

XXIII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

EMPREGO DO ÍNDICE PADRONIZADO DE PRECIPITAÇÃO E REDES NEURAIS ARTIFICIAIS PARA PREVISÃO DE SECA METEOROLÓGICA

Saulo Freire Crosland Guimarães¹ ; Veber Costa²

RESUMO – Este estudo busca implementar uma ferramenta para previsão de secas meteorológicas, empregando-se o índice padronizado de precipitação (SPI) e redes neurais artificiais (RNAs). Para tanto, utilizou o posto pluviométrico União dos Palmares, no estado de Alagoas, o qual apresenta uma série contínua de 81 anos, para estimação do SPI, considerando-se em escalas de agregação de um, três, seis e doze meses, as quais relacionam-se com os mecanismos de propagação da seca meteorológica para as outras variáveis do ciclo hidrológico. A capacidade de previsão das RNAs para valores de SPI, para as referidas escalas de agregação, foi avaliada em horizontes de previsão de um a seis meses. Para a análise dos resultados obtidos utilizou-se o coeficiente de determinação. O estudo mostrou que as RNAs utilizadas foram aptas a prever, com relativo grau de acurácia, a ocorrência de secas em escalas temporais de três, seis e doze meses com horizonte de previsão mensal. Em escalas temporais de longo prazo, como o SPI-12, esse horizonte de previsão é maior, chegando a quatro meses.

ABSTRACT– This study aims to implement a tool for forecasting meteorological droughts, using the standardized precipitation index (SPI) and artificial neural networks (ANNs). For this, the rainfall gauging station of União dos Palmares, in the state of Alagoas, which encompasses a period-of-record of 81 years, was utilized. Aggregation scales of 1, 3, 6 and 12 months, which are related to the mechanisms of drought propagation to other variables of the water cycle, were considered. The predictive abilities of RNAs, for the referred aggregation scales, was evaluated for lead times of one to six months. For assessing the results obtained, the coefficient of determination was used. The study showed that the utilized RNAs were able to forecast, with a relative degree of accuracy, the occurrence of droughts at three, six and twelve month aggregation scales, for the lead time of 1 month. On long-term time scales, such as the SPI-12, this forecast horizon is longer, reaching four months.

Palavras-Chave – Secas meteorológicas; Redes neurais artificiais; Índice Padronizado de Precipitação.

1) Instituto Mineiro de Gestão das Águas – IGAM, Prédio Minas, Primeiro andar, sala 06. Rod. Papa João Paulo II, 4101 - Serra Verde, Belo Horizonte – MG. Fone: (31) 3915-1576. Email: saulo.guimaraes@meioambiente.mg.gov.br

2) Universidade Federal de Minas Gerais - UFMG – Departamento de Engenharia Hidráulica e Recursos Hídricos, Bloco 1- Sala 4506. Av. Antônio Carlos 6627, Pampulha, Belo Horizonte - MG. Fone: (31) 3409-1002. Email: veber@ehr.ufmg.br.

INTRODUÇÃO

A Organização Mundial de Meteorologia - World Meteorological Organization (1986) - define seca meteorológica como a deficiência na precipitação considerada normal em uma dada área. Essa deficiência, dependendo da intensidade, duração e cobertura espacial, afeta diretamente atividades humanas como abastecimento, agricultura, pecuária, geração de energia por fonte hídrica e indústria. Os impactos socioeconômicos podem ser de grande magnitude, uma vez que podem provocar deslocamentos populacionais em locais afetados, baixos rendimentos em colheitas e perda de criações animais. Desse modo, são de grande importância estudos que possam contribuir com ações de planejamento dos recursos hídricos frente às secas, em especial para mitigação de seus efeitos na exploração da água.

Nesse sentido, a aplicação de índices que possam caracterizar as secas tem se tornado relevante. O Índice Padronizado de Precipitação, ou SPI (do acrônimo em inglês para “*Standard Precipitation Index*”), por exemplo, tem sido amplamente utilizado na detecção de secas ao redor do mundo. O SPI é de fácil implementação, pois necessita apenas de dados de precipitação como variável de entrada. Ademais, é um índice versátil, pois pode ser aplicado em diversas escalas temporais, as quais, indiretamente, refletem os efeitos da escassez de precipitação nas variáveis do ciclo hidrológico, tais como a água no solo ou nos cursos d’água.

