Earth Science Informatics (2020) 13:1163–1173
https://doi.org/10.1007/s12145-020-00496-z

RESEARCH ARTICLE

Optimal level of wavelet decomposition for daily inflow forecasting

Paula Karenina de Macedo Machado Freire[1] · Celso Augusto Guimarães Santos[1]

Received: 30 March 2020 / Accepted: 27 July 2020 / Published online: 6 August 2020

Abstract
A methodology to select the maximum level of wavelet decomposition to forecast seven days of daily inflows by a hybrid model wavelet-based artificial neural network (WANN) is proposed. The wavelet decomposition was employed to decompose an input time series into approximation and detail components, and the approximations were used as inputs to artificial neural networks (ANN) for WANN hybrid models. In this study, it was used daily inflows from January 1931 to December 2010 to three Brazilian reservoirs with different discharge patterns, and evaluated the accuracy of the WANN models when using seven different mother-wavelets, including Haar, Daubechies, Biorthogonal, Biorthogonal Reverse, Symlet, Coiflet and Discrete Meyer. It was found that the model performance is dependent on the input sets and the selected mother-wavelets. Based on the obtained results, it was observed that the maximum level of decomposition was five, because upper than this level, independently on the inflow magnitude, there is no guarantee that the WANN hybrid models would perform better than the ANN model.

Keywords Optimal level · Streamflow forecast · Wavelet filter · DWT

Introduction

The inflow forecasting is one of the most active research areas in surface water hydrology. Currently, there have been a great number of relevant studies, and many methods and models could be used to perform inflow forecasting (e.g., Fernando and Jayawardena 1998; Toth et al. 1999; Shamseldin and O'Connor 2001; Xiong and O'Connor 2002; Moradkhani et al. 2004; Goswami et al. 2005; Kentel 2009; Kagoda et al. 2010; Jothiprakas and Maga 2012; Liu et al. 2012; Danandeh et al. 2014; Terzi and Ergin 2014; Akrami et al. 2014; Yaseen et al. 2015; Afan et al. 2020).

Inflow forecasting is a need for an adequate water management and reservoir operation. Currently, the Brazilian National System Operator (ONS), which is responsible for the operation of the hydroelectric power plants reservoirs in Brazil, uses stochastic models to subsidize their work. However, these models have limited precision and, therefore, it is necessary to develop more efficient tools to plan and operate such a system (Hidalgo et al. 2012; Santos and Silva 2014; Freire et al. 2019).

Communicated by: H. Babaie

✉ Celso Augusto Guimarães Santos
celso@ct.ufpb.br

[1] Department of Civil and Environmental Engineering, Federal University of Paraíba, João Pessoa, Paraíba 58051-900, Brazil

Artificial neural networks (ANN) have already shown good results in inflow forecasting (Karunanithi et al. 1994; Campolo et al. 1999; Danh et al. 1999; Lauzon et al. 2000; Sivakumar et al. 2002; Cigizoglu 2003; Kumar et al. 2004; Cigizoglu and Kisi 2005; Cheng et al. 2005; Farias et al. 2013; Farias and Santos 2014). Therefore, they could be used as an alternative to stochastic models, or in conjunction with other models to improve the operation of the interconnected systems.

Recently, the ANNs forecasting results have been improved by pre-processing the input data through some type of signal filter such as wavelet transform, which transforms the raw input time series into high frequency (details) and low frequency (approximation) components. The recent studies show that the use of signal pre-processed by wavelet transform improves the results obtained by the regular ANN in the inflows forecasting (Cannas et al. 2005; Kisi 2009; Adamowski and Sun 2010; Pramanik et al. 2010; Krishna et al. 2011; Tiwari and Chatterjee 2011; Krishna 2013; Maheswaran and Khosa 2013; Wei et al. 2013; Santos et al. 2014, 2019; Honorato et al. 2018).

The main objective here is to investigate the influence of the raw signal decomposition level choice on wavelet-based daily inflow forecasting. Thus, this paper describes the method for discrete wavelet decomposition of time series, the daily inflow time series registered in three different reservoirs used as study cases, and the forecasting results using WANN models with inputs of different decomposition levels.

Material

Study area

The selected reservoirs are Sobradinho, 14 de Julho, and Itaipu (Fig. 1). Sobradinho Reservoir is in the São Francisco River, in north-eastern Brazil, in the Bahia State, the 14 de Julho Reservoir is in the Antas River, Cotiporã city, Rio Grande do Sul State, southern Brazil, and Itaipu Reservoir is in the Paraná River located on the border between Brazil and Paraguay.

Sobradinho has a hydroelectric plant that is located in the São Francisco River at 748 km far from its mouth, with a drainage area of 498,968 km^2. This reservoir has the second largest artificial lake in the world, with about 320 km long, a water surface of 4214 km^2 and a storage capacity of 34.1 billion m^3 at the depth of 392.50 m. This lake is the largest water reservoir of north-eastern Brazil, and through the São Francisco Hydroelectric Company, it regulates the São Francisco River downstream inflow. The 14 de Julho Hydroelectric Plant has a generation capacity of 100 MW, with a maximum height of 33.5 m and a flooded area of 6 km^2. Itaipu has an installed generation capacity of 14 GW, with 20 generating units providing 700 MW each with a hydraulic design head of 118 m. This lake is the seventh largest in Brazil (1350 km^2), but has the best rate of use of water to produce energy among the largest Brazilian reservoirs.

Inflow data

The data used in this paper correspond to the natural daily inflow into those reservoirs, for the period from 1 January 1931 to 31 December 2010, which were obtained from the ONS, which is also responsible for developing forecasting and scenario generation of average daily natural flow, weekly and monthly to all hydroelectric development sites in Brazil.

Figure 2 shows the daily hydrograph of the inflows into the studied reservoirs. Sobradinho Reservoir, with an average inflow of 2656 m^3/s, a maximum of 18,525 m^3/s and a minimum of 400 m^3/s. The 14 de Julho Reservoir, with an average inflow of 285 m^3/s, a maximum of 6912 m^3/s and a minimum of 2 m^3/s. Itaipu Reservoir, with an average inflow of 10,209 m^3/s, a maximum of 42,322 m^3/s and a minimum of 2512 m^3/s. These data comprehend 80 years (29,220 days) measured from the 1931 to 2010. The first 77 years of the inflow data (28,124 days, 96% of the whole data set) were used for the calibration, which was divided into three sets for the ANN

