Financial Time Series Forecasting Using Hybrid Wavelet-Neural Model

Jovana Božić and Djordje Babić
School of Computing, University Union, Serbia

Abstract: *In this paper, we examine and discuss results of financial time series prediction by using a combination of wavelet transform, neural networks and statistical time series analytical techniques. The analyzed hybrid model combines the capabilities of wavelet packet transform and neural networks that can capture hidden but crucial structure attributes embedded in the time series. The input data is decomposed into a wavelet representation using two different resolution levels. For each of the new time series, a neural network is created, trained and used for prediction. In order to create an aggregate forecast, the individual predictions are combined with statistical features extracted from the original input. Additional to the conclusion that the increase in resolution level does not improve the prediction accuracy, the analysis of obtained results indicates that the suggested model presents satisfactory predictor. The results also serve as an indication that denoising process generates more accurate results when applied.*

Keywords: *Time-series forecasting, wavelet packet transform, neural networks.*

Received November 23, 2014; accepted January 20, 2016

1. Introduction

The financial time series are inherently a non-stationary, noisy and chaotic [23]. They are a combination of long and short memory processes imbedded in one complex signal, explaining why their prediction can present a true challenge [23], especially knowing that due to its congenital complexity, traditional statistical methods perform poorly in this field. All of this actually suggests that there is no complete information base on which we can perform successful forecast. Moreover, the general assumption made in these cases is that the historical data of one time series integrates all important features necessary for successful prediction. Despite this complicated scenario, our goal in this paper is to investigate the use of one specific wavelet transform and Artificial Neural Networks (ANNs) for the prediction of financial time series.

Over the last few decades, it has become obvious that linear models do not adequately represent nonlinear series, while wavelet analysis theory has emerged as a powerful tool in the mathematical analysis field [10]. Simply said, the wavelet transform produces a functional local decomposition of a signal in both the time and frequency domains and is not restrained by the assumption of stationarity [10]. Numerous publications describe the application of wavelets in the field of finance [7, 19]. The transform we use in this paper is Wavelet Packet Transform (WPT). Each wavelet transform offers the capability of capturing key features of an underlying process with a limited number of coefficients, but this choice is driven by the fact that with the WPT the most complex and detailed signal analysis is obtained. The fundamental and novel contribution of this paper is to use one particular processing technique to decompose financial time series into a set of approximation and detail series which are fed into the neural networks in the model's next phase. The traditional approaches to time series prediction, such as Box-Jenkins or ARIMA method, assume that the time series used for the prediction process are linear and stationary [1, 4]. Therefore, these methods are obviously not a good tool for financial time series prediction. On the other hand, during the past few decades the ANNs have shown great applicability in time series prediction [21, 22]. Studies have compared the performance of neural networks to Autoregressive Integrated Moving Average ARIMA [14], with all research agreeing that ANNs perform better than ARIMA models. Several unique features of ANNs make them an attractive forecasting tool: they are multivariate, nonparametric statistical methods that can map any nonlinear function without a priori assumption about the data, yet maintain desired accuracy [22]. Numerous articles illustrate the practical considerations of ANNs' applicability [1, 8, 18]. By combining wavelet transform with ANNs we get a new kind of modelling method with great prediction ability for high frequency financial data. With this synergy, we gain advantages from both of the methods-the multiscale analysis supplied by wavelet theory and powerful learning and training capability of the neural network. Amongst many studies that investigate the concept of mixing wavelets and neural networks, we reference several in our research [3, 5, 15, 16].

In this paper, we use a time series prediction using hybrid wavelet-neural model. In order to expose the complex underlying structures for deeper evaluation, the time-series is first subjected to a wavelet-based decomposition process using decomposition levels of two and three. The decomposed signal components are then used as input elements to a new group of neural networks where valuable information is captured during the process of the training phase. In the third stage of the model, the newly obtained time series along with the statistical features extracted from the original input are directed into the final neural network where the prediction is made. Comparison of the prediction results of the two models are based on four evaluative parameters: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Real Mean Squared Error (RMSE). The results show that the hybrid model did not perfectly fulfil the prediction tasks, meaning there is a certain mismatch between original and predicted values. However, having in mind the great number of parameters affecting financial series trends and the large number of scenarios taking place on a daily basis in global stock markets, this mismatch is quite reasonable. Strategy applied shows that financial time series follow a trend that is not completely random, and therefore the presented topology is capable in grasping complex interactions and chaotic components in financial time series. Also, the increase in resolution level, perhaps unexpectedly, brings no improved results.

