

Deriving reservoir operational behavior with artificial neural networks: the case of Luiz Gonzaga dam, Brazil

Análise do comportamento operacional de reservatórios com redes neurais artificiais: o caso de Luiz Gonzaga, Brasil

Ana Paula Dalcin¹ , Olavo Correa Pedrollo¹ , Juliano Santos Finck¹ ,
Guilherme Fernandes Marques¹ 

¹Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. E-mails: ana.dalcin@ufrgs.br, pedrollo.olavo@gmail.com, juliano.finck@gmail.com, guilherme.marques@ufrgs.br

How to cite: Dalcin, A. P., Pedrollo, O. C., Finck, J. S., & Marques, G. F. (2021). Deriving reservoir operational behavior with artificial neural networks: the case of Luiz Gonzaga dam, Brazil. *Revista de Gestão de Água da América Latina*, 18, e4. <https://doi.org/10.21168/rega.v18e4>

ABSTRACT: Reservoirs are operated following specific policies, constrained by hydrological and structural conditions. When modeling anthropized water systems with reservoirs, the incorporation of existing operating policies is important to improve model capability. However, operating policies are not always available or easy to identify within large-scale multi-reservoir systems, where operation derives from large number of variables and constraints rather than a clear-cut local objective function. This study applies Artificial Neural Networks (ANNs) with the objective of analyzing if local variables (inflow, storage level, and evaporation) of a sub-system part of a large-scale coordinated multi-reservoir system are sufficient predictors of the operational behavior (release decisions) in a daily time step. The sub-system includes the Luiz Gonzaga and Sobradinho reservoirs. Results pointed to a Nash–Sutcliffe efficiency coefficient (NS) of 0.67 to 0.74 and a coefficient of determination (r^2) of 0.75, showing that we can predict the sub-system operational behavior most of the time but with some outflow peaks under predicted.

Keywords: Reservoir operating policy; Reservoir operation emulation; Anthropized water systems simulation.

RESUMO: Reservatórios são operados de acordo com políticas específicas, condicionadas por condições hidrológicas e estruturais. Em simulações hidrológicas de sistemas hídricos antropizados com reservatórios, a incorporação de regras operacionais é fundamental para melhorar a capacidade de modelagem. No entanto, regras de operação nem sempre estão disponíveis ou são fáceis de identificar em sistemas multireservatórios de grande escala, onde a operação deriva de um grande número de variáveis e restrições, em vez de uma função objetivo local bem definida. Este estudo aplica Redes Neurais Artificiais (RNAs) com o objetivo de analisar se variáveis locais (vazão, armazenamento e evaporação) de um subsistema parte de um sistema multireservatório integrado de grande escala são preditores suficientes do seu comportamento operacional (decisões de despacho) em um intervalo de tempo diário. O subsistema inclui os reservatórios de Luiz Gonzaga e Sobradinho. Os resultados apontaram para um coeficiente de eficiência Nash-Sutcliffe (NS) de 0,67 a 0,74 e um coeficiente de determinação (r^2) de 0,75, mostrando que podemos prever o comportamento operacional do subsistema na maior parte do tempo, mas com alguns picos de vazão não previstos.

Palavras-chave: Regra de operação de reservatórios; Emulação da operação de reservatórios; Simulação de sistemas hídricos antropizados.

1 INTRODUCTION

Reservoir operation has several objectives, including narrowing the gap between water availability (system conditions) and demand (system objectives). For hydropower reservoirs, the operation generally follows two main objectives: (a) ensure the delivery of energy following contractual agreements and (b) provide the energy at minimum costs (Wurbs, 1996). Operating rule curves are often used as guideline to indicate the optimal reservoir release or storage volumes at any

Received: December 18, 2020. Revised: March 03, 2021. Accepted: April 17, 2021.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

particular time, allowing it to achieve the desired objectives according to the hydrological conditions (Loucks & van Beek, 2017). For multi-reservoirs systems, coordinated operating approaches usually set reservoirs' release decisions, substituting single reservoir's operating rule curves. This is the case of the Brazilian hydropower system, where the generation and transmission of energy occurs through a national interconnected hydro-thermo-wind system (Organização Nacional de Sistema Elétrico, 2020).

When modeling anthropized water systems with reservoirs, the inclusion of existing operating policies is important to improve model capability. However, in such large-scale, integrated systems, modeling a single reservoir or sub-system presents a challenge given it is driven by a complex interaction of variables and dynamic conditions, rather than a clear-cut local objective function. As operating rules are not always accessible or easy to incorporate into simulation models (Ehsani et al., 2016), the modeler should thrive to identify how these variables draw a boundary condition for system's operation. Recent works have successfully brought data-driven artificial intelligence (AI) models in the field of reservoir operations to identify predictors and extract operating rules from historical data. Among AI models, artificial neural networks (ANN) and support vector machine or regression (SVM or SVR) are the two main models used in this field (Zhang et al., 2018). Yang et al. (2019) point out that although the Recurrent Neural Network (RNN), a class of ANN, still suffers from limitations including longer run times, gradient vanishing, and exploding problems, it is well-suited for simulation of reservoir operation with dynamic process and high dependence on the historical information.

Jain et al. (1999) implemented ANNs to map functional relations between inflow, storage, demand and release of a reservoir operating policy. In Ehsani et al. (2016), ANNs were applied with the objective of developing a general reservoir operation scheme to be used in the simulation of large-scale hydrological models. Data of twelve dams were used to calibrate and validate an ANN model, whose output was the daily release (R_t) and inputs were the inflow up to the past two days (I_t, I_{t-1}, I_{t-2}), the release in the previous two days (R_{t-1}, R_{t-2}) and storage data in the previous day (S_{t-1}). The results pointed to a Nash-Sutcliffe efficiency coefficient (NS) of 0.86, and a coefficient of determination (r^2) of 0.85, concluding that a limited number of input variables is suitable to extract reservoir operations. The trained ANN representing a general reservoir operation was then used to evaluate the hydrological impact of dam sizes and their distribution pattern within a drainage basin.

Recently, Yang et al. (2019) explored the applicability of three RNNs (NARX, LSTM and GA-NARX) to learn operating policies from historical operation data and construct a reservoir operation model to simulate the reservoir outflow. To build the RNN-based reservoir operation model, the inflows of the previous two days ($Q_{in}^{t-2}, Q_{in}^{t-1}$), current day (Q_{in}^t) and the future two days ($Q_{in}^{t+1}, Q_{in}^{t+2}$), and the initial storage (S^{t-1}) were chosen as predictor variables, having the reservoir daily outflow (Q_{out}^t) on the current day as target variable. The results showed that RNNs simulated the outflow of three reservoirs satisfactorily with high NS (greater than 0.65) and low absolute percent bias (less than 10%), with the GA-NARX performing the best one among three RNNs. The RNN-based reservoir operation model was then combined with a geomorphology-based hydrological model (GBHM) to forecast the inflow and build a real-time reservoir operation system.

Although recent advances in emulating reservoir operation have contributed to improve representation of broader operation in simulation models, there is still a gap in understanding if and how reservoir dependent variables can be used to derive large-scale and multi-reservoirs' operation policies. For the Brazilian case, high capacity hydropower plants and their reservoirs are integrated in a hydro-thermal-wind power system through an extensive transmission network. From the total system 164.9 GW installed capacity, 108.5 GW are hydropower, distributed in reservoirs across the country (Organização Nacional de Sistema Elétrico, 2020). This extensively integrated power system exploits natural hydrological variability in the country, reducing more expensive thermal generation in dry seasons/years by transferring hydropower from regions where storage levels are higher. The operation is centralized by an Independent System Operator – ISO, which produces multiple time scale (from monthly down to half hour dispatch) operation considering hydrological conditions, energy demand, fossil fuel prices, scheduled maintenance, deficit cost, entry of new projects and the availability of generation and transmission equipment (Organização Nacional de Sistema Elétrico, 2020).

