International Journal of Engineering and Applied Sciences

www.eaas-journal.org

AN INTERACTIVE WAVELET ARTIFICIAL NEURAL NETWORK IN TIME SERIES PREDICTION

[1] JAIRO MARLON CORRÊA, [2] ANSELMO CHAVES NETO, [3] LUIZ ALBINO TEIXEIRA JÚNIOR, [4]SAMUEL BELLIDO RODRIGUES, [5]EDGAR MANUEL CARREÑO FRANCO

[1,2,4] Graduate Program in Numerical Methods in Engineering Federal University of Paraná, Brazil
[3] Federal University of Latin America Integration, Brazil.
[5] State University of West Paraná, Paraná, Brazil

E-mail: [1]samuelb@utfpr.edu.br, [2]arineicls@gmail.com , [3]luiz.junior@unila.edu.br, [4]alvaro.faria@open.ac.uk, [5]emfra.uinoeste@gmail.com

ABSTRACT

An interactive mathematical methodology for time series prediction that integrates wavelet de-noising and decomposition with an Artificial Neural Network (ANN) method is put forward here. In this methodology, the underlying time series is initially decomposed into trend and noise components by a wavelet de-noising method. Both trend and noise components are then further decomposed by a wavelet decomposition algorithm generating Wavelet Components (WCs) for each one. Each WC is individually modeled by an ANN method in order to produce both in-sample and out-of-sample (point) forecasts. At each time t, the forecasts of the WCs of the trend and noise components are added to provide in-sample and out-of-sample forecasts of the underlying time series. This methodology, applied to the well-known Canadian lynx time series that, as pointed by [1], *exhibit non-linear and non-Gaussian characteristics, is shown to outperform other methods traditionally applied to this data.*

Keywords: ***Time Series (TS), wavelet de-noising (WN), wavelet decomposition (WD), Artificial Neural Network (ANN)***

1. INTRODUCTION

According to the traditional decomposition (as in [2]), a time series y_t ($t=1,\dots,T$) can be expanded, at each time t, as follows: $y_t = \tilde{y}_t + \varepsilon_t$, where $\tilde{y}_t$ and ε_t are the deterministic and independent stochastic components, respectively. From the Wavelet Theory, there are two commonly adopted ways of decomposing y_t ($t=1,\dots,T$); they are referred to as *wavelet decomposition of level r* (as in [3]), and *wavelet de-noising* (as in [4]). In one hand, by means of a wavelet decomposition of level r, the underlying time series y_t ($t=1,\dots,T$) can be separated into $r+1$ WCs - namely, a WC of approximation, denoted by A_t ($t=1,\dots,T$), and r WCs of detail, denoted, respectively, by $D_{1,t},\dots,D_{r,t}$ ($t=1,\dots,T$). Mathematically talking, it means that each state y_t can be orthogonally expanded such as: $y_t = A_t + D_{1,t} + \dots + D_{r,t}$, for all $t=1,\dots,T$. On the other hand, through the *wavelet de-noising* approach, y_t ($t=1,\dots,T$) can be decomposed, at each time t, as follows: $y_t = \tilde{y}_{t,w} + \varepsilon_{t,w}$, where $\tilde{y}_{t,w}$ and $\varepsilon_{t,w}$ are referred to as the deterministic (trend) and stochastic (noise) components of y_t, respectively. Importantly, the sequence $\varepsilon_{t,w}$ ($t=1,\dots,T$) of wavelet noises is usually assumed to be stochastically independent; however, the conventional wavelet de-noising algorithms cannot guarantee this strong statistical property, once they are based on heuristics (and not on statistical tests). In effect, in time series forecasting process, it is absolutely plausible to assume that the sequence $\varepsilon_{t,w}$ ($t=1,\dots,T$) has structures of auto-dependence so that predictive method can be appropriately adopted to produce its forecasts. In this paper, a case study is presented, in which the r' WCs (from a wavelet decomposition of level r') of $\varepsilon_{t,w}$ *($t=1,\dots,T$)* were individually modeled by using an ANN, possessing good forecasting power.

