

COMPARATIVO DE REDES NEURAS ARTIFICIAIS E REGRESSÃO LINEAR MÚLTIPLA PARA PREENCHIMENTO DE FALHAS DE PRECIPITAÇÃO DIÁRIA EM REGIÃO SEMIÁRIDA

João Maria de Andrade¹
José Benito de Andrade Vieira²
Ulisses Alencar Bezerra³

RESUMO

A falta de informações e dados hidrográficos em uma bacia hidrográfica, principalmente dados de precipitação, podem causar enormes prejuízos na elaboração de projetos e dimensionamento de estruturas hidráulica, além de interferir na representação dos processos hidrológicos. Uma solução é o preenchimento destas falhas a partir de informações dos postos adjacentes. O objetivo deste estudo foi comparar a técnica de regressão linear múltipla (RLM) e redes neurais artificiais (RNA) para o preenchimento de falhas em séries diárias de precipitação na bacia hidrográfica do rio Piancó, na Paraíba. Os modelos propostos neste estudo foram desenvolvidos utilizando o *software RStudio*. Para se avaliar o seus desempenhos, foram utilizados os seguintes índices: coeficiente de eficiência de Nash e Sutcliffe (NSE), Raiz do Erro Quadrático Médio (RMSE) e coeficiente de correlação (r). Por meio dos resultados, observou-se que os modelos apresentaram baixo desempenho em decorrência dos fatores climatológicos da bacia e disposição dos postos pluviométricos. De forma geral, a rede neural apresentou uma maior eficiência nos preenchimentos das lacunas dos dados em relação à regressão linear múltipla.

Palavras-chave: Bacia do rio Piancó, Inteligência artificial, *Software* livre.

¹Graduado pelo Curso de Engenharia Civil da Universidade Federal de Campina Grande - UFCG, andradejmn@gmail.com

²Graduado pelo Curso de Engenharia Civil da Universidade Federal de Campina Grande - UFCG, josebenitoeng@gmail.com;

³Doutorando em Engenharia Civil e Ambiental, Universidade Federal de Campina Grande (UFCG) – Campina Grande (PB), Brasil. Fone: +55 83 3332-0243 E-mail: ulisses.alencar17@gmail.com;

INTRODUÇÃO

De acordo com Collischonn (2006), o ciclo hidrológico representa a dependência, o movimento e a troca da água nos seus diferentes comportamentos no meio ambiente, sendo formado por diversas componentes, sendo a precipitação a que apresenta maior variabilidade no espaço e no tempo. A disponibilidade e acessibilidade dos dados de precipitação é essencial para avaliar as condições hidrológicas de uma região (SUN et al., 2018). Longas lacunas de dados de precipitação podem ser críticas, produzindo resultados tendenciosos em suas descrições (SINGH; XIAOSHENG, 2019) e interferindo na representação de modelos hidrológicos e estatísticos para representação dos processos hidrológicos da bacia (DEPINÉ et al., 2014).

A falta de informações e dados hidrológicos em uma bacia hidrográfica, principalmente dados de precipitação, pode causar enormes prejuízos na elaboração de projetos e dimensionamento de estruturas hidráulicas, já que as estruturas podem ser sub/superdimensionadas. No Brasil os principais bancos de dados que contém dados pluviométricos são da Agência Nacional de Águas (ANA) e do Instituto Nacional de Meteorologia (INMET), entretanto, muitos postos pluviométricos apresentam interrupção de informações em suas séries temporais, seja por problemas com o operador do posto ou problemas técnicos no equipamento (BERTONI; TUCCI, 2013).

Uma solução é o preenchimento das falhas nos registros dos postos pluviométricos a partir de informações dos postos adjacentes situados na mesma zona climática. Esse processo requer um estudo de cada estação disponível, bem como sua correlação com as outras estações (BILENKI JUNIOR, 2018). A eficiência dos preenchimentos das falhas está condicionada à qualidade do modelo utilizado, a disponibilidade espacial de outros postos e fatores climatológicos da região (DEPINÉ et al., 2014).

Segundo Bertoni e Tucci (2013), a Ponderação Regional, Regressão Linear Simples ou Múltipla são alguns métodos para preenchimento de séries de precipitação. Além desses métodos, podem ser implementados algoritmos baseados em Redes Neurais Artificiais (HASAN, 2013; COUTINHO, 2018).

