



MODELAGEM DA DISTRIBUIÇÃO ESPACIAL DA PRECIPITAÇÃO ANUAL DO RIO GRANDE DO NORTE: ABORDAGEM BASEADA EM MACHINE LEARNING

Alíbia Deysi Guedes da Silva ¹; Sara Fernandes Flor de Souza ²; Rebecca Luna Lucena ³; João Santiago Reis ⁴

¹ Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Geografia - GEOCERES, Caicó - RN, Brasil. alibiadeysi.prof@gmail.com
<https://orcid.org/0000-0002-8476-8599>

² Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Geografia - GEOCERES, Caicó - RN, Brasil. sara.flor@ufrn.br
<https://orcid.org/0000-0001-6829-3434>

³ Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Geografia - GEOCERES, Caicó - RN, Brasil. rebecca.luna@ufrn.br
<https://orcid.org/0000-0003-4670-265X>

⁴ Universidade Federal do Rio Grande do Norte, Programa de Pós-Graduação em Geografia - GEOCERES, Caicó - RN, Brasil. joao.reis@ufrn.br
<https://orcid.org/0000-0002-3516-4334>

RESUMEN. El mapeo de la precipitación pluviométrica es esencial para la planificación territorial, sin embargo, el territorio brasileño presenta baja cobertura de estaciones pluviométricas y estaciones climatológicas, resultando en la ausencia de informaciones térmicas y pluviométricas en diversas localidades. En ese contexto, esta investigación tuvo como objetivo principal realizar el modelado espacial de los datos de precipitación del estado de Rio Grande do Norte, mediante técnicas de aprendizaje automático (machine learning). Los procedimientos metodológicos se basaron en la interpolación espacial por medio del algoritmo Random Forest, procesado en la plataforma Google Earth Engine. Para la calibración del modelo predictivo, se utilizó un conjunto de 15 covariables predictoras, considerando factores relacionados al relieve, vegetación y dependencia espacial, y base de datos de precipitación total anual de 198 puestos pluviométricos, considerando el año 2023. El desempeño del modelo predictivo fue evaluado como satisfactorio, presentando $r^2 = 0,72$ y MAE = 114,86. Los resultados sugieren que la aridez fue la covariable más importante, seguida por la posición geográfica, en la explicación de la distribución espacial de la precipitación del estado, en cambio, factores relacionados al relieve fueron menos relevantes. La variabilidad espacial de la precipitación se mostró similar a los datos de precipitación registrados y a la normal climatológica.

Palabras clave: Aprendizaje automático; Random Forest; Google Earth Engine; Pluviosidad.

RESUMO. O mapeamento da precipitação pluviométrica é essencial para o planejamento territorial, no entanto, o território brasileiro apresenta baixa cobertura de estações pluviométricas e estações climatológicas, resultando na ausência de informações termopluviométricas em diversas localidades. Nesse contexto, esta pesquisa teve como objetivo principal realizar a modelagem espacial dos dados de precipitação do estado do Rio Grande do Norte, mediante técnicas de aprendizado de máquina (*machine learning*). Os procedimentos metodológicos basearam-se na interpolação espacial por meio do algoritmo *Random Forest*, processado na plataforma *Google Earth Engine*. Para calibração do modelo preditivo, foi utilizado um conjunto de 15 covariáveis predictoras, considerando fatores relacionados ao relevo, vegetação e dependência espacial, e banco de dados de precipitação total anual de 198 postos pluviométricos, considerando o ano de 2023. O desempenho do modelo preditivo foi avaliado como satisfatório, apresentando $r^2 = 0,72$ e MAE = 114,86. Os resultados sugerem que a aridez foi a covariável mais importante, seguida pela posição geográfica, na explicação da distribuição espacial da precipitação do estado, em contrapartida, fatores relacionados ao relevo foram menos relevantes. A variabilidade espacial da precipitação apresentou-se semelhante aos dados de precipitação registrados e à normal climatológica.

Palavras-chave: Aprendizado de máquina; Random Forest; Google Earth Engine; Pluviosidade.



ABSTRACT. Mapping rainfall is essential for territorial planning. However, the Brazilian territory has low coverage of rainfall stations and climatological stations, resulting in the absence of thermos-pluviometric information in several locations. In this context, the main goal of this research was to carry out spatial modeling of rainfall data for the state of Rio Grande do Norte, using machine learning techniques. The methodological framework was based on spatial interpolation using the Random Forest algorithm, processed on the Google Earth Engine platform. To calibrate the predictive model, a set of 15 predictor covariates was used, considering factors related to relief, vegetation and spatial dependence, and a total annual rainfall database from 198 rainfall stations, considering the year 2023. The performance of the predictive model was assessed as satisfactory, with $r^2 = 0.72$ and $MAE = 114.86$. The results suggest that the aridity index was the most important covariate, followed by geographical position, in explaining the spatial distribution of rainfall in the state, while factors related to relief were less relevant. The spatial variability of rainfall was similar to the recorded rainfall data and the climatological normal.

