Revista Brasileira de Climatologia
Brazilian Journal of Climatology
OPEN ACCESS
UFGD Editora

TÉCNICAS INDIVIDUAIS E COMBINADAS PARA PREENCHIMENTO DE FALHAS EM DADOS DIÁRIOS DE PRECIPITAÇÃO NO MUNICÍPIO DE SÃO GONÇALO (RJ)

Individual and combined techniques for gap filling in daily rainfall data in the city of São Gonçalo (RJ)

Técnicas individuales y combinadas para relleno de huecos en datos de lluvia diaria en la ciudad de São Gonçalo (RJ)

Anna Regina Corbo Costa
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – Cefet/RJ
anna.costa@cefet-rj.br

Suellen Araujo Franco dos Santos
Faculdade de Formação de Professores – Universidade do Estado do Rio de Janeiro (UERJ/FFP)
suellenaraujo95@gmail.com

Ana Valéria Freire Allemão Bertolino
Faculdade de Formação de Professores – Universidade do Estado do Rio de Janeiro (UERJ/FFP)
anabertolino@uol.com.br

Ana Beatriz de Souza Pinto
Defesa Civil de São Gonçalo - RJ
ana_beatriz@ymail.com

Resumo: A precipitação é uma variável importante na caracterização climatológica de uma localidade. A existência de registros faltantes em uma série temporal de chuva reduz a quantidade de informações disponíveis para a análise, muitas vezes limitando a utilidade da série. O presente trabalho tem como objetivo avaliar as técnicas individuais de preenchimento: ponderação pelo inverso da distância (IDW), razão normal (RN), regressão linear múltipla (RLM) e a combinação destas mesmas técnicas utilizando média simples (CMS) e a combinação ótima com variância mínima (CVM) para dados diários de chuva da Estação Climatológica da UERJ/FFP, no município de São Gonçalo – RJ. Para isto, foram utilizadas observações de cinco estações vizinhas para preenchimento de séries temporais com 35, 87 ou 351 dias seguidos de falhas. As simulações indicaram que a estimativa individual obtida pela regressão linear múltipla foi a mais ajustada com $R^2 > 0.9$ nas três configurações de período de falha analisadas. Já dentre os métodos combinados, a combinação por

média simples foi a mais acurada com R^2 = 0.97 e RMSE de 0.55mm. Os resultados indicam que a superestimação obtida pelos métodos de preenchimento não é estatisticamente significativa para α=5%. Este estudo visa a levantar dados locais disponibilizados por diferentes órgãos e a analisar os métodos de preenchimento mais indicados de modo a minimizar os impactos causados por séries parcialmente completas em estudos climatológicos para o município.
Palavras-chave: Séries temporais. Preenchimento de falhas. Combinação de métodos.

Abstract:

Precipitation is an important variable for climatological characterization of a location. The existence of missing records in a rainfall time series reduces the amount of information available for analysis, often limiting the usefulness of the series. This study evaluated the individual filling techniques inverse distance weighting (IDW), normal ratio (NR) and multiple linear regression (MLR) along with the combination of these techniques using simple average and the optimal combination with minimum variance for daily rainfall data from the UERJ/FFP Climatological Station, located in the municipality of São Gonçalo, Rio de Janeiro State. For this, observations from five neighboring stations were used to fill time series with 35, 87 or 351 consecutive days of failures. The simulations indicated that the individual estimate obtained by multiple linear regression had the best adjustment with $R^2 > 0.9$ in the three failure period configurations analyzed. Among the combined methods, the combination by simple average was the most accurate, with R^2 = 0.97 and RMSE of 0.55 mm. The results indicated that the overestimation obtained by the gap filling methods was not statistically significant at α = 5%. This analysis was based on local data made available by different agencies to determine the most suitable gap filling methods to minimize the impacts caused by incomplete series in climatological studies for the municipality.
Keywords: Time series. Gap filling. Combination of methods.

Resumen: La precipitación es una variable importante en la caracterización climatológica de una localidad. La existencia de registros faltantes en una serie temporal de lluvia disminuye la cantidad de informaciones para el análisis, lo que limita la utilidad de ésa serie. El objetivo de esa investigación es evaluar las técnicas individuales de interpolación espacial: distancia inversa ponderada (IDW, razón normal (RN), regresión lineal múltiple (RLM) y la combinación de estas técnicas utilizando el promedio simple (CMS) y la combinación óptima con varianza mínima (CVM) para datos diarios de lluvia en la Estación Climatológica de la Universidad del Estado de Río de Janeiro (UERJ/FFP), en la ciudad de São Gonçalo (Río de Janeiro, Brasil). Para ello, se han utilizado observaciones recogidas en cinco estaciones vecinas para interpolación espacial de series temporales con 35, 87 o 351 días consecutivos de fallos. Las simulaciones han indicado que la estimación individual obtenida por RLM fue la más ajustada, con $R^2 > 0,9$ en las tres configuraciones de período de falla analizadas. Entre los métodos combinados, CMS fue el más acurado con R^2 = 0,97 y RMSE de 0, 55mm. Los resultados indican que la sobreestimación obtenida por los métodos de interpolación espacial no es estadísticamente significativa para α = 5%. Esta investigación tiene, como objetivo, recoger datos locales disponibles producidos por distintas instituciones, además de analizar los métodos de interpolación espacial más indicados de manera a disminuir los impactos que tienen las series parcialmente completadas en los estudios climatológicos en el municipio.
Palabras clave: series temporales; interpolación de fallos; combinación de métodos.

Submetido em: 08/08/2023
Aceito para publicação em: 19/09/2024
Publicado em: 17/10/2024

1. INTRODUÇÃO

A precipitação é uma das variáveis mais utilizadas em estudos de carácter climatológico e hidrológico. Ao longo do tempo, o estudo desta variável fornece subsídios para a gestão inteligente dos recursos hídricos e para o planejamento urbano. Para isso, é ideal que a série de dados não apresente falhas, de modo a obter um resultado confiável e representativo, o que muitas vezes não acontece na prática. A ocorrência de períodos sem informações nos dados de precipitação é comum e estas falhas podem ocorrer por fatores diversos. Esses valores em falta são um importante problema na análise e modelagem de processos hidrológicos. Diversos estudos têm se concentrado em desenvolver estratégias para estimar razoavelmente estes valores de modo a preencher tais falhas (Little; Rubin, 2019; Naghettini; Pinto, 2017; Ruezzene *et al.*, 2020).

Em geral, as técnicas de preenchimento de falhas em séries históricas consistem em estimar os dados faltantes usando dados de estações meteorológicas vizinhas através de processos de interpolação espacial. Quando se trata da precipitação, a maior parte dos estudos se concentra em analisar falhas em séries históricas de dados mensais ou anuais (Correia *et al.*, 2016; Junqueira; Amorim; Oliveira, 2018), considerando a alta variabilidade inerente ao dado diário. No entanto, uma vez que o dado diário é de extrema importância na análise de eventos extremos, existem alguns estudos voltados para validar técnicas de preenchimento para este tipo de dado (Brubacher; Oliveira; Guasselli, 2020; Depiné *et al.*, 2014; Longman *et al.*, 2020).

Podemos citar, como métodos baseados em interpolação espacial, a média aritmética (MA), ponderação regional (PR), a interpolação pelo inverso do quadrado da distância (IDW, do inglês) e a razão normal (RN) como os mais tradicionais para a estimativa de falhas em dados de chuva (Bielenki Junior *et al.*, 2018; Kim; Ryu, 2016; Little; Rubin, 2019; Tucci, 2001;).

Métodos baseados em regressão linear múltipla (RLM) também são bastante utilizados na literatura. A partir da RLM, as informações de precipitação em uma dada estação são linearmente correlacionadas com os dados correspondentes em mais de uma estação vizinha. A RLM consta, na literatura, como um dos métodos de melhor desempenho

inclusive quando comparado a técnicas baseadas em inteligência artificial e redes neurais (Brubacher; Oliveira; Guasselli, 2020; Depiné *et al.*, 2014; Ruezzene *et al.*, 2021).

Bates e Granger (1969) sugeriram combinar diferentes previsões ao invés de se concentrar na seleção do melhor modelo de estimação para uma variável. Desde então, numerosos métodos de combinação foram propostos e a literatura indica que a combinação linear de previsões comumente é mais acurada do que as previsões individuais que a compõem (Atiya, 2020; Martins; Werner, 2014).

Para Mendonça e Medrano (2016), um dos principais argumentos para o uso de combinação de previsões é o fato de que o verdadeiro processo de geração da variável na natureza é desconhecido. Para os autores, é provável que mesmo o modelo de formulação mais complicada esteja mal especificado, sendo, na melhor das hipóteses, uma razoável aproximação local do verdadeiro modelo.