As redes neurais artificiais (RNAs) são métodos inseridos dentro do grupo da inteligência artificial, que tentam simular como o cérebro processa informações para pensar e raciocinar. Nos últimos anos, as RNAs têm sido amplamente utilizadas nas pesquisas relacionadas a recursos hídricos, principalmente nas que envolvem modelagem de fenômenos não lineares, como: modelos chuva-vazão (ABRAHART, 2003), previsão da vazão afluente em

reservatórios (MOHAMMADI; ESLAMI, 2005) e previsão de precipitação (RAMÍREZ; FERREIRA; VELHO, 2006).

O objetivo deste estudo foi comparar a técnica de regressão linear múltipla (RLM) e redes neurais artificiais para o preenchimento de falhas de séries diárias de precipitação. A pesquisa foi aplicada utilizando os dados dos postos pluviométricos disponibilizados na bacia hidrográfica do rio Piancó, na Paraíba, utilizando os dados diários de precipitação dos anos de 1966, 1967, 1972, 1973, 1976, 1978, 1979, 1982, 1987.

METODOLOGIA

Área de Estudo

O estudo foi desenvolvido na bacia hidrográfica do Rio Piancó (Figura 1), localizada no estado da Paraíba, na qual apresenta uma área total de 9.205 km² e é a principal sub-bacia da bacia hidrográfica do Rio Piancó-Piranhas-Açu. Esta bacia nasce no alto sertão paraibano, estendendo-se por todo o estado e pelo Rio Grande do Norte, desaguando no Oceano Atlântico. A região de estudo apresenta clima semiárido e valores anuais médios de precipitação e temperatura iguais a 821 mm e 24°C, respectivamente (LIMA, 2004).

Dados de Precipitação

Os dados de precipitação foram obtidos em sete postos pluviométricos localizados na bacia, adquiridos no portal Hidroweb (ANA, 2019). A escolha dos dados dos períodos utilizados para aplicação dos modelos para preenchimento de falhas foi conduzida conforme Lourenço (2012), optando-se por excluir anos que apresentassem falhas de leituras ou suspeita de leituras errôneas e/ou de problemas no sistema de banco de dados. Na Tabela 1 apresentam-se os postos e período de dados utilizados neste trabalho. Na Figura 1, há a localização dos postos da área de estudo com seus respectivos códigos.

Tabela 1. Postos utilizados no estudo.

Código	Nome do posto	Anos
737006 (P1)	Piancó	
738020 (P2)	Conceição	
738015 (P3)	Manaíra	1966, 1967, 1972, 1973, 1976, 1978,
738013 (P4)	Princesa Isabel	1979, 1982, 1987.
738012 (P5)	Boa Ventura	
738014 (P6)	Nova Olinda	
738018 (P7)	Ibiara	

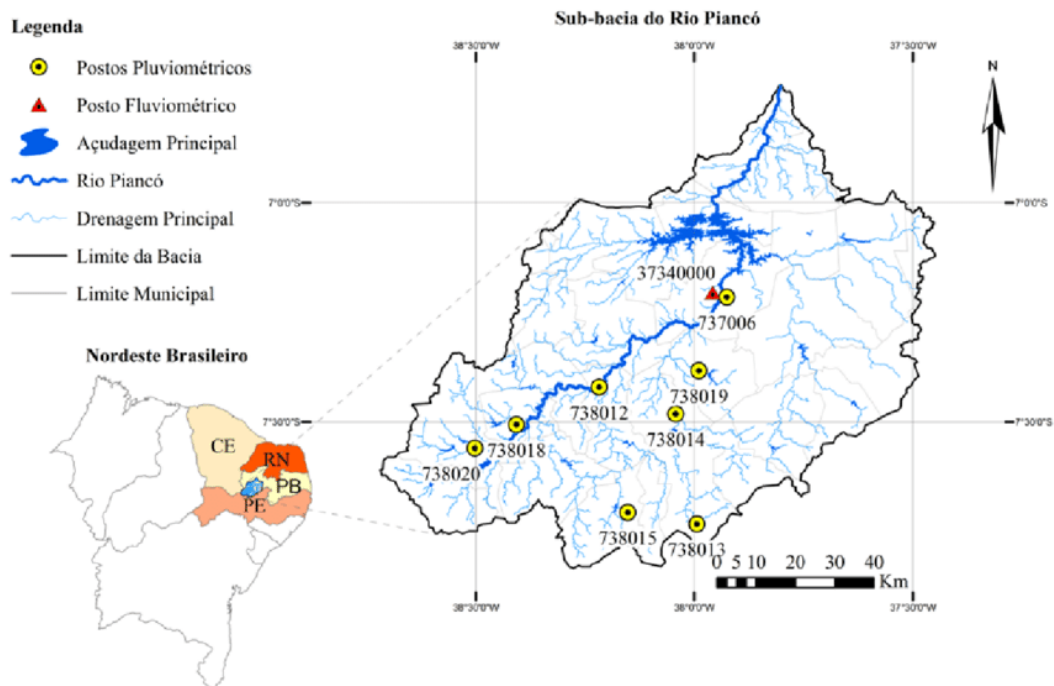


Figura 1. Localização dos postos utilizados na bacia do rio Piancó. (Fonte: LOURENÇO, 2014).

