

Revisão bibliográfica dos principais métodos de preenchimento de falhas em dados de precipitação

Bibliographic review of the main methods of filling in precipitation data

Camila Bermond Ruezzene, Pós-graduanda em Ciências da Engenharia Ambiental, Universidade de São Paulo.

camila.ruezzene@gmail.com

Renato Billia de Miranda, Doutor em Ciências da Engenharia Ambiental, Universidade de São Paulo.

eng.renato.miranda@gmail.com

Frederico Fabio Mauad, professor associado, Universidade de São Paulo.

mauadffm@gmail.com

Resumo

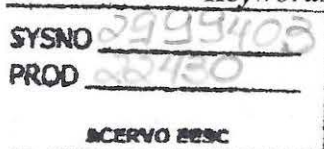
O objetivo deste estudo foi de realizar levantamento bibliográfico dos principais métodos de preenchimentos de falhas em dados de precipitação. Os métodos de preenchimento contribuem para a geração de dados mais confiáveis e nesta revisão foram abordados os métodos de regressão linear em que estima-se os parâmetros da regressão para determinar dados em falta; ponderação regional preenche a série por ponderação com pelo menos três postos vizinhos; interpolação do inverso da distância é definido a vizinhança e os parâmetros de ponderação e calcula-se os dados ausentes por interpolação utilizando ponderação; razão normal é realizado a ponderação dos dados registrados pelas estações vizinhas e redes neurais artificiais determina-se a arquitetura da rede e estima-se parâmetros e dados perdidos. A reconstrução de séries históricas por esses métodos possibilita a realização de estudos detalhados das alterações climáticas e fornecem subsídios para o planejamento de gestão de recursos hídricos nas diversas regiões do país.

Palavras-chave: Regressão linear; Ponderação regional; Redes neurais artificiais.

Abstract

The objective of this study was to carry out a bibliographical survey of the main methods of filling of faults in precipitation data. Filling methods contribute to the generation of more reliable data and in this review we have addressed the linear regression methods in which the regression parameters are estimated to determine missing data; regional weighting completes the series by weighting with at least three neighboring posts; interpolation of the inverse of the distance is defined the neighborhood and the weighting parameters and calculates the missing data by interpolation using weighting; normal reason is carried out the weighting of the data recorded by neighboring stations and artificial neural networks determines the network architecture and estimates lost parameters and data. The reconstruction of historical series by these methods makes it possible to carry out detailed studies of climate change and provides subsidies for the planning of water resources management in the various regions of the country.

Keywords: Linear regression; Regional weighting; Artificial neural networks.



1. Introdução

Para o estudo de fenômenos atmosféricos e processos climáticos necessita-se de séries históricas confiáveis e completas, no entanto, uma das grandes dificuldades que surge ao se utilizar dados medidos a partir de estações meteorológicas está relacionado a problemas nos dispositivos que acarreta a erros de medição gerando dados inconsistentes, assim como a ocorrências de falhas nos mesmos (BIER; FERRAZ, 2017; COUTINHO et al., 2018). Conforme Wanderley, Amorim e Carvalho (2014) citam 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, esta são técnicas que viabilizam a construção de um novo conjunto de dados por meio de conjuntos 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, para o 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; OLIVIERA JÚNIOR et al., 2014).

Existem uma série de métodos de preenchimento, desde técnicas simples como exemplo, a média aritmética entre estações vizinhas, até métodos mais elaborados, como redes neurais. Esses métodos envolvem desde fórmulas estatísticas até algoritmos, afim de gerar estimativas para dados faltantes (BIER; FERRAZ, 2017). Desta forma, o objetivo deste estudo é realizar levantamento bibliográfico dos principais métodos de preenchimentos de falhas em dados de precipitação mais utilizados na literatura.

2. Métodos de preenchimento de falha

Os estudos de métodos para o preenchimento de falhas em séries temporais contribuem significativamente para a geração de dados mais confiáveis, promovendo continuamente o avanço de pesquisas nas áreas afins. A seguir serão apresentados os principais métodos de preenchimento de falhas.

2.1 Método da Regressão Linear (RL)

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 atribuído ao acaso (variação residual). Admitindo ser uma reta a linha teórica de regressão, a função entre X e Y é da seguinte forma conforme a equação (1):

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

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 o método dos mínimos quadrados e este é 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 (NAGUETTINNI; PINTO, 2007).