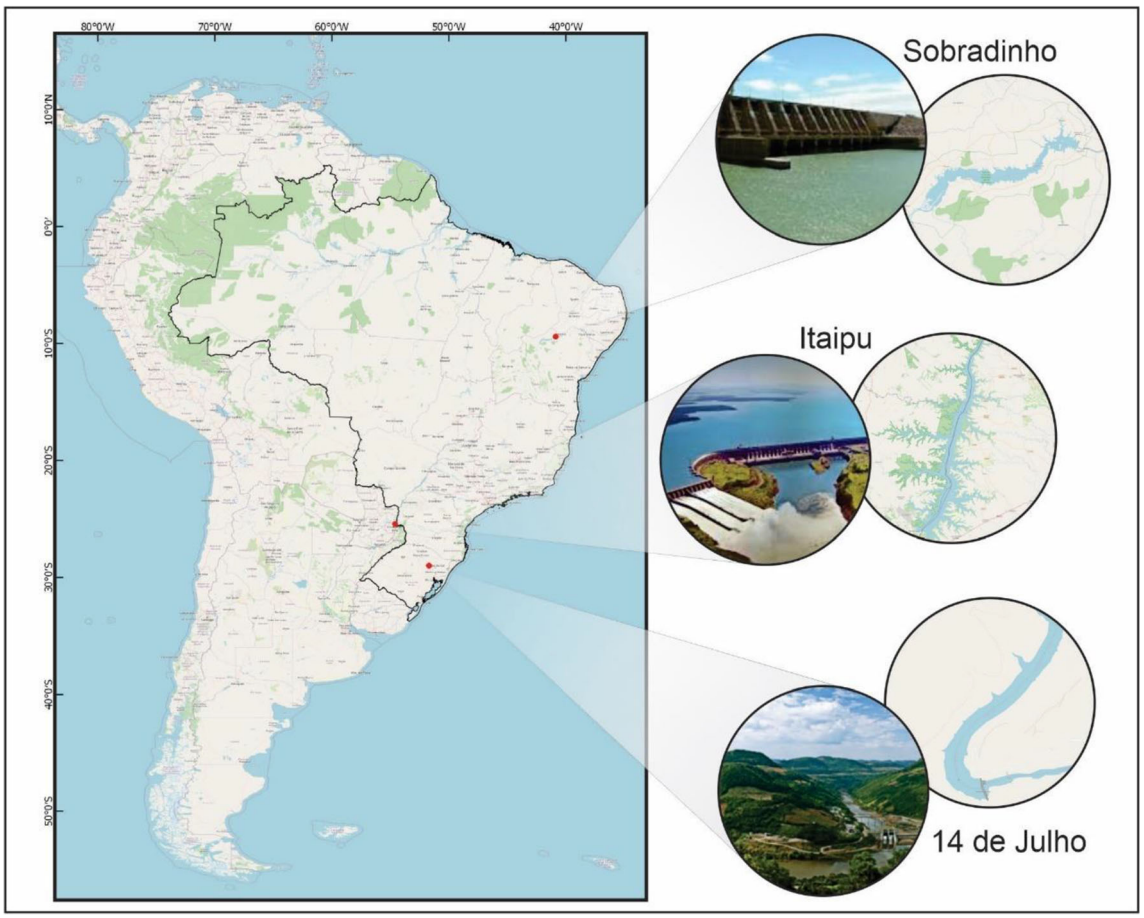

Fig. 1 Location of the selected reservoirs

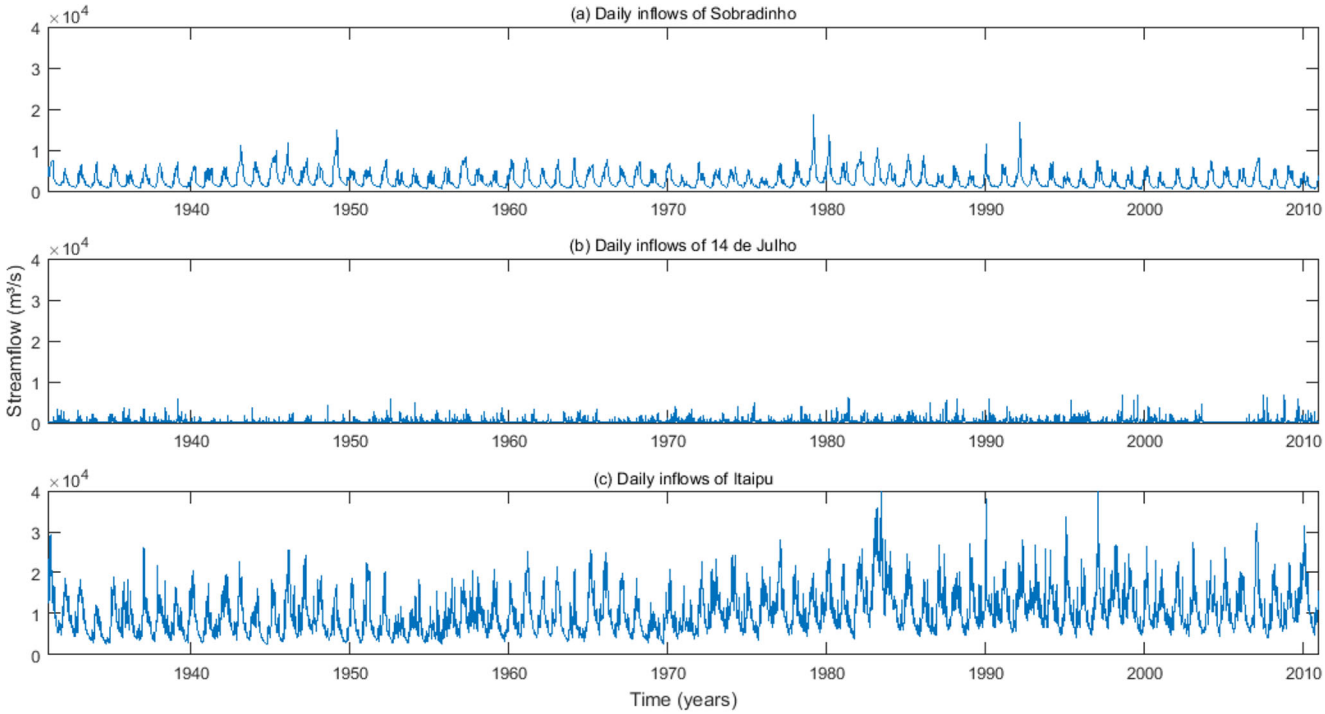

Fig. 2 Daily hydrograph of the three studied reservoirs: (**a**) Sobradinho, (**b**) 14 de Julho and (**c**) Itaipu (1931–2010)

training, validation and testing. The remaining three years (1096 days, 4% of the whole data set) was used for the final test. The statistical indices of these data sets are presented in Table 1 for each period.

Methods

Discrete wavelet transform (DWT)

The DWT is generally used in the decomposition and filtering of time series (Wang and Ding 2003; Ravansalar et al. 2015), because it does not cause coefficient redundancies between the scales, and the information about the time location of certain events is not lost in the process (Daubechies 1990; Alessio 2016).

For the calculation of DWT, the simplest and most efficient method was introduced by Mallat (1989), in which the scale and position parameters are chosen based on power of 2. This simple algorithm turns the DWT function a bypass filter, by calculating quickly the wavelet coefficients and thus decomposing the input signal into low and high frequency components (Misiti et al. 1996). The approximations correspond to the low frequency components and represent the

Table 1 Descriptive statistics for the daily inflow data used in the study

Reservoir	Period	Descriptive statistics							
		Minimum (m^3/s)	Maximum (m^3)/s	Average (m^3/s)	Mode (m^3/s)	Median (m^3/s)	Standard deviation (m^3/s)	Coefficient of variation	Coeficient of skewness
Sobradinho	1931–2010	400	18,525	2656	1054	1868	2041	0.77	1.78
	1931–2007	400	18,525	2676	1381	1871	2058	0.77	1.78
	2008–2010	436	6205	2127	1054	1634	1424	0.67	0.93
14 de Julho	1931–2010	2	6912	285	65	146	433	1.52	4.96
	1931–2007	4	6912	283	65	144	429	1.52	4.87
	2008–2010	2	6835	329	2	188	531	1.61	5.99
Itaipu	1931–2010	2512	42,322	10,209	6421	8883	5129	0.50	2512.00
	1931–2007	2512	42,322	10,127	6421	8793	5118	0.51	1.19
	2008–2010	4931	31,522	12,327	8633	11,079	4955	0.40	0.85

general behavior of the series, whereas the details correspond to the high frequency components and could be understood as the noises present in the series, depending on the level of decomposition (Santos et al. 2013, 2019; Freire et al. 2019).

The decomposition could continue in an iterative process, with the approximations being decomposed in turn; then, the original signal is broken down into several lower resolution components. This process is called wavelet decomposition tree as illustrated in Fig. 3. Thus, the maximum number of decompositions ($l_{máx}$) of each wavelet subfamily for the studied reservoirs was determined according to criterions: (a) the size of the series – known as criterion on the signal; and (b) the wavelet subfamily used – known as entropy criterion (Misiti et al. 2006):

$$l_{max} = \frac{\log\left(\frac{l_x}{(l_w - 1)}\right)}{\log 2} \quad (1)$$

where l_x is the size of the time series ($l_x = 29.220$) and l_w is the filter size associated with the orthogonal or biorthogonal wavelet.

