Integrating Hidden Markov and Multinomial Models for Hydrological Drought Prediction under nonstationarity.

Marcus Suassuna Santos *,[1]

Louise J. Slater [2]

[1] Department of Hydrology, Geological Survey of Brazil

[2] School of Geography and the Environment, University of Oxford, Oxford, UK

***Corresponding Author**: SBN Quadra 02, Bl. H, Central Brasília, Brasília - DF - Brazil - CEP: 70040-904; marcus.santos@cprm.gov.br

Abstract:

Understanding the drivers of drought variability is crucial for developing effective adaptation and management strategies. This study develops a two-step modelling approach to characterize and predict hydrological droughts in a nonstationary context. First, a multivariate Hidden Markov Model (HMM) is used to classify low-water level time series into Dry, Normal, and Wet years, identifying Dry years as hydrological droughts. Second, a Multinomial Logistic Regression model (MLR) is proposed to predict low-water level class transitions, incorporating external variables into the transition matrix estimates. Precipitation thresholds for annual minima are also derived, with uncertainties and sensitivities assessed via bootstrap resampling. Our framework was successfully applied to the Paraguay River basin (PRB), where long-term changes in hydrological variables are frequent. The HMM transition matrix reveals a long persistence of years in each water level class and an inhomogeneity between two periods (1901-1960 and 1961-2024). The second period exhibits more extended runs of wet, dry, and non-dry years, suggesting a change in the driving dynamics. A multi-annual hydrological drought lasting for 13 years (1961-1973) was identified, followed by a stretch of 46 years (1974-2019) with no droughts in the study area. Our simulations indicate that the 46-year period with no drought had only a 4%

probability of occurrence. Precipitation is the primary predictor of regime shifts, but the class transition probabilities and precipitation thresholds are non-homogeneous and conditional on the current low-water level regime. We identify precipitation thresholds for initiating transitions between Dry, Wet and Normal years, conditioned on the current water levels: in a normal year, precipitation below 1,040mm triggers a hydrological drought, while in a drought year, precipitation above 1180mm triggers a return to normal conditions. The research advances nonstationary extreme event analysis by proposing an efficient new approach to estimate inhomogeneity in hydrological drought occurrence; identify long persistence of hydrological drought episodes and their associated probabilities; define precipitation thresholds; and reveal the importance of coupled drivers of low water level shifts.

Keywords: Hydrological drought, low-water level regimes, Prediction, Multisite Hidden Markov model, Multinomial model, Attribution.

1 INTRODUCTION

Drought impacts extend across various sectors, affecting ecosystems, hydropower generation, navigation, agriculture, and communities dependent on reliable water supplies (Hao et al., 2018; Mishra and Singh, 2010; Van Loon, 2015). In the nonstationary context of 21st century anthropogenic climate change – where natural systems no longer fluctuate within a stationary envelope of variability (Milly et al., 2008) – extreme hydrological droughts may occur more frequently (Gu et al., 2023; Hagenlocher et al., 2023). This shift challenges traditional approaches to water management, as the statistics of hydrological variables, such as streamflow and precipitation, may no longer be time-invariant (Montanari & Koutsoyiannis, 2014).

In this context, the main objective in the domain of drought-related research and water management lies in deciphering the drivers of fluctuations in low flows and low water levels (Hao et al., 2018; Rolim and Filho, 2020; Salas et al., 2018; Slater et al., 2021; Smakhtin, 2001). Understanding the interplay among climatic patterns, land-use transformations, and hydrological responses is an ongoing research question in hydrology (Brunner et al., 2021; Slater et al., 2021). By tackling these attribution challenges, we deepen our comprehension of drought-related phenomena and can help alleviate the impacts of hydrological droughts, such as abnormally low water levels in rivers, lakes, and groundwater (Van Loon, 2015), while contributing to the development of adaptation strategies and management plans (Brunner et al., 2021).

One of the pillars of drought management is the development of drought monitoring and early warning systems (Funk and Shukla, 2020). Effective drought management strategies should enable an appropriate understanding of the mechanisms driving drought occurrence, in order to prepare communities and adequately communicate drought risks. Here, the aim of this research is to propose a framework describing the long-term fluctuations of hydrological droughts, identifying the mechanisms driving drought occurrence, and integrating these insights into a predictive model for drought dynamics. Achieving this objective could contribute to the development of targeted strategies for hydrological drought mitigation, such as the identification of drought thresholds.

Traditional drought prediction methods, such as autoregressive models, Markov chains, and machine learning techniques (such as artificial neural networks or support vector machines), have been widely used under stationary assumptions (Salas et al. 2003; Mishra and Singh, 2010; Hao et al., 2018). However, these methods often fail to account for the nonstationary nature of hydrological systems, which are increasingly influenced by climate change and human activities (Milly et al., 2008; Montanari and Koutsoyiannis, 2014). Recent advances in nonstationary modelling, such as generalized additive models (GAMs), Bayesian hierarchical models, and hybrid approaches combining physical and data-driven techniques, have shown promise in addressing these challenges (Slater and Villarini, 2018; Slater et al., 2021; Brunner et al., 2021; Rolim and Filho, 2021; Gu et al., 2023). These methods incorporate time-varying parameters and external drivers (e.g., climate indices, land-use changes) to better capture the dynamics of hydrological droughts under changing conditions.

Despite these advancements, there is still a need for approaches that can effectively describe regime-shifting behaviour and predict hydrological droughts fluctuation, particularly in large and complex basins where multiple drivers interact. To address this gap, we develop a new framework based on a two-step modelling approach using a multivariate Hidden Markov Model (HMM) and a Multinomial Logistic Regression model (MLR) (Suchetana et al., 2019), applying a bootstrap strategy to assess uncertainty (Bishop, 2006). HMM models are widely used in hydrology for capturing regime transitions and temporal dependencies in hydrological data (Hughes & Guttorp, 1994; Thyer & Kuczera, 2000, 2003; Akintuğ, & Rasmussen, 2005; Gelati et al., 2010; Bracken et al., 2014). However, to the best of our knowledge, no studies have combined HMM with MLR for predicting drought class transitions. Using HMM, we classify the low-water level time series into dry, normal, and wet years and assess the frequency of each class in the historical data. Dry years are identified as hydrological droughts. To predict low-water level class transitions in a nonstationary context, we propose a novel MLR framework. With this, we estimate probabilities of class transitions by inputting external variables into the transition matrix estimates. Uncertainties and sensitivity of the class transition are estimated using a bootstrap approach.

We apply this framework to the Paraguay River Basin (PRB), where recent droughts have caused widespread impacts on ecosystems, agriculture, navigation and on the communities dependent on reliable water supplies (Libonati et al., 2022; Santos et al., 2021; Tomas et al., 2021), and where prolonged multiannual drought episodes have been reported (Collischonn et al., 2001; Santos et al., 2021). The PRB encompasses the South American Pantanal, a World Heritage Area and one of the world's most extensive wetlands, spanning approximately 180,000 km², with immense ecological importance and significant cultural and socio-economic interest (Assine et al., 2014; Collischonn et al., 2001; Marengo et al., 2021; Tomas et al., 2019; UNESCO, 2020).

Based on these objectives, our research questions are as follows. Can a unified statistical classification of low-water levels time series be used to derive a unique regime classification? What is the underlying cause of shifts in low-water level regimes in the PRB? How are these regimes influenced by precipitation, land use, and climate teleconnection patterns? How can these drivers be better integrated into drought prediction models?

We describe the main features of the PRB and the data used in Section 2. Next, we explain the methodology used, emphasising the theoretical aspects of the Multisite HMM and the MLR that enabled us to achieve the results presented in Section 3. In Section 4, we present the main findings of the research, including a detailed description of the evolution of the low-water levels transition matrix in the area and the complexities related to drought manifestations in the PRB. We also explain how we define quantitative thresholds for drought definitions based on different precipitation thresholds, depending on the current state of the catchment. Additionally, we provide relevant discussion on the main findings of the research in section 4. Finally, we summarise the main results of this study in the conclusions presented in Section 5.

2 DATA AND METHODS

2.1 Study area and data

The study focuses on the PRB, as depicted in Figure 1. We considered the Porto Murtinho stream gauge at the basin outlet (code 67100000), which has a drainage area of 575,000 km².

The PRB is located in a sedimentary basin formed due to the continental plate's distension and the Pantanal subsidence resulting from the Andes' formation (Assine et al., 2014). This sedimentary structure, with its flat topography and the formation of a distributive fluvial system, favours the occurrence of very slow flows and the presence of extensive water bodies and connection between lakes, which contributes to a large time of concentration and high evaporation rates (Assine et al., 2014; Collischonn et al., 2001; Pupim et al., 2017).

These physical characteristics result in a very low runoff ratio. The entire PRB, with the Porto Murtinho stream gauge as its outlet, records an average annual precipitation of approximately 1,150 mm over the catchment area based on CHIRPS data from 1981 to 2024. The mean annual flow is 2,350 m^3/s (129 mm), resulting in an estimated annual runoff ratio of 0.11. Wigley and Jones (1985) suggested that catchments with such low runoff ratios experience substantial runoff changes even with minor alterations in precipitation. Additionally, the large size of the PRB means that even slight variations in rainfall can lead to significant shifts in hydrological regimes (Collischonn et al., 2001).

Figure 1: Paraguay River Basin: the top right panel shows the location of the upper Paraguay River in the middle of South America. The larger panel on the left displays the river catchment in detail. The stream gauges are indicated as coloured triangles, with colours reflecting the time series length. The sub-catchments and other graphical elements are shown in the legend. The daily water levels are displayed on the map. The time series plots indicate daily and annual minimum water levels highlighted with darker lines and points, respectively. Only data from 2010 onward are shown in the time series plots to illustrate seasonal patterns.

