

## INFLUÊNCIA DA SELEÇÃO DE ESTAÇÕES PLUVIOMÉTRICAS DE APOIO NO PREENCHIMENTO DE FALHAS EM SÉRIES HISTÓRICAS DE PRECIPITAÇÃO

Anderson de Oliveira **Pinheiro**<sup>1</sup>, Demetrius David da **Silva**<sup>2</sup>, Michel Castro **Moreira**<sup>3</sup>

(1 – Universidade Federal de Viçosa (UFV), Programa de Pós-Graduação em Engenharia Agrícola, [anderson.pinheiro@ufv.br](mailto:anderson.pinheiro@ufv.br), <https://orcid.org/0009-0005-6672-6618>; 2 – Universidade Federal de Viçosa (UFV), Departamento de Engenharia Agrícola, [demetrius@ufv.br](mailto:demetrius@ufv.br), <https://orcid.org/0000-0001-9666-7421>; 3 – Universidade Federal de Viçosa (UFV), Departamento de Engenharia Agrícola, [michelcm@ufv.br](mailto:michelcm@ufv.br), <https://orcid.org/0000-0001-8024-7705>)

**Resumo:** Séries históricas de precipitação são fundamentais para a condução de estudos hidrológicos, no entanto, sua qualidade pode ser comprometida pela presença de falhas nos registros. Nesses casos, o uso de métodos de preenchimento de falhas é essencial para garantir a consistência e a continuidade dos dados. Embora diferentes métodos sejam utilizados para o preenchimento de falhas em séries pluviométricas, ainda são limitados os estudos que avaliam de forma sistemática a influência da seleção das estações de apoio na precisão das estimativas. Sendo assim, este estudo teve como objetivo avaliar a influência da seleção das estações pluviométricas de apoio no preenchimento de falhas em séries históricas de precipitação. Foram selecionadas nove estações pluviométricas na região Norte de Minas Gerais, sendo oito utilizadas como apoio para o preenchimento de falhas da estação central. Foram simulados percentuais de falhas de 10% e 40%, e os valores faltantes foram preenchidos com base no método de Regressão Linear, sendo os preenchimentos avaliados pelos índices RMSE, SMAPE e NSE. Os resultados indicaram bom desempenho no preenchimento de falhas utilizando o método de Regressão Linear e a necessidade de se realizar uma análise espacial detalhada da região, levando em consideração a topografia, a fim de selecionar estações próximas e livres de barreiras orográficas, garantindo maior confiabilidade no preenchimento das séries pluviométricas.

**Palavras-chave:** Análise hidrológica, dados ausentes, série temporal.

## INFLUENCE OF THE SELECTION OF SUPPORTING RAINFALL STATIONS ON THE GAP-FILLING OF HISTORICAL PRECIPITATION SERIES

**Abstract:** Historical precipitation series are essential for conducting hydrological studies; however, their quality may be compromised by missing data in the records. In such cases, the use of gap-filling methods is crucial to ensure data consistency and continuity. Although different methods are used to fill gaps in rainfall time series, studies that systematically evaluate the influence of the selection of supporting stations on the accuracy of the estimates are still limited. Therefore, this study aimed to assess the influence of the selection of supporting rain gauge stations on gap filling in historical precipitation series. Nine rainfall stations were selected in the northern region of Minas Gerais, Brazil, eight of which were used as supporting stations for filling gaps in the central station. Missing data rates of 10% and 40% were simulated, and the missing values were filled using the Linear Regression method. The performance of the fillings was evaluated through the RMSE, SMAPE, and NSE indices. The results indicated good performance of the Linear Regression method for gap filling and highlighted the importance of conducting a detailed spatial analysis of the region, considering topographic features, to select nearby stations free from orographic barriers, ensuring greater reliability in the completion of rainfall series.

**Keywords:** Hydrological analysis, missing data, time series.

## INFLUENCIA DE LA SELECCIÓN DE ESTACIONES PLUVIOMÉTRICAS DE APOYO EN EL RELLENO DE VACÍOS EN SERIES HISTÓRICAS DE PRECIPITACIÓN

**RESUMEN:** Las series históricas de precipitación son fundamentales para la realización de estudios hidrológicos; sin embargo, su calidad puede verse comprometida por la presencia de vacíos en los registros. En estos casos, el uso de métodos de relleno de datos faltantes es esencial para garantizar la consistencia y continuidad de la información. Aunque se utilizan diferentes métodos para el relleno de faltantes en series pluviométricas, aún son limitados los estudios que evalúan de manera sistemática la influencia de la selección de las estaciones de apoyo en la precisión de las estimaciones. Por lo tanto, este estudio tuvo como objetivo evaluar la influencia de la selección de las estaciones pluviométricas de apoyo en el relleno de faltantes en series históricas de precipitación. Se seleccionaron nueve estaciones pluviométricas en la región norte

del estado de Minas Gerais (Brasil), de las cuales ocho se utilizaron como apoyo para completar los vacíos de la estación central. Se simularon porcentajes de vacíos del 10% y 40%, y los valores faltantes se estimaron mediante el método de Regresión Lineal. Los rellenos se evaluaron utilizando los índices RMSE, SMAPE y NSE. Los resultados indicaron un buen desempeño del método de Regresión Lineal y resaltaron la necesidad de realizar un análisis espacial detallado de la región, considerando la topografía, con el fin de seleccionar estaciones cercanas y libres de barreras orográficas, garantizando así una mayor confiabilidad en el relleno de las series pluviométricas.

**Palabras clave:** Análisis hidrológico, datos faltantes, serie temporal.

## 1. Introdução

A ocorrência de falhas em séries históricas de precipitação representa um desafio recorrente na análise hidrológica, uma vez que a escassez de dados compromete a avaliação do regime pluviométrico e limita a realização de estudos hidrológicos e de planejamento de recursos hídricos. Essas lacunas podem ser advindas de falhas técnicas nos instrumentos de medição, interrupções operacionais das estações meteorológicas ou condições adversas de campo que inviabilizam a coleta de dados (Kajornritm *et al.*, 2012).

