INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.400

TIME SERIES FORECASTING BY USING A NEURAL ARIMA MODEL BASED ON WAVELET DECOMPOSITION

Eliete Nascimento Pereira
Federal University of Paraná, Brazil
E-mail: elietenp@gmail.com

Cassius Tadeu Scarpin
Federal University of Paraná, Brazil
E-mail: cassiusts@gmail.com

Luíz Albino Teixeira Júnior
Latin American Integration University (UNILA),Brazil
E-mail: luiz.a.t.junior@gmail.com

Submission: 30/10/2015
Revision: 23/11/2015
Accept: 26/11/2015

ABSTRACT

In the prediction of (stochastic) time series, it has been common to suppose that an individual predictive method – for instance, an Auto-Regressive Integrated Moving Average (ARIMA) model – produces residuals like a white noise process. However, mainly due to the structures of auto-dependence not mapped by a given individual predictive method, this assumption might be easily violated, in practice, as pointed out by Firmino et al. (2015). In order to correct it (and accordingly to produce forecasts with more accuracy power), this paper puts forward a Wavelet Hybrid Forecaster (WHF) that integrates the following numerical techniques: wavelet decomposition; ARIMA models; Artificial Neural Networks (ANNs); and linear combination of forecasts. Basically, the proposed WHF can map simultaneously linear – by means of a linear combination of ARIMA forecasts – and non-linear – through a linear combination of ANN forecasts – auto-dependence structures exhibited by a given time series. Differently of other hybrid methodologies existing in literature, the WHF forecasts are produced carrying into account implicitly the information from the

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

frequency presenting in the underlying time series by means of the Wavelet Components (WCs) obtained by the wavelet decomposition approach. All numerical results show that WHF method has achieved remarkable accuracy gains, when comparing with other competitive forecasting methods already published in specialized literature, in the prediction of a well-known annual time series of sunspot.

Keywords: *Wavelet decomposition, ARIMA model, Artificial neural networks, Linear combination of forecasts*

1. INTRODUCTION

Over the years, several forecasting methods have been proposed with the aim of producing increasingly accurate predictions of (stochastic) time series. In general, they could be split into two great exclusive categories: the individual (or single) predictive methods, such as the well-known *Auto-Regressive Integrated Moving Average* (ARIMA) models, as in Hamilton (1994); and the combination of individual predictive methods, proposed initially by Bates and Granger (1969). Indeed, the collection of single predictive methods might yet be regrouped into two exclusive classes: the statistical one (here it lies, for instance, the (linear) ARIMA models), and the machine learning one (here it lies the (non-linear) *Artificial Neural Networks* (ANNs), as in HAYKIN, 2001). By hybrid forecasting methods, it means those ones that always carry out the modelling of a given time series, denoted by $y_t(t = 1, \dots, T)$, according to the following three steps: in Step 1, a single forecasting method from statistical class/machine learning class is applied to $y_t(t = 1, \dots, T)$ for producing its forecasts as well as its residuals; in Step 2, the forecasting errors generated in Step 1 are predicted by using an individual forecaster from machine learning class/statistical class; and, in step 3, the forecasts provided in Step 1 are “corrected” by the predictions of the residuals produced in Step 2 such that to generate the hybrid forecasts of the underlying time series $y_t(t = 1, \dots, T)$. In effect, a hybrid predictive method can be referred to as particular case of combined forecasters.

In the process of prediction of time series, it has often been to assume that a single predictive method produces residuals like a white noise process (i.e., unpredictable). However, mainly due to the linear or non-linear structures of auto-dependence not captured by a single predictive method adopted by the decision

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

maker, this supposition may be trivially broken, most applications in real world. For instance, the (linear) ARIMA models are able to statistically guarantee, based on a significance level α, only uncorrelated residuals (but not statistical independence, as it is often assumed); this because, from mathematical point of view, those ones may be visualized as linear filters, as pointed out by Hamilton (1994). In turn, Zhang (2003) shows in his numerical experiments – on which three very popular time series were forecasted by employing an hybrid forecaster – that the residuals produced by the ARIMA models could be efficiently predicted by using non-linear forecasters (namely, the *Multi-Layer Perceptron* ANNs (MLP-ANNs), as in HAYKIN, 2001)). In his research, it can be verified that the ARIMA model forecasts were in fact remarkably improved when received the sum of the prediction of their respective residuals. Finally, in the numerical experiments in Firmino et al. (2015), each time series was modelled by means of several plausible ARIMA models whose forecasts were added by the predictions of their residuals produced by different ANNs; the results achieved exhibit a remarkable accuracy gain in the hybrid forecasts when compared with other traditional methods, in all adherence statistics.

In turn, wavelet decompositions of levels *r* have shown to be very efficient in dealing with time series forecasting due to its multi-scaling property, as highlighted in Teixeira Jr et al. (2015). Basically, from a given time series $y_t(t = 1, \dots, T)$, through a wavelet decomposition of level *r*, its *r*+1 *Wavelet Components* (WCs) – namely, one WC of approximation at scale m_0, represented by $y_{A_{m_0},t}$(t=1,...,T), and *r* WCs of detail at scales $m_0, m_0 + 1, \dots, m_0 + (r - 1)$, denoted by $\tilde{y}_{D_{m_0},t}$, $\tilde{y}_{D_{m_0+1},t}$, …, $\tilde{y}_{D_{m_0+(r-1)},t}$ $(t = 1, \dots, T)$, respectively – are produced. In one hand, the WCs can be interpreted as frequency patterns of the underlying time series $y_t(t = 1, \dots, T)$ represented at time domain. Each WC tends to exhibit lesser noise in its stochastic pattern than the underlying time series. On the other hand, more formally speaking, each WC is an orthogonal projection of $y_t(t = 1, \dots, T)$ onto a pairwise orthogonal subspace (as in Kubrusly (2011)), recognized as the “wavelet subspaces”, of the space l^2 (defined in Section 2.1). In this way, accuracy gains may be achieved when

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

the wavelet methods (as the wavelet decomposition) are adopted in the forecasting process.

