

XXIV SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

APLICAÇÃO DO MODELO DE *MACHINE LEARNING M5 MODEL TREE* PARA PREVISÃO DE NÍVEIS DIÁRIOS NA BACIA DO RIO URUGUAI

*Camila Dalla Porta Mattiuzi*¹, *Artur Jose Soares Matos*², *Franco Turco Buffon*³

Palavras-Chave – Previsão Hidrológica, *M5 model tree*, Bacia do Rio Uruguai

INTRODUÇÃO

A previsão da variação dos níveis fluviométricos é essencial para o monitoramento e planejamento das decisões em nível de bacia hidrográfica, principalmente em situações de eventos hidrológicos extremos. A previsão de níveis pode ser realizada a partir de modelos conceituais, que são aqueles baseados nos processos físicos que ocorrem na bacia hidrográfica, ou modelos *data driven*, os quais capturam as relações entre variáveis sem o conhecimento prévio sobre o comportamento do sistema em estudo (Solomatine *et al.*, 2008). Dentre os modelos *data driven* pode-se citar o modelo de *machine learning M5 model tree* (Quinlan, 1992), cuja principal característica é aprendizado automático e a construção de uma árvore de decisão com regressões lineares a partir de um conjunto de dados fornecido.

No âmbito do monitoramento de níveis fluviométricos ao longo do território brasileiro destaca-se o Sistema de Alerta Hidrológico (SAH), o qual é implementado e operado pela SGB/CPRM (Serviço Geológico do Brasil), que inclui o monitoramento em tempo real de chuvas e níveis de rios em diversas estações hidrometeorológicas no Brasil, bem como a previsão dos níveis durante eventos hidrológicos extremos, para subsidiar a tomada de decisões por parte dos órgãos relacionadas à mitigação dos impactos (CPRM, 2021).

Na Bacia do Rio Uruguai foi implementado um SAH em dezembro de 2018, o qual realiza a previsão de níveis durante eventos de inundação para seis municípios no Rio Grande do Sul: Uruguaiana, Itaqui e São Borja (rio Uruguai), Alegrete (rio Ibirapuitã), Manoel Viana (rio Ibicuí), e Rosário do Sul (rio Santa Maria), beneficiando uma população total de 349 mil pessoas (CPRM, 2020).

Neste contexto, o objetivo deste trabalho é desenvolver uma aplicação do modelo de *machine learning M5 model tree* para realizar a previsão dos níveis na estação Passo São Borja, no Rio Uruguai, para horizontes de 1, 2 e 3 dias, utilizando dados de níveis das estações hidrometeorológicas e precipitação por satélite.

METODOLOGIA

M5 model tree

As variações dos níveis em rios são processos naturais de complexa representação, o que muitas vezes torna difícil a construção de um modelo global que capture adequadamente o comportamento

1) CPRM-SGB: Banco da Província, 105, Morro Santa Tereza, Porto Alegre, RS, 90.840-030, (51) 3406-7300, camila.mattiuzi@cprm.gov.br

2) CPRM-SGB: Avenida Brasil, 1731, Funcionários, Belo Horizonte, MG, 90.840-030, (31) 3878-0307, artur.matos@cprm.gov.br

3) CPRM-SGB: Banco da Província, 105, Morro Santa Tereza, Porto Alegre, RS, 90.840-030, (51) 3406-7300, franco.buffon@cprm.gov.br

do sistema como um todo. No rio Uruguai, eventos de cheias e estiagens ocorrem ao longo do ano sem periodicidade bem definida. Em casos como este uma alternativa é a utilização de modelos modulares, nos quais, durante o processo de calibração, os dados de entrada são divididos em subconjuntos e a partir destes são construídos os modelos locais (Solomatine *et al.*, 2008; Sattari *et al.*, 2013).

