

Geosciences

Use of artificial intelligence for seasonal precipitation forecasting in the Minas Gerais state, Brazil

Uso de inteligência artificial para previsão de precipitação sazonal no estado de Minas Gerais, Brasil

Matheus José Gomes^I , Juliana Aparecida Anochi^{II} ,
Marília Harumi Shimizu^{II} 

^I Universidade Federal de Itajubá, Itajubá, MG, Brasil

^{II} Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP

ABSTRACT

Precipitation forecast is crucial for water resource management, agriculture, and climate disaster mitigation. This study investigates the use of artificial intelligence (AI) to predict seasonal precipitation in the state of Minas Gerais, Brazil. Utilizing neural networks trained on high temporal and spatial resolution datasets (GPCC, GPCP, MERGE-CPTEC, and CHIRPS), the research compares AI-based forecasts with numerical physics-based models. The results indicate that averaging multiple runs of each neural network model and using higher spatial resolution datasets, particularly CHIRPS, yield more accurate predictions, with lower Root Mean Square Errors (RMSE) and Relative Errors (RE). AI applications have shown promising results, offering forecasts with reduced computational processing time compared to traditional models. The study emphasizes the need to enhance computational infrastructure and explores the potential of advanced AI techniques to improve extreme climatic events forecasts, benefiting sectors such as agriculture and water resource management.

Keywords: Artificial intelligence; Seasonal precipitation; Forecast models

RESUMO

A previsão de precipitação é essencial para a gestão de recursos hídricos, agricultura e prevenção de desastres climáticos. Este estudo explora o uso de inteligência artificial (IA) para prever a precipitação sazonal no estado de Minas Gerais, Brasil. Utilizando redes neurais treinadas com dados de alta resolução temporal e espacial (GPCC, GPCP, MERGE-CPTEC e CHIRPS), a pesquisa compara a eficácia das previsões de IA com modelos numéricos de previsão físico-dinâmicos. Os resultados mostram que a média de múltiplas execuções de cada modelo de rede neural e o uso de dados de maior resolução espacial, especialmente do CHIRPS, oferecem previsões mais precisas, com menores erros quadráticos

médios (RMSE) e erros relativos (RE). A aplicação da IA mostrou resultados promissores, fornecendo previsões com menor tempo de processamento computacional em comparação com os modelos tradicionais. O estudo destaca a importância de aprimorar a infraestrutura computacional e o potencial de novas técnicas de IA para melhorar a previsão de eventos climáticos extremos, beneficiando setores como agricultura e gestão de recursos hídricos.

Palavras-chave: Inteligência artificial; Precipitação sazonal; Modelos de previsão

1 INTRODUCTION

Precipitation forecast plays a crucial role in water resource management, agriculture, and climate disaster mitigation. Extreme climatic events such as droughts and floods are becoming more frequent and intense due to climate change, necessitating improvements in forecasting methodologies to ensure socioeconomic safety (Funk et al., 2015). In Brazil, and particularly in Minas Gerais – a region transitioning between tropical and subtropical climates – seasonal forecasts face challenges, partly due to the climate variability influenced by phenomena such as the El Niño-Southern Oscillation (ENSO), which have moderate effects in the region (Minuzzi, 2005).

Traditional climate forecast approaches, which rely on physics-based dynamical models, use physical equations to describe atmospheric processes (Lam, 2023). Although these models have made progress, they still have limitations in spatial and temporal accuracy and require high computational capacity to process large data volumes. In this context, artificial intelligence (AI) emerges as a promising alternative. AI models, such as neural networks, can handle large datasets, identify complex patterns, and generate forecasts more quickly (Anochi et al., 2021).

Renowned meteorological organizations, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) and the United Kingdom Met Office, have adopted AI techniques to improve the accuracy of climate forecasts, showing encouraging results in handling nonlinear data and enhancing forecast precision (Conti,

2024). In this study, we propose the use of a neural network for seasonal precipitation forecasting in the state of Minas Gerais, aiming to explore AI's potential to overcome some limitations of traditional physics-based methods. These limitations include the high computational cost required to process large data volumes, the limited spatial and temporal resolution of forecasts, and the difficulty of representing complex nonlinear relationships in the atmosphere that can be more effectively captured by AI models trained on historical data.

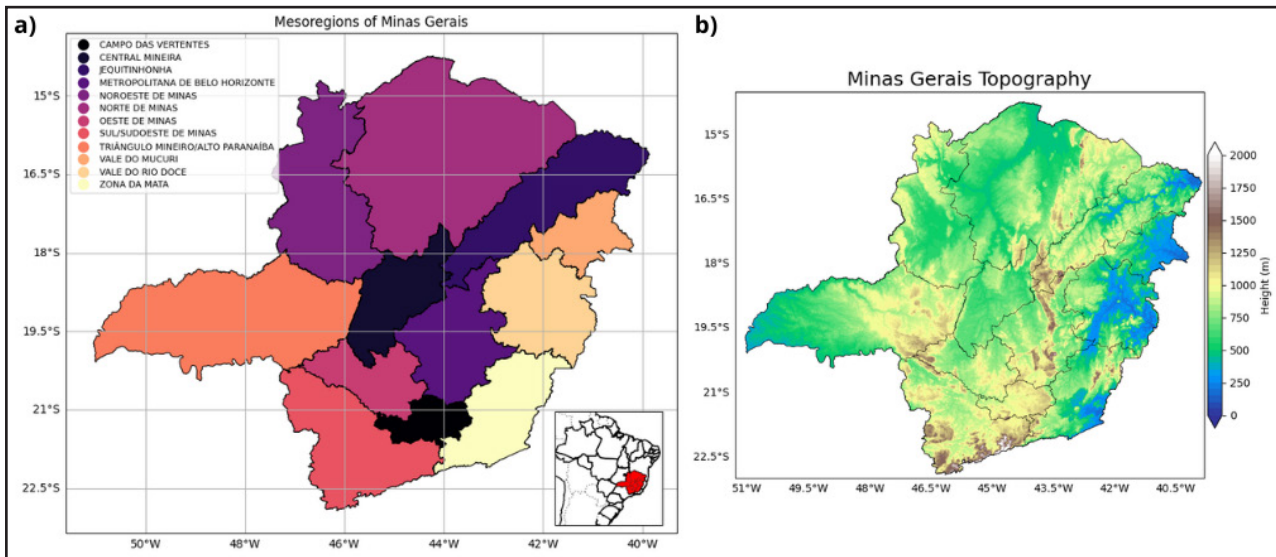
This paper is structured as follows: Section 2 describes the data used and the applied methodology; Section 3 presents the results obtained with the AI models; and finally, Section 4 discusses the main findings and limitations.

2 MATERIAL AND METHODS

2.1 Study area

The state of Minas Gerais, located in the southeast of Brazil, is known for its vast territorial expanse and a wide diversity of landscapes and climates. With an area of approximately 586,521 km², Minas Gerais borders the states of Bahia, Goiás, Mato Grosso do Sul, Espírito Santo, São Paulo, and Rio de Janeiro (IBGE, 2022). This geographical diversity results in significant variations in altitude, ranging from lower areas in the northeast and west of the state to mountainous regions in the south, such as the Serra da Mantiqueira, where peaks can reach around 2,800 meters in elevation (Figure 1.b).

Figure 1 – Mesoregions and Topography of the State of Minas Gerais



Source: Authors (2024). Caption: a) Mesoregions of the state of Minas Gerais according to IBGE (1990). b) Topography of the State of Minas Gerais in meters based on ETOPO 2022 topography data (Amante et al., 2009) with a spatial resolution of $0.0167^\circ \times 0.0167^\circ$

The mountainous topography plays a significant role in the precipitation pattern of the state. The central and southern regions of Minas, with higher elevations, tend to receive more rainfall due to orographic phenomena that force the ascent of moist air, resulting in precipitation. In contrast, the northern and western regions, with lower altitudes, experience drier climatic variations and receive less precipitation.

Minas Gerais is situated between tropical and subtropical climate zones, which leads to considerable variability in climatic conditions throughout the year. This pattern is influenced by global climate phenomena, such as the El Niño-Southern Oscillation (ENSO), which can cause periods of prolonged drought or intense rainfall, impacting both agriculture and water resources.