A previsão de um evento de seca é uma importante ferramenta na gestão dos recursos hídricos, pois auxilia em seu planejamento e mitigação dos efeitos causados pela seca. O estudo de métodos que possam prever secas se faz importante nesse contexto. Vários métodos vêm sendo empregados nesse sentido, com destaque, métodos de inteligência artificial como redes neurais artificiais (RNAs) suporte à regressão vetorial (SVR), *Wavelets* acopladas e *Hidden Markov Models* (HMM) (Belayneh *et al.*, 2014, 2015, 2016, Belayneh e Adamowski, 2012, 2013; Guedes, 2011; Kousari *et al.*, 2017)

Estudos como os de Guedes (2011), Belayneh *et al.* (2014, 2015) e Deo e Şahin (2015) apontam as RNAs como ferramentas eficientes e de grande utilidade na modelagem de fenômenos de comportamento não-linear. Portanto, o presente estudo apresenta-se com o objetivo de empregar modelos baseados em redes neurais artificiais na previsão de SPI, de escalas temporais de um, três, seis e doze meses, as quais englobam intervalos de agregação associados a diferentes tipologias de secas, com previsões de um a seis meses de antecedência.

MATERIAL E MÉTODOS

Caracterização da região

A estação pluviométrica do presente estudo encontra-se inserida na Região Hidrográfica Atlântico Nordeste Oriental e localiza-se no Estado de Alagoas, na abrangência da Sub-bacia 39, que intercepta parte de três estados brasileiros, a saber: Paraíba (PB), Pernambuco (PE) e Alagoas (AL), conforme é mostrado na Figura 1.

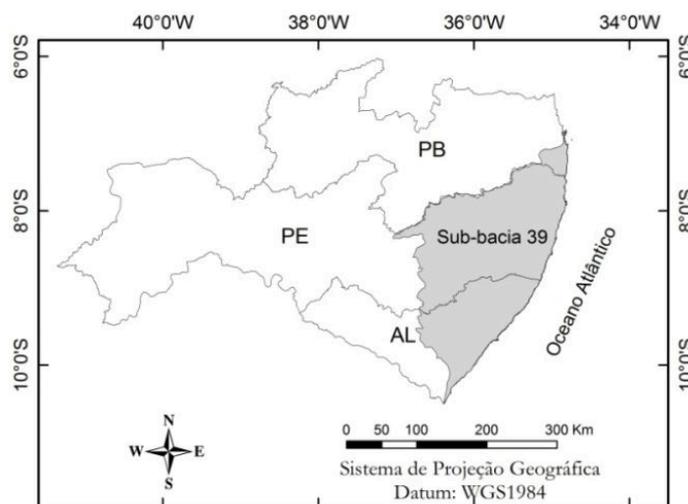


Figura 1 - Interseção da Sub-bacia 39 com Estados brasileiros. Fonte: SALGUEIRO (2015)

Coleta de dados

As informações de precipitação foram obtidas junto ao Sistema Nacional de Informação em Recursos Hídricos, por meio da plataforma online Hidroweb, da Agência Nacional de Águas (ANA). Devido à necessidade de longas séries históricas, de trinta anos no mínimo (McKee *et al.*;1993), e sendo desejáveis séries maiores que sessenta anos (Guttman, 1994), para se implementar o SPI, foi selecionado o posto pluviométrico União do Palmares, de código 00936053, que possui oitenta e um anos de observações.

Preparação dos dados de precipitação

Para o cálculo do SPI, primeiramente definiram-se as escalas temporais de estudo, a saber: mensal, trimestral, semestral e anual. A partir dos dados obtidos do posto pluviométrico, foram elaborados quatro novos vetores, contendo, cada um, o valor acumulado de precipitações dos meses anteriores, de acordo com cada escala temporal pré-estabelecida. Por exemplo, em uma escala mensal, os valores presentes no vetor são os mesmos da série obtida na base da ANA. Para uma escala trimestral os valores presentes no vetor correspondem à precipitação total acumulada no período dos

três meses anteriores, de modo que o mês de dezembro apresenta o total acumulado de precipitação para os meses de outubro, novembro e dezembro, e assim sucessivamente.