No método de regressão linear simples as precipitações de um posto com falhas e de um posto vizinho são relacionadas. Na regressão linear múltipla as informações pluviométricas do posto com falhas são relacionadas com as correspondentes observações de vários postos vizinhos (BERTONI; TUCCI, 2013).

Diversos trabalhos demonstram o método de regressões lineares múltipla como uma ferramenta eficiente para o preenchimento de falhas de séries temporais meteorológicas, como precipitação, temperatura, umidade, dentre outros. No trabalho de Coutinho et al. (2018) ao analisar os modelos de regressões lineares múltiplas e redes neurais, pôde-se verificar que os métodos de regressões lineares apresentaram resultados satisfatórios com altos índices de correlação e baixos erros médios comparados com os dados reais.

Mello, Kolhs e Oliveira (2017) utilizaram os métodos de regressão linear simples e múltipla, ponderação regional, e ponderação regional com base em regressões lineares para o preenchimento de falhas de dados pluviométricos, com o objetivo de verificar qual técnica melhor se adequaria para a região de Joinville (SC). O método estatístico que apresentou os melhores resultados foi a regressão linear múltipla demonstrando o resultado mais próximo do real e um menor desvio relativo médio, seguido do método de ponderação regional e o de ponderação regional com base em regressões lineares. Ainda conforme este trabalho, as análises confirmaram que os melhores resultados foram encontrados pelo método mais sofisticados estatisticamente, logo nem sempre há dados disponíveis no mesmo período em outras estações próximas para realizar o preenchimento. Desta forma, a única alternativa naquele caso, seria a utilização de regressão linear simples.

No estado de Goiás, Oliveira et al. (2010) empregaram em séries de precipitações totais anuais em 6 postos com 22 anos contínuos de dados, comparações com cinco metodologias de preenchimento de falhas sendo: regressão linear e potencial múltipla, ponderação regional, ponderação regional com base em regressões lineares e vetor regional. Os autores puderam concluir com os testes que os melhores resultados foram demonstrados pela técnica de regressão linear, seguido pela técnica que combinava o vetor regional e regressão potencial e a ponderação regional como a terceira melhor técnica.

Junqueira, Amorim e Oliveira (2018) ao analisarem a regressão linear simples e múltipla para preenchimento de falhas em dados pluviométrico na bacia hidrográfica do Rio das Mortes, utilizando dados mensais da Hidroweb (ANA, 2018) no período de 1998 a 1995 correspondentes a 28 anos de dados contínuos, encontraram maiores variações nos valores da média observada, correspondendo a cerca de 7,6% dos valores estimados menores do que o observado para a regressão linear simples, e ainda ressalta que se fosse utilizado este método para dimensionamento de reservatório para abastecimento público ou irrigação, haveria um subdimensionamento da obra e erros nos cálculos de disponibilidade hídrica na região. Logo, a regressão linear múltipla foi o método que apresentou os melhores resultados com os valores observados.

Por fim, cabe destacar que existem algumas ferramentas que auxiliam no melhoramento da regressão linear para o preenchimento de falhas, dentre estas pode-se citar o algoritmo *Intermittent Sliding Window Period (ISWP)* empregado por Hema e Kant (2017) em que ao utilizarem essa técnica detectaram reduções do erro absoluto médio e a raiz do erro quadrático médio, sendo essas formas de medir o desempenho de determinados modelos de preenchimento de falhas.

2.2 Ponderação regional (PR)

2.2.1 Ponderação regional por médias

Considera-se um grupo de N postos, X_1, X_2, \dots, X_{N-1} e Y (ao menos 3 postos que possuem no mínimo 10 anos de dados e que seja semelhante a região climática do posto que será preenchido), este último representando a série dependente do grupo homogêneo. O valor estimado para preenchimento da série do posto “Y”, em determinado mês, é determinado pela equação (2) conforme descrito por Recursos Hídricos e Ambientais - RHA (2009):

$$y = \frac{1}{N-1} \cdot \left[\left(\frac{\bar{y}}{\bar{x}_1} \right) \cdot x_1 + \left(\frac{\bar{y}}{\bar{x}_2} \right) \cdot x_2 + \dots + \left(\frac{\bar{y}}{\bar{x}_{N-1}} \right) \cdot x_{N-1} \right] \quad (2)$$

Em que:

y = total mensal precipitado, estimado (preenchido ou estendido) para o posto “Y”, no referido mês;

\bar{y} = total médio precipitado na estação “Y”, no mês em referência, correspondente ao período comum de observação;

\bar{x}_i = total médio precipitado para a estação “Xi” do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

x_i = total mensal observado na estação “Xi”, no mês em que o total de precipitação na estação “Y” deve ser preenchido ou estendido.

