

XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

IDENTIFICAÇÃO DE GRUPOS HOMOGÊNEOS DE ESTAÇÕES PLUVIOGRÁFICAS NO BAIXO SÃO FRANCISCO UTILIZANDO ANÁLISE DE CLUSTER

Anderson de Oliveira Pinheiro¹; Demetrius David da Silva² & Michel Castro Moreira²

Abstract: The regionalization of rainfall stations plays an important role in understanding the spatial patterns of precipitation and in the efficient planning of water resources. This study aimed to identify rainfall stations with similar precipitation patterns in the Lower São Francisco region by segmenting them into groups with comparable rainfall behavior. To this end, cluster analysis techniques were applied, using both K-means and hierarchical methods based on Euclidean distance and monthly precipitation data recorded over a 30-year period. The optimal number of clusters was determined using the Elbow Method and the Silhouette Index, which provided greater reliability to the resulting classification. The results revealed the formation of two distinct clusters: one composed of stations located near the coast, characterized by higher annual precipitation totals and greater variability, influenced by oceanic atmospheric systems; and another formed by inland stations, marked by lower rainfall volumes and a more regular seasonal pattern, typical of semi-arid regions. The adopted methodology proved effective in representing the region's pluviometric heterogeneity, allowing the identification of two well-defined climatic groups: coastal and inland.

Resumo: A regionalização de estações pluviométricas tem papel importante para a compreensão dos padrões espaciais da precipitação e para o planejamento eficiente dos recursos hídricos. Este estudo teve como objetivo identificar as estações pluviométricas com padrões de chuvas semelhantes na região do Baixo São Francisco, por meio da segmentação das estações em grupos com comportamentos pluviométricos semelhantes. Para isso, foram aplicadas técnicas de análise de agrupamento (*cluster analysis*), utilizando os métodos K-means e hierárquico, com base na distância euclidiana e nos valores mensais de precipitação registrados ao longo de 30 anos. A definição do número ideal de grupos se deu por meio dos métodos do Cotovelo (*elbow method*) e da Silhueta (*silhouette index*), o que garantiu maior confiabilidade à classificação obtida. Os resultados evidenciaram a formação de dois clusters distintos: um composto por estações localizadas próximas ao litoral, caracterizado por maiores totais anuais de precipitação e variabilidade acentuada, influenciado por sistemas atmosféricos oceânicos; e outro formado por estações situadas no interior, com menores volumes de chuva e padrão sazonal mais regular, típico de regiões semiáridas. A metodologia adotada demonstrou-se eficaz na representação da heterogeneidade pluviométrica da região, permitindo a identificação de dois grupos climáticos bem definidos: o litoral e o interior.

Palavras-Chave – Semelhança entre estações; Regiões Pluviométricas; Agrupamento de estações.

1) Mestrando em Engenharia Agrícola, Universidade Federal de Viçosa, anderson.pinheiro@ufv.br.

2) Professor do Departamento de Engenharia Agrícola, Universidade Federal de Viçosa, demetrius@ufv.br; michelcm@ufv.br.

1. INTRODUÇÃO

O estudo da precipitação desempenha um papel central no planejamento e gerenciamento dos recursos hídricos, especialmente em regiões com alta variabilidade climática. A compreensão do comportamento espaço-temporal das chuvas permite não apenas o monitoramento de eventos extremos, como secas e enchentes, mas também contribui significativamente para a formulação de políticas públicas, o zoneamento agrícola e o uso racional da água (Montano e Souza, 2016; Nascimento *et al.*, 2015).

A análise espacial da precipitação é uma ferramenta essencial na caracterização hidrológica de bacias, fornecendo subsídios importantes para a gestão territorial e ambiental. As chuvas, por serem responsáveis para o abastecimento hídrico das bacias hidrográficas, exercem influência direta sobre o clima (Silva *et al.*, 2023). A variabilidade regional da precipitação implica na necessidade de identificar zonas pluviometricamente homogêneas, ou seja, áreas com características similares de precipitação, o que auxilia na distinção de regiões com maior aporte ou déficit hídrico (Menezes *et al.*, 2015).

Dessa forma, a delimitação de regiões homogêneas de precipitação é essencial para o entendimento das dinâmicas hidrológicas, principalmente em regiões que possuem comportamentos hidrológicos distintos. O Baixo São Francisco é uma região localizada no nordeste do Brasil, correspondente a foz do rio São Francisco e que por isso tem contado direto com o mar, sendo uma região que contrasta o clima da região do semiárido com a região litorânea, abrangendo áreas com distintos regimes pluviométricos (Medeiros *et al.*, 2014).

Para identificar padrões espaciais de precipitação e agrupar áreas com características semelhantes, diversos métodos estatísticos e de aprendizado de máquina têm sido empregados. Entre eles, destaca-se o algoritmo de agrupamento k-means, muito utilizado por sua eficiência e simplicidade na identificação de regiões com comportamento pluviométrico similar (Medeiros *et al.*, 2014). Esse tipo de abordagem tem se mostrado promissora em diversos estudos voltados à regionalização da precipitação e ao planejamento hidrológico, como nos trabalhos de (Alam e Paul, 2020; Firat *et al.*, 2012; Nkoni *et al.*, 2024).

