Nov. 2015. Vol. 7. No. 04 ISSN2305-8269
International Journal of Engineering and Applied Sciences

www.eaas-journal.org

A WAVELET SARIMA-ANN LINEAR COMBINATION WITH MULTIPLE STAGES IN TIME SERIES PREDICTION

[1]EMERSON LAZZAROTTO, [2]LUIZ ALBINO TEIXEIRA JUNIOR, [3]LILIANA MADALENA GRAMANI, [4]ANSELMO CHAVES NETO, [5]EDGAR MANUEL CARREÑO FRANCO

[1,5]Asstt. Prof., Eng. and Exact Sciences Center, State University of West Paraná, Foz do Iguaçu, Brazil
[2]Assoc. Prof., ILATIT, Federal University of Latin America Integration, Foz do Iguaçu-PR, Brazil
[3,4]Assoc. Prof., Department of Mathematic and Statistic, Federal University of Paraná, Curitiba-PR, Brazil
E-mail: [1]emerson.lazzarotto@gmail.com, [2]luiz.a.t.junior@gmail.com, [3]l.gramani@gmail.com, [4]anselmo@ufpr.br, [5]emfra1@gmail.com

ABSTRACT

This paper proposes a hybrid methodology for combining forecasts to (stochastic) time series referred to as Wavelet Linear Combination (WLC) SARIMA-ANN with Multiple Stages. Firstly, the wavelet decomposition of level p is performed, generating (approximations of the) p+1 wavelet components (WCs). Then, the WCs are individually modeled by means of a Box and Jenkins model and an artificial neural network in order to capture, respectively, plausible linear and non-linear structures of auto-dependence for, then, being linearly combined, providing hybrid forecasts for each one. Finally, all of them are linearly combined by the WLC of forecasts (to be defined). For evaluating, it was used the Box and Jenkins (BJ) models, artificial neural networks (ANN), and its traditional Linear Combination (LC1) of forecasts; and ANN integrated with the wavelet decomposition (ANN-WAVELET), BJ model integrated with the wavelet decomposition (BJ-WAVELET), and its conventional Linear Combination (LC2) of forecasts. All predictive methods applied to the monthly time series of average flow of the Itaipu plant, located in Foz do Iguaçu, Brazil. In all analysis, the proposed hybrid methodology has provided higher predictive performance than the other ones.

Keywords: ***Time Series, Wavelet Decomposition, Box and Jenkins Models, Artificial Neural Networks, Linear Combination of Forecasts.***

1. INTRODUCTION

Over the years, many prediction methods (or predictive) have been proposed in order to design, each time more accurately, time series (stochastic). Generally speaking, they can be grouped into two disjoint classes: individual predictive methods (which are divided into the statistical approach and artificial intelligence), and the combination of individual predictive methods [1], [2], [3], [4], [5]; [6], [7], [8], and [9]. It is important to note that the term "combination of predictive individual methods" can be used in a broad sense, referring both to the approach of combining predictions [3], as the combination of predictive Bayesian densities [9]. In this article, however, it is used only to refer to the combination of forecasts.

In the literature, there is a range of individual predictive methods, which are widely used in projection time series, for example, Box and Jenkins [10] and artificial neural networks (ANNs) [11] - which capture respectively, linear and non-linear self-reliance structures. Thus, since the time series, generally, exhibit both, using a model of Box and Jenkins or an ANN, according to [12] and [13], may involve loss of information (Linear, by ANN, and non-linear, by the model of Box & Jenkins) relevant to achieving greater predictive power. Therefore, predictions generated from the combination of plausible individual models, not biased and significantly different tend to be more accurate, because, according to [9], are, in fact, aggregators of information from different sources.

Wavelet Analysis (or Wavelet theory), in turn, derives important helper methods of preprocessing consisting basically in making decomposition, filtering or smoothing of temporal data before its effective modeling [14]. Various approaches used, in an integrated way, individual predictive methods and methods of wavelet preprocessing. Some of them can be viewed in: [15], [16], [17] and [18].

As exposed and the fact that there are a relatively small number of method of combining individual predictive models that make the processing of the data, it is proposed the: "linear combination wavelet SARIMA-ANN with multiple stages". Basically, the proposed method can be described in the following basic steps: (1) measurements are obtained in a time series to be modeled; (2) it is conducted at a p level of decomposition, thus generating first approximation component and p detail components; (3) it is modeled each $p + 1$ Wavelet components, (WC) from step (2) through a model of Box and Jenkins (to map a plausible linear structure), and an artificial neural network (to map a non-linear structure plausible); (4) combine the predictions for each WC originating from both the individual predictors mentioned by linear traditional forecasts, where the adjustment of adaptive parameters is made by means of non-linear programming (so long as they are $p + 1$ Wavelet components, linear combinations are required and, there for, $p + 1$ nonlinear programming problem, in this step, and (5) linearly combine the predictions of each hybrid WC generated in step (4) generating the predicted the desired time series. Note that, in step (5), it is necessary to solve over a non-linear program, totaling $p + 2$ nonlinear programs (which is what justifies the term "multi-stage" in the method name).

Besides the introduction, the article is divided into seven sections. Section 2 presents basic concepts of wavelet theory. In sections 3 and 4, are introduced, respectively, models of Box & Jenkins and artificial neural networks. The definition of linear combination is written in section 5. Section 6 describes in detail the methodology proposed. Finally, in the sections 7 and 8, are set, respectively, the main results and the conclusions.

