

XXIII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

ANÁLISE DA RELAÇÃO ENTRE ÍNDICES CLIMÁTICOS E VAZÕES USANDO UMA ABORDAGEM DE APRENDIZADO DE MÁQUINA

Larissa Zaira Rafael Rolim¹; Francisco de Assis de Souza Filho²; Renata Locarno Frota¹; Gabriela de Azevedo Reis¹ & Victor Costa Porto¹

RESUMO – Uma operação eficiente de reservatórios exige uma compreensão de como as vazões dos reservatórios estão mudando sob diferentes condições hidrológicas e climáticas, especialmente em localidades onde a ocorrência de eventos extremos é comum. Ao longo da última década, o uso de técnicas de Inteligência Artificial e Mineração de Dados teve um crescimento significativo nas áreas de operação de reservatórios e previsão sazonal. Este trabalho propõe a utilização de um classificador de floresta aleatório para identificar os preditores mais importantes dentre os índices climáticos que possam estar relacionados à vazão do reservatório Orós. Foram analisados 22 índices climáticos e ranqueados utilizando a técnica de aprendizado em máquina. A anomalia do Niño 3 foi a principal informação climática relacionada com a região, seguida do Índice do Atlântico Sul Tropical e do Índice de Niño Oceânico. Desta forma, estas informações climáticas estão correlacionadas com as condições hidrológicas locais no estudo de caso, podendo conter informações que melhorem a previsão sazonal da região.

ABSTRACT– An efficient reservoir operation requires an understanding of how reservoir flows are changing under different hydrological and climatic conditions, especially in locations where the occurrence of extreme events is common. Over the last decade, the use of Artificial Intelligence and Data Mining techniques has experienced significant growth in the areas of reservoir operation and seasonal forecasting. This work proposes the use of a random forest classifier to identify the most important predictors among the climatic indexes that may be related to the Orós reservoir inflow. It was analyzed 22 climatic indexes and ranked using the machine learning technique. The Niño 3 anomaly was the main climatic information related to the region, followed by the Tropical South Atlantic Index and the Oceanic Niño Index. Thus, this climatic information is correlated with the local hydrological conditions in our case studies, and may contain information that improves the seasonal forecast of the region.

Palavras-Chave – Índices climáticos, florestas aleatórias, aprendizado em máquina.

¹ Doutorando em Eng. Civil – Recursos Hídricos (UFC): larissazairarr@gmail.com; renata.locarno@hotmail.com; gabrielareisazevedo@gmail.com; victorporto@gmail.com.

² Professor do Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Campus do Pici, Bloco 713 – assis@ufc.br

