the average, and λ is an exponent. The exponent λ was fixed at 0.5 to mitigate the heteroscedasticity in model residuals, which is otherwise typically present when $\lambda = 1$ (McInerney et al., 2017).

The Shuffled Complex Evolution (SCE-UA; Duan et al., 1992, 1993) was used as the calibration algorithm for the inference of the models' parameters. The SCE-UA has some parameters that must be configured by the user. For this study, we used 7,000 iterations, 5 complexes, and 1 simplex. Each catchment took, on average, 20 min for one calibration. Considering that four models were applied on two periods in 508 catchments, a total of 4,064 calibrations were carried out.

Given the already high computational burden, no further sensitivity or uncertainty analyses were carried out. Such analyses may contribute to a more comprehensive assessment of which model components are necessary and which ones are superfluous in each individual catchment (e.g., Bai et al., 2009). Here, such insights were inferred from model comparisons, as illustrated in the following section, using the rationale, typical in the application of a top-down framework (Sivapalan et al., 2003), that when additional complexity does not result in improved performance it is unnecessary since the usefulness of such resultant data is poor.

3.3.2. Analysis of Streamflow Simulations

Model performance was assessed using model validation results. The first assessment of model performance used the F_{NS} objective function defined in Equation 4. For subsequent analyses, we removed those catchments for which all models performed poorly which was defined here by F_{NS} lower than the specified threshold fixed at 0.5. This low model performance may have been a consequence of data errors or model inadequacy. In either case, it would have been problematic to associate a model structure to a given catchment if the model performance had been exceedingly low. Thus, the subsequent analyses were not considered meaningful if at least one of the four model structures did not perform adequately.

While F_{NS} provides an absolute performance, in order to compare models, we were also interested in the relative performance between models. For this purpose, we calculated a relative performance indicator, which was defined as follows:

$$\Delta F_{NS}^{(Mi,Mj)} = \frac{\left(1 - F_{NS}^{(Mi)}\right) - \left(1 - F_{NS}^{(Mj)}\right)}{1 - F_{NS}^{(Mi)}} \quad (5)$$

where M stands for model, i and j are model indices (F, fast reservoir; UF, unsaturated and fast reservoir; FS, fast and slow reservoir; UFS, unsaturated, fast, and slow reservoir). Hence, for example, $F_{NS}^{(M_F)}$ would correspond to the F_{NS} of model M_F in the validation period.

The indicator $\Delta F_{NS}^{(Mi,Mj)}$ was calculated for pairs of models that differ for the addition of a single reservoir, with the index i referring to the model without the reservoir, and with the index j referring to the model with the reservoir. The indicator $\Delta F_{NS}^{(Mi,Mj)}$ was therefore calculated for the following model pairs: M_F–M_{UF}, M_F–M_{FS}, M_{UF}–M_{UFS}, and M_{FS}–M_{UFS}. A positive value of $\Delta F_{NS}^{(Mi,Mj)}$ indicates that adding a reservoir improves model performance. The denominator was used to give greater weight to the improvement in performance the higher the F_{NS} values. For example, an improvement from $F_{NS}^{(Mi)} = 0.45$ to $F_{NS}^{(Mj)} = 0.50$ (i.e., $\Delta F_{NS}^{(Mi,Mj)} = 0.09$) has a smaller $\Delta F_{NS}^{(Mi,Mj)}$ than an improvement from $F_{NS}^{(Mi)} = 0.75$ to $F_{NS}^{(Mj)} = 0.80$ (i.e., $\Delta F_{NS}^{(Mi,Mj)} = 0.20$).

3.3.3. Analysis of Model Parameters and States

In order to gain further insights into model behavior in relation to the different catchments, we considered the most complex model M_{UFS} and performed the following analyses.

- *Analysis of selected model parameters*. We analyzed the inferred values of the partitioning parameter D in relation to the observed signatures. This parameter divides the flow between (sub)superficial flow (fast reservoir) and baseflow (slow reservoir). The higher its value, the greater the proportion of water flows to the slow reservoir.
- *Analysis of selected states*. We analyzed the inferred states of two reservoirs, UR and SR, according to the water balances in Table 1. Specifically, we considered the difference ΔS_u and ΔS_s between the maximum and the minimum storage for each calibration period in UR and SR, respectively. Larger values of ΔS indicate a larger accumulation of water in the reservoir.

These diagnostics on model parameters and states were selected because of considered informative of model behavior. In particular, besides being potentially informative on the correspondence of model parameters to signatures—and, therefore, on whether model components correspond to their intended process representation—these diagnostics may also indicate excessive or misused model complexity and therefore complement the previous model comparisons. For example, the calibrated value of the parameter D may be such that it effectively excludes a reservoir; or a relatively small ΔS value may indicate a relatively constant reservoir state and therefore that a reservoir element may be superfluous. The employment of these diagnostics in the model evaluation process is described in the following section.

3.4. Development of Key Hypotheses

Our selection of model structures and signatures is motivated by the following hypotheses:

1. The three-element model structure M_{UFS}—which, as discussed in Section 3.1, is broadly consistent with the "best compromise" models identified in previous studies—is representative of the main catchment functions of partition, storage, and release. However, not all of these functions may be present in all catchments, meaning that simpler models may work comparatively well. In particular
    a. The SR model component, which is intended to simulate groundwater processes, is not needed in catchments with low baseflow.
    b. The UR component, which is intended to model the nonlinear partitioning of precipitation between effective rainfall and evaporation depending on catchment storage, is not needed in very wet catchments as they should be close to saturation.

Based on the considerations above, we expected that a simpler model should work well in wet catchments with low baseflow and that more complex models are needed in the other scenarios.

In terms of the analyses of model parameters and states, we expected that

2. Catchments with a high baseflow index have a greater percentage of flow into the slow reservoir, resulting in larger D values.
3. Larger values of ΔS_u indicate a larger variation of water in the unsaturated soil, which we expected to happen in arid or highly seasonal catchments. In wet catchments, instead, this reservoir should always be close to saturation.
4. Larger values of ΔS_s indicate a larger variation of water in the groundwater storage, which is expected in catchments with a higher baseflow index.

Besides relating relative model performance to signatures, we also expected that a posteriori, it is possible to relate relative model performance to landscape properties. In particular:

5. We expected that relative model performance can be related to properties such as soil type, soil depth, and topography, which in turn affect the signatures of streamflow response.

For testing these hypotheses, we used a baseflow index to classify catchments into three classes:

- Low baseflow (LIB), when $I_B < I_{B,Low}$.
- Medium baseflow (MIB), when $I_{B,Low} \leq I_B < I_{B,High}$.
- High baseflow (HIB), when $I_B \geq I_{B,High}$,

where the following threshold values were used: $I_{B,Low} = 0.40$, and $I_{B,High} = 0.60$. The motivation for these thresholds was to have an intermediate class (MIB) where the partitioning of flow would represent an equal partitioning of flow between baseflow and quick flow (i.e., around $I_B = 0.50$), and two extreme classes with catchments characterized by mostly quick flow (LIB) or mostly low flow (HIB).

