




Artigo

Preenchimento de Falhas em Séries Temporais de Precipitação Diária no Rio Grande do Sul

João Paulo Brubacher¹ , Guilherme Garcia de Oliveira¹ , Laurindo Antonio Guasselli¹ 

¹Laboratório de Geoprocessamento e Análise Ambiental, Programa de Pós-Graduação em Sensoriamento Remoto, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.

Recebido em: 27 de Maio de 2019 - Revisado em: 19 de Novembro de 2019 - Aceito em: 22 de Março de 2020

Resumo

O preenchimento de falhas em séries temporais de precipitação é um importante processo para aplicações em hidrologia, visando o aproveitamento de longas séries, evitando que as mesmas sejam descartadas. Desse modo, este estudo teve como objetivo realizar o preenchimento de falhas em séries históricas de precipitação diária no Rio Grande do Sul (RS), auxiliando no aproveitamento desses dados em estudos que necessitem de análises de longo prazo. Para tanto, foram utilizadas séries históricas de 287 estações, no período entre 1987 e 2016 e aplicados os métodos de Regressão Linear Múltipla (RLM) e Redes Neurais Artificiais (RNA), comparando e avaliando os valores preenchidos. Um algoritmo foi desenvolvido para realizar as seguintes operações: i) identificar os dias com falhas em cada estação; ii) identificar as estações que podem ser utilizadas para o preenchimento de cada falha; iii) identificar todas as combinações de entrada para o preenchimento de falhas em cada estação; iv) realizar o ajuste/treinamento dos modelos RLM e RNA; v) realizar a validação dos modelos com base no período sem falhas de cada estação. Os principais resultados indicam que a maior densidade de estações pluviométricas favorece o processo de preenchimento de falhas em séries históricas de precipitação, melhorando a qualidade da série preenchida. O preenchimento de falhas apresentou maior coeficiente de determinação e menor erro médio absoluto usando o modelo RLM em relação às RNA, possivelmente em função da forte correlação linear entre os dados de precipitação de cada local em relação a sua vizinhança. O modelo de RLM apresentou um coeficiente de determinação (R^2) médio de 0,697, enquanto que o modelo RNA obteve média de 0,675. Levando-se em conta a análise por meio do erro médio absoluto (EMA), a média dos valores foi de 2,27 mm para a RLM, enquanto para a RNA o erro ficou em 2,31 mm. Conclui-se, considerando o conjunto de dados de precipitação diária do RS, que houve uma pequena superioridade do método RLM em relação à RNA.

Palavras-chave: regressão linear múltipla, redes neurais artificiais, estações pluviométricas.

Gap Filling in Historical Data of Daily Precipitation in Rio Grande do Sul

Abstract

The gap filling in time series of precipitation is an important process for applications in hydrology, aiming the use of long series, avoiding that they are discarded. Thus, this study had as objective to gap filling in historical series of daily precipitation in Rio Grande do Sul, assisting in the use of these data in studies that require long term analysis. For that, we used historical series of 287 stations in the period between 1987 and 2016 and comparing the Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) methods, comparing to evaluating the filled values. An algorithm was developed to perform the following operations: i) identify the days with missing datas in each station; ii) identify the stations that can be used to fill each missing data; iii) identify all combinations of input to fill in failures in each station; iv) perform the adjustment/training of the MLR and ANN models; v) perform validation of the models based on the period without failure of each station. The main results indicate that the higher density of rainfall stations favors the process of filling of faults in historical series of precipitation, improving the quality of the filled series. Gap filling showed a higher coefficient of determination and lower mean absolute error using the MLR model in relation to RNA, possibly due to the strong linear correlation of the precipitation data of each site in relation to their neighborhood. The MLR model presented an average coefficient of determination (R^2) of 0.697, while the RNA model obtained an

average of 0.675. Considering the analysis by means of the mean absolute error (MAE), the mean values were 2.27 mm for MLR, while for RNA the error was 2.31 mm. It is concluded, considering the daily rainfall data set of RS, that there was a slight superiority of the RLM method in relation to RNA.

Keywords: multiple regression, artificial neural networks, rainfall stations.

1. Introdução

Estudos com séries temporais de precipitação para espacializar chuvas são cada vez mais importantes considerando suas aplicações em estudos hidrológicos, fundamentais para o desenvolvimento social e econômico (Ichiyanagi *et al.*, 2018; Ballari *et al.*, 2018). Séries de precipitação têm aplicações diretas em climatologia, agricultura, hidrologia e gestão de desastres, bem como no planejamento e gestão de recursos ambientais e urbanos (Mekis *et al.*, 2018).

Estudos de previsão hídrica, modelagem, variações e tendências climáticas, por exemplo, principalmente no que se refere a análises de longo prazo, fazem com que a utilização de séries temporais de dados (séries contínuas e sem falhas) seja uma condição fundamental para a execução de uma análise consistente e confiável (Costa *et al.*, 2013). Métodos para preencher falhas podem contribuir na compreensão da variação climática e na previsão de fenômenos que venham a se repetir por influência do clima (Fernandez, 2007). O preenchimento de falhas é necessário para aproveitar séries longas de precipitação, evitando que as mesmas sejam descartadas (Brubacher *et al.*, 2012).

Entre os métodos para o preenchimento de falhas destacam-se a ponderação a partir de Regressão Linear Múltipla (RLM), em que a precipitação no posto com falhas é correlacionada estatisticamente com a precipitação de postos vizinhos. A RLM é um modelo estatístico adequada no tratamento de séries temporais de dados, configurando-se como uma metodologia estatística de previsão de valores de uma ou mais variáveis de resposta (variáveis dependentes) a partir de um conjunto de variáveis explicativas independentes (Naghetini e Andrade Pinto, 2007).

A técnica de RLM correlaciona as observações correspondentes entre vários pontos vizinhos, e as estimativas dos parâmetros da equação podem ser obtidas graficamente ou por meio de critérios de mínimos quadrados para calibração dos valores de forma iterativa (Bertoni e Tucci, 2001). Mello *et al.* (2017) utilizaram Regressão Linear Múltipla para o preenchimento de falhas, com valores obtidos mais próximos do real em comparação com outros métodos, tendo o resultado mais próximo do real em 60,7% das análises e um desvio relativo médio de 14,4%.