A ausência de registros contínuos é especialmente crítica em regiões semiáridas, como o Norte do Estado de Minas Gerais, onde a disponibilidade hídrica depende fortemente da variabilidade climática e da distribuição temporal da precipitação, afetando diretamente a gestão de recursos hídricos, a agricultura e o desenvolvimento socioeconômico local (Afamondji *et al.*, 2025).

Para lidar com séries históricas de precipitação incompletas, diversas abordagens têm sido aplicadas, destacando-se os modelos de regressão linear, que permitem estabelecer relações estatísticas entre os dados das estações pluviométricas e estimar os valores ausentes. Essa técnica oferece uma alternativa eficiente para corrigir lacunas e assegurar a continuidade das séries históricas, fator essencial para análises hidrológicas confiáveis (Oliveira *et al.*, 2010; Hurtado *et al.*, 2021).

Estudos recentes, no entanto, ressaltam que a escolha das estações de apoio deve ser baseada na similaridade dos padrões pluviométricos ao invés da mera proximidade geográfica,

uma vez que esta estratégia garante maior acurácia na estimativa dos dados faltantes (Aguilera *et al.*, 2020; Camuffo *et al.*, 2022).

Apesar desses avanços, ainda existem lacunas na literatura quanto à influência da seleção das estações pluviométricas de apoio na qualidade do preenchimento de falhas em séries históricas de precipitação. Estudos como os de Duarte *et al.* (2022), Martins *et al.* (2023), Fagandini *et al.* (2024) e Hirca e Türkkán (2024) concentram-se na aplicação e comparação de métodos de estimativa, enquanto a etapa de seleção das estações de apoio, que pode afetar significativamente a precisão das estimativas, é frequentemente tratada de forma simplificada ou baseada apenas na proximidade geográfica. Dessa forma, torna-se necessário investigar de forma mais sistemática como diferentes critérios de seleção de estações podem influenciar o desempenho dos métodos de preenchimento de falhas.

Diante desse contexto, torna-se relevante compreender de que maneira a seleção das estações pluviométricas de apoio pode influenciar a precisão do preenchimento de falhas em séries históricas de precipitação. Assim, o presente trabalho teve como objetivo avaliar a influência da seleção das estações pluviométricas de apoio no preenchimento de falhas em séries históricas de precipitação, considerando aspectos como a proximidade geográfica, a altitude e a correlação entre a estação com falhas e cada uma das estações de apoio avaliadas.

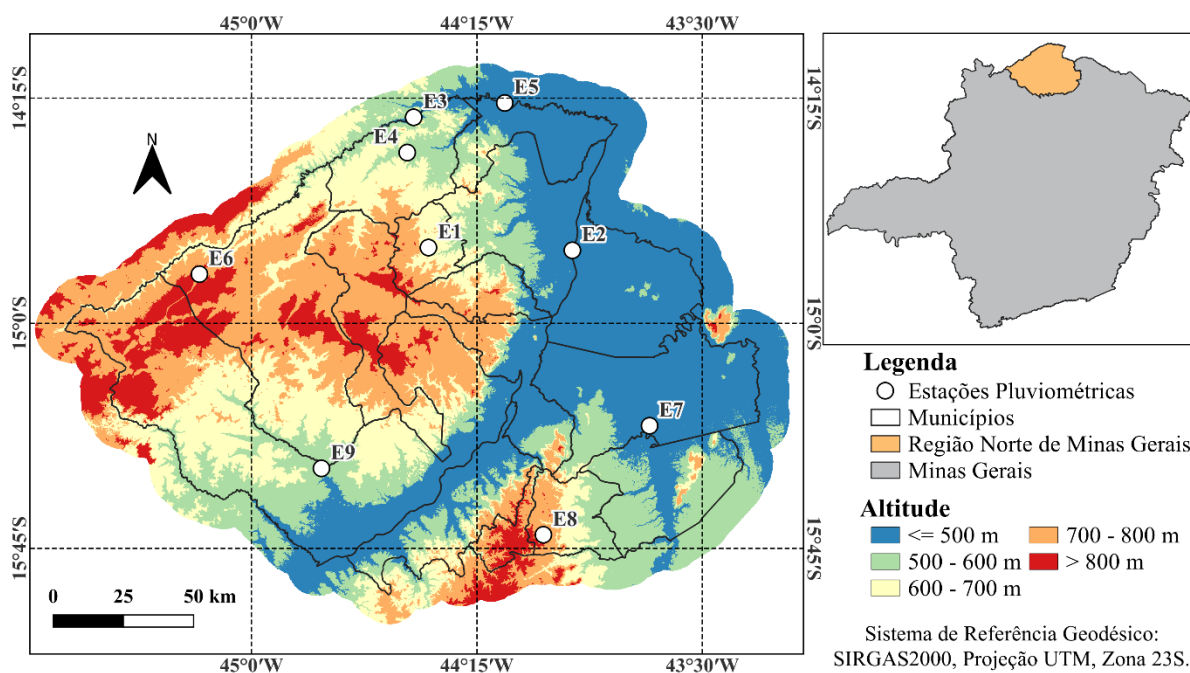
## **2. Material e métodos**

### **2.1. Área de estudo e base de dados**

O estudo foi conduzido utilizando nove estações pluviométricas localizadas no norte de Minas Gerais, conforme ilustrado na Figura 1. A região apresenta um clima semiárido caracterizado por verões chuvosos e invernos secos, com alta variabilidade temporal e espacial da precipitação. A precipitação média anual varia entre 900 e 1.200 mm, concentrando-se principalmente nos meses de verão, enquanto o período seco se estende por cerca de cinco a seis meses, influenciando diretamente a disponibilidade hídrica e as atividades agrícolas locais (Reboita *et al.*, 2015).

As estações selecionadas para este estudo apresentam séries temporais completas, com registros mensais contínuos ao longo de 30 anos, abrangendo o período-base de 1995 a 2024. A base de dados utilizada foi obtida por meio do HidroWeb, plataforma mantida pela Agência

Figura 1: Região Norte de Minas Gerais, com destaque para a localização das estações pluviométricas utilizadas no estudo.