Additionally, we classified the catchments into three classes of aridity index, using the following approach:

- Low aridity (LIA), when $I_A < I_{A,Low}$.
- Medium aridity (MIA), when $I_{A,Low} \leq I_A < I_{A,High}$.
- High aridity (HIA), when $I_A \geq I_{A,High}$,

where the following threshold values are used: $I_{A,Low} = 0.90$, and $I_{A,High} = 1.10$. The motivation for these thresholds was to separate catchments that are energy limited (LIA) from those that are water limited (HIA) through an intermediate class (MIA) that contains the separation threshold ($I_{A,Low} = 1$). We argue that such hypotheses and classifications using baseflow and aridity are pertinent because, as discussed in Section 2, hydrological process variability in the study area is closely related to these signatures.

The combination of three classes of baseflow (I_B) and aridity (I_A) indices leads to nine signature classes. These nine classes are used to broadly distinguish catchment response behavior.

4. Results

4.1. Catchment Classes

Figure 3 maps the catchments based their signatures classification, where panel a indicates the three classes associated with baseflow index (I_B), panel b shows the three classes associated with the aridity index, and panel c shows the nine classes associated with the combination of these two signatures. While I_B and I_A had a clear regional pattern, the nine classes combining I_B and I_A were more mixed. The aridity index was lower in the southern and northern regions of the country, increasing toward the northeast. The latter region was characterized by the greatest seasonality and the lowest annual precipitation volume in the country. The baseflow index values were higher in the southwest and inland parts of the northeastern region (the transition from the semiarid to the Amazon). The catchments with low aridity (classes LIA–LIB, LIA–MIB and LIA–HIB) were located in the

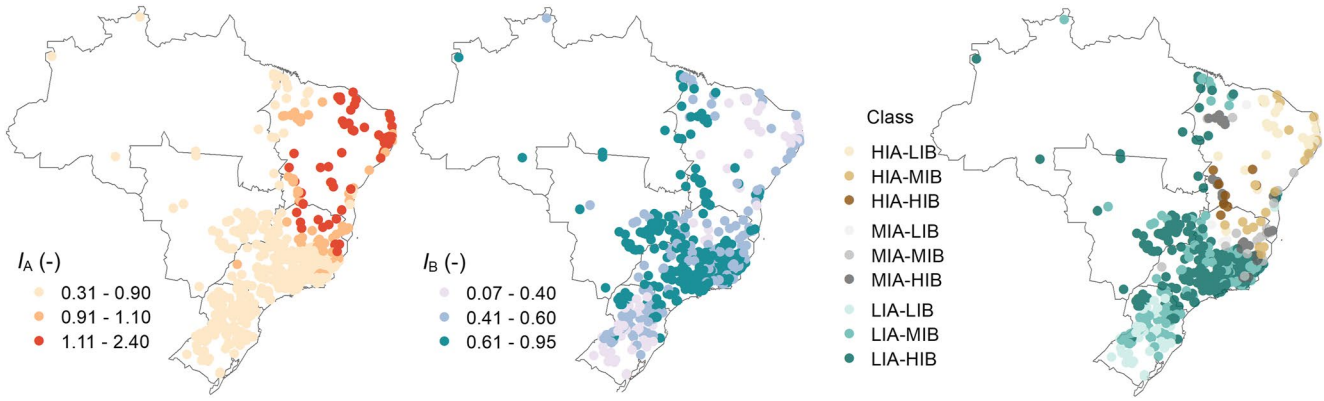

Figure 3. Spatial variability of hydrological signatures and definition of catchment classes. (a) Aridity index (I_A). (b) Baseflow index (I_B). (c) Catchments classes according to I_A and I_B. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B. Gray lines indicate the regions of Brazil.

south, southwest, central-west, and northern regions. The LIA–LIB class was located mainly in the south, while the LIA–MIB and LIA–HIB classes were distributed in the other regions. The classes with medium I_A (MIA–LIB, MIA–MIB, and MIA–HIB) were in the inland northeast and in the upper part of the southeast region, which are in climate transition zones. The high I_A classes were mainly in the northeastern region. The HIA–HIB was in the interior part, while HIA–MIB and HIA–HIB classes were closer to the coast.

4.2. Analysis of Streamflow Simulations

Figure 4 maps the validation performance of the four model structures in terms of Nash and Sutcliffe efficiency (F_{NS}). This figure illustrates the following points:

- Model performance generally increased from M_F to M_{UFS}, which we attributed to the fact that more complex models are more flexible in terms of ability to adapt to different catchment dynamics.
- However, there were catchments where model M_F performed relatively well ($F_{NS} > 0.75$), which illustrated that even such a simple model could provide good performance in some catchments. These catchments were located in the southern region of Brazil and had I_A values between 0.48 and 0.63 and I_B between 0.42 and 0.67.
- There were few catchments where none of the models provided a sufficiently good performance (here specified as $F_{NS} < 0.5$), which left us with 466 catchments (from the initial sample of 508 catchments) for further analyses.

Figure 5 shows the analysis of model performance for the nine catchment classes based on the combination of baseflow and aridity indices. Looking at the extreme catchment classes (all low–high combinations, excluding middle ranges, hence LIA–LIB, LIA–HIB, HIA–LIB, and HIA–HIB), it can be observed that these classes behave differently in terms of relative model performance.

- Model M_F achieved the best average performance in the class LIA–LIB, meaning that this class was the easiest to model. In LIB (hence LIA–LIB, HIA–LIB), M_{UF} and M_{UFS} were both better than M_F and M_{FS}, which means that adding UR was more effective than adding SR.
- In HIB (hence LIA–HIB, HIA–HIB), M_{FS} and M_{UFS} were both better than M_F and M_{UF}, which means that adding SR was more effective than adding UR.
- In LIA–LIB, there was no benefit in adding a groundwater reservoir, whereas in HIA–LIB there was a benefit in adding a groundwater reservoir.
- HIA–LIB exhibited a high percentage of failure for all four model structures. For the most complex one, M_{UFS}, 41% of the catchments presented $F_{NS} < 0.5$.
- The percentage of failure for model structures without SR increased from classes with low I_B to classes with high I_B.