Precipitation forecast is particularly relevant in Minas Gerais, as economic activities like agriculture, mining, and hydroelectric energy production are highly dependent on stable climatic conditions. Analyzing and forecasting precipitation in the region has direct implications for water resource management and the mitigation of natural disasters, such as floods and severe droughts.

2.2 Data

The precipitation data used in this study were obtained from four main sources, covering different spatial and temporal resolutions, as listed below:

- Global Precipitation Climatology Centre (GPCC): monthly data with a spatial resolution of $1.0^{\circ} \times 1.0^{\circ}$ (Huffman et al., 2019) from 1981 to 2022.
- Global Precipitation Climatology Project (GPCP): monthly data with a spatial resolution of $0.5^{\circ} \times 0.5^{\circ}$ from 1981 to 2022 (Adler et al., 2003).
- MERGE-CPTEC: monthly data with a spatial resolution of $0.1^{\circ} \times 0.1^{\circ}$ (Rozante et al., 2010), available for 2001 to 2020, a product generated and provided by the Center for Weather Forecast and Climate Studies (CPTEC), attributed to the National Institute for Space Research (INPE).
- Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS): monthly data with a spatial resolution of $0.05^{\circ} \times 0.05^{\circ}$ (Funk et al., 2015), covering the period from 1981 to 2022.

All precipitation values were expressed as seasonal totals or anomalies, representing the accumulated precipitation for the three-month period in millimeters per season (mm/season).

For comparative purposes between the AI-generated forecasts and those obtained from physics-based dynamical models, we also used seasonal forecasts from the North American Multi-Model Ensemble (NMME), with a spatial resolution of $1.0^{\circ} \times 1.0^{\circ}$ (Kirtman et al., 2014), from December 2021 to February 2022.

2.3 Methods

2.3.1 Analysis of Precipitation Data over the State of Minas Gerais

After acquiring the precipitation data, each dataset's seasonal climatology was calculated from 1991 to 2020 to establish a reference pattern, except for the

MERGE-CPTEC dataset, in which the 2001-2020 period was considered. Based on this baseline, seasonal precipitation anomalies were determined, where positive and negative values indicate periods with precipitation above and below the climatological average, respectively.

This analysis was essential to verify the consistency in the representation of precipitation across the different datasets, ensuring the reliability of the results before training the AI model. Consistency is crucial because data from different sources, such as weather stations, reanalysis, or satellites, can show variations in their measurements due to differences in collection methods, spatial and temporal resolutions, or algorithms used to estimate precipitation.

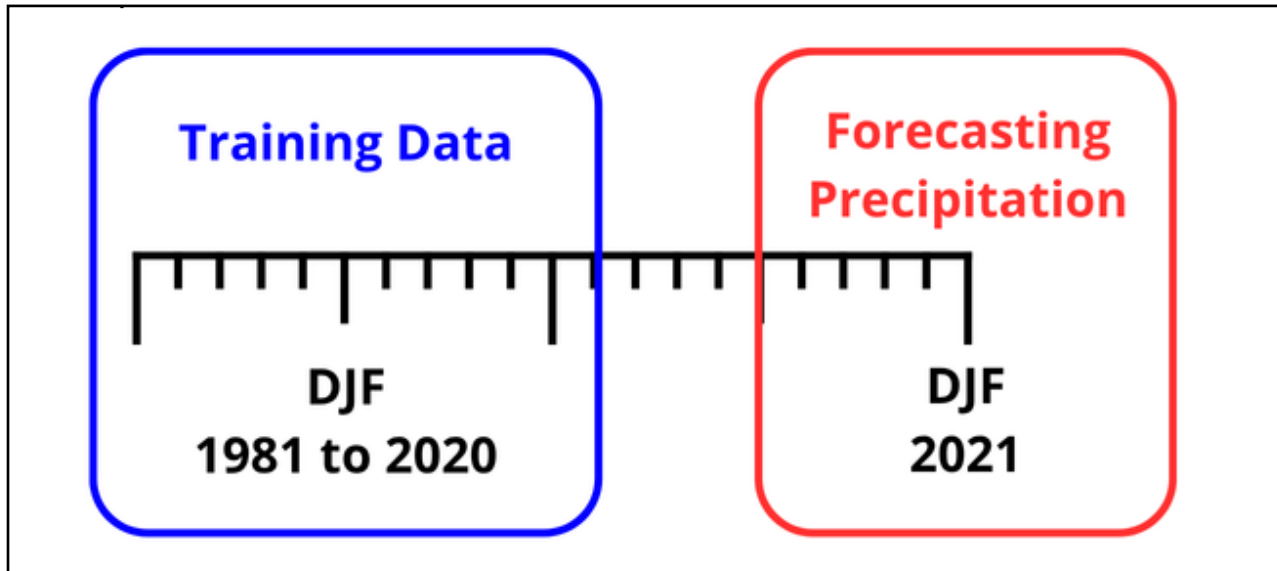
2.3.2 Training and validation of AI Models

After acquiring, processing, and visualizing the data, the neural network was trained to construct the forecasting models. The training and validation data consist of precipitation values for December, January, and February (DJF) from 1981 to 2020, except for the MERGE-CPTEC dataset, in which the 2001-2020 period was considered, according to data availability (Figure 2). Each dataset was divided into 90% for training and 10% for validation to ensure model generalization. After the training phase, the neural networks were applied to forecast precipitation for the DJF 2021/2022 season, which was reserved exclusively as an independent test period to evaluate the models' predictive performance.

Although this study primarily focused on the DJF season, which corresponds to the main rainy period over Minas Gerais, additional tests were also conducted for the MAM, JJA, and SON seasons. However, since these seasons present substantially lower precipitation totals and less pronounced spatial variability, the performance of the AI models was more difficult to evaluate and less relevant for hydrometeorological applications. Therefore, the DJF season was selected for detailed presentation and

discussion, as it represents the most critical period for water resource management and agriculture in the state.

Figure 2 – Temporal scheme of the data used in the neural network



Source: Authors (2024). Caption: The neural network was trained using precipitation data for the DJF (December–January–February) season from 1981 to 2020 (shown in blue). After training, the model was used to forecast precipitation for the DJF season of 2021 (shown in red)

Four AI models were developed, as summarized in Frame 1. Although the number of temporal samples (41 annual seasons) is limited, each sample consists of a spatial field containing hundreds to thousands of grid points. Therefore, the model learns not only from the temporal evolution but also from the complex spatial precipitation patterns and their interrelationships within each year. This substantially increases the effective number of training instances. To further improve model generalization, 90% of the data was used for training and 10% for validation, and a set of 20 runs was generated for each dataset to reduce random initialization effects and stabilize results.

Frame 1 – Parameters of the AI models implemented in this study

Model name	Training Data	Resolution	Training Period	Number of Epochs	Number of runs	Execution Time (Hours)
RN - GPCC	Global Precipitation Climatology Centre (GPCC)	1.0°x1.0°	1981 to 2020	50	20	0.75
RN - GPCP	Global Precipitation Climatology Project (GPCP)	0.5°x0.5°	1983 to 2020	50	20	1.5
RN - MERGE/CPTEC	MERGE/CPTEC	0.1°x0.1°	2001 to 2020	50	20	15
RN - CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)	0.05°x0.05°	1981 to 2020	50	20	110

Source: Authors' organization

The AI model adopted in this study is a feed-forward fully connected neural network, also known as a multilayer perceptron (MLP). The MLP structure allows each grid point to contribute to the learning of large-scale spatial-temporal relationships, which are essential for seasonal forecasting. The AI model training was implemented using the TensorFlow library in Python, a widely used tool for large-scale numerical computation and machine learning. The choice of this Machine Learning methodology was based on the study by Anochi and co-authors (2021). The neural network model is structured in layers, through which the input data is processed by multiple neurons until reaching the output layer. During training, the data is repeatedly processed through the network, adjusting the internal weights to optimize the model's accuracy. This continuous adjustment process is known as an "epoch."

The initial network configuration followed the structure proposed by Anochi et al. (2021) for seasonal precipitation forecasting, with two hidden layers and 25 neurons in each. Each neural network model was trained using the Adam optimizer to minimize the Mean Square Error (MSE) (Marmolin, 1986), defined in Equation (1). The activation function used was the Rectified Linear Unit (ReLU), which is widely employed in neural networks due to its similarity to the functioning of the human brain (Rizzo et al., 2020).