O algoritmo *M5 model tree* (Quinlan, 1992) pode ser considerado um modelo modular pois assume que não existe uma função constante em todo o domínio do problema, mas que ela pode ser aproximada em subdomínios menores que são definidos e caracterizados a partir de funções de regressão linear (Sattari *et al.*, 2013). Em outras palavras, o algoritmo divide automaticamente o conjunto de dado de entrada em subconjuntos conforme um critério de decisão e constrói uma estrutura de árvore de decisão, que é clara e objetiva, porém com funções de regressão linear nas pontas, ao invés de atributos numéricos contínuos. Regressões lineares são criadas assumindo que existe uma relação linear entre entradas e saídas do modelo, no entanto a construção do *M5 model tree* vai na direção de não-linearidade, pois o resultado é um modelo localmente linear, porém globalmente não-linear (Solomatine *et al.*, 2008), ou seja, é considerado um modelo piecewise linear (Solomatine e Xue, 2004).

A construção da árvore de decisão utiliza a técnica *divide-and-conquer*: partindo de um conjunto de dados de treinamento um atributo é selecionado e colocado no nó inicial da árvore (*root*), e uma sub-árvore (*branch*) é construída para cada valor possível, então o conjunto de dados de entrada é dividido em sub-conjuntos; o processo é repetido recursivamente para cada sub-árvore utilizando somente o conjunto de dados referente àquela sub-árvore, e quando todas as amostras estão classificadas o desenvolvimento da sub-árvore para (Quinlan, 1992; Solomatine e Xue, 2004). O atributo utilizado para dividir os subconjuntos é chamado de *splitting criterion*, e no *M5 model tree* este critério é baseado no desvio padrão do subconjunto em cada nó, no qual é escolhido o atributo que maximiza a redução do erro (Quinlan, 1992). Após a construção da árvore, o modelo avalia se existem muitas sub-árvores, o que poderia gerar um modelo excessivamente preciso com pouca generalização; para deixar o modelo mais robusto são combinadas sub-árvores através do processo de *pruning*. Por fim, é utilizado o processo de *smoothing* para evitar descontinuidades entre as folhas das árvores, ou seja, entre os modelos de regressão linear. (Wang e Witten, 1996; Solomatine e Xue, 2004)

O algoritmo *M5 model tree* tem sido utilizado em diversos estudos na área das ciências hidrológicas, principalmente para construir modelos chuva-vazão (Khan e See, 2006; Battarchaya e Solomatine, 2005; Solomatine e Dulal, 2003), previsão de vazão e secas (Adnan *et al.*, 2018; Deo *et al.*, 2017; Sattari *et al.*, 2013; Stravs e Brilly, 2007; Solomatine *et al.*, 2007; Solomatine e Xue, 2004), modelagem de concentração de sedimentos em suspensão (Adnan *et al.*, 2021; Goyal, 2014; Kumar *et al.*, 2012) e modelagem de níveis de água subterrânea (Rezaie-balf *et al.*, 2017, Nalarajan e Mohandas, 2015), entre outros.

A principal vantagem do *M5 model tree* é que ele fornece uma representação do sistema que é reproduzível e compreensível por parte dos usuários e tomadores de decisão, pois é possível visualizar os nós, sub-árvores e regressões lineares. Outras vantagens são: os processos de treinamento e calibração do modelo são rápidos, o modelo sempre converge, *model trees* são mais compactos que outras de árvore de decisão, e existe a possibilidade de construção de modelos mais simples ou mais complexos a partir do aumento/diminuição no número de sub-árvores e folhas (regressões lineares).

Área de estudo

A Bacia Hidrográfica do Rio Uruguai está localizada na fronteira entre Brasil, Uruguai e Argentina e possui área de drenagem de 385 mil km², das quais aproximadamente 45% está em território brasileiro, sendo 73% está no estado do Rio Grande do Sul e 27% em Santa Catarina (MMA,

2006). O Rio Uruguai possui cerca de 2.200 km de extensão, é formado pela confluência dos rios Pelotas e Canoas e tem seu curso na direção leste-oeste dividindo os estados do Rio Grande do Sul e Santa Catarina; após a confluência com o rio Peperi-Guaçu, o curso muda para a direção sudoeste, delimitando a fronteira entre o Brasil e a Argentina; após a afluência do rio Quaraí, o Rio Uruguai deixa o território brasileiro, e passa a definir a fronteira entre o Uruguai e a Argentina, até a sua foz, no estuário do Rio da Prata (MMA, 2006; CPRM, 2020). A Figura 1 apresenta o mapa de localização da Bacia Hidrográfica do Rio Uruguai.