In such context, modeling a sub-system requires identification of boundary conditions to represent several other decisions and objectives not explicitly modeled. The present paper brings in a contribution to this problem by investigating if more easily available local variables (inflow, reservoir level, and evaporation) are able to explain a sub-system operational behavior with reasonable accuracy, so it can be represented by models with less complexity and cost, also avoiding relying on highly uncertain exogenous variables (e.g., fossil fuel prices). We applied feed-forward Artificial Neural Networks (ANNs) to learn from daily historical

operation data and model the daily release (outflow). The chosen sub-system includes two large hydropower reservoirs connected to the country's integrated power system, Luiz Gonzaga and Sobradinho, located in the São Francisco River Basin, Brazil. The reservoirs are operated under both hydropower generation and local urban water supply purposes. The Luiz Gonzaga reservoir has recently received a water transfer project to deliver water to semiarid areas of the northeastern region of Brazil. Given the relevance of some key input variables in this strategic system are still unknown, its investigation will support the emulation of operating rules as boundary conditions, helpful to optimize future water transfers, improve the understanding of regional water use trade-offs and minimize conflicts.

2 STUDY AREA

The São Francisco River Basin is located in the Northeast region of Brazil (Figure 1A), draining 639,219 km² (7.5% of the country) with an average flow of 2,850 m³/s (Comitê da Bacia Hidrográfica do Rio São Francisco, 2016). The hydropower generation is a relevant aspect of this basin, which has 9 hydropower plants part of the integrated hydro-thermal-wind power system, adding up to total hydropower capacity of 10,556.2 MW (Organização Nacional de Sistema Elétrico, 2020). Sobradinho (SB) is the largest reservoir of the basin with a total storage of 34,116 hm³. Further downstream, another important reservoir, Luiz Gonzaga (LG), is operated in coordination, with a total storage of 10,782 hm³ (Figure 1B).

The São Francisco River Basin has also recently received a large-scale Water Transfer Project (PISF) to deliver water to neighboring semiarid areas. The water transfer system is divided in two branches, North and East. The Eastern branch draws water from Luiz Gonzaga reservoir (São Francisco Basin) and delivers it to the Boqueirão reservoir (Alto Paraíba Basin) covering a total distance of 220 km (Agência Nacional de Águas, 2016). As the operation of the PISF water transfers project depends on the state of Luiz Gonzaga (LG) reservoir (mainly storage and releases), any further analysis of the water transfer demands simulation of Luiz Gonzaga operation, which depends on the operation of other reservoirs connected to the integrated hydro-thermal-wind power system.

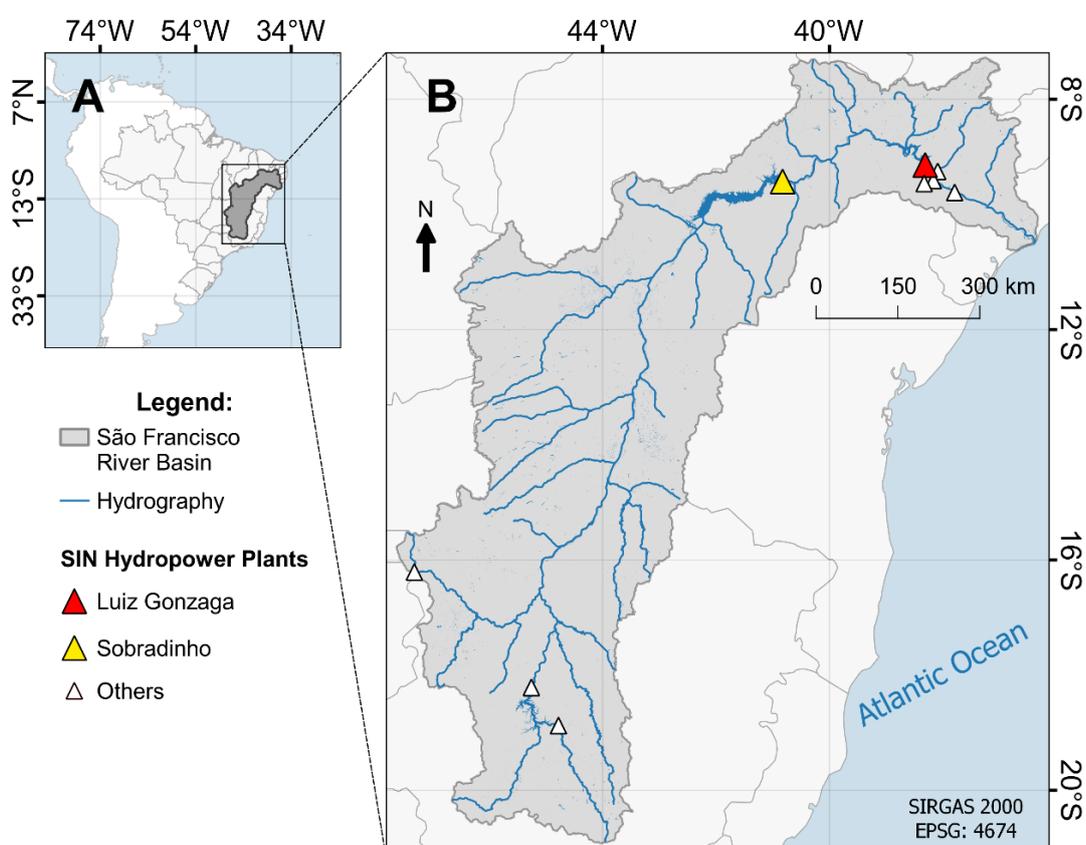


Figure 1. São Francisco River Basin, Luiz Gonzaga and Sobradinho reservoirs location (Agência Nacional de Águas, 2020)

3 METHODOLOGY

3.1 Artificial Neural Networks

Artificial Neural Networks (ANN) are data-driven models capable of distributed and parallel processing, composed of interconnected input nodes, processing elements (neurons), and output nodes (Hecht-Nielsen, 1990). In a feed-forward or multilayer perceptron (MLP) ANN, nodes and neurons are arranged in a unidirectional manner: input layer, hidden layers and output layer (Tangri et al., 2008). The input layer has only input nodes, whose function is to distribute the input signals via connections to each hidden neuron of the next hidden layer. Each connection has a weight (w) adjusted via training, and each neuron has an activation function. Neurons process the sum of signals from previous layers and independent terms (also known as biases) with the activation function, and each output neuron sends its output signal to the respective output node. Figure 2 illustrates a single hidden layer.

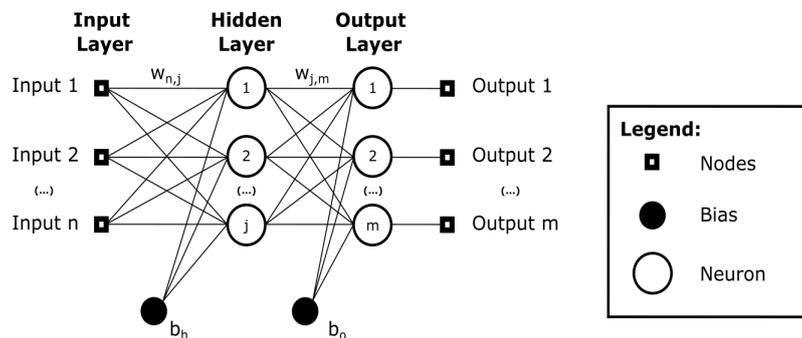


Figure 2. ANN general (n - j - m) architecture.

The neural net used in this study consists of three layers: an input layer, a hidden layer, and an output layer. The activation function used in each perceptron is unipolar sigmoid and the input variables were linearly scaled to fit the limits [0,1]. We used a training algorithm based on backpropagating the errors (Rumelhart et al., 1986) and subsequently adjusting the weights based on the delta rule (Widrow & Hoff, 1960). The following accelerating methods were applied: updating weights based on the errors of all training samples, applying momentum factor and dynamic learning rate (Vogl et al., 1988).

