

COMPARAÇÃO DE MÉTODOS PARA PREENCHIMENTO ESPAÇO-TEMPORAL DE FALHAS EM DADOS HISTÓRICOS DE PRECIPITAÇÃO NO RIO GRANDE DO SUL

João Paulo Brubacher¹, Guilherme Garcia de Oliveira², Laurindo Antonio Guasselli³

^{1,2,3} Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500 / Prédio 44202, Bairro Agronomia, Porto Alegre/RS. E-mail: ¹jp.brubacher@hotmail.com, ²g.g.oliveira@gmail.com, ³laurindo.guasselli@ufrgs.br

RESUMO

Este estudo teve como objetivo comparar a aplicação dos métodos de Regressão Múltipla (RM) e de Redes Neurais Artificiais (RNA) para o preenchimento espaço-temporal de falhas em séries históricas de precipitação no Rio Grande do Sul. Foram utilizadas séries históricas de precipitação diária de 287 estações, no período de 1987 a 2016. Foi desenvolvido um algoritmo para organização dos dados e realização do processo de preenchimento a partir de RM e RNA. Os principais resultados indicam que, em 76% das estações, o modelo RM apresentou melhor desempenho do que as RNAs, tendo obtido R^2 médio de 0,697. Além disso, observa-se que a maior densidade de estações na porção nordeste do RS favorece o processo de estimativa de chuva, melhorando a qualidade da série preenchida.

Palavras-chave — regressão múltipla, redes neurais artificiais, preenchimento de falhas.

ABSTRACT

In this study, we compared the application of the Multiple Regression (MR) and Artificial Neural Networks (ANN) methods for the spatial-temporal gap filling in historical rainfall series in Rio Grande do Sul. We used 287 historical series of daily precipitation in the period from 1987 to 2016. We developed an algorithm to organize the data and perform the fill process with the MR and ANN models. The main results indicate that, in 76% of the stations, the MR model presented better performance than the ANN's, with an average R^2 of 0.697. In addition, it is observed that the higher density of stations in the Northeast RS favors the rain estimation process, improving the quality of the filled series.

Key words — multiple regression, artificial neural networks, gap filling.

1. INTRODUÇÃO

Estudos com séries de precipitação aplicados para a espacialização de chuvas são realizados no mundo inteiro, devido à importância dessa variável para o desenvolvimento social e econômico [1] [2]. A utilização dos dados de chuvas tem aplicações diretas em climatologia, agricultura,

hidrologia e gestão de desastres, bem como no planejamento e gestão de recursos ambientais, naturais e urbanos [3].

Dessa forma, estudos de previsão hídrica, modelagem e variabilidade climática, por exemplo, principalmente no que se refere a análises de longo prazo, faz com que a utilização de uma sequência de dados (séries contínuas e sem falhas) se mostre como uma condição fundamental [4].

O estudo de métodos para o preenchimento de dados em séries temporais busca contribuir para o avanço de pesquisas na compreensão da variação climática e na previsão de fenômenos que venham a se repetir por influência do clima [5]. Há cerca de 100 anos não existiam muitas estações pluviométricas disponíveis, dessa forma, esse tipo de pesquisa torna-se necessário de modo a aproveitar longas séries de precipitação, evitando que, em função de algumas falhas, as mesmas sejam descartadas [6].

Entre os métodos utilizados para o preenchimento de falhas destacam-se: i) o método de ponderação a partir de Regressão Múltipla (RM), segundo a qual a precipitação no posto com falhas é correlacionada estatisticamente com a precipitação de postos vizinhos; ii) o modelo de Redes Neurais Artificiais (RNA), que se trata de uma abordagem computacional alternativa inspirada em estudos do cérebro e sistemas nervosos.

A RM é um modelo estatístico referente ao tratamento de séries temporais de dados, de modo que se trata de uma metodologia estatística de previsão de valores de uma ou mais variáveis de resposta (Dependentes) através de um conjunto de variáveis explicativas (Independentes) [7]. A técnica de RM correlaciona as observações correspondentes entre vários pontos vizinhos, de forma que 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 [8].

As RNAs, por outro lado, baseiam-se nas estruturas de armazenamento e processamento do sistema nervoso biológico, de forma que consistem em um grande número de processamentos simples com muitas interconexões [9]. Tratam-se de modelos matemáticos empíricos, com capacidade de armazenar e disponibilizar conhecimento experimental, por meio de treinamento e da resposta aos estímulos (variáveis de entrada) [10] [11].