This paper is organized as follows. The next section presents a review of theory behind wavelets and neural networks. The third section covers all aspects of the introduced hybrid model, including a detailed description of the procedure for its design. In the fourth section, the experimental results are presented to demonstrate the effectiveness of the presented hybrid strategy. The fifth section presents analysis of the results. The final section proposes conclusions and recommendations for future research.

2. The use of Wavelets and Neural Networks

This Section explains the value of wavelets in time-series prediction together with neural networks.

2.1. The Value of Wavelet Transform

One of the drawbacks of the Fourier analysis is that although it is possible to determine the frequencies present in the signal with this analysis, it is not possible to establish when they actually occur [17]. This premise is surpassed with the introduction of wavelet transform, representation of the signal whose root is comprised of wavelets (or mother wavelets or analyzing wavelets). Wavelet transform, through the mechanism of mother wavelets translation, offers precise information about time and frequency resolution [10].

Discrete Wavelet Transform (DWT) is one of the most convenient tools for signal analysis [10]. With its help, we can present the signal with a limited number of coefficients that capture information at different frequencies at distinct time moments. One of the variations of the DWT is WPT, where both approximation and detail coefficients are decomposed (2^n sets of coefficients are produced, unlike $n+1$ in the case of a DWT. With this kind of signal presentation, the most complex and detail signal presentation is gained. Figure 1 shows level 2 wavelet packet decomposition.

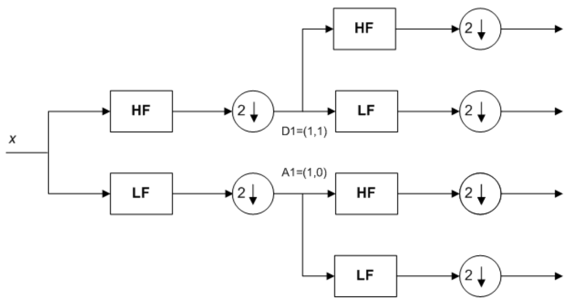


Figure 1. Wavelet packet decomposition of level two.

As for the wavelet used in previous studies, most signal processing researchers adopted Daubechies wavelets [20] and one of the most commonly used in time series analysis is Db40.With previously specified signal processing technique, the input time series can be analyzed at multiple time resolutions; the signal can be smoothened until the long-term trend is identified and the fluctuations around the trend can be investigated at multiple time scales. After the decomposition, the individual time series can give a detailed and more easily analyzed view of the inner underlying processes.

2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) are a class of nonlinear models that can extract crucial parameters from complex high-dimensional time series and approximate any nonlinear function with a high level of accuracy as a result [1, 9]. They are capable of discovering the underlying pattern or auto-correlational structure in the time series even when the underlying law is unknown or hard to determine, making them a powerful forecasting tool in many different fields. Despite the fact that neural networks have been successfully implemented in prediction process on numerous occasions, designing a predictor for specific financial time series with neural network is a challenging and nontrivial task. In comparison with Box Jenkins, ARIMA models and other regressive models, a larger number of factors play a role in the neural networks design.

One of the most popular and most successfully implemented neural network model is the feed forward multilayer network or Multi-Layer Perceptron (MLP) [24]. This type of network consists of several layers that contain nodes (artificial neurons). The first layer is the input level and receives external information, while the last layer is the output level and produces model solutions. Hidden layers lie in between these two. All of the nodes in one layer are connected to the nodes in the adjacent layer by an interconnection strength called weights. These weights are set through a training algorithm, where the goal is to minimize the difference between the network`s target and actual output. Figure 2 shows MLP network with one hidden layer.