Apesar de ser uma metodologia muito utilizada em dados econométricos para previsões mais assertivas, há poucos relatos na literatura do uso da combinação em previsões em dados hidrológicos. Papacharalampous e Tyralis (2020) utilizam um método de combinação de previsões baseado na mediana das previsões individuais para um conjunto muito grande de dados de vazão. Já Poletti *et al.* (2019) descrevem a utilização de um modelo híbrido que explora dados observados em campo e modelos de previsão para melhorar a acurácia das previsões. Todavia, os trabalhos que se utilizam de combinação de modelos em dados hidrológicos são voltados para modelos estatísticos de previsão futura e não de preenchimento de falhas do passado, objetivo principal de nosso estudo.

O município de São Gonçalo está localizado na região metropolitana do Rio de Janeiro e possui um histórico expressivo de ocorrências de enchentes e deslizamentos associados à ocorrência de chuva (Queiroz; Nunes, 2022; Silva; Lamego; Santos, 2017; Tórnio; Kede, 2021).

Queiroz e Nunes (2022) enfatizam a necessidade de atenção ao município que é constantemente afetado por diversos transtornos relacionados à precipitação em conjunto com a complexidade e vulnerabilidade de sua estrutura e dinâmica ambiental, sendo associadas à ocupação desenfreada do município. Silva, Lamego e Santos (2017) ressaltam a percepção da população em relação aos problemas que ocorrem relacionados a eventos de precipitação, com a sua maioria tendo passado por algum transtorno, como alagamentos e

deslizamentos de terra. Tórnio e Kede (2021) apontam a ocorrência de casos de alagamentos e deslizamentos durante a estação chuvosa, concentrando-se entre os meses de janeiro e abril. Esses casos ocorrem majoritariamente em bairros centrais de São Gonçalo, que possuem comércio ativo e/ou são mais populosos dentro do município.

Desta forma, este trabalho tem, como objetivo, analisar o preenchimento de falhas em dados diários de precipitação através das técnicas individuais de interpolação espacial como inverso da distância (IDW), regressão linear múltipla (RLM), razão normal (RN) e o preenchimento utilizando a combinação dos métodos citados via média simples e método da variância mínima, para o município de São Gonçalo. Além disto, verificou-se a acurácia destes, para as condições da localidade, através de métricas estatísticas de erro.

2. METODOLOGIA

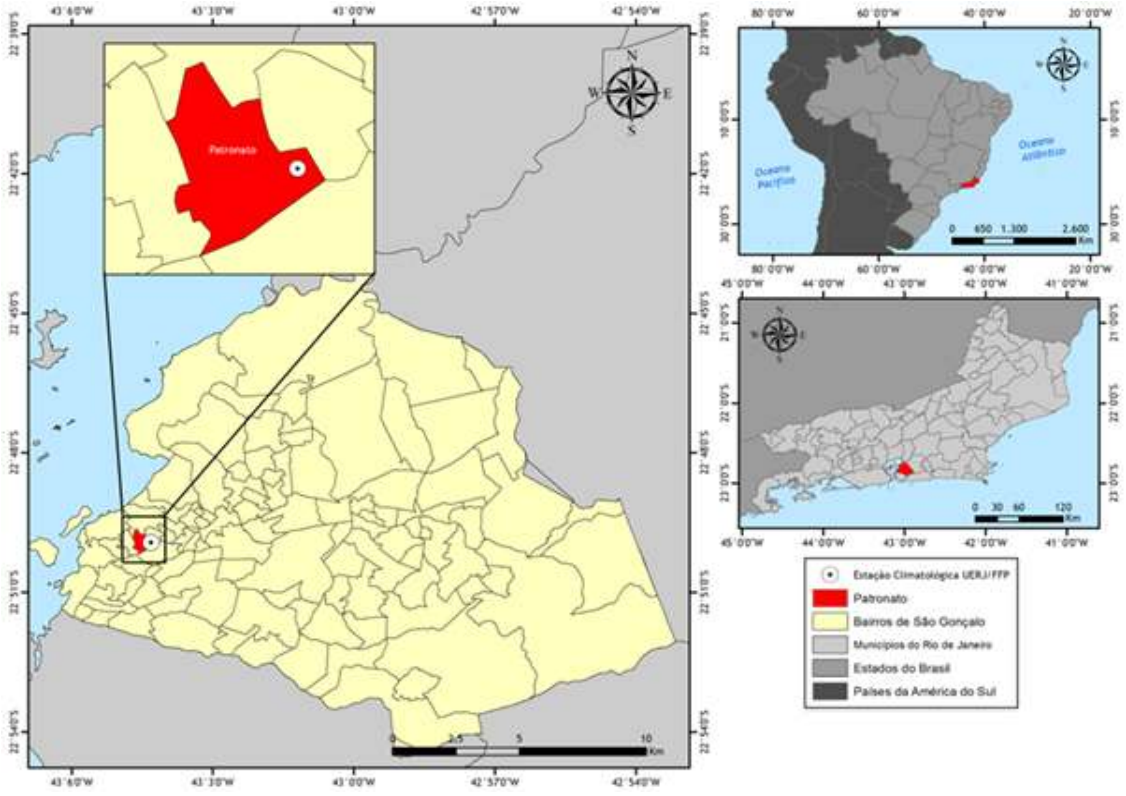
2.1. Área de estudo

O município de São Gonçalo pertence à Região Metropolitana do Rio de Janeiro. Possui uma extensão territorial de 248,160 km² (IBGE, 2021). Localiza-se entre as coordenadas 22º 49' 55,97" S e 43º 4' 25,52" O, tendo como limites, ao norte, Itaboraí e Baía de Guanabara; ao sul, Maricá e Niterói; a leste, Itaboraí e Maricá e a oeste, a Baía de Guanabara. Na Figura 1, observa-se o bairro Patronato, onde está localizada a Estação Climatológica Urbana da UERJ/FFP.

No contexto da região, ocorre um clima Aw, conforme a classificação de Köppen, tendo clima quente com chuvas de verão e outono. Já o período seco apresenta-se entre maio e outubro com chuvas menores do que 100 mm. Há maior ocorrência de chuvas entre novembro e abril apresentando-se valores superiores a 100 mm (Pereira *et al.*, 2021).

A região, historicamente, carece de dados de monitoramento ambiental. Ao contrário de municípios fluminenses vizinhos, como Niterói e Maricá, São Gonçalo não possuía, antes de 2012, estações meteorológicas integradas às redes governamentais de monitoramento, o que é um grande entrave na busca por dados históricos para a região.

Figura 1 - Localização do município de São Gonçalo-RJ e a Estação Climatológica Experimental Urbana.



Fonte: Elaborado pelos autores (2024).

2.2. Dados pluviométricos

Para a realização dos estudos, foram selecionadas 18 estações pluviométricas da região pertencentes à UERJ - Faculdade de Formação de Professores (UERJ-FFP), ao Centro Nacional de Monitoramento e Alerta de Desastres Naturais (CEMADEN), ao Instituto Estadual do Ambiente (INEA) e a Defesa Civil de São Gonçalo. Os dados pluviométricos do CEMADEN foram obtidos através do banco de dados público *online*, ao passo que os dados das estações pluviométricas da UERJ-FFP, do INEA e da Defesa Civil de São Gonçalo foram cedidos pelas equipes gestoras.

A estação escolhida para ter seus dados preenchidos foi a Estação Experimental Climatológica Urbana da UERJ-FFP. Esta estação é manual, não integrada à rede governamental e possui uma série temporal histórica de 18 anos com algumas falhas no decorrer do período, o que motivou a elaboração deste estudo. As outras 17 estações se encontram no mesmo município ou em município vizinho (Niterói) e distam até 12 km da estação de estudo. A estação da UERJ-FFP possui dados desde 2004, o INEA possui estações

de monitoramento desde 2012 e o CEMADEN e a Defesa Civil, desde 2019. Na Tabela 1, podem-se observar os nomes de cada estação, o município de localização, distância em relação à estação climatológica da UERJ-FFP, instituto a que pertencem e datas originais de início e fim dos dados brutos.

Tabela 1- Estações utilizadas na primeira fase da criação da base de dados pluviométricos.