Thus, the *Wavelet Hybrid Methodology* (WHF) proposed here aims to produce hybrid forecasts aggregating different information from different sources (i.e., numerical methods) with higher level of accuracy than other methodologies presenting in literature. The WHF integrates the following approaches: wavelet decomposition of level *r*, in Section 2.1; ARIMA models, in Section 2.2; ANN methods, in Section 2.3. Section 3 describes detail all steps to carried out the WHF (in particular, presents the *Linear Combination* (LC) of forecasts proposed here). In general lines, the proposed WHF can capture, at the same time, linear (by means of a LC of ARIMA models) and non-linear (by using a LC of ANNs) auto-dependence structures exhibited by a given time series to be forecasted, as well as its information on spectral frequency. In a different way of other hybrid predictive methods in time series literature, the WHF hybrid forecasts implicitly adheres spectral information through the WCs and perform an alternative linear combination of forecasts, as it will be seen. In order to illustrate the proposed methodology, the Canadian lynx and the British pound/US dollar exchange rate time series are modelled and hence predicted, following Zhang (2003).

The current paper is split into 5 sections. Beyond the introduction (Section 1), in Section 2, a review of the used methodologies is presented; Section 3 describes the steps of the proposed WHF; in Section 4, the main numerical results are delayed and comments upon; and, finally, in Section 5, the paper is closed.

2. REVIEW OF LITERATURE

The purpose of this section is to present a brief review of some basic concepts which are needed for defining the WHF method, described in Section 3. It starts, in Section 2.1, by describing the wavelet decomposition of level *r*, which is the algorithm adopted in initial step of the ARIMA and ANN methods. This is followed by the basic concepts on ARIMA and ANN models, in Sections 2.2 and 2.3, respectively.

2.1. Wavelet decomposition of level *r*

Based on Kubrusly and Levan (2006), and Teixeira Jr et al.(2015), if a subset $\{\omega_{m,n}(.)\}_{\substack{m,n\in Z \\ m\geq m_0}} \cup \{\phi_{m_0,n}(.)\}_{n\in Z}$, wherein m_0 takes a fixed integer value, is in fact an

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

orthonormal basis for the space l^2, wherein $l^2 := \{y; \sum_{t\in\mathbb{Z}} |y_t|^2 < \infty\}$, then a vector $y_t (t \in \mathbb{Z})$ belonging to l^2 can be orthogonally decomposed, in terms of $\{\omega_{m,n}(.)\}_{\substack{(m,n\in\mathbb{Z}\\ m\geq m_0}} \cup \{\phi_{m_0,n}(.)\}_{n\in\mathbb{Z}}$, as in (1).

$$y_t = y_{A_{m_0},t} + \sum_{m=m_0}^{+\infty} y_{D_m,t} \quad (1)$$

Where: $y_{A_{m_0},t} := \sum_{n\in\mathbb{Z}} a_{m_0,n}\, \phi_{m_0,n}(.)$ is recognized as the WC of approximation at level m_0 of the state y_t, with $a_{m_0,n} := \langle y_t; \phi_{m_0,n}(.)\rangle$; and $y_{D_m,t} := \sum_{n\in\mathbb{Z}} d_{m,n}\, \omega_{m,n}(.)$ is referred to as the WC of detail at level m of y_t, with $d_{m,n} := \langle f(.); \omega_{m,n}(.)\rangle$. The orthogonal decomposition in (1) is usually called a *wavelet decomposition* of y_t.

Tautologically, any finite scalar-valued complex time series $y_t (t = 1, \ldots, T)$ may be interpreted as an infinite sequence $y_t (t \in \mathbb{Z})$ in l^2, defined as follows: $y_t = y_t$, if $t \in \{1, 2, \ldots, T\}$; and $y_t = 0$, if $t \in \mathbb{Z} - \{1, 2, \ldots T\}$. In effect, each state y_t can be orthogonally expanded by means of a wavelet decomposition, as the identity in (1). However, in practical terms, it cannot model individually all WCs generated by the expansion (1). Consequently, an adaptation is required in order to obtain a finite number of WCs. In this perspective, Donoho and Johnstone (1994) have proposed the *wavelet decomposition of level r*, wherein $r > 0$ and $r \in \mathbb{Z}$, defined for each state y_t, as below

$$y_t = A_{m_0,t} + \sum_{m=m_0}^{m_0+(r-1)} D_{m,t} + \varepsilon_t \quad (2)$$

where m_0 is the level parameter (which is commonly assumed to be equal to *r*); $A_{m_0,t} := \sum_{n=0}^{2^{(M-m_0)}-1} a_{m_0,n}\, \phi_{m_0,n}(t)$ and $D_{m,t} := \sum_{n=0}^{2^{(M-m_0)}-1} d_{m,n}\, \omega_{m,n}(t)$ (that consist, respectively, of the approximations of the WCs $y_{A_{m_0},t}$ and $y_{D_m,t}$ from Equation (1), wherein m is such that $m = m_0, \ldots, m_0 + (r-1)$); M is a parameter that takes an

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

integer value such that $T = 2^M$; and ε_t is the error vector of approximation of the state y_t (where ε_t is usually assumed to be equals zero).

Importantly, if *T* is not an integer power of 2, it is often to fill in the underlying time series with zeros until its length *T* is increased up to the next integer power of 2. This procedure may be always carried out due to the fact that the zeros added up do not affect the calculation of the WCs $A_{m_0,t}$ and $D_{m,t}$ in (2), what means the previously auto-correlation exhibited by $y_t(t = 1, \ldots, T)$, as well its WCs, is preserved (HAVEN; LIU; SHEN, 2012).