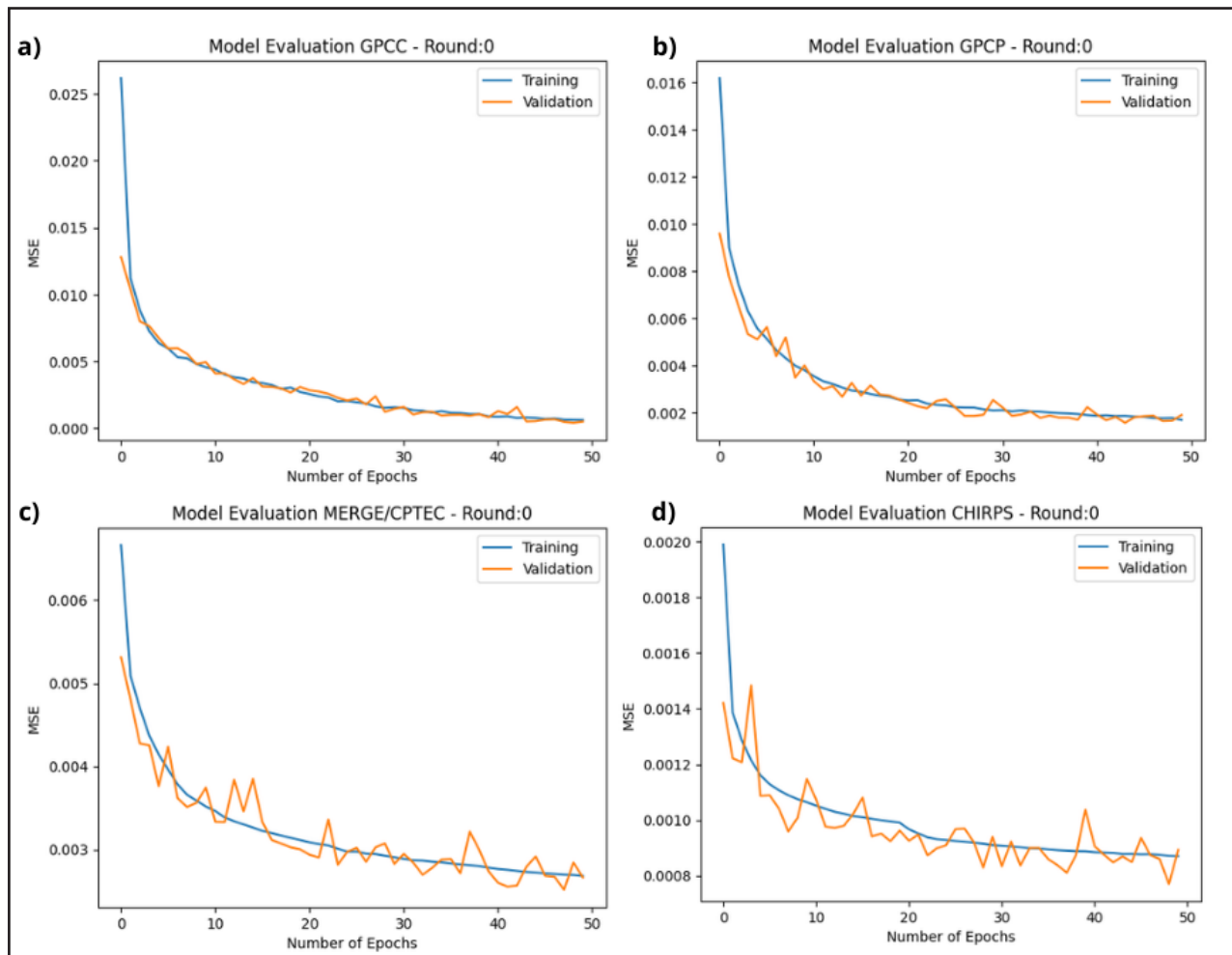
$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2 \quad (1)$$

where: n is the total number of grid points (pixels) in the analyzed spatial domain; x is the observed data; and y is the model's prediction

A sensitivity analysis was conducted by varying the number of hidden layers (2–3), neurons (20–50 per layer), and epochs (30–50), resulting in ten variations of the initial model. The final configuration (two hidden layers with 50 neurons and 50 epochs) showed the most stable convergence behavior. The number of 50 epochs was selected based on the convergence of the loss function, as additional iterations produced no significant improvement in accuracy.

For each of the four models (RN-GPCC, RN-GPCP, RN – MERGE/CPTEC, RN – CHIRPS), we performed 20 training runs to reduce the variability associated with the random initialization of neural network weights. The final forecast of each model corresponds to the average of these 20 runs, which helps stabilize the results and mitigate random errors present in individual runs. After each set of training runs, the MSE over the iterations was calculated, as defined in Equation 1, and illustrated in Figure 3.

Figure 3 – Evaluation of the first round of the models, based on the MSE for training and validation data



Source: Authors (2024). Caption: a) MSE of the first round for RN-GPCC model; b) MSE of the first round for RN-GPCP model; c) MSE of the first round for RN-MERGE/CPTEC model; d) MSE of the first round for RN-CHIRPS model

Additionally, we evaluated the execution time for each model, from the initial processing to the final plotting. The execution time was directly proportional to the data resolution. The RN-GPCP was the fastest model, followed by RN-GPCC. Due to the higherspatial resolution, RN-MERGE-CPTEC and RN-CHIRPS showed significantly higher execution times, as can be seen in Frame 1.

2.3.3 Evaluation of the Predictions by AI Models

After training the AI models, the precipitation forecast for each AI model was generated for the period between December 2021 and February 2022, referred to as DJF/2022. The evaluation of these predictions was conducted using the observed data for the same period, employing metrics such as bias (Eq. 2) and the Root Mean Square Error (RMSE) (Chai & Rexler, 2014) (Eq. 3).

$$bias = \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - x_i)^2}{n}} \quad (3)$$

The comparison between the forecasts of each model was also made by calculating the Relative Error (RE), as defined by Gleckler et al. (2008) (Eq. 4). The Relative Error (RE) provides a normalized metric to compare the performance of each model against the “typical” behavior of the entire set of models. While the RMSE quantifies the absolute error magnitude, the RE expresses the deviation of a model’s RMSE from the median RMSE of all models, thereby indicating whether the model performs above or below the collective average. A negative RE value (close to -1) indicates better-than-average performance, whereas a positive RE indicates higher errors relative to the median. This metric allows a direct and consistent comparison of relative forecast performance across models trained with different datasets and spatial resolutions. The results for RE and RMSE were represented in a heatmap for each mesoregion of the state of Minas Gerais.

$$RE = \frac{E - \bar{E}}{\bar{E}} \quad (4)$$

where: E is the RMSE of the predicted data, and $\bar{E}$ is the median of all RMSEs of the model forecasts

Finally, the precipitation forecasted by the AI models was categorized for each mesoregion to identify rainy, dry and normal rainfall patterns. In this study, the terms

Rainy, Dry and Normal do not represent statistical extremes in the strict sense but rather relative categories based on the distribution of precipitation for each grid point. The classification was derived from the upper and lower quartiles (75th and 25th percentiles) of the climatological record, following the methodology proposed by Shimizu et al. (2017) for South America. Therefore, the categories identify regions with precipitation substantially above or below the climatological mean, without implying the occurrence of extreme events as defined by higher percentile thresholds (e.g., 95th or 99th percentiles). The forecasted precipitation was classified into three categories: Above Normal (corresponding to the upper quartile), Below Normal (corresponding to the lower quartile), and Normal.

3 RESULTS

3.1 Observed Precipitation Patterns

Figure 3 shows the observed precipitation anomalies for DJF/2022 for each dataset. The anomalies for all datasets were predominantly positive compared to the climatological average. For GPCC (Figure 4.a), higher wet anomalies are concentrated in the center of the state (Metropolitana de Belo Horizonte mesoregion), as well as in GPCP (Figure 4.b). In MERGE/CPTEC (Figure 4.c) and CHIRPS (Figure 4.d), the peak of positive anomalies is shifted to the west compared to the other datasets, with the highest values occurring in the Central Mineira and Triângulo Mineiro regions, respectively.

