Predictive analysis of precipitation in South America using machine learning and deep learning techniques

Matheus C. Domingos, Valdivino A de Santiago Júnior, Elcio H. Shiguemori, Juliana A. Anochi

Laboratório de Inteligência Artificial para Aplicações AeroEspaciais e Ambientais (LIAREA)
Programa de Pós-Graduação em Computação Aplicada - CAP
Instituto Nacional de Pesquisas Espaciais - INPE

matheus.domingos@inpe.br, valdivino.santiago@inpe.br, elciohs@gmail.com, juliana.anochi@inpe.br

Introduction

Climate forecasts play an essential role in understanding and managing the impacts of global climate change. The atmosphere, composed of gases, particles and vapors, continuously interacts with the Earth's surface, generating a complex dynamic of energy and mass exchange. Traditional methods based on differential equations, which model atmospheric behavior based on boundary conditions, already offer valuable insights, but the search for greater precision is ongoing. In this work, deep learning and machine learning techniques, such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Convolutional Neural Network 1D (CNN-1D), Random Forest and XGBoost are applied with the aim of improving the quality of precipitation forecasts, integrating large volumes of data and capturing complex patterns that contribute to more accurate modeling of atmospheric behavior[2]. The developed climate forecast, focused on the 2019 autumn season, was compared with observed data from the GPCP (Global Precipitation Climatology Project), seeking a more accurate correspondence between what was predicted and what was observed. Results show that LSTM was the best model for the metrics of root mean squared error (RMSE), mean squared error (MSE) and coefficient of determination (R^2).

Objetive

This study aims to improve the accuracy of fall precipitation forecasts 2019 season, employing advanced deep learning and machine learning techniques. Put integrating large volumes of climate data and identifying complex patterns in the atmosphere dynamics, we seek to improve the modeling of atmospheric behavior. Comparing predictions with observed GPCP data seeks to evaluate forecast results and generate valuable insights to understand and manage environmental impacts.

Methodology

The data covers the period from January 1980 to March 2020 obtained by GPCP, covering historical climate measurements that allow detailed seasonal analysis [1]. In the figure 1, contains the variables used in the dataset.

Variables	*Variable Units*
Surface Pressure (surface)	millibars
Air Temperature (surface)	degC
Air Temperature at 850 hPa	degC
Specific Humidity at 850 hPa	grams/kg
Meridional wind component at 850 hPa	m/s
Zonal wind component at 500 hPa & m/s	m/s
Zonal wind component at 850 hPa	m/s
Precipitation	mm/day

Figure 1: *Variables composing the dataset*

To separate the seasons, the months of December, January and February (DJF) were considered as summer and the months of March, April and May (MAM) as autumn. The strategy adopted for forecasting is based on using the previous season to predict the next one, that is, summer data is used to predict autumn precipitation.

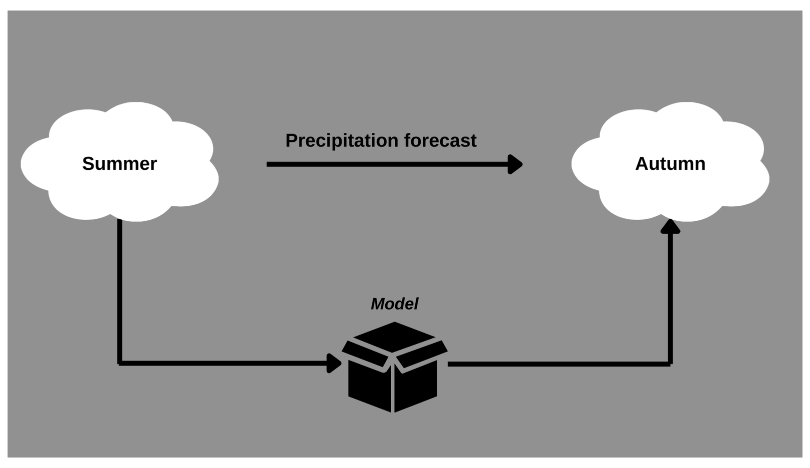


Figure 2: *Illustration of the methodology*

After separating the data by season, the datasets were divided into three parts:

- **75%** of the data was used to train the models;
- **25%** was allocated for validation;
- The test suite is reserved from the others, containing autumn 2019

The models were trained with historical data up to 2018. Subsequently, predictions were made and analyzed based on data for 2019, allowing an assessment of the effectiveness of the models in the context of climate prediction for subsequent seasons.