2.2. ARIMA models

Let $y_t(t = 1, \ldots, T)$ be a time series exhibiting structure of auto-correlation. According to Liu (2006), $y_t(t = 1, \ldots, T)$ is an ARIMA (*p,d,q*) process, if only if, it can be represented as in Equation (3):

$$\nabla^d y_t = \phi_0 + \sum_{j=1}^{p} \phi_j \nabla^d y_{t-j} + \sum_{j=1}^{q} \theta_i \, e_{t-i} + e_t \qquad (3)$$

Where: B is a backward operator, defined by $B^k y_t := y_{t-k}$, wherein *k* runs over in the set $\mathbb{Z}$; $\nabla^d := (1 - B)^d$ is the difference operator, with *d* representing its order; $(\phi_i)_{i=0}^{p}$ and $(\theta_j)_{j=1}^{q}$ are the lists of model complex parameters, with $\phi_p \neq 0$ and $\theta_q \neq 0$, where they must satisfy both the invertibility and the stationarity conditions (as in HAMILTON, 1994); e_t is the innovation consisting of a state of the random variable ε_t of a white noise stochastic process, denoted by $\varepsilon_t(t = 1, \ldots, T)$, with zero mean and null auto-covariance; *p* and *q* are, respectively, the orders of the auto-regressive part, denoted by $\sum_{j=1}^{p} \phi_j \nabla^d y_{t-j}$, and of the moving average part, represented by part $\sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$.

According to Liu (2006), in order to obtain the best possible ARIMA model, three basic steps should be carried out: (i) test the plausible values for the parameters *p*, *d*, and *q*, in Equation (3), which can be obtained through the profile

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

analysis of the plots of simple and partial auto-correlation functions of the residuals (as in HAMILTON, 1994); (ii) define the method to be used to estimate the ARIMA parameters (the most common is the Maximum Likelihood Estimation method (as in Liu (2006)); and, (iii) make a diagnostic check to choose the most parsimonious and adequate model to be used for generating both the in-sample and the out-of-sample forecasts of $y_t(t = 1, \dots, T)$.

2.3. Artificial Neural Network

Artificial Neural Networks (ANNs) are well-known to be flexible computing frameworks for modeling and forecasting a broad range of stochastic time series exhibiting either linear or non-linear auto-dependence structures. Contrary to many linear statistical forecasting models, stationarity is not required by ANN methods (see e.g. TEIXEIRA JR. *et al*., 2015).

Another important aspect of ANNs is that they are universal approximates of compact (i.e., closed and bounded) support functions, as showed in Khashei and Bijari (2011). In effect, since observations from a time series $y_t(t = 1, \dots, T)$ that exhibit dependency on past values may be seen as points of the domain of an unknown compact support function, it follows that the ANNs are capable of approximating them (for modeling or forecasting) with a high degree of accuracy.

According to Zhang (2003), the predictive power of ANNs comes from the parallel processing of the information exhibited by the data. In addition, AAN models are largely determined by the stochastic characteristics inherent in the time series.

In this context, the feedforward multilayer perceptron ANNs (see e.g. HAYKIN, 2001) are the most widely used neural prediction models for time series forecasting. Particularly, an artificial network composed by three layers (namely, input, hidden an output layers) of simple processing units numerically connected by acyclic links. The relationship between the output y_t and the *L*-lagged inputs, represented by $y_{t-k}\ (k = 1, \dots, L)$, has the following Equation (4):

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j\, g\left(\beta_{0j} + \sum_{i=1}^{L} \beta_{ij}\, y_{t-i}\right) + \varepsilon_t \qquad (4)$$

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

where $\alpha_j (j = 0, 1, \ldots, q)$ and $\beta_{ij} (i = 0, 1, \ldots, L;\ j = 0, 1, \ldots, q)$ are the ANN parameters, called connection weights; L 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 the transfer function, here, a logistic function - although it is possible to adopt other functions (see e.g HAYKIN, 2001). The logistic function is widely employed as the hidden layer transfer function in neural network forecasting and is mathematically defined by Equation (5):

$$g(x_t) = \frac{1}{1+\exp(-x_t)} \tag{5}$$

where $x_{t,j} = \beta_{0j} + \sum_{i=1}^{k} \beta_{ij}\, y_{t-i}$ and $exp(.)$ is the exponential function with Euler's basis (as in Haykin, 2001). Due to $g(.)$ being a non-linear transfer function, the ANN model in (5), in fact performs a non-linear mapping of the past observations $y_{t-k}\ (k = 1, \ldots, L)$ to produce a forecast for y_t.

3. PROPOSED METHODOLOGY

Let $y_t (t = 1, \ldots, T)$ be a time series for which *k* steps-ahead forecasts are required in a forecasting horizon *h*. For this purpose, the WHF proposed here is carried out accordingly by the following six steps.

Step 1: A wavelet decomposition of level *r* of $y_t (t = 1, .., T)$ is performed, producing its *r+1* WCs – namely, one WC of approximation at level m_0, denoted by $A_{m_0,t} (t = 1, \ldots, T)$, and *r* WCs of detail at levels from m_0 to $m_0 + (r-1)$, denoted, respectively, by $D_{m_0,t}$, $D_{m_0+1,t}$, ….,$D_{m_0+(r-1),t} (t = 1, \ldots, T)$.

Step 2: The *r+1* WCs generated in Step 1 are individually modelled through ARIMA model in order to yield the following lists of in-sample and out-of-sample forecasts: $\hat{A}_{m_0,t} (t = t' + 1, \ldots, T + h)$, and $\hat{D}_{m_0,t}$, …,$\hat{D}_{m_0+(r-1),t} (t = t' + 1, \ldots, T + h)$, where *h* represents the forecasting horizon, and t', the value of degrees of freedom lost in the ARIMA modelling. It is important to point out that $\hat{A}_{m_0,t}$ and $\hat{D}_{m,t}$, for

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

each $m = m_0, \ldots, m_0 + (r-1)$, denote, respectively, the prediction (from ARIMA model 1) of $A_{m_0,t}$ and the prediction (from ARIMA model *m+1*) of $y_{D_m,t}$.

Step 3: The forecasts of the *r+1* WCs from Step 2 are linearly combined by means of the LWC, defined in Equation (6).

$$\hat{L}_{C,t} = \hat{A}_{m_0,t} \times \hat{p}_{A_{m_0}} + \sum_{m=m_0}^{m_0+(r-1)} \hat{D}_{m,t} \times \hat{p}_{D_m} \quad (6)$$

Where $\hat{L}_{C,t}$ is the combined linear forecast of y_t; and $\hat{p}_{A_{m_0}}$, $\hat{p}_{D_{m_0}}$, …, $\hat{p}_{D_{m_0+(r-1)}}$ are the optimal LWC parameters, which are obtained by solving the NLP (RAGSDALE, 2004) described below.