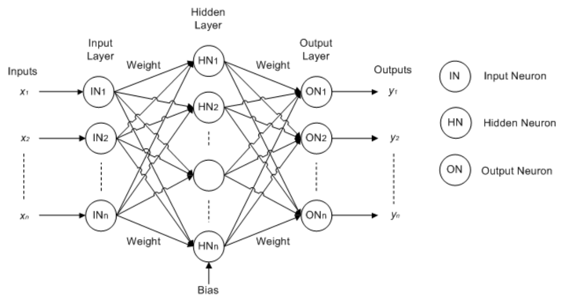


Figure 2. Neural network with one hidden layer.

The design of neural network is a difficult task concerning the number of factors that influence its performance. Although some studies [12] propose certain methods for neural network design, no study has been reported to analytically determine general architecture rules for successful neural network design. One of the most sensitive parameters are number of layers, number of neurons in each layer, activation function (function that produces an output based on input values entering the node) and learning algorithm (the way the weights are set). The number of input and output layers depends on the problem’s nature (in most papers, the suggested value is one). As for the hidden layers (internal information processing layers), it has been pointed out that one hidden layer network is able to approximate most of the nonlinear functions [11]. The dimension of each layer, rather the number of neurons in each layer, is one of the most essential parameters for the network’s proficiency and successful performance [12]. Increased training time and reduced generalization ability of the NN can be a result of too many hidden nodes. On the other hand, if it is too few, the network’s ability to learn will be reduced. In most cases, because of the lack of a systematic approach to neural network design and established guidelines for it, trial and error approaches are suggested for most previously stated architecture parameters.

To summarize, neural networks are difficult to design, require high processing and training time, and give unstable results in many situations. However, if their architecture is correctly planned (which demands a significant level of invested time and resources), they present a powerful tool that can perform many demanding tasks that linear programs cannot and can therefore be successfully implemented in many applications.

2.3. Statistical Feature Extraction

Besides basic wavelet features, we implemented an additional seven statistical features, all of them commonly used in finance and economics, into our model in order to improve the prediction process. They are given in the Table 1, where *x* presents the time series and value of current observation, *n* presents the total number of observations, σ the standard deviation, *t* represents the moment of time between 1 and *n*, and p_1 to p_n are probabilities of the signal.

Table 1. Statistical features and their definitions.

Statistical feature	Definition
Mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Mean Absolute Deviation	$MAD(x_1..x_n) = \frac{1}{n}\sum_{i=1}^{n}\left\|x_i - \bar{x}\right\|$
Variance	$VAR(x_1..x_n) = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$
Skewness	$SKEW(x_1..x_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^3$
Kurtosis	$KURT(x_i - x_n) = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i - \bar{x}}{\sigma}\right]^4\right\} - 3$
Turning Points	$(x_{t+1} - x_t)(x_t - x_{t-1}) < 0$
Shannon entropy	$H(x) = -\sum_{i=1}^{n} p_i \log_2 p_i$

3. Hybrid Modelling Strategy

The main concept behind the prediction method presented in this paper is to decompose the financial time series, using one particular wavelet transform into a range of frequency scales and to pull these individual components through separate neural networks, making an aggregation forecast in the final neural network. The entire process of this algorithm involves a series of steps, including: statistical feature extraction, pre-processing step, wavelet analysis, neural networks training and modelling, and final forecasting. We approach this issue by dividing the model into three separate stages:

- An exchange rate is processed and subjected to a wavelet-based decomposition process in order to detect underlying processes (features) for further evaluation.
- All individual decomposed components are fed into a set of neural networks in order to capture valuable information.
- The outputs from the neural network, rather the predicted values of each component, along with a set of statistical features (calculated on the original

time series) are fed into the final neural network after which the prediction is to be made.

The data preparation phase includes the statistical parameters calculation and normalization of the financial time series. Each statistical feature for a specific sample is calculated based on 10 previous samples. The normalization process has to be done in order to avoid the effect of outsized values on the model and to fasten the calculation. It is carried out in such manner that both the inputs and targets fall in the range [-1, 1].