Figura 1. Localização da Bacia Hidrográfica do Rio Uruguai



Na calha principal do Rio Uruguai, antes da afluição do rio Chapecó, entre os municípios de Águas de Chapecó (SC) e Alpestre (RS), está localizada a UHE Foz do Chapecó, a qual iniciou a operação em 2010, possui quatro unidades geradoras e tem uma potência instalada de 855 megawatts. A barragem tem 598m de extensão e 48m de altura; o reservatório da usina tem uma área de 79,2km², dos quais 40km² correspondem à própria calha do Rio Uruguai. (Foz do Chapecó, 2021)

A primeiro ponto de monitoramento do nível do Rio Uruguai pertencente ao SAH-Uruguai à jusante da UHE Foz do Chapecó é a estação hidrometeorológica Iraí (74100000), distante 28km da usina. A distância entre Iraí e a estação Passo São Borja é de aproximadamente 560km²; as áreas de drenagem das estações Iraí e Passo São Borja são 61.900 km² e 125.000 km², respectivamente. Conforme análise das séries de vazões nas duas estações no período que vai de jan/2015 a ago/2020, a qual está apresentada na Figura 2, o tempo de reação dos níveis entre as duas estações é de aproximadamente 3 dias. A vazão média em Iraí é 45% da vazão média de Passo São Borja; o resumo das vazões está apresentado na Tabela 1.

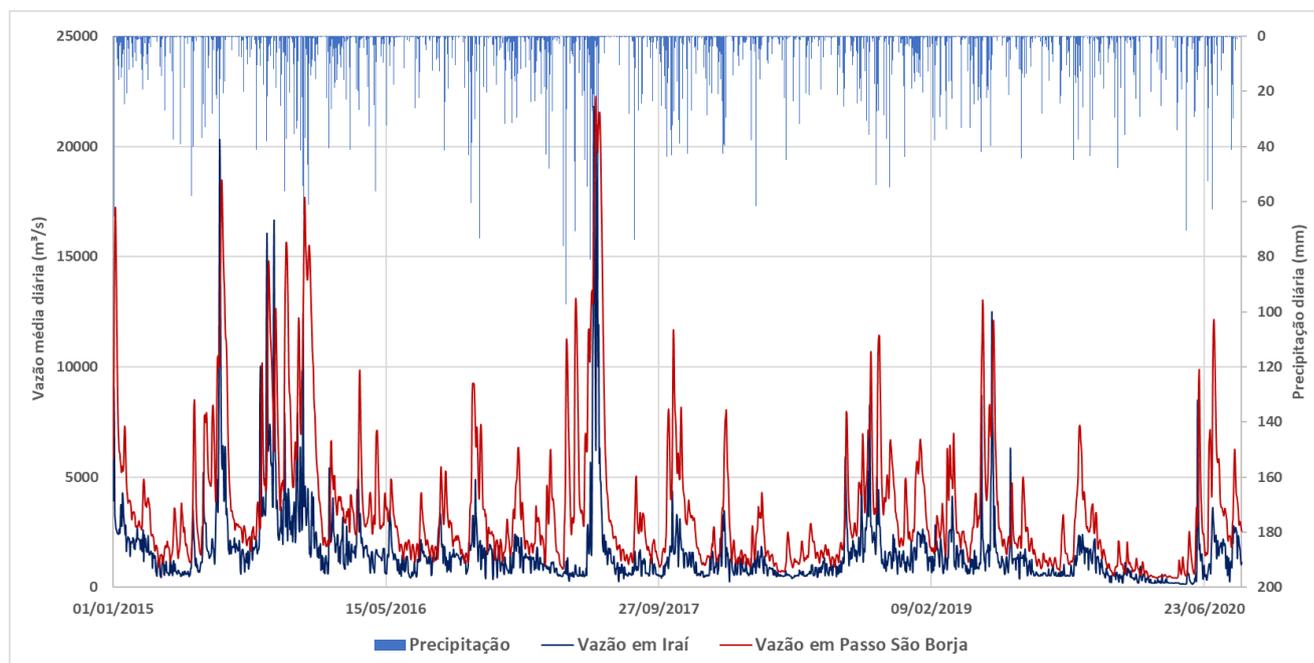
Para calibrar o modelo de previsão dos níveis para Passo São Borja para os três dias subsequentes, ou seja, o horizonte de previsão é t+1, t+2 e t+3, foram utilizados os dados da própria anteriores a t, inclusive, e os dados de níveis da estação Iraí anteriores a t, inclusive, como entrada para o modelo, sendo que os dados de Iraí refletem a operação da UHE Foz do Chapecó. A precipitação média dos dias anteriores sobre a área de drenagem entre Passo São Borja e Iraí, que corresponde a 63.000 km², também foi utilizada como dado de entrada do modelo, para representar a contribuição da precipitação nos níveis da estação Passo São Borja. Ao todo foram identificadas 21 variáveis, sendo selecionadas àquelas que apresentavam melhor correlação com o nível a ser previsto.