Modelo de Regressão Linear Múltipla

Na regressão linear múltipla as informações pluviométricas do posto dependente são correlacionadas com as correspondentes observações das estações vizinhas. Montgomery, Runger e Hubele (2010) explicam que o modelo de regressão linear múltipla utiliza mais de um regressor, ou seja, mínimo de duas variáveis independentes, ou seja, a variável resposta ou dependente pode estar relacionada a n variáveis independentes ou regressoras, sendo descrito pela equação:

$$Y = b_0 + b_1 \times X_1 + b_2 \times X_2 + b_n \times X_n + e \quad (1)$$

Sendo:

Y o valor estimado para a variável dependente;

b_0 a constante de regressão (interseção do plano);

b_1, b_2, b_n os coeficientes parciais de regressão;

X_1, X_2, X_n as variáveis independentes;

e o erro aleatório.

Logo, neste caso a estação contendo a falha é preenchida com o valor predito pelo modelo de regressão que utiliza as estações vizinhas como variáveis independentes.

Modelo Redes Neurais Artificiais

Nesta pesquisa optou-se por aplicar Rede Neural Perceptron de Múltiplas Camadas (MLP) com a utilização do algoritmo de aprendizado *Backpropagation* ou retropropagação de erro (RPROP), juntamente com as funções de ativação *Sigmoidais* para camada oculta e saída (VALENÇA e LUDERMIR; HAYKIN, 1999). Na Figura 2 mostra a estrutura do modelo de rede neural desse estudo. O número de neurônios na camada oculta N é obtido através de um procedimento baseado em tentativa e erro. Sendo assim utilizaram-se redes com 1x1; 1x2; 1x3; 1x4 ... 4x4 neurônios, na camada oculta. A estrutura com 2x1 foi a escolhida, por atingir a convergência no treinamento da rede mais rapidamente e por apresentar as melhores métricas de desempenho entre as redes utilizadas. Vale salientar que cada rede foi rodada 5 vezes.

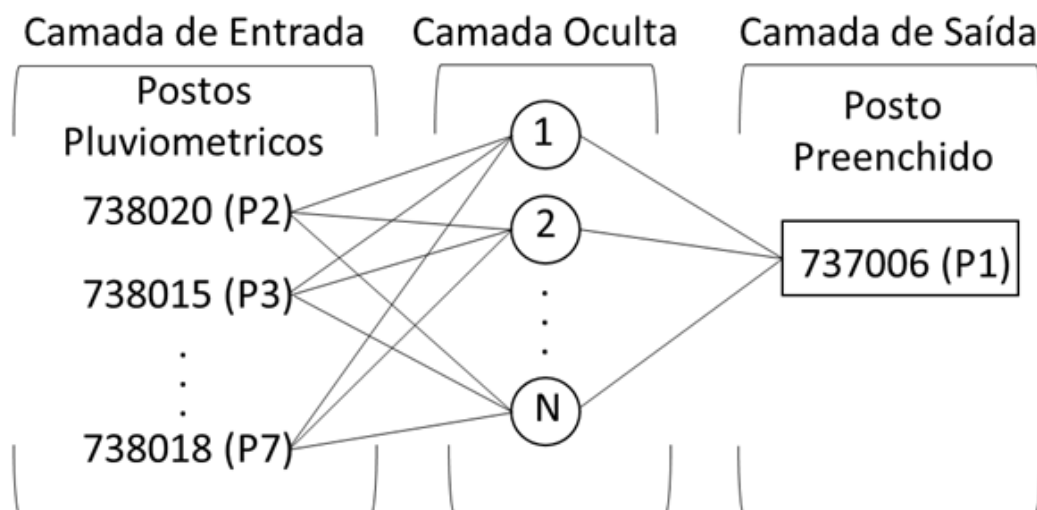


Figura 2: Estrutura da RNA para o preenchimento de falhas de precipitação.