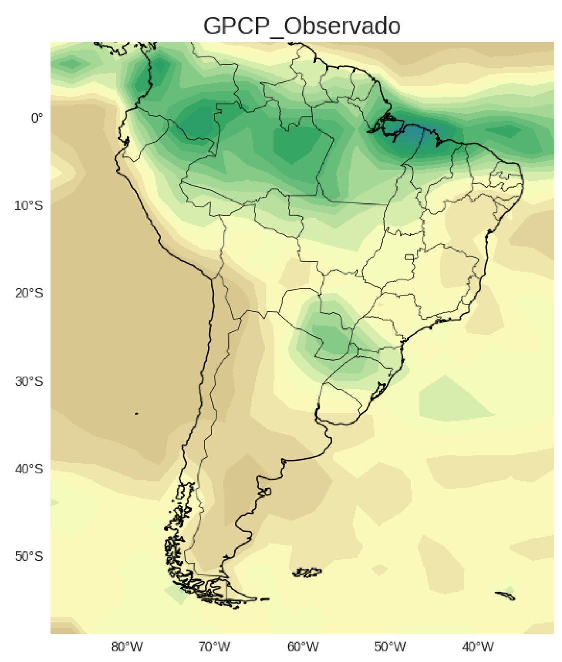
Results

In figure 3, we compare the performance of Deep Learning and Machine Learning models in predicting precipitation. Deep Learning models (LSTM, CNN1D, GRU) showed lower MSE and RMSE, with LSTM standing out with MSE of 0.89 and RMSE of 0.94, indicating greater accuracy. In contrast, Machine Learning models (Random Forest and XGBoost) showed slightly higher errors. The R^2 was also higher in the Deep Learning models, with the LSTM reaching 0.90, confirming its superiority in capturing atmospheric patterns.

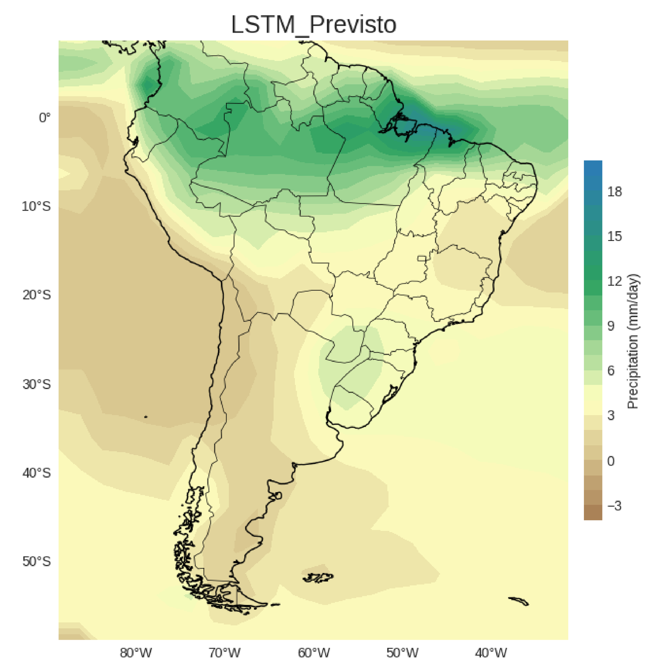
Metrics	*Deep Learning (LSTM)*	*Deep Learning (CNN1D)*	*Deep Learning (GRU)*	*Machine Learning (Random Forest)*	*Machine Learning (XGBoost)*
MSE	*0.89*	*0.98*	*01.02*	*1.15*	*1.14*
RMSE	*0.94*	*0.99*	*01.01*	*01.07*	*01.07*
R²	*0.90*	*0.89*	*0.88*	*0.87*	*0.86*

