XLIII CILAMCE
21-25, November, 2022
FOZ DO IGUAÇU - PR - BRAZIL

Hybrid models for time series forecasting of the dam monitoring data

Eliete N. da Silva[1], Cassius T. Scarpin[2], Sheila R. Oro[3], Ubirajara F. Moreno[4], Nestor Roqueiro[4]

[1]*Postdoctoral student in Numerical Methods in Engineering, Universidade Federal do Paraná (UFPR)*
Jardim das Américas, Centro Politécnico, CEP 81531-980, Curitiba, Paraná, Brasil
elietenp@gmail.com
[2]*Department of General and Applied Administration, Universidade Federal do Paraná (UFPR)*
Av. Prefeito Lothário Meissner, 632, Jardim Botânico, CEP 80210-170, Curitiba, Paraná, Brasil
cassiusts@gmail.com
[3]*Postdoctoral student in Automation and Systems Engineering, Universidade Federal de Santa Catarina (UFSC)*
Linha Santa Bárbara, CEP 85601-970, Francisco Beltrão, Paraná, Brasil
sheilaro@utfpr.edu.br
[4]*Department of Automation and Systems, Universidade Federal de Santa Catarina (UFSC)*
Campus Universitário Reitor João David Ferreira Lima, Trindade, CEP 88040-900, Florianópolis, Santa Catarina, Brasil
ubirajara.f.moreno@ufsc.br, nestor.roqueiro@ufsc.br

Abstract. Time series forecasting is a result that contributes to analysis and decision-making and can be applied in several areas of knowledge. In the area of dam structural safety, this practice is little explored, although the equipment used in the monitoring feeds an extensive database. This work aims to apply a hybrid methodology for forecasting time series, integrated into the processing of data collected by monitoring instruments of a concrete dam. Wavelet Decomposition will perform the time series processing, then to separate the series components, an Autoregressive Integrated Moving Average model will be fitted. Residuals resulting from this mathematical/statistical model will be modeled through Artificial Neural Networks of Radial Basis Functions. The linear combination of these models will generate the time series forecast itself. The combination weights will be defined by solving a nonlinear programming problem. The investigated approaches will be compared and selected according to the smallest mean absolute percentage error measure. The partial series models do not need to have high performance for the prediction proposed to be satisfactory. The proposed approaches for predicting the test set have MAPE of less than 0.57%, while ARIMA and ANN-RBF models used separately reached values of up to 4.33%. The results indicate gains in forecasting assertiveness, aiding decision-making, which aims to create preventive measures to ensure dam safety.

Keywords: Wavelet Decomposition. Artificial Neural Networks of Radial Basis Functions. Concrete dam. Piezometer.

1 Introduction

The representation of a process through a mathematical model is a fundamental action when looking to evaluate its behavior over time. For modeling, generally, it is necessary to combine knowledge from different areas, such as engineering, computing, mathematics, and statistics.

When the process in question is composed of a set of values collected sequentially in time regarding a quantitative variable, it is called a time series [1]. The prediction of time series has been applied in several real contexts, such as the stock market [2], rainfall [3], and structural monitoring [4], among others.

Among time series forecasting methods, the following can be highlighted: Box and Jenkins models (B&J) and Artificial Neural Networks (ANN). The first is robust and capable of modeling data with linear and stationary characteristics [1]. On the other hand, the ANN can model data with complex structures of nonlinear autocorrelation [5]. Each real-time series of instruments reading has specific characteristics Therefore, testing is necessary to identify the most appropriate forecasting method; there is no general rule for defining a method beforehand.

The literature presents papers that have hybrid methods to reduce the risk of using an inappropriate forecasting method. In general, these studies combine models B&J, ANN, and data pre-processing [6], [7], [8]. The acceptance of hybrid methods is because real-time series usually combine linear and nonlinear features. Different methods are then needed to capture these various characteristics [7] and [9].