ISSN2305-8269
International Journal of Engineering and Applied Sciences

www.eaas-journal.org

There is a body of literature on time series predicting with a number of mathematical methodologies that integrate a wavelet preprocessing algorithm (commonly, the wavelet decomposition of level r or the wavelet de-noising) with an ANN forecaster (see e.g. [5]). Those methodologies (referred to here as *wavelet ANN* methods) usually adopt one of the two following approaches: (1) performing an initial wavelet decomposition of level r of y_t ($t=1,\dots,T$) that generates $r+1$ WCs, followed by modelling each WC individually with an ANN method to provide forecasts of the WCs that are simply added in order to produce in-sample and out-of-sample forecasts of y_t ($t=1,\dots,T$), which are represented by $\hat{y}_t$ ($t=1,\dots,T+h$), being the predicting horizon of h instants; or (2) applying an initial wavelet de-noising algorithm to y_t ($t=1,\dots,T$) in order to obtain the de-noised time subseries $\tilde{y}_{t,w}$ ($t=1,\dots,T$), followed by modelling $\tilde{y}_{t,w}$ $(t = 1,\dots,T)$ by employing an ANN method (with the de-noised noises $\varepsilon_{t,w}$ ($t=1,\dots,T$) being removed), producing $\hat{y}_t$ ($t=1,\dots,T+h$). In [3] and [4], it may be seen, respectively, that both forecasting methodologies (1) and (2) have achieved remarkable predictive accuracy gains.

Although it is well-known that *wavelet ANN* methods usually outperform traditional methods not based on wavelet preprocessing, those are still useful in determining the best wavelet ANN approach to be adopted. Thus, unlike current wavelet ANN approaches described above, the methodology proposed in this paper integrates, in an interactive way, both wavelet decomposition of level r and wavelet de-noising with an ANN method, as we shall see.

This paper is structured in four sections. Section 1 sets the context of the proposed wavelet ANN method and introduces the notations adopted here. Section 2 describes in detailed the proposed methodology. Section 3 shows the main numerical results of the application to the Canadian lynx data, including a comparative analysis with other ten methodologies existing in time series literature. Finally, Section 4 concludes the paper.

2. PROPOSED METHODOLOGY:

Let y_t ($t=1, \dots, T$) represent a time series for which k steps-ahead prediction are required in a forecasting horizon equals h. The wavelet ANN method proposed here follows the four steps below:

(1) A wavelet de-noising algorithm is applied to y_t *($t=1,\dots,T$)* producing the (wavelet) time subseries $\tilde{y}_{t,w}$ *($t=1,\dots,T$)* and $\varepsilon_{t,w}$ *($t=1,\dots,T$)* (referred to, respectively, as trend and noise components) such that $y_t = \tilde{y}_{t,w} + \varepsilon_{t,w}$, for $t=1,\dots,T$;

(2) A Wavelet Decomposition (WD) of level r of the trend component $\tilde{y}_{t,w}$ *($t=1,\dots,T$)* and a WD of level r' of the noise component $\varepsilon_{t,w}$ *($t=1,\dots,T$)*, where $r = r'$ or $r \neq r'$, are performed to generate $r + 1$ WCs of $\tilde{y}_{t,w}$ *($t=1,\dots,T$)* and $r' + 1$ WCs of $\varepsilon_{t,w}$ *($t=1,\dots,T$)*. Accordingly, $r' + r + 2$ WCs (i.e., wavelet time subseries of y_t *($t=1,\dots,T$)*) are produced here;

(3) The WCs from Step (2) are separately modeled by using a *multilayer perceptron ANN* (see e.g. **[5]**) to provide in-sample and out-sample forecasts. It is convenient to point out that only one optimum ANN is required to model individually all the $r' + r + 2$ WCs generated in Step 2; and

(4) For each time t, the forecasts of the WCs from Step (3) are added in order to produce the in-sample and out-sample forecasts of y_t *($t=1,\dots,T$)*.

The four steps above are illustrated in the diagram in Figure 1.

Note on the top of Figure 1 that a training sample (or in-sample) is chosen from a given “total” time series such that the optimum ANN is determined from this one, and used for generating the in-sample and out-of-sample forecasts.

An interactive computational algorithm is implemented whose aim is to find the optimal parameters of Steps (1), (2) and (3) above concerning the proposed wavelet ANN method. Optimal parameters here refer to the ones associated with the wavelet and ANN approaches that implies in-sample forecasts of the original time series y_t *($t=1,\dots,T$)* with the smallest Mean Squared Error (MSE) Note that, in Step (1), the parameters to be optimized are: the level p of the wavelet decomposition and the Wavelet Orthonormal Basis (WOB) (see e.g. [6]), the thresholding rule, and the threshold (see e.g. [7]). In Step (2), both r and r', as well as the two WOBs involved, are the parameters to be optimized. Finally, in Step (3), the ANN parameters are the preprocessing, the activation function and the number of neurons in the hidden and in the output layers, the window length and the training algorithm (see e.g. [5]).