Os dados hidrológicos fluviométricos foram obtidos a partir do portal Hidroweb da Agência Nacional de Águas (ANA, 2021); optou-se por trabalhar com dados de níveis médios diários ao invés de vazões médias diárias por dois motivos: para evitar possíveis erros de extrapolação de curva-chave, principalmente em vazões mínimas e máximas, e para que o resultado do modelo seja fornecido em nível, atendendo ao padrão de monitoramento no SAH. A precipitação acumulada média diária ao longo da área de contribuição entre as estações de Passo São Borja e Iraí foi obtida a partir do produto MERGE do INPE (Instituto Nacional de Pesquisas Espaciais) (Rozante *et al.*, 2010), o qual é gerado utilizando dados do IMERG/GPM (Global Precipitation Measurement Integrated Multi-satellite Retrievals) incluindo dados de estações pluviométricas; o comportamento da precipitação média pode ser observado na Figura 2. O modelo foi construído na aplicação desenvolvida por Witten e Frank (2002); para a calibração do modelo foram utilizados dados de 01/01/2015 a 31/12/2018, e para validação dados de 01/01/2019 a 31/08/2020.

Tabela 1. Vazões observadas nas estações Iraí e Passo São Borja (janeiro de 2015 a agosto de 2020)

	Vazão mínima (m ³ /s)	Vazão máxima (m ³ /s)	Vazão média (m ³ /s)	Área de Drenagem (km ²)
Iraí (74100000)	145	21.665	1.605	61.900
Passo São Borja (75780000)	423	22.268	3.595	125.000

Figura 2. Hidrograma de Iraí e Passo São Borja e precipitação



RESULTADOS

Conforme apontado anteriormente, uma vantagem da construção de *M5 model trees* é a rapidez na calibração; neste estudo foram calibrados um total de 22 de modelos diferentes para a previsão de nível em Passo São Borja para horizontes de 1, 2 e 3 dias. As variáveis de entrada dos modelos foram os níveis em dias anteriores em Passo São Borja, níveis em dias anteriores em Iraí e precipitação média acumulada na bacia em dias anteriores. As métricas de avaliação dos modelos selecionados estão apresentadas na Tabela 2, tendo sido calculadas a partir do subconjunto de dados de validação.

Para avaliar a performance dos modelos foram escolhidos os índices de Nash–Sutcliffe (NSE) e Kling–Gupta (KGE). Valores de NSE e KGE próximos a 1 indicam ótima correspondência entre os dados simulados e observados (Knoben *et al.*, 2019). Conforme métricas apresentadas na Tabela 2, os valores de NSE variaram de 0,97 a 0,91, decrescendo conforme o aumento no horizonte de previsão; o mesmo comportamento foi observado no KGE, que variou de 0,98 a 0,88. Estes resultados apontam para uma boa correspondência entre os dados observados e os resultados da simulação.

