


Artigo

Previsão Sazonal de Vazões para a Bacia do Orós (Ceará, Brasil) Utilizando Redes Neurais e a Técnica De Reamostragem dos K-vizinhos

Carla Beatriz Costa de Araújo¹ , Francisco de Assis de Souza Filho²,
Luiz Martins de Araújo Júnior³, Cleiton da Silva Silveira²

¹*Departamento de Geologia, Universidade Federal do Ceará, Fortaleza, CE, Brasil.*

²*Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, CE, Brasil.*

³*Instituto de Engenharias e Desenvolvimento Sustentável, Universidade da Integração Internacional da Lusofonia Afro-Brasileira, Redenção, CE, Brasil.*

Recebido em: 25 de Novembro de 2018 - Revisado em: 27 de Fevereiro de 2019 - Aceito em: 28 de Maio de 2019

Resumo

Este trabalho tem por objetivo realizar um comparativo de previsão de vazões para a bacia do Orós (Ceará, Brasil) utilizando redes neurais artificiais (RNA) e a técnica de reamostragem dos k-vizinhos. Os modelos foram desenvolvidos a partir da série histórica de 100 anos de dados hidrometeorológicos (temperatura da superfície do mar e vazões). Ambos utilizam como preditores climáticos as temperaturas dos oceanos Atlântico Norte, Atlântico Sul e Pacífico Equatorial e realizam a previsão em julho do regime de vazões do período chuvoso do ano seguinte (janeiro a junho). O modelo k-vizinhos foi elaborado a partir da identificação dos anos vizinhos mais próximos para reamostragem da aproximação, já o modelo de RNA foi formulado a partir dos pesos sinápticos e bias obtidos na etapa de treinamento da rede. Os modelos foram comparados considerando as previsões de vazões para a etapa de teste, utilizou-se como parâmetros comparativos: o coeficiente de eficiência Nash-Sutcliffe (E), o coeficiente de determinação (R^2), o diagrama de Taylor (2001) e a razão de máxima verossimilhança. Para todas as variáveis comparativas o modelo neuronal apresentou melhores valores, indicando que este representa de forma mais eficiente o comportamento das vazões para o reservatório.

Palavras-chave: modelos, previsão de vazões, açude Orós.

Seasonal Flow Forecast for the Orós Dam (Ceará, Brazil) Using Neural Networks and the Resampling Technique of K-neighbors

Abstract

The objective of this work is to perform a comparative flow forecast for the Orós basin (Ceará, Brazil) using artificial neural networks (RNA) and k-neighbors re-sampling technique. The models were developed from the historical series of 100 years of hydrometeorological data (sea surface temperature and flows). Both use as predictors the temperatures of the North Atlantic, South Atlantic and Equatorial Pacific oceans, and forecast July in the next year's rainy season (January to June). The k-neighbors model was elaborated from the identification of the closest neighbor years for the resampling of the approximation, since the RNA model was formulated from the synaptic and bias weights obtained in the training phase of the network. The Nash-Sutcliffe (E) efficiency coefficient, the coefficient of determination (R^2), the Taylor diagram (2001) and the coefficient of determination (R^2) were used for the validation step. maximum likelihood ratio. For all comparative variables, the neural model presented better values, indicating that this represents more efficiently the behavior of the flows to the reservoir.

Keywords: models, streamflow forecasting, Oros reservoir.

1. Introdução

Modelos de previsão sazonal de vazões que utilizam preditores climáticos podem ser utilizados como ferramenta para auxiliar as tomadas de decisões na gestão de recursos hídricos e no gerenciamento de desastres hidrometeorológicos (secas e cheias). Torna-se então relevante o desenvolvimento e o aprimoramento de metodologias para previsões de aflúências de longo alcance que podem ser usadas para operação de reservatórios e como ferramentas que auxiliem no processo de alocação de água para demandas concorrentes, cujas previsões e incertezas devem ser consideradas na operação e planejamento do sistema (Souza Filho e Lall, 2003).

A previsão de vazões é um problema multivariado e de grande complexidade, existindo diferentes metodologias para modelagem do problema. As abordagens para previsão de vazões são divididas predominantemente em duas categorias, sendo estas: métodos estatísticos e métodos dinâmicos (integração dos modelos climático e hidrológico). Os métodos estatísticos se baseiam nos registros históricos de observação dos preditores (ex.: vazão, precipitação) e preditores climáticos relevantes (ex.: temperatura da superfície do mar, pressão atmosférica), ou um índice relacionado diretamente com a estimativa de vazões utilizando técnicas estatísticas.

No Brasil, segundo Alexandre (2012), os modelos estatísticos de previsão de aflúências têm sido uma ferramenta importante no suporte à tomada de decisões, no planejamento e na gestão de recursos hídricos aplicados ao Sistema Interligado Nacional (SIN). Silveira *et al.* (2017) citam que modelos estatísticos foram propostos em diversos estudos, comprovando que o uso de índices climatológicos como variáveis explicativas em modelos matemáticos proporciona capacidade de explicar regimes de vazões interanuais, destacando os trabalhos de Uvo e Graham (1998), Souza Filho e Lall (2003, 2004), Silva e Molion (2004), Rohn *et al.* (2003) e Rocha *et al.* (2007).