Fonte: Autores.

Nacional de Águas e Saneamento Básico (ANA), que disponibiliza dados hidrológicos para todo o território nacional (<https://www.snirh.gov.br/hidroweb>).

A estação E1 - Miravânia (01444003) foi selecionada como sendo a estação em análise para fins de preenchimento de falhas, por estar centralmente posicionada em relação às demais, as quais foram utilizadas como estações de apoio. A Tabela 1 apresenta os dados das nove estações pluviométricas utilizadas no estudo, incluindo a localização, altitude e distância entre a estação de Miravânia e as estações de apoio ao preenchimento de falhas.

Tabela 1: Estações pluviométricas utilizadas no estudo

Estação	Código	Nome	Latitude	Longitude	Distância (km)*	Altitude (m)
E1	01444003	Miravânia	-14,7475	-44,4111	-----	647
E2	01443001	Manga	-14,7567	-43,9322	51,6	457
E3	01444000	São Gonçalo	-14,3136	-44,4603	48,3	485
E4	01444001	Capitânea	-14,4311	-44,4817	35,83	536
E5	01444004	Juvenilia	-14,2658	-44,1572	59,91	449
E6	01445000	Cajueiro	-14,8361	-45,1733	82,64	826
E7	01543002	Colônia Jaíba	-15,3411	-43,6753	102,83	476
E8	01544030	Varzelândia	-15,7042	-44,0286	113,56	747
E9	01544032	Usina do Pandeiro - montante	-15,4831	-44,7672	89,35	506

\* Distância entre a estação de Miravânia e as estações de apoio ao preenchimento de falhas.

Fonte: Autores.

## 2.2. Preenchimento de falhas

Para avaliar a influência da seleção das estações pluviométricas de apoio no preenchimento de falhas, aplicou-se o método de regressão linear simples (Equação 1). Essa técnica permite ajustar uma equação entre a estação em análise e cada uma das estações de apoio, possibilitando estimar valores ausentes em uma estação a partir dos registros de outra próxima e correlacionada (Bleidorn *et al.*, 2022).

$$P_y = \beta_0 + \beta_1 P_{xi} \quad (1)$$

em que:

- $P_y$  = precipitação estimada na estação de estudo, em mm;
- $P_{xi}$  = precipitação observada na estação de apoio i, com i variando de 2 a 9, de acordo com a estação de apoio, em mm;
- $\beta_0$  = intercepto da equação, adimensional; e
- $\beta_1$  = coeficiente angular, adimensional.

Como todas as estações pluviométricas analisadas possuíam séries históricas completas, foram realizadas simulações de falhas na estação E1 com o objetivo de avaliar o desempenho do método de regressão linear no preenchimento de lacunas a partir das oito estações de apoio.

As falhas foram geradas artificialmente nas séries mensais, por meio de sorteios aleatórios de meses a serem removidos, de modo a representar dois cenários distintos de

ausência de dados: 10% e 40% do total da série. Esse procedimento foi repetido com diferentes combinações de meses em cada simulação, assegurando a aleatoriedade do processo e a robustez da análise.

Na sequência, ajustaram-se modelos de regressão linear entre a estação E1 e cada uma das oito estações de apoio (E2 a E9), e os valores correspondentes aos meses removidos na estação E1 foram então estimados com base nas equações obtidas, conforme feito por Dias e Albuquerque (2021).

Para cada percentual de falhas, o procedimento de sorteio e ajuste foi repetido 100 vezes, com o valor do contador aleatório inicial (*seed*) fixado em 123, a fim de garantir a reprodutibilidade dos resultados. A partir dessa configuração, diferentes amostras foram geradas ao longo das repetições, permitindo avaliar a variabilidade dos erros de estimativa e aumentar a robustez estatística da análise (Sleziak *et al.*, 2020). Os valores preenchidos das falhas (valores simulados) foram então comparados aos valores observados originais (valores que foram retirados na simulação das falhas), a fim de avaliar o desempenho do método de preenchimento e a estação de apoio que melhor serviria como base para o preenchimento.

Esse procedimento seguiu o princípio da validação cruzada (*cross-validation*), muito utilizado em estudos hidrometeorológicos para testar a capacidade preditiva de modelos quando se dispõe dos dados observados reais para comparação (Baudouin *et al.*, 2020).

O ajuste das regressões, o processo de validação cruzada e as avaliações de desempenho apresentadas a seguir foram realizados na linguagem R, por meio do ambiente de desenvolvimento RStudio.

### 2.3. Avaliação de desempenho

A avaliação do desempenho das estimativas foi realizada por meio de três métricas normalmente utilizadas em estudos hidrológicos (Gao *et al.*, 2023; Hosseinzadeh *et al.*, 2023; Onyutha, 2024): RMSE, SMAPE e NSE, permitindo uma análise abrangente da precisão e eficiência do preenchimento de falhas.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Y_o - Y_e)^2}{n}} \quad (2)$$

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_o - Y_e|}{\frac{|Y_o| + |Y_e|}{2}} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Y_o - Y_e)^2}{\sum_{i=1}^n (Y_o - \bar{Y}_e)^2} \quad (4)$$

em que:

- RMSE = Raiz do erro quadrático médio;
- SMAPE = Erro percentual absoluto médio simétrico;
- NSE = coeficiente de Nash-Sutcliffe;
- $Y_o$  = precipitação observada, em mm;
- $Y_e$  = precipitação estimada, em mm;
- $\bar{Y}_e$  = precipitação média estimada, em mm; e
- n = número de observações, adimensional.

O RMSE (Root Mean Squared Error) mede a magnitude média dos erros ao elevar as diferenças entre os valores estimados e observados ao quadrado. Essa característica torna o RMSE especialmente sensível a grandes discrepâncias, penalizando de forma mais intensa os erros extremos. Por esse motivo ele é indicado para identificar situações em que o método apresenta desvios significativos, proporcionando uma medida robusta da precisão absoluta das estimativas (Willmott; Matsuura, 2005; Kanda *et al.*, 2018).