2. WAVELET ANALYSIS

Take the ordered pair $(l^2, \langle ; \rangle)$,where l^2 is defined by the collection of all infinite sequences of quadratically summable complex numbers (i.e.,$l^2 \coloneqq \{f: \mathbb{Z} \to \mathbb{C} / \sum_{t\in\mathbb{Z}} |f(t)|^2 < \infty\}$, where $\mathbb{Z}$ and $\mathbb{C}$ are, respectively, integers and the set of complex numbers) and the map $\langle ; \rangle : l^2 \to \mathbb{C}$ is an intern product. An element $\omega(\cdot)$ is l^2 a vector l^2 −wavelet or simply wavelet vector, with inner product $\langle ; \rangle : l^2 \to \mathbb{C}$, if and only if, the doubly indexed sequence $\{\omega_{m,n}(\cdot)\}_{(m,n)\in\mathbb{Z}\times\mathbb{Z}}$ consists of an orthonormal basis for l^2, where the parameter m is called dyadic scale parameter and n, called translation parameter. In turn, one element $\phi(\cdot) \in l^2$ is a vector l^2-scale (or simply vector scale), inner product$\langle ; \rangle : l^2 \to \mathbb{C}$, if, and only if, the collection $\{\phi_{m,n}(\cdot)\}_{(m,n)\in\mathbb{Z}\times\mathbb{Z}}$, where, for all $m, n \in \mathbb{Z}$, $\phi_{m,n}(\cdot) = 2^{m/2}\phi(2^m(\cdot) - n)$ is such that $\langle \phi_{l,i}(\cdot); \phi_{j,k}(\cdot)\rangle = 0$, always that $l = j$ and $i \neq k$, and $\langle \phi_{l,i}(.); \phi_{j,k}(.)\rangle \neq 0$, otherwise.

Based on [19] and [20] a vector $f(.)$ in l^2 can be decomposed orthogonally uniquely in terms of a base orthonormal wavelet, represented by$\{\phi_{m_0,n}(.)\}_{n\in\mathbb{Z}} \cup \{\omega_{(m,n)}(.)\}_{(m,n)\in\{m\}_{m=m_0}^{+\infty}\times\mathbb{Z}}$, as in (1).

$$f(\cdot) = f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{+\infty} f_{W_m(\omega)}(\cdot) \quad (1)$$

Where: $f_{V_{m_0}(\phi)}(\cdot) \coloneqq \sum_{n\in\mathbb{Z}} a_{m_0,n}\,\phi_{m_0,n}(\cdot)$ is the approximation component of level m_0, for $a_{m_0,n} \coloneqq \langle f(\cdot), \phi_{(m_0,n)}(\cdot)\rangle$ (that is, the usual inner product between the signal vector $f(\cdot)$ and the vector scale level m_0 and n); and $f_{W_m(\omega)}(\cdot) \coloneqq \sum_{n\in\mathbb{Z}} d_{m,n}\,\omega_{m,n}(\cdot)$ is the level m of detail component, and that $d_{m,n} \coloneqq \langle f(\cdot), \omega_{(m,n)}(\cdot)\rangle$ (that is, the usual inner product between $f(\cdot)$ and the wavelet vector of level m an n). $a_{m_0,n}$ and $d_{m,n}$ are, respectively, the wavelet coefficients of approximation and detail. In (1), there is the wavelet decomposition, in terms of $\{\phi_{m_0,n}(.)\}_{n\in\mathbb{Z}} \cup \{\omega_{(n,m)}(.)\}_{(m,n)\in\{m\}_{m=m_0}^{+\infty}\times\mathbb{Z}}$, of $f(\cdot)$ about l^2.

3. BOX & JENKINS MODELING

According to [10], if a time series $(y_t)_{t=1}^{T}$ is an auto-regressive moving average process with integrated with means of the order of p and q (*ARMA(p,q)*), if, and only if, it can be represented as in (2).

$$y_t = \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Where: $\varphi_{i}{}_{i=1}^{p}$ and $\theta_{j}{}_{j=1}^{q}$ consists on lists of complex parameters that satisfy the conditions of invertibility and stationary [21]; $\varepsilon_{t-j}{}_{j=0}^{q}$ is a realization of a stochastic white noise process centered on zero [22]. In (4), there is an alternative representation of the model in (3), which is given in terms of two polynomials.

$$(1 - \varphi_1 B - \cdots \varphi_p B^p) y_t = (1 - \theta_1 B - \cdots \theta_q B^q)\varepsilon_t \quad (3)$$

Where: B is a unilateral shift [20] such that: $B^k y_t = y_{t-k}$, k being a strictly positive integer. In

class of models ARMA(*p*,*q*), it follows that: AR*(p)* represents the polynomial $\varphi(B) = \left(1 - \varphi_1 B - \cdots \varphi_p B^p\right)$ of order p of the autoregressive portion; and MA(q), the polynomial $\theta(B) = (1 - \theta_1 B - \cdots \theta_q B^q)$ of the order q of moving averages. In turn, if a time series$(y_t)_{t=1}^T$ present seasonal effects (stationary or non-stationary homogeneous) or non-homogeneous stationary on average, then a class plausible models is generally given in (4).

$$\varphi(B)(1 - \Phi_1 B^S - \cdots \Phi_P B^{PS})\nabla^d \nabla_S^D y_t = \theta(B)(1 - \Theta_1 B^S - \cdots \Theta_Q B^{QS})\epsilon_t \quad (4)$$