Nome	Instituto	Município	Distância FFP (km)	Início da série	Fim da série
UERJ-FFP	UERJ	São Gonçalo	0	01/03/2005	16/03/2020
Charitas	CEMADEN	Niterói	11.5	01/01/2019	29/12/2021
Praia João Caetano		Niterói	9.4	01/01/2019	31/12/2021
Morro do Castro		Niterói	4.1	06/01/2019	31/12/2021
Pita		São Gonçalo	1.8	01/01/2019	31/12/2021
Visconde de Itaboraí		Niterói	7.5	01/01/2019	31/12/2021
Arsenal	Defesa Civil São Gonçalo	São Gonçalo	6.9	15/07/2019	31/12/2021
Boa Vista		São Gonçalo	1.8	01/04/2019	31/12/2021
Engenho Pequeno II		São Gonçalo	2.5	01/04/2019	31/12/2021
Engenho Pequeno IV		São Gonçalo	3.3	24/05/2019	31/12/2021
Gradim I		São Gonçalo	1.9	01/04/2019	31/12/2021
Novo México		São Gonçalo	5.0	23/05/2019	31/12/2021
Sete Pontes I		São Gonçalo	2.0	01/04/2019	31/12/2021
Tenente Jardim		São Gonçalo	3.6	07/06/2019	31/12/2021
Colubandê	INEA	São Gonçalo	10.2	22/05/2012	04/04/2017
Colubandê-Alcântara		São Gonçalo	6.9	04/04/2018	16/02/2021
Ipiíba		São Gonçalo	11.0	01/01/2013	31/10/2021
Largo da Ideia		São Gonçalo	16.4	10/01/2013	31/12/2021

Fonte: Elaborado pelos autores (2024).

O monitoramento da estação experimental UERJ-FFP é diário e os dados vêm sendo analisados e publicados em diversos trabalhos (Bertolino *et al.*, 2007; Silva, 2007; Pereira, 2009; Delazeri, 2015; Oliveira, 2015; Bertolino *et al.*, 2016; Tórnio, 2019; Pereira *et al.*,

2021). Já as estações do INEA e da Defesa Civil apresentam dados com intervalo de 15 minutos; e os dados do CEMADEN, horários. Deste modo, todos os dados foram uniformizados e tratados como o total acumulado em 24 horas (dados diários) até as 12 TMG compondo, assim, a base de dados deste estudo.

Os dados brutos foram tratados e analisados comparativamente em relação aos períodos de existência e ao percentual de falha da série histórica de cada estação. A escolha do intervalo a ser trabalhado foi condicionada à presença de dados vizinhos. Observa-se, a partir da Tabela 1, que muitas das estações disponíveis iniciaram seu monitoramento no ano de 2019, o que viabiliza somente o preenchimento de dados diários e em um período de avaliação relativamente curto, uma vez que a estação UERJ-FFP possui dados somente até março de 2020. Deste modo, o intervalo escolhido para avaliação dos métodos foi de 01/04/2019 a 16/03/2020, totalizando 351 observações diárias contínuas. As estações vizinhas com falhas nesse período foram excluídas do estudo. Na Tabela 2, sumarizam-se as informações das estações selecionadas e na Figura 2, é possível observar a espacialização de todas as estações disponíveis e as estações efetivamente utilizadas no estudo.

Tabela 2 - Estações selecionadas para preenchimento de falhas da estação UERJ-FFP.

Nome	Sigla	Instituto	Município	Distância FFP (km)	Início da série	Fim da série
Engenho Pequeno II	ENG	Defesa Civil São Gonçalo	São Gonçalo	2.5	01/04/2019	31/12/2021
Gradim I	GRA		São Gonçalo	1.9	01/04/2019	31/12/2021
Sete Pontes I	SET		São Gonçalo	2.0	01/04/2019	31/12/2021
Ipiíba	IPB	INEA	São Gonçalo	11.0	01/01/2013	31/10/2021
Colubandê - Alcântara	COL		São Gonçalo	6.9	04/04/2018	16/02/2021

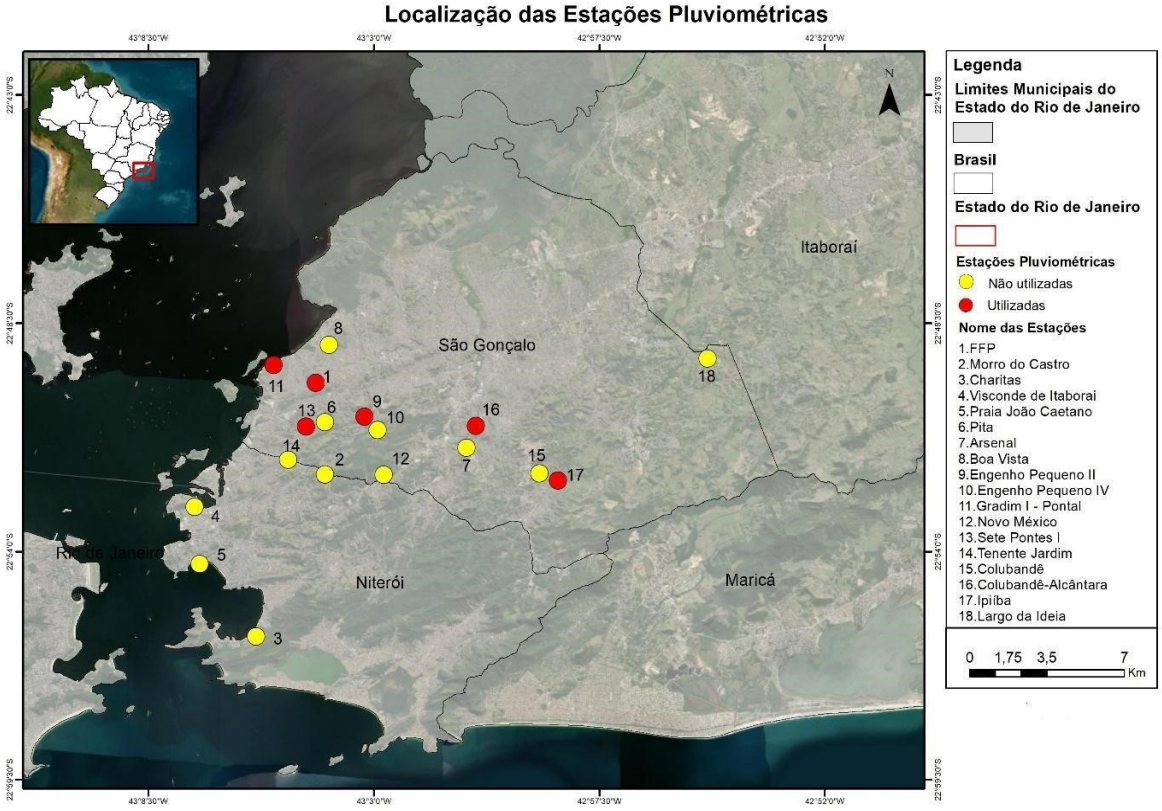
Fonte: Elaborado pelos autores (2024).

Com os dados padronizados, foi realizada uma análise exploratória da base de dados observados, de modo a analisar a homogeneidade da precipitação na região de estudo. Posteriormente, foram realizadas as aplicações dos métodos individuais e combinados de preenchimento visando a estimar os dados da estação UERJ-FFP no intervalo que

compreende a série completa (351 dias). Para a análise da acurácia dos métodos, os dados estimados foram confrontados com os dados observados através de métricas de erro.

Além disso, buscou-se analisar a precisão dos métodos individuais quando a série histórica fosse parcialmente incompleta. Para isto, foram consideradas duas condições de falha: quando a série tivesse 10% de dados faltantes (35 dias seguidos sem registro) e quando tivesse 25% de ausências (87 dias seguidos). Para cada um dos períodos de falha (10% e 25%), foram escolhidos aleatoriamente 40 intervalos ininterruptos dentro da série observada para a estação UERJ-FFP. A avaliação de cada uma das técnicas de preenchimento foi realizada por meio das médias das métricas de erro para cada um dos métodos, calculada com base na estimativa da precipitação *versus* valor observado para cada um dos 40 períodos aleatoriamente escolhidos. A base de dados foi consolidada em planilha eletrônica e as rotinas necessárias para o desenvolvimento do projeto foram implementadas em linguagem R.

Figura 2 - Localização das estações pluviométricas utilizadas no estudo.



Fonte: Elaborado pelos autores (2024).

2.3. Métodos individuais para preenchimento de falhas

Na elaboração deste estudo, utilizaram-se os seguintes métodos espaciais de preenchimento de falhas: Método da Ponderação pelo Inverso da Distância (IDW), Método da Razão Normal (RN) e o Método da Regressão Linear Múltipla (RLM). Estes métodos consistem em estimar dados faltantes através da combinação linear de observações de estações meteorológicas vizinhas, diferindo na forma como os pesos da combinação linear são gerados.