Cálculo do SPI

Para o cálculo do SPI associado, utilizou-se a linguagem de programação R. A cada um dos quatro vetores de precipitação acumulada foi ajustada uma distribuição de probabilidades. Neste trabalho, assim como na formulação original do método, utilizou-se a distribuição Gama, a qual é dada por:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad (1)$$

na qual $\alpha > 0$ é o parâmetro de forma, $\beta > 0$ o parâmetro de escala, $x > 0$ representa a quantidade de precipitação e $\Gamma(\alpha)$ é a função Gama, definida como:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (2)$$

Os parâmetros α e β são estimados pela máxima verossimilhança. A probabilidade acumulada do evento de precipitação observado é calculada com os parâmetros estimados, logo:

$$G(x) = \int_0^x g(x) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-x/\beta} dx \quad (3)$$

Fazendo-se $t = x/\beta$, esta equação torna-se uma função Gama incompleta, logo:

$$G(x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (4)$$

A distribuição Gama não é definida em zero. Por outro lado, a ocorrência de valores de precipitação nulos, principalmente, para escalas temporais de curto prazo, é comum. Devido a isso, ajustou-se aos dados uma distribuição mista de probabilidades, como apresentado por Edwards e McKee (1997), a qual é formalmente expressa por:

$$H(x) = q + (1-q)G(x) \quad (5)$$

na qual q é a frequência empírica de valores nulos. A probabilidade acumulada, $H(x)$, é então transformada em uma variável aleatória normal padrão Z , com média zero e variância unitária, que é o valor do SPI.

Definição da arquitetura da rede neural artificial

Para implementação da rede neural utilizou-se o pacote “neuralnet”, disponível no R. Como abordado por Fritsch *et al.* (2016), esse pacote pode implementar o “...treinamento de redes neurais usando *backpropagation*, *backpropagation* resiliente (Riedmiller, 1994) com ou sem *backtracking* de peso (Riedmiller e Braun, 1993) ou a versão globalmente convergente modificada por (Anastasiadis *et al.*, 2005)”. O pacote permite personalizar a métrica de erro para treinamento e a função de ativação (FRITSCH *et al.*, 2016). A arquitetura utilizada neste estudo para as RNAs foi a *Multilayer Perceptron* (MLP) e o algoritmo de treinamento o *backpropagation* (Rprop). Estabeleceu-se uma parcela de 70% do comprimento da série para fins de treinamento do modelo e 30% para validação.

A quantidade de dados de entrada, camadas escondidas e de neurônios por camada, bem como *threshold*, a função de ativação, o limite máximo de iterações e o método de cálculo do erro foram estabelecidos por meio de processos de tentativa e erro. Para a escolha dos parâmetros mais adequados para o modelo, levou-se em conta a convergência e o coeficiente de determinação apresentado em cada resultado. Inicialmente, foram testados modelos mais simples, passando-se para modelos mais complexos, com o número de neurônios variando entre 1 e 20. Ademais, testes para se definir a quantidade de vetores de entrada foram realizados, variando-se de 1 a 5 *inputs* na camada de entrada. A configuração com 2 vetores apresentou o melhor desempenho, haja vista que o uso de mais vetores de entrada, não acarretou ganho substancial no coeficiente de determinação.

O tipo de aprendizado de rede definido foi o supervisionado. Nesse tipo de aprendizado, a rede recebe os dados de entrada e a saída pré-definida. Os dados de entrada utilizados foram valores de SPI anteriores ao mês que se deseja prever. Os dados de saída são os valores de SPI de previsão. Neste trabalho, foram considerados horizontes de previsão de um a seis meses.

Avaliação dos modelos de previsão

A Figura 2 apresenta o gráfico relacionando o coeficiente de determinação calculado para cada escala de agregação de SPI (SPI-1, SPI-3, SPI-6 e SPI-12), em função do *lead* time definido.

