

<https://doi.org/10.1590/2318-0331.262120210011>

Modeling coordinated operation of multiple hydropower reservoirs at a continental scale using artificial neural network: the case of Brazilian hydropower system

Simulando operação coordenada de múltiplos reservatórios de geração de energia em escala continental usando redes neurais artificiais. Estudo de caso: sistema hidrelétrico brasileiro

João Paulo Lyra Fialho Brêda¹ , Rodrigo Cauduro Dias de Paiva¹ , Olavo Corrêa Pedrollo¹ ,
Otávio Augusto Passaia²  & Walter Collischonn¹ 

¹Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

²Fundação Cearense de Meteorologia e Recursos Hídricos, Fortaleza, CE, Brasil

E-mails: joaopaulofb@gmail.com (JPLFB), rodrigocpaiva@gmail.com (RCDP), pedrollo.olavo@gmail.com (OCP), otapassaia@gmail.com (OAP), waltercollischonn@gmail.com (WC)

Received: January 27, 2021 - Revised: March 29, 2021 - Accepted: April 02, 2021

ABSTRACT

Reservoirs considerably affect river streamflow and need to be accurately represented in environmental impact studies. Modeling reservoir outflow represents a challenge to hydrological studies since reservoir operations vary with flood risk, economic and demand aspects. The Brazilian Interconnected Energy System (SIN) is an example of a unique and complex system of coordinated operation composed by more than 160 large reservoirs. We proposed and evaluated an integrated approach to simulate daily outflows from most of the SIN reservoirs (138) using an Artificial Neural Network (ANN) model, distinguishing run-of-the-river and storage reservoirs and testing cases whether outflow and level data were available as input. Also, we investigated the influence of the proposed input features (14) on the simulated outflow, related to reservoir water balance, seasonality, and demand. As a result, we verified that the outputs of the ANN model were mainly influenced by local water balance variables, such as the reservoir inflow of the present day and outflow of the day before. However, other features such as the water level of 4 large reservoirs that represent different regions of the country, which infers about hydropower demand through water availability, seemed to influence to some extent reservoirs outflow estimates. This result indicates advantages in using an integrated approach rather than looking at each reservoir individually. In terms of data availability, it was tested scenarios with (WITH_Qout) and without (NO_Qout and SIM_Qout) observed outflow and water level as input features to the ANN model. The NO_Qout model is trained without outflow and water level while the SIM_Qout model is trained with all input features, but it is fed with simulated outflows and water levels rather than observations. These 3 ANN models were compared with two simple benchmarks: outflow is equal to the outflow of the day before (STEADY) and the outflow is equal to the inflow of the same day (INFLOW). For run-of-the-river reservoirs, an ANN model is not necessary as outflow is virtually equal to inflow. For storage reservoirs, the ANN estimates reached median Nash-Sutcliffe efficiencies (NSE) of 0.91, 0.77 and 0.68 for WITH_, NO_ and SIM_Qout respectively, compared to a median NSE of 0.81 and 0.29 for the STEADY and INFLOW benchmarks respectively. In conclusion, the ANN models presented satisfactory performances: when outflow observations are available, WITH_Qout model outperforms STEADY; otherwise, NO_Qout and SIM_Qout models outperform INFLOW.

Keywords: Reservoir outflow estimate; Machine learning.

RESUMO

Reservatórios afetam consideravelmente a vazão dos rios e por isso precisam ser adequadamente representados em estudos de impactos ambientais. Simular defluência de reservatórios representa um desafio para estudos hidrológicos já que sua operação depende do risco de inundações e aspectos econômicos e de demanda. O Sistema Interligado Nacional (SIN) é um exemplo de um sistema único e complexo de operação coordenada composto por mais de 160 grandes reservatórios. Então foi proposto e avaliado uma abordagem integrada para simular defluências diárias da maioria dos reservatórios do SIN (138) usando um modelo de Redes Neurais Artificiais (RNA), distinguindo reservatórios de armazenamento e a fio d'água e testando casos em que defluência e nível estavam disponíveis ou



não como dados de entrada. Além disso, foi investigada a relação entre as variáveis de entrada propostas (14) e a defluência simulada, sendo as variáveis relativas ao balanço hídrico do reservatório, sazonalidade e demanda indireta. Como resultado, foi verificado que as saídas do modelo de RNA foram principalmente influenciadas pelas variáveis locais de balanço hídrico, como a afluência do mesmo dia e a defluência do dia anterior. No entanto, outras variáveis como o nível de 4 grandes reservatórios representativos de diferentes regiões do país (que infere sobre a demanda pela disponibilidade de água para geração de energia) aparentemente teve uma influência considerável nas saídas do modelo. Esse resultado aponta para vantagens em se usar uma abordagem integrada ao invés de olhar cada reservatório isoladamente. Em termos de disponibilidade de dados, foram testados cenários com (WITH_Qout) e sem (NO_Qout and SIM_Qout) defluência e níveis observados como variáveis de entrada do modelo de RNA. Esses 3 modelos de RNA foram comparados com dois benchmarks simples: a defluência é igual a defluência do dia anterior (STEADY) e a defluência é igual a afluência do mesmo dia (INFLOW). Em reservatórios de armazenamento, as estimativas dos modelos atingiram uma mediana do coeficiente de Nash-Sutcliffe (NSE) de 0.91, 0.77 e 0.68 para os modelos WITH_, NO_ e SIM_Qout respectivamente, comparado com uma mediana do NSE de 0.81 e 0.29 para o STEADY e INFLOW respectivamente. Concluindo, os modelos RNA apresentaram performances satisfatórias: quando dados de defluências dos dias anteriores estão disponíveis, WITH_Qout é superior ao STEADY; caso contrário, NO_Qout e SIM_Qout são superiores ao INFLOW.

Palavras-chave: Estimativa de escoamento do reservatório; Aprendizado de máquina.

INTRODUCTION

River dams are structures that affect the natural streamflow and needed to be accurately represented on hydrological studies (Zajac et al., 2017). Reservoirs assume a regulation function, modifying river flow duration curves and attenuating flow peaks (Ayalew et al., 2013; Li et al., 2010; Vogel et al., 2007; Volpi et al., 2018). These structures hold water on land, approximately tripling the mean age of the river water worldwide, which has impact not only on the natural hydrograph but on the sediment flux and the re-oxygenation of surface water (Vörösmarty et al., 1997). Regarding water volume, reservoir operations and water irrigation are responsible to reduce the global river discharge in approximately 2.1% (Biemans et al., 2011).

Worldwide, Lehner et al. (2011) estimated that there are nearly 3 million impoundments larger than 0.1 ha, in consequence, only 36% of rivers longer than 1,000 km are free flowing rivers, i.e. present natural connectivity (Grill et al., 2019). Nowadays, nearly 60,000 large dams (defined as over 15 m height or impounding more than 3 million cubic meters) are listed in the World Register of Dams database of which 20% to 25% are intended for hydropower purposes (International Commission on Large Dams, 2020). And the number of hydropower reservoirs are consistently growing. In 2014, over 3,000 major dams with capacity over 1 MW were planned to be built, mostly in developing countries (Zarfl et al., 2014). The large number of existing and planned reservoirs and their cumulative effect on streamflow justifies an explicit representation of hydropower reservoirs for accurate streamflow estimates.

There are several difficulties in estimating reservoir operation and outflows, as they are influenced by fluctuations on demand, downstream conditions, costs of other sources of energy, bed floor leakage and irregular inflow series. These challenges are more pronounced at large-scale context (national, continental, global). In several cases, operating systems deal with a cascade of reservoirs which demand complex optimization technics to maximize energy production (Liu et al., 2011; Pereira & Pinto, 1985; Zahraie & Karamouz, 2004).

Some studies have proposed simplified operation schemes to estimate reservoir outflows in a large-scale context. For sake of general applicability, these outflow simulations only use inflow and

storage as input (Hanasaki et al., 2006; Shin et al., 2019), although extra information about the reservoir purpose, such as water demand for irrigation and maximum discharge for flood control, might be needed on the optimization process (Haddeland et al., 2006) or as a limiting condition (Zhao et al., 2016). In general, these simplified operation schemes are composed by few linear equations, tested on global/continental hydrological models with a monthly time step and they yield adequate results for a limited data situation and a large-scale context.