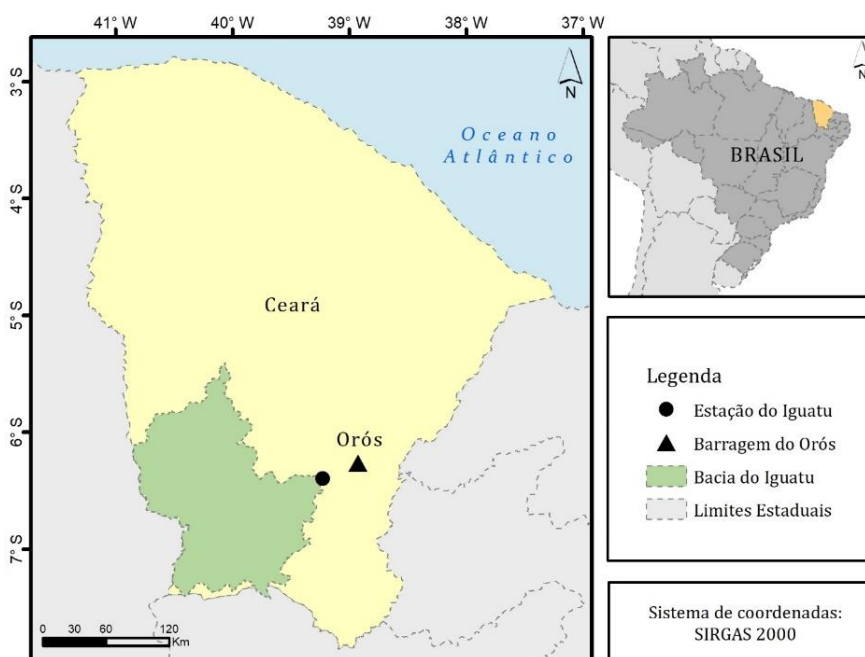
1. INTRODUÇÃO

O Nordeste brasileiro (NEB) é caracterizado por um clima semiárido, existindo uma grande variabilidade na distribuição de precipitação na escala temporal e espacial. A região é suscetível a ocorrência de eventos extremos (secas e cheias) que podem afetar os hidrossistemas e controlar a disponibilidade hídrica, tendo assim um grande impacto social e econômico. A variabilidade climática da região é afetada por diversos fenômenos como a oscilação nos padrões da Temperatura da Superfície do Mar (TSM) ou fenômenos como El Niño-Oscilação Sul (ENOS) e a Oscilação Decadal do Pacífico (ODP).

As mudanças climáticas e as ações antropogênicas podem ser observadas por meio de modificações no ciclo hidrológico com impacto direto nos padrões das variáveis hidroclimatológicas, principalmente nos eventos de extremos (IPCC, 2013).

A previsão de afluências tem uma influência direta na operação dos hidrossistemas, podendo ser uma ferramenta importante no nível de planejamento para auxílio de tomada de decisão. Diversos estudos apontam que a incorporação de variáveis exógenas, como os índices climáticos na previsão sazonal, que podem melhorar o modelo da previsão devido à redução de incerteza (SOUZA FILHO e LALL, 2003; CARDOSO *et al.*, 2006; SILVEIRA *et al.*, 2017). Assim, nesse trabalho é proposto a utilização de um classificador de floresta aleatório para identificar os preditores mais importantes relacionados à vazão do reservatório Orós (Figura 1).

Figura 1 - Localização da Barragem do Orós.



2. METODOLOGIA

Base de Dados

Os dados fluviométricos utilizados nesse trabalho foram obtidos no site da Agência Nacional de Águas (ANA), no Sistema de Informações Hidrológicas (HIDROWEB). O posto utilizado foi o do Iguatu, referentes as fluviometria mensais dos anos 1912-2016. O preenchimento das falhas do posto foi feito utilizando o modelo SMAP (Soil Moisture Accounting. Procedure). Os índices climáticos utilizados foram resumidos na Tabela 1.

Tabela 1 - Informação dos índices de climático selecionados.

Índice	Nome	Fonte
Niño1.2Anom	Niño 1+2 SST Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Niño12/
Niño3Anom	Niño 3 SST Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Niño3/
Niño3.4Anom	Niño 3.4 SST Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Niño34/
Niño4Anom	Niño 4 SST Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Niño4/
BEST	Bivariate enso Index (Niño 3.4 HadleySST + SOI)	https://www.esrl.noaa.gov/psd/people/cathy.smith/best/
SOIfromCRU	Southern Oscillation Index (utiliza dados do CRU - Climate Research Unit)	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/SOI/
TNI	Trans-Niño Index (dados HadISST e NCEP OI SST)	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/TNI/
AMO	Atlantic multidecadal oscillation (AMO) SST Index	http://www.esrl.noaa.gov/psd/data/timeseries/AMO/
PDO	Pacific Decadal Oscillation (PDO) Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/PDO/
NP	North Pacific Index	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NP/
SAHELRAINFALL	SahelStandardized Rainfall	https://www.esrl.noaa.gov/psd/data/climateindices/list/
NAO	North Atlantic Oscillation	https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/NAO/
EPNP	East Pacific/North Pacific Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
WP	Western Pacific Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
TNA	Tropical Northern Atlantic Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
TSA	Tropical Southern Atlantic Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
WHWP	Western Hemisphere warm pool	https://www.esrl.noaa.gov/psd/data/climateindices/list/
MEI	Multivariate ENSO Index (MEI)	https://www.esrl.noaa.gov/psd/data/climateindices/list/
ONI	Oceanic Niño Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
QBO	Quasi-Biennial Oscillation	https://www.esrl.noaa.gov/psd/data/climateindices/list/
PNA	Pacific North American Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/
CAR	Caribbean SST Index	https://www.esrl.noaa.gov/psd/data/climateindices/list/