The input data are decomposed into a certain number of sub-time series components with the help of wavelet decomposition. Our main considerations regarding which wavelet transform and mother wavelet we use are:

- The input time series is discretely sampled so there is a dyadic relationship between resolution scales, leading to the use of a DWT, rather its variation of a WPT.
- Wavelet function which respects the complex nature of the financial time series leads to the use of Daubechies wavelet function.

The choice of optimal decomposition level, generally depending on the researcher's experience and time series nature, is one of the most important factors of the model's performance in the first stage. In this study, we use two and three decomposition levels (poor results are obtained when using decomposition levels greater than three). Figure 3 shows a representation of the model. For simplicity, the model has been considered only for a decomposition level of two.

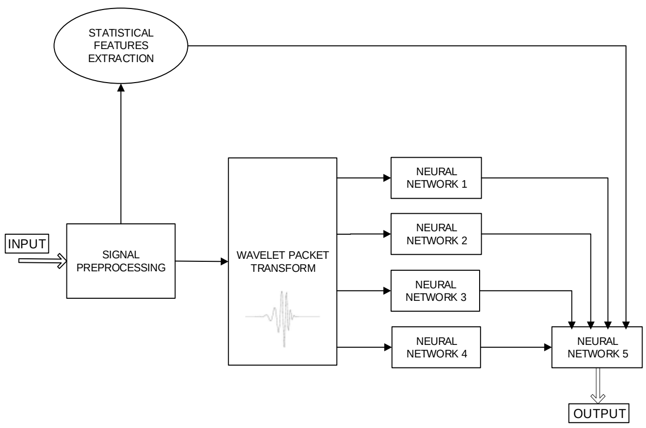


Figure 3. Hybrid model with Wavelet packet transform; the 2nd level of resolution.

The concept of using wavelet transform in time series analysis offers the advantage of separating the smooth part (approximation series) and the irregular and noisy (detail series) part of the signal, which are both more stable to handle and easier to predict due to the filtering effect of the transform. As a result of this stage, the objective is to exploit these series as input signals to a set of neural networks in the following stage.

The second stage of the model consists of a set of feed forward neural networks that uses lagged detail and smooth coefficients gained from decomposition in the previous step as input. One of the papers where a similar idea is presented is [7]. Separate neural networks are built for each decomposition level, meaning that a 2-level wavelet packet decomposition results in four, while a 3-level wavelet packet decomposition results in eight neural networks.

What can be noticed here is that we have a small amount of control over the complexity of the architecture in this stage. One possible way of handling this complexity is to test networks with different designs and compare them in order to choose the optimal one. Having this in mind, the problem can be tackled by varying a large number of design factors that influence the prediction result. This is the reason why we vary the number of input nodes and why we design the hidden layer as simply as possible. Not only does our research confirm that this is the best approach, but our findings are also supported by literature which states that the simplest model is the least likely to over fit/under fit and the most likely to generalize well on the unseen data [13].

All series are split into two sets: training and testing. There are no specific rules for data division between these sets, and most researchers use a trial and error approach. For the training phase, we use a method known as the „sliding window" technique where the *n*-tuple input goes through the entire training set while a single output is used as the target value. Once trained, the networks are used for prediction.

As for the transfer functions, we use a linear one for the node in the output layer and a tan sigmoid function for the nodes in the hidden layer. This function is the most commonly used function in forecasting problems and pattern detection because it outperforms the alternatives when deviations are calculated from the average behaviour [12]. All of the networks are trained using the Scaled Conjugate Gradient algorithm, a supervised learning algorithm that shows linear convergence on the most of the problems [6] and provides faster learning.

In the final stage, the individual predictions along with statistical features that are calculated on the original time series are combined to generate an aggregate forecast. All of those values serve as inputs into the last neural network, where the final output is the one-step-ahead predicted sample. To design the last neural network, we use the same parameters as we did for the networks from the secondary stage of the model, with the only difference being that the number of input samples is fixed. Table 2 summarizes the architecture of the final network in the system, depending on the level of resolution used in the second stage.

