RESEARCH ARTICLE OPEN ACCESS

A Combination of Wavelet Artificial Neural Networks Integrated with Bootstrap Sampler in Time Series Prediction

Tásia Hickmann*, Liliana M. Gramani**, Eloy Kaviski***, Luiz A. T. Júnior****, Samuel B. Rodrigues*****
*(Department of Mathematics and Statistics, Federal Technological University of Paraná, Brazil)
** (Department of Mathematics, Federal University of Paraná, Brazil)
*** (Department of Hydraulics and Sanitation, Federal University of Paraná, Brazil)
****(Department of Statistics, Federal Latin-American Integration University, Brazil)
*****(Department of Mathematics and Statistics, Federal Technological University of Paraná, Brazil)

ABSTRACT
In this paper, an iterative forecasting methodology for time series prediction that integrates wavelet de-noising and decomposition with an Artificial Neural Network (ANN) and Bootstrap methods is put forward here. Basically, a given time series to be forecasted is initially decomposed into trend and noise (wavelet) components by using a wavelet de-noising algorithm. Both trend and noise components are then further decomposed by means of a wavelet decomposition method producing orthonormal Wavelet Components (WCs) for each one. Each WC is separately modelled through an ANN in order to provide both in-sample and out-of-sample forecasts. At each time t, the respective forecasts of the WCs of the trend and noise components are simply added to produce the in-sample and out-of-sample forecasts of the underlying time series. Finally, out-of-sample predictive densities are empirically simulated by the Bootstrap sampler and the confidence intervals are then yielded, considering some level of credibility. The proposed methodology, when applied to the well-known Canadian lynx data that exhibit non-linearity and non-Gaussian properties, has outperformed other methods traditionally used to forecast it.

***Keywords* - A**rtificial neural networks, bootstrap sample, forecasts, time series, wavelet de-noising, wavelet decomposition.

I. Introduction

Based on [1], an arbitrary time series y_t (t=1,…,T) can be expanded, at each time t, as follows: $y_t = \tilde{y}_t + \varepsilon_t$, wherein $\tilde{y}_t$ and ε_t are the deterministic and the independent stochastic components, respectively. From the theory of Wavelet Analysis, there are two commonly adopted ways of decomposing a given time series. They are usually referred to as wavelet decomposition of level r, proposed by [2], and wavelet de-noising, proposed by [3]. In one hand, by means of a wavelet decomposition of level r, y_t (t=1,…,T) can be separated into r+1 WCs - namely, a WC of approximation A_t (t=1,…,T), and r WCs of detail $D_{1,t}$,…,$D_{r,t}$ (t=1,…,T). Mathematically talking, it follows that $y_t = A_t + D_{1,t} + \ldots + D_{r,t}$, for all t=1,…,T. On the other hand, through wavelet de-noising, y_t (t=1,…,T) can be decomposed, at each time t, as $y_t = \tilde{y}_{t,w} + \varepsilon_{t,w}$, wherein $\tilde{y}_{t,w}$ and $\varepsilon_{t,w}$ consist, respectively, of the deterministic and independent stochastic WCs of the state y_t. It is usual to assume that the collection $\varepsilon_{t,w}$(t=1,…,T) of (wavelet) is independent; however, the conventional wavelet de-noising algorithms cannot guarantee this statistical property, because they are based on heuristics, and not statistical tests. Accordingly, it is absolutely plausible to suppose that the wavelet noises have either a linear or a non-linear structure of auto-dependence (as in [1]) so that a linear or a non-linear time series methodology may be appropriately adopted . In this paper, a case study is presented where $\varepsilon_{t,w}$ (t=1,…,T), as well as its r' WCs (from a wavelet decomposition of level r'), is modeled by ANN methods possessing good forecasting power.

There is a body of literature on time series modelling with a number of forecasting methodologies that integrate a wavelet preprocessing algorithm (commonly, the wavelet decomposition of level r or the wavelet de-noising) with an ANN algorithm (see e.g., [4]). Those methodologies, known generically 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,…,T) that generates r+1 WCs, followed by modelling each WC individually with an ANN method to generate forecasts of the WCs that are simply added to produce forecasts of y_t (t=1,…,T); or (2) applying an initial wavelet de-noising algorithm to y_t (t=1,…,T) to obtain the de-noised series $\tilde{y}_{t,w}$(t=1,…,T), followed by modelling $\tilde{y}_{t,w}$ $(t = 1, \dots, T)$ with an ANN method with the de-noised error $\varepsilon_{t,w}$(t=1,…,T) being removed. In [2] and [3], it may be seen, respectively, that both approaches (1) and (2) achieve remarkable forecasting 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 automatically determining the best wavelet ANN method to be adopted, as well as proposing way of calculating interval predictions.