The PRB has experienced several significant hydrological droughts in its history, highlighting the basin's vulnerability to water deficits and extremely low water levels. From 2020 to 2024, the PRB experienced a sequence of five consecutive dry years, with the most severe drought on record occurring in 2024 (Serviço Geológico do Brasil, 2024), and the third-worst drought in 2021 (Santos et al., 2021). Additionally, a prolonged drought lasting over 10 years affected the basin between the 1960s and the early 1970s (Santos et al., 2021). These events have had profound impacts on multiple sectors, including navigation (hindering the movement of larger vessels), sanitation (requiring floating pumps to maintain water supply), and the environment (increasing the risk of the *Decoada* phenomenon, which involves high fish mortality due to reduced water flow and degraded aquatic habitats) (Santos et al., 2021). The

wide-ranging influence of these events underscores the importance of understanding and predicting hydrological droughts in the PRB, particularly under nonstationary conditions.

The rainy season in the PRB is centred in the Austral summer (from December to February), while reduced rainfall occurs during the Austral winter (from June to August). This characteristic is linked to the activity of the South America Monsoon System (SAMS) (Cavalcanti et al., 2015; Marengo et al., 2012). Rainfall seasonality is reflected in daily water level plots, presented in Figure 1. All stream gauges' daily water levels time series show a single annual flow peak and a trough (Figure S.1).

The El Niño-Southern Oscillation (ENSO) is reported to exert a climatological influence on the PRB. El Niño (La Niña) events in the area are associated with increased (reduced) occurrences of extreme precipitation in the austral spring in the Upper La Plata Basin, where the PRB is located (Cavalcanti et al., 2015). The Pacific Decadal Oscillation (PDO) is also a relevant teleconnection pattern in South America. Kayano and Andreoli (2007) described the PDO as a modulator of ENSO impacts, where the PDO index either amplifies or dampens ENSO's influence in South America, depending on whether the two phenomena are in the same or opposite phases, respectively. In this context, the PDO provides conditions that enhance the observability of ENSO-related teleconnections in the region.

The Atlantic Multidecadal Oscillation (AMO) also may influence the study area, with positive (negative) AMO anomalies leading to reduced (increased) precipitation in the region (Chiessi et al. 2009). Recent studies suggest that the predictability of extreme precipitation and drought events in the Pantanal region has improved significantly due to a reduced lead-time between sea surface temperature (SST) warming and its associated impacts on precipitation. This shorter lag time - approximately two months - accounts for 80% of the observed variation in precipitation patterns, as driven by key oceanic indices such as ENSO, PDO, AMO, and ATL3 (Thielen et al., 2020). However, there are also reports of inconsistencies between river levels and rainfall in the PRB and ENSO (Marengo et al., 2021).

The internal catchment dynamics induce further complexity into the predictability of hydrological variables in the PRB, including low-flows. Groundwater systems strongly control hydrological drought characteristics, particularly in equatorial and warm temperate climates (Van Lanen et al., 2013). The PRB is characterised by interconnected lakes typical of the Pantanal wetland region (Pupim et al., 2017), and these natural reservoirs induce a natural capacity for flow regularisation, limiting abrupt hydrological transitions, and also drought duration. Due to the basin's internal dynamics and hydrological characteristics, the propagation of rainfall effects on river levels does not occur directly, often attenuating climatic forcings.

For this study, water level data was collected from the Brazilian National Water Resources Information System. The National Water and Sanitation Agency (ANA) makes all Brazilian hydrological data freely available from the *HidroWeb* system (ANA, 2017). The data used in this study was accessed in November 2024. Figure 1 shows the location of the 26 stream gauges used in this study, along with the length of the annual minimum water level records, ranging from 45 to 124 years. The Ladário stream gauge (66825000), located in the main channel of the Paraguay River, has the most extended time series in the study area. This stream gauge is often used as a regional hydrological index for droughts and floods in the PRB (Santos et al., 2021). Summing up the annual minimum low-water levels in all 26 stream gauges, in this research we analyse 1,527 events.

Rainfall data were collected from CHIRPS (Funk et al., 2015) and GPCC (Schneider et al., 2020). CHIRPS data is available from 1981 and is currently one of the data sources used by the Early Warning System in the PRB, maintained by the Geological Survey of Brazil (Santos et al., 2021). Additional rainfall information was gathered from the GPCC, which provides long-term global precipitation estimates from 1891 onwards. This time window covers some relevant hydrological phenomena in the study area that CHIRPS data does not, for instance, the long-term drought observed from 1961 to 1973. CHIRPS and GPCC data were compared (Figure 2(b)) and no significant bias was found between the two datasets.

In Figure 2(a), we present the time series of accumulated rainfall throughout the hydrological year, which is composed of CHIRPS data for the period 1981-2024 and CHIRPS-equivalent

estimates for the period prior to 1981 (1941-1980). These estimates were generated using a linear regression model, where GPCC data were used to predict CHIRPS rainfall values. Details on the composition of rainfall data is provided in Supplementary Material (Figure S.5). In Figure 2(a), we also display a locally weighted regression line (*loess*) to highlight the long-term oscillations in the historical series. The *loess* fit line indicates a drier period during the 1960s and rainier period during the 1980s, returning to an average level during the 1990s until approximately 2018. The linear regression line in Figure 2(a) reveals a statistically significant downward trend over the entire period, suggesting a long-term decrease in precipitation.

Figure 2: (a) Precipitation time series between 1940 until 2024, based on CHIRPS data from 1981 onwards, and CHIRPS estimates from GPCC based on linear regression using GPCC data as input from 1940 until 1980 using linear regression. A locally weighted regression line (loess) shows long-term oscillations in the historical series, alongside a linear regression line representing the long-term trend. The negative trend observed in the graph is statistically significant (p-value << 0.05) at an average rate of 9.70 mm reduction per decade; (b) Scatterplot of cumulative monthly precipitation, constructed using the common data period from 1981 to 2020. The colours indicate the year quarters (the 1st and the 4th quarter indicate the rainy season).

We also evaluated groundwater and land-use data. Groundwater data were derived from GRACE-FO RL06.3 Mascon Solutions (RL0603), accessed through the Center for Space Research at the University of Texas at Austin (Save et al., 2016). Additionally, land-use data were obtained from Dias et al. (2016), which offers detailed classifications of land cover and agricultural land use for the study area.

2.2 Methodology

We classify low-water levels into three categories: Dry, Normal, and Wet years, with the first category corresponding to hydrological drought. To perform this classification, we use a HMM framework, which also evaluates the transition probabilities between low-water level regimes in the PRB using data from multiple monitoring sites.

We adopt a two-step modeling approach: the HMM captures temporal regime shifts, while MLR identifies the primary factors driving these shifts (Suchetana et al., 2019). As inputs to the MLR, we evaluate the role of precipitation and teleconnection patterns – namely the El

Niño-Southern Oscillation (ENSO), Atlantic Multidecadal Oscillation (AMO), and Pacific Decadal Oscillation (PDO) – as predictors of low-flow class transitions in the study area. This methodology integrates these drivers to establish quantitative thresholds for the drought occurrence.

A multisite HMM is preferred over multiple univariate HMMs to avoid estimating independent water level regimes and creating a complex network of hidden states. We assume that the use of one unique underlying state provides better insights into the underlying processes, thus allowing for more accurate predictions using teleconnections with the PRB. We use the *depmixS4* package in R (Visser and Speekenbrink, 2010) for our analysis and provide reproducible R scripts for all results achieved in this study in https://github.com/msuassuna/RegimeSwitchesPantanal.

Before fitting the HMM, the water level data is scaled by subtracting the mean and dividing by the standard deviation. Missing values are imputed using the *depmixS4* package, which employs the likelihood function to perform this procedure. This approach ensures that no additional information is introduced during the imputation process, as the missing values are estimated solely based on the likelihood of the current state (Visser and Speekenbrink, 2010).

We determine the number of low-water level regimes using the Bayesian Information Criterion (BIC). Information criteria, such as the Akaike Information Criterion (AIC) and BIC, are widely used to address biases in maximum likelihood estimation. These criteria balance model fit, represented by the likelihood, with model complexity, represented by the number of parameters, by introducing penalty terms to discourage overfitting (Bishop, 2006; Burnham and Anderson, 2002). Compared to AIC, BIC applies a stricter penalty for model complexity, favouring more parsimonious models (Bishop, 2006). Further details on the BIC calculation are provided in the Supplementary Material.

For the multisite HMM, we use the maximum number of gauges available in each period. From 1900 to 1932, data was available only from the Ladário stream gauge (ID 66825000), which was the sole gauge used during that period. From 1933 to 1963, data from Porto Murtinho and Cuiabá stream gauges (67100000 and 66260001) is included. Those three sites

determine the regional HMM before 1963. The most extensive time series provide important insights into the basin's behaviour, making this period essential for drawing meaningful conclusions about long-term droughts observed in the region during the 1960s and 1970s. Moreover, these three stations generally reflect the regional behaviour of the basin and are used as indicators of the PRB state (Santos et al., 2021). Comparison with other stream gauges in the region shows that their use as indicators is consistent with the regional regime transition classes. Nonetheless, we also evaluate whether this procedure introduces bias to our understanding of the catchment's behaviour.

A relevant assumption for the multisite HMM is that the residuals are Gaussian. This assumption is reasonable for the scaled water level data. The Shapiro-Wilk normality test (Royston, 1995) is applied to the residuals of the HMM to check if the assumption holds. We also evaluate the coherence between each stream gauge's low-water level regime and the regional variation. ANOVA tests are performed to check if the regimes' magnitudes in different sites have the same regional ones. We verify whether each site's low-flow class magnitudes are consistent with the regional patterns.

We compare the regional transition matrices between the two halves of the time series: 1900-1960 and 1961-2024, to evaluate if they are homogeneous in time. The homogeneity was assessed by estimating the 0.95 confidence intervals of the transition probabilities using a bootstrap approach. Non-overlapping intervals indicate that the transition matrices are significantly different. We also simulate the water-level regime using the estimates of the transition matrices to assess the frequency of past extreme drought events in the study area.