The value of l_w ranged from 2 to 102, depending on the subfamily used, then the calculated $l_{máx}$ ranged from 8 to 14. Thus, the value 8 was chosen to be the maximum level of signal decomposition, because this value caters for all studied families. For example, Fig. 3 shows the approximations (low frequency) and the details (high frequency), both in eight levels of decomposition for the data of daily natural inflows to Sobradinho Reservoir. It is notorious by a simple visual checking that approximations above level eight would be distant from the hydrograph form of the raw signal.

Artificial neural network (ANN)

In this paper, the ANN inputs were formed by the inflow observed in the current day t (Q_t) and in the previous four days ($Q_{t\text{-}1}$, $Q_{t\text{-}2}$, $Q_{t\text{-}3}$ and $Q_{t\text{-}4}$); thus, the input layer had five neurons. As for the WANN, the input was formed by the signal approximation of such raw signal. The output layer of both models had only one neuron, which corresponds to the forecasted inflow seven days ahead (Q_{t+7}).

The architectures of ANN and proposed WANN were a feed-forward network, with 20 hidden neurons whose activation function was the sigmoid (hidden layer) and a linear function for the output layer. The Levenberg-Marquardt algorithm was used as the learning algorithm, because it is considered

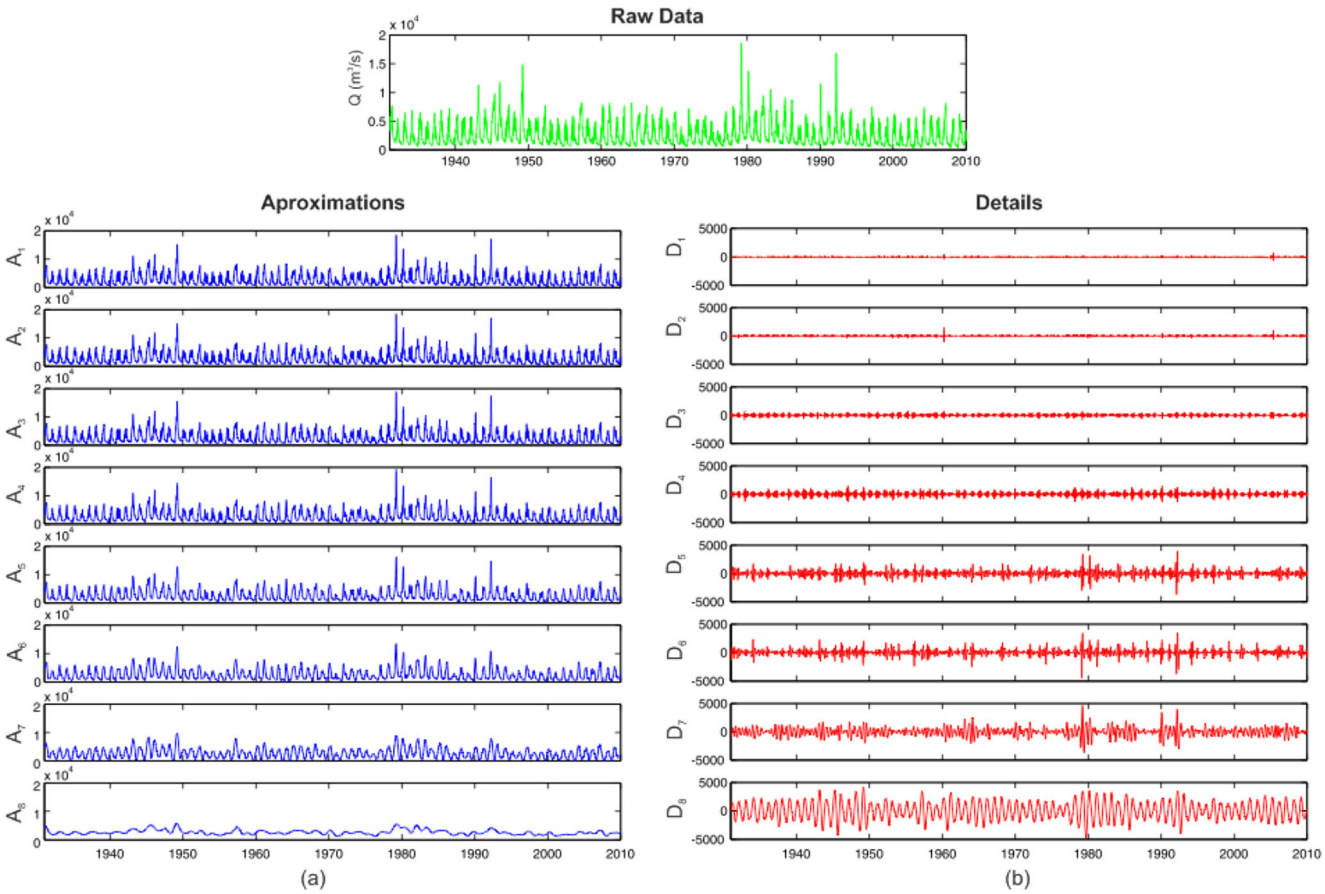

Fig. 3 Multiple decomposition up to the chosen maximum level of the Sobradinho Reservoir inflow time series using one selected mother-wavelet: approximations *in the left*, and details *in the right*

one of the fastest methods for training (Renno et al. 2015). The error verification on the training data set in the ANN and WANN was done by calculating the mean square error (*MSE*):

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(Q_{c_t} - Q_{o_t}\right)^2 \quad (2)$$

where Q_{ct} and Q_{ot} are respectively the calculated and observed inflows at time *t*.

Performance evaluation

In the present paper, three statistical indices were used to evaluate the accuracy of the forecasting results: (a) the root mean square error (*RMSE*), (b) the Nash-Sutcliffe coefficient (*NASH*) and (c) the correlation coefficient (*R*):

$$RMSE = \sqrt{\frac{1}{N}\Sigma\left(Q_c - Q_o\right)^2} \quad (3)$$

$$NASH = 1 - \frac{\Sigma\left(Q_o - Q_c\right)^2}{\Sigma\left(Q_o - \overline{Q_o}\right)^2} \quad (4)$$

$$R = \frac{\Sigma\left(Q_o - \overline{Q_o}\right)\left(Q_c - \overline{Q_c}\right)}{\sqrt{\Sigma\left(Q_o - \overline{Q_o}\right)^2}\sqrt{\Sigma\left(Q_c - \overline{Q_c}\right)^2}} \quad (5)$$

where Q_c is the calculated inflow; Q_o is the observed inflow; $\overline{Q_c}$ is the mean calculated inflow; and $\overline{Q_o}$ is the mean observed inflow.

The root mean square error is the square root of the mean square error (*MSE*), whose optimal value is *RMSE* = 0. The Nash-Sutcliffe coefficient is considered one of the most important statistical criteria to evaluate the accuracy of hydrological models, which can range from –∞ to 1, whose optimal value is *NASH* = 1. The correlation coefficient can range from −1 to 1 and indicates the degree of collinearity between forecasted and observed values; if *R* = 1, a perfect positive linear relationship exists.

Results and discussion

In order to improve the ANN efficiency, the input and output data were normalized and, then, scaled before training, in the interval [−1, 1] (Demuth and Beale 2005). For each simulation, 70% of the original data were used in the training, 15% for validation and 15% for testing. The forecasted results using the ANN model presented *RMSE* = 457.5096 m^3/s, *NASH* = 0.8967, and *R* = 0.9479, for Sobradinho Reservoir; *RMSE* = 511.7921 m^3/s, *NASH* = 0.0658, and *R* = 0.2624, for 14 de Julho Reservoir; and *RMSE* = 2299.3335 m^3/s, *NASH* = 0.7844, and *R* = 0.8862, for Itaipu Reservoir.