Objective: minimize MSE.
Subject to:

$$L_{C,t} = \hat{A}_{m_0,t} \times p_{A_{m_0}} + \sum_{m=m_0}^{m_0+(r-1)} \hat{D}_{m,t} \times p_{D_m}$$

$$MSE = \frac{\sum_{t=t'+1}^{T} (e_t)^2}{(T-t')}.$$

Where, $e_t = y_t - L_{C,t}$ $(t = t'+1, \ldots, T)$ is the list of (in-sample) forecasting errors; and $p_{L_{m_0}}$ and p_{D_m} $(m = m_0, \ldots, m_0 + (r-1))$ are the decision variables (to be optimized and substituted in Equation (6)).

Step 4: The list $e_t(t = t'+1, \ldots, T)$ is decomposed by a wavelet decomposition of level *k*, providing its *k+1* WCs – i.e., one WC of approximation at level m'_0, denoted by $A'_{m'_0,t}(t = t'+1, \ldots, T)$, and *k* WCs of detail at levels from m'_0 to $m'_0 + (k-1)$, denoted, respectively, by $D'_{m'_0,t}, D'_{m'_0+1,t}, \ldots, D'_{m'_0+(k-1),t}(t = t'+1, \ldots, T)$.

Step 5: Each *k+1* WC produced in Step 4 is used as input patterns in the ANN in order to generate the combination of the following list of in-sample and out-of-sample forecasts $\hat{e}_t(t = t''+1, \ldots, T)$, where t'', is the degrees of freedom lost until here.

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

Step 6: The combined linear forecasts $\hat{L}_{Ct}$ (from Step 3) and the combined non-linear forecast $\hat{N}_t$ (from Step 5) are simply summed for each instant *t* (i.e., $\hat{y}_{H,t} = \hat{L}_{C,t} + \hat{N}_t$), producing the in-sample and out-of-sample hybrid forecasts, denoted by $\hat{y}_{H,t}\ (t = t'' + 1, \ldots, T + h)$, of the underlying time series $y_t(t = 1, \ldots, T)$.

Following the above seven steps of the WHF, it is worth noting some important aspects. Firstly, Steps 1 and 4 are aimed to obtain a finite number of temporal subseries (WCs) with better stochastic pattern than the underlying time series. It is possible due to the fact each WC has a stationary spectral frequency, as noted in Mallat (2009); while $y_t(t = 1, \ldots, T)$ can be seen as the result of the sum of spectral components with different frequencies, as pointed by Levan and Kubrusly (2003).

Secondly, in Step 2, the goal is to identify a plausible linear auto-dependence structure (i.e., a linear forecaster), the ARIMA models were chosen here, as they are widely adopted for this purpose, as highlighted by Hamilton (1994). Thirdly, in Step 3, after identifying a plausible linear structure for each WC, their predictions are linearly combined by means of the WLC and then generated the combined linear forecasts, which decode linear information exhibited by $y_t(t = 1, \ldots, T)$.

Consequently, the forecast $\hat{L}_{Ct}$ is in fact a forecast of L_t and allows for being interpreted as an aggregator of distinct linear information from different linear sources (i.e., the *r+1* different linear ARIMA models).

Fourthly, in Step 4, each WC produced here is endowed with noise (like a white noise) and non-linear auto-dependence structure of $y_t(t = 1, \ldots, T)$, as the forecasting errors $e_t(t = t' + 1, \ldots, T)$ are, following (HAMILTON, 1994), outcomes from a linear filter (in this case, the WLC of ARIMA models, in Step 3).

The WLC of ARIMA models is not able to map properly non-linear auto-dependence structures. In effect, it is easily noticed that each WCs of the residuals needed to be modelled by a non-linear forecaster. Unlikely, important information could be lost in the predictive process. Thus, in Step 5, each WC was combined by a

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

(non-linear) ANN model. The software library used here allows for the parameters may be set by the decision maker.

Fifthly, in Step 6, the forecast produced in Step 5 generate the list $\hat{N}_{C,t}(t = t'' + 1, \dots, T)$ of in-sample and out-of-sample combined forecasts that provide information on the non-linear auto-dependence structure of $y_t(t = 1, \dots, T)$. Similarly to the combined linear forecast $\hat{L}_{C,t}$, $\hat{N}_{C,t}$ can be interpreted as an aggregator of non-linear information from different nonlinear sources (namely, the r'+1 ANN models). Sixthly, the hybrid forecasts $\hat{y}_{h,t}(t = t'' + 1, \dots, T + h)$, endowed with both linear and non-linear information, can be seen as a version of $y_t(t = t'' + 1, \dots, T)$ filtered by both a linear filter (namely, the WLC of ARIMA models, in Step 3) and a non-linear filter (i.e., the WLC of ANNs, in Step 5). Accordingly, the list of forecasting errors $e_{H,t}(t = t'' + 1, \dots, T)$, note that $e_{H,t} := y_t - \hat{y}_{h,t}$, can be in fact classified as a noise process.

4. NUMERICAL RESULTS

For evaluating the proposed hybrid methodology, an annual time series of sunspot concerning the period from 1700 to 1987 (resulting 288 points), plotted in Figure 1, the number of lynx caught per year in the Mackenzie Lake in Canada's northern district in period 1821 to 1934 (114 observations) in Figure 2, the weekly British pound/US dollar exchange rate from the year 1980 to 1993 (731 points), in Figure 3, have been projected here. As pointed out in Zhang (2003), prediction has practical importance for decision makers of several areas, as geophysicists, environment scientists, and climatologists.

In this paper, a comparison between the actual values and their forecasts is performed accordingly for a forecasting horizon out-of-sample, which is used for all approaches considered here. The E-Views 8 software was used for the ARIMA modelling; while the MLP ANN modelling was performed in MATLAB R2013a software.

In turn, the adjustment of LWC parameters in Steps 3 was done through the solver package of Excel 2013. Finally, the wavelet decompositions were

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

implemented in MATLAB R2013a software. Regarding the accurate performances, the MAE and MSE (see HAMILTON, 1994)), were taken into evaluation here.