2.2.2 Ponderação regional por correlação

Conforme detalhado por RHA (2009) em um exemplo de um mês qualquer representado pela equação (3):

$$y = \frac{s_y}{N-1} \cdot \left[r_1 \cdot \frac{x_1 - \bar{x}_1}{s_{x_1}} + r_2 \cdot \frac{x_2 - \bar{x}_2}{s_{x_2}} + \dots + r_{N-1} \cdot \frac{x_{N-1} - \bar{x}_{N-1}}{s_{x_{N-1}}} \right] + \bar{y} \quad (3)$$

Em que:

y = total mensal precipitado, estimado (preenchido ou estendido) para o posto “Y”, no referido mês;

\bar{y} = total médio precipitado na estação “Y”, no mês em referência, correspondente ao período comum de observação;

s_y = desvio padrão do total precipitado na estação “Y”, no mês em referência, correspondente ao período comum de observação;

\bar{x}_i = total médio precipitado para a estação " X_i " do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

s_{xi} = desvio padrão do total precipitado na estação " X_i ", do grupo homogêneo, no mês em referência, correspondente ao período comum de observação;

x_i = total mensal observado na estação " X_i ", no mês em que o total de precipitação na estação " Y " deve ser preenchido ou estendido;

r_i = correlação linear entre a série de total precipitado na estação " Y " e a correspondente série na estação " X_i ", considerando o período comum de observação no mês em referência.

De acordo com Bertoni e Tucci (2013), o método da ponderação regional é uma forma de preenchimento que pode ser considerada simplificada e normalmente utiliza-se séries anuais ou mensais de precipitações, objetivando à homogeneização das informações do período e à análise estatística das precipitações e pode ser utilizado também para a extensão de séries pluviométricas. Classifica-se o método conforme o tipo de estatística utilizado no critério do peso das estações: correlação ou médias.

Alguns estudos demonstram que metodologias para o preenchimento de falhas de dados de precipitação em alguns casos não são tão eficientes devido os dados de precipitação de diferentes estações vizinhas serem menos correlacionadas, e assim gera-se estimativas menos relacionadas com a série original e maiores erros nos resultados, como é o caso do estudo de Bier e Ferraz (2017) que realizou comparações de metodologias de preenchimento de falhas com os métodos de: regressão linear múltipla, ponderação regional, interpolação do inverso da distância, método da razão normal, método tradicional do reino unido e média aritmética simples. Dentre todos os métodos analisados o mais indicado foi o de ponderação regional. Tal resultado apresentou melhores estimativas, mas não se sobressaiu significativamente melhor em relação aos outros métodos. Os autores ressaltam que tais estimativas representam de forma razoável as variações mensais da precipitação, detectando picos de chuvas mensais entre as séries originais e as estimadas, demonstrando assim que é possível gerar boas estimativas para dados mensais e anuais de precipitação.

Metodologias mais simples como de ponderação regional e regressão linear múltipla apresentam bons desempenhos para o preenchimento de falhas em relação aos métodos de vetor regional e ponderação regional com base em regressões lineares, ressaltando assim que essas metodologias não devem ser empregadas sem que aja uma prévia análise regional de seu desempenho (OLIVEIRA et al., 2010).

Segundo Junqueira, Amorim e Oliveira (2018) ao realizar comparações com diferentes metodologias de preenchimento de falhas de dados pluviométricos, pôde contatar 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) e para os métodos de regressão linear e regressão múltipla e interpolação do inverso da distância subestimaram a precipitação.

Conforme o estudo de Diaz, Pereira e Nobre (2018) ao compararem o método de ponderação regional e os dados estimados pelo TRMM-Tropical Rainfall Measuring Mission, ambos apresentaram desempenhos compatíveis, tanto em proximidades com os valores quanto na variabilidade dos regimes sazonais de chuvas, os resultados de correlação também corroboraram a relação dos dados estimados por ponderação regional e

os pelo satélite TRMM indicando uma alta significância estatística, demonstrando assim a validação de ambos os métodos para o preenchimento de falha.