Where:$\Phi_m \in \mathbb{C}$ and $\Theta_n \in \mathbb{C}$, for *m=1,…,P* and *n=1,…,Q*, are the complex parameters regarding seasonal polynomials, respectively, the autoregressive part and average seasonal furniture, and meet the conditions of invertibility and stationary [21]; *D* is the order of the operator $\nabla_S^D$of seasonal differences (required in cases of non-homogeneous stationary seasonal component $(y_t)_{t=1}^T$), which is defined by $\nabla_S^D := (1 - B^S)^D$; *S* is the seasonal period (if annual, then *S = 12*); *d* is the order of the operator $\nabla := (1 - B)$ of the simple differences (required in cases of non-homogeneous stationary on average $(y_t)_{t=1}^T$); and polynomials $\varphi(B) := \left(1 - \varphi_1 B - \cdots \varphi_p B^p\right)$ and $\theta(B) := (1 - \theta_1 B - \cdots \theta_q B^q)$ refer to simple auto-regressive and parts moving average (equation (3)), respectively. The class Box and Jenkins models in (4), is denoted by *SARIMA (P,D,Q) × (p,d,q)*.

4. ARTIFICIAL NEURAL NETWORKS

The *Artificial Neural Networks* (or simply ANNs) are very flexible computing frameworks for modeling and forecasting a broad range of stochastic time series, because they just require they exhibit either linear and non-linear auto-dependence structures. As is the case of most statistical linear models, the stationarity property are not required by ANN approaches (as in [10]). Another important aspect is that the ANNs are universal approximators of compact (i.e., closed and bounded) support functions, as pointed out by [23]. Thus, since a time series y_t $(t = 1, \dots, T)$ that depends on its own past may be seen as points a compact support, it follows that the ANNs are capable to approximate (for modeling or forecasting) it with a high degree of accuracy. According [12], their predictive power comes from the parallel processing of the information from the data. In addition, the ANN models are largely determined by the stochastic characteristics inherent in the time series.

In perspective, the *feed-forward multi-layer perceptron ANNs* (referred to from now on as ANNs) are the most widely used prediction model for time series forecasting. Particularly, a single hidden layer ANN is characterized by a network composed by three layers of simple processing units numerically connected by acyclic links. The relationship between the output at instant *t*, denoted by y_t, and the *p*-lagged inputs, represented by the sequence y_{t-k} $(k = 1, \dots, p)$, has got the following mathematical representation:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} \, y_{t-i}\right) + \varepsilon_t \quad (5)$$

where α_j $(j = 0, 1, \dots, q)$ and β_{ij} $(i = 0,1,, \dots, p;\ j = 0, 1, \dots, q)$ are the (single hidden layer) ANN parameters, which are often called the connection weights; p is the number of input nodes; q is the number of hidden nodes; ε_t is the approximation error at time *t*; and $g(.)$ is here a logistic function, although it would be possible to adopt another transfer function (please, see [11] for more details). The logistic function is widely used as the hidden layer transfer function in forecasting processes and its mathematical representation is given by

$$g(x_t) = \frac{1}{1 + \exp(-x_t)} \quad (6)$$

where $x_t := \beta_{0j} + \sum_{i=1}^{p} \beta_{ij} \, y_{t-i}$ and $exp(.)$ is the exponential function with Euler's basis (as in [11]). Due to $g(.)$ is a non-linear transfer function, the ANN model, in (5), in fact performs a non-linear mapping from the past observations y_{t-k} $(k = 1, \dots, p)$ to the future state y_t. Equivalently, the model in (5) can be rewritten, as follows:

$$y_t = f\left(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w\right) + \varepsilon_t \quad (7)$$

where w denotes a vector of all ANN parameters and $f\left(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w\right)$ is the model determined by the network structure and connection weights. Indeed, the neural network is equivalent to a non-linear auto-regressive model.

In practice, w is an unknown vector of ANN parameters and hence needs to be adjusted. So, in order to find the optimal solution $\hat{w}$, accounting for some criteria, for the vector of ANN parameter w, some optimization algorithm must be employed. Although there are several methodologies in specialized literature, maybe the Levemberg-Marquardt's algorithm (as in [11]) might be considered most used for this assignment. The minimization in-sample squared error mean (i.e.,

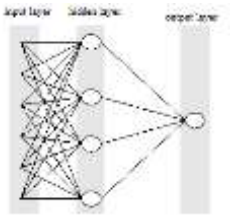
$min_w \sum_{t=1}^{T} \varepsilon_t{}^2$) is usually used as numerical criteria. Thus, it is desired that the solution $\hat{w}$ of this optimization problem is the argument that minimizes the $\sum_{t=1}^{T} \varepsilon_t{}^2$. Once obtained $\hat{w}$, it has

$$f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \hat{w}) + \hat{\varepsilon}_t \quad (8)$$

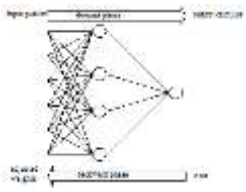
where $f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, \hat{w})$ is the final ANN output at instant t which consists of forecast, denoted by $\hat{y}_t$, of the state y_t, and $\hat{\varepsilon}_t$ is its forecasting error.

In Figure 1(a), it has become a more common architecture illustration of an ANN (feedforward) multilayer perceptron (MLP) with three layers: the input layer (input), hidden layer (or intermediate) and output layer (output).

