Revista Brasileira de Climatologia
Brazilian Journal of Climatology

OPEN ACCESS

DOI: 10.5380/abclima

PREENCHIMENTO DE FALHAS EM DADOS DE PRECIPITAÇÃO ATRAVÉS DE MÉTODOS TRADIONAIS E POR INTELIGÊNCIA ARTIFICIAL

FILLING FAILURES IN PRECIPITATION DATA THROUGH TRADITIONAL METHODS AND ARTIFICIAL INTELLIGENCE

LLENAR FALLAS EN LOS DATOS DE PRECIPITACIÓN A TRAVÉS DE MÉTODOS TRADICIONALES E INTELIGENCIA ARTIFICIAL

Camila Bermond Ruezzene
Universidade de São Paulo
camila.ruezzene@usp.br

Renato Billia de Miranda
Universidade de São Paulo
rbm1706@gmail.com

Adriano Rogério Bruno Tech
Universidade de São Paulo
adriano.tech@usp.br

Frederico Fábio Mauad
Universidade de São Paulo
mauadffm@sc.usp.br

Resumo: A falta de informações quanto à distribuição de dados pluviométricos torna-se uma limitação frequente para se modelar e compreender o regime pluviométrico e a variabilidade espaço temporal da precipitação. Desta forma, o objetivo deste trabalho foi realizar a comparação entre métodos tradicionais e redes neurais artificiais para o preenchimento de falhas em dados de precipitação no município de São Carlos entre 1979 a 1989. Foram utilizados os dados de precipitação de 9 estações de monitoramento e empregados quatro técnicas de preenchimento de falhas: método razão normal, ponderação distância inversa, regressão múltipla e redes neurais. Para validação e avaliação do desempenho dos métodos foi aplicado o coeficiente de determinação (R^2), erro absoluto médio (MAE), erro quadrático médio (RMSE), coeficiente de Nash-Sutcliffe (Nash), índice de concordância (D), índice de confiança (C) e técnicas não paramétricas através do teste de Mann-Witney e Kruskal-Wallis. Todos

os métodos apresentaram ótimos desempenhos, exceto na estação A7 que pode ser devido a condições ambientais diferentes das demais estações. Os métodos que se sobressaíram foram de regressão múltipla e redes neurais, e esses resultados poderão ser utilizados como suporte para estudos mais detalhados das possíveis alterações do clima e as suas eventuais implicações para os setores econômico, social e ambiental.

Palavras-chave: Coeficiente de Nash-Sutcliffe. Método razão normal. Ponderação distância inversa. Regressão múltipla. Redes neurais.

Abstract: The lack of information regarding the distribution of rainfall data becomes a frequent limitation for modeling and understanding the rainfall regime and the temporal variability of rainfall. Thus, the objective of this work was to compare traditional methods and artificial neural networks to fill gaps in precipitation data in the municipality of São Carlos between 1979 and 1989. Precipitation data from 9 monitoring stations were used and four fault filling techniques were used: normal ratio method, inverse distance weighting, multiple regression and neural networks. To validate and evaluate the performance of the methods, the determination coefficient (R^2), mean absolute error (MAE), mean square error (RMSE), Nash-Sutcliffe coefficient (Nash), agreement index (D), of the index were applied (C) and non-parametric techniques using the Mann-Witney and Kruskal-Wallis test. All methods showed excellent performances, except at station A7. The methods that stood out were multiple regression and neural networks, and these results can be used as support for more detailed studies of possible climate changes and their possible implications for the economic, social and environmental sectors.

Keywords: Nash-Sutcliffe coefficient. Normal ratio method. Inverse distance weighting. Multiple regression. Neural networks.

Resumen: La falta de información sobre la distribución de los datos de lluvia se convierte en una limitación frecuente para modelar y comprender el régimen de lluvia y la variabilidad espacial temporal de la lluvia. Así, el objetivo de este trabajo fue comparar métodos tradicionales y redes neuronales artificiales para llenar vacíos en los datos de precipitación en el municipio de São Carlos entre 1979 y 1989. Se utilizaron datos de precipitación de 9 estaciones de monitoreo y se utilizaron cuatro técnicas de llenado de fallas: método de relación normal, ponderación de distancia inversa, regresión múltiple y redes neuronales. Para validar y evaluar el rendimiento de los métodos, se aplicaron el coeficiente de determinación (R^2), el error absoluto medio (MAE), el error cuadrado medio (RMSE), el coeficiente Nash-Sutcliffe (Nash), el índice de acuerdo (D), del índice (C) y técnicas no paramétricas utilizando la prueba de Mann-Witney y Kruskal-Wallis. Todos los métodos mostraron excelentes desempeños, excepto en la estación A7. Los métodos que se destacaron fueron la regresión múltiple y las redes neuronales, y estos resultados se pueden utilizar como soporte para estudios más detallados de posibles cambios climáticos y sus posibles implicaciones para los sectores económico, social y ambiental.

Palabras-clave: Coeficiente de Nash-Sutcliffe. Método de relación normal. Ponderación de distancia inversa. Regresión múltiple. Redes neuronales.

Submetido em: 27/07/2020
Aceito para publicação em: 12/08/2021
Publicado em: 22/09/2021

INTRODUÇÃO

A falta de informações quanto à distribuição de dados pluviométricos torna-se uma limitação frequente para se modelar e compreender as características pluviométricas e a variabilidade espaço temporal da precipitação, sendo que séries temporais sem falhas e contínuas é uma condição essencial para análises ambientais e de cenários socioeconômicos consistentes e confiáveis. A obtenção desses dados podem ser através de diversos dispositivos, como estações meteorológicas, sensores, satélites, balões e radares (COUTINHO *et al*., 2018).

No entanto, a utilização de dados de estações meteorológicas está suscetível a muitas dificuldades, já que, devido à ausência de observador, a problemas de acesso à estação, a danificações nos equipamentos, é possível haver erros de medição, gerando dados inconsistentes e levando à ocorrência de falhas nas séries históricas (COUTINHO *et al*., 2018; BIER; FERRAZ, 2017; MWALE; ADELOYE; RUSTUM, 2012). Wanderley, Amorim e Carvalho (2014) citam que a utilização de séries que apresentem falhas pode levar a conclusões errôneas, exibindo padrões diferentes dos observados sem a existência de lacunas, comprometendo, assim, as análises da variabilidade temporal e espacial de processos hidrometeorológicos, além de prejudicar na caracterização climática de uma determinada região.

Apesar dessas lacunas serem usuais, há um grande interesse e necessidade de métodos que preencham falhas de dados, pois são técnicas que viabilizam a construção de um novo conjunto de dados por meio de séries pontuais de dados previamente conhecidos, permitindo assim estudos mais detalhados sobre anomalias de chuva, tendências, intensidade e a utilização para possíveis modelos hidrológicos. A inclusão desses dados pluviométricos preenchidos deve traduzir informações mais próximas possíveis do real, considerando a grande importância para o âmbito socioambiental, planejamento urbano, previsão de desastres naturais, geração de energia, agricultura, atividades turísticas e gestão de recursos hídricos como um todo (BIER; FERRAZ, 2017; HUANG; WANG; XUE, 2015; MELLO; KOLHS; OLIVEIRA, 2017; MISHRA, 2013; OLIVEIRA JÚNIOR *et al*., 2014).

Existe uma série de métodos de preenchimento, desde técnicas simples como, a média aritmética entre estações vizinhas, até métodos mais elaborados, como redes neurais. Esses métodos envolvem desde fórmulas estatísticas até algoritmos, a fim de gerar estimativas para

dados faltantes contribuindo significativamente para a geração de dados mais confiáveis, promovendo continuamente o avanço de pesquisas nas áreas afins. (BIER; FERRAZ, 2017).

Dentre os métodos mais utilizados no meio científico aplicado ao preenchimento de falha em dados hidrometeorológicos, destaca-se o método da Razão Normal (BIER; FERRAZ, 2017; KHOSRAVI *et al*., 2015), Ponderação Distância Inversa (BIELENKI JUNIOR *et al*., 2018; BIER; FERRAZ, 2017; HEMA; KANT, 2017), Regressão Múltipla (COUTINHO *et al*., 2018; BIER; FERRAZ, 2017; HEMA; KANT, 2017; KHOSRAVI *et al*., 2015; MILOVANOVIĆ *et al*., 2017; MELLO, KOHLS; OLIVEIRA, 2017; OLIVEIRA *et al*., 2010; TRUJILLO *et al*., 2015) e Redes Neurais (COUTINHO *et al*., 2018; DEPINÉ *et al*., 2014; CORREIA *et al*., 2016; NKUNA; ODIYO, 2011; WANDERLEY; AMORIM; CARVALHO, 2014), sendo demonstrados ótimos desempenhos para os dados estimados.