Outro método é o modelo de Redes Neurais Artificiais (RNA), uma abordagem computacional alternativa inspirada em estudos do cérebro e sistema nervoso (Haykin, 2008). Baseia-se nas estruturas de armazenamento e processamento do sistema nervoso biológico, que consistem em um grande número de processamentos simples com muitas interconexões (Leahy *et al.*, 2008). São mode-

los matemáticos empíricos, com capacidade de armazenar e disponibilizar conhecimento experimental, por meio de treinamento e de resposta aos estímulos (variáveis de entrada) (Tokar e Markus, 2000; Gupta, 2013).

Coutinho *et al.* (2018) aplicaram RNA para estimar e preencher falhas em séries de dados meteorológicos, comparando as variáveis mensais com modelos de Regressões Múltiplas. Os resultados obtidos por meio da RNA mostram alta correlação linear (r) com os dados registrados de temperatura máxima do ar (r de 0,94 a 0,98), obtendo-se um erro médio percentual (EMP) entre 1,05% e 2,32%. Em relação à umidade relativa do ar, o coeficiente r se manteve entre 0,77 e 0,94, sendo que o uso de RNA para a estimativa dessa variável resultou em um EMP entre 1,85% e 2,41%. Esses resultados destacam a MLP como sendo eficaz para as variáveis testadas.

Depiné *et al.* (2014) aplicaram RNA para preenchimento de falhas horárias em séries históricas de precipitação, e analisaram o desempenho dos modelos a partir do coeficiente de *Nash-Sutcliffe* (NS) e do quantil 95% dos erros (E95). Os coeficientes de *Nash-Sutcliffe* variaram entre 0,91 e 0,99 para o treinamento dos modelos e entre 0,90 e 0,99 para a verificação. O quantil 95% dos erros foi no máximo de 1,5 mm no treinamento e de 1,4 mm na verificação.

Diversos autores trabalham com preenchimento de falhas de precipitação utilizando diversos métodos: (a) RLM (Ramos-Calzado *et al.*, 2008; Presti *et al.*, 2010; Simolo *et al.*, 2010; Tardivo e Berti, 2012; Woldesenbet *et al.*, 2017; Vieira *et al.*, 2018); (b) RNA (Teegavarapu e Chandramouli, 2005; Bustami *et al.* 2007; Aly *et al.*, 2009; Oliveira *et al.*, 2014; Correia *et al.*, 2016; Gimenez, 2017). No entanto, na maioria das aplicações, as áreas analisadas se restringem a poucas estações pluviométricas envolvidas, tanto pelas dificuldades na obtenção de dados quanto pelo tamanho da área de estudo. Quem tem a intenção de trabalhar com esse tipo de dado enfrenta dificuldades semelhantes.

Os resultados promissores obtidos por técnicas de RLM e RNA justificam a escolha desses métodos. Entretanto, diferentemente de outros autores, esse estudo teve como objetivo propor um algoritmo para preencher falhas em séries históricas de precipitação diária para o estado do Rio Grande do Sul, utilizando todas as estações pluviométricas com dados disponíveis, comparando a aplicação dos métodos RLM e RNA. A maior contribuição do algoritmo criado se refere à: i) capacidade de identificar, automaticamente, para cada estação, as falhas e as estações vizinhas disponíveis para o preenchimento de falhas nos dados diários; ii) aplicação de dois modelos utilizando o

maior conjunto de dados disponíveis para cada dia, considerando restrições pré-estabelecidas de distância, para evitar que se use dados de precipitação diária de estações muito distantes, e que não fazem parte da mesma região climática.

2. Materiais e Métodos

2.1. Área de estudo

O estado do Rio Grande do Sul (RS) localiza-se na região Sul do Brasil, faz fronteira ao norte com o estado de Santa Catarina, a oeste com Argentina e ao sul com Uruguai. O RS possui uma área de 281,7 mil km² (Fig. 1). Situa-se em uma zona climática de transição, em que as principais características climáticas refletem a participação de Sistemas Atmosféricos Extratropicais (massas e frentes polares) e Sistemas Intertropicais (massas tropicais

e correntes perturbadas). Embora os primeiros exerçam o controle dos tipos de tempo em 90% dos dias do ano, proporcionando também a distribuição mensal e anual das chuvas (Rossato, 2011).

Segundo Britto *et al.* (2008) o regime pluviométrico no RS é caracterizado pela interação dos sistemas meteorológicos com os elementos da paisagem geográfica, como o relevo, a latitude, a maritimidade e a continentalidade. A pluviosidade é condicionada principalmente pela dinâmica atmosférica em relação ao relevo que, através de sua compartimentação, promove a distribuição espacial da precipitação. No Estado, ocorrem quatro sub-regiões com valores de pluviosidade total semelhantes. No inverno, chove mais na sub-região do litoral e no centro-sul, relacionado ao sistema atmosférico frontal; na primavera, chove mais na sub-região noroeste, onde dominam os Complexos Convectivos de Mesoescala (CCMs); no verão, chove mais na sub-região nordeste associado a chu-