Nesse contexto, este estudo teve como objetivo comparar a aplicação dos métodos de RM e RNA para o preenchimento espaço-temporal de falhas em dados históricos de precipitação no Rio Grande do Sul.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo

O estado do Rio Grande do Sul (RS) localiza-se na região Sul do Brasil, fazendo fronteira ao norte com o estado de Santa Catarina, ao leste com o Oceano Atlântico, a oeste com Argentina e ao sul com Uruguai. O RS possui uma área de 281,7 mil km² (Figura 1). O regime pluviométrico do estado do estado 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 [12].

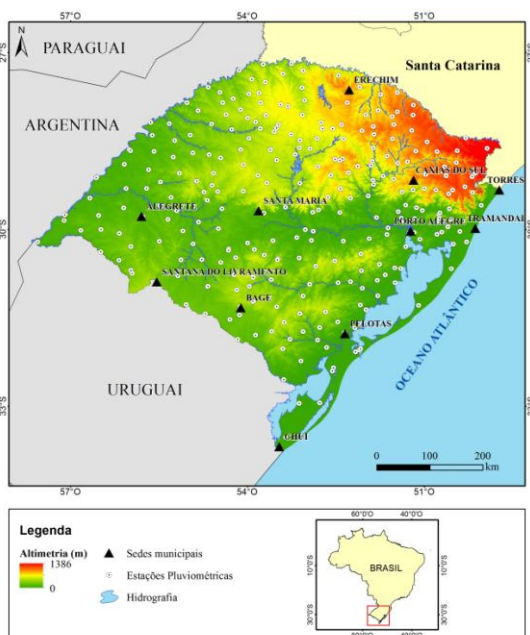


Figura 1. Localização da área de estudo.

2.2. Materiais e Métodos

A primeira etapa consistiu na aquisição e organização dos dados de precipitação disponíveis no Hidroweb, da Agência Nacional de Águas, e no Instituto Nacional de Meteorologia. Ao todo, foram utilizados dados de 287 estações pluviométricas no RS (Figura 1), com séries completas ou parciais no período entre 01/01/1987 e 31/12/2016 (30 anos). Os dados foram organizados em uma matriz, de modo que cada linha representasse um dia e cada coluna uma estação.

Após, na etapa seguinte, foi desenvolvido um algoritmo no *software* MATLAB para o processo de preenchimento de falhas usando os métodos RM e RNA. Para fins de comparação entre os métodos, foram definidas algumas regras e critérios em comum: i) para uma estação vizinha ser

considerada apta, o coeficiente de correlação (r) da precipitação ocorrida nos dois locais deve ser igual ou superior a 0,6 e a distância máxima 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 RM e RNA, em cada estação, é sempre o mesmo.

O algoritmo desenvolvido 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 encontrada; iii) identifica todas as combinações de entrada para o preenchimento de falhas em cada estação; iv) realiza o ajuste de um modelo RM; v) realiza o treinamento de uma RNA; vi) realiza a validação dos modelos com base no período sem falhas da estação.

Para a regressão múltipla, as informações pluviométricas do posto Y são correlacionadas com as observações correspondentes de vários postos vizinhos, por meio de uma equação linear (Equação 1).

$$Yc = a_1 x1_i + a_2 x2_i + \dots + a_{n-1} xn_i + a_n \quad (1)$$

Em que: Yc é a precipitação estimada; n é o número de postos considerados; a_1, a_2, \dots, a_n são os coeficientes a serem otimizados por mínimos quadrados; $x1_i, x2_i, \dots, xn_i$ são as precipitações medidas nos postos vizinhos.

O modelo de RNA também foi escrito no *software* MATLAB. O treinamento da rede foi realizado pelo método retropropagativo com múltiplas camadas [13], usando a Regra Delta [14] para atualização dos pesos sinápticos da rede. Os valores dos atributos de entrada foram escalonados a partir de transformações lineares. A função de ativação utilizada nas camadas do modelo foi a sigmoideal. Uma série de validação cruzada paralela à série de treinamento foi utilizada para evitar o superajustamento do modelo.

O número de neurônios na camada oculta foi testado, variando entre n_i-2 e n_i+4 para cada configuração de dados de entrada (n_i = número de *inputs*). Uma vez que a inicialização dos pesos é aleatória, foram executadas 20 iterações para cada configuração do modelo. Em cada iteração, foi definido um número máximo de 5.000 ciclos de aprendizagem para interrupção, caso a rede não atinja o limiar de convergência pré-estabelecido.

Após realizado o processo de preenchimento de falhas, foi realizada uma análise dos resultados, de modo a identificar a distribuição espacial das falhas por estações, os locais no RS com maior número de estações vizinhas para o preenchimento, e o desempenho dos modelos RM e RNA medido pelo coeficiente de determinação (R^2) para cada estação.

