Geoscience Data Journal

WILEY

Geoscience Data Journal RMetS

DATA ARTICLE OPEN ACCESS

SSP-CABra—Streamflow Scenarios Projections for Brazilian Catchments

André Almagro[1] | Paulo Tarso S. Oliveira[1,2] | André S. Ballarin[2]

[1]Faculty of Engineering, Architecture and Urbanism, and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil | [2]Department of Hydraulics and Sanitation, São Carlos School of Engineering, University of São Paulo, São Carlos, Brazil

Correspondence: Paulo Tarso S. Oliveira (paulotarsoms@gmail.com)

Received: 15 March 2025 | **Revised:** 27 July 2025 | **Accepted:** 17 August 2025

Funding: 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 406647/2022-4 and 446044/2023-7) and Fundação de Apoio ao Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul—Fundect (Grant 71/032.795/2022). P.T.S.O. acknowledges the CNPq (Grant 309048/2023-0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil—CAPES (Capes Print).

Keywords: catchment hydrology | climate change | hydrological modelling | water security

ABSTRACT

Climate change has significant impacts on hydrological fluxes worldwide, with pronounced effects in Brazil, including intense and recurrent droughts and floods. Accurate streamflow prediction is therefore essential for advancing water resources engineering, improving water resources management, and informing climate adaptation strategies. Here, we introduce the Streamflow Scenarios Projections for Brazilian Catchments (SSP-CABra), which provides long-term to daily streamflow simulations for 735 Brazilian catchments. These simulations are generated using five hydrological models of varying complexity, forced by 10 bias-corrected CMIP6-based climate simulations for historical (1980–2013) and future (2015–2100; SSP2-4.5 and SSP5-8.5) periods. SSP-CABra addresses a critical gap in large-scale hydrological modelling for Brazil, offering valuable insights for researchers and policymakers. Despite its broad applicability, the dataset includes models with varying performance across regions; users should assess model skill locally to ensure appropriate use, particularly for decision-making or extreme event analyses. Still, by leveraging multiple models and climate scenarios, SSP-CABra supports not only the mitigation of climate change impacts on water security, but also advances the understanding of model performance and regional hydrological behaviour.

Dataset Details:

Identifier: https://doi.org/10.5281/zenodo.14976730.

Creator: André Almagro, Paulo Tarso S. Oliveira, André S. Ballarin.

Dataset correspondence: andre.almagro@ufms.br.

SSP-CABra: Streamflow Scenario Projections for Brazilian Catchments.

Publisher: Zenodo.

Publication year: 2025.

(Resource type): Dataset.

(Version): Version 1.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Geoscience Data Journal* published by Royal Meteorological Society and John Wiley & Sons Ltd.

1 | Introduction

Climate change has emerged as a critical global concern, riving shifts in climate dynamics that intensify hydrological extremes such as droughts and floods, with far-reaching consequences for water resources management, agriculture practices, and ecosystems worldwide (Ficklin et al. 2022). These changes not only amplify the frequency and severity of extreme events but also compromise water availability, a challenge further exacerbated by rising water demand to support a continuously growing population (Destouni et al. 2013). In Brazil, projections indicate significant shifts in the frequency and intensity of extreme events, including droughts and floods (Ballarin, Wendland, et al. 2024; Ballarin, Vargas Godoy, et al. 2024; Mattos et al. 2025), which added to the ever-increasing water demand (ANA 2019), will likely increase risks to national water security (Ballarin, Sousa Mota Uchoa, et al. 2023). Furthermore, rising temperatures and more frequent heatwaves are expected to intensify water stress, mainly in the North, Northeast, and Central regions, where the most severe and prolonged heatwave events are projected by the end of the century (Ballarin, Oliveira, et al. 2024).

For instance, the unprecedented floods that occurred in the Rio Grande do Sul state, Brazil, between April and early May 2024 affected approximately 2.4 million people and resulted in 213 deaths or missing persons. This extreme event brought around 444 mm of accumulated precipitation in just 8 days, 652 mm after 35 days and up to 900 mm in some areas (Collischonn et al. 2024; Pillar and Overbeck 2024; Simoes-Sousa et al. 2025). Conversely, in 2023, the Amazon region experienced an extreme dry and warm condition (Espinoza et al. 2024). Given these trends, a deeper understanding of climate change impacts on national water resources is essential for informing sustainable water management practices and mitigating adverse effects.

To assess catchments potential responses to climate change, hydrological models are commonly used for streamflow projection based on climate models-based variables, such as precipitation, temperature, or evapotranspiration. Several hydrological models have been applied across different catchments at various spatial and temporal scales, ranging from long-term streamflow projections in large datasets of catchments (Ballarin, Sousa Mota Uchoa, et al. 2023) to daily-scale simulations in both large (Siqueira et al. 2021) and small (Sone et al. 2022) catchments. However, there is no consensus on how predicted streamflow values may vary depending on the hydrological model used or the spatial and temporal resolution considered. Furthermore, no existing dataset provides streamflow projections from multiple hydrological models under diverse climate change scenarios in Brazil, a country characterised by a wide range of hydrological responses across its varied climatic and geographic regions (Almagro et al. 2024). Such a dataset would be valuable for assessing the convergence and divergence of streamflow predictions across different hydrological models and could serve as an important tool for research across multiple disciplines, as well as for practical applications by stakeholders. In addition, future streamflow projections under different climate scenarios would enable the development of effective water management strategies seeking to mitigate the adverse water-related impacts of global warming on Brazil's water resources.

To fill this gap, here we propose the Streamflow Scenarios Projections for the Brazilian Catchments (SSP-CABra). This dataset aims to provide high-resolution streamflow predictions for 735 Brazilian catchments, leveraging the extensive data available in the Catchments Attributes for Brazil (CABra) dataset (Almagro, Oliveira, Meira Neto, et al. 2021). Our approach integrates five different hydrological models—ranging from simple functional forms to complex data-driven models—and multiple bias-corrected CMIP6-based outputs available in the Climate Change Dataset for Brazil (CLIMBra; Ballarin, Sone, et al. 2023) to simulate streamflow under a range of climate change scenarios. By incorporating a diverse range of models and scenarios, SSP-CABra generates robust predictions to support water resource management and policy decisions in Brazil.

2 | Data Description and Development

2.1 | Hydrological Models' Input Data

To simulate the trajectory of daily streamflow time series nationwide through the 21st century, we used the recently developed large-sample datasets on catchment attributes (CABra) and climate change projections (CLIMBra). CABra is a comprehensive dataset that compiles long-term hydrological and physiographic attributes for 735 Brazilian catchments, integrating data on climate, streamflow, groundwater, geology, soil, topography, land cover, and hydrologic disturbances. We used CABra's data to calibrate and evaluate hydrological models for these 735 catchments.

To predict streamflow across these catchments for the historical and future periods, we used bias-corrected projections from 10 CMIP6 (Coupled Model Intercomparison Project—Phase 6) climate models—for which the required inputs were available—forced by the Shared Socioeconomic Pathways SSP2-4.5 and SSP5-8.5 CMIP6-based scenarios. CLIMBra provides both raw and bias-corrected climate projections from up to 19 CMIP6 models for Brazil, offering high-resolution ($0.25° \times 0.25°$) daily time series of key meteorological variables, such as precipitation, temperature, and potential evapotranspiration, which were used as inputs to run the hydrological models present in the SSP-CABra dataset. CLIMBra products are also available for all CABra's catchments at a spatially averaged daily spatiotemporal resolution. By integrating these datasets within both data-driven and process-based frameworks, SSP-CABra enables the characterisation of spatiotemporal dynamics of streamflow under different CMIP6 climate scenarios (Figure 1).