Séries de treinamento e de verificação

A fase de treinamento e verificação dos modelos propostos foram definidos com base em 70% dos dados usados para treinamento e 30% para verificação da RLM e da RNA. Na etapa de verificação, os valores observados em campo do posto 737006 (P1) foram retirados, para que cada modelo pudesse realizar a predição e assim obter os valores estimados para o preenchimento, e posteriormente compará-los com os valores reais.

Índices de desempenho

Para se avaliar o desempenho dos modelos propostos, foram utilizados os seguintes índices: o coeficiente de eficiência de *Nash-Sutcliffe* (NSE) (Nash e Sutcliffe, 1970), Raiz do Erro Quadrático Médio (RMSE) (BARNSTON et al, 1992) e coeficiente de correlação (r) (FIGUEIREDO FILHO e SILVA JÚNIOR, 2009).

Softwares utilizados

Os métodos foram desenvolvidos utilizando o *software* R (R DEVELOPMENT CORE TEAM, 2019) com a interface *RStudio*. Na interface do *RStudio*, a função *lm* e o pacote *neuralnet* (GÜNTHER; FRITSCH, 2010) foram aplicados para o desenvolvimento dos modelos de RLM e RNA respectivamente.

RESULTADOS E DISCUSSÃO

Nas Figuras 3 e 4 são apresentadas as comparações entre os dados observados e estimados pelos modelos RLM e RNA, para o período de verificação.

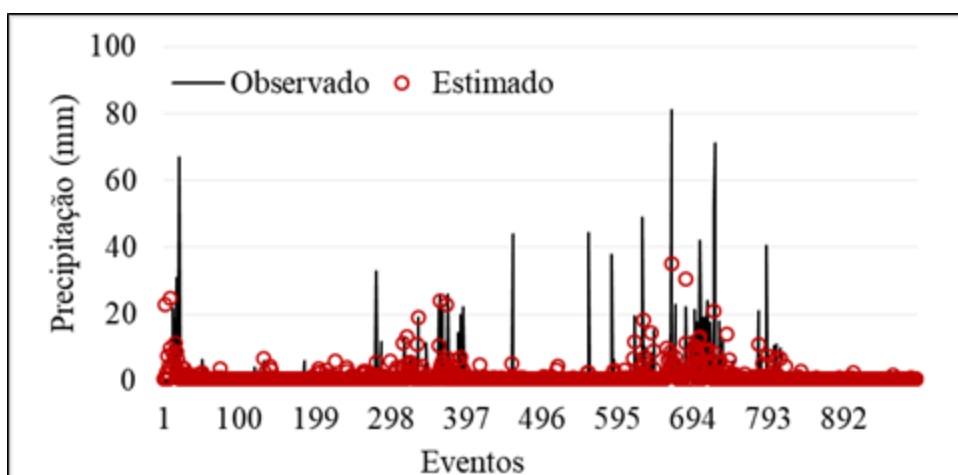


Figura 3. Comparação entre dados estimados pela RLM para o período de verificação.

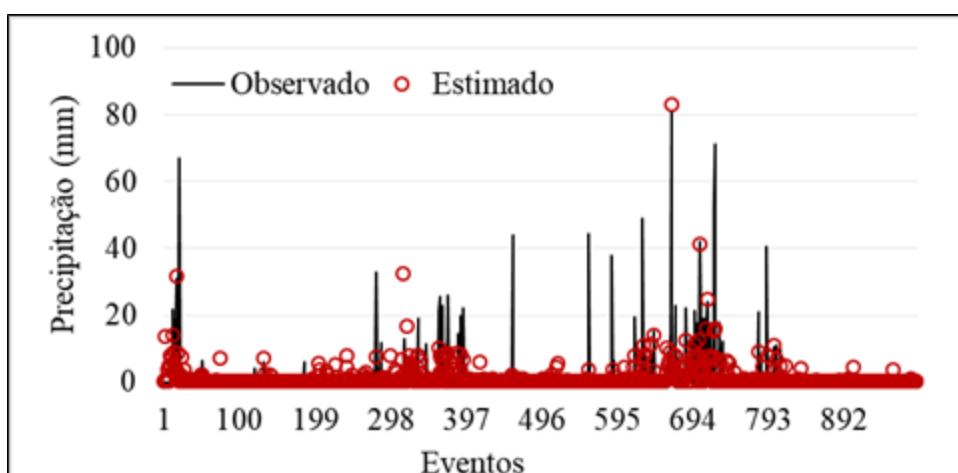


Figura 4. Comparação entre dados estimados pela RNA para o período de verificação.