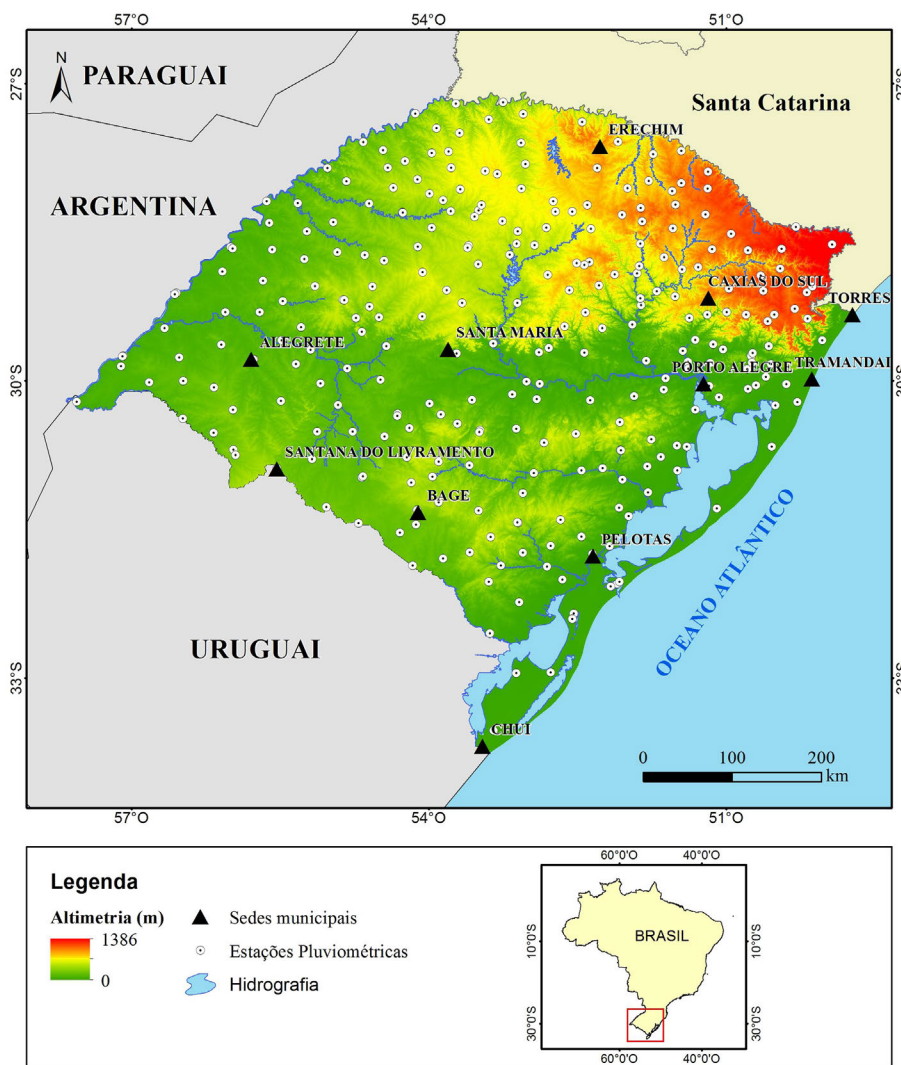


Figura 1 - Área de estudo, e localização das estações pluviométricas sobre imagem de altimetria.

vas convectivas relacionadas às Zonas de Convergência do Atlântico Sul (ZCAS); no outono, chove mais na sub-região sudoeste e o sistema dinâmico mais influente é o bloqueio atmosférico. Registram-se, dessa forma, segundo Rossato (2011), significativas diferenças entre os valores de precipitação das áreas geográficas menos chuvosas (1290 mm anuais) e áreas mais chuvosas (2000 mm anuais).

2.2. Métodos

Para atingir o objetivo proposto, a primeira etapa consistiu na aquisição e organização dos dados de precipitação disponíveis no Hidroweb, da Agência Nacional de Águas (ANA, 2019), e no Instituto Nacional de Meteorologia. Foram utilizados dados de 287 estações pluviométricas (Fig. 1), com séries completas ou parciais no período entre 01/01/1987 e 31/12/2016, totalizando 30 anos de registros. Os dados foram organizados em uma matriz, em que cada linha representa um dia e cada coluna uma estação.

Foi desenvolvido um algoritmo no *software* MATLAB© para o preenchimento de falhas usando os métodos RLM e RNA. Visando comparar os dois métodos e limitar o preenchimento para estações com disponibilidade de dados em locais próximos (reduzindo as chances de buscar valores de precipitações em outras regiões climáticas), foram definidas regras e critérios: i) para uma estação vizinha ser considerada apta para o preenchimento de falhas, o coeficiente de correlação linear (r) da precipitação ocorrida nos dois locais deve ser igual ou superior a 0,6 e a distância máxima entre as estações não pode ultrapassar 100 km; ii) o número mínimo de estações vizinhas para executar o processo de preenchimento é três; iii) o conjunto amostral de treinamento e de validação dos modelos de RLM e RNA, em cada estação, é sempre o mesmo, visando uma comparação idônea entre os métodos utilizados. É importante destacar também que o preenchimento de falhas só foi realizado nas séries históricas com ao menos 85% dos dados disponíveis.

O algoritmo proposto possui as seguintes operações: i) identifica os dias com falhas em cada estação; ii) identifica as estações que podem ser utilizadas para o preenchimento de cada falha, iii) identifica todas as possíveis combinações de entrada para o preenchimento de falhas em cada estação; iv) realiza o ajuste de um modelo RLM; v) realiza o treinamento de uma RNA; vi) realiza a validação dos modelos com base no período sem falhas da estação, identificando a melhor combinação de variáveis explicativas e o modelo mais adequado para cada estação, de acordo com o índice de desempenho desejado.

Para a Regressão Linear Múltipla, as informações pluviométricas do posto Y são correlacionadas com as observações correspondentes de vários postos vizinhos, por meio de uma equação linear (Eq. 1).

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \beta_i.X_i + \varepsilon_0 \quad (1)$$

em que Y é a variável dependente, nesse caso, a precipitação estimada; X_1, X_2, \dots, X_i , são as variáveis independentes, nesse caso os postos considerados; $\beta_0, \beta_1, \beta_2, \dots, \beta_{2i}$, são os coeficientes da regressão; e indica o erro residual.

O treinamento da RNA foi realizado pelo método retropropagativo com múltiplas camadas (Rumelhart *et al.*, 1986), usando a Regra Delta (Widrow e Hoff, 1960) para atualizar os pesos sinápticos da rede. Os valores dos atributos de entrada foram escalonados a partir de transformações lineares (Eq. (2)).

$$\frac{(y_t - p_o)}{S_o} = RNA \left(\frac{(x_t - p_i)}{S_i} \right) \quad (2)$$

em que x_t e y_t são a entrada e a saída das variáveis, respectivamente; S_o e p_o são a escala e a posição dos parâmetros de saída do modelo; S_i e p_i são a escala e a posição dos parâmetros de entrada do modelo; RNA é a Rede Neural Artificial.

A função de ativação para os neurônios foi a sigmoide unipolar, não linear, contínua e derivável em todo o seu domínio. Uma série de validação cruzada paralela à série de treinamento foi utilizada para evitar o superajustamento do modelo. O modelo de RNA utilizado é representado pela Eq. (3).

$$RNA \left(\frac{(x_t - p_i)}{S_i} \right) = f_o \left(\sum_h w_{ofh} \sum_i w_{hi} \left(\frac{x_t - p_i}{S_i} \right) + b_h \right) + b_o + e_o \quad (3)$$

em que w_h, b_h, f_h, w_o, b_o e f_o são os pesos sinápticos (w); tendência (b); funções de ativação (f), respectivamente, das camadas ocultas (h); das camadas de saída (o); e_o é o erro esperado da camada de saída.

Foi escolhida a estrutura de RNA com apenas uma camada oculta, assim como nos estudos de Depiné *et al.* (2014) e Oliveira *et al.* (2014), tendo em vista que essa arquitetura de RNA tem a capacidade de solucionar problemas complexos e gerar uma grande quantidade de conexões sinápticas, conforme o teorema apresentado por Hetch-Nielsen (1987) e reforçado por Hornik *et al.* (1989). O número de neurônios na camada oculta foi testado, variando entre $ni - 2$ e $ni + 4$ para cada configuração de dados de entrada (ni = número de *inputs*). Essa variação de número de neurônios foi escolhida com base em aplicações anteriores de RNA em hidrologia, como em Oliveira *et al.* (2013) e Depiné *et al.* (2014).