In order to improve the forecasting efficiency, as aforementioned, the raw signal was pre-processed using 54 wavelet subfamilies to break it down into approximations and details up to the maximum level set (i.e. 8). Then, the approximations were used as ANN inputs to forecast the inflows seven days ahead, totalling 432 WANN models (i.e., 54 × 8), for each reservoir.

Figure 4 shows the performance of the WANN models, based on *RMSE*, in which the dots below the red line means

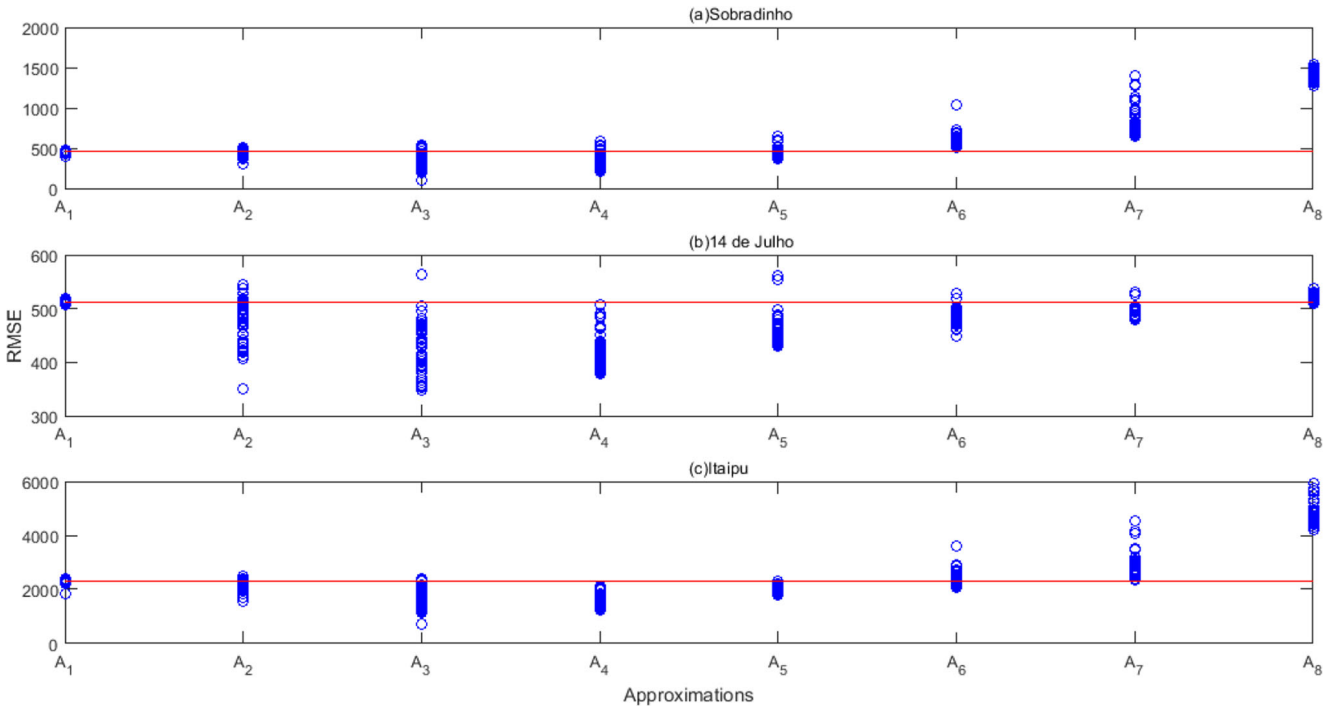

Fig. 4 Performance of the WANN models based on *RMSE* index

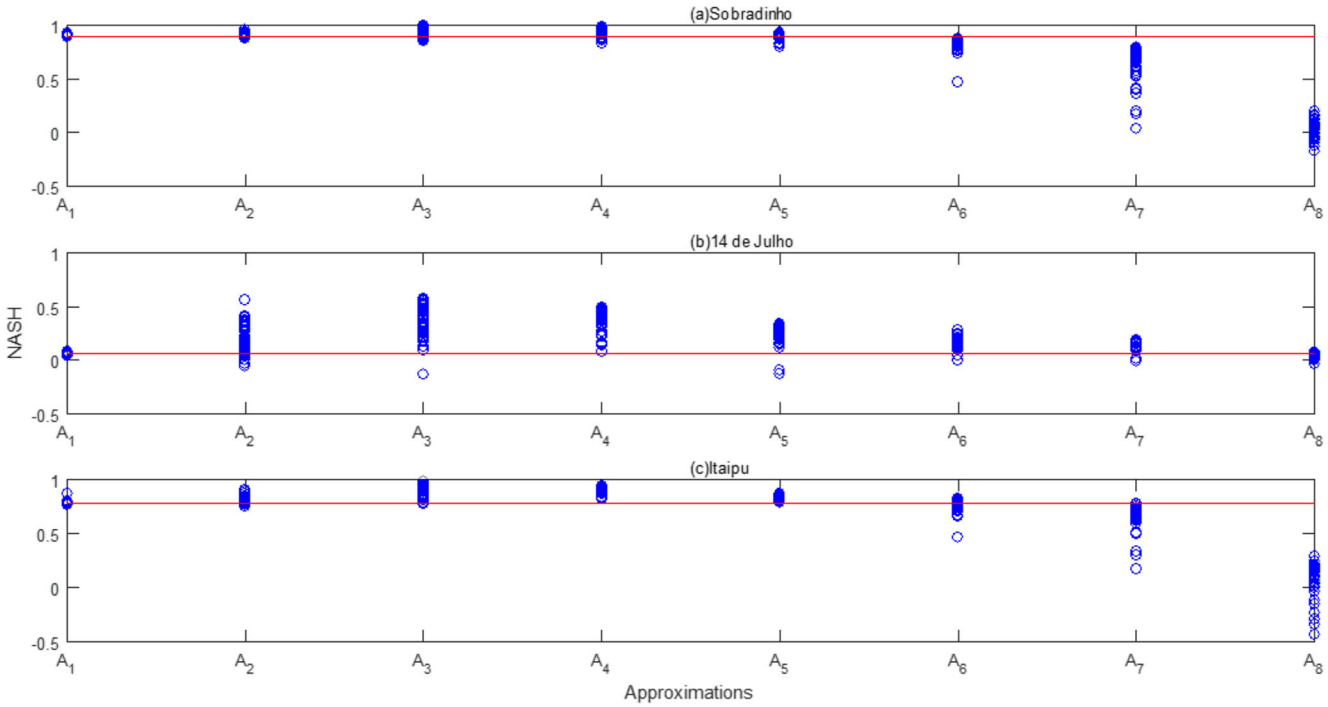

Fig. 5 Performance of the WANN models based on *NASH* index

that the WANN models were successful in relation to the regular ANN model, whereas the dots above the red line means that the WANN models were not successful. It is observed that with the use of the approximations A_1 to A_5 as inputs, the *RMSE* decreases in many cases, independently on the studied reservoir. After that, the error substantially increases, and one may note that it is not worthwhile to use the A_6, A_7, A_8 or even higher approximations, because those approximations were not able to provide any forecasting improvement. Exceptionally, for reservoirs with small inflow volumes (e.g., 14 de Julho Reservoir), the A_6 and A_7 approximations could be used. The same is observed when analysing the *NASH* (Fig. 5) and *R* (Fig. 6) indices. In both figures, the dots above the red line means again that the WANN models

