Computer Modelling and New Technologies, 2010, Vol.14, No.2, 20–28
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

FORECASTING TRAFFIC LOADS: NEURAL NETWORKS vs. LINEAR MODELS

I. Klevecka

Institute of Telecommunications / Riga Technical University
12 Azenes St., Riga, LV-1048, Latvia
Phone: +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 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. It has been proven that empirically observed packet-switched traffic is characterized by self-similarity which comes along with such statistical effects as long-range dependence and a slowly decaying variance. That led to a belief (quite ungrounded) that traditional linear methods are not suitable for solving a forecasting task because of their focusing on short-range dependant processes.

In the analysis of dynamic behaviour of IP-networks the mechanism of neural networks is gaining more and more acceptance. Neural networks provide additional opportunities in modelling non-linear phenomena and recognizing chaotic behaviour of processes. On the other hand, neural networks are often criticized for a very large number of parameters to define in empirical way, difficulties in producing and replicating a stable solution and the risk of losing generalization ability due to over-training. Besides, neural networks are very time-consuming methods which also require powerful technical facilities. Therefore, it is important to define if there is a necessity to apply neural networks and, if so, under which conditions.

Numerous papers dedicated to the application of neural networks to forecasting packet-switched traffic have been published. However, most of them solve a trivial task of forecasting with more or less success and did not pay much attention to evaluating the statistical properties of analysed traffic traces. None of them also carry out a complex comparative analysis of forecasts produced by neural networks and traditional linear methods.

Taking that into account the main goals of the research were specified as:

- to test the ability of traditional linear models such as autoregressive-integrated moving average models (ARIMA) and exponential smoothing to produce the short-term forecasts of real network traffic;
- to compare the quality of the forecasts produced by traditional linear methods with those which are produced by means of neural networks (a multilayer perceptron);
- to specify the conditions under which the mechanism of neural networks has to be applied to forecasting the traffic of telecommunications networks.

From practical point of view the traffic of packet-switched network is of most interest. However, there are only few research papers dedicated to the prediction of conventional telephone traffic by means of neural networks. Therefore, the task of verifying the ability of neural networks to predict telephone traffic was set as well.

2. Background of Forecasting Network Traffic

Solving the task of traffic modelling and forecasting, 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 [0][0]. 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. Consider a discrete time stochastic process or time series $\{x(t)\}$, $t\in Z$, where $x(t)$ is the traffic volume – measured in packets, bits or bytes – at time instance t. Under the assumption of stationarity, $\{x(t)\}$ is called *exactly* second-order self-similar with Hurst parameter H ($0.5 < H < 1$) if for all $k \geq 1$ [1][0]

$$\gamma(k) = \frac{\sigma^2}{2}\left((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right), \tag{2.1}$$

where $\gamma(k)$ – the autocovariance function of $\{x_t\}$; k – time shift (lag); H – Hurst exponent.

Objects possessing self-similar quality are called fractals. For aggregated processes[2], $\gamma(k) = \gamma^{(m)}(k)$ for all $m \geq 1$, where m – the aggregation period. Thus, second-order self-similarity assumes that correlation structure exactly or asymptotically preserves under time aggregation.

Two important statistical features of self-similar processes are long-range dependence and a slowly decaying variance.

Let $r(k) = \gamma / \sigma^2$ denote the autocorrelation function. Then, $\{x(t)\}$ is called the stationary process with long-range dependence, if under the assumption $0.5 < H < 1$, $r(k)$ asymptotically behaves as $ck^{-\beta}$ for $0 < \beta < 1$ (and $\beta = 2 - 2H$), where $c > 0$ is a constant. It this case $r(k)$ is assumed to be non-summable [0]:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty . \tag{2.2}$$

That is, the autocorrelation function of long-range dependent processes decays slowly – i.e., hyperbolically, in contrast to short-rage dependant processes with autocorrelation function decaying quickly.