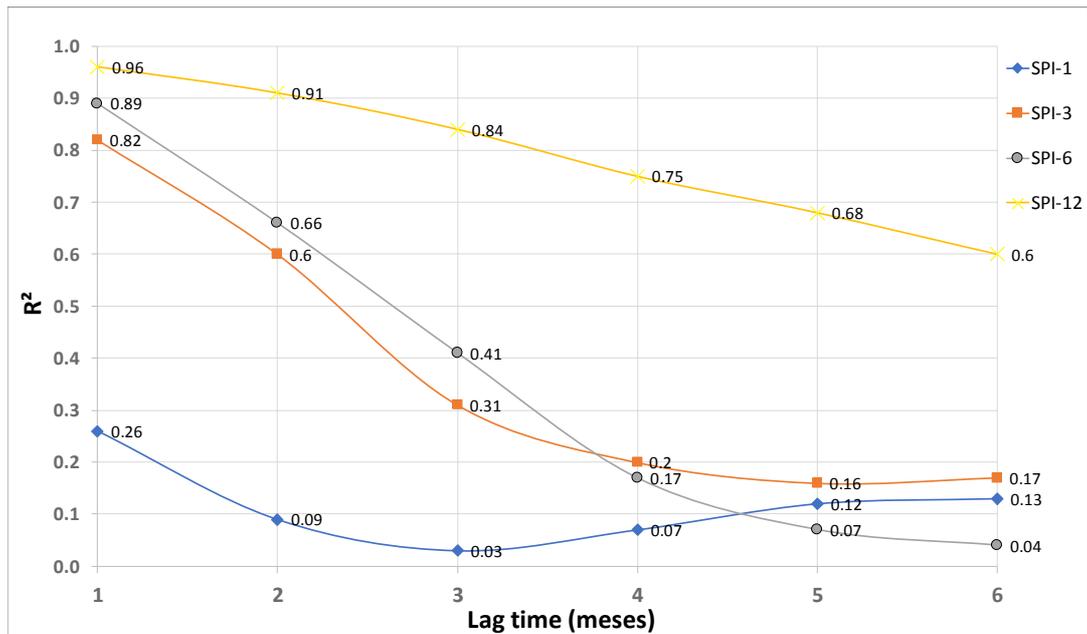


Figura 2 - Coeficiente de determinação por escala temporal em função do lead time na etapa de validação

Em nenhuma das simulações o coeficiente de determinação do SPI-1 superou o valor de 26%. Esse desempenho era esperado, uma vez que, na escala de agregação mensal, a componente aleatória do processo é substancial em relação àquela associada à persistência. Devido a isso, espera-se, em escalas temporais de curto prazo, uma maior dificuldade no treinamento do modelo. Logo, as previsões, nessas escalas de agregação, são pouco acuradas e de difícil identificação para eventos de seca mais significativos. Esse padrão de resultados corrobora com aqueles encontrados por Guedes (2011) e Khadr (2016). À medida em que a escala de agregação é aumentada, o grau de suavização da série também aumenta, e, conseqüentemente, a série apresenta padrões menos aleatórios, devido a diminuição da variabilidade temporal melhorando, desse modo, a capacidade de treinamento do modelo. Percebe-se a componente aleatória preponderante para escalas de curto prazo, como SPI-1 e SPI-3. A partir do SPI-6, oscilações perceptíveis em escalas menores têm menor impacto, deste modo aprimorando a capacidade de identificação de eventos críticos. A medida em que as escalas aumentam, eventos mais prolongados e severos são perceptíveis.

A Figura 3 apresenta os resultados de SPI-12 observados e previstos, com previsão de um mês de antecedência. Percebe-se um ajuste adequado da série prevista aos valores observados. Conforme esperado, à medida em que se aumenta o *lead time*, ou a antecedência da previsão, o coeficiente de determinação diminui. Os melhores ajustes ocorrem em escalas de médio a longo prazo, como no SPI-12. Observou-se um coeficiente de determinação de 96% para a escala temporal anual, (SPI-12), com previsão de um mês de antecedência. Segundo Guedes (2011), SPI de escalas temporais de longo prazo estão associados à secas hidrológicas e hidrogeológicas, afetando a disponibilidade de água

para consumo, escoamento superficial e recarga de aquíferos. A figura , por sua vez, apresenta as séries temporais de SPI-1 observadas e previstas, com previsão de um mês de antecedência. Percebe-se a baixa aderência entre as duas, sobretudo nos pontos mínimos, que apresentam os valores de SPI mais relevantes no estudo de secas.