3. RESULTADOS E DISCUSSÕES

Na Figura 2 é apresentada a distribuição espacial das estações pluviométricas, considerando 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 correlação (2c) e a distância (2d) média entre as estações vizinhas e o local a ser realizado o preenchimento, quando atendidos os requisitos estipulados na metodologia. Assim, na Figura 2c e 2d são mostradas apenas as estações em que foi realizado o processo de preenchimento de falhas.

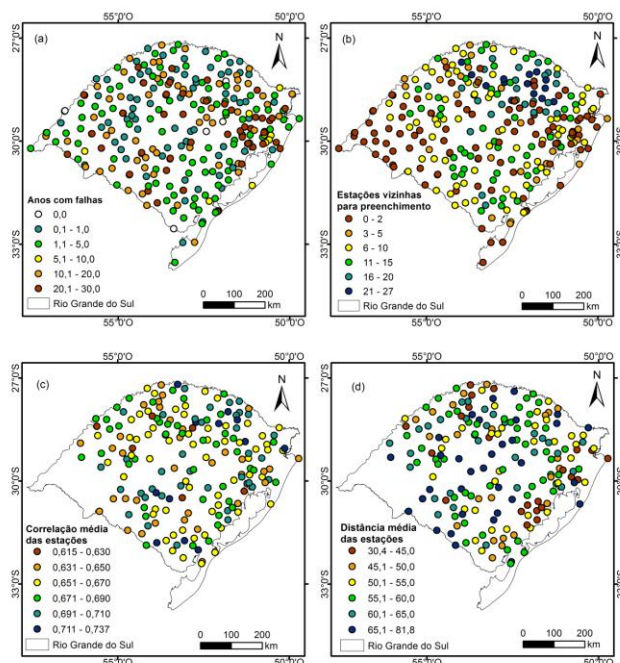


Figura 2. Distribuição espacial das estações pluviométricas no RS, conforme a quantidade de falhas (a), o número de estações vizinhas para preenchimento de falhas (b), a correlação média da precipitação (c) e a distância média (d) das estações vizinhas em relação a cada local.

Em relação às falhas, observa-se que apenas 2,4% das estações apresentaram série completa no período de 30 anos. De modo geral, a maior parte das estações (60,3%) apresentam menos de 5 anos de falhas em seus dados. Por outro lado, cabe ressaltar que 16,4% das estações apresentam menos de 10 anos de dados no período analisado, o que corrobora com a necessidade de preenchimento de falhas.

A análise da Figura 2b mostra que nas estações localizadas na porção nordeste do RS, em parte da serra e do planalto gaúcho, a quantidade de estações vizinhas para preenchimento dos dados é maior, indicando uma rede mais densa de estações. Já, no sudoeste (Campanha Gaúcha), existe um menor número de estações vizinhas que atendam aos critérios estabelecidos, fazendo com que não fosse possível realizar o preenchimento de falhas nessas estações. Por esse motivo, essas estações foram excluídas nos mapas de correlação e distância (2c e 2d).

Na Figura 3 é representado o desempenho dos modelos RM (3a) e RNA (3b), tendo como base o coeficiente de determinação (R^2) obtido entre a série preenchida e a série original, considerando o mesmo conjunto de validação por

estação. Na Figura 3a, a quantidade de estações que apresentam um R^2 superior a 0,6 chega a 53,6%, enquanto na 3b, 48,8%. O resultado do preenchimento de falhas foi mais satisfatório (R^2 acima de 0,8) em 15,3% das estações, considerando RM, e 10,1%, nos modelos baseados em RNA.

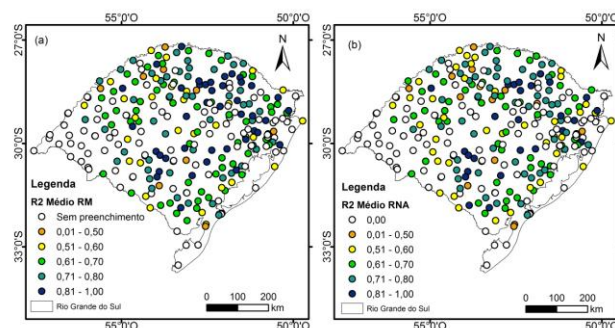


Figura 3. Desempenho dos modelos RM e RNA no processo de preenchimento de falhas nas estações pluviométricas do RS.

No contexto espacial, nota-se que o desempenho do preenchimento de falhas nos dados de precipitação é superior na porção nordeste do RS. Ainda que o relevo nessa área apresente 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 utilizadas no estudo, o preenchimento foi realizado em 195, em função das restrições selecionadas na metodologia. Observa-se que o desempenho do modelo RM foi superior às RNAs em 76% dos locais (Figura 4). Em relação aos valores médios de R^2 , o modelo de RM apresentou o valor de 0,697, enquanto a média dos valores de para o modelo RNA foi de 0,675. Demonstrando, também nesse aspecto, uma superioridade do método RM em relação ao método RNA.