O SMAPE (Symmetric Mean Absolute Percentage Error) expressa o erro em termos percentuais, variando entre 0% e 200%, e adota uma formulação simétrica que evita distorções em séries com valores muito baixos ou próximos de zero. Essa métrica é útil em análises de precipitação, nas quais podem ocorrer valores nulos ou muito pequenos, permitindo uma interpretação mais equilibrada da precisão relativa das estimativas e facilitando a comparação entre diferentes séries ou métodos (Yang e Xing, 2021).

Já o NSE (Nash-Sutcliffe Efficiency) quantifica a eficiência do modelo de preenchimento ao comparar a variância dos erros de estimativa com a variância dos dados observados. Valores de NSE próximos de 1 indicam alta eficiência preditiva, ou seja, forte concordância entre os valores simulados e os observados. Essa métrica é muito empregada em estudos hidrológicos e em modelagem de séries temporais de precipitação, pois avalia a capacidade do modelo de reproduzir de maneira fiel o comportamento real dos dados,

fornecendo uma medida integrada da performance do método (Agarwal *et al.*, 2023; Duc; Sawada, 2023).

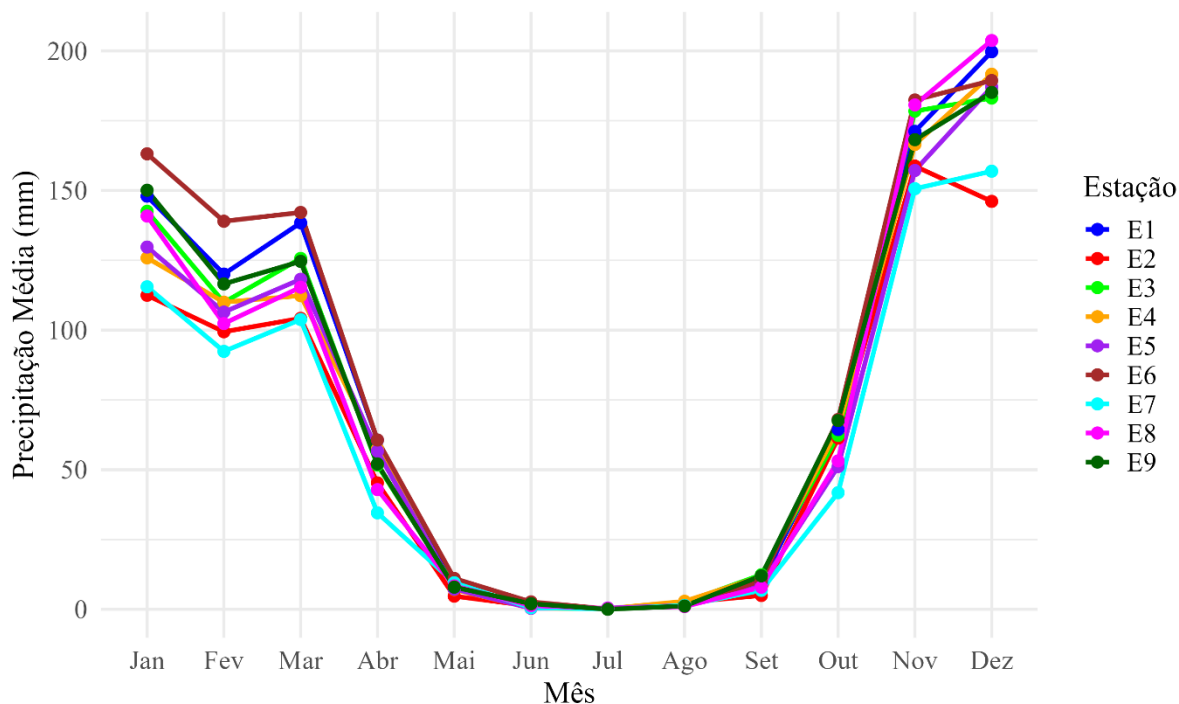
Após a avaliação das métricas de RMSE, SMAPE e NSE, o teste de Kruskal-Wallis foi aplicado para verificar a existência de diferenças significativas entre os valores médios das métricas de desempenho (RMSE, SMAPE e NSE) obtidas para as diferentes estações utilizadas no processo de correção. A escolha desse teste deve-se ao fato de ele ser um método não paramétrico, adequado para comparar três ou mais grupos independentes quando os dados não atendem aos pressupostos de normalidade e homocedasticidade. O procedimento baseou-se nas somas de postos (*ranks*) dos dados agrupados, avaliando a hipótese nula de que todas as amostras provêm da mesma distribuição populacional (Kruskal; Wallis, 1952).

Quando o teste de Kruskal-Wallis indicou diferença estatisticamente significativa ( $p < 0,05$ ), foi aplicado o teste de Dunn como procedimento *post-hoc* para identificar quais pares de estações apresentaram diferenças significativas entre si, também ao nível de significância de 5%. Esse teste realiza comparações múltiplas entre os grupos, utilizando as somas de postos e ajustando os valores de  $p$  (com correções como Bonferroni ou Holm) para controlar o erro do tipo I decorrente das múltiplas comparações (Dunn, 1964).

### 3. Resultados e discussão

Na Figura 2 apresentam-se as precipitações médias mensais, no período de 1995 a 2024, das estações pluviométricas utilizadas no estudo. Conforme apresentado na Figura 2, as precipitações médias mensais registradas nas estações pluviométricas evidenciam um padrão sazonal bastante característico do clima semiárido do Norte de Minas Gerais. Durante os meses de maio a setembro, observa-se uma redução expressiva dos valores de precipitação, refletindo o período seco típico da região, quando a ocorrência de chuvas é escassa e irregular. Por outro lado, entre outubro e abril, as precipitações atingem valores consideravelmente mais elevados, correspondendo ao período chuvoso, que concentra a maior parte do volume pluviométrico anual.

Figura 2: Precipitações médias mensais no período de 1995 a 2024 das estações pluviométricas utilizadas no estudo.



Fonte: Autores.