Silveira *et al.* (2017) citam que estudos como os de Souza Filho e Lall (2003) e Foley *et al.* (2002) são realizados desde a década de 1990 e apontam que o uso de variáveis climáticas pode reduzir incertezas devido à alta correlação dessas variáveis com variáveis hidrológicas. No Brasil, os trabalhos de Carriello *et al.* (2005), Soares *et al.* (2006) e Kim e Vissotto (2003) analisaram a relação entre a temperatura da superfície do mar (TSM) e a hidrografia do país. Em termos de bacias regionais específicas, os trabalhos de Cardoso e Silva Dias (2006) e Berri *et al.* (2002), foram implementados em bacias no sudeste brasileiro; Foley *et al.* (2002) e Barros *et al.* (2004) analisaram bacias da Região Norte do Brasil que mostraram a boa capacidade representativa das variações da TSM (temperatura da superfície do mar) associadas à variabilidade das vazões no Brasil. Souza Filho e Lall (2003) realizaram uma investigação da dependência da vazão com relação à

TSM com diferentes defasagens para a Região Nordeste do Brasil, especificamente para bacias do Estado do Ceará, e obtiveram correlações que chegaram a valores de 0,9, realizando previsões para cenários com até 18 meses de antecedência.

Além do uso de métodos estatísticos, nos últimos anos tem sido desenvolvido modelos de previsão de vazões utilizando redes neurais artificiais (RNA). Zealand *et al.* (1999) exploram a capacidade das RNA e compara o desempenho desta ferramenta com outras abordagens utilizadas para previsão de vazões, neste trabalho os resultados obtidos mostraram que as RNA superaram os modelos convencionais.

Coulibaly *et al.* (2000) treinaram uma RNA, que faz a predição de um passo à frente da série temporal de múltiplas camadas para previsão em tempo real de aflúências no reservatório de Quebec. O desempenho do modelo utilizado é comparado com modelos estatísticos conceituais, os resultados mostraram que o método é eficaz na precisão da previsão, servindo como uma alternativa quando se deseja uma previsão dinâmica adaptável.

Podem ser citados ainda os trabalhos de Uvo *et al.* (2000), Wang *et al.* (2006), Menezes Junior e Barreto (2008), Gomes *et al.* (2010) e Araújo *et al.* (2015) que utilizam RNA para previsão de aflúências.

As RNA extraem seu poder de aprendizagem de um determinado fenômeno primeiramente através de sua estrutura maciçamente paralelamente distribuída, e segundo de sua habilidade de reproduzir padrões para os quais não foi posta em contato durante o processo de aprendizagem (Haykin, 2001). Esta capacidade de generalização é talvez, uma das mais importantes características das RNA, e uma das maiores vantagens da sua utilização de fenômenos complexos, multivariados, e não-lineares, quando comparado aos métodos de modelagem baseados em regressões estatísticas de séries históricas. Os trabalhos de Araújo *et al.* (2015), Araújo (2015), Dantas Neto *et al.* (2014) e Dantas Neto *et al.* (2016) podem ser consultados para entender o funcionamento e o desenvolvimento da formulação para as redes neurais artificiais do tipo perceptron multicamadas.

Neste trabalho é feito o comparativo de previsão de vazões para a bacia do reservatório Orós (Ceará, Brasil) entre um modelo de utilizando RNA, que foi desenvolvido a partir do melhoramento do modelo de Araújo *et al.* (2015), e o método estatístico de Souza Filho e Lall (2003) que usa uma técnica chamada k-vizinhos para reamostrarem dos possíveis cenários de vazões que mantém a consistência de espaço-tempo em diferentes locais e subperíodos.

2. Materiais e Métodos

Para desenvolvimento da pesquisa, inicialmente foi escolhida a região de estudo com base na importância do

local para o cenário de recursos hídricos no Estado do Ceará e feito o levantamento de dados da série histórica de vazões e dos preditores climáticos (temperaturas da superfície do mar). Em seguida, calibrou-se os dois modelos: RNA e utilizando a técnica de reamostragem k-vizinhos. Por fim, foi realizada uma análise comparativa dos modelos com os dados de teste, utilizando as seguintes variáveis: o coeficiente de eficiência Nash-Sutcliffe (E), o coeficiente de determinação (R^2), o diagrama de Taylor (2001) e a razão de máxima verossimilhança.

2.1. Região de estudo

O local escolhido para desenvolver os modelos e realizar o comparativo foi a barragem do Orós, que faz parte da Bacia do Alto Jaguaribe, localizada no leito do Rio Jaguaribe na cidade de Orós, Estado do Ceará (Fig. 1), com capacidade máxima de armazenamento de 2.100.000.000 m³. Esta tem como finalidades: perenização do rio Jaguaribe, irrigação do Médio e Baixo Jaguaribe, piscicultura, culturas agrícolas de áreas de montante, turismo e abastecimento da Região Metropolitana de Fortaleza - RMF. Até que fosse construído o açude Castanhão, o reservatório Orós era o maior reservatório em capacidade de acumulação de água do estado do Ceará.

2.2. Levantamento dos dados

Inicialmente foi feita a coleta de dados de temperatura da superfície do mar (considerando o período de 6 meses antes a estação chuvosa) e vazões médias anuais. Para o modelo utilizando as RNA, foram utilizados os seguintes preditores:

- Nino3: TSM Leste do Pacífico Tropical (5N-5S;150W-90W)
- TAS: Índice Atlântico Sul Tropical - Anomalia da TSM (0-20S;10E-30W)

- TAN: Índice Atlântico Norte Tropical - Anomalia da TSM (5.5N-23.5N;15W-57.5W)
- PDO: Oscilação Decadal do Pacífico - Anomalia TSM Oceano Pacífico Norte
- AMO: Oscilação Atlântica Multidecadal

Os preditores climáticos Nino3, TAS e TAN foram derivados da malha de dados de TSM desenvolvidos por Kaplan *et al.* (1998) e encontram-se disponíveis em: <http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED>. Os índices climáticos PDO e AMO foram

determinados a partir do sítio do *Earth System Research Laboratory* (ESRL) do *National Oceanic and Atmospheric Administration* (NOAA) em <http://www.esrl.noaa.gov/psd/data/climateindices/list>.