In its turn, the variance of aggregated self-similar processes decays more slowly as compared to the magnitude inverse to the sample size. For $0.5 < H < 1$, $H \neq 0.5$, it holds [0]

$$\sigma^2(X^{(m)}) \sim m^{-\beta} \tag{2.3}$$

with $0 < \beta < 1$ (and $H = 1 - \beta/2$).

It implies that for rather large m, a self-similar process is visually more uneven and irregular (i.e. possesses the property of high variance) than a short-range dependant process.

The degree of long-range dependence is usually evaluated by means of a Hurst exponent [0]. In network traffic theory the notions of *self-similarity* and *long-range dependence* are often interchangeable but it is worth noting that not all self-similar processes are long-range dependant and vice versa. However, *asymptotic* second-order self-similarity assumes long-range dependence by the restriction $0.5 < H < 1$, and vice versa [0].

Due to the influence of long-range dependence, a forecasting process of self-similar traffic is more complicated as compared to the prediction of traditional telephone traffic which is characterized by short-range

---

[1] The notion of *asymptotic* self-similarity also exists nad can be found in [0].

[2] To formulate scale-invariance, in traffic theory the aggregated process $x^{(m)}$ at aggregation level m is defined as $x^{(m)}(i) = \frac{1}{m}\sum_{t=m(i-1)+1}^{mi} x(t)$ [0]. That is, $\{x(t)\}$is portioned into non-overlapping blocks of size m, their values are averaged, and i is used to index these blocks.

dependence. Non-linear neural networks have won popularity in the prediction of packet-switched traffic. However, numerous research papers dedicated to the application of neural networks usually miss the fact that a fractal nature of packet-switched traffic has a prominent influence only in the case of measurements on a large scale – over the aggregation periods varying from milliseconds to approximately 5–15 minutes (see Fig. 2.1).

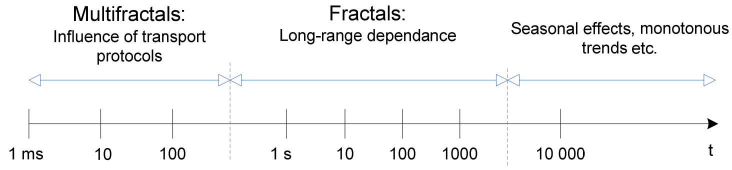

Figure 2.1. Statistical effects of packet-switched traffic depending on a time scale [0, with author's changes]

Such a fine sampling scale is often unreasonable from the point of view of time series forecasting. In this case the selection of the relevant 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 on a large scale. Besides, an aggregation/sampling period also determines a forecasting horizon for which reliable forecasts can be produced. In other words, the possible forecasting horizons for time series aggregated over the period of one second or 24 hours are different. At present neural networks are not suitable for real-time forecasting; therefore aggregation on a fine scale does not make sense. Taking that into account and following the ITU-T Recommendation E.492 [0], it is advised to average measurements of network traffic over 15-minutes and/or one-hour intervals. Over these sampling periods, human behaviour and technical progress are those factors that influence the statistical properties of traffic more than self-similarity (Fig. 2.2).