2.2 | Hydrological Models

To generate streamflow simulations for both historical and future CMIP6-based scenarios, we used five hydrological models ranging from simple formulations to advanced techniques of deep learning. Specifically, here we use the following

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

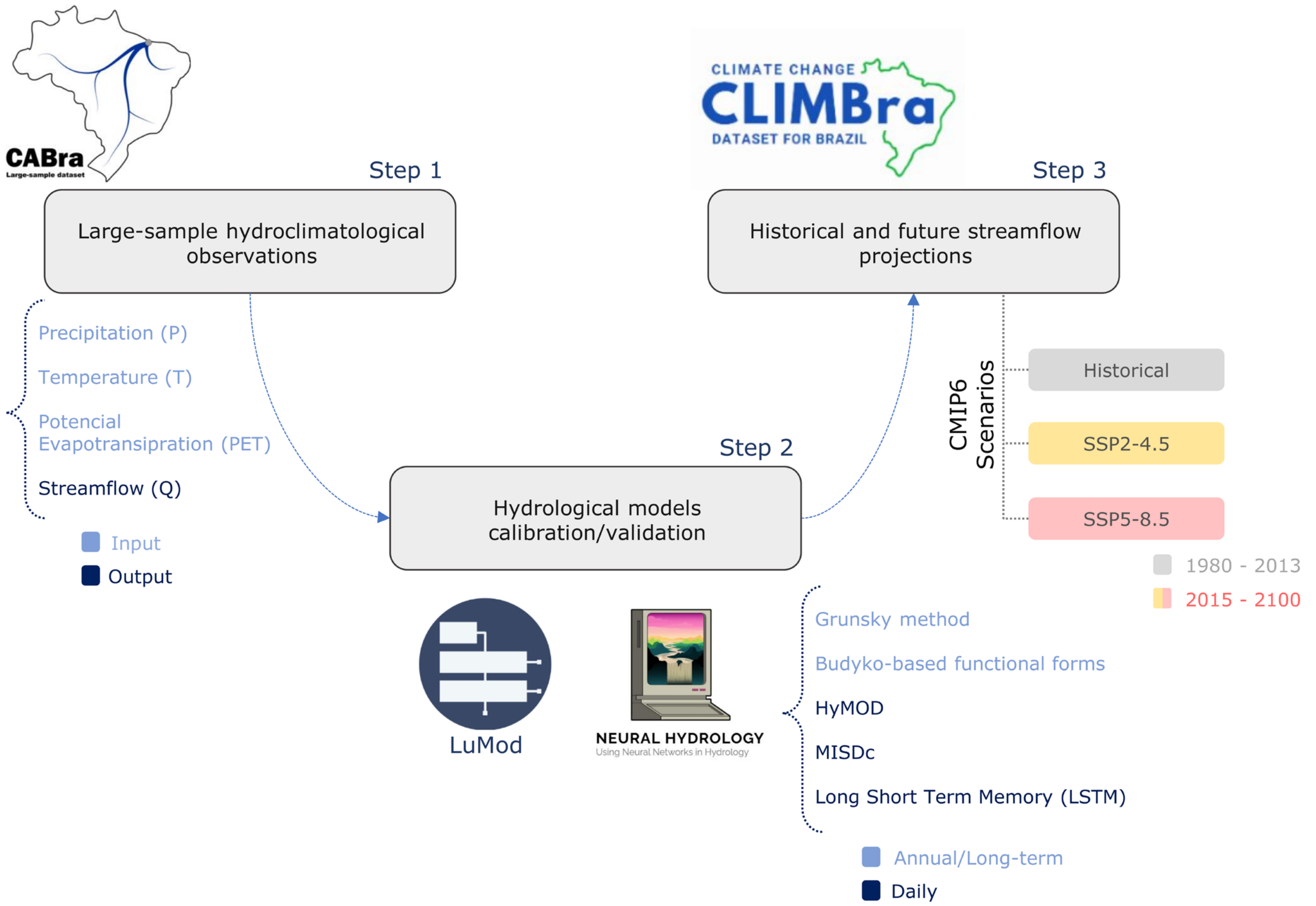

FIGURE 1 | Overview of the core steps for developing historical and future streamflow simulations for Brazilian catchments. (1) Downloading and pre-processing large-sample and climate change datasets, (2) calibration and validation of the hydrological models to simulate streamflow observations, and (3) projecting historical and future (*Shared Socioeconomic Pathways*—SSP2-4.5 and SSP5-8.5) streamflow based on the calibrated hydrological models forced by bias-corrected, CMIP6-based climate simulations.

models: Budyko-based functional forms (Budyko 1974; Meira Neto et al. 2020; Ballarin et al. 2022), Grunsky method (Marchezepe et al. 2023, 2025), HYdrological MODel (HyMOD) (Wagener et al. 2001), Modello Idrologico Semi-Distribuito in Continuo (MISDc) (Brocca et al. 2011) and Long Short-Term Memory (LSTM) neural network (Kratzert et al. 2018). A brief description of each hydrological model used in this study is included in the following subsections and summarised in Table 1.

It is important to note that the models operate in different temporal resolutions, leading to output generated on distinct timescales. Specifically, Budyko-based functional forms are provided on a long-term basis (historical and near, intermediate, and distant future). Outputs from the Grunsky method are available at an annual timescale. For the other three models—HyMOD, MISDc, and LSTM—simulations are produced at a daily timestep.

2.2.1 | Budyko-Based Functional Forms

Meira Neto et al. (2020) proposed Budyko-based functional forms which describes streamflow components as a function of ϕ (Equations 1 and 2). The mathematical reasoning of this formulation is based on the theoretical framework presented by L'vovich (1979) and Budyko (1974). The former describes the water balance as a two-stage process relating precipitation P with catchment wetting W, evaporation E and the streamflow

TABLE 1 | Overview of models' characteristics.

Model	Temporal resolution	Inputs	Calibration/ validation period
Budyko-based	Long-term[a]	P, PET, Q	[b]
Grunsky	Annual	P, T, Q	[b]
HyMOD	Daily	P, PET, Q	20/10 years
MISDc	Daily	P, PET, Q	20/10 years
LSTM	Daily	P, PET, Q, static attributes	20/10 years (?)

Abbreviations: P, precipitation (mm); PET, potential evapotranspiration (mm); Q, streamflow (mm).
[a]Approximately 30-year period, encompassing both historical (1980–2010) and near (2015–2040), intermediate (2040–2070), and distant (2070–2100) future periods.
[b]Budyko-based functional forms and the Grunsky method were modelled without following the traditional calibration/validation steps. More details can be found in Ballarin et al. (2022) and Marchezepe et al. (2025).