Tabela 2. Resultados da calibração dos modelos de previsão de nível em Passo São Borja para 1, 2 e 3 dias

	Nº de sub-árvores	Nº de variáveis	Correlação	MAE	RMSQ	NSE	KGE
t+1	1	9	0.985	21	33	0.97	0.98
t+2	8	15	0.973	32	45	0.94	0.93
t+3	5	14	0.958	42	57	0.91	0.88

MAE = mean absolute error (erro médio absoluto)

RMSQ = root mean square error (raiz do erro médio quadrático)

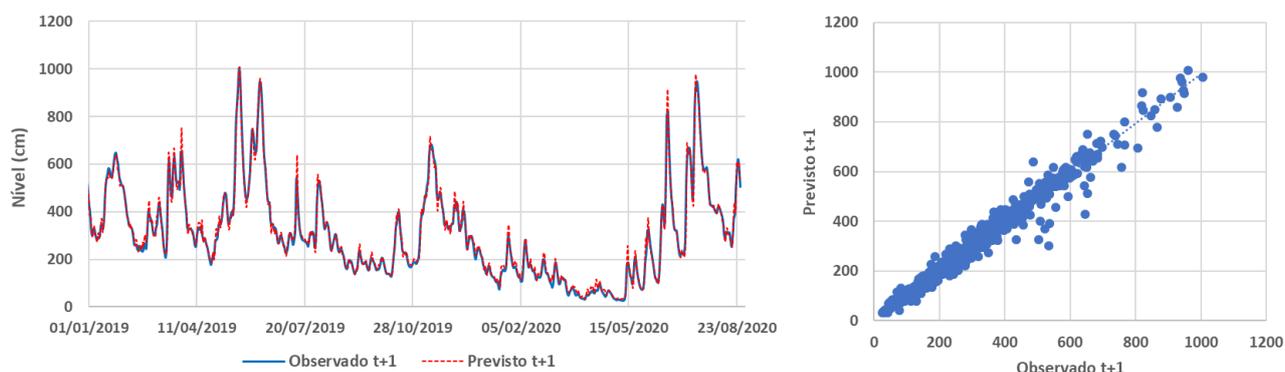
NSE = Nash–Sutcliffe

KGE = Kling–Gupta

Previsão t+1

O modelo de previsão t+1 escolhido é composto por apenas uma árvore de decisão com uma equação de regressão linear contendo nove variáveis; os resultados da validação para a previsão t+1 estão apresentados na Figura 3. Percebe-se boa correlação entre os dados simulados (previstos) e observados (correlação = 0.985 e erro médio absoluto na ordem de 21cm). Nota-se que houve superestimação nos valores previstos em alguns picos de cheia.

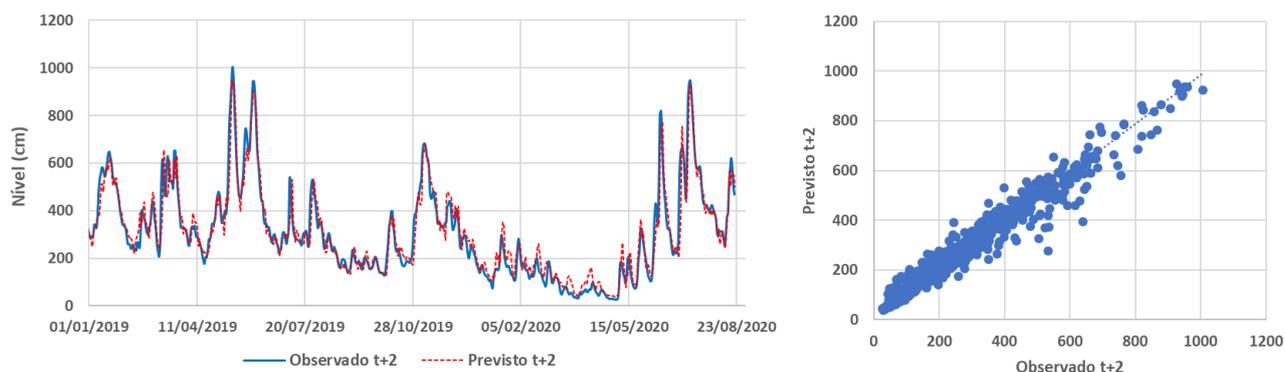
Figura 3. Níveis previstos em t+1 e observados no subconjunto de validação e diagrama de dispersão



Previsão t+2

O modelo de previsão t+2 é formado por oito equações de regressão linear contendo quinze variáveis; os resultados da validação para a previsão t+2 estão apresentados na Figura 4. Percebe-se novamente boa correlação entre os dados simulados e observados (correlação = 0.973 e erro médio absoluto na ordem de 32cm). Entretanto, ao contrário do que houve no modelo t+1, nota-se que houve subestimação nos valores previstos em alguns picos de cheia, e superestimação em picos previstos no período de estiagem (entre março e maio de 2020).