On the other hand, machine learning techniques are interesting alternatives to evaluate specific reservoirs and their operation when enough data is provided. Machine learning techniques have ability to represent highly non-linear relations and can autonomously detect patterns and provide predictions. Artificial neural networks (ANN), for example, has proved useful for optimizing reservoir operations (Carneiro & Farias, 2013; Chaves & Chang, 2008; Senthil Kumar et al., 2013) and estimating reservoir inflows (Paz et al., 2008; Valipour et al., 2013). Different machine learning techniques were used to simulate operation of specific reservoirs, such as decision-tree methods (Yang et al., 2016), supporting vector machine, ANN of a single layer and deep learning (Zhang et al., 2018). In a broader scale, Ehsani et al. (2016) proposed a general reservoir operation scheme (GROS) based on ANN, suitable for large-scale modelling. The authors coupled an ANN with a water balance model to simulate reservoir storage and release from reservoir inflow, using the output variables as input for the next steps. GROS presented better performance compared to other simplified methods to estimate operation.

Brazil has a complex and unique operation system where the largest reservoirs were created for hydropower purposes. Nowadays over 60% of the energy production comes from hydraulic sources and this number was much higher (over 90%) no more than 15 year ago (www.aneel.gov.br/dados/geracao). Brazil is the second country in installed hydropower capacity, with nearly 110 GW (International Hydropower Association, 2019). It has more than 160 large hydropower plants on its National Interconnected Energy System (SIN), distributed in different climates. The National Operator of the Energy System (ONS) orientate hydroelectric companies to operate reservoirs connected on SIN based on the country's energy demand, cost of alternative

energy sources and water availability. Thus, Brazil's interconnected system of hydropower reservoirs ends up being an interesting example that require an integrated approach for modeling outflows rather than using local and specific features only.

In this paper, we simulate daily outflows of most of the hydropower reservoirs connected to the SIN using ANN and assess the model capacity to represent a coordinated system. This work characterizes a proof of concept that machine learning techniques can model individual reservoirs of a complex hydropower system in a large-scale context. It was proposed several input features related not only to local reservoir water balance, seasonality, and demand, but also information from other reservoirs. The relevance of each variable was evaluated distinguishing run-of-the-river and storage reservoirs and testing cases whether outflow and water level data were available as input. We have not focused on proposing an optimized operation but simulated daily outflows that can be useful for environmental impact studies on specific river reaches or coupling to hydrological models to understand effects on basin scale.

METHODOLOGY

The national interconnected hydrothermal energy system (SIN)

All Brazilian major dams are operated considering the SIN, which concentrate almost 68% of the national electrical production capacity (Operador Nacional do Sistema Elétrico, 2019). Due to the significantly diverse hydrological characteristics, currently the SIN is divided as regional interconnected subsystems (South, Southeast/Central-west, Northeast, and North), where all the reservoirs within each subsystem are treated as a single equivalent reservoir. Currently, there are more than 160 reservoirs in the SIN (Figure 1). It is coordinated by the ONS, which tries to minimize spills and maximize the hydro electrical energy production, in order to avoid

the utilization of the expensive and air-polluting thermal energy, while it also guarantees consumptive water uses and environmental restrictions (Operador Nacional do Sistema Elétrico, 2019). It is a hard task, because it considers the randomness of the affluent flows, the expansion of the system, flows downstream, future demands, the current reservoirs storage, etc. (Zambon, 2015). A system wide operation strategy prevails over individual ones, and the operation of a given hydropower plant affects other units downstream. The system allows hydropower plants to dispatch and transfer the energy to another region, where the reservoirs are low in storage, avoiding the use of local thermal plants.

Artificial neural network model

ANN is a self-learning technique that estimates an output variable giving a proper amount of data. It is composed by an input layer (formed by the input features), one or more hidden layers, and an output layer, which contains the information learned by the neural network. Input features multiplied by coefficients (synaptic weights) feed an activation function, resulting on “neurons” (nodes) that build a first hidden layer. Following the same steps, neurons of the first hidden layer are multiplied by synaptic weights to generate neurons of a next layer and so on. A trained ANN have optimized weight matrices, which is usually obtained through an iterative method called backpropagation (Rumelhart et al., 1986). This method compares observations to ANN outputs, generating deltas that are propagated backwards on the network to correct the weight matrices, from the output to the input layer. ANN are often referred to arrangements with few hidden layers, as neural networks composed by many hidden layers are often called deep-learning techniques (Shen, 2018).

In this paper, it is proposed an ANN of a “nearly” single hidden layer (Figure 2) and sigmoidal activation functions to predict outflow in time t (OUT_t). The input variables are submitted to feature scaling (normalization) in order to guarantee equal range between features and enable weight comparison. We selected

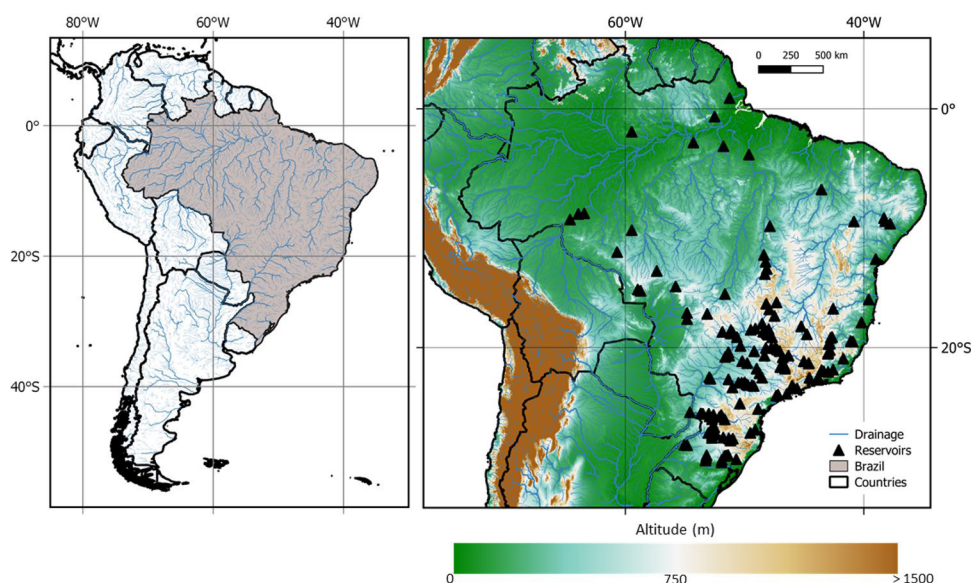


Figure 1. Drainage Network in South America and location of SIN reservoirs in Brazil.

