FORECASTING BY NEURAL NETWORKS IN THE WAVELET DOMAIN

Ion RĂILEAN, Sorin MOGA, Monica BORDA
Technical University of Cluj-Napoca, 15 Constantin Daicoviciu Street,Cluj Napoca,400020,Romania
Telecom Bretagne, Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3 – France
Email : raileanion@yahoo.com

Abstract: This paper presents a forecasting method for time series. This method combines the wavelet analysis and several forecasting techniques such as Artificial Neural Networks (ANN), linear regression and random walk. The proposed method is tested using three real time series: the first contains historical data recorded during eight weeks from a WiMAX network and the other two are based on financial series. It is shown that AI with wavelet analysis can be an efficient and versatile approach in time series prediction for small periods time interval (up to 1 month). For long time interval, the best method used is Linear Regression technique. Also we compared the results obtained using various types of wavelets. The results show that Daubechies 1 (db1) and Reverse biorthogonal 1 (rbio1.1) give the best results.

Key words: *Time series, Wavelet transform, forecasting, Neural Networks*

I. INTRODUCTION

Forecasting or prediction is the process of estimation in unknown situations, based on the analysis of some factors that are believed to influence the future values, or based on the study of the past data behavior over time, in order to take decisions. A model is the representation of reality as it is seen by individuals who want to use this model to analyze and understand the reality and based on this information, to make short-term to long-term forecasts. Modeling and forecasting have applications in domains such as marketing, finance, telecommunications or organizational behavior.

Time-series forecasting is an important area of forecasting where the historical values are collected and analyzed in order to develop a model describing the behavior of the series. Next, this model is used to extrapolate the time series into the future where the measurements are not available.

A time-series represent a set of historical data, measured typically at successive times, each data being associated to a value. Time-series are interesting because many business operations are represented through time-series.

Modeling financial time-series is interesting and useful, with many applications. The sales history of a certain product for example represents a time-series to be forecasted. The sales forecasts are very useful in the economic domain because they are used to optimize the inventory levels. The prediction of foreign currency risk or stock market volatility is also of high interest.

Network traffic prediction plays a fundamental role in characterizing network performance and it is of significant interests in many network applications, such as adaptive applications, admission control or network management. Models that accurately catch the statistical characteristics of actual traffic are useful for analysis and simulation, and they help us to understand the network dynamics and to design and control the network. The main idea of traffic forecasting is to precisely predict traffic in the future considering the measured traffic history. The choice of the prediction method is based on the prediction interval, prediction error and computational cost.

In order to come with a suitable conclusion regarding what prediction technique to use for this scope, different types of forecasting methods have been studied. There are numerous existing models for time series forecasting, which can be grouped into four categories [1]:

(1) case-based reasoning: is a means for solving a new problem by using or adapting solutions to old problems - its essence in analogy. The basic principle of CBR is that similar problems have similar solutions

(2) rule based forecasting: its application depends upon features of the time series

(3) statistical models: exploit historical data. They contain early traditional models such as the single regressive model, exponential smoothing, ARIMA model

(4) based on soft computing models: such as neural networks and their amelioration or mixture with other methods

Between all of the above forecasting models, ANNs have been shown to produce better results [2], [3]. The performance of the neural networks against a standard statistical time series predictor is presented in [8], where a comparison between ARIMA and ANNs has been shown. This fact has been demonstrated again in [4], [5], [6]. In [7], the advantage of the artificial neural networks over

Manuscript received on October 01, 2009; revised on November 10, 2009

traditional rule-based systems is proved. Several distinguishing features of ANNs make them valuable and attractive for a forecasting task. First of all, they can be treated as multivariate nonlinear nonparametric statistical methods [9], [10], [11]. After that, ANNs can generalize, and are universal functional approximators [9]. It has been shown that a network can approximate any continuous function to any desired accuracy [12], [13], [14], [15]. Finally, ANNs are nonlinear. The traditional approaches to time series prediction, such as Box-Jenkins or ARIMA method, assume that the time series are generated from linear processes, but they may be totally inappropriate if the underlying mechanism is nonlinear [9].

In this paper, the proposed forecasting model is based on time series decomposition in wavelet domain and three prediction models: ANN, linear regression and random walk

The rest of the paper is organized as follows: in Section 2 we present the forecasting framework. Section 3 provides a presentation of the wavelet decomposition fundamentals and the à trous Wavelet Transform. A description of the three prediction techniques that were used is given in Section 4. Finally, the last two sections are dedicated to results and conclusions.

II. FORECASTING FRAMEWORK

The main idea of prediction method using wavelets is to decompose the original signal into a range of frequency scales and then to apply the forecasting methods to these individual components.

The elaboration process of a forecasting algorithm implies a series of steps such as: preprocessing step, wavelet analysis, modeling, forecasting and reconstruction.

The framework of the forecasting model using the Stationary Wavelet Transform is shown in Figure 1. It implies a series of steps such as:

1) Use the Stationary Wavelet Transform to decompose the real data.

2) Take the information from each level of decomposition, and apply the desired prediction technique in order to build the forecasting models.

3) Select the testing data sets and apply them to the forecasting models. Predict each level of decomposition.

4) Use the Inverse Stationary Wavelet Transform in order to obtain the final predicted signal.

Of course the real data passes through a preprocessing step before applying the wavelet transform. The preprocessing step includes data clearing, such as identification of the potential errors in data sets, handling missing values, and removal of noises or other unexpected results that could appear during the acquisition process. In this stage the input data is also analyzed in order to find if it contains large spikes and valleys indicating periodicities.

III. THE WAVELET ANALYSIS

The transform of a signal is just another form of representing the signal, which does not change the information content present in the signal.

The multi-resolution analysis (MRA) is a signal processing technique that takes into account the signal's representation at multiple time resolutions. Using wavelet multi-resolution analysis, the collected measurements can be smoothed until the overall long-term trend is identified. The fluctuations around the obtained trend are further analyzed at multiple time scales. The level of decomposition depends on the length of the data set (the number of values).

At each temporal resolution two categories of coefficients are obtained: approximation coefficients and detail coefficients. Generally the MRA is implemented based on Mallat's algorithm, which corresponds to the computation of the Discrete Wavelet Transform (DWT).

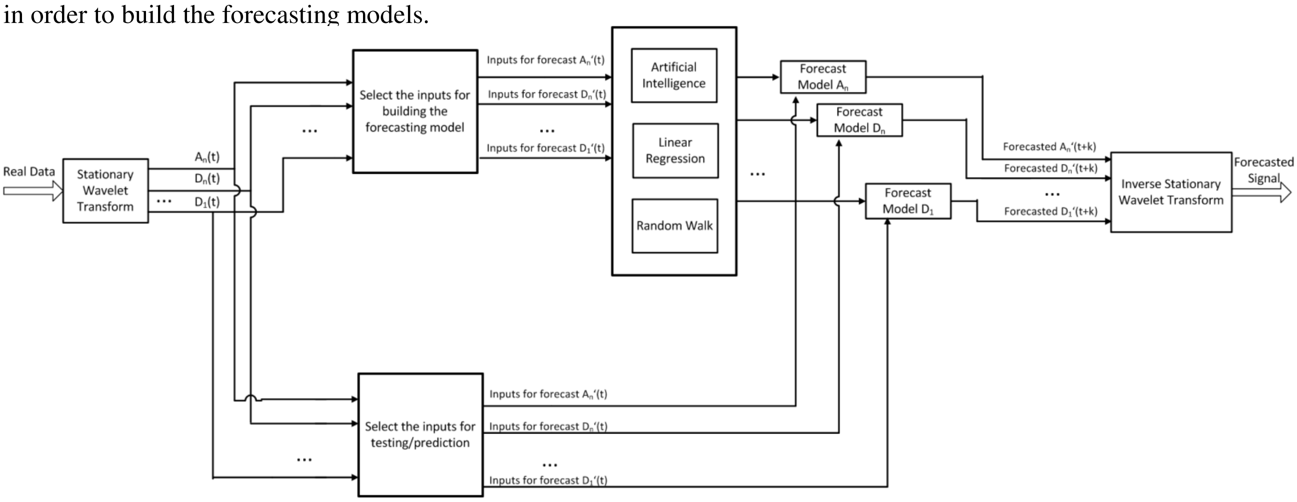

Figure 1. Forecasting Framework.