2.3.1. Método da ponderação pelo inverso da distância (IDW)

O método da ponderação pelo inverso da distância é um estimador determinístico onde o valor estimado depende inversamente da distância euclidiana entre o posto a ser estimado e suas estações vizinhas. Deste modo, quanto mais próxima a estação, maior o peso de sua precipitação observada na estimativa. De modo complementar, quanto maior a distância entre as estações, menor o seu fator de ponderação na estimativa. Por depender somente da distância entre as estações, este método é considerado de fácil aplicação (Ruezzene *et al.*, 2021). No entanto, uma limitação deste método está na qualidade da interpolação pode diminuir se a distribuição de pontos amostrais é irregular. O valor estimado $\hat{p}$ é descrito pelas equações a seguir:

$$\hat{p} = \frac{\sum_{i=1}^{n}(w_i \cdot x_i)}{\sum_{i=1}^{n} w_i}$$

$$w_i = \frac{1}{d_i^k}$$

onde n é a quantidade total de postos vizinhos, w_i é o fator de ponderação do modelo para cada observação dos postos vizinhos *i*; x_i é a precipitação registrada no posto vizinho *i* e d_i é a distância entre cada posto vizinho i para o posto com falhas. Neste trabalho, foi utilizado o expoente k=2, uma vez que é o expoente mais recorrente na literatura.

2.3.2. Método da razão normal (RN)

Segundo Young (1992), este método trata-se de uma ponderação de dados registrados em estações vizinhas onde o fator de ponderação neste caso está associado à

correlação entre os dados registrados no posto a ser preenchido e os postos vizinhos. Neste caso, registros de postos com maior correlação linear teriam maior peso na estimativa. O método é expresso pela equação:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i \cdot w_i}{\sum_{i=1}^{n} w_i}$$

onde, $\hat{p}$ é o valor de precipitação a ser estimado, n é a quantidade total de postos vizinhos, w_i é o fator de ponderação do modelo para cada observação dos postos vizinhos *i*; x_i é a precipitação do posto vizinho *i*. O peso w_i, neste modelo, é descrito por:

$$w_i = r_i^2 \left(\frac{n_i - 2}{1 - r_i^2} \right)$$

onde, r_i é a correlação linear entre o posto a ser preenchido e o posto vizinho *i* e n_i é o número de dias em que ocorreu a sobreposição de dados entre o posto a ser preenchido e o posto vizinho *i*. No presente estudo, n_i será o próprio tamanho da série envolvida no cálculo do coeficiente de correlação. A principal limitação deste método está no fato de que a metodologia não analisa a distribuição espacial dos postos vizinhos mais bem correlacionados. Deste modo, discrepâncias espaciais podem ocorrer, eventualmente, em razão de uma coincidente alta correlação entre os dados (Young, 1992).

2.3.3. Método da Regressão Linear Múltipla (RLM)

O preenchimento de falhas utilizando regressão linear múltipla foi proposto por Tabony (1983), onde foram adotados, como variáveis explicativas, os dados de postos vizinhos próximos do posto com falhas. No RLM, o peso associado a cada observação pode ser determinado utilizando correlações parciais entre estações, no qual se elimina o efeito das estações vizinhas. De modo específico, a regressão linear múltipla é um modelo de ajuste no qual assumimos que existe uma relação linear entre uma variável $\hat{p}$ (variável dependente) e n variáveis independentes (explicativas) x_1, x_2, ..., x_n. Este método é definido pela equação a seguir:

$$\hat{p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon_0$$

onde, $\hat{p}$ representa valor a ser estimado; x_1, x_2,...,x_n são o valor das observações das n variáveis; β_0, β_1, β_2,..., β_n são coeficientes de regressão e ε_0 indica o erro residual. Os erros

residuais devem ser normalmente distribuídos e os coeficientes de regressão são estimados numericamente pelo método dos mínimos quadrados.

Apesar de este método ser amplamente utilizado no preenchimento de falhas, seu uso pressupõe certas hipóteses como uma correlação entre o posto teste e os postos vizinhos estatisticamente significativa com coeficiente de correlação r normalmente distribuído. Para isso, r é convertido usando a transformação Z de Fisher. Além disso, os postos vizinhos devem ser classificados e sua importância deve ser pesada em relação ao seu valor de Z de Fisher. Uma desvantagem dos métodos lineares está na própria presunção de relações lineares entre as variáveis, o que pode produzir resultados considerados grosseiramente imprecisos (Murara, 2019).

2.4. Combinação de modelos

Neste estudo, usaremos dois métodos para combinar previsões: a mais comum que se utiliza da média simples (CMS) e a combinação ótima de variância mínima (CVM). A média simples entre os modelos-base é trivial e não será descrita aqui. Já o método CVM será explicitado a seguir conforme descrito por Timmermann (2006).

Por simplicidade, sejam P_{1t} e P_{2t} duas previsões para a mesma variável de interesse Y_t no tempo t e F_t uma combinação das previsões, para $0 \leq w \leq 1$ peso, representada por:

$$F_t = (1 - w)P_{1t} + wP_{2t}$$

Assumindo que os erros ε_{it} das *i* previsões individuais são não viesados, o erro da previsão combinada é dado por:

$$\varepsilon_t = (1 - w)\varepsilon_{1t} + w\varepsilon_{2t}$$

O termo $\varepsilon_t = Y_t - F_t$ é também não viesado e possui a seguinte variância:

$$V(\varepsilon_t) = (1 - w)^2\sigma_1^2 + w^2\sigma_2^2 + 2w(1 - w)\rho\sigma_1\sigma_2$$

onde ρ é o valor da correlação linear entre os erros das previsões obtidas em P_{1t} e P_{2t} e σ^2_1 e σ^2_2 correspondem à variância dos erros de previsão de P_{1t} e P_{2t}, respectivamente. Derivando a expressão de modo a encontrar o valor de *w* que minimiza a variância de ε_t, obtemos a expressão:

(1)

$$\hat{w} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Assumindo que os erros de previsão obtidos em P_{1t} e P_{2t} são independentes, tem-se que a correlação linear (ρ) entre esses erros é nula. Considerando-se a independência desses erros, é possível reescrever a Equação 1, formulação conhecida como método da variância mínima simplificado.

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

onde σ^2_1 e σ^2_2 correspondem à variância dos erros de previsão de P_{1t} e P_{2t}, respectivamente.

2.5. Avaliação de desempenho dos métodos

Para avaliar os métodos de interpolação dos dados de precipitação, as séries diárias $\hat{p}$ estimadas por cada uma das técnicas foram comparadas com a série temporal observada p de modo a identificar o modelo mais ajustado. Para isto, foram aplicadas as seguintes métricas estatísticas: coeficiente de determinação (R^2) que visa a verificar o ajuste linear entre os dados estimados e os observados; erro absoluto médio (MAE); erro quadrático médio (RMSE) e coeficiente de eficiência de Nash-Sutcliffe (*Nash*). Considere que as expressões para as métricas são desenvolvidas para uma série temporal $\hat{p} = (\widehat{y_1}, \cdots, \widehat{y_n})$ de dados estimados para o posto teste com n dias de falhas.

O coeficiente de determinação (R^2) calcula a relação linear entre duas variáveis avaliando, assim, a qualidade do ajuste de um modelo $\hat{p}$ indicando o quanto ele foi capaz de explicar os dados de referência p. Quanto maior o valor encontrado, melhor o ajuste da série estimada à série observada o que é expresso pela equação a seguir:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \qquad onde\ 0 \leq R^2 \leq 1$$

onde y_i é o valor observado e $\hat{y}_i$ é o valor estimado na i-ésima observação, n é o número de observações e $\bar{y}$ é a média das observações.

O erro absoluto médio (MAE) consiste na média da diferença absoluta dos valores estimados e dos valores observados. É uma métrica robusta e de simples implementação, onde pressupõe-se que quanto mais próximos de zero ou igual a zero seu valor esteja, mais

precisos sejam os valores estimados. No caso de dados de precipitação, seu valor será expresso em milímetros e sua equação é dada por:

$$MAE = \frac{\sum_{I=1}^{n} |\hat{y}_i - y_i|}{n}$$

onde y_i é o valor observado, $\hat{y}_i$ é o valor estimado na i-ésima observação e n é o número total de observações.

A raiz do erro médio quadrático (RMSE) fornece a raiz da média de diferença quadrática entre os valores previstos e os valores observados. Quanto mais próximo de zero, melhores os valores estimados. Seu valor será expresso no trabalho em milímetros e sua expressão de cálculo é dada por:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$$

onde y_i é o valor observado, $\hat{y}_i$ é o valor estimado na i-ésima observação e n é o número total de observações.

Por fim, o coeficiente de Nash-Sutcliffe (*Nash*) é utilizado em hidrologia para avaliar o desempenho de modelos hidrológicos e é expresso por:

$$Nash = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \qquad onde -\infty < Nash \ll 1$$

onde, *Nash é o* coeficiente Nash-Sutcliffe (adimensional), y_i é o valor observado, $\hat{y}_i$ é o valor estimado, $\bar{y}$ é a média das observações e n é o número de observações.