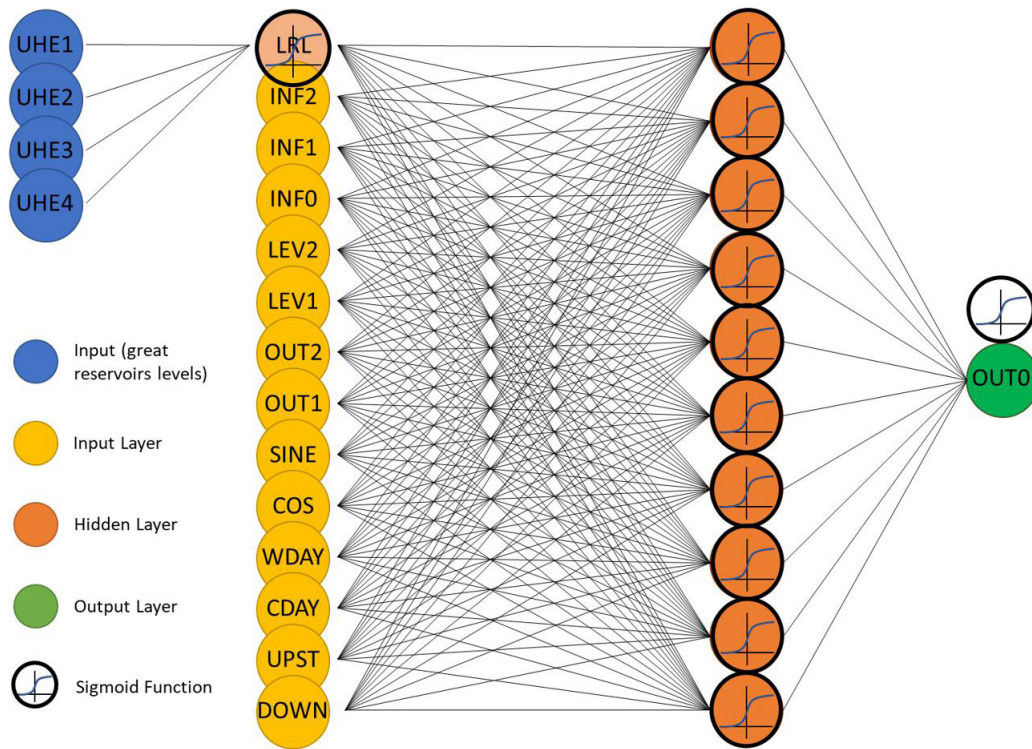


Figure 2. Artificial Neural Network proposition with one hidden layer (+LRL) and a single output which is the outflow in time t (the bias constant was not illustrated).

17 input features (yellow and blue circles on Figure 2) to build the ANN of each reservoir which are explained on the next paragraphs:

- i) water balance variables: inflow (INF), reservoir water surface level (LEV) – which is supposed to be an equivalent alternative to storage – and outflow (OUT);
- ii) time related variables: time of the year (SINE and COS), weekday (WDAY) and continuous time (CDAY);
- iii) state of other reservoirs: upstream and downstream reservoir water levels (UPST and DOWN, respectively) and energy availability (LRL) inferred by SIN reservoir levels (UHE).

The features related to the water balance are often the most relevant variables, sometimes used exclusively as input in an ANN for operation prediction. It was selected as input: inflow on time t (INF0), inflow one and two days before (INF1 and INF2, respectively), water level and outflow one and two days before (LEV1, LEV2, OUT1, OUT2).

Time variables are important to account for seasonality and demand. The time-of-the-year feature consists of a day within a year from 1 to 365 and was adapted to circular representations (i.e. sine and cosine of $2\pi \frac{day}{365}$) in order to provide continuity from one year to the other (e.g. from December 31st to January 1st). The weekday feature is important on hydropower operation since there is a significant reduction of energy demand on weekends as industries stop working. WDAY assumed a value of 1 if it is weekend or holiday and 0 if it is a workday. Continuous time refers to the day since the beginning of a reservoir operation and it was considered to infer eventual changes on the operation due to changes on the energy demands.

Finally, operation of Brazilian hydropower reservoirs is integrated and optimized to generate most energy for the

interconnected system. Thus, it was considered information of other reservoirs as well. It was evaluated the water level of one reservoir upstream and other downstream, when applicable, in order to account for level restrictions, safety operations (e.g. flood control) and maximum energy generation for the cascade system of reservoirs. These variables are the UPST and DOWN features. In addition, the accumulated water volume of other hydropower reservoirs indicates the water availability for energy generation and, consequently, the energy demand for that specific reservoir. For example, if most of the great hydropower plants are operating in high water levels, consequently the country's energy demand is likely to be met, and the operator decision of a specific reservoir might be to reserve water for times of scarcity. Then, it was selected 4 large storage reservoirs from different regions of Brazil (North, Northeast, Southeast and South) as a proxy of the current potential of hydropower energy generation: Tucuruí, Sobradinho, Marimondo and Foz do Areia. These reservoirs are important in terms of energy production and are old enough to provide a long time series of inflow, outflow and water level. The water level in time $t-1$ from these 4 reservoirs (UHE1, 2, 3 and 4) are input to a first node named LRL (large reservoirs level), which consist of the only neuron that composes the first hidden layer, but it can be interpreted as a neuron on the input layer (Figure 2) if we consider that this ANN has only a single hidden layer.

Data acquisition and ANN training

The ANN input data and output observations were obtained through the Brazilian Reservoir Monitoring System (SAR) database from the Brazilian National Water Agency (ANA). The SAR

database provided inflow, outflow, and water level time series from 159 reservoirs connected to SIN by the day it was accessed (Nov/2019) (Table 1). Then, it was selected reservoirs that had at least 8 years of not necessarily consecutive data registered with information of all input features (Figure 3). Thus, 21 reservoirs were discarded, remaining 138 of which 71 are classified by the ONS as run-of-the-river and 67 as storage reservoir.

Data was randomly split into training data (60%), cross-training data (20%) and validation data (20%). The training data are used on the optimization process, but the weight matrices are selected based on the model fitness to the cross-training data. The backpropagation algorithm is fed exclusively with the training data while the cross-training data helps to select unbiased and not overfitted weight matrices on previous algorithm iterations. Then, the validation data, which was left untouched, indicates the ANN performance.

A complexity test was performed in order to identify an adequate number of neurons in the hidden layer. We selected two large and important reservoirs in terms of energy production for the complexity test, each representing a different type of hydropower facility: run-of-the-river reservoirs (Itaipu) and storage reservoirs

(Furnas). It was assessed ANN arrangements with 1, 2, 3, 5, 7, 10, 15 and 20 hidden neurons and in each configuration the ANN was trained 10 times in order to obtain a more reliable and representative result and not depend on the algorithm starting point.

After deciding on an adequate number of hidden neurons, the ANN was specifically trained for each reservoir with the training data. It was used a momentum term of 0.96 (Rumelhart et al., 1986) and an initial learning rate of 0.0001 that is adapted based on the error evolution (Vogl et al., 1988), both techniques applied to accelerate the convergence of the gradient descent. The backpropagation algorithm was run three times to improve chances of reaching a good set of weight matrices, selecting the best results based on the cross-training sample.

ANN assessment

Information about the reservoir level and previous releases are essential to estimate the present outflow. However, this data is not always available, or the latency period is too long for immediate applications. So, we have shaped this ANN model to fit a short

Table 1. Table 1. List of the 159 SIN reservoirs labeled by the first year of data available: i) superscript 1 [< 1991]; ii) superscript 2 [$1991/2000$]; iii) superscript 3 [$2001/2010$]; iv) superscript 4 [> 2010]. The reservoir names in uppercase letter (total of 138) were assessed in this article and the lowercase reservoirs (total of 21) were discarded due to the limited amount of data available.