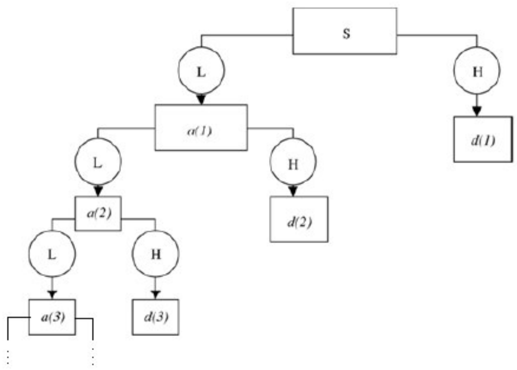

Figure 2. System for the computation of the SWT.

The disadvantage of the Mallat's algorithm is the decreasing length of the coefficient sequences with the increasing of the iteration index due to the utilization of the decimators. Another way to implement a MRA is the use of the à trous wavelet transform, also known as Shensa's algorithm [27] presented in fig.2, which corresponds to the computation of the Stationary Wavelet Transform (SWT). In this case the utilization of decimators is avoided but at each iteration different low-pass and high-pass filters are used.

The *à trous wavelet* transform decomposes a signal X_t as follows:

$$X_t = a_{p,t} + \sum_{j=1}^{p} d_{j,t} \qquad (1)$$

where $a_{p,t}$ represents the smooth version (the approximation at the p^{th} level of decomposition) of the original signal, while $d_1, \ldots, d_p$ represents the details of X_t at scale 2^{-j}. This equation can be seen as a multiple linear regression model also, where the original signal is expressed in terms of its coefficients.

IV. FORECASTING MODELS

A. Artificial Neural Networks

The Artificial Neural Network is a mathematical model that simulates the structure and functions of the real biological neural networks.

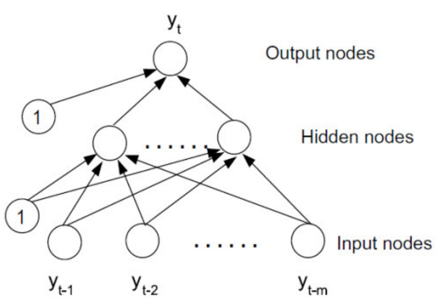

Figure 3. A three layer feed-forward network [16].

The principal characteristic is the distributed information or knowledge in connection or weight between simple elements, the artificial neuron. This element execute in/out function in two steps: weighted sum, and transfer function. This artificial neuron is associated with other similar elements, and interconnected in the network, the neural network.

As in most of the systems, the neural networks have their own advantages and disadvantages [28]:

• Advantages:

– a neural network can perform tasks that a linear program cannot

– when an element of the neural network fails, it can continue without any problem by their parallel nature

– a neural network learns and does not need to be programmed

– it can be implemented in any application

– it can be implemented without any problem

• Disadvantages

– the neural network needs training to operate

– requires high processing time for large neural networks

A neural network is characterized by three things:

1. Its architecture: the pattern of nodes and connections between them.
2. Its learning algorithm, or training method: the method for determining the weights of the connections.
3. Its activation function: the function that produces an output based on the input values received by a node.

Although many types of neural networks models have been proposed, the most popular one for time series forecasting is the feed-forward neural network [16] (Figure 3). The transfer function of the model can be written as:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j}\right) + \varepsilon_t \qquad (2)$$

where m is the number of input nodes, n is the number of hidden nodes, f is the transfer function of the network.

B. Genetic Algorithms

Genetic Algorithms (GA) are search algorithms based on the mechanics of natural selection and genetics as observed in the biological world. They use both direction ("survival of the fittest"), and randomization to robustly explore a function.

The genetic algorithm provide efficient, effective techniques for optimization and machine learning applications [40], based on natural selection, the process that drives biological evolution [30]. It repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

There are several steps in order to apply a GA:

a) encoding technique: gene, chromosome

b) initialization procedure: creation
c) evaluation function: environment
d) selection of parents: reproduction
e) genetic operators: mutation, recombination
f) parameter settings: practice and art

The population members are strings or chromosomes, which usually are binary representations of solution vectors. A GA undertakes to select the subsets (usually pairs) of solutions from a population, called parents, to combine them to produce new solutions called children or offspring. The rules of combination to yield children are based on the genetic notion of crossover, which consists of interchanging solution values of particular variables, together with occasional operations such as random value changes, called mutations. The children produced by the mating of parents, and that pass a survivability test, are then available to be chosen as parents for the next generation. The standard genetic algorithm is illustrated in the Figure 4.

The Artificial Intelligence approach in the forecasting model

The general scheme of the application of the AI is shown in Figure 5.

In order to understand our approach in using the AI modeling, we are going to explain the most important steps shown in the Figure 5. Also, it should be pointed out that the Artificial Intelligence was applied for day and week ahead prediction. From AI techniques, neural networks have been used as a forecasting model.

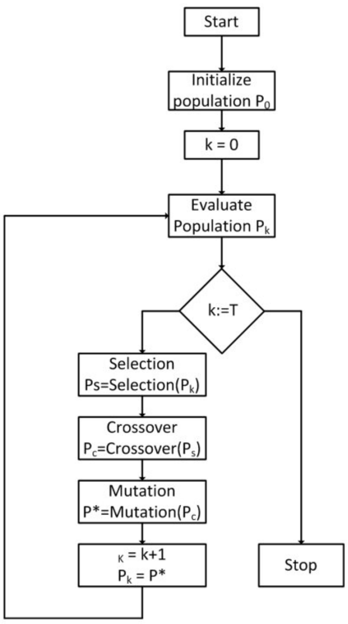

Figure 4. Standard Genetic Algorithm.

We had three data sets for analyzing. However, the processing is the same for all of them, this is why we are going to make the next explanations using only one of the data sets presented. The first data was obtained by monitoring the incoming and outgoing WiMAX traffic from 67 Base Stations during 8 weeks: from March 17th till May 11th. For each BS we have its own data set. It consists of numerical values representing the total number of packets from the uplink or downlink traffic during a time interval of 15 minutes. It results that for a given BS we have the following number of samples: 96 samples/per day, 672 per week, and a total number of 5376 samples for each Base Station.