Khosravi *et al*. (2015) identificaram por meio do erro absoluto médio que o método da razão normal e o de coordenadas geográfica convencional obtiveram ótimas estimativas para dados de precipitações anuais em 24 estações localizadas em diferentes regiões do Irã.

No estudo realizado por Bielenki Junior *et al*. (2018) é proposta uma alternativa de preenchimento de falhas em séries de precipitação mensal por meio dos métodos de interpolação do inverso da distância e polígono de Thiessen, acoplada a uma ferramenta computacional de sistema de informação geográfica em uma bacia hidrográfica do rio das Cinzas, localizada no estado de Paraná. Puderam concluir que a ferramenta alcançou resultados satisfatórios, confirmado por meio das baixas medidas de erros calculadas.

Segundo Junqueira, Amorim e Oliveira (2018), ao realizarem comparações com diferentes metodologias de preenchimento de falhas de dados pluviométricos, constataram que o método de ponderação regional, média aritmética e ponderação regional com base na regressão superestimaram a precipitação para a bacia hidrográfica do Rio das Mortes (MG). Entretanto, os métodos de regressão linear, regressão múltipla e interpolação do inverso da distância subestimaram a precipitação.

Coutinho *et al*. (2018) utilizaram uma ferramenta de preenchimento de falhas por redes neurais em séries meteorológicas em quatro estações no estado do Rio de Janeiro no período de 2002 a 2014, e concluíram que o método de redes neurais é uma ferramenta eficaz para preencher e estimar com segurança variáveis meteorológicas, pois os dados estimados estavam próximos da realidade dos dados reais.

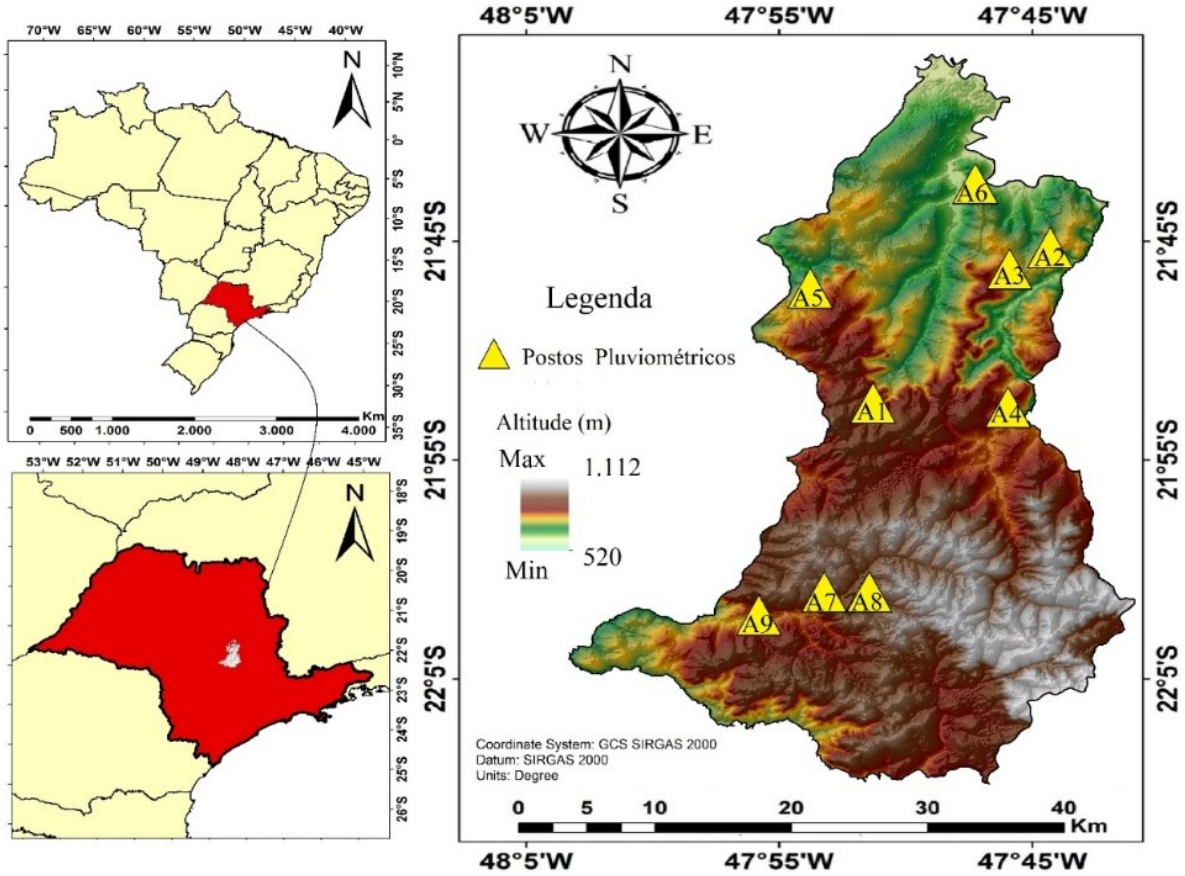
Sendo assim, o objetivo foi realizar a comparação entre métodos tradicionais e de inteligência artificial por meio das redes neurais para o preenchimento de falha em dados de precipitação, no município de São Carlos entre 1979 e 1989.

MATERIAIS E MÉTODOS

Caracterização da área de estudo

A área de estudo para a análise da variação temporal das chuvas foi a porção central de São Paulo onde está localizado o município de São Carlos (figura 1), que apresenta uma população estimada de 251.983 habitantes e área territorial de 1.136,907 km². (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA- IBGE, 2019).

Figura 1 - Localização da área de estudo e as estações meteorológicas utilizadas nesse estudo.



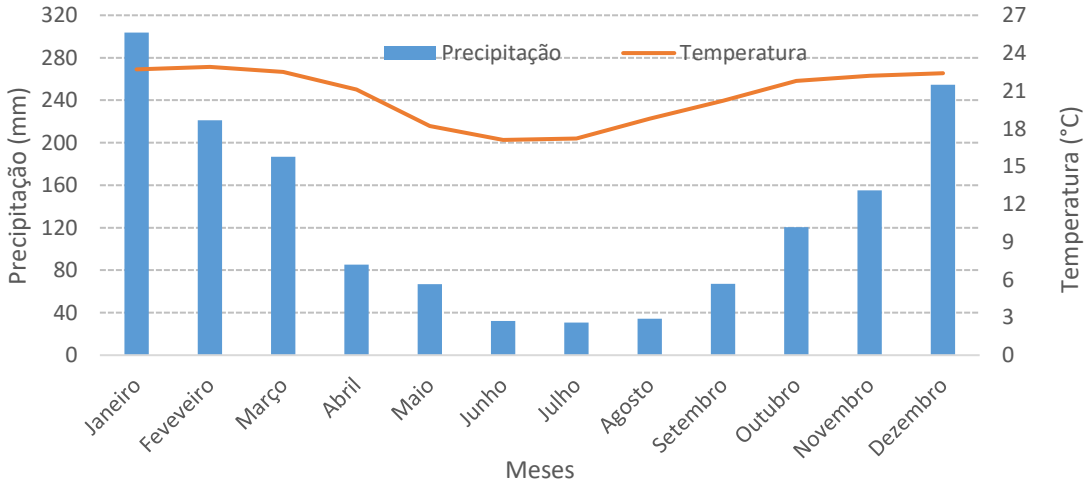
Fonte: Elaborado pelos autores (2020)

O clima predominante da região, de acordo com a classificação de Köppen é Cwa, definido por clima Subtropical de invernos secos e verões chuvosos, controlado por sistemas

tropicais e equatoriais. Essa região central tem uma sazonalidade com estação úmida com 79,69% do total pluviométrico anual entre os meses de outubro a março (figura 2), concentrada principalmente nos meses de dezembro a fevereiro, e a estação seca, correspondente aos meses de abril a setembro, apresentando uma precipitação média de 1493 mm, anualmente (MACHADO; MATTOS, 2001).

Os mecanismos responsáveis pela ocorrência de chuvas nessa região são devidos aos sistemas frontais e às variações vinculadas à Zona de Convergência do Atlântico Sul, aos mecanismos dinâmicos e termodinâmicos, além dos processos convectivos que promovem altos valores acumulados de precipitação dependendo da velocidade com que ocorre o deslocamento dos sistemas frontais (TEIXEIRA; SATYAMURTY, 2011; VÁSQUEZ, *et al.*, 2018).

Figura 2 - Climograma do Município de São Carlos do período de 1991 a 2010.



Fonte: Instituto Nacional de Meteorologia - INMET (2021).