ANNs with a three-layer structure are universal approximators (Hornik, 1991) but overfitting during training can compromise ANN generalization capability to new data. To attempt preventing overfitting, we applied a cross-validation approach by partitioning the dataset into three sets: training, validation, and verification (Hecht-Nielsen, 1990). The cross-validation approach consists of training the model with the training set, checking the errors after each cycle with a separate set (the validation set) and ceasing training when this validation set error stops improving. Weight adjustment is only undertaken with the training set error. The final model performance is checked with the verification set. An overall flowchart explaining the cross-validation method applied in this study is presented in Appendix A.

Training algorithms based on the descending gradient depend highly on the initial conditions. An approach to reach a nearer global solution is to repetitively train the ANN with different initial conditions. We thus applied an evolutionary algorithm (EA) to search for the initial synaptic weights that yielded the smallest validation set error. Additionally, we investigated the number of hidden neurons necessary to present a validation set performance similar to an oversized ANN, preventing the development of overly complex models. In Lucchese et al. (2020), a complexity analysis was performed based on a set of repetitions to select the network configuration that resulted in the best validation performance. Here, the complexity analysis was carried out according to the optimal initial weights obtained by the EA result (oriented search).

The EA approach consisted in, first, running a preliminary inspection to find the smallest validation error regardless of the number of hidden neurons. Then, for each complexity configuration analyzed, the next run searched for 20 individuals (set of initial weights) that yielded validation errors smaller than the preliminary search to compound the first generation. Subsequent generations were created from the 10 fittest individuals (smallest validation errors) in the previous generation plus 10 mutations for each one of these individuals

until the converged criteria was satisfied. The complexity configuration that improved the performance without increasing the complexity was chosen to compose the model.

3.2 ANN Architecture Proposal and Input data

The variables (input and output nodes) proposed to build the ANN model were based on the local state variables, which are likely to explain the operation of the LG reservoir in a daily time step (Figure 3), and they could be used to build boundary conditions to a LG simulation. Given the presence of a significantly larger reservoir immediately upstream (Sobradinho – SB), which releases provide most of the inflow to LG, it was chosen to provide an extra exogenous variable as predictor to LG operation. Different ANN configurations are then proposed, associating the variables selected with the release decisions at LG. The ANN models are then run and their performance is evaluated, in order to check on the contribution of each predictor variable to the behavior of LG reservoir and produce a LG reservoir operating function.

The M1 configuration has 4 input variables used as predictors of LG release decisions in the current day ($Q_{out, LG, t}$): the LG reservoir inflow in the current day ($Q_{in, LG, t}$), the LG evaporation in the current day ($e_{LG, t}$), and both SB and LG reservoirs storage levels at the previous day ($L_{SB, t-1}$; $L_{LG, t-1}$).

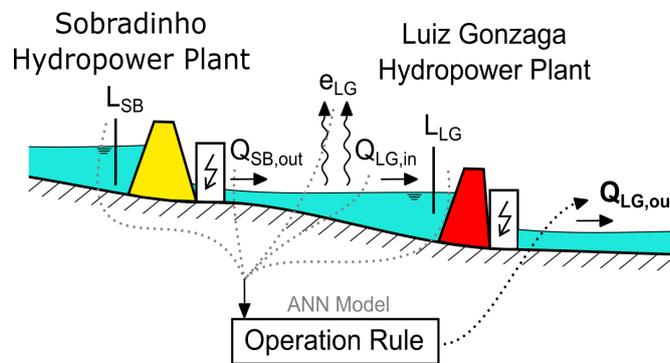


Figure 3. Model conceptualization.

We conducted a correlation analysis and determined that SB outflow affects the LG inflow with a lag time of 3 days based on a $Q_{in, LG, t} \times Q_{out, SB, t}$ (see correlogram in Figure 4). From this result, the second ANN configuration (M2) additionally computed the outflows of the Sobradinho reservoir in the current and four previous days ($Q_{out, SB, t}$; $Q_{out, SB, t-1}$; $Q_{out, SB, t-2}$; $Q_{out, SB, t-3}$; $Q_{out, SB, t-4}$) as predictors of the LG release decisions ($Q_{out, LG, t}$). These flows act as future inflows forecast for the LG reservoir operation. In total, 9 input variables were used in the M2 configuration: the reservoir inflow in the current day ($Q_{in, LG, t}$), the evaporation in the current day ($e_{LG, t}$), both reservoirs previous time levels ($L_{SB, t-1}$; $L_{LG, t-1}$), and the outflow of the Sobradinho reservoir in the current and previous times ($Q_{out, SB, t}$; $Q_{out, SB, t-1}$; $Q_{out, SB, t-2}$; $Q_{out, SB, t-3}$; $Q_{out, SB, t-4}$).

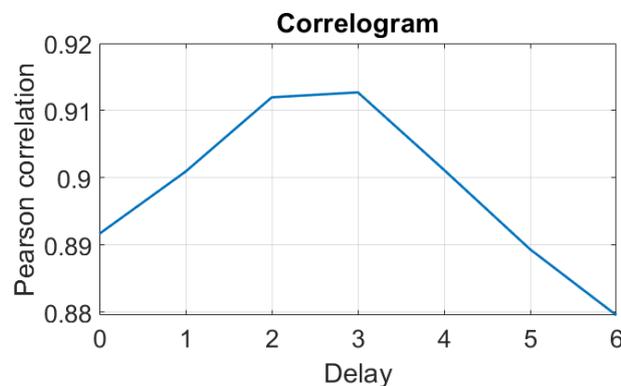


Figure 4. Correlogram between SB outflow (t+Delay) and LG inflow (t) .

Table 1 summarizes the main parameters used to build the ANN configurations. The index *LG* represents Luiz Gonzaga reservoir, the index *SB* represents the Sobradinho reservoir, and the index *t*

represents the time-step (daily). We have linearly scaled down patterns and target to (0, 1) and used sigmoid unipolar activation function.

Table 1. ANN parameters

Parameter	Description
Input variables (M1)	$L_{SB,t-1}; L_{LG,t-1}; Q_{in,LG,t}; e_{LG,t}$
Input variables (M2)	$L_{SB,t-1}; L_{LG,t-1}; Q_{in,LG,t}; e_{LG,t};$
	$Q_{out,SB,t}; Q_{out,SB,t-1}; Q_{out,SB,t-2}; Q_{out,SB,t-3}; Q_{out,SB,t-4}$
Output variable	$Q_{out,LG,t}$
Architecture	4 - j - 1 (M1) and 9 - j - 1 (M2)
Number of hidden neurons (j)	chosen according to the the complexity analysis
Activation function	sigmoid unipolar; $f(x) = \frac{1}{1 + e^{-x}}; [0,1]$
Input data scaling	linear (amplitude)
Data time step	daily

Energy generation objectives are usually dependent on the hydropower plant firm energy, demands, energy prices and contractual market agreements. As this information is dynamic in the Brazilian energy market and it involves other variables and constraints (e.g., fossil fuel prices, deficit cost, availability of generation and transmission equipment, etc.), its representation is beyond the scope of this paper, and should be investigated in future developments.

3.3 Division of Training, Validation and Verification Sets

The available data period spans from 01/01/1991 to 01/01/2019, resulting in 10,228 valid registers (red and blue lines of Figure 5), where a register is an individual input-output pair. Regarding the evaporation time series, the closest evaporation monitoring station to the study area (Paulo Afonso station) had its missing values filled by other three near monitoring stations (Água Branca, Cabrobó and Pão de Açúcar). Table 2 summarizes the input variables information.

Table 2. Input variables information.

