Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Forecasting Severe Meteorological Droughts in the Central Amazon Basin: A Supervised Learning Approach

Matheus V. Cerqueira[1], Angelica N. Caseri[2], Francisco A. Rodrigues[3]
Instituto de Ciências Matemáticas e de Computação (ICMC) - Universidade de São Paulo, São Carlos, SP

This study presents an early warning system based on observational data to forecast meteorological drought conditions in the Central Amazon Basin (CAB) at seasonal lead times. Our approach leverages machine learning algorithms to capture teleconnection patterns between regional precipitation deficit and dominant modes of sea surface temperature (SST) variability.

The proposed methodology frames the problem as a supervised learning task for time series forecasting in which, given a temporal domain T, we define a target series (in this case, a drought index for the CAB) $\{Y_t\}_{t\in T}$ and a set of candidate predictor series represented by $\{\mathbf{X}_t\}_{t\in T}$.

Using gridded precipitation data from the Global Precipitation Climate Center [5], we computed the 12-month Standardized Precipitation Index (SPI-12) for each grid point within the CAB area (11.25°S to 1.25°N, 71.25°W to 53.75°W) for each month from October 1953 to December 2020. Based on these SPI-12 values, we derived a monthly time series representing the proportion of grid points experiencing severe drought (SPI-12 $<$ -1.5) in the CAB as a proxy for the drought affected area, following the methodology proposed by [4]. This series, referred to as the Drought Index, serves as the observed target variable $\{y_t\}_{t\in T}$ and can be verified in the right panel of Figure 1.

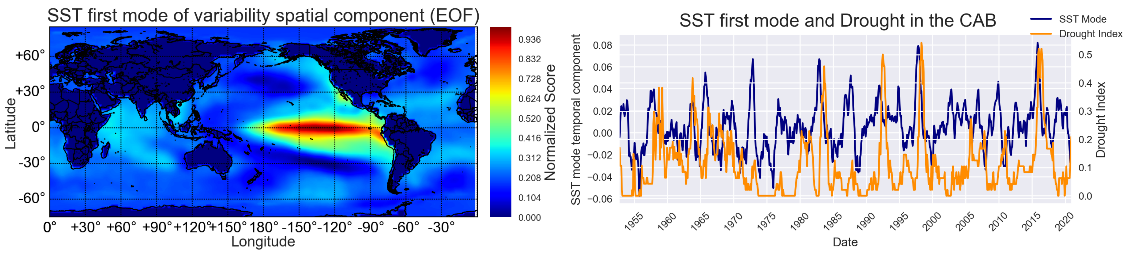

Figure 1: Spatial pattern (rotated EOF, left panel) and temporal signal (principal component, right panel) of the first SST mode of variability along with the observed DI series. Source: Authors.

To obtain candidate regressor variables, we applied rotated Empirical Orthogonal Function (EOF) analysis to the global SST data from the NOAA ERSST V5 project [3]. In this approach, Principal Component Analysis is employed to extract the dominant modes of SST variability. The obtained eigenvectors (EOFs)—representing spatial patterns of SST anomaly variability—are then rotated using the VARIMAX criterion to enhance interpretability. Subsequently, the original SST anomaly matrix is projected onto each rotated EOF, yielding corresponding time series (principal components) that quantify the temporal evolution of each mode [2]. These modes can be associated with known climate variability patterns, such as the El Niño-Southern Oscillation (ENSO),

[1]matheus.victal@usp.br
[2]angelica.caseri@gmail.com
[3]francisco@icmc.usp.br

based on their spatiotemporal characteristics. This association enables the analysis of the relationships between drought indices and the oscillations of these phenomena. Figure 1 displays the ENSO-related spatial pattern and temporal evolution of the mode that best explains SST anomaly variability. The derived principal components of these modes serve as candidate predictor series for our model, corresponding to the observed matrix $\{\mathbf{x}_t\}_{t\in T}$ in our supervised learning task.

Finally, by selecting lags from the DI series and the 30 most explanatory SST modes, a CatBoost model [1] was fitted to forecast DI values 6 months ahead, using 60% of the data for training and 40% for testing. The selection criterion was based on Pearson's correlation and Granger Causality relative to the DI values 6 months ahead. The forecast results of the final model, along with its performance in both the training and test sets evaluated by the RMSE and MAE metrics, are shown in Figure 2. This proof of concept for our proposed methodology yielded promising results in forecasting the DI of the CAB 6 months in advance, demonstrating that observational data-driven approaches can contribute to the anticipation of extreme events in this critical climatic hotspot.

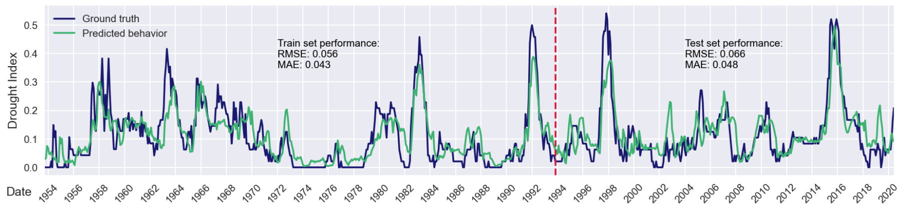

Figure 2: Results obtained from the fitted CatBoost model. The curves represent the observed and predicted DI values 6 months ahead of the reference date on the abscissa. Source: Authors

Acknowledgments

The authors would like to thank the São Paulo Research Foundation (FAPESP - n°2024/08278-2), CAPES and CNPq for providing funding for this research.

References

[1] A. V. Dorogush, V. Ershov, and A. Gulin. "CatBoost: gradient boosting with categorical features support". In: **arXiv preprint arXiv:1810.11363** (2018).

[2] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. "Empirical orthogonal functions and related techniques in atmospheric science: A review". In: **International Journal of Climatology: A Journal of the Royal Meteorological Society** 27.9 (2007), pp. 1119–1152. DOI: 10.1002/joc.1499.

[3] B. Huang et al. **NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5**. https://www.esrl.noaa.gov/psd/. (2017, Accessed: 20/10/2024). DOI: 10.7289/V5T72FNM.

[4] M. Noorbakhsh, C. Connaughton, and F. A. Rodrigues. "Discovering causal factors of drought in Ethiopia". In: **Proceedings of the 10th International Conference on Climate Informatics**. 2020, pp. 72–78. DOI: 10.1145/3429309.3429320.

[5] U. Schneider et al. **GPCC Full Data Monthly Product Version 2022 at 2.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data**. https://doi.org/10.5676/DWD_GPCC/FD_M_V2022_250. (2022, Accessed: 20/10/2024). DOI: 10.5676/DWD_GPCC/FD_M_V2022_250.