Conforme apresentado no climograma da figura 2 a temperatura média anual é de 21ºC com menores temperaturas para o mês de junho (17,1 ºC) e maiores temperaturas para o mês de fevereiro (22,9 ºC), com variações entre mais amenas e altas, conforme a época do ano (CEPAGRI, 2015; PEEL; FINLAYSON; MCMAHON, 2007; PERUSSI; VECCHIA, 2016). Os solos desta região são basicamente arenosos, e em função dos seus aspectos fisiográficos, é constituído por textura e estrutura variada, tendo como principais Latossolos, Argissolos, Planossolos, Gleissolos e Organossolos. Em direção a Serra Geral, encontra-se maior presença de solos argilosos (OLIVEIRA, 1999; ZANCHETTA, TANNUS; PINHEIRO, 2007).

Dados de precipitação

Foram utilizados dados históricos de precipitação mensal de 1979 a 1989, selecionados a partir de postos pluviométricos no município de São Carlos e controlados pelo Departamento de Águas e Energia Elétrica - DAEE. Esses dados estão disponíveis, em forma de séries temporais, no portal da HIDROWEB da plataforma da Agência Nacional de Águas, sendo utilizados os dados dessa plataforma por já estarem consistidos. Na Tabela 1 são apresentados os postos com seus respectivos códigos, altitude, coordenadas geográficas (latitude e longitude), além do período em anos de cada série analisada.

Tabela 1 - Postos pluviométricos localizados no município de São Carlos.

Nomenclatura adotada	Código	Altitude	Latitude	Longitude	Série
A1	2147031	800	21°53'00''	47°53'00''	1979-1989
A2	2147059	620	21°44'53''	47°46'06''	1979-1989
A3	2147070	620	21°46'00''	47°48'00''	1979-1989
A4	2247002	610	21°53'08''	47°47'09''	1979-1989
A5	2247006	620	21°47'00''	47°57'00''	1979-1989
A6	2247060	900	21°41'48''	47°48'59''	1979-1989
A7	2247109	820	22°01'00''	47°54'00''	1979-1989
A8	2247182	830	22°01'00''	47°53'00''	1979-1989
A9	2247206	660	22°02'00''	47°58'00''	1979-1989

Fonte: Adaptado ANA (2019)

Foi adotado o período de 1979 a 1989 por ser o único período que não havia falhas nos dados para treinamento e assim, possibilitou uma melhor representatividade das características presentes em cada estação e a comparação dos dados reais e estimados por cada método proposto com uma série contínua de dados. A partir do treinamento para esse período e identificando o melhor método de preenchimento será possível preencher falhas para os anos após 1989 quando for necessário.

Análise e sistematização dos dados

Durante as medições podem ocorrer erros e até mesmo falhas na coleta dos dados. Desta forma, para melhorar a confiabilidade dos resultados obtidos, foram realizadas análises minuciosas dos dados e exclusão dos dados inconsistentes, como no caso de valores negativos

ou excedentes que não condissessem com a realidade da região.

Após a identificação das respectivas falhas mensais em cada estação, foram removidos os dados faltantes das demais estações, isto é, se em uma determinada estação *x* não havia dados de um certo mês, estes também foram removidos nas demais estações para que se possa obter uma série homogênea e com o mesmo número de dados para todas as estações de acordo com a metodologia de Coutinho *et al.* (2018), desta forma, considerando todos os dados disponíveis, foram retirados 15,96% dos dados mensais de São Carlos.

Preenchimentos de falhas

A reconstrução de registros de dados incompletos em séries temporais foi assunto de diversos estudos, onde variadas técnicas foram implementadas e comparadas ao redor do mundo, a exemplo de: Regressão Múltipla (JUNQUEIRA, AMORIM; OLIVEIRA, 2018; XIA *et al.,* 1999), Ponderação de Distância Inversa (DE SILVA; DAYAWANSA; RATNASIRI, 2007; TEEGAVARAPU; CHANDRAMOULI, 2005; SIMOLO *et al*., 2009), Método da Razão Normal (KHOSRAVI *et al*. 2015; YOUNG, 1992;) e Redes Neurais Artificiais (CORREIA et al., 2016; COUTINHO *et al*., 2018; DEPINÉ *et al*., 2014; WANDERLEY, AMORIN; CARVALHO, 2014; XIAO *et al*., 2017). Através de um conhecimento aprofundado dos métodos mais eficazes para preenchimento de falhas de dados de precipitação, é possível realizar a reconstrução de séries históricas que poderão ser úteis para caracterização do clima da região, modelos hidrológicos, planejamento urbano, previsão de desastres naturais, geração de energia, agricultura, atividades turísticas dentre outros.

Regressão múltipla (RM)

Considerando que exista uma relação funcional nos valores entre *Y* e *X*, essa função deverá explicar parcela significativa da variação de Y com X (o valor médio de uma das variáveis em função da outra). No entanto, parte da variação permanece não explicada e é atribuída ao acaso (variação residual). Admitindo-se ser uma reta a linha teórica de regressão, a função entre *X* e *Y* é dada conforme a Equação 1.

$$Y = \alpha + \beta X + e \qquad (1)$$

Em que: *Y* é a variável dependente, *X* é a variável independente, α e β são os coeficientes do modelo e "*e*" representa os erros ou resíduos da regressão.

Os dados observados da amostra são estimados por meio dos coeficientes α e β da reta teórica. Aplica-se, então, o método dos mínimos quadrados, um dos procedimentos objetivos mais adequados, segundo o qual a reta a ser adotada deverá ser aquela que torna mínima a soma dos quadrados dos erros ou resíduos da regressão (NAGHETTINI; PINTO, 2007).

Na regressão múltipla as informações pluviométricas do comportamento de uma variável dependente *Y* se dá em função de duas ou mais variáveis independentes X_j, $j = 1, ..., p$. Logo, um possível modelo para avaliar essa relação pode ser dado pela Equação 2.

$$Yi = \beta 0 + \beta 1Xi1 + \beta 2Xi2 + \ldots + \beta pXip + ei\,,\ i = 1, \ldots, n \qquad (2)$$

Em que: *n* é o número de observações, Y_i é a observação da variável dependente para o *i-ésimo* indivíduo, $X_i = (X_{i1}, X_{i2}, ..., X_{ip})$ é um vetor de observações das variáveis independentes para o *i-ésimo* indivíduo, $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_p)$ é um vetor de coeficientes de regressão (parâmetros) e e_i é um componente de erro aleatório. Assume-se que esses erros são independentes e seguem distribuição normal com média zero e variância desconhecida σ^2.

Ponderação de Distância Inversa (PDI)

A técnica de ponderação de distância inversa é aplicada por meio de uma combinação linear das observações dentro de um raio de pesquisa, ocorrendo uma influência decrescente com o aumento da distância. Conforme Hubbard (1994), o método de PDI para o preenchimento de falha é calculado conforme a Equação 3.

$$D_x = \frac{\sum_{i=1}^{n}(D_i/d_i)}{\sum_{i=1}^{n}(1/d_i)} \qquad (3)$$

Em que: *Dx* é o dado mensal em falta a ser preenchido na estação teste, D_i corresponde ao dado da estação vizinha de ordem "i" no mês de ocorrência da falha na estação teste e d_i é a distância entre a estação teste e a estação vizinha de ordem "i".

Neste método o peso dado durante a interpolação refere-se à influência de um ponto amostrado em relação a outro, que diminui conforme aumenta a distância ao ponto que

deseja estimar. As amostras próximas à localização do ponto a ser estimado recebem peso maior que os pontos amostrados de localização mais distantes.

Método da Razão Normal (MRN)

Segundo Young (1992), o método da razão normal é a ponderação de dados com registros em estações vizinhas que pode ser calculada conforme a Equação 4.

$$\boldsymbol{D_x = \frac{\sum_{i=1}^{n} D_i w_i}{\sum_{i=1}^{n} w_i}} \qquad (4)$$

Em que: *Dx* é o dado mensal que necessita ser preenchido na estação teste, *Di* corresponde ao dado da estação vizinha de ordem "i" no mês de ocorrência da falha na estação teste e *wi* ao peso ponderativo atribuído a cada estação vizinha de ordem "i", que é descrito pela Equação 5.

$$\mathbf{w_i} = \boldsymbol{r^2_i \left(\frac{n_i - 2}{1 - r^2_i}\right)} \qquad (5)$$

Em que: $\boldsymbol{r_i}$ é a correlação entre a estação teste e a estação vizinha de ordem "i" e $\boldsymbol{n_i}$ o número de meses em que ocorreu sobreposição de dados entre a estação teste e a estação vizinha de ordem "i". Ou seja, é o tamanho da série de dados envolvida para o cálculo do coeficiente de correlação.

Redes Neurais Artificiais (RNA)

As redes neurais são calculadas por meio de funções matemáticas e são propensas naturalmente a realizar o armazenamento do conhecimento e torná-lo útil, assemelhando-se ao cérebro humano. São calculadas funções não lineares, que podem ser apropriadas para análises complexas, como estimar dados de precipitação pluviométrica (CORREIA *et al*., 2016; COUTINHO *et al*., 2018; DEPINÉ *et al*., 2014; DI PIAZZA *et al*., 2011; WANDERLEY, AMORIM; CARVALHO, 2014;).