Keywords: Machine learning; Random Forest; Google Earth Engine; Rainfall.

1. INTRODUÇÃO

O semiárido brasileiro é caracterizado por apresentar irregularidades espaço-temporais das precipitações, considerada como a principal variável reguladora das condições socioambientais na região. A quantidade e distribuição das chuvas estão relacionados à atuação e intensidade de diferentes sistemas atmosféricos, como a Zona de Convergência Intertropical – ZCIT, os Distúrbios Ondulatórios de Leste – DOLs, os Vórtices Ciclônicos de Altos Níveis – VCANs, que atuam por vezes associados a padrões de teleconexão, a exemplo do *El Nino* Oscilação Sul – ENOS (Gomes; Zanella, 2023). Entretanto, fatores geográficos como localização, topografia, maritimidade/continentalidade, orientação das vertentes, etc., podem atuar como modificadores dos padrões pluviométricos (Correia-Filho *et al.*, 2019). O Rio Grande do Norte (RN), um dos onze estados que compõem o semiárido brasileiro, apresenta também precipitação desigual quanto ao volume e distribuição espacial, manifestando tipos climáticos que vão do semiárido ao úmido. Nesse sentido, assume-se que a relação entre os fatores geográficos do clima e os sistemas atmosféricos atuantes seja responsável por essa diversidade pluviométrica.

Os dados espacialmente contínuos de elementos meteorológicos, como a precipitação, são importantes para a compreensão da dinâmica socioambiental, logo são necessários para o planejamento territorial. Todavia, as informações disponíveis, como os dados de estações pluviométricas e meteorológicas em superfície, apresentam desigualdades na cobertura espacial e registros incompletos. Ademais, os produtos de precipitação obtidos de satélites orbitais, que embora estejam disponíveis gratuitamente, apresentam resoluções espaciais grosseiras e erros aleatórios, assim limitando seus usos e aplicações locais, por exemplo o CHIRPS com pixel de 5km e o TRMM de 25km (Zhang *et al.*, 2023).

Em termos operacionais, o uso de ferramentas de modelagem para estimar o volume de precipitação é uma abordagem poderosa. Nesse contexto, a interpolação consiste em um processo matemático-estatístico que prevê valores em locais não amostrados a partir dos dados reais (Landim, 2000). Os principais interpoladores utilizados estão agrupados em método determinístico (ou não geoestatísticos), como a interpolação por ponderação inversa da distância (IDW); método estocástico (ou geoestatístico), como a Krigagem (*Kriging*); e combinado, como a Krigagem de regressão (Hengl *et al.*, 2018).

Estudos utilizando técnicas de aprendizado de máquina (*machine learning* - ML) têm se destacado na modelagem estatístico-computacional para predição de dados espaciais. Algoritmos como o *Random Forest* (RF), *K-Nearest Neighbor* (KNN), Redes Neurais Artificiais (RNN) e *Support Vector Machine* (SVM) são amplamente utilizados com



desempenho satisfatório em diversas áreas do conhecimento (Hengl *et al.*, 2018). Os algoritmos de ML apresentam alta capacidade em aprender padrões complexos e realizar correlações em conjuntos de variáveis que não possuem distribuição normal. O RF é um modelo não-paramétrico que utiliza método *bagging* (*Bootstrap Aggregating*) para a construção de um conjunto de árvores de decisão a partir de amostras aleatórias (Breiman, 2001). Algoritmos baseados em árvores são constantemente utilizados para tarefas de classificação e regressão. O RF suporta um grande número de covariáveis, e atualmente está disponível na plataforma de processamento em nuvem *Google Earth Engine*.

Diante do exposto, a presente pesquisa tem como objetivo analisar a modelagem espacial da distribuição de precipitação anual de 2023 para o estado do Rio Grande do Norte, através do algoritmo de machine learning Random Forest, compreendendo a importância dos elementos de relevo, posição geográfica e sistemas atmosféricos para a espacialização das chuvas no estado. Estudos dessa natureza viabilizam o entendimento espacial da intensidade e padrões das precipitações. Ressalta-se que o produto gerado e demais informações obtidas são relevantes para a eficiência do planejamento territorial dos municípios e do estado.