SIN RESERVOIRS					
CAMARGOS ¹	ESPORA ³	SALTO OSORIO ²	ILHA POMBOS ²	CAÇU ³	CURUA-UNA ²
ITUTINGA ¹	LSOLTEIRA ²	SALTO CAXIAS ²	TOCOS ²	B.COQUEIROS ³	BALBINA ²
FUNIL-MG ³	HENRY BORDEN ²	BARRA GRANDE ³	LAJES ²	FOZ DO RIO CLARO ³	COARACY NUNES ²
FURNAS ¹	BILLINGS ²	CAMPOS NOVOS ³	SANTANA ²	SALTO ³	<i>dardanelos⁴</i>
M.MORAES ¹	GUARAPIRANGA ²	MACHADINHO ³	VIGARIO ³	S.R.VERDINHO ³	<i>ferreira gomes⁴</i>
L.C.BARRETO ¹	PONTE NOVA ²	ITÁ ²	GUILM.AMORIM ²	SEGREDO ²	<i>garibaldi⁴</i>
JAGUARA ¹	EDGARD SOUZA ²	PASSO FUNDO ²	SÁ CARVALHO ²	SANTA CLARA-PR ³	<i>Jirau⁴</i>
IGARAPAVA ²	B. BONITA ²	MONJOLINHO ³	SALTO GRANDE CM ²	FUNDÃO ³	<i>passo são joão⁴</i>
VOLTA GRANDE ¹	BARIRI ²	FOZ CHAPECÓ ³	PORTO ESTRELA ³	JORDÃO ²	<i>simplicio⁴</i>
P.COLOMBIA ¹	IBITINGA ²	QUEBRAQUEIXO ³	CANDONGA ³	SALTO SANTIAGO ²	<i>sto antonio do jari⁴</i>
CACONDE ¹	PROMISSÃO ²	SÃO JOSÉ ³	BAGUARI ³	JAGUARI ²	<i>belo monté⁴</i>
E. DA CUNHA ¹	N.AVANHANDAVA ²	ERNESTINA ²	AIMORES ³	FUNIL ²	<i>cachoeira caldeirão⁴</i>
LIMOEIRO ¹	TRÊS IRMÃOS ²	PASSO REAL ²	MASCARENHAS ²	SANTA CECILIA ²	<i>colider⁴</i>
MARIMBONDO ¹	JUPIÁ ²	JACUI ²	ROSAL ²	PICADA ³	<i>itiquira ii³</i>
A.VERMELHA ¹	PORTO PRIMAVERA ²	ITAUBA ²	STA.CLARA-MG ³	SOBRAGI ²	<i>pimental⁴</i>
S.DO FACÃO ¹	JURUMIRIM ²	D.FRANCISCA ²	IRAPE ³	BOA ESPERANÇA ²	<i>são domingos⁴</i>
EMBORCAÇÃO ²	PIRAJU ³	CASTRO ALVES ³	ITAPEBI ³	SERRA DA MESA ²	<i>são manoe⁴</i>
NOVA PONTE ²	CHAVANTES ²	MONTE CLARO ³	PEDRA DO CAVALO ²	CANABRAVA ³	<i>teles pires⁴</i>
MIRANDA ²	OURINHOS ³	14 DE JULHO ³	RETIRO BAIXO ³	SÃO SALVADOR ³	<i>fontes²</i>
C.BRANCO-1 ³	SALTO GRANDE CS ²	G.PSOUZA ²	TRÊS MARIAS ²	PEIXE ANGICAL ³	<i>nilo peçanha²</i>
C.BRANCO-2 ³	CANOAS II ²	SALTO PILÃO ³	QUEIMADO ³	LAJEADO ³	<i>pereira passos⁴</i>
CORUMBA-3 ³	CANOAS I ²	MANSO ³	SOBRADINHO ²	ESTREITO ³	<i>anta⁴</i>
CORUMBA-4 ³	CAPIVARA ²	ITIQUIRAI ³	LUIZGONZAGA ²	TUCURUI ²	<i>mauá⁴</i>
CORUMBA ²	TAQUARUÇU ²	PONTE DE PEDRA ³	MOXOTO ²	SANTO ANTONIO ⁴	<i>batalha⁴</i>
ITUMBIARA ¹	ROSANA ²	JAUURU ³	P.AFONSO1,2,3 ²	GUAPORE ³	
C.DOURADA ¹	ITAIPU ²	PARAIBUNA ²	P.AFONSO4 ²	SAMUEL ²	
SÃO SIMÃO ¹	G.B.MUNHOZ ²	SANTA BRANCA ²	XINGÓ ²	RONDONII ³	

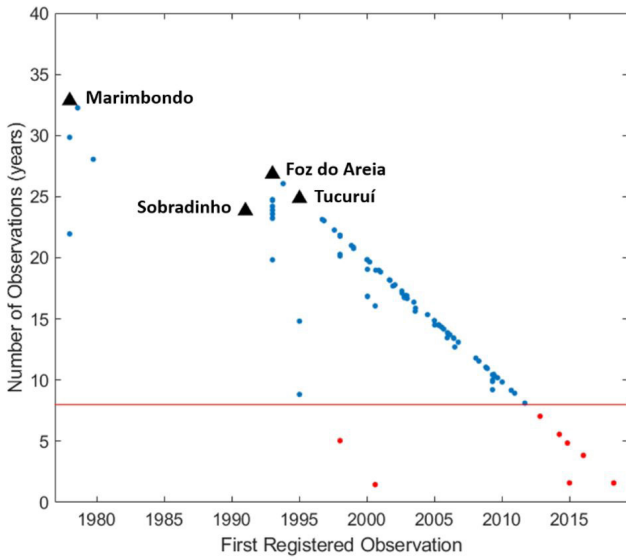


Figure 3. Data quantification of all 159 SIN reservoirs available on the SAR database. Blue dots represent reservoirs selected on this study, red dots are the discarded reservoirs and the red line represents a threshold of eight years of data. The black triangles represent four specific reservoirs named above.

information and large-scale context. For example, this ANN model could have included variables such as energy price and demand or even weather forecast for flooding control, but we rather have proposed water and time related variables that are usually available or even can be simulated by hydrological models. ANN can be used to simulate different scenarios and make projections, and reservoir levels and outflows could be products from previous modeling steps. In order to evaluate the best alternative to simulate reservoir outputs, we proposed 3 different ANN configurations:

WITH_Qout – Reservoirs outflow in previous days are available. Thus, it is trained a complete ANN, with all the features proposed.

NO_Qout – Observed outflow in recent previous days is unavailable. Then, the ANN is trained without the OUT1, OUT2, LEV1 and LEV2 features.

SIM_Qout – Observed outflow in recent previous days is unavailable. The ANN is trained with all the proposed features, including OUT1, OUT2, LEV1 and LEV2 as input. However, OUT1 and OUT2 are those simulated on the last steps and LEV1 and LEV2 are calculated based on a water balance equation using the reservoir specific stage-volume curve. The stage-volume curve was approximated by a 2nd degree polynomial equation using observations of level difference, inflow and outflow. This approach is similar to GROS (Ehsani et al., 2016) however presenting a different ANN structure and more input features.

These three ANN models are compared with two simple benchmarks: outflow of time t is equal to outflow in time $t-1$, which we called steady hypothesis (STEADY); and outflow is equal to inflow (INFLOW). The performances are evaluated in terms of the normalized root mean squared error (NRMSE) and Nash-Sutcliffe efficiency (NSE) which are given by:

$$NRMSE(\%) = \sqrt{\frac{\sum_{i=1,n} (Q_o^i - \bar{Q}^i)^2}{n}} \div \bar{Q}_o \times 100\% \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1,n} (Q_o^i - \bar{Q}^i)^2}{\sum_{i=1,n} (Q_o^i - \bar{Q}_o)^2} \quad (2)$$

where Q is the simulated outflow; Q_o is the observed outflow; \bar{Q}^i is the mean simulated outflow; \bar{Q}_o is the mean observed outflow and n is the number of samples.

17 features were proposed on this ANN arrangement, water balance, time or demand related. We evaluated the influence of every feature on the ANN output in order to understand which input variables are more relevant. We adopted the Weight method (Garson, 1991 apud Gevrey et al., 2003), which basically multiply normalized synaptic weights from layer to layer. In this specific case, there is just one hidden layer, thus there are only two weight matrices (and a third to build the LRL neuron which was not evaluated, see Figure 2). The weight method was conducted as follow: i) Consider the first weight matrix (Θ_h), with dimensions $n \times h$ (n° of features \times n° of neurons); ii) Θ_h is normalized dividing each component by the sum of the coefficients related to each neuron (column); iii) the second weight matrix (Θ_o), with dimensions $h \times 1$ (n° of neurons \times n° of outputs), is also normalized; iv) finally the input features influence on the output is given by the dot product of the normalized weight matrices.

$$\Theta_h = \begin{bmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1h} \\ \theta_{21} & & \ddots & \theta_{2h} \\ \vdots & & & \vdots \\ \theta_{n1} & \theta_{n2} & \dots & \theta_{nh} \end{bmatrix} \quad \Theta_{xh} = [\theta_{x1} \quad \theta_{x2} \quad \dots \quad \theta_{xh}] \quad \Theta_o = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_h \end{bmatrix} \quad (3)$$

$$W_x = \left[\frac{\theta_{x1}}{\sum_{i=1,n} \theta_{i1}} \quad \frac{\theta_{x2}}{\sum_{i=1,n} \theta_{i2}} \quad \dots \quad \frac{\theta_{xh}}{\sum_{i=1,n} \theta_{ih}} \right] \cdot \frac{\Theta_o}{\sum_{i=1,h} \theta_i}$$

where x is the number of input features; h is the number of hidden neurons; Θ_h is the first weight matrix, from the input to the hidden layer; Θ_{xh} is a specific line of Θ_h that correspond to the weights of feature x ; and Θ_o is the second weight matrix, from the hidden to the output layer. As the proposed ANN only has one output, the second weight matrix (Θ_o) has just one column. Thus, the weight influence of feature x in the output (W_x – synaptic weight factor) is given by the dot product of the normalized line correspondent of x on the first weight matrix and the normalized column of the second weight matrix (Equation 3).

