Proceedings of the 9th International Conference "Reliability and Statistics in Transportation and Communication" (RelStat'09), 21–24 October 2009, Riga, Latvia, p. 170-177. ISBN 978-9984-818-21-4
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

FORECASTING NETWORK TRAFFIC: A COMPARISON OF NEURAL NETWORKS AND LINEAR MODELS

Irina Klevecka

Institute of Telecommunications / Riga Technical University
Azenes Str. 12, Riga, LV-1048, Latvia
Ph.: +371 26006970. E-mail: klevecka@inbox.lv

The main aim of the research was to produce the short-term forecasts of traffic loads by means of neural networks (a multilayer perceptron) and traditional linear models such as autoregressive-integrated moving average models (ARIMA) and exponential smoothing. The traffic of a conventional telephone network as well as a packet-switched IP-network has been analysed. The experimental results prove that in most cases the differences in the quality of short-term forecasts produced by neural networks and linear models are not statistically significant. Therefore, under certain circumstances, the application of such complicated and time-consuming methods as neural networks to forecasting real traffic loads can be unreasonable.

Keywords: *telecommunications, packet-switched networks, traffic forecasting, neural networks, ARIMA, exponential smoothing*

1. Introduction

The object of the research is the time series characterizing the real traffic of both traditional telephone networks (POTS) and packet-switched IP-networks. The reliable forecasts of traffic generated by users (subscribers) allow planning the capacity of transmission channels, avoiding the overload and sustaining the optimal level of quality of service.

A rapid development of packet-switched networks and the transformation of traditional telephone networks into multi-service systems offer new opportunities to a user (subscriber) and expand his / her scope of activities. Though, not only the architecture of telecommunications networks but also the statistical nature of traffic has been changed what implies a strong influence of such self-similarity.

In the analysis of dynamic behaviour of IP-networks such a complicated and time-consuming non-linear method as neural networks is gaining more and more acceptance. However, the solution of the task of traffic forecasting is not trivial. As our research proves, under the certain circumstances the reliable forecasts of the traffic of packet-switched networks can be produced by applying traditional linear methods as well.

2. Peculiarities of Forecasting Network Traffic

In order to model and forecast the dynamics of network traffic, we usually assume that its values are expressed by discrete time series. A discrete time series is defined as a vector $\{x_t\}$ of observations made at regularly spaced time points $t = 1, 2, \ldots, N$. Unlike the observations of a random variable, the observations of a time series are not statistically independent. This relation sets up the specific base for forecasting an analysed variable (i.e. for producing the estimate $\hat{x}(N+L)$ of an unknown value $x(N+L)$ taking into account the historical values $x(t_1), x(t_2), \ldots, x(t_N)$).

The methods of traffic forecasting are defined by the ITU-T recommendations E.506 and E.507 [1; 2]. Even the recommendations are partly obsolete and are supposed to be used for forecasting the traffic of ISDN-networks, some of the methods still can be applied to modern telecommunications networks. In particular, these methods are autoregressive-integrated moving average (ARIMA) models and exponential smoothing.

As it has been already mentioned, the empirically observed traffic of packet-switched networks is self-similar in a statistical sense, over a wide range of time scales. Self-similarity in wide sense means that a covariance structure is preserved when the time series is aggregated. Objects with this self-similar quality are called fractals. Two important features of self-similar traffic are long-range dependence and a slowly decaying variance [3]. These statistical effects of self-similar traffic make its forecasting more complicated than the prediction of the traffic of traditional telephone traffic which is characterized by short-range dependence.

Non-linear neural networks have won popularity in time series forecasting, among the tasks of which is also the prediction of packet-switched traffic. They provide additional opportunities in modelling non-linear phenomena and recognizing chaotic behaviour of processes. However, as the current research shows, in some cases traditional linear models succeed in forecasting the traffic of packet-switched networks not less than such resources- and time-consuming methods as neural networks.