Variable	Description	Unit	Station Id	Reference
L_{SB}	Sobradinho Level	m	19121	(Agência Nacional de Águas, 2020)
L_{LG}	Luiz Gonzaga Level	m	19122	(Agência Nacional de Águas, 2020)
$Q_{in,SB}$	Sobradinho Inflow	m ³ /s	19121	(Agência Nacional de Águas, 2020)
$Q_{in,LG}$	Luiz Gonzaga Inflow	m ³ /s	19122	(Agência Nacional de Águas, 2020)
$Q_{out,LG}$	Luiz Gonzaga Outflow	m ³ /s	19122	(Agência Nacional de Águas, 2020)
e_{LG}	Luiz Gonzaga Evaporation	mm	82986 - Paulo Afonso	(Instituto Nacional de Meteorologia, 2020)
			82989 - Água Branca	
			82886 - Cabrobó	
			82990 - Pão de Açúcar	

A drought in 2012 resulted in a reservoir operating policy change from 2013 (Silva et al., 2020). To select a representative period with both low and high flows, the verification period was set from 31/07/2000 to 05/06/2007 and from 05/01/2015 to 02/02/2017 (blue line), while the remaining period (red line) compounded the training and validation sets (Figure 4).

Training and validation registers were sampled from the red line period in two steps. First, we assigned input-output pairs to the training set that contained at least one maximum or minimum value of each input and output variables domain. As the ANNs are not suitable to extrapolate the domain of the training set, assigning extreme values to the training set attempts to prevent extrapolations of the domain. Secondly, based the systematic sampling technique (Cochran, 1977), the remaining pairs were sorted in relation to the output variable and alternately assigned to each set, leading to the approximate ratio composition of 50% for training and 50% for the validation.

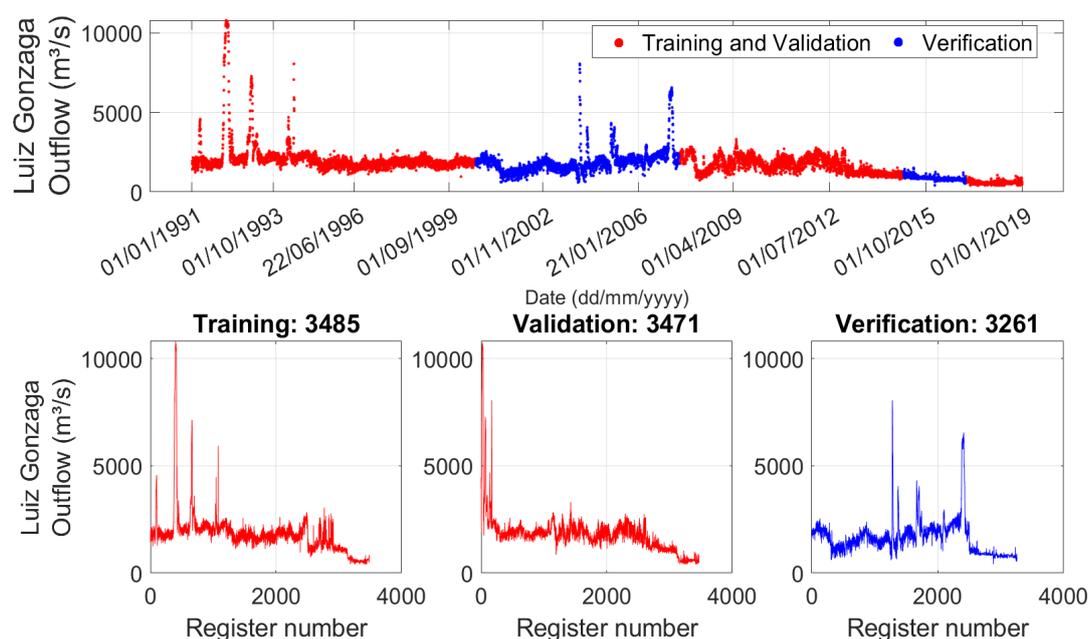


Figure 5. Total available data and organization of the training, validation and verification sets.

3.4 Evaluation metrics

The performance analysis of each data set was based on the Nash-Sutcliffe coefficient (1), coefficient of determination (2), mean error (3), mean absolute error (4), root mean square error (5), and maximum error (6).

$$NS = 1 - \frac{\sum_{i=1}^I (O_i - P_i)^2}{\sum_{i=1}^I (O_i - \bar{O})^2} \quad (1)$$

$$r^2 = \left(\frac{\sum_{i=1}^I (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^I (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^I (P_i - \bar{P})^2}} \right)^2 \quad (2)$$

$$ME = \frac{\sum_{i=1}^I (O_i - P_i)}{I} \quad (3)$$

$$MAE = \left| \frac{\sum_{i=1}^I (O_i - P_i)}{I} \right| \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^I (O_i - P_i)^2}{I}} \quad (5)$$

$$Emax = \max(O_i - P_i) \quad (6)$$

Where the index i represents the register number, O_i is the observed output at i ; P_i is the predicted output at i ; \bar{O} is the mean of the observed output; \bar{P} is the mean of the predicted output; I is the total number of samples. Both NS and r^2 coefficients vary from 0 to 1, where values closer to 1 indicate better fitting (Krause et al., 2005).

4 RESULTS

4.1 Complexity Analysis

Figure 6 shows that one hidden neuron would produce a final error higher than if the complexity of the network was increased. However, a number of hidden neurons higher than 2 would not improve the performance considerably. We selected 2 neurons to compose the hidden layer for both ANN proposals (M1 and M2).

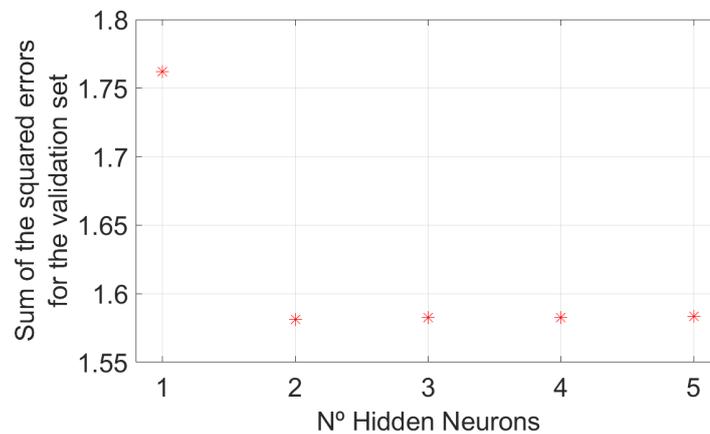


Figure 6. Complexity analysis.

Figure 7 presents the trained ANN obtained in this study, where the connections represent the synaptic weights, whose thickness are associated to the magnitudes relative to other connections impinging to the same neuron; and color to the signs of the values. The thicker the connections, the greater the contribution to the respective neuron. The red lines indicate positive weights and the blue, negative ones.

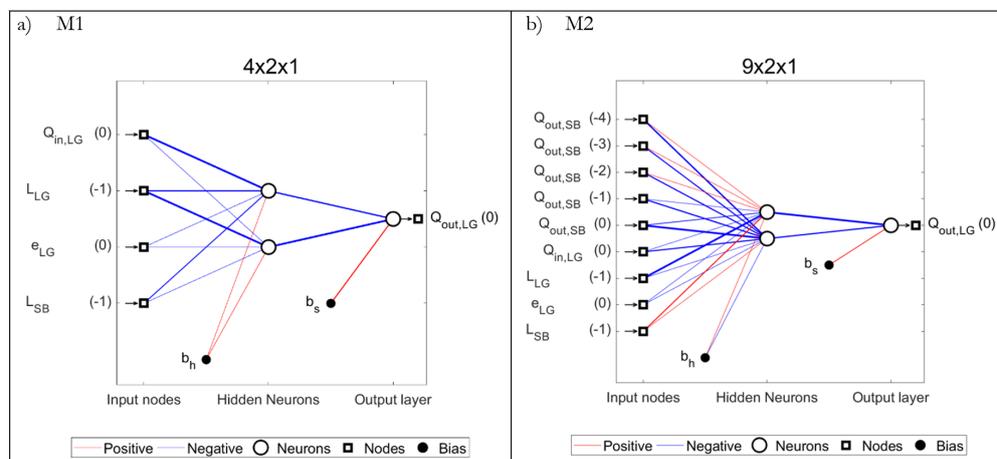


Figure 7. Representation of the nodes and neurons connections of the respective trained models.