The first layer of the neural network is the input layer, being the only exposed to the input patterns. The input layer imparts the values of input patterns to the neurons of the intermediate layer so that they extract patterns and transmit the results to the output layer (last layer of the ANN). The setting of the number of neurons in the hidden layer is empirically performed.



(a) ANN Feedforward of 3 layers.



(b) Phases of the Backpropagation Algorithm.

Figure 1 - Artificial Neural Network Feedforward MLP and Algorithm Backpropagation.

The main neural network training algorithm is backpropagation, which the synaptic weights setting occurs by means of an optimization process carried out in two phases: forward and backward, as shown in Figure 1(b). In the forward phase, the answer provided by the ANN for a given input pattern is calculated. In the backward phase, the deviation (or error) between the response of ANN and the desired response is used in the adjustment process of synaptic weights. Throughout the ANN training, the various input patterns and associated desired responses are presented to the ANN, such that the synaptic weights are such as to minimize the sum of squared errors (MSE). The prediction of future values of a time series, by an ANN starts with assembling the training patterns (input/output pairs), which depends on the definition of the size of the time window L (for past values of own time series you want to foresee and to the explanatory variables) and h forecast horizon. In an auto-regressive (linear or nonlinear) process, the input pattern is formed by past values of the series itself that you want to predict.

In turn, the pattern of desired output is the value of observation of the time series over the forecast horizon. In Figure 2, it is illustrated how is generally built the training set in the case of the prediction be based on the last 4 values passed. Note that the construction of the network training patterns consists of moving the input and output windows over the entire time series so that each pair of windows (input/output) functions as a training pattern and must be presented repeatedly until the learning algorithm converges.

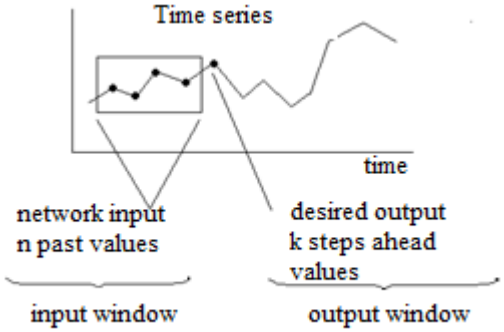


Figure 2 - Installation of the training set.

5. COMBINATION OF FORECASTS

Assume that $(y_t)_{t=1}^{T}$ is a time series and that $\{M_j\}_{j=1}^{m}$, $m \geq 2$, is a collection of predictive methods plausible to its modeling in order to generate forecasts. In this process, you can choose any method $\{M_j\}_{j=1}^{m}$, based on some selection criteria, or take k methods in $\{M_j\}_{j=1}^{m}$,where $k \leq m$, and combine them. Take k methods in the collection $\{M_j\}_{j=1}^{m}$, where $k \leq m$. Assume that ∇^k is the set with all the forecasts (in and out of the sample) from the k methods chosen in $\{M_j\}_{j=1}^{m}$,considering a forecast with h steps ahead. Consider also that ∇_C is the set with all predictions derived from certain combination of forecasts belonging to the set ∇^k.

By combining the forecasts (general), it is understood to be a form $\hat{y}_C: \nabla^k \rightarrow \nabla_C$ such that maps to an array of predictions $(\hat{y}_{t,k})_{k=1}^{K} \in \nabla^k$ in a combined prediction $\hat{y}_{t,C} \in \nabla_C$ where t is an integer in $1 \leq t \leq T + h$, where h forecasting horizon and T is the cardinality of the training sample. In particular, the linear combination of predictions consists of a form $\hat{y}_{CL}: \nabla^k \rightarrow \nabla_{CL}$ that maps a vector estimates $(\hat{y}_{t,k})_{k=1}^{K} \in \nabla^k$ in a linearly combined in a prediction $\hat{y}_{t,CL} \in \nabla_{CL}$, which is defined in (**9**).

$$\hat{y}_{t,CL} = \left[\sum_{i=1}^{K} (\rho_i) \times \hat{y}_{t,i}\right] + \beta \tag{9}$$

Where: ρ_i is the weight associated with the adaptive prediction $\hat{y}_{t,i}$, t being an integer in the time interval $1 \leq t \leq T + h$, which is obtained by the method M_i; and $\beta$are, respectively, the multiplicative and additive adaptive constants.

6. METHODOLOGY PROPOSAL

For $\chi_{[0,T]}$ a map such that $\chi_{[0,T]} \coloneqq 1$, if $t \in \{1, \dots, T\}$ and $\chi_{[0,T]} \coloneqq 0$ if $t \in \mathbb{Z} - \{1, \dots, T\}$ and $y(\cdot) \coloneqq (y_t)_{t=1}^{T}$ is a time series of size T, where $T > 1$. Thus, the composition $\chi_{[0,T]} \circ y(\cdot) = \tilde{y}(\cdot): \mathbb{Z} \longrightarrow \mathbb{R}$ can be visualized as a sequence $\tilde{y}(\cdot) \coloneqq (\dots, 0, 0, y_1, y_2, \dots, y_T, 0, 0, \dots)$ in l^2. Assuming that $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}(\cdot)\}_{(m,n) \in \{m\}_{m=m_0}^{+\infty} \times \mathbb{Z}}$ is a base for orthonormal wavelet for l^2, which follows, based on Section 2, the time series $y(\cdot)$ admits to be decomposed, approximately, as in (10).