Table 2. The architecture details of the neural network from the last stage.

Method	Resolution Level	Number of Inputs	Number of Outputs	NN 1architecture
Wavelet Packet	2	11	1	11:6:1
Wavelet Packet	3	15	1	15:8:1

To conclude, based on a previously introduced model, the goal is to demonstrate that the one-step-ahead predictions for specific financial time series can be estimated with reasonable accuracy. The predictive power of these forecasts is compared by using a set of statistical parameters. In the next section, we show that the model's ability to capture dynamical behaviour differs with the wavelet resolution level, but not in a way that we expect.

4. Results

In this Section we test the proposed forecasting model by using a specific financial time series. The financial time series used here is the official exchange rate of Euro against the domestic exchange rate (republic serbia dinar) between July 2003 and September 2007 (total length of 1024 samples). We use this particular financial time series because we believe that this period of exchange rate is the most suitable for testing because it contains various dynamic changes in the exchange rate. The first 80% of it is used as sample data for the training phase; while the remaining 20% is used for evaluation of each neural network (this kind of division gives best performance results). The graph of the EUR/RSD exchange rate is illustrated in Figure 4. For all tests and simulations, a special MATLAB code has been generated.

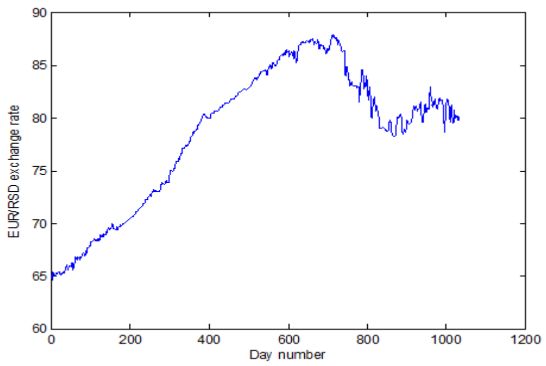


Figure 4. EUR/RSD exchange rate from July 2003 till September 2007.

As for the prediction performance, the hybrid wavelet neural model is evaluated by using four statistical parameters: MAE, MAPE, MSE and RMSE. These parameters are defined in the following manner:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \quad (1)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \hat{y}_t\right|}{y_t} * 100\% \quad (2)$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_y - \hat{y}_t)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_y - \hat{y}_t)^2} \quad (4)$$

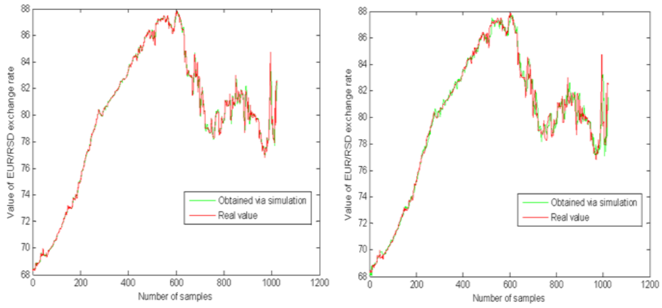
Where y_t is the real value and $\hat{y}_t$ is the predicted value. Although, most of the above expressions are self-explanatory, it is useful to point out the following: MAE measures the deviations between the actual and predicted values, MAPE is the average absolute percentage error, MSE is the average of the squared errors between the predicted and real value, and RMSE presents how good a variance of the estimate is. Obviously, the closer that these values are to zero, the more accurate is the prediction performance.

Following the presented modeling strategy, the EUR/RSD exchange data is preprocessed and decomposed into two different resolution levels by WPT. The goal is to make underlying temporal processes of the original exchange rate more traceable and easier for further analysis. Afterwards, the new set of sub-time series is fed into a set of neural networks.