The first step consists in choosing the information for training and for testing. We used the information from the first 7 weeks for training the networks: 6 weeks for ANN inputs, and the information from the 7th week used as a targeting data. Our testing data consists of weeks 2 till 8: the samples from weeks 2 - 7 are applied at ANNs inputs. The obtained output is compared to the traffic from the last 8th week.

Wavelet Transform for ANNs

Before applying the data to ANNs, we should preprocess it. The preprocessing is done using wavelet transform.

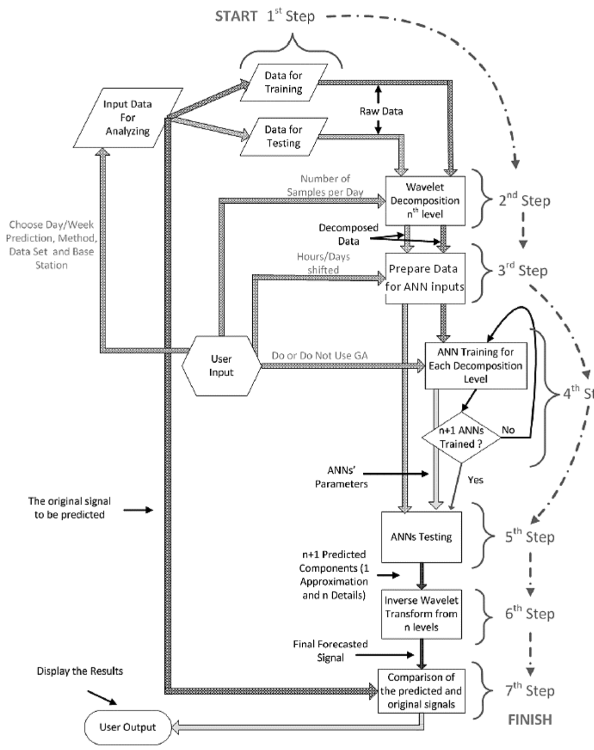

Figure 5. Block Diagram of the AI Modeling.

According to [31], for a discrete signal, in order to be able to apply the Stationary Wavelet Transform, if a decomposition at level n is needed, 2 must divide evenly into the length of the signal. So, for our data, if we have 96 samples per day (in case of 1 day ahead forecasting), then n is 5. Anyway, according to [32], there should be some periodicities in the WiMAX traffic, which are better noticed if we modify the sampling from 15 minutes to 45 and 90 minutes, by making a downsampling of the signal with 3 and 6. It results in 32 or 16 samples per day. In this case, our maximum level of decomposition cannot exceed 4, or 3 respectively. It results that we have to train and model 3, 4, or 5 neural networks.

As it has been mentioned at the beginning of this section, we are going to have a prediction for day and week ahead forecasting.

a) Wavelet Transform for Day Forecasting

We had two approaches in day prediction. The first method consists in selection of days which are similar to the one we want to forecast. For example, if we want to predict how the traffic on Wednesday is, then for processing we take only the data from all Wednesdays during weeks 1 to 7: results in the training phase that we have 6 days for future ANNs inputs, and one day for ANNs targets. The advantage in this technique, is that usually the user's behavior is modeled during certain week days, but the disadvantage is that the number of subscribers is always changing.

The second idea for day prediction, is to take into consideration the entire information until the day we want to be forecasted. Let's take the same example: prediction of Wednesday from the last 8th week. For training our ANNs, we used at theirs inputs the given level of wavelet decomposition corresponding to the information from the Monday, 1st week, until Tuesday from 7th week. The targeting data consists of the wavelet decomposition obtained from Wednesday 7th week. In testing phase, we took at ANNs inputs the decomposition from Monday 2nd week, till Tuesday 8th week. The obtained output after inverse wavelet transform, was compared to the real traffic from the 3rd day of the 8th week. The main difference between this method and the previous one, is that we do always know about the amount of traffic coming from new subscribers.

b) Wavelet Transform for Week Forecasting

We predicted the traffic from the last 8th week. We used the wavelet decomposition from the first 7 weeks for ANN training (6 for ANN inputs, and 1 for ANN target), and the wavelet transform from weeks 2 - 8 for testing (2 - 7 for inputs, while the traffic from the last week is compared to the one obtained after inverse wavelet transform from all the details and approximation taken from ANNs outputs).

Design of the Neural Network

The 2 most important types of ANNs are Feed-Forward Neural Networks, and Recurrent Neural Networks. Feed-Forward ANNs were applied in our forecasting techniques, because according to [33] ,[9], this model is relatively accurate in forecasting, despite being quite simple and easy to use. Anyway, during our tests, the recurrent network forecast performance was lower than that of the feed-forward model. It might be because of the fact that recurrent networks pass the data from back to forward as well as from forward to back, and can become “confused” or unstable. Further designing of the ANN implies the establishment of the number of layers, and the number of neurons in each layer. In [34] is pointed out the fact that the choice of the number of layers is made knowing that one hidden layer network is able to approximate most of the nonlinear functions demanded by practice. This fact has been observed by us in earlier studies on Artificial Neural Networks. That's why we chose a single hidden layer ANN. Concerning the dimension of each neuron layer the situation is as follows: input and output layers are imposed by the problem to be solved, while the dimension of the hidden layer is essential for efficiency of the network. Now, let's discuss our analysis according to the two types of prediction used.

a) ANN for Day Forecasting

For the output number of neurons, taking into consideration that we want to predict a single element (data from one day, or data from one week), gives us the idea of using just one neuron for output, which contains in case of day prediction an array of length 96, 32, or 16 samples. In case of week forecasting, we have 672, 224, or 112 values.

Regarding the number of neurons for the input layer, we have several options. The first option was to make a temporal synchronization during the whole day. One of the methods to achieve this is by taking an entire day from hours 00:00, till 24:00 as each input of the ANNs. By this we made sure that the morning, noon, and evening periods from known data, were responsible for the same periods of the day from the forecasting sequence. In this case, we have 6 neurons for input.

Another option for choosing the information for our neurons is if we think about the behavior of each person. According to [35], [36] our life cycle is divided into 3 parts: 8 hours of sleep, 8 hours of work, and the rest 8 hours for rest. But there are some industries where the use of these hours is shifted. The same survey described in the mentioned articles noted that there was a range of different 12 hours systems: in companies as steel, chemicals, aluminum, oil, food, engineering. Based on these articles, we used 8, 12, and 4 hours shifting between the information taken for ANN's neurons. In this case, the numbers of input neurons are 16, 11, or 31 respectively, by applying the next formula:

$$NrOfInputs = [NrOfDays - 1] \cdot \frac{24h}{NrHoursShifted} + 1 \quad (3)$$

An example of the data selection in 8 hours shifting is presented in Figure 6.

b) ANN for Week Forecasting

In case of week prediction, we have also single neuron output, and variable number of inputs. The difference is that we will make a shifting not by hours, but by days.

We have implemented also some methods for several weeks forecasting. The first method is similar to the previous: one week ahead forecasting. The only difference is that in this case we will take as a target 2 weeks (or more in case if we have enough information), and consequently less data for ANNs inputs.