The history of the readings of the instruments used in the structural monitoring of dams forms an extensive time series. At the Itaipu Hydroelectric Power Plant, monitoring is carried out with the support of more than 2700 instruments. By analyzing instrumentation data, for example, it is possible to identify causes and consequences of reactions of the foundation, allowing preventive maintenance to be carried out, thus increasing the safety of the dam and its useful life [10]. This study considered only one piezometer, installed in the foundation of a stretch of the Itaipu dam, responsible for measuring the sub-pressure in this region [11].

In this work, a proposal of a hybrid model composed of wavelet decomposition (WD), autoregressive integrated moving average (ARIMA), and radial basis function neural network (RBF-NN) is developed. The choice of RBF-NN was because of its simple structure, accuracy in nonlinear approximation, rapid learning of complex patterns and trends present in the data, as well as its rapid adaptation to changing data over time. This model can be applied to stationary and non-stationary series, which have linear or nonlinear autocorrelation structures.

2 Time series models

The wavelet decomposition can deal with non-stationary time series, for this reason, it is being applied in time series analysis. The WD results in new hierarchical series that are more "well-behaved" for modeling and forecasting. The approximation series has little noise (unexplained part of the time series), while the low order detail series is more susceptible to noise [12].

The original time series is decomposed into different components, approximation, and low frequency, which maintains the general trend of the series and many details components that contain high-frequency components and are separated at different frequencies and scales [13]. These components are established through dilations and translations according to Equation (1), and these operations generate a system of orthonormal functions, related to time and frequency [14].

The properties of translation and dilation are represented in a single function $\psi_{j,k}(t)$, being that the change of j and k generate the so-called Wavelets daughters from the mother wavelet $\psi_0(t)$, $t \in Z$. Where j is the contraction factor ($j \neq 0$), if its value increases and of dilation otherwise (scaling), and k is the translation term. The successive variation of these parameters promotes the wavelet transform. Therefore, the initial signal of the time series is decomposed into distinct frequency elements, and each of them is analyzed according to a scale resolution [15].

$$\psi_{j,k}(t) = \frac{1}{\sqrt{j}}\psi_0\left(\frac{t-k}{j}\right) \tag{1}$$

The autoregressive integrated moving average method can be applied to model non-stationary time series with linear autocorrelation between the data. The ARIMA method is one of the most established linear models for predicting time series, and it is widely adopted in hybrid models to increase the prediction capacity [7].

A non-stationary time series can be represented by the ARIMA (p, d, q), according to Equation (2). The differentiation of the series is performed to achieve stationarity on average.

$$\emptyset(B)\nabla^d Z_t = \theta(B)a_t \tag{2}$$

Where, B, is the delay operator; $\phi(B) = 1 - \sum_{i=1}^{p} \emptyset_i B^i$, is the characteristic autoregressive polynomial; $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$, is the moving average characteristic polynomial; and a_t, is the random noise.

Complex nonlinear time series and dynamic systems can be modeled using ANN It imposes few assumptions for its application, can generalize even though it is applied to non-stationary data, and uses fewer parameters than other methods. These parameters are adjusted iteratively and optimized by learning the historical patterns [7].

The structure of the radial basis function neural network (Fig. 1) is composed of some neurons (nodes) in the hidden layer; the lag is used to form the patterns and their dimensions. The smaller the structure of the network, the less cost and computational speed it has besides expanding its generalization capacity, avoiding overfitting the RBF network. Computational experiments are performed to designate them and to establish a better structure for RBF-NN. After the definition of the input patterns, the training and testing sets of the network must be separated. The training will define the structure and parameters of the neural network, and the test patterns will reveal the adjustment error.