Diante disso, o presente trabalho teve como objetivo identificar estações com comportamentos pluviométricos semelhantes na região do Baixo São Francisco, por meio do método de agrupamento K-means, visando delinear regiões com características pluviométricas homogêneas.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo e conjunto de dados

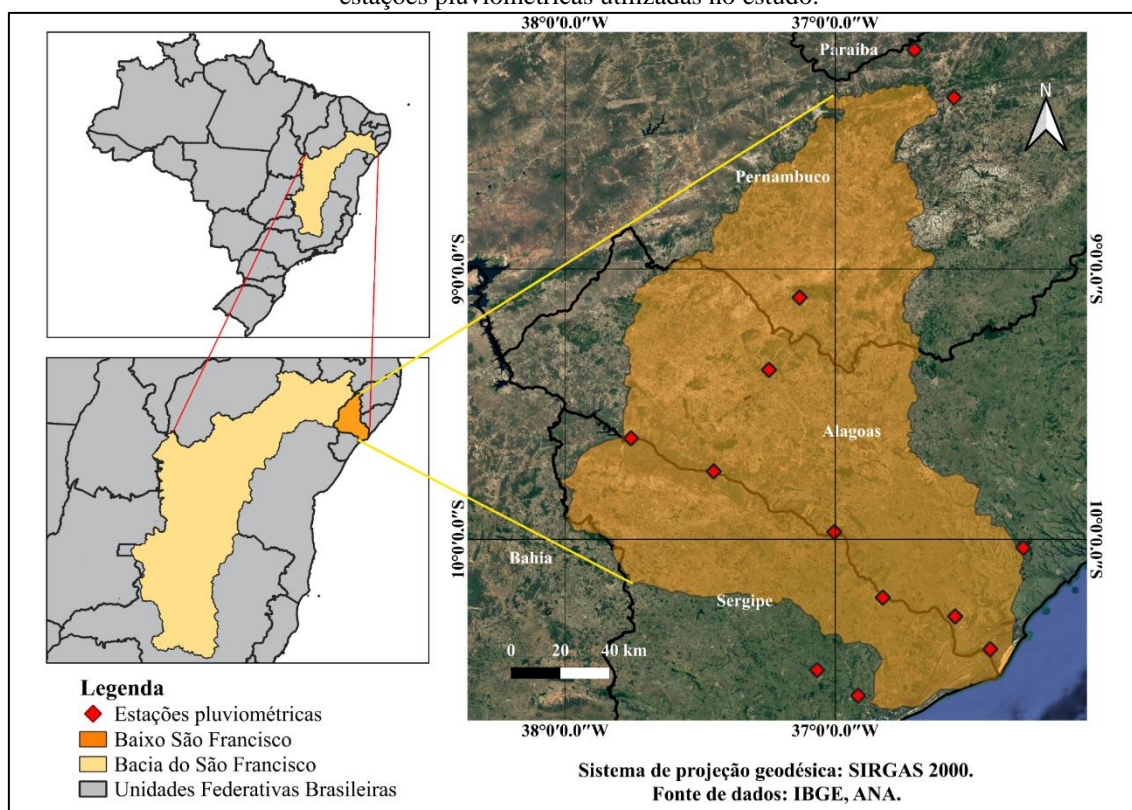
A área de estudo abrange a região do Baixo São Francisco, conforme ilustrado na Figura 1, incluindo as estações pluviométricas localizadas até aproximadamente 16 km além dos limites geográficos tradicionalmente atribuídos a essa região. Essa ampliação territorial se deu pelo comportamento natural das precipitações, que não respeitam divisões administrativas e são controladas por fatores fisiográficos, como relevo e altitude, e climáticos, como a atuação de sistemas atmosféricos regionais (Jangarelli *et al.*, 2022). Assim, a ampliação da área de estudo teve como objetivo delimitar a região adjacente que compartilha padrões semelhantes de variabilidade na precipitação regional.

Foram selecionadas estações pluviométricas cujos dados estão disponíveis na base da Agência Nacional de Águas e Saneamento Básico (ANA), por meio da plataforma HidroWeb (<http://www.snirh.gov.br/hidroweb>). A série histórica analisada compreende o período de 1995 a

2024, com registros mensais. Foram selecionadas as estações pluviométricas que apresentassem no máximo 5% de falhas nos dados de precipitação para o período considerado, sendo identificadas 13 estações pluviométricas (Figura 1).

As falhas identificadas nas séries temporais foram preenchidas por meio do método da ponderação regional, com base nas três estações geograficamente mais próximas com dados disponíveis no mesmo período. Esse método consiste em calcular os valores ausentes com base em pesos proporcionais às normais climatológicas de cada estação de referência, garantindo consistência com as condições climáticas e geográficas locais. Estudos demonstram que a ponderação regional é eficaz na estimativa de dados ausentes em séries pluviométricas, apresentando resultados satisfatórios em diferentes regiões do Brasil (Diaz *et al.*, 2018).

Figura 1: Localização da região do Baixo São Francisco e das estações pluviométricas utilizadas no estudo.



Fonte: Autores.