2. MATERIAL E MÉTODOS

2.1. Área de estudo

O estado do Rio Grande do Norte (Figura 1) é uma unidade federativa do Brasil localizada na quina setentrional da região Nordeste, entre as coordenadas de $-4^{\circ}49'53''$ e $-6^{\circ}58'57''$ de latitude $-34^{\circ}58'03''$ e $-38^{\circ}36'12''$ de longitude. Abrangendo uma área de 52.797 km², com uma população de 3.302.729 habitantes (IBGE, 2022), limita-se com o estado do Ceará a oeste, a Paraíba ao sul, e com o oceano atlântico ao norte e leste do seu território. O estado apresenta diversidade ambiental, como as diferentes formas de relevo, que inclui as planícies costeiras, tabuleiros, chapadas, inselbergs e uma extensa área aplainada, denominada de depressão sertaneja. Ainda se registram elevações que ultrapassam os 700 m de altitude, como a Serra de Santana e a Serra do Martins, nas regiões central e oeste, respectivamente. A vegetação predominante é típica de Caatinga e varia entre as fases hiperxerófila em áreas de semiaridez rigorosa, e hipoxerófila em áreas de altitude ou com condições específicas de solo (Araújo Filho *et al.*, 2022). No litoral norte, há padrões de vegetação de restinga, dunas e manguezais; no setor litorâneo leste há presença de mata Atlântica. A respeito do clima, este pode ser dividido em três domínios climáticos principais: semiárido, subúmido seco e úmido. O semiárido está presente em 155 dos 167 municípios do estado, sendo mais rigoroso na região central do estado, incluindo o litoral setentrional, caracterizados por irregulares espaço-temporais das chuvas e balanço hídrico deficitário. O subúmido seco ocorre nas regiões do oeste e agreste potiguar. E a região mais úmida do Rio Grande do Norte localiza-se no litoral oriental, onde as chuvas são mais frequentes (Lucena *et al.*, 2018; Silva *et al.*, 2022).

2.2. Conjunto de dados utilizado na modelagem

Os dados de precipitação utilizados para treinamento dos modelos preditivos foram extraídos da plataforma da Empresa de Pesquisa Agropecuária do Estado do Rio Grande do Norte (EMPARN), a qual reúne informações provenientes de telepluviômetros e de estações pluviométricas automáticas (PCDs - Plataforma de Coleta de Dados). Foram selecionados dados provenientes das medições referentes ao ano de 2023, classificado como um ano normal (habitual), segundo dados da EMPARN. Durante a organização da base de dados, foi realizada

a retirada de registros incompletos, restando 198 postos com valores de precipitação total anual, distribuídos ao longo do estado do Rio Grande do Norte (Figura 1).

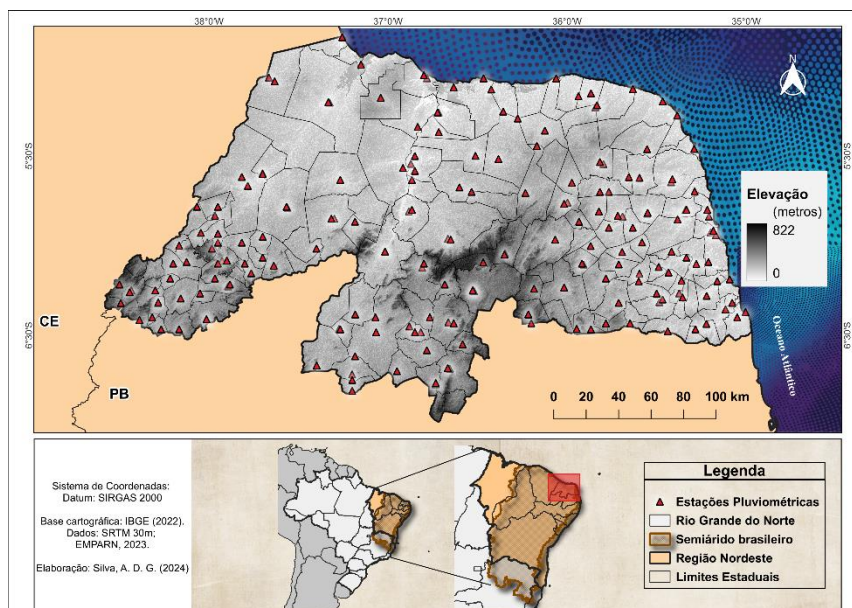


Figura 1. Mapa de localização e distribuição das estações pluviométricas.
Fonte: pesquisa (2024)

A plataforma de processamento em nuvem Google Earth Engine (GEE) foi a principal ferramenta para a execução dos modelos preditivos de machine learning, por meio da linguagem *JavaScript*. A precipitação foi considerada como variável alvo para o modelo de regressão, ou seja, variável dependente. A seleção das covariáveis independentes baseou-se nos critérios das características de relevo, posição geográfica, condições atmosféricas e cobertura vegetal, as quais contribuem para entender a intensidade da precipitação, são elas: elevação, declive, aspecto, sombreamento, NDVI, SAVI, latitude e longitude, temperatura de superfície, distâncias, temperatura do ar, índice de aridez e umidade relativa do ar (Quadro 1).

Conforme Hengl *et al.* (2018), o algoritmo RF não faz autocorrelação espacial, dessa forma, para que ocorra conexão espacial, a adição de informações sobre a posição geográfica e a distância euclidiana entre os pontos observados e o centro da área de estudo, bem como, a influência da maritimidade são opções de covariáveis para considerar a dependência espacial do fenômeno estudado. Também no GEE, as distâncias foram calculadas considerando os limites leste e norte como distintos, bem como, o centro do estado.