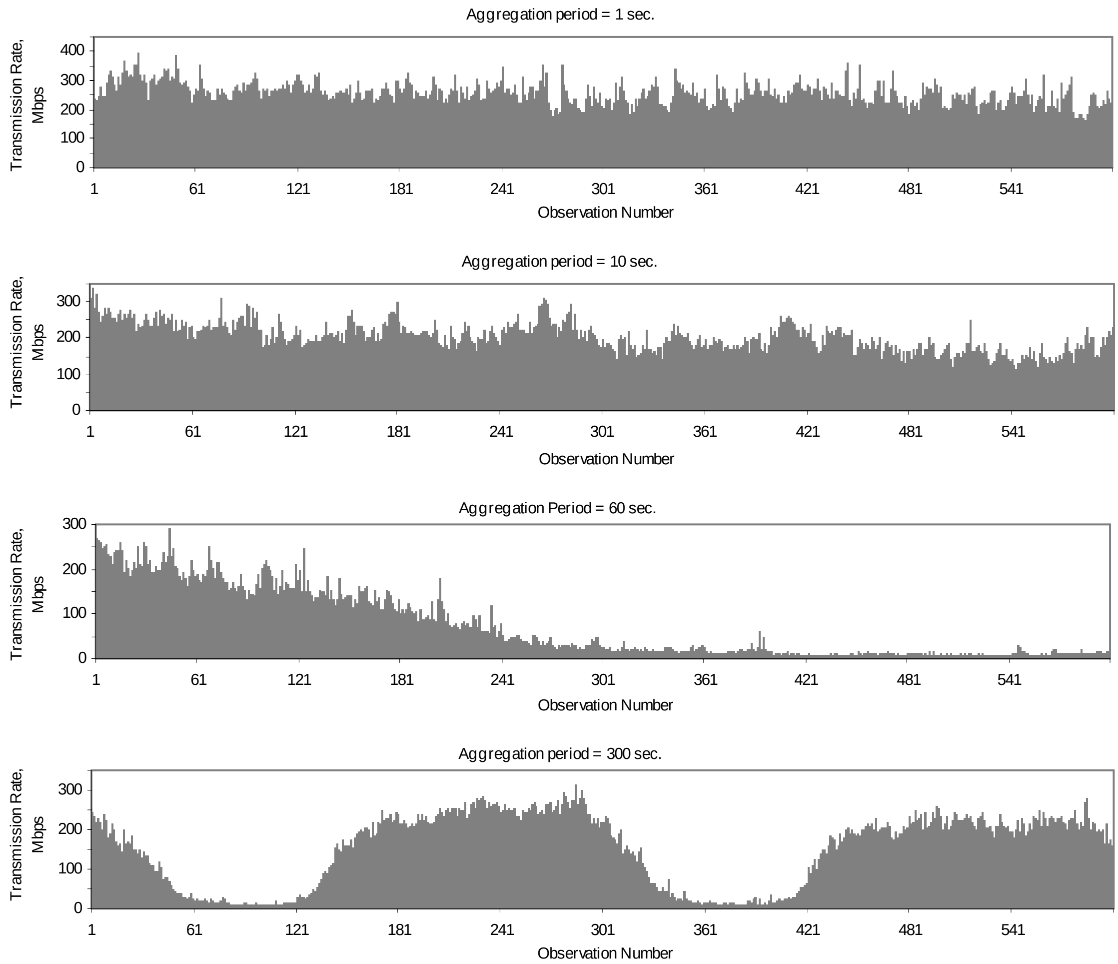

Figure 2.2. Measurements of real packet-switched traffic over different aggregation periods

Therefore, we can often speak about the possibility of applying traditional linear methods of time series forecasting.

The main accent of this research was put on the application of neural networks (i.e. a multilayer 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 is usually built into the operation of a neural network in implicit way when a static neural network is provided with dynamic properties [0]. In this case 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.

For a neural network to be dynamic, the memory must be given, 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. Short-term memory is usually build into the structure of a neural network 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.

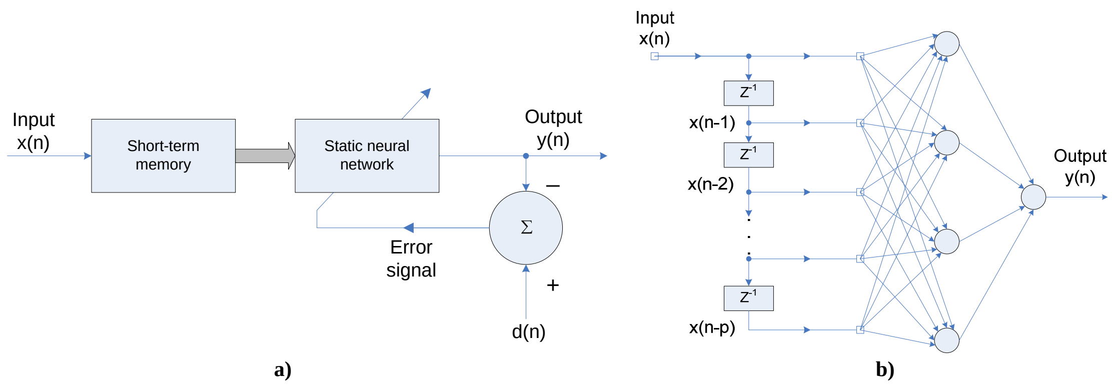

Figure 3.1. Temporal processing using neural networks:
a) nonlinear filter built on a static neural network [0]; b) time lagged feedforward network (TLFN) [0] [0]

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 $d(n)$ [0].