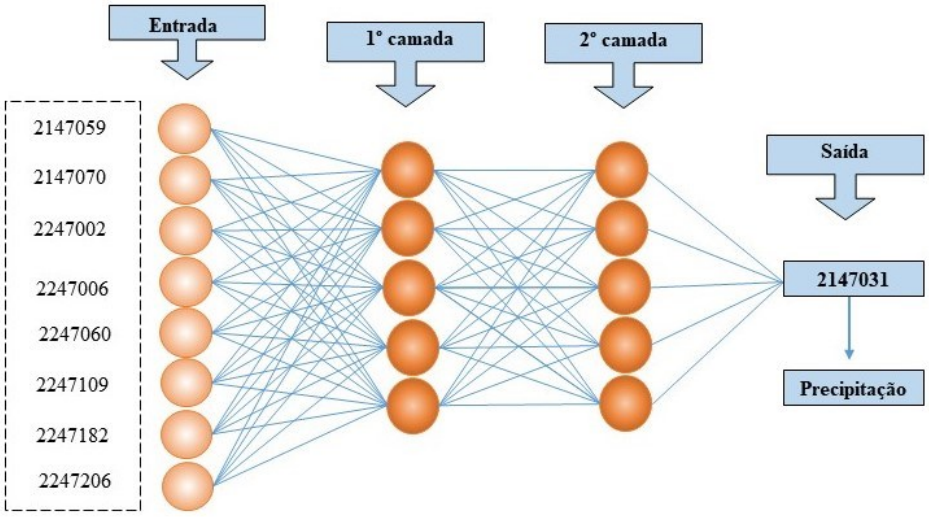
A RNA é uma ferramenta que simula o sistema neurológico do ser humano. Operando em camadas e neurônios, ela recebe do usuário uma ou mais séries de parâmetros que são denominadas como entrada da rede e uma única série de parâmetros denominada alvo da

rede. A rede atribui pesos a cada tipo de dado de entrada para obter uma saída que é, então, comparada com o valor do alvo, ou seja, a saída esperada. Esse processo é repetido diversas vezes, de modo que, a cada iteração, a rede altera os valores dos pesos atribuídos em busca de uma correlação cada vez mais precisa entre os valores de saída e o alvo (HEMA; KANT, 2017; NKUNA; ODIYO, 2011).

Neste estudo foram utilizadas as redes neurais do tipo *Perceptron Multicamada* (PMC) devido a sua maior versatilidade e aplicabilidade nesta temática. Esse tipo de rede tem a grande capacidade de extrair especificidades de dados complexos, identificando padrões e modelos matemáticos complicados para interpretação por computadores. Esse tipo de rede pode ser utilizado para estimar informações e novas condições desejadas, e para obter respostas acuradas das análises em questão (WANDERLEY, AMORIM; CARVALHO, 2014).

Na Figura 3 pode ser observado um exemplo da arquitetura da rede neural correspondente a uma estação. As redes foram treinadas no *software* MATLAB, versão R2015a da empresa *MathWorks*. Além disso, o tipo de rede empregada foi a *Feed-forward backpropagation* com a função de performance sendo o erro quadrático médio. Para essa pesquisa, a arquitetura da rede foi definida com 8 entradas correspondentes às estações localizadas no município, 2 camadas, 10 neurônios, 1 saída e a função de ativação *tan-sigmoid*. Para o treinamento e validação da rede, utilizou-se 70% dos dados para treinamento, 15% para teste e 15% para validação, conforme estabelecido pelo *software* Matlab.

Figura 3 - Configuração da rede neural para o município de São Carlos de 1979 a 1989.



Fonte: Elaborado pelos autores (2020)

Validação e avaliação do desempenho dos métodos

Para avaliar o desempenho e os erros dos métodos de preenchimento de falhas em dados de precipitação foram aplicados testes estatísticos como, coeficiente de determinação (R^2) para verificação da relação entre os dados estimados e medidos; erro absoluto médio (*MAE)*, erro quadrático médio (*RMSE*), índice de concordância (D) e índice de confiança (C), e coeficiente de eficiência de Nash-Sutcliffe (Nash), sendo utilizados por diversos autores (BIER; FERRAZ, 2017; COUTINHO *et al*., 2018; GOYAL, 2014; PEREIRA *et al.* 2014; JUNQUEIRA, AMORIM; OLIVEIRA, 2018; WANDERLEY; AMORIM; CARVALHO, 2014).

Coeficiente de determinação (R^2)

O coeficiente de determinação (R^2) (Equação 6) avalia a qualidade do ajuste de modelo indicando o quanto ele foi capaz de explicar os dados de referência. Foram calculados os coeficientes para cada rede criada, comparando os valores de precipitação diários reais com os estimados. Quanto maior o valor encontrado, melhor ele se ajusta ao modelo.

$$R^2 = \frac{\sum_{i-1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i-1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i-1}^{n} e_i^2/(n-1)}{\sum_{i-1}^{n}(Y_i - \bar{Y})^2/(n-1)} \qquad (6)$$

Em que: R^2 é o coeficiente de determinação (%),Y_i é o valor observado da variável dependente, $\hat{Y}_i$ é o valor estimado da variável dependente e $\bar{Y}$ é a média da variável dependente.

Erro absoluto médio (MAE)

Segundo Alves *et al.* (2012) o MAE refere-se ao desvio médio absoluto dos valores interpolados em relação aos valores observados. É considerado como uma medida precisa e robusta para verificação de modelos numéricos. O ideal é que os seus valores sejam os mais próximos de zero ou igual a zero (equação 7).

$$MAE = \frac{\sum_{j=1}^{n}|O_j - x_j|}{n} \qquad (7)$$

Em que: MAE é o erro absoluto médio (mm), O_j são os valores observados nas estações

de medição, x_j correspondem aos valores estimados pelo método de preenchimento e “n” ao número de observações.

Erro médio quadrático (RMSE)

Através do RMSE, é possível verificar a magnitude média dos erros estimados. O valor obtido sempre é positivo e, quanto mais próximo de zero, melhores os valores estimados. Esse parâmetro pode ser calculado pela Equação 8.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(O_j - x_j)^2}{n}} \quad (8)$$

Em que: RMSE é o erro quadrático médio (mm), O_j são os valores observados nas estações de medição, x_j correspondem aos valores estimados pelo método de preenchimento, e “n” ao número de observações.

Índice de confiança (c) e de concordância (d)

Por meio do índice de confiança é possível verificar a precisão e exatidão dos resultados obtidos. E o índice de concordância é utilizado em diferentes simulações de um mesmo fenômeno. Os valores deste índice variam de 0, que representa nenhum acordo e 1, uma ótima concordância. Pode-se analisar na Tabela 2 os critérios para avaliação do desempenho. Estes parâmetros podem ser obtidos a partir das Equações 10 a 12.

$$D = 1 - \frac{\sum_{j=1}^{n}(O_j - x_j)^2}{\sum_{j=1}^{n}(|x_j - \bar{O}| + |O_j - \bar{O}|)^2} \quad (10)$$

$$r = \frac{\frac{\sum_{j=I}^{N}(x_j - \bar{x}) \times (O_j - \bar{O})}{N}}{\frac{\sqrt{\sum_{j=1}^{N}(x_j - \bar{x})^2}}{N} \times \frac{\sqrt{\sum_{j=1}^{N}(O_j - \bar{O})^2}}{N}} \quad (11)$$

$$C = (r \times D) \quad (12)$$

Em que: D refere-se ao índice de concordância e C ao índice de confiança (ambos adimensionais); O_j são os valores observados nas estações de medição, $\bar{O}$ média dos valores

estimados, x_j correspondem aos valores estimados pelo método de preenchimento, r é o coeficiente de correlação (adimensional) utilizado para o cálculo do índice de confiança, $\bar{x}$ aos valores estimados médios e "n" ao número de observações e n ao número de observações.

Tabela 2 - Valores do Índice de confiança (C) para avaliação e análise de desempenho dos modelos.

Valor do C	Desempenho
> 0,85	Ótimo
0,76 a 0,85	Muito bom
0,66 a 0,75	Bom
0,61 a 0,65	Intermediário
0,51 a 0,60	Tolerável
0,41 a 0,50	Ruim
≤ 0,40	Péssimo

Fonte: Coutinho et al. (2018).