Dando importância à influência de fatores topográficos nas dinâmicas climáticas (Pereira, 2013; Souza, Almeida e Correa, 2015), utilizou-se o modelo digital de elevação (MDE) da missão SRTM, e foram extraídos os elementos: altitude, declividade, sombreamento e aspecto do relevo. A temperatura de superfície (*Land Surface Temperature* – LST) reflete a energia térmica emitida pela superfície terrestre, comumente usada para predição de precipitação (Alvarez *et al.*, 2014), foi gerado conforme equação descrita em Jesus e Santana (2017).

O índice de aridez (Ia) consiste em uma medida para quantificar o grau de aridez de um local e está diretamente relacionado ao volume de precipitação pluviométrica, tornando-se crucial para compreender as condições climáticas (Dajoz, 1978). Adotou-se a metodologia de De Martonne, que considera a relação entre a precipitação e a temperatura anuais. Os dados foram obtidos na plataforma da EMPARN. Tendo em vista a resposta da vegetação à entrada de água pluvial (Jensen, 2009), índices de vegetação foram utilizados para treinar o algoritmo. Os índices NDVI



e SAVI foram calculados a partir das imagens do LANDSAT-8. Para o ano de 2023, a coleção foi reduzida pela técnica da mediana dos pixels, resultando em uma imagem síntese, da qual se extraíram as bandas do espectro visível vermelha e infravermelha próxima (B4 e B5).

Abreviação	Nome	Tipo de dado	Descrição
Temp	Temperatura	Climático	Temperatura coletada em superfície
Elevation	Altitude	Topográfico	Altitude em relação ao nível do mar
Slope	Declividade	Topográfico	Declividade do terreno
hillshade	Sombreamento	Topográfico	Simulação de sombreamento causado pela direção da vertente em relação a iluminação solar
Aspect	Aspecto	Topográfico	Direção da orientação das vertentes
Longitude e latitude	Coordenadas geográficas	Coordenada geográfica	Localização terrestre a partir de um sistema espacial de referência.
dist_Leste, dist_Norte e dist_Centr	Distâncias relativas	Distância	Distância espacial
Ia	Índice de Aridez	Climático	Grau de aridez do local
UR	Umidade relativa do ar	Climático	Quantidade de vapor de água presente no ar
NDVI	Índice de Vegetação por Diferença Normalizada	Bioclimático	Estimativa da quantidade de vegetação
SAVI	Índice de Vegetação Ajustado ao Solo	Bioclimático	Estimativa da quantidade de vegetação
LST	Temperatura média da superfície	Climático	Temperatura da superfície estimada por satélite

Quadro 1. Descrição das covariáveis utilizadas na modelagem.

Fonte: pesquisa (2024)

2.3. Processamento dos dados (Modelagem)

O conjunto de covariáveis foi empilhado e integrado em uma única estrutura. Os 198 pontos de estações pluviométricas foram tratados como amostras, extraindo-se os valores para cada camada das covariáveis. As amostras foram subdivididas aleatoriamente em 80% para treinamento e 20% para avaliar a performance do modelo. O conjunto de treino foi aplicado ao RF, o qual foi ajustado para o modo *regression*, com número de 500 árvores e sem limites para o valor máximo de nós. Os dados de teste foram utilizados para avaliar o desempenho dos modelos gerados. Como critérios de desempenho, empregou-se duas medidas estatísticas: erro absoluto médio (*Mean Absolute Error* - MAE) e o coeficiente de determinação (r^2). Em suma, o MAE representa a média dos valores absolutos dos erros entre os valores estimado e real, em que quanto menor o valor, melhor o ajuste do modelo aos dados observados. O r^2 , por sua vez, refere-se à correlação entre o dado real e o previsto, variando entre 0 e 1.

Para contornar a aleatoriedade inerente ao algoritmo na etapa de seleção de amostras de treino e validação, e com o intuito de reduzir incertezas da predição, o *script* desenvolvido executou todas as etapas da modelagem em 100 repetições (Paes *et al.*, 2023), ou seja, o modelo foi treinado e avaliado diversas vezes usando uma semente (*seed*) aleatória a cada iteração. Além disso, os resultados foram armazenados para cálculo do resultado médio. Sendo assim, o mapa de predição final é constituído do valor médio das estimativas realizadas em cada pixel de cada repetição. Como medida da incerteza das rodadas de predição, foi gerado um mapa com o desvio padrão dos valores de cada pixel. O resultado das métricas de desempenho também corresponde ao valor médio das rodadas. Tendo em vista que múltiplas covariáveis e suas combinações são importantes para aprimorar a performance da modelagem, utilizou-se o



recurso *Feature Importance* extraído do RF para verificar a importância de cada camada no modelo de predição e sua influência na variável alvo.

3. RESULTADOS E DISCUSSÃO

O algoritmo Random Forest realizou a predição espacial da precipitação de 2023 para o estado do RN de maneira satisfatória. Considerando a análise das métricas avaliativas de desempenho do modelo preditivo, os valores estimados explicam 72% da precipitação observada, conforme resultado de r^2 . Quanto à diferença média entre os valores observados e preditos, o modelo apresenta erro médio absoluto (MAE) 114,86 mm de precipitação (Figura 2a).