4.2 Performance Analysis

Figure 8 and Figure 9 present the comparison between observed and predicted outflows for the training, validation and verification sets, while Table 3 presents the evaluation metrics for both proposals (M1 and M2).

The ANN model was able to satisfactorily reproduce the main reservoir release behavior for both configurations, M1 and M2 (Figure 8a and Figure 9a). Blue and red lines have similar slopes for the training and validation sets (Figure 8b and Figure 9b), indicating that everyday occurrences and outflow peaks could be learned from the training set and generalized to the verification set. The inflows forecast of the ANN M2 model improved its overall performance.

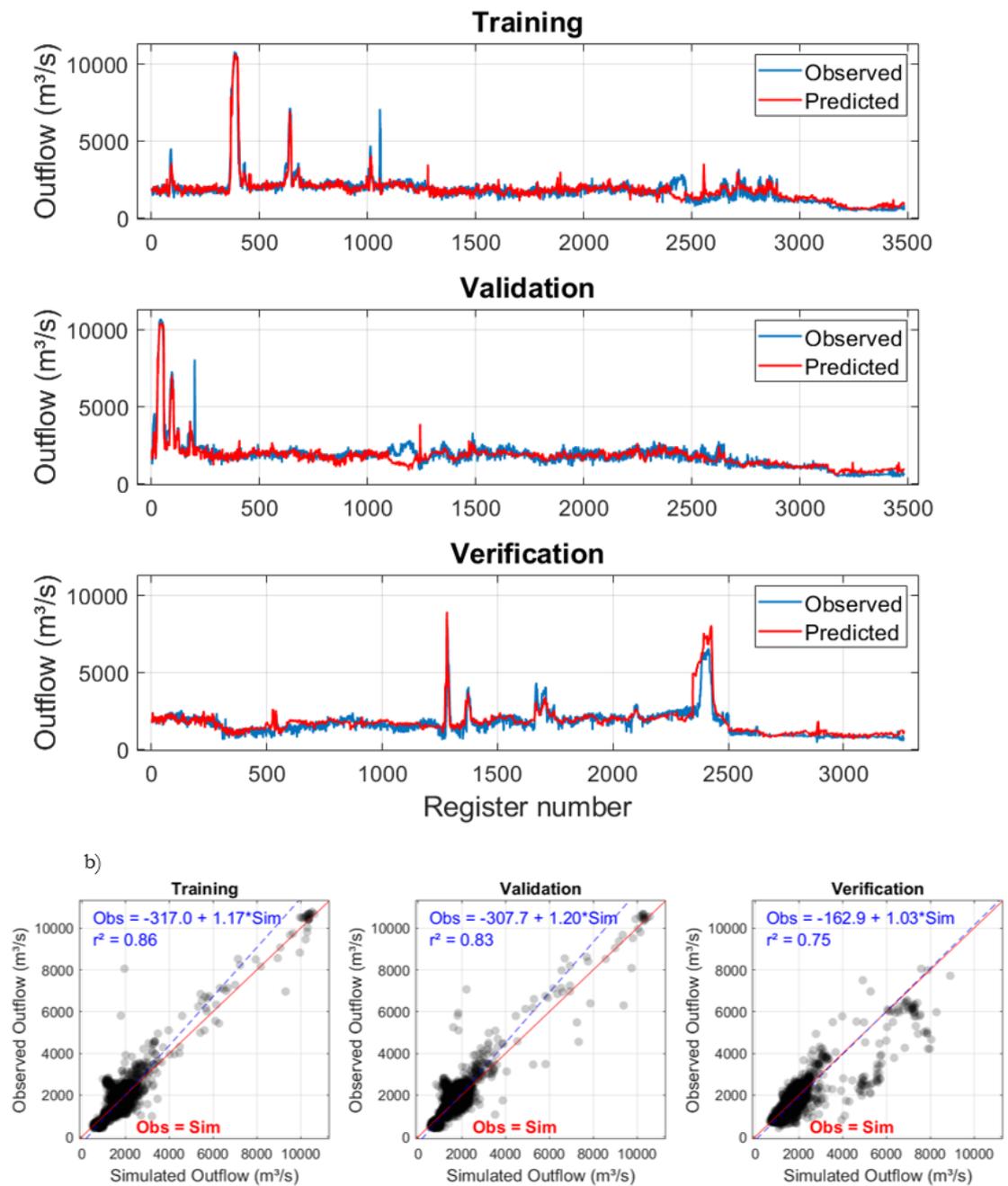


Figure 8. Comparison between observed and predicted outflow (ANN configuration M1).

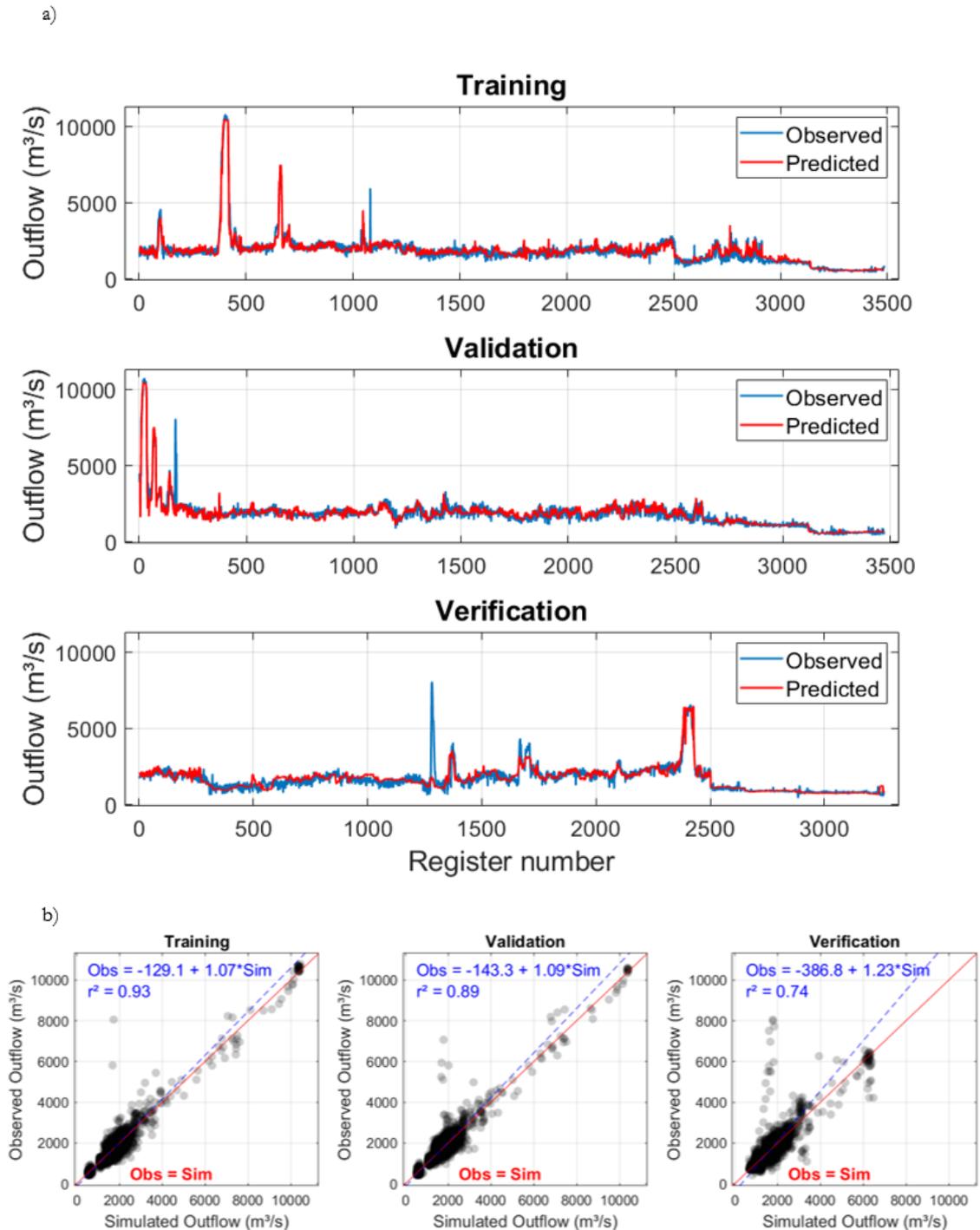


Figure 9. Comparison between observed and predicted outflow (ANN configuration M2).