O coeficiente de *Nash* varia de menos infinito a 1, onde o valor de 1 indica um ajuste perfeito entre o modelo e os dados observados. Um valor próximo a 0 indica que o modelo é similar a usar a média dos valores observados, e valores negativos indicam que o modelo possui um desempenho pior do que usar a média dos valores observados. De acordo com Silva *et al.* (2008), ao encontrar valores de coeficiente maiores que 0.75, o modelo é considerado adequado; entre 0.36 e 0.75, considerado aceitável, e valores inferiores a 0.36, o modelo é considerado inaceitável.

3. RESULTADOS

3.1. Análise exploratória dos dados

As medidas descritivas de posição e dispersão das séries utilizadas no presente trabalho são apresentadas na Tabela 3. O desvio-padrão e a forte assimetria positiva das distribuições são justificados pela natureza do dado diário de precipitação, com muito mais dias sem registro de chuva do que com chuva em um período de quase um ano. De fato, em mais da metade dos dias não houve registro de chuva, como indica a mediana de todas as séries. Esses resultados sugerem uma certa homogeneidade regional nos dados entre as estações.

Tabela 3 - Medidas descritivas das séries diárias de precipitação das estações nas estações UERJ-FFP, Engenho Pequeno – ENG, Gradim – GRA, Sete Pontes – SET, Ipiíba – IPB e Colubandê/Alcântara - COL (Série Completa - de 01/04/2019 a 16/03/2020).

Medida Descritiva	ENG	GRA	SET	IPB	COL	FFP
Número de Observações	351	351	351	351	351	351
Dias de Chuva	110	121	120	115	140	114
Mínimo (mm)	0	0	0	0	0	0
1º Quartil (mm)	0	0	0	0	0	0
Mediana (mm)	0	0	0	0	0	0
Média (mm)	2.15	4.05	3.93	2.73	3.75	3.28
3º Quartil (mm)	0.6	1.6	1.6	0.4	1.2	1.0
Máximo (mm)	83.2	119.4	73.6	90.4	125.2	87.4
Desvio-padrão (mm)	7.72	11.67	8.90	9.46	12.09	10.07
Curtose	58.19	39.16	27.72	43.25	45.23	28.48
Assimetria	6.88	5.36	4.78	5.98	5.99	4.89

Fonte: Elaborado pelos autores (2024).

De modo a avaliar a homogeneidade dos dados, a correlação existente entre as observações de estações vizinhas é um outro fator importante a se considerar. A Figura 3 mostra matriz de correlações das séries temporais nas estações estudadas. Os dados de precipitação nas estações vizinhas foram comparados em pares. Os valores do coeficiente R de correlação linear de Pearson variam entre 0.79 e 0.96 sugerindo uma alta correlação linear entre as séries duas as duas.

Diante dos resultados obtidos, infere-se que todas as estações analisadas no presente estudo estão localizadas em uma região homogênea em termos de precipitação e sob as mesmas condições climáticas. As precipitações registradas nas estações mostram padrões semelhantes, o que viabiliza o preenchimento de falhas.

Figura 3 – Matriz de correlação para as séries diárias de precipitação nas estações UERJ-FFP, Engenho Pequeno – ENG, Gradim – GRA, Sete Pontes – SET, Ipiíba – IPB e Colubandê/Alcântara - COL (de 01/04/2019 a 16/03/2020).



Fonte: Elaborado pelos autores (2024).

3.2. Simulação e preenchimento de falhas para série completa

Aplicou-se o teste de normalidade de Shapiro-Wilk para todos os dados observados da estação UERJ-FFP e para os dados estimados pelos modelos. Foi obtido p-valor menor que o nível de significância de 0.05, indicando que os dados não seguem uma distribuição normal. A não-normalidade era esperada diante da natureza diária do dado de precipitação neste estudo.

Para a aplicação do método da ponderação pelo inverso da distância (IDW), considerou-se no modelo o peso w_i de cada posto vizinho x_i como o inverso do quadrado da distância ele e o posto a ser preenchido. Na Tabela 4, apresentam-se as distâncias e seus respectivos pesos para cada uma das cinco estações vizinhas deste estudo.

Tabela 4 – Pesos utilizados no método IDW para as estações de Engenho Pequeno – ENG, Gradim – GRA, Sete Pontes – SET, Ipiíba – IPB e Colubandê/Alcântara – COL.

Estação	Distância (km)	Peso
ENG	2.54	0.155418
GRA	1.93	0.269394
SET	1.98	0.254423
IPB	11.00	0.008264
COL	6.95	0.020678

Fonte: Elaborado pelos autores (2024).

Visando à utilização do método da Regressão Linear Múltipla (RLM), aplicou-se a transformada Z de Fisher aos coeficientes de correlação R obtidos para os postos vizinhos de modo a analisar a sua normalidade pelo teste de Shapiro-Wilk. O valor p obtido foi de 0.872 indicando que seguem a distribuição normal, o que satisfaz a hipótese para o uso do método da RLM como método de ajuste.

Na Tabela 5, é possível observar o desempenho de cada um dos métodos de preenchimento em estudo para a estação UERJ-FFP no preenchimento da série completa com 351 dados faltantes. Dentre os métodos individuais, o RLM obteve o melhor desempenho na maior parte das métricas de validação, exceto o coeficiente de *Nash*. Todos os métodos apresentaram coeficiente de determinação e de *Nash* maiores que 0.8, indicando assim que tanto o método IDW como o método da Razão Normal (RN) também são boas opções para séries com longos períodos de falhas (aproximadamente um ano). Ao analisar a verificação de desempenho pelo MAE e RMSE, também foram encontrados erros pequenos, um pouco abaixo de 3mm para todos os métodos de preenchimento.

De modo a analisar o desempenho das combinações de previsões, foram analisadas a combinação das três estimativas por média simples (CMS) e as estimativas duas a duas pela combinação ótima da variância mínima (CVM). Neste segundo método de combinação, foram obtidas três estimativas: CVM1, que analisa a combinação do modelo IDW com o modelo RM; CVM2, onde é analisada a combinação do modelo IDW com o modelo RN; e CVM3 que expressa a combinação do modelo RM com o modelo RN. Na Tabela 5, é possível observar o desempenho de todos os modelos combinados deste estudo para a série de 351 dias de falhas. Nela, fica expressa a superioridade das estimativas por combinação uma vez que todos os modelos obtiveram R^2 e *Nash* maior que 0.9 e RMSE inferior a 1mm.

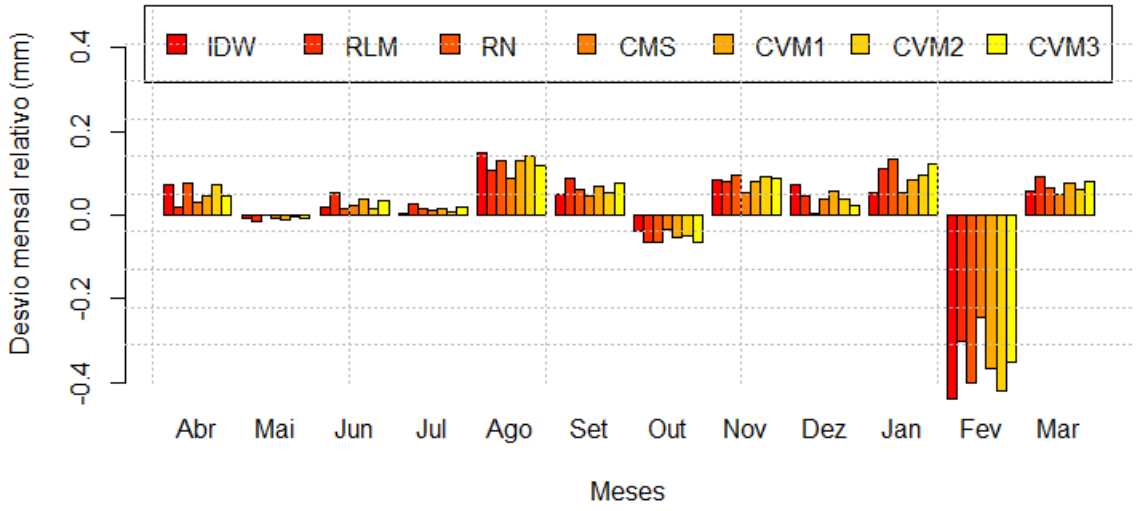
Tabela 5 - Desempenho dos métodos em estudo através do coeficiente de determinação (R^2), erro absoluto médio (MAE), erro quadrático médio (RMSE) e coeficiente de eficiência de Nash-Sutcliffe (*Nash*) para a estação UERJ-FFP (Série Completa - 351 dias de falha).

Método	R^2	Nash	RMSE (mm)	MAE (mm)
IDW	0.9252	0.8489	2.73	0.84
RLM	0.9366	0.8351	2.52	0.83
RN	0.9253	0.9000	2.73	0.90
CMS	0.9709	0.9682	0.74	0.55
CVM1: IDW-RM	0.9338	0.9201	0.91	0.82
CVM2: IDW-RN	0.9266	0.9201	0.93	0.87
CVM3: RM-RN	0.9336	0.9282	0.92	0.86

Fonte: Elaborado pelos autores (2024).