In this phase, we also tackle the denoising effect on the prediction performance. The goal is to check if the process of noise removal will improve the quality of the overall final forecast. We apply denoising in the first stage with the WPT used. Because the process of noise removal is quite complex, we notice that the model performance is particularly sensitive to the threshold parameter and that it is very important to determine the correct threshold value and apply denoising to the detail coefficients [1, 2]. The process of soft threshold has been tested for various values from 0.01 to 0.06 with step of 0.005, and the best results are obtained for the threshold value of 0.02. After filtration, wavelet packet reconstruction is performed to obtain a de-noised signal that will serve as an input signal in the next phase. It can be noticed that the noise from the original time series is removed without the influence of sudden glitches which means that most of the original signal is preserved. This feature is one of the biggest advantages of the wavelet packet method.

In the next stage, we investigate the approach of forecasting each individual sub-time series in the set of neural networks. All networks are trained separately (by using corresponding wavelet coefficients), and the objective is to perform a one-day-ahead prediction for each time series. Depending on the resolution level used in the first stage, we train 4 or 8 neural networks sown in Figure 3. As for the design of each neural network, we apply the same architecture for each of them. In order to determine the optimal number of input nodes, we tested the ANNs with input layer that consists of 1, 2, 3, and 4 nodes. The standard for determining the optimal number of input nodes is RMSE. As for the internal architecture, we design all NNs with a single-hidden layer. This is done because the NN with a single hidden layer can approximate any function with arbitrary precision and because input layer has fewer nodes. For the number of nodes in the hidden layer, we apply a principle most often used in literature-Okam's razor principl-where the number of hidden nodes equals half of the input and output nodes (we notice that increasing the number of hidden nodes or even adding more layers does not improve networks' performance). Also having in mind that we want to predict a single value, we use one neuron in output layer.

The prediction results of each neural network are individually combined with statistical parameters, calculated on the original EUR/RSD time series, thus establishing the inputs for the final neural network. Figure 5 compares the real time series and the output of the last neural network (the simulated time series). There is a large degree of overlap in input and output values for both resolution levels, with no major discrepancies.



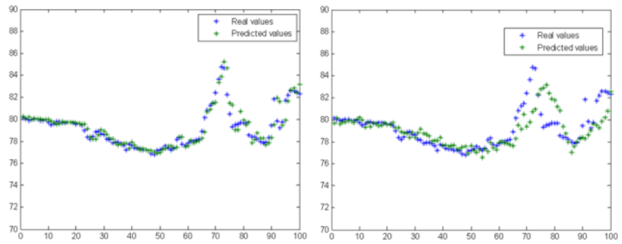
a) Wavelet packet transform used, 2 resolution levels.
b) Wavelet packet transform used, 3 resolution levels.

Figure 5. Visual comparison of real and simulated signal.

The forecasting analysis is performed based upon the results from the one-step-ahead prediction of the presented hybrid model. We measure the performance metrics to investigate how well the model captures the underlying trend of the movement of the EUR/RSD exchange rate. Table 3 shows the performance metrics achieved by our model. We illustrate this performance in Figure 6 where the prediction of 100 samples for both resolution levels is shown.

Table 3. Performance metrics for EUR/RSD exchange rate depending on parameters used in model's stages.

Wavelet Transform	Level of Resolution	Wavelet	Number of NNs in the Second Stage	Number of Inputs in the NN in the 3rd Stage	Number of Samples Predicted	MAE	MAPE	MSE	RMSE
WPT	2	Db40	4	11	100	0.0049	0.00006	0.0045	0.00756
WPT	3	Db40	8	15	100	0.006	0.00008	0.00746	0.00974



a) Wavelet packet transform used, 2 resolution levels.
b) Wavelet packet transform used, 3 resolution levels.

Figure 6. Comparative view of 100 real and predicted samples.

Based on the results, the learning algorithm applied on coefficients handles the underlying structures in a satisfying manner so it can be concluded that the WPT decomposes the signal in an overall accurate and precise way. When it comes to denoising applied, the results indicate that both models are running well, with slightly more accurate and stable results in the case where noise is removed.