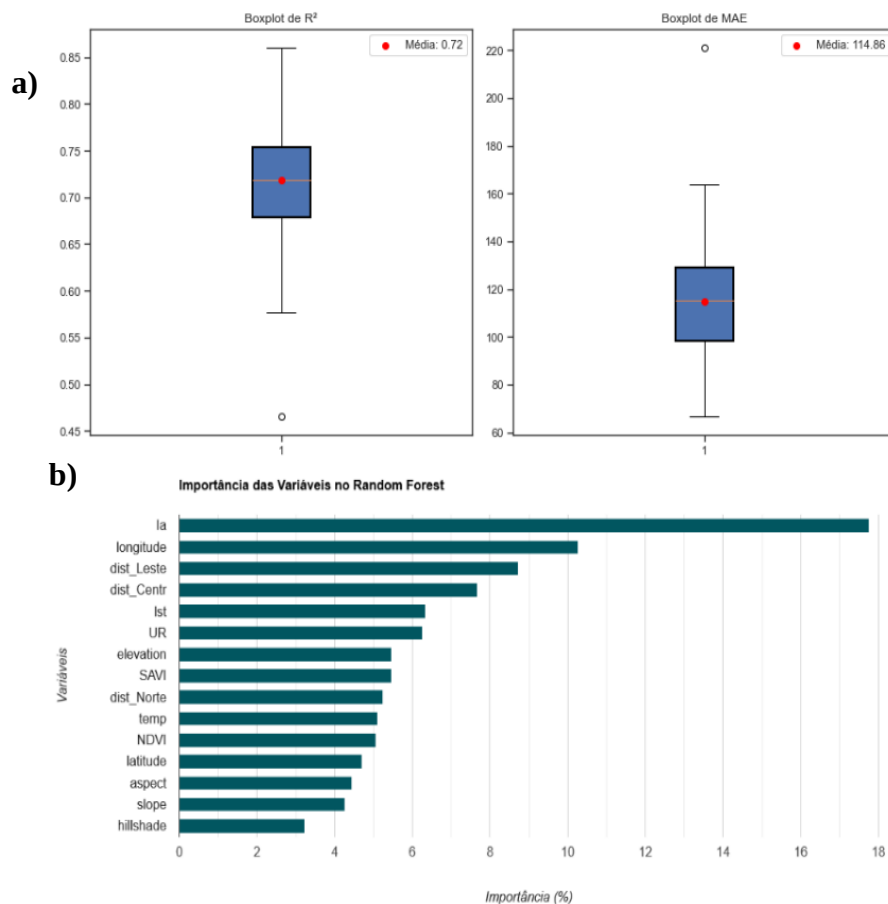


Figura 2. Valores de r^2 e MAE (a) dos 100 modelos preditivos gerados e importância das covariáveis (b). Fonte: pesquisa (2024)

As 15 covariáveis utilizadas na predição foram organizadas em ordem de importância conforme apresentado na Figura 2b. Dentre elas, o índice de aridez apresentou a maior influência para os resultados da modelagem, seguido pela longitude e as distâncias centro e leste. As variáveis topográficas, como o aspecto, declividade e sombreamento apresentaram as menores relevâncias no processamento do algoritmo. As covariáveis elencadas como menos importantes pouco influenciam o modelo gerado através do RF, constituindo menor peso durante a definição dos nós das árvores de decisão. Este resultado sugere que, considerando os fatores geográficos do clima, o contexto posicional, está entre os principais fatores para se compreender a variabilidade da chuva no estado. Ou seja, a posição geográfica, em especial a longitude e distância do leste mostraram-se mais importantes do que fatores relacionados ao relevo, na distribuição espacial das chuvas no território do RN.

Diante do mapa médio gerado (Figura 3), percebe-se a variabilidade espacial das precipitações estimadas. Os maiores valores predominam no litoral leste, considerada a região mais chuvosa do estado. Ressalta-se os resultados dos municípios de Natal e Baía Formosa em que foram computados 1326,8 mm e 1437,1 mm, respectivamente (Figura 3a). Tais informações são condizentes com a literatura, uma vez que os efeitos da brisa marítima e de sistemas atmosféricos, como os DOLs, contribuem para a regularidade e maior volume de precipitação, proporcionando um clima litorâneo úmido (Silva; Lúcio, 2015, Neves; Alcântara; Souza, 2016). Em contrapartida, Touros apresentou precipitação real de 253,6 mm e foi superestimada pelo modelo em 572,1 mm, contudo, cabe considerar a extensão e disposição territorial do município situado numa faixa de transição climática (Figura 3b).