Foram utilizados como dados de entrada as médias do período de maior afluência da estação (Fev-Jun) e, para os índices, a média do trimestre do ano anterior com maior correlação entre a vazão (Abr-Jun).

Florestas aleatórias

Florestas aleatórias (FA) são agrupamentos de árvores de decisão treinadas, formando um modelo robusto não paramétrico capaz de manipular dados multidimensionais não-lineares, com ruído para classificação, combinando variáveis selecionadas dos dados originais aleatoriamente (bootstrap) (BREIMAN, 2001). Esta técnica foi escolhida devido à sua ampla e bem-sucedida aplicação em operações de reservatórios (BESSLER *et al.*, 2003; CORANI *et al.*, 2009; YANG *et al.*, 2016) e na previsão de afluências (Yang *et al.* 2017)

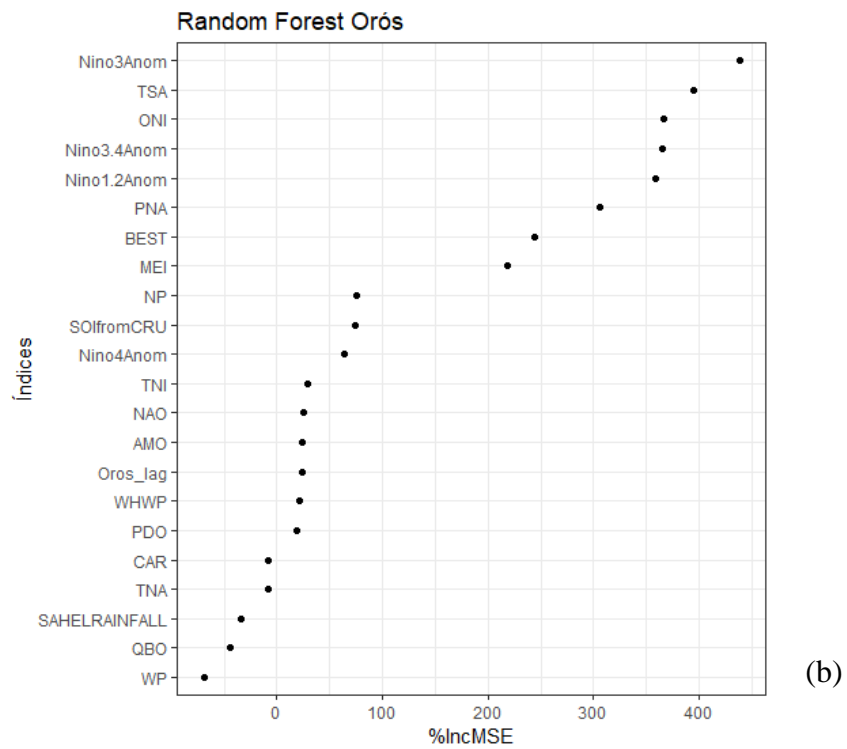
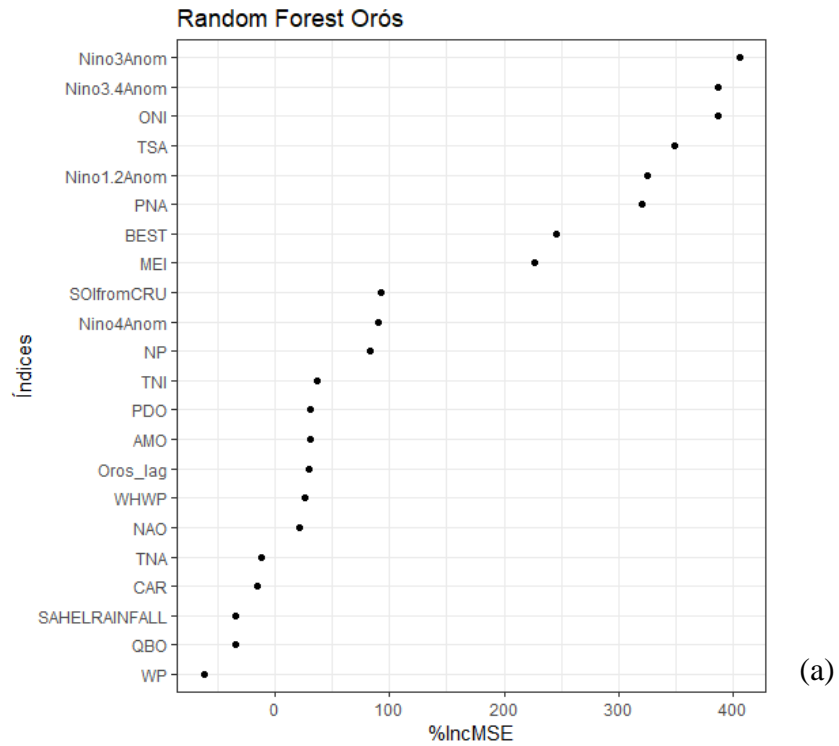
A estratégia do algoritmo FA é explorar a importância dos preditores usando subconjuntos de dados e preditores *bootstrapped* para o crescimento de árvores de decisão. Na FA cada nó da árvore de decisão é dividido usando o melhor entre um subconjunto de preditores escolhido aleatoriamente nesse nó. Para cada iteração de *bootstrap*, a floresta aleatória aponta os dados não incluídos (chamados de *out-of-bag*) utilizando a árvore cultivada na amostra de *bootstrap*. Depois de agregar todas as previsões de *out-of-bag*, a taxa de erro global é estimada (BREIMAN, 2001).