Coeficiente de Nash-Sutcliffe (Nash)

O coeficiente de eficiência de Nash-Sutcliffe (Equação 13) é um dos métodos estatísticos mais importante e usuais em hidrologia para avaliar desempenho de modelos hidrológicos, conforme descrito por Pereira *et al.* (2014). Este coeficiente pode variar entre -∞ a 1, sendo que ao obter o valor correspondente a 1, tem-se o ajuste ideal dos dados estimados.

$$Nash = 1 - \frac{\sum_{i=1}^{n}(X_{obs,i} - X_{sim})^2}{\sum_{i=1}^{n}(X_{obs,i} - \overline{X}_{obs})^2} \quad (13)$$

Em que, *Nash é o* coeficiente Nash-Sutcliffe (adimensional), X_{obs} é o dado de precipitação observado, X_{sim} é o dado de precipitação simulado pelo modelo, $\overline{X}_{obs}$ é a média dos dados observados no período de simulação e n, o número de eventos.

Foi adotada a classificação conforme Silva *et al.* (2008): ao se encontrar valores de coeficiente maiores que 0,75, considera-se que o modelo é adequado e bom; entre 0,36 e 0,75, o modelo é considerado aceitável, e valores inferiores a 0,36, o modelo é considerado inaceitável.

Análise descritiva e inferencial e testes de normalidade dos dados

O pressuposto de normalidade de Anderson Darling foi aplicado por apresentar como vantagem ser um teste mais sensível e por dar mais peso aos pontos das caudas da distribuição (ESPINOSA; CALIL JÚNIOR; LAHR, 2004). Também se verificou através do teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) por ser um dos mais eficientes para identificação de dados não normais.

Para ambos os testes foi utilizado nível de significância de 0,05, sendo que, ao encontrar p-valores menores que nível de significância, rejeita-se H_0, correspondendo aos dados que seguem uma distribuição normal. Desta forma, quando tais dados não apresentaram normalidade, foi realizada a análise dos dados com técnicas não paramétricas e análise da mediana dos dados.

As hipóteses para o teste de Anderson Darling e Shapiro-Wilk são:

H_0: os dados seguem uma distribuição normal.

H_1: os dados não seguem uma distribuição normal.

Testes não paramétricos Mann-Whitney (MW) e Kruskal-Wallis (KW)

Foi realizado o teste não paramétrico de Mann-Whitney com nível de significância de 0,05 para testar se havia diferenças significativas entre o valor estimado pelos métodos e o dado real de referência da precipitação. Ao encontrar p-valores maiores que o nível de significância, aceita-se H_0 e rejeita-se a H_1, em que a distribuição das duas amostras são as mesmas (TRIOLA, 2008).

Teste de hipótese:

H_0: os dados provêm de amostras iguais.

H_1: os dados provêm de amostras diferentes.

Também foi aplicado o teste não paramétrico de Kruskal-Wallis proposto por Kruskal e Wallis (1952). Adotou-se o nível de significância de 0,05 para determinar se as medianas entre os 4 métodos de preenchimentos utilizados diferem significativamente entre si. Desta forma, se o p-valor encontrado for menor ou igual ao nível de significância, rejeita-se a hipótese nula e se conclui que nem todas as medianas são iguais.

H_0: as diferenças entre alguma das medianas são estatisticamente significativas.

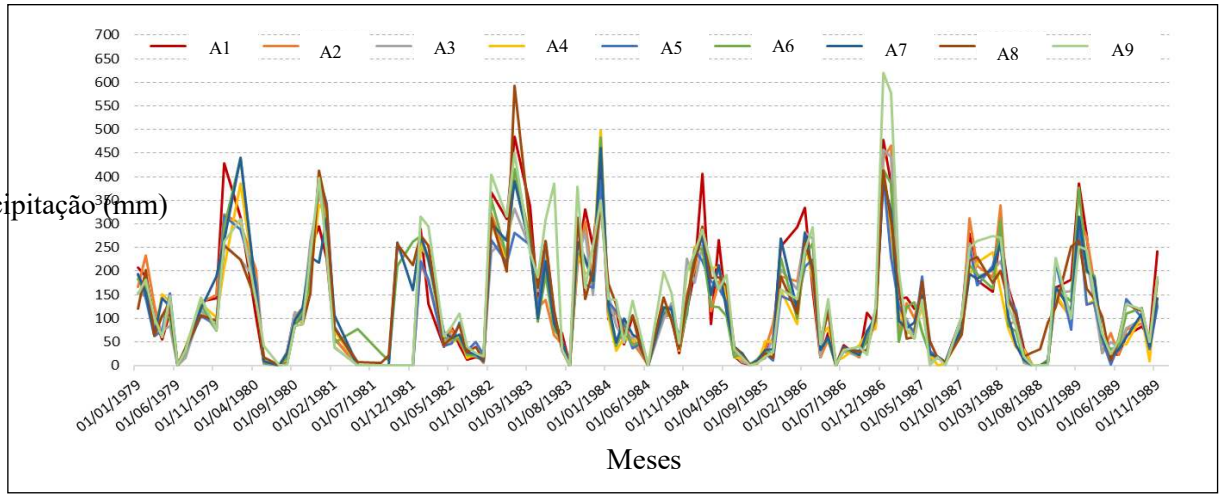
H_1: as diferenças entre alguma das medianas não são estatisticamente significativas.

RESULTADOS E DISCUSSÃO

As precipitações pluviais assumem um forte destaque na compreensão do clima da região. A Zona de Convergência do Atlântico Sul (ZCAS) é a responsável pelos níveis elevados de chuvas diárias e consecutivas na área de estudo, sendo que, ela realiza o transporte de umidade para essa região por meio da junção entre os sistemas de Massa Equatorial Continental e Massa Tropical Atlântica que agem sobre a porção oeste da região Amazônica e a porção sudeste do oceano Atlântico localizado abaixo da linha do Equador. Desta maneira, quando ocorre o seu bloqueio ou manifestação, há a ocorrência de precipitações na área de estudo provocando estiagens ou aguaceiros (SANCHES; VECCHIA; SOUZA, 2020).

Na Figura 4 é possível analisar a variação temporal da precipitação no período de 1979 a 1989, correspondente a todas as nove estações utilizadas nesta pesquisa para o município de São Carlos.

Figura 4 - Variação temporal da precipitação das nove estações nos anos de 1979 a 1989.



Fonte: Elaborado pelos autores (2020)

Pode-se verificar a semelhança do comportamento da precipitação entre as estações, identificando períodos de maiores picos para os anos de 1982 e 1986, e maiores valores acumulados para a estação A8 e A9. Essa análise completa de todas as séries foi de suma importância para que se pudesse identificar possíveis explicações com relação à sua

variabilidade temporal em cada estação. Tais verificações corroboram com o encontrado por Lima e Amorim (2014) para essa mesma localidade no período de 1980 a 2009.

As diferenças de precipitações acumuladas entre as estações podem ser devido à distribuição espaço temporal das chuvas nas diversas localidades, influenciadas pela orografia e as conjunturas geoambientais como, relevo, hidrografia, vegetação, uso e ocupação, intervindo, assim, nas características peculiares de cada localidade como também constatado por Santos *et al*. (2018).

Os maiores acumulados de precipitação pluvial identificados nos meses de outubro a dezembro e janeiro a março por fazerem parte do semestre hidrológico com maiores concentrações ocorridas nesse período devido a atuação da ZCAS e dos processos frontogenéticos (SANCHES; VECCHIA; SOUZA, 2020).

Santos, Souza e Vecchia (2017) identificaram que a variabilidade da precipitação ao longo dos anos na região de São Carlos/SP manteve próximo da média habitual no que se refere à precipitação, com uma média aproximada de 69% de Anos Normais em que o índice pluviométrico ficou muito próximo ou igual à média da série histórica, 15% de Anos Tendentes a Chuvosos, 9,5% de Anos Tendentes a Secos e 4,7% de Anos Secos.

Foi encontrado por Souza e Galvani (2017) em seus estudos localizados na bacia do Jacaré-Guaçu onde se encontra parte do município de São Carlos, que o relevo tem forte influência sobre a variabilidade de precipitações nos sentidos oeste para leste e noroeste para sudeste. Com a variação altimétrica disposta entre 420 e 840 m de leste a oeste, a amplitude pluvial foi de 326 mm e de noroeste para sudeste a amplitude da precipitação foi de 302 mm, disposta sobre uma variação topográfica de 670 m e 830 m de altitude. É de suma importância conhecer as características da realidade climática e geográfica em que estão inseridas as estações meteorológicas a serem estudadas como foi realizado preliminarmente neste estudo.

A compreensão e a determinação da dinâmica espaço temporal da precipitação e de seus principais fatores atuantes na bacia hidrográfica foram relevantes por contribuírem de forma significativa nos diversos métodos de preenchimento de falhas em dados de precipitação que devem ser levados em consideração na escolha do melhor método, contribuindo de forma considerável para a gestão de recursos hídricos, ciclo hidrológico e planejamento urbano das cidades, dentre outros.

