Revista Brasileira de Climatologia
Brazilian Journal of Climatology
OPEN ACCESS

ALGORITMO DE MACHINE LEARNING PARA MAPEAMENTO DAS CHUVAS EXTREMAS EM UMA REGIÃO DO NORDESTE BRASILEIRO

Machine Learning Algorithm for Mapping Extreme Rainfall in a Region of Northeast Brazil

Algoritmo de Aprendizaje Automático para el Mapeo de Lluvias Extremas en una Región del Noreste de Brasil

Alexsander Souza da Costa
Universidade Federal da Grande Dourados
alexsander0souza@gmail.com

Elias Silva de Medeiros
Universidade Federal da Grande Dourados
eliasmedeiros@ufgd.edu.br

Vanderson Hafemann Fragal
Universidade Federal da Grande Dourados
vandersonfragal@ufgd.edu.br

Alessandra Querino da Silva
Universidade Federal da Grande Dourados
alessandrasilva@ufgd.edu.br

Carolina Cristina Bicalho
Universidade Estadual de Mato Grosso do Sul
carolinabicalho@gmail.com

Resumo: A precipitação máxima diária é um evento climático extremo que desempenha um papel significativo na compreensão das mudanças climáticas. Neste estudo, o objetivo foi analisar a precipitação máxima diária no estado da Paraíba, como forma de compreender as mudanças climáticas e apresentar ferramentas que auxiliem na minimização dos desastres relacionados a esse fenômeno. Para isso, foi utilizado um algoritmo de aprendizagem de máquina supervisionado para construir mapas em alta resolução da precipitação máxima. A base de dados para este estudo foi construída a partir de informações fornecidas pela Agência Executiva de Gestão das Águas da Paraíba, contendo dados de 123 estações pluviométricas, abrangendo o período de 1994 a 2021. Para modelagem dos

dados foram realizados testes estatísticos e ajustada a distribuição Gumbel, com o objetivo de determinar os resultados relevantes. O algoritmo escolhido para a construção dos mapas em alta resolução foi o Support Vector Machine, uma técnica capaz de classificar e realizar regressão de dados, além de permitir a interpolação de dados não rotulados. Por meio deste estudo, foi possível obter estimativas da precipitação máxima diária para todo o estado da Paraíba em períodos de retorno de 2, 5, 10 e 50 anos. Essas estimativas têm grande importância para o planejamento de áreas de risco e para o desenvolvimento de sistemas de alerta de inundações, visando proteger a população e reduzir os impactos econômicos e sociais associados a eventos extremos de chuva.
Palavras-chave: Precipitação. Desastre Ambiental. Evento Climático. Gumbel.

Abstract: Daily maximum precipitation is an extreme weather event that plays a significant role in understanding climate change. In this study, the objective was to analyze daily maximum precipitation in the state of Paraíba, Brazil, in order to comprehend climate change and present tools to aid in minimizing disasters related to this phenomenon. For this purpose, a supervised machine learning algorithm was used to construct high-resolution maps of maximum precipitation. The database for this study was built from information provided by the Executive Agency for Water Management of Paraíba, containing data from 123 rain gauge stations spanning the period from 1994 to 2021. Statistical tests were performed to model the data and the Gumbel distribution was adjusted to determine relevant results. The Support Vector Machine algorithm was chosen for constructing high-resolution maps, a technique capable of classifying and performing data regression, as well as allowing for the interpolation of unlabeled data. Through this study, estimates of daily maximum precipitation were obtained for the entire state of Paraíba for return periods of 2, 5, 10, and 50 years. These estimates are of great importance for risk area planning and for the development of flood alert systems, aiming to protect the population and reduce the economic and social impacts associated with extreme rainfall events.
Keywords: Precipitation. Environmental Disaster. Climatic Event. Gumbel Distribution.

Resumen: La precipitación máxima diaria es un evento climático extremo que desempeña un papel significativo en la comprensión del cambio climático. En este estudio, el objetivo fue analizar la precipitación máxima diaria en el estado de Paraíba, Brasil, con el fin de comprender el cambio climático y presentar herramientas para ayudar a minimizar los desastres relacionados con este fenómeno. Para ello, se utilizó un algoritmo de aprendizaje automático supervisado para construir mapas de alta resolución de la precipitación máxima. La base de datos para este estudio se construyó a partir de información proporcionada por la Agencia Ejecutiva de Gestión del Agua de Paraíba, que contiene datos de 123 estaciones de medición de lluvia que abarcan el período de 1994 a 2021. Se realizaron pruebas estadísticas para modelar los datos y se ajustó la distribución Gumbel para determinar resultados relevantes. Se eligió el algoritmo de Support Vector Machine para la construcción de mapas de alta resolución, una técnica capaz de clasificar y realizar regresión de datos, así como permitir la interpolación de datos no etiquetados. A través de este estudio, se obtuvieron estimaciones de la precipitación máxima diaria para todo el estado de Paraíba para períodos de retorno de 2, 5, 10 y 50 años. Estas estimaciones son de gran importancia para la planificación de áreas de riesgo y para el desarrollo de sistemas de alerta de inundaciones, con el objetivo de proteger a la población y reducir los impactos económicos y sociales asociados con eventos extremos de lluvia.
Palabras clave: Precipitación. Desastre Ambiental. Evento Climático. Distribución Gumbel.

Submetido em: 31/07/2023
Aceito para publicação em: 24/02/2024
Publicado em: 25/02/2024

1. INTRODUÇÃO

Eventos climáticos extremos são caracterizados por apresentarem um comportamento atípico em relação aos níveis considerados normais em um determinado local ou região. Esses eventos podem ocorrer em intervalos temporais de horas e até mesmo de décadas, ocasionando secas prolongadas, chuvas torrenciais, altas temperaturas, geadas, dentre outros (MARENGO; TORRES; ALVES, 2017; CARVALHO *et al.*, 2020). Dentre os eventos extremos climáticos, a precipitação máxima diária é um dos aspectos mais relevantes para a compreensão das mudanças climáticas. Esse parâmetro se refere à quantidade acumulada de chuva registrada em um intervalo de tempo, e é utilizado para avaliar a intensidade e frequência desse evento.

Os altos volumes de chuvas em intervalos de 24 horas podem ocasionar diversos transtornos à população de uma região, que vai desde inundações em áreas urbanas até perdas na produção agrícola em áreas rurais (MEDEIROS; ALVES; SOUZA, 2019). Um estudo realizado no município de Colatina, Espírito Santo, apontou que os períodos com maiores intensidades de chuvas aumentam consideravelmente os riscos de desastres naturais, tais como inundações, enchentes e alagamentos (PINTO *et al.*, 2017). Diante disso, uma análise detalhada das chuvas máximas é de fundamental importância para subsidiar ações que possam minimizar, e até mesmo impedir, desastres causados por este tipo de evento, oportunizando medidas de planejamento de áreas de risco e alertas de inundações (SOUZA *et al.*, 2022a).