No modelo k-vizinhos foram utilizados somente os preditores: Nino3, TAS e TAN, pois a incorporação dos novos índices (PDO e AMO) não apresentou resultados satisfatórios para a previsão de vazões.

Os dados de vazões anuais medidas foram obtidos nos postos fluviométricos da região do Orós, disponível em: <http://hidroweb.ana.gov.br/>, e através da modelagem chuva-vazão. A série histórica utilizada possui 100 anos, dados estes colhidos entre os anos de 1912 e 2011. O período de 1977 a 2011 possui dados observados de vazões obtidas em postos fluviométricos, para os outros anos, o modelo hidrológico adotado foi o SMAP (*Soil Moisture Accounting Produce*) utilizando a regionalização dos parâmetros para bacias hidrográficas do Ceará desenvolvida por Barros *et al.* (2013).

Dos dados entrada-saída disponíveis, foram separados aleatoriamente um percentual de 85% para o treinamento/calibração (média de 207 m³/s e desvio padrão de 244 m³/s) e 15% para a fase de teste (média de 144 m³/s e desvio padrão de 124 m³/s) de ambos os modelos. Os anos utilizados para a fase de teste do modelo foram: 1923, 1933, 1938, 1939, 1941, 1945, 1948, 1951, 1958, 1960, 1982, 1989, 1996, 2000 e 2006.

Figura 1 - Localização do reservatório Orós (Ceará, Brasil). Fonte: Adaptado de Ferreira *et al.* (2015).

2.3. Modelo redes neurais artificiais

O modelo de RNA deste trabalho é um melhoramento do modelo de Araújo *et al.* (2015), que tem por objetivo fazer a previsão da vazão anual afluyente ao açude Orós utilizando as anomalias de temperatura das superfícies do Oceano Atlântico Sul, Oceano Atlântico Norte e Oceano Pacífico do ano anterior a previsão, podendo ser representado de forma simplificada pela Eq. (1):

$$Q = f(TAS, TAN, NINO3) \quad (1)$$

Segundo Araújo *et al.* (2015), o modelo que forneceu os melhores resultados foi o que utiliza como dados de entrada a média das anomalias de temperatura dos meses de abril, maio e junho (AMJ) do ano anterior ao da previsão, para o novo modelo foram empregados os mesmos meses, porém, foram inseridos mais 12 anos da série histórica de TSM e vazões e incluídos mais dois preditores climáticos (PDO e AMO), sendo então representado pela Eq. (2):

$$Q = f(TAS, TAN, NINO3, PDO, AMO) \quad (2)$$

A primeira etapa do desenvolvimento do modelo de RNA é a normalização dos dados, uma vez que, a função de ativação (função sigmoide) utilizada para os cálculos dos sinais dos neurônios da camada de saída fornece resultados no intervalo (0, 1). Os valores dos dados de entrada e saída foram normalizados entre 0,15 e 0,85, considerando-se uma relação linear entre os valores de máximos e mínimos de cada uma das variáveis envolvidas (anomalias de temperaturas e vazões), conforme apresentado na Tabela 1.

Conforme descrito por Araújo *et al.* (2015), a definição do modelo neuronal consiste na definição da arquitetura da rede (definição do número de camadas de neurônios e do número de neurônios em cada camada), da escolha da função de ativação dos neurônios, e definição dos pesos sinápticos e bias que melhor representem o comportamento do fenômeno estudado. Sendo desenvolvidas em duas etapas: etapa de treinamento e etapa de teste.

Para definir o modelo apresentado neste artigo, foram testadas inúmeras arquiteturas de rede, porém, mais adiante apresenta-se apenas os resultados do modelo que apresentou melhor desempenho nas fases de treinamento e teste, definido com base no coeficiente de determinação

(R^2). Em todas as arquiteturas consideradas foi utilizada a função sigmoide para a ativação dos neurônios, devido à comprovada eficiência desta função para a modelagem de fenômenos complexos com as redes neurais artificiais apresentado nos trabalhos de Araújo *et al.* (2015), Araújo (2015), Dantas Neto *et al.* (2014) e Dantas Neto *et al.* (2016).

Após o estabelecimento do modelo (pesos e limiares) foi realizada a análise dos resíduos (erros) gerados na etapa de treinamento. O objetivo foi verificar se os resíduos do modelo se comportam segundo alguma distribuição de probabilidade, para então definir um intervalo de confiança, considerando-se um dado nível de confiança. Os resíduos foram calculados conforme Eq. (3):

$$d = Q_{observado} - Q_{modelado} \quad (3)$$

onde d = resíduo; $Q_{observado}$ = vazão observado na série histórica; $Q_{modelado}$ = vazão gerada pelo modelo de previsão.

A resposta final do modelo da RNA, considerando a incerteza deste, pode ser dada pela seguinte equação:

$$Q_{RNA} = Q_{modelado} + \varepsilon \quad (4)$$

onde: ε = erro considerando uma distribuição de probabilidade; $Q_{modelado}$ = vazão gerada pelo modelo de previsão em função dos preditores climáticos; Q_{RNA} = resposta final do modelo.