$$\begin{aligned} \tilde{y}(\cdot) \simeq \tilde{\tilde{y}} &= \sum_{n=1}^{n_{m_0}} a_{m_0,n} \phi_{m_0,n}(\cdot) \\ &+ \sum_{m=m_0}^{m_0+(p-1)} \sum_{n=1}^{n_m} d_{m,n} \omega_{m,n}(\cdot) \end{aligned} \tag{10}$$

The expansion, in (10), is called wavelet decomposition level p. The value adopted for m_0 level parameter is usually the same of p. Rewriting (10), it follows that:

$$\tilde{y}(\cdot) \simeq \tilde{\tilde{y}} = y'_{V_{m_0(\phi)}}(\cdot) + \sum_{m=m_0}^{m_0+(p-1)} y'_{W_{m(\omega)}}(\cdot) \tag{11}$$

Where, in l^2: $y'_{V_{m_0(\phi)}}(\cdot) \coloneqq \sum_{n=1}^{n_{m_0}} a_{m_0,n} \phi_{m_0,n}(\cdot)$ is an approach to the WC of approximation $y_{V_{m_0(\phi)}}(\cdot)$; and $y'_{W_{m(\omega)}}(\cdot) \coloneqq \sum_{n=1}^{n_m} d_{m,n} \omega_{m,n}(\cdot)$, of WC of detail $y_{W_{m(\omega)}}(\cdot)$.

After obtaining the $p+1$ approximation to the WCs, in (11), it is carried out the modeling of each model by means of a Box and Jenkins (linear predictor) and an artificial neural network (non-linear predictor), mapping, respectively, self-reliance linear and non-linear structures. In order to produce the following sequences forecasts, in and out of the sample, and h denotes the forecast horizon and T'' and T' are the degrees of freedom lost by ANN models and Box & Jenkins respectively: $\hat{y}_{V_{m_0}(\phi),BJ}(\cdot) \coloneqq (\hat{y}_{V_{m_0}(\phi),BJ,t})_{t=T'}^{T'+h}$; and $\hat{y}_{V_{m_0}(\phi),ANN}(\cdot) \coloneqq (\hat{y}_{V_{m_0}(\phi),ANN,t})_{t=T''}^{T''+h}$; and $\hat{y}_{W_m(\omega),BJ}(\cdot) \coloneqq (\hat{y}_{W_m(\omega),BJ,t})_{t=T'}^{T'+h}$; and $\hat{y}_{W_m(\omega),ANN}(\cdot) \coloneqq (\hat{y}_{W_m(\omega),ANN,t})_{t=T''}^{T''+h}$, for each integer value in the interval $m_0 \leq m \leq m_0 + (p-1)$.

Taking the training sample originating from the individual forecasts of predictive methods Box and Jenkins and artificial neural networks approximation for each component is made a linear combination of the predictions, as in (12).

$$\begin{aligned} \hat{y}_{V_{m_0}(\phi),CL,t} \coloneqq (&\hat{y}_{V_{m_0}(\phi),BJ,t} \times \rho_{V_{m_0}(\phi),BJ}) \\ &+ \hat{y}_{V_{m_0}(\phi),ANN,t} \times \rho_{V_{m_0}(\phi),ANN} \\ &+ \beta_{V_{m_0}(\phi),CL} \end{aligned} \tag{12}$$

Where: $\hat{y}_{V_{m_0}(\phi),BJ,t}$ is the prediction of Box & Jenkins model for $y_{V_{m_0}(\phi),t}$; $\hat{y}_{V_{m_0}(\phi),ANN,t}$ is the prediction of the artificial neural network model for $y_{V_{m_0}(\phi),t}$; $\rho_{V_{m_0}(\phi),BJ}$ is adaptive weight associated to $\hat{y}_{V_{m_0}(\phi),BJ,t}$; $\rho_{V_{m_0}(\phi),ANN}$ is adaptive weight associated to $\hat{y}_{V_{m_0}(\phi),ANN,t}$; $\beta_{V_{m_0}(\phi),CL}$ is the additive adaptive constant; and finally, $\hat{y}_{V_{V_{m_0}(\phi),CL,t}}$ is the forecasting linearly combined for the approximation $y_{V_{m_0}(\phi),t}$.

The adjustment of the adaptive parameters in (**12**) occurs by means of a mathematical programming problem (MPP) [24] which is described below. Objective function (*OF*)

$$OF: \min_{\rho_{V_{m_0}(\phi),BJ}; \rho_{V_{m_0}(\phi),ANN}; \beta_{V_{m_0}(\phi),CL}} \{MSE_{V_{m_0}(\phi),CL}\}$$

Subject to

$$\begin{cases} \hat{y}_{V_{m_0}(\phi),CL,t} := \hat{y}_{V_{m_0}(\phi),BJ,t} \times \rho_{V_{m_0}(\phi),BJ} \\ +\hat{y}_{V_{m_0}(\phi),ANN,t} \times \rho_{V_{m_0}(\phi),ANN} + \beta_{V_{m_0}(\phi),CL} \\ MSE_{\mathrm{V_{m_0}}(\phi),CL} := \dfrac{\sum_{t=1}^{T}\left(y_{V_{m_0}(\phi),t} - \hat{y}_{V_{m_0}(\phi),CL,t}\right)^2}{T} \\ \rho_{V_{m_0}(\phi),BJ} + \rho_{V_{m_0}(\phi),ANN} = 1 \\ \rho_{V_{m_0}(\phi),ANN}, \rho_{V_{m_0}(\phi),ANN} \geq 0 \end{cases}$$

