DESEMPENHO DE DIFERENTES BASES DE DADOS GLOBAIS NA ESTIMATIVA DA PRECIPITAÇÃO NO MUNICÍPIO DE PALMARES-PE

Luis Felipe Francisco Ferreira da Silva [1] ; Ivanildo Batista da Silva Júnior [2]; Gabriel Graciano de Mendonça[3]; Antônio Samuel Alves da Silva[4]

RESUMO- O monitoramento hidrológico é uma ferramenta de suma importância para a gestão hídrica, especialmente em bacias hidrográficas estratégicas. Neste contexto, o presente estudo avaliou o desempenho de 11 bases de dados globais de precipitação (produtos de satélite e reanálises) para o município de Palmares, inserido na bacia do rio Una, Pernambuco. O objetivo foi validar as estimativas de precipitação mensal dessas bases globais, confrontando-as com os dados observados em grade da base BRDWGD (1982-2023). A metodologia baseou-se na extração espacial das séries temporais e na aplicação de métricas estatísticas de erro e eficiência (r, R^2, RMSE, ME, d, KGE e PBIAS). Em seguida, aplicou-se a análise de agrupamento hierárquico multivariada método de Ward para classificar os produtos com base em suas similaridades e perfis de desempenho. Os resultados revelaram expressivas variações na capacidade das fontes em representar o volume pluviométrico local. A análise de agrupamento permitiu delimitar grupos com assinaturas estatísticas semelhantes, destacando bases de alta acurácia (como MSWEP, CPC e CHIRPS) e evidenciando o viés sistemático de subestimativa severa em reanálises (como ERA5) e satélites específicos. Conclui-se que a aplicação de uma validação estatística local sobre diferentes produtos globais é uma etapa indispensável para garantir escolhas mais assertivas e seguras no planejamento hidroambiental.

ABSTRACT– Hydrological monitoring is a tool of paramount importance for water management, especially in strategic watersheds. In this context, the present study evaluated the performance of 11 global precipitation datasets (satellite products and reanalyses) for the municipality of Palmares, located in the Una River basin, Pernambuco state, Brazil. The objective was to validate the monthly precipitation estimates of these global datasets by comparing them with observed gridded data from the BRDWGD (1982-2023) database. The methodology was based on the spatial extraction of time series and the application of statistical error and efficiency metrics (r, R^2, RMSE, ME, d, KGE, and PBIAS). Subsequently, a multivariate Hierarchical Cluster Analysis was applied to classify the products based on their similarities and performance profiles. The results revealed significant variations in the capacity of the sources to represent the local rainfall volume. The cluster analysis allowed the delimitation of groups with similar statistical signatures, highlighting high-accuracy datasets (such as MSWEP, CPC, and CHIRPS) and evidencing the systematic bias of severe underestimation in reanalyses (such as ERA5) and specific satellites. It is concluded that the

[1]) Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife - PE, CEP: 52171-900. Fone: (82) 9 9934-9312 E-mail: luis.francisco@ufrpe.br
[2]) Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife - PE, CEP: 52171-900. E-mail:Ivanildo..batista13@gamil.com
[3]) Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife - PE, CEP: 52171-900. E-mail:gabriel.gracianom@ufrpe.br
[4]) Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco (UFRPE). Rua Dom Manoel de Medeiros, s/n, Dois Irmãos, Recife - PE, CEP: 52171-900. E-mail:antonio.sasilva@ufrpe.br

application of local statistical validation on different global products is an indispensable step to ensure more assertive and safe choices in hydro-environmental planning.

Palavras-Chave – Precipitação mensal; Sensoriamento remoto; Validação estatística; Análise multivariada; Bacia do rio Una.

1- INTRODUÇÃO

O Nordeste brasileiro é uma região vulnerável a extremos de variabilidade climática, onde a demanda hídrica é severamente influenciada tanto pela elevação das temperaturas quanto pela irregularidade das taxas pluviométricas Durães (2024). Nesse contexto, a mensuração precisa dos volumes precipitados para o planejamento hidrológico é de suma importância. Com isso, a expansão de redes de monitoramento e o refinamento de dados orbitais tornam-se essenciais para mitigar os impactos decorrentes da escassez e da má distribuição hídrica Su *et al.* (2026).