ISSN2305-8269
International Journal of Engineering and Applied Sciences

www.eaas-journal.org

3. NUMERICAL RESULTS:

In this section the well-known annual time series of Canadian lynx was used to show the effectiveness and the power of the proposed wavelet ANN method. In this experiment, only one-step-ahead predictions (i.e., $k=1$) were considered with a forecasting horizon of 14 time periods (i.e., $h=14$). Those choices were made purely by convenience in accordance with the other methods that were considered for comparison. The underlying time series, shown in Figure 2, consist of the number of lynx trapped per year in the Mackenzie River district of Northern Canada and cover the period from 1821 to 1934 with a total of 114 observations. Note that despite not exhibiting trend the data set shows irregular cyclical behavior.

According to [1], this data set has also been extensively analyzed in time series literature with a focus on non-linear and non-Gaussian modeling. Following the research of other authors, the logarithms (to base 10) of Canadian lynx time series were adopted in all projections and analysis here.

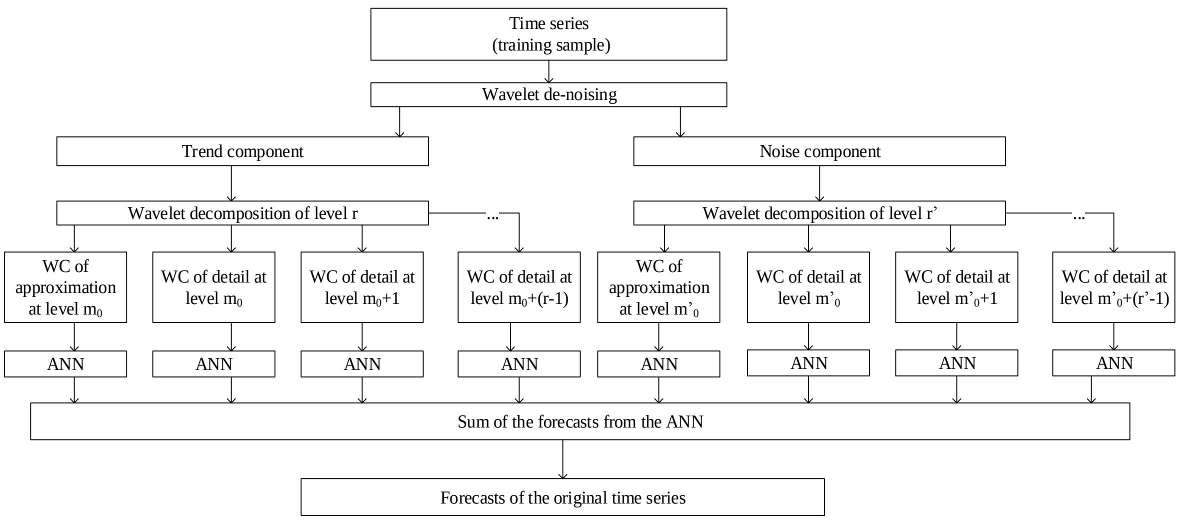

Figure 1 – Flowchart of the four steps of the proposed wavelet ANN methodology.

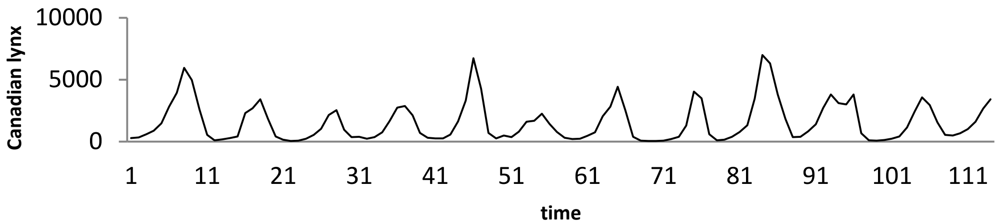

Figure 2 – Total annual time series of Canadian lynx (1821-1934).

For evaluating the predictive performances, the out-of-sample Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE) and the Mean Squared Error (MSE) (see e.g. [1]) were calculated. The time series was split into a training sample of size 100 *(t = 1, ..., 100)* and a testing sample of size 14 *(t=101, ..., 114)*. The training sample was used exclusively to obtain the optimal parameters of the wavelet ANN method described in Section 2; whereas the test sample was only used to evaluate its accuracy.