Avaliação do desempenho do preenchimento de falha para o município de São Carlos

Foi aplicado o teste de normalidade de Anderson Darling e Shapiro-Wilk para todos os dados reais e estimados e ambos demonstraram p-valores menores que o nível de significância de 0,05, concluindo que os dados não seguem uma distribuição normal. É essencial a aplicação do teste de normalidade, pois a partir disso é possível identificar métodos inferenciais adequados ao conjunto de dados.

Neste estudo foram utilizadas técnicas não paramétricas e para fins de comparação entre os dados reais e estimados, a medida utilizada foi a mediana, pois a média que é tradicionalmente aplicada nos estudos, só pode ser trabalhada em distribuições numéricas normais e esta pode ser influenciada por valores muito maiores ou menores do que o conjunto de dados representa. Sendo assim, a mediana expressa de forma coerente a relação entre ponto médio dos dados que não apresentam normalidade.

A Tabela 3 apresenta as análises de desempenho do método da razão normal. Foi o obtido o melhor desempenho para a estação A3 com valores de R^2 0,961, o MAE foi de 15,803 mm, o RMSE com 21,589 mm, o *Nash* foi de 0,959 e foi obtido um ótimo desempenho (Tabela 3). As estações A2, A4 e A5 também demonstraram altos valores de R^2 e Nash acima de 0,9. Com relação a avaliação dos erros, também apresentaram baixos valores, com ótimo desempenho, apontando, assim, que o modelo gerado por esse método é bom e aplicável para todas as estações, com exceção da estação A7 que apresentou os piores resultados em todas as análises, indicando que não se deve utilizar para essa estação este método de preenchimento de falha devido não representar de forma correta os valores que são estimados.

Tabela 3 - Desempenho do método da Razão Normal através do R^2, MAE, RMSE, Nash, D, C e desempenho gerado para todas as estações, de 1979 a 1989.

Estação	R^2 (%)	MAE (mm)	RMSE (mm)	Nash	D	C	Desempenho
A1	0,884	32,750	46,237	0,860	0,956	0,899	Ótimo
A2	0,917	23,527	33,415	0,913	0,975	0,934	Ótimo
A3	0,914	23,070	31,577	0,912	0,976	0,933	Ótimo
A4	0,822	32,960	46,989	0,808	0,947	0,859	Ótimo
A5	0,854	26,793	40,179	0,843	0,959	0,886	Ótimo
A6	0,805	34,538	50,684	0,802	0,940	0,843	Muito bom
A7	0,291	82,110	106,884	0,114	0,724	0,391	Péssimo
A8	0,836	28,834	44,515	0,833	0,953	0,872	Ótimo
A9	0,792	40,080	64,007	0,756	0,917	0,816	Muito Bom

Fonte: Elaborado pelos autores (2020)

O método de ponderação da distância inversa como é possível observar na Tabela 4, demonstrou bons resultados em relação aos valores de R^2, tendo as estações A2 e A3 com valores maiores que 0,9, assim como também observado pelo coeficiente de *Nash*. Já para a estação A7 foi verificado um péssimo desempenho: os menores valores de R^2 e coeficiente de Nash, respectivamente, iguais a 0,291 e 0,114. Ao realizar a verificação de erros pelo MAE e RMSE foram encontrados erros acima de 80 mm, o que poderia levar a conclusões errôneas para essa estação com a aplicação desse método de preenchimento. As estações A9 e A6 apresentaram o segundo e terceiro maior valor de erros encontrados pela estimativa, mas o seu desempenho, ainda assim, foi considerado como muito bom, com valores de R^2 e coeficiente de Nash acima de 0,7.

Tabela 4 - Desempenho do método de ponderação da distância inversa através do R^2, MAE, RMSE, Nash, D, C e desempenho gerado para todas as estações de, 1979 a 1989.

Estação	R^2 (%)	MAE (mm)	RMSE (mm)	Nash	D	C	Desempenho
A1	0,883	30,807	44,482	0,871	0,961	0,903	Ótimo
A2	0,947	18,795	26,466	0,946	0,985	0,959	Ótimo
A3	0,961	15,803	21,589	0,959	0,99	0,97	Ótimo
A4	0,929	19,85	29,503	0,924	0,98	0,945	Ótimo
A5	0,937	17,83	26,891	0,929	0,983	0,951	Ótimo
A6	0,891	23,264	37,737	0,89	0,97	0,915	Ótimo
A7	0,418	69,849	91,463	0,352	0,794	0,513	Tolerável
A8	0,831	30,751	44,946	0,83	0,952	0,867	Ótimo
A9	0,86	33,726	53,952	0,826	0,944	0,876	Ótimo

Fonte: Elaborado pelos autores (2020)

A Tabela 5 apresenta as análises de desempenho do método de regressão múltipla. Os melhores valores de R^2 para esse método foram encontrados nas estações A2 e A3 com 0,945 e 0,942, respectivamente, e o coeficiente de Nash com valores correspondentes a 0,942 e 0,938 para as estações citadas anteriormente. O pior resultado continua sendo para a estação A7 com péssimo desempenho, baixo valor de R^2 e coeficiente de Nash considerado como péssimo. Os demais apresentaram ótimo desempenho com relação aos dados estimados.

Tabela 5 - Desempenho do Método de regressão múltipla através do R^2, MAE, RMSE, Nash, D, C e desempenho gerado para todas as estações de 1979 a 1989.

Estação	R^2 (%)	MAE (mm)	RMSE (mm)	Nash	D	C	Desempenho
A1	0,898	27,409	39,577	0,886	0,972	0,921	Ótimo
A2	0,945	18,944	26,483	0,942	0,986	0,959	Ótimo
A3	0,942	17,847	25,740	0,938	0,985	0,956	Ótimo
A4	0,914	21,834	31,486	0,906	0,977	0,934	Ótimo
A5	0,908	22,431	30,725	0,899	0,975	0,929	Ótimo
A6	0,889	24,846	37,961	0,875	0,970	0,915	Ótimo
A7	0,378	65,402	89,603	-0,645	0,727	0,447	Péssimo
A8	0,859	28,710	40,926	0,836	0,961	0,890	Ótimo
A9	0,867	30,163	47,224	0,846	0,963	0,897	Ótimo

Fonte: Elaborado pelos autores (2020)

É possível verificar na tabela 6 o desempenho do método redes neurais para as nove estações. Foram encontrados ótimos desempenhos e valores acima de 0,9 para R^2 e para o coeficiente de Nash, assim como menores valores de erros para quase todas as estações analisadas exceto, novamente, para a estação A7, que apresentou um desempenho ruim de acordo com a classificação, altos valores de erros, baixo valor de R^2, com 0,462, e coeficiente de *Nash* considerado como inaceitável. Vale ressaltar que para a estação A2 foi encontrado um perfeito ajuste com relação ao coeficiente de *Nash*, indicando ótima relação entre o dado real e simulado assim como, os menores valores de erro

Tabela 6 - Desempenho do Método de redes neurais através do R^2, MAE, RMSE, Nash, D, C e desempenho gerado para todas as estações de 1979 a 1989.

Estação	R^2 (%)	MAE (mm)	RMSE (mm)	Nash	D	C	Desempenho
A1	0,922	25,071	33,073	0,910	0,979	0,940	Ótimo
A2	0,960	10,296	17,567	0,977	0,994	0,974	Ótimo
A3	0,941	15,446	24,034	0,941	0,986	0,956	Ótimo
A4	0,941	15,362	22,616	0,953	0,988	0,959	Ótimo
A5	0,922	18,692	25,654	0,927	0,982	0,943	Ótimo
A6	0,941	19,117	25,668	0,942	0,986	0,957	Ótimo
A7	0,462	60,688	81,215	0,461	0,774	0,527	Ruim
A8	0,922	20,951	29,884	1	0,979	0,939	Ótimo
A9	0,941	16,320	27,364	0,949	0,988	0,958	Ótimo

Fonte: Elaborado pelos autores (2020)

Realizadas as análises de desempenho dos métodos utilizados nesta pesquisa, utilizou-se também o teste de Mann-Whitney (MW) que tem a função de identificar se existe diferença significativa entre o dado real e o estimado. Desta forma, na Tabela 7 tem-se o resultado desse teste com os seus respectivos p-valores para cada método e os p-valores correspondentes ao teste de Kruskal-Wallis (KW) para verificar se as medianas entre os métodos diferem entre si.

Tabela 7 - p-valor do teste de Mann-Whitney (MW) e Kruskal-Wallis (KW).

Estação	MRN-MW	PDI-MW	RM-MW	RN-MW	KW
A1	0,953	0,928	0,730	0,468	0,896
A2	0,713	0,506	0,997	0,970	0,883
A3	0,659	0,475	0,856	0,775	0,937
A4	0,454	0,166	0,931	0,879	0,48
A5	0,617	0,372	0,547	0,426	0,905
A6	0,744	0,590	0,778	0,479	0,977
A7	0,561	0,250	0,225	0,055	*0,003
A8	0,963	0,851	0,784	0,590	0,836
A9	0,765	0,815	0,733	0,841	0,934

Nota: *significância estatística

Fonte: Elaborado pelos autores (2020)

Para os métodos de MRN, PDI, RM e RN foi constatado que não existe diferença significativa entre os dados reais e estimados para cada estação devido a todos os p-valores

encontrados serem acima de 0,05. Portanto, não existe diferença significativa entre o valor estimado pelos métodos e o dado real de referência da precipitação, corroborando que os métodos são eficientes e viabilizam do preenchimento de falhas de dados precipitação.