Com isto, incorpora-se a resposta final o erro intrínseco ao modelo, que depende da acurácia e da dispersão deste. Podendo-se então, trabalhar não apenas com um único valor referente a saída da rede, mas com um intervalo de valores, dado um nível de confiança e com base na distribuição de probabilidade do erro.

2.4. Modelo k-vizinhos

Nesse modelo foi utilizado uma abordagem semi-paramétrica, semelhante a utilizada em Souza Filho e Lall (2003). Esse método usa a informação de preditores climáticos, previamente conhecidos, que possuem uma boa correlação com a vazão sobre o local de estudo. Dados de vazões anuais são considerados a fim de se identificar os anos vizinhos mais próximos para reamostragem da aproximação que mantém a consistência de espaço e tempo (Silveira, 2014).

O esquema de reamostragem condicional se baseia em uma regressão linear múltipla, em que a vazão média de 12 meses é normalizada a fim de assegurar que o resíduo dessa regressão tenha um padrão de distribuição e uma variância aproximadamente constante. A normalização é feita pela diminuição do valor da precipitação anual pela sua média e dividida pelo desvio padrão da mesma. A matriz dos preditores climáticos é uma matriz da média

Tabela 1 - Valores máximos e mínimos das variáveis envolvidas na modelagem da previsão de vazões com utilização das redes neurais artificiais.

Variável/ valor	Entrada Anomalia de temperatura da superfície do mar (°C)	Saída Vazão (m ³ / s)
Máximo	2,38	1247,05
Mínimo	-1,87	9,64

sazonal das anomalias de temperatura do mar para AMJ, já citadas, para igual período de tempo da variável reduzida defasadas de um ano. Através dessa regressão os pesos de cada preditor climático são obtidos identificando a contribuição de cada preditor na geração da vazão.

Entretanto, a regressão não foi usada de maneira direta na previsão, sendo a mesma usada somente na identificação do peso do cálculo da distância entre a condição do ano de previsão (vetor dos preditores anuais x^*) e dos anos da série histórica (valor dos preditores em sua série histórica x_i). O cálculo da distância Euclidiana (d_i) é determinado por:

$$d_i = \sum_{j=1}^p \left\{ (x_j^* - x_{i,j}) \gamma_j \right\}^2 \quad (5)$$

em que x^* = vetor $1 \times p$; x_i = vetor $1 \times p$ de preditores durante o ano i -ésimo usado no modelo ajustado; γ = resultado da regressão, é um vetor $p \times 1$ dos coeficientes.

Com o vetor distância calculado tem-se a informação dos anos vizinhos mais próximos, a partir daí faz-se um ordenamento desses valores, de maneira crescente, sendo então a reamostragem baseada no número (k) de vizinhos arbitrados, que neste caso são 30, ou seja, as estatísticas das vazões previstas serão baseadas nos 30 anos mais próximos. O ajuste dos coeficientes γ foi feito utilizando os 85 anos separados aleatoriamente para calibração do modelo. Na etapa de teste, as vazões modeladas foram calculadas considerando os 30 anos mais próximos presentes na série de calibração.

2.5. Avaliação de desempenho dos modelos

Em busca de identificar os modelos que melhor representam a previsão de vazões para o reservatório Orós, foi avaliado o desempenho de cada um dos modelos utilizando as seguintes variáveis comparativas: o coeficiente de eficiência Nash-Sutcliffe (E), o coeficiente de determinação (R^2), o diagrama de Taylor (2001) e a razão de máxima verossimilhança.

O coeficiente de eficiência Nash-Sutcliffe é usado para avaliar o poder preditivo de modelos hidrológicos e foi calculado conforme a Eq. (6) (Nash e Sutcliffe, 1970):

$$E = \frac{\sum (Q_o - Q_m)^2}{\sum (Q_o - \bar{Q}_o)^2} \quad (6)$$

onde: Q_o = vazão observada, \bar{Q}_o = vazão média observada, Q_m = vazão calculada pelo modelo.

Os valores de eficiência de Nash-Sutcliffe podem variar de $-\infty$ a 1. Uma eficiência de 1 ($E = 1$) corresponde a um “modelo perfeito”, onde os valores modelados correspondem igualmente aos dados observados. Uma eficiência de 0 ($E = 0$) indica que as previsões do modelo são tão precisas quanto a média dos dados observados, já uma

eficiência negativa ($E < 0$) sinaliza que a média observada funciona como um preditor melhor do que o modelo.

Em diversos processos de modelagem o coeficiente de determinação (R^2) é utilizado como avaliação de desempenho do modelo. A obtenção do valor de R^2 pode ser feita através da Eq. (7).

$$R^2 = \left(\frac{Cov(X, Y)}{DP(X)DP(Y)} \right)^2 \quad (7)$$

onde: $Cov(X, Y)$ = covariância entre duas variáveis aleatórias X e Y ; $DP(X)$ = desvio padrão de X ; $DP(Y)$ = desvio padrão de Y . Sabe-se que os valores de R^2 variam entre 0 e 1 e que quanto maior for R^2 , mais explicativo é o modelo, melhor ele se ajusta à amostra.

O diagrama de Taylor é uma ferramenta gráfica que permite sejam apresentados em um mesmo gráfico com um único ponto três métricas de avaliação: coeficiente de correlação (R), desvio padrão e raiz do erro médio quadrático ($RMSE$) (Eq. (8)). Segundo Pereira *et al.* (2014), além de definir graficamente os melhores métodos de interpolação, essa ferramenta permite fazer escolhas entre eles, dentro de um conjunto menor e mais definido de informações.

$$RMSE = \left[\frac{1}{N} \sum (Q_m - Q_o)^2 \right]^{1/2} \quad (8)$$

onde: Q_o = vazão observada, Q_m = vazão calculada pelo modelo e N = número de dados observados.