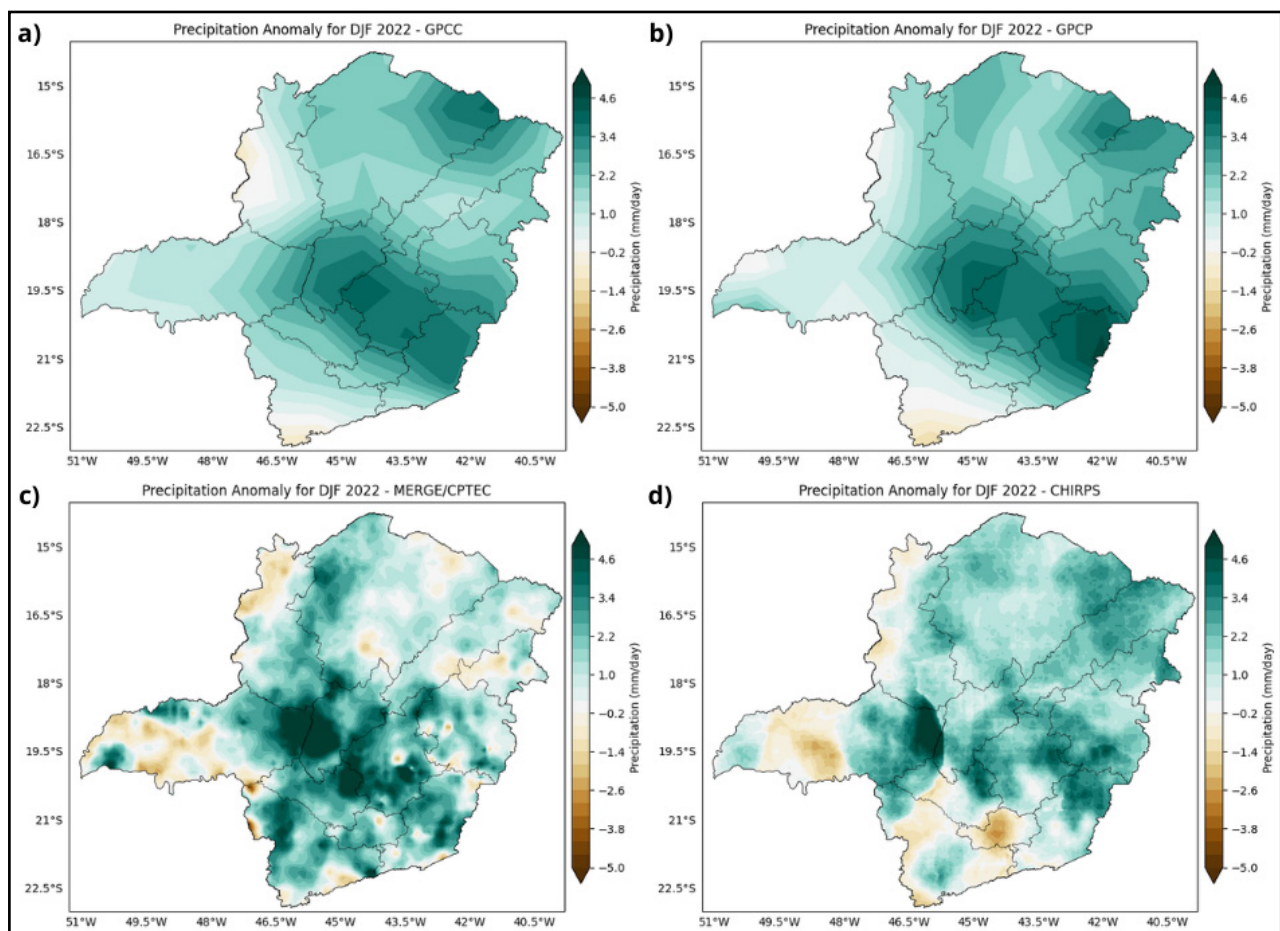
In general, GPCC and GPCP display smoother, more generalized patterns with milder anomalies, likely due to coarser resolution, while MERGE-CPTEC and CHIRPS exhibit finer spatial detail with more distinct positive and negative anomalies, capturing localized variability and extreme values more effectively.

The differences in the visual smoothness of the anomaly maps can be explained by the spatial resolution and data composition of each dataset. The GPCC and GPCP products, with coarser resolutions ($1.0^{\circ} \times 1.0^{\circ}$ and $0.5^{\circ} \times 0.5^{\circ}$), tend to produce

smoother and more generalized precipitation patterns. In contrast, the MERGE/CPTEC and CHIRPS datasets have much finer spatial resolutions ($0.1^\circ \times 0.1^\circ$ and $0.05^\circ \times 0.05^\circ$) and incorporate satellite-based estimates combined with a denser network of surface stations. As a result, they capture localized variations and small-scale features more effectively, which explains the coexistence of positive and negative anomalies over short spatial distances in panels (c) and (d).

These results align with previous studies that identified significant precipitation anomalies in 2022 (Marengo et al., 2023). Due to the magnitude and impact of these anomalies, the AI models were used to predict 2022 precipitation patterns, focusing on the rainy season, which spans from December 2021 to February 2022.

Figure 4 – Precipitation anomaly for DJF 2022



Source: Authors (2024). Caption: a) GPCP anomalies; b) GPCP anomalies; c) MERGE/CPTEC anomalies; d) CHIRPS anomalies. Positive values indicate above-average seasonal totals and negative values indicate below-average totals relative to the climatology

3.2 Precipitation Forecast Results by AI Models

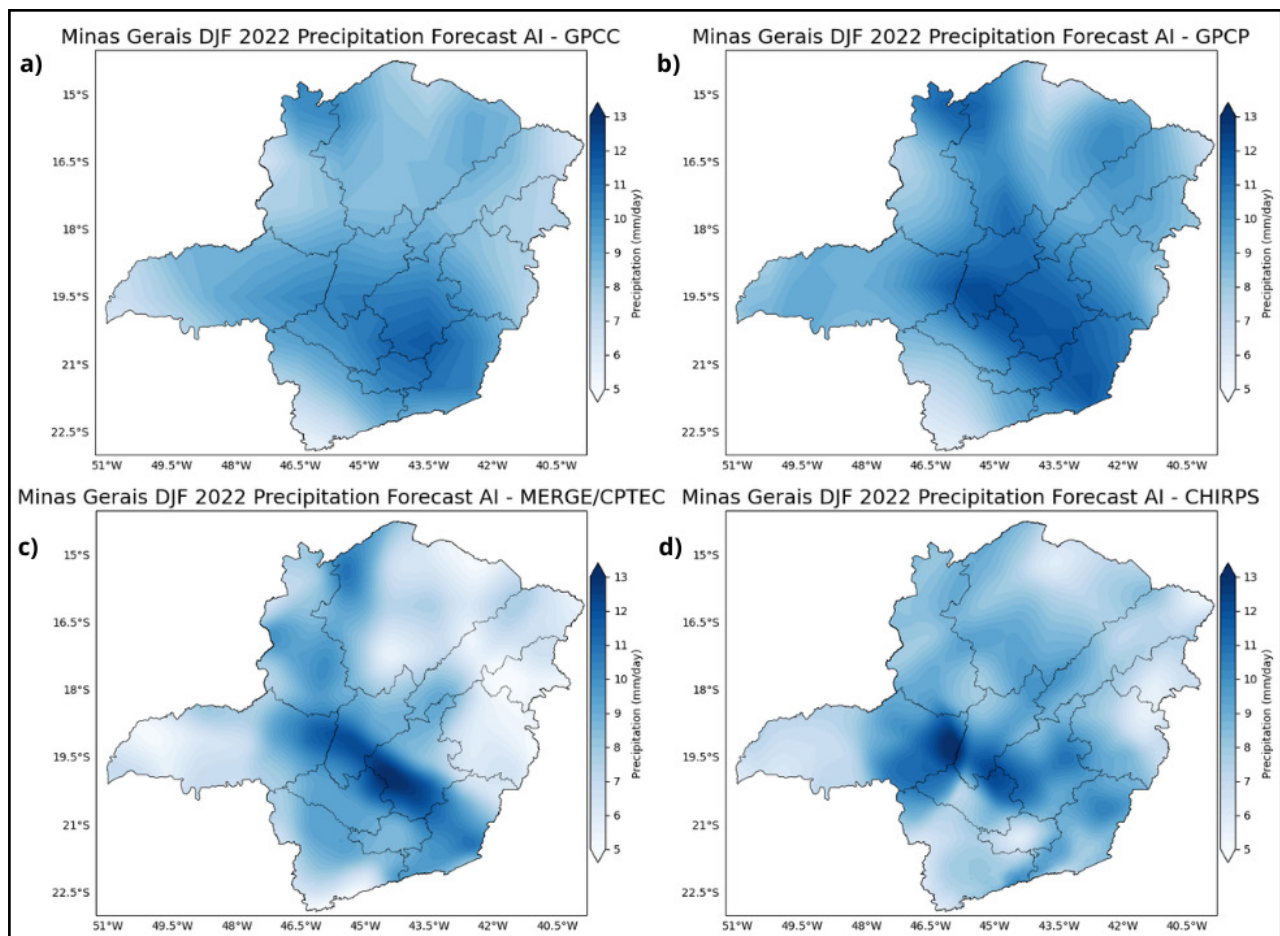
Figure 4 shows the precipitation forecasts from the AI set of models, where each result represents the mean output of multiple training runs conducted for the same precipitation dataset. Each neural network was trained 20 times with different random weight initializations, and the ensemble mean of these runs was adopted as the final forecast to minimize internal bias and random variability. It is important to emphasize that no averaging or combination was performed across the four models trained with distinct datasets (GPCC, GPCP, MERGE/CPTEC, and CHIRPS). The purpose of this comparison is exclusively to assess how the characteristics and spatial resolutions of the input datasets influence the forecasting performance of the AI models.