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

components direct runoff Q_D and baseflow Q_B. The latter assumes the partition of P into E and Q—which is the simple sum of the streamflow components Q_B and Q_D—as a function of ϕ, suggesting the balance between the atmosphere's water demand and supply as the main controlling mechanism of long-term water balance partitioning. With simple algebraic manipulations of these formulations, Meira Neto et al. (2020) proposed the functional forms. Here, we used the functional forms fitted to the CABra dataset as presented by Ballarin et al. (2022):

$$f_D(\phi) = \exp\left(-\phi^{1.10} + \ln[0.38]^{\frac{1}{1.23}}\right)^{1.23} \quad (1)$$

$$f_B(\phi) = \exp\left(-\phi^{1.20} + \ln[0.62]^{\frac{1}{0.74}}\right)^{0.74} \quad (2)$$

2.2.2 | Grunsky Method

The Grunsky method consists of a one-variable functional form that relates mean annual precipitation to streamflow (Grunsky and Manson 1908). Santos and Hawkins (2011) proposed a generalised approach using the mean annual precipitation ($\overline{P}$) and a coefficient α, which may be adapted according to the catchment's characteristics, to calculate interannual streamflow ($\overline{Q}$) (Equations 3 and 4).

$$\overline{Q} = \alpha \overline{P}^2, \quad \overline{P} < 1/(2\alpha) \quad (3)$$

$$\overline{Q} = \overline{P} - \frac{1}{4\alpha}, \quad \overline{P} > 1/(2\alpha) \quad (4)$$

where $\overline{Q}$ is the mean annual streamflow, $\overline{P}$ is the mean annual precipitation, and α is the Grunsky coefficient.

Following Santos and Hawkins (2011), Marchezepe et al. (2025) proposed a new formulation (Equation 5) to estimate the α coefficient as a function of mean annual temperature ($\overline{T}_A$, in °C) and mean annual precipitation ($\overline{P}_A$, in mm) based on catchments from the CABra dataset. This equation captures the variation of α across the Brazilian territory and can be used alongside Equations (3 and 4) to improve the generalised Grunsky method for estimating annual streamflow.

$$\alpha = -1.102\text{e}-05 \cdot \overline{T}_A + 4.851\text{e}-08 \cdot \overline{P}_A + 3.819e-04 \quad (5)$$

2.2.3 | HYdrological MODel

HyMOD is a lumped conceptual model that provides reliable simulation results in a computationally efficient manner as it requires minimal input flux data (precipitation and potential evapotranspiration), and has only six parameters to calibrate (Wagener et al. 2001). In summary, HyMOD consists of a nonlinear reservoir (soil moisture component) that generates excess precipitation, which is then routed through two parallel series of linear reservoirs representing fast and slow responses, ultimately propagating the flow to the basin outlet (Bastola and Misra 2014). The model's total outflow is the sum of these slow and fast components, while actual evapotranspiration is simulated as the minimum of potential evapotranspiration and soil moisture. Here, we used an improved version of the HyMOD proposed by Roy et al. (2017), which is available in the LuMod Python package (documentation available at: https://pypi.org/project/lumod/, accessed on July 7, 2025).

2.2.4 | Modello Idrologico Semi-Distribuito in Continuo

MISDc is a semi-distributed hydrological model designed for streamflow simulation (Brocca et al. 2011). The model incorporates multiple sub-catchments to account for spatial variability across the watershed, allowing each sub-catchment to have distinct characteristics (e.g., land use, slope, soil type). MISDc is a parsimonious and reliable rainfall-runoff model that estimates total streamflow while considering soil wetness conditions prior to a rainfall event. In this paper, we used a one-layer, lumped version of the MISDc, available in the python package LuMod. Similar to HyMOD, the MISDc model was calibrated for each catchment.

2.2.5 | Long Short-Term Memory

The LSTM is a type of recurrent neural network (RNN) that is capable to capture long-term dependencies via a cell state and multiple gating mechanisms that control the information at each time step (Kratzert et al. 2018). Here we employed the NeuralHydrology package (Kratzert et al. 2022), which implements a deep learning-based hydrological model leveraging the capabilities of LSTM networks, to simulate and predict hydrological processes such as streamflow, precipitation, and other water-related variables. This model is part of the growing field of neuro hydrology, where deep learning methods are applied to hydrological and water resources modelling.

In addition to demonstrating that deep learning surpasses conventional hydrological modelling, recent studies have shown that the traditional approach of training an individual model for each catchment does not yield the best streamflow estimates. Kratzert et al. (2024) and Nearing et al. (2021) demonstrated that LSTMs trained with data from multiple watersheds outperformed both traditional hydrological models and locally trained LSTMs. This highlights the potential of developing a single model using data from various watersheds (a regional approach) capable of capturing universal hydrological patterns and transferring them to individual basins, resulting in an improved performance (Kratzert et al. 2018, 2019). Therefore, unlike the HyMOD and MISDc models, which are calibrated separately for each catchment, the LSTM model was trained and tested using data from all catchments simultaneously.

2.3 | Hydrological Model Calibration and Evaluation

2.3.1 | HyMOD and MISDc—A Local Approach

The calibration of the conventional hydrological models (HyMOD and MISDc) was performed using daily time series of precipitation, potential evapotranspiration, and streamflow from October 1, 1980 (start of the 1981 water year) to September 30, 2000 (end of the 2000 water year). The validation period

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

extended from October 1, 2000 (start of the 2001 water year) to September 30, 2010 (end of the 2010 water year). Model parameters—six for HyMOD and nine for MISDc—were optimised through a Monte Carlo simulation routine, addressing both model calibration and uncertainty analysis. A total of 800 simulations were performed, using the KGE as the objective function, and the best-performing set of parameters was selected as the calibrated result. It is important to note that a local calibration procedure was applied for each of the 735 catchments, resulting in a set of calibrated parameters for each of those catchments.

2.3.2 | LSTM—A Regional Approach

Considering the deep learning model, we trained a single LSTM network for all catchments using a regional approach, leveraging its ability to capture hydrological behaviour by learning universal patterns from diverse training data (Nearing et al. 2021; Kratzert et al. 2024). The process involved training, validation, and testing phases, with meteorological features (e.g., precipitation and potential evapotranspiration) as dynamic inputs and catchment attributes (e.g., aridity index, area, slope, permeability, soil depth, sand fraction, and cover fraction of forest, grass and crops) as static inputs. The training/validation period (1981–2005) covered 80% of the data, while testing (2005–2010) comprised the remaining 20%.

Hyperparameters were optimised to balance computational efficiency and accuracy, using a hidden size of 256 cells, a dropout rate of 0.4, a batch size of 32, a learning rate of 0.01, and MSE as the loss function—consistent with previous LSTM hydrological studies (Kratzert et al. 2018; Klotz et al. 2022). The loss function was based on the Nash-Sutcliffe Efficiency—NSE. Unlike conventional hydrological models, only a single deep learning network was trained for the entire set of catchments, resulting in a regional LSTM model capable of simulating daily streamflow across Brazilian catchments.

2.4 | Hydrological Models Performance

Figure 2 presents a spatial evaluation of hydrological model performance across Brazilian catchments, in terms of the Kling-Gupta Efficiency (KGE) metric for both calibration/train and validation/test periods. Regarding the conventional hydrologic modelling, using a local calibration approach, the MISDc performed better than the HyMOD, with higher median KGE (mKGE = 0.73 and 0.61, respectively). In turn, the regional approach using the data-driven LSTM model presented mKGE = 0.72. As previously seen in Almagro, Oliveira, and Brocca (2021) model performance varies significantly across the regions. Catchments in the southern and southeastern regions generally exhibit higher KGE values. The good performance of the LSTM model in those catchments during training can be related to the relatively denser hydro-climatic monitoring network and more stable hydrological regimes in these areas, which provide more reliable data for model training. Conversely, northeastern catchments show lower KGE values, reflecting the challenges posed by data scarcity, more complex rainfall-runoff processes—with strong reservoir regulation—and higher climate variability.