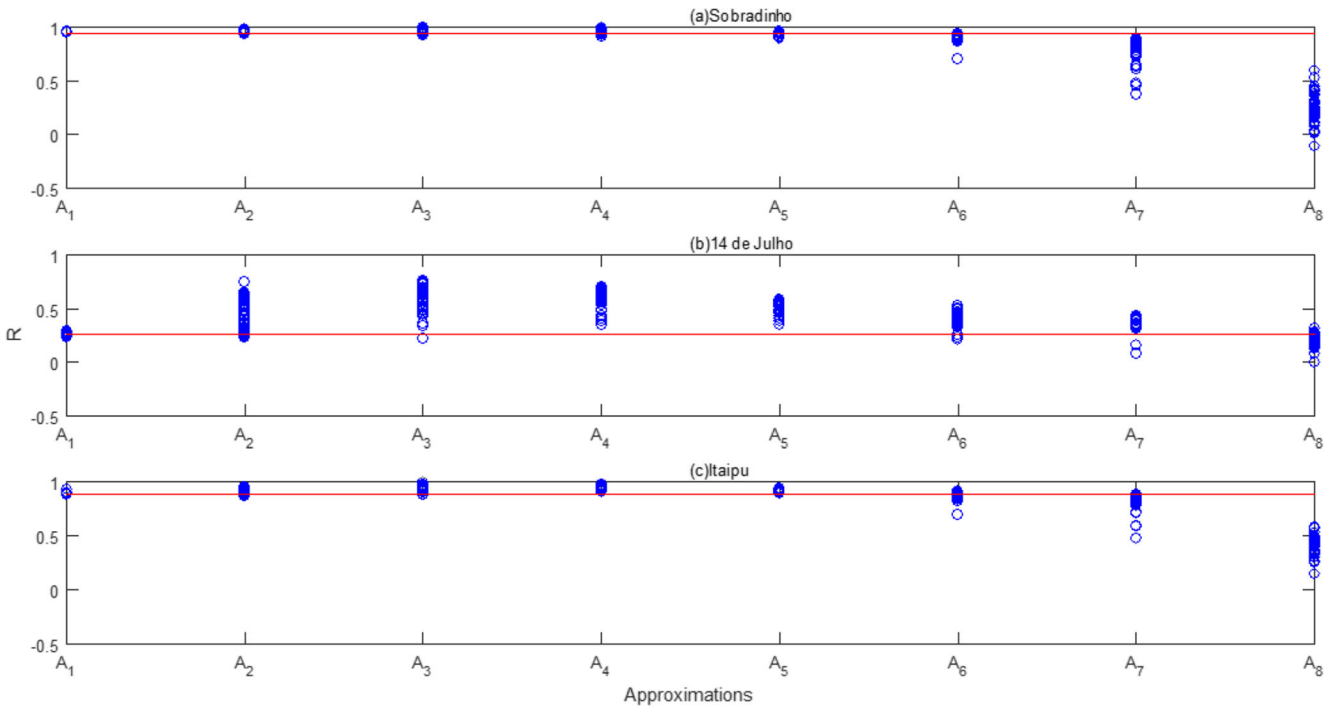

Fig. 6 Performance of the WANN models based on *R* index

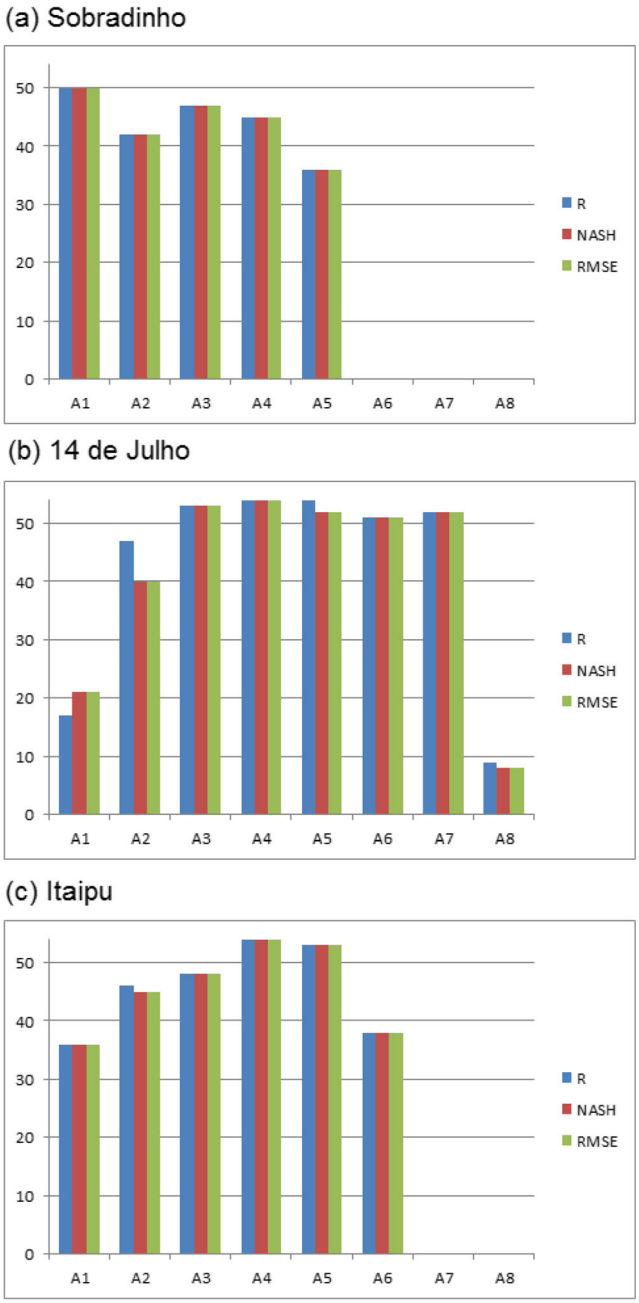

Fig. 7 Quantitative success of WANN models against the ANN model, based on the performance indices (*RMSE*, *NASH* and *R*) for each reservoir (**a**) Sobradinho, (**b**) 14 de Julho and (**c**) Itaipu

were successful in relation to the regular ANN model and the dots below the red line means that they were not successful in relation to the ANN model. Then, it is confirmed that the use of A_6, A_7, A_8 or higher approximations is not indicated to be used as ANN inputs, because such a procedure would not improve the performance of the WANN models. The A_6 and A_7 approximations could be used exceptionally for inflow time series composed by small volumes.

Figure 7 shows the quantitative success of WANN models in relation to ANN model, according to the results shown in Figs. 4, 5 and 6. From the quantitative point of view, the best approximation was A_4, which obtained 100.00% success for all analysed indices for 14 de Julho Reservoir and Itaipu Reservoir. For the Sobradinho Reservoir, the best approximation was A_1, which obtained 92.59% success for all indices, as 54 wavelet subfamilies were evaluated and this approximation was successful in 50 subfamilies. The second-best approximation for Sobradinho Reservoir was A_3 with 87.04% success for all indices, followed by A_4 and A_2 with 83.33% and 77.78% success, respectively. Finally, A_5 presented 66.67% success, whereas the other approximations (A_6, A_7, A_8) did not have success as stated earlier. For the 14 de Julho Reservoir, the second-best approximation was A_3 with 98.15% success for all indices, followed by A_5 with 96.30% success for *RMSE* and *NASH* indices and 100.00% success for the *R* index, followed by A_7 and A_6 with 96.30% and 94.44% success for all indices, respectively. Finally, A_2, A_1 and A_8 presented 74.07%, 38.89% and 14.81% success for *RMSE* and *NASH* indices and 87.04%, 31.48% and 16.67% success for the *R* index, respectively. For the Itaipu Reservoir, the second-best approximation was A_5 with 98.15% success for all indices, followed by A_3 and A_2 with 88.89% and 83.33% success, respectively. Finally, A_6 and A_1 presented 70.37% and 66.67% success, respectively, whereas the other approximations (A_7, A_8) did not have success as aforementioned.