The RN-GPCC model (Figure 5.a) indicates highest precipitation values over central Minas Geras (Metropolitana de Belo Horizonte region), as well as the RN-MERGE/CPTEC (Figure 5.c). The RN-GPCP (Figure 5.b) shows slight higher precipitation values than RN-GPCC. Meanwhile, the RN-CHIRPS (Figure 5.d) shows highest rainfall to the west of the state (Triângulo Mineiro/Alto Paranaíba region).

Figure 6 shows the bias between observed and the set of AI models forecasted precipitation for DJF 2022. Precipitation forecast for the RN-GPCC (Figure 6.a) has the lowest error, followed by RN-GPCP (Figure 6.b). RN-GPCC and RN-GPCP display relatively low spatial variability in bias, with broader, more generalized patches of positive and negative biases. This suggests that the set of AI models performed more consistently but may have limited sensitivity to finer spatial variations in precipitation. Meanwhile, RN-MERGE/CPTEC (Figure 6.c) shows the largest errors, both positive and negative, with the highest errors occurring where the highest precipitation was forecasted (Figure 4.c). The forecast using CHIRPS data, represented by RN-CHIRPS (Figure 5.d), has a lower error compared to RN-MERGE/CPTEC. RN-MERGE-CPTEC

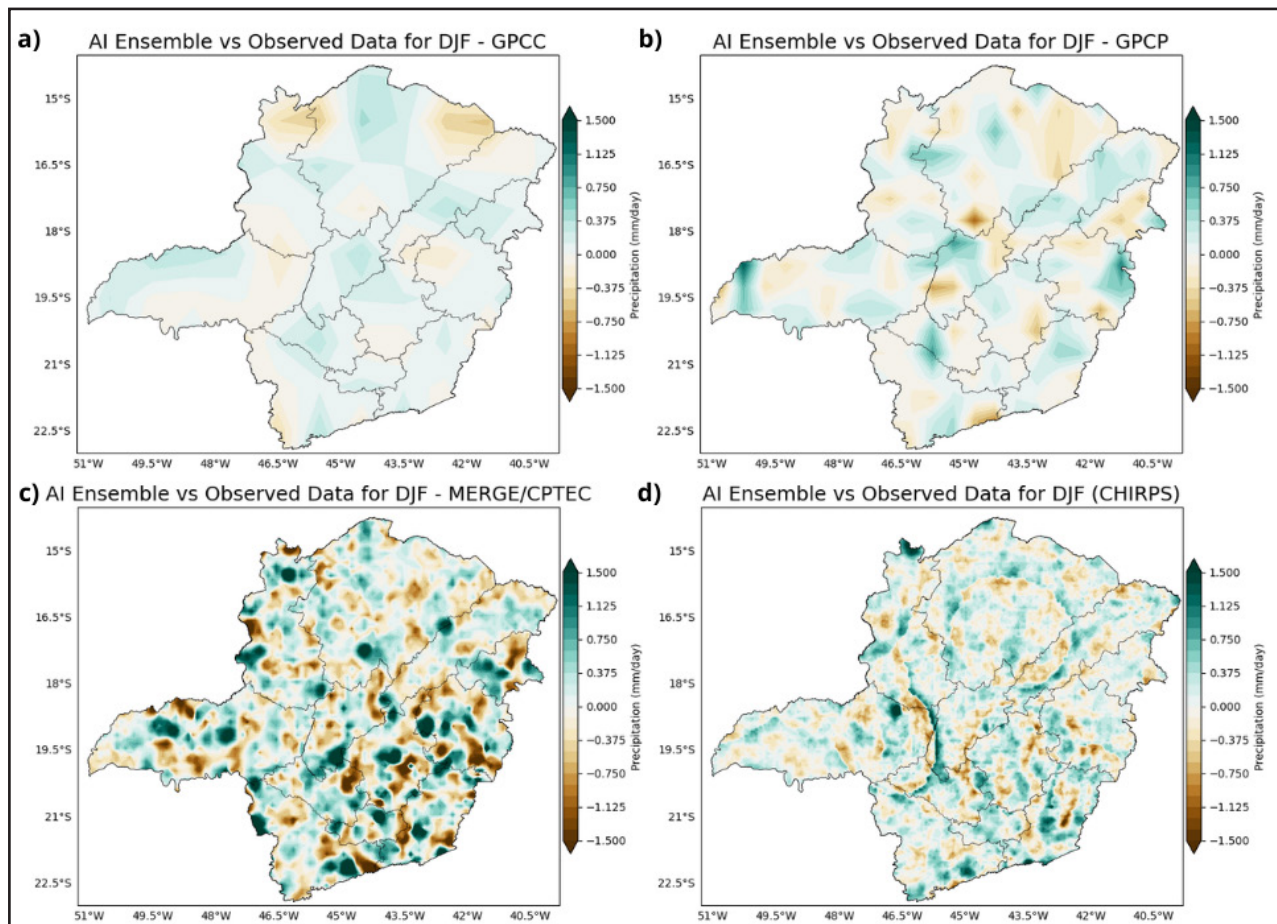
and RN-CHIRPS show a higher degree of localized bias, with a patchier distribution of positive and negative biases, especially evident in RN-CHIRPS. This indicates that while these higher-resolution datasets capture more spatial detail, the set of AI models has more difficulty accurately forecasting precipitation at this finer scale, leading to a more heterogeneous bias distribution.

Figure 5 – DJF 2022 Precipitation Forecast over Minas Gerais from the Suite of AI Model



Source: Authors (2024). Caption: a) Forecast for DJF 2022 in the mesoregions of Minas Gerais using the set of AI models for GPCC data; b) Forecast for DJF 2022 in the mesoregions of Minas Gerais using the set of AI models for GPCP data; c) Forecast for DJF 2022 in the mesoregions of Minas Gerais using the set of AI models for MERGE/CPTEC data; d) Forecast for DJF 2022 in the mesoregions of Minas Gerais using the set of AI models for CHIRPS data

Figure 6 – Error between observed data and the set of AI models forecast



Source: Authors (2024). Caption: a) Error between observed data and the precipitation forecast model, generated from the set of AI models, for GPCC data in the mesoregions of Minas Gerais. b) Error between observed data and the precipitation forecast model, generated from the set of AI models, for GPCP data in the mesoregions of Minas Gerais. c) Error between observed data and the precipitation forecast model, generated from the set of AI models, for MERGE/CPTC data in the mesoregions of Minas Gerais. d) Error between observed data and the precipitation forecast model, generated from the set of AI models, for CHIRPS data in the mesoregions of Minas Gerais