It is important to keep in mind that self-similarity of packet-switched traffic, which comes along with a slowly decaying variance and long-range dependence, has a prominent influence only in the case of measurements in a very small scale – over the aggregation period varying from milliseconds to approximately 15 minutes (see Figure 2.1). From the point of view of time series forecasting such a fine scale is often unreasonable. In this case the selection of the adequate statistical model can be complicated due to a strong influence of autocorrelation between distant observations of a times series as well as due to the influence of extraneous noises and anomalous outliers which unavoidably entail the measurements in such a fine scale. Therefore, according to the ITU Recommendation E.492 [4] the measurements of network traffic should be averaged over 15- minutes and / or one-hour intervals. In this case we can often speak about the possibility of applying traditional methods of time series forecasting.

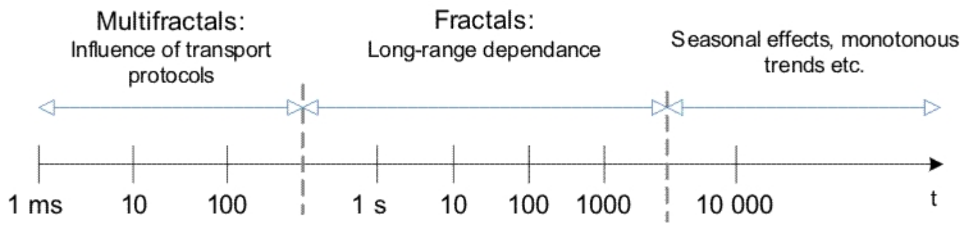


Figure 2.1. Statistical effects of packet-switched traffic depending on a time scale [5, with author's changes]

A reference period also determines a forecasting horizon for which reliable forecasts can be produced. Conditionally speaking, the possible forecasting horizons for time series aggregated over the period of one second or 24 hours are different.

The main accent of this research was put on the application of neural networks (i.e. a multi-layer perceptron) for forecasting the changes of the traffic of both traditional telephone networks and packet-switched IP-networks. The forecasts produced by non-linear models were compared to those which were produced by traditional linear models. For the purpose of this comparison the models of ARIMA and exponential smoothing were chosen (as the methods recommended by the ITU-T). If the comparative analysis of forecasts produced by neural networks and linear models do not reveal any statistically significant differences, then the application of such a complicated and time-consuming method as neural networks does not make sense.

3. The Methods of Traffic Forecasting

3.1. Neural Networks

Neural networks are massively parallel, distributed processing systems representing a new computational technology built on the analogy to the human information processing system. A neural network consists of a large number of simple processing elements called neurons or nodes. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight. The weights represent information being used by the network to solve a problem.

Neural networks are suitable for solving various tasks including time series forecasting. The temporal structure of an analysed sample can be built into the operation of a neural network in implicit or explicit way [6]. In the first case the temporal structure of the input signal is embedded in the spatial structure of the network. The input signal is usually uniformly sampled, and the sequence of synaptic weights of each neuron connected to the input layer of the network is convolved with a different sequence of input samples.

Explicit representation of temporal structure, when time is given its own particular representation, is used rarely. Therefore, only the implicit representation of time, whereby a static neural network is provided with dynamic properties, has been implemented in this research.

For a neural network to be dynamic, it must be given memory, which may be divided into short-term and long-term memory. Long-term memory is built into a neural network through supervised learning, whereby the information content of the training data set is stored in the synaptic weights of the network. However, if the task has a temporal dimension, we need some form of short-term memory to make the network dynamic. One way of building short-term memory into the structure of a neural network is through the use of time delays which can be implemented at the synaptic level inside the network or at the input layer of the network.