3.1 Modeling

The four steps of the proposed wavelet ANN method were implemented by an interactive computational algorithm in MATLAB R2013a software. In Step (1), the wavelet decomposition of level *r* took integer values from 1 to 6; regarding the WOBs, the Haar (as in [6]), the Daubechies (as in [8]), the Coifelet and the Symelet (as in [7]) families were tested with the hard and soft thresholding rules (see e.g. [7]) as well as Stein's Unbiased Risk Estimate (SURE) and universal thresholds (see [9], [10], respectively). In turn, in

 ISSN2305-8269
International Journal of Engineering and Applied Sciences

www.eaas-journal.org

Step (2), r and r' also took values from 1 to 6, and the same WOBs tested in Step (1) were used here as parameters to be optimized. Finally, in Step (3), the ANN parameters to be tested were: the premnmx and the score transformations, in the preprocessing stage; the linear and the hyperbolic tangent for the activation function in the hidden and the output layers; the window length took integer values from 1 to 20; and the Levemberg-Marquardt's algorithm was used for training (see e.g. [5] for more details about the ANN parameters). Following all the interactions carried out by MATLAB R2013a software, the best configuration achieved for the proposed wavelet ANN method is detailed below, for each step.

Step 1: Haar's WOB, wavelet decomposed of level 2, soft thresholding rule, and universal threshold;

Step 2: wavelet decomposition of level 2, with the Daubechies's WOB with null moment equals 10, for the trend component; and wavelet decomposition of level 2, with the Daubechies's WOB with null moment equals 12, for the noise component; and

Step 3: for modeling each of the six WCs obtained in Step (2), the optimum ANN holds the following configuration: premnmx preprocessing; window of 12; a hidden layer of 14 artificial neurons with hyperbolic activation function; and an output layer of one artificial neuron with linear activation function.

Table 1 below shows the MSE, MAE and MAPE adherence statistics regarding the out-of-sample forecasts of 11 competing predictive methods. The proposed optimum wavelet ANN method is highlighted at the bottom of the table. The meaning of the acronyms associated with each predictive method can be found in Appendix I.

Figure 3 shows the plots of the actual observed values and the out-of-sample predictions produced by the proposed optimum wavelet ANN method. Observe that the predictive accuracy was so high that it is difficult to distinguish between the two ones.

Table 1 - Comparison for the log-transformed Canadian lynx time series.

Authors	**Predictive Methods**	**h=14**		
		MSE	**MAE**	**MAPE**
Zhang (2003), [1]	ARIMA model	0.020486	0112255	-
	ANN	0.020466	0.112109	-
	hybrid method	0.017233	0.103972	-
Kajitani (2005), [11]	SETAR	0.01400	-	-
Aladag (2009), [12]	Hybrid	0.00900	-	-
Khashei and Bijari(2010), [13]	ANN(p,d,q)	0.01361	0.089625	-
Khashei and Bijari (2011), [14]	ANNs/ARIMA	0.00999	0.085055	-
Zheng and Zhong (2011), [15]	BS-RBF	0.002809	-	1.42%
	BS-RBFAR	**0.002199**	-	**1.18%**
Khashei and Bijari (2012), [16]	ARIMA/PNN model	0.01146	0.084381	-
	ANN/PNN model	0.01487	0.079628	-
Karnaboopathy and Venkatesan (2012), [17]	FRAR	0.00455	-	-
Adhikari and Agrawal (2013), [18]	ARIMA	0.01285	-	3.28%
	SVR	0.05267	-	5.81%
	Ensamble	0.00715	-	2.07 %
Ismail and Shabri (2014), [19]	SVR	0.0085	0.07460	-
	LSSVR	0.00300	**0.04180**	-
authors	**Proposed method**	**0.00017**	**0.010396**	**0.36%**

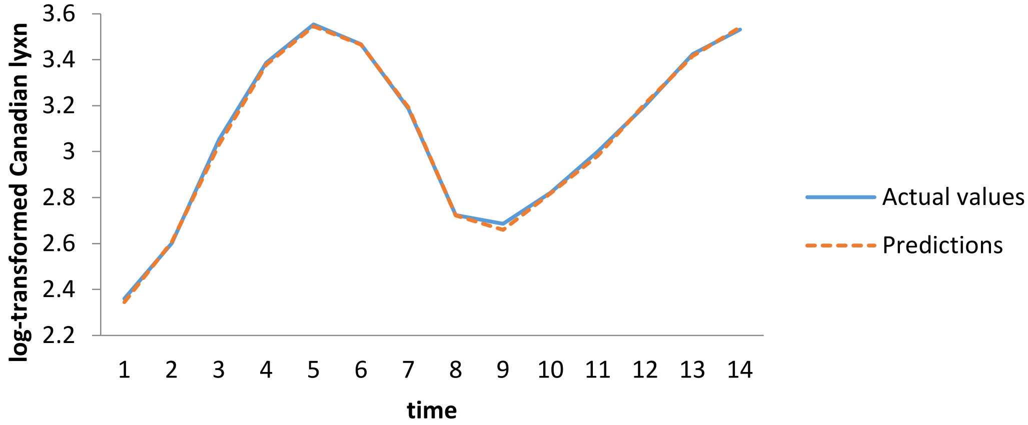

Figure 3- Out-of-sample actual values and predictions of the proposed wavelet ANN method.