The verification set result in Figure 8b and Figure 9b shows that the ANN model satisfactorily predicts the operation behavior for events near the average for both models (cluster of points near 2000 m³/s). While ME coefficient indicates an underestimation for M1 (ME = -109.5 m³/s), the M2 positive mean error (ME = 5.1 m³/s) indicates an overall overestimation. However, for observed LG outflows above 3,000 m³/s, M1 resulted in a higher dispersal around the 1:1 line, showing a higher trend for overestimating some peaks, while M2 underpredicted some extreme events. One possible explanation is the training prioritizes fitting to more frequent events rather than the extreme ones due to most outflows being under 3000 m³/s.

Intervening factors driving the operation of the whole hydropower integrated system may produce isolated disturbances that cannot be explained by the local variables modeled. It can be noticed by some peaks that were not predicted, particularly the peak of the 1100th sample of the training sets and the 200th sample

peak of the validation sets (Figures 8a and 9a). For the verification set, the M2 model could not predict the 1300th event, which reflected on the reduction of the performance of the Emax coefficient. It indicates that, while the additional input variables (inflows' forecast) allow better prediction of everyday events (also producing an overall better performance), it does not help in reproducing events that cannot not be explained by the local input variables (level, inflow and evaporation).

Table 3. Performance analysis.

Performance	Training		Validation		Verification	
	M1	M2	M1	M2	M1	M2
NS	0.852	0.935	0.850	0.898	0.669	0.745
r ²	0.852	0.935	0.853	0.899	0.751	0.751
ME (m ³ /s)	-3.2	-0.3	58.084	28.7	-109.5	5.1
MAE (m ³ /s)	269.2	189.9	272.5	193.6	281.5	218.4
RMSE (m ³ /s)	398.0	269.1	394.4	318.7	478.2	420.1
Emax (m ³ /s)	4854.3	4255.9	6103.5	6331.9	2586.8	6286.4

4.3 Input variables relevance

The interactions between explanatory variables and the complexity of functions implemented by neural networks make analytical studies of each input variable contribution to explain the output a challenge (Lek et al., 1996). We proposed a leave-one-out performance reduction analysis as an indicative approach of the relevance of the input variable to explain the resulting output (Lek et al., 1996; Tan & Beklioglu, 2006). The model M1 was retrained leaving one input variable out at each simulation. The simulations were performed in triplicate runs to avoid and catch any possible result instability. For each run, the resulting sets' performance was computed and used as indicative of the contribution of the missing input variable to explain the observed operation at LG. Table 4 shows the resulting r² and NS performance reduction for the verification sets. All run triplicates presented stable results with less than 0.3% of difference.

Table 4: Performance reduction analysis for each input variable removed.

Input variable removed	r ²	Performance reduction (%)	NS	Performance reduction (%)
None	0.751	-	0.669	-
$Q_{in, LG, t}$	0.280	62.7	0.267	60.1
$L_{LG, t-1}$	0.726	3.3	0.600	10.3
$L_{SB, t-1}$	0.735	2.1	0.598	10.6
$e_{LG, t}$	0.740	1.5	0.640	4.3

Removal of any input variable caused a performance decrease. The LG inflow ($Q_{in, LG, t}$) had the greatest performance reduction (62.7% for r² and 60.1% for NS), which indicates it is an influent predictor to the LG reservoir operation. The LG level at previous day ($L_{LG, t-1}$) and the SB level at previous day ($L_{SB, t-1}$) presented similar performance reduction, which indicates the SB reservoir may constraint the LG operation in some circumstances. As pointed out by the resolution n° 2081/2017 (Agência Nacional de Águas, 2017), a minimum storage must be maintained in LG reservoir when the SB reservoir is operating under low (20 to 60% of the active storage) and very low storage conditions (less than 20% of the active storage). According to the historical operation, the SB reservoir operation in the very low storage conditions corresponds to 15% of the available data, while the low storage condition corresponds to 48%. The resulting weights produced by the models incorporate this behavior.

The evaporation ($e_{LG, t}$) is the least influent predictor but still important given both reservoirs are located in a semi-arid region. When retraining the M1 model without the evaporation input, the

performance (r^2) has reduced in 4.3% for the verification set. The official modeling tools applied by the ISO for dispatching reservoir power production throughout the national integrated system consider the evaporation when formulating operational decisions (Eletrobras, 2020), which explains the importance of this variable to the final result.

4.4 Reservoir Operating Function

We finally combined the resulting connection weights to produce a reservoir operating function. Table 5 presents the final values of the scaling parameters and weights for the M1 model, which can be used to produce a non-linear function to simulate LG outflow releases. The M1 model was chosen due to its simpler representation (not requiring incorporation of forecast models), while still able to perform well.

Table 5. Variable weights for the M1 configuration.

Layer	Variable ($X_{in,t}$)	Scale		Weight (W)	
		min	max	j1	j2
Input	$L_{SB,t-1}$	342.6	432.1	1.0275	-3.4502
	$L_{LG,t-1}$	269.4	334.6	9.2536	-22.1410
	$e_{LG,t}$	0	24.5	0.0554	-0.8347
	$Q_{in,LG,t}$	255.6	12102	-2.3648	-2.3766
	$b_{in,j}$	-		-3.3019	10.654
Output	$Q_{out,LG,t}$	366.3	11875	-8.0675	-14.105
	b_{out}	-		5.5659	

The operating function is represented by Equations 7 to 12, having the LG outflow (X_{out}) represented as a function of the input variables ($X_{in,t}$), and the trained weights (W).

$$X_{esc,in} = \frac{X_{in,t} - \min(X_{in,t})}{\max(X_{in,t}) - \min(X_{in,t})} \quad (7)$$

$$l_j = \sum_{in} (X_{esc,in} * W_{in,j}) + b_{in,j} \quad \forall j \quad (8)$$

$$h_j = \left(\frac{l}{1 + e^{-l_j}} \right) \quad \forall j \quad (9)$$

$$r = \sum_j (h_j * W_{j,out}) + b_{out} \quad (10)$$

$$s = \left(\frac{l}{1 + e^{-r}} \right) \quad (11)$$

$$X_{out} = s * [(\max(X_{out}) - \min(X_{out}))] + \min(X_{out}) \quad (12)$$

The index “in” represents the input variables ($L_{SB,t-1}$; $L_{LG,t-1}$; $Q_{in,LG,t}$; $e_{LG,t}$), while the index “out” represents the output variable ($Q_{out,LG,t}$). X_{in} is the value of the input variables, $X_{esc,in}$ is the scaled value of the input variables, b represents the independent terms, W is the weight of each input/output variable, h_j is the signal in the hidden neuron j , s is the signal of the output neuron, and X_{out} is the value of the output variable after scale reverting.