2.2. Análise de agrupamento

Para identificar as estações com mesmas características no regime pluviométrico, aplicou-se a análise de agrupamento (*cluster analysis*), utilizando os valores mensais de precipitação de cada estação pluviométrica. A análise de *cluster* é uma técnica estatística multivariada que tem como objetivo principal identificar estruturas de similaridade dentro de um conjunto de dados, agrupando observações em subconjuntos homogêneos denominados *clusters*. O processo baseia-se em medidas de distância ou dissimilaridade, sendo considerado nesse trabalho a distância euclidiana, entre vetores de características, organizando as unidades amostrais de forma que aquelas pertencentes a um mesmo grupo sejam mais semelhantes entre si do que em relação às pertencentes a outros grupos (Singh *et al.*, 2017).

Em procedimentos desse tipo, é comum reescalar os dados para colocá-los em uma mesma unidade, especialmente quando as variáveis apresentam diferentes ordens de magnitude (Amorim *et al.*, 2023). No entanto, como a análise neste estudo foi realizada exclusivamente com base em valores mensais de precipitação (mm), que já compartilham a mesma unidade e escala, não foi necessário aplicar a padronização dos dados.

Com o objetivo de identificar o número ideal de *clusters* para a separação das estações, foram empregados dois métodos amplamente utilizados: o Método do Cotovelo (*Elbow Method*) e o Índice de Silhueta (*Silhouette Index*). O Índice de Silhueta, proposto por (Rousseeuw, 1987), avalia a qualidade do agrupamento ao medir, para cada ponto, o quão semelhante ele é aos membros de seu próprio *cluster* em comparação aos de outros *clusters*. O valor do índice varia de -1 a 1, sendo que os valores próximos de 1 indicam que o ponto está bem alocado dentro do seu grupo e suficientemente distante de outros grupos. Dessa forma, o número ideal de *clusters* é aquele que maximiza o valor médio do Índice de Silhueta para todos os pontos, refletindo uma melhor separação entre os grupos (Rousseeuw, 1987).

Já o Método do Cotovelo é baseado na análise da soma total das distâncias quadráticas dentro dos *clusters*, também conhecida como *within-cluster sum of squares* (WCSS). À medida que o número de *clusters* aumenta, essa soma tende a diminuir, pois os dados passam a ser representados por grupos mais específicos. No entanto, há um ponto em que esse ganho se torna marginal, a chamada “curva do cotovelo”. Este ponto de inflexão representa o número ótimo de *clusters*, pois além dele os ganhos de variância explicada são mínimos (Thorndike, 1953).

Logo após determinado o número ótimo de *clusters*, aplicou-se o algoritmo K-means, que tem como objetivo minimizar a soma das distâncias euclidianas quadráticas entre cada ponto e o centróide de seu respectivo grupo. Esse algoritmo é utilizado em estudos de regionalização climática e hidrológica devido à sua eficiência computacional e capacidade de identificar padrões espaciais coerentes (Macqueen, 1967; Pike e Lintner, 2020). Foi empregada a análise de agrupamento hierárquico utilizando a distância euclidiana, com o objetivo de identificar padrões de similaridade entre as estações hidrometeorológicas. A estrutura dos agrupamentos foi representada por meio de um dendrograma, permitindo visualizar a formação de grupos homogêneos e o grau de proximidade entre as estações (Sun *et al.*, 2015).

Ao final, foi analisado o comportamento dos grupos formados com base na série anual, considerando a distribuição, a dispersão e a amplitude dos valores ao longo do tempo, com o objetivo de compreender as particularidades e distinções entre os *clusters* identificados. (Mcgill *et al.*, 1978; Reimann *et al.*, 2005).

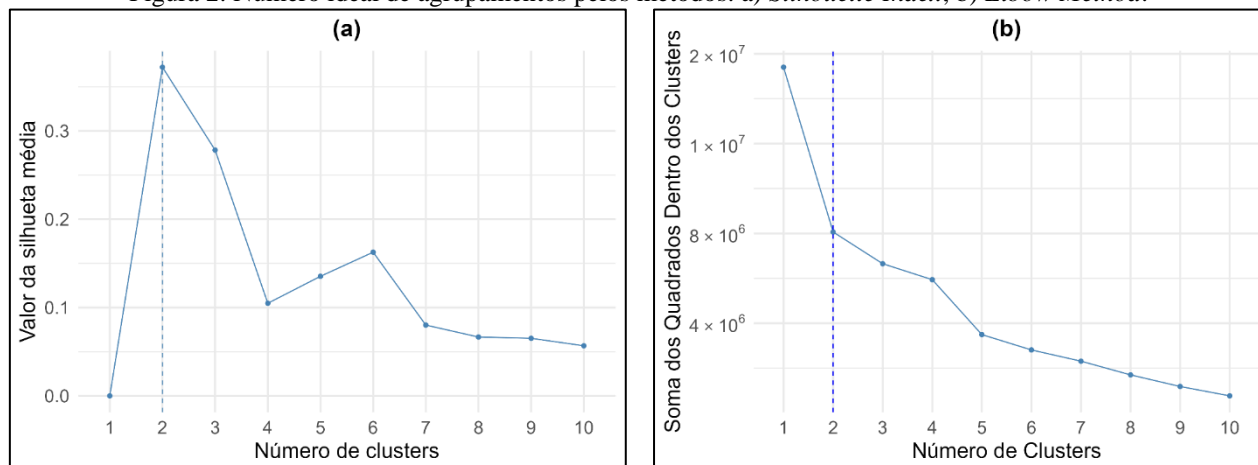
Todas as etapas do trabalho, desde a aquisição dos dados, o preenchimento de falhas nas séries históricas, a identificação da quantidade ideal de *clusters*, a aplicação do método K-means até a análise dos resultados, foram desenvolvidas utilizando a linguagem R por meio do ambiente RStudio.