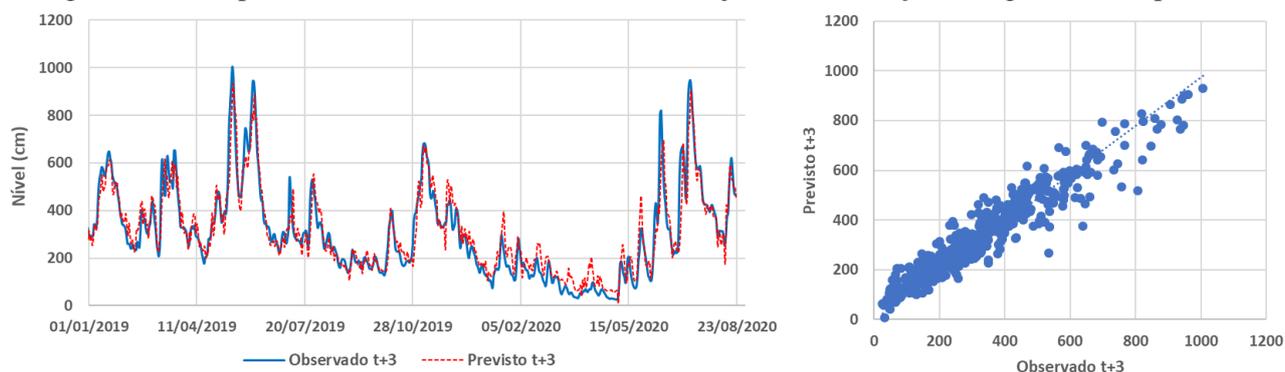
Figura 4. Níveis previstos em t+2 e observados no subconjunto de validação e diagrama de dispersão



Previsão t+3

O modelo de previsão t+3 é formado por cinco equações de regressão linear contendo quatorze variáveis; os resultados da validação para a previsão t+3 estão apresentados na Figura 5. Percebe-se boa correlação entre os dados simulados e observados (correlação = 0.958 e erro médio absoluto na ordem de 42cm). Assim como no modelo t+2, percebe-se que na previsão t+3 houve subestimação nos valores previstos em alguns picos de cheia, e superestimação em picos previstos no período de estiagem.

Figura 5. Níveis previstos em t+3 e observados no subconjunto de validação e diagrama de dispersão



Discussões

O comportamento dos níveis no rio Uruguai depende de diversos fatores: a bacia hidrográfica possui declividades mais elevadas na região das cabeceiras (Alto Uruguai), e áreas mais planas na região do Médio/Baixo Uruguai (CPRM, 2020), a precipitação média anual também é distinta ao longo da bacia, variando de 2.100mm nas cabeceiras até 1.400mm na região de fronteira com o Uruguai (Pinto *et al.*, 2011), sendo bem distribuída ao longo do ano, não havendo um período de estiagem ou cheias definido. Conforme resultados dos modelos calibrados, mesmo com toda a complexidade existente na bacia do rio Uruguai, o algoritmo do *M5 model tree* foi capaz de capturar as relações existentes entre níveis e precipitação, e gerar modelos para previsão de níveis na estação Passo São Borja para até 3 dias.

Os modelos ajustados utilizaram uma quantidade menor de variáveis do que a quantia que foi disponibilizada inicialmente (9, 15 e 14 para t+1, t+2 e t+3, respectivamente, de um total de 21). Nos três modelos selecionados observou-se um descolamento no ajuste nos picos de níveis; neste caso, o acréscimo dos dados de uma estação intermediária entre Iraí e Passo São Borja pode ser investigado para melhorar as previsões em subidas.