Uma vez que a inicialização dos pesos é aleatória, foram executadas 20 iterações para cada configuração,

com taxa de aprendizagem redefinida para 0,01 no início de cada uma. Isso é necessário para minimizar a chance do treinamento estacionar em um mínimo local. A aceleração da convergência ocorre pela variação da taxa de aprendizado durante os ciclos de treinamento, a partir das soluções apresentadas em Vogl et al. (1988). Foi definido um número máximo de 5.000 ciclos de aprendizagem para interrupção, caso a rede não atinja o limiar de convergência pré-estabelecido.

Os resultados foram analisados para identificar a distribuição espacial das falhas por estações, os locais com

maior número de estações vizinhas para o preenchimento, e o desempenho dos modelos RLM e RNA medido pelo coeficiente de determinação (R^2) e pelo erro médio absoluto (EMA) para cada estação.

3. Resultados e Discussões

A distribuição espacial das estações pluviométricas, Fig. 2, considera o total de falhas por estação (2a), a quantidade de estações vizinhas que atendem aos critérios mínimos para o preenchimento em cada local (2b), a cor-

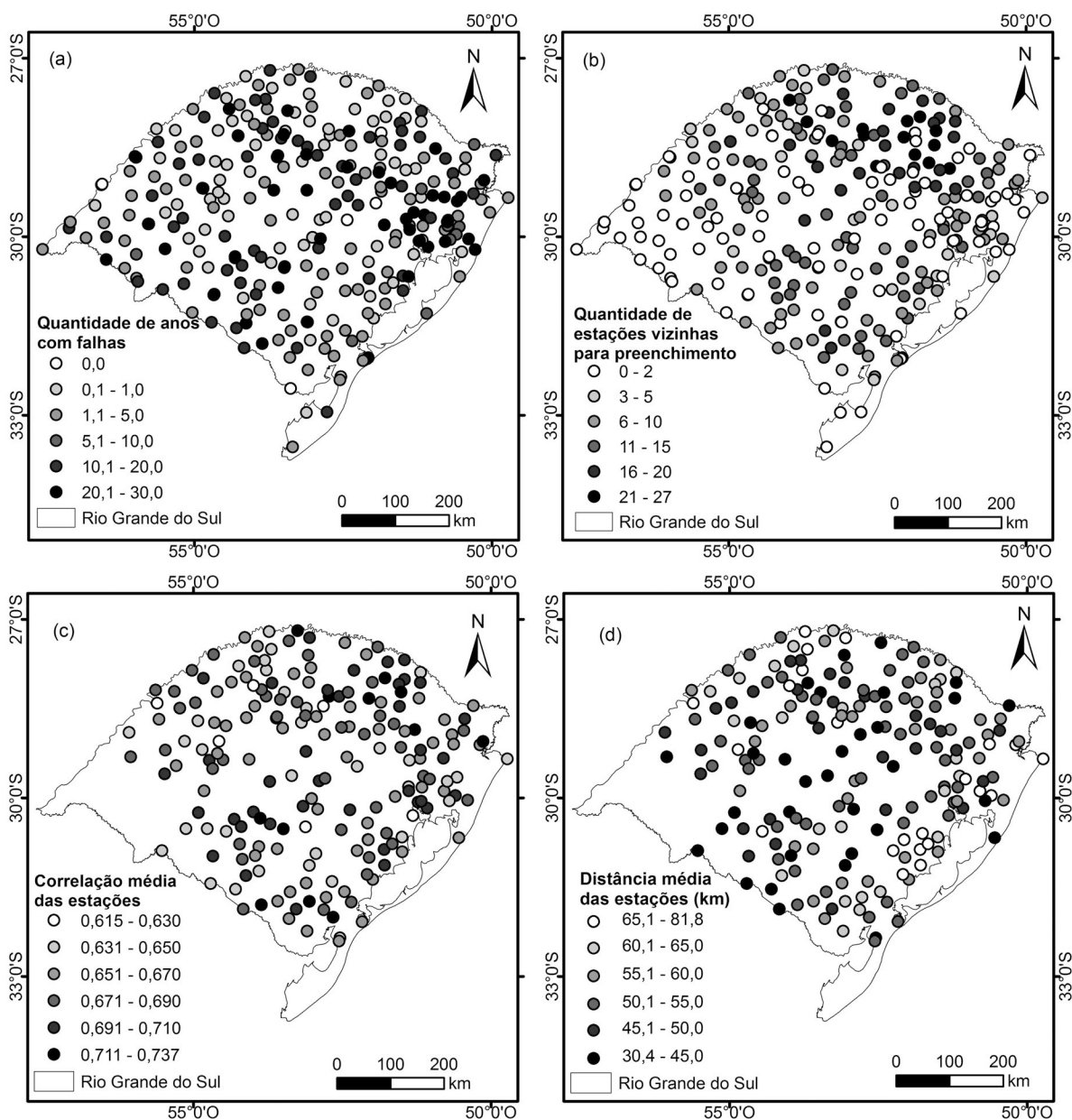


Figura 2 - Distribuição espacial das estações pluviométricas no RS, conforme: (a) quantidade de anos com falhas em cada estação; (b) número de estações vizinhas utilizadas para o preenchimento de falhas; (c) o coeficiente de correlação linear de Pearson média entre a precipitação das estações vizinhas em relação a cada local; (d) a distância média das estações vizinhas em relação a cada local.

relação (2c) e a distância média entre as estações vizinhas (2d) e o local a ser realizado o preenchimento, quando atendidos os requisitos estipulados na metodologia. As Fig. 2c e 2d mostram apenas as estações em que foi realizado o processo de preenchimento de falhas.

Em relação às falhas, apenas 7 estações das 287 analisadas, ou seja, apenas 2,4%, apresentaram séries completas no período de 30 anos. A maioria das estações (60,3%) apresentam menos de 5 anos de falhas e 16,4% das estações apresentam menos de 10 anos de dados no período entre 1987 e 2016.