O foco da presente pesquisa foi conduzir um estudo sobre o comportamento pluviométrico a partir de dados de chuvas diárias registradas no estado da Paraíba, localizado na Região Nordeste do Brasil (NEB). A NEB apresenta uma grande variabilidade espacial e temporal da precipitação pluviométrica, sendo comum a ocorrência de eventos extremos que podem ser atribuídos ao clima local e à atuação de diversos sistemas meteorológicos (ABREU; MUTTI; LIMA, 2019). Conforme o Atlas de Desastres Naturais (CEPED, 2013), o Estado da Paraíba contabilizou 215 ocorrências oficiais de enxurradas entre os anos de 1991 e 2012. Esses eventos afetaram mais de 350 mil pessoas. Além disso, o estado registrou 136 casos oficiais de inundação, impactando, aproximadamente, em 241 mil pessoas. O estudo desses eventos extremos é fundamental para o desenvolvimento de estratégias de adaptação e

mitigação das mudanças climáticas, visando minimizar seus impactos socioeconômicos e ambientais.

Os métodos estatísticos mais comuns adotados para modelar dados de chuvas máximas incluem a análise descritiva e o ajuste de distribuições de probabilidade. Entre as distribuições mais utilizadas, destaca-se a distribuição de Gumbel, que é uma das distribuições de valor extremo mais amplamente aplicadas na modelagem de eventos extremos de chuvas, pois os resultados indicam um bom ajuste da distribuição aos dados (MEDEIROS *et al.*, 2019; SOUZA *et al.*, 2022b). MEDEIROS *et al.* (2019) modelaram dados de chuvas máximas diárias no município de Jataí, Goiás, e utilizando de uma série histórica de 39 anos, concluíram que a distribuição Gumbel mostrou-se adequada para ajustar aos dados referentes aos meses mais chuvosos na região. SOUZA *et al.* (2022a) ajustaram a distribuição Gumbel aos dados de chuvas máxima diária em 238 estações pluviométricas localizadas no estado da Paraíba, verificando que essa distribuição se ajustou adequadamente para todas as estações quando considerado os meses do primeiro semestre.

No entanto, esses estudos supracitados apresentam algumas limitações. Uma das principais limitações é que eles fornecem estimativas apenas para os locais onde existem estações pluviométricas. Isso significa que eles não consideram a variabilidade espacial da precipitação e podem subestimar o risco de eventos extremos em áreas sem estações pluviométricas. Já para estimar valores de precipitação máxima em locais onde não há medição direta ou onde a densidade de estações pluviométricas é baixa, sendo utilizada a interpolação espacial. SILVA *et al.* (2012) mostraram que a krigagem ordinária pode ser aplicada para interpolação da precipitação pluviométrica mensal no estado do Espírito Santo, Brasil. A krigagem é uma técnica comumente aplicada em geoestatística para estimar valores em locais não amostrados com base em observações em pontos amostrados (IMAI *et al.,* 2003). Entretanto, a utilização dessa metodologia geoestatística pressupõe um campo aleatório gaussiano, o que não pode ser considerado no caso de valores extremos. Como alternativa, tem-se os algoritmos baseados em *Machine Learning*, que por sua vez são flexíveis aos pressupostos estatísticos.

As técnicas de *Machine Learning* são preferíveis aos métodos geoestatísticos quando modelada variáveis que apresentam alta variabilidade espacial, não sendo necessário a suposição de estacionariedade de primeira e segunda ordem e, consequentemente, o ajuste

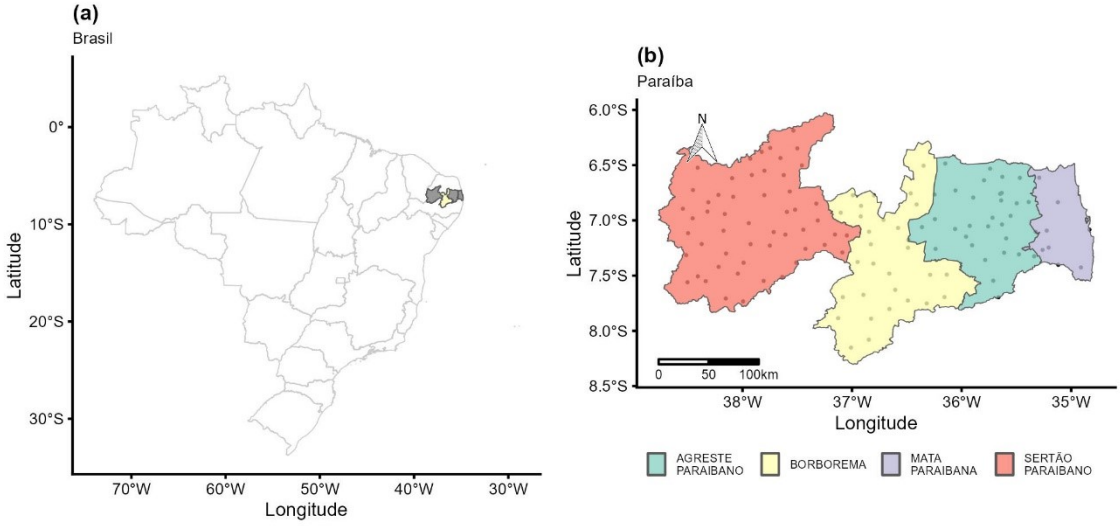
de variogramas teóricos (SEKULIĆ *et al.*, 2020). Adicionalmente, a krigagem pode exigir uma alta demanda computacional no processo do ajuste do modelo geoestatístico, não sendo garantido que esse modelo ajuste adequadamente a qualquer tipo de dado (HENGL *et al.*, 2018).

Neste contexto, a presente pesquisa teve por objetivo a aplicação de uma técnica de *Machine Learning* para mapeamento e modelagem da precipitação máxima diária em todo o Estado da Paraíba, Brasil. Ressalta-se que as estimativas dos níveis de retorno e demais informações obtidas nesse estudo possuem grande relevância e potencial para serem utilizadas em planejamentos (estadual e municipal) mais eficientes. Estudos como os aqui conduzidos são de grande importância, pois viabilizam a obtenção de previsões de danos causados por chuvas extremas em locais onde não existam estações pluviométricas. Isso é relevante visto que grandes quantidades de chuvas em um período curto de tempo podem resultar em inundações e enchentes, até mesmo em regiões com menor capacidade hídrica.

2. METODOLOGIA

Os dados foram obtidos de 123 estações pluviométricas localizadas no Estado da Paraíba (Figura 1b) fornecidos pela Agência Executiva de Gestão das Águas do Estado da Paraíba (AESA - http://www.aesa.pb.gov.br/).

Figura 1 - Mapas do Brasil com a divisão das unidades federativas (a) e do Estado da Paraíba com destaque às mesorregiões e da distribuição das 123 estações pluviométricas (b).



Fonte: Elaborado pelos autores (2023).

As estações pluviométricas foram selecionadas pelo critério de registro da precipitação diária, sendo assim, estações que não apresentam registros em todos os dias do ano foram descartadas para essa análise. Para construir a base de dados foi extraída a precipitação máxima diária anual (PMDA) para cada uma das 123 localizações. Ao todo, foram obtidas 3.444 observações, sendo 28 registros para cada estação, compreendendo o período de janeiro de 1994 a dezembro de 2021.