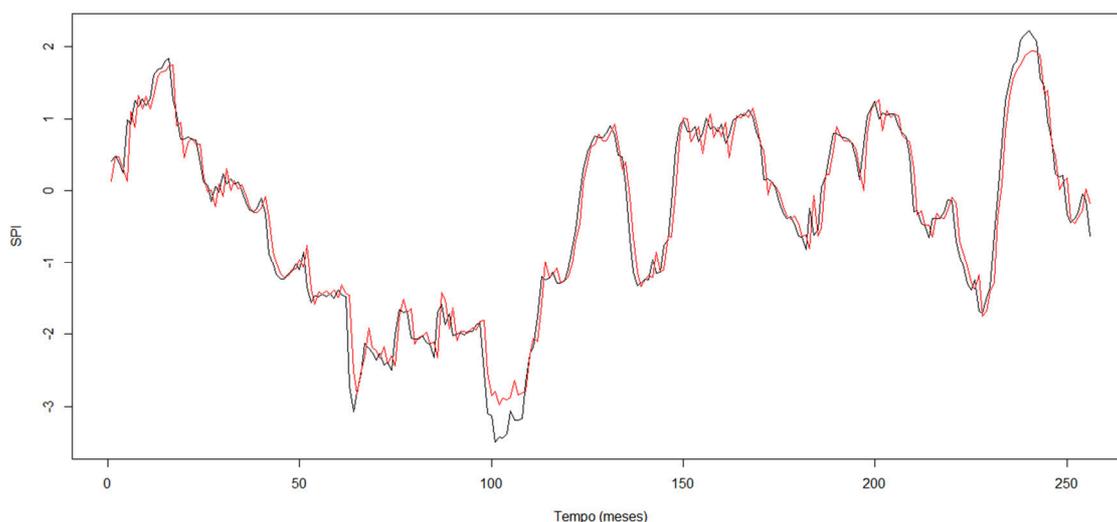


Figura 3 - SPI-12 observado e previsto com *lead time* de um mês.

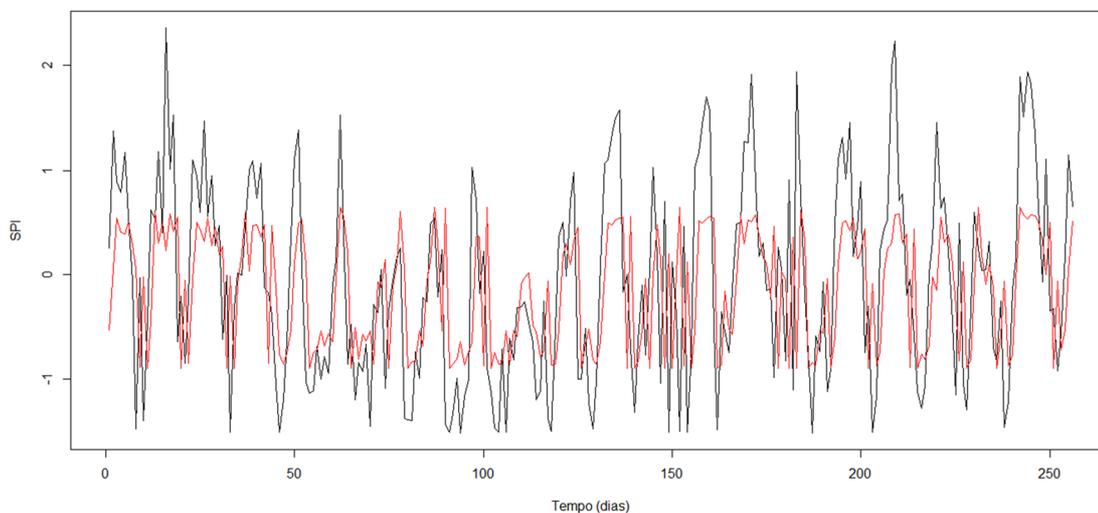


Figura 4 - SPI-1 observado e previsto com *lead time* de um mês.

CONCLUSÕES

O estudo das secas é uma importante iniciativa para a gestão dos recursos hídricos. A aplicação de RNAs para previsão do SPI, neste estudo, apresentou resultados compatíveis com os encontrados na literatura. Avaliou-se a aplicação de RNAs para previsão de SPI de escalas temporais de um, três, seis e doze meses, SPI-1, SPI-3, SPI-6 e SPI-12, respectivamente, com horizonte de previsão de um a seis meses. Os resultados a que este estudo chegou mostram que as RNAs são aptas a prever com

relativo grau de acurácia, a ocorrência de secas em escalas temporais de três, seis e doze meses com horizonte de previsão de um mês. Especificamente, para a escala de doze meses (SPI-12), chegou-se a coeficientes de determinação superiores a 0.75, para horizontes de previsão de até quatro meses. Esse fato evidencia a potencialidade do método proposto como ferramenta auxiliar na gestão dos recursos hídricos, seja na operação de reservatórios para abastecimento, geração de energia, gestão da agricultura e culturas animais, ou em aplicações associadas à mitigação dos danos socioeconômicos causados pela seca.