2.3 Ponderação do inverso da distância (IDW)

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 IDW para o preenchimento de falha é calculado conforme a equação (4):

$$D_x = \frac{\sum_{i=1}^n (D_i/d_i)}{\sum_{i=1}^n (1/d_i)} \quad (4)$$

Em que:

D_x 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”.

O método de ponderação de distâncias inversa (SHEPARD, 1968) é um dos métodos mais utilizados para estimar dados faltantes em hidrologia e ciências geográficas (TEEGAVARAPU; CHANDRAMOULI, 2005), este estimador analisa de forma abrangente as propriedades estatísticas das observações da amostra, pois não fornece uma avaliação do erro de predição, sendo este uma ferramenta de interpolação determinística (LY et., 2013). Desta forma, a interpolação tem como função estimar a precipitação em um determinado local onde não há amostragens, com determinadas coordenadas utilizando observações de dados pluviométricos disponíveis. Este método aplica combinações lineares de observações entorno de um raio de pesquisa e com influência decrescente com o aumento da distância.

Existem algumas limitações em relação a esse método como é o caso em que o mesmo para obter sucesso deve-se haver principalmente a existência de autocorrelação espacial positiva, onde haverá provável similaridade nas estações próximas do que em locais mais distantes uns dos outros, então presumisse que a autocorrelação negativa pode tornar-se uma limitação para que se utilize esse método, sendo que apenas a distância não é um critério suficiente para designar a similaridade entre as séries temporais de precipitação. Outro fator é a arbitrariedade quanto a escolha das estações vizinhas, onde há uma recomendação de três ou quatro estações, mas não há um padrão quanto a essa escolha e este método em alguns casos acaba superestimando o número de dias chuvosos em estações meteorológicas próximas (SIMOLO et al., 2009; TEEGAVARAPU; CHANDRAMOULI, 2005).

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. Com base nos resultados, pôde concluir que a ferramenta alcançou resultados satisfatórios, confirmado por meio das baixas medidas de erros calculadas e a automatização trouxe vantagens como menor demanda no tempo de processamento, garantindo assim a eficiência dos resultados.

Segundo Bier e Ferraz (2017) ao realizarem comparações de metodologias de preenchimento de falhas de dados de precipitação, os autores puderam constatarem que o método de interpolação do inverso da distância obteve resultados bastante próximos ao encontrado com o melhor método correspondente ao de ponderação regional, ao realizarem comparações das análises estatísticas dos erros entres os métodos.

2.4 Método da Razão Normal (MRN)

Segundo Young (1992) o método da razão normal trata-se de uma ponderação de dados com registros de dados de estações vizinhas que pode ser calculada pela seguinte equação (5):

$$D_x = \frac{\sum_{i=1}^n D_i w_i}{\sum_{i=1}^n w_i} \quad (5)$$

Em que:

D_x é o dado mensal que necessita 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 W_i ao peso ponderativo atribuído a cada estação vizinha de ordem “i”, que é descrito pela seguinte equação (6):

$$w_i = r_i^2 \left(\frac{n_i - 2}{1 - r_i^2} \right) \quad (6)$$

Em que:

r_i a correlação entre a estação teste e a estação vizinha de ordem “i” e 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.

O método da razão normal foi inicialmente proposto por Paulhus e Koller (1952), e foi modificado por Young (1992), e trata de uma ponderação dos dados registrados em estações vizinhas para o preenchimento de falhas.

Bier e Ferraz (2017) realizou comparações de metodologias de preenchimento de falhas em dados meteorológicos para estações no sul do Brasil no período entre 2002 e 2012, utilizando dados do Instituto Nacional de Meteorologia (INMET), estações pertencentes a ANA e a Companhia Estadual de Energia Elétrica (CEEE), em seus resultados encontrou o método da razão normal com resultados satisfatórios e baixos erros absolutos médios (18,1%) nas estimativas para o preenchimento de falhas.

No estudo de Khosravi et al. (2015) ao empregarem comparações de métodos de preenchimento de falhas que identificasse a melhor estimativa de observações anuais de precipitação em 24 estações localizadas em diferentes regiões do Irã, pôde encontrar com base no menor valor do erro absoluto médio para cada estação os métodos de coordenadas geográfica convencional e o da razão normal com melhores resultados para este estudo.