Em relação à distribuição espacial das estações, de acordo com a Fig. 2b, na região nordeste do RS há uma rede mais densa de estações que atendem aos critérios mínimos estabelecidos para o preenchimento de falhas. No sudoeste (Campanha Gaúcha), há um número menor de estações vizinhas que atendam aos critérios estabelecidos, não permitindo realizar o preenchimento de falhas. Deste modo, essas estações foram excluídas nos mapas de correlação e distância (2c e 2d). Em média, para cada estação, existem sete estações vizinhas que podem ser utilizadas no preenchimento de falhas. Na porção nordeste, em alguns casos, foram utilizadas mais de 20 estações vizinhas no preenchimento.

As estações que apresentam coeficiente médio de correlação linear superior a 0,7, ao relacionar sua precipitação diária com a registrada em estações vizinhas (Fig. 2c), representam 30 estações do conjunto amostrado, ou seja, cerca de 15,3%. A correlação média da precipitação diária entre as estações está associada principalmente à flexibilidade dos critérios utilizados no processo de preenchimento, permitindo o uso de estações distantes até 100 km. A variabilidade das condições pluviométricas associadas às características do relevo do Rio Grande do Sul faz com que os dados diários apresentem diferenças consideráveis entre estações, sendo que nos dados mensais essa característica acaba sendo compensada, aumentando a correlação espacial dos dados.

Segundo Britto *et al.* (2008), os meses de inverno são os mais chuvosos no litoral e no centro-sul do RS, devido à passagem de sistemas frontais, enquanto no sudoeste, o outono é mais chuvoso, devido à configuração de bloqueios atmosféricos. No noroeste chove mais nos meses de primavera, por consequência dos complexos convectivos de mesoescala. Essa análise trata dos padrões sazonais de precipitação no estado, que tende a apresentar uma homogeneidade maior que a chuva diária, caracterizada por grande aleatoriedade dentro uma mesma cidade.

Na distância média das estações, Fig. 2d, a proporção das estações vizinhas que apresentaram distância média menor que 50 km ficou em 25,1%, o equivalente a 49 estações. Há uma maior correlação linear dos dados de precipitação quanto mais próxima for a relação com a sua vizinhança, de modo que as estações localizadas mais próximas ajudam a explicar os resultados obtidos.

O desempenho médio dos modelos RLM (Fig. 3a) e RNA (Fig. 3b) tem como base o coeficiente de determinação (R^2) entre a série preenchida e a série original, e considera o mesmo conjunto de validação por estação. No caso dos modelos de RLM, a proporção de estações que apresentaram R^2 superior a 0,6 foi de 53,6%, enquanto que para os modelos baseados em RNAs (Fig. 3b) foi 48,8%. O resultado do preenchimento de falhas foi superior a 0,8 em 15,3% das estações, considerando RLM, e 10,1%, nos modelos baseados em RNA. Em relação EMA Médio para RLM (Fig. 3e) e EMA Médio para RNA (Fig. 3f), as estações apresentaram resultados semelhantes, no entanto os menores erros foram encontrados no método de RLM, de forma que, das estações preenchidas, 75% das estações apresentaram erros menores que o RNA.

Em relação aos valores máximos de RLM (Fig. 3c) e RNA (Fig. 3d), as estimativas acompanham os valores médios, ou seja, também nesse parâmetro a RLM apresentou valores superiores a RNA. Dessa forma, mesmo que em determinadas estações a discrepância não seja tão evidente, a superioridade da RLM foi predominante.

No contexto espacial, o desempenho do preenchimento de falhas foi superior na porção nordeste do RS. Ainda que o relevo nessa área se apresente com vales mais profundos e linhas de escarpas erosivas, com maior declividade do terreno, o que favorece mudanças mais abruptas nos valores de precipitação, a maior densidade das estações fez com que o processo de transferência de dados entre as estações apresentasse maior precisão.

Das 287 estações, o preenchimento foi realizado em 195, em função das restrições definidas na metodologia. O desempenho do modelo RLM foi superior às RNA em 76% dos locais (Fig. 4). Em relação aos valores médios de R^2 , o modelo de RLM apresentou o valor de 0,697, enquanto a média dos valores para o modelo RNA foi de 0,675. Levando-se em conta a análise por meio do EMA, a média dos valores foi de 2,27 para RLM, enquanto para o RNA o erro ficou em 2,31. Ou seja, para esse tipo de estimativa há uma superioridade do método RLM em relação ao método RNA, para esse conjunto de dados e área de estudo.

Quando consideradas as médias de chuvas mensais, os valores de preenchimento por meio de RNA apresentaram semelhanças com os dados obtidos em estações vizinhas. Esses resultados podem ter sido decorrência de que, ao avaliar mensalmente os dados de chuva, as discrepâncias relativas ao relevo e ao tipo de chuva são suavizadas (Depiné *et al.*, 2014).

Mello *et al.* (2017) desaconselham o preenchimento de falhas quando da análise da distribuição de eventos extremos, dados diários ou de um evento específico. O preenchimento pode gerar um dado não condizente com o registrado no evento analisado. No entanto, os modelos RLM e RNA demonstraram que quanto mais densa for a rede pluviométrica, o preenchimento de falhas pode per-