A utilização de sensores remotos tem se consolidado como uma alternativa indispensável para suprir a escassez de estações pluviométricas *in situ*, permitindo o preenchimento de lacunas espaciais em áreas onde a rede de observação convencional é insuficiente ou inexistente Su *et al.* (2026); Xiong *et al*. (2023). O monitoramento hidrológico contínuo destaca-se como uma ferramenta essencial para a gestão hídrica local, especialmente em áreas críticas e vulneráveis a eventos extremos, como o município de Palmares, inserido na bacia do rio Una, que tem sofrido recorrentemente com desastres hidrológicos nos últimos anos. Para contornar as limitações de resolução espacial e temporal, o uso de produtos globais de precipitação tem surgido como uma alternativa robusta na literatura recente Banimfreg *et al.* (2025). Contudo, a aplicação prática desses produtos remotos exige uma validação rigorosa em escala local antes de sua integração ao manejo dos recursos hídricos. A avaliação dessas ferramentas frequentemente revela inconsistências significativas entre os *datasets*, ressaltando a necessidade de comparar as estimativas com dados de referência para assegurar a confiabilidade dos produtos Chagas *et al.* (2022); Fassoni-Andrade *et al.* (2021).

Diante dessa necessidade, este estudo avalia o desempenho de onze bases de dados globais de precipitação, englobando produtos de satélite e reanálises: CCS, CHIRPS, CPC, CRU, ERA5, GPCC, GPCP, MERRA2, MSWEP, PERSIANN e TERRA. Para fins de validação e análise de acurácia das estimativas de precipitação mensal, a base em grade BRDWGD Xavier *et al.* (2015) foi rigorosamente adotada como o dado observado de referência. A metodologia consistiu na extração espacial dos dados para o município de Palmares, seguida pela aplicação de métricas estatísticas de aderência e erro (r, R^2, RMSE, ME, d, KGE e PBIAS) para avaliar a capacidade dos produtos em representar os volumes pluviométricos locais. Adicionalmente, empregou-se a análise multivariada de agrupamento

hierárquico (método de Ward) para classificar os produtos com base na similaridade de seus desempenhos, permitindo a identificação de grupos com tendências de subestimativa ou superestimativa. Através disso, busca-se fornecer subsídios científicos sólidos para a escolha assertiva de produtos remotos voltados ao monitoramento climático e à gestão hidrológica regional

2- MATERIAIS E MÉTODOS

2.1 Área de Estudo

O estudo foi realizado no município de Palmares, estrategicamente localizado na mesorregião da Mata Sul do estado de Pernambuco que estar inserido na bacia hidrográfica do rio Una. A área de estudo abrange uma extensão territorial de 339,388 km², caracterizada por um relevo predominantemente ondulado IBGE (2020). De acordo com a classificação climática de Köppen, a região apresenta clima do tipo *As,* tropical com estação seca de verão, sob forte influência do regime de chuvas concentradas no período de outono-inverno. Essa dinâmica pluviométrica reverbera historicamente nas condições socioambientais e no planejamento urbano e hidrológico local. Para representar espacialmente essa extensão territorial, utilizou-se o contorno municipal disponibilizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), sobre o qual foram interceptados os dados em grade, resultando na atribuição de 4 pontos (pixels) incidentes sobre a cidade para extrair os dados de precipitação média local.