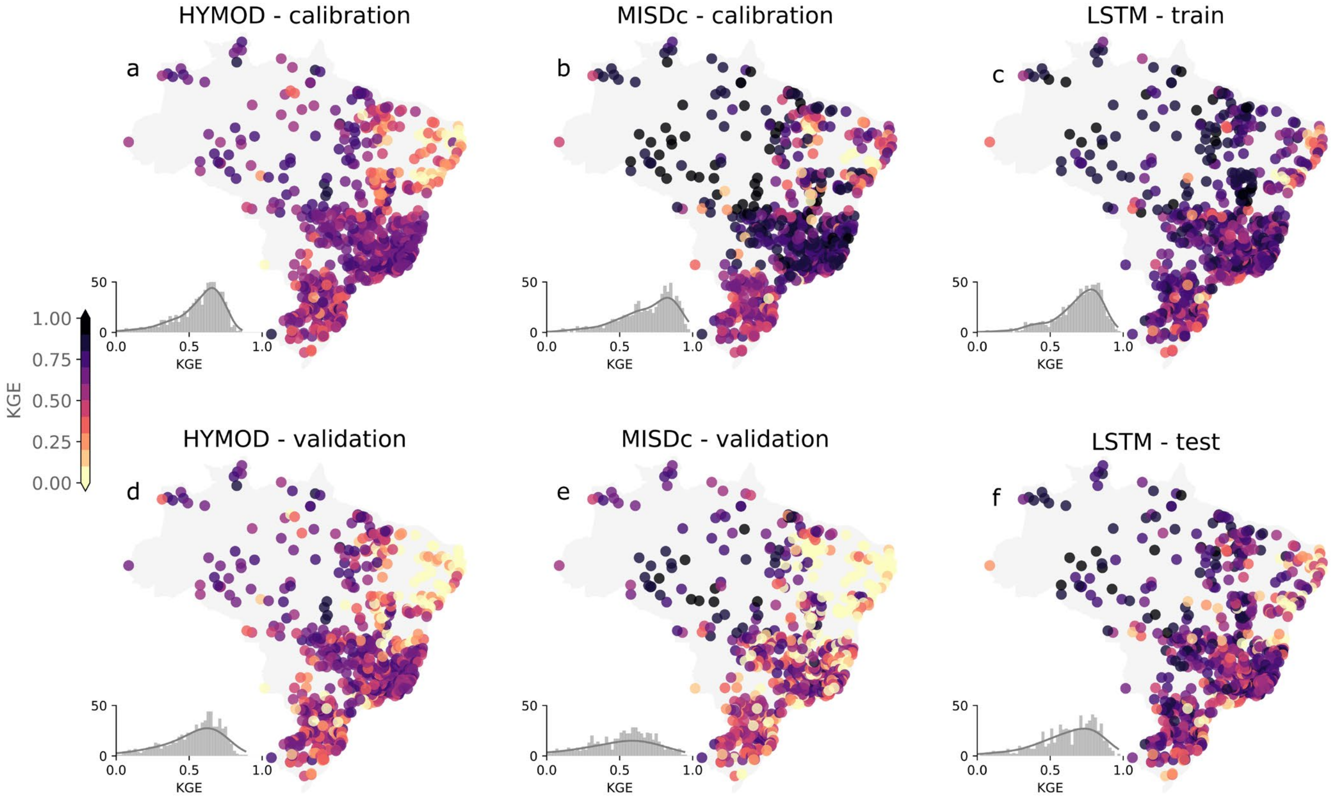

FIGURE 2 | Performance evaluation of the hydrological models in calibration/training and validation/testing periods. (a) HyMOD in calibration, (b) MISDc in calibration, (c) LSTM in training, (d) HyMOD in validation, (e) MISDc in validation, and (f) LSTM in test. The evaluation is based on the Kling-Gupta Efficiency on simulating the daily streamflow, ranging from 0 (poor performance) to 1 (perfect match with observed data).

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The validation/test evaluation, however, reveals a decline in model performance across most regions, as expected. While some southern catchments maintain relatively high KGE values, there is a noticeable decrease in the northeast and parts of the central west. During the validation period, HyMOD (mKGE = 0.56) presented a better performance in simulating the daily streamflow of Brazilian catchments than the MISDc model (mKGE = 0.49). This drop suggests a mismatch between the climate conditions of the calibration and validation periods, limitations of the models in capturing complex hydrological processes, and extreme events, which are more prevalent in these regions due to rainfall variability. Although we can also note a decrease in the performance of the LSTM model, the regional approach proved to be the most accurate one in simulating the daily streamflow in Brazilian catchments. There was a slight drop in efficiency during the test period (mKGE = 0.66) but still consistent with the values obtained during the training phase, highlighting the ability of deep learning models in representing the hydrological processes across a wide and diverse set of catchments. Besides, although relatively low, LSTM showed better values of mKGE for the northeastern catchments, especially in the Sao Francisco River basin, where the streamflow regulation by reservoirs plays an important role.

Nevertheless, it is important to highlight that all models exhibited limitations in certain regions. For instance, during the validation period, model performance was considered weaker than the mean-flow benchmark (KGE < −0.41, see Knoben et al. 2019) for 18 catchments with HyMOD, 83 with MISDc, and 10 with LSTM, indicating substantial challenges in accurately simulating even mean daily streamflow dynamics in some Brazilian basins. These numbers significantly rise when considering only positive KGE values as a benchmark: 50, 142, and 24 for HyMOD, MISDc, and LSTM, respectively. Details about models' performance can be found in the dataset repository. These outliers could point to specific hydrological complexities, such as intermittent streamflow, strong anthropogenic influences, or poorly represented physical processes within the models.

To evaluate the performance of the calibrated models in simulating streamflow using climate simulations, we compared long-term streamflow observations with the calibrated-models simulations forced by the bias-corrected historical CLIMBra simulations (Figure 3). This assessment is a crucial step in climate change-related analyses, as hydrological models must be capable of reproducing observed streamflow characteristics when driven by historical climate simulations—rather than historical observations—to ensure that this combination can be used to assess future scenarios. In general, all evaluated models exhibited a good performance (mean absolute error [MAE] < 1). The Grunsky method and Budyko-based functional forms exhibited the better performance (MAE = 0.32 and 0.38, respectively), which can be attributed to two factors. First, those arithmetic equations were specifically calibrated