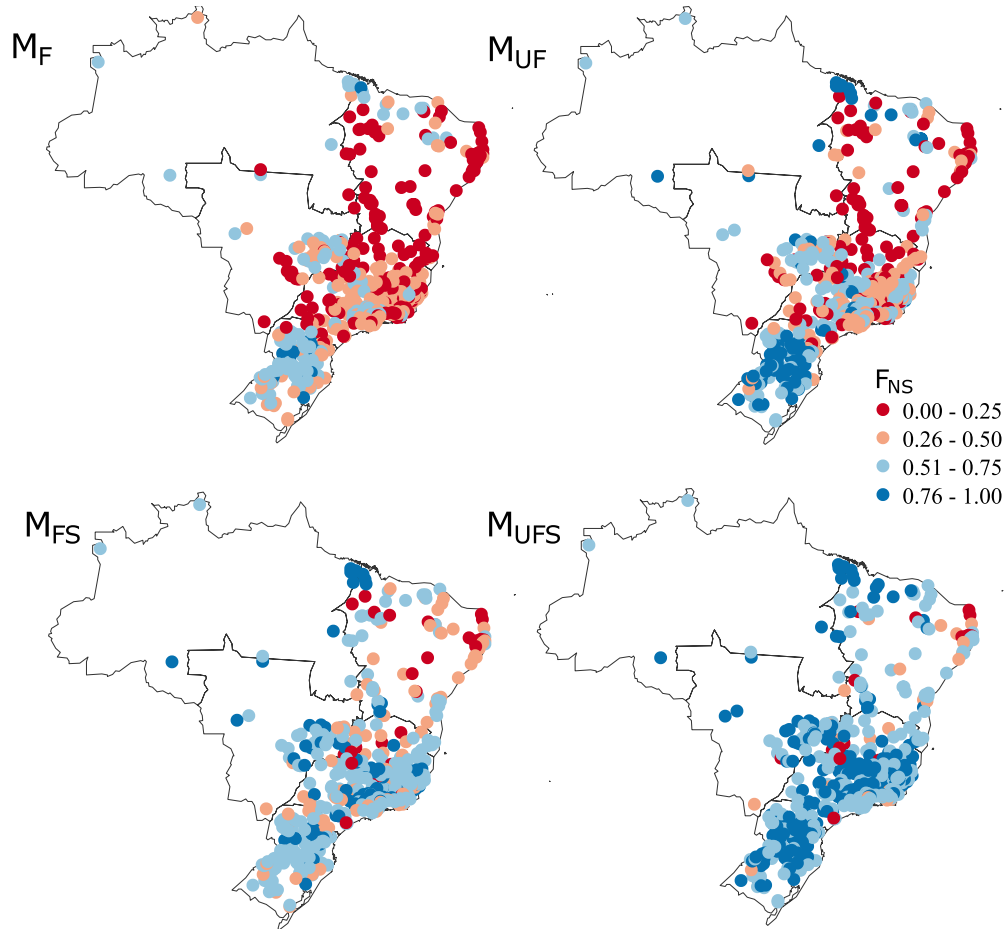

Figure 4. Spatial distribution of Nash and Sutcliffe efficiency (F_{NS}) values in the validations of each of the four model structures evaluated. Gray lines indicate the regions of Brazil.

Figure 6 shows the spatial distribution of the differential improvement in going from a simpler to a more complex structure, defined by $\Delta F_{NS}^{(Mi,Mj)}$ (Equation 5). Comparing with Figures 3, Figure 6 shows that

- The inclusion of UR did not improve performance in catchments with high I_B (in the southeast and northeast interior regions), especially when we compared M_F with M_{UF}. Regions with smaller I_B (south and coastal part of the northeast) showed higher improvement.
- In catchments with low I_B (southern region), the inclusion of SR degraded the performance or just slightly improved it. On the other hand, for catchments with high I_B, the inclusion of the SR improved the performance of both simpler and more complex models.

Figure 7 shows the differential improvement, $\Delta F_{NS}^{(Mi,Mj)}$, stratified for the nine catchment classes. This figure complements Figure 6 which, by providing absolute model performances, does not show whether the improvement is consistent across all catchments. Figure 7 leads to the following results:

- In LIA–LIB, adding UR generally led to an improvement, whereas adding SR could lead to both an improvement and a deterioration (mostly deterioration in the case of $\Delta F_{NS}^{(M_{UF},M_{UFS})}$).
- In HIB (hence LIA–HIB, MIA–HIB, and HIA–HIB), while the inclusion of UR and SR reservoirs improved the simulation, clearly the impact of the slow reservoir was higher than that of the inclusion of the unsaturated zone reservoir.
- In HIA, the inclusion of the SR became more important than that of UR when the baseflow was dominant. On the other hand, the inclusion of UR led to an improvement in the majority of the catchments.

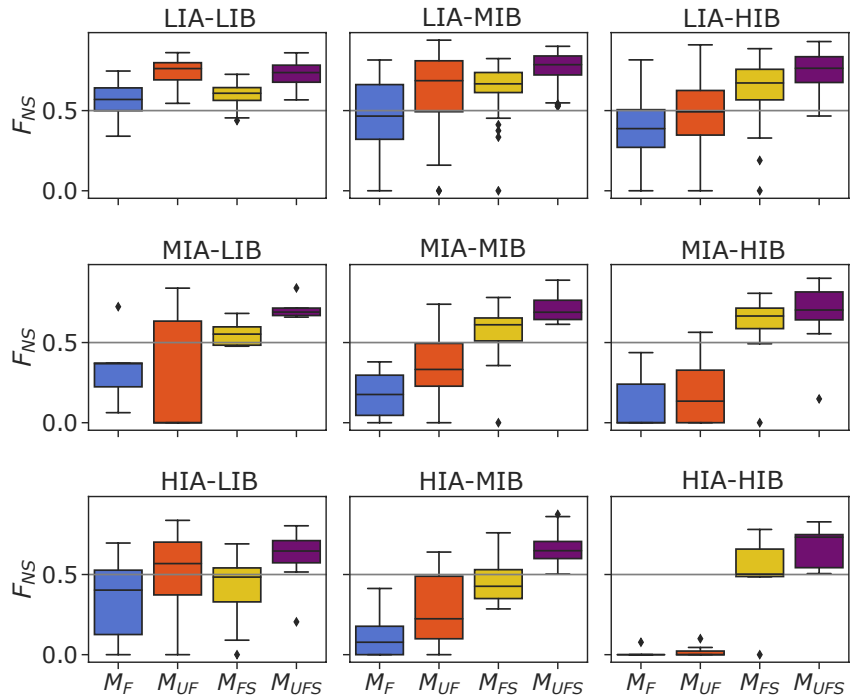

Figure 5. Model performance in the validation of each model structure for the nine catchment classes. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B.

4.3. Analysis of Hydrographs on Selected Catchments

As it is difficult to assess how much model simulations differ from F_{NS} values alone, some emblematic examples of differences in model simulations for the four more extreme catchment classes are provided hereafter. Figure 8 exemplifies the hydrographs simulated by the four model structures in the four catchments that belong to the four extreme categories LIA–LIB (Figure 8a), LIA–HIB (Figure 8b), HIA–LIB (Figure 8c), and HIA–HIB (Figure 8d):