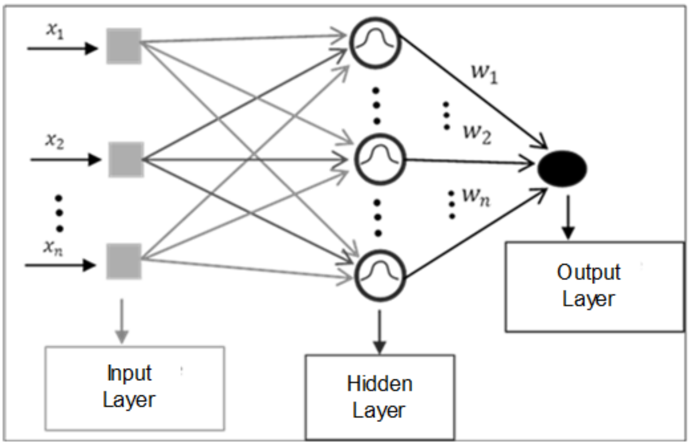

Figure 1. Radial basis functions neural network structure.

3 Proposed hybrid methodology

The new hybrid method called Wavelet ARIMA-Radial Basis Function (WA-RBF) is composed of the following phases: WD of level two of data series; ARIMA modeling of each decomposition series; calculation of the residuals of each ARIMA model; RBF-NN modeling of each residue series; combining all settings. A representation of these phases is presented in the flowchart in Figure 2.

The identification of the structure and the definition of the parameters of the ARIMA model is performed through the training set (data within the sample). Once the model is established, it is applied to the test set (data outside the sample), generating the fit and prediction from the defined structure.

The RBF-NN modeling of the residues of each decomposition component was investigated, changing the amount of data in the pattern and the number of centroids. This methodology allows making the forecast one step ahead and n steps ahead.

The linear combination of the adjustments performed is defined in equation (3).

$$\hat{y}_{WA-RBF,t} = \eta_1 \hat{A}_{2,t} + \eta_2 \widehat{D}_{1,t} + \eta_3 \widehat{D}_{2,t} + \hat{e}(\hat{A}_{2,t}) + \hat{e}(\widehat{D}_{1,t}) + \hat{e}(\widehat{D}_{2,t}) \tag{3}$$

Where, $\hat{A}_{2,t}$, $\widehat{D}_{1,t}$, and $\widehat{D}_{2,t}$ are the adjusted values of the decomposition series; $\hat{e}(\hat{A}_{2,t})$, $\hat{e}(\widehat{D}_{1,t})$, and $\hat{e}(\widehat{D}_{2,t})$, the adjusted values of the residuals of the decomposition series, and η are the weights of the linear combination, defined by the nonlinear objective function programming problem.

The objective function minimizes the mean squared error (MSE) and all parameters defined in the training process are applied to the test set.

Also, this methodology generates two classes of models regarding the fit and type of forecast (one step and n forward steps). Furthermore, this methodology generates two classes of models regarding fit and forecast type (one-step and n-steps forward).

We use the methodology by dynamically changing the training set and consequently, the test set, indicated by 1_step. After defining the first prediction, a new training set is formed, without the time series's first element and including the previous prediction's actual value. This new setting generates the next prediction. The first element of this new training set is eliminated and includes the original value of the second prediction. This procedure is repeated until the forecast horizon is satisfied. The structures of the models used in the methodology are the same used previously.

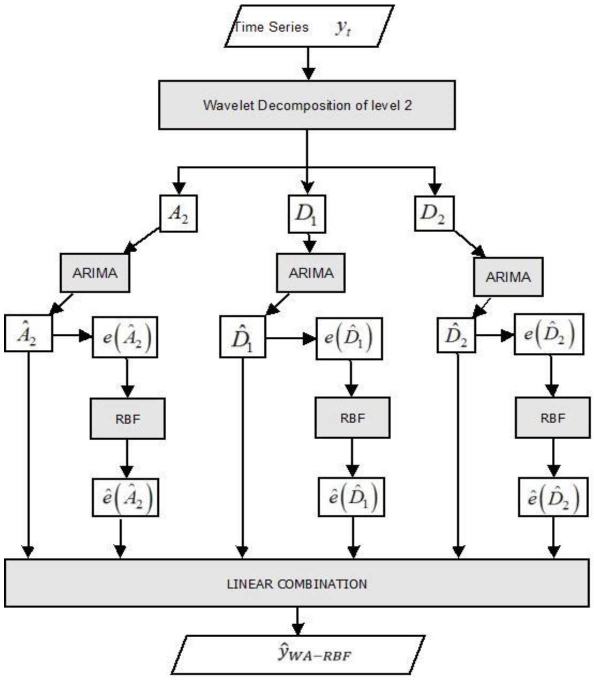

Figure 2. Stages of the proposed hybrid methodology.