A partir dos valores encontrados nas Figuras 3 e 4, foram calculados o NSE, RMSE e r do estudo. A seguir, a Tabela 2 apresenta os índices de desempenho analisados nesse estudo

para os modelos de regressão linear múltipla e redes neurais artificiais, para o período de treinamento e verificação dos modelos.

Tabela 2. Índices de desempenho de treinamento e verificação dos modelos RLM e RNA.

MODELO	NSE	RMSE	<i>r</i>
Treinamento			
RLM	0,22	9,00	0,47
RNA	0,30	8,54	0,55
Verificação			
RLM	0,26	5,72	0,51
RNA	0,36	5,31	0,60

Ao se analisar a Tabela 2, observa-se que a RNA apresenta desempenho um pouco superior a RLM. O coeficiente NSE apresentou valores inferiores a 0,36, ficando abaixo da faixa aceitável (0,36 e 0,75), segundo Collischonn (2001). Avaliando o RMSE foi observado que a RNA apresentou menor erro 5,31 em relação a RLM que apresentou 5,72. Os valores do *r* apresentaram-se entre 0,47 a 0,60 tanto para o período de treinamento como para o período de verificação. De acordo com Callegari-jaques et al (2003), valores do coeficiente de correlação entre 0,3 e 0,6 a correlação é dita como regular.

De modo geral, os resultados apresentam baixo desempenho, isso pode ser justificado pela baixa correlação entre os postos pluviométricos analisados, para os dados diários de precipitação, em função de sua distância e das características climáticas da área de estudo. A distribuição e organização da precipitação na região semiárida é de natureza convectiva e ocorre em áreas isoladas (BARBOSA et al, 2005), o que pode afetar o preenchimento de falhas para região. Em estudo realizado por Depiné et (2014), o autor observou que o tamanho da bacia, distribuição espacial de chuva, qualidade dos dados de entrada e da escolha e de configuração da rede a ser utilizada afeta eficiência das redes neurais artificiais. Assim, esses fatores associados interferiram no preenchimento de falhas dos postos pluviométrico da bacia hidrográfica.

Ainda assim, foi observado que o modelo baseado em RNA apresentou melhores índices de desempenho nos preenchimentos das lacunas dos dados em relação a regressão linear múltipla, tanto para o período de treinamento como para o período de verificação. Isso pode ser explicado devido à grande capacidade que as RNAs apresentam em lidar com incertezas, dados não linearidades e descontinuidades (HAYKIN et al, 2007).

CONCLUSÃO

Tendo como objetivo o preenchimento de falhas em séries diárias de precipitação na bacia hidrográfica do rio Piancó, foram desenvolvidos dois modelos, um baseado em regressão linear múltipla e outro baseado em rede neurais artificiais. Por meio dos resultados, observou-se que os modelos apresentaram baixo desempenho em decorrência dos fatores climatológicos da bacia e disposição dos postos pluviométricos.

Comparando as duas técnicas utilizadas, o modelo de rede neural apresentou uma eficiência um pouco maior nos preenchimentos das lacunas dos dados em relação à regressão linear múltipla. Fato que se deve pôr as RNA ter a capacidade de lidar com dados não lineares, principalmente para variáveis que possuem grande quantidade de dados, sendo o caso das séries diárias de precipitação e bacias hidrográficas com inúmeros postos pluviométricos.

Portanto a busca de melhores modelos para preenchimento de falhas e técnicas que possam lidar com grande quantidade de informações é de suma importância para que se possa ter um preenchimento de falhas mais próximo do real, já que essas informações subsidiam projetos que necessitam de séries climáticas consistentes e sem falhas.

AGRADECIMENTOS

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo incentivo a pesquisa e concessão de bolsas Pós-Graduação.

REFERÊNCIAS

ABRAHART, R. J. Neural network rainfall-runoff forecasting based on continuous resampling. **Journal of Hydroinformatics**, 2003.

BARBOSA, T. F.; CORREIA, M. F. Sistemas convectivos intensos no semi-árido brasileiro: O controle da grande escala. **Revista Brasileira de Meteorologia**, v. 20, n. 3, p. 395-410, 2005.

BARNSTON, A. G. Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score. **Weather and Forecasting**, v. 7, n. 4, p. 699-709, 1992.

BERTONI, J. C.; TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. **Hidrologia: ciência e aplicação**. 4. ed. Porto Alegre: Ed. Universidade/UFRGS/ABRH, 2013. 943 p.