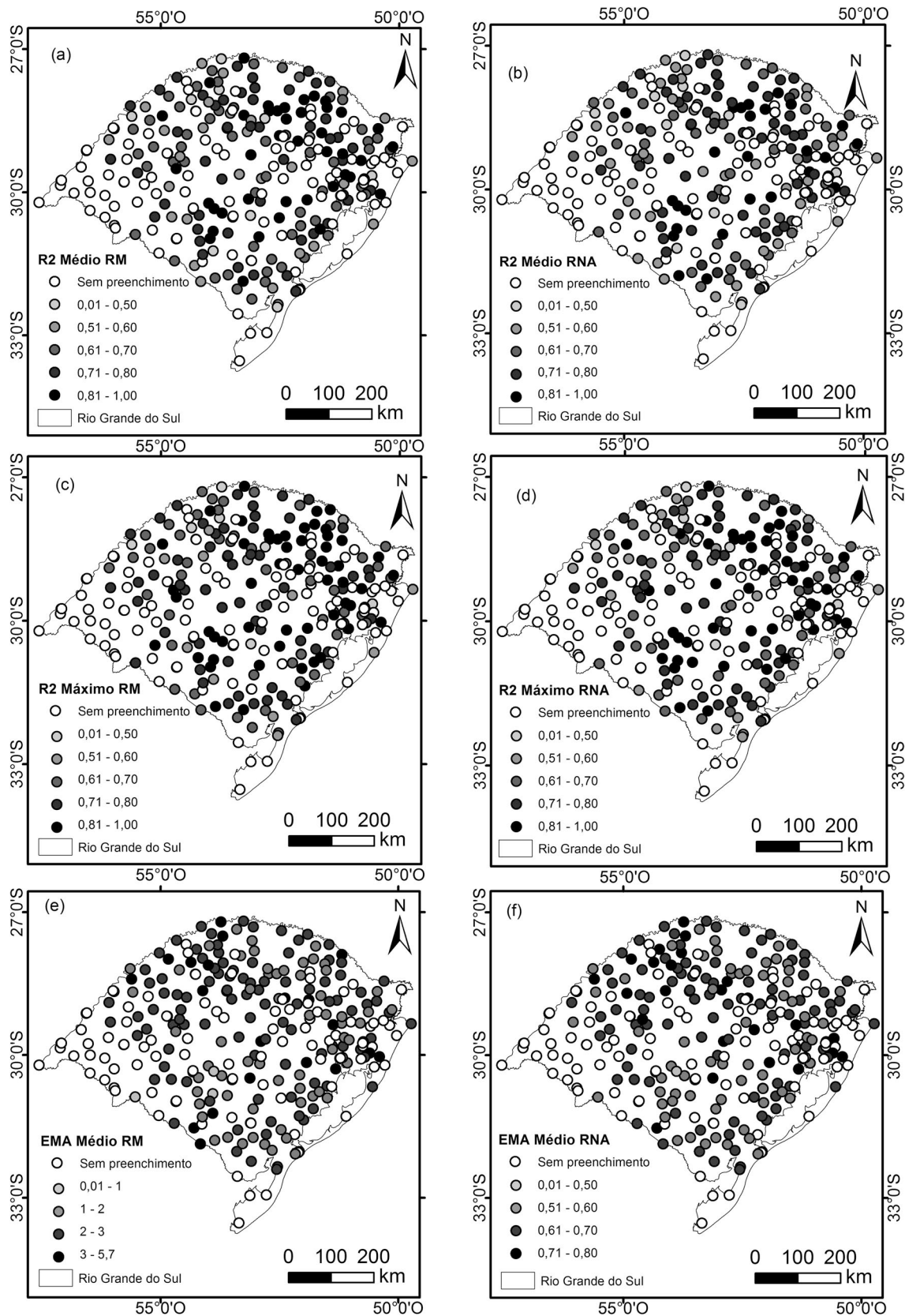


Figura 3 - Desempenho dos modelos RLM e RNA no processo de preenchimento de falhas nas estações pluviométricas do RS: (a) R² médio RLM; (b) R² médio RNA; (c) R² máximo RLM; (d) R² máximo RNA; (e) EMA Médio RLM; (f) EMA Médio RNA.

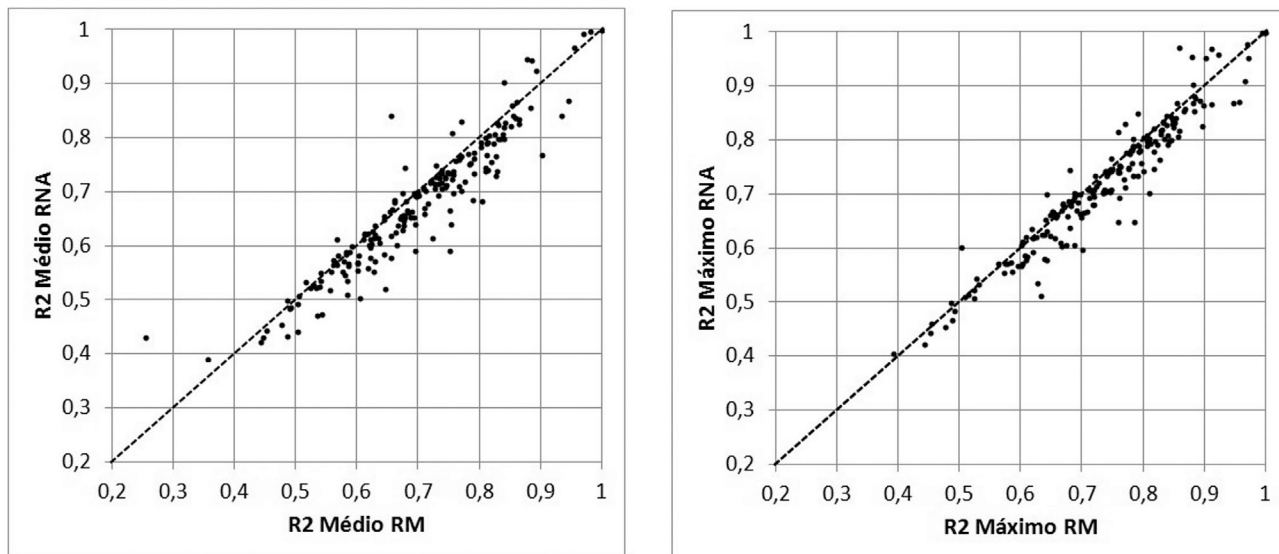


Figura 4 - Dispersão do coeficiente de determinação (R^2) Médio e Máximo no preenchimento de falhas com RLM e com RNA.

mitir a utilização de séries que, anteriormente, teriam de ser descartadas.

A alta variabilidade das chuvas e as muitas formas em que a precipitação pode ocorrer são extremamente desafiadoras (Mekis *et al.*, 2018). Então, ao mensurar o preenchimento de dados faltantes em estações, deve-se levar em conta a capacidade de cada método em lidar com as discrepâncias espaciais.

Não se trata apenas de decidir qual o melhor método, mas sim, com base na análise dos dados, verificar a eficiência do método para a complexidade espacial. Considerando a estrutura das RNA e sua alta capacidade para solucionar problemas complexos e não-lineares, o resultado obtido foi diferente do esperado. A hipótese inicial era de que as RNA apresentassem um valor de preenchimento de precipitação mais próximo das estações de validação, ainda que resultasse em um conjunto extenso de pesos sinápticos na rede.

Nos testes realizados por Ventura *et al.* (2013), os coeficientes de correlação variaram entre 0,79 e 0,96 e o erro médio absoluto entre de 0,62 e 1,22, mostrando um bom desempenho da rede neural para uma série de dados com valores ausentes. Nesse trabalho, os valores ficaram entre 0,43 e 0,73, o que demonstra que a área de estudo pode influenciar nos resultados, justificando o uso do RNA.