Sabe-se que a proximidade do oceano atlântico pode influenciar nos níveis de chuvas de uma região. Diante disso, foi calculada a distância de cada estação pluviométrica até o litoral utilizando a fórmula de haversine, sendo essa distância incorporada no conjunto de dados como uma covariável. A fórmula de haversine calcula a distância mais curta entre dois pontos em uma esfera usando suas latitudes e longitudes medidas ao longo da superfície (DAUNI *et al.*, 2019). O ponto selecionado como litoral foi o Ponto de Seixas (7°09'07.2"S e 34°47'34.4"W). O haversine do ângulo central (Equação 1) é calculado pela seguinte fórmula:

$$\left(\frac{d}{r}\right) = haversine(\emptyset_2 - \emptyset_1) + cos\ cos\ (\emptyset_1)\ \ cos\ cos\ (\emptyset_2)\ haversine(\gamma_2 - \gamma_1). \quad (1)$$

Na Equação (1) tem-se que *r* é o raio da terra (6.371 km), *d* é a distância entre dois pontos, ϕ_1, ϕ_2 é a latitude dos pontos e γ_1, γ_2 é a longitude dos dois pontos, respectivamente. Resolvendo *d* aplicando a função seno inversa (Equação 2), obtemos:

$$d = 2rsin^{-1}\sqrt{sin^2\left(\frac{\emptyset_2 - \emptyset_1}{2}\right) + cos\ cos\ (\emptyset_1)\ \ cos\ cos\ (\emptyset_2)\ sin^2\left(\frac{\gamma_2 - \gamma_1}{2}\right)}. \quad (2)$$

A distribuição Gumbel é uma distribuição de probabilidade contínua que é comumente utilizada para modelar dados extremos, como temperaturas máximas, cotas de enchentes, entre outros. Os parâmetros de locação da distribuição Gumbel são os parâmetros que determinam a posição da curva de distribuição Gumbel. A função de densidade de probabilidade da distribuição Gumbel (Equação 3) é dada por:

$$f(x) = \left(\frac{1}{\beta}\right) e^{\left(\frac{x-\mu}{\beta}\right)} e^{-e^{\left(\frac{x-\mu}{\beta}\right)}},\ x\ \in R. \quad (3)$$

Os dois parâmetros são μ e β, em que μ é o parâmetro de locação que determina a posição da curva de distribuição e β é o parâmetro de escala que determina a largura da curva de distribuição. A função da distribuição acumulada (Equação 4) é dada pela seguinte equação:

$$F(x) = e^{-e^{\left(-\frac{x-\mu}{\beta}\right)}},\ x \in R. \quad (4)$$

Nesse estudo foi aplicado o teste de estacionariedade Wald-Wolfowitz (WALD; WOLFOWITZ, 1940). Esse é um teste estatístico utilizado para verificar se uma série temporal é estacionária ou não. A estacionariedade é uma propriedade importante de uma série temporal, pois significa que as suas propriedades estatísticas são constantes ao longo do tempo. É baseado em uma estatística não paramétrica e considera a hipótese nula (H_0) de que a série é estacionária contra uma hipótese alternativa (H_1) de não estacionariedade (SOUZA *et al.*, 2022b). O teste consiste na contagem do número de mudanças de direção (*run*) na série temporal. Se a série temporal é estacionária, espera-se que haja uma quantidade razoável de mudanças de direção, e as mudanças de direção serão distribuídas de forma aleatória. Se a série temporal não é estacionária, espera-se que haja uma quantidade relativamente grande ou pequena de mudanças de direção, ou que as mudanças de direção não sejam distribuídas de forma aleatória (BACK, 2001).

Para verificação de independência na série temporal foi utilizado o teste Ljung-Box (LJUNG; BOX, 1978). Esse teste é baseado na ideia de que, em uma série temporal independente, as autocorrelações dos resíduos devem ser próximas de zero para todos os atrasos considerados. Ele verifica se há algum padrão ou dependência entre os valores em uma série ao longo do tempo, baseado na hipótese nula (H_0) de que as observações são independentes. Se a hipótese nula é rejeitada, isso indica que há autocorrelação presente na série. O teste de Ljung-Box é amplamente utilizado em análises financeiras, econômicas e climáticas para determinar se uma série temporal precisa ser transformada antes de ser usada em análises mais avançadas (PEREIRA *et al.*, 2021).

Após serem verificadas a estacionariedade e independência de cada série, a próxima análise consistiu na obtenção das estimativas dos parâmetros da distribuição Gumbel por meio do método da máxima verossimilhança, uma vez que esse método pressupõe que as observações são independentes. Ajustado os parâmetros da distribuição Gumbel, determinou-se as estimativas dos níveis de retorno da precipitação máxima para um período de 10 anos (T=10), sendo essa estimativa obtida pela seguinte expressão:

$$\hat{x}_p = \hat{\beta} - \hat{\mu} \ln\left[\ln\left(\frac{1}{1-p}\right)\right]. \quad (5)$$

Na Equação (5), $\hat{\mu}$ e $\hat{\beta}$ representam as estimativas dos parâmetros da distribuição Gumbel. A probabilidade *p = 1/T*, sendo p = 0,10 para um período de 10 anos. Estimado o nível de retorno da precipitação para cada estação pluviométrica, a análise seguinte consistiu da interpolação espacial utilizando o algoritmo de *Machine Learning* denominado por Support Vector Machines (SVM).

2.1. Support Vector Machines

O SVM é um algoritmo de aprendizado de máquina supervisionado e utilizado para classificação e regressão. O algoritmo busca encontrar o hiperplano (uma superfície de decisão) que melhor separa as classes dos dados de treinamento (MAMMONE; TURCHI; CRISTIANINI, 2009). O SVM é chamado de "Support Vector" porque usa os dados de treinamento mais próximos do hiperplano para definir a sua posição e orientação. Adicionalmente, pode ser usado para classificação binária (em que os dados pertencem a apenas duas classes) e para classificação multiclasse (em que os dados pertencem a mais de duas classes). O SVM também pode ser usado para problemas de regressão, conforme os dados de precipitação do nosso estudo, cujo objetivo é encontrar uma função que melhor se ajuste aos dados de treinamento (MAMMONE; TURCHI; CRISTIANINI, 2009).

Dentro do SVM existem vários argumentos que podem ser ajustados para obter um melhor desempenho do modelo (CHANG; LIN, 2011). Diante disso, foram utilizadas 10 partições na validação cruzada, com o intuito de determinar a melhor combinação das estimativas dos parâmetros “epsilon”, “cost” e “gamma”, quando aplicado o *kernel* “radial basis”. Para detalhes da implementação desse algoritmo consulte o estudo de Suganyadevi e Babulal (2014).

2.2 Shiny

O Shiny (https://shiny.posit.co/) é uma estrutura de desenvolvimento (ou pacote) de aplicativos Web interativos para a linguagem de programação R. Ele permite que os usuários criem interfaces interativas para análise de dados, visualização, modelagem estatística e outras tarefas relacionadas (CASTRO *et al.*, 2023). O Shiny combina a facilidade de uso da

linguagem R com a flexibilidade da Web, permitindo que os usuários construam aplicativos Web que respondam às interações dos usuários em tempo real.