The attribution part is a crucial point of the quantitative hydrological drought definition and prediction in a nonstationary context. At this step, we search for the variables that allow improvements in the predictions of low-water level regime transitions. Climatic drivers such as precipitation and teleconnection patterns are tested as predictors of long-term variability of low-water levels. We first verify if the rainfall data reflects similar low-water level regimes by repeating the HMM classification analysis using rainfall. We also check for a significant change in the precipitation amounts during different low-water level classes. Considering the substantial quantity of lakes and the natural capacity of water storage in the PRB, we

hypothesize that hydrological drought cannot be inferred from the climatic variables alone. If this is the case (based on the outcome of the hypothesis test), then the definition of thresholds should be conditional on the current catchment state, as the internal dynamics of the catchment are assumed to play a relevant role in the formation or persistence of low-water levels classes. Thus, we evaluate different precipitation/teleconnection thresholds during different low-water level regimes to check if that hypothesis holds. The results of this test are presented in section 3.1.2. These results lead us to select a non-homogeneous MLR model, in which the parameters are conditional on the current catchment state. Finally, the time series is resampled to estimate confidence intervals for the precipitation thresholds and assess the models' sensitivity to input variables.

2.2.1 Multisite Hidden Markov Models

The Hidden Markov Model (HMM) is a state-space framework with latent variables that can be interpreted as a mixture distribution (Naghettini & Pinto, 2007), where the choice of component for each observation depends on the state of the previous observation. (Bishop, 2006). The model can be specified by adding a latent variable $\mathbf{z}_n$ for the observed time series $\mathbf{x}_n$ (bold letters for vectors). In this study, $\mathbf{x}_n$ refers to the time series of annual minimum water level in the PRB, and $\mathbf{z}_n$, to an unobserved hidden variable that reflects a *low-water level regime* or *low-water level class*.

One can assume that $\mathbf{z}_n$ is a discrete variable, giving rise to the structure known as a discrete state space model (Bishop, 2006). If one assumes that $\mathbf{z}_n$ depends on the previous observation only, and thus $\mathbf{z}_n$ forms a Markov Chain, it gives rise to the structure of the HMM. HMMs thus consist of a set of hidden states ($\mathbf{z}_n$), a set of observable variables ($\mathbf{x}_n$), a set of transition probabilities between the hidden states described by the transition matrix **A**, and a set of emission probabilities, defined by Bishop (2006) as the probabilities of observing $\mathbf{x}_n$ conditional on the latent state $\mathbf{z}_n$ ($p(\mathbf{x}_n|\mathbf{z}_n, \Phi)$), where Φ is a set of parameters governing the conditional distribution of $\mathbf{x}_n$.

In this framework, the probabilities that describe the occurrence of $\mathbf{z}_n$ depend only on the state of the previous latent variable $\mathbf{z}_{n-1}$, through a conditional distribution $p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A})$,

which defines the Markovian framework, where only the last observations affect the probabilities of the next one. The matrix **A** is a composition of those conditional probabilities, and because they are probabilities, each element A_{jk} of the transition matrix satisfy the condition $0 \leq A_{jk} \leq 1$ with $\sum_k A_{jk} = 1$. In this formulation j and k stand for the latent state in $n-1$ and n, respectively.

The model is completed by defining the emission probabilities ($p(\mathbf{x}_n|\mathbf{z}_n, \Phi)$), In this specific application of the HMM framework to low-water level data, Φ is a set of parameters that describes the distribution of low-water levels for each low-water regime. The following equation then describes the joint formulation of latent and observed variables:

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \prod_{1}^{K} p(\mathbf{z}_{1k} = 1) \left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{m=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \Phi)$$

where $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$, and $\boldsymbol{\theta} = \{\mathbf{A}, \Phi\}$ describes the parameters that define the model. In this research, the main goal of HMMs is to estimate the most likely sequence of hidden states ($\mathbf{z}_n$) as well as the transition matrix (**A**) given the observed data ($\mathbf{x}_n$), employing the capabilities of HMMs as a tool for time series classification and simulation.

The estimation of HMM parameters in this study was performed using the *depmixS4* package (Visser and Speekenbrink, 2010) written in the R programming language. The *depmixS4* package employs the Expectation-Maximization (EM) algorithm. The EM algorithm iteratively refines the model parameters by alternating between two key steps: the Expectation step, where the expected value of the complete-data log-likelihood is computed based on the observed data and current parameter estimates; and the Maximization (M) step, where parameters are updated to maximize this expected log-likelihood (Visser and Speekenbrink, 2010). Further technical details on the EM algorithm and its implementation in *depmixS4* can be found in Visser and Speekenbrink (2010).

2.2.2 A Multinomial Model with Bootstrap

A Multinomial Logistic Regression (MLR) was used to capture the probabilities of class transitions. The MLR model can define probabilities and thresholds that link the covariates'

dependence and the expected sequence of low-water level regimes obtained from the HMM. We fit different models to each of the three low-water level regimes, so the model setup is comparable to a nonhomogeneous HMM (Robertson et al., 2004).

The logistic regression model arises from the desire to model the posterior probabilities of the $k = 1, \dots, K$ classes via linear functions, ensuring that they sum to one and remain in [0, 1] (Hastie et al., 2009). The probabilities in the logistic model are estimated by rearranging the terms of the *log-odds* or *logit* transformations, resulting in the formulation described in the following function:

$$P(\mathbf{z}_n = k|\mathbf{Y}_n) = \frac{\exp\left(\beta_{k0} + {\beta_k}^T \mathbf{Y}_n\right)}{1 + \sum_{l=1}^{K-1} \exp\left(\beta_{l0} + {\beta_l}^T \mathbf{Y}_n\right)}$$

Where β_{k0} is the intercept and β_k is the angular coefficient of the *log-odds* ratio; k, in this study, denotes the low-water level classes (Dry, Normal, and Wet years); the index k refers to the model for each low water level class; $\mathbf{Y}_n$ relates to the set of predictors of the low-water level class transition. Logistic regression models are usually fit by maximum likelihood (Hastie et al., 2009), given $\mathbf{Y}_n$ as the set of predictors of the hidden state $\mathbf{z}_n$. Typically, many models are fit in a search for a parsimonious set of predictors $\mathbf{Y}_n$ (Hastie et al., 2009).

In this study, the MLR model was fitted using the *nnet* package in R (Venables and Ripley, 2002), which implements MLR via neural networks. To assess statistical accuracy, we employed a bootstrap resampling approach. Subsamples of the same size as the original training set were drawn with replacement, and a MLR was fitted to each subsample. This iterative procedure produced a MLR ensemble, enabling the estimation of confidence intervals for precipitation thresholds that trigger transitions between low-water level classes. Furthermore, this approach allowed us to quantify uncertainty: greater dispersion among the fitted MLR lines corresponds to higher uncertainty in the precipitation thresholds.

3 RESULTS

3.1.1 Multisite HMM analysis

We develop a new approach to objectively classify hydrological Dry, Normal and Wet years across the PRB. To do this, we analyse the scaled water level data from 26 stream gauges to obtain a single low-flow class for the entire region annually, using a multisite HMM. The resulting class represents the average low-water state within the PRB, although specific stream gauges may exhibit behaviour that deviates from the regional mean.

The optimal number of classes was objectively defined based on the BIC. Figure S.2 in the Supplementary Material depicts BIC estimates for different numbers of low-water level regimes, and shows that the improvement in model fit decreases rapidly beyond three drought categories. Thus, three low-water level regimes is the optimal number of classes.

To illustrate how the regional classification of low-water level regimes is reflected locally, we present detailed results at the Ladário stream gauge (Figure 3a). The average minimum water levels in a Dry, Normal, and Wet year are -10.6 cm, 108 cm, and 197 cm, respectively. It is important to note that the classification is based on regional hydrological conditions across the entire basin, and local conditions at Ladário may occasionally deviate from the regional average due to spatial variability within the basin. The mean values vary significantly across the different drought regimes (Student's t-Test, $p<<0.05$). At the same time, the estimated standard deviations show slight variation, with an average of 35 cm within each low-water level class. Figure 3(b) depicts a trimodal fit of a Gaussian distribution, suggesting the mixed distribution closely resembles the observed behaviour of the empirical kernels of low water level classes – further supporting the robustness of the classification. Details on how the mixed distribution is estimated are provided in Supplementary Material (Figure S.3).

We detect a growing persistence of low-flow regimes in recent decades (Figure 3). From 1961 to 1973, we observe 12 consecutive years classified as Dry, indicative of a multiannual hydrological drought. Subsequently, from 1976 to 1999, we find a continuous run of 23 Wet years. Remarkably, from 1974 to 2019, no Dry years were observed in the study area over the entire 46-year period. Such prolonged runs were not reported in the first half of the time series between 1900 and 1960, where a more regular pattern of alternating low-flow regimes is observed.

Figure 3: (a) Annual minimum water level time series at the Ladário stream gauge on the PRB. The low-water level classes were estimated for the entire PRB. The mean and standard deviation of Dry, Normal and Wet years are displayed in the legend; (b) Mixed Trimodal Probability Distribution Function at the Ladário stream gauge. The means and standard deviations of low water levels are indicated. The frequencies of Dry, Normal, and Wet years are 23.8%, 46.7%, and 29.5%, respectively. The mixed distribution (shown as a grey polygon) closely resembles the observed behaviour of the empirical kernels of low water level classes.

We provide a detailed basin-scale overview of the regional low-water levels by summarising the different regimes using violin plots (Figure 4). As a single low-flow classification was utilised to represent the average behaviour of the catchment, specific stream gauges may exhibit behaviour that deviates from the regional mean. The regional regimes resulting from the regional HMM are used to estimate local low-water level averages. If the regional low-water regimes are consistent with the local classes, one should observe an increase in the low-flow mean values in each stream gauge from left to right in Figure 4.

Figure 4: Spatial pattern of mean water levels during different regimes: 1 – Dry years in red; 2 – Normal years in green; and 3 - wet years in blue. The stream gauge triangles coloured in blue follow the average regional pattern; those in orange, do not. The main channel of the Paraguay river determines the regional pattern of Dry, Normal, and Wet years illustrated in this map, where the Ladário (Base Naval) stream gauge is located.