Figure 3: *Comparison of deep learning models (LSTM, CNN1D, and GRU) and machine learning models (Random Forest and XGBoost) based on performance metrics.*

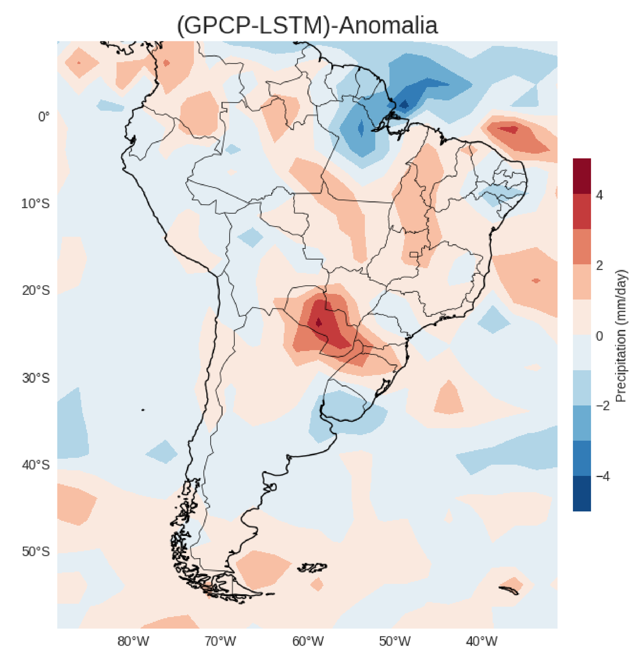
The following graphs present a detailed analysis of the anomaly, forecast, and result maps. The anomaly map highlights discrepancies between actual observations and expected values, while the forecast map shows the estimates generated by the models.



(a) GPCP Precipitation - Autumn



(b) Seasonal Climate Prediction for Autumn



(c) Error map for autumn forecast

Figure 4: *Results map*

The maps demonstrate positive results, with relatively low errors, indicating that the models used were effective in predicting precipitation. The accuracy of the forecasts is reflected in the reduced errors, highlighting the models' ability to successfully capture variations in precipitation.

Conclusions

Analysis of the results revealed that Deep Learning models, especially LSTM, significantly outperformed Machine Learning models in terms of accuracy in predicting precipitation. LSTM presented the lowest mean squared error (MSE) and root mean squared error (RMSE) values, as well as the highest coefficient of determination (R^2), highlighting its superior ability to capture complex atmospheric patterns and provide forecasts more accurate.

The maps generated corroborate these results, showing low errors and confirming the effectiveness of the models in predicting precipitation. This evidence highlights the relevance of advanced deep learning techniques in improving the accuracy of climate predictions. Using these approaches can offer valuable insights for water resources management, environmental planning and strategic decision making. The adoption of Deep Learning methods such as LSTM can, therefore, represent a significant advance in the analysis and prediction of climate events, contributing to better adaptation and response to variability in precipitation.

Acknowledgment

I would like to thank the National Institute for Space Research (INPE), the Coordination for the Improvement of Higher Education Personnel (CAPES) for their support and encouragement in carrying out this work.

References

[1] Robert F. Adler, Mathew R. P. Sapiano, George J. Huffman, Jian-Jian Wang, Guojun Gu, David Bolvin, Long Chiu, Udo Schneider, Andreas Becker, Eric Nelkin, Pingping Xie, Ralph Ferraro, and Dong-Bin Shin. The global precipitation climatology project (gpcp) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9(4), 2018.

[2] Vinicius Schmidt Monego, Juliana Aparecida Anochi, and Haroldo Fraga de Campos Velho. South america seasonal precipitation prediction by gradient-boosting machine-learning approach. *Atmosphere*, 13(2):243, 2022.