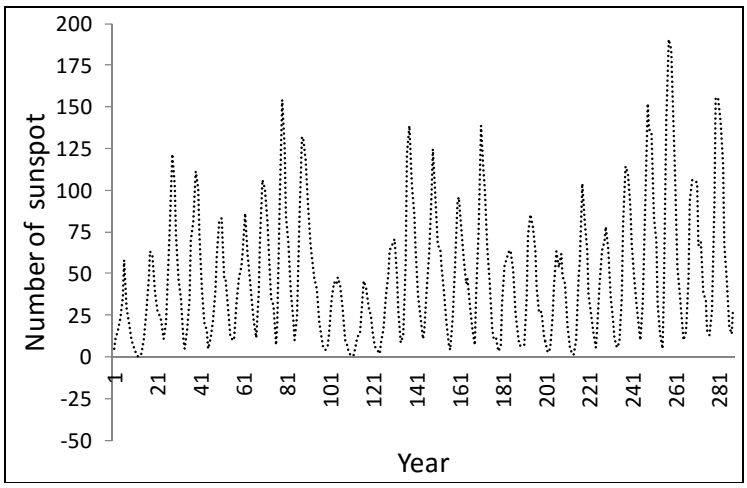


Figure 1: Sunspot time series (1700-1987).

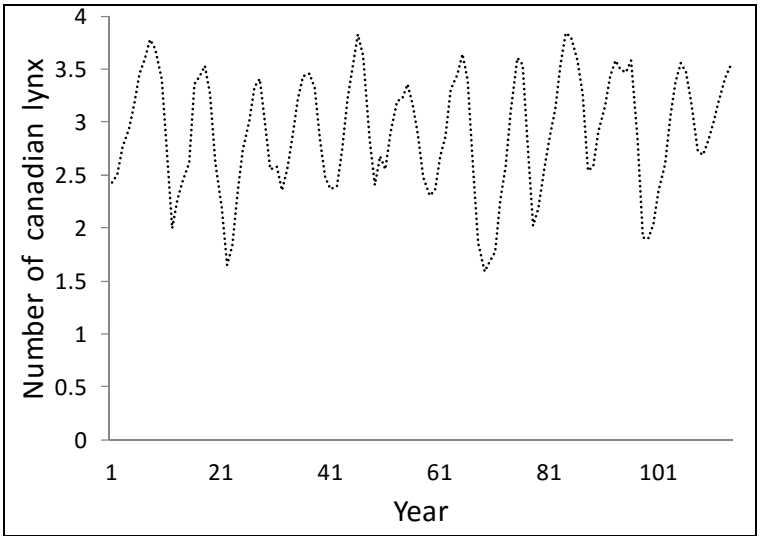


Figure 2: Canada's lynx time series (1821 to 1934).

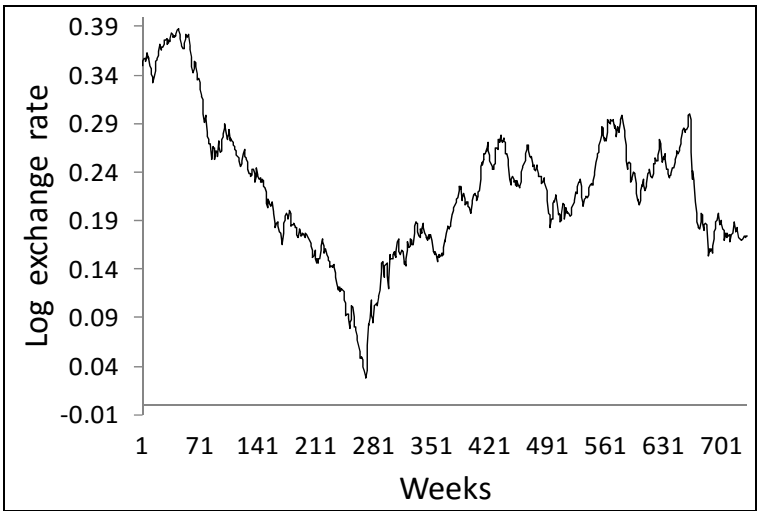
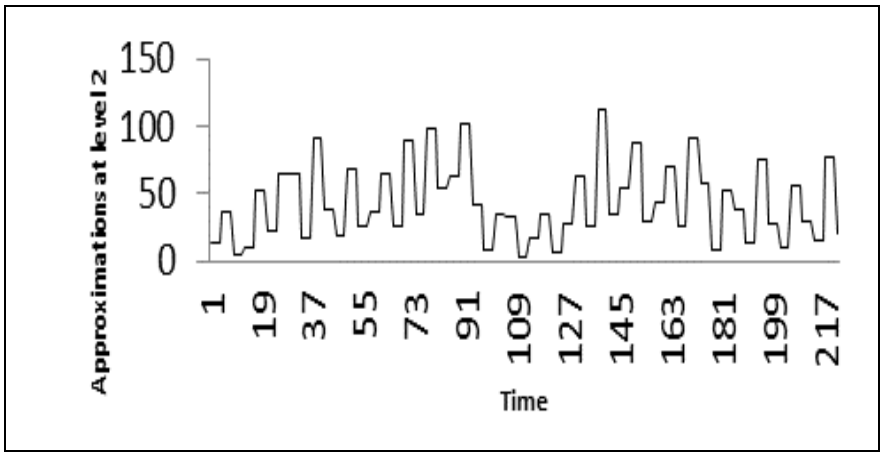


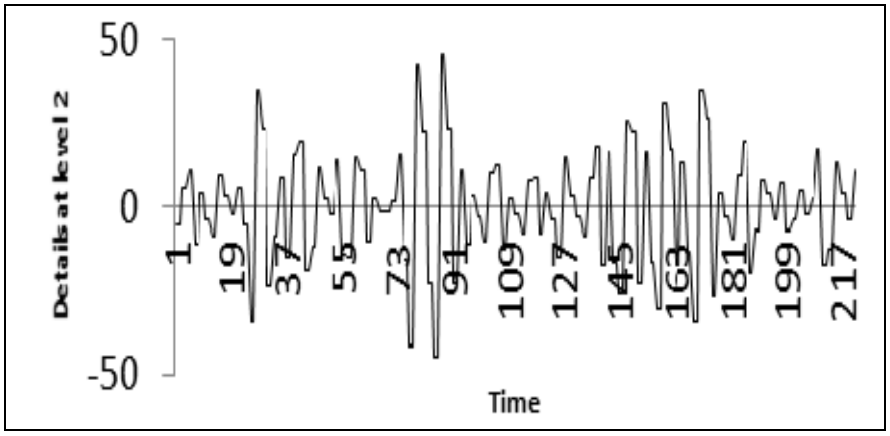
Figure 3: Exchange rate British pound/US dollar (1980 to 1993).