Then, the linear combination is made of the predictions for each level of detail component m, as in (13).

$$\begin{aligned} \hat{y}_{W_m(\omega),CL,t} := \hat{y}_{W_m(\omega),BJ,t} \times \rho_{W_m(\omega),BJ} \\ + \hat{y}_{W_m(\omega),ANN,t} \\ \times \rho_{W_m(\omega),ANN} + \beta_{W_m(\omega),CL} \end{aligned} \quad (13)$$

Where: $\hat{y}_{W_m(\omega),BJ,t}$ is the prediction of the Box & Jenkins model for $y_{W_m(\omega),t}$; $\hat{y}_{W_m(\omega),ANN,t}$ is the prediction of the artificial neural network model for $y_{W_m(\omega),t}$; $\rho_{W_m(\omega),BJ}$ is the adaptive weight associated with the prediction $\hat{y}_{W_m(\omega),BJ,t}$; $\rho_{W_m(\omega),ANN}$ is the adaptive weight associated with the prediction $\hat{y}_{W_m(\omega),ANN,t}$; $\beta_{W_m(\omega),CL}$ is the additive adaptive constant; and $\hat{y}_{W_m(\omega),CL,t}$ is forecasting linearly combined for the approximation of $y_{W_m(\omega),CL,t}$.

The MPP used, in the adjust of the combination, in (13), is the following.

$$OF: \min_{\rho_{W_m(\omega),BJ}; \rho_{W_m(\omega),ANN}; \beta_{W_m(\omega),CL}} \{MSE_{W_m(\omega),CL}\}$$

Subject to

$$\begin{cases} \hat{y}_{W_m(\omega),CL,t} := \hat{y}_{W_m(\omega),BJ,t} \times \rho_{W_m(\omega),BJ} \\ +\hat{y}_{W_m(\omega),ANN,t} \times \rho_{W_m(\omega),ANN} + \beta_{W_m(\omega),CL} \\ MSE_{W_m(\omega),CL} := \dfrac{\sum_{t=1}^{T}\left(y_{W_m(\omega),t} - \hat{y}_{W_m(\omega),CL,t}\right)^2}{T} \\ \rho_{V_{m_0}(\phi),BJ} + \rho_{V_{m_0}(\phi),ANN} = 1 \\ \rho_{W_m(\omega),ANN}, \rho_{W_m(\omega),ANN} \geq 0 \end{cases}$$

This procedure is repeated on individual shaping of each of the components p of the wavelet detail. That is, until now are necessary $p+1$ stages of adjustments. Finally, the hybrid linearly combined predictions for the wavelet components of approximation and detail are combined, also linearly, as in (14).

$$\begin{aligned} \hat{y}_{CL,wave,t} := \left(\rho_{V_{m_0}(\phi)} \times \hat{y}_{V_{m_0}(\phi),CL,t}\right) \\ + \left(\sum_{m=m_0}^{m_0+(p-1)} \rho_{W_m(\omega)} \times \hat{y}_{W_m(\omega),CL,t}\right) + \beta_{CL,wave} \end{aligned} \quad (14)$$

Where: $\hat{y}_{V_{m_0}(\phi),CL,t}$ is the forecasting linearly for $y_{V_{m_0}(\phi),t}$; $\hat{y}_{W_m(\omega),CL,t}$ is the forecasting linearly for $y_{W_m(\omega),t}$; $\rho_{V_{m_0}(\phi)}$ is the adaptive weight associated with the prediction $\hat{y}_{V_{m_0}(\phi),CL,t}$; $\rho_{W_m(\omega)}$ is the adaptive weight associated with the prediction $\hat{y}_{W_m(\omega),CL,t}$; $\beta_{CL,wave}$ is the additive adaptive constant; and $\hat{y}_{CL,wave,t}$ is the forecasting linearly combined for the point y_t.

The formulation of the MPP used in adjusting the linear combination of the wavelet components, (14), is described below.

$$OF: \min_{\rho_{V_{m_0}(\phi)}; \rho_{W_m(\omega)}; \beta_{CL,wave}} \{MSE_{CL,wave}\}$$

Subject to

$$\begin{cases} \hat{y}_{CL,wave,t} := \left(\rho_{V_{m_0}(\phi)} \times \hat{y}_{V_{m_0}(\phi),CL,t}\right) + \\ \left(\sum_{m=m_0}^{m_0+(p-1)} \rho_{W_m(\omega)} \times \hat{y}_{W_m(\omega),CL,t}\right) + \beta_{CL,wave} \\ MSE_{CL,wave} := \dfrac{\sum_{t=1}^{T}\left(y_t - \hat{y}_{CL,t}\right)^2}{T} \end{cases}$$

The hybrid forecast linearly combined for time series $\tilde{y}$ (and to the original time series y, at time of $T+h$) is achieved by replacing the optimal values of the adaptive weights and additive adaptive constant at (14).

7. COMPUTER EXPERIMENT

To illustrate the hybrid proposed methodology, it was modeled time series of the average monthly flow of Itaipu, which period is from January 1970 until December 2010 (i.e., 492 observations). Their choice was due to its relevance and difficulty to modeling. For experimental purposes, 80% of the first data were used in the training of individual models (Box & Jenkins and artificial neural networks), as well as to obtain the optimum parameters of the adaptive forecasts combinations; 10% of the data were used in subsequent validation sample; and 10% of remaining in the test sample. To do this, short-term predictions were made a step forward, in a horizon of 49 steps ahead forecast. Residuals statistics to assess the performance of predictive methods were MAE (mean absolute error) and MAPE (mean absolute percentage error). In Figure 3, there is the graph of the time series with all the 492 observations.