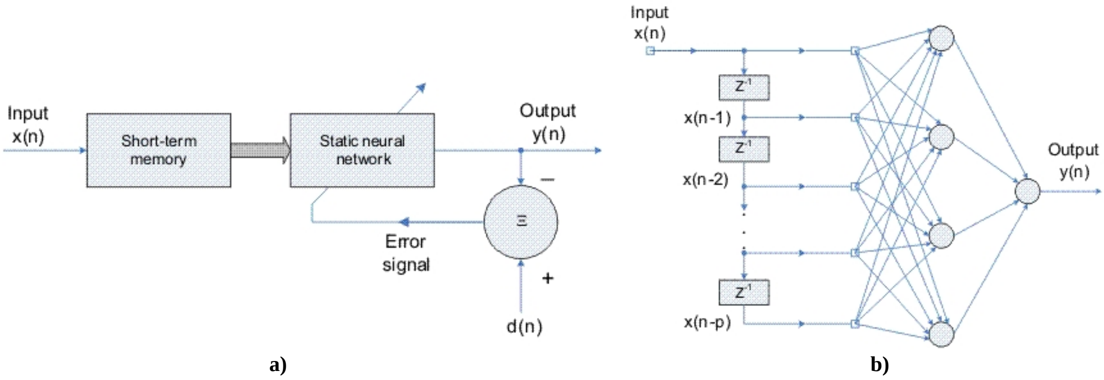


Figure 3.1 Temporal processing using neural networks:
a) nonlinear filter built on a static neural network [6]; b) time lagged feed-forward network (TLFN) [6] [7]

Temporal pattern recognition requires processing of patterns that evolve over time, with the response at a particular instant of time depending not only on the present value of the input but also on its past values. Figure 3.1(a) shows the block diagram of a nonlinear filter built on a static neural network. Given a specific input signal consisting of the current value $x(n)$ and the p past values $x(n-1)$, …, $x(n-p)$ stored in a delay line memory of order p, the free parameters of the network are adjusted to minimize the training error (the mean square error) between the output of the network, $y(n)$, and the desired response $d(n)$ [6].

The structure shown in Figure 3.1(a) can be implemented at the level of a single neuron or a network of neurons. A time lagged feed-forward network is shown in Figure3.1(b). It consists of a tapped delay memory of order p and a multilayer perceptron (MLP). A standard back-propagation algorithm can be used to train this type of neural networks.

3.2. ARIMA Models

The processes of auto-regression, moving average and their combinations refer to the class of linear models, as all the relations between the observations and random errors of a time series are expressed by means of linear mathematical operations.

In contrast to simulated traffic, the real traffic usually incorporates seasonal and / or cyclic components. In this case one should pay his / her attention to the seasonal modifications of ARIMA.

The ARIMA is the Box-Jenkins variant of conventional ARMA models, which is predestinated for applications to non-stationary time series that become stationary after their differencing. In the case of seasonal ARIMA models, seasonal differencing is also applied in order to eliminate a seasonal component of period s.

If d and D are non-negative integers, then $\{x_t\}$ is a seasonal ARIMA(p,d,q)(P,D,Q) process given by [8]:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D x_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \qquad (3.1)$$

where
s – period of a cyclic component;
B – delay operator;
$\phi(B)$ – autoregressive operator of order p;
$\theta(B)$ – moving-average operator of order q;
$\Phi(B^s)$ – seasonal autoregressive operator of order P;
$\Theta(B^s)$ – seasonal moving-average operator of order Q;
∇ – differencing operator given by $\nabla=\nabla_1=1-B$;
∇_s – seasonal differencing operator given by $\nabla_s=1-B^s$
ε_t – white noise.

The operators $\phi(B)$, $\theta(B)$, $\Phi(B^s)$ and $\Theta(B^s)$ have to satisfy the conditions of stationarity and reversibility. The indexes p, P, q and Q are introduced here in order to remind about different orders of the operators. The description of the ARIMA process incorporating two and more periodic components is analogous to this.

3.3. Exponential Smoothing

The method of exponential smoothing is the generalization of moving average technique. It allows building the description of a process whereby the latest observations are given largest weights in comparison with earlier observations, and the weights are exponentially decreasing.

There exist different modifications of exponential smoothing which are suitable for modelling and forecasting the time series incorporating linear / non-linear trends and / or seasonal fluctuations. Such models are based on the decomposition of time series.