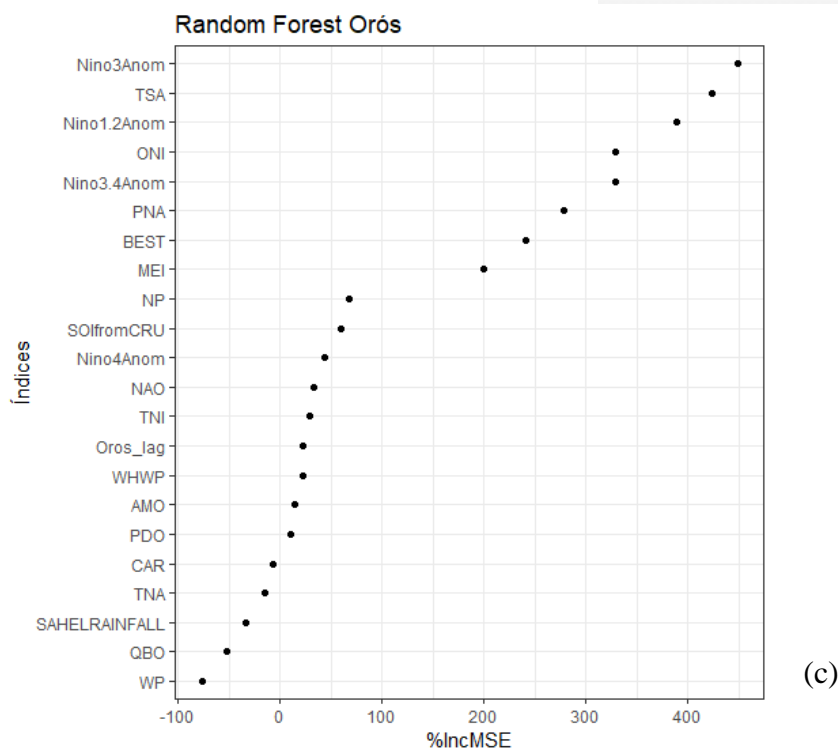
A floresta aleatória prevê a importância de uma variável observando o quanto o erro de previsão aumenta quando os dados *out-of-bag* para essa variável são permutados, enquanto todos os outros são deixados fixos (Catani *et al.*, 2013). Os principais parâmetros das FA são: o número de árvores (*ntree*); a complexidade das árvores; o número de variáveis preditoras selecionadas aleatoriamente em cada divisão (*mtry*); e o tamanho do conjunto de treinamento. Foi utilizado o pacote “*random forest*” do *software* livre R para a classificação da importância dos índices climáticos.