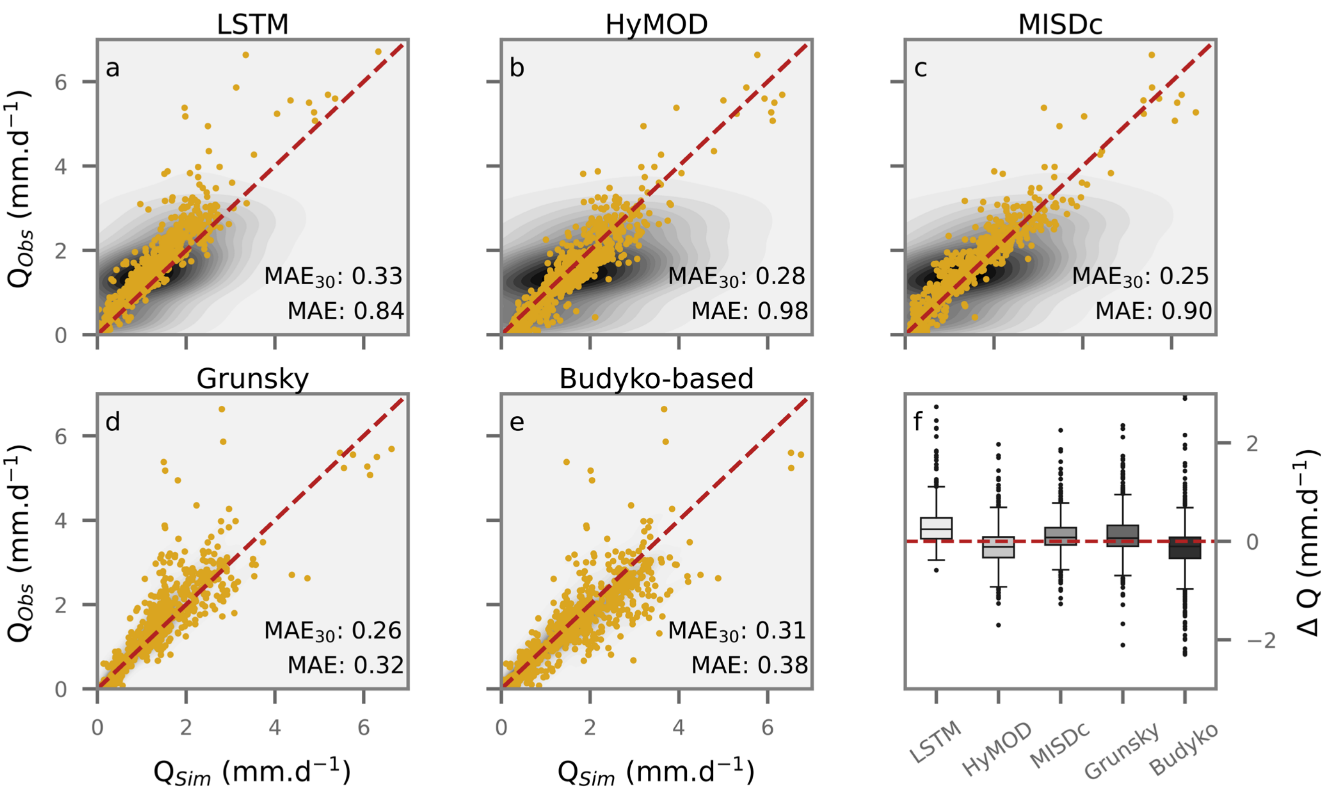

FIGURE 3 | Performance of the hydrological models forced by bias-corrected CMIP6 simulations to describe historical streamflow observations. (a) LSTM, (b) HyMOD, (c) MISDc, (d) Grunsky method, (e) Budyko-based functional forms. A bivariate density plot represents the observation-simulation pairs considering the 10 CMIP6 models separately. Dark regions indicate high density. Dark yellow points indicate the pairs obtained with the CMIP6 multi-model ensemble mean simulations. 1:1 line is displayed as a red dashed line. Mean absolute error for the multi-model ensemble is displayed in the bottom right corner. (f) Box-plot of the differences between simulated (multi-model ensemble mean) and observed streamflow for all evaluated hydrological models.

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to reproduce long-term (or annual) mean streamflow, whereas the other evaluated conceptual or data-driven methods were primarily designed to represent water components in a finer temporal scale. Second, daily-scale models exhibited a poor performance for certain catchments, which amplified these discrepancies. For instance, when we remove, for each hydrologic model, the 30 catchments with the largest discrepancies between observed and simulated long-term streamflow and recalculate the mean absolute error (MAE_{30}), we find that all models exhibit similar performance ($MAE \cong 0.3$). This confirms that the higher catchments were models poorly performed were the main responsible for the lower MAE. Finally, among the daily-scale models—LSTM, HyMOD, and MISDc—LSTM exhibited the better performance when forced by different CMIP6 models, as indicated by the density plot with the smallest spread, followed by MISDc and HyMOD, respectively.

2.5 | Projected Changes in Streamflow Between Historical (1980–2013) and Future (2070–2100) Periods

To illustrate the application of the SSP-CABra, we analysed projected changes in streamflow dynamics between historical (1980–2013) and future (2070–2100) periods under the SSP5-8.5 scenario (Figures 4 and 5). First, we computed the relative differences in long-term mean streamflow between these periods (Figure 4). In general, a consistent trend of increasing streamflow is observed in the South and Northeast regions, a decrease in the North region, and greater uncertainty in the projections for the Southeast and Central-West regions. Such spatial patterns align with expected changes projected for rainfall and long-term water availability across the country (Ballarin, Sousa Mota Uchoa, et al. 2023).

These highly variable between-catchment pattern of changes highlight the influence of hydrological models on both the magnitude and direction of future streamflow. For example, the percentage of catchments with increasing trends across Brazil ranged from 50.5% to 85.6%, depending on the hydrological model. As expected, the discrepancy is higher in regions with overall lower model performance. Similarly, the choice of climate models also affected the results, with the proportion of catchments showing high agreement (> 70%) among CMIP6 models varying between 46% and 66.1%, suggesting that hydrological models may produce conflicting projections depending on the bias-corrected CMIP6 dataset used as input.

As an additional example, we present the streamflow climatology (long-term daily mean) for one catchment of each hydrological group proposed by Almagro et al. (2024), comparing historical and future LSTM-based projections (Figure 5). Catchments were randomly selected for each hydrological group (A1–A6), but only those with KGE > 0.8 were considered. The hydrological groups were defined based on similarities in hydrological signatures and catchment attributes using a k-means clustering approach, optimised by the Elbow method and supported by principal component analysis. These six groups mainly reflect an aridity gradient and differences in streamflow seasonality. For additional