Os aplicativos Shiny são construídos usando uma combinação de código R e elementos HTML (HyperText Markup Language) (CASTRO *et al.*, 2023). A Interface do usuário (UI) é definida usando funções R que criam os elementos como botões, gráficos e tabelas. Os elementos da UI podem ser estilizados usando CSS (Cascading Style Sheets) para personalizar a aparência dos aplicativos. Com a utilização do Shiny, foi possível criar uma página na web com os resultados das análises, disponível em https://mapatcc.shinyapps.io/MapaTCC/.

3. RESULTADOS E DISCUSSÃO

Na Tabela 1 temos os resultados da análise descritiva da PMDA. O estudo foi dividido entre quatro mesorregiões no qual foram analisadas 8 estações pluviométricas na Mata Paraibana, 33 no Agreste Paraibano, 29 na Borborema e 53 no Sertão Paraibano.

Tabela 1 - Resumo estatístico da Precipitação Máxima Diária Anual de 1994 a 2021 em 123 estações pluviométricas localizadas no estado da Paraíba, Brasil.

Mesorregião	Estações (n)	Máximo	Média	Mediana	Desvio Padrão	Coeficiente Variação
Mata Paraibana	8	190,0	92,13	84,30	35,09	38,09%
Agreste Paraibano	33	196,0	63,49	59,05	26,41	41,59%
Borborema	29	218,2	71,25	68,20	31,93	44,81%
Sertão Paraibano	53	284,6	84.,03	80,05	28,82	34,30%

Fonte: Elaborado pelos autores (2023).

O maior valor da precipitação máxima diária foi registrado no Sertão Paraibano com 284,6 mm. Ao analisarmos as estimativas da média e mediana, os maiores índices foram observados na mesorregião da Mata Paraibana, sendo, respectivamente, 92,1 e 84,3. Um estudo sobre a precipitação máxima mensal no principal município localizado nesta mesorregião, João Pessoa, capital do estado, evidenciou uma média 69,5 mm no mês de julho para o período de 1980 a 2019 (SOUZA *et al.*, 2022b).

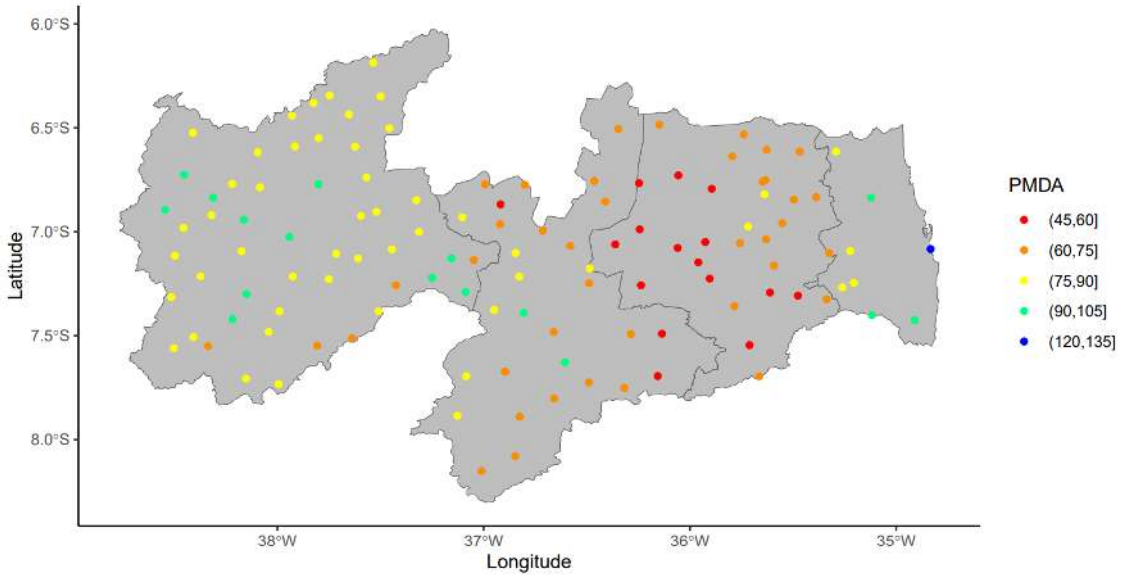
Uma possível justificativa para a ocorrência dos níveis diários de chuvas extremas diz respeito a interferência dos fenômenos atmosféricos. Uma pesquisa realizada no Estado de Goiás, Brasil, investigou as influências dos fenômenos El Niño e La Niña sobre a precipitação

máxima diária, concluindo que os maiores índices ocorreram em anos do El Niño moderado e fraco (MARCUZZO; ROMERO, 2013).

A região do Agreste Paraibano conta com um valor de desvio padrão de 26,4, sendo o menor encontrado entre as regiões. Já em relação ao coeficiente de variação, Borborema conta com a maior variabilidade (44,81%). Um estudo em 2003 mostrou que a variabilidade se torna mais pronunciada na estação seca, que apresenta coeficientes de variação entre 61,37% e 161,08%. A ausência total de chuvas em alguns anos no período da seca, pode ser a explicação dessa variabilidade (SILVA *et al.*, 2003).

A Figura 2 apresenta os resultados da média pluviométrica anual (PMDA) obtidos a partir das estações pluviométricas instaladas em diferentes regiões do estado da Paraíba, Brasil.

Figura 2 - Distribuição espacial da Precipitação Máxima Diária Anual no Estado da Paraíba avaliada em 123 estações pluviométricas no período de 1994 a 2021.