4. CONCLUSIONS

It can be seen, in Table 1, that the proposed wavelet ANN method achieved remarkably better results than any of the ten other predictive methods on all three out-of-sample performance measures. In fact, the proposed wavelet ANN method outperformed the second best method, the BS-RBFAR of [15], by 92.27% in terms of the MSE and by 69.50% regarding the MAPE statistic. It also outperformed the LSSVR of [19] by 75.14% on the MAE measure. In addition, it is clear from Figure 3 that the observed values and the predictions produced by the proposed method over the out-of-sample period are strongly correlated, implying that a high predictive power was achieved in the 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 MATLAB R2013a software.

Acknowledgements: The authors would like to express gratitude for the support provided by the Coordination for the Improvement of Higher Education Personnel (CAPES) of the Brazilian Education Ministry.

REFERENCES

[1] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing*, vol. 50, p. 159–175, jan. 2003.

[2] J. D. Hamilton, *Time Series Analysis*, vol. 39, n^{o} 1. Princeton University Press, 1994.

[3] D. L. Donoho e J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, vol. 81, n^{o} 3, p. 425–455, 1994.

[4] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, e D. Picard, "Universal Near Minimaxity of Wavelet Shrinkage", 1995.

[5] S. S. Haykin, *Redes Neurais - 2ed.*, 2^{o} ed. Porto Alegre: Bookman, 2001.

[6] C. S. Kubrusly e N. Levan, "Abstract wavelets generated by Hilbert space shift operators", *Adavances Math. Appl.*, vol. 16, p. 643–660, 2006.

[7] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Third Edit. 2009.

[8] I. Daubechies, *Ten Lectures on Wavelets*. 1992.

[9] D. Donoho e I. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage", *J. Am. Stat. Assoc.*, vol. 90, n^{o} 432, p. 1200–1224, 1995.

[10] D. L. Donoho e I. M. Johnstone, "Minimax estimation via wavelet shrinkage", *Ann. Stat.*, vol. 26, n^{o} 3, p. 879–921, jun. 1998.

International Journal of Engineering and Applied Sciences

www.eaas-journal.org

[11] Y. Kajitani, K. W. Hipel, e A. I. Mcleod, "Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian lynx data", *J. Forecast.*, vol. 24, n° 2, p. 105–117, mar. 2005.

[12] C. H. Aladag, E. Egrioglu, e C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology", *Appl. Math. Lett.*, vol. 22, n° 9, p. 1467–1470, set. 2009.

[13] M. Khashei e M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting", *Expert Syst. Appl.*, vol. 37, n° 1, p. 479–489, 2010.

[14] M. Khashei e M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting", *Appl. Soft Comput. J.*, vol. 11, n° 2, p. 2664–2675, 2011.

[15] F. Zheng e S. Zhong, "Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing", *World Academy of Science Engineering and Technology*, p. 1125–1129, 2011.

[16] M. Khashei e M. Bijari, "A new class of hybrid models for time series forecasting", *Expert Syst. Appl.*, vol. 39, n° 4, p. 4344–4357, 2012.

[17] R. Karnaboopathy e D. Venkatesan, "Data mining in canadian lynx time series", *J. Reliab. Stat. Stud.*, vol. 5, n° 1, p. 1–06, 2012.

[18] R. Adhikari e R. K. Agrawal, "A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting", *Int. J. Comput. Appl.*, vol. 32, n° 7, p. 8, 2013.

[19] S. Ismail e A. Shabri, "Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data", *J. Teknol.*, vol. 1, p. 11–15, 2014.

Appendix I: The meanings of the acronyms in Table 1.
ARIMA: *Auto-Regressive Integrated with Moving Average*.
ANN: *Artificial Neural Networks.*
SETAR: *Smoothing Exponential Transition Auto-Regressive.*
BS-RBF: *Radial Basis Function (RBF) neural network based on Binomial Smoothing (BS).*
BS-RBFAR: *Radial Basis Function (RBF) neural network and Auto-Regression (AR) model based on Binomial Smoothing (BS) technique.*
PNN: *Probabilistic Neural Network.*
FRAR: *Full Range Auto-regressive Model.*
SVR: *Support Vector Regression.*
LSSVR: *LEAST SQUARE SUPPORT VECTOR MACHINE.*