Na Tabela 2 são apresentados os coeficientes de correlação e de determinação obtidos a partir dos modelos de regressão linear simples ajustados entre a estação E1 e cada uma das oito estações de apoio, juntamente com a distância entre elas. Os resultados indicam uma forte correlação entre a estação central (E1) e as estações vizinhas, evidenciando a representatividade espacial das séries pluviométricas e a adequação do método de regressão linear para o preenchimento de falhas na estação E1.

Entre as estações analisadas, destaca-se a E4, que apresentou o maior coeficiente de correlação e a menor distância geográfica em relação à E1, estando aproximadamente 110 m em altitude inferior. Em seguida, a E3 apresentou o segundo maior valor de correlação, com altitude cerca de 160 m inferior. A estação E5 obteve o terceiro maior coeficiente e a quarta menor distância, estando 100 m abaixo em altitude, enquanto a E2 ocupou a quarta posição em correlação, terceira em proximidade e altitude 190 m inferior.

As estações E6 e E7 apresentaram correlações intermediárias, ocupando a quinta e sexta posição, com diferenças altimétricas de 170 m abaixo e 180 m acima da E1, respectivamente.

Já as estações E8 e E9 apresentaram as menores correlações e as maiores distâncias, estando cerca de 100 m acima e 150 m abaixo da estação E1.

Observa-se na Figura 1 que as estações E2, E3, E4 e E5 estão localizadas a nordeste da estação E1, em áreas de menor altitude e sem a presença de barreiras orográficas expressivas entre elas, o que contribui para a maior similaridade pluviométrica observada entre essas estações. Essa configuração espacial favorece o deslocamento relativamente contínuo das massas de ar úmidas sobre a região, permitindo que os eventos de precipitação ocorram de forma mais simultânea entre essas estações. Em contrapartida, as demais estações encontram-se separadas da E1 por áreas com relevo mais acentuado, o que pode interferir na circulação local do ar e no transporte de umidade, contribuindo para diferenças na ocorrência e na intensidade das chuvas (Marra *et al.*, 2022). Esse padrão ajuda a explicar os menores valores de correlação observados entre essas estações e a estação de referência.

Embora exista uma tendência geral de as estações mais próximas apresentarem maiores valores de correlação em decorrência da maior semelhança nas condições climáticas e topográficas locais, a relação entre distância e correlação nem sempre é linear. Em alguns casos, fatores como diferenças altimétricas, padrões de vento, ou características fisiográficas da bacia podem influenciar a semelhança entre as séries de precipitação, fazendo com que estações mais distantes apresentem correlações comparáveis ou até superiores à outras estações mais próximas (Oriani *et al.*, 2020; Gheysouri *et al.*, 2025). Isso evidencia que a proximidade espacial por si só nem sempre garante um melhor alinhamento entre as séries, como observado nos casos em que a estação mais consistente com o padrão de precipitação não segue a ordem de distância em relação à estação de referência.

Nesse contexto, um dos fatores que pode explicar essas diferenças é a presença de elevações do terreno entre os locais de algumas estações. Essas feições do relevo podem dificultar ou modificar o deslocamento das massas de ar úmidas, fazendo com que determinados eventos de precipitação atinjam algumas estações e não outras, mesmo quando estão relativamente próximas. Como resultado, estações separadas por essas barreiras podem registrar padrões de chuva distintos, enquanto aquelas localizadas em áreas mais abertas tendem a apresentar maior semelhança entre suas séries pluviométricas.

De acordo com Fagandini *et al.* (2024) e Oliveira *et al.* (2010), a seleção de estações de apoio deve priorizar aquelas que apresentem forte relação linear com a estação de referência,

avaliada pelo coeficiente de correlação de Pearson ( $r$ ). É indicado que o coeficiente de determinação ( $r^2$ ) seja igual ou superior a 0,7, valor que indica que pelo menos 70% da variabilidade da série com falhas pode ser explicada pela estação de apoio (Fagandini *et al.*, 2024). Assim, conforme os resultados apresentados na Tabela 2, todas as estações analisadas atendem ao limite mínimo de correlação sugerido, podendo ser consideradas adequadas para o preenchimento das falhas na série histórica da estação E1.

Tabela 2: Resultados da análise de regressão linear e da distância entre a estação de Miravânia (E1) e as estações de apoio (E2 a E9)

Estação	E2	E3	E4	E5	E6	E7	E8	E9
R	0,88	0,89	0,92	0,89	0,85	0,86	0,85	0,84
R <sup>2</sup>	0,77	0,80	0,84	0,79	0,73	0,74	0,73	0,70
Distância (km)	51,6	48,3	35,83	59,91	82,6	102,83	113,56	89,35

Fonte: Autores.

Na Tabela 3 são apresentados os valores do RMSE, SMAPE e NSE para os cenários de 10% e 40% de falhas na série da estação em análise. Os valores médios das métricas para 10% e 40% foram semelhantes, com valores piores para o cenário de 40% de falhas em todas as situações. Além disso observou-se maior variabilidade das métricas no cenário de 10%, o que pode ser atribuído à maior quantidade de dados disponíveis para o ajuste da regressão linear simples. Com mais pontos de observação, o modelo tende a capturar melhor as oscilações naturais da série, resultando em uma dispersão maior dos indicadores de desempenho. Em contrapartida, no cenário de 40% de falhas, a menor quantidade de dados disponíveis reduz a capacidade do modelo de representar essas variações, o que tende a gerar resultados mais uniformes, porém menos representativos da variabilidade real da precipitação (Cunha Júnior; Firmino, 2022).

Os preenchimentos de falhas realizados utilizando-se as estações de apoio E2, E3, E4 e E5 destacaram-se pelos menores valores de RMSE, SMAPE e os maiores valores de NSE, refletindo a alta correlação entre essas estações e a estação E1. Valores semelhantes dessas métricas também foram encontrados por Dias e Albuquerque (2021), mostrando que essa correlação reflete a associação entre os eventos pluviométricos registrados em duas estações, indicando a ocorrência de períodos de aumento e diminuição da precipitação. Valores elevados de correlação sugerem que os picos e as reduções de precipitação em uma estação tendem a ser

replicados temporalmente na outra, evidenciando uma resposta conjunta aos mesmos sistemas atmosféricos de grande escala e ao deslocamento das massas de ar que influenciam a dinâmica pluviométrica sobre Minas Gerais (Mallick *et al.*, 2022).