Just as in the case of ARIMA models, the task of forecasting real network traffic requires applying the seasonal modifications of exponential smoothing. In this research the model of exponential smoothing with additive seasonality was implemented to constant-level processes. Its mathematical expression is given by [9]:

$$\begin{aligned} S_t &= \alpha(x_t - I_{t-p}) + (1-\alpha)S_{t-1} \\ I_t &= \delta(x_t - S_t) + (1-\delta)I_{t-p} \end{aligned}, \tag{3.2}$$

where
α – smoothing parameter for the level of the series;
S_t – smoothed level of the series, computed after x_t is observed;
δ – smoothing parameter for seasonal factors;
I_t – smoothed seasonal index at the end of the period t;
p – number of periods in the seasonal cycle.

In this case the forecast is calculated as follows [9]:

$$\hat{x}_t(l) = S_t + I_{t-s+l}, \tag{3.3}$$

where
$\hat{x}_t(l)$ – forecast for l periods ahead from origin t.

Network traffic measured over long time periods (several years) usually incorporates not only seasonal fluctuations but also a linear trend. Then it is necessary to use seasonal trend modifications of exponential smoothing, the description of which can be found in [9].

4. Practical Research

Sixteen time series of different length and different aggregation periods have been analysed in the process of the research. The measurements were taken on the transportation level and characterize three variables – the transmission rate of outgoing international traffic of the IP-network, the transmission rate of total outgoing traffic of the IP-network and the intensity of the total serviced load of the conventional telephone network.

The main aim of the practical research was to analyse the statistical properties of the time series and to develop such a neural network which is suitable for modelling the underlying process and producing a reliable forecast for a pre-defined forecasting horizon. The selection of the relevant neural network closely followed an advanced algorithm introduced in [10].

A secondary task of the research was evaluating the influence of the size of the basic sample on the effectiveness and reliability of forecasting. The size of the basic sample was equal to 9, 12, and, in some cases, 18 weeks. The forecasting period (i.e. the size of a testing sample) varied from one to 14 days. The period of averaging the errors of one-step-ahead forecasts was set up to one day.

Following the ITU-T Recommendation E.492 [4] the initial traffic measurements were averaged over 15-minutes and one-hour periods. The main characteristics of the time series are shown in Table 4.1.

All the analysed time series are characterized by the presence of seasonal components with periods of 24 hours and one week. That was revealed by applying a Fourier analysis. The estimates of the

Hurst exponent vary from 0.65 for telephone traffic to 0.85 for packet-switched traffic. Such values indicate the persistence of analysed time series and exploit the potentialities of their further forecasting.

Table 4.1. The main characteristics of analysed time series

Network Type	Traffic Type	Aggregation Period	Labelling	Basic Sample	Size of Basic Sample	Size of Testing Sample
IP-network	Outgoing international traffic	15 min.	A	May 12 – Jul. 13, 2008	9 weeks (6048 obs.)	1- 14 days
			B	May 12 – Aug. 3, 2008	12 weeks (8064 obs.)	
		1 hour	C	May 12 – Jul. 13, 2008	9 weeks (1512 obs.)	
			D	May 12 – Aug. 3, 2008	12 weeks (2016 obs.)	
IP-network	Total outgoing traffic	15 min.	E	Feb. 4 – Apr. 6, 2008	9 weeks (6048 obs.)	1- 14 days
			F	Feb. 4 – Apr. 27, 2008	12 weeks (8064 obs.)	
			G	Feb. 4 – June 8, 2008	18 weeks (12096 obs.)	
		1 hour	H	Feb. 4 – Apr. 6, 2008	9 weeks (1512 obs.)	
			I	Feb. 4 – Apr. 27, 2008	12 weeks (2016 obs.)	
			J	Feb. 4 – June 8, 2008	18 weeks (3024 obs.)	
Telephone network (POTS)	Total serviced traffic	15 min.	K	Jan. 8 – Mar. 11, 2007	9 weeks (6048 obs.)	1- 14 days
			L	Jan. 8 – Apr. 1, 2007	12 weeks (8064 obs.)	
			M	Jan. 8 – May 13, 2007	18 weeks (12096 obs.)	
		1 hour	N	Jan. 8 – Mar. 11, 2007	9 weeks (1512 obs.)	
			O	Jan. 8 – Apr. 1, 2007	12 weeks (2016 obs.)	
			P	Jan. 8 – May 13, 2007	18 weeks (3024 obs.)	