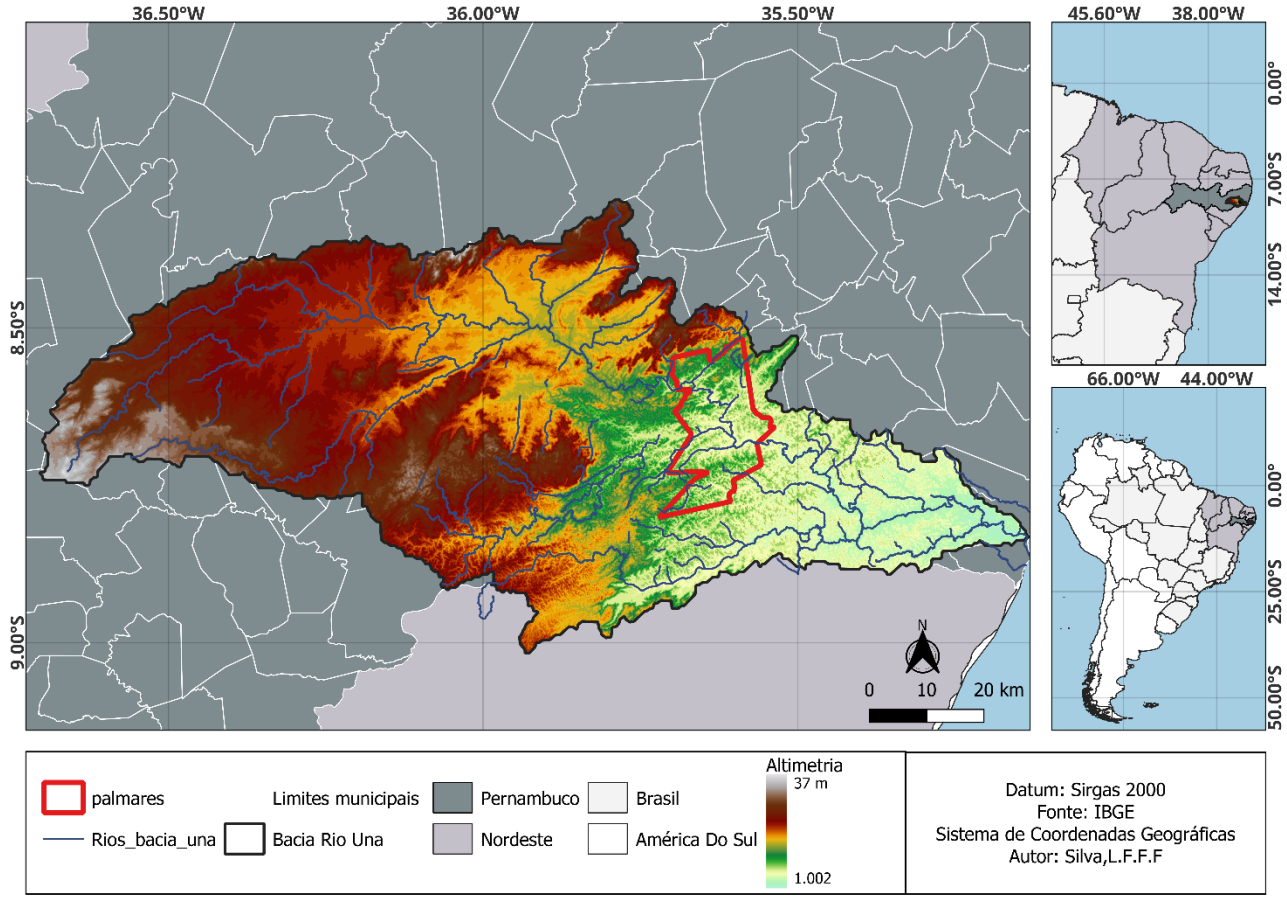


Figura 1- Mapa de localização da Bacia do Rio Una e sua altimetria com destaque para a cidade de Palmares- PE.