Conforme Pereira *et al.* (2014), a construção do diagrama de Taylor é feita através da representação de $\frac{1}{2}$ ou $\frac{1}{4}$ de círculo. Os eixos da abscissa (x) e da ordenada (y) têm as medidas do desvio-padrão, sendo sobre o eixo x colocado o valor do desvio padrão dos dados observados. A distância radial da origem à posição representando o modelo é o desvio padrão dos dados gerados pelo modelo e o azimute da origem ao modelo é proporcional a R . Ressalta-se ainda que a representação gráfica deste diagrama permite que sejam comparados diversos modelos em uma mesma imagem (Taylor, 2001).

As variáveis anteriormente apresentadas para avaliação do desempenho, consideram apenas resultados determinísticos, não levando em conta que a previsão de vazões é processo aleatório, podendo ser representado por alguma distribuição de probabilidade. Deve-se ainda considerar que as previsões utilizadas para o gerenciamento dos recursos hídricos estão associadas a um nível de confiança e a um risco de falha, não sendo feito uso de um valor puramente determinístico. Uma forma de avaliar o desempenho dos modelos considerando as distribuições de probabilidade é a razão de máxima verossimilhança.

O princípio da verossimilhança afirma que: “Uma inferência estatística deve ser consistente com a hipótese

de que a melhor explicação de um conjunto de dados é fornecida por um valor de que maximiza a função de verossimilhança (Lindgren, 1976)". Segundo Silveira (2014), intuitivamente, maximizar a verossimilhança significa obter a população com a maior probabilidade de ter gerado a amostra.

Para determinar a razão de máximo verossimilhança é necessário definir a função de verossimilhança para as observações y_1, \dots, y_n , correspondentes às variáveis Y_1, \dots, Y_n , que é dada por:

$$L(\theta) = f(y_1, \dots, y_n; \theta) \quad (9)$$

onde f é uma densidade de probabilidade conjunta do vetor das observações (y_1, \dots, y_n) e θ é um vetor de parâmetros a ser estimado.

Se Y é a variável resposta, partindo da hipótese de que as observações são mutuamente independentes e identicamente distribuídas, a função L pode ser escrita como (Silveira, 2014):

$$L(\theta; y_1, \dots, y_n) = f(y_1; \theta) f(y_2; \theta) \dots f(y_n; \theta) \quad (10)$$

No caso dos modelos deste trabalho, o vetor de parâmetros de interesse é:

$$\theta = (\beta, \sigma^2) \quad (11)$$

onde β é o valor esperado e σ e o desvio padrão.

E definindo-se L como função de β e σ^2 tem-se:

$$L(\beta, \sigma^2) = \prod_{i=1}^n f_{y_i}(y_i; \beta, \sigma^2) \quad (12)$$

Conforme apresentado por Silveira (2014), pode-se então calcular do desempenho na previsão, utilizando essa metodologia com a climatologia em comparação aos dados observados totais, é realizado conforme a Eq. (13):

$$Desempenho = \left(\frac{L(\beta, \sigma^2)_{modelo}}{L(\beta, \sigma^2)_{climatologia}} \right)^{1/n} \quad (13)$$

onde n é o número de anos da série histórica utilizada. Quando Desempenho > 1 , significa que houve uma melhora na previsão em relação à climatologia. Ao contrário, quando Desempenho < 1 , significa que houve uma piora na previsão.

3. Resultados e Discussão

Dentre as arquiteturas testadas para o modelo de RNA, a que utiliza 5 nós de entrada, 13 neurônios na primeira camada intermediária, 16 na segunda camada intermediária e 1 neurônio na camada de saída (A:5-13-16-1) foi escolhida por apresentar os maiores coeficientes de determinação para as etapas de treinamento e teste (Fig. 2).

Figura 2 - Arquitetura do modelo de previsão de vazão para o açude Orós utilizando RNA.

Na [Fig. 3](#) é exibida a evolução do coeficiente de determinação em função do número de iterações, para as fases de treinamento (alteração dos pesos sinápticos) e teste. Observa-se que o melhor desempenho do modelo de RNA é obtido com 200.000 iterações, apesar de os valores de R^2 aumentarem para as iterações seguintes na etapa de treinamento, na etapa de teste o desempenho tende a piorar, indicando a ocorrência de “over-fitting”, ou treinamento excessivo. Na [Fig. 4](#) é possível observar a correlação, na fase de treinamento, entre os valores calculados pela rede neural e os valores reais da vazão para o açude Orós.

No modelo de RNA é possível obter o percentual de contribuição de cada uma das variáveis de entrada na resposta final gerada pela rede, estas contribuições são determinadas com base no algoritmo proposto por [Garson \(1991\)](#), que envolve essencialmente o particionamento dos pesos das conexões entre as camadas ocultas e a de saída em componentes associadas com cada um dos sinais de entrada. Para o modelo neuronal deste trabalho, estes valores podem ser observados na [Tabela 2](#). Os resultados indicam a Oscilação Decadal do Pacífico tem a menor contribuição no processo de previsão das vazões afluentes ao açude Orós utilizando RNA, e conforme o modelo do trabalho de [Araújo et al. \(2015\)](#), as anomalias de temperatura da superfície do mar do Oceano Atlântico são as que mais influenciam na resposta.