3. RESULTADOS

Foram construídos três modelos fixando o número de variáveis preditoras selecionadas aleatoriamente em cada divisão e variando o número de árvores (250, 500 e 750 árvores) e três modelos fixando o número de árvores e variando o *mtry* (5, 7 e 9 variáveis).

Figura 2 - Importância das variáveis aplicando (a) $mtry = 5$ (b) $mtry = 7$ e (c) $mtry = 9$





Para os modelos onde foram variados o número de árvores, não houve alteração na ordem de importância dos índices. A Figura 2 ilustra o ganho de informação dos três modelos com a variação do *mtry* e aplicando o número de árvores igual a 500.

A anomalia do Niño 3 foi a variável climática de maior importância em todos os 3 modelos. Esta anomalia é amplamente conhecida por influenciar o regime de precipitações da região nordeste. Em segundo lugar de importância, têm-se o Índice do Atlântico Sul Tropical, o qual já foi indicado por Saravanan e Chang (2000) como um fenômeno climático que afeta o clima no NEB.

4. CONCLUSÕES

Foi aplicada neste estudo uma técnica bastante popular em aprendizado em máquina para selecionar os principais índices climáticos, buscando incorporá-los para realizar a previsão sazonal do reservatório Orós. A técnica indicou que a variabilidade climática das afluências é influenciada pela anomalia Niño 3, seguida do índice do Atlântico Sul Tropical e demais índices relacionados ao fenômeno ENSO. Esta técnica é de fácil compreensão e pode torna-se uma importante ferramenta na seleção de variáveis climatológicas que serão incorporadas na previsão sazonal.

REFERÊNCIAS

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

CARDOSO, Giovanni Brígido Bezerra; MARTINS, E. S. P. R.; SOUZA FILHO, F. A. **Uso de otimização/simulação e previsão de aflúências na operação tática dos reservatórios do sistema Jaguaribe-Metropolitano**. 2003. Tese de Doutorado. Dissertação de Mestrado. Universidade Federal do Ceará. Departamento de Engenharia Hidráulica e Ambiental. 141p.

CATANI, F. et al. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. **Natural Hazards and Earth System Sciences**, v. 13, n. 11, p. 2815-2831, 2013.

SARAVANAN, R.; CHANG, Ping. Interaction between tropical Atlantic variability and El Niño–Southern oscillation. **Journal of Climate**, v. 13, n. 13, p. 2177-2194, 2000.

SILVEIRA, Cleiton da Silva et al. Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information. **RBRH**, Porto Alegre, v. 22, e30, 2017. Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2318-03312017000100233&lng=en&nrm=iso>. access on 13 Apr. 2019. Epub June 26, 2017. <http://dx.doi.org/10.1590/2318-0331.011715186>.

SOUZA FILHO, Francisco Assis; LALL, Upmanu. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. **Water Resources Research**, v. 39, n. 11, 2003.

STOCKER, Thomas F. et al. *Climate change 2013: The physical science basis*. 2013.

YANG, Tiantian et al. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. **Water Resources Research**, v. 53, n. 4, p. 2786-2812, 2017.

YANG, Tiantian et al. Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. **Water Resources Research**, v. 52, n. 3, p. 1626-1651, 2016.