2.5 Método de Redes Neurais Artificiais (RNAs)

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

Para estimar os dados de precipitação iremos utilizar as redes neurais do tipo Perceptrons Multicamadas (MLP) como representado na Figura 1 a representação estrutural da rede neural composta de sucessivas camadas de neurônios artificiais, sendo formada por nós de entrada, camadas intermediárias e por fim uma camada de saída. No esquema da Figura 9 temos representadas, i entradas, j neurônios que fazem parte da camada intermediária (h), e o k a camada de saída dos neurônios (s). Os pesos sinápticos são constituídos pelos parâmetros w_{hi} e w_{sjk} , das camadas, intermediária e de saída, respectivamente, logo os parâmetros b_{hj} e b_{sk} correspondem aos termos de tendência. Dado as amostras com entradas e saídas, as redes neurais são capazes de automaticamente mapear as suas interações, aprendem este tipo de técnica e realizam o armazenamento deste aprendizado em seus parâmetros (DEPINÉ et al., 2014; MAQSOOD; ABRAHAM, 2007).

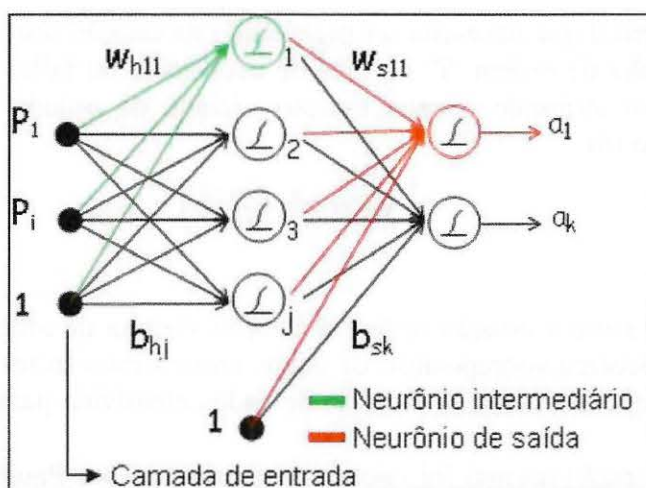


Figura 1. Representação de uma rede neural artificial. Fonte: Depiné et al. (2014).

As camadas ocultas ou também chamadas de intermediárias são primordiais para que se possa introduzir a não linearidade que as tornam ainda mais poderosas por não possuírem apenas entradas e saída como no caso das redes neurais *perceptrons* simples, mostrando-se superiores em relação ao método das redes mais simples (MAQSOOD; ABRAHAM, 2007).

As redes neurais artificiais são um conjunto de técnicas computacionais baseadas na estrutura de aprendizagem do cérebro humano, do qual adquire conhecimento a partir de experiências. A utilização das RNAs para estudo de gerenciamento de recursos hídricos e hidrologia surgiu no início da década de 1990 e em meados desta década o seu uso dobrou, desta forma, constantemente tem se trabalhado com este método como uma ferramenta para modelar os processos meteorológicos e hidrológicos, sendo a inteligência artificial como um dos métodos mais eficientes para preenchimento de falha (WANDERLEY; AMORIM; CARVALHO, 2014).

Ao se trabalhar com ferramentas estatísticas não lineares como é o caso das redes neurais, torna-se as análises mais eficazes para relacionar as complexas entradas e saídas de simulações como precipitação-precipitação, precipitação-vazão, estimativa e transporte de sedimento, previsão de enchentes, qualidade da água, previsão de níveis de reservatório,

assim como simulações climáticas das alterações sobre os recursos hídricos, agrícolas e meteorológicos (COUTINHO et al., 2018; MAQSOOD; ABRAHAM, 2007).

Diversos autores têm trabalhado com o método de redes neurais para preenchimento de falhas em dados de precipitação, assim como o estudo de Depiné et al. (2014) que utilizou o método para preenchimento horário de dados de precipitação de nove estações localizadas na bacia do Arroio Taboão no Rio Grande do Sul entre os anos de 2000 a 2010, dividindo-se o seu experimento em três diferentes alternativas com relação a disposição e extensão da série, sendo a primeira alternativa com 365 dias de dados, a segunda com séries curtas de 12 a 30 dias e a terceira foi semelhante a segunda acrescentando apenas as análises por períodos sazonais. O estudo obteve melhores resultados do preenchimento na terceira alternativa em todos os postos analisados, reproduzindo de forma eficiente os picos de verão.

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.