Para o modelo de RNA, foi realizada a análise dos resíduos, com o objetivo de se determinar uma distribuição de probabilidade para os mesmos, e em seguida se definir o nível de confiança para os valores obtidos pelo modelo. A partir da realização de testes de aderência, observou-se que os resíduos do modelo se adequavam a Distribuição de Probabilidade Normal, e que considerando um nível de confiança de 95%, o valor estimado da vazão

Figura 4 - Correlação entre os valores calculados pela rede neural e os valores reais da vazão para o açude Orós na fase de treinamento.

(Q_{est}) afluente ao açude Orós pode ser dada em função do valor calculado pela rede (Q_{mod}) de acordo com a seguinte expressão:

$$Q_{mod} - 66,75 \leq Q_{est} \leq Q_{mod} + 29,16 \quad (14)$$

O comparativo entre os modelos de RNA e o de [Souza Filho e Lall \(2003\)](#) foi feito com base nos resultados da etapa de teste (15 anos da série histórica), uma vez que, nesta etapa é verificada a eficácia da previsão para valores fora do conjunto de treinamento/calibração. Na [Fig. 5](#) é apresentada a comparação entre os valores calculados pelos modelos neuronal e o de [Souza Filho e Lall \(2003\)](#) e

Figura 3 - Variação do coeficiente de determinação (R^2) com o número de iterações para o modelo de RNA.

Tabela 2 - Contribuição das variáveis estudadas na previsão de vazões no açude Orós para o modelo de RNA.

Entrada	1- NINO3	2 - TAS	3 - TAN	4 - PDO	5 - AMO
Contribuição (%)	22,32	21,64	22,61	4,98	28,44

Figura 5 - Correlação entre os valores calculados pelo modelo neuronal (RNA) e o de Souza Filho e Lall (2003) (K-vizinhos) e os valores observados da vazão para o açude Orós na fase de teste.

os valores observados da vazão para o açude Orós para a série de teste. Já na Fig. 6 pode-se analisar a variabilidade dos valores gerados pelos modelos, sem levar em conta sua distribuição de probabilidade. É possível observar que na análise dos valores determinísticos, o modelo que utiliza RNA apresenta um comportamento mais aproximado com a série histórica, já a mediana produzida pelo modelo de Souza Filho e Lall (2003) gera valores próximos a média das vazões observadas.

Com base no Gráfico de Taylor (Fig. 7), sabe-se que quanto mais próximo o valor representativo de um método à referência, em geral, melhor o método. Verifica-se que o modelo de RNA apresentou maior coeficiente de correlação, menor *RMSE* e uma maior proximidade ao desvio-padrão da série histórica de vazões em comparação ao modelo de Souza Filho e Lall (2003).

Analisando-se a eficiência Nash-Sutcliffe, para o modelo de RNA o valor obtido foi de 0,5, considerado aceitável, já para o modelo k-vizinhos (Souza Filho e Lall, 2003), o valor foi de $E = 0,2$, indicando que o modelo apresenta-se um pouco melhor do que utilizar a média histórica da série de vazões para realizar a previsão de vazões (Tabela 3).

Para a análise probabilística dos valores gerados pelos modelos, calculou-se a razão de verossimilhança. Para ambos os modelos e para série histórica, foi constatado que a função de densidade de probabilidade, que melhor se adequou aos dados, foi a da distribuição Nor-

Figura 6 - Distribuição dos dados observados e modelados por meio do gráfico bloxpot.

Figura 7 - Aplicação do Diagrama de Taylor para representação dos modelos de previsão de vazões.

mal. Para o modelo de [Souza Filho e Lall \(2003\)](#), a distribuição de probabilidade de cada um dos valores gerados foi feita considerando os 30 anos vizinhos da série histórica de vazões. No modelo de RNA utilizou-se a distribuição de probabilidade dos resíduos. A determinação do valor do desempenho indicada na [Tabela 3](#) foi feita com base na Equação 11, verifica-se que o modelo de RNA apresenta valor satisfatório (Desempenho > 1) e o modelo k-vizinhos indica uma piora na previsão (Desempenho < 1).

Na [Fig. 8](#) é possível verificar a distribuição dos dados gerados pelos modelos, considerando a característica estocástica destes. Verifica-se que para todos os anos da série de teste, os valores observados encontram-se dentro o intervalo máximo e mínimo do modelo de [Souza Filho Lall \(2003\)](#). Isto ocorre, pois, este intervalo varia em função dos anos vizinhos, levando a maiores valores em comparação ao modelo de RNA, que é limitada a análise dos resíduos.

Tabela 3 - Variáveis comparativas entre os modelos.

Modelo	R ²	Nash-Suttcliffe	Desempenho
RNA	0,7	0,5	1,3
k-Vizinhos	0,3	0,2	1,0

Figura 8 - Distribuição das vazões da série histórica de teste e dos valores gerados pelos modelos de [Souza Filho e Lall \(2003\)](#) e de RNA, considerando a distribuição de probabilidade dos valores e o intervalo de confiança, respectivamente.

4. Conclusões

Neste trabalho foi realizado o comparativo entre modelos de previsão de vazões para a bacia hidrográfica da barragem Orós no Estado do Ceará utilizando redes neurais artificiais (melhoramento do modelo de Araújo et al. (2015)) e o modelo proposto por Souza Filho e Lall (2003) a partir de uma série histórica de 100 anos de dados hidrometeorológicos (anomalias de temperatura da superfície do mar e vazões).

Para o modelo neuronal, foi identificado que dentre os preditores climáticos utilizados como variáveis de entrada, a Oscilação Decadal do Pacífico tem menor contribuição nas repostas geradas pelo modelo.