RESULTS

The complexity test indicates the number of hidden neurons that should be used in the ANN in order to provide an efficient and accurate performance. Figure 4 demonstrates the ANN performance considering different number of neurons in the hidden layer in terms of NRMSE. The NRMSE have converged to 20% in Furnas and to 4.75% in Itaipu, which suggest that an even larger number of hidden neurons would not improve the ANN

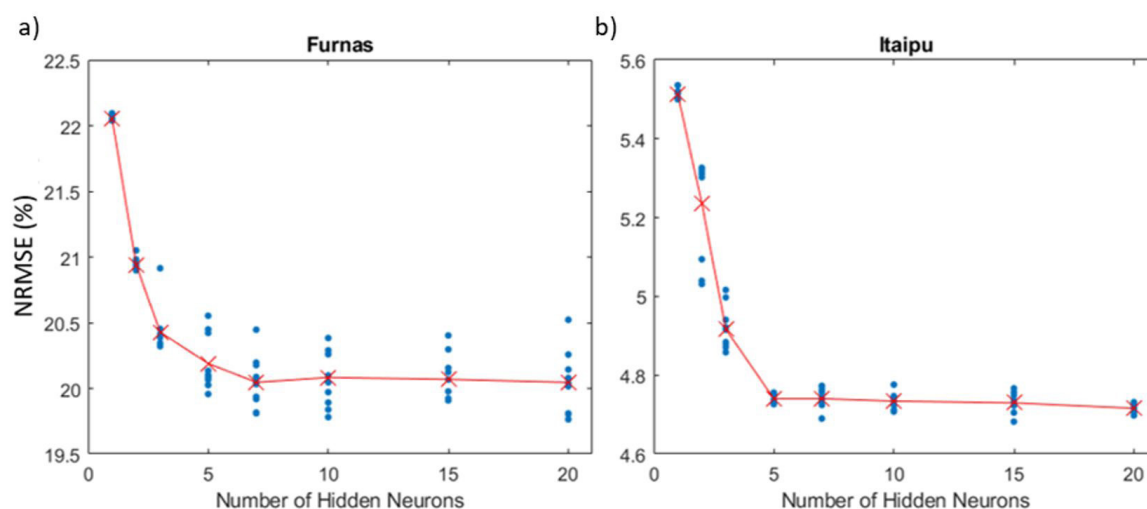


Figure 4. Complexity test which indicates the ideal number of neurons on the hidden layer on two reservoirs. (a) Furnas, which is a storage reservoir; and (b) Itaipu, which is a run-of-the-river reservoir.

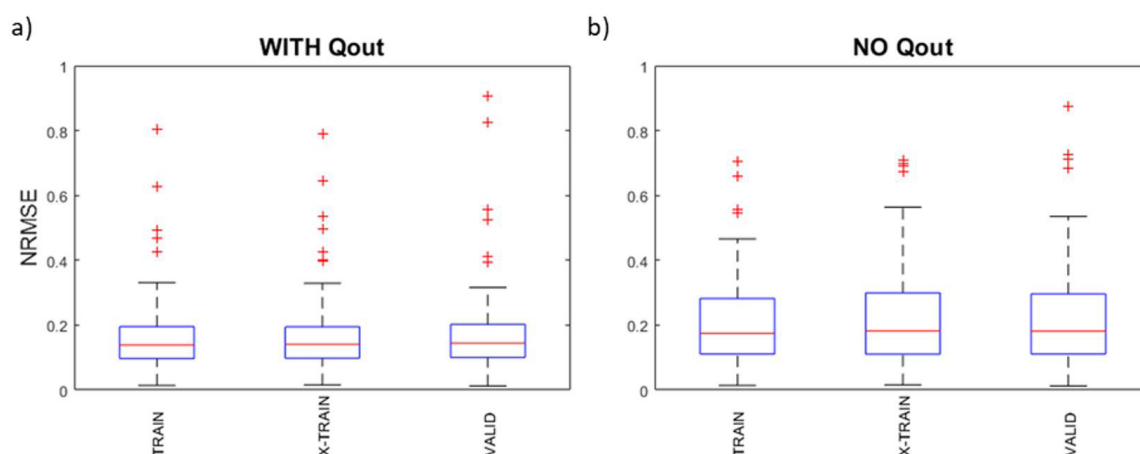


Figure 5. Performance of the training, cross-training and validation data for the (a) WITH_Qout; and (b) NO_Qout ANN models.

performance. The run-of-the-river reservoir has converged with lesser hidden neurons: 5 in Itaipu compared to 7 in Furnas. This was expected since there is almost no range of storage volume for reservoir operation in a run-of-the-river reservoir and, by definition, outflow is mostly governed by inflow. It is common practice to select the smallest number of hidden neurons that provide a good performance. Since the same ANN arrangement are applied to all reservoirs, we adopted an ANN composed by 10 hidden neurons as a cautious alternative.

The ANN training seemed to succeed. First results show that usual regression problems such as overfitting were avoided, since training (TRAIN), cross-training (X-TRAIN) and validation (VALID) data presented similar performances (Figure 5). Also, outflow estimation was reasonable as NRMSEs median were around 14% and 18% for WITH_Qout and NO_Qout ANN models, respectively.

The influence of every input feature on the predicted variable was evaluated in terms of synaptic weights through the Weight method. Although this method has no physical meaning,

it infers about input features that are more influent on the ANN output. Figure 6 illustrate the Weight method results as a box plot using the 138 selected SIN reservoirs as samples. The input variables DOWN and UPST were not considered in this specific test since they are not applicable to all reservoirs.

For the WITH_Qout model, the inflow in time t (INF0) is the feature that has the most influence on the ANN predicted variable (OUT0), followed by the outflow in time $t-1$ (OUT1). However, this sample analysis is biased since outflows of run-of-the-river reservoirs are largely dominated by inflow. Analyzing each type of reservoir individually, we can see that outflow of run-of-the-river reservoirs are indeed governed by inflow; but on storage reservoirs, OUT1 has more influence on the predicted variable than INF0.

For the NO_Qout model, outflow and level are not input variables, thus INF0 largely influences the ANN output. However, looking at storage reservoirs, other features that are related to time or demand presented a high weight factor as well. The LRL feature, for example, is indirectly related to hydropower demand