4 Application of methodology

The methodology was applied to the time series of the piezometer instrument, installed in a key block of the Itaipu Dam, to monitor the contact between two basaltic flows of the dam's foundation. The analysis was performed from the values of manual readings, carried out between the years 2001 to 2015, totaling 180 data. The training set adopted for the modeling consisted of 174 data, and the test set used to make the prediction contained six values, the equivalent of a semester of readings. Therefore, six months was the forecast horizon adopted.

The piezometer series was considered non-stationary by the Dickey-Fuller test ($t_{calc} = -2.5 > -2.9 = t_{5\%}$). Also, the graph of the integrated periodogram of the piezometer time series (Fig. 3) indicated that the sequence of measurements was not entirely random, with a systematic part to be modeled.

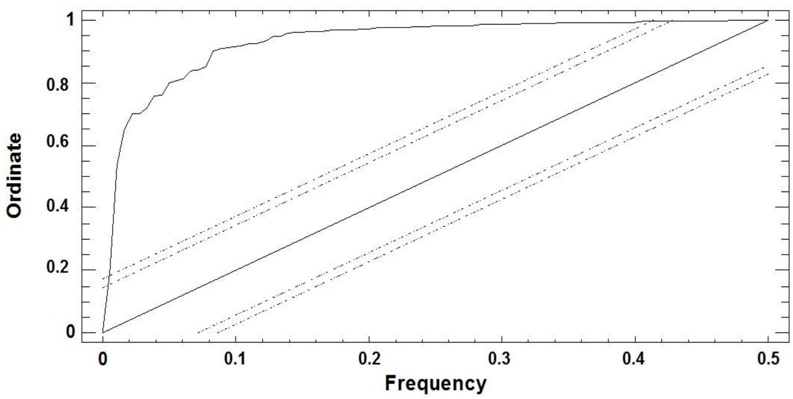

Figure 3. Integrated periodogram of the 2001-2015 series.

First, a DW of level two was performed in the piezometer series using the Wavelet Db45, approximation component A_2, which promoted a smoothing of the initial series; and two detail components D_1 and D_2. The detail components have an oscillatory behavior, as seen in Figure 4, with the highest peaks occurring in regions where the approximation component has greater smoothing than the original series. The ranges of decomposition values are smaller when compared to the approximation component.

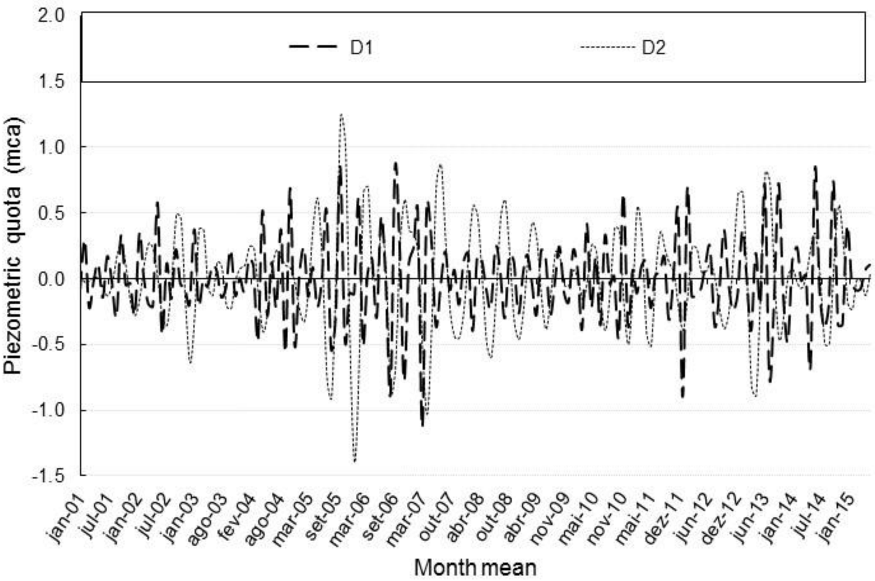

Figure 4. Detail wavelet components D_1 and D_2.