Para o teste de Kruskal-Wallis, os p-valores apontaram que não existe diferença significativa entre as medianas dos métodos de preenchimento, exceto para a estação A7 que apresentou p-valor de 0,003, indicando que existe diferença significativa entre os métodos para esta estação.

De forma geral, pode-se observar que todos os métodos apresentaram ótimos desempenhos, exceto para a estação A7 com relação aos testes estatísticos aplicados. Foram obtidos valores acima de 0,7, considerados como fortes para o R^2 e coeficiente de Nash, sendo encontrado um total de 88% desse valor em todos os métodos e estações aplicados. Logo, ao verificarmos os valores significativos acima de 0,7 com relação ao teste de Mann-Whitney, para o MRN, apenas 55,5% dos dados apresentaram valores acima desse limiar. Foram encontrados 33,3% para a PDI. Para a RM foram verificados que 77,7% dos dados foram significativamente fortes, e com relação à RN, 44,4% dos dados indicaram valores acima de 0,7. Por fim, para o teste de Kruskal-Wallis, 77,7% das análises entre os quatro métodos para cada estação apresentaram p-valores acima de 0,7, indicando que as medianas não diferem entre si, ressaltando, assim, que a utilização de qualquer um desses métodos para o preenchimento de falha será eficaz, com exceção para a estação A7, que apresentou baixos desempenhos.

Desta forma, ressalta-se que as peculiaridades que compõe essa região possivelmente influenciaram nos resultados obtidos na estação A7 que apresentou baixos desempenhos em todos os métodos empregados para o preenchimento. Tal resultado também pode ser devido ao efeito orográfico que ocasiona condições diferentes das demais devido à ascensão das massas de ar e expansão adiabática (SANCHES; VECCHIA; SOUZA, 2020).

Foi verificado nesse estudo que os métodos de regressão múltipla e redes neurais são os mais adequados para o preenchimento de falha, assim como verificado nos estudos de Junqueira, Amorim e Oliveira (2018), Mello, Kohls e Oliveira (2017), Oliveira *et al*. (2010) e Silva e Jardim (2017) de e em suas aplicações de técnicas de preenchimento de falha em dados de precipitação mensal e anual.

Silva e Jardim (2017) ressaltam que a transição entre o período seco e úmido representa uma maior dispersão nos dados acumulados de chuva, por apresentarem períodos com baixa precipitação e valores acima da normalidade, resultando em anomalias que ocorrem de forma desproporcional entre as estações avaliadas, o que pode ter dificultado as estimativas deste trabalho para algumas estações.

Moraes e Arraes (2012), em suas análises de metodologia de preenchimento de falha para dados faltantes de precipitação para o Estado do Paraná, utilizaram os métodos de redes neurais, regressão linear e o k-vizinho mais próximo, e encontraram para todos os métodos um coeficiente de determinação (R^2) maior que 0,9. Além disso, o modelo gerado pelas redes neurais foi o mais representativo para substituição de dados faltantes, obtendo menores valores de erro absoluto médio e médio viés, o segundo melhor modelo foi o de regressão linear seguido pelo k-vizinho.

Foram encontrados os melhores resultados para as estimativas de redes neurais, assim como Correia *et al*. (2016) que concluíram em suas análises que a utilização de redes neurais para o preenchimento de falha em dados mensais de precipitação é eficaz, tendo apresentado os melhores valores do coeficiente de Nash, úteis para subsidiar pesquisas que necessitam de séries climáticas consistentes e sem falhas.

Outros parâmetros climatológicos e geoambientais poderiam ser inseridos no método de redes neurais, já que podem ter relação com a precipitação e ser utilizados para o preenchimento de falhas, tais como: temperatura, umidade, velocidade do vento, relevo, altitude, latitude e longitude. Esta é uma das vantagens da utilização de inteligência artificial ao compararmos com demais métodos tradicionais.

A ausência de dados e estações pluviométricas em diversas regiões do país reforça a importância de aproveitar ao máximo os dados existentes. Nesse sentido, tais defasagens podem ser supridas em determinados casos através de eficientes métodos de preenchimento de falhas como foi encontrado nesse estudo, possibilitando assim, a reconstrução de séries históricas de chuvas.

CONCLUSÕES

A partir das análises dos quatro métodos utilizados no estudo, as redes neurais e regressão múltipla apresentaram os melhores resultados, sendo os mais indicados para

preenchimento de falha em dados de precipitação. De maneira geral, para o município de São Carlos, todos os métodos apresentaram ótimos desempenhos nas nove estações analisadas com exceção de uma (estação A7), que apresentou baixos desempenhos em todos os métodos, devido, provavelmente, a condições ambientais diferentes das demais estações.

Esses resultados permitiram gerar séries de dados de precipitação mais consistentes que poderão ser utilizadas como suporte para estudos mais detalhados das possíveis alterações do clima e as suas eventuais implicações para os setores econômico, social e ambiental. Além disso, permitem englobar as mudanças do clima nos processos de planejamento dos recursos hídricos com outras políticas setoriais, incorporando os possíveis impactos que as mudanças nos padrões da estação chuvosa podem acarretar ao meio em escalas locais e globais.

Os resultados obtidos nesta pesquisa a partir de um conhecimento mais aprofundado dos métodos mais eficazes para preenchimento de falhas de dados de precipitação, possibilitaram a reconstrução de séries históricas que poderão também ser úteis para caracterização do clima da região, assim como para possíveis previsões climáticas que possam estar se repetindo mais constantemente com o passar dos anos, contribuindo para avanços de pesquisas nessas temáticas.

AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa.

REFERÊNCIAS

AGÊNCIA NACIONAL DE ÁGUAS-ANA. **Hidroweb: Sistemas de Informações hidrológicas.** Disponível em:<http://www3.ana.gov.br/portal/ANA/panorama-das-aguas/quantidade-da-agua>. Acesso em: 15 de fev.2020.

ALVES, Elis Dener Lima; BIUDES, Marcelo Sacardi; VECCHIA Francisco Arthur Silva. Interpolação espacial na climatologia: análise dos critérios que precedem sua aplicação. **Geonorte**, v. 1, n. 5, p. 606–618, 2012.

BIELENKI JUNIOR, Claudio *et al.* Alternative methodology to gap filling for generation of monthly rainfall series with GIS approach. **Revista Brasileira de Recursos Hídricos**, v. 23, p. 1-10, 2018.

BIER, Anderson Augusto; FERRAZ, Simone Erotildes Teleginski. Comparação de metodologias de preenchimento de falhas em dados meteorológicos para estações no sul do Brasil. **Revista Brasileira de Meteorologia**, v. 32, n. 2, p. 215-226, 2017.

CEPAGRI. CENTRO DE PESQUISAS METEOROLÓGICAS E CLIMÁTICAS APLICADAS À AGRICULTURA. **Clima dos municípios paulistas.** Campinas, 2015. Disponível em:< https://www.cpa.unicamp.br/cepagri/previsao >. Acesso em: 13 de fev. 2020

CORREIA, Tamíres Partélli *et al.* Aplicação de redes neurais artificiais no preenchimento de falhas de precipitação mensal na região serrana do Espírito Santo. **Geociências**, v. 35, n. 4, p. 560-567, 2016.

COUTINHO, Eluã Ramos *et al.* Application of artificial neural networks (ANNs) in the gap filling of meteorological time series. **Revista Brasileira de Meteorologia**, v. 33, n. 2, p. 317-328, 2018.

DE SILVA, Ranjith Premalal; DAYAWANSA, Nihal; RATNASIRI, M.D. A comparison of methods used in estimating missing rainfall data. **Journal of Agricultural Sciences**, v. 3, n. 2, p. 101-108, 2007.

DEPINÉ, Haline *et al.* Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 1, p. 51-63, 2014.

DI PIAZZA, Annalisa *et al.* Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. **International Journal of Applied Earth Observation and Geoinformation**, v. 13, n. 3, p. 396-408, 2011.

ESPINOSA, Mariano Martínez; CALIL JÚNIOR, Carlito; LAHR, Francisco Antonio Rocco. Métodos paramétricos e não paramétricos para determinar o valor característico em resultados de ensaio de madeira. **Scientia Forestalis**, v. 66, n. 50, p. 76-83, 2004.

GOYAL, Manish Kumar. Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India. **Theoretical and Applied Climatology**, v. 118, p. 25-34, 2014.