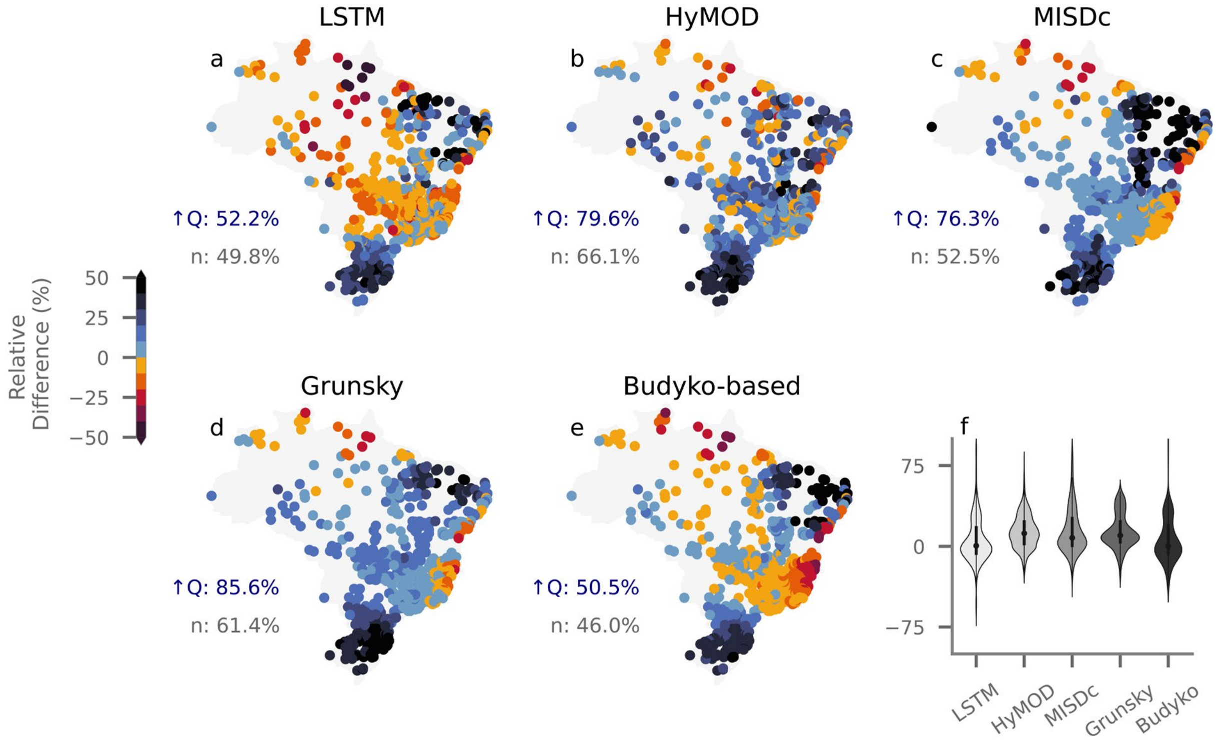

FIGURE 4 | Projected changes between historical and distant future long-term mean streamflow. (a) LSTM, (b) HyMOD, (c) MISDc, (d) Grunsky method, (e) Budyko-based functional forms. A dark blue text in the bottom left corner of each panel indicates the percentage of catchments with positive projected changes (multi-model ensemble mean), whereas the grey text indicates the percentage of catchments with at least 70% of between-CMIP6-model agreement in the sign of change. (f) Violin-plot representing the distribution of the projected changes displayed in panels (a–e) for each hydrological model.

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

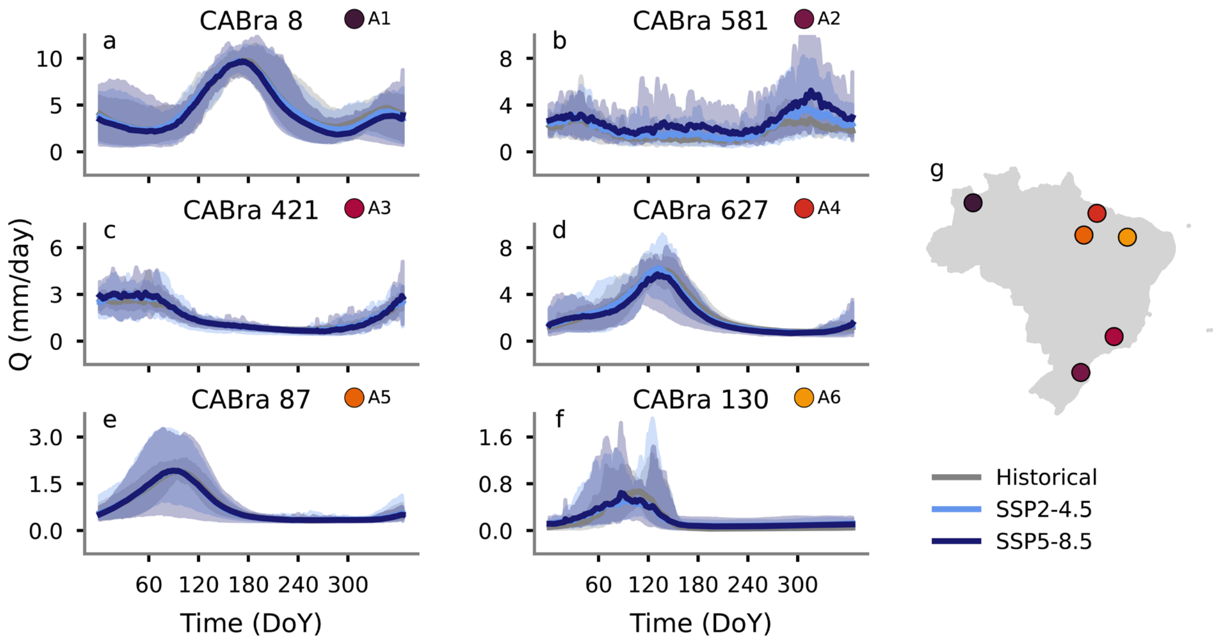

FIGURE 5 | Projected changes between historical and distant future streamflow daily climatology for catchments of six different hydrological groups. (a) CABra 8, (b) CABra 581, (c) CABra 421, (d) CABra 627, (e) CABra 87, and (f) CABra 130. A grey line represents daily (day of year—DoY) streamflow climatology simulated for the historical period by the CMIP6 multimodel ensemble mean, whereas a light blue and dark blue line indicates the simulations for the distant future period forced by the SSP2-4.5 and SSP5-8.5 scenarios, respectively. Shaded area represents the CMIP6 simulations' range (minimum and maximum). (g) Spatial distribution of each evaluated catchment within the Brazilian territory. The hydrological group for each catchment is displayed in the top right corner of each panel.

details on the hydrological characteristics and dynamics of each group, readers are referred to Almagro et al. (2024).

The results highlight the country's highly heterogeneous hydrological regimes, demonstrating the dataset's relevance not only for water resource management practices on a nationwide scale, but also for specific, regional-based studies aiming to incorporate local, specific information or global studies seeking to enhance the comprehension of hydrological processes. For instance, while some catchments exhibit pronounced seasonality, with daily streamflow exceeding 10 mm in certain months (Figure 5a), others show a more stable hydrological pattern (Figure 5c), exhibiting even intermittent regimes (Figure 5f). Such differences emphasise the valuable contribution that SSP-CABra holds globally for hydrological studies in a climate change context.

3 | Dataset Access

3.1 | General Description and Dataset Location

The SSP-CABra dataset provides streamflow simulations for 735 Brazilian catchments present in the CABra dataset. Simulations are based on a set of hydrological models forced with bias-corrected meteorological CMIP6-based outputs from 10 different models retrieved from the CLIMBra dataset. Simulations are available for both historical (1980–2013) and future (2015–2100; SSP2-4.5 and SSP5-8.5) periods at different time scales. For Budyko-based functional forms and Grunsky method derived simulations, streamflow simulations are available for long-term and intra-annual timescales, respectively. For the other hydrological models—HyMOD, MISDc, and LSTM—simulations are available at a daily scale. Streamflow time series are provided as text file format (.txt) and are freely available at Zenodo (https://doi.org/10.5281/zenodo.14976731).

3.2 | Dataset Format and Accessibility

Daily streamflow simulations are provided in m^3/s for LSTM, HyMOD and MISDc models; total annual streamflow are in mm/year for the Grunsky method; and the long-term mean daily streamflow for the Budyko-based functional forms are in mm/day. Data is available for each catchment separately and is organised in folders representing the CMIP6-climate models available in the CLIMBra dataset, the scenario (*hist* for the historical period, and *ssp245* or *ssp585* for the scenarios SSP2-4.5 and SSP5-8.5, respectively), and the hydrological model (FUNCTIONAL FORMS, GRUNSKY, LSTM, HYMOD, and MISDC), respectively. For instance, future streamflow simulations using the LSTM method forced by the SSP2.4–5 simulations from the EC-EARTH3 model are stored in the following path: EC-EARTH3/ssp245/LSTM. Inside this folder, one will find 735 files named 'CABra_ID_streamflow_sim.txt', where CABra_ID refers to the identification of each catchment within the CABra dataset.

3.3 | Dataset Limitations

Streamflow time series were generated by forcing different hydrological models and functional forms with bias-corrected CMIP6 meteorological time series from the CLIMBra dataset. CLIMBra

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

employed the delta quantile mapping (Cannon et al. 2015) and a observational-based dataset (Xavier et al. 2016) to reduce systematic deviations between observed and simulated historical data while keeping changes projected by climate models. Nevertheless, it is worth noting that CLIMBra's bias-corrected outputs is able to represent statistical properties of observations, but not necessarily the timing of observed events. Hence, historical streamflow simulations from the SSP-CABra should not be used to assess historical recorded events, such as droughts and floods.