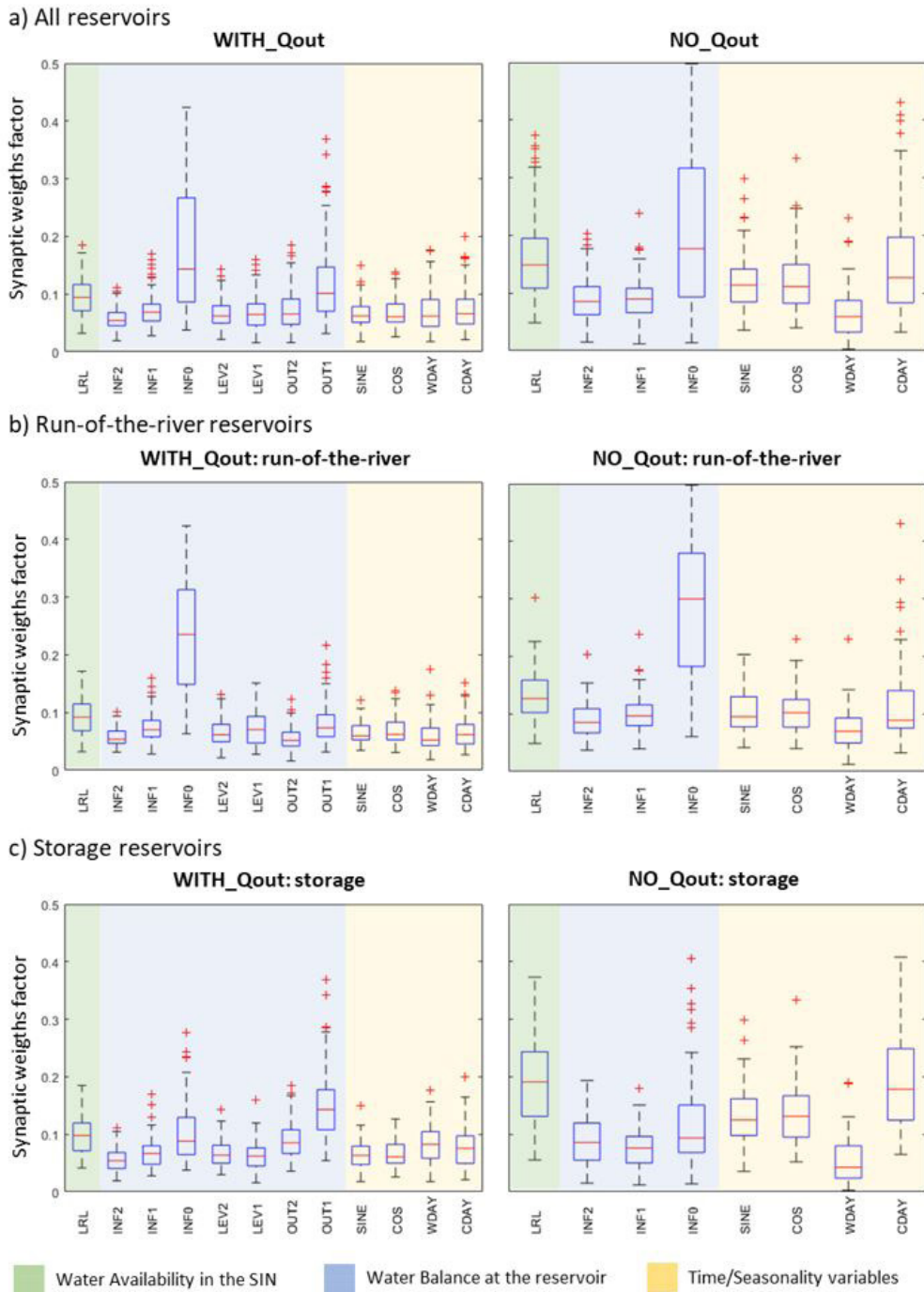


Figure 6. Inference of each input variable influence on the ANN output (outflow in time t) for 138 SIN reservoirs based on the synaptic weights. LRL – Large Reservoirs Level; INF x – Inflow from x days before; OUT x – Outflow from x days before; LEV x – Water level from x days before; SINE and COS – variables related to the day of the year; WDAY – variable related to the day of the week; CDAY – Number of days since the beginning of the reservoir operation.

based on the water level of large SIN reservoirs. This feature presented a relatively high synaptic weight factor, which indicates the benefits of an integrated operation to simulate hydropower reservoirs in Brazil.

The ANN was able to predict relatively well outflow for most SIN reservoirs (Figure 7). When outflow and level observations were available as input to the ANN (WITH_Qout), the sample median NRMSE (NSE) was 14% (0.95) compared to 25% (0.83) and 24% (0.85) of the STEADY and INFLOW benchmarks,

respectively, and the RMSE upper (NSE lower) quartile was 19% (0.90) compared to 38% (0.74) and 59% (0.08). If outflow and level observations were unknown, the ANN performance deteriorates. In general terms, SIM_Qout presented a performance similar to INFLOW, but for storage reservoirs SIM_Qout performance was much superior. NO_Qout presented a superior performance in average compared to INFLOW and STEADY, but for storage reservoirs exclusively, STEADY was a slightly better. Given this ANN arrangement and input features, these results indicate that it is

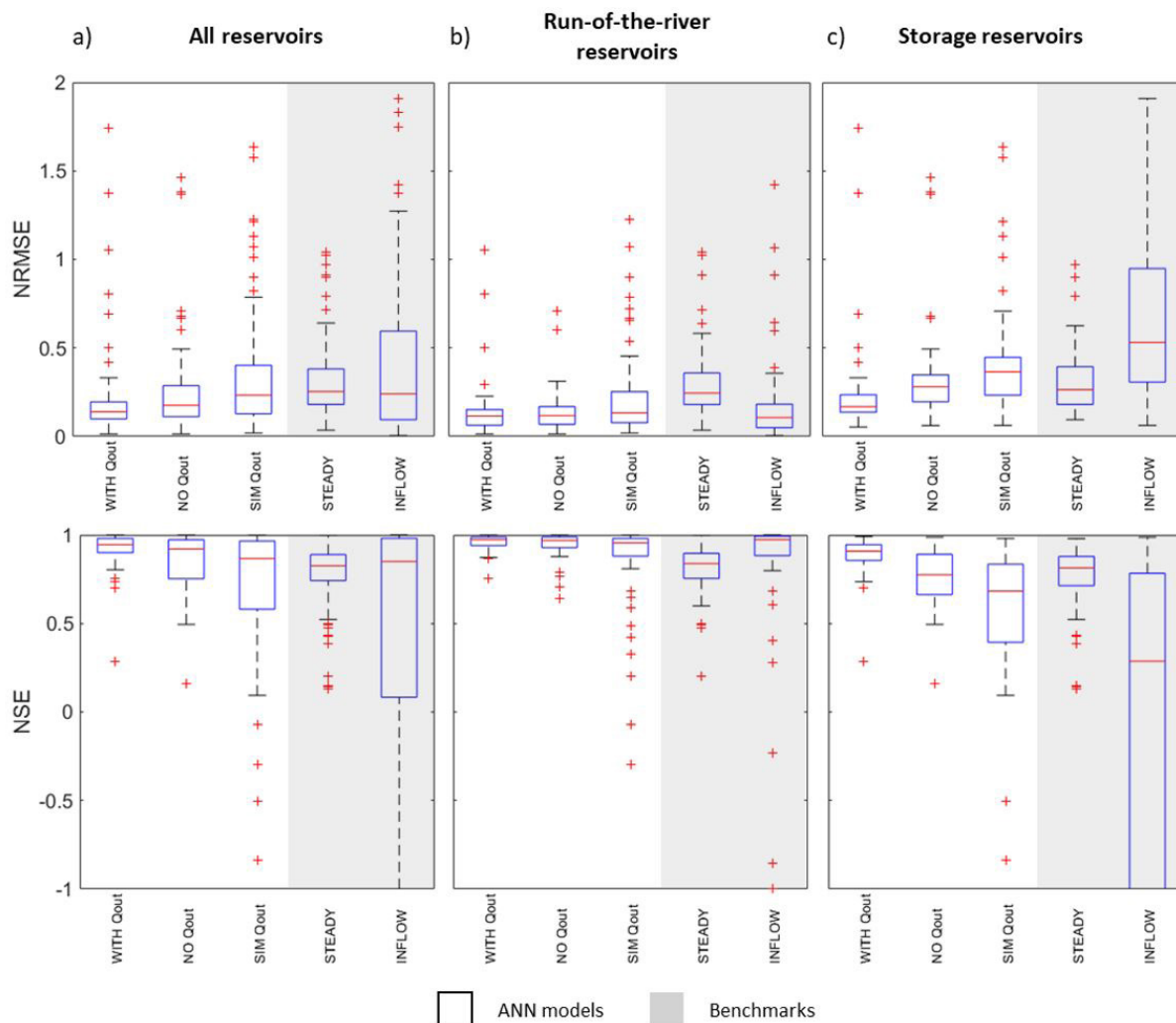


Figure 7. Performance of the complete series (through NRMSE and NSE) of each approach to represent outflow considering (a) all reservoirs; (b) only the run-of-the-river reservoirs; and (c) the storage reservoirs.

better to train an ANN without non-observed variables (OUT and LEV – NO_Qout), rather than training an ANN with all features and use simulated variables as input (SIM_Qout). The latter was adopted by GROS (Ehsani et al., 2016), which provided adequate results in large-scale, however it is important to remark that the authors used a different ANN structure with more hidden layers and less input features.