The specification of the developed neural network is displayed in Table 4.2.

Neural networks belong to the so-called heuristic methods. It means that the relevant values of most parameters of neural networks have to be evaluated in experimental way.

Table 4.2. The main parameters of the developed neural network[1]

Stage	Parameter / Procedure	Parameter Value / Procedure Description
Selection of network topology	Type of topology	Time-lagged feed-forward network (multilayer perceptron)
	Number of hidden layers	1
	Number of hidden neurons	Varying from 1 to 10
	Number of output neurons	1
	Activation function	Hidden layer – hyperbolic tangent; output layer – linear function
Training	Number of training epochs	600
	Training algorithm	Back propagation (100 epochs) & conjugate gradient descent (1000 epochs)
	Error function	Mean square error
	Learning rate	0.1
	Momentum term	0.3
	Method of initialisation of weights and biases	Randomised values from a uniform distribution with a range of [-0.5;0.5]
	Number of times to randomise weights and biases	100
	Methods to prevent over-learning	Cross-validation, weight regularization [13]
	The size of training, validation and test subsets	In 3:1:1 proportion
	Stopping criterion	Training error is invariable during 50 epochs
In-sample and out-of-sample evaluation	The parameters of in-sample evaluation	R, MAE, RMSE, MAPE, AIC, BIC
	Diagnostic testing of residuals	Lagrange multiplier type test [14] , χ^2 - test
	The parameters of out-of-sample evaluation	RMSE, MAE, MAPE, the Diebold-Mariano criterion

[1] Notations: R – correlation coefficient, MAE – mean square error, RMSE – root mean square error, MAPE – mean absolute percentage error, AIC – Akaike's information criterion, BIC – Bayesian information criterion.

The appropriate architecture of a neural network was defined as follows. According to the universal approximation theorem [11] the number of hidden layers was equal to one. The size of the input window was equal to the largest period of the cyclic component identified by means of a Fourier analysis. The number of output neurons was equal to one and implied a one-step ahead forecasting. In order to identify the relevant number of hidden neurons all the architectures with the number of hidden neurons varying form one to ten have been tested and verified.

A two-stage training process was implemented. During the first stage a multilayer perceptron was trained by applying the back propagation during one hundred epochs, with learning rate 0.1 and momentum 0.3. It usually gives the opportunity to locate the approximate position of a reasonable minimum. During the second stage, a long period of conjugate gradient descent (1000 epochs) is used, with a stopping window of 50, to terminate training once convergence stops or over-learning occurs. Once the algorithm stops, the best network from the training run is restored.

The final neural network was chosen in compliance with the method suggested in [10]. According to that, among competing neural networks the model with uncorrelated residuals and the smallest value of the information criterion (IC) has to be chosen for further forecasting.

The final forecasts produced by neural networks were compared to those which are produced by the seasonal ARIMA and seasonal exponential smoothing. The Diebold-Mariano test [12] has been applied in order to evaluate statistically significant differences between these forecasts. The test is non-parametric and can be used even if forecasting errors do not comply with the classic requirements, i.e. they are non-normally distributed, auto-correlated or serially correlated.

For the sake of space saving, only one empirical example illustrating the production of the forecasts for the time series (B) is shown here. However, the main conclusions have been drawn taking into account the whole set of produced forecasts and the complete results of verification.