3. RESULTADOS E DISCUSSÕES

Conforme ilustrado na Figura 2, que apresenta os resultados dos métodos da silhueta e do cotovelo aplicados para a definição dos agrupamentos, ambos os métodos convergiram para a mesma solução. Ao avaliar a qualidade do particionamento por meio da coesão interna e da separação entre os grupos, o método da silhueta indicou a formação de dois agrupamentos como a mais adequada. Da mesma forma, ao examinar a variação da soma das distâncias quadráticas intra-cluster (WCSS), o método do cotovelo identificou o ponto de inflexão da curva também na divisão em dois grupos. Essa

convergência reforça a consistência da escolha, evidenciando que a divisão em dois clusters é a mais apropriada para representar a heterogeneidade espacial da precipitação na região.

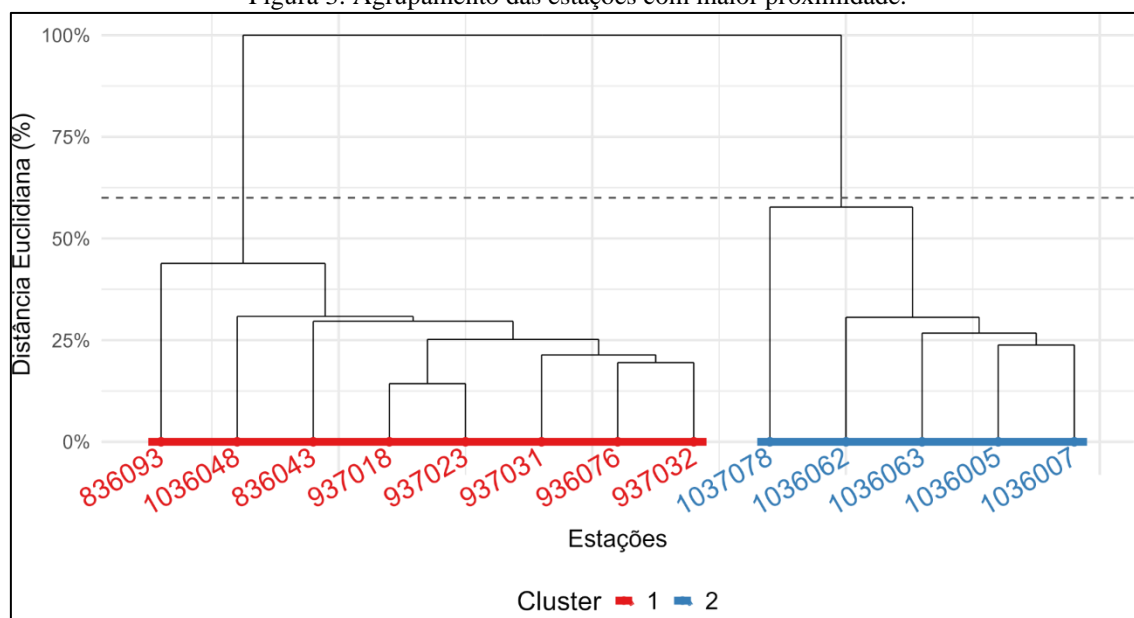
Figura 2: Número ideal de agrupamentos pelos métodos: a) *Silhouette Index*; b) *Elbow Method*.



Fonte: Autores.

Na Figura 3 apresenta-se o dendrograma com os agrupamentos por estação pluviométrica utilizada no estudo, na qual observa-se uma clara divisão das estações em dois grupos distintos, com uma separação evidente ocorrendo a aproximadamente 60% da distância Euclidiana, o que reforça a consistência dos agrupamentos definidos anteriormente em dois grupos.

Figura 3: Agrupamento das estações com maior proximidade.



Fonte: Autores.