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

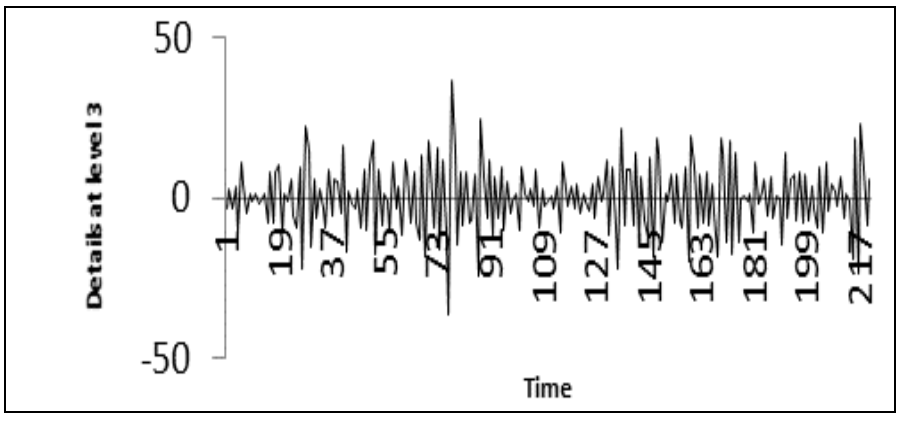
In Step 1, a Wavelet decomposition of level 2 with the Haar orthonormal basis of the training sample of time series of Sunspot was performed. The plots of the WCs can be seen in Figure 4.



(a) WC of approximation at level 2, $A_{2,t}(t = 1, \ldots, 221)$.



(b) WC of detail at level 2, $D_{2,t}(t = 1, \ldots, 221)$.



(c) WC of detail at level 3, $D_{3,t}(t = 1, \ldots, 221)$.

Figure 4: WCs of the training sample of the Sunspot annual time series.

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

Regarding Step 2, it has:

a) An ARIMA(28,1,32) model, with logarithm transformation, for predicting the WC of approximation at level 2. Let $log(A_{2,t}) - log(A_{2,t-1}) = z_t$ be, for each instant *t,* the first difference of $A_{2,t}$ (which is the WC of approximation at level 2 of the state y_t). Thereby, the mentioned predictor is algebraically given by Equation (7):

$$z_t = -0.794332z_{t-4} - 0.719947z_{t-8} + 0.105334z_{t-20} + \\ -0.242021z_{t-28} - 0.339762a_{20} - 0.660078a_{32} \quad (7)$$

b) b) An ARIMA (20,0,0) model for predicting the WC of detail at level 2. Assume that $z'_t = D_{2,t}$ for each instant t. So, the mentioned model is (algebraically) defined by Equation (8):

$$z'_t = -0.802609z'_{t-1} - 0.737875z'_{t-2} - 0.547102z'_{t-3} - 0.771224z'_{t-4} + \\ -0.567598z'_{t-5} - 0.767603z'_{t-6} + -0.551931z'_{t-7} - 0.534644z'_{t-8} + \\ -0.359850z'_{t-9} - 0.128574z'_{t-20} \quad (8)$$

c) c) An ARIMA(20,1,28) model in order to forecast the WC of detail at level 3. Assume that $z''_t = D_{3,t} - D_{3,t-1}$, for each instant t. So, the mentioned model is (algebraically) defined by Equation (9):

$$z''_t = -1.875623z''_{t-2} - 0.876238z''_{t-4} - 0.473194z''_{t-16} - 0.848674z''_{t-18} + \\ -0.369861z''_{t-20} - 2.661716a_{t-4} + 2.723923a_{t-8} - 1.128414a_{t-12} + \\ +0.003598a_{t-28} + 0.074823a_{t-28} \quad (9)$$

The method used here for estimating the complex parameters in items (a), (b) and (c) above was the Likelihood approach, as in Hamilton (1994).

The optimal adaptive parameters from Step 3 are equal to $\hat{p}_{A_2} = 1{,}014978$, $\hat{p}_{D_2} = 0.88357$, $\hat{p}_{D_3} = 0.942313$.

Following, in *Step 4*, the predicting errors $e_t(t = 35, \ldots, 221)$, being $t' = 34$, produced in Step 3, was decomposed by means of a wavelet decomposition of level 2, with orthonormal basis of Haar (see MALLAT, 2009)).

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

The best configurations of the ANN required in Step 5 to model the three WCs of the forecasting errors $e_t(t = 35, \ldots, 221)$, has five data of each decomposition in input layer and seven neurons in hidden layer. Finally, in Step 6, the forecasts produced in Steps 3 and 5 are simply summed, generating the hybrid forecasts (t=$36, \ldots, 288$), where $t'' = 35$, of the Sunspot time series. The plot of the out-of-sample actual values and their hybrid forecasts can be visualized in Figure 5.