The structure shown on Figure 3.1(a) can be implemented at the level of a single neuron or a network of neurons. A time lagged feedforward network is shown on Figure 3.1(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 autoregression, 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 [0]:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D x(t) = \theta_q(B)\Theta_Q(B^s)\varepsilon(t), \qquad (3.1)$$

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 $\nabla_s = 1 - B^s$
$\varepsilon(t)$ – 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 [0]:

$$\begin{aligned} S(t) &= \alpha\cdot[x(t) - I(t-p)] + (1-\alpha)\cdot S(t-1) \\ I(t) &= \delta\cdot[x(t) - S(t)] + (1-\delta)\cdot I(t-p) \end{aligned}, \qquad (3.2)$$

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 [0]:

$$\hat{x}_t(l) = S(t) + I(t-s+l), \qquad (3.3)$$

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 [0].

4. Practical Research

The object of the research is the time series of different length and aggregation period which characterize the real traffic of both traditional telephone networks (POTS) and packet-switched IP-networks. The main aim was to analyse the statistical properties of time series and to develop such a neural network which is suitable for modelling an 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 [0].

The measurements were taken on the transportation level and represent three variables:

- the transmission rate of outgoing international traffic of the IP-network;
- the transmission rate of total outgoing traffic of the IP-network;
- the intensity of the total serviced load of the conventional telephone network.

Following the ITU-T Recommendation E.492 [0] the initial traffic measurements of each variable were averaged over 15-minutes and one-hour periods. The size of the basic sample was equal to 9 and 12 weeks for the first variable, and to 9, 12 and 18 weeks for two other variables. Thus, sixteen time series were produced in total. The forecasting horizon (i.e. the size of a testing sample) for each time series varied from one to 14 days with the step of one day.

All the analysed time series are characterized by the presence of seasonal components with periods of 24 hours and one week. It was revelled 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. The specification of the developed neural network is displayed in Table 4.1.

Table 4.1. The main parameters of the developed neural network[3]

Stage	Parameter / Procedure	Parameter Value / Procedure Description
Selection of network topology	Type of topology	Time-lagged feedforward 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 [0]
	Stopping criterion	Training error is invariable during 50 epochs
In-sample and out-of-sample evaluation	The parameters of in-sample evaluation	R, RMSE, MAE, MAPE, AIC, BIC
	Diagnostic testing of residuals	Lagrange multiplier type test [0] , χ^2 - test
	The parameters of out-of-sample evaluation	RMSE, MAE, MAPE, the Diebold-Mariano criterion

Neural networks belong to so called heuristic methods. It means that appropriate values of most parameters of the developed neural network had to be evaluated in experimental way. The architecture of

[3] Notes: 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.

a neural network was defined as follows. According to the universal approximation theorem [0] 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 appropriate 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 backpropagation 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 [0]. 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 quality of the forecasts was estimated by means of such standard parameters as root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). Besides, the Diebold-Mariano test [0] was applied in order to evaluate relative accuracy of forecasts and to reveal any statistically significant differences between the forecasts produced by neural networks and traditional linear methods such as a seasonal ARIMA and seasonal exponential smoothing. The main advantage of this test is that it 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, autocorrelated 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). They are 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 (1344 observations) are shown on Figure 4.1.