The performances of outflow predictions were strongly dependent of the type of the hydropower reservoir. Errors were much smaller for run-of-the-river reservoirs compared to storage reservoirs, which was expected since outflow of the former is easier to be estimated. Indeed, the INFLOW assumption presented a performance as good as WITH_Qout for run-of-the-river reservoirs; however, for storage reservoirs its performance considerably worsens. This indicates that storage reservoirs must be well represented in a hydrological model in order to provide accurate estimation of discharge downstream rather than only simulate natural streamflow. Furthermore, NO_Qout provided results similar to STEADY for storage reservoirs. While the former is a good option if no outflow and level data is available on the simulation period, the latter becomes a simple alternative

to represent storage reservoir releases if previous days outflow is known.

Figure 8 illustrates and exemplifies the ANN results through outflow hydrographs of four storage reservoirs: Furnas (a), Jurumirim (b), Sobradinho (c) and Tucuruí (d). It can be seen that ANN provide better estimates compared to INFLOW. The ANN were able to represent reservoirs with high regularization capacity that significantly impacts natural streamflow regimes. In fact, WITH_Qout outflow considerably approximates to observation, while outflows from NO_Qout and SIM_Qout provide adequate seasonal tendencies but are rarely accurate in a daily scale. The weekday feature can be detected on the ANN outflow hydrographs as a 7 days cycle where outflow reduces on the last day; this variable seems to be important to predict outflow especially in Furnas. Particularly on Tucuruí, ANN NRMSE varied from 9% (WITH_Qout) to 17% (SIM_Qout), while INFLOW presented a NRMSE of 33%. In terms of Nash Sutcliff coefficient, ANN performances varied from 0.99 (WITH_Qout) to 0.95 (SIM_Qout), while INFLOW was 0.81. These results indicate that Tucuruí has a relatively well-defined operation and the ANN was capable to capture that.

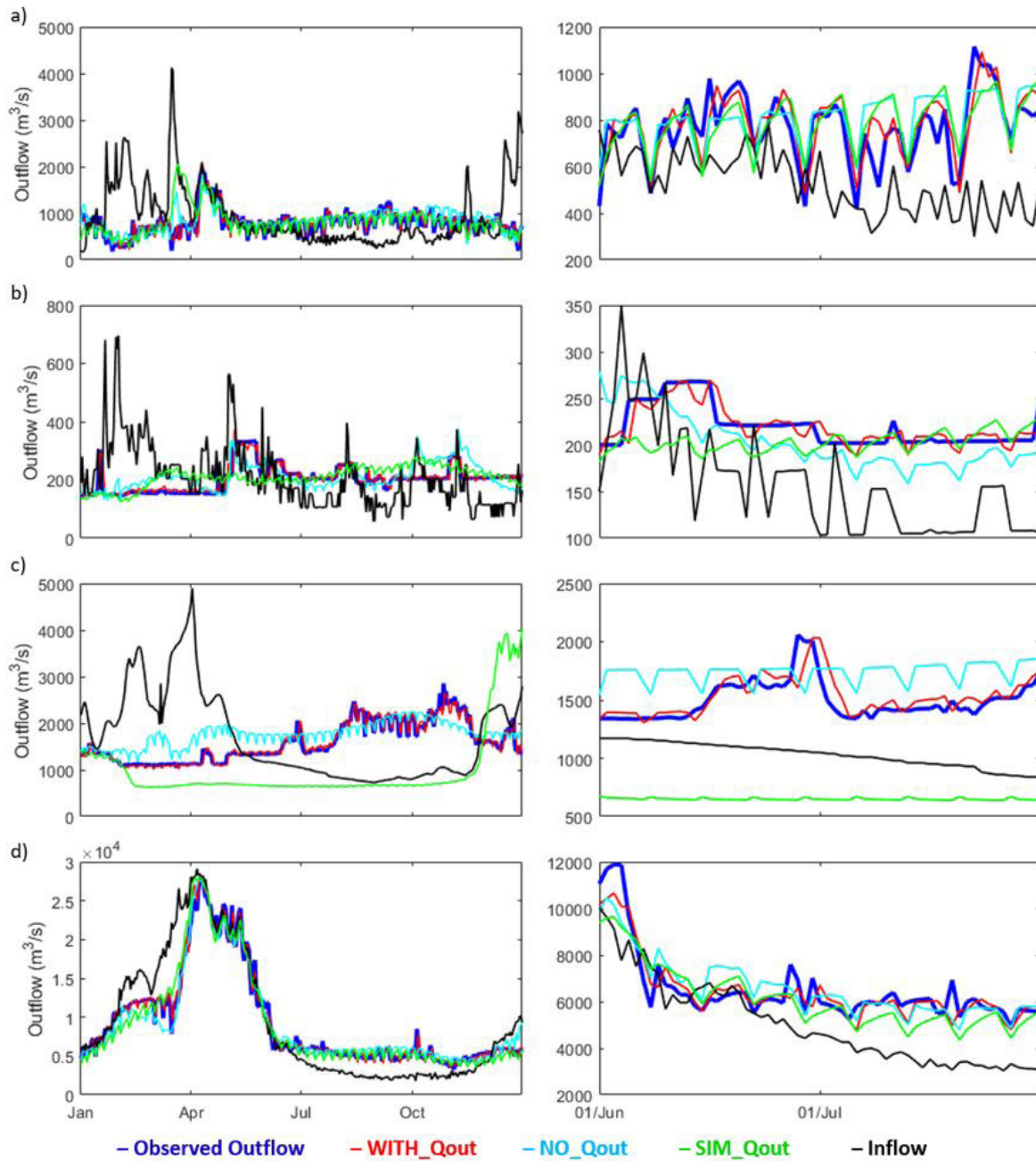


Figure 8. Hydrographs illustrating the results of each ANN model (WITH_, NO_ and SIM_Qout) and the observed outflow and inflow. The hydrographs are for the year 2008 and the following reservoirs: (a) Furnas; (b) Jurumirim; (c) Sobradinho; (d) Tucuruí.

CONCLUSION

This paper offered a first evaluation of the potential of using machine learning techniques to simulate a complex coordinated system of hydropower reservoirs such as the Brazilian SIN. It was proposed an ANN model to predict daily outflow from most of hydropower reservoirs connected to SIN giving water balance, time, and demand input variables.

We used 14 input features and assessed their influence on the model output. As expected, model outflow is mainly influenced by water balance variables, such as the outflow of previous days and inflow. However, features as the water levels in 4 large representative reservoirs (LRL), which infers about hydropower

demand through the water availability in different regions of the country, seemed to influence to some extent reservoirs outflow predictions, indicating advantages of an integrated assessment of SIN reservoirs.

The ANN was trained with (WITH_Qout) and without (NO_Qout and SIM_Qout) reservoir water level and outflow observations as input features to represent usual situations of outflow estimates. The ANN results were compared to two simple benchmarks: outflow is equal to the outflow of the day before (STEADY) and outflow is equal to inflow (INFLOW). There is a significant difference between estimating outflow on run-of-the-river and storage reservoirs since the latter considerably modify the

natural streamflow hydrograph. Using ANN for run-of-the-river reservoirs seemed unnecessary as inflow is almost equal to outflow. However, for storage reservoirs, the ANN model presented a superior performance compared to the benchmarks. When outflow data is available, the WITH_Qout ANN model (median NSE=0.91) outperforms STEADY (median NSE=0.81). When outflow data is not available, ANN performance deteriorates, dropping to NSE medians of 0.77 (NO_Qout) and 0.68 (SIM_Qout), however it is still much superior to INFLOW (median NSE = 0.29).

In conclusion, we have simulated daily outflow from 138 reservoirs individually but using input features that include information from other reservoirs as well and the model performance has been superior to the benchmarks. These results indicate that an integrated approach benefits simulation of a coordinate hydropower system and machine learning techniques are interesting tools for estimating reservoir outflow. However, other ANN arrangements, other possible input features and/or deep learning techniques might improve outflow predictions (Zhang et al., 2018) giving this large amount of data and high non-linearity. The ANN model reasonably approximates to reservoir operations and becomes an interesting tool for large scale analysis of streamflow impacts. However, this is a general model and does not substitute specific reservoirs models that include important local information such as water supply demands, flood control, and environmental legislation for management purposes.

