RIGA TECHNICAL UNIVERSITY
Faculty of Electronics and Telecommunications
Institute of Telecommunications

Irina KLEVECKA

Doctoral student of the study program “Telecommunications and Computer Networks”

NEURAL NETWORKS
FOR SHORT-TERM FORECASTING OF NETWORK TRAFFIC

Summary of Promotion Thesis

Scientific Supervisor
Dr.Sc.Ing. J. LELIS

Riga 2011

Klevecka I. Neural Networks for Short-Term Forecasting of Network Traffic. Summary of Promotion Thesis.-R.:RTU, 2011.-32 p.

Printed in accordance with resolution of Promotion Council RTU-P-08 of Riga Technical University, Minute No.1 dated February 22, 2011.

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».

PROMOTION THESIS
IS SUBMITTED TO RIGA TECHNICAL UNIVERSITY IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF ENGINEERING SCIENCES

The public defense of the promotion thesis, submitted for the degree of Doctor of Engineering Sciences, takes place at the Faculty of Electronics and Telecommunications of Riga Technical University, 12 Azenes St., Room 210, Riga, Latvia, on the 9th June of 2011 at 15:00.

OFFICIAL REVIEWERS

Dr.habil.sc.ing. V. Štrauss, Senior Scientist
Institute of Polymer Mechanics of University of Latvia, Riga, Latvia

Dr.habil.sc.ing. M.L. Šneps-Šnepe, Professor
Ventspils University College, Ventspils, Latvia

Dr.sc.ing. I. Jackiva, Professor
Transport and Telecommunication Institute, Riga, Latvia

CONFIRMATION

I confirm that the promotion thesis, submitted to Riga Technical University in fulfillment of the requirement for the degree of Doctor of Engineering Sciences, is my own work. The promotion thesis has not been submitted for a scientific degree to any other university.

Irina Klevecka…………………………….

Date: ……………………………………….

The promotion thesis is written in English and consists of two volumes. The first volume contains introduction, four chapters, main conclusions and recommendations and the list of bibliography, 174 pages in total. The second volume contains 96 appendices, 116 pages in total. The bibliography includes 207 sources.

GENERAL DESCRIPTION OF THESIS

- **Topicality of Subject Matter**

The *subject* of the promotion thesis is short-term network traffic forecasting by means of neural networks. *Forecasting* is a special research study aimed at the evaluation of future development of objects or phenomena. Forecasts have to precede strategic planning, assess the potential directions of development, and take into account the consequences of fulfilment or failure of plans.

There are three main parameters, the prediction of which plays an important role in design, optimization and performance of modern telecommunications networks. They are:

1) Traffic loads produced by users or subscribers,
2) Number of users / subscribers or telephone lines, and
3) Demand of inhabitants for telecommunications services.

These parameters are closely interrelated and their reliable forests are determined not only by accurate computing methods but also by financial capabilities of an operator. However, from both a theoretical and practical point of view, forecasts of traffic dynamics raise the most interest. It is also one of the strategic engineering tasks specified by ITU-T.

The topicality of the problem of forecasting lies in the fact that knowledge of network performance facilitates network management, in particular – helps to develop the algorithm of preventing an overload of transmission channels. Accurate and reliable forecasts allow planning the capacity of a network on time and sustain the required level of quality of service. Besides, the properties of network traffic directly influence both capital costs of equipment and expected income of an operator.

The emergence of packet-switched Internet networks as well as transformation of traditional telephone networks into multi-service systems provides new opportunities to a user in the sphere of his/ her activities. It has changed not only the architecture of a network but also statistical nature of teletraffic, which is now characterized by the effects of self-similarity and strong long-range dependence. Therefore, new approaches to the analysis and forecasting of states and parameters of packet-switched networks are strongly required. A *non-linear neural network* is one of these methods, which is rapidly gaining recognition in time series forecasting.

We can often hear that *neural networks are more art than science*. This is primarily due to the lack of a functional algorithm for applying neural networks to time series forecasting. Because of that, the active expert assessment is still necessary at all the stages of implementation, which prevents the automation of a forecasting process. It is also important to understand that, in contrast to some linear time series methods, neural networks have not originally been developed to meet the challenges of forecasting.