The ARIMA models for each series of decomposition are, ARIMA (8,1,4) for A_2, ARIMA (1,0,2) for D_1, and ARIMA (2,0,0) for D_2. The residuals of each modeling were calculated and showed no random behavior among the data in the series. Therefore, the RBF was used at this stage to model the nonlinearity between the residuals.

The RBF structure obtained for the residual series $e(A_2)$ and $e(D_1)$ had four elements in the input pattern while $e(D_2)$ had five elements. All series had seven hidden layer neurons.

Then, the modeling was performed through the combination of the series resulting from the ARIMA and RBF models, attributes weigh only the modeling of the wavelet components of the series, the modeling of the residuals of these series are only summed, it is defined by $\hat{y}_{IWA-RBF,t}$ in equation (6).

$$\hat{y}_{1WA-RBF,t} = 1.000\hat{A}_{2,t} + 1.019\hat{D}_{1,t} + 0.969\hat{D}_{2,t} + \hat{e}(\hat{A}_{2,t}) + \hat{e}(\hat{D}_{1,t}) + \hat{e}(\hat{D}_{2,t}) \quad (6)$$

The assertiveness of the model's adjustments can be seen in Table 1 and Figure 5. The proposed model 1WA-RBF obtained the best adjustment, followed by ARIMA and, finally, the RBF-NN. Both models present a lag about the original series.

Table 1. Descriptive statistics of the model adjustment.

Model	MSE	MAE
1WA-RBF	8.2 x10^{-3}	6.68 x10^{-2}
ARIMA	7.8 x10^{-1}	6.07 x10^{-1}
RBF-NN	9.0 x10^{-1}	6.72 x10^{-1}

Source: The authors.

After defining the models through the training set, the prediction of the test set was obtained by each of the models. The test prediction established by the classic models in the literature (ARIMA and RBF-NN) did not obtain good results, with the RBF showing a trend. In contrast, the ARIMA resulted in an extrapolation. Still, the ARIMA model can be expected to make a better prediction than the RBF, since the training set modeling was better, but the prediction did not confirm this.

As a reference for the analysis of the behavior of the series, the historical minimum and maximum values of the piezometer, are 63.99mca and 100.59mca, respectively. Figure 6 illustrates the prediction of the test set referring to the proposed methodology. The "6_Steps" specification indicates that the series size and training data are not changed, so the patterns are not modified. The prediction of the first standard of the test set is informed to the network to compose the next standard; this is performed for the entire forecast horizon (6 months).