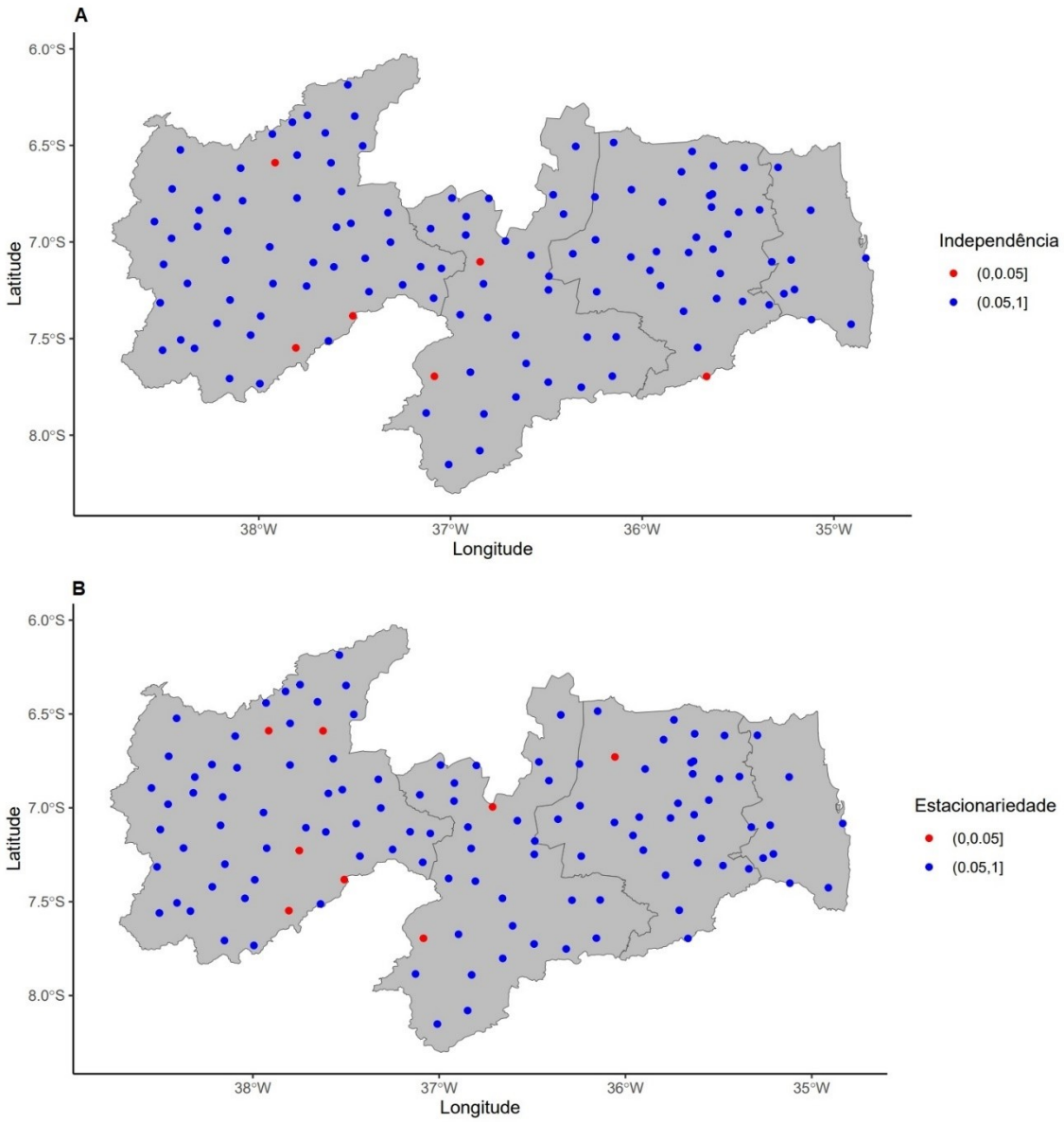
Fonte: Elaborado pelos autores (2023).

Analisando os dados, observa-se que a região do Agreste Paraibano apresenta as menores médias pluviométricas, com valores que variam entre 45 e 60 mm. A região de Borborema também apresenta valores baixos de PMDA, com valores variando entre 60 e 75 mm. Por outro lado, a região do Sertão Paraibano apresenta valores intermediários de PMDA, com valores que variam entre 60 e 105 mm. Já a região da Mata Paraibana apresenta os

maiores valores de PMDA, com o maior valor registrado na estação pluviométrica localizada no litoral, com valores entre 120 e 135 mm. Uma possível explicação para essa diferença é o fato de a região ser litorânea, o que pode influenciar na quantidade de chuvas registradas.

A Figura 3 apresenta os resultados obtidos a partir do teste de independência de Ljung-Box (A) e estacionariedade Wald-Wolfowitz Runs Test (B).

Figura 3 - Distribuição espacial do Teste de Independência (A) e Estacionariedade (B) no Estado da Paraíba avaliada em 123 estações pluviométricas no período de 1994 a 2021.

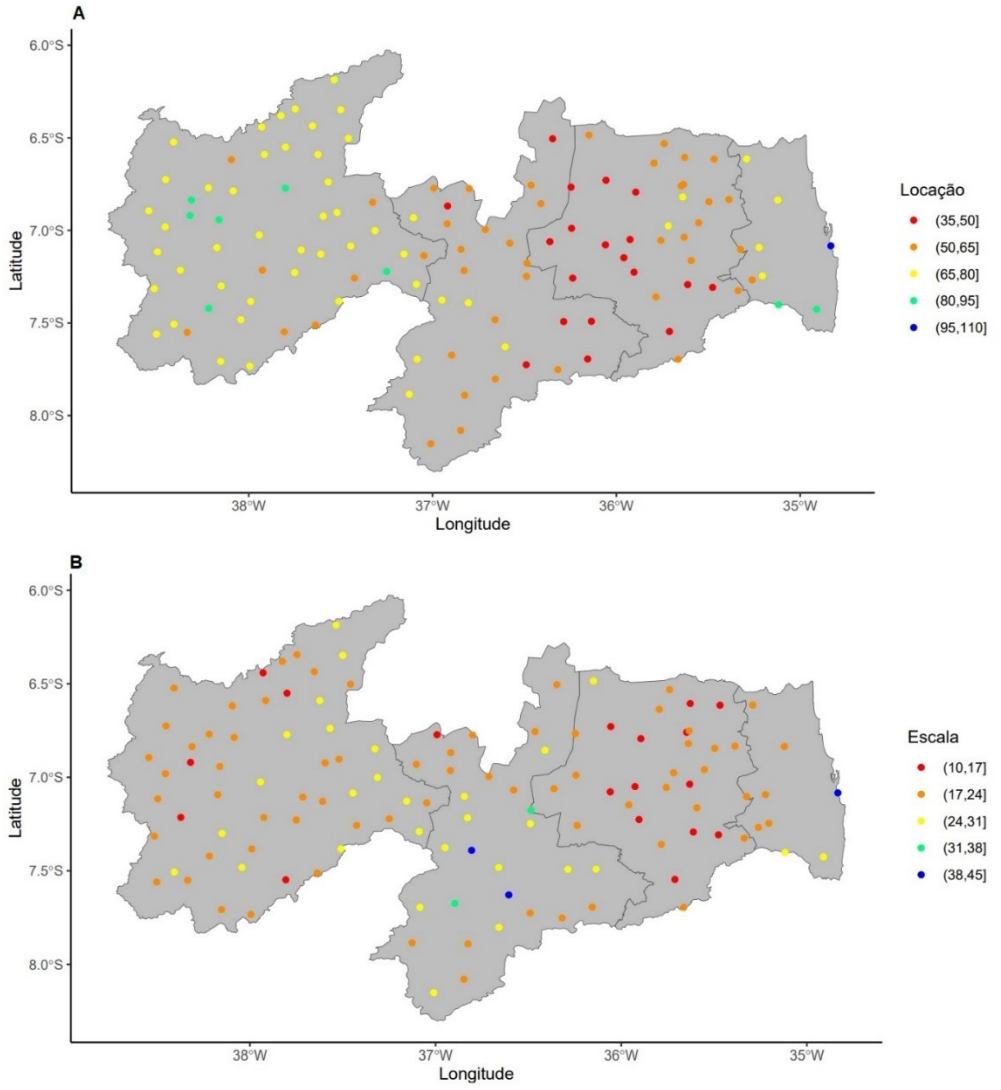


Fonte: Elaborado pelos autores (2023).