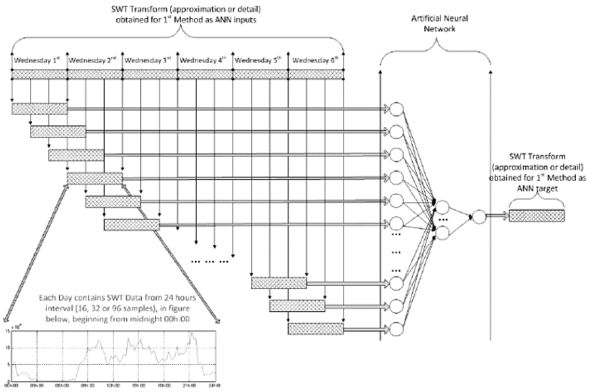

Figure 6. An example of 8 hours shifting.

The next method uses sliding with retraining the network with the real data. The entire information is divided into smaller parts. Each of this sequence will predict a small part from the final forecasted signal. The information for neural networks retraining is always taken from the real data. An example of 9 weeks prediction with 3 weeks sliding is shown in the Figure 7.

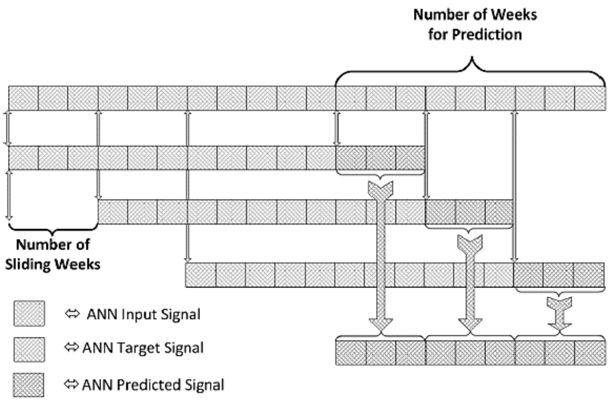

Figure 7. An example of 9 weeks prediction with 3 weeks sliding.

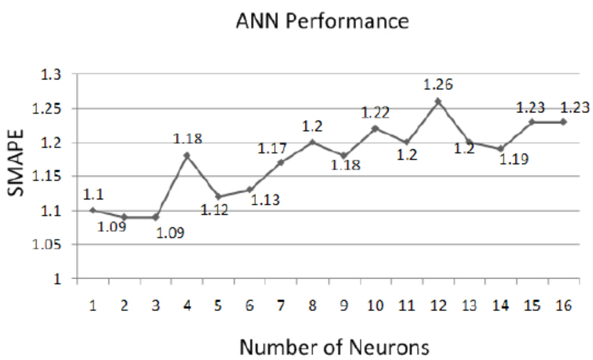

Figure 8. Obtained SMAPE values as a function of neurons number in the hidden layer for ANNs used in Day Prediction.

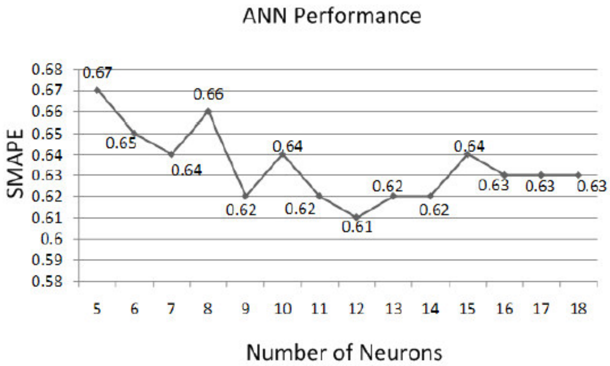

Figure 9. Obtained SMAPE values as a function of neurons number in the hidden layer for ANNs used in Week Prediction.

The last method, was making a forecasting using sliding with unknown data. The only difference consists in the fact that the information used for the next simulation and retraining is taken not from the original signal, but from the previously predicted one.

If the number of input and output neurons are kind of "decided" by the information we have, then the hidden layer stays completely at our choice of designing.

In [34] is said that the network dimension must satisfy at least two criteria: 1) the network must be able to learn the input data; 2) the network must be able to generalize for similar input data that was not used in training set. The accomplishment degree of these requirements depends on the network complexity, training data set, and the number of iterations for training.

After simulations and tests, presented in Figures 8 and 9, we got the conclusion that the best value for the number of neurons in the hidden layer is 3 for day forecasting, and 12 for week forecasting.

Genetic Algorithm optimization

The training of the neural networks can be done in several ways. One way is to "train" the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. For this method we used the Neural Toolbox from Matlab. The other idea for training the networks was to apply the GA in order to find the optimal weights between the input and hidden layer. Because of the processing time, which is proportional to the number of links between layers, this technique was used only in day prediction for first method.

The designing of training the ANN using the Genetic Algorithms is as follows:

- each individual contains a set of weights for all the links between layers
- each gene represents a single weight
- we had a population size of 100 individuals, meaning 100 different possibilities at each generation for the network

• the number of generations is 100: less generations resulted in not finding an acceptable solution for our problem, while more generations resulted in a longer time processing. However, above this value, we didn't manage to observe better performance of the final results

• the fitness function represented the summation between the two training data sets, calculated as follows:

$$[Fitness] = \frac{1}{[NrOfSamples]}\sum_{i=1}^{[NrOfSamples]}(x_i^{forecasted1} - x_i^{original1}) + \frac{1}{[NrOfSamples]}\sum_{i=1}^{[NrOfSamples]}(x_i^{forecasted2} - x_i^{original2}) \tag{4}$$

Now, we will explain what do *forecasted1, forecasted2, original1*, and *original2* signals mean. As it has been mentioned before, we have 8 weeks of data. The first method uses 1 day from each week. The GA method proposed by us, requires 2 different signals for training. For the calculation of the first part in fitness function, at the input of the ANN we used the samples from wavelet transform of the weeks 1st till 5th. As a target, the wavelet transform of the day from the week number 6 was used. So, the forecasted1 signal represents the ANN output when ww apply at its inputs the information from the weeks 1 - 5, while original1 is the real signal from the 6th week. The forecasted2 signal (used in the computation of the second part of the fitness function) represents the output of the ANN if we apply at its inputs the data from weeks 2 - 6, and compare it to the original2 signal, which is the real information from week number 7.

B. Linear regression

Linear regression is a statistical tool for modeling where the output is linear combination of inputs.

$$y = \beta_0 + \sum_{i=1}^{n}\beta_i x_i \tag{5}$$

where y represents the output data, β is the weight vector, β_0 is called bias of the model, and x represents the input data ($x_0 = 1$ for bias).

The parameters of the linear regression model are usually estimated using the least-squares method.

C. Random walk

The random walk process is the simplest model from the time series models and it is characterized by the fact that the changes in the time series follow a random direction that is unpredictable.

This model is given by:

$$y(t) = y(t-1) + \alpha \tag{6}$$

meaning that the prediction of the next values will equal the previous values plus the average change one period to the next, α. This model assumes that, from one period to the next, the original time series takes a random "step" away from its last recorded position.