Note that unlike current wavelet ANN approaches described above, the method proposed in this paper integrates, in an interactive way, both wavelet decomposition and wavelet de-noising with the ANN methods as we shall see. In addition, it generates predictive densities by means a conventional Bootstrap sampler and the confidence intervals then further are produced.

The current paper is divided into four sections. Section 1 sets the context of the proposed methodology and introduces notation adopted in the work. Section 2 describes in detailed the proposed methodology. Section 3 shows the statistical results of the application of the proposed method to the time series of Canadian lynx data including a comparative analysis with other methodologies. Finally, Section 4 closes the paper.

II. Proposed Methodology

Assume that y_t (t=1, …, T) represents a time series for which h steps-ahead point and interval forecasts are required. The method proposed here follows the following five steps: (1) A wavelet de-noising algorithm is applied to y_t (t=1,…,T) producing the wavelet series $\tilde{y}_{t,w}$ (t=1,…,T) and $\varepsilon_{t,w}$ (t=1,…,T) (referred to as trend and noise components, respectively) such that $y_t = \tilde{y}_{t,w} + \varepsilon_{t,w}$ for t=1,…,T; (2) A Wavelet Decomposition (WD) of level $r$of the trend component $\tilde{y}_{t,w}$ (t=1,…,T) and a WD of level r' the noise component $\varepsilon_{t,w}$ (t=1,…,T), where $r = r'$ or $r \neq r'$, are performed to generate $r + 1$ WCs of $\tilde{y}_{t,w}$ (t=1,…,T) and $r' + 1$ WCs of $\varepsilon_{t,w}$ (t=1,…,T). Accordingly, $r' + r + 2$ WCs (time wavelet subseries of y_t (t=1,…,T)) are produced here; (3) each WC from Step (2) is separately modeled by a multilayer perceptron ANN method (see e.g. [5]) to produce h steps-ahead in-sample and out-sample forecasts; (4) for each time t, the forecasts of the WCs from Step (3) are added to provide the in-sample and out-sample forecasts of y_t (t=1,…,T), namely $\hat{y}_t$ (t=1,…,T+h); and (5) for each instant out-of-sample, generate the empirical densities by using the Bootstrap sampler (as in [1]).