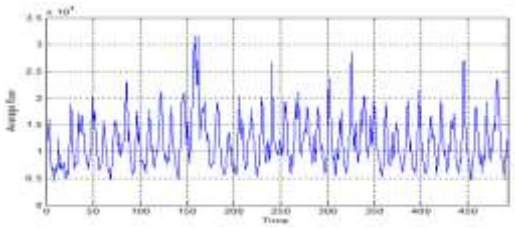
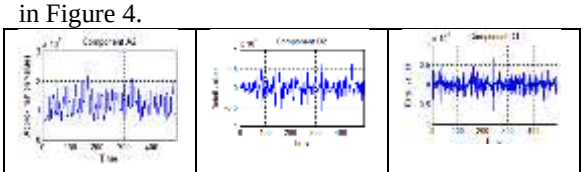


Figure 3 - Total time series of monthly average flow of the Itaipu plant.

7.1 Wavelet Decomposition

The wavelet decomposition level 2 (in MATLAB software 2013) was implemented and is presented in Figure 4.



(a)Approximation of the WC of approximation level 2 (A2).	(b)Approximation of the WC of detail level 2 (D2)	(c)Approximation of the WC of detail level 1 (D1)

Figure 4 - Approximation of WCs of the total time series average monthly flow of Itaipu plant.

The choice of level 2 was made to preserve parsimony in the number of individual models to be estimated. In relation to the base orthonormal Haar wavelet, it was used for providing the best forecasts.

7.2 Modeling

For comparison, the modeling of time series was carried out above using the methods of artificial neural networks (ANN) and of Box & Jenkins (BJ), with traditional approach. Moreover, the predictive methods were used BJ and ANN integrated with the two-level wavelet decomposition (respectively denoted by ANN-WAVELET and BJ-WAVELET) [15]. Finally, were used the estimates of linear combinations of approaches BJ and ANN (CL1) and the ANN-WAVELET and BJ-WAVELET (CL2) as the approach of [7]. The process of adjusting the adaptive weights CL1 and CL2 occurred by means of a PPM, which the objective function has the minimum mean square error of the training sample residuals (as usual) and the convex weights (i.e., not normalized and negative) [9].

Table 1 - MAPE values, in samples of training, validation and test.

Methods	**MAPE(%)**		
	Training	*Validation*	*Test*
ANN	16.35	15.19	21.78
BJ	16.60	19.46	21.25
CL 1	14.95	14.82	19.13
ANN-WAVELET	2.11	2.47	2.80
BJ-WAVELET	2.74	3.13	3.63
CL 2	2.06	2.38	2.64
Proposed Method	1.27	1.27	1.36

Table 2 - MAE values, in samples of training, validation and test.

Methods	**MAE**		
	Training	*Validation*	*Test*
ANN	1,886.87	1,682.38	2,389.97
BJ	2,032.22	2,185.50	2,560.19
CL 1	1,877.10	1,752.25	2,274.33
ANN-WAVELET	224.00	232.73	290.79
BJ-WAVELET	291.92	278.34	396.26
CL 2	218.29	222.45	275.75
Proposed Method	131.83	121.71	141.95

In Table 1 and Table 2, note that CL1 was superior to the individual ANN and BJ predictive methods, where the optimal linear adaptive weights associated with ANN and BJ predictions were respectively equal to 0.618388907 and 0.323249491. The same can be checked using the combination CL2 method, when compared with the methods based on ANN-WAVELET and BJ-WAVELET. The optimal linear adaptive weights associated with forecasts derived from ANN-WAVELET and BJ-WAVELET, respectively, equal to 0.910473771 and 0.089013323. Therefore, we used the solver of the EXCEL software 2007. The optimization algorithm used in both cases was the evolutionary and the default settings were kept unchanged. It is possible to check the level of wavelet decomposition level 2, with the orthonormal basis of Haar, resulted in predictive material gains. Finally, the combination linear wavelet hybrid multistage was higher than the others in all other three samples.

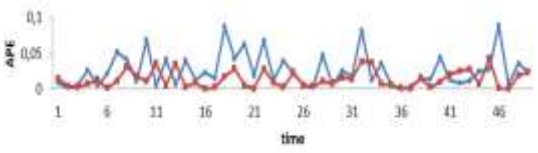


Figure 5 - Temporal evolution of out-of-sample of Absolut Percentage Error (APE) values of the CL2 model (blue line) and the proposed method (red line).

In Figure 5, there is a comparison between the proposed method and the CL2 approach (which was the best in Table 1 between the benchmark methods). Note that, in every 49 months, in the test sample, the present method showed stable over the

time course of the APE values of CL2 (for example, in 6 months the APE values of the CL2 violate the range of 6% while this is at no time in the proposed method). Moreover, in most instants, the APE proposed combination shows values lower than those of the CL2.