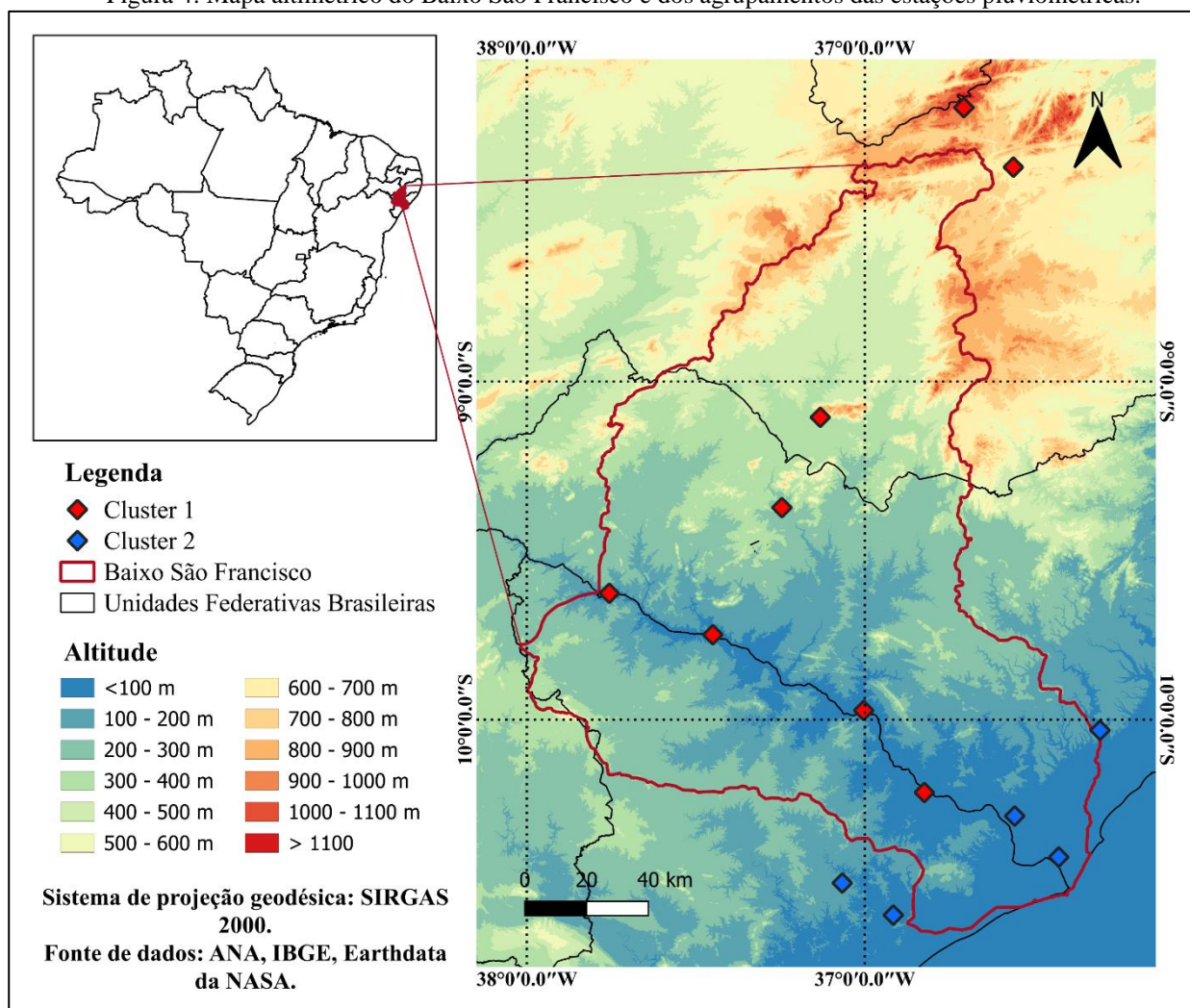
A Figura 4 apresenta a distribuição espacial das estações agrupadas com base na similaridade dos padrões de precipitação. Observa-se que as estações localizadas mais ao interior, afastadas da zona costeira, formam um grupo relativamente homogêneo, cobrindo a maior parte da área de estudo. Por outro lado, as estações situadas próximas ao litoral constituem um segundo grupo, espacialmente mais restrito, mas também caracterizado por uma homogeneidade interna. Essa diferenciação pode estar associada à influência de sistemas atmosféricos oriundos do Oceano Atlântico, como as Zonas

de Convergência Intertropical e as brisas marítimas, que impactam com maior intensidade as regiões litorâneas (Ferreira e Mello, 2005; Molion e Bernardo, 2002).

Além disso, fatores fisiográficos, como relevo e altitude, contribuem para a formação de regimes pluviométricos distintos entre áreas costeiras e interiores, uma vez que influenciam diretamente a dinâmica atmosférica local, como a ocorrência de chuvas orográficas e a modulação de ventos úmidos vindos do oceano. (Luiz-Silva e Oscar-Junior, 2022). De maneira semelhante, estudos realizados na ilha de Maurício demonstraram que variáveis como elevação, orientação do relevo e proximidade da costa exercem influência direta na distribuição espacial da precipitação, evidenciando a importância desses elementos na heterogeneidade climática regional, (RAJA *et al.*, 2019). Essa diferenciação torna-se evidente na Figura 4, ao se observar uma transição altimétrica entre 100 e 200 metros, faixa na qual ocorre uma mudança significativa na composição dos agrupamentos (*clusters*), indicando que a altitude atua como um fator limitante para a separação de regiões com distintos regimes pluviométricos.

Santos e Souza (2012), ao aplicarem métodos multivariados, observaram resultados semelhantes, nos quais as regiões litorâneas apresentaram um padrão de agrupamento distinto em relação às demais áreas, formando *clusters* bem definidos e homogêneos, o que evidencia a eficácia da técnica na identificação de zonas climáticas com características próprias.

Figura 4: Mapa altimétrico do Baixo São Francisco e dos agrupamentos das estações pluviométricas.