- In catchment 83345000 (the number is the gauge station code adopted in the Brazilian national data set) in the Hydrological Region South Atlantic, belonging to the LIA–LIB class, all models were broadly able to follow the hydrograph dynamics. Hence, in this catchment, even the single-reservoir model M_F provided a relatively good performance, as also apparent form Figure 8, where the simulations of the four models appear similar. A closer examination indicates that M_F and M_{FS} performed similarly to each other, and M_{UF} and M_{UFS} performed similarly as well. As these pairs of models differed from the addition of a groundwater reservoir, this similarity illustrates how in this catchment class the addition of a groundwater reservoir did not result in an appreciable improvement in model simulations. The addition of UR, instead, appeared to be marginally beneficial.
- In catchment 6095000 (in the Hydrological Region Paraná), belonging to the LIA–HIB class, it can be observed that both M_F and M_{UF} suffered from similar deficiencies, by not being able to capture both the fast and the slow hydrograph dynamics. M_{FS} and M_{UFS}, instead, could match the fast and slow hydrograph modes and had similar behavior. In this catchment, therefore, the addition of SR was essential, whereas the addition of UR did not appear to have an influence.
- In catchment 3524000 (in the Hydrological Region Eastern Atlantic), belonging to the HIA–LIB class, it is clearly visible how M_F was not suitable, with a negative F_{NS}. Although $M_{FS,}$ which includes a groundwater reservoir, had a much better performance than M_F, M_{UF} was the simplest model that provided adequate performance. The addition of a groundwater reservoir, exemplified by M_{UFS}, led to a similar behavior. Therefore, in this catchment, the addition of UR appeared to be essential, whereas the addition of SR had a minor influence.
- In catchment 33205000 (in the Hydrological Region Eastern Atlantic), belonging to the HIA–HIB class, one can see how one needs the full range of model complexity. Only model M_{UFS} can match the full range of hydrograph responses, whereas comparatively, all other models demonstrate strong deficiencies.

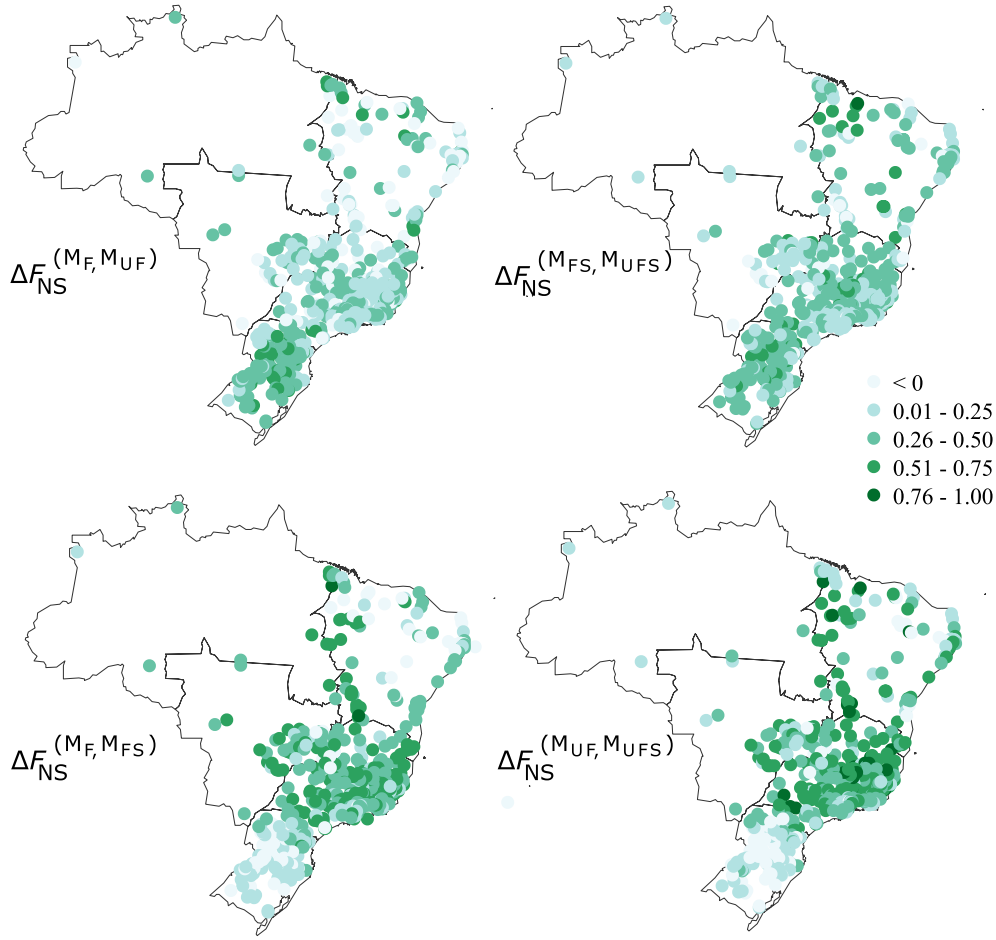

Figure 6. Spatial distribution of the four differential improvements analyzed, $\Delta F_{NS}^{(Mi,Mj)}$, stratified for the nine catchment classes. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B. Gray lines indicate the regions of Brazil.

The hydrographs of these catchments exemplify similar behavior to the other catchments belonging to the same classes.

4.4. Analysis of Model Parameters and States

The values of the parameter D of M_{UFS} for each class of catchments (top panel), and for the two calibration periods, can be seen in Figure 9. The higher its value, the greater proportion of water flows to the slow reservoir. The parameter D value was higher for catchments with a high baseflow index. This result indicates that the model represented as intended the division of the flow into fast and slow components, which confirmed the importance of adding the slow reservoir for catchments with high I_B. It is also noted that for the two periods the distribution of the parameter values was similar, that is, they were not sensitive to the calibration period. Thus, a period of 10 years in this case was sufficiently long to capture hydrological temporal variability. Independently of catchment dryness, there was a strong relationship between parameter D value and baseflow index (Figure 9b). As this parameter can be directly compared with I_B, this is a strong indication that the model's internal dynamics were consistent with their intended process representation.

Figure 10 presents the difference between the maximum and the minimum inferred states of UR and SR according to the nine catchments classes for one calibration period (both periods presented similar results). For the UR, we can see that

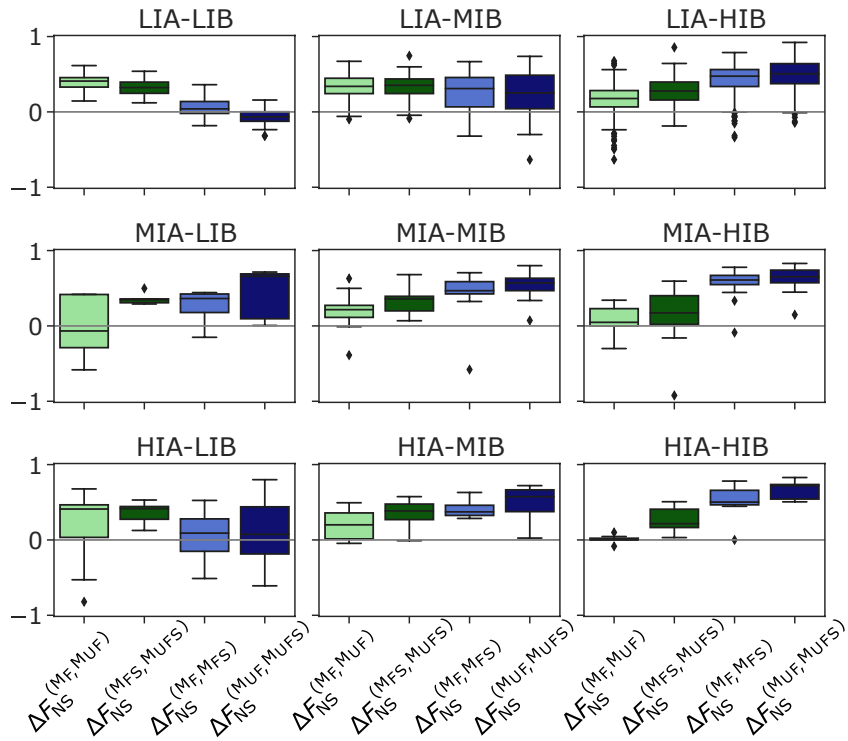

Figure 7. Differential improvements analyzed, $\Delta F_{NS}^{(Mi,Mj)}$, stratified for the nine catchment classes. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B.