Additionally, we observe that the ability of the model to capture dynamical behavior is changing with the resolution level used. Although one would think that for lower resolution levels, resulting in noise and irregular sub-time series, the model would show less accurate results, we actually get less accurate results with the higher resolution time series (in other words, when we use the more smooth series). This phenomenon can be summarized in the following way-the performance of the model deteriorates with an increase in the resolution level, with level 2 being the optimal decomposition. An increase of the resolution level greater than 3 yields poor results and a meaningless prediction as a result.

Evident from the preceding is that we here analyze an exchange rate with frequent, high jumps and peaks, which can corrupt the prediction process to a large degree. The model tests the EUR/RSD exchange rate over a very particular time period, during which the Republic of Serbia experienced a difficult financial crisis in correlation with a generally volatile global economy. Due to this volatility, we believe that the trade market itself is not valid enough and that the historic data cannot depict all of the information required. This explains why we have boosted the model's accuracy with the introduction of statistic features in the final stage. Having all of the preceding in mind, the model largely manages to predict the one day value of the EUR/RSD exchange rate.

5. Conclusions and Further Research

In this paper, we have analyzed prediction strategy combining wavelet transform, neural networks and statistical features for financial time series prediction. According to our findings, the model presents promising candidates for such prediction. Moreover, it seems that a WPT has a large capacity to capture global behaviour in a financial time series and thus offers rich information that is used in the second stage for the purpose of training, modelling and forecasting.

Another conclusion we have drawn is opposite to our expectations-increasing the resolution does not improve the system performance, indicating that a prediction may not necessarily be more accurate if the signal decomposition level is higher.

In the past years, numerous hybrid models have been investigated, and they all show significant promise. Future research could possibly study the predictive power for long term forecasts of the same model or the utilization of other outside imported economic indicators. Also having in mind that the predictive power of wavelet neural hybrid model is highly sensitive to a large number of parameters, we had to do tedious experiments and trial-and-error procedures in order to obtain valid results. These major weaknesses can perhaps be avoided by changing the selection of appropriate number of hidden nodes, training times and lags, and determining and setting systematic rules for these tricky tasks.

Acknowledgements

This paper is partly a result of the technology development project funded by the Ministry of Education and Science of the Republic of Serbia entitled "Performance optimization of energy-efficient computing and communication systems (TR 32023)".

References

[1] Al-wazan H., Ibraheem K., and Salim A., “A Hybrid Algorithm to Forecast Enrolment Based on Genetic Algorithms and Fuzzy Time Series,” *The International Arab Journal of Information Technology*, vol. 11, no. 3, pp. 250-257, 2014.

[2] Bozic J. and Babic D., “Predicting the EUR/RSD Exchange Rate Using Wavelets and Neural Network,” *in Proceedings of International Conference on Applied Internet and Information Technologies*, Zrenjanin, pp. 108-112, 2013.

[3] Bozic J., Vukotic S., and Babic D., “Prediction of the RSD Exchange Rate by Using Wavelets and Neural Networks,” *in Proceedings of the 19th Telecommunications Forum*, Belgrade, pp. 703-706, 2011.

[4] Box G., Jenkins G., and Reinsel G., *Time Series Analysis: Forecasting and Control*, John Wiley and Sons, 2008.

[5] Fan S., Ji T., Gordon W., and Rickard B., “Forecasting Baltic Dirty Tanker Index by Applying Wavelet Neural Networks,” *Journal of Transportation Technologies*, vol. 3, no. 1, pp. 68-87, 2013.

[6] Orozco J. and Garcia C., “Detecting Pathologies From Infant Cry Applying Scaled Conjugate Gradient Neural Networks,” *in Proceedings of the European Symposium on Artificial Neural Networks*, Bruges, pp. 349-354, 2003.

[7] Gencay R., Selcuk F., and Whitcher B., *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, 2002.

[8] Gil P., Cortes J., Hernandez S., and Aquino V., “A Neural Network Scheme for Long-Term Forecasting of Chaotic Time Series,” *Neural Processing Letters*, vol. 33, no. 3, pp. 215-233, 2011.