However, when analysing the values of these indices, it was observed that (a) for the Sobradinho Reservoir, the *RMSE* ranged from 400.0489 to 473.3114 m^3/s for A_1, from 302.5705 to 504.5766 m^3/s for A_2, from 94.3232 to 537.4780 m^3/s for A_3, from 203.4392 to 579.0129 m^3/s for A_4 and from 371.0106 to 642.2430 m^3/s for A_5; (b) for the 14 de Julho Reservoir, the *RMSE* ranged from 506.9577 to 519.2857 m^3/s for A_1, from 350.1133 to 543.8587 m^3/s for A_2, from 347.3893 to 564.1586 m^3/s for A_3, from 378.6133 to 507.4274 m^3/s for A_4 and from 429.7828 to 562.1094 m^3/s for A_5; and (c) for the Itaipu Reservoir, the *RMSE* ranged from 1825.7296 to 2396.7418 m^3/s for A_1, from 1555.2176 to 2491.0735 m^3/s for A_2, from 728.4548 to 2380.0655 m^3/s for A_3, from 1229.3971 to 2114.6630 m^3/s for A_4 and from 1771.7477 to 2323.0118 m^3/s for A_5, as can be seen in Table 2. Table 2 also shows the variation of all performance indices for all ANN input configurations. By analyzing Table 2, it could be noted that (a) for the Sobradinho Reservoir, the A_3 approximation improved the forecasting by up to 79% decrease in *RMSE*, 11% increase in *NASH* and 5% increase in *R*, whereas the A_4 approximation improved the forecasting by up to 56% decrease in *RMSE*, 9% increase in *NASH* and 4% increase in *R*; while A_1 improved only *RMSE* by 13%, *NASH* by 3% and *R* by 1%; (b) for the 14 de Julho Reservoir, the A_3 approximation improved the forecasting by up to 32% decrease in *RMSE*, 766% increase in *NASH* and 189% increase in *R*, whereas the A_2 approximation improved the forecasting by up to 32% decrease in *RMSE*, 755%

Table 2 Variation of the performance indices for the ANN (raw data) and WANN (A_1 to A_8) models for each reservoir

Reservoir	Inputs	*RMSE*			*NASH*			*R*		
		Minimum	Median	Maximum	Minimum	Median	Maximum	Minimum	Median	Maximum
Sobradinho	raw data	457.5096	457.5096	457.5096	0.8967	0.8967	0.8967	0.9479	0.9479	0.9479
	A_1	400.0489	442.3452	473.3114	0.8895	0.9035	0.9210	0.9444	0.9516	0.9609
	A_2	302.5705	415.3662	504.5766	0.8744	0.9149	0.9548	0.9366	0.9574	0.9775
	A_3	94.3232	333.6860	537.4780	0.8575	0.9451	0.9956	0.9271	0.9727	0.9978
	A_4	203.4392	298.5179	579.0129	0.8346	0.9560	0.9796	0.9159	0.9781	0.9898
	A_5	371.0106	427.2326	642.2430	0.7965	0.9099	0.9321	0.8935	0.9541	0.9655
	A_6	505.9062	569.5803	1039.0498	0.4674	0.8399	0.8737	0.6985	0.9171	0.9355
	A_7	650.6293	795.9037	1395.4518	0.0393	0.6875	0.7912	0.3736	0.8322	0.8937
	A_8	1273.6083	1394.7793	1541.7995	−0.1728	0.0402	0.1997	−0.1121	0.2446	0.5911
14 de Julho	raw data	511.7921	511.7921	511.7921	0.0658	0.0658	0.0658	0.2624	0.2624	0.2624
	A_1	506.9577	512.4453	519.2857	0.0382	0.0634	0.0834	0.2288	0.2568	0.2924
	A_2	350.1133	489.7175	543.8587	−0.0549	0.2893	0.5628	0.2292	0.4112	0.7506
	A_3	347.3893	414.6564	564.1586	−0.1352	0.3868	0.5696	0.2217	0.6237	0.7593
	A_4	378.6133	416.5873	507.4274	0.0817	0.3810	0.4887	0.3521	0.6246	0.7000
	A_5	429.7828	455.1488	562.1094	−0.1269	0.2611	0.3412	0.3481	0.5136	0.5850
	A_6	448.7201	482.0309	529.4274	0.0003	0.1713	0.2819	0.2154	0.4189	0.5312
	A_7	478.4639	495.7818	531.8130	−0.0087	0.1233	0.1835	0.0841	0.3516	0.4344
	A_8	509.1452	518.2585	538.8718	−0.0357	0.0420	0.0754	−0.0038	0.2118	0.3130
Itaipu	raw data	2299.3335	2299.3335	2299.3335	0.7844	0.7844	0.7844	0.8862	0.8862	0.8862
	A_1	1825.7296	2276.0376	2396.7418	0.7658	0.7884	0.8641	0.8758	0.8885	0.9304
	A_2	1555.2176	2138.2865	2491.0735	0.7470	0.8136	0.9014	0.8655	0.9030	0.9495
	A_3	728.4548	1564.8443	2380.0655	0.7690	0.9001	0.9784	0.8784	0.9492	0.9893
	A_4	1229.3971	1548.5311	2114.6630	0.8177	0.9022	0.9384	0.9058	0.9501	0.9687
	A_5	1771.7477	1878.2263	2323.0118	0.7800	0.8562	0.8720	0.8846	0.9254	0.9338
	A_6	2070.5670	2197.1710	3620.6149	0.4655	0.8032	0.8252	0.6945	0.8976	0.9087
	A_7	2336.6894	2713.6642	4515.0494	0.1689	0.6997	0.7774	0.4761	0.8413	0.8830
	A_8	4194.9657	4549.3134	5936.9070	−0.4371	0.1562	0.2825	0.1419	0.4286	0.5788

increase in *NASH* and 186% increase in *R*; while A_4 improved only *RMSE* by 26%, *NASH* by 643% and *R* by 167%; and (c) for the Itaipu Reservoir, the A_3 approximation improved the forecasting by up to 68% decrease in *RMSE*, 25% increase in *NASH* and 12% increase in *R*, while A_4 improved only *RMSE* by 47%, *NASH* by 20% and *R* by 9%.

Thus, a qualitative analysis of the successes of WANN models in relation to the ANN model is necessary and can be observed in Table 3. The negative sign in the *RMSE* column of Table 3 means an improvement in such index; then, the best value is close to 0.0. On the other hand, the plus sign in the *NASH* and *R* columns shows an improvement of such indices, and the best values are close to 1.0. From the qualitative point of view, the best approximation was A_3, for all analysed indices of each analysed reservoir, which showed an improvement range from 0.06 to 79.38% for *RMSE*, from 0.01 to 11.03% for *NASH* and from 0.06 to 5.26% for *R* in the Sobradinho Reservoir; for the 14 de Julho Reservoir this approximation showed an improvement range from 1.37 to 32.12% for *RMSE*, from 38.76 to 765.60% for *NASH* and from 30.37 to 189.39% for *R*; for the Itaipu Reservoir the improvement range of this approximation was from 6.13 to 68.32% for *RMSE*, from 3.27 to 24.72% for *NASH* and from 1.80 to 11.62% for *R*. The second-best approximation was A_4 for Sobradinho and Itaipu reservoirs with an improvement range from 6.13 to 55.53% and from 8.03 to 46.53% for *RMSE*, from 1.37 to 9.24% and from 4.24 to 19.62% for *NASH* and from 0.66 to 4.41% and from 2.20 to 9.31% for *R*, respectively. For the 14 de Julho Reservoir, the second-best approximation was A_2 with an improvement range from 0.08 to 31.59% for *RMSE*, from 2.29 to 755.31% for *NASH* and from 0.44 to 186.06% for *R*. The A_4 approximation ranked third for the 14 de Julho reservoir with an improvement range from 0.85 to 26.02% for *RMSE*, from 24.11 to 642.74% for *NASH* and from 34.20 to 166.77% for *R* and the A_1 approximation ranked fifth for the

Table 3 Percentage of improvement of the forecasting performance using the WANN models compared to the ANN model