- HIA classes (hence HIA–LIB, HIA–MIB, and HIA–HIB) had the largest UR storages. These drier catchments usually presented a nonlinear response and therefore the model had to store more water in UR in order to represent this nonlinearity.
- LIA–LIB presented the smallest UR storage values. As shown before, these catchments were the easiest to model, benefitting the least from extra model elements compared to other classes.
- HIB classes (LIA–HIB, MIA–HIB, and HIA–HIB) presented larger storage values than classes with smaller I_B and similar I_A values.

As for the SR storages, we can see that HIB classes (hence LIA–HIB, MIA–HIB, and HIA–HIB) presented larger UR storage when comparing with the classes with the same I_A range.

5. Discussion

Our model validation results showed a connection between the performance of the four model structures and the two hydrological signatures of aridity index (I_A) and baseflow index (I_B). These results are broadly consistent with the hypotheses defined in Section 3.4, as detailed below.

The easiest catchments to model were those with low I_A and low I_B (hypothesis 1a and 1b). In these catchments, the simplest model M_F presented F_{NS} higher than 0.5 in 74.5% of the catchments. This result can be interpreted by considering the landscape characteristics of these catchments, their perceived dominant processes, and their model representation (hypothesis 5). These catchments are mainly located in the southern region, they are highly responsive to rainfall and are characterized by the highest percentages of clay and lowest soil depth (Figure S1). High clay content indicates limited infiltration and therefore small groundwater flow, whereas limited soil depth associated with relatively wet conditions indicates small fluctuations in storage dynamics in the unsaturated zone. We have explored the relationship using correlation analysis and did not find a significant correlation between baseflow index values and soil characteristics. Nevertheless, these characteristics could explain why

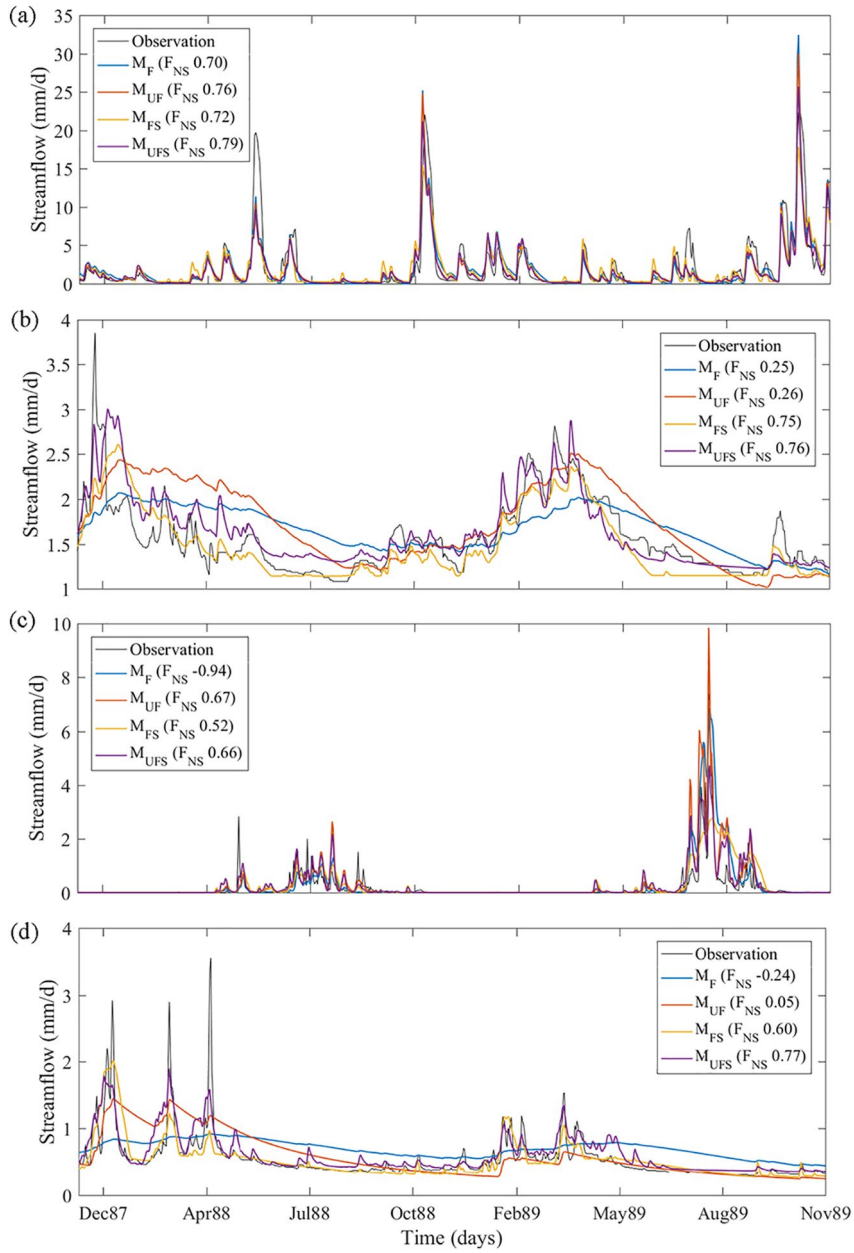

Figure 8. Observed and simulated hydrographs (validation using Nash–Sutcliff efficiency) of four catchments that belong to the four extreme categories. (a) Low aridity and low baseflow catchment (ID 83345000, in the Hydrological Region South Atlantic). (b) Low aridity and high baseflow catchment (ID 60950000, in the Hydrological Region Paraná). (c) High aridity and low baseflow catchment (ID 35240000, in the Hydrological Region Eastern Atlantic). (d) High aridity and high baseflow catchment (ID 33205000, in the Hydrological Region Eastern Atlantic).

the groundwater and unsaturated model compartments are superfluous in these catchments, and a single bucket model is relatively well performing. This result generalizes other studies, where single bucket models were found to perform well in wet catchments with low baseflow, such as Maimai (Kavetski & Fenicia, 2011) and Plynlimon (Kirchner, 2009).