D. Measuring the accuracy

The widely used statistical measures of error that are used to identify a method or the optimum value of the parameter within a method are:

• **Mean absolute error** (MAE) value is the average absolute error value; closer this value is to zero the better the forecast is.

$$MAE = \frac{1}{T}\sum_{1}^{T}|y_{real} - y_{forecast}| \tag{7}$$

• **Mean Square Error** (MSE) and its normalized function (NMSE) is a measure of the absolute error; as the prediction accuracy increases, the MSE becomes smaller. If NMSE > 1, it means that the prediction performance is worse than that of the trivial predictor.

$$MSE = \frac{\sum_1^T (y_{real} - y_{forecast})^2}{\sum_1^T (y_{real} - \bar{y})^2} \tag{8}$$

• **ANOVA** (analysis of variance) - computing the coefficient of determination (R^2) determines the goodness of the regression. The model is considered to be statistically significant if it can account for a large fraction of the variability in the response (large values for R^2).

$$R^2 = \frac{SS_R}{SS_T} \tag{9}$$

$$SS_T = \sum_i (y_i - \bar{y})^2 \tag{10}$$

$$SS_R = \sum_i (\hat{y}_i - \bar{y})^2 \tag{11}$$

in which y_i, $\hat{y}_i$, and $\bar{y}$ are the original, modeled, and medium values respectively.

• **Mean Absolute Percent Error** (MAPE) - calculates the mean absolute error in percent between the real and forecasted signals:

$$MAPE = \frac{1}{T}\sum_{1}^{T}\left|\frac{y_{real} - y_{forecast}}{y_{real}}\right| \times 100\% \tag{12}$$

• **Symmetric Mean Absolute Percent Error** (SMAPE) - calculates the symmetric absolute error in percent between the actual *X* and the forecast *F* across all observations t of the test set of size *n* for each time series *s*.

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|X_t - F_t|}{(X_t + F_t)/2} \qquad (13)$$

• **Root Mean Square Error** (RMSE): measure of the differences between values predicted by a model and the values actually observed from the time-series being modeled or estimated.

$$RMSE = \sqrt{MSE} \qquad (14)$$

V. RESULTS

The tests were made using Matlab software and its WaveLab 850 toolbox, which is a collection of functions used to implement a variety of algorithms related to wavelet analysis. In this work we used historical data collected continuously during a period of eight weeks at the level of each base station (BS) composing a WiMAX network. These values represent the amount of traffic measured in bytes or in packets, for all the links of all the BSs in the WiMAX network during the period of collection. Also, we applied the method on a series of financial data representing the total number of EUR-USD currency exchanges during 15 minutes interval during 15 weeks. The third dataset contains a series composed by the EUR - USD exchange rate at a time interval of 15 minutes. The results of our simulations have to answer the next questions:

• what is the best wavelet transform applied
• what are the optimal configurations for our neural networks
• what is the best implemented method for week prediction 83577

1) Wavelet Transform

In order to answer the first question, regarding wavelet transform, we have made a comparison between Daubechies, Coiflet, Symlet, Biorthogonal and Reverse Biorthogonal wavelets for the 3 data sets described. The experimental results for the WiMAX traffic are presented in Table 1.

The results prove that the Haar wavelet, which is the simplest of the Daubechies wavelets, and the *rbio1.1* wavelet, are the best wavelets regarding this aspect. The results also indicate that with the increase of the filters' length, the performance of the wavelet deteriorates.

For the second set containing a series formed by the EUR-USD exchange currency the results are shown in Table 2. In this case we can observe that the best forecasting performance is obtained using the wavelets *coif2* and *sym2*.

For the third data set the results regarding the wavelet type used are presented in Table 3. Analyzing these values we cannot conclude which type of wavelet to use. It depends of the errors we consider to be most important. Also, these values have been obtained after simulating only a single database. We need much more data and simulations in order to be sure about the decision regarding the wavelet transform.

Wavelet	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
coif 1		1.09	0.20786	11.72	2.80	1.01353	0.890	0.001991	0.94416
coif 2	1.493	1.22	0.22480	12.95	2.83	0.86053	0.837	0.001875	0.70730
db 1	1.168	1.08	0.23290	8.06	2.43	0.75595	0.812	0.001663	0.72075
db 2	1.364	1.15	0.24113	10.52	2.69	0.82710	0.855	0.001876	0.76411
db 3	1.358	1.12	0.20827	9.76	2.64	0.80595	0.857	0.001834	0.75529
db 4	1.490	1.11	0.21242	10.61	2.58	0.78548	0.834	0.001833	0.74397
db 5	1.435	1.11	0.21546	12.56	2.75	0.82029	0.823	0.001904	0.76039
bior 3.1	0.695	1.13	0.31003	9.86	2.52	0.98410	0.860	0.001824	0.69559
rbio 1.1	1.200	1.08	0.21793	10.00	2.61	0.78186	0.820	0.001739	0.87994
rbio 2.2	1.482	1.19	0.31499	10.29	2.71	0.83093	0.981	0.001797	0.75639
rbio 3.3	1.952	1.21	0.25805	10.33	2.88	0.93533	0.907	0.002168	1.05153
sym 2	1.365	1.26	0.21120	13.20	2.89	0.87092	0.895	0.001916	0.72906

Table 1. Comparison of the Wavelets used for WiMAX Traffic Forecasting.

Wavelet	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
coif 1	0.688	0.556	0.115242	1.627	1.220	0.358626	0.516	0.082165	0.424088
coif 2	0.455	0.522	0.079319	1.089	1.038	0.296780	0.453	0.073231	0.379912
db 1	0.625	0.520	0.083934	1.356	1.126	0.317560	0.454	0.071385	0.369044
db 2	0.715	0.578	0.108890	1.610	1.219	0.355066	0.497	0.081286	0.419912
db 3	0.586	0.585	0.115659	1.499	1.188	0.361879	0.531	0.086407	0.446132
db 4	0.871	0.600	0.111495	1.604	1.239	0.374560	0.527	0.086363	0.445989
db 5	0.808	0.587	0.112176	1.557	1.225	0.370033	0.546	0.091220	0.471000
bior 3.1	0.628	0.534	0.113714	1.552	1.173	0.339066	0.433	0.071297	0.368176
rbio 1.1	0.615	0.519	0.095835	1.286	1.096	0.320857	0.457	0.071670	0.370462
rbio 2.2	0.541	0.555	0.108747	1.440	1.149	0.340681	0.491	0.079011	0.408154
rbio 3.3	0.772	0.595	0.099352	1.418	1.167	0.348099	0.455	0.071692	0.370538
sym 2	0.476	0.499	0.089000	1.117	1.037	0.296242	0.453	0.073626	0.381319

Table 2. Comparison of the Wavelets used for WiMAX financial data (exchange rate).

Wavelet	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
coif 1	1.4612	0.0087	0.669804	11.7330	3.2000	0.008724	0.2745	0.314869	0.410545
coif 2	2.4350	0.0089	0.687557	10.2959	3.0502	0.008953	0.2739	0.313528	0.408785
db 1	0.5011	0.0093	0.719850	11.0782	3.1762	0.009412	0.2675	0.302545	0.394475
db 2	0.9468	0.0112	0.872283	12.6357	3.4910	0.011402	0.2772	0.317884	0.414524
db 3	1.7864	0.0071	0.549051	8.3235	2.6839	0.007193	0.2725	0.312330	0.407254
db 4	3.9849	0.0069	0.530839	10.6729	2.9384	0.006964	0.2726	0.312841	0.407943
db 5	3.3660	0.0068	0.525482	10.6936	2.8954	0.006887	0.2731	0.313664	0.409014
bior 3.1	2.4196	0.0085	0.660927	12.4006	3.2603	0.008647	0.2727	0.311866	0.406642
rbio 1.1	0.5584	0.0103	0.797138	11.3382	3.2730	0.010407	0.2681	0.303005	0.395087
rbio 2.2	0.8704	0.0080	0.617233	10.6246	3.0167	0.008035	0.2722	0.311225	0.405800
rbio 3.3	0.7793	0.0087	0.672329	10.6935	3.0813	0.008800	0.2652	0.299092	0.389660
sym 2	0.9531	0.0089	0.689317	11.0653	3.1447	0.009030	0.2758	0.316808	0.413070

Table 3. Comparison of the Wavelets used for financial data (exchange rate).