Deve-se ainda, ampliar a área de abrangência do estudo, uma vez que se utilizou neste trabalho apenas uma estação pluviométrica. A proposta de se trabalhar com mais estações pluviométricas é realizar uma análise mais ampla da ocorrência das secas, avaliando-se, também, suas características espaço-temporais. No presente estudo, verificou-se o desempenho de RNAs *feedforward* com algoritmo de treinamento *backpropagation*. É válido, também, um estudo comparativo de outras arquiteturas, como a RNA recorrente e a reticulada, com outros algoritmos de treinamento, a fim de se avaliar e comparar seu desempenho com a configuração definida no presente estudo. De mesmo modo, é importante a avaliação de outros métodos de previsão como o *Support Vector Regression* (SVR), *Wavelets* acopladas e *Hidden Markov Model* (HMM). Ademais, a comparação desta metodologia com outras fundamentadas em métodos lineares e tradicionais de modelagem, como a abordagem Box-Jenkins, é válida e deve ser realizada futuramente.

REFERÊNCIAS

ANASTASIADIS, A. D.; MAGOULAS, G. D.; VRAHATIS, M. N. (2005). "New globally convergent training scheme based on the resilient propagation algorithm". *Neurocomputing*, v. 64, n. 1–4 SPEC. ISS., p. 253–270.

BELAYNEH, A. et al. (2014). "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models". *Journal of Hydrology*, v. 508, p. 418–429.

BELAYNEH, A. et al. (2016). "Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction". *Atmospheric Research*, v. 172–173, p. 37–47.

BELAYNEH, A.; ADAMOWSKI, J. (2012). "Standard Precipitation Index Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Support Vector Regression". *Applied Computational Intelligence and Soft Computing*, v. 2012, p. 1–13.

BELAYNEH, A.; ADAMOWSKI, J. (2013). "Drought forecasting using new machine learning methods". *Journal of Water and Land Development*, v. 18, n. I–VI, p. 3–12.

BELAYNEH, A.; ADAMOWSKI, J.; KHALIL, B. (2015). "Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods". *Sustainable Water Resources Management*, v. 2, n. 1, p. 1–15.

DEO, R. C.; ŞAHIN, M. (2015). "Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia". *Atmospheric Research*, v. 161–162, p. 65–81.

EDWARDS, D. C.; MCKEE, T. B. (1997). "Characteristics of 20th Century drought in the United States at multiple time scales". *Atmospheric Science Paper No. 634*, May 1–30, v. 298, n. 704, p. 174.

FRITSCH, S. et al. Training of neural networks. Package "neuralnet". The R project for statistical computing. p. 13, 2016.

GUEDES, R. V. S. Previsão de seca na bacia hidrográfica do açude epitácio pessoa com base em redes neurais artificiais. Dissertação de Mestrado - Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, Campina Grande. 2011.

GUTTMAN, N. B. (1994). "On the Sensitivity of sample L moments to sample size". *Journal of Climate*, v. 7, p. 4.

GUTTMAN, N. B. (1998). "Comparing the Palmer Drought Index and the Standardized Precipitation Index". *Journal of the American Water Resources Association*, v. 34, n. 1, p. 113–121.

KHADR, M. (2016). "Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia)". *Ain Shams Engineering Journal*, v. 7, n. 1, p. 47–56, 2016.

KOUSARI, M. R. et al. (2017). "Introducing an operational method to forecast long-term regional drought based on the application of artificial intelligence capabilities". *Theoretical and Applied Climatology*, v. 127, n. 1–2, p. 361–380.

MCKEE, T. B.; DOESKEN, N. J.; KLEIST, J. (1993). "The relationship of drought frequency and duration to time scales" in. *AMS 8th Conference on Applied Climatology*, n. January, p. 179–184.

WORLD METEOROLOGICAL ORGANIZATION - WMO. (1986). "Annual Report of the World Meteorological Organization". Official records of the World Meteorological Organization. United Nations.

REN, G. et al. (2012). "SPI user guide". *Journal of Climate*, v. 21, n. 6, p. 1333–1348, 2012.

RIEDMILLER, M.; BRAUN, H. (1993). "A direct adaptive method for faster backpropagation learning: The RPROP algorithm" in *IEEE International Conference on Neural Networks - Conference Proceedings*, v. 1993–Janua, p. 586–591.

SALGUEIRO, J. H. P. DE B. Eventos extremos máximos de precipitação na sub-bacia 39 e o caso do alerta na cidade de Recife - PE. Tese de Doutorado - Universidade Federal de Pernambuco, Centro de Tecnologia e Ciências, Recife. 2015.

SANTOS, M. S. Caracterização espaço-temporal de secas utilizando ondaletas e o standardized precipitation index: uma aplicação para a parcela mineira da bacia do Rio São Francisco. Dissertação de Mestrado - Universidade Federal de Minas Gerais, PPG - SMARH, Belo Horizonte. p. 163, 2011.