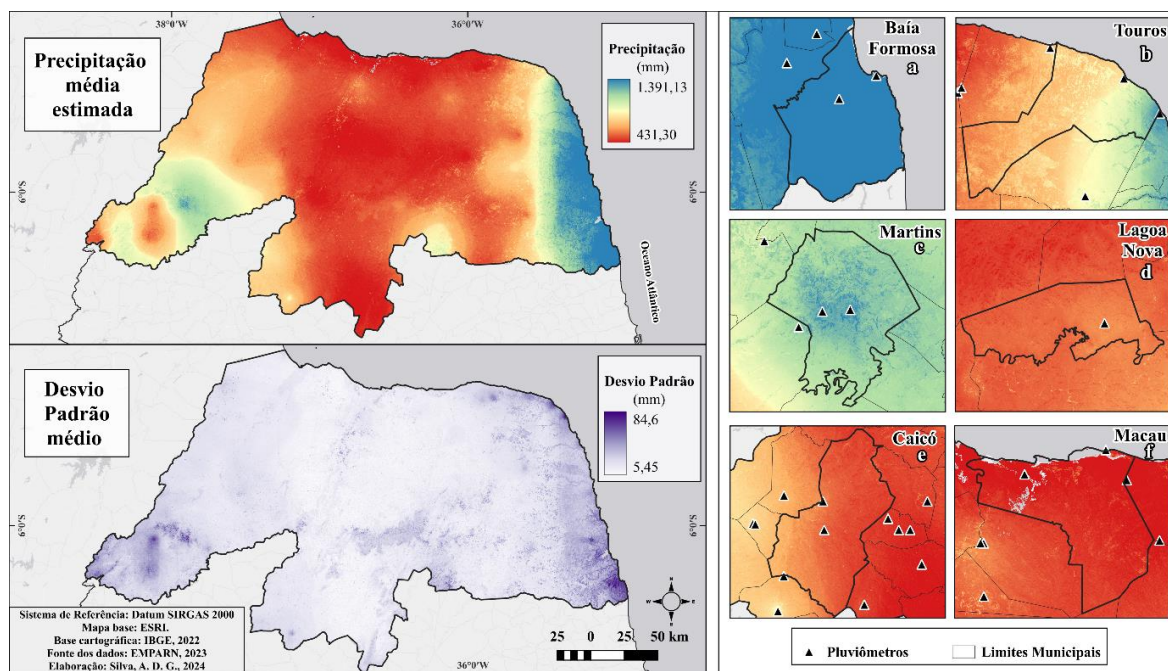


Figura 3. Mapa médio de precipitação estimada e desvio padrão das predições.

Fonte: pesquisa (2024)

Os resultados também apontam maiores valores de precipitação em algumas áreas serranas no sudoeste do estado, computando 1214,3 mm no município de Martins, que se localiza a 735 m de altitude (Figura 3c). A Serra de Martins atua como barreira para a convecção dos ventos e da umidade, formando as chuvas orográficas (Medeiros; Cestaro; Queiroz, 2021). Neste caso, as condições topográficas como a altitude e a posição a barlavento são importantes no aumento nos totais pluviométricos e característica climática subúmida. Porém, este padrão não é observado em Lagoa Nova, situada na Serra de Santana a 719 m de altitude, na mesorregião central, que teve valor estimado em 701 mm pela predição (Figura 3d). Esta observação evidencia que as condições topográficas por si só não condicionam o quantitativo de precipitação no estado do RN, assim como explicitado pelo ranking de importância de variáveis. Ademais, a modelagem conseguiu distinguir as serras consideradas subúmidas, Serra de Martins, das serras secas, Serra de Santana. Dados climatológicos do RN têm mostrado que a região central do estado, onde está situada a serra de Santana é mais seca que a região oeste, onde está localizada a serra de Martins (Amorim *et al.*, 2020; Monteiro *et al.*, 2024).

Baixas precipitações foram analisadas na porção central, localizado predominantemente na depressão sertaneja, e ao norte na bacia sedimentar da costa litorânea - áreas em que historicamente a semiaridez é mais severa (Silva *et al.*, 2022). No centro-sul, a exemplo do



município de Caicó, a precipitação estimada foi 635,2 mm (Figura 3e). Valores semelhantes foram encontrados por Lucena *et al.* (2018) e Silva *et al.* (2022), que utilizaram médias temporais de precipitação. Todavia, distanciando-se a oeste do efeito sotavento do planalto da Borborema, ocorreu no município de Timbaúba dos Batistas o aumento da precipitação estimada de 831,5 mm e valor real de 862,4 mm. Já no centro-norte, as chuvas estimadas correspondem a 465,3 mm e 454 mm nos municípios Macau e Porto do Mangue, respectivamente (Figura 3f). Segundo Teixeira (2008), o litoral norte do RN, apresenta redução das chuvas em virtude da direção predominante dos ventos associada ao formato côncavo da margem continental.