The low-water levels in the main channel of the Paraguay River conform to the regional classification. The Ladário stream gauge serves as an illustration of the catchment's average behaviour. Other stream gauge records that follow the average regional pattern are coloured in blue (21 sites), while those in orange deviate (5 sites). The ones that deviate from the regional pattern exhibit two types of behaviour. Some sites fail to separate low-water level classes in the same manner as observed in the region (e.g. stream gauge 66260001, located in the northeastern part of the catchment in the Cuiabá River basin). In contrast, some other stream gauges merely exhibit slight differences in drought regime magnitudes relative to the regional behaviour (e.g. in the southern sub-catchments of the Miranda River). A detailed visualization of the annual minimum water levels across all stream gauges, categorized by low-water level class, is provided in the Supplementary Material (Figure S.4).

Overall, the main channel of the Paraguay River in the west and parts of the Correntes, Aquidauana, and Taquari rivers in the east represent the regional average low-flow classes,

443 with the Ladário stream gauge serving as an example. The Cuiabá River basin differs from
444 the regional average behaviour. Southeast catchments exhibit similar behaviour to the
445 regional average, yet not significantly, according to the ANOVA test ($p > 0.05$).

446 The stream gauges that conform to regional patterns are identified by the blue markers in
447 Figure 4. A three-component Gaussian mixture distribution is a possible fit in those sites,
448 with varying mean values and a fixed standard deviation. For instance, the Ladário stream
449 gauge time series, which covers the entire period from 1900 to 2024, represents drought
450 regimes using sub-samples with a frequency of 23.8%, 46.7%, and 29.5% in Dry, Normal,
451 and Wet years, respectively. The means and standard deviations for each drought regime in
452 this specific stream gauge are shown in Figure 4.

453 To illustrate the result of a Gaussian fit of the trimodal distribution, we select the
454 hydrological drought observed in 2021. The probability of observing an event of -60 cm in
455 this site in a random year results from a 23.8% probability of a severe drought in a random
456 year, multiplied by a probability of 7.6% of a water level of -60 cm within this low-water
457 level class. With that, we estimate a probability of 1.8% of observing a low-water level in the
458 2021 event. The event corresponds to a return period of 55 years in the fitted trimodal
459 Gaussian distribution. However, this probability estimate is not conditioned on the observed
460 low-water level class in 2020. We later analyse the transitional matrix of low-water level
461 classes to consider these conditional probabilities.

462 The transition matrices from the multisite HMM show the transition probabilities between
463 three possible states corresponding to the different low-water level regimes (Figure 5). As
464 indicated, the transition probabilities were estimated for the entire time series (Figure 5a;
465 1900-2024) and its halves (Figure 5b and Figure 5c, from 1900-1960, and 1961-2024,
466 respectively). Detailed information on those transition probabilities is provided in Table 1,
467 which displays the results of the transition probabilities and their confidence intervals
468 evaluated with Bootstrap. The table includes lower and upper endpoint matrices representing
469 the 0.95 confidence intervals for the drought transition matrices.

The transition matrices indicate low probabilities of direct transitions between Dry and Wet years. Specifically, during the first half of the time series, the probability of transitioning from a Dry to a Wet year was 0.06, and from a Wet to a Dry year 0.09. In contrast, both probabilities dropped to zero in the second half of the time series. This suggests that in the latter period, abrupt transitions between Dry and Wet years (and vice versa) became virtually non-existent, and transitions between low-water level classes occurred predominantly in a gradual manner.

The persistence of low-water level regimes, as illustrated in Figure 3, highlights significant temporal changes in the transition probabilities. In the first half of the time series (Figure 5b), the probability of a Dry year persisting into another Dry year is 43%, while transitions to Normal (51%) or Wet years (6%) remain more likely than in later periods. Similarly, Wet years during this period show lower persistence (36%) and a higher chance of transitioning into a Normal year (56%), indicating a more dynamic hydrological regime with frequent changes between states.

In contrast, the second half of the time series (Figure 5c) reveals an increase in persistence for Dry years, where the probability rises to 87%, nearly doubling the previous estimate. At the same time, the persistence of Wet years also increases significantly to 77%, suggesting that the hydrological system has shifted to clustering extreme events (both Dry and Wet years) with fewer transitions to intermediate (Normal) conditions. Transitions between Dry and Wet years become notably rare, with probabilities close to zero, as discussed earlier.

Figure 5: Transition diagrams representing the transition matrices between low-water level classes in the PRB. The transition probabilities indicated in the figure correspond to the elements of the transition matrices estimated for different periods: (a) entire record (1900-2024), (b) first half (1900-1960) and (c) second half (1961-2024) of the time series.

The persistence of severe droughts in the region has changed significantly over time (Table 1). The 95% lower endpoint probability of this persistence in the second half of the time series (0.85) is higher than the upper endpoint matrix in the first half of the time series (0.46),

as well as compared with the entire time series (0.67). The 95% confidence intervals do not overlap; hence, the probabilities' differences may be considered statistically significant.

Recent extreme drought events in the PRB illustrate this change in persistence. The severe drought observed in 2021 followed another severe drought in 2020. Based on the transition matrix for the entire period, the probability of observing a severe drought in 2021 was estimated at 67%. However, if the transition matrix for the second half of the time series is considered, the lower endpoint matrix indicates a probability of a severe drought of 88% in 2021. Furthermore, based on the transition matrix for the entire period, the average duration of a dry run would be three years, whereas if the drought transition of the second half is considered, the expected drought length would be eight years.

While unconditional estimates help evaluate average probabilities and return periods of droughts, conditional probabilities are more applicable in the operational monitoring drought conditions in the study area. Here, we propose an evaluation of the 2021 drought based on the situation observed in 2020, and compare it with the unconditional probability. The unconditional probability of observing a water level of - 60 cm in the Ladário stream gauge is 1.8% (as demonstrated in section 3.1.1). However, if we assume that a dry year was observed in 2020 and the transition matrix for the entire period is considered, the probability of another dry year in 2021 would be 67%, which, multiplied by 7.9% (the probability of a water level of - 60 cm within this low-flow class), returns a probability of observing this water level of 5% (i.e. a return period of 20 years). This probability is much higher than the unconditional estimate. If the transition matrix for the second half is used, a conditional probability of 6.7% would be estimated (return period of 11 years).

Table 1: Transition matrix and 95% confidence level of the lower and upper matrix endpoints, estimated using bootstrapping. Different transition matrices were estimated for different periods: total length (1900-2024), first (1900-1960) and the second half (1961-2024) of the time series.

		1900 – 2024			1900 - 1960			1961 - 2024		
		Dry	Normal	Wet	Dry	Normal	Wet	Dry	Normal	Wet
Transition matrix estimates	Dry	0.65	0.31	0.03	0.43	0.51	0.06	0.87	0.13	0.00
	Normal	0.18	0.64	0.18	0.22	0.63	0.15	0.15	0.60	0.25
	Wet	0.02	0.31	0.67	0.09	0.56	0.36	0.00	0.23	0.77
	Dry	0.63	0.29	0.03	0.40	0.48	0.04	0.85	0.12	0.00

		1900 – 2024			1900 - 1960			1961 - 2024		
		Dry	Normal	Wet	Dry	Normal	Wet	Dry	Normal	Wet
Matrix Lower Endpoint	Normal	0.17	0.63	0.17	0.20	0.61	0.13	0.13	0.57	0.22
	Wet	0.02	0.29	0.65	0.07	0.52	0.31	0.00	0.20	0.75
Matrix Upper Endpoint	Dry	0. 67	0.34	0.04	0.46	0.55	0.07	0.88	0.15	0.00
	Normal	0.20	0.66	0.19	0.24	0.65	0.17	0.18	0.63	0.27
	Wet	0.03	0.33	0.68	0.11	0.59	0.40	0.00	0.25	0.80

We also evaluate the probability of observing a prolonged period without dry years in the PRB using simulations of low-flow classes with transition matrix estimates from Table 1. Our motivation for conducting these simulations is based on the observation that the PRB experienced 46 years (1974-2019) without a single dry year, and we aim to assess the likelihood of this occurrence. Using the transition matrices displayed in Figure 5, we generated 2,000 samples of a 46-year time series and counted the number of hydrological droughts observed within the period.

Figure 6 summarizes the results of our drought simulations. The three histograms illustrate the frequency distribution of simulated drought years using the transition matrix displayed in in Figure 5(a) (Figure 6a), in Figure 5(b) (Figure 6b), and in the Figure 5(c) (Figure 6c). For the transition matrix in the first half of the time period, from 1900–1960 (Figure 6b), the number of Dry years closely follows a Poisson distribution, with an average occurrence of 10.6 droughts in 46 years. This corresponds to a probability of approximately 23.1% for any given year to be in a Dry state, which is consistent with the steady-state probabilities derived from the Markov chain model shown in Figure 5b. Importantly, the likelihood of observing a 46-year period with no severe droughts is extremely low (< 0.01).

In contrast, the second half of the time series, from 1961-2024 (Figure 6c), reveals a notable shift. The simulations indicate a higher frequency of Dry years, with a more right-skewed distribution, reflecting the increased persistence of drought conditions observed in the transition matrix (Figure 5c). This shift suggests that the hydrological regime has become more prone to extended drought periods. Using the transition matrix from 1961–2024 (Figure 5c), we estimate an average of 12.9 Dry years over the 46 years, corresponding to a probability of occurrence of 28.1%. While this probability aligns closely with the steady-state condition of the Markov chain (29.4% chance of observing severe droughts), the frequency

distribution of Dry years does not follow a Poisson distribution. Instead, it approximates a zero-inflated Gaussian-like distribution.

Figure 6: Histograms depicting the number of years of hydrological droughts (Dry years) in a 46-year period using (a) the average transition matrix for the entire time series; (b) the transition matrix for the 1900-1960 period; and (c) the transition matrix for the 1961-2024 period. The 46-year period was chosen to reproduce the 1974-2019 period, during which no hydrological droughts were observed. In histogram (c), the red bar represents the frequency of simulated series in which *zero* hydrological droughts were observed, based on the transition matrix for the 1961-2024 period.