RMSE for the IA models and for the dynamic NMME models are shown in a heatmap (Figure 7) for the mesoregions of Minas Gerais. Each cell's color represents the RMSE value, with blue indicating lower errors and red indicating higher errors. It is noticeable that the highest errors occur in the southern (Sul/Sudoeste Minas region) and central Minas Gerais (Metropolitana de Belo Horizonte) in RN-GPCC, and the lowest in the east of the state. (Vale do Mucuri mesoregion). Similar to the previous figure, the highest error of RN-GPCP is in South/Southwest Minas, with the

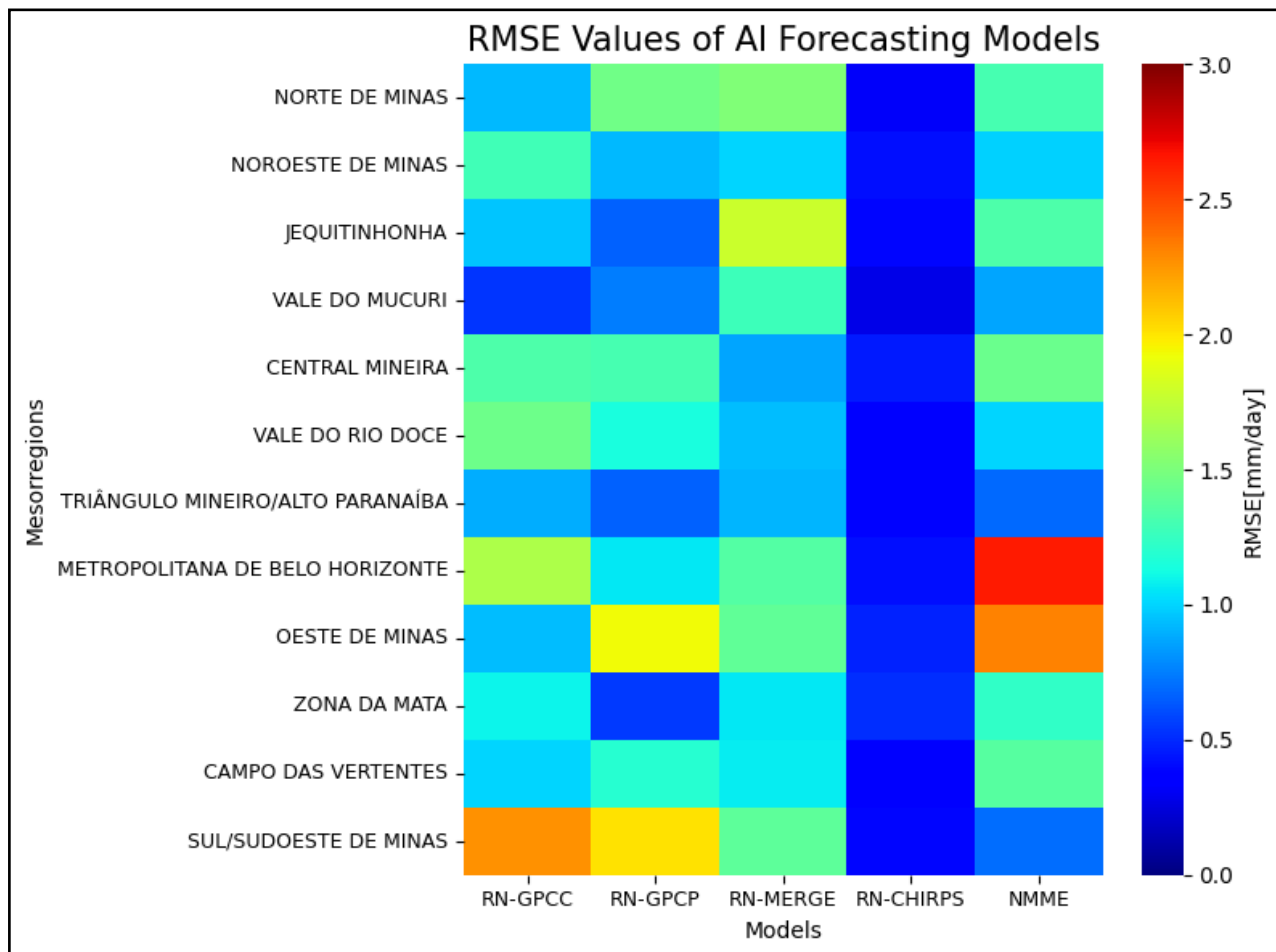
second highest value in Western Minas, and the lowest in Zona da Mata. The RN-MERGE/CPTEC shows lower RMSE values, with the smallest in Central Mineira and the highest in Vale do Jequitinhonha.

In general, RN-GPCC, RN-GPCP, and RN-MERGE models show low to moderate RMSE values across most regions. On the other hand, the RN-CHIRPS consistently demonstrates the lowest RMSE values across nearly all mesoregions, indicating it as the most accurate AI model in terms of minimizing error, aligning with its higher spatial resolution and ability to capture finer regional details. The case study results for DJF/2022 suggest that the CHIRPS-trained model achieved the best performance, with the lowest RMSE values in almost all mesoregions. Although it is not possible to draw a definitive conclusion from a single test period, this result highlights the significant potential of using high-resolution data to improve the accuracy of AI models.

The NMME model shows higher RMSE values in several mesoregions (particularly in Metropolitana de Belo Horizonte and Sul/Sudoeste de Minas), suggesting that it struggles with forecasting accuracy in these areas. This contrast highlights the advantage of the AI models, especially RN-CHIRPS, in improving forecast precision.

Similar to what is shown in the RMSE, we have the Relative Error (RE) (Figure 8), with the same mesoregions having the best and worst evaluation values. For instance, in RN-GPCP, the error is greater in the South/Southwest of Minas. It is also possible to observe that the best evaluation among the mesoregions comes from the RN-CHIRPS prediction, followed by RN-MERGE and the models with lower spatial resolution. The mesoregion with the smallest error was the Vale do Mucuri, while the largest errors were recorded in the southern mesoregions of the state. The NMME is not included in this comparison due to the calculation method, which uses the median of the RMSE instead of the mean, resulting in a discrepancy compared to the other data.

Figure 7 – Root Mean Square Error (RMSE) of the set of AI models with each data set, for each mesoregion of Minas Gerais