A estrutura simplificada de parâmetros do modelo RLM foi capaz de obter resultados mais próximos ao esperado. O que pode ser explicado pelos altos índices espaciais de correlação linear entre as estações, principalmente na região nordeste. O que indica não ser necessário usar um modelo complexo para estimar valores de precipitação em um local a partir de estações vizinhas. Resultados semelhantes foram encontrados por Oliveira *et al.* (2010) com o método RLM obtendo melhor desem-

penho no preenchimento de falhas em relação aos demais métodos testados. No entanto, ressaltam não poder afirmar que essa metodologia possa ser aplicada sem prévia análise regional de desempenho.

Junqueira *et al.* (2018) também observaram melhores resultados no método RLM, em comparação com outros métodos, sendo o que mais se aproximou dos valores utilizados para validação. Esse método, mais simples em comparação com IDW (utilizado pelos autores) ou RNA, tem se mostrado eficiente em relação a esse tipo de demanda.

No entanto, em relação a estudos com outras variáveis meteorológicas, como precipitação e umidade do ar, Coutinho *et al.* (2018) identificaram que as RNA foram responsáveis pelos melhores resultados para o preenchimento de falhas, evidenciando que o seu uso não deve ser descartado, mas sim avaliado de acordo com a necessidade.

Em áreas com grande densidade de estações, os dois métodos demonstraram bons resultados, com R^2 médio acima de 0,8. Ainda assim, o padrão foi superior para o R^2 médio, tendo em vista que o relevo não foi relevante em relação aos dois métodos utilizados. O RLM apresentou o melhor desempenho, com os melhores resultados nas estações analisadas.

Uma das principais contribuições desse estudo é o desenvolvimento de uma abordagem que considera os melhores resultados obtidos por meio dos dois métodos utilizados, sendo que aplicação se deu em uma larga área de estudo com complexidade de sistemas atmosféricos atuantes e que são diretamente influenciados pelo relevo, fatores que dificultam o preenchimento de dados faltantes, mas ainda assim obteve resultados satisfatórios na maioria das lacunas.

A importância dessa pesquisa diz respeito a proposição de um método para o preenchimento confiável de falhas, mesmo com a complexidade dos sistemas atmosféricos atuantes e da influência do relevo, o que permite uma série de estudos com base nesses dados. Com essa finalidade, o método utilizado se mostrou útil, mesmo que o resultado tenha sido diferente do esperado.

A hipótese inicial era de que as RNA apresentariam valor de R^2 superior a RLM, ou seja, que os valores corrigidos apresentassem maior correlação com a série de validação, devido a sua capacidade de aprendizagem e possibilidade de lidar com dados complexos. No entanto, para o Rio Grande do Sul e com base nos critérios adotados, o método RLM apresentou valores superiores em relação R^2 .

4. Considerações Finais

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

- A maior densidade de estações pluviométricas na porção nordeste do RS favoreceu preenchimento de falhas em séries históricas, melhorando a qualidade da série preenchida;
- O preenchimento de falhas em dados diários de precipitação apresentou maior coeficiente de determinação com o modelo RLM em relação às RNA, possivelmente em função da alta correlação linear espacial entre os valores de precipitação de estações vizinhas, indicando não ser necessário usar um modelo mais complexo.
- Destaca-se o estudo contemplando todo o estado do Rio Grande do Sul, com uma área de mais de 281 mil km², o que diverge de trabalhos que preenchem falhas de precipitação em escalas normalmente em bacias hidrográficas. Além disso, o preenchimento de falhas das séries de precipitação em grandes escalas amplia as possibilidades de estudos hidrológicos regionais.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001; e da FAPERGS, processo 17/2551-0000894-4, Edital 01/2017.

Referências

- ALY, A.; PATHAK, C.; TEEGAVARAPU, R.S.V.; ALQUIST, J.; FUELBERG, H. Evaluation of improvised spatial interpolation methods for infilling missing precipitation records. In: **World Environment Water Resources Congress 2009: Great Rivers**, Kansas City, Missouri, p. 5914-5923, 2009.
- BALLARI, D.; GIRALDO, R.; CAMPOZANO, L.; SAMANIEGO, E. Spatial function al data analysis for regionalizing precipitation seasonality and intensity in a sparsely

- monitored region: Unveiling the spatio-temporal dependencies of precipitation in Ecuador. **Int. J. Clim.**, v. 38, p. 3337-3354, 2018.
- BERTONI, J.C.; TUCCI, C.E.M. Precipitação. In: Tucci, C.E.M. (Org.) **Hidrologia: Ciência e Aplicação**. Porto Alegre: Universidade Federal do Rio Grande do Sul, p. 177-242, 2001.
- BRITO, F.; BARLETTA, R.; MENDONÇA, M. Regionalização sazonal e mensal da precipitação pluvial máxima no estado do Rio Grande do Sul. **Rev. Bras. Clim.**, v. 2, n. 2, p. 83-99, 2008.
- BRUBACHER, J.P.; OLIVEIRA, G.G.; GUASSELLI, L.A. Identificação e análise de eventos extremos de precipitação na bacia hidrográfica do rio dos Sinos/RS. In: **I Congresso Brasileiro Sobre Desastres Naturais**, Rio Claro, 2012. CD-ROM.
- BUSTAMI, R.; BESSAIH, N.; BONG, C.; SUHAILI, S. Artificial Neural Network for Precipitation and Water Level Predictions of Bedup River. **Int. J. Comput. Sci.**, v. 34, n. 2, p. 228-233, 2017.
- COUTINHO, E.R. Aplicação de Redes Neurais Artificiais (RNAs) no Preenchimento de Falhas de Séries Temporais Meteorológicas. **Rev. bras. meteor.**, v. 33, n. 2, p. 317-328, 2018.
- CORREIA, T.P.; DOHLER, R.E.; DAMBROZ, C.S.; BINNOT, D.H.B. Aplicação de redes neurais artificiais no preenchimento de falhas de precipitação mensal na região serrana do Espírito Santo. **Revista Geociências**, São Paulo, UNESP, v. 35, n. 4, p. 560-567, 2016.
- COSTA, M.N. M.; BECKER, C.T.; BRITO, J.I.B. Análise das séries temporais de precipitação do semiárido paraibano em um período de 100 anos - 1911 a 2010. **Rev. Bras. Geogr. Física**, v. 6, n. 4, p. 680-696, 2013.
- DEPINÉ, H.; CASTRO, N. M.R.; PINHEIRO, A.; PEDROLO, O. Preenchimento de falhas de dados horários de precipitação utilizando redes neurais artificiais. **Rev. Bras. Recur. Hídricos**, v. 19, n. 1, p. 51-63, 2014.
- FERNANDEZ, M.N. **Preenchimento de Falhas em Séries Temporais**. Dissertação (mestrado em Engenharia Oceânica) - Fundação Federal Universidade de Rio Grande, Rio Grande, 106 p., 2007.
- GIMENEZ, D.F.S. **Análise da Precipitação Pluvial e Ocorrências de Desastres Naturais no Estado de São Paulo (1976-2012)**. Dissertação de Mestrado em Geografia, Universidade Estadual de Campinas, Campinas, 98 p., 2017.
- GUPTA, N. Artificial neural network. **Network and Complex Systems**, v. 3, n. 1, p. 24-28, 2013.
- HAYKIN, S.O. **Neural Networks and Learning Machines**. 3. ed. Upper Saddle River: Prentice Hall, 889 p. 2008.
- HECHT-NIELSEN, R. **Kolmogorov's Mapping Neural Network Existence Theorem**. Proceedings of the First IEEE-International Joint Conference on Neural Networks, San Diego, California, p. 11-14, 1987.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, v. 2, n. 5, p. 359-366, 1989.
- ICHIYANAGI, K.; YAMANAKA, M.D.; MURAJI, Y.; VAIDYA, B.K. Precipitation in Nepal between 1987 and 1996. **Int. J. Clim.**, v. 27, p. 1753-1762, 2018.