As a result of verification procedures, three models have been chosen for further forecasting of the time series (B) – a time lagged multilayer perceptron with one hidden neuron MLP 672-1-1, a seasonal model $SARIMA(1,0,6)(0,1,1)_{672}$, and the model of exponential smoothing with additive seasonality and parameters $\alpha = 0.19$ и $\gamma = 0.00$. The final forecasts produced by these models over a forecasting horizon up to two weeks are shown in Figure 4.1.

The standard estimates of the quality of the produced pseudo-forecasts are shown in Figure 4.2. The values of these parameters do not differ significantly and vary depending on a forecasting period. Therefore, it is hard to say which model performs better than others. The Diebold-Mariano test [12] has been implemented in order to identify statistically significant differences between produced forecasts for three forecasting horizons such as 24 hours, a week and two weeks. As we can see in Table 5.1, there are

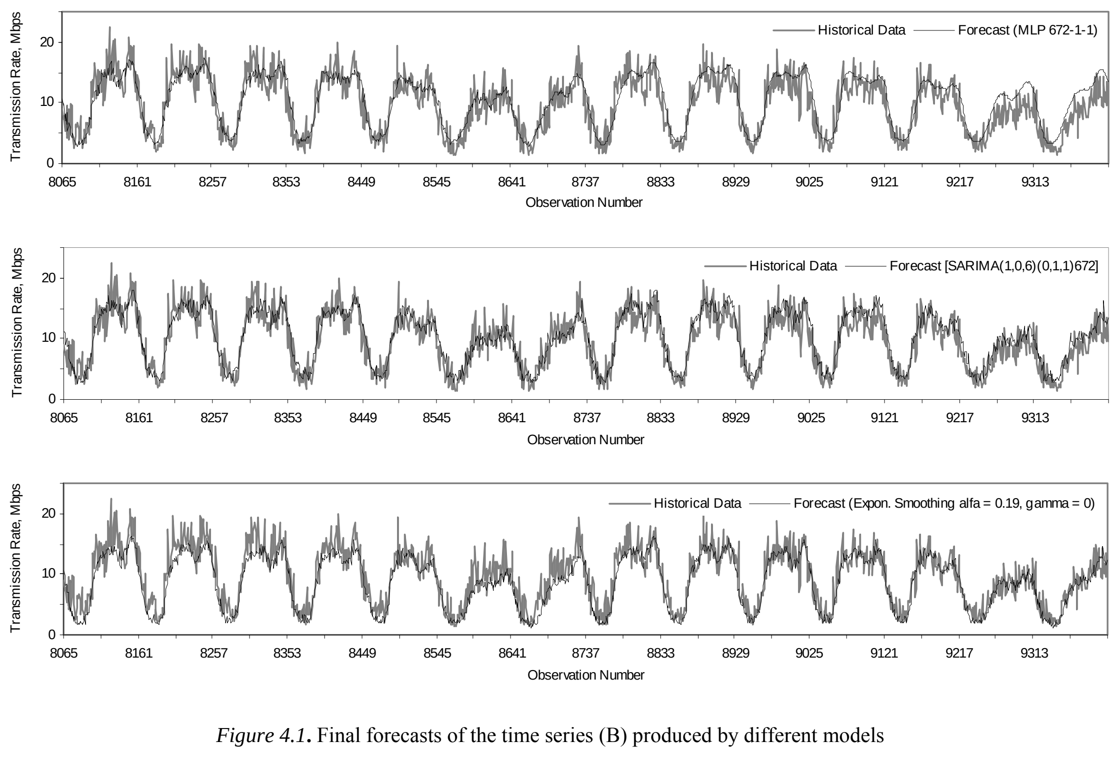
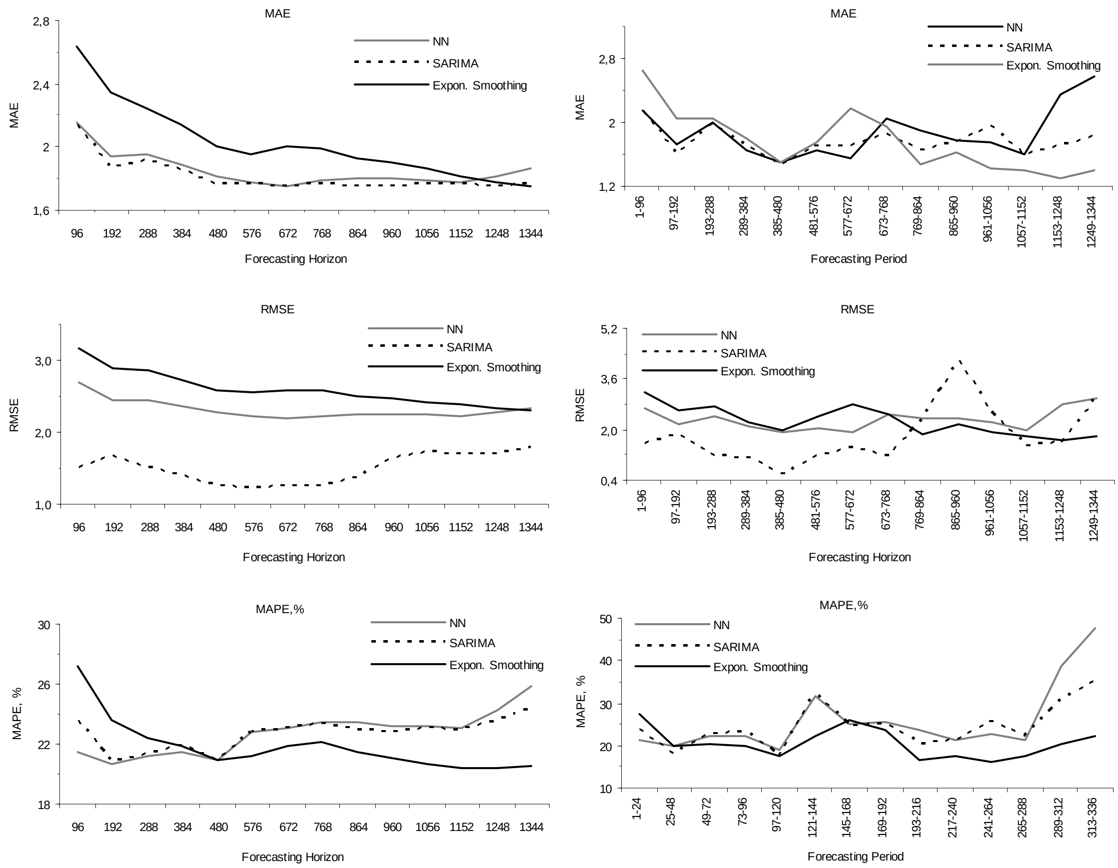