2) Optimal Configuration for the Neural Networks

Our next task, was to find out what is the best ANN configuration used in forecasting. Because we have only one financial database, the next experiments regarding this aspect were done based on WiMAX traffic. In case of day prediction, we had the next options to choose from: station number (from 67 possible), number of samples per day (16, 32 or 96), number of hours shifted (4, 8, 12 or 24), the day of the week (from Monday till Sunday), and applying or not the Genetic Algorithm optimization. 8232 simulations were considered from 11256 possible after combining all of the above possibilities (except of the last one referring to the GA). The results for RSQ and SMAPE values are shown in the Table 4.

Method	Hours shifted	Measured Value	96 samples	32 samples	16 samples
I Method	4 hours	RSQ	1.212	1.249	1.317
		SMAPE	0.905	0.927	0.802
	8 hours	RSQ	**1.159**	**1.120**	**1.205**
		SMAPE	**0.886**	**0.858**	**0.756**
	12 hours	RSQ	1.330	1.219	1.287
		SMAPE	0.904	0.924	0.814
	24 hours	RSQ	1.178	1.201	1.263
		SMAPE	0.910	0.946	0.780
II Method	4 hours	RSQ	0.715	0.739	0.711
		SMAPE	0.868	0.843	0.732
	8 hours	RSQ	**0.820**	**0.782**	**0.792**
		SMAPE	**0.861**	**0.849**	**0.720**
	12 hours	RSQ	0.634	0.767	0.690
		SMAPE	0.866	0.857	0.738
	24 hours	RSQ	0.741	0.781	0.703
		SMAPE	0.869	0.876	0.752

Table 4. Medium RSQ and SMAPE values in one day prediction using ANN.

The forecasting technique using GA Optimized Neural Networks was applied for the same data set also. Another 8232 simulations have been obtained. The medium RSQ and SMAPE values are presented in the Table 5.

Hours shifted	Measured Value	96 samples	32 samples	16 samples
4 hours	RSQ	1.127	1.200	1.176
	SMAPE	1.001	**0.913**	0.867
8 hours	RSQ	**1.108**	**1.129**	**1.099**
	SMAPE	**0.983**	0.924	**0.820**
12 hours	RSQ	1.211	1.155	1.168
	SMAPE	1.008	0.951	0.851
24 hours	RSQ	1.087	1.189	1.215
	SMAPE	1.067	0.950	0.872

Table 5. Medium RSQ and SMAPE values in one day prediction using Genetically Optimized ANN.

By comparing the results from the Tables 4 and 5, we can see that applying second method prediction (all days selection), gives us better results with about 2-6% in comparison to the first forecasting method. Regarding the use of GA optimization, we can see that the results for SMAPE are worse with about 10% compared to those obtained using ordinary ANN training. But the RSQ values are closer to the ideal 1. In both cases, the optimal number for shifted hours is 8. Another aspect is the number of samples per day. Even if the best results were obtained using 16 samples per day, we should point out that in this case we are not able to predict the sudden increases and peaks of the traffic. So, for this aspect we would recommend to stay with 96 samples per day.

In case of week prediction, we had 1449 simulations by changing the station number (from 1 till 67), number of samples per day (16, 32 or 96), and number of shifted days (from 1 till 7). The results for RSQ and SMAPE values are presented in Table 6.

Samp/Day	Days shifted	1	2	3	4	5	6	7
96	RSQ	**1.243**	**1.195**	1.659	1.281	1.308	**1.077**	1.252
	SMAPE	**0.895**	1.036	1.057	1.247	1.081	**0.819**	**0.894**
32	RSQ	**1.318**	**1.276**	1.415	1.392	1.401	**1.106**	1.380
	SMAPE	0.992	1.0004	**0.989**	1.129	1.001	**0.857**	**0.896**
16	RSQ	**1.663**	**1.397**	1.891	1.805	1.742	**1.215**	1.712
	SMAPE	**0.840**	1.012	0.961	1.093	1.036	**0.845**	**0.793**

Table 6. Medium RSQ and SMAPE values in one week prediction.

ANN Type	RSQ	SMAPE	MAPE	MSE	RMSE	MAE	SMAPE L	MAPE L	MAE L
No Sliding	1.137	**0.946**	48728	2.15	1.43	52578	**0.472**	93.68	37188
Known Sliding	**0.960**	1.006	**25218**	**1.991**	**1.399**	**50084**	0.509	**79.60**	**35652**
UnKnown Sliding	0.845	1.052	33783	3.255	1.738	56620	0.722	150.59	56111

Table 7. ANN Methods Comparison (un-normalized MAPE and MAE values).

We can see that the best results are obtained using 1, 2 or 6 days shifting with 16 sample per day. However, as in the one day prediction method, by using 16 samples per day we are not able to seize the peaks of the signal.

Regarding the 3 methods using neural networks for week prediction, the results are shown in the Table 7. SMAPE L, MAPE L, and MAE L values are calculated from the medium value of all the samples during one week. We implemented 2 weeks ahead forecasting. The second ANN technique with the retraining with Known Data for ANN inputs, gives better results. This is because we predict small intervals of information considering all the time the known data during training process: both inputs and target. An example shown the comparison between 2nd and 3rd methods for prediction are presented in Figures 10 and 11.

3) Methods' Comparison for week forecasting

In case of WiMAX traffic, the comparison was done using rbio1.1 wavelet, 2 weeks ahead forecasting. The results are shown in the Table 8. We can observe that ANN performs better than the other prediction techniques. We can see also a very good forecasting from the linear regression model.

Forecasting Model	SMAPE L	MAPE L	MAE L
ANN No Sliding	0.472	0.00109	0.43557
ANN Known Sliding	0.509	0.00093	0.41758
ANN UnKnown Sliding	0.722	0.00176	0.65721
Linear Regression	0.523	0.00306	0.38054
Random Walk using Wavelets	4.440	0.00301	1.34098

Table 8. Forecasting Techniques Comparison used for WiMAX Traffic.

For the financial data, we used the coif 2 wavelet. The results are presented in Table 9.

Forecasting Model	SMAPE L	MAPE L	MAE L
ANN No Sliding	0.169	0.020747	0.107956
ANN Known Sliding	0.153	0.017824	0.095736
ANN UnKnown Sliding	0.267	0.034494	0.181274
Linear Regression	0.191	0.024351	0.110956
Random Walk using Wavelets	0.940	0.267021	1.417362
Random Walk without Wavelets	0.803	0.228032	1.160505

Table 9. Forecasting Techniques Comparison (Financial Traffic).

In case of data traffic, both WiMAX and financial, ANN has good results for small future time intervals, several weeks at most. But if we are interested in the tendency of the traffic, meaning prediction for several months, than the Linear Regression model is the one that should be taken into consideration.

