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.

Climate Clustering of Brazil Using Extreme Indices

Emanuel Bissiatti[1]
PUC-Rio, Rio de Janeiro, RJ
Reinaldo Marques[2], Paulo César[3]
UNIFAL, Alfenas, MG
Matheus Tavares[4]
INPE, Cachoeira Paulista, SP
Danilo Couto[5]
USP, São Paulo, SP

Brazil is a vast country with a diverse range of biomes and ecosystems. Due to this complexity, predicting the impact of extreme weather events in each region is challenging. Additionally, Brazil experiences a high frequency of extreme events, such as floods and droughts, which result from various climatic factors and have distinct effects on agriculture and society. To analyze the recurrence and intensity of these events, it is essential to cluster regions with similar historical patterns of extreme weather.

This work calculated 46 monthly extreme indices - like spei [2], mean wind, total of precipitation, txn (less max temperature of the month), etc. - for each grid of $0.1° \times 0.1°$ in Brazil, approximately 7700 points on your territory of 1961 to 2019. These indices are calculated using [4] dataset. To reduce the dimensionality and non-linear dependency of the data, Principal Component Analysis (PCA) is applied to the indices. After that, four strategies are used to cluster the regions: K-means on federal units with normalized original indices, k-means on federal units with PCA indices, k-means on hydrographic basins with normalized original indices and k-means on hydrographic basins with PCA indices [3]. The optimal number of clusters is found using the elbow method [1]. The results show that the number of clusters fluctuated between 5 and 8. The results of the clustering are shown in the Figure 1. Note that colors have no meaning or interpretation, they are used to differentiate the clusters.

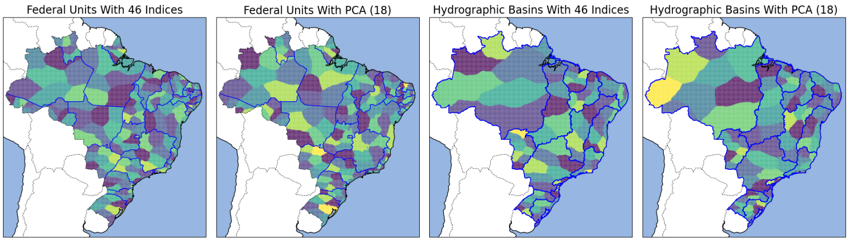
Federal Units With 46 Indices
Federal Units With PCA (18)
Hydrographic Basins With 46 Indices
Hydrographic Basins With PCA (18)

Figure 1: Clustering of Brazilian regions using extreme indices. Source: The author.

The dataset used in this study consist a 3D matrix with dimensions $7700 \times 46 \times 696$, where 7700 represents the number of points in Brazil, 46 corresponds to the number of indices and 696 represents the number of months (covering 58 years). Reshape this matrix to $5,359,200 \times 46$ and apply PCA to extract principal components of the extreme indices. 18 components are used to explain 90% of the variance of the data. After that, reshaped the matrix to $7,700 \times 12,528$.

[1] emanuelbissiatti@outlook.com
[2] reinaldo.marques@unifal.edu.br
[3] cesarmoraesdemenezespaulo@gmail.com
[4] matheus.tavares@inpe.br
[5] danilo.oceano@gmail.com

On final, k-means clustering was applied to each strategies in every subregion (federal units or hydrographic basins) to cluster the regions with similar behavior of extreme indices. The elbow method is applied in all subregion to find the optimal number of clusters. Because of this, the number of clusters fluctuated between 5 and 8, depending of the territory used. The elbow method select the number of clusters. While running this method, note that when the curve of the sum of squared distances is always smooth, the elbow is not clear. The state of Amazonas, biggest state in Brasil, has same number of cluster of your little neighbors like Acre, the convergence of this state is not clear. Please note, federal unit are political division and not climatic which makes climate clustering difficult. In contrast, hydrographic basins provide better regionalization since water drives climate, leading to more homogeneous climatic behavior within the same basin.

To validate these results, it is necessary to compare the number of extreme events in each cluster and their impact on agriculture and society. To achieve this, rural insurance data from the PSR (Rural Subsidy Program) on Paraná basin were used. This dataset contains most rural insurance contracts in Brazil for various types of natural disasters, and this basin was chosen because it has more data than the other basins. As shown in Figure 2, the proportion of natural disasters differs among clusters: the brown cluster has the highest occurrence of drought events and low rainfall, whereas the cyan cluster has more hail events than droughts. In this context, the clustering approach effectively distinguishes these types of insured losses. Additionally, most clusters exhibit a higher frequency of drought events compared to other types of disasters, as drought is the primary cause of agricultural losses in Brazil. According to PSR data, more than 50% of losses are attributed to drought.

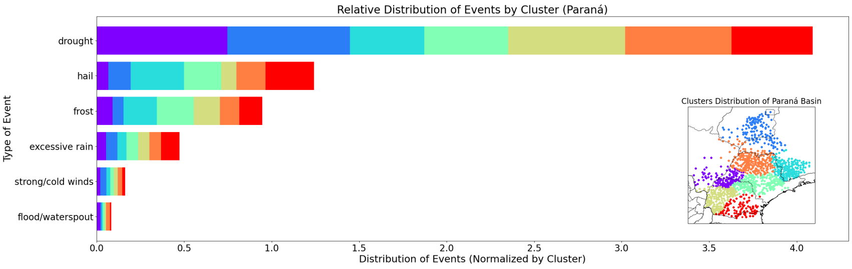

Figure 2: Relative crop insurance by clusters on Paraná basin. Source: The author.

References

[1] J. S. Ramírez et al. “Evaluation of Unsupervised Machine Learning Algorithms with Climate Data”. In: **Ingeniería y Desarrollo** 40.2 (2022), pp. 131–165. DOI: 10.14482/inde.40.02.622.553.

[2] S. M. Vicente-Serrano et al. “A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index”. In: **Journal of climate** 23.7 (2010). DOI: 10.1175/2009JCLI2909.1.

[3] J. Wang et al. “Exploring potential drivers of terrestrial water storage anomaly trends in the Yangtze River Basin (2002–2019)”. In: **Journal of Hydrology: Regional Studies** 58 (2025), p. 102264. DOI: 10.1016/j.ejrh.2025.102264.

[4] A. C. Xavier et al. “New improved Brazilian daily weather gridded data (1961–2020)”. In: **International Journal of Climatology** 42.16 (2022), pp. 8390–8404. DOI: 10.1002/joc.7731.