A análise do desempenho dos modelos, utilizando as variáveis comparativas: coeficiente de eficiência Nash-Sutcliffe (E), coeficiente de determinação (R^2), o diagrama de Taylor (2001) e a razão de máxima verossimilhança, mostrou que o modelo de RNA apresentou uma maior eficiência na predição das vazões do reservatório Orós, pois para todas as variáveis analisadas foram obtidos melhores valores no modelo neuronal. Ressalta-se que apesar de o modelo das RNA apresentar melhor desempenho, as alterações para inclusão de novos registros históricos requerem que sejam desenvolvidos novos modelos, o que não é necessário no modelo de Souza Filho e Lall (2003). Observa-se também que o modelo k-vizinhos representa melhor a característica estocástica da previsão de vazões.

Por fim, a implementação de modelos de previsão de vazões possibilita uma melhor gestão dos recursos hídricos, permitindo estimar diferentes cenários de disponibilidade hídrica meses antes do período chuvoso, proporcionando um melhor planejamento da operação do reservatório e de ações de resposta eventos a eventos extremos (secas e cheias).

Agradecimentos

O presente trabalho foi realizado com apoio da Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) - Segurança Hídrica, Alimentar e Energética na Bacia Estendida do Rio São Francisco: 441457/2017-1 que também financiou os custos da publicação deste artigo. Ademais, as pesquisas contaram também com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - SECAS E CHEIAS - Impactos das Mudanças Climáticas em Extremos Hidrológicos (Secas e Cheias): 88887.123932/2015-00.

Referências

ALEXANDRE, A.M.B. **Previsão de Vazões Mensais para o Sistema Interligado Nacional Utilizando Informações**

- Climáticas**. Tese de Doutorado em Engenharia Civil, Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, 293 p, 2012.
- ARAÚJO, C.B.C; DANTAS NETO, S.A.; SOUZA FILHO, F.A. **Streamflow forecasting for the dam Orós/Ce from hydrometeorological data using perceptrons**. Revista Brasileira de Meteorologia, v. 30, n.1, p. 37-46, 2015.
- ARAÚJO, C.B.C. **Aplicação das Redes Neurais Artificiais do Tipo Perceptron na Estimativa de Recalques em Estacas**. Dissertação de Mestrado em Engenharia Civil, Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 227p, 2015.
- BARROS, R.C.; ROLIM, P.A.M.; ALMEIDA, R.M.B.; MELO, A.P. Influência do ENOS na variabilidade da vazão, na Bacia do Alto Rio Negro. In: **Anais 13º Congresso Brasileiro De Meteorologia**, Sociedade Brasileira de Meteorologia, Fortaleza, 2004.
- BARROS, F.V.F.; MARTINS, E.S.P.R.; SOUZA FILHO, F.A. Regionalização de parâmetros do modelo chuva-vazão SMAP das bacias hidrográficas do Ceará. In: SOUZA FILHO, F.A.S.; CAMPOS, J.N.; AQUINO, S.H. (eds) **Gerenciamento de Recursos Hídricos no Semiárido**. Expressão Gráfica e Editora, Fortaleza, 2013.
- BERRI, G.J.; GHETTO, MA.; GARCÍA, N.O. The influence of ENSO in the flows of the upper Paraná River of South America over the past 100 years. **Journal of Hydro-meteorology**, v. 3, n. 1, p. 57-65, 2002.
- CARDOSO, A.O.; SILVA DIAS, P.L. The relationship between ENSO and Paraná River flow. **Advances in Geosciences**, v. 6, p. 189-193, 2006.
- CARRIELLO, F.; SOARES, J.V.; FERREIRA, N.J. A resposta hidrológica das sub-bacias brasileiras e sua relação com o fenômeno ENSO-El Niño/Oscilação Sul. In: **Anais do 12º Simpósio Brasileiro de Sensoriamento Remoto**. Instituto Nacional de Pesquisas Espaciais, São José dos Campos. p. 2479-2486, 2005. Available from: <<http://urlib.net/ltid.inpe.br/sbst/2004/11.19.19.09>>. Access on: 07 nov. 2018.
- COULIBALY, P.; ANCTIL, F.; BOBÉE, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. **Journal of Hydrology**, v. 230, p. 244-257, 2000.
- DANTAS NETO, S.A.; SILVEIRA, M.V.; AMÂNCIO, L.B.; ANJOS, G.M. Pile settlement modelling with multilayer perceptrons. **Electronic Journal of Geotechnical Engineering**, v. 19, p. 4517-4528, 2014.
- DANTAS NETO, S.A.; SILVEIRA, M.V.; ANJOS, G.M.; MOURA, A.S. Settlement prediction model for continuous flight auger, steel and bored piles using artificial neural networks. **Revista Geotecnia**, v. 136, p. 27-47, 2016.
- FERREIRA, K.C.D; LOPES, F.B.; ANDRADE, E.M.; MEIRELES, A.C.M; SILVA, G.S. Adaptação do índice de qualidade de água da National Sanitation Foundation ao semiárido brasileiro. **Revista Ciência Agrônômica**, v. 46, n. 2, p. 277-286, 2015.
- FOLEY, J.A.; BOTTA, A.; COE, M.T.; COSTA, M.H. El Niño-Southern oscillation and the climate, ecosystems and rivers of Amazonia. **Global Biogeochemical Cycles**, v. 16, n. 4, p. 79,1-79,17, 2002.
- GARSON, G.D. Interpreting neural network connection weights. **Artificial Intelligence Expert**, v. 6, p. 47-51, 1991.