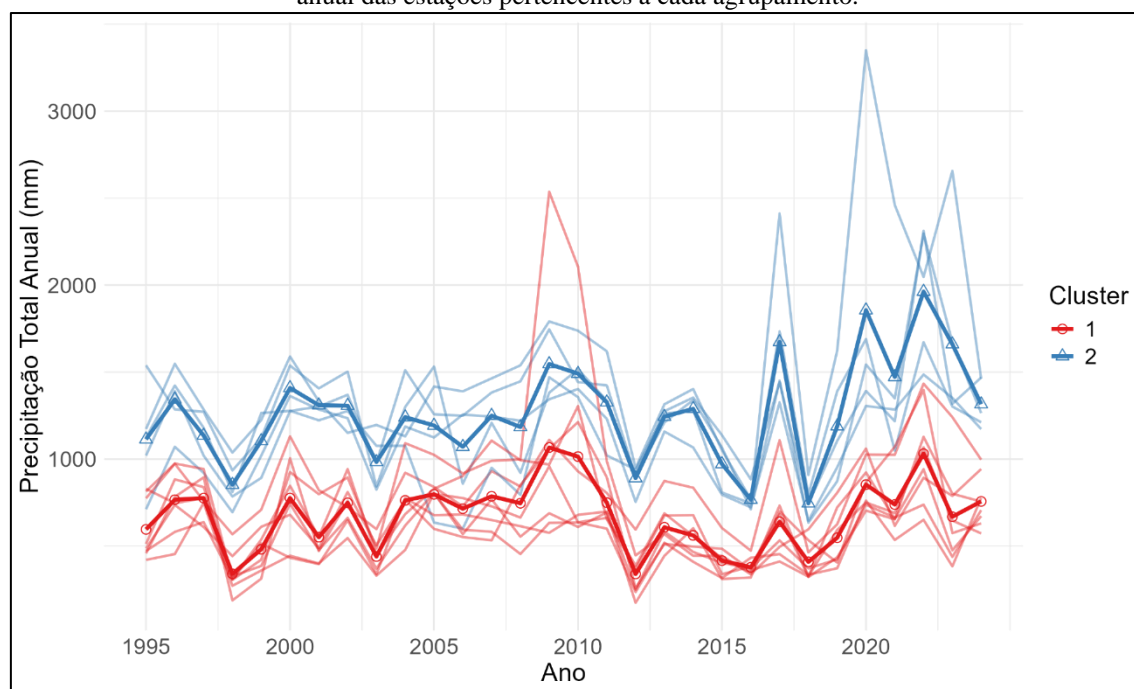
Fonte: Autores.

A Figura 5 apresenta a média do total anual de precipitação para as estações agrupadas em cada *cluster*, permitindo observar padrões regionais distintos na distribuição da chuva. As estações pertencentes ao *Cluster 2*, localizadas mais próximas ao litoral, registram os maiores volumes pluviométricos anuais, corroborando com o trabalho de Bernador et al. (2013), que identificaram o mesmo comportamento pluviométrico no estado da Bahia. Esse comportamento pode ser explicado pela maior influência de sistemas meteorológicos oceânicos, como as brisas marítimas e os Distúrbios Ondulatórios de Leste, que intensificam a precipitação nas áreas costeiras do Nordeste, (Gomes et al., 2015). A proximidade com o oceano favorece a umidade disponível e os mecanismos convectivos associados ao relevo local, resultando em maior volume de chuvas nessas regiões, (Bergeman e Jakob, 2016).

Por outro lado, as estações classificadas no *Cluster 1*, situadas no interior da bacia do Baixo São Francisco, apresentam menores totais pluviométricos anuais. Essa redução é reflexo da atuação do regime semiárido típico do Sertão nordestino, onde predominam elevadas temperaturas, baixa umidade relativa do ar e precipitação escassa e irregular (Oliveira et al., 2022). A distância do litoral reduz a influência direta de sistemas oceânicos, limitando os volumes de precipitação nessas áreas continentais (Ogino et al., 2016).

Apesar dessa diferença nos valores absolutos, observa-se que o padrão de variabilidade temporal entre os clusters apresenta certa semelhança. Isso indica que, embora os volumes totais variem significativamente entre litoral e interior, a sazonalidade e a resposta às grandes forçantes atmosféricas, como El Niño e La Niña, tendem a afetar ambos os grupos de forma sincronizada, (Oliveira et al., 2017). Dessa forma, a análise dos totais anuais evidencia tanto a influência da localização geográfica sobre o volume da precipitação quanto a presença de um regime climático comum que afeta toda a região do Baixo São Francisco.

Figura 5: Evolução da precipitação total anual por estação, agrupadas por *cluster*. As linhas mais transparentes representam os totais anuais individuais de cada estação, enquanto as linhas mais destacadas correspondem à média anual das estações pertencentes a cada agrupamento.