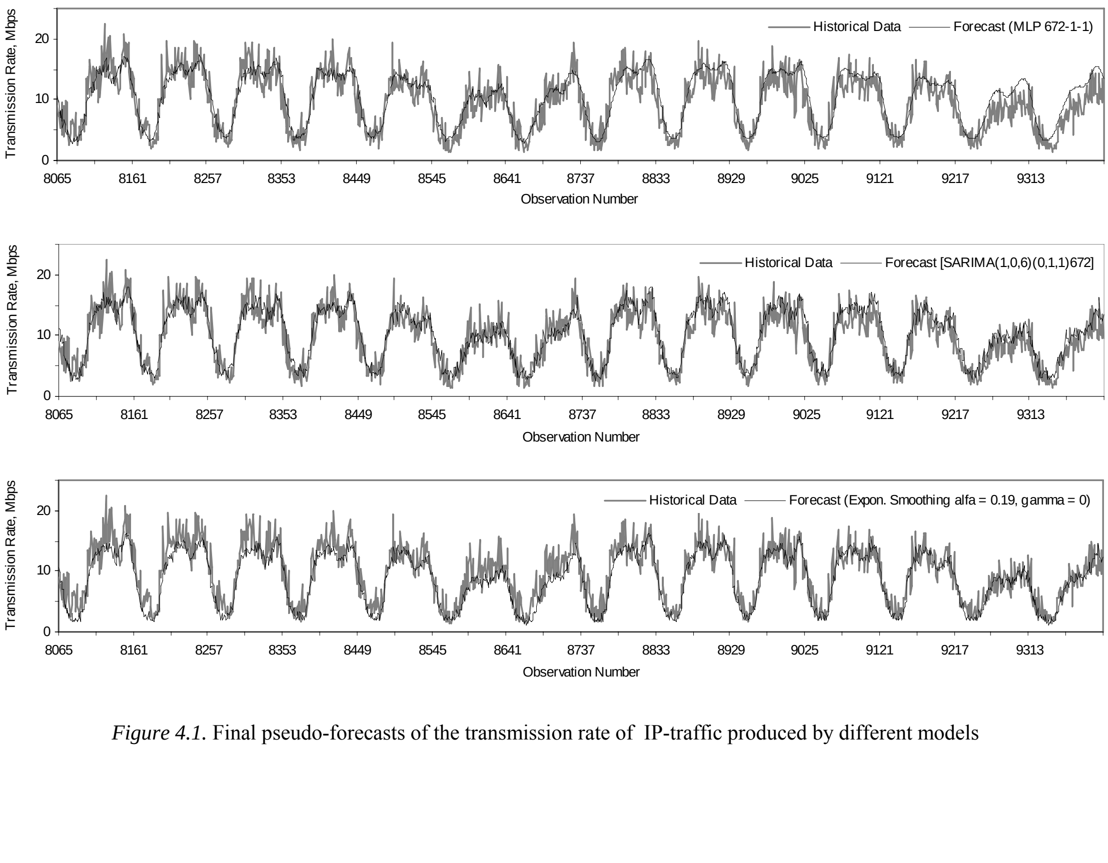

Figure 4.1. Final pseudo-forecasts of the transmission rate of IP-traffic produced by different models

We can see in Table 4.2 that the values of the standard estimates of the quality of produced forecasts do not differ significantly. Therefore it is hard to say which model performs better than others. The Diebold-Mariano test [0] was implemented in order to identify statistically significant differences between produced forecasts for three forecasting horizons such as 24 hours (96 observations), a week (672 observations) and two weeks (1344 observations). It is shown in Table 4.3 that there are 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 over the forecasting horizons of 24 hours and one week. Nevertheless, over the horizon of two weeks, the Diebold-Mariano test does not reveal any statistically significant differences between forecasts produced by a neural network and the model of exponential smoothing. We can also see that SARIMA outperforms seasonal exponential smoothing independently of a forecasting horizon. Therefore, in this particular case, it is reasonable to select the SARIMA model for further forecasting. It is a simpler and much less time-consuming method as compared to non-linear neural networks but provides relatively the same preciseness of forecasts.

Regarding other analysed time series, in most cases the comparison of forecasts produced by neural networks and linear models did not reveal any statistically significant differences.