Dry catchments, instead, were generally harder to model. For some dry catchments, especially with low I_B, all models performances were considered insufficient ($F_{NS} < 0.5$). This result is consistent with other studies (e.g., Coxon et al., 2014; Massmann, 2020; Parajka et al., 2013; Poncelet et al., 2017) that found that catchment aridity is one of the main characteristics that affect model performance. It can be hypothesized that dry catchments are

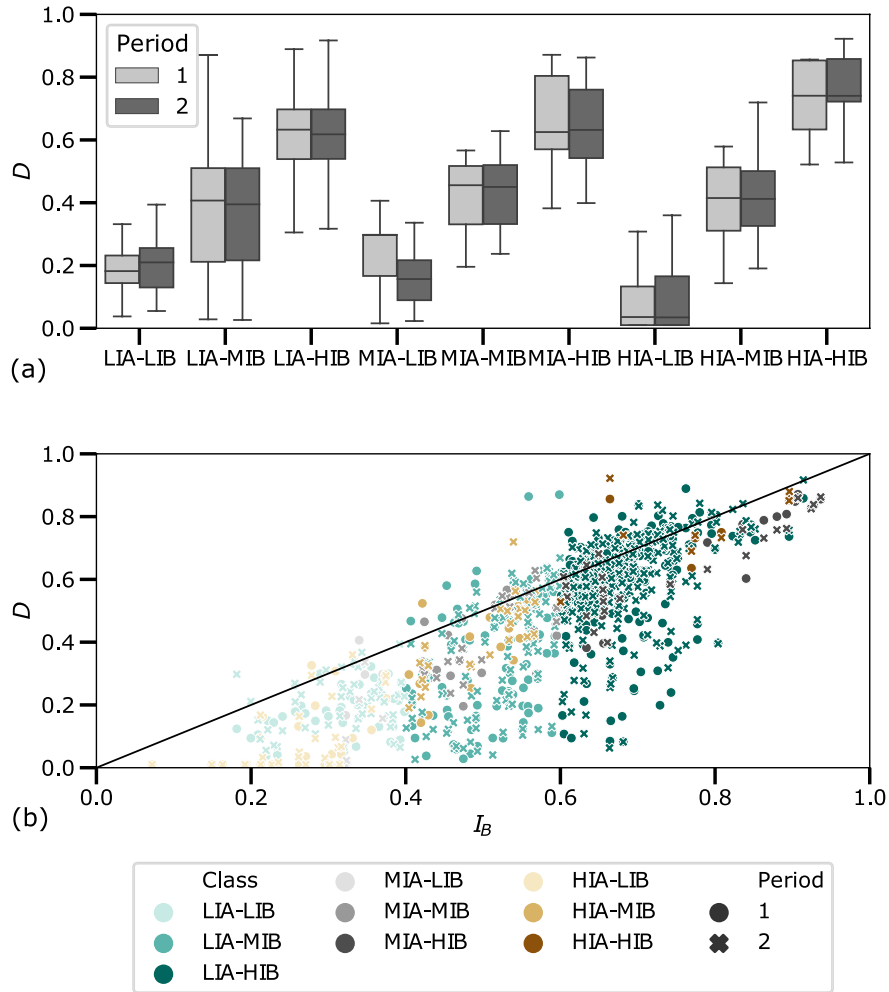

Figure 9. (a) Distribution of parameter D values (i.e., the proportion of flow directed to the slow reservoir) for the nine catchment classes. Period 1 (light gray) corresponds to the calibration between 1985 and 1994 and period 2 (dark gray) to the calibration from 1995 to 2004. (b) Scatterplot between baseflow index and the parameter D. The black line indicates a 1:1 line. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B.

more difficult to simulate because of a more complex water balance than wetter catchments. The impact of evapotranspiration and changes in groundwater is reduced in wetter catchments, facilitating the representation of the water balance by hydrological models (Poncelet et al., 2017). Those catchments may also have long periods with flow close to zero or zero, which makes modeling especially difficult (e.g., McInerney et al., 2019).

The inclusion of the unsaturated zone reservoir improved the model performance in dry catchments (hypothesis 1b), which we interpret in terms of a highly nonlinear response compared with wet catchments, that is, the water balance depends less on precipitation and more on evapotranspiration and underground storage. The UR is therefore important to represent this nonlinearity. This reservoir appears to be more important for catchments with low I_B (south and coast of the northeast), as we can see in Figure 6, with a mean F_{NS} for models with UR above the threshold of 0.50 and models without with a mean F_{NS} below 0.50. The inclusion of the UR does not improve performance in catchments with high I_B (in the southeast and northeast inland regions), especially when we compare M_F with $M_{UF,}$ with both structures with a 100% failure for HIA–HIB. This result is also evident in HIB catchments, where the inclusion of UR did not improve the time and shape of the hydrographs (Figure 8). This result may be explained by taking into account that in such high I_B catchments, groundwater processes are more dominant than soil related processes.

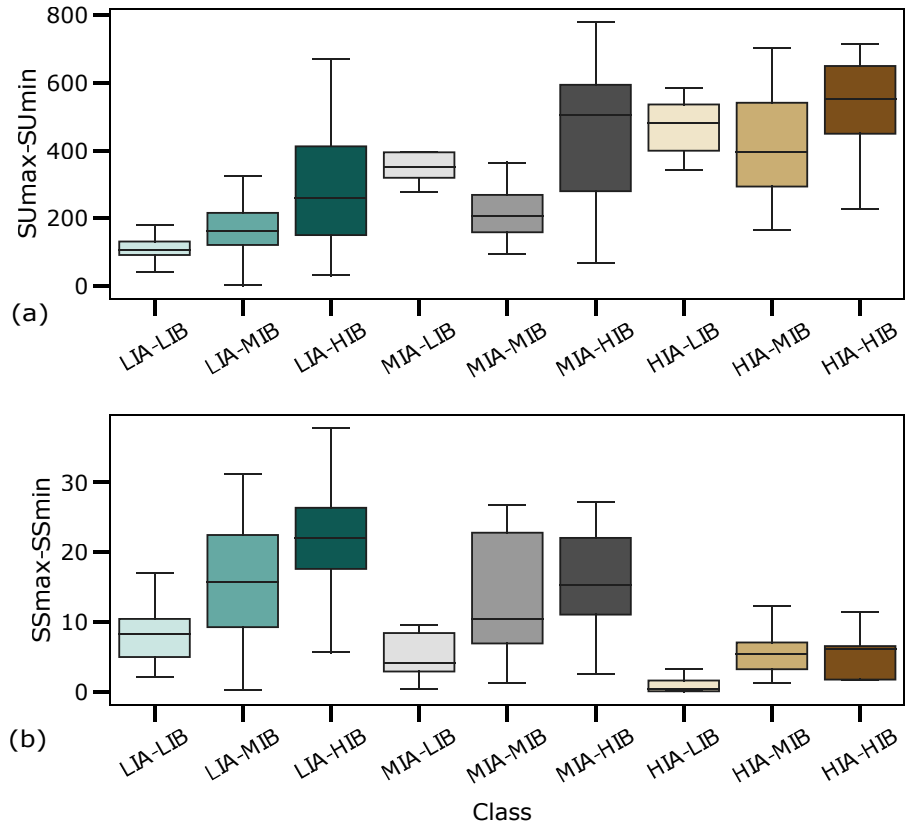

Figure 10. Difference between the maximum and the minimum inferred states of (a) the unsaturated zone reservoir (UR) and of (b) the slow reservoir (SR) according to the nine catchment classes. LIA, low I_A; MIA, medium I_A; HIA, high I_A; LIB, low I_B; MIB, medium I_B; and HIB, high I_B.