The zero-inflated Gaussian distribution combines two distinct components: a standard Gaussian distribution centered at approximately 17 Dry years and a point mass at zero years Figure 6c. The presence of multiple components within this distribution suggests that low-water level regimes may result from different climatological processes or mechanisms. This statistical behaviour highlights the potential complexity of drought dynamics in the region. However, elucidating the underlying factors that lead to the simulation result of a zero-inflated Gaussian distribution derived from a high-persistence transition matrix requires further investigation, which falls outside the scope of this study.

This simulation estimates a 4.2% probability of observing zero Dry years within a 46-year period, based on the transition matrix with long persistence and the zero-inflated Gaussian distribution. This result indicates that the 46-year drought-free period observed between 1974 and 2019 is statistically unlikely, with a probability of only 4.2%. However, this value remains significantly higher than what would be expected under different conditions or if one assumes a Poisson distribution with an average of 12.9 drought years over the same period.

Our simulations further suggest that a prolonged period without any drought, such as the one observed in the PRB between 1974 and 2019, is unlikely to recur statistically, even if the underlying conditions for drought events remain unchanged, keeping the transition matrix described in the Figure 5c. The three transition matrices presented in Figure 5 support this finding, indicating that long periods without hydrological droughts are unlikely given the observed shift towards higher drought persistence in recent decades. Consequently, this statistical improbability suggests that future conditions in the PRB are likely to include an

increased frequency of drought events, as prolonged intervals without droughts become less feasible.

3.1.2 Drivers of long-term fluctuations

To evaluate whether the low-flow class transitions in the PRB are determined or induced by precipitation, we used a combination of GPCC and CHIRPS data, as shown in Figure 2. We applied the same HMM framework to the precipitation data to evaluate whether rainfall exhibited a similar regime-switching behaviour to that of the low-water levels. This analysis aimed to determine if changes in precipitation could elicit a direct hydrological response in the basin. However, the analysis did not indicate any regime-switching behaviour in rainfall patterns (Figure S.6 and S.7).

To further investigate this point, we imposed the same low-water level classes derived from the multisite HMM onto the rainfall data and evaluated how precipitation varied across each low-water level regime. The Kernel density estimates provide a smoothed version of the histograms (Figure 7a), with the empirical probabilities of observing different classes and how they vary with the precipitation. The position of each estimate on the graph is set so that the empirical probabilities sum to one along the precipitation values. It can be observed that when precipitation was lower than 1,050 mm/year in a random year, the probability of observing a Dry year in the past (56%) was higher than the sum of the other two classes of low-water levels (24% and 20%, respectively). The probability of Dry years rapidly decreases with increasing precipitation. For instance, if precipitation was 1,150 mm/year in the past, the probability of observing a severe drought was reduced to 25%. With 1,200 mm/year, it reduced to values lower than 13%.

Figure 7: (a) Density kernels of the three low water level classes by annual precipitation; (b) violin plots of total annual precipitation for the three low-water level classes. Rainfall data from linear regression using GPCC data (1940 to Jan 1981) and CHIRPS (1981-onwards).

Precipitation during severe hydrological drought years (Class 1) is notably distinct from the other two classes, as illustrated in Figure 7(b). The mean annual precipitation during Dry years is estimated at 1,030 mm (±96 mm/year), which is significantly lower compared to

Normal years (1,193 ± 96 mm/year) and Wet years (1,219 ± 101 mm/year). The separation between Dry years and the other classes is well-defined, as confirmed by the statistical significance of the t-tests (detailed results of the t-tests, including mean values, standard deviations, and confidence intervals, are presented in the Supplementary Material). However, the distinction between Normal and Wet years is less clear, with overlapping precipitation ranges and no statistically significant difference observed between their mean values. This result suggests that while precipitation is a strong indicator of drought conditions, it may be less effective in differentiating between Normal and Wet years when considered in isolation. These findings emphasize the importance of exploring additional hydrological and climatological factors to enhance the classification of low-flow regimes.

Another relevant aspect of rainfall analysis is the persistence of precipitation in a specific drought regime. We evaluated the autocorrelation function of the precipitation time series and found no significant serial correlation (Figure S.8), contrary to what happens with water level data. This result suggests that persistency in a specific drought regime is not reproduced by precipitation data alone.

Finally, we examine the relationship between precipitation and the Dry, Normal, or Wet years derived from the low-water classes (Figure 8). Our analysis highlights the importance of considering the preceding year's drought regime when interpreting precipitation patterns and transitions between low-water level classes (Figure 8). The t-tests (Supplementary Material) reveal clear distinctions in precipitation totals depending on the initial hydrological state, improving the distinction presented in Figure 7.

For instance, when transitioning from a Dry year, precipitation required to shift to a Normal year is significantly higher (mean: 1,300 mm, $p < 0.001$) compared to remaining in a Dry regime (mean: 1,035 mm). This marked difference underscores the importance of substantial rainfall in breaking persistent drought conditions. Similarly, for transitions from Normal years, precipitation totals differ significantly when shifting to Wet years (mean: 1,267 mm) compared to remaining in a Normal regime (mean: 1,161 mm, $p = 0.048$). However, transitions from Wet to Normal years or vice versa show less pronounced distinctions, with

overlapping precipitation ranges and statistically non-significant differences ($p > 0.4$), suggesting greater variability in these transitions.

Figure 8: Water year cumulative precipitation and low-water level class transitions. The cumulative rainfall required to trigger a drought transition depends on the initial low-water level class. When the initial state is Wet (c), there is considerable uncertainty regarding the likelihood of transition (based on rainfall) to a Normal year or another Wet year. If the current year is Dry (a), the precipitation required to transition to a Normal year is higher than the precipitation required to transition from a Normal year to a Wet year (b).

Our analysis highlights that precipitation serves as a good predictor of drought regimes, particularly for identifying Dry years. However, rainfall alone does not fully explain the persistence of specific drought regimes. Hydrological features such as groundwater and lake storage play a crucial role in sustaining low-water level conditions, as they are intrinsic components of the basin's natural regulation system. This understanding of precipitation dynamics highlights the basin's intrinsic regulatory processes and reinforces the need to incorporate both internal and external factors into predictive models of low-water level regimes. This approach was implemented in this study using the MLR described in Section 3.1.3.

Concerning teleconnection patterns, weak relationships between low water level classes and teleconnection indices (ENSO34, PDO, and AMO) were observed (Supplementary Material – Figure S.9). For PDO, Wet years (Class 3) tend to align with positive values, while Dry and Normal years (Classes 1 and 2) are associated with negative phases. In contrast, AMO shows that extreme conditions (Dry and Wet years) occur predominantly under negative AMO values, whereas Normal years align with neutral conditions. These patterns suggest that PDO and AMO may modulate precipitation variability in the PRB, particularly during extreme events, but their influence appears buffered by the basin's internal hydrological processes such as groundwater storage and flow regulation.

Land-use changes are hypothesized to influence hydrological processes, potentially driving nonstationarity in river discharge and water levels through mechanisms such as infiltration, runoff, and evapotranspiration. Using land-use data from Dias et al. (2016), our preliminary analyses reveal that the relationships between land-use variables and minimum flows are complex and inconsistent. For example, while total cultivated area shows a positive correlation with minimum flows, soybean cultivation exhibits the opposite signal, suggesting the influence of additional confounding variables or non-linear interactions. These findings highlight the need for more detailed investigations to reliably assess the role of land-use changes in hydrological behaviour. Consequently, land-use data were excluded from the predictive models in this study, and the results are provided in the Supplementary Material for reference (Figure S.11 and S.12).

3.1.3 Nonhomogeneous Multinomial prediction model

Building on the above findings, we develop a nonhomogeneous MLR prediction model that uses precipitation and the current low-water level class as predictors. To assess the model's robustness and uncertainty, we derive confidence intervals for the precipitation thresholds by resampling the dataset 500 times while maintaining a homogenous distribution of low-water levels classes in the training and validation datasets. This bootstrap approach not only provides estimates of the thresholds but also evaluates the model's goodness-of-fit by quantifying the variability in the predictions. We find that the number of years in Dry, Normal, and Wet classes is 21, 35, and 29, respectively. As all these class counts approximate multiples of 7, we used 5/7 (71%) of the samples to train and 2/7 (29%) to validate the models. To preserve proportional representation, the training samples were balanced to reflect the original distribution of Dry, Normal, and Wet years in the time series (3/5/4).

The results provide the estimated transition probability as a function of the annual cumulative precipitation during dry years (Figure 9). As we fit the MLR model 500 times, the results show 500 lines that represent the probability of changing the low-water level class to a Normal year (in green) and the probability of maintaining a dry year based on the precipitation (in orange). The grey shading represents the 95% confidence area where the MLR fitted lines intersect. Its width reflects the uncertainty associated with the model's

predictions and serves as a measure of the model's goodness-of-fit. Narrower confidence intervals indicate higher confidence in the model's predictions, while wider intervals reflect greater uncertainty. In Dry conditions, the transition to a Normal year occurs if precipitation reaches 1,180 mm, with a narrow 95% confidence interval ranging from 1,120 to 1,240 mm. In this condition, the low dispersion of the lines across the 500 simulations (Figure 9) suggests that the model is robust and consistent in its predictions.

Figure 9: Results of the nonhomogeneous MLR model illustrating: (a) the probability of transition as a function of the total annual precipitation during dry years. In this class, the transition to a normal year happens if a precipitation of 1,221 mm occurs (with a narrow 95% confidence interval of 1,209 and 1,234 mm). If the current low-water level class is normal (b), the threshold for transitioning to a dry year is 1,037 mm (1,020-1,054 mm, 95% confidence interval), and to a wet year it is 1,477 mm, with a wider interval (1,257-1,697 mm, 95% CI). From a wet year (c) the width of the confidence interval is poorly defined (800-1,700 mm, 95% CI). The grey shading represents the 95% confidence area, reflecting the uncertainty in where the fitted MLR lines intersect.