HEMA, Nagaraja; KANT, Krishna. Reconstructing missing hourly real-time precipitation data using a novel intermittent sliding window period technique for automatic weather station data. **Journal of Meteorological Research**, v. 31, p. 774-790, 2017.

HUANG, Qixing; WANG, J. L.; XUE, X. Interpreting the influence of rainfall and reservoir infilling on a landslide. **Landslides**, v. 13, p. 1139-1149, 2015.

HUBBARD, Katharine. Spatial variability of daily weather variables in the high plains of the USA. **Agricultural and Forest Meteorology**, v. 68, n. 1, p. 29-41, 1994.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. **Censo 2010**. Disponível em:< https://cidades.ibge.gov.br >. Acesso em: 10 maio 2020.

Instituto Nacional de Meteorologia – INMET. **Normais climatológicas do Brasil 1991-2010**. Disponível em: <https://clima.inmet.gov.br/GraficosClimatologicos/DF/83377>. Acesso em: 19 de maio 2021.

JUNQUEIRA, Rubens; AMORIM, Jhones da Silva; OLIVEIRA, Alisson Souza de. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Sustentare**, v. 2, n. 1, p. 198-210, 2018.

KHOSRAVI, Gholamreza *et al.* A Modified distance-weighted approach for filling annual precipitation gaps: application to different climates of Iran. **Theoretical and Applied Climatology**, v. 119, n. 1-2, p. 33-42, 2015.

KRUSKAL, William Henry; WALLIS, Wilson Allen. Use of ranks in on-criterion variance analyses. **Journal os the American Statistical Association**, v. 47, n. 260, p. 583-621, 1952.

LIMA, Altieris Porfírio; AMORIM, Margarete Cristiane de Costa Trindade. Análise de episódios de alagamentos e inundações urbanas na cidade de São Carlos a partir de notícias de jornal. **Revista Brasileira de Climatologia**, v. 15, n. 10, p. 182-204, 2014.

MACHADO, Ronalton Evandro; MATTOS, Arthur. Construção e instalação de um lisímetro com sistema de drenagem. **Revista Brasileira de Agrometeorologia**, v. 9, n. 5, p. 147–151, 2001.

MELLO, Yara Rúbia de; KOHLS, Werner; OLIVEIRA, Therezinha Maria Novais de. Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. **Boletim de Geografia**, v. 35, n. 1, p. 112-121, 2017.

MISHRA, Anoop Kumar. Effect of rain gauge density over the accuracy of rainfall: a case study over Bangalore, India. **Springerplus**, v. 2, p. 311-316, 2013.

MORAES, Rafael Aldighieri; ARRAES, Christiano Luna. Análise de uma metodologia para preenchimento de valores faltantes em dados de precipitação, para o Estado do Paraná. **Científica Ciências Exatas Tecnológicas**, v. 11, n. 1, p. 25-30, 2012.

MWALE, Faidess Dumbizgani; ADELOYE, Adebayo Johnson; RUSTUM, Rabee. Infilling of missing rainfall and streamflow data in the Shire River basin, Malawi – a self organizing map approach. **Physics and Chemistry of the Earth**, v. 50, p. 34-43, 2012.

NAGHETTINI, Mauro; PINTO, Éber José de Andrade. Correlação e Regressão**.** *In:* NAGHETTINI, Mauro; PINTO, Éber José de Andrade (org). **Hidrologia Estatística.** Belo Horizonte: CPRM, 2007, cap. 9, p. 362-363.

NKUNA, Tinyiko; ODIYO, John Ogony. Filling of missing rainfall data in Luvuvhu River Catchment using artificial neural networks. **Physics and Chemistry of the Earth**, v. 36, p. 830-835, 2011.

OLIVEIRA JÚNIOR, José Francisco *et al.* Análise da precipitação e sua relação com sistemas meteorológicos em Seropédica. **Floresta e Ambiente**, v. 21, n. 2, p. 140-149, 2014.

OLIVEIRA, João Bertoldo de. Solos do Estado de São Paulo: descrição das classes registradas no mapa pedológico. **Boletim Científico IAC 45**, Campinas, n. 45, 108 p, 1999.

OLIVEIRA, Luiz Fernando Coutinho de *et al.* Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 11, p. 1186-1192, 2010.

PEEL, Murray C; FINLAYSON, Brian L.; MCMAHON, Thomas. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, p. 1633-1644, 2007.

PEREIRA, Donizete dos Reis *et al.* Hydrological simulation using SWAT model in headwater basin in southeast Brazil. **Engenharia Agrícola**, v. 34, n. 4, p. 789-799, 2014.

PERUSSI, Rafael; VECCHIA, Francisco Arthur da Silva Análise do comportamento térmico de um sistema de cobertura verde utilizando plataformas de teste. **Pesquisa em Arquitetura e Construção**, v. 7, n. 4, p. 25-236, 2016.

SANCHES, Rafael Grecco; VECCHIA, Francisco Arthur da Silva; SOUZA, Paulo Henrique de. Comportamento e distribuição das chuvas na Região de São Carlos/SP a partir de dados de Estações Climatológicas no período de 1993-2014. **Ateliê Geográfico -Goiânia-GO**, v. 14, n. 01, p. 89 -111, 2020.

SANTOS, Bruno César dos; SOUZA, Paulo Henrique de; VECCHIA, Francisco Arthur da Silva. A caracterização da precipitação do ano hidrológico de 2013-2014 na região de São Carlos/SP e sua repercussão no espaço geográfico. **Revista Brasileira de Climatologia**, v. 21, n.13, p. 135-152, 2017.

SANTOS, Roziane Sobreira dos *et al.* Caracterização de extremos mensais de precipitação em Cacoal (RO). **Revista Brasileira de Climatologia**, v. 22, n. 14, p. 267-280, 2018.

SHAPIRO, Samuel Sanford; WILK, Martin. An analysis of variance test for normality (complete samples). **Biometrika**. 1965;52(3) 591-611.

SILVA, Aion Angelu Ferraz; JARDIM, Carlos Henrique Aplicação de técnicas de preenchimento de falhas de dados de pluviosidade mensal e anual para o noroeste do estado de Minas Gerais - Brasil. **Geografias Artigos Científicos**, v. 15, n. 2, p. 83-106, 2017.

SILVA, Pollyana Maria *et al.* Modelagem da hidrografia de cheia em uma bacia hidrográfica da região Alto Rio Grande. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 12, n. 3, p. 258-265, 2008.

SIMOLO, Claudia; BRUNETTI, Michele; MAUGERI, Maurizio. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. **International Journal of Climatology**, v. 29, n. 11, p. 1564-1576, 2009.

SOUZA, Vladimir; GALVANI, Emerson. Distribuição Espaço Temporal Da Precipitação Pluvial E Sua Interação Com O Relevo Na Bacia Do Rio Jacaré Guaçu (SP). **Ciência e Natura**, v. 39, p. 110-124, 2017.

TEEGAVARAPU, Ramesh S. V.; CHANDRAMOULI, Viswanathan. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. **Journal Of Hydrology**, v. 312, n. 1-4, p. 191-206, 2005.

TEIXEIRA, Mateus da Silva, SATYAMURTY, Prakki. Trends in the Frequency of Intense Precipitation Events in Southern and Southeastern Brazil during 1960–2004. **Journal of Climate**, v. 24, n. 7, p. 1913–1921, 2011.

TRIOLA Mário. **Introdução a estatística.** 10. ed. Rio de Janeiro: LTC, 2008.

VÁSQUEZ, Isela Leonor *et al.* Historical analysis of interannual rainfall variability and trends in southeastern Brazil based on observational and remotely sensed data. **Climate Dynamics**, v. 50, n. 3, p. 801–824, 2018.

WANDERLEY, Henderson Silva; AMORIM, Ricardo Ferreira Carlos de; CARVALHO, Fred Oliveira de. Interpolação espacial de dados médios mensais pluviométricos com redes neurais artificiais. **Revista Brasileira de Meteorologia**, v. 29, n. 3, p. 389-396, 2014.

XIA, Youlong; FABIAN, Peter; STOHL, Andreas; WINTERHALTER, Martin. Forest climatology: estimation of missing values for Bavaria, Germany. **Agricultural and Forest Meteorology**, v. 96, n.1–3, p. 131-144, 1999.

YOUNG, Kenneth C. Three-way model for interpolating for monthly precipitation values. **Monthly Weather Review**, v. 120, n. 11, p. 2561-2569, 1992.

ZANCHETTA, Denise; TANNUS, João Luis Sanches; PINHEIRO, Leandro de Souza. Análise biofísica dos processos envolvidos na invasão biológica de sementes de *Pinus elliottii* na Estação Ecológica de Itirapina - SP e alternativas de manejo. **Climatologia e Estudos da Paisagem**, v. 2, n. 1, p. 72-89, 2007.