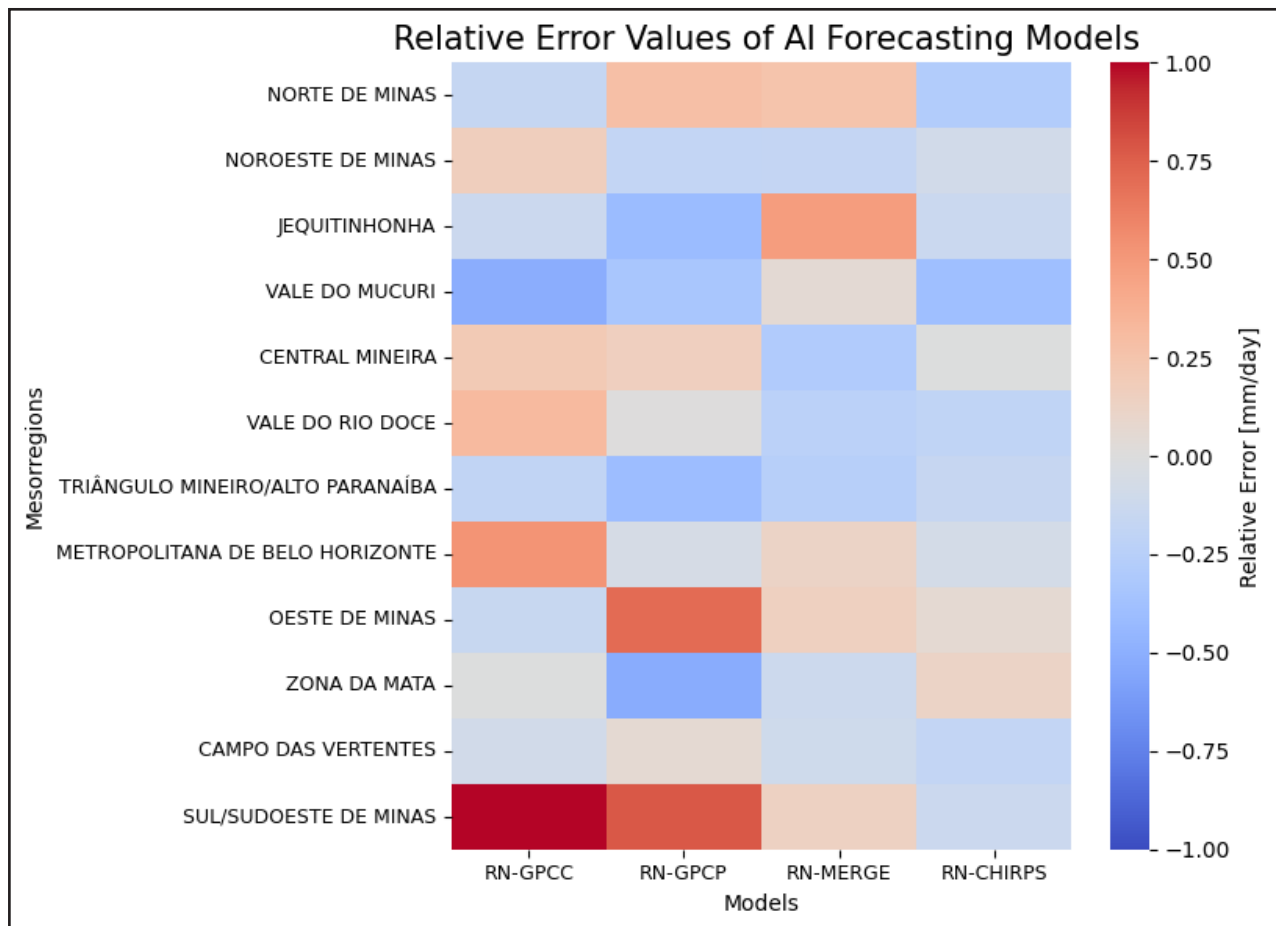


Source: Authors (2024). Caption: RMSE Heatmap of the neural networks and the NMME in relation to their respective observed data, for all mesoregions of Minas Gerais

Figure 9 shows the classification of the predicted precipitation for DJF 2022, divided into Rainy, Dry, and Normal categories. Most of the state of Minas Gerais was classified as Rainy across all datasets, except in the regions of Triângulo Mineiro, Oeste, and Sul/Sudoeste de Minas, where precipitation was below average, thus aligning with the previously observed data. In the case of RN-MERGE/CPTEC and RN-CHIRPS, there are even areas that can be classified as Dry, such as the extreme south of Sul/Sudoeste de Minas and the western part of Triângulo Mineiro in the first case, and also in Campo das Vertentes in the second case. These areas coincide with negative anomalies

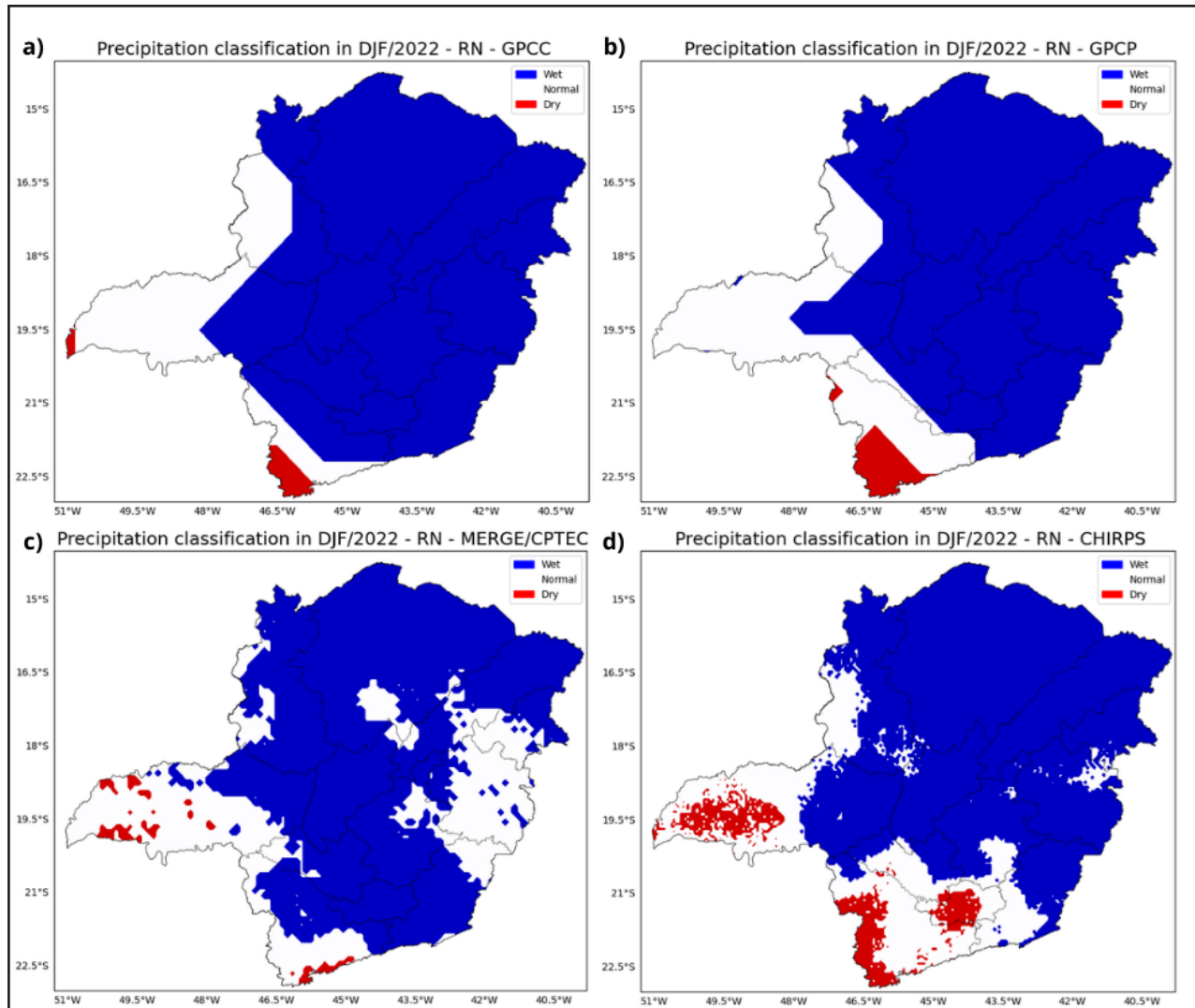
observed in Figure 4, confirming that the AI models successfully reproduced the spatial distribution of drier conditions and captured regional precipitation contrasts associated with the 2021/2022 rainy season.

Figure 8 – Relative Error Heatmap of the set of AI models based on each dataset for the mesoregions of Minas Gerais



Source: Authors (2024). Caption: Relative Error Heatmap of the Neural Networks and the NMME compared to their respective observed data, for all mesoregions of the state of Minas Gerais

Figure 9 – Classification of Precipitation Forecast over Minas Gerais from the Suite of AI Models



Source: Authors (2024). Caption: a) Upper and lower quartiles of RN-GPCC for the mesoregions of the state of Minas Gerais; b) Upper and lower quartiles of RN-GPCP for the mesoregions of the state of Minas Gerais; c) Upper and lower quartiles of RN-MERGE/CPTEC for the mesoregions of the state of Minas Gerais; d) Upper and lower quartiles of RN-CHIRPS for the mesoregions of the state of Minas Gerais

4 FINAL CONSIDERATIONS

This study investigated the use of artificial intelligence models, specifically neural networks, to improve precipitation forecasting and identify climate anomalies over the state of Minas Gerais, Brazil. The results obtained highlight the potential of

this approach, consistent with findings reported by Materia et al. (2024). The analysis used high-resolution spatial data from multiple sources, such as GPCC, GPCP, MERGE-CPTEC, and CHIRPS, enabling a comprehensive comparison across datasets. The same AI configuration, a feed-forward multilayer perceptron with two hidden layers of 50 neurons each, ReLU activation, and the Adam optimizer, was applied to all datasets. Therefore, the performance differences reflect the quality and resolution of the precipitation inputs rather than variations in the AI method itself.