The transition probabilities from a Normal year to a Dry or Wet year can be seen as follows in Figure 9b. The 500 lines indicate the probability of changing to a Dry year (in orange); to a Wet year (in blue); or of maintaining a Normal year (green). The width of the confidence intervals reflects the model's uncertainty in predicting these transitions. For example, the 95% confidence interval for transitioning to a Dry year spans from 1,013 to 1,070 mm (centred at 1,040 mm), while the interval for transitioning to a Wet year is wider, ranging from 1,257 to 1,340 mm (centred at 1,300 mm). This wider interval indicates greater uncertainty in predicting transitions to Wet years during a Normal year. If the current condition of the catchment is a Wet year (Figure 9c), the 95% confidence interval is poorly defined, ranging from 150 to 1,616 mm, reflecting the model's difficulty in predicting transitions under wet conditions. These results highlight the model's strengths (Dry years) and limitations (Wet years), providing a comprehensive assessment of its performance across different hydrological regimes.

The differences between the thresholds that predict a Dry year (hydrological drought) are worth noticing. If the current state is a Normal year, this threshold is 1,040 mm (spanning form 1,013 to 1,070 mm), but if the current state is a Dry year, the threshold is 1,180 (spanning from 1,209 to 1,234 mm), increasing by 140 mm. This precipitation volume can be

724 interpreted as compensating for the lack of moisture from one year to another. This result
725 corroborates Figure 8, where we observed a more evident separation between precipitation
726 within each low-water level class. By employing rainfall in a MLR model, we are able to
727 provide a more quantitative understanding of the effect of precipitation on the probability of
728 transition among low-water level classes. This procedure allows us to define specific
729 hydrological drought thresholds based on precipitation inputs.

730 **4 DISCUSSION**

731 Our analysis demonstrates that hidden states derived from the HMM framework provide
732 valuable insights into the temporal dynamics of annual minima in the PRB. The classification
733 into low-water level regimes revealed significant differences in both the expected values and
734 variability of yearly low-water levels, accompanied by nonhomogeneous transition
735 probabilities between classes. These findings underscore the nonstationary nature of the low-
736 water level time series, characterized by evolving statistical properties – such as changing
737 mean values and standard deviations – and the shift in drought-generating mechanisms over
738 time. This dynamic complexity poses significant challenges for traditional time series
739 analyses, requiring adaptive frameworks capable of capturing these transitions.

740 The low transition probabilities observed between extreme regimes (e.g., Dry to Wet years)
741 can be attributed to the hydrological characteristics of the PRB. The basin's interconnected
742 lakes and wetlands, typical of the Pantanal region, act as natural reservoirs with a pronounced
743 capacity for flow regularization (Assine et al., 2014; Pupim et al., 2017). These buffering
744 mechanisms mitigate abrupt transitions, emphasizing the basin's ability to dampen external
745 climatic variability and sustain low-flow regimes over extended periods. This natural
746 regulation likely explains the heightened persistence of extreme hydrological drought,
747 presenting both opportunities and challenges for drought management in the region.

748 The interplay between external climatic drivers and internal hydrological dynamics further
749 underscores the complexity of low-water level behaviour in the PRB. As Collischonn et al.
750 (2001) observed, precipitation patterns alone fail to fully account for flow regime variability
751 due to the significant role of storage and evapotranspiration dynamics. Our findings align

752 with this view, highlighting the need for predictive models that integrate both external drivers
753 and internal catchment processes. Such integrated approaches can improve our ability to
754 forecast hydrological responses and develop effective management strategies.

755 Future projections of hydrological behaviour in the PRB are challenging due to uncertainty of
756 climate models and the inadequate representation of land surface processes (Bravo et al.,
757 2014). However, existing studies point to rising temperatures, reduced rainfall, delayed
758 rainfall onset, and shorter rainy seasons as key drivers of future changes (Marengo et al.,
759 2016). These factors are expected to exacerbate water deficits, particularly in the central and
760 northern portions of the basin. Our findings corroborate these projections, suggesting that
761 hydrological droughts are likely to become more frequent in the future compared to the
762 1974–2018 period.

763 The results from Marengo et al. (2016) coupled with the findings of this study further
764 highlight a critical concern: the reduced likelihood of transitions from Dry to Normal years,
765 driven by the diminished probability of above-average rainfall capable of offsetting
766 cumulative deficits. This dynamic suggests that, once the basin enters a dry regime, it may
767 become "locked" in this state, with limited possibility of rapid recovery. Simulating future
768 scenarios that incorporate non-homogeneous thresholds conditioned on the basin's current
769 state could offer deeper insights into the challenges of escaping prolonged drought conditions
770 and inform adaptive management strategies tailored to mitigate these risks.

771 Wigley and Jones (1985) hypothesized that catchments with low runoff ratios are highly
772 sensitive to even minor changes in precipitation. While this hypothesis could partially apply
773 to the PRB, given its low runoff ratio of 0.11, the relatively low variability in basin-wide
774 precipitation suggest that this hypothesis is less applicable in the study area. With a
775 coefficient of variation of approximately 10%, significant deviations from mean rainfall
776 values (1,160 mm, ±120 mm) are uncommon. Consequently, internal hydrological processes,
777 such as groundwater and lake storage, tend to dominate the variability of low-water levels.
778 GRACE satellite data (Save et al., 2016) supports the importance of groundwater storage in
779 the basin, despite its temporal limitations (2002–2024), as it reveals a recent downward trend
780 in groundwater storage, which aligns with the intensification of drought conditions in recent

781 years (Figure S.10). While GRACE data remain underutilized in predictive models due to
782 their limited temporal coverage, their potential for enhancing our understanding of internal
783 basin dynamics warrants further exploration in large catchments.

784 The weak correlations between droughts in the PRB and large-scale teleconnection patterns,
785 such as ENSO, PDO, and AMO, further illustrate the complexity of climatic influences on
786 the PRB. Consistent with previous studies (Marengo et al., 2021; Thielen et al., 2020), our
787 findings indicate that these indices alone cannot reliably explain drought events. For instance,
788 while El Niño events are often associated with extreme hydrological phenomena in other
789 regions, the PRB experiences both droughts during La Niña and wet years during El Niño.
790 This lack of a consistent teleconnection signal highlights the need for refined analyses that
791 integrate combined SST indices and local hydrological conditions to improve predictability.

792 Our statistical simulations add another layer of insight by estimating the probability of the
793 unprecedented 46-year period without drought observed between 1974 and 2019. The results
794 suggest that such a prolonged absence of droughts is statistically unlikely to recur, given the
795 transition matrix and a zero-inflated Gaussian distribution, which estimate the probability of
796 occurrence at just 4.2%. These findings emphasize the increasing likelihood of hydrological
797 droughts due to reduced probabilities of precipitation exceeding critical thresholds, with
798 significant implications for water resource management.

799 While our framework provides insights into the nonstationary dynamics of low-water level
800 regimes in the PRB, uncertainties and limitations should be acknowledged. First, the
801 accuracy of our predictions is inherently limited by the quality and resolution of the input
802 data. For instance, while CHIRPS precipitation data offer broad spatial coverage, they may
803 not fully capture local variability, particularly in regions with complex topography or sparse
804 observational networks. Another important limitation is the difficulty of reconstructing
805 precipitation during severe drought periods, such as the 1960s. While the hydrological
806 droughts of that period are comparable in severity to those observed between 2020 and 2024,
807 the estimated precipitation values for years like 1964, 1967, and 1971 are not as low as those
808 estimated for more recent droughts. This discrepancy raises uncertainties about the
809 precipitation thresholds derived in this study, as they may not fully capture the climatic

conditions of past extreme events. A more dedicated effort to reconstruct past climate conditions, potentially using paleoclimatic data or advanced statistical techniques, could improve the accuracy of these thresholds. However, such an approach falls outside the scope of this study and represents a promising direction for future research.

Similarly, the limited temporal coverage of GRACE satellite data (2002–2024) restricts our ability to assess long-term trends in groundwater storage, which plays a critical role in the basin's hydrological dynamics. Although recent efforts have been made to reconstruct GRACE-like data for earlier periods (e.g., Palazzoli et at., 2025; Yin et al., 2023), reconstructions often struggle to accurately represent extreme hydrological events, raising uncertainties about their reliability for analysing past droughts.

Furthermore, the MLR framework, while effective in capturing nonstationary transitions, may struggle to account for nonlinear relationships and feedback mechanisms between variables. For instance, the effects of land use on hydrological behaviour are complex and may involve non-linear interactions that require further investigation. As shown in the Supplementary Material, land-use changes can influence hydrological responses, but their impacts are often difficult to quantify and may vary depending on local conditions. Given these complexities, land-use data were not included in the main predictive models to avoid potential misinterpretation of results. This limitation highlights the need for future studies to explore modelling alternatives, such as physically based hydrological models or machine learning techniques, to better capture the complexity of the PRB's hydrological system, including the role of land-use changes.

5 CONCLUSIONS

This study presents a new two-step modelling framework combining HMM and MLR to analyse and predict low-water level class transitions indicative of dry, normal and wet years. The approach was applied to the PRB, a region that has suffered from the worst droughts in its historical record (1900-2024) in 2024 and 2021, and therefore serves as an ideal testbed to demonstrate the potential of this framework for hydrological drought prediction. By linking a unified regional low-flow classification to precipitation, we derived quantitative

precipitation-based thresholds for hydrological drought, conditioned on the basin's current low-water level state. This approach provides valuable insights into the temporal dynamics of annual low-water minima and the mechanisms governing low-flow regimes, revealing the nonstationary nature of the time series and nonhomogeneous transition probabilities between classes. The integration of both external climatic drivers and the current low-water class improves the predictability of these events, offering a practical tool for monitoring and managing droughts.