Despite exhibiting a better performance than raw climate simulations in representing observed climate, bias-corrected simulations can still present some deficiencies, such as physically unrealistic values or systematic errors derived from the different climate models used in this study—especially regarding extreme events—that are transferred to the streamflow simulations (White and Toumi 2013; Switanek et al. 2017; Casanueva et al. 2020; Abdelmoaty et al. 2021). Hence, for some specific applications, we recommend that additional performance analysis should be conducted.

Finally, we emphasise that not all models performed well in simulating daily and long-term streamflow across Brazilian catchments. Some exhibited weak performance in a considerable number of catchments during the validation period. These patterns underscore the importance of considering model performance when interpreting projected changes, as lower calibration and validation metrics may indicate higher uncertainty and reduce the reliability of the outputs for certain hydrological applications. We recommend that additional tests be conducted—particularly for studies focused on extreme events—and that performance metrics be carefully evaluated prior to any application of the dataset, to assess whether a given model is suitable for a specific catchment. Acknowledging these uncertainties is essential for the appropriate use of the projections. Future work should further explore model limitations and potential improvements, especially in data-scarce or hydrologically complex regions.

4 | Conclusions

In this paper, we introduce the Streamflow Scenarios Projections for Brazilian Catchments (SSP-CABra), a novel dataset that provides daily streamflow simulations for 735 Brazilian catchments included in the CABra dataset. These streamflow simulations are generated using five hydrological models with varying levels of complexity, forced by 10 bias-corrected CMIP6-based climate simulations obtained from the CLIMBra dataset. The dataset encompasses both historical (1980–2013) and future (2015–2100) periods, covering two climate change scenarios: SSP2-4.5 and SSP5-8.5. Streamflow simulations in SSP-CABra are available at different temporal scales depending on the hydrological model used. For Budyko-based functional forms, streamflow estimates are available as long-term mean annual values, while the Grunsky method provides simulations at intra-annual timescales. HyMOD, MISDc, and LSTM yield simulations at a daily time step, offering a high-resolution perspective on hydrological variability. SSP-CABra addresses a critical gap in large-scale hydrological modelling for Brazil, providing important insights for researchers and policymakers. By integrating multiple hydrological models and climate scenarios, this dataset improves the understanding of hydrological processes across Brazil's diverse catchments, supporting informed decision-making and strategies to mitigate the impacts of climate change on water security.

Acknowledgements

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 406647/2022-4 and 446044/2023-7) and Fundação de Apoio ao Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul—Fundect (Grant 71/032.795/2022). P.T.S.O. acknowledges the CNPq (Grant 309048/2023-0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil—CAPES (Capes Print).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at https://zenodo.org/records/14976731, reference number Version 1. *Code availability*: Daily hydrological simulations were performed using the LuMod and the Neural hydrology Python packages. LuMod—detailed in https://zaul_ae.gitlab.io/lumod-docs/—encompasses a set of widely used lumped hydrological models, including HyMOD and a one-layer version of the MISDc. Neural Hydrology—described in https://neuralhydrology.github.io/—is a python package for deep learning research designed for hydrological applications (Kratzert et al. 2022). Streamflow estimations based on the Budyko-based functional forms and the Grunsky method were also conducted on Python using the Equations (1–5). For a detailed description of these methods, readers are referred to Ballarin et al. (2022) and Marchezepe et al. (2025).

References

Abdelmoaty, H. M., S. M. Papalexiou, C. R. Rajulapati, and A. AghaKouchak. 2021. "Biases Beyond the Mean in CMIP6 Extreme Precipitation: A Global Investigation." *Earth's Future* 9, no. 10: 1–17. https://doi.org/10.1029/2021EF002196.

Almagro, A., A. A. Meira Neto, N. Vergopolan, T. Roy, P. A. Troch, and P. T. S. Oliveira. 2024. "The Drivers of Hydrologic Behavior in Brazil: Insights From a Catchment Classification." *Water Resources Research* 60, no. 8: e2024WR037212. https://doi.org/10.1029/2024WR037212.

Almagro, A., P. T. Oliveira, A. A. Meira Neto, T. Roy, and P. Troch. 2021. "CABra: A Novel Large-Sample Dataset for Brazilian Catchments." *Hydrology and Earth System Sciences*: 1–40. https://doi.org/10.5194/hess-2020-521.

Almagro, A., P. T. S. Oliveira, and L. Brocca. 2021. "Assessment of Bottom-Up Satellite Rainfall Products on Estimating River Discharge and Hydrologic Signatures in Brazilian Catchments." *Journal of Hydrology* 603: 126897. https://doi.org/10.1016/j.jhydrol.2021.126897.

ANA. 2019. *Manual of Consumptive Water Use in Brazil*. Agência Nacional de Águas—Brasília.

Ballarin, A. S., P. T. S. Oliveira, B. K. Marchezepe, et al. 2022. "The Impact of an Open Water Balance Assumption on the Understanding of the Factors Controlling the Long-Term Streamflow Components." *Water Resources Research* 58: e2022WR032413. https://doi.org/10.1029/2022WR032413.

Ballarin, A. S., P. T. S. Oliveira, J. G. S. M. Uchôa, et al. 2024. "Escalating Threat of Human-Perceived Heatwaves in Brazil." *Environmental*

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Research Communications 6, no. 11: 111004. https://doi.org/10.1088/2515-7620/ad9140.

Ballarin, A. S., J. S. Sone, G. C. Gesualdo, et al. 2023. "CLIMBra—Climate Change Dataset for Brazil." *Scientific Data*: 1–31. https://doi.org/10.1038/s41597-023-01956-z.

Ballarin, A. S., J. G. Sousa Mota Uchoa, M. S. dos Santos, et al. 2023. "Brazilian Water Security Threatened by Climate Change and Human Behavior." *Water Resources Research* 59: e2023WR034914. https://doi.org/10.1029/2023WR034914.

Ballarin, A. S., M. R. Vargas Godoy, M. Zaerpour, et al. 2024. "Drought Intensification in Brazilian Catchments: Implications for Water and Land Management." *Environmental Research Letters* 19, no. 5: 054030. https://doi.org/10.1088/1748-9326/ad3e18.

Ballarin, A. S., E. Wendland, M. Zaerpour, S. Hatami, A. A. Meira Neto, and S. M. Papalexiou. 2024. "Frequency Rather Than Intensity Drives Projected Changes of Rainfall Events in Brazil." *Earth's Future* 12, no. 1: e2023EF004053. https://doi.org/10.1029/2023EF004053.

Bastola, S., and V. Misra. 2014. "Evaluation of Dynamically Downscaled Reanalysis Precipitation Data for Hydrological Application." *Hydrological Processes* 28, no. 4: 1989–2002. https://doi.org/10.1002/hyp.9734.

Brocca, L., F. Melone, and T. Moramarco. 2011. "Distributed Rainfall-Runoff Modelling for Flood Frequency Estimation and Flood Forecasting." *Hydrological Processes* 25, no. 18: 2801–2813. https://doi.org/10.1002/hyp.8042.

Budyko, M. I. 1974. *Climate and Life*. Academic Press.

Cannon, A. J., S. R. Sobie, and T. Q. Murdock. 2015. "Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?" *Journal of Climate* 28, no. 17: 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1.

Casanueva, A., S. Herrera, M. Iturbide, et al. 2020. "Testing Bias Adjustment Methods for Regional Climate Change Applications Under Observational Uncertainty and Resolution Mismatch." *Atmospheric Science Letters* 21, no. 7: 1–12. https://doi.org/10.1002/asl.978.