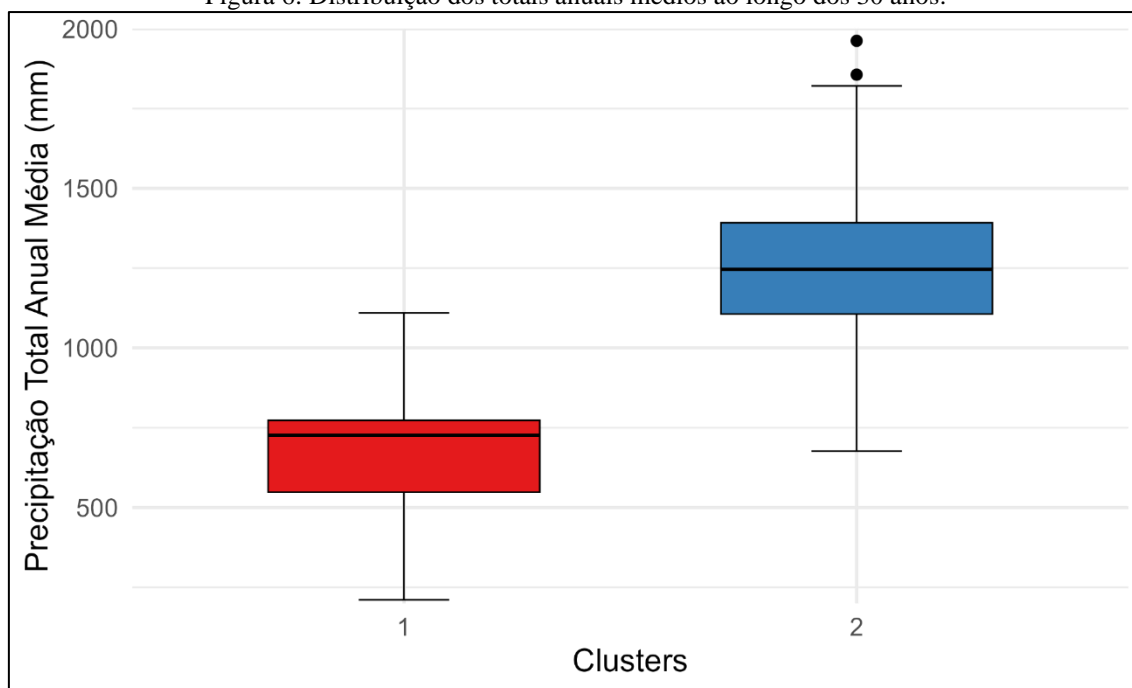


Fonte: Autores.

A Figura 6 apresenta o *boxplot* da precipitação total anual para os dois *clusters* identificados, evidenciando de forma clara a distinção entre os grupos. Conforme também observado na Figura 5, as estações localizadas próximas ao litoral, agrupadas no *Cluster 2*, apresentam medianas mais elevadas, amplitudes interquartis maiores e uma maior irregularidade na variabilidade anual da precipitação, atribuída à alternância entre anos mais secos e mais chuvosos. Essa variabilidade pode ser influenciada por fatores como a atuação de sistemas atmosféricos e a topografia local, (Ferreira e Mello, 2005; RAJA *et al.*, 2019). Por outro lado, as estações do *Cluster 1*, situadas no interior da região, apresentam menores valores medianos e menor variabilidade, refletindo uma maior regularidade nas chuvas, onde a distribuição anual da precipitação é menos suscetível às variações climáticas de grande escala.

A diferença nos padrões de variabilidade entre os *clusters* evidencia o contraste climático marcante entre o litoral úmido e o sertão semiárido, fenômeno reconhecido na literatura climatológica. Esse padrão também foi observado por García-Garizábal (2017) que, ao analisar a variabilidade pluviométrica na região costeira árida do Equador, identificou agrupamentos distintos de precipitação ao longo da costa, com características significativamente diferentes em relação às demais regiões.

Figura 6: Distribuição dos totais anuais médios ao longo dos 30 anos.



Fonte: Autores.

4. CONCLUSÕES

A partir da análise dos resultados pode-se concluir que:

- As estações situadas próximas ao litoral apresentaram maiores volumes pluviométricos, evidenciando a influência direta de sistemas meteorológicos oceânicos;
- As estações localizadas no interior da região, em área de clima semiárido, apresentaram menores totais anuais e menor variabilidade, condizente com a natureza irregular das chuvas nesse domínio climático; e

- Foram identificados pela análise de *cluster* dois agrupamentos de estações pluviométricas para a área de estudo.

AGRADECIMENTOS – Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal Nível Superior – CAPES, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais – FAPEMIG – pelo apoio no desenvolvimento do trabalho.

REFERÊNCIAS

- ALAM, Mohammad Samsul; PAUL, Sangita. (2020). *A comparative analysis of clustering algorithms to identify the homogeneous rainfall gauge stations of Bangladesh*. Journal of Applied Statistics, v. 47, n. 8, p. 1460-1481.
- BERGEMANN, Martin; JAKOB, Christian. (2016). *How important is tropospheric humidity for coastal rainfall in the tropics?*. Geophysical Research Letters, v. 43, n. 11, p. 5860-5868.
- DE OLIVEIRA RODRIGUES, Luana et al. (2017). *Influência dos eventos de El Niño e La Niña no regime de precipitação do Agreste de Pernambuco*. Revista Brasileira de Geografia Física, v. 10, n. 6, p. 1995-2009.
- DE AMORIM, Lucas BV; CAVALCANTI, George DC; CRUZ, Rafael MO. (2023). *The choice of scaling technique matters for classification performance*. Applied Soft Computing, v. 133, p. 109924.
- DIAZ, Caio Cesar Farias; DOS SANTOS PEREIRA, João Antônio; NÓBREGA, Ranyére Silva. (2018). *Comparação de dados estimados pelo método da ponderação regional (pr) e dados estimados pelo trmm para o preenchimento de falhas de precipitação na bacia hidrográfica do rio Pajeú*. Revista Brasileira de Climatologia, v. 22.
- DOURADO, Camila da Silva; OLIVEIRA, Stanley Robson de Medeiros; AVILA, Ana Maria Heuminski de. (2013). *Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia*. Bragantia, v. 72, p. 192-198.
- FERREIRA, Antonio Geraldo; MELLO, Namir Giovanni da Silva. (2005). *Principais sistemas atmosféricos atuantes sobre A Região Nordeste do Brasil e a influência dos Oceanos Pacífico e Atlântico no clima da região*.
- FIRAT, Mahmut et al. (2012). *Classification of annual precipitations and identification of homogeneous regions using k-means Method*. Teknik Dergi, v. 23, n. 115, p. 1609-1622.
- GARCÍA-GARIZÁBAL, Iker. (2015). *Rainfall variability and trend analysis in coastal arid Ecuador*. International Journal of Climatology, v. 37, n. 13, p. 4620-4630.
- GOMES, Helber Barros et al. (2015). *Easterly wave disturbances over Northeast Brazil: an observational analysis*. Advances in Meteorology, v. 2015, n. 1, p. 176238.
- JANGARELLI, George Antônio; JÚNIOR, Jório Bezerra Cabral; DO NASCIMENTO, Melchior Carlos. (2022). *Variabilidade espacial e temporal da precipitação pluvial na região fisiográfica do Baixo São Francisco*. Revista de Geografia-PPGEO-UFJF, v. 12, n. 2, p. 260-280.
- LUIZ-SILVA, Wanderson; OSCAR-JÚNIOR, Antonio Carlos. (2022). *Climate extremes related with rainfall in the State of Rio de Janeiro, Brazil: a review of climatological characteristics and recorded trends*. Natural Hazards, v. 114, n. 1, p. 713-732.