BIELENKI JUNIOR, C., SANTOS, F. M. D., POVINELLI, S. C. S., & MAUAD, F. F. Alternative methodology to gap filling for generation of monthly rainfall series with GIS approach. **Revista Brasileira de Recursos Hídricos**, v. 23, 2018.

CALLEGARI-JACQUES, S.M. 2003. Bioestatística: princípios e aplicações. Porto Alegre, **Artmed**. 255 p.

COLLISCHONN, W. Simulação hidrológica em grandes bacias. Tese de doutorado. **Instituto de Pesquisas Hidráulicas**. UFRGS.2001.

COLLISCHONN, B. **Uso de precipitação estimada pelo satélite TRMM em modelo hidrológico distribuído**. Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental. Mestrado em Recursos Hídricos e Saneamento Ambiental. Porto Alegre -RS. 2006. 131p.

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, p. 317-328, 2018.

DEPINÉ, H., CASTRO, N. D. R., PINHEIRO, A., & PEDROLLO, O 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.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. da. Desvendando os mistérios do coeficiente de correlação de Pearson (r). **Revista Política Hoje**, v. 18, n. 1, p. 115-46. 2009.

GÜNTHER, F; FRITSCH, S. neuralnet: Training of neural networks. **The R journal**, v. 2, n. 1, p. 30-38, 2010.

HAYKIN S. **Neural networks: a comprehensive foundation**. 2nd ed. Upper Saddle River: Prentice Hall, 1999.

HAYKIN, S. **Redes neurais artificiais: princípios e prática**. 2. ed. Bookman, 2007.

HASAN, M.M. e CROKE, B.F.W. Filling gaps in daily rainfall data: a statistical approach. 20th **International Congress on Modelling and Simulation**, Adelaide, Australia, n. December, p. 1–6, 2013. Disponível em: <<http://www.mssanz.org.au/modsim2013/A9/hasan.pdf>>.

LIMA, C. A. G. Análise e sugestões para diretrizes de uso das disponibilidades hídricas superficiais da bacia hidrográfica do rio Piancó, situada no estado da Paraíba. Tese de Doutorado, **Universidade Federal de Campina Grande**, Campina Grande, Brasil, 2004.

LOURENÇO, A. M. G. **Modelos Chuva-Vazão baseados em Redes Neurais Artificiais para Rios Intermitentes no Semiárido Paraibano**. Monografia – Universidade Federal de Campina Grande, 46 p. 2012.

LOURENÇO, A. M. G. et al. Análise e modelagem hidrológica no ambiente R. Anais Do XII Simpósio de Recursos Hídricos Do Nordeste, (April), 1–10, 2014.

MOHAMMADI, K.; ESLAMI, H. R. Comparison of regression, ARIMA and ANN models for reservoir inflow forecasting using snowmelt equivalent (a case study of Karaj). **J. Agric. Sci. Technol**, v. 7, p. 17-30, 2005.

MONTGOMERY, D. C., RUNGER, G. C., & HUBELE, N. F. (2009). **Engineering statistics**. John Wiley & Sons.

NASH, J. E., SUTCLIFFE, J. V., River flow forecasting through conceptual models part 1 – a discussion of principles. **Journal of Hydrology** 10, 282–290, 1970.

R Development Core Team. (2018). *R: "A Language and Environment for Statistical Computing"*. Acesso em 20 de Fevereiro de 2019, disponível em R Foundation for Statistical Computing.: <https://www.r-project.org/>

RAMÍREZ, M. C.; FERREIRA, N. J.; VELHO, H. F. C. Linear and Nonlinear Statistical Downscaling for Rainfall Forecasting over Southeastern Brazil. **Weather and Forecasting**, v. 21, n. 6, p. 969–989, dez. 2006.

SINGH, V.; XIAOSHENG, Q. Data assimilation for constructing long-term gridded daily rainfall time series over Southeast Asia. **Climate Dynamics**, p. 1-25, 2019.

SUN, Q., MIAO, C., DUAN, Q., ASHOURI, H., SOROOSHIAN, S., & HSU, K. L. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. **Reviews of Geophysics**, v. 56, n. 1, p. 79-107, 2018.

VALENÇA, M. J. S., LUDERMIR, T. B. Introdução às redes neurais. In: GALVÃO, C. O. et al. *Sistemas Inteligentes*. Porto Alegre: **Editora da Universidade Federal do Rio Grande do Sul**, 1999. p.19-57.