De acordo com Reboita *et al.* (2015), durante o verão predominam ventos de nordeste em quase todo o estado, enquanto no extremo norte os ventos sopram de leste. Essa configuração favorece o deslocamento das chuvas no sentido sudoeste-nordeste, atingindo inicialmente a estação E1 e, em seguida, as estações E3, E4 e E5. A estação E1 recebe precipitação associada aos ventos de leste que primeiro atingem a estação E2, o que explica a elevada autocorrelação entre as séries temporais dessas estações.

A estação E9 apresentou o pior desempenho nas métricas RMSE, SMAPE e NSE devido à presença de barreiras orográficas que, segundo Nicolas e Boss (2025), restringem o compartilhamento de eventos pluviométricos com estações próximas, mesmo estando relativamente próximas. A estação E7, por sua vez, apresentou desempenho comparável e/ou superior à estação E6, mesmo estando mais distante, evidenciando a influência da topografia e da direção predominante dos ventos na distribuição das chuvas.

Contudo, não é possível estabelecer uma relação direta entre a semelhança altimétrica e a eficiência no preenchimento de falhas. Observa-se que a estação E8, apesar de apresentar a menor diferença de altitude em relação à estação de referência, obteve um dos piores desempenhos no preenchimento. Em contrapartida, a estação E4, que possui a segunda altitude mais próxima, apresentou os melhores resultados. Já a estação E9, com a terceira maior proximidade altimétrica, obteve um dos piores desempenhos, enquanto a estação E3 apresentou boa eficiência mesmo com diferença intermediária de altitude. As estações E7 e E6, por sua vez, registraram resultados insatisfatórios, ao passo que as estações E2 e E5, apesar das maiores diferenças altimétricas, apresentaram bons desempenhos no preenchimento, conforme mostrado na Tabela 3.

Na área de estudo, observa-se que as estações separadas por barreiras orográficas também apresentaram maiores distâncias entre si, o que sugere que os efeitos do relevo podem ter atuado de forma combinada com a distância na redução da correlação entre as estações de apoio e a estação de estudo, bem como na diminuição da eficiência na estimação das precipitações.

Tabela 3: Valores médios e desvio padrão (entre parênteses) para RMSE, SMAPE e NSE entre os dados observados na estação E1 – Miravânia e os preenchidos em cada estação de apoio (E2 a E9), com letras diferentes indicando diferenças significativas segundo os testes de Kruskal-Wallis e de Dunn

Falhas	E2	E3	E4	E5	E6	E7	E8	E9
	<b>RMSE</b>							
10%	47,38bc (8,85)	45,65b (9,35)	39,76a (7,28)	46,55b (8,69)	52,66d (10,17)	51,55cd (9,66)	51,52cd (11,37)	52,92d (14,12)
40%	49,29c (3,85)	45,97b (4,06)	40,34a (2,65)	47,76bc (4,03)	53,22d (3,42)	52,54d (3,94)	53,97d (4,77)	54,98d (5,74)
	<b>SMAPE</b>							
10%	96,23ab (11,76)	96,29ab (11,98)	94,63a (11,95)	97,77ab (11,76)	101,18b (11,83)	101,01b (11,77)	99,20ab (12,21)	100,29b (11,15)
40%	96,04a (5,48)	95,37a (5,34)	93,55a (5,34)	96,07a (5,41)	100,56b (5,48)	99,97b (5,53)	98,85b (5,61)	99,32b (5,49)
	<b>NSE</b>							
10%	0,76b (0,06)	0,78b (0,07)	0,83a (0,08)	0,76b (0,09)	0,70d (0,10)	0,71cd (0,12)	0,71cd (0,12)	0,70d (0,13)
40%	0,76c (0,04)	0,79b (0,03)	0,84a (0,02)	0,78bc (0,03)	0,72d (0,04)	0,73d (0,05)	0,71d (0,05)	0,70d (0,06)

A distribuição espacial da precipitação nas estações reflete a influência combinada de sistemas atmosféricos de grande escala, como as frentes frias e o Anticiclone Subtropical do Atlântico Sul (ASAS), os quais promovem o transporte de umidade do oceano Atlântico e modulam o regime de chuvas sobre o Sudeste do Brasil (Assis *et al.*, 2013). No norte de Minas Gerais, durante o verão, o transporte de umidade proveniente do Atlântico, impulsionado pelos ventos predominantes de leste e nordeste, é o principal responsável pela formação de chuvas frontais. Essa circulação favorece a chegada das precipitações primeiramente na estação E2, situada mais a Leste, e, posteriormente, na estação E1, localizada mais ao interior, devido ao deslocamento oeste da umidade advectada. Já as estações E3, E4 e E5, localizadas mais ao norte, são mais influenciadas pela expansão meridional da Zona de Convergência do Atlântico Sul (ZCAS) e por fluxos úmidos provenientes da Amazônia, o que explica a maior variabilidade observada nessa porção da área de estudo.

Conforme evidenciado por Pezzi *et al.* (2023), a Zona de Convergência do Atlântico Sul (ZCAS) atua como um rio atmosférico de origem oceânica, transportando volumes significativos de umidade da Amazônia em direção ao Sudeste do Brasil e modulando a ocorrência de precipitação sobre as estações E1 a E5.

Os resultados apresentados na Tabela 3 indicam que a estação E4 obteve desempenho significativamente superior em relação às demais, conforme apontado pelos testes de Kruskal-

Wallis e Dunn, demonstrando maior capacidade de representar os dados observados na estação E1 – Miravânia. As estações E2, E3 e E5 apresentaram resultados estatisticamente semelhantes ( $p > 0,05$ ), sugerindo que também possuem bom potencial para o preenchimento de falhas, embora com desempenho ligeiramente inferior à E4. Já as estações E7 e E8 mostraram comportamento intermediário, enquanto E6 e E9 apresentaram os menores índices de desempenho, sem diferença significativa entre si.