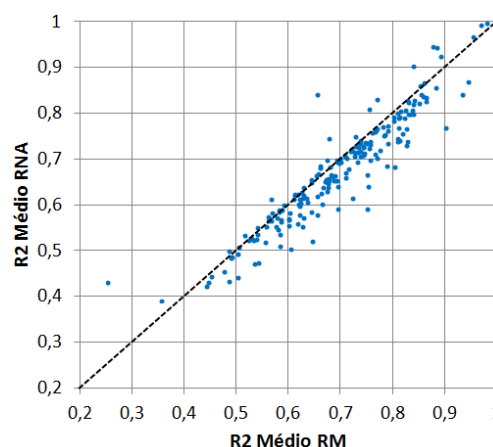


Figura 4. Dispersão entre o coeficiente de determinação R^2 obtido no preenchimento de falhas com RM e com RNA.

Considerando a estrutura das RNAs e sua alta capacidade para solucionar problemas complexos e não-lineares, o resultado obtido neste estudo foi surpreendente. Isso porque esperava-se que as RNAs apresentassem um valor preenchido de precipitação mais semelhante ao medido pela estação, ainda que isso resultasse em conjunto extenso de pesos sinápticos na rede.

O que aconteceu, entretanto, foi que a estrutura simplificada de parâmetros do modelo RM foi capaz de obter resultados mais semelhantes à chuva medida. A melhor explicação para isso está relacionada aos altos índices espaciais de correlação linear entre as estações, principalmente na região nordeste, indicando não ser necessário usar um modelo complexo para estimar valores de precipitação em um local a partir de estações vizinhas.

4. CONCLUSÕES

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

i) A maior densidade de estações pluviométricas na porção nordeste do RS favorece o processo de preenchimento de falhas em séries históricas de precipitação, melhorando a qualidade da série preenchida;

ii) O preenchimento de falhas em dados diários de precipitação apresentou maior coeficiente de determinação usando o modelo RM em relação às RNAs, 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.

5. 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.

6. REFERÊNCIAS

[1] Ichianagi, K.; Yamanaka, M.D.; Muraji, Y.; Vaidya, B.K. "Precipitation in Nepal between 1987 and 1996". *International journal of Climatology*. v. 27, pp. 1753-1762, 2018.

[2] Ballari, D.; Giraldo, R.; camposano, 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". *International journal of Climatology*. v. 38, pp. 3337-3354, 2018.

[3] Mekis, E.; Donaldson, N.; Reid, J.; zuconi, 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. pp. 1-25, 2018.

[4] 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. *Revista Brasileira de Geografia Física*. v.06, n.04, pp. 680-696, 2013.

[5] Fernandez, M.N. Preenchimento de Falhas em Séries Temporais. Dissertação (mestrado em Engenharia Oceânica). Fundação Federal Universidade de Rio Grande. pp. 106, 2007.

[6] 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: Congresso Brasileiro sobre Desastres Naturais, 1., Rio Claro. *Anais...* pp.1-10 (CDROM), 2012.

[7] Naghettini, M., Andrade Pinto, E.J. Hidrologia estatística. Serviço Geológico do Brasil. Belo Horizonte, pp. 552, 2007.

[8] Bertoni, J. C.; Tucci, C. E. M. Precipitação. In: TUCCI, C. E. M. Hidrologia: Ciência e aplicação. Porto Alegre: Ed. da Universidade UFRGS, ABRH, 2001.

[9] Riad, S; Mania, J; Bouchaou, L.; Najjar, Y. Rainfall-Runoff Model Using an Artificial Neural Network Approach. *Mathematical and Computer Modelling*. v.40, n.7-8, pp. 839-846, 2004.

[10] Gupta, N. Artificial neural network: Network and Complex Systems. v. 3, n. 1, pp. 24-28, 2013.

[11] Oliveira, G.G.; Pedrollo, O.C., Castro, N.M.R. O Desempenho das Redes Neurais Artificiais (RNAs) para Simulação Hidrológica Mensal. *Revista Brasileira de Recursos Hídricos*. v. 19, n.2 , pp. 251-265, 2014.

[12] Britto, F.; Barletta, R.; Mendonça, M. Regionalização sazonal e mensal da precipitação pluvial máxima no estado do Rio Grande do Sul. *Revista Brasileira de Climatologia*, v. 02, n. 2, pp. 35-51, 2006.

[13] Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, pp. 533-536, 1986.

[14] Widrow, B.; Hoff, M. E. Adaptive Switching Circuits. 4. 1960 IRE Wescon Convention Record, New York: v.4, pp. 96-104, 1960.