To apply the results in a simulation model, the example procedure in Figure 10 can be followed. The reservoirs' level initial conditions are the first input, with the outflow at Luiz Gonzaga ($Q_{out,LG,t}$) calculated based on Equations 7 to 12 and Table 4. Given the resulting reservoir level depends on the release decision, the level at Luiz Gonzaga at each time t can be calculated based on the mass balance Equation 13 and the

reservoir level \times storage curve (14) (Câmara de Comercialização de Energia Elétrica, 2020), serving as input to the next time-step run. Figure 11 presents the simulation result considering a synthetic inflow and evaporation time-series based on the observed years from 2000 to 2019. For Sobradinho, we generated level time series simulating low and very low conditions, which restricted the LG operation.

$$S_t = S_{t-1} + Q_{in,t} - e_t - Q_{out,t} \quad (13)$$

$$L_t = 374 + (1.40 \times 10^{-3}) \cdot S_t - (5.35 \times 10^{-8}) \cdot S_t^2 + (1.16 \times 10^{-12}) \cdot S_t^3 - (9.55 \times 10^{-18}) \cdot S_t^4 \quad (14)$$

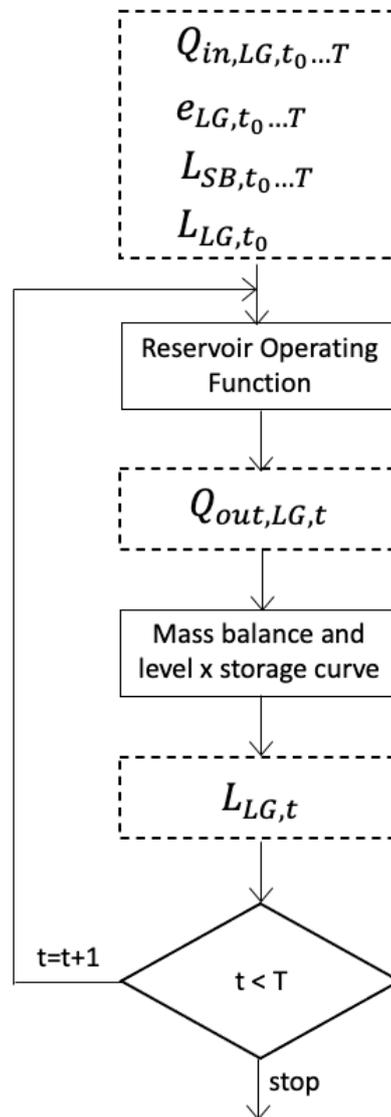


Figure 10. Simulation approach.

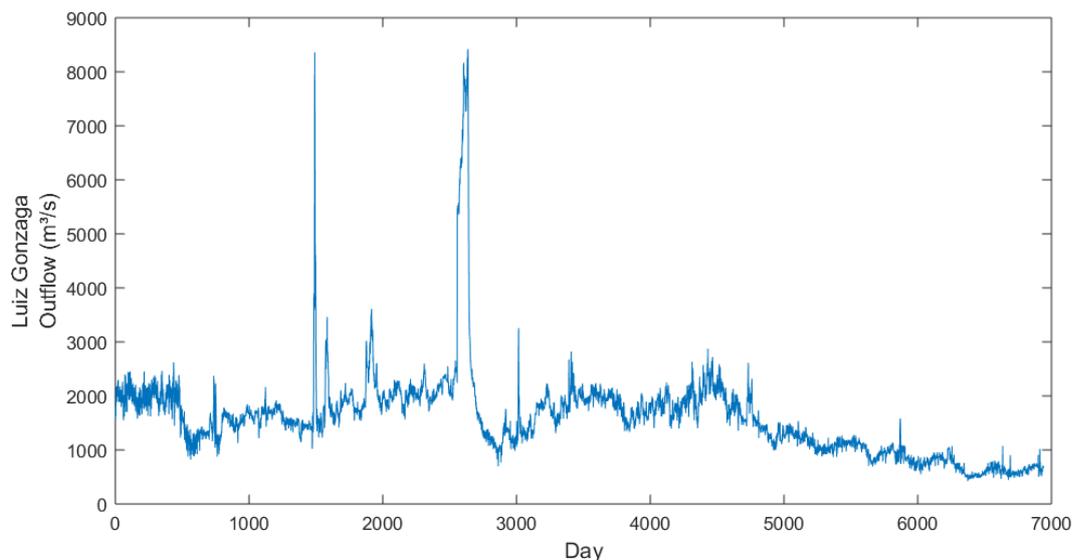


Figure 11. Simulation result.

The simulated outflow average obtained was $1786 \text{ m}^3/\text{s}$ with a standard deviation of $1002 \text{ m}^3/\text{s}$. As comparison to the observed outflow for the same period whose inflow and evaporation input time-series were originated, the values are close to the observed average and standard deviation ($1779 \text{ m}^3/\text{s}$ and $970 \text{ m}^3/\text{s}$).

5 CONCLUSION

In this study we investigated the applicability of ANNs to emulate a sub-system reservoir operation part of a large-scale hydropower system. We conclude that the ANNs are able to reflect and capture the main historical reservoir operation behavior, and the sub-system local variables inflow, level, and evaporation are able to explain about 75% of the operating behavior, producing accurate results for most part of the observed events, especially when the system is closer to average flow conditions. Although the ANN M2 model presented better overall performance, the fewer M1 model variables are able to explain the operational behavior with reasonable accuracy and less complexity.

Some outflow peaks could not be predicted by the ANN models, which suggests that intervening factors driving the operation of the whole integrated system may produce isolated disturbances that cannot be explained by the local variables. Variables representing energy demand, fuel prices, deficit cost, entry of new projects and availability of generation and transmission equipment, which are not easily available as time-series, could be investigated in future studies regarding the increment in the model complexity and performance.

Despite the limitations, our methods and results are useful to construct a daily-time reservoir operation model, which is helpful to simulate the reservoir outflow under different scenarios. It is especially important to the study area, where reservoirs for hydropower generation and more recently inter-basin water supply purposes have a strategic relevance, so that understanding and identifying water use trade-offs by applying simulation models improves the capacity of managing future water use conflicts.

REFERENCES

- Agência Nacional de Águas – ANA. (2016). *Nota técnica conjunta 2/2016/COSER/SRE/SAS*. Brasília. Agência Nacional de Águas - ANA. (2017). *Resolução nº 2.081, de 04 de dezembro de 2017*. Brasil. Retrieved from: <https://arquivos.ana.gov.br/resolucoes/2017/2081-2017.pdf>
- Agência Nacional de Águas – ANA. (2020). *SAR - Sistema de acompanhamento de reservatórios*. Retrieved from <https://www.ana.gov.br/sar/>
- Câmara de Comercialização de Energia Elétrica – CCEE. (2020). *Hydroedit - apoio à leitura de arquivos*. Retrieved in 2020, December 10, from https://www.ccee.org.br/portal/faces/pages_publico/o-que-fazemos/como_ccee_atua/precos/deck_de_precos?_afzLoop=21599906344280&_adf.ctrl-state=4wc76zq6g_1#!%40%40%3F_afzLoop%3D21599906344280%26_adf.ctrl-state%3D4wc76zq6g_5

- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, USA: John Wiley & Sons, Inc.
- Comitê da Bacia Hidrográfica do Rio São Francisco – CBHSF. (2016). *Plano de Recursos Hídricos 2016/2025*. Brazil. Retrieved in 2020, December 10, from <https://cbhsaofrancisco.org.br/a-bacia/>
- 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>
- Eletrobras. (2020). *Newave and Decomp models - reservoir operating policies*. Retrieved in 2020, February 1, from http://www.cepel.br/pt_br/produtos/decomp-modelo-de-planejamento-da-operacao-de-sistemas-hidrotermicos-interligados-de-curto-prazo.htm
- Hecht-Nielsen, R. (1990). *Neurocomputing*. United States: Addison-Wesley.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251-257. [http://dx.doi.org/10.1016/0893-6080\(91\)90009-T](http://dx.doi.org/10.1016/0893-6080(91)90009-T)
- Instituto Nacional de Meteorologia – INMET. (2020). *Dados meteorológicos*. Retrieved in 2020, February 1, from <http://www.inmet.gov.br/portal/>
- Jain, S. K., Das, A., & Srivastava, D. K. (1999). Application of ANN for Reservoir Inflow Prediction and Operation. *Journal of Water Resources Planning and Management*, 125(5), 263-271. [http://dx.doi.org/10.1061/\(ASCE\)0733-9496\(1999\)125:5\(263\)](http://dx.doi.org/10.1061/(ASCE)0733-9496(1999)125:5(263))
- Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89-97. <http://dx.doi.org/10.5194/adgeo-5-89-2005>
- Lek, S., Belaud, A., Baran, P., Dimopoulos, I., & Delacoste, M. (1996). Role of some environmental variables in trout abundance models using neural networks. *Aquatic Living Resources*, 9(1), 23-29. <http://dx.doi.org/10.1051/alr:1996004>
- Loucks, D. P., & van Beek, E. (2017). *Water resource systems planning and management*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-44234-1>
- Lucchese, L. V., de Oliveira, G. G., & Pedrollo, O. C. (2020). Attribute selection using correlations and principal components for artificial neural networks employment for landslide susceptibility assessment. *Environmental Monitoring and Assessment*, 192(2), 129. <http://dx.doi.org/10.1007/s10661-019-7968-0>
- Organização Nacional de Sistema Elétrico – ONS. (2020). *Brazil hydro-thermal-wind system*. Retrieved in 2020, February 1, from <http://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>
- 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>
- Silva, T., Callado, N., Souza, V., & Vasconcelos, M. (2020). Respostas da qualidade da água e fitoplânctons à redução de vazão e recepção de cargas de sedimentos no reservatório de Xingó/AL. *Revista de Gestão de Água Da América Latina*, 17(1), 15-0. <http://dx.doi.org/10.21168/rega.v17e15>
- Tan, C. O., & Beklioglu, M. (2006). Modeling complex nonlinear responses of shallow lakes to fish and hydrology using artificial neural networks. *Ecological Modelling*, 196(1-2), 183-194. <http://dx.doi.org/10.1016/j.ecolmodel.2006.02.003>
- Tangri, N., Ansell, D., & Naimark, D. (2008). Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. *Nephrology, Dialysis, Transplantation*, 23(9), 2972-2981. <http://dx.doi.org/10.1093/ndt/gfn187>
- 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>
- Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. *IRE Wescon Convention Record*, 8(4), 96-104.
- Wurbs, R. A. (1996). *Modeling and analysis of reservoir system operation*. Prentice Hall.
- Yang, S., Yang, D., Chen, J., & Zhao, B. (2019). Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *Journal of Hydrology (Amsterdam)*, 579(May), 124229. <http://dx.doi.org/10.1016/j.jhydrol.2019.124229>
- 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 (Amsterdam)*, 565(August), 720-736. <http://dx.doi.org/10.1016/j.jhydrol.2018.08.050>

Author contributions:

Ana Paula Dalcin: performed the activities of scientific construction of the article, data search, development of the model, result analysis and paper writing.

Olavo Correa Pedrollo: participated in the scientific construction of the article, guided the analysis of the results and reviewed the text manuscript.

Juliano Santos Finck: performed the activities of model development, implementation of the evolutionary algorithm and model calibration, result analysis and paper writing.

Guilherme Fernandes Marques: participated in the scientific construction of the article, guided the analysis of the results and reviewed the text manuscript.

APPENDIX A

The flowchart (Figure 1A) summarizes the ANN modeling procedures of training with cross-validation and testing with the verification set.

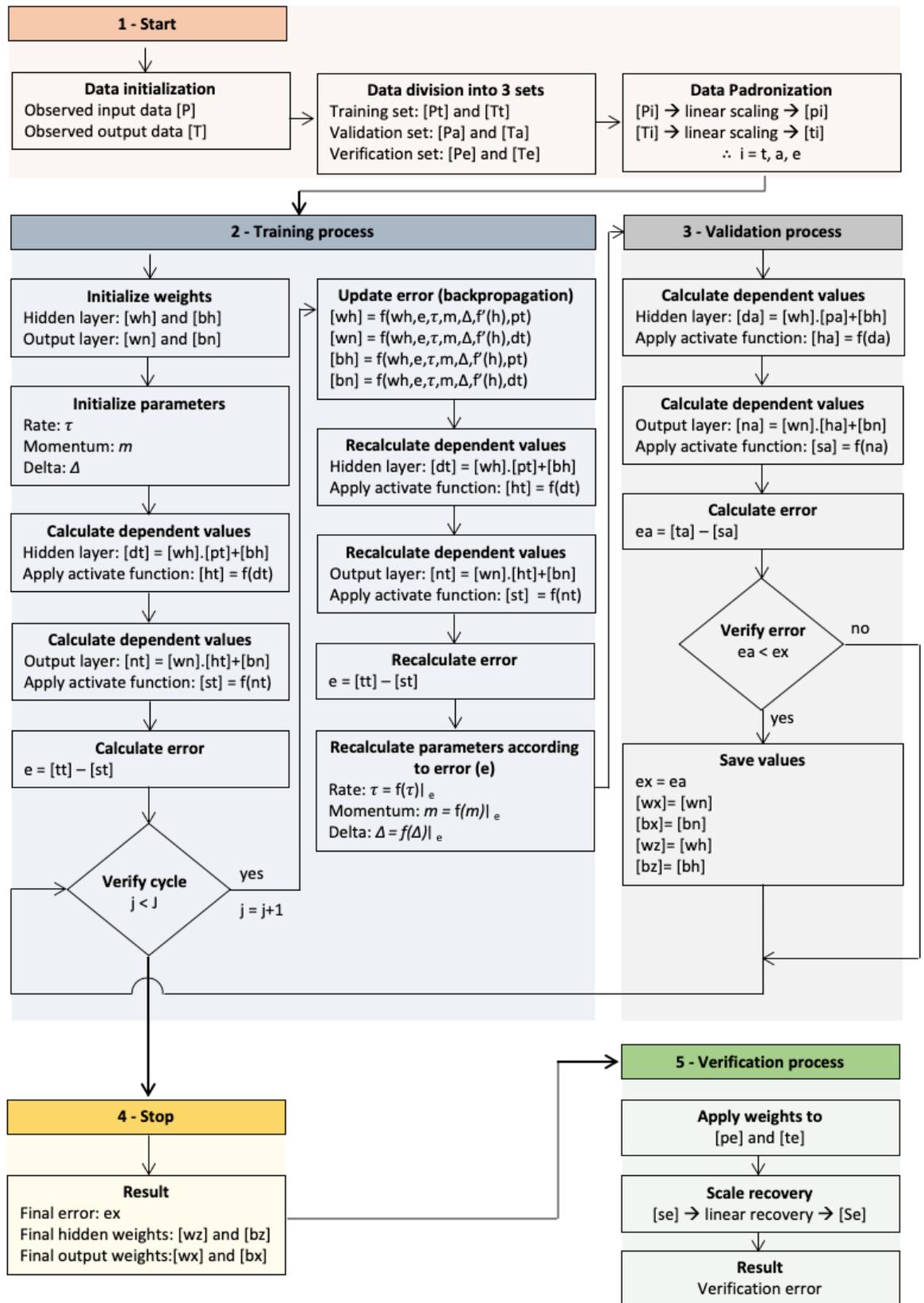


Figure 1A. Flowchart of the Training, Validation and Verification Processes.