2.2 Bases de Dados

Para a avaliação do desempenho na estimativa pluviométrica, foram selecionadas onze bases de dados globais (englobando estimativas por satélites e modelos de reanálise), com suas respectivas resoluções espaciais: CCS (0,04°), CHIRPS (0,05°), CPC (0,5°), CRU (0,5°), ERA5 (0,25°), GPCC (0,25°), GPCP (2,5°), MERRA2 (0,5° x 0,625°), MSWEP (0,1°), PERSIANN (0,25°) e TERRA (1/24°). Como dado observacional de referência, utilizou-se a base em grade BRDWGD (*Brazilian Daily Weather Gridded Data*), desenvolvida por Xavier *et al*. (2015), amplamente validada para o território brasileiro. Considerando as diferenças de resolução espacial entre os produtos, todas as bases foram harmonizadas para a resolução do BRDWGD (0.1 x 0.1) mediante reamostragem espacial por interpolação bilinear Jensen (2015), permitindo a comparação direta entre os conjuntos de dados. A obtenção dos dados médios para o município de Palmares foi realizada via geoprocessamento em ambiente R, onde a interseção espacial da malha municipal resultou na seleção de 4 *pixels* incidentes sobre a área de estudo.

2.3 Análise Estatística e Agrupamento Multivariado

A acurácia dos onze produtos globais em estimar a precipitação mensal foi avaliada por meio da comparação quantitativa direta com a série da base BRDWGD. Para atestar a confiabilidade das estimativas, foram calculados sete métricas de desempenho. Considerando Oi como o valor observado (BRDWGD), Si como o valor estimado pelo produto remoto, $\bar{O}$ e $\bar{S}$ como as suas respectivas médias, e n como o número total de observações na série histórica, as métricas aplicadas são definidas pelas equações de (1) a (7).

Coeficiente de Correlação de Pearson (r):

Mede o grau de correlação linear entre os dados estimados e observados.

$$r = \frac{\sum_{i=1}^{n}(Oi-\bar{O})(Si-\bar{S})}{\sqrt{\sum_{i=1}^{n}(Oi-\bar{O})^2}\sqrt{\sum_{i=1}^{n}(Si-\bar{S})^2}} \quad (1)$$

Coeficiente de Determinação (R^2):

Indica a proporção da variância nos dados observados que é explicada pelo produto remoto.

$$R^2 = \left(\frac{\sum_{i=1}^{n}(Oi-\bar{O})(Si-\bar{S})}{\sum_{i=1}^{n}(Oi-\bar{O})^2\sqrt{\sum_{i=1}^{n}(Si-\bar{S})^2}}\right)^2 \quad (2)$$

Raiz do Erro Quadrático Médio ($RMSE$):

Quantifica a magnitude média dos erros, penalizando discrepâncias maiores.

$$\text{RSME} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Si-Oi)^2} \quad (3)$$

Erro Médio (ME):

Indica a tendência média do produto em superestimar (valores positivos) ou subestimar (valores negativos) os dados reais.

$$ME = \frac{1}{n}\sum_{i=1}^{n}(Si-Oi) \quad (4)$$

Índice de Concordância de Willmott (d):

Avalia o grau de exatidão das estimativas, variando de 0 (nenhuma concordância) a 1 (concordância perfeita).

$$d = 1 - \frac{\sum_{i=1}^{n}(Si-Oi)^2}{\sum_{i=1}^{n}(|Si-\bar{O}|+|Oi-\bar{O}|)^2} \quad (5)$$

Eficiência de Kling-Gupta (KGE):

Métrica robusta que integra a correlação, o viés e a variabilidade em um único índice.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (6)$$

Sendo α= $\frac{\sigma s}{\sigma o}$ (razão entre o desvio padrão dos dados simulados e observados) e $\beta = \frac{\bar{S}}{\bar{0}}$ razão entre os acumulados

Viés Percentual (*PBIAS*):

Mede a tendência percentual média dos dados simulados em relação aos observados.

$$PBIAS = \left(\frac{\sum_{i=1}^{n}(Si-Oi)}{\sum_{i=1}^{n} Oi}\right) x\ 100 \quad (7)$$

Em relação à escala de variação, os coeficientes *r*, R^2 e *d* variam de 0 a 1 (com *r* podendo variar de -1 a 1), onde o valor 1 indica concordância perfeita. A eficiência de Kling-Gupta (KGE) varia de $-\infty$ a 1, sendo 1 o valor ideal. As métricas de erro RMSE e ME são expressas na mesma unidade da variável avaliada (mm), com o valor 0 representando a ausência de erro. Por fim, o PBIAS quantifica o viés em porcentagem (%), onde 0% indica estimativa perfeita, valores positivos apontam superestimativa e valores negativos indicam subestimativa.