Figure 4.1. Final forecasts of the time series (B) produced by different models

no statistically significant differences between forecasts produced by the neural network and SARIMA over the forecasting horizons of 24 hours and one week. However, as a forecasting horizon increases, the quality of forecasts produced by the neural network deteriorates. Thus, the SARIMA model outperforms the neural network over a forecasting horizon of two weeks. On the other hand, forecasts produced by the neural network perform better than those produced by exponential smoothing independently of a forecasting horizon.



a) **b)**

Figure 4.2. Out-of-sample evaluation of the quality of forecasts
a) depending on the forecasting horizon; b) for successive forecasting periods of 96 observations (24 hours)

Table 4.3. The evaluation of statistically significant differences between final forecasts of the time series (B) by means of the Diebold -Mariano test [2]

Forecasting Horizon (L) & Step (h) / **Models**	**L=96, h=1**	**L=672, h=1**	**L=1344, h=1**
Neural Network vs. SARIMA	-0.44 (0.67)	0.80 (0.42)	-4.83 **(0.00)**
Neural Network vs. Seasonal Exponential Smoothing	2.40 **(0.02)**	37.09 **(0.00)**	14.97 **(0.00)**
SARIMA vs. Seasonal Exponential Smoothing	2.59 **(0.01)**	36.24 **(0.00)**	20.30 **(0.00)**

[2] The significance level of the Diebold-Mariano test is shown in brackets and highlighted in bold for the values less than 0.05.