O método de redes neurais também pode ser aplicado para preenchimento de falhas mensais como foi o caso do estudo de Correia et al. (2016) que realizou a pesquisa na região serrana do Espírito Santo com quatro estações pluviométricas da ANA entre 1947 a 2014, admitindo-se três diferentes porcentagens de falhas nos dados, sendo 7, 15 e 30%, foi verificado que com falhas de 7% o preenchimento foi eficiente, logo o autor concluiu que para o banco de dados com 15 e 30% de falhas o resultado não foi satisfatório.

Segundo Wanderley, Amorim e Carvalho (2014) as variações das redes neurais para o preenchimento de falhas de dados pluviométricos são devido ao procedimento adotado para teste, o treinamento e a validação das redes, o mesmo ainda ressalta que quando há uma menor variabilidade espacial da precipitação os resultados dos dados estimados apresentam mais precisão.

3. Síntese das principais metodologias referentes ao preenchimento de falhas

No Quadro 1 pode-se observar de forma resumida os principais métodos de preenchimento de falhas que foram propostos para utilização neste trabalho, de forma detalhada com a descrição, quando utilizá-los, vantagens e desvantagens e algumas referências correspondentes a cada um deles.

Quadro 1. Visão geral dos métodos de preenchimento de falha.

| Técnica | Descrição | Quando usar? | Vantagens | Desvantagens | Referência |
|--------------------------|--|--|---|---|---------------------------|
| Regressão linear | Estimar parâmetros da regressão e usá-los para determinar dados em falta | Conjuntos de dados exibindo padrões temporais significativos | Os dados estimados preservam o desvio da média e da forma dos dados disponíveis | Poderia distorcer o número de graus de liberdade. | Kuligoski e Barros (1998) |
| Ponderação regional (PR) | Preencher série por | Quando há correlação | Cálculo simplificado | As estações vizinhas devem | Bertoni e Tucci |



| Técnica | Descrição | Quando usar? | Vantagens | Desvantagens | Referência |
|--|--|---|---|--|---------------------------|
| | ponderação com base em pelo menos três postos vizinhos | entre as estações | | ter no mínimo 10 anos de dados | (2013) |
| Interpolação do inverso da distância (IID) | Definir a vizinhança e os parâmetros de ponderação. Em seguida, calcula-se os dados ausentes por interpolação espacial usando ponderação | As estações são altamente correlacionadas | Apresenta melhores estimativas de inferência estatística (quando bem utilizado) | Problemática com a existência de autocorrelação negativa | Hydrology (1996) |
| Método da razão normal (MRN) | Ponderar os dados registrados em estações vizinhas | Quando houver correlação entre as estações vizinhas | Atribuição de pesos para cada estação, dando mais consistência aos resultados | Necessidade de estações vizinhas | Young(1992) |
| Redes neurais artificiais (RNA) | Determinar a arquitetura da RNA, estimar parâmetros e dados perdidos | Quando suposições sobre o mecanismo de dados perdidos não podem ser feitas e no caso de relações não-lineares entre variáveis | Capacidade de modelar padrões complexos sem um conhecimento prévio do processo subjacente | Numerosos parâmetros para estimar e origina em resultados irreais quando tal ruído está disponível nos dados | Raman e Sunilkumar (1995) |

Fonte: Adaptado de Aissia, Chebana e Ouarda (2017).

A escolha dos métodos mais adequados deve contemplar as necessidades do estudo em conformidade com a realidade climática e geográfica em que se estão inseridas e a partir daí é possível que seja gerado séries de dados de precipitação mais consistentes.

4. Conclusão

Através de um conhecimento mais aprofundado dos métodos mais eficazes para preenchimento de falhas de dados de precipitação, possibilitará a reconstrução de séries históricas que poderão ser úteis para caracterização do clima da região, suporte para estudos mais detalhados das possíveis alterações climáticas e as suas eventuais implicações para os setores econômico, social e ambiental, permitindo englobar as mudanças do clima nos processos de planejamento dos recursos hídricos, 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 e nas tomadas de decisão intersetoriais nas diversas regiões do país.