Os resultados indicam que somente 4,87% das estações foram rejeitadas no teste de independência, enquanto 95,13% das estações estão dentro dos parâmetros estabelecidos e são consideradas independentes. Em relação ao teste de estacionariedade, conforme os resultados apresentados, 8 estações pluviométricas foram rejeitadas no estudo, sendo que algumas delas também haviam sido rejeitadas no teste de independência anteriormente realizado.

A Figura 4 apresenta os resultados dos parâmetros de locação e escala da distribuição Gumbel. Conforme os resultados, somente uma estação obteve valores de μ (mu) entre 95 e 110, enquanto três estações obtiveram valores de β (beta) entre 38 e 45.

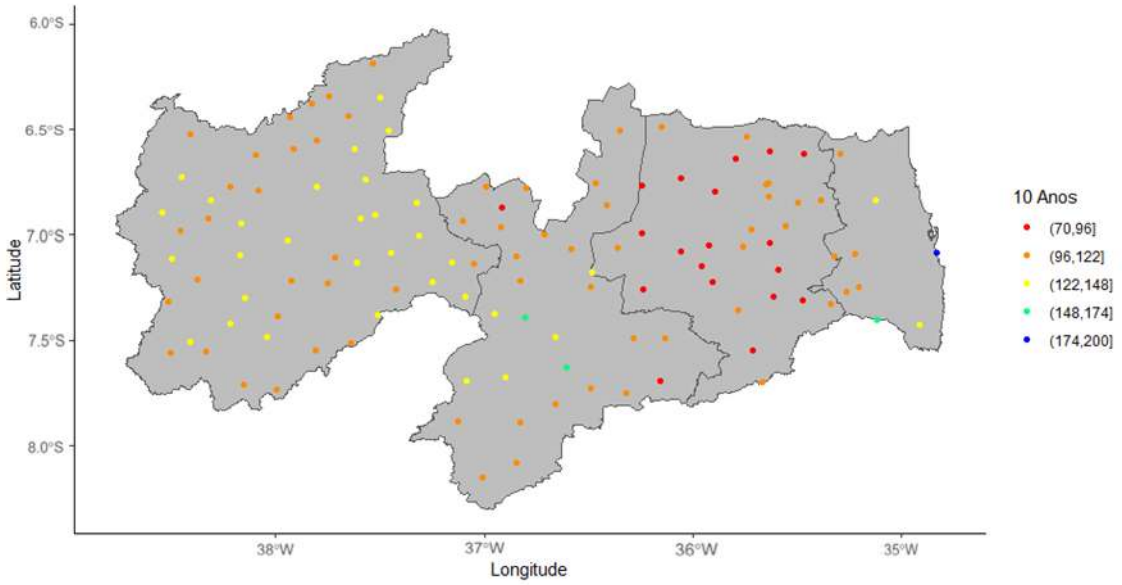
Figura 4 - Distribuição espacial dos Parâmetros de Locação (A) e Escala (B) da distribuição Gumbel no Estado da Paraíba avaliada em 123 estações pluviométricas no período de 1994 a 2021.



Fonte: Elaborado pelos autores (2023).

A Figura 5 representa os níveis de retorno relacionados a um período de 10 anos. Os menores níveis de retorno foram registrados na mesorregião do Agreste Paraibano, com valores variando entre 70 e 96 mm nesse período de 10 anos. Isso significa que para um período de retorno de 10 anos é esperado que ocorra uma precipitação máxima diária de pelo menos 70 mm. Em estações pluviométricas localizadas na Mata Paraibana e no Sertão estão presentes as maiores estimativas da precipitação máxima diária. Adicionalmente, têm-se algumas localidades da Borborema com estimativa diária superior a 148 mm em um único dia do ano. No Estado da Paraíba no período de 1991 a 2012 foram registros 215 ocorrências de enxurradas (inundações bruscas), sendo 32% observados no Sertão e em apenas 7 municípios da mesorregião da Borborema não registraram enxurradas severas (CEPED, 2013). Ainda de acordo com o Atlas Brasileiro de Desastres Naturais foi constatado que ao longo do período examinado que ocorreram 215 incidentes de enxurradas que impactaram significativamente mais de 350 mil indivíduos. Esses eventos resultaram na desestruturação das moradias de 22 mil pessoas, além de desalojar outras 35 mil. Além disso, lamentavelmente, 33 pessoas perderam suas vidas em decorrência dessas ocorrências.

Figura 5 - Distribuição espacial das estimativas do Nível de Retorno de 10 anos no Estado da Paraíba



Fonte: Elaborado pelos autores (2023).

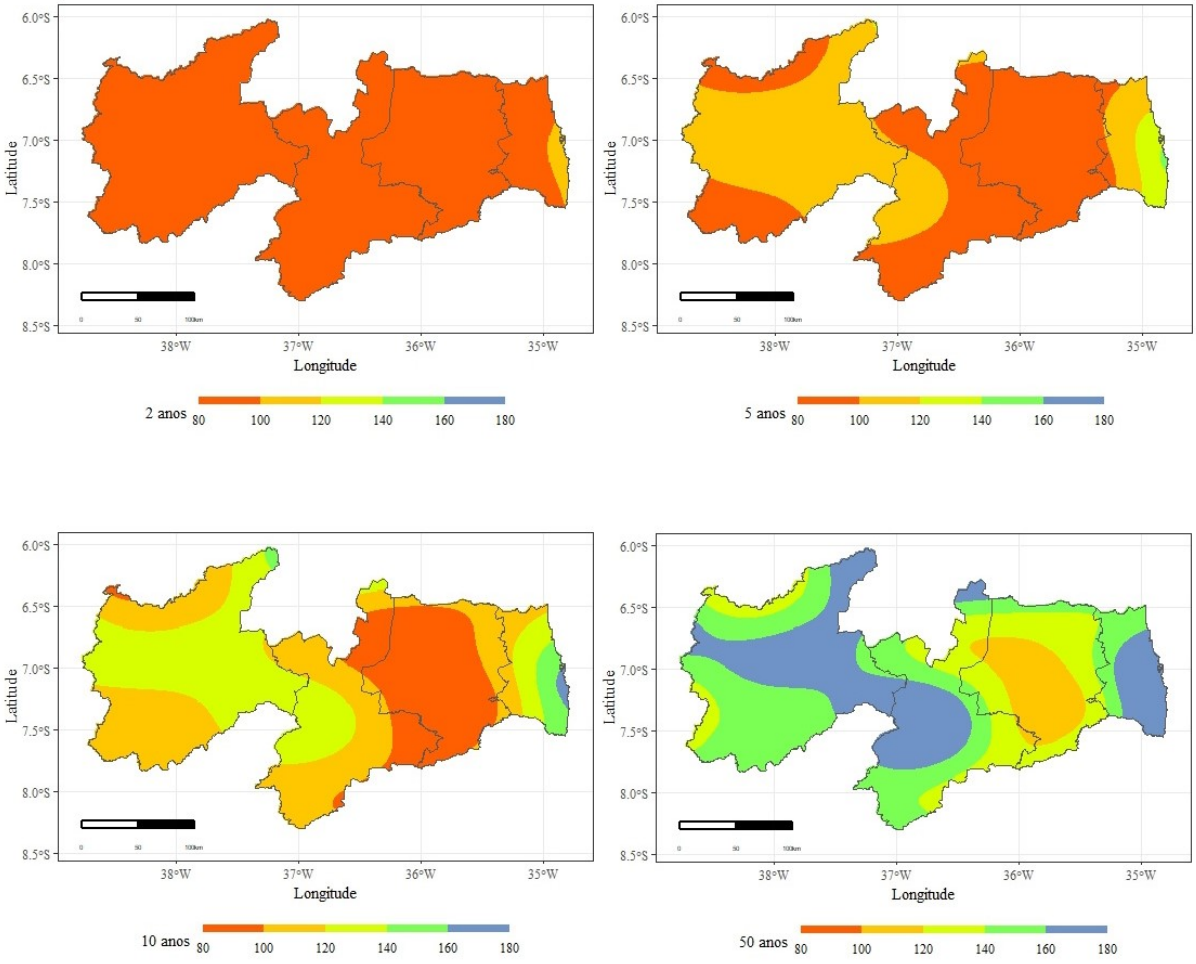
Com base nas estimativas da PMDA para os períodos de retorno de 2, 5, 10 e 50 anos, foram realizadas interpolações utilizando o algoritmo SVM. A seguir, são apresentados os