Por fim, visando uma avaliação sistêmica, os resultados dessas métricas estatísticas foram submetidos a uma Análise de Agrupamento Hierárquico *Cluster Analysis*. Utilizou-se o método de Ward como medida de agrupamento e a distância euclidiana como medida de dissimilaridade pacote *factoextra* Kassambara e Mundt (2020) para classificar e agrupar os produtos em *clusters* de acordo com a similaridade de seus desempenhos aderência, subestimativa ou superestimativa frente à base BRDWGD para a variável de precipitação.

3- RESULTADOS E DISCUSSÕES

3.1 Desempenho na Estimativa da Precipitação Mensal

A avaliação dos onze produtos globais frente à base observada (BRDWGD) evidenciou variações expressivas na representação pluviométrica de Palmares (Figura 1).

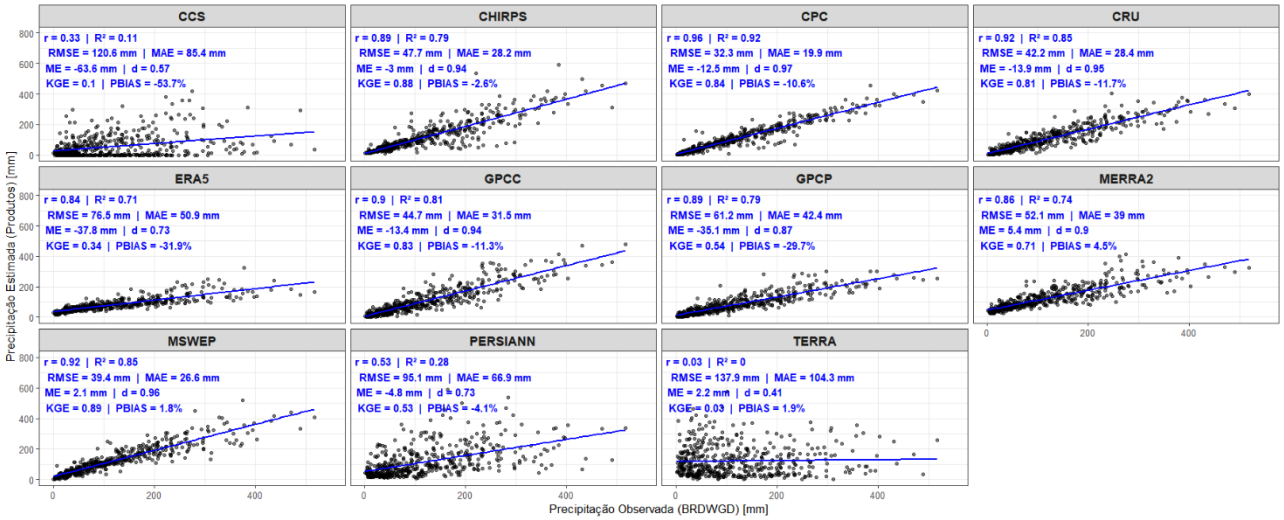


Figura 2- Dispersão e métricas estatísticas estimado pelos produtos globais versus o observado (BRDWGD) para Palmares-PE.

Os melhores desempenhos foram de MSWEP, CPC, CHIRPS, CRU e GPCC. O MSWEP destacou-se pelo melhor equilíbrio geral (KGE = 0,89; PBIAS = 1,8%), enquanto o CPC obteve a maior correlação (r = 0,96) e o menor erro (RMSE = 32,3 mm).

Em nível intermediário, MERRA2, GPCP e ERA5 apresentaram forte correlação linear (r > 0,84), mas sofreram altas penalizações por subestimarem o volume. ERA5 e GPCP registraram PBIAS próximos a -30%, exigindo correção de viés calibração prévia ao uso hidrológico.