- MCGILL, Robert; TUKEY, John W.; LARSEN, Wayne A. (1978). *Variations of box plots*. The american statistician, v. 32, n. 1, p. 12-16.
- MEDEIROS, Paulo Petter et al. (2014). *Características ambientais do Baixo São Francisco (AL/SE): efeitos de barragens no transporte de materiais na interface continente-oceano*. Geochimica Brasiliensis, v. 28, n. 1, p. 65-65.
- MENEZES, Franciane Pantoja; FERNANDES, Lindemberg Lima; ROCHA, Edson José Paulino. (2015). *O uso da estatística para regionalização da precipitação no Estado do Pará, Brasil*. Revista Brasileira de Climatologia, v. 16.
- MOLION, L.C.B.; BERNARDO, S.O. (2002). *Uma revisão da dinâmica das chuvas no nordeste brasileiro*. Revista Brasileira de Meteorologia, v.17, p.1-10.
- MONTAÑO, Marcelo; SOUZA, Marcelo Pereira de. (2016). *Integração entre planejamento do uso do solo e de recursos hídricos: a disponibilidade hídrica como critério para a localização de empreendimentos*. Engenharia Sanitaria e Ambiental, v. 21, n. 03, p. 489-495.
- NASCIMENTO, D. A. et al. (2015). *A análise pluviométrica na gestão dos recursos hídricos: Exame de dados de 2015 de uma estação pluviométrica da Amazônia Central*. POISSON, E. Recursos Hídricos em Foco, v. 1, p. 1.
- NKONI, Godiraone et al. (2024). *Use of multivariate techniques to regionalize rainfall patterns in semiarid Botswana*. Discover Environment, v. 2, n. 1, p. 77.
- OLIVEIRA, Willian dos Santos et al. (2022). *Spatial mapping of annual rainfall in the São Francisco River Basin*. Revista Ambiente & Água, v. 17, n. 3, p. e2762.
- OGINO, Shin-Ya et al. (2016). *How much is the precipitation amount over the tropical coastal region?*. Journal of Climate, v. 29, n. 3, p. 1231-1236.
- PIKE, Maxwell; LINTNER, Benjamin R. (2020). *Application of clustering algorithms to TRMM precipitation over the tropical and South Pacific Ocean*. Journal of Climate, v. 33, n. 13, p. 5767-5785.
- RAJA, Nussaibah Begum; AYDIN, Olgu. (2019). *Regionalization of precipitation in Mauritius: A statistical approach*. Meteorological Applications, v. 26, n. 4, p. 711-719.
- REIMANN, Clemens; FILZMOSER, Peter; GARRETT, Robert G. (2005). *Background and threshold: critical comparison of methods of determination*. Science of the total environment, v. 346, n. 1-3, p. 1-16.
- ROUSSEEUW, Peter J. (1987). *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. Journal of computational and applied mathematics, v. 20, p. 53-65.
- SANTOS, Winícius Araújo; DE SOUZA, Francisco de A. Saviano. (2012). *Identificação de regiões pluviometricamente homogêneas no nordeste do brasil usando análise multivariada*. Revista Brasileira de Climatologia, v. 10.
- SILVA, Gabriela Larissa; CARVALHO, Daniela Junqueira; KOIDE, Sergio. (2023). *Distribuição espacial de chuva e o efeito sobre a simulação hidrológica—estudo de caso na bacia do riacho fundo—DF*. Revista Eletrônica de Gestão e Tecnologias Ambientais.
- SINGH, Saurabh K.; YAT-MAN LO, Edmond; QIN, Xiaosheng. (2017). *Cluster Analysis of monthly precipitation over the western maritime continent under climate change*. Climate, v. 5, n. 4, p. 84.
- SUN, Xun et al. (2015). *Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany*. Water Resources Research, v. 51, n. 8, p. 6586-6601.
- THORNDIKE, Robert L. (1953). *Who belongs in the family?*. Psychometrika, v. 18, n. 4, p. 267-276.