The results indicated that averaging multiple runs within each model and using high-resolution input data were key factors in improving forecast accuracy. Among the models evaluated, the network trained with CHIRPS data achieved the lowest RMSE and RE values. However, since the RE was calculated for the mean values across mesoregions, the results may have masked larger errors. The highest errors were observed in mesoregions that experienced drier conditions during DJF/2022, which are also characterized by significant variations in altitude.

Computational constraints were identified as one of the main challenges of this study, primarily due to the high processing time and memory requirements associated with training models on fine-resolution datasets such as CHIRPS and MERGE-CPTEC. These limitations restricted the use of more advanced hyperparameter optimization methods and the testing of additional seasons. Nevertheless, the application of machine learning techniques for precipitation forecasting remains an appealing alternative, as they require substantially less computational time than physics-based dynamical models.

The division of Minas Gerais into mesoregions, while not directly influencing neural network performance, proved useful for interpreting the spatial distribution of errors and understanding how topographic and climatic heterogeneity affect precipitation predictability. Although the CHIRPS-trained neural network exhibited the lowest RMSE and RE values across most mesoregions, this result should be interpreted cautiously, as

the evaluation was based on a single independent season (DJF 2021/2022). Therefore, the present analysis represents an exploratory case study aimed at assessing the models' behavior under specific climatic conditions. Additional tests across multiple seasons and years are required to statistically validate model performance and confirm the potential benefits of using high-resolution datasets such as CHIRPS.

The categorical classification of precipitation for DJF 2022 further demonstrated the ability of the AI models to reproduce the observed spatial distribution of rainfall extremes. Most of Minas Gerais was correctly classified as Rainy, while regions such as Triângulo Mineiro, Oeste, and Sul/Sudoeste de Minas were identified as drier than normal, in agreement with the observed anomalies. The models trained with MERGE/CPTEC and CHIRPS data provided greater spatial detail, successfully detecting localized dry areas, such as the extreme south of Sul/Sudoeste de Minas and parts of Campo das Vertentes, corresponding to negative anomalies observed in the reference datasets. These results suggest that the AI models were able to capture regional contrasts and relative extremes within the seasonal precipitation patterns, highlighting their potential for identifying areas prone to hydrometeorological anomalies.

Looking forward, it is necessary to improve computational infrastructure and explore even more advanced artificial intelligence algorithms capable of efficiently processing large volumes of data, such as the methods used by Chen et al. (2024). This would also enable the application of systematic hyperparameter optimization techniques. The use of methods such as Grid Search or Bayesian Optimization would ensure a more exhaustive search for the ideal model architecture, overcoming the limitation of the manual search conducted in this study and potentially increasing the accuracy of the forecasts. Additionally, the continuous development of techniques for integrating different data sources could further enhance the accuracy and usefulness of meteorological forecasts, benefiting various sectors such as agriculture, water resource management, and preparation for extreme weather events.

REFERENCE

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, 4(6), 1147-1167.
- Amante, C., & Eakins, B. W. (2009). *ETOPO1 - 1 arc-minute global relief model: procedures, data sources and analysis*. (NOAA Technical Memorandum NESDIS NGDC-24). National Geophysical Data Center, National Oceanic and Atmospheric Administration. <https://doi.org/10.7289/V5C8276M>
- Anochi, J. A., de Almeida, V. A., de Campos Velho, H. F. (2021). Machine learning for climate precipitation prediction modeling over South America. *Remote Sensing*, 13(13), 2468. <https://doi.org/10.3390/rs13132468>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific model development discussions*, 7(1), 1525-1534. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chen, L., Zhong, X., Li, H., Wu, J., Lu, Bo., Chen, D., Xie, S., Wu, L., Chao, Q., Lin, C., & Qi, Y. (2024). A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, 15(1), 6425. <https://doi.org/10.1038/s41467-024-50714-1>
- Conti, S. (2024). Artificial intelligence for weather forecasting. *Nature Reviews Electrical Engineering*, 1(1), 8-8. <https://doi.org/10.1038/s44287-023-00009-2>
- 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. *Scientific data*, 2(1), 1-21. <https://doi.org/10.1038/sdata.2015.66>
- Huffman, G. J., Adler, R. F.; Behrangi, Ali; Bolvin, David T.; Nelkin, Eric; Song, Yang. Algorithm Theoretical Basis Document (ATBD) for Global Precipitation Climatology Project Version 3.1 Precipitation Data. 2020. Disponível em: https://disc.gsfc.nasa.gov/datasets/GPCPMON_3.2/summary. Acesso em: 11 mar. 2023.
- Instituto Brasileiro de Geografia e Estatística (2022). Área territorial brasileira 2022. <https://cidades.ibge.gov.br/brasil/mg/panorama>. Acesso em: 05 nov. 2024.
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M. P., Becker, E., Peng, P., Tripp, P., Huang, P., DeWitt, D. G., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., et al. (2014). The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585-601. <https://doi.org/10.1175/BAMS-D-12-00050.1>

- Lam, R., Gonzalez-Sanchez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Vinyals, O., Stott, J., Prtizel, A., Mohamed, S., & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677), 1416-1421. <https://doi.org/10.1126/science.ad2336>
- Marengo, J. A., Seluchi, M. E., Cunha, A. P., Cuartas, L. A., Goncalves, D., Sperling, V. B., ... & Moraes, O. L. (2023). Heavy rainfall associated with floods in southeastern Brazil in November–December 2021. *Natural Hazards*, 116(3), 3617-3644. <https://doi.org/10.1007/s11069-023-05827-z>
- Marmolin, H. Subjective MSE measures. *IEEE transactions on systems, man, and cybernetics*, 16(3), 486-489. <https://doi.org/10.1109/TSMC.1986.4308985>
- Materia, S., García, L. P., van Straaten, C., O, S., Mamalakis, A., Cavicchia, L., Coumou, D., de Luca, P., Kretschmer, M., & Donat, M. (2024). Artificial intelligence for climate prediction of extremes: State of the art, challenges, and future perspectives. *WIREs Climate Change*, 15(6), e914. <https://doi.org/10.1002/wcc.914>
- Minuzzi, R. B., Sediya, G. C., Ribeiro, A., & da Costa, J. M. N. (2005). El Niño: ocorrência e duração dos veranicos do Estado de Minas Gerais. *Revista brasileira de engenharia agrícola e ambiental*, 9, 364-371. <https://doi.org/10.1590/S1415-43662005000300011>
- Rizzo, I. V., & Canato, R. L. C. (2020). Inteligência artificial: funções de ativação. *Prospectus*, 2(2), 51-65.
- Rozante, J. R., Moreira, D. S., de Gonçalves, L. G. G., Vila, D. A. (2010). Combining TRMM and surface observations of precipitation: technique and validation over South America. *Weather and forecasting*, 25(3), 885-894. <https://doi.org/10.1175/2010WAF2222325.1>
- Shimizu, M. H., Ambrizzi, T., & Liebmann, B. (2017). Extreme precipitation events and their relationship with ENSO and MJO phases over northern South America. *International Journal of Climatology*, 37(6), 2977-2989. <https://doi.org/10.1002/joc.4893>

Authorship contributions

1 – Matheus José Gomes

Undergraduate student in Atmospheric Sciences at the Federal University of Itajubá (UNIFEI)
<https://orcid.org/0009-0008-9760-7891> • matheusjosegomes1@gmail.com

Contribution: Methodology, Investigation, Data curation, Writing - original draft preparation, Writing - review & editing

2 – Juliana Aparecida Anochi

PhD in Applied Computing from the National Institute for Space Research (INPE)

<https://orcid.org/0000-0003-0769-9750> • juliana.anochi@inpe.br

Contribution: Conceptualization, Methodology, Investigation, Supervision, Writing - review & editing

3 – Marilia Harumi Shimizu

PhD in Meteorology from the National Institute for Space Research (INPE)

<https://orcid.org/0000-0003-0895-555X> • marilia.shimizu@gmail.com

Contribution: Conceptualization, Methodology, Investigation, Supervision, Writing - review & editing

How to quote this article

Gomes, M. J., Anochi, J. A., Shimizu, M. H. (2026). Use of artificial intelligence for seasonal precipitation forecasting in the Minas Gerais state, Brazil. *Ciência e Natura*, Santa Maria, 48, e90228. DOI 10.5902/2179460X90228. Disponível em: <https://doi.org/10.5902/2179460X90228>.