Esse comportamento pode estar associado, no contexto da área de estudo, à influência conjunta da distância e das barreiras orográficas. Observa-se que as estações mais distantes apresentam barreiras orográficas em relação à estação de estudo, o que resultou em menores coeficientes de correlação e menor eficiência no ajuste das regressões para reproduzir as chuvas. Por outro lado, as estações mais próximas, que não sofrem influência dessas barreiras e apresentam maiores correlações, proporcionaram ajustes mais eficientes.

Ao analisar os diferentes cenários de falhas (10% e 40%), observa-se que, embora a redução na quantidade de dados disponíveis afete os resultados, os métodos de preenchimento apresentaram resultados consistentes, com valores próximos e convergindo para médias semelhantes, especialmente nas estações com maior correlação e comportamento pluviométrico semelhante. No cenário de 10% de falhas, os valores de RMSE e SMAPE permaneceram baixos, e o NSE indicou bom ajuste, evidenciando que pequenas lacunas podem ser preenchidas com elevada confiabilidade.

No cenário de 40% de falhas, verificou-se um aumento esperado nos erros e uma leve redução na eficiência do preenchimento, em razão da menor representatividade das séries de referência. Ainda assim, o desempenho manteve-se próximo ao observado no cenário de 10% de falhas, o que sugere que o método pode ser aplicado, desde que as estações de apoio sejam selecionadas com base na proximidade geográfica e na ausência de barreiras orográficas que influenciem o regime de chuvas.

#### 4. Conclusões

Com base nos resultados obtidos, pode-se concluir que:

- No preenchimento de falhas em séries históricas de precipitação deve-se priorizar a utilização de estações de apoio que apresentem elevada correlação com a estação com

falhas em seus registros, uma vez que a correlação indica a similaridade no comportamento pluviométrico das séries;

- A proximidade geográfica e a semelhança altimétrica entre as estações contribuem para o aumento da correlação, especialmente quando não há barreiras orográficas significativas entre elas, o que favorece a ocorrência de chuvas simultâneas e a consistência dos dados; e
- Embora os resultados médios entre os cenários de 10% e 40% de falhas tenham sido próximos, o aumento do percentual de dados ausentes reduz a capacidade do modelo de capturar a variabilidade natural da precipitação, resultando em estimativas mais uniformes, porém menos representativas da dinâmica real das chuvas.

### Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

### Referências

- Afamondji, K. M. C., Mohsine, I., Karmouda, N., Bouramtane, T., Tahiri, M., Kacimi, I., Tweed, S., Driouech, F & Leblanc, M. (2025). Climate Extreme Indices Influencing GRACE Total Water Storage (TWS) in Semi-Arid Africa. *Journal of Arid Environments*, 229, 105357. <https://doi.org/10.1016/j.jaridenv.2025.105357>
- Agarwal, S., Mukherjee, D., & Debbarma, N. (2023). Determination of annual rainfall in north-east India using deterministic, geospatial, and machine learning techniques. *Water Policy*, 25(12), 1113-1124. <https://doi.org/10.2166/wp.2023.078>
- Aguilera, H., Guardiola-Albert, C., & Serrano-Hidalgo, C. (2020). Estimating extremely large amounts of missing precipitation data. *Journal of Hydroinformatics*, 22(3), 578-592. <https://doi.org/10.2166/hydro.2020.127>
- Assis, W. L., Parizzi, M. G., & Ribeiro, S. M. (2013). Identificação dos sistemas atmosféricos responsáveis por inundações e deslizamentos no município de Belo Horizonte: análise dos

períodos chuvosos de 2011-2012 e 2012-2013. *Revista Geografias*, 9(2), 38-55.

<https://doi.org/10.35699/2237-549X..13360>

Baudouin, J. P., Herzog, M., & Petrie, C. A. (2020). Cross-validating precipitation datasets in the Indus River basin. *Hydrology and Earth System Sciences*, 24(1), 427-450.

<https://doi.org/10.5194/hess-24-427-2020>

Bleidorn, M. T., Pinto, W. D. P., Schmidt, I. M., Mendonça, A. S. F., & Reis, J. A. T. D. (2022). Methodological approaches for imputing missing data into monthly flows series. *Revista Ambiente & Água*, 17, e2795. <https://doi.org/10.4136/ambi-agua.2795>

Camuffo, D., Becherini, F., della Valle, A., & Zanini, V. (2022). A comparison between different methods to fill gaps in early precipitation series. *Environmental Earth Sciences*, 81(13), 345. <http://doi.org/10.1007/s12665-022-10467-w>

Cunha Júnior, R. O., & Firmino, P. R. A. (2022). Simulação de valores ausentes em séries temporais de precipitação para avaliação de métodos de imputação. *Revista Brasileira de Climatologia*, 30, 691-714. <https://doi.org/10.55761/abclima.v30i18.15243>

Dias, A. S., & de Albuquerque Soares, W. (2021). Uso de metodologias de preenchimento de falhas para estimativas de dados de precipitação. *Research, Society and Development*, 10(5), e57610515383-e57610515383. <https://doi.org/10.33448/rsd-v10i5.15383>

Duarte, L. V., Formiga, K. T. M., & Costa, V. A. F. (2022). Comparison of methods for filling daily and monthly rainfall missing data: statistical models or imputation of satellite retrievals?. *Water*, 14(19), 3144. <https://doi.org/10.3390/w14193144>

Duc, L., & Sawada, Y. (2023). A signal-processing-based interpretation of the Nash–Sutcliffe efficiency. *Hydrology and Earth System Sciences*, 27(9), 1827-1839.

<https://doi.org/10.5194/hess-27-1827-2023>

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241-252.