Por fim, TERRA, CCS e PERSIANN registraram desempenhos críticos e insatisfatórios. O TERRA teve correlação praticamente nula (r = 0,03) e o maior RMSE (137,9 mm). Já o CCS subestimou as chuvas locais em mais da metade (PBIAS = -53,7% e KGE = 0,17), inviabilizando a aplicação direta desses modelos no monitoramento da região.

Para complementar a análise individual das métricas e compreender o comportamento sistêmico das bases de dados, aplicou-se a Análise de Agrupamento Hierárquico (método de Ward). Essa técnica estatística multivariada classificou os onze produtos em quatro grupos (*clusters*) distintos, com base na similaridade de seus desempenhos. O número ideal de *clusters* foi definido através do critério heurístico de inspeção visual do dendrograma, alocando o corte na maior distância de ligação salto entre os nós, visando a máxima homogeneidade interna e heterogeneidade externa Hair *et al* (2009).

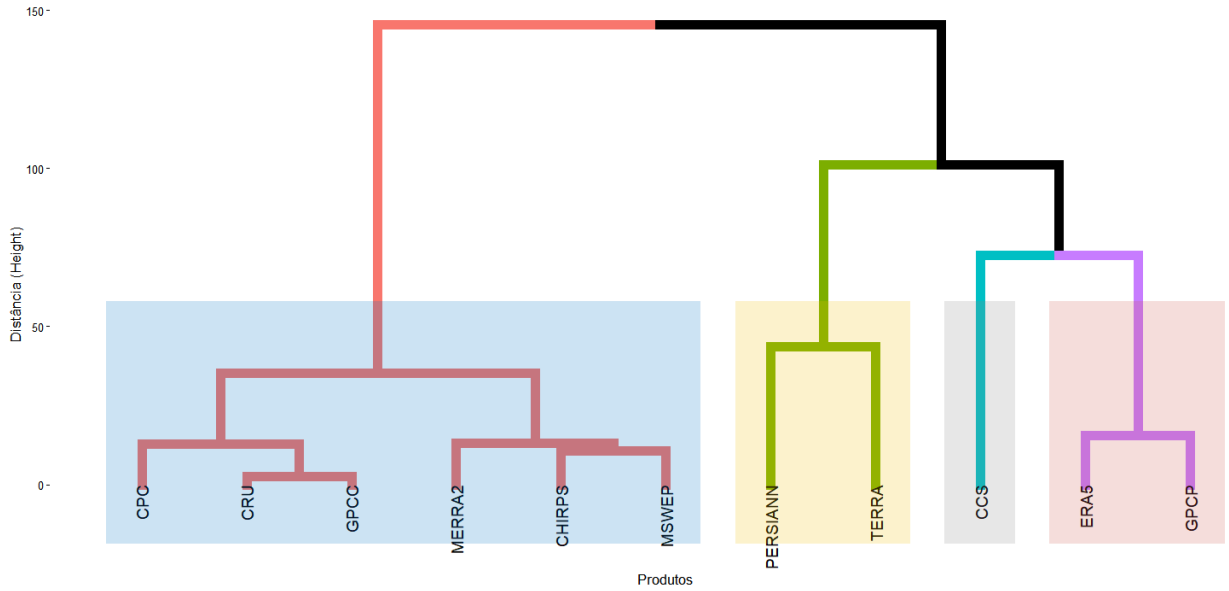


Figura 3- Dendrograma de Similaridade dos produtos de precipitação mensal.

O primeiro grupo (destacado no bloco azul) reuniu a maior quantidade de produtos: CPC, CRU, GPCC, MERRA2, CHIRPS e MSWEP. Este *cluster* principal engloba as bases de dados que apresentaram as melhores performances globais para a área de estudo. A similaridade metodológica que une esse grupo é caracterizada pela alta aderência aos dados reais (índices KGE e d elevados), forte correlação linear e os menores erros residuais da análise.

O segundo grupo (destacado no bloco rosa à direita) uniu a reanálise ERA5 e o produto GPCP. A proximidade estatística entre essas duas bases reside no fato de ambas apresentarem o mesmo padrão de viés: uma forte correlação temporal com os dados da base BRDWGD, porém acompanhada de uma expressiva subestimativa sistemática do volume mensal.