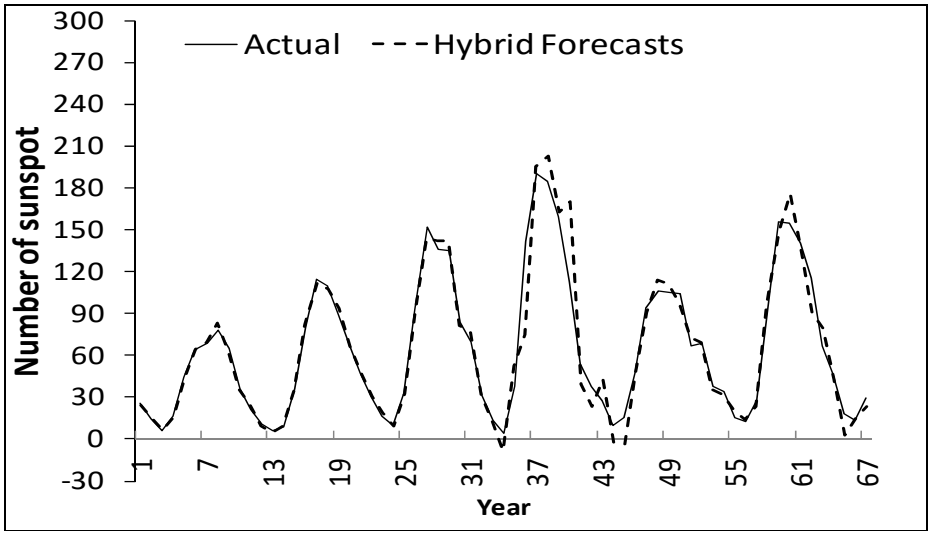


Figure 5: Hybrid forecasts and actual values of out-of-sample.

Table 1 exhibits the out-of-sample MSE and MAE adherence statistics regarding the forecasts of competing predictive methods; and the proposed WHF method (highlighted at the bottom of the Table 1).

Table 1: MAE and MSE of out-of-sample values sunspot time series.

REFERENCES	METHODS	67 steps-ahead	
		MSE	MAE
(ZHANG, 2003)	ARIMA	306.0	13.03
	ANN	351.1	13.54
	Hybrid	280.1	12.78
(KHASHEI; BIJARI, 2011)	ANN/ARIMA	218.6	11.45
(ADHIKARI; AGRAWAL, 2013)	SVM	792.9	-
	Ensemble ANN	280.5	-
	Proposed WHF	**69.28**	**3.90**

Source: The authors.

The models and parameters to the exchange rate time series are given in the Table 2, according to steps of the proposed method.

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

Table 2: Models and parameters to exchange rate methodology.

	Variable	Values
ARIMA	$z_t = A_{2,t}$	$0.998140\, z_{t-1} + 0.382051\, a_{t-4}$
	$z'_t = D_{2,t}$	$-0.663467\, z'_{t-2} - 0.557623\, z'_{t-3} - 1.009415\, a_{t-1} + +0.688248\, a_{t-2} - 0.120479\, a_{t-3} - 0.545579\, a_{t-4}$
	$z''_t = D_{3,t} - D_{3,t-1}$	$-0.797656\, z''_{t-2} - 0.512195\, z''_{t-4} - 0.353988\, z''_{t-6} + +0.114708\, z''_{t-10} - 1.181057\, a_{t-2} + 0.182107\, a_{t-12}$
Adaptative Parameters	$\hat{\phi}_{A_2}$	1.014978
	$\hat{\phi}_{D_2}$	0.88357
	$\hat{\phi}_{D_3}$	0.942313

Source: The authors.

The predicting errors, being produced in Step 3, was decomposed by means of a wavelet decomposition of level 2, with orthonormal basis of Daubechies 45. The ANN input patterns has five observation of each decomposition and the hidden layer has six neurons.

The plot of the out-of-sample actual values and their hybrid forecasts of exchange rate can be visualized in Figure 6.

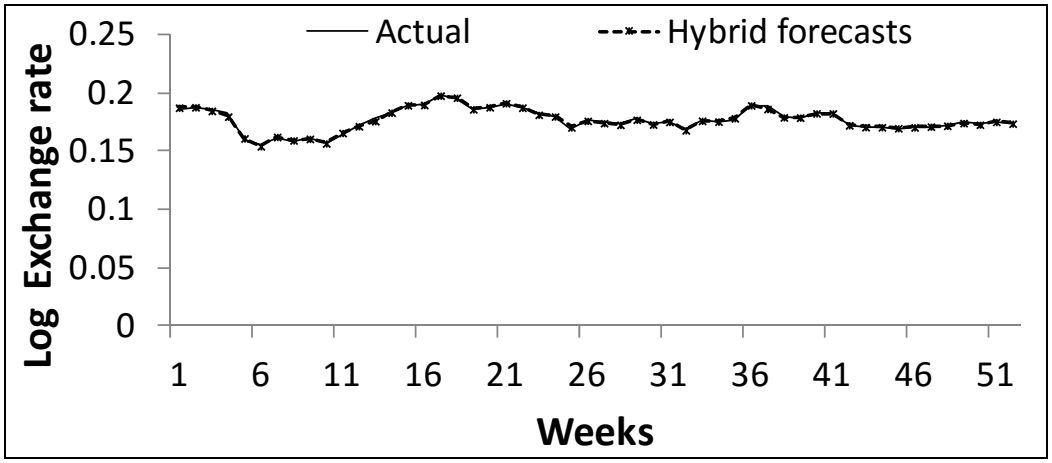


Figure 6: Hybrid forecasts and actual values of out-of-sample.

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

Table 3 exhibits the out-of-sample MSE and MAE adherence statistics regarding the forecasts of competing predictive methods; and the proposed WHF method for the time series of exchange rate.

Table 3: MAE and MSE of out-of-sample values exchange rate time series.

REFERENCES	METHODS	52 steps-ahead	
		MSE	MAE
(ZHANG, 2003)	ARIMA	4.52977×10^{-5}	0.005397
	ANN	4.52657×10^{-5}	0.0052513
	Hybrid	4.35907×10^{-5}	0.0051212
(KHASHEI; BIJARI, 2011)	ANN/ARIMA	3.64774×10^{-5}	0.0049691
	Proposed WHF	**4.23** $\times 10^{-6}$	**0.000907**

Source: The authors.

The necessary information to application of the proposed method for Canadian lynx time series is shown in Table 4.

Table 4: Models and parameters to Canadian lynx methodology.