Despite of the result of this particular comparative study the comparison of forecasts produced for other time series by applying neural networks, SARIMA models and exponential smoothing, in most cases did not reveal any statistically significant differences.

5. Conclusions

The results of the research show that in most cases the differences in quality between short-term forecasts of network traffic produced by neural networks and linear models are not statistically significant. Therefore, contrary to popular belief, the use of such complicated and time-consuming methods as neural networks is not always appropriate. This issue requires further research with the aim of specifying the conditions under which the mechanism of neural networks has to be applied to forecasting the traffic of telecommunications networks.

Some other important conclusions have been made as well:

- A long-range dependence and a slowly decaying variance of the packet-switched traffic are apparent only for the measurements taken over a very fine scale, usually over the periods up to 10-15 minutes. If according to the ITU-T recommendations the measurements of traffic are averaged over largest periods than the traditional linear methods can be implemented for forecasting purposes as well.
- Unlike simulated time series, real time series usually incorporate prominent seasonal fluctuations which are the result of human behaviour. A neural network can model and forecast seasonal time series without prior de-seasonalisation. In this case the most important parameter to define is the size of the input window which has to be equal to the largest period of a seasonal component.
- The task of forecasting network traffic incorporating periodic fluctuations requires focusing on the seasonal modifications of linear models. Just as in the case of neural networks, the correct identification of the periods of seasonal components is important for successful modelling and forecasting.

References

1. *ITU-T (CCITT) Recommendation E.506 (rev.1). Forecasting International Traffic*. Geneva, 1992.
2. *ITU-T Recommendation E.507. Models for Forecasting International Traffic*. ITU, 1993 (1988).
3. Sheluhin, O.I., Smolskiy, S.M. and Osin, A.V. *Self-Similar Processes in Telecommunications*. John Willey & Sons, 2007.
4. I*TU-T Recommendation E.492. Traffic Reference Period (Telephone Network and ISDN. Quality of Service, Network Management and Traffic Engineering).* ITU, 02/1996.
5. Fowler, T.B. A Short Tutorial on Fractals and Internet Traffic, *The Telecommunications Review*, No 10, 1999, pp.1-15.
6. Haykin, S. *Neural Networks: A Comprehensive Foundation (2nd ed.).* Prentice Hall, 1998.
7. Mozer, M. C. Neural Network Architectures for Temporal Pattern Processing. In: *Time Series Prediction: Forecasting the Future and Understanding the* Past */ Eds. A.S. Weigend and N. A. Gershenfeld.* Perseus Book Publishing, 1993, pp. 243-264.
8. Box, G. and Jenkins, G. *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day, 1970.
9. Gardner, E.S.Jr. Exponential Smoothing: The State of Art, *Journal of Forecasting*, No 4, 1985, pp.1-28.
10. Klevecka, I. An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks. In: *Scientific Proceedings of Riga Technical University (Series "Telecommunications and Electronics"),* No 9, 2009. (Submitted for publication)
11. Cybenko, G. Approximation by Super positions of Sigmoidal Function, *Mathematics of Control, Signals, and Systems*, No 2, 1989, pp.303-314.
12. Diebold, F.C. and Mariano, R.S. Comparing Predictive Accuracy, *Journal of Business & Economic Statistics,* Vol. 13, No 3, July 1995, pp.253-263.
13. Weigend, A.S., Rumelhart, D.E. and Huberman, B.A. Generalization by Weight-Elimination with Application to Forecasting. In: *Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems,* No 3, 1990, pp. 875-882.
14. Medeiros, M.C., Teräsvirta, T. and Rech, G. Building Neural Network Models for Time Series: A Statistical Approach, *Journal of Forecasting*, No 25, 2006, pp. 49-75.