No entanto, uma preocupação é o acúmulo da pequena variação diária obtida pelos métodos de preenchimento nos consolidados mensais. De modo a verificar o impacto desta variação nos acumulados mensais, analisou-se a diferença entre o acumulado mensal observado na estação FFP e cada acumulado mensal previsto pelos métodos em estudo. A Figura 4 mostra graficamente os desvios percentuais entre cada método de preenchimento e o observado, para cada mês, dentro do período de análise de 351 dias. No gráfico, é possível observar que, na maior parte dos meses, a variação foi positiva para os sete métodos. De fato, dentre os métodos individuais, o desvio médio foi de 0.6% (1.1 mm) para o método IDW, 2%(3.5mm) para o método RLM e 1.05% (1.9mm) para o método RN.

Figura 4 - Gráfico de variação percentual entre o observado na estação FFP e o previsto pelos métodos IDW, RLM, RN e pelos métodos de combinação CMS (Média Simples), CVM1 (variância mínima entre os métodos IDW e RLM), CVM2 (variância mínima entre os métodos IDW e RN) e CVM3 (variância mínima entre os métodos RLM e RN) para o acumulado mensal entre 01/04/2019 e 16/03/2020.



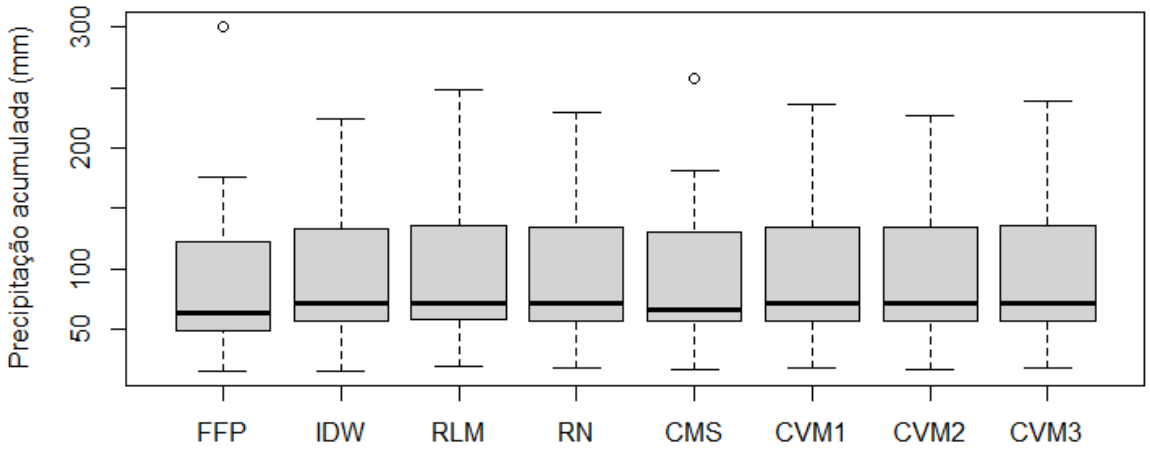
Fonte: Elaborado pelos autores (2024).

Ainda na Figura 4, observa-se que a variação percentual dos acumulados mensais obtidos pelas estimativas dos métodos combinados em relação ao observado na estação FFP foi menor do que a variação dos métodos individuais em todos os meses. O método da combinação por média simples (CMS) e a combinação por variância mínima entre o IDW e RN (CVM2) apresentaram um desvio médio no período inferior a 1% (0.87% e 0.84% de variação, respectivamente). Dentre os métodos combinados, o que apresentou maior variação foi a combinação por variância mínima entre os métodos RLM e RN (CVM3) com desvio médio de 1.57% ou 2.7mm. Os meses de maior variabilidade nos acumulados mensais, para todas as sete técnicas de preenchimento, foram os meses de agosto, onde os métodos superestimaram a pluviosidade, e em fevereiro, onde houve expressiva subestimação da variável.

Apesar da alta variabilidade para o mês de fevereiro de 2020, o diagrama boxplot comparando os acumulados mensais observado na UERJ-FFP e o previsto para a mesma estação pelos métodos IDW, RLM e RNA além das combinações CMS, CVM1, CVM2 e CVM3 sugerem que o dado mensal não é muito discrepante entre o que foi observado e o que foi previsto, conforme observado na Figura 5. O mesmo diagrama indica que quase todos os métodos de interpolação tendem a superestimar a chuva, com exceção do método

combinado por média simples. Em particular, o método CMS foi o único a identificar o mês de fevereiro como um *outlier* tal qual o dado observado. Os demais apresentaram variabilidade maior do que foi registrado em campo. A partir dos dados que constam no gráfico, pode-se concluir que meses em que ocorreram eventos muito intensos podem ser prejudicados no processo de preenchimento de falhas, dependendo do método escolhido.

Figura 5 – *Boxplot* dos acumulados mensais de precipitação observados na estação UERJ-FFP e acumulados mensais previstos pelos métodos IDW, RLM, Razão Normal, Combinação por Média Simples (CMS) e por variância mínima (CVM1, CVM2, CVM3) para período entre 01/04/2019 e 16/03/2020.



Fonte: Elaborado pelos autores (2024).

De modo a testar se a média mensal dos dados observados na estação FFP e as médias mensais previstas por todos os métodos de preenchimento utilizados, diferiram significativamente para o período em estudo, foi realizado um teste de hipótese para diferença de médias com 5% de significância. Concluiu-se que não há evidências de que a diferença entre as médias mensais seja significativa.

3.3. Simulação e avaliação dos métodos de preenchimento de séries parcialmente completas

Essa seção trata da análise das simulações realizadas quando a série histórica possui falhas parciais. Aqui, trabalhou-se com dois períodos a saber: 10% (série curta) e 25% de falhas (série intermediária). Para cada percentual de falhas, foram selecionados

aleatoriamente 40 trechos da série da estação UERJ-FFP e aplicados os três métodos de preenchimento em cada trecho. Deste modo, ao total, foram realizadas nesta etapa 2 x 40 x 3 = 240 simulações.

Na Tabela 6, apresenta-se o desempenho médio de cada um dos métodos de preenchimento em estudo para a série intermediária (com 25% de dias ausentes). O método RN apresentou o melhor R^2 e *Nash,* ao passo que a regressão linear múltipla apresentou o melhor RMSE e MAE. Neste caso, todos os métodos apresentaram R^2 e *Nash* superiores a 0.85, indicando que todos são boas opções para séries com períodos de falhas de alguns meses.

Tabela 6 - Desempenho médio dos métodos em estudo através do coeficiente de determinação (R^2), erro absoluto médio (MAE), erro quadrático médio (RMSE) e coeficiente de eficiência de Nash-Sutcliffe (*Nash*) para a estação UERJ-FFP (Série Intermediária – 25% falhas).

Método	R^2	Nash	RMSE (mm)	MAE (mm)
IDW	0.8831	0.8324	0.97	1.03
RLM	0.9213	0.9109	0.82	0.71
RN	0.9868	0.9690	0.96	0.92

Fonte: Elaborado pelos autores (2024).

Consta, da Tabela 7, o desempenho médio de cada um dos métodos de preenchimento em estudo para uma série curta (com 10% de dias ausentes). A regressão linear múltipla apresentou o melhor desempenho em todas as métricas. O método IDW possui coeficiente de *Nash* médio de 0.55, o que leva a classificar o modelo como aceitável. Já o método RN apresentou R^2 = 0.82 e coeficiente de *Nash* de 0.69, o que o tornaria igualmente aceitável. Ao analisar a verificação de desempenho pelo MAE e RMSE, a regressão linear apresentou erros inferiores a 1mm, ao passo que os métodos IDW e RN apresentaram erros superiores a 1 mm.

Tabela 7 - Desempenho médio dos métodos em estudo através do coeficiente de determinação (R^2), erro absoluto médio (MAE), erro quadrático médio (RMSE) e coeficiente de eficiência de Nash-Sutcliffe (*Nash*) para a estação UERJ-FFP (Série Curta – 10% de falha).

Método	R^2	Nash	RMSE (mm)	MAE (mm)
IDW	0.9055	0.5533	0.97	1.26
RLM	0.9775	0.9745	0.55	0.47
RN	0.8296	0.6942	1.20	1.44

Fonte: Elaborado pelos autores (2024).

Os resultados aqui apresentados vão de encontro aos obtidos por Brubacher, Oliveira e Guasselli (2020) para dados diários de precipitação no Rio Grande do Sul. Os autores concluíram que a RLM foi superior a métodos de preenchimento de falhas com base em redes neurais artificiais, apresentando maior R^2 e menor MAE. Os autores justificam o melhor desempenho da RLM por conta da forte correlação linear entre os dados de precipitação em relação a sua vizinhança, o que também é observado para o município de São Gonçalo.