resultados da validação para o período de retorno de 10 anos. Por meio de 10 partições da validação cruzada, o melhor modelo, que apresentou a menor estimativa do erro quadrático médio, foi constituído pelas estimativas de cost, gamma e epsilon iguais a 36, 0,2 e 0,10, respectivamente.

Utilizando o modelo selecionado para os dados apresentados na Figura 5, foi obtido um valor do coeficiente de determinação igual a 65,2%. Este resultado evidencia uma predição adequada do algoritmo SVM, o qual pode agora ser aplicado para interpolação da PMDA em diferentes períodos de retorno.

A Figura 6 representa os mapas em alta resolução para os tempos de retorno de 2, 5, 10 e 50 anos obtidos utilizando o algoritmo SVM.

Figura 6 - Mapa em alta resolução para os Períodos de Retorno de 2, 5, 10 e 50 anos no Estado da Paraíba.



Fonte: Elaborado pelos autores (2023).

Ao analisar o mapa de 10 anos, é possível observar que os locais com valores entre 140 mm e 180 mm estão localizados próximos ao litoral, isso é um grande ponto para explicar esses valores extremos. A região do Agreste Paraibano é onde se encontram os menores valores, 80 a 100 mm. Esse resultado está de acordo com o fato de o clima predominante na região de todo agreste nordestino ser o clima semiárido, apresentando temperaturas elevadas todo o ano e chuvas irregulares, por vezes, escassas.

Uma das vantagens do SVM é a sua capacidade de lidar com dados de alta dimensionalidade e com um número pequeno de amostras de treinamento (HEARST *et al.*, 1998). Além disso, o SVM é menos propenso a *overfitting* do que outros algoritmos de aprendizado de máquina, como as redes neurais (SANTOS, 2002). O termo *overfitting*, em termos gerais, é um fenômeno indesejado que ocorre durante o processo de treinamento de modelos de aprendizado de máquina ou inteligência artificial. Esse fenômeno acontece quando um modelo se ajusta excessivamente aos dados de treinamento específicos em vez de generalizar padrões mais amplos e representativos presentes nos dados. O SVM tem sido amplamente utilizado em aplicações como reconhecimento de imagem, classificação de documentos, detecção de *spam*, entre outras.

A verificação da capacidade preditiva de um modelo através da computação do seu risco antecipado é uma etapa crucial no processo de seleção de um preditor eficaz para prever eventos futuros. Entretanto, é recomendável evitar a utilização da mesma base de dados na avaliação, na qual o modelo foi construído, já que geralmente isso resulta em estimativas subestimadas do risco antecipado real do modelo. Para contornar esse problema, técnicas de validação cruzada (como K-fold, leave-one-out, Hold-Out e Bootstrap) são sugeridas, uma vez que permitem a divisão da base de dados em amostras de treinamento e validação, permitindo assim que a criação do modelo e a avaliação do seu risco sejam conduzidas em bases de dados diferentes (CUNHA, 2019).

4. CONSIDERAÇÕES FINAIS

Os resultados evidenciam que as regiões situadas no litoral apresentam os maiores valores médios de chuva diária, o que se configura como um fator influenciador para esse resultado. As análises espaciais realizadas através dos testes de independência e de estacionariedade revelaram que apenas em um número reduzido de estações pluviométricas

foi rejeitado, indicando que as séries históricas apresentam estacionariedade e observações independentes.

Os resultados obtidos revelam uma distribuição heterogênea das estimativas dos níveis de retorno da precipitação máxima diária na Paraíba, especialmente ao longo de um período de 10 anos. Observou-se que os menores níveis de retorno foram registrados na mesorregião do Agreste Paraibano, variando entre 70 e 96 mm. Por outro lado, a região da Mata Paraibana apresentou estimativas de precipitação diária superiores a 174 mm. Portanto, esses resultados podem fornecer subsídios relevantes para o planejamento de políticas públicas visando à mitigação dos impactos causados por eventos de chuvas intensas.

É importante ressaltar que essa pesquisa apresenta algumas limitações que devem ser consideradas ao interpretar os resultados. O método de Máquinas de Vetores de Suporte (SVMs), embora seja uma ferramenta poderosa para modelagem e previsão, foi aplicado não em dados observados, mas sim em pseudo-dados, ou seja, estimativas dos níveis de retorno da precipitação máxima diária. Essa abordagem pode introduzir incertezas e limitações à precisão dos resultados apresentados, uma vez que as estimativas podem não capturar totalmente a complexidade e a variabilidade dos eventos climáticos reais. Portanto, é fundamental exercer cautela ao interpretar e utilizar as informações fornecidas pelo mapa de precipitação máxima diária anual apresentado para o Estado da Paraíba, especialmente em contextos de tomada de decisão e planejamento de políticas públicas relacionadas ao gerenciamento de eventos climáticos extremos.

AGRADECIMENTOS

Os autores expressam seus agradecimentos à Agência Executiva de Gestão das Águas do Estado da Paraíba pelo fornecimento dos dados utilizados nesta pesquisa e à Universidade Federal da Grande Dourados por conceder ao primeiro autor a oportunidade de participar como voluntário no âmbito do programa de Iniciação Científica, por meio do qual o projeto em questão teve início.

REFERÊNCIAS

ABREU, L. P.; MUTTI, P. R.; LIMA, K. C. Variabilidade espacial e temporal da precipitação na bacia hidrográfica do Rio Parnaíba, Nordeste do Brasil. **Revista Brasileira de Meio Ambiente**, v. 7, n. 2, p. 82-97, 2019. DOI: https://doi.org/10.5281/zenodo.3524759.

BACK, A. J. Aplicação de análise estatística para identificação de tendências climáticas. **Pesquisa Agropecuária Brasileira**, v. 36, n. 5, p. 717-726, 2001. DOI: https://doi.org/10.1590/S0100-204X2001000500001.