Table 4.2. Standard estimates of the quality of pseudo-forecasts (forecasting horizon =1344 observations or 14 days)

Models \ Parameters	RMSE	MAE	MAPE, %
Neural Network	2.34	1.87	25.91
SARIMA	2.18	1.76	24.45
Seasonal Exponential Smoothing	2.30	1.75	20.48

Table 4.3. The evaluation of statistically significant differences between final pseudo-forecasts by means of the Diebold- Mariano test

Models \ Forecasting Horizon	96 obs. (24 h)		672 obs. (7 days)		1344 obs. (14 days)	
	DM	p_{DM}	DM	p_{DM}	DM	p_{DM}
Neural Network vs. SARIMA	-0.44	0.67	0.80	0.42	-4.83	0.00
Neural Network vs. Seasonal Exponential Smoothing	3.95	0.00	6.05	0.00	-0.75	0.45
SARIMA vs. Seasonal Exponential Smoothing	3.58	0.00	6.95	0.00	3.02	0.00

Notes: DM – the Diebold-Mariano statistics, p_{DM} – the significance level of a DM statistics

Conclusions

Both traditional linear methods and neural methods are accurate in producing short-term forecasts of traditional telephone traffic and packet-switched traffic. The results of the research show that in most cases the differences in quality between 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 reasonable.

It is important to keep in mind that the strong influence of fractal nature of packet-switched traffic is apparent only for the measurements taken on a very large-scale, usually over the periods up to 5–15 minutes. If according to the ITU-T recommendations the measurements of real network traffic are averaged over largest periods, the seasonal variations (due to the human behaviour) and monotonous trends (due to the influence of technical progress) are usually becoming those factors which affect the statistical properties of packet-switched traffic to a greater extent than self-similarity. Despite the fact that traffic traces can still poses the property of some "burstiness", the influence of long-range dependence is usually weakened. In this case, linear models can be applied with much success as well.

In the course of the research it was also revealed that a neural network can model and forecast seasonal time series without prior deseasonalization. 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 as well. 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.
3. *Self-Similar Network Traffic and Performance Evaluation / Eds. K. Park and W. Willinger.* Wiley-Interscience, 2000.
4. Sheluhin, O. I., Smolskiy, S. M., Osin, A. V. *Self-Similar Processes in Telecommunications*. John Willey & Sons, 2007.
5. *ITU-T Recommendation E.492.* Traffic Reference Period (Telephone Network and ISDN. Quality of Service, Network Management and Traffic Engineering). ITU, 1996.
6. Fowler, T. B. A Short Tutorial on Fractals and Internet Traffic, *The Telecommunications Review*, No 10, 1999, pp. 1–15.
7. Haykin, S. *Neural Networks: A Comprehensive Foundation, 2nd Ed.* Prentice Hall, 1998.
8. 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.
9. Jenkins, G. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.
10. Gardner, E. S. Jr. Exponential Smoothing: The State of Art, *Journal of Forecasting*, No 4, 1985, pp. 1–28.
11. Klevecka, I. An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks, *Scientific Journal of Riga Technical University. Telecommunications and Electronics,* No 8, 2009. (In print)
12. Cybenko, G. Approximation by Superpositions of Sigmoidal Function, *Mathematics of Control, Signals, and Systems*, No 2, 1989, pp. 303–314.
13. Diebold, F. C., Mariano, R. S. Comparing Predictive Accuracy, *Journal of Business & Economic Statistics,* Vol. 13, No 3, 1995, pp. 253–263.
14. Weigend, A. S., Rumelhart, D. E., Huberman, B. A. Generalization by Weight-Elimination with Application to Forecasting. In: *Proceedings of the Conference on Advances in Neural Information Processing Systems,* No 3, 1990, pp. 875–882.
15. Medeiros, M. C., Teräsvirta, T., Rech, G. Building Neural Network Models for Time Series: A Statistical Approach, *Journal of Forecasting*, No 25, 2006, pp. 49–75.

Received on the 1st of December, 2009