O terceiro grupo (destacado no bloco amarelo) associou os produtos PERSIANN e TERRA. Esta ramificação reflete diretamente os desempenhos insatisfatórios de ambas as bases para a região de Palmares, caracterizados por uma baixa correlação e pela maior dispersão de erros (altos valores de RMSE).

Por fim, o produto CCS formou um quarto *cluster* isolado (bloco cinza). A elevada distância euclidiana que separa o CCS de todos os demais produtos no dendrograma confirma o seu comportamento estatístico atípico e discrepante, marcado por uma subestimativa severa (reduzindo a chuva local em mais da metade) e uma eficiência global quase nula para a representação do regime pluviométrico local.

4- CONCLUSÃO

A avaliação das onze bases globais de precipitação no município de Palmares-PE demonstrou que a escolha de produtos remotos para a bacia do rio Una exige uma validação local criteriosa. Os resultados evidenciaram que os produtos MSWEP, CPC e CHIRPS apresentaram os melhores desempenhos para a estimativa do volume pluviométrico mensal, consagrando-se como as bases mais robustas e recomendadas para a região. Por outro lado, alerta-se que reanálises amplamente utilizadas, como ERA5 e MERRA2, subestimaram sistematicamente os volumes captados, exigindo a aplicação de técnicas de correção de viés prévias ao seu uso em modelagens hidrológicas. Ademais, a aplicação direta de produtos como CCS, PERSIANN e TERRA é fortemente desencorajada para a área de estudo, dado o isolamento estatístico, falhas críticas na quantificação do volume e inaptidão em representar adequadamente a climatologia local. Conclui-se que o uso integrado de métricas quantitativas de erro e análises multivariadas de agrupamento é uma etapa indispensável para evitar incertezas e embasar escolhas seguras para o planejamento e a gestão hidroambiental na Zona da Mata pernambucana.

AGRADECIMENTOS

Agradecimento a agência de fomento CAPES, pela concessão de bolsa de Pós-Graduação e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio à pesquisa por meio da bolsa de produtividade em pesquisa (nº 306590/2024-7).

REFERÊNCIAS

BANIMFREG, B. *et al.* (2025). "Abordagem inovadora para QPE baseada em medições em climas áridos: comparação entre redes neurais e métodos tradicionais". Journal of Big Data 12(1), p. 177.

CHAGAS, V. B. P.; CHAFFE, P. L. B.; BLÖSCHL, G. (2022). "Climate and land management accelerate the Brazilian water cycle". Nature Communications 13(1), p. 5136.

DURÃES, M. F. (2024). "Avaliação da agricultura irrigada brasileira: o que esperar?".
FASSONI-ANDRADE, A. C. *et al.* (2021). "Amazon hydrology from space: scientific advances and future challenges". Reviews of Geophysics 59(4), p. e2020RG000728.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. (2009). "Multivariate Data Analysis" (7th ed.). Prentice Hall.

IBGE (2020). Brasil em Síntese: Território. Instituto Brasileiro de Geografia e Estatística.
JENSEN, J. R. (2015). "Introductory Digital Image Processing: A Remote Sensing Perspective" (4th ed.). Pearson.

KASSAMBARA, A.; MUNDT, F. (2020). "factoextra: Extract and Visualize the Results of Multivariate Data Analyses". R package version 1.0.7.

SU, J. *et al.* (2026). "Precipitation observing network gaps limit climate change impact assessment". Nature, pp. 1-7.

XAVIER, A. C.; SCANLON, B. R.; KING, C. W.; ALVES, A. I. (2015). "New improved Brazilian daily weather gridded data (1961–2014)". International Journal of Climatology 36(1), pp. 55-68.

XIONG, J. *et al.* (2023). "ET-WB: water-balance-based estimations of terrestrial evaporation over global land and major global basins". Earth System Science Data 15(10), pp. 4571-4597.