Collischonn, W., J. P. L. F. Brêda, S. Wongchuig, et al. 2024. *Unprecedented April–May 2024 Rainfall in South Brazil Sets New Record*. Vol. 29. RBRH. Associação Brasileira de Recursos Hídricos. https://doi.org/10.1590/2318-0331.292420240088.

Destouni, G., F. Jaramillo, and C. Prieto. 2013. "Hydroclimatic Shifts Driven by Human Water Use for Food and Energy Production." *Nature Climate Change* 3, no. 3: 213–217. https://doi.org/10.1038/nclimate1719.

Espinoza, J.-C., J. C. Jimenez, J. A. Marengo, et al. 2024. "The New Record of Drought and Warmth in the Amazon in 2023 Related to Regional and Global Climatic Features." *Scientific Reports* 14, no. 1: 8107. https://doi.org/10.1038/s41598-024-58782-5.

Ficklin, D. L., S. E. Null, J. T. Abatzoglou, K. A. Novick, and D. T. Myers. 2022. "Hydrological Intensification Will Increase the Complexity of Water Resource Management." *Earth's Future* 10, no. 3: e2021EF002487. https://doi.org/10.1029/2021EF002487.

Grunsky, C. E., and M. C. E. Manson. 1908. *Reports on the Water Supply of San Francisco, California, 1900 to 1908, inclusive*. Transactions of the American Society of Civil Engineers. http://hdl.handle.net/2027/uc1.$b78665.

Klotz, D., F. Kratzert, M. Gauch, et al. 2022. "Uncertainty Estimation With Deep Learning for Rainfall–Runoff Modeling." *Hydrology and Earth System Sciences* 26, no. 6: 1673–1693. https://doi.org/10.5194/hess-26-1673-2022.

Knoben, W. J. M., J. E. Freer, and R. A. Woods. 2019. "Technical Note: Inherent Benchmark or Not? Comparing Nash–Sutcliffe and Kling–Gupta Efficiency Scores." *Hydrology and Earth System Sciences* 23, no. 10: 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Kratzert, F., M. Gauch, D. Klotz, and G. Nearing. 2024. "HESS Opinions: Never Train a Long Short-Term Memory (LSTM) Network on a Single Basin." *Hydrology and Earth System Sciences* 28, no. 17: 4187–4201. https://doi.org/10.5194/hess-28-4187-2024.

Kratzert, F., M. Gauch, G. Nearing, and D. Klotz. 2022. "NeuralHydrology—A Python Library for Deep Learning Research in Hydrology." *Journal of Open Source Software* 7, no. 71: 4050. https://doi.org/10.21105/joss.04050.

Kratzert, F., D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger. 2018. "Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences* 22, no. 11: 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Kratzert, F., D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing. 2019. "Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets." *Hydrology and Earth System Sciences* 23, no. 12: 5089–5110. https://doi.org/10.5194/hess-23-5089-2019.

L'vovich, M. I. 1979. *World Water Resources and Their Future*. American Geophysical Union.

Marchezepe, B. K., A. Almagro, A. S. Ballarin, and P. T. S. Oliveira. 2023. "Streamflow Prediction in Ungauged Catchments by Using the Grunsky Method." *International Soil and Water Conservation Research* 11, no. 4: 700–712. https://doi.org/10.1016/j.iswcr.2023.06.004.

Marchezepe, B. K., A. Almagro, A. S. Ballarin, and P. T. S. Oliveira. 2025. "Improved Grunsky Method for Streamflow Prediction in Ungauged Catchments." *Hydrological Sciences Journal* 70: 535–545. https://doi.org/10.1080/02626667.2024.2446278.

Mattos, T. S., A. S. Ballarin, M. C. Lucas, A. Almagro, and P. T. S. Oliveira. 2025. "Uneven Shifts in Flood and Drought Flows in a Brazilian Water Supply Catchment." *Journal of Hydrology: Regional Studies* 60: 102498. https://doi.org/10.1016/j.ejrh.2025.102498.

Meira Neto, A. A., T. Roy, P. T. S. de Oliveira, and P. A. Troch. 2020. "An Aridity Index-Based Formulation of Streamflow Components." *Water Resources Research* 56, no. 9: 1–14. https://doi.org/10.1029/2020WR027123.

Nearing, G. S., F. Kratzert, A. K. Sampson, et al. 2021. "What Role Does Hydrological Science Play in the Age of Machine Learning?" *Water Resources Research* 57, no. 3: e2020WR028091. https://doi.org/10.1029/2020WR028091.

Pillar, V. D., and G. E. Overbeck. 2024. "Learning From a Climate Disaster: The Catastrophic Floods in Southern Brazil." *Science* 385, no. 6713: eadr8356. https://doi.org/10.1126/science.adr8356.

Roy, T., H. V. Gupta, A. Serrat-Capdevila, and J. B. Valdes. 2017. "Using Satellite-Based Evapotranspiration Estimates to Improve the Structure of a Simple Conceptual Rainfall–Runoff Model." *Hydrology and Earth System Sciences* 21, no. 2: 879–896. https://doi.org/10.5194/hess-21-879-2017.

Santos, F. L., and R. H. Hawkins. 2011. "Generalized Mediterranean Annual Water Yield Model: Grunsky's Equation and Long-Term Average Temperature." *Journal of Hydrologic Engineering* 16, no. 11: 874–879. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000402.

Simoes-Sousa, I. T., C. M. L. Camargo, J. Tavora, A. Piffer-Braga, J. T. Farrar, and T. M. Pavelsky. 2025. "The May 2024 Flood Disaster in Southern Brazil: Causes, Impacts, and SWOT-Based Volume Estimation." *Geophysical Research Letters* 52, no. 4: e2024GL112442. https://doi.org/10.1029/2024gl112442.

Siqueira, P. P., P. T. S. Oliveira, D. Bressiani, A. A. Meira Neto, and D. B. B. Rodrigues. 2021. "Effects of Climate and Land Cover Changes on Water Availability in a Brazilian Cerrado Basin." *Journal of Hydrology: Regional Studies* 37: 100931. https://doi.org/10.1016/j.ejrh.2021.100931.

Sone, J. S., T. F. Araujo, G. C. Gesualdo, et al. 2022. "Water Security in an Uncertain Future: Contrasting Realities From an Availability-Demand

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Perspective." *Water Resources Management* 36: 2571–2587. https://doi.org/10.1007/s11269-022-03160-x.

Switanek, B. M., A. P. Troch, L. C. Castro, et al. 2017. "Scaled Distribution Mapping: A Bias Correction Method That Preserves Raw Climate Model Projected Changes." *Hydrology and Earth System Sciences* 21, no. 6: 2649–2666. https://doi.org/10.5194/hess-21-2649-2017.

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian. 2001. "A Framework for Development and Application of Hydrological Models." *Hydrology and Earth System Sciences* 5, no. 1: 13–26. https://doi.org/10.5194/hess-5-13-2001.

White, R. H., and R. Toumi. 2013. "The Limitations of Bias Correcting Regional Climate Model Inputs." *Geophysical Research Letters* 40, no. 12: 2907–2912. https://doi.org/10.1002/grl.50612.

Xavier, A. C., C. W. King, and B. R. Scanlon. 2016. "Daily Gridded Meteorological Variables in Brazil (1980–2013)." *International Journal of Climatology* 36, no. 6: 2644–2659. https://doi.org/10.1002/joc.4518.

20496060, 2025, 4, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/gdj3.70029 by University Of Sao Paulo - Brazil, Wiley Online Library on [29/09/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License