Our findings underscore the significant role of internal drought-controlling mechanisms, such as groundwater and lake storage, in sustaining low-flow regimes and buffering external precipitation variability. While precipitation serves as a predictor of low-water level class transitions, internal processes dominate, introducing strong serial correlation in the low-water level time series and complicating transitions between low-water level classes. Despite the role of hydrological persistence, our analysis highlights that precipitation thresholds are critical for assessing regime transitions, especially in the context of rising temperatures, reduced rainfall, and delayed rainy seasons projected for the PRB. These climatic changes are reducing the likelihood of transitioning from Dry to Normal years, increasing the persistence of drought conditions (a particularly pressing issue given severe recent droughts in the region and especially the worst drought in recorded history observed in 2024).

The multistep MLR-HMM approach showcased here provides a robust framework for analysing nonstationary hydrological systems and predicting drought regimes under evolving climatic conditions. While limited by data constraints, such as the temporal coverage of GRACE satellite observations, the methodology demonstrates the value of integrating internal and external dynamics into predictive models. Future research could explore the integration of scenario-based simulations and state-dependent thresholds to address the challenges posed by prolonged drought persistence. By leveraging these advancements, researchers and practitioners can develop adaptive strategies to mitigate the socio-environmental risks associated with increasing drought frequency and persistence in the PRB and beyond.

6 ACKNOWLEDGEMENTS

We would like to express our sincere gratitude for the support and contributions provided by the Geological Survey of Brazil (SGB/CPRM) in enabling the development of this research. We extend our appreciation to Artur Matos, Head of the Early Warning Systems; Andrea Germano, the Head of the Department of Hydrology; and Alice Castilho, Director of Hydrology and Land Management at SGB/CPRM, for granting M.S. the autonomy to pursue research on the topic. Furthermore, we would like to thank all our colleagues at SGB/CPRM, particularly those involved in the maintenance and operation of the Brazilian Hydrometeorological Data Network. Their dedication and efforts have been instrumental not only to the success of this study but also to any other study that relies on hydrological data from Brazil. L.S. wishes to thank UKRI (MR/V022008/1), NERC (NE/S015728/1), and the Oxford Fell Fund.

7 DATA AVAILABILITY

The data used in this study are sourced from multiple datasets. The river level data was obtained from HidroWeb (ANA, 2017), which provides comprehensive hydrological information for water resources management at: http://www.snirh.gov.br/hidroweb. Rainfall data was acquired from the Climate Hazards InfraRed Precipitation with Stations (CHIRPS; https://data.chc.ucsb.edu/products/CHIRPS-2.0/) and the Global Precipitation Climatology Centre (GPCC; https://opendata.dwd.de/climate_environment/GPCC/html/download_gate.html). Information on land use was obtained from (Dias et al., 2016), which provides detailed land cover and land-use classification for the study area (http://www.biosfera.dea.ufv.br/pt-BR/banco/uso-do-solo-agricola-no-brasil-1940-2012---dias-et-al-2016). Sea surface temperature (SST) indices, including TNA, AMO, PDO, and ONI, were obtained from the National Oceanic and Atmospheric Administration (https://psl.noaa.gov) dataset. Details regarding the specific data subsets, processing methods, and data manipulation can be found in the references cited. The GRACE-FO RL06.3 Mascon Solutions (RL0603) data was accesses from the Center for Space Research from the University of Texas at Austin through the following link https://www2.csr.utexas.edu/grace/RL06_mascons.html.

8 REFERENCES

Akintuğ, B., Rasmussen, P.F., 2005. A Markov switching model for annual hydrologic time series. Water Resour. Res. 41, 1–10. https://doi.org/10.1029/2004WR003605

ANA, 2017. Hidroweb: Sistemas de informações hidrológicas [WWW Document]. Agência Nac. Águas. URL http://www.snirh.gov.br/hidroweb/%5Cnhttp://hidroweb.ana.gov.br/default.asp

Assine, M.L., Corradini, F.A., Pupim, F. do N., McGlue, M.M., 2014. Channel arrangements and depositional styles in the São Lourençoo fluvial megafan, Brazilian Pantanal wetland. Sediment. Geol. 301, 172–184. https://doi.org/10.1016/j.sedgeo.2013.11.007

Bishop, C.M., 2006. Pattern recognition and Machine Learning. Springer.

Bracken, C., Rajagopalan, B., Zagona, E., 2014. A hidden Markov model combined with climate indices for multidecadal streamflow simulation. Water Resour. Res. 50, 7836–7846. https://doi.org/10.1002/2014WR015567

Bravo, J.M., Collischonn, W., da Paz, A.R., Allasia, D., Domecq, F., 2014. Impact of projected climate change on hydrologic regime of the Upper Paraguay River basin. Clim. Change. https://doi.org/10.1007/s10584-013-0816-2

Brunner, M.I., Slater, L., Tallaksen, L.M., Clark, M., 2021. Challenges in modeling and predicting floods and droughts: A review. Wiley Interdiscip. Rev. Water 8, 1–32. https://doi.org/10.1002/wat2.1520

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference : a practical information-theoretic approach, 2nd ed. Springer-Verlag New York, New York.

Cavalcanti, I.F.A., Carril, A.F., Penalba, O.C., Grimm, A.M., Menéndez, C.G., Sanchez, E., Cherchi, A., Sörensson, A., Robledo, F., Rivera, J., Pántano, V., Bettolli, L.M., Zaninelli, P., Zamboni, L., Tedeschi, R.G., Dominguez, M., Ruscica, R., Flach, R., 2015. Precipitation extremes over La Plata Basin - Review and new results from

observations and climate simulations. J. Hydrol. 523, 211–230. https://doi.org/10.1016/j.jhydrol.2015.01.028

Collischonn, W., Tucci, C.E.M., Clarke, R.T., 2001. Further evidence of changes in the hydrological regime of the River Paraguay: Part of a wider phenomenon of climate change? J. Hydrol. 245, 218–238. https://doi.org/10.1016/S0022-1694(01)00348-1

Dias, L.C.P., Pimenta, F.M., Santos, A.B., Costa, M.H., Ladle, R.J., 2016. Patterns of land use, extensification, and intensification of Brazilian agriculture. Glob. Chang. Biol. 22, 2887–2903. https://doi.org/10.1111/gcb.13314

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. Sci. Data 2, 1–21. https://doi.org/10.1038/sdata.2015.66

Funk, C., Shukla, 2020. Drought early warning and forecasting theory and practice. Elsevier, https://doi.org/10.1016/B978-0-12-814011-6.00017-8.

Gelati, E., Madsen, H., Rosbjerg, D., 2010. Markov-switching model for nonstationary runoff conditioned on El Niño information. Water Resour. Res. 46, 1–13. https://doi.org/10.1029/2009WR007736

Gu, L., Yin, J., Slater, L.J., Chen, J., Do, H.X., Wang, H.M., Chen, L., Jiang, Z., Zhao, T., 2023. Intensification of global hydrological droughts under anthropogenic climate warming. Water Resour. Res. https://doi.org/10.1029/2022WR032997

Hagenlocher, M., Van Loon, A. F., Beek, T. J., Blauhut, V., Baldassarre, G. D., Dutra, E., ... & Mishra, V. (2023). Tackling growing drought risks: The need for a systemic perspective. Earth's Future, 11(8), e2023EF003019. https://doi.org/10.1029/2023EF003019

Hao, Z., Singh, V.P., Xia, Y., 2018. Seasonal drought prediction: advances, challenges, and future prospects. Rev. Geophys. 56, 108–141. https://doi.org/10.1002/2016RG000549

946 Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: Data
947 Mining, Inference, and Prediction, 2nd ed. Springer-Verlag New York.
948 https://doi.org/10.1007/b94608

949 Hughes, J.P., Guttorp, P., 1994. A class of stochastic models for relating synoptic
950 atmospheric patterns to regional hydrologic phenomena. Water Resour. Res. 30, 1535–
951 1546. https://doi.org/10.1029/93WR02983

952 Libonati, R., Geirinhas, J. o. L., Silva, P.S., Russo, A., Rodrigues, J.A., Belém, L.B.C.,
953 Nogueira, J., Roque, F.O., Dacamara, C.C., Nunes, A.M.B., Marengo, J.A., Trigo, R.M.,
954 2022. Assessing the role of compound drought and heatwave events on unprecedented
955 2020 wildfires in the Pantanal. Environ. Res. Lett. 17. https://doi.org/10.1088/1748-
956 9326/ac462e

957 Marengo, J.A., Alves, L.M., Torres, R.R., 2016. Regional climate change scenarios in the
958 Brazilian Pantanal watershed. Clim. Res. 68, 201–213. https://doi.org/10.3354/cr01324

959 Marengo, J.A., Cunha, A.P., Cuartas, L.A., Deusdará Leal, K.R., Broedel, E., Seluchi, M.E.,
960 Michelin, C.M., De Praga Baião, C.F., Chuchón Ângulo, E., Almeida, E.K.,
961 Kazmierczak, M.L., Mateus, N.P.A., Silva, R.C., Bender, F., 2021. Extreme Drought in
962 the Brazilian Pantanal in 2019–2020: Characterization, Causes, and Impacts. Front.
963 Water 3. https://doi.org/10.3389/frwa.2021.639204

964 Marengo, J.A., Liebmann, B., Grimm, A.M., Misra, V., Silva Dias, P.L., Cavalcanti, I.F.A.,
965 Carvalho, L.M.V., Berbery, E.H., Ambrizzi, T., Vera, C.S., Saulo, A.C., Nogues-Paegle,
966 J., Zipser, E., Seth, A., Alves, L.M., 2012. Recent developments on the South American
967 monsoon system. Int. J. Climatol. 32, 1–21. https://doi.org/10.1002/joc.2254

968 Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W.,
969 Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water
970 management? Science, 319(5863), 573-574. https://doi.org/10.1126/science.1151915

971 Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216.
972 https://doi.org/10.1016/j.jhydrol.2010.07.012

973 Montanari, A., & Koutsoyiannis, D. (2014). Modeling and mitigating natural hazards:
974 Stationarity is immortal! Water Resources Research, 50(12), 9748-9756.
975 https://doi.org/10.1002/2014WR016092