Reservoir	Inputs	*RMSE*			*NASH*			*R*		
Sobradinho	A_1	−0.46%	to	−12.56%	0.11%	to	2.71%	0.02%	to	0.83%
	A_2	−1.16%	to	−33.87%	0.26%	to	6.48%	0.14%	to	3.11%
	A_3	−0.06%	to	−79.38%	0.01%	to	11.03%	0.06%	to	5.26%
	A_4	−6.13%	to	−55.53%	1.37%	to	9.24%	0.66%	to	4.41%
	A_5	−1.25%	to	−18.91%	0.29%	to	3.94%	0.04%	to	1.85%
	A_6		—			—			—	
	A_7		—			—			—	
	A_8		—			—			—	
14 de Julho	A_1	−0.02%	to	−0.94%	0.44%	to	26.69%	0.38%	to	11.43%
	A_2	−0.08%	to	−31.59%	2.29%	to	755.31%	0.44%	to	186.06%
	A_3	−1.37%	to	−32.12%	38.76%	to	765.60%	30.37%	to	189.39%
	A_4	−0.85%	to	−26.02%	24.11%	to	642.74%	34.20%	to	166.77%
	A_5	−2.74%	to	−16.02%	76.84%	to	418.53%	32.66%	to	122.95%
	A_6	−1.63%	to	−12.32%	45.92%	to	328.36%	24.09%	to	102.44%
	A_7	−1.28%	to	−6.51%	36.22%	to	178.88%	18.45%	to	65.57%
	A_8	−0.03%	to	−0.52%	0.74%	to	14.65%	2.36%	to	19.30%
Itaipu	A_1	−0.33%	to	−20.60%	0.05%	to	10.15%	0.04%	to	4.98%
	A_2	−0.12%	to	−32.36%	0.07%	to	14.91%	0.03%	to	7.14%
	A_3	−6.13%	to	−68.32%	3.27%	to	24.72%	1.80%	to	11.62%
	A_4	−8.03%	to	−46.53%	4.24%	to	19.62%	2.20%	to	9.31%
	A_5	−3.31%	to	−22.95%	1.79%	to	11.16%	0.93%	to	5.37%
	A_6	−0.60%	to	−9.95%	0.33%	to	5.20%	0.18%	to	2.53%
	A_7		—			—			—	
	A_8		—			—			—	

Sobradinho reservoir with an improvement range from 0.46 to 12.56% for *RMSE*, from 0.11 to 2.71% for *NASH* and from 0.02 to 0.83% for *R*.

Conclusions

The use of the wavelet transform to eliminate the noise presented in the raw signal showed to be extremely important to improve the ANN forecast performance; i.e., the WANN models performed significantly better than the ANN model to forecast the Sobradinho, the 14 de Julho and the Itaipu reservoir inflows seven days ahead.

Seven wavelet families were analysed, for which the maximum decomposition level of each wavelet subfamily was calculated for the inflow data of the Sobradinho Reservoir, the 14 de Julho Reservoir and the Itaipu Reservoir based on the signal criterion (size of the series) and entropy criterion (wavelet subfamily). Thus, the maximum level of decomposition was chosen equal to eight, because such decomposition caters for all the studied subfamilies.

A total of 432 WANN models were tested against a regular ANN for each reservoir, and it was observed that the best forecastings of the WANN models were for the approximations between level A_1 and A_5, from which the A_4 approximation was the most successful, followed by the A_3 approximation for 14 de Julho Reservoir and the Itaipu Reservoir. Although for the Sobradinho Reservoir, the A_1 approximation obtained the highest amount of success followed by the A_3 approximation. The A_3 approximation was chosen as the best approximation to be used as ANN inputs, because such an approximation provided the best forecasting results for all reservoirs: (a) with *RMSE* ranging from 94.3232 to 537.4780 m^3/s, *NASH* ranging from 0.8575 to 0.9956 and *R* ranging from 0.9271 to 0.9978 for Sobradinho Reservoir; (b) with *RMSE* ranging from 347.3893 to 564.1586 m^3/s, *NASH* ranging from 0.1352 to 0.5696 and *R* ranging from 0.2217 to 0.7593 for the 14 de Julho Reservoir; and (c) with *RMSE* ranging from 728.4548 to 2380.0655 m^3/s, *NASH* ranging from 0.7690 to 0.9784 and *R* ranging from 0.8784 to 0.9893 for Itaipu Reservoir, while indices for A_1 approximation ranged from 400.0489 to 473.3114 m^3/s for *RMSE*, from 0.8895 to 0.9210 for *NASH*, and 0.9444 to 0.9609 for *R* for Sobradinho Reservoir and the indices for A_4 approximation ranged from 378.6133 to 507.4274 m^3/s and from 1229.3971 to 2114.6630 m^3/s for *RMSE*, from 0.0817 to 0.4887 and from

0.8177 to 0.9384 for *NASH*, and from 0.3521 to 0.7000 and from 0.9058 to 0.9687 for *R* for the 14 de Julho and Itaipu reservoirs, respectively.

Finally, it can be concluded that by decomposing a daily inflow time series up to the fifth level and using the A_5 approximation as ANN inputs, i.e., eliminating the D_1, D_2, D_3, D_4 and D_5 details, it is possible to often obtain better forecasting results than using the raw data as input data (regular ANN model). However, if the D_1, D_2 and D_3 details could be assumed as noise of the raw signal, then the A_3 approximation could be used as ANN inputs, and such procedure would provide the best WANN forecasting results, regardless of the river discharge patterns and the chosen mother-wavelet or wavelet subfamily.

Funding information This study was funded by National Council for Scientific and Technological Development, Brazil (304213/2017–9). This study was also financed in part by the Brazilian Agency for the Improvement of Higher Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES) – Fund Code 001 and the Federal University of Paraíba.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91

Afan HA, Allawi MF, El-Shafie A et al (2020) Input attributes optimization using the feasibility of genetic nature inspired algorithm: application of river flow forecasting. Sci Rep 10:4684. https://doi.org/10.1038/s41598-020-61355-x

Akrami SA, El-Shafie A, Naseri M, Santos CAG (2014) Rainfall data analyzing using moving average (MA) model and wavelet multi-resolution intelligent model for noise evaluation to improve the forecasting accuracy. Neural Comput & Applic 25(7–8):1853–1861. https://doi.org/10.1007/s00521-014-1675-0

Alessio SM (2016) Discrete wavelet transform (DWT). In: Digital signal processing and spectral analysis for scientists. Signals and Communication Technology. Springer, Cham, pp 645–714. https://doi.org/10.1007/978-3-319-25468-5_14

Campolo M, Andreussi P, Soldati A (1999) River flood forecasting with a neural network model. Water Resour Res 35(4):1191–1197. https://doi.org/10.1029/1998WR900086

Cannas B, Fanni A, Sias G, Tronci S, Zedda MK (2005) River flow forecasting using neural networks and wavelet analysis. Geophys Res Abstr 7:08651

Cheng C, Chau K, Sun Y, Lin J (2005) Long-term prediction of discharges in Manwan reservoir using artificial neural network models. In: Wang J, Liao XF, Yi Z (eds) Advances in neural networks – ISNN 2005. Lecture Notes in Computer Science, vol 3498, Springer, Berlin, Heidelberg, pp 1040–1045. https://doi.org/10.1007/11427469_165

Cigizoglu HK (2003) Estimation, forecasting and extrapolation of flow data by artificial neural networks. Hydrological Sci J 48(3):349–361. https://doi.org/10.1623/hysj.48.3.349.45288

Cigizoglu HK, Kisi Ö (2005) Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. Nord Hydrol 36(1):1–16

Danandeh Mehr A, Kahya E, Sahin A, Nazemosadat MJ (2014) Successive-station monthly streamflow prediction using different artificial neural network algorithms. Int J Environ Sci Technol. https://doi.org/10.1007/s13762-0140613-0