8. CONCLUSIONS

In this article, it was proposed a new hybrid approach combining forecasts, which uses in an integrated manner, models of Box & Jenkins and artificial neural networks, the Wavelet Analysis and Nonlinear Programming. In general, it can be described in three basic stages: (1) it expands via wavelet decomposition on level *p*, a time series (namely, monthly average flow of Itaipu) is generating *p + 1* approximations of WCs (being an approximation of WC and *p* detail WCs); (2) Modelling up individually the WCs obtained in (1) by means of a model of Box and Jenkins (to capture linear structures) and an artificial neural network (to capture non-linear structures); and (3) linearly combine the predictions generated in (2), using the *p+2* linear combinations wavelet numerical predictions adjustable via nonlinear programming, producing hybrid combined predictions for the time series to be provided. To compare it, were considered the following prediction methods: Box and Jenkins (BJ), artificial neural networks (ANN) and its traditional linear combination (CL1); and integrated artificial neural network with wavelet decomposition (ANN-WAVELET), Box and Jenkins integrated with wavelet decomposition (BJ-WAVELET) and its traditional linear combination (CL2).

Analyzing the results, it follows that, in Table 1 and Table 2, the method proposed in accordance with the statistics MAPE and MAE, was superior to all others, the training samples, validation and testing - which shows more power learning (the sample workout) and generalization (in the validation and test samples). In Figure 5, in turn, obtained from the second best method of Table 1 and Table 2, (namely, CL2), greater evolution temporal stability of the APE values (absolute percentage error). Furthermore, it has been that the proposed method generated forecasts that are strongly correlated with the data (showing good performance), confirming its accuracy in computational experiment. Ultimately, it is noted that although the theoretical basis of the proposed methodology, at least in part, be based on mathematical highly complex content (e.g., the Wavelet Theory), the software mentioned in the text, enable its use in actual applications so as to be operated so as, in relative terms, simple.

REFERENCES

[1] J.M. Bates and C.W.J. Granger, "The Combination of Forecasts," *Operational Research Quarterly*, vol. 20, pp. 451-468, Dec. 1969.

[2] P. Newbold and C.W.J. Granger, "Experience with Forecasting Univariate Time Series and the Combination of Forecasts," *Journal of the Royal Statistical Society, Series A*, vol. 137, no. 2, pp. 131-165, 1974.

[3] S. Makridakis and R.L. Winkler, "Averages of forecasts: Some empirical Results.," *Management Science*, vol. 29, no. 9, pp. 987-996, 1983.

[4] S. Gupta and P.C. Wilton, "Combination of forecasts: An extension," *Management Science*, vol. 33, no. 3, pp. 356-372, 1987.

[5] S. Makridakis, "Why Combining works?," *International Journal of Forecasting*, vol. 5, no. 4, pp. 601-603, 1989.

[6] B.E. Flores and E.M. White, "Subjective versus objective combining of forecasts: an experiment.," *Journal of Forecasting*, vol. 8, no. 3, pp. 331-341, July/September 1989.

[7] B.E. Flores and E.M. White, "A Framework for the combination of forecasts," *Journal of the Academic of Marketing Science*, vol. 16, no. 3-4, pp. 95-103, 1998.

[8] H. Zou and Y. Yang, "Combining Time Series Models for Forecasting," *International Journal of Forecasting*, vol. 20, no. 1, pp. 69-84, 2004.

[9] A. Faria and E. Mubwandarinkwa, "Multimodality on the geometric combination of bayesian forecasting models," *International Journal of Statistics and Management System*, vol. 3, no. 1-2, pp. 1-25, 2008.

[10] J.D. Hamilton, *Time Series Analysis*. Princenton: Princeton University Press, 1994.

[11] S. Haykin, *Redes Neurais Artificiais: princípios e práticas*, 2nd ed. Porto Alegre: Bookman, 2001.

[12] G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003.

[13] K.F. Wallis, "Combining forecasts : forty years later," *Applied Financial Economics*, vol. 21, no. 1-2, pp. 33-41, 2011.

[14] R.R.B. Aquino et al., "Application of wavelet and neural network models for wind speed and power generation forecasting in a Brazilian experimental wind park," *International Joint Conference on Neural Networks*, pp. 172-178, June 2009.

[15] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425-455, Aug. 1994.

[16] D.L. Donoho, I.M. Jonhstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?," *Journal of the Royal Statistical Society, Series B*, vol. 57, no. 2, pp. 301-369, 1995.

[17] C. Lei and L. Ran, "Short-term wind speed forecasting model for wind farm based on wavelet decomposition," *Third Conference on Electric Utility Deregulation and Restructuring and Power Technologies - DRPT*, pp. 2525-2529, April 2008.

[18] T. Ogden and E. Parzen, "Data dependent wavelet thresholding in nonparametric regression with change-point applications," *Computational Statistics and Data Analysis*, vol. 22, no. 1, pp. 53-70, 1996.

[19] C.S. Kubrusly and N. Levan, "Abstract Wavelets Generated by Hilbert Space Shift Operators," *Adv. Math. Sci. Appl.*, vol. 16, pp. 643-660, 2006.

[20] C.S. Kubrusly, *The Elements of OperatorTheory*, 2nd ed. Boston: Birkhauser, 2011.

[21] G.E.P. Box and G.M. Jenkins, *Time series analysis forecasting and control*. San Francisco: Holden-Day, 1976.

[22] P.A. Morettin and C.M.C. Toloi, *Análise de Séries Temporais*, 2nd ed. São Paulo: Edgard Blucher, 2006.

[23] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function," *Math. Control Signals Systems*, vol. 2, no. 4, pp. 303-314, 1989.

[24] C.T. Ragsdale, *Spreadsheet Modeling & Decision Analysis: A practical Introduction to Management Science*, 4th ed.: South-Western, 2004.