976 Naghettini, M., Pinto, E.J.D.A., 2007. Hidrologia Estatística, 1st ed. CPRM, Belo Horizonte.

977 Palazzoli, I., Ceola, S. & Gentine, P. GRAiCE: reconstructing terrestrial water storage
978 anomalies with recurrent neural networks. Sci Data 12, 146 (2025).
979 https://doi.org/10.1038/s41597-025-04403-3

980 Pupim, F. do N., Assine, M.L., Sawakuchi, A.O., 2017. Late Quaternary Cuiabá megafan,
981 Brazilian Pantanal: Channel patterns and paleoenvironmental changes. Quat. Int. 438,
982 108–125. https://doi.org/10.1016/j.quaint.2017.01.013

983 Rolim, L.Z.R., Filho, F. de A. de S., 2020. Shift detection in hydrological regimes and
984 pluriannual low-frequency streamflow forecasting using the hidden Markov model.
985 Water (Switzerland) 12. https://doi.org/10.3390/w12072058

986 Royston, P., 1995. Remark AS R94: A remark on algorithm AS 181: The W-test for
987 normality. Appl. Stat. https://doi.org/10.2307/2986146

988 Salas, J.D.., Obeysekera, J., Vogel, R.M., 2018. Techniques for assessing water infrastructure
989 for nonstationary extreme events: a review. Hydrol. Sci. J. 63, 325–352.
990 https://doi.org/10.1080/02626667.2018.1426858

991 Salas, J. D., Ramirez, J. A., Burlando, P., & Pielke, R. A. (2003). Stochastic simulation of
992 precipitation and streamflow processes. In T. D. Potter & B. R. Colman (Eds.), chapter
993 33 *Handbook of weather, climate, and water: Atmospheric chemistry, hydrology, and*
994 *societal impact* (Vol. 2, 607, pp. 607–640). New Jersey: John Wiley & Sons.

995 Santos, M.S., Henriques, M.P., Matos, A.J.S., 2021. Relatório de atividades: Departamento
996 de Hidrologia: Sistema de Alerta Hidrológico da bacia do rio Paraguai: relatório anual
997 de operação 2021. CPRM, Brasília.

Save, H., S. Bettadpur, and B.D. Tapley (2016), High resolution CSR GRACE RL05 mascons, J. Geophys. Res. Solid Earth, 121, doi:10.1002/2016JB013007.

Schneider, U., Becker, A., Finger, P., Rustemeier, E., Ziese, M., 2020. GPCC Full Data Monthly Product Version 2020 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data. GPCC Full Data Dly. Version 2020. https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025

Serviço Geológico do Brasil [CPRM]. (2024). Boletim de monitoramento da bacia do Rio Paraguai: Seca de 2024 (Boletim nº 17, 30 de outubro de 2024). Serviço Geológico do Brasil. https://www.sgb.gov.br/sace/boletins/Paraguai/20241030_17-20241030%20-%20173325.pdf

Slater, L.J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T., Murphy, C., Wilby, R.L., 2021. Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management. Hydrol. Earth Syst. Sci. 25, 3897–3935. https://doi.org/10.5194/hess-25-3897-2021

Smakhtin, V.U., 2001. Low flow hydrology: a review. J. Hydrol. 240, 147–186.

Sordo-Ward, A., Bejarano, M.D., Iglesias, A., Asenjo, V., Garrote, L., 2017. Analysis of current and future SPEI droughts in the La Plata Basin based on results from the regional Eta Climate Model. Water (Switzerland) 9. https://doi.org/10.3390/w9110857

Suchetana, B., Rajagopalan, B., Silverstein, J.A., 2019. Investigating regime shifts and the factors controlling Total Inorganic Nitrogen concentrations in treated wastewater using non-homogeneous Hidden Markov and multinomial logistic regression models. Sci. Total Environ. 646, 625–633. https://doi.org/10.1016/j.scitotenv.2018.07.194

Thielen, D., Schuchmann, K.L., Ramoni-Perazzi, P., Marquez, M., Rojas, W., Quintero, J.I., Marques, M.I., 2020. Quo vadis Pantanal? Expected precipitation extremes and drought dynamics from changing sea surface temperature. PLoS One 15, 1–25. https://doi.org/10.1371/journal.pone.0227437

Thyer, M., Kuczera, G., 2003. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 2. Real data analysis. J. Hydrol. 275, 27–48. https://doi.org/10.1016/S0022-1694(02)00411-0

Thyer, M., Kuczera, G., 2000. Modeling long-term persistence in hydroclimatic time series using a hidden state Markov Model. Water Resour. Res. 36, 3301–3310. https://doi.org/10.1029/2000WR900157

Tomas, W.M., Berlinck, C.N., Chiaravalloti, R.M., Faggioni, G.P., Strüssmann, C., Libonati, R., Abrahão, C.R., do Valle Alvarenga, G., de Faria Bacellar, A.E., de Queiroz Batista, F.R., Bornato, T.S., Camilo, A.R., Castedo, J., Fernando, A.M.E., de Freitas, G.O., Garcia, C.M., Gonçalves, H.S., de Freitas Guilherme, M.B., Layme, V.M.G., Lustosa, A.P.G., De Oliveira, A.C., da Rosa Oliveira, M., de Matos Martins Pereira, A., Rodrigues, J.A., Semedo, T.B.F., de Souza, R.A.D., Tortato, F.R., Viana, D.F.P., Vicente-Silva, L., Morato, R., 2021. Distance sampling surveys reveal 17 million vertebrates directly killed by the 2020's wildfires in the Pantanal, Brazil. Sci. Rep. 11, 1–8. https://doi.org/10.1038/s41598-021-02844-5

Slater, L. J., & Villarini, G. (2018). Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach. Geophysical Research Letters, 45, 6504–6513. https://doi.org/10.1029/2018GL077945

Tomas, W.M., de Oliveira Roque, F., Morato, R.G., Medici, P.E., Chiaravalloti, R.M., Tortato, F.R., Penha, J.M.F., Izzo, T.J., Garcia, L.C., Lourival, R.F.F., Girard, P., Albuquerque, N.R., Almeida-Gomes, M., Andrade, M.H. d. S., Araujo, F.A.S., Araujo, A.C., Arruda, E.C. d., Assunção, V.A., Battirola, L.D., Benites, M., Bolzan, F.P., Boock, J.C., Bortolotto, I.M., Brasil, M. da S., Camilo, A.R., Campos, Z., Carniello, M.A., Catella, A.C., Cheida, C.C., Crawshaw, P.G., Crispim, S.M.A., Junior, G.A.D., Desbiez, A.L.J., Dias, F.A., Eaton, D.P., Faggioni, G.P., Farinaccio, M.A., Fernandes, J.F.A., Ferreira, V.L., Fischer, E.A., Fragoso, C.E., Freitas, G.O., Galvani, F., Garcia, A.S., Garcia, C.M., Graciolli, G., Guariento, R.D., Guedes, N.M.R., Guerra, A., Herrera, H.M., Hoogesteijn, R., Ikeda, S.C., Juliano, R.S., Kantek, D.L.Z.K., Keuroghlian, A.,

Lacerda, A.C.R., Lacerda, A.L.R., Landeiro, V.L., Laps, R.R., Layme, V., Leimgruber, P., Rocha, F.L., Mamede, S., Marques, D.K.S., Marques, M.I., Mateus, L.A.F., Moraes, R.N., Moreira, T.A., Mourão, G.M., Nicola, R.D., Nogueira, D.G., Nunes, A.P., Nunes da Cunha, C. da, Oliveira, M.D., Oliveira, M.R., Paggi, G.M., Pellegrin, A.O., Pereira, G.M.F., Peres, I.A.H.F.S., Pinho, J.B., Pinto, J.O.P., Pott, A., Provete, D.B., dos Reis, V.D.A., dos Reis, L.K., Renaud, P.C., Ribeiro, D.B., Rossetto, O.C., Sabino, J., Rumiz, D., Salis, S.M., Santana, D.J., Santos, S.A., Sartori, Â.L., Sato, M., Schuchmann, K.L., Scremin-Dias, E., Seixas, G.H.F., Severo-Neto, F., Sigrist, M.R., Silva, A., Silva, C.J., Siqueira, A.L., Soriano, B.M.A., Sousa, L.M., Souza, F.L., Strussmann, C., Sugai, L.S.M., Tocantins, N., Urbanetz, C., Valente-Neto, F., Viana, D.P., Yanosky, A., Junk, W.J., 2019. Sustainability agenda for the Pantanal wetland: Perspectives on a collaborative interface for science, policy, and decision-making. Trop. Conserv. Sci. 12. https://doi.org/10.1177/1940082919872634

UNESCO, 2020. Pantanal Conservation Area [WWW Document].

Van Lanen, H.A.J., Wanders, N., Tallaksen, L.M., Van Loon, A.F., 2013. Hydrological drought across the world: Impact of climate and physical catchment structure. Hydrol. Earth Syst. Sci. 17, 1715–1732. https://doi.org/10.5194/hess-17-1715-2013

Van Loon, A.F., 2015. Hydrological drought explained. Wiley Interdiscip. Rev. Water 2, 359–392. https://doi.org/10.1002/wat2.1085

Venables WN, Ripley BD (2002). Modern applied statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Visser, I., Speekenbrink, M., 2010. depmixS4: An R package for hidden markov models. J. Stat. Softw. 36, 1–21. https://doi.org/10.18637/jss.v036.i07

Wigley, T.M.L., Jones, P.D., 1985. Influences of precipitation changes and direct CO2 effects on streamflow. Nature. https://doi.org/10.1038/314149a0

Yin, J., Slater, L. J., Khouakhi, A., Yu, L., Liu, P., Li, F., Pokhrel, Y., and Gentine, P.: GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from

1940 to present, Earth Syst. Sci. Data, 15, 5597–5615, https://doi.org/10.5194/essd-15-5597-2023, 2023.