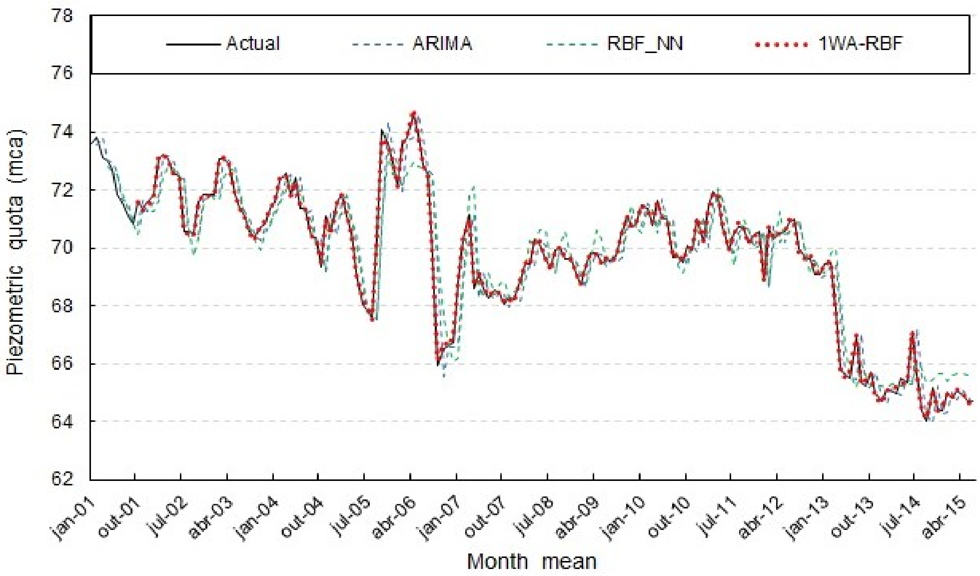

Figure 5. Adjustment of the models.

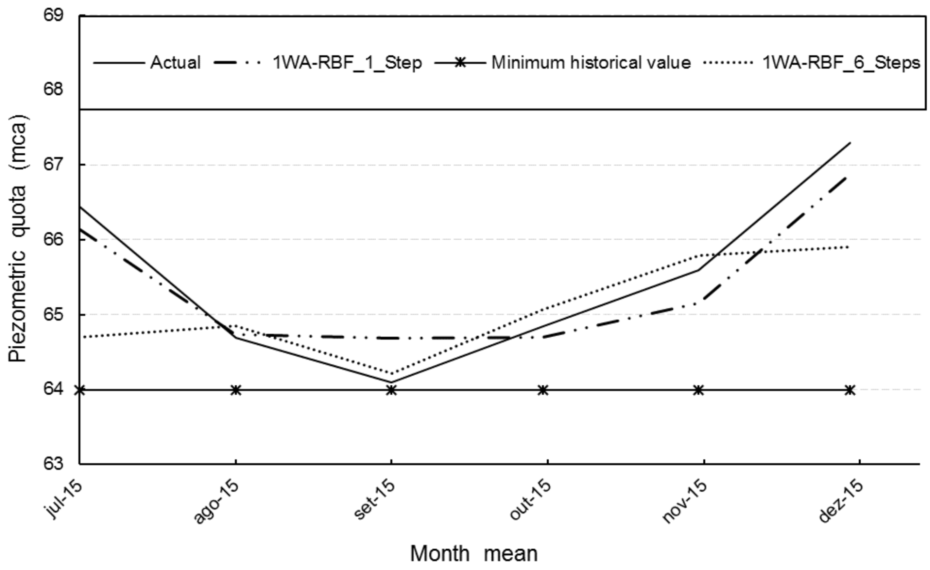

Figure 6. Predict a step with a sliding window of the test set of the proposed methodology.

The inclusion and removal of data from the series, represented by the model $\hat{y}_{IWA-RBF\_6\_steps}$, showed better results in the July and December 2015 forecasts, as seen in Figure 6. Furthermore, the models based on the proposed methodology could follow the series data trend. The predictions made one step ahead, by the model$\hat{y}_{IWA-RBF_1_steps}$, were generally better, according to the MAPE statistics presented in Table 2.

Table 2. MAPE statistics for the forecast of each element of the test set

Model	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6
$\hat{y}_{IWA-RBF\_6\_steps}$	2.621%	0.256%	**0.204%**	0.357%	0.302%	2.052%
$\hat{y}_{IWA-RBF\_1\_steps}$	**0.450%**	**0.069%**	0.921%	**0.215%**	0.669%	**0.633%**
ARIMA	2.704%	0.154%	0.698%	0.548%	1.766%	4.326%
RBF	1.619%	1.174%	2.084%	1.027%	**0.010%**	2.616%

5 Conclusions

In this article, we combine the application of wavelet decomposition with ARIMA models and radial basis function neural networks to perform non-stationary time series prediction. The proposed methodology was applied to the historical values of a piezometer used to monitor a stretch of the Itaipu Dam. The wavelet decomposition made it possible to improve the modeling of the instrument's time series.