[9] Hippert H., Pedreira C., and Souza R., “Neural Networks for Short-Term Load Forecasting: A Review and Evaluation,” *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44-54, 2001.

[10] Holschneider M., *Wavelets: An Analysis Tool*, Oxford University Press, 1999.

[11] Incze A., Ileana I., and Rotar C., “The Optimization of Feed Forward Neural Networks Structure Using Genetic Algorithms,” *in Proceedings of the International Conference on Theory and Applications of Mathematics and Informatics*, Thessaloniki, pp. 223-234, 2004.

[12] Kaastr I. and Boyd M., “Designing a Neural Network for Forecasting Financial and Economic Time Series,” *Neurocomputing*, vol. 10, no. 3, pp. 215-236, 1996.

[13] Kamruzzaman J., Begg R., and Sarker R., *Artificial Neural Networks in Finance and Manufacturing*, Idea Group Inc, 2006.

[14] Kamruzzaman J. and Sarker R., “Comparing ANN Based Models with ARIMA for Prediction of Forex Rates,” *Asor Bulletin*, vol. 22, no.2, pp. 2-11, 2003.

[15] Minu K., Lineesh M., and Jessy C., “Wavelet Neural Networks for Nonlinear Time Series Analysis,” *Applied Mathematical Sciences*, vol. 4, no. 50, pp. 2485-2595, 2010.

[16] Murtagh F., Starck J., and Renaud O., “On Neuro-wavelet Modeling,” *Decision Support Systems*, vol. 37, no. 4, pp. 475-484, 2004.

[17] Ortega L. and Khashanah K., “A Neuro-wavelet Model for the Short-Term Forecasting of High-Frequency Time Series of Stock Returns,” *Journal of Forecasting*, vol. 33, no. 2, pp. 134-146, 2013.

[18] Pacelli V., Bevilacqua V., and Azzollini M., “An Artificial Neural Network Model to Forecast Exchange Rates,” *Journal of Intelligent Learning Systems and Applications*, vol. 3, no. 2, pp. 57-69, 2008.

[19] Renaud O., Stark J., and Murtagh F., “Prediction Based on a Multiscale Decomposition,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 1, no. 2, pp. 217-232, 2003.

[20] Vonesch C., Blu T., and Unser M., “Generalized Daubechies Wavelet Families,” *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4415-4429, 2007.

[21] Yao J., Li Y., and Tan C., "Option Price Forecasting using Neural Networks," *Omega*, vol. 28, no. 4, pp. 455-466, 2000.
[22] Yao J. and Tan, C., "A Case Study on Using Neural Networks to Perform Technical Forecasting of Forex," *Neurocomputing*, vol. 34, no. 1-4, pp. 79-98, 2000.
[23] Yu L., Wang S., Huang W., and Lai K., "Are Foreign Exchange Rates Predictable? Scientific Inquiry," *A Journal of International Institute for General Systems Studies*, vol. 8, no. 2, pp. 207-228, 2007.
[24] Zhang G. and Qi M., "Neural Network Forecasting for Seasonal and Trend Time Series," *European Journal of Operational Research*, vol. 160, no. 2, pp. 501-514, 2005.

Jovana Bozic obtained her Diploma Engineer degree in electrical engineering in 2007 from the School of Electrical Engineering, University of Belgrade, Serbia. She is currently a PhD candidate at the School of Computing, Department of Signal processing in telecommunications, of the Union University. Her main research interests are in the areas of time series prediction, artificial neural networks and wavelet-based signal processing. She has published several papers in national and international conferences and journals.

Djordje Babic obtained his Diploma Engineer degree in 1999 at the School of Electrical Engineering, University of Belgrade. He defended his doctoral thesis in the field of signal processing in telecommunications at Tampere University of Technology, Finland, in 2004. From 1999 to 2004, he was employed at the Institute of Telecommunications, Tampere University of Technology. Since 2008 he has been employed at the School of Computing, Belgrade, as an associate professor in the field of networked computer systems. He conducts research in the field of signal processing in telecommunications, multirate signal processing. He has published over 30 articles in different international journals and at international conferences.