REFERENCES

- Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2013). Exploring the effect of reservoir storage on peak discharge frequency. *Journal of Hydrologic Engineering*, 18(12), 1697-1708. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000721](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000721).
- Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., Von Bloh, W., & Gerten, D. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research*, 47(3), 1-15. <http://dx.doi.org/10.1029/2009WR008929>.
- Carneiro, T., & Farias, C. (2013). Otimização Estocástica Implícita e Redes Neurais Artificiais para Auxílio na Operação Mensal dos Reservatórios Coremas - Mãe d'Água. *Revista Brasileira de Recursos Hídricos*, 18(4), 115-124. <http://dx.doi.org/10.21168/rbrh.v18n4.p115-124>.
- Chaves, P., & Chang, F. J. (2008). Intelligent reservoir operation system based on evolving artificial neural networks. *Advances in Water Resources*, 31(6), 926-936. <http://dx.doi.org/10.1016/j.advwatres.2008.03.002>.
- Ehsani, N., Fekete, B. M., Vörösmarty, C. J., & Tessler, Z. D. (2016). A neural network based general reservoir operation scheme. *Stochastic Environmental Research and Risk Assessment*, 30(4), 1151-1166. <http://dx.doi.org/10.1007/s00477-015-1147-9>.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3), 249-264. [http://dx.doi.org/10.1016/S0304-3800\(02\)00257-0](http://dx.doi.org/10.1016/S0304-3800(02)00257-0).
- Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., Nilsson, C., Olden, J. D., Opperman, J. J., Petry, P., Reidy Liermann, C., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R. J. P., Snider, J., Tan, F., Tockner, K., Valdujo, P. H., van Soesbergen, A., & Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, 569(7755), 215-221. PMID:31068722. <http://dx.doi.org/10.1038/s41586-019-1111-9>.
- Haddeland, I., Skaugen, T., & Lettenmaier, D. P. (2006). Anthropogenic impacts on continental surface water fluxes. *Geophysical Research Letters*, 33(8), 2-5. <http://dx.doi.org/10.1029/2006GL026047>.
- Hanasaki, N., Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models. *Journal of Hydrology*, 327(1-2), 22-41. <http://dx.doi.org/10.1016/j.jhydrol.2005.11.011>.
- International Commission on Large Dams – ICOLD. (2020). *World register of dams*. Retrieved in 2021, January 27, from <http://www.icold-cigb.org>
- International Hydropower Association – IHA. (2019). *Hydropower Status Report 2018*. London.
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, 9(9), 494-502. <http://dx.doi.org/10.1890/100125>.
- Li, X., Guo, S., Liu, P., & Chen, G. (2010). Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty. *Journal of Hydrology*, 391(1-2), 124-132. <http://dx.doi.org/10.1016/j.jhydrol.2010.07.011>.
- Liu, P., Guo, S., Xu, X., & Chen, J. (2011). Derivation of Aggregation-Based Joint Operating Rule Curves for Cascade Hydropower Reservoirs. *Water Resources Management*, 25(13), 3177-3200. <http://dx.doi.org/10.1007/s11269-011-9851-9>.
- Operador Nacional do Sistema Elétrico – ONS. (2019). *Sumário executivo PAR/PEL 2020/2024*. Rio de Janeiro.
- Paz, A., Uvo, C., Bravo, J., & Pedrollo, O. (2008). Previsões de curto prazo de vazão afluente ao reservatório de furnas utilizando redes neurais artificiais. *Revista Brasileira de Recursos Hídricos*, 13(2), 77-88. <http://dx.doi.org/10.21168/rbrh.v13n2.p77-88>.
- Pereira, M. V. F., & Pinto, L. M. V. G. (1985). Stochastic optimization of a multireservoir hydroelectric system: a decomposition approach. *Water Resources Research*, 21(6), 779-792. <http://dx.doi.org/10.1029/WR021i006p00779>.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536. <http://dx.doi.org/10.1038/323533a0>.
- Senthil Kumar, A. R., Goyal, M. K., Ojha, C. S. P., Singh, R. D., Swamee, P. K., & Nema, R. K. (2013). Application of ANN, fuzzy logic and decision tree algorithms for the development of reservoir operating rules. *Water Resources Management*, *27*(3), 911-925. <http://dx.doi.org/10.1007/s11269-012-0225-8>.
- Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558-8593. <http://dx.doi.org/10.1029/2018WR022643>.
- Shin, S., Pokhrel, Y., & Miguez-Macho, G. (2019). High-resolution modeling of reservoir release and storage dynamics at the continental scale. *Water Resources Research*, *55*(1), 787-810. <http://dx.doi.org/10.1029/2018WR023025>.
- Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2013). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, *476*, 433-441. <http://dx.doi.org/10.1016/j.jhydrol.2012.11.017>.
- Vogel, R. M., Sieber, J., Archfield, S. A., Smith, M. P., Apse, C. D., & Huber-Lee, A. (2007). Relations among storage, yield, and instream flow. *Water Resources Research*, *43*(5), 1-12. <http://dx.doi.org/10.1029/2006WR005226>.
- Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., & Alkon, D. L. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics*, *59*(4-5), 257-263. <http://dx.doi.org/10.1007/BF00332914>.
- Volpi, E., Di Lazzaro, M., Bertola, M., Viglione, A., & Fiori, A. (2018). Reservoir effects on flood peak discharge at the catchment scale. *Water Resources Research*, *54*(11), 9623-9636. <http://dx.doi.org/10.1029/2018WR023866>.
- Vörösmarty, C. J., Sharma, K. P., Fekete, B. M., Copeland, A. H., Holden, J., Marble, J., & Lough, J. A. (1997). The storage and aging of continental runoff in large reservoir systems of the world. *Ambio*, *26*(4), 210-219. <http://dx.doi.org/10.2307/4314590>.
- Yang, T., Gao, X., Sorooshian, S., & Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resources Research*, *52*(3), 1626-1651. <http://dx.doi.org/10.1002/2015WR017394>.
- Zahraie, B., & Karamouz, M. (2004). Hydropower reservoirs operation: a time decomposition approach. *Scientia Iranica*, *11*(1), 92-103.
- Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F., & Beck, H. (2017). The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, *548*, 552-568. PMID:28649141. <http://dx.doi.org/10.1016/j.jhydrol.2017.03.022>.
- Zambon, R. C. (2015). A operação dos reservatórios e o planejamento da operação hidrotérmica do Sistema Interligado Nacional. *Revista USP*, *104*(104), 133. <http://dx.doi.org/10.11606/issn.2316-9036.v0i104p133-144>.
- Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L., & Tockner, K. (2014). A global boom in hydropower dam construction. *Aquatic Sciences*, *77*(1), 161-170. <http://dx.doi.org/10.1007/s00027-014-0377-0>.
- Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., & Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, *565*, 720-736. <http://dx.doi.org/10.1016/j.jhydrol.2018.08.050>.
- Zhao, G., Gao, H., Naz, B. S., Kao, S. C., & Voisin, N. (2016). Integrating a reservoir regulation scheme into a spatially distributed hydrological model. *Advances in Water Resources*, *98*, 16-31. <http://dx.doi.org/10.1016/j.advwatres.2016.10.014>.

Authors contributions

João Paulo Lyra Fialho Brêda: Most of the writing, programming, planning, and execution of the paper.

Rodrigo Cauduro Dias de Paiva: Proposed the subject and methodology. Reviewed the manuscript.

Olavo Corrêa Pedrollo: Specialist on Artificial Neural Network that helped with the programming.

Otávio Augusto Passaia: Wrote the section about the hydropower reservoirs. Reviewed the manuscript.

Walter Collischonn: Specialist in hydropower reservoirs. Proposed many input features. Reviewed the manuscript.

Editor-in-Chief: Adilson Pinheiro

Associated Editor: Carlos Henrique Ribeiro Lima