	Variable	Values
ARIMA	$z_t = A_{2,t}$	$0.996053\, z_{t-1} - 0.593094 z_{t-4} + 0.591410 z_{t-5} + -0.61611\, a_{t-4} + 0.484902\, a_{t-8}$
	$z'_t = D_{2,t}$	$-0.900507\, z'_{t-1} - 0.663425\, z'_{t-2} - 0.584009\, z'_{t-3} + -0.996099 z'_{t-4} - 0.861253 z'_{t-5} - 0.801834 z'_{t-6} + -0.728475 z'_{t-7} - 0.629844 z'_{t-8} - 0.522753 z'_{t-9} + -0.672373 a_{t-2} - 0.283073 a_{t-8}$
	$z''_t = D_{3,t} - D_{3,t-1}$	$-1.578805 z''t - 2 - 2.125563 z''_{t-4} - 2.317076 z''_{t-6} + -1.720207 z''_{t-8} - 1.129782 z''_{t-10} + -0.470963 z''_{t-12} + -0.111867 z''_{t-18} + -0.999949 a_{t-8}$
Adaptative Parameters	β_{A_2}	0.998526
	β_{D_2}	1.001407
	β_{D_3}	0.990531

Source: The authors.

The Canadian lynx error was decomposed by means of a wavelet decomposition of level 2, with Haar orthonormal basis. The ANN input patterns has five observation of each decomposition and the hidden layer has five neurons. The results can be seen in Figure 7 and the adherence statistics in Table 5

ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

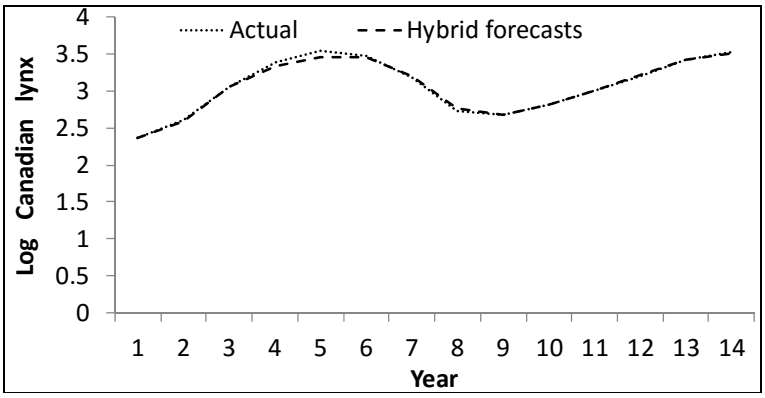


Figure 7: Hybrid forecasts and actual values of out-of-sample.

Table 5: MAE and MSE of out-of-sample values Canadian lynx time series.

REFERENCES	METHODS	14 steps-ahead	
		MSE	MAE
Zhang (2003)	ARIMA	0.020486	0.112255
	ANN	0.020466	0.112109
	Hybrid	0.017233	0.103972
Khashei e Bijari (2011)	ANN/ARIMA	0.00999	0.085055
(ADHIKARI; AGRAWAL, 2013)	SVM	0.05267	-
	Ensemble ANN	0.00715	-
	Proposed WHF	**0.000981**	**0.018828**

Source: The authors.

5. CONCLUSIONS

Comparisons in Table 1, 3 and 5 clearly proved that the proposed WHF method achieved remarkably better results than any of other predictive methods cited in this paper, on out-of-sample performance measures.

According to Figure 5, 6 and 7, the observed values and the predictions produced by the proposed method over the out-of-sample period are strongly correlated, meaning that a high predictive power was achieved in the Sunspot, Exchange rate and Canadian lynx data application.

It is also worth pointing out that despite the relative complexity of the mathematical techniques that integrate the proposed methodology, described in Section 2, its implementation is indeed relatively straightforward with use of appropriate software such as E-Views 8 and MATLAB R2013a software.

REFERENCES

ADHIKARI, R.; AGRAWAL, R. K. (2013) A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting. **International Journal of Computer Applications**, v. 32, n. 7, p. 8..

BATES, J. M.; GRANGER, C. W. J. (1969) **The Combination of ForecastsJournal**

http://www.ijmp.jor.br v. 7, n. 1, January - March 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i1.324

of the Operational Research Society.

DONOHO, D. L.; JOHNSTONE, J. (1994) M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, n. 3, p. 425–455.

FIRMINO, P. R. A.; DE MATTOS NETO, P. S. G.; FERREIRA, T. A. (2015) E. Error modeling approach to improve time series forecasters. **Neurocomputing**, v. 153, p. 242–254.

HAMILTON, J. D. (1994) **Time Series Analysis**,1ed. New Jersey : Princeton University Press.

HAVEN, E.; LIU, X.; SHEN, L. (2012) De-noising option prices with the wavelet method. **European Journal of Operational Research**, v. 222, n. 1, p. 104–112.

HAYKIN, S. S. (2001) **Redes Neurais,** 2ed. Porto Alegre: Bookman.

KHASHEI, M.; BIJARI, M. A (2011) New Hybrid Methodology for Nonlinear Time Series Forecasting. **Modelling and Simulation in Engineering**, v. 2011, p. 1–5.

KUBRUSLY, C. S. (2011) **The Elements of Operator Theory**, 2 ed. New York: Birkhäuser.

KUBRUSLY, C. S.; LEVAN, N. (2006) Abstract wavelet generated by hilbert space shift operators. **Adavances in mathematical Sciences and applications**, v. 16, p. 643–660.

LEVAN, N.; KUBRUSLY, C. S. (2003) A wavelet “time-shift-detail” decomposition. **Mathematics and Computers in Simulation**, v. 63, n. 2, p. 73–78.

LIU, L.-M. (2006) **Time Series Analysis and Forecasting**. second ed. Chicago, IL: Scientific Computing Associates Corporation.

LUTKEPOHL, H.(2006) Forecasting with VARMA Models. In: **Handbook of Economic Forecasting**. [s.l.] Elsevier, v. 1, p. 287–325.

MALLAT, S. (2009) **A Wavelet Tour of Signal Processing: The Sparse Way**, 3 ed. Burlington: Elsevier Inc.

RAGSDALE, C. (2004) **Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science**. Fourth edi ed. [s.l.] South-Western.

TEIXEIRA JR, L. A. et al. (2015) Artificial Neural Network and Wavelet decomposition in the Forecast of Global Horizontal Solar Radiation. **Sobrapo**, v. 35, n. 1, p. 1–16.

ZHANG, G. P. (2003) Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159–175.