CARVALHO, A. L.; SANTOS, D. V.; MARENGO, J. A.; COUTINHO, S. M. V.; MAIA, S. M. F. Impacts of extreme climate events on Brazilian agricultural production. **Sustainability in Debate,** v. 11, n. 3, p. 197–224, 2020. DOI: https://doi.org/10.18472/SustDeb.v11n3.2020.33814.

CASTRO, C.; COSTEIRA, H.; GOUVEIA, J.; MARTINS, P.; SILVA, J. **Tutorial R Shiny**. Braga, Portugal: Departamento de Matemática e Aplicações da Universidade do Minho, 2023.

CEPED. Atlas Brasileiro de Desastres Naturais: 1991-2012. **Centro Universitário sobre Estudos e Pesquisa sobre Desastres**, Florianópolis, v. Paraíba, n. 2, p. 105, 2013. Disponível em: https://s2id.mi.gov.br/paginas/atlas/. Acesso em: 11 nov. 2023.

CHANG, C. C.; LIN, C. J. LIBSVM: a library for support vector machines. **ACM transactions on intelligent systems and technology**, v. 2, n. 3, p. 1-27, 2011.

CUNHA, J. P. Z. **Um estudo comparativo das técnicas de validação cruzada aplicadas a modelos mistos**. 2019. Tese de Doutorado. Universidade de São Paulo.

DAUNI, P.; FIRDAUS, M. D.; ASFARIANI, R.; SAPUTRA, M. I. N.; HIDAYAT, A. A.; ZULFIKAR, W. B. Implementation of Haversine formula for school location tracking. **Journal of Physics: Conference Series.** v. 1402, n. 7, p. 077028. DOI: https://doi.org/10.1088/1742-6596/1402/7/077028.

HEARST, M. A.; DUMAIS, S. T.; OSUNA, E.; PLATT, J.; SCHOLKOPF, B. Support vector machines. **IEEE Intelligent Systems and their applications**, v. 13, n. 4, p. 18-28, 1998.

HENGL, T.; NUSSBAUM, M.; WRIGHT, M. N.; HEUVELINK, G. B.; GRÄLER, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. **PeerJ**, v. 6, n. e5518, p. e5518, 2018. DOI: https://doi.org/10.7717/peerj.5518.

IMAI, N. N.; VICENTE, J.; LIMA, D. L.; TACHIBANA, V. M.; SILVA, E. A.; VOLL, E.; OLIVEIRA, H. E. S. Análise comparativa da interpolação por krigagem ordinária e krigagem por indicação no caso de ervas daninhas em cultura de soja. In: **XXI Congresso Brasileiro de Cartografia. Publicação em CD-Rom sem paginação.** 2003.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297–303, 1978. DOI: https://doi.org/10.2307/2335207.

MAMMONE, A.; TURCHI, M.; CRISTIANINI, N. Support vector machines. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 1, n. 3, p. 283-289, 2009.

MARCUZZO, F. F. N.; ROMERO, V. Effects of El Niño and La Niña in maximum daily precipitation of the State Goiás. **Revista Brasileira de Meteorologia**, v. 28, n. 4, p. 429-440, 2013. DOI: https://doi.org/10.1590/S0102-77862013000400009.

MARENGO, J. A.; TORRES, R. R.; ALVES, L. M. Drought in Northeast Brazil-past, present, and future. **Theoretical and Applied Climatology**, v. 129, p. 1189-1200, 2017. DOI: https://doi.org/10.1007/s00704-016-1840-8.

MEDEIROS, E. S. D.; ALVES, M. A.; SOUZA, S. A. D. Return level of estimation of maximum daily precipitation in the municipality of Jataí, Goiás. **Ciência e Natura**, v. 41, p. e36, 2019. DOI: http://doi.org/10.5902/2179460X35639.

PEREIRA, T. P.; ALBUQUERQUE, C. S.; MATTOS, V. L. D.; KONRATH, A. C.; NAKAMURA, L. R.; CARMO, C. V. D. Análise de desempenho do teste ljung-box na aplicação do modelo sarima nos dados mensais da demanda de energia elétrica da classe consumidora comercial do RS. **Encontro nacional de modelagem computacional e encontro de ciência e tecnologia de materiais.** Recife, Brasil, 2021. DOI: https://doi.org/10.29327/154013.24-22.

PINTO, W. D. P.; MONTE, E. Z.; LIMA, G. B.; & MONTEBELLER, C. A. Using logistic regression techniques to evaluate of maximum flow rates of the Rio Doce, city of Colatina, ES, Brazil. **Ciência e Natura**, v. 39, n. 3, p. 526-538, 2017.

SANTOS, E. M. D. **Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência**. Dissertação (Mestrado em Informática) – Centro de Informática, Universidade Federal da Paraíba. Campina Grande, p. 110. 2002.

SEKULIĆ, A.; KILIBARDA, M.; HEUVELINK, G. B.; NIKOLIĆ, M.; BAJAT, B. Random forest spatial interpolation. **Remote Sensing**, v. 12, n. 10, p. 1687, 2020. DOI: https://doi.org/10.3390/rs12101687.

SILVA, J. W. D.; GUIMARÃES, E. C.; TAVARES, M. Temporal variability of monthly rains in the climatological station of Uberaba-MG, Brazil. **Ciência e Agrotecnologia**, v. 27, n. 3, p. 665-674, 2003. DOI: https://doi.org/10.1590/S1413-70542003000300023.

SILVA, K. R. D.; CECÍLIO, R. A.; XAVIER, A. C.; PEZZOPANE, J. R. M.; GARCIA, G. D. O. Interpolação espacial da precipitação no Estado do Espírito Santo. **Floresta e Ambiente**, v. 18, n. 4, p. 417-427, 2012.

SOUZA, L. S.; MEDEIROS, E. S.; GALBETTI, M. V.; BICALHO, C. C.; SANTOS, C. A. C.; SHINMA, T. A.; SILVA, A. Q.; OLIVEIRA, L. A. Spatial Modeling of Intense Rainfall in the State of Paraíba. **Revista Brasileira de Meteorologia**, v. 37, n. 3, p. 373-383, 2022a. DOI: https://doi.org/10.1590/0102-77863730059.

SOUZA, L. S.; MEDEIROS, E. S.; SILVA, A. Q.; OLIVEIRA, L. A. Modeling the maximum daily rainfall in the municipality of João Pessoa-PB, Brazil, using the Extreme Value Theory. **Revista Brasileira de Climatologia**, v. 30, p. 488-503, 2022b. DOI: https://doi.org/10.55761/abclima.v30i18.14886.

SUGANYADEVI, M. V.; BABULAL, C. K. Support vector regression model for the prediction of loadability margin of a power system. **Applied Soft Computing**, v. 24, p. 304-315, 2014. DOI: https://doi.org/10.1016/j.asoc.2014.07.015.

WALD, A.; WOLFOWITZ, J. On a test whether two samples are from the same population. **The Annals of Mathematical Statistics**, v. 11, n. 2, p. 147-162, 1940. DOI: https://doi.org/10.1214/aoms/1177731909.