For the data set containing the EUR-USD exchange rates, the simulations were done using db1 wavelet, with the results presented in Table 10.

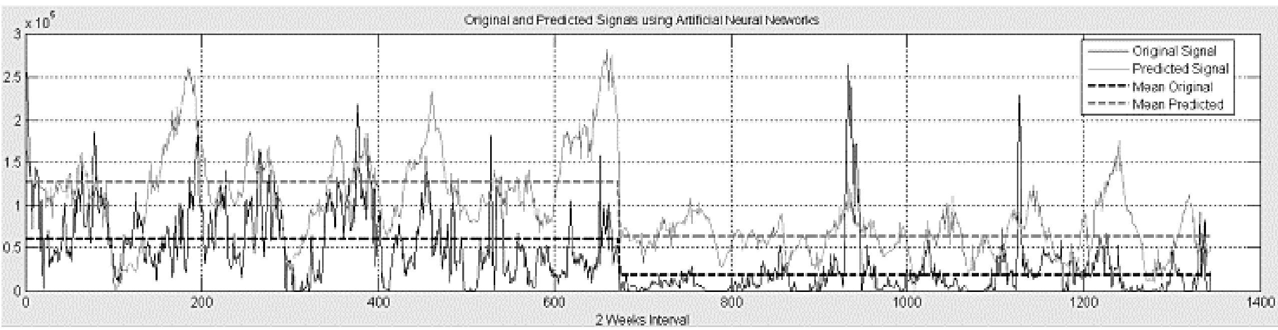

Figure 10. Example of a Known Sliding Prediction for 2 Weeks.

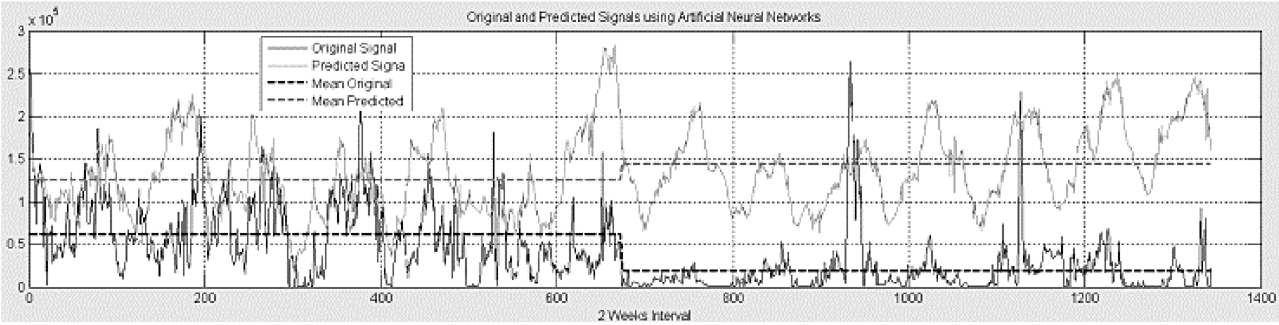

Figure 11. Example of an UnKnown Sliding Prediction for 2 weeks.

Forecasting Model	SMAPE L	MAPE L	MAE L
ANN No Sliding	0.00524	0.403206	0.005356
ANN Known Sliding	0.01244	0.956496	0.012580
ANN UnKnown Sliding	0.00957	0.727808	0.009649
Linear Regression	0.02159	1.632790	0.021632
Random Walk using Wavelets	1.17423	56.494513	0.746563
Random Walk without Wavelets	0.08567	6.173454	0.081114

Table 10. Forecasting Techniques Comparison (EUR-USD Exchange Rates).

In case of the Exchange Rates, the only suitable model is the one using neural networks. We are not able to apply Linear Regression or Random Walk models for future tendency prediction.

VI. CONCLUSIONS

In this paper we proposed several data traffic prediction algorithms based on Wavelet Analysis. One of them was the technique combining AI modeling. The raw data is decomposed into distinct timescales using Wavelet Transform. The forecasted wavelet signals are obtained independently at each decomposition level using ANN approach. Genetic Optimized Neural Networks were also implemented. The other forecasting models implemented are: Linear Regression (LR), and Random Walk (RW). So, we had the following prediction methods:

- 2 for Day Forecasting (ANN models)
- 3 for Week Forecasting: ANN approach, LR, and RW models

We had 3 sets of data for this evaluation.

1) The first data set represents the WiMAX traffic from 8 weeks time interval. It has been shown that the best results are obtained when applying db1 and rbio1.1 wavelet transform on the signals.

A comparison between the 2 methods for Day Forecasting was made. The second method, that takes into consideration the entire information gives more accurate results than the single day selection technique.

In order to improve our forecasting method, we implemented several approaches regarding the way we use our data for ANNs inputs: for one day prediction, we made a time synchronization of 4, 8, 12 and 24 hours; while for one week prediction, we implemented a synchronization of 1, 2, ..., 7 days. Our simulations show, that for day forecasting the best results are obtained using 8 hours shifting. The results are better with 1-3% in comparison to the other cases. For 1 week prediction, best results are obtained if we apply 1, 2, and 6 days shifting. In this case a performance with 5-15% was noticed.

To increase further the model's performance, we applied Genetic Optimization techniques to find out optimal values for ANN's weights. This optimization increased the SMAPE error with 2-7%, but decreased the RSQ error, setting it closer to the ideal 1 value.

For week prediction, it has been shown that the smallest error values are obtained when applying ANNs (No Sliding Technique). Anyway, we should remind again that if we are interested in tendency prediction of future traffic, than the Linear Regression model is more welcome to this aspect.

2) The second data set represents the number of EUR-USD exchanges during 15 weeks time interval. For this data, our results show that the best wavelet transforms are coif2, and sym2. The techniques for prediction which give better results are the same: ANNs and LR models. As in the case of the WiMAX traffic, we should apply LR model for long time interval, if we are interested in global medium of the traffic, and not in possible peaks.

3) The third data set consists of EUR-USD exchange rates during 15 weeks. According to our simulations, no preferable wavelet transform was found for this data set. But we should take into consideration that we had small amount of information, and no certain generalization can be concluded here. Comparing the proposed prediction models, the smallest error values have been obtained using Artificial Neural Networks (No Sliding). Its performance was much higher for this data set than the results obtained using other forecasting techniques. In comparison to the other data sets, the LR model cannot be used for long time interval forecasting in this case.

The forecasting techniques in this paper can be effectively used for building prediction models for time series. But in order to have higher performance and to reduce the prediction errors, we would recommend to have longer range of data for analysis, to take into consideration the localization of the region where the information has been taken, and to analyze what types of business is developed in the given region. These aspects represent the start points for our future work in this domain.