Danh NT, Phien HN, Gupta A (1999) Neural network models for river flow forecasting. Water SA 25(1)

Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. Inf Theory, IEEE Trans 36(5):961–1005

Demuth H, Beale M (2005) Neural network toolbox: for use with Matlab. The MathWorks, Inc., Natick, USA

Farias CAS, Santos CAG (2014) The use of Kohonen neural networks for runoff–erosion modeling. J Soils Sediments 14(7):1242–1250. https://doi.org/10.1007/s11368-013-0841-9

Farias CAS, Santos CAG, Lourenço AMG, Carneiro TC (2013) Kohonen neural networks for rainfall-runoff modeling: case of Piancó River basin. J Urban Environ Eng 7(1):176–182. https://doi.org/10.4090/juee.2013.v7n1.176182

Fernando DAK, Jayawardena AW (1998) Runoff forecasting using RBF networks with OLS algorithm. J Hydrol Eng 3:203–209. https://doi.org/10.1061/(ASCE)1084-0699(1998)3:3(203)

Freire PKMM, Santos CAG, Silva GBL (2019) Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. Appl Soft Comput 80:494–505. https://doi.org/10.1016/j.asoc.2019.04.024

Goswami M, O'Connor KM, Bhattarai KP, Shamseldin AY (2005) Assessing the performance of eight real-time updating models and procedures for the Brosna River. Hydrol Earth Syst Sci 9:394–411. https://doi.org/10.5194/hess-9-394-2005

Hidalgo I, Fontane D, Arabi M, Lopes J, Andrade J, Ribeiro L (2012) Evaluation of optimization algorithms to adjust efficiency curves for hydroelectric generating units. J Energy Eng 138(4):172–178

Honorato AGSM, Silva GBL, Santos CAG (2018) Monthly streamflow forecasting using neuro-wavelet techniques and input analysis. Hydrol Sci J 63(15–16):2060–2075. https://doi.org/10.1080/02626667.2018.1552788

Jothiprakas V, Maga RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. J Hydrol 450–451:293–307. https://doi.org/10.1016/j.jhydrol.2012.04.045

Kagoda PA, Ndiritu J, Ntuli C, Mwaka B (2010) Application of radial basis function neural networks to short-term streamflow forecasting. Phys Chem Earth 35:571–581. https://doi.org/10.1016/j.pce.2010.07.021

Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. J Comput Civ Eng 8(2):201–220. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(201)

Kentel E (2009) Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. J Hydrol 375:481–488. https://doi.org/10.1016/j.jhydrol.2009.06.051

Kisi O (2009) Neural networks and wavelet conjunction model for intermittent stream-flow forecasting. J Hydrol Eng 14(8):773–782

Krishna B (2013) Comparison of wavelet based ANN and regression models for reservoir inflow forecasting. J Hydrol Eng. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000892

Krishna B, Satyaji Rao YR, Nayak PC (2011) Time series modeling of river flow using wavelet neural networks. J Water Resour and Prot 03(1):50–59. https://doi.org/10.4236/jwarp.2011.31006

Kumar DN, Raju KS, Sathish T (2004) River flow forecasting using recurrent neural network. Water Resour Manag 18(2)

Lauzon N, Rousselle J, Birikundavyi S, Trung HT (2000) Real-time daily flow forecasting using black-box models, diffusion processes, and neural networks. Can J Civ Eng 27(4):671–682. https://doi.org/10.1139/cjce-27-4-671

Liu Y, Weerts AH, Clark M, Hendricks Franssen H-J, Kumar S, Moradkhani H, Seo D-J, Schwanenberg D, Smith P, van Dijk AIJM, van Velzen N, He M, Lee H, Noh SJ, Rakovec O, Restrepo P (2012) Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol Earth Syst Sci 16:3863–3887. https://doi.org/10.5194/hess-16-3863-2012

Maheswaran R, Khosa R (2013) Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River. J Hydroinf 15(3):1022–1041. https://doi.org/10.2166/hydro.2013.135

Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions Pattern Analysis and Machine Intelligence 11(7):674–693

Misiti M, Misiti Y, Oppenheim G, Poggi JM (1996) Wavelet toolbox:for use with MATLAB. The MathWorks Inc., Natick (MA), p 262

Misiti M, Misiti Y, Oppenheim G, Poggi JM (2006) Wavelet toolbox User's guide - for use with MATLAB. 3 ed. Massachusets (USA): The MathWorks, Inc.

Moradkhani H, Hsu K, Gupta HV, Sorooshian S (2004) Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. J Hydrol 295:246–262. https://doi.org/10.1016/j.jhydrol.2004.03.027

Pramanik N, Panda RK, Singh A (2010) Daily river flow forecasting using wavelet ANN hybrid models. J Hydroinf 13(1):49–63. https://doi.org/10.2166/hydro.2010.040

Ravansalar M, Rajaee T, Ergil M (2015) Prediction of dissolved oxygen in river Calder by noise elimination time series using wavelet transform. J Exp Theor Artif In 28(4):689–706. https://doi.org/10.1080/0952813X.2015.104253

Renno C, Petito F, Gatto A (2015) Artificial neural network models for predicting the solar radiation as input of a concentrating photovoltaic system. Energy Convers Manag 106:999–1012. https://doi.org/10.1016/j.enconman.2015.10.033

Santos CAG, Silva GBL (2014) Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models. Hydrol Sci J 59(2):312–324. https://doi.org/10.1080/02626667.2013.800944

Santos CAG, Freire PKMM, Torrence C (2013) A Transformada wavelet e sua Aplicação na Análise de Séries Hidrológicas (the wavelet transform and its application for hydrological time series analysis). Brazilian J Water Resour 18(3):271–280. https://doi.org/10.21168/rbrh.v18n3.p271-280

Santos CAG, Freire PKMM, Silva GBL, Silva RM (2014) Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias reservoir. Proc IAHS 364:100–105. https://doi.org/10.5194/piahs-364-100-2014

Santos CAG, Freire PKMM, Silva RM, Akrami SA (2019) Hybrid wavelet neural network approach for daily inflow forecasting using tropical rainfall measuring mission data. J Hydrol Eng 24(2):04018062. https://doi.org/10.1061/(asce)he.1943-5584.0001725

Shamseldin AY, O'Connor KM (2001) A non-linear neural network technique for updating of river flow forecasts. Hydrol Earth Syst Sci 5:577–598. https://doi.org/10.5194/hess-5-577-2001

Sivakumar B, Jayawerdana AW, Fernando TMKG (2002) River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. J Hydrol 265(1–4):225–245. https://doi.org/10.1016/S0022-1694(02)00112-9

Terzi Ö, Ergin G (2014) Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. Neural Comput & Applic 25:179–188. https://doi.org/10.1007/s00521-013-1469-9

Tiwari MK, Chatterjee C (2011) A new wavelet–bootstrap–ANN hybrid model for daily discharge forecasting. J Hydroinf 13(3):500–519. https://doi.org/10.2166/hydro.2010.142

Toth E, Brath A, Montanari A (1999) Real-time flood forecasting via combined use of conceptual and stochastic models. Phys Chem Earth B 24:793–798

Wang W, Ding J (2003) Wavelet network model and its application to the prediction of the hydrology. J Nat Sci 1:67–71

Wei S, Yang H, Song JX, Abbaspour K, Xu ZX (2013) A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. Hydrol Sci J 58(2):374–389. https://doi.org/10.1080/02626667.2012.754102

Xiong LH, O'Connor KM (2002) Comparison of four updating models for real-time river flow forecasting. Hydrol Sci J 47:621–639. https://doi.org/10.1080/02626660209492964

Yaseen ZM, El-Shafie A, Afan HA, Hameed M, Mohtar WHMW, Hussain A (2015) RBFNN versus FFNN for daily river flow forecasting at Johor River, Malaysia. Neural Comput Appl 27:1533–1542. https://doi.org/10.1007/s00521-015-1952-6