- GOMES, L.F.C.; MONTENEGRO, S.M.G.L.; VALENCIA, M.J.S. Modelo baseado na técnica de redes neurais para previsão de vazões na Bacia do Rio São Francisco. **Revista Brasileira de Recursos Hídricos**, v. 15, p. 5-15, 2010.
- HAYKIN, S. **Redes Neurais Artificiais: Princípios e Práticas**. 2ª ed. Editora Bookman, Porto Alegre, 2001.
- KAPLAN, A.; CANE, M.; KUSHNIR, Y.; CLEMENT, A.; BLUMENTHAL, M.; RAJAGOPALAN, B. Analyses of global sea surface temperature 1856-1991. **Journal of Geophys Research**, v. 103, n. C9, p. 18,567-18,589, 1998.
- KIM, I S.; VISSOTTO, S. Relações entre variações das vazões nas bacias hidrográficas do brasil e os fenômenos El Niño e La Niña. In: **Anais do 15º Simpósio Brasileiro de Recursos Hídricos**. Associação Brasileira de Recursos Hídricos, Curitiba, p. 1-5, 2003.
- LINDGREN, B. W. **Statistical Theory** (3rd ed). Macmillan, New York, 1976.
- MENEZES JUNIOR, J.M.P.; BARRETO, G.A. Long-term time series prediction with the NARX network: na empirical evaluation. **Neurocomputing**, v. 71, p. 3335-3343, 2008.
- NASH J.E.; SUTCLIFFE J.V. River flow forecasting through conceptual models part i - a discussion of principles. **Journal of Hydrology**, v. 10, 3 ed., p. 282-290, 1970.
- PEREIRA, D.P.; LIMA, J.S. de S; XAVIER, A.C.; PASSOS, R.R.; FIEDLER, N.C. Taylor Diagram application for the evaluation of spatial interpolators in attributes of soil under cultivation with eucalyptus. **Revista Árvore**, v. 38, n. 5, p. 899-905. 2014.
- ROCHA, E.J.P.; ROLIM, P.A.M.; SANTOS, D.M. Modelo estatístico hidroclimático para previsão de níveis em Altamirapa. In: **Anais 17º Simpósio Brasileiro de Recursos Hídricos**, São Paulo. Associação Brasileira de Recursos Hídricos, São Paulo, p. 1-8, 2007.
- ROHN, M.C.; KAVISKI, E.; CUNHA, L.M. Estimativa de variáveis hidrológicas a partir de indicadores do fenômeno EL NIÑO. In: **Anais 15º Simpósio Brasileiro De Recursos Hídricos**, Associação Brasileira de Recursos Hídricos, Curitiba, p. 1-8, 2003.
- SILVA, D.F.; MOLION, L.C.B. Influência da variabilidade climática interanual na hidrologia da Bacia do Rio São Francisco. In: **Anais Congresso Brasileiro de Meteorologia**. Sociedade Brasileira de Meteorologia, Fortaleza, p. 1-5, 2004.
- SILVEIRA, C.S. **Modelagem Integrada de Meteorologia e Recursos Hídricos em Múltiplas Escalas Temporais e Espaciais: Aplicação no Ceará e no Setor Hidroelétrico Brasileiro**. Tese de Doutorado, Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, 2014.
- SILVEIRA, C.S.; ALEXANDRE, A.M.B.A.; SOUZA FILHO, F.A.; VASCONCELOS JUNIOR, F.C.; CABRAL, S.L. Monthly streamflow forecast for National Interconnected System (NIS) using periodic auto-regressive endogenous models (PAR) and exogenous (PARX) with climate information. **Revista Brasileira de Recursos Hídricos**, v. 22, p. 31-40, 2017.
- SOARES, J.V.; CARRIELLO, F.; FERREIRA, N.J.; RENNÓ, C.D. Mapping the hydrologic response of the brazilian hydrologic regions and their variability associated with El Niño and La Niña. **Revista Ambiente & Água**, v. 1, n. 1, p. 21-36, 2006.
- SOUZA FILHO, F.A.; LALL, U. Modelo de previsão de vazões sazonais e interanuais. **Revista Brasileira de Recursos Hídricos**, v. 9, n. 2, p. 61-74, 2004.
- SOUZA FILHO, F.A.; LALL, U. Seasonal to interannual ensemble streamflow forecasts for Ceará, Brazil: application of multivariate, semiparametric algorithm. **Water Resources Research**, v. 39, n. 11, p. SWC1-SWC13, 2003.
- TAYLOR, K.E. Summarizing multiple aspects of model performance in a single diagram. **Journal of Geophysical Research**, v. 106, n. 7, p. 7183-7192, 2001.
- UVO, C.B.T.U.; BERNDTSSON, R. Forecasting discharge in Amazonia using artificial neural networks. **International Journal of Climatology**, v. 20, p. 1495-1507, 2000.
- UVO, C.B.; GRAHAM, N.E. Seasonal runoff forecast for Northern South America: a statistical model. **Water Resources Research**, v. 34, n. 12, p. 3515-3524, 1998.
- WANG, W.; VAN GELDER, H.A.J.; VRIJLING, J.K; MA, J. Forecasting daily streamflow using hybrid ANN models. **Journal of Hydrology**, v. 324, p. 383-399, 2006.
- ZEALAND, C.M.; BURN, D.H.; SIMONOVIC, S.P. Short term streamflow forecasting using artificial neural networks. **Journal of Hydrology**, v. 214, p. 32-48, 1999.