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

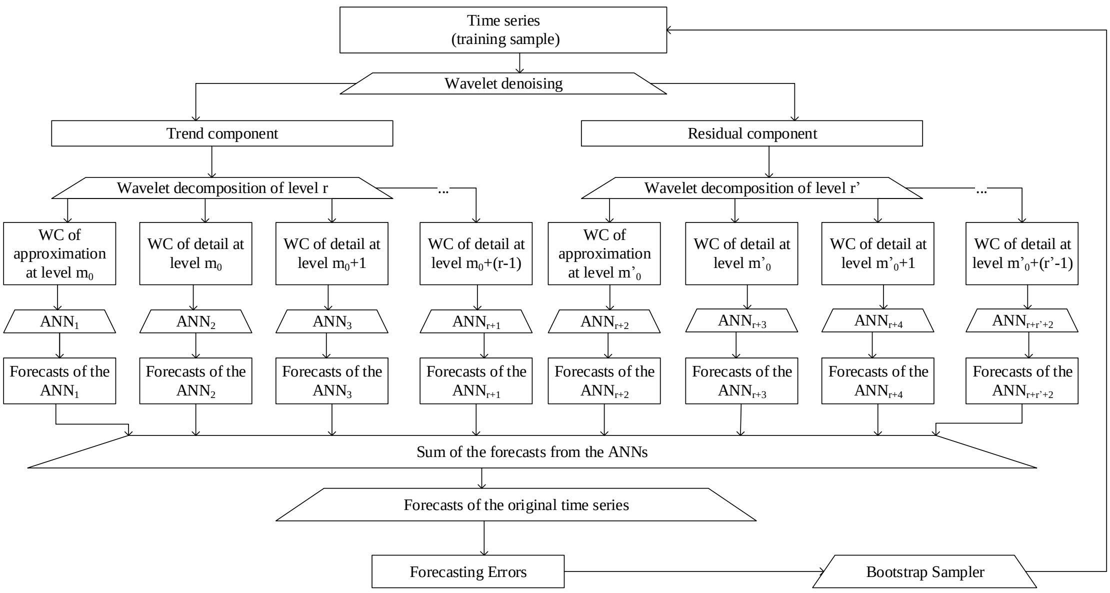

Figure 1 – Flowchart of the four steps of the proposed methodology.

Note from to box on the top of Figure 1 that a training sample (or in-sample) is chosen from the time series such that the model can be determined from that sample and used for testing in the out-of-sample. An interactive computational algorithm is used to determine the optimal parameters of the proposed method. Optimal parameters here refer to the ones associated with the model that produces in-sample forecasts with the smallest mean squared error (MSE) as in [1]. Note that in Step (1), the parameters to be optimized are: (i) the level p of the wavelet decomposition and the wavelet orthonormal basis (WOB) (as in [6]), (ii) the thresholding rule, and (iii) the threshold (see e.g. [7]). In Step (2), both r and r', in addition to 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. [4]).

III. 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 method. In this experiment, only one-step-ahead predictions 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 Fig. 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 shows irregular cyclical behavior unsuitable to be modelled by a linear model.

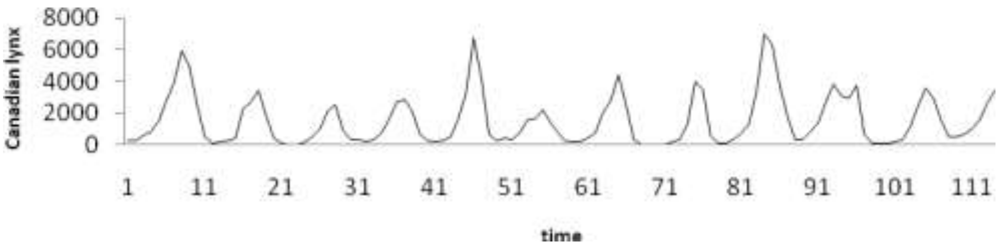

Figure 2–Annual time series of Canadian lynx (1821-1934).

According to [5], 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.

For evaluating its predictive performances, the out-of-sample mean absolute error (MAE), the mean absolute percentage error (MAPE) and the mean squared error (MSE) 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. The five steps of the proposed method were implemented by an interactive computational algorithm in MATLAB R2013a.

3.1Modeling

The training sample of the log-transformed annual Canadian lynx (non-linear) time series is represented by y_t (t=1,…,100). In Step (1) the wavelet decomposition of level r took integer values from 1 to 6. For obtaining the best 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] and [10], respectively). In 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. Finally, in Step (3), the ANN parameters to be tested were the premnmx and the score 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 10; and the Levemberg-Marquardt's algorithm was used for training (as in [4]). Following all the interactions carried out by MATLAB, the best configuration achieved for the proposed method is detailed below.

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

Step 2: wavelet decomposition of level 2, with the Daubechies's WOB with null moment 10 (db10), for the trend component; and wavelet decomposition of level 2, with the Daubechies's WOB with null moment of 12, for the noise component;

Step 3:for modeling the six WCs from Step (2), the best simultaneous configuration is a premnmx preprocessing with a window of 12, a hidden layer of 14 neurons with hyperbolic activation function and an output layer of one neuron with linear activation function.

Table 1 below shows the MSE, the MAE and the MAPE statistics regarding the out-of-sample forecasts of 11 competing predictive methods. The optimum proposed 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.

Table 1- Comparison for the log-transformedCanadian lynx time series.

Authors	PredictiveMethods	h=14		
		MSE	MAE	MAPE
Zhang (2003), [5]	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 andVenkatesan (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**	-
Current	**Proposed method**	**0.00017**	**0.010396**	**0.36%**

Fig. 3 shows the plots of the actual observed values and the out-of-sample point and interval predictions produced by the proposed method. Note that the predictive accuracy was so high that it is difficult to distinguish between the two. Furthermore, there is no violation of the out-of-sample predictive intervals.