REFERENCES

[1] Chen Rong Qiu, Zhan Rong, “A case based reasoning system for individual demand forecasting”, 4th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM '08. 12-14 Oct. 2008, pp.1-6,
[2] Scott Armstrong, Monica Adyaa, ”An application of rule based forecasting to a situation lacking domain knowledge”, International Journal of Forecasting, Vol. 16, pp. 477-484, 2000,
[3] Amit Mitra, Sharmishtha Mitra, “Modeling exchange rates using wavelet decomposed genetic neural networks.” Statistical Methodology,vol. 3, issue 2, april 2006, pp. 103-124 2006,
[4] R.J.Abrahart, L.See, “Neural Network vs. ARMA Modelling: constructing benchmark case studies of river flow prediction,” http://divcom.otago.ac.nz/sirc/ GeoComp/GeoComp98/05/ gc_05.htm.
[5] Refenes AP, “Constructive learning and its application to currency exchange rate forecasting”, In: Trippi RR, Turban E, editors. Neural networks in finance and investing. Chicago: Probus Publishing Company, 1993. pp. 465-93.
[6] G. Ibarra-Berastegi, A. Elias, R. Arias, A. Barona, "Artificial Neural Networks vs Linear Regression in a Fluid Mechanics and Chemical Modelling Problem: Elimination of Hydrogen Sulphide in a Lab-Scale Biofilter," IEEE/ACS International Conference on Computer Systems and Applications, 2007, pp.584-587,
[7] Clarence N. W. Tan., “Incorporating Artificial Neural Network into a Rule-based Financial Trading System”, The First New Zealand International Two Stream Conference on Artificial Neural Networks and Expert Systems (ANNES), University of Otago, Dunedin, New Zealand, November 24-26, 1993, IEEE Computer Society Press, ISBN 0-8186-4260-2, 1993.
[8] Ahmed Mohammed, Ahmed Omran,”The arima versus articial neural networks modeling”
[9] G. Zhang, B. Eddy Patuwo, M. Y. Hu, “Forecasting with artificial neural networks: The state of the art”, International Journal of Forecasting 14, 1998, pp 35-62
[10] H. White,“Learning in artificial neural networks: A statistical perspective”, Neural Computation 1, 1989, pp. 425-464
[11] B.D. Ripley, “Statistical aspects of neural networks”, Barndoff-Nielsen, O. E., Jensen, J. L., W. S. (Eds), Networks and Chaos-Statistical and Probabilistic Aspects, Chapman and Hall, London, 1993, pp40-123
[12] B. Irie, S. Miyake, “Capabilities of three-layered perceptrons”, Proceedings of the First Forum on Application of Neural Networks to Power Systems, 1988, pp. 41-45
[13] K. Hornik, M. Stinchcombe, H. White, “Multilayer feedforward networks are universal approximators”, Neural Networks 2, 1989, pp. 359-366
[14] K. Funahashi, “On the approximate realization of continuous mappings by neural networks”, Neural Networks 2, 1989, pp. 183-192
[15] K. Hornik, “Some new results on neural network approximation”, Neural Networks 6, 1993, pp. 1069-1072
[16] G. Peter Zhang, Min Qi, “Neural network forecasting for seasonal and trend time series”, European Journal of Operational Research 160, 2005, pp. 501-514
[17] K. Papagiannaki, N. Taft, Zhi-Li Zhang, and C. Diot, ‘’Long-term forecasting of internet backbone traffic’’, IEEE Transactions on Neural Networks 2005, vol.16, pp. 1110-1124, 2005,
[18] Maria Papadopouli, Elias Raftopoulos, Haipeng Shen, ‘’Evaluation of short-term traffic forecasting algorithms in wireless networks’’, Next Generation Internet Design and Engineering, 2006-NGI '06, vol. 8 pp. 109, 2006,
[19] Huang Hong-Qiong; Tang Tian-Hao, “Short-term Traffic Flow Forecasting Based on ARIMA-ANN”, IEEE International Conference on Control and Automation, 2007. ICCA 2007, May 30 2007-June 1 2007, pp. 2370 – 2373,
[20] Ming Zhou; Zheng Yan; Yixin Ni; Gengyin Li, “An ARIMA approach to forecasting electricity price with accuracy improvement by predicted errors”, IEEE Power Engineering Society General Meeting, 6-10 June 2004, vol.1,pp. 233 - 238,
[21] Zhou, M.; Yan, Z.; Ni, Y.X.; Li, G.; Nie, Y ,“Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach”,Generation, Transmission and Distribution, IEEE Proceedings, 16 March 2006, vol.153, pp. 187 - 195,
[22] J.C.Palomares-Salas, et al, ‘’ARIMA vs. Neural networks for wind speed forecasting” Computational Intelligence for Measurement Systems and Applications”, IEEE CIMSA '09 International Conference, 11-13 May 2009 pp. 129 – 133,
[23] P.Perez et all “An Study on Data Mining Methods for Short-Term Forecasting of the Extra Virgin Olive Oil Price in the Spanish Market”, Eighth International Conference on Hybrid Intelligent Systems, 2008- HIS '08, 10-12 Sept. 2008, pp. 943 – 946,
[24] J.Sosa, C.Sosa, B.Paz, ”ARIMA models in the rain attenuation prediction in a Mexican tropical area”, Antennas and Propagation Society International Symposium, 16-21 July 2000, vol.2, pp. 546 – 549,
[25] Jian-chang Lu; Xingping Zhang; Wei Sun , “A Real-time Adaptive Forecasting Algorithm for Electric Power Load”, Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES , pp. 1 – 5,
[26] Huifang Feng Yantai Shu Shuyi Wang Maode Ma,”SVM-Based Models for Predicting WLAN Traffic”,ICC '06. IEEE International Conference on Communications, 2006. , vol.2, pp.597-602,
[27] M.J.Shensa, “Discrete Wavelet Transform.Wedding the a trous and Mallatorithms,”IEEE Transactions and Signal Processing, 40 (1992), pp.2464-2482,
[28] K. Kannan, D. Venkatesan, “A genetic algorithm-based articial neural network model for the optimization of machining processes”, Neural Computing & Applications 2008, pp. 135-140
[29] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, CA, 1970,
[30] Genetic Algorithm and Direct Search Toolbox. The MathWorks Inc., 2004-2009.
[31] Yves Misiti, Michel Misiti. Wavelet Toolbox 4 Users Guide. MathWorks Inc., 1997 2009,
[32] Cristina Stolojescu, Alina Cusnir, Sorin Moga, Alexandru Isar, “Forecasting WiMAX BS Traffic by Statistical Processing in the Wavelet Domain”, Proceedings of the International Symposium on Signals, Circuits and Systems - ISSCS 2009, 9-20 July 2009, Iasi, Romania, pp.177-183,
[33] Milton S. Boyd, Giovani Dematos, “Feedforward versus recurrent neural networks for forecasting monthly japanese yen exchange rates”, Financial Engineering and the Japanese Markets, Kluwer Academic Publishers 1996, vol.3, pp.59-75,
[34] Arpad Incze, Ioan Ileana, Corina Rotar, “The optimization of feed forward neural networks structure using genetic algorithms”, Proceedings of the International Conference on Theory and

Applications of Mathematics and Informatics - ICTAMI 2004, Thessaloniki, Greece, pp.223-234
[35] P. Knauth, “Designing better shift systems”, Appl. Ergon 27 (1) (1996), pp. 39–44.
[36] Lawrence Smith, Philip Tucker, ”Distribution of rest days in 12 hour shift systems: impacts on health, wellbeing, and on shift alertness”, Occupational and Environmental Medicine 1999,(56), pp.206-214
[37] http://www.statsoft.com/textbook/sttimser.html#1general, economics.uwaterloo.ca/2007-Winter/603notes3.pdf ,
[39] http://home.ubalt.edu/ntsbarsh/statdata/ Forecast.htm#rboxjenkm
[40] http://www.compapp.dcu.ie/~ray/teaching/CA559/notes/ GA.pdf