<http://doi.org/10.1080/00401706.1964.10490181>

Fagandini, C., Todaro, V., Tanda, M. G., Pereira, J. L., Azevedo, L., & Zanini, A. (2024). Missing rainfall daily data: a comparison among gap-filling approaches. *Mathematical Geosciences*, 56(2), 191-217. <https://doi.org/10.1007/s11004-023-10078-6>

Gao, Y., Taie Semiromi, M., & Merz, C. (2023). Efficacy of statistical algorithms in imputing missing data of streamflow discharge imparted with variegated variances and

- seasonalities. *Environmental Earth Sciences*, 82(20), 476. <http://doi.org/10.1007/s12665-023-11139-z>
- Gheysouri, M., Sigaroodi, S. K., Salajegheh, A., Choubin, B., & Liu, B. (2025). Orographic changes in precipitation using gradient-based IMERG data assessment in the Alborz Mountains, Iran. *Advances in Space Research*, 75(11), 7800-7816. <https://doi.org/10.1016/j.asr.2025.03.016>
- Hosseinzadeh, P., Nassar, A., Boubrahimi, S. F., & Hamdi, S. M. (2023). MI-based streamflow prediction in the upper Colorado river basin using climate variables time series data. *Hydrology*, 10(2), 29. <https://doi.org/10.3390/hydrology10020029>
- Hirca, T., & Türkkän, G. E. (2024). Assessment of different methods for estimation of missing rainfall data. *Water Resources Management*, 38(15), 5945-5972. <https://doi.org/10.1007/s11269-024-03936-3>
- Hurtado, S. I., Zaninelli, P. G., Agosta, E. A., & Ricetti, L. (2021). Infilling methods for monthly precipitation records with poor station network density in Subtropical Argentina. *Atmospheric Research*, 254, 105482. <https://doi.org/10.1016/j.atmosres.2021.105482>
- Kajornrit, J., Wong, K. W., & Fung, C. C. (2012). A comparative analysis of soft computing techniques used to estimate missing precipitation records. <https://hdl.handle.net/10419/72486>
- Kanda, N., Negi, H. S., Rishi, M. S., & Shekhar, M. S. (2018). Performance of various techniques in estimating missing climatological data over snowbound mountainous areas of Karakoram Himalaya. *Meteorological Applications*, 25(3), 337-349. <https://doi.org/10.1002/met.1699>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621. <http://doi.org/10.1080/01621459.1952.10483441>
- Mallick, J., Talukdar, S., Almesfer, M. K., Alsubih, M., Ahmed, M., & Islam, A. R. M. T. (2022). Identification of rainfall homogenous regions in Saudi Arabia for experimenting and improving trend detection techniques. *Environmental Science and Pollution Research*, 29(17), 25112-25137. <http://doi.org/10.1007/s11356-021-17609-w>

- Marra, F., Armon, M., & Morin, E. (2022). Coastal and orographic effects on extreme precipitation revealed by weather radar observations. *Hydrology and Earth System Sciences*, 26(5), 1439-1458. <https://doi.org/10.5194/hess-26-1439-2022>
- Martins, L.L., Martins, W. A., Rodrigues, I. C. D. A., Freitas Xavier, A. C., Moraes, J. F. L. D., & Blain, G. C. (2023). Gap-filling of daily precipitation and streamflow time series: a method comparison at random and sequential gaps. *Hydrological Sciences Journal*, 68(1), 148-160. <https://doi.org/10.1080/02626667.2022.2145200>
- Nicolas, Q., & Boos, W. R. (2025). Sensitivity of tropical orographic precipitation to wind speed with implications for future projections. *Weather and Climate Dynamics*, 6(1), 231-244. <https://doi.org/10.5194/wcd-6-231-2025>
- Oliveira, L. F., Fioreze, A. P., Medeiros, A. M., & Silva, M. A. (2010). Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 14, 1186-1192. <https://doi.org/10.1590/S1415-43662010001100008>
- Onyutha, C. (2024). Pros and cons of various efficiency criteria for hydrological model performance evaluation. *Proceedings of IAHS*, 385, 181-187. <https://doi.org/10.5194/piahs-385-181-2024>
- Oriani, F., Stisen, S., Demirel, M. C., & Mariethoz, G. (2020). Missing data imputation for multisite rainfall networks: a comparison between geostatistical interpolation and pattern-based estimation on different terrain types. *Journal of Hydrometeorology*, 21(10), 2325-2341. <https://doi.org/10.1175/JHM-D-19-0220.1>
- Pezzi, L. P., Quadro, M. F., Souza, E. B., Miller, A. J., Rao, V. B., Rosa, E. B., ... & António, J. F. (2023). Oceanic SACZ produces an abnormally wet 2021/2022 rainy season in South America. *Scientific Reports*, 13(1), 1455. <http://doi.org/10.1038/s41598-023-28803-w>
- Sleziak, P., Holko, L., Danko, M., & Parajka, J. (2020). Uncertainty in the number of calibration repetitions of a hydrologic model in varying climatic conditions. *Water*, 12(9), 2362. <https://doi.org/10.3390/w12092362>
- Reboita, M. S., Rodrigues, M., Silva, L. F., & Alves, M. A. (2015). Aspectos climáticos do estado de Minas Gerais. *Revista brasileira de Climatologia*, 17. <https://doi.org/10.5380/abclima.v17i0.41493>

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

Yang, R., & Xing, B. (2021). A comparison of the performance of different interpolation methods in replicating rainfall magnitudes under different climatic conditions in Chongqing Province (China). *Atmosphere*, 12(10), 1318.

<https://doi.org/10.3390/atmos12101318>

**Publisher:** Universidade Federal de Jataí. Instituto de Geografia. Programa de Pós-graduação em Geografia. Publicação no Portal de Periódicos UFJ. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

**Contribuições dos autores:** Anderson de Oliveira Pinheiro: processamento dos dados, análise dos resultados, escrita; Demetrius David da Silva: escrita, revisão e edição; Michel Castro Moreira: escrita, revisão e edição.

**Financiamento:** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Brasil (CAPES), sob o número de concessão 88887.955847/2024-00 (Código de Financiamento 001).

**Conflito de interesse:** Os autores declaram que não possuem interesses financeiros ou não financeiros relevantes relacionados a este trabalho.