Referências

- AISSIA, M.-A. B.; CHEBANA, F.; OUARDA, T. B. M. J. Multivariate missing data in hydrology – Review and applications. **Advances in Water Resources**, v. 110, 2017. p. 299-309.
- BERTONI, J. C.; TUCCI, C.E.M. Precipitação. In.: Hidrologia: Ciência e Aplicação, Org. Carlos E. M. Tucci, 4ª ed., 5. Reimpr., Porto Alegre: Ed. Universidade/UFRGS: ABRH, 2013. 943 p.
- BIELENKI JUNIOR, C. et al. Alternative methodology to gap filling for generation of monthly rainfall series with GIS approach. **RBRH**, v. 23, 2018. p. 1-10.
- BIER, A. A.; FERRAZ, S. E. T. Comparação de metodologias de preenchimento de falhas em dados meteorológicos para estações no sul do Brasil. **Revista Brasileira de Meteorologia**, v. 32, n. 2, 2017. p. 215-226.
- CORREIA, T.P. 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, 2016. p. 560-567.
- COUTINHO, E. R. et al. Application of artificial neural networks (ANNs) in the gap filling of meteorological time series. **Revista Brasileira de Meteorologia**, v. 33, n. 2, 2018. p. 317-328.
- DEPINÉ, H. et al. Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais. **Revista Brasileira de Recursos Hídricos**, v. 19, n. 1, 2014. p. 51-63.
- DIAZ, C. C. F.; PEREIRA, J. A. S.; NÓBREGA, R. S. Comparação de dados estimados por dois diferentes métodos para o preenchimento de falhas de precipitação pluvial na bacia do Rio Pajeú, Pernambuco, Brasil. **Revista Brasileira de Climatologia**, v. 22, n. 14, 2018. p. 324-339.
- HEMA, N.; KANT, K. 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, n. 4, 2017. p. 774-790.
- HUANG, Q. X.; WANG, J. I.; XUE, X. Interpreting the influence of rainfall and reservoir infilling on a landslide. **Landslides**, v. 13, n. 5, 2015. p. 1139-1149.
- HUBBARD, K.G. Spatial variability of daily weather variables in the high plains of the USA. **Agricultural and Forest Meteorology**, v. 68, n. 1, 1994. p. 29-41.
- HYDROLOGY handbook. New York: ASCE, 1996.
- JUNQUEIRA, R.; AMORIM, J. S.; OLIVEIRA, A. S. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Sustentare**, v. 2, n. 1, 2018. p. 198-210.
- KHOSRAVI, G. 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, 2015. p. 33-42.
- KULIGOWSKI, R. J.; BARROS, A. P. Using artificial neural networks to estimate missing rainfall data 1. **JAWRA – Journal American Water Resources Association**, v. 34, n. 6, 1998. p. 1437-1447.

LY, S.; CHARLES, C.; DEGRÉ, A. Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. A review. **Biotechnology, Agronomy, Society and Environment**, v.17, n. 2, p. 392-406, dez. 2013.

MAQSOOD, I.; ABRAHAM, A. Weather analysis using ensemble of connectionist learning paradigms. **Applied Soft Computing**, v. 7, n. 3, 2007. p. 995-1004.

MELLO, Y. R.; KOHLS, W.; OLIVEIRA, T. M. N. Uso de diferentes métodos para o preenchimento de falhas em estações pluviométricas. **Boletim de Geografia**, v. 35, n. 1, 2017. p. 112-121.

NAGHETTINI, M.; PINTO, E.J.A. **Hidrologia Estatística**. Belo Horizonte. CPRM, 2007.

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

OLIVEIRA, L. F. C. 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, 2010. p. 1186-1192.

PAULHUS, J. L. H.; KOHLER, M.A. Interpolation of missing pre-cipitation records. **Monthly Weather Review**, v. 80, n. 5, p. 129-133, 1952.

RAMAN, H.; SUNILKUMAR, N. Multivariate modelling of water resources time series using artificial neural networks. **Hydrological Science Journal**, v. 40, n. 2, 1995. p. 145-163.

RHA. Recursos Hídricos e Ambientais. **Qualificação de Dados Hidrológicos e Reconstituição de Vazões Naturais no País**. Relatório Final do Contrato N° 016/ANA/2009. Agência Nacional de Águas. 2011. 442 p.

SHEPARD, D. A two-dimensional interpolation function for irregularly spaced data. In: **Proceedings of the Twenty-Third National Conference of the Association for Computing Machinery**, 1968. p. 517-524.

SIMOLO, C. et al. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. **International Journal of Climatology**, v. 29, n. 11, 2009. p. 1564-1576.

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

WANDERLEY, H. S.; AMORIM, R. F. C.; CARVALHO, F. O. Interpolação espacial de dados médios mensais pluviométricos com redes neurais artificiais. **Revista Brasileira de Meteorologia**, v. 29, n. 3, 2014. p. 389-396.

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