Observa-se que no conjunto de dados da calibração existem períodos de estiagem, porém a estiagem mais pronunciada foi incluída no conjunto de dados de validação (entre março e maio de 2020), o que pode explicar o ajuste inferior que os modelos t+2 e t+3 tiveram para simular níveis baixos. Neste caso, sugere-se ampliar a calibração para abranger o período de estiagem, e atualizar o conjunto de validação com dados mais recentes.

CONCLUSÕES

A previsão de níveis fluviométricos é uma ferramenta importante para auxiliar o monitoramento e tomadas de decisão, principalmente durante eventos extremos como inundações e estiagens. Neste trabalho foi aplicado o algoritmo de *machine learning M5 model tree* para desenvolver modelos de previsão de níveis médios diários para a estação Passo São Borja, localizada na calha principal do Rio Uruguai, para horizontes de 1, 2 e 3 dias.

Foram analisadas inicialmente 21 variáveis de entrada, que consistiram de dados de níveis médios nas estações fluviométricas de Iraí e Passo São Borja, e precipitação média diária de satélite, todos para dias anteriores a t, inclusive. Dos 22 modelos construídos, foram selecionados 3 modelos para prever os horizontes de 1, 2 e 3 dias, cujos valores de eficiência de Nash-Sutcliffe variaram de 0,97 a 0,91, e os valores de eficiência de Kling-Gupta variaram de 0,98 a 0,88, o que indica bons ajustes dos modelos de previsão.

Os modelos de previsão resultantes apresentaram boas correlações e baixas dispersões: para t+1 houve um excelente ajuste, com apenas alguns valores picos de cheia superestimados na validação;

para t+2 houve novamente um excelente ajuste, desta vez com a subestimativa de alguns picos de cheia, e finalmente, o modelo t+3 apresentou bom ajuste, porém com a ocorrência de superestimava de picos em épocas de estiagem.

Para melhorar o desempenho dos modelos sugere-se adicionar uma estação intermediária no rio Uruguai, bem como estender o período de calibração para abranger mais eventos de estiagem.

M5 model trees possuem uma estrutura de modelos lineares que são de fácil compreensão pelos hidrólogos e tomadores decisões; estes modelos também têm como vantagens serem compactos, de rápida calibração e convergência, conseguem trabalhar com uma grande quantidade de dados com baixo consumo computacional; além disso, estes modelos podem ser aplicados em valores que extrapolam as faixas dos conjuntos de calibração, devendo este último fator ser observado com cautela (Quinlan, 1992).

REFERÊNCIAS

- ADNAN, R.M.; PARMAR, K.S.; HEDDAM, S.; SHAHID, S.; KISI, O. (2021). “*Suspended Sediment Modeling Using a Heuristic Regression Method Hybridized with Kmeans Clustering*”. *Sustainability*, v. 13, 31p.
- ADNAN, R.M.; YUAN, X.; KISI, O. (2018). “*Stream Flow Forecasting of Poorly Gauged Mountainous Watershed by Least Square Support Vector Machine, Fuzzy Genetic Algorithm and M5 Model Tree Using Climatic Data from Nearby Station*”. *Water Resources Management*, v. 32, p. 4469–4486.
- AGÊNCIA NACIONAL DE ÁGUAS – ANA. (2021). “*HIDROWEB: Séries históricas de estações*”. Disponível em: <http://www.snirh.gov.br/hidroweb/serieshistoricas>. Acesso em: maio. 2021.
- BHATTACHARYA, B.; SOLOMATINE, D.P. (2005). “*Neural networks and M5 model trees in modeling water level discharge relationship*”. *Neurocomputing*, v. 63, p. 381–396.
- CPRM – COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. (2020). “*Sistema de Alerta Hidrológico na Bacia do rio Uruguai*”. Relatório anual, 16p.
- CPRM – COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. (2021). “*SACE - Sistema de Alerta de Eventos Críticos*”. Disponível em: <https://www.cprm.gov.br/sace/> Acesso em: maio 2021
- DEO, R.C.; KISI, O.; SINGH, V.P. (2017). “*Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model*”. *Atmospheric Research*, v. 184, p. 149-175.
- FOZ DO CHAPECÓ. (2021). “*Foz do Chapecó: Usina*”. Disponível em: <http://www.fozdochapeco.com.br/usina/>. Acesso em: maio de 2021
- GOYAL, M.K. (2014). “*Modeling of Sediment Yield Prediction Using M5 Model Tree Algorithm and Wavelet Regression*”. *Water Resources Management*, v. 28, p. 1991–2003.
- KHAN, A.S.; SEE, L. (2006). “*Rainfall Runoff modeling using data driven and statistical methods*”. *International Conference on Advances in Space Technologies (IEEE)*.
- KNOBEN, W.J.; FREER, J.E.; WOODS, R.A. (2019). “*Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores*”. *Hydrol. Earth Syst. Sci.*, v. 23, p. 4323–4331.
- KUMAR, A.R.S.; OJHA, C.S.P.; GOYAL, M.K.; SINGH, R.D.; SWAMEE, P.K. (2012) “*Modeling of Suspended Sediment Concentration at Kasol in India Using ANN, Fuzzy Logic, and Decision Tree Algorithms*”. *Journal of Hydrologic Engineering*, v. 17, p. 394-404.