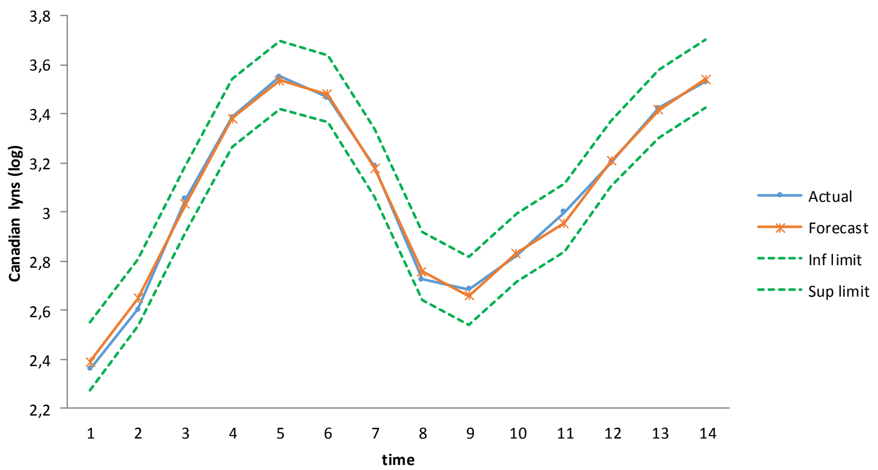

Figure 3- Out-of-sample actual values and predictions of the proposed method.

IV. Conclusions

It can be seen from Table 1 that the proposed method proposed here obtained remarkably better results than any of the ten other predictive methods on all three out-of-sample performance measures. In fact, the proposed 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; in addition, the predictive intervals exhibited efficiency once no real states have violated their inferior and superior limits.

Ultimately, it is also worth pointing out that, despite the relative complexity of the mathematical techniques that integrate the proposed method, its operational implementation is indeed relatively straightforward with use of appropriate software such as MATLAB.

V. Acknowledgements

The authors would like to express gratitude for the support provided by the Itaipu hydroelectric company, the Post-graduate Program in Numerical Methods in Engineering (PPGMNE), the Center for Advanced Studies in Dam Safety (CEASB), the Itaipu Technological Park (PTI) and the Coordination for the Improvement of Higher Education Personnel (CAPES) of the Brazilian Education Ministry.

References

[1] J. D. Hamilton, Time series analysis, *Princeton University Press*, *39(1)*, 1994.

[2] D. L. Donoho and J. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, *8(3)*, 1994, 425–455.

[3] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, Universal Near Minimaxity of Wavelet Shrinkage, *Springer*, 1997, 183-218.

[4] S. S. Haykin, *Redes neurais* (Porto Alegre: Bookman, 2001).

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

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

[7] S. Mallat, *A wavelet tour of signal processing* (The Sparse Way, 2009).

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

[9] D. Donoho and I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. Am. Stat. Assoc.*, *90(432)*, 1995, 1200–1224.

[10] D. L. Donoho and I. M. Johnstone, Minimax estimation via wavelet shrinkage, *Ann. Stat.*, *26(3)*, 1998, 879–921.

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

[12] C. H. Aladag, E. Egrioglu, and C. Kadilar, Forecasting nonlinear time series with a hybrid methodology, *Appl. Math. Lett.*, *22(9)*, 2009, 1467–1470.

[13] M. Khashei and M. Bijari, An artificial neural network (p, d, q) model for time series forecasting, *Expert Syst. Appl.*, *37(1)*, 2010, 479–489.

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

[15] F. Zheng and 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*, 2011, 1125–1129.

[16] M. Khashei and M. Bijari, A new class of hybrid models for time series forecasting, *Expert Syst. Appl.*, 39(4), 2012, 4344–4357.

[17] R. Karnaboopathy and D. Venkatesan, Data mining in canadian lynx time series, *J. Reliab. Stat. Stud*., *5(1)*, 2012, 1–06, 2012.

[18] R. Adhikari and R. K. Agrawal, A homogeneous ensemble of artificial neural networks for time series forecasting, *Int. J. Comput. Appl.*, *32(7)*, 2013.

[19] S. Ismail and A. Shabri, Time series forecasting using least square support vector machine for Canadian Lynx Data, *J. Teknol.*, *1*, 2014, 11–15.

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 BasisFunction (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.*