Já Longman *et al.* (2020), ao analisarem o preenchimento de dados diários de precipitação no Havaí utilizando os RLM, RN, IDW e outras técnicas de interpolação espacial, encontraram que o método da Razão Normal obteve o menor erro para 1, 15 e 30 dias seguidos com falhas. Para os autores, a correlação entre a estação a ser preenchida e as estações vizinhas é mais importante do que a proximidade na determinação da qualidade do preenchimento, isso em vista que as estações estavam distribuídas por todo o estado, o que implica grandes distâncias e grande variabilidade climática. A diferença nos resultados do presente trabalho, para períodos curtos, pode ser explicada pelas escalas diferentes utilizadas uma vez que foi realizado em uma região menor e com regime de chuva muito mais coeso.

Martins et al. (2023), ao comparar o desempenho de sete métodos individuais de preenchimento de falhas em dados diários precipitação na região das bacias hidrográficas dos rios Piracicaba, Capivari e Jundiaí no estado de São Paulo (Brasil), concluem que os melhores métodos foram IDW e regressão linear simples para falhas aleatórias na série de dados. Quando analisados em relação às lacunas em até sete dias consecutivos, a regressão

linear simples e a RLM apresentaram boa performance, o que vai de encontro ao obtido aqui onde a RLM apresenta os melhores indicadores para períodos curtos de falhas.

4. CONSIDERAÇÕES FINAIS

A combinação de estimativas mostrou-se muito promissora, com métricas de acurácia melhores do que as dos métodos de estimativa analisados individualmente (IDW, RLM e RN), para a série com 351 dias de falhas. Em particular, o método de combinação por média simples (CMS) foi o de melhor desempenho tanto na estimativa diária quanto no acumulado mensal estimado, com variação média em torno de 1.5mm no acumulado mensal de chuva em relação ao observado em campo na estação UERJ-FFP.

Já na análise de período menores com falhas, dentre os três métodos individuais de preenchimento de falhas, o RLM apresentou os melhores resultados. De maneira geral, para o município fluminense de São Gonçalo, todos os métodos apresentaram bons resultados para o preenchimento da estação UERJ-FFP em séries com falhas em períodos intermediários (87 dias seguidos de falhas) e longos (351 dias de falhas). Já para períodos curtos (35 dias), os métodos IDW e Razão Normal apresentaram resultados com distintas interpretações ao comparar o R^2 com coeficiente de *Nash*.

Analisando o problema de séries com falhas de outra perspectiva, dentre os modelos individuais com melhor desempenho, o RN apresenta metodologia mais simples de ser implementada por operadores, uma vez que possui menos hipóteses a serem satisfeitas, como a RLM. A RN depende da correlação linear entre as estações e seu uso trouxe resultados bem promissores para períodos de falhas intermediário e/ou longo. Já dentre os métodos de previsão combinada, a combinação por média simples foi o método de melhor desempenho dentre todos e possui uma metodologia igualmente fácil, dependendo dos métodos base a serem utilizados. Deste modo, a utilização deste tipo de combinação consistiria em um ferramental para a melhoria de desempenho de métodos individuais de preenchimento.

O município de São Gonçalo carece de dados de monitoramento, assim como de estudos que sirvam de suporte ao planejamento urbano e dos recursos hídricos. O estudo teve como objetivo fazer um levantamento dos dados disponíveis. Além disto, os resultados

apresentados permitiram obter séries históricas de precipitação mais consistentes apresentando um ferramental de suporte que poderá ser utilizado, no futuro, em estudos climatológicos e suas possíveis consequências em setores econômico, social e ambiental do município.

Por fim, os resultados obtidos nesta pesquisa, a partir de um conhecimento mais detalhado de métodos eficazes para preenchimento de falhas de dados de precipitação, reforçam a importância da análise estatística prévia na utilização de modelos de ajuste. Além disso, estes resultados poderão também ser utilizados na caracterização local do clima, contribuindo para avanços de pesquisas dentro desse tema.

REFERÊNCIAS

ATIYA, A. F. Why does forecast combination work so well? **Int. J. Forecast.**, v. 36, n. 1, p. 197-200, nov. 2020. DOI 10.1016/j.ijforecast.2019.03.010. Disponível em: https://doi.org/10.1016/j.ijforecast.2019.03.010. Acesso em: 28 set. 2024.

BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. **J Oper Res Soc**, v. 20, n. 4, p. 451-468, dec. 1969. DOI 10.2307/3008764. Disponível em: https://doi.org/10.2307/3008764. Acesso em: 28 set. 2024.

BERTOLINO, A. V. F. A.; COSTA, A. R. C.; BERTOLINO, L. C.; FIALHO, E. S. Análise da dinâmica climatológica no município de São Gonçalo RJ - triênio 2004 - 2007. **Revista Tamoios**, São Gonçalo, v. 4, p. 1-13, jul./dez. 2007. Disponível em: https://www.e-publicacoes.uerj.br/tamoios/article/view/624. Acesso em: 28 set. 2024.

BERTOLINO, A. V. F. A.; BERTOLINO, L. C.; MERAT, G. S.; LEMES, M. W. Movimentos de massa no município de São Gonçalo. In: SANTOS, M. G. **Biodiversidade e sociedade no Leste Metropolitano do Rio de Janeiro**. Rio de Janeiro: Eduerj, 2016, p. 243-263.

BIELENKI JUNIOR, C.; SANTOS, F. M.; POVINELLI, S. C. S.; MAUAD, F. F. Alternative methodology to gap filling for generation of monthly rainfall series with GIS approach. **RBRH**, Porto Alegre, v. 23, p. e29, 2018. DOI 10.1590/2318-0331.231820170171. Disponível em: https://doi.org/10.1590/2318-0331.231820170171. Acesso em: 30 set. 2024.

BRUBACHER, J. P.; OLIVEIRA, G. G.; GUASSELLI, L. A. Preenchimento de falhas em séries temporais de precipitação diária no Rio Grande do Sul. **Rev. Bras. Meteorol.**, São Paulo, v. 35, p. 335-344, 2020. DOI 10.1590/0102-7786352035. Disponível em: https://doi.org/10.1590/0102-7786352035. Acesso em: 30 set. 2024.

CORREIA, T. P.; DOHLER, R. E.; DAMBROZ, C. S.; BINOTI, D. H. B. Aplicação de redes neurais artificiais no preenchimento de falhas de precipitação mensal na região serrana do Espírito Santo. **Geociências**, Rio Claro, v. 35, n. 4, p. 560-567, 2016. Disponível em:

https://www.periodicos.rc.biblioteca.unesp.br/index.php/geociencias/article/view/12000. Acesso em 30 set. 2024.

DELAZERI, E. M. **Análise pluviométrica e sua relação com as enchentes em São Gonçalo, uma análise espacial por georreferenciamento das ocorrências dos anos de 2006 a 2014**. 2015. Monografia (Graduação em licenciatura em Geografia) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro, São Gonçalo-RJ, 2015.

DEPINÉ, H.; CASTRO, N. M. R.; PINHEIRO, A.; PEDROLLO, O. Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais**. RBRH**, Porto Alegre, v. 19, n. 1, p. 51-63, jan./mar. 2014. Disponível em: http://hdl.handle.net/10183/229616. Acesso em: 30 set. 2024.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cidades e Estados**: São Gonçalo-RJ. 2021. Disponível em: https://www.ibge.gov.br/cidades-e-estados/rj/sao-goncalo.html. Acesso em: 25 mar. 2023.

JUNQUEIRA, R.; AMORIM, J. S.; OLIVEIRA, A. S. de. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Sustentare**, Três Corações, v. 2, n. 1, p. 198-210, 2018. DOI 10.5892/st.v2i1.4982. Disponível em: http://dx.doi.org/10.5892/st.v2i1.4982. Acesso em: 30 set. 2024.

KIM, J.; RYU, J. H. A heuristic gap filling method for daily precipitation series. **Water resources management**, v. 30, p. 2275-2294, 2016. DOI 10.1007/s11269-016-1284-z. Disponível em: https://doi.org/10.1007/s11269-016-1284-z. Acesso em: 30 set. 2024.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical analysis with missing data**. New York: John Wiley & Sons, 2019.

LONGMAN, R. J.; NEWMAN, A. J.; GIAMBELLUCA, T. W.; LUCAS, M. Characterizing the uncertainty and assessing the value of gap-filled daily rainfall data in Hawaii. **J Appl Meteorol Climatol.**, v. 59, n. 7, p. 1261-1276, 2020. DOI 10.1175/JAMC-D-20-0007.1. Disponível em: https://doi.org/10.1175/JAMC-D-20-0007.1. Acesso em: 30 set. 2024.