- NALARAJAN, N.A.; MOHANDAS, C. (2015). “*Groundwater Level Prediction using M5 Model Trees*”. Journal of The Institution of Engineers (India): Series A; v. 96, p; 57-62.
- PINTO, E.J. de A.; AZAMBUJA, A.M.S. de; FARIAS, J.A.M.; PICKBRENNER, K.; SALGUEIRO, J. P.de B.; SOUSA, H. R. (2011). “*Atlas pluviométrico do Brasil: isoetas mensais, isoetas trimestrais, isoetas anuais, meses mais secos, meses mais chuvosos, trimestres mais secos, trimestres mais chuvosos*”. Brasília/DF.
- QUINLAN, J.R. (1992). “*Learning with continuous classes*”. Proc. AI, 5th Australian Joint Conference on Artificial Intelligence, Adams, A. and Sterling, L. (eds.), World Scientific: Singapore, p. 343–348.
- REZAIIE-BALF, M.; NAGANNA, S.R.; GHAEMI, A; DEKA, P.C. (2017). “*Wavelet coupled MARS and M5 model tree approaches for groundwater level forecasting*”. Journal of Hydrology, v. 553, p. 356-373.
- ROZANTE, J.R.; MOREIRA, D.S.; GONÇALVES, L.G.; VILA, D. (2010). “*Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America*”. Weather and Forecasting, v. 25, p. 885-894.
- SATTARI, M.T.; PAL, M.; APAYDIN, H.; OZTUK, F. (2013). “*M5 Model Tree Application in Daily River Flow Forecasting in Sohu Stream, Turkey*”. Water Resources, v. 40, n. 3, p. 233–242.
- SOLOMANTINE, D.P.; DULAL, K.N. (2003). “*Model trees as an alternative to neural networks in rainfall runoff modeling*”. Hydrological Sciences Journal, v. 48, p. 455–472.
- SOLOMATINE, D.; SEE, L.M.; ABRAHART, R.J. (2008). “*Data-Driven Modelling: Concepts, Approaches and Experiences, in Practical Hydroinformatics*” em Computational Intelligence and Technological Developments in Water Applications. Org. por Abrahart, Robert J., See, Linda M., Solomatine, Dimitri P. Springer, pp. 17-32.
- Solomatine, D.P.; Maskey, M.; Shrestha, D.L. (2007). “*Instance-based learning compared to other data-driven methods in hydrologic forecasting*”. Hydrological Processes, v. 21.
- SOLOMATINE, D.P.; XUE, Y. (2004). “*M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China*”. ASCE J. Hydrologic Engineering, v. 9, p. 491–501.
- STRAVS, L.; BRILLY, M. (2007). “*Development of a low flow forecasting model using the M5 machine learning method*”. Hydrological Sciences, v. 52, p. 466–477.
- WANG, Y.; WITTEN, I. H. (1996). “*Induction of model trees for predicting continuous classes*”. Hamilton, New Zealand: University of Waikato, Department of Computer Science, 13p.
- WITTEN I.H.; FRANK, E. (2002). “*Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*”. Morgan Kaufmann Publishers, San Francisco, 416p.