O mapa de desvio padrão (Figura 3) revela que os locais com maior variabilidade entre as 100 rodadas de predição correspondem aos pontos com registros de precipitação elevada, situados no litoral leste e em áreas de altitude no extremo oeste do estado. Isto é, são áreas em que a incerteza da predição foi maior. O litoral leste apresentou o valor máximo de precipitação registrado pelo pluviômetro localizado no extremo sudeste de 1847,2 mm (dado real), sendo estimado pelo modelo preditivo o valor de 1470,1 mm, e desvio padrão de 73,3 mm entre as rodadas de predição. Em regiões de altitude, na Serra do Martins um pluviômetro registrou 1608 mm (dado real), e a predição resultou em 1237,8 mm, com desvio padrão de 81,8 mm entre as repetições. Para locais que apresentam menor quantitativo de precipitação foi observado menor desvio padrão entre os resultados do modelo preditivo, verificando os municípios de Caicó, Cruzeta, Florânia, Macau e Porto do Mangue (Tabela 1). Até mesmo em relação às precipitações medianas, como em Apodi e Mossoró, no baixo oeste e Pau dos Ferros, no alto oeste, os desvios são baixos.

As informações da Tabela 1 comparam os resultados com a normal climatológica de 1981-2010 e demonstram que a precipitação predita e o desvio padrão estão próximos aos dados históricos, assim como dos registros reais, indicando consistência e boa precisão nas predições do modelo. Com relação aos produtos CHIRPS, embora apresentem uma boa variabilidade dos dados para os municípios, observa-se maiores diferenças aos valores reais de precipitação, como em Florânia (+155,2 mm), Caicó (+151,5 mm), Mossoró (-120,2 mm) e Martins (-563,3 mm).

Localidade	Normal climatológica (mm)	Precipitação 2023 (mm)	Precipitação predita (mm)	Desvio padrão das predições (mm)	CHIRPS 2023 (mm)
Apodi	829,6	664,2	706	20,7	673,4
Baía Formosa	-	1752,1	1437,1	72,8	1408,6
Caicó	629,5	633,5	635,2	10,1	785
Cruzeta	667,3	529,2	559,1	10,4	622,6
Florânia	713,4	580	575,4	10,3	735,2
Lagoa Nova	-	669	701	21,1	607,6
Macau	518,4	459,2	465,3	12,1	564,4
Martins	-	1546,3	1214,3	72,8	983
Mossoró	835,4	767,1	730	13,5	646,9
Natal	1691,2	1454,7	1326,8	44,7	1144,7
Pau dos Ferros	-	811	848,1	21,1	752,1
Porto do Mangue	-	417,9	454	16,1	624,5
Timbaúba dos Batistas	-	862,4	831,5	15,1	782,1
Touros	-	253,6	572,1	81	1043,6

Tabela 1. Dados de precipitação histórica, real e predita e desvio padrão médio.

Fonte: pesquisa (2024)

4. CONCLUSÕES



De maneira geral, embora existam limitações quanto à disponibilidade de dados espacializados para o estado do Rio Grande do Norte, a modelagem realizada pelo Random Forest a partir do conjunto de dados utilizados foi considerada satisfatória. O modelo preditivo elaborado através da abordagem por machine learning conseguiu produzir bons resultados, considerando comparações dos valores estimados com dados de precipitação total anual do ano de 2023, da normal climatológica e, até mesmo, com os produtos CHIRPS.

A partir da classificação de importância de covariáveis para a performance do modelo preditivo é possível concluir que as melhores respostas ocorreram para a aridez e a posição geográfica, e estas, por sua vez, possuem relação com o balanço hídrico (Ia) e à exposição da localidade aos ventos úmidos provenientes de leste (brisa marítima e DOLs), respectivamente.

Conclui-se que a precisão e reprodutibilidade do modelo preditivo subestimou valores de precipitação em regiões que historicamente apresentam maiores registros de valores totais anuais. Além dos valores preditos mais próximos dos observados, os resultados evidenciaram menor desvio padrão, entre as rodadas de predição, em localidades com menores quantitativos de precipitações, a exemplo da região central.

Ressalta-se a utilização do Google Earth Engine, plataforma gratuita de computação avançada em grande escala, como alternativa essencial para realização deste estudo na medida em que possibilita processamento de grandes conjuntos de dados em nuvem.

REFERÊNCIAS

- ALVAREZ, O. et al. Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. **International Journal of Climatology**, v. 34, n. 7, p. 2258–2268, 18 out. 2013.
- AMORIM, A. C. B.; LUCIO, P. S.; SILVA, C. M. S.; RODRIGUES, D. T.; BRITO NETO, F. A. Regionalização e Análise da Tendência da Precipitação do Rio Grande do Norte associados a padrões de TSM. **Revista Brasileira de Meteorologia**, v. 35, p. 1-12, 2020.
- ARAÚJO FILHO, J. C.; MARQUES, F. A.; AMARAL, A. J.; CUNHA, T. J.; SOUZA JÚNIOR, V. S.; GALVÃO, P. V. M. Solos do semiárido: características e estoque de carbono. In: GIONGO, V.; ANGELOTTI, F. **Agricultura de baixa emissão de carbono em regiões semiáridas**. Embrapa. Brasília/DF, 2022, 256 p.
- BREIMAN, L. **Random forests. Machine Learning**. Boston, v.45, n.1, p. 5-32, 2001.
- CORREIA FILHO, W. L. F., OLIVEIRA-JÚNIOR, J. F. de, Santiago, D. B., Terassi, P. M. B., Teodoro, P. E., Gois, G., ... Santos, P. J. (2019). Rainfall variability in the Brazilian northeast biomes and their interactions with meteorological systems and ENSO via CHELSA product. **Big Earth Data**, 3(4), 315–337. <https://doi.org/10.1080/20964471.2019.1692298>
- DAJOZ, R. **Ecologia geral**. Trad. De Francisco M. Guimarães. Petrópolis: Vozes, 1978. 472p.
- GOMES, F. I. B. P.; ZANELLA, M. E. Histórico, causas e características da semiaridez do Nordeste do Brasil. **Geografares**, Vitória, Brasil, v. 3, n. 37, p. 209–233, 2023. DOI: 10.47456/geo.v3i37.41515.
- JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em Recursos Terrestres**. Tradução de J. C. N. Epiphânio. São José dos Campos, SP: Parênteses, 2009. 598