The inclusion of the slow reservoir improved the performance of both dry and wet catchments with baseflow indices greater than 0.60 (HIB classes, hypothesis 1a). For HIB classes, the number of catchments with F_{NS} greater than 0.50 changed from 40% for M_{UF} to 87% for M_{UFS}. This indicates that when a process becomes increasingly dominant in the total streamflow, its respective model representation becomes more necessary. The percentage of failure decreased from 54.6% with M_{UF} to 6.25% with M_{UFS} for LIA–HIB catchments. The regions with larger improvements were the ones with higher I_B and greater soil depth (hypothesis 5), that is, those in the southwest and inland of the northeastern region (Figure 6). Catchments with greater soil depth probably exhibit a greater contribution from the baseflow to the total runoff, making it necessary to add the slow reservoir. For the LIB classes (I_B smaller than 0.40) the performance worsened with SR (as seen in Figure 7), especially in the southern region of Brazil, which had the smallest values of I_B and soil depth. This is an example where increased model complexity can degrade model performance in validation. This result reinforces the hypothesis that the slow reservoir becomes superfluous where the processes it is intended to represent are not dominant. Results similar to ours, where parallel models performed better in catchments with dominant groundwater flows, are consistent with previous studies (e.g., van Esse et al., 2013), although in the UK it has been shown that this is not always the case (Lee et al., 2005). High values of I_B have been related to lower streamflow responses to changes in precipitation (e.g., Sawicz et al., 2011), suggesting the need for an independent reservoir with a slow and constant flow. Small values of I_B may indicate catchments with rapid response (Addor et al., 2017) and greater streamflow variability, reducing the need for the slow reservoir. Nevertheless, in HIA–LIB, there is a benefit in adding a groundwater reservoir, which could be attributed to the impact of permeability in dry catchments.

The model structures without SR, hence M_F and M_{UF}, performed worse in areas with a higher percentage of sand (comparing Figure 4 and Figure S1). We hypothesize that this result is due to those catchments having higher soil permeability with larger groundwater flow, with the inclusion of a slow reservoir being necessary

to model its behavior (hypothesis 5). On the other hand, these models performed better in catchments with high clay content, which can be explained by low soil permeability and therefore limited groundwater flow. Although such a model performance pattern may lead to the hypothesis of a connection between sand/clay percentages and baseflow index, our independent analyses could not determine such a relationship, indicating that the connection of patterns of model performance and landscape properties is more difficult to establish than the one with streamflow signatures.

The analysis of the partitioning parameter D of M_{UFS} showed that this parameter is strongly related to I_B (Figure 9). This result shows that this model represents, as intended, the partitioning of the flow into fast and slow components, lending confidence in the hypothesis that this parameter is relatively well identifiable (hypothesis 2) and confirming the importance of adding the slow reservoir for catchments with high I_B (hypothesis 1a).

By analyzing the UR storage dynamics, we found that the variability in UR storage was higher in the HIA class than in other classes (hypothesis 3). This result explains why this reservoir is particularly necessary in arid catchments (hypothesis 1b). SR storage dynamics confirm a larger accumulation of water in the groundwater storage in catchments with a high baseflow index (hypothesis 4), which clarify the function of this storage in such conditions, and its necessity (hypothesis 1a).

It should be noted that the relative performance between models ($\Delta F_{NS}^{(Mi,Mj)}$) for simple and complex models was quite similar, indicating a correspondence between the element of the model and the hydrological process it represents. Our results provide increased evidence that the addition of complexity given by the number of parameters being calibrated does not necessarily improve processes representation, as has been previously reported (de Boer-Euser et al., 2017; Fenicia et al., 2008, 2014; Orth et al., 2015; van Esse et al., 2013). Therefore, the correct representation of the dominant processes in each catchment is more important than the complexity of the model structure. Models M_{UF} and M_{UFS}, for instance, have the same number of parameters (i.e., 5), but their performance in baseflow-dominated catchments is quite different, with mean F_{NS} close to zero for M_{UF} and close to 0.60 for M_{UFS} in LIA-HIB catchments. These results clearly show that the overall model architecture, and the processes that it is intended to represent, plays a more important role than merely the number of parameters in model performance.

6. Conclusions

This study investigated the relationships between model structures and hydrological signatures in order to interpret dominant hydrological processes at the catchment scale. To this end, we analyzed the performance of four different structures of conceptual hydrological models in 508 catchments located in Brazil. The use of such a large set of catchments provides the basis for robust generalizations.

- There was a relatively clear correspondence between classes of signatures and hydrological model structures. In particular, wet catchments with low baseflow were the easiest to model, showing the highest performance of the single bucket model structure (F_{NS} higher than 0.5 in 74.5% of the catchments). The addition of a groundwater reservoir was particularly impactful in catchments with high baseflow. Similarly, the addition of an unsaturated zone reservoir had a large influence in dry catchments.
- A posteriori, the catchment classes defined based on model performance could be associated with landscape properties and the associated dominant processes. Catchments simulated well by a single bucket were characterized by small soil depths and low permeability, whereas the need for a groundwater reservoir was evident in catchments with high permeability. This correspondence provides reassurance that suitable models are broadly representative of dominant catchment-scale processes.
- Analyzing the correspondence between model parameters, internal states, and signatures on the most complex model structure, we found that the partitioning parameter D is strongly related to the baseflow index, whereas the variation in the storage of the unsaturated and slow reservoir can be associated with the aridity index and baseflow index, respectively. These relationships provide confidence that model components are broadly representative of the processes they are designed to represent.

These results lay out a path for a better understanding of the correspondence between model structures, hydrological signatures, and landscape properties. Future work may reinforce the understanding of these connections through a larger sample of catchments, signatures, and model structures.

Data Availability Statement

The hydrometeorological data used for catchment classification and hydrological modeling in the study are freely available at Chagas et al. (2020) via http://doi.org/10.5281/zenodo.3964745.

Acknowledgments
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 88881.146046/2017-01. Paula C. David would like to thank Eawag for the EPP fellowship. V. Chagas would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the scholarship.