- JUNQUEIRA, R.; AMORIM, J.S.; OLIVEIRA, A.S. Comparação entre diferentes metodologias para preenchimento de falhas em dados pluviométricos. **Sustentare**, v. 2, n. 1, p. 198-210, 2018.
- LEAHY, P.; KIELY, G.; CORCORAN, G. Structural optimisation and input selection of an artificial neural network for river level prediction. **J. Hydrol.**, v. 355, p. 192-201, 2008.
- MEKIS, E.; DONALDSON, N.; REID, J.; ZUCCONI, A.; JEFFERY, H.; LI, Q.; NITU, R.; MELO, S. An overview of surface-based precipitation observations at environment and climate change Canada. **Atmosphere-Ocean Journal**, v. 56, p. 1-25, 2018.
- 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, p. 112-121, 2017.
- NAGHETTINI, M.; ANDRADE PINTO, E.J. **Hidrologia Estática**. Belo Horizonte: CRPM, 552 p., 2007.
- OLIVEIRA, L.F.C.; FIORENZE, A.P.; MEDEIROS, A.M.M.; SILVA, M.A.S. Comparação de Metodologias de preenchimento de falhas de séries históricas de precipitação pluviométrica. **Rev. Bras. Eng. Agríc. Ambient.**, v. 14, n. 11, p. 1186-1192, 2010.
- OLIVEIRA, G.G.; PEDROLLO, O.C.; CASTRO, N.M.R.; BRAVO, J.M. Simulações hidrológicas com diferentes proporções de área controlada na bacia hidrográfica. **Rev. Bras. Recur. Hídricos**, v. 18, n. 3, p. 193-204, 2013.
- OLIVEIRA, G.G.; PEDROLLO, O.C.; CASTRO, N.M.R. O Desempenho das Redes Neurais Artificiais (RNAs) para Simulação Hidrológica Mensal. **Rev. Bras. Recur. Hídricos**, v. 19, n. 2, p. 251-265, 2014.
- PRESTI, R.L.; BARCA, E.; PASSARELLA, G. A methodology for treating missing data applied to daily rainfall data in the Candelero River Basin. **Environ. Monit. Assess.**, v. 160, p. 1-22, 2010.
- RAMOS-CALZADO, P.; CAMACHO J.G.; PÉREZ-BERNAL, F.; PITA, M.F. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. **Int. J. Clim.**, v. 28, n. 11, p. 1525-1534, 2008.
- ROSSATO, M.S. **Os Climas do Rio Grande do Sul: Variabilidade, Tendências e Tipologia**. Tese de Doutorado em Geografia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 253 p., 2011.
- RUMELHART, D.E.; HINTON, G.E.; WILLIAMS, R.J. Learning representations by back-propagating errors. **Nature**, n. 323, p. 533-536, 1986.
- SIMOLO, C.; BRUNETTI, M.; MAUGERI, M.; NANNI T. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. **Int. J. Clim.**, v. 30, n. 10, p. 1564-1576, 2010.
- TARDIVO, G.; BERTI A. A dynamic method for gap filling in daily temperature datasets. **J. Appl. Meteorol. Climatol.**, v. 51, n. 6, p. 1079-1086, 2012.
- TEEGAVARAPU, R.S.V.; CHANDRAMOULI, V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. **J. Hydrol.**, v. 312, p. 191-206, 2005.
- TOKAR, S.; MARKUS, M. precipitation-runoff modeling using artificial neural networks and conceptual models. **J. Hydrol. Eng.**, v. 5, n. 2, p. 156-161, 2000.
- VENTURA, T.M.; OLIVEIRA A.G.; MARQUES, H.O.; OLIVEIRA, R.S.; MARTINS, C.A.; FIGUEIREDO, J.M.; BONFANTE, A.G. Uma abordagem computacional para preenchimento de falhas em dados micro meteorológicos. **Rev. Bras. Ciênc. Ambientais**, n. 27, p. 61-70, 2013.
- VIEIRA, S.A.; OSORIO, D.M.M.; QUEVEDO, D.M.; ADAM, K.N.; PEREIRA, N.A.F. Metodologia de imputação de dados hidrometeorológicos para análise de séries históricas – bacia do rio dos Sinos, RS, Brasil. **Rev. Bras. Clim.**, v. 23, p. 189-204, 2018.
- VOGL, T.P.; MANGIS, J.K.; ZIGLER, A.K.; ZINK, W.T.; ALKON, D.L. Accelerating the convergence of the back-propagation method. **Bio. Cybern.**, v. 59, p. 256-264, 1988.
- WIDROW, B.; HOFF, M.E. **Adaptive Switching Circuits**. 4 IRE Wescon Conv. Record, New York: v. 4, p. 96-104, 1960.
- WOLDESENBET, T.A.; ELAGIB, N.A.; RIBBE, L.; HEINRICH, J. Gap filling and homogenization of climatological datasets in the headwater region of the Upper Blue Nile Basin, Ethiopia. **Int. J. Clim.**, v. 37, n. 4, p. 2122-2140, 2017.

Endereços de Internet

Endereços de internet ANA - AGÊNCIA NACIONAL DE ÁGUAS. **Portal Hidroweb: Sistema Nacional de Informações sobre Recursos Hídricos**. Disponível em: <https://www.snirh.gov.br/hidroweb/>. Acesso em: 18 out. 2018.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.