p. (Prentice Hall Series in Geographic Information Science) Tradução de: Remote Sensing of the environment: an earth resource perspective.

JESUS, J. B. D.; SANTANA, I. D. M. Estimation of Land Surface Temperature In Caatinga Area Using LANDSAT 8 Data. **Journal of Hyperspectral Remote Sensing**, v. 7, n. 3, p. 150, 16 nov. 2017. DOI: 10.29150/jhrs.v7.3.p150-157

HENGL, T. et al. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, v. 6, p. e5518, 29 ago. 2018. DOI: 10.7717/peerj.5518

LANDIM, P.M.B. **Introdução aos métodos de estimação espacial para confecção de mapas**. Rio Claro: UNESP, 2000. 20p.

LUCENA, R. L.; CABRAL JÚNIOR, J. B.; STEINKE, E. T. Comportamento Hidroclimatológico do Estado do Rio Grande do Norte e do Município de Caicó. **Revista Brasileira de Meteorologia**, v. 33, n. 3, p. 485–496, 2018.

MEDEIROS, J. F. DE; CESTARO, L. A.; QUEIROZ, L. S. CARACTERIZAÇÃO CLIMÁTICA DA SERRA DE MARTINS-RN. *Revista de Geociências do Nordeste*, p. 92–100, 2 ago. 2021. <https://doi.org/10.21680/2447-3359.2021v7n2ID21397>

MONTEIRO, F., LUCENA, R. L., FARIA, R. M., & REIS, J. S. (2024). Análisis espacial de las precipitaciones en el Estado de Río Grande do Norte, Brasil. **Revista Geográfica De América Central**, 1(72). <https://doi.org/10.15359/rgac.72-1.2>

NEVES, D. J. D.; ALCÂNTARA, C. R.; SOUZA, E. P. DE. Estudo de Caso de um Distúrbio Ondulatório de Leste sobre o Estado do Rio Grande do Norte - Brasil. *Revista Brasileira de Meteorologia*, v. 31, n. 4, p. 490–505, dez. 2016.

PAES, E. C.; VELOSO, G. V.; SILVA, D. L. A.; FERNANDES-FILHO, E. I.; FONTES, M. P. F.; SOARES, E. M. B. Use of modeling to map potentially toxic element and assess the risk to human health in soils affected by mining activity. *Catena*, v. 220, p. 1-19, 2023. DOI: 10.1016/j.catena.2022.106662

PEREIRA, T. Influência do relevo na precipitação das regiões hidrográficas do litoral norte de alagoas. **GEOUSP-espaco e tempo**, n.33, p.239-253, 2013.

SILVA, B. K. N.; LUCIO, P. S. Characterization of risk/exposure to climate extremes for the Brazilian Northeast—case study: Rio Grande do Norte. **Theoretical and applied climatology**, v. 122, n. 1–2, p. 59–67, 2015. DOI: <https://doi.org/10.1007/s00704-014-1275-z>.

SILVA, A. D. G. da; SANTOS, A. L. B.; SANTOS, J. M.; LUCENA, R. L. Balanço hídrico climatológico e classificação climática do estado do Rio Grande do Norte. **Revista Brasileira de Climatologia**, [S. l.], v. 30, n. 18, p. 798–816, 2022.

SOUZA, J.O.P.; ALMEIDA, J.D.M.; CORREA, A.C.B. Caracterização e espacialização em bacia hidrográfica com relevo complexo: Sertão Central pernambucano –bacia do Riacho do Saco. **Revista de Geografia (UFPE)**, v.32, n.2, p.106-126, 2015.

TEIXEIRA, R. F. B. O fenômeno da brisa e sua relação com a chuva sobre Fortaleza-CE. **Revista Brasileira de Meteorologia**, v.23, n.3, 2008. p. 282-29.

ZHANG, T. *et al.* Merging multisatellite precipitation products using stacking method and the censored-shifted gamma ensemble model output statistics in china's Beimiaoji basin. **Journal of hydrology**, v. 618, p. 129263–129263, 1 mar. 2023. DOI: 10.1016/j.jhydrol.2023.129263