References

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

Atkinson, S. E., Woods, R. A., & Sivapalan, M. (2002). Climate and landscape controls on water balance model complexity over changing timescales. *Water Resources Research*, *38*(12), 1314. https://doi.org/10.1029/2002WR001487

Bai, Y., Wagener, T., & Reed, P. (2009). A top-down framework for watershed model evaluation and selection under uncertainty. *Environmental Modelling & Software*, *24*(8), 901–916. https://doi.org/10.1016/j.envsoft.2008.12.012

Beven, K. (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

Carlotto, T., & Chaffe, P. L. B. (2019). Master Recession Curve Parameterization Tool (MRCPtool): Different approaches to recession curve analysis. *Computers & Geosciences*, *132*, 1–8. https://doi.org/10.1016/j.cageo.2019.06.016

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., & Siqueira, V. A. (2020). CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil. *Earth System Science Data*, *12*(3), 2075–2096. https://doi.org/10.5194/essd-12-2075-2020

Coxon, G., Freer, J., Wagener, T., Odoni, N. A., & Clark, M. (2014). Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. *Hydrological Processes*, *28*(25), 6135–6150. https://doi.org/10.1002/hyp.10096

David, P. C., Oliveira, D. Y., Grison, F., Kobiyama, M., & Chaffe, P. L. B. (2019). Systematic increase in model complexity helps to identify dominant streamflow mechanisms in two small forested basins. *Hydrological Sciences Journal*, *64*(4), 455–472. https://doi.org/10.1080/02626667.2019.1585858

de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., et al. (2017). Looking beyond general metrics for model comparison—Lessons from an international model intercomparison study. *Hydrology and Earth System Sciences*, *21*(1), 423–440. https://doi.org/10.5194/hess-21-423-2017

Duan, Q., Gupta, V. K., & Sorooshian, S. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985

Duan, Q., Gupta, V. K., & Sorooshian, S. (1993). A Shuffled Complex Evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521. https://doi.org/10.1007/bf00939380

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., et al. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology*, *320*(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Eckhardt, K. (2008). A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. *Journal of Hydrology*, *352*(1–2), 168–173. https://doi.org/10.1016/j.jhydrol.2008.01.005

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, *47*, W11510. https://doi.org/10.1029/2010WR010174

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., & Freer, J. (2014). Catchment properties, function, and conceptual model representation: Is there a correspondence? *Hydrological Processes*, *28*(4), 2451–2467. https://doi.org/10.1002/Hyp.9726

Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, *52*, 954–989. https://doi.org/10.1002/2015WR017398

Fenicia, F., McDonnell, J. J., & Savenije, H. H. G. (2008). Learning from model improvement: On the contribution of complementary data to process understanding. *Water Resources Research*, *44*, W06419. https://doi.org/10.1029/2007WR006386

Freeze, R. A. (1974). Streamflow generation. *Reviews of Geophysics*, *12*(4), 627–647. https://doi.org/10.1029/RG012i004p00627

Gupta, H. V., Perrin, C., Bloschl, G., Montanari, A., Kumar, R., Clark, M., & Andreassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, *18*(2), 463–477. https://doi.org/10.5194/hess-18-463-2014

Jakeman, A. J., & Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model? *Water Resources Research*, *29*(8), 2637–2649. https://doi.org/10.1029/93WR00877

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*, W11511. https://doi.org/10.1029/2011WR010748

Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, *45*, W02429. https://doi.org/10.1029/2008WR006912

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13–24. https://doi.org/10.1080/02626668609491024

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*, e2019WR025975. https://doi.org/10.1029/2019WR025975

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Lee, H., McIntyre, N., Wheater, H., & Young, A. (2005). Selection of conceptual models for regionalisation of the rainfall-runoff relationship, *Journal of Hydrology*, *312*(1–4), 125–147. https://doi.org/10.1016/j.jhydrol.2005.02.016

Le Moine, N., Andréassian, V., Perrin, C., & Michel, C. (2007). How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. *Water Resources Research*, *43*, W06428. https://doi.org/10.1029/2006WR005608

Ley, R., Hellebrand, H., Casper, M. C., & Fenicia, F. (2016). Comparing classical performance measures with signature indices derived from flow duration curves to assess model structures as tools for catchment classification. *Hydrology Research*, *47*(1), 1–14. https://doi.org/10.2166/nh.2015.221

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., & Bergstrom, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, *201*(1–4), 272–288. https://doi.org/10.1016/s0022-1694(97)00041-3

Lyne, V., & Hollick, M. (1979). Stochastic time-variable rainfall runoff modelling. In *Proceedings of the Hydrology and Water Resources Symposium* (pp. 89–92).

Marengo, J. A., Liebmann, B., Grimm, A. M., Misra, V., Silva Dias, P. L., Cavalcanti, I. F. A., et al. (2012). Recent developments on the South American Monsoon System. *International Journal of Climatology*, *32*, 1–21. https://doi.org/10.1002/joc.2254

Massmann, C. (2020). Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments. *Hydrological Processes*, *34*(1), 4–20. https://doi.org/10.1002/hyp.13566

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/s0022-1694(04)00421-4

McInerney, D., Kavetski, D., Thyer, M., Lerat, J., & Kuczera, G. (2019). Benefits of explicit treatment of zero flows in probabilistic hydrological modeling of ephemeral catchments. *Water Resources Research*, *55*, 11035–11060. https://doi.org/10.1029/2018WR024148

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*, 2199–2239. https://doi.org/10.1002/2016WR019168

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nathan, R. J., & Mcmahon, T. A. (1990). Evaluation of automated techniques for base-flow and recession analyses. *Water Resources Research*, *26*(7), 1465–1473. https://doi.org/10.1029/WR026i007p01465

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., & Zappa, M. (2015). Does model performance improve with complexity? A case study with three hydrological models. *Journal of Hydrology*, *523*, 147–159. https://doi.org/10.1016/j.jhydrol.2015.01.044

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins—Part 1. Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, *17*(5), 1783–1795. https://doi.org/10.5194/hess-17-1783-2013

Perrin, C., Michel, C., & Andreassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1–4), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7

Poncelet, C., Merz, R., Merz, B., Parajka, J., Oudin, L., Andréassian, V., & Perrin, C. (2017). Process-based interpretation of conceptual hydrological model performance using a multinational catchment set. *Water Resources Research*, *53*, 7247–7268. https://doi.org/10.1002/2016WR019991

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

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

Schoups, G., Vrugt, J. A., Fenicia, F., & Giesen, van deN. C. (2010). Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. *Water Resources Research*, *46*, W10530. https://doi.org/10.1029/2009WR008648

Seiller, G., Anctil, F., & Roy, R. (2017). Design and experimentation of an empirical multistructure framework for accurate, sharp and reliable hydrological ensembles. *Journal of Hydrology*, *552*, 313–340. https://doi.org/10.1016/j.jhydrol.2017.07.002

Singh, S. K., Bárdossy, A., Götzinger, J., & Sudheer, K. P. (2012). Effect of spatial resolution on regionalization of hydrological model parameters. *Hydrological Processes*, *26*(23), 3499–3509. https://doi.org/10.1002/hyp.8424

Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*(11), 2101–2111. https://doi.org/10.1002/hyp.1425

Spieler, D., Mai, J., Craig, J. R., Tolson, B. A., & Schütze, N. (2020). Automatic model structure identification for conceptual hydrologic models. *Water Resources Research*, *56*, e2019WR027009. https://doi.org/10.1029/2019WR027009

van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., & Lobligeois, F. (2013). The influence of conceptual model structure on model performance: A comparative study for 237 French catchments. *Hydrology and Earth System Sciences*, *17*(10), 4227–4239. https://doi.org/10.5194/hess-17-4227-2013

Wagener, T., Sivapalan, M., Troch, P. A., & 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

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

Young, P. (2003). Top-down and data-based mechanistic modelling of rainfall–flow dynamics at the catchment scale. *Hydrological Processes*, *17*(11), 2195–2217. https://doi.org/10.1002/hyp.1328