The results indicated that the proposed hybrid methodology, performed both in multi-step and one-step, presented better predictions when compared to the traditional ARIMA and RBF-NN methods.

The methodology was able to assertively model and predict the piezometer time series. Considering that it is a non-stationary series with a nonlinear correlation between the data the ARIMA and RBF-NN methods could not model the characteristics of this time series. The WA-RBF model developed by the proposed methodology resulted in predicted values more in line with reality. This result is important for the team responsible for monitoring and dam safety, as it provides more robust information to support decision-making and can generate savings since any physical intervention in monitoring requires time and high expenses.

Acknowledgments. The authors are grateful for the support provided by the Itaipu Technological Park (PTI), Center for Advanced Studies in Dam Safety (CEASB), Graduate Program in Numerical Methods in Engineering (PPGMNE-UFPR), Graduate Program in Automation and Systems Engineering (PPGEAS-UFSC) and Coordination for the Improvement of Higher Education Personnel (CAPES).

Authorship statement. The authors hereby confirm that they are the solely liable persons responsible for the authorship of this work and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors or has the permission of the owners to be included here.

References

[1] G. E. P. Box; G. M. Jenkins; G. C. Reinsel and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
[2] F. Affonso; T. M. R. Dias and A. L. Pinto, “Financial times series forecasting of clustered stocks”. *Mobile Networks and Applications*, vol. 26, n. 1, pp. 256–265, 2021.
[3] S. Sharma; P. D. Ranjeet; I. S. Mirdha and R. S. Sikarwar, “Precipitation trend analysis by Mann-Kendall test of different districts of Malwa”. *Agroclimatic Zone. Environment and Ecology*, vol. 36, n. 2A, pp. 664–671, 2018.
[4] A. Silik; M. Noori; W. A. Altabey; R. Ghiasi and Z. Wu, "Analytic wavelet selection for time-frequency analysis of big data form civil structure monitoring". In: *International Workshop on Civil Structural Health Monitoring*. Springer, Cham, pp. 431–455, 2021.
[5] S. S. Haykin, *Redes Neurais*. 2ed., Bookman, 2001.
[6] E. Lazzarotto; L. A. Teixeira Júnior; L. M. Gramani; A. Chaves Neto and E. M. C. Franco, “A wavelet SARIMA-ANN linear combination with multiple stages in time series prediction”. *International Journal of Engineering and Applied Sciences*, vol. 7, pp. 8–16, 2015.
[7] M. Khashei and M. Bijari, “Hybridization of the probabilistic neural networks with feed-forward neural networks for forecasting”. *Engineering Applications of Artificial Intelligence*. vol. 25, n. 6, pp. 1277–1288, 2012.
[8] J. M. Corrêa; A. Chaves Neto; L. A. Teixeira Júnior; E. M. C. Franco and A. E. Faria, “Time series forecasting with the WARIMAX-GARCH method”. *Neurocomputing*. vol. 216, pp. 805–815, 2016.
[10] H. Jiurian and S. Bingfeng, “The application of ARIMA-RBF model in urban rail traffic volume forecast”, In: *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, pp. 1662–1665, 2013.
[11] J. F. A. Silveira, *Instrumentação e Comportamento de Fundações de Barragens de Concreto*. Oficina de Textos, 2003.
[12] Superintendência de Engenharia da Itaipu Binacional. *Usina Hidrelétrica de Itaipu - Aspectos de Engenharia*. Itaipu Binacional, 2009.
[13] A. B. Mabrouk; N. B. Abdallah and Z. Dhifaoui. Wavelet decomposition and autoregressive model for time series prediction. *Applied Mathematics and Computation*, vol. 199, pp. 334–340, 2008.
[14] Y. Seo; S. Kim; O. Kisi and V. P. Singh. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, vol. 520, pp. 224–243, 2015.
[15] C. S. Kubrusly. *The Elements of Operator Theory*, 2 ed., Birkhäuser, 2011.
[16] S. Mallat. *A Wavelet Tour of Signal Processing*, 3 ed., Academic Press, 2008.