The attempts to predict the traffic loads of both a conventional telephone network and a packet-switched Internet network by means of neural networks have been made many times in the past. However, most of these research papers solve a trivial task of time series approximation, with more or less success, without taking into account a general theory of neural networks and time series forecasting. Thus, the *main problems* of applying neural networks to network traffic forecasting are:

- the absence of a consistent and efficient algorithm for applying the method of neural networks in time series forecasting;
- the absence of efficient and comprehensive criteria for selecting the final forecasting model and evaluating its quality.

- **Objects of Research**

The objects of this research are the time series of different lengths and aggregation rates, which describe:

- traffic of a conventional circuit-switched telephone networks (i.e. POTS);
- traffic of a packet-switched IP network.

The real measurements of IP network traffic were taken at the transport layer. All the measurements were brought to a form suitable for further statistical analysis and forecasting.

- **Main Goal and Tasks**

The main goal of the promotion thesis is to address and solve the problems related to the use of neural networks in time series forecasting, and, after *real data* testing and comparing the produced results, give practical recommendations on the effective application of neural networks and statistical models to network traffic forecasting.

The *main tasks* of the thesis are formulated as follows:

1) *Develop a functional algorithm* for implementing neural networks to solve a short-term forecasting task, which would provide a maximum *automation* of the process of selecting an optimal model and guarantees an appropriate quality of forecasts.
2) *Give practical recommendations* on
- selecting the models and methods to produce the operative forecasts (for 24 hours ahead) and short-term forecasts (for up to two weeks ahead);
- verifying a forecasting model;
- producing operative and short-term forecasts (or *ex ante* forecasts);
- assessing the accuracy of forecasts (or *ex ante* forecasts).
3) *Produce the empirical operative and short-term forecasts* of traffic of conventional telephone networks and packet switched IP networks by applying the method of neural networks, evaluate the accuracy of forecasts and compare them with the forecasts produced by traditional linear models and "naïve" methods. The topicality of this problem is mainly related to the fact that complexity of neural networks has provoked strong opinion about their advantages over simpler linear methods in solving a forecasting task. However, none of the scientific papers or books, known to the author, has accomplished a comprehensive comparison of the results produced by non-linear neural networks and traditional linear methods.

- **Hypothesis to Defend**

The author advances the hypothesis that:

- the proposed *algorithm of solving a forecasting task with neural networks* allows automating the identification of neural network solutions, which are capable of producing reliable forecasts;
- reliable operative and short-term forecasts can be produced by using not only non-linear neural networks, but also simpler linear models such as ARIMA and exponential smoothing, if an aggregation / sampling period of packet-switched network traffic is properly selected (usually over some minutes);
- neural networks outperform linear methods in the case of predicting traffic of real telephone networks, if the observations are taken over relatively small read-out periods (e.g., over 15-minute periods following ITU-T Recommendation E.492).

- **Main Methods of Research**

In order to fulfil the indicated tasks, the following main methods were applied:

1) *Neural networks* – the method of artificial intelligence and universal approximator, which allows revealing non-linear dependencies of stochastic processes.
2) *Autoregressive integrated moving average models* (ARIMA) and *exponential smoothing* – the classic linear methods of time series forecasting, which are useful for modelling and forecasting short-range dependent processes.
3) *Spectral analysis* helps to identify the periodic and quasi-periodic components of time series. It is also useful in evaluating the Hurst exponent.
4) *Correlation analysis* allows identifying the statistical dependencies between members of a time series taken with a time shift (autocorrelation) or statistical bonds between two processes (cross-correlation).
5) *Regression analysis* is helpful in identifying the reliable model for a trend component, if any.
6) *The methods of non-parametric statistics* are useful in testing the hypothesis of stationarity or normality of time series in the absence of clear knowledge about the probability distribution of a process.

- **Structure**

The thesis consists of two volumes. The first volume contains four main chapters:

- *Chapter 1* examines the main aspects and prerequisites of network traffic forecasting.
- *Chapter 2* describes in detail the methods applied in the practical studies, such as non-linear neural networks, ARIMA models, exponential smoothing models and “naïve” methods.
- *Chapter 3* highlights the main contribution of the author to the theory of time series forecasting. The description of the proposed advanced algorithm and the main aspects of its practical realization are given in detail.
- *Chapter 4* analyzes the results of practical studies.

The first volume