MARTINS, V. L. M.; WERNER, L. Comparação de previsões individuais e suas combinações: um estudo com séries industriais. **Production**, São Paulo, v. 24, n. 3, p. 618–627, jul. 2014. DOI 10.1590/S0103-65132013005000066. Disponível em: https://doi.org/10.1590/S0103-65132013005000066. Acesso em: 30 set. 2024.

MARTINS, L. L.; MARTINS, W. A.; RODRIGUES, I. C. D. A.; XAVIER, A. C.; MORAES, J. F. L. D.; BLAIN, G. C. Gap-filling of daily precipitation and streamflow time series: a method comparison at random and sequential gaps. **Hydrological Sciences Journal**, v. 68, n. 1, p. 148-160, 2023. DOI:10.1080/02626667.2022.2145200 Disponível em: https://doi.org/10.1080/02626667.2022.2145200. Acesso em: 30 set. 2024.

MENDONÇA, M. J. C.; MEDRANO, L. A. T. **Um modelo de combinação de previsões para arrecadação de receita tributária no Brasil**. **Texto para discussão**. Rio de Janeiro: Ipea, 2016. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/6610/1/td_2186.pdf. Acesso em: 30 set. 2024.

MURARA, P. G. Técnicas de interpolação para a pesquisa em climatologia regional e agroclimatologia. **Revista Brasileira de Climatologia**, *[S. l.]*, v. 25, 2019. DOI 10.5380/abclima.v0i0.66602. Disponível em: https://doi.org/10.5380/abclima.v0i0.66602. Acesso em: 30 set. 2024.

NAGHETTINI, M.; PINTO, E. Regional Frequency Analysis of Hydrologic Variables. **Fundamentals of Statistical Hydrology**, p. 441-495, 2017. DOI 10.1007/978-3-319-43561-9_10, Disponível em: https://doi.org/10.1007/978-3-319-43561-9_10. Acesso em: 30 set. 2024.

OLIVEIRA, A. C. B. **Dinâmica e espacialização dos movimentos de massa no município de São Gonçalo - RJ**. 2015. Monografia (Graduação em licenciatura em Geografia) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro. São Gonçalo-RJ, 2015.

PAPACHARALAMPOUS G., TYRALIS, H. Hydrological time series forecasting using simple combinations: Big data testing and investigations on one-year ahead river flow predictability. **J. Hydrol**, v. 590, p. 125205, 2020. DOI 10.1016/j.jhydrol.2020.125205. Disponível em: https://doi.org/10.1016/j.jhydrol.2020.125205. Acesso em: 30 set. 2024.

PEREIRA, G. C. **Correlação dos dados de intensidade de chuva com as enchentes mais expressivas ocorridas no município de São Gonçalo – RJ no período de 2005 a 2008**. 2009. Monografia (Graduação em licenciatura em Geografia) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro. São Gonçalo-RJ, 2009.

PEREIRA, V. C. M.; BERTOLINO, A. V. F. A.; KEDE, M. L. F. M.; CORBO, A. R.; BERTOLINO, L. C.; FIALHO, E. S. Contribuições para a análise da dinâmica climatológica no município de São Gonçalo/RJ 2008-2018. **Revista Tamoios**, São Gonçalo, v. 17, n. 2, p. 111-136, 2021. DOI 10.12957/tamoios.2021.58214. Disponível em: https://doi.org/10.12957/tamoios.2021.58214. Acesso em: 30 set. 2024.

POLETTI, M. L.; SILVESTRO, F.; DAVOLIO, S.; PIGNONE, F.; REBORA, N. Using nowcasting technique and data assimilation in a meteorological model to improve very short range hydrological forecasts, **Hydrol. Earth Syst. Sci**., v. 23, p. 3823–3841, 2019. DOI 10.5194/hess-23-3823-2019. Disponível em: https://doi.org/10.5194/hess-23-3823-2019. Acesso em: 30 set. 2024.

QUEIROZ, W. P.; NUNES, A. C. Proposta de ações do programa VIGIDESASTRES, os sistemas de informação e vigilância no município de São Gonçalo-RJ, Brasil. **Pesquisa & Educação a Distância**, América do Norte, v. 2, n. 28, 2022. Disponível em: http://revista.universo.edu.br/index.php?journal=2013EAD1&page=article&op=view&path%5B%5D=9739&path%5B%5D=5196. Acesso em: 30 Set. 2024.

RUEZZENE, C. B.; MIRANDA, R. B.; TECH, A. R. B.; MAUAD, F. F. Revisão bibliográfica dos principais métodos de preenchimento de falhas em dados de precipitação. **Brazilian Journal of Business**, Curitiba, v. 2, n. 3, p. 2057-2071, jul./set. 2020. DOI 10.34140/bjbv2n3-015. Disponível em: https://doi.org/10.34140/bjbv2n3-015. Acesso em: 30 Set. 2024.

RUEZZENE, C. B.; MIRANDA, R. B.; TECH, A. R. B.; MAUAD, F. F. Preenchimento de falhas em dados de precipitação através de métodos tradicionais e por inteligência artificial. **Revista Brasileira de Climatologia**, *[S. l.]*, v. 29, p. 524–544, 2022. Disponível em: https://ojs.ufgd.edu.br/rbclima/article/view/15563. Acesso em: 30 set. 2024.

SILVA, J. C. **Relação entre dados pluviométricos e as ocorrências de movimentos de massa no município de São Gonçalo - RJ nos anos de 2005 e 2006**. 2007. Monografia (Graduação em licenciatura em Geografia) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro. São Gonçalo-RJ, 2007.

SILVA, K. R.; LAMEGO, C. R. S.; SANTOS, M. C. F. dos. Percepções de alunos do ensino médio sobre impactos socioambientais causados pela chuva no município de São Gonçalo, RJ. *In*: SIMPÓSIO DE GESTÃO AMBIENTAL E BIODIVERSIDADE, 6., 2017, Três Rios. **Anais** [...]. Três Rios: Sigabi, 2017. Disponível em: https://www.itr.ufrrj.br/sigabi/wp-content/uploads/6_sigabi/Sumarizado/JEFERSON_SILVA.pdf. Acesso em: 30 set. 2024.

SILVA, P. M. de O.; MELLO, C. R.; SILVA A. M.; COELHO, G. Modelagem da hidrógrafa de cheia em uma bacia hidrográfica da região Alto Rio Grande. **Rev. bras. eng. agríc. ambient.**, Campina Grande, v. 12, n. 3, p. 258-265, mai. 2008. DOI 10.1590/S1415-43662008000300006. Disponível em: https://doi.org/10.1590/S1415-43662008000300006. Acesso em: 30 set. 2024.

TABONY, R. C. The estimation of missing climatological data. **Journal of Climatology**, v. 3, n. 3, p. 297-314, 1983. Disponível: https://doi.org/10.1002/joc.3370030308. Acesso em: 30 set. 2024.

TIMMERMANN, A. Forecast combinations. *In*: ELLIOTT, G.; GRANGER, C.; TIMMERMANN, A. (Ed.). **Handbook of economic forecasting**. 1st ed. v. 1. p. 135-196. Amsterdam: Elsevier, 2006. DOI 10.1016/S1574-0706(05)01004-9. Disponível em: https://doi.org/10.1016/S1574-0706(05)01004-9. Acesso em: 30 set. 2024.

TÓRNIO, C. A. A. **Dinâmicas atmosféricas e a gênese das chuvas no município de São Gonçalo (RJ) no período de 2005 a 2016**. 2019. Monografia (Graduação em licenciatura em Geografia) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro, São Gonçalo-RJ, 2019.

TÓRNIO, C. A. A.; KEDE, M. L. F. M. Os impactos das chuvas nos municípios de São Gonçalo (RJ) e Niterói (RJ) no decênio 2010-2019. *In*: SIMPÓSIO BRASILEIRO DE CLIMATOLOGIA GEOGRÁFICA, 14., 2021, João Pessoa. **Anais** [...]. João Pessoa: ABClima, 2021. Disponível em: https://www.xivsbcg.com/anais. Acesso em: 30 set. 2024.

TUCCI, C. E.M. (Org). **Hidrologia:** Ciência e Aplicação. 2° Ed. Porto Alegre: ED. UFRGS: ABRH, 2001. 943p.

YOUNG, K. C. Three-way model for interpolating for monthly precipitation values**. Mon. Weather Rev.**, v. 120, n. 11, p. 2561-2569, 1992. DOI 10.1175/1520-0493(1992)120<2561:ATWMFI>2.0.CO;2. Disponível em: https://doi.org/10.1175/1520-0493(1992)120<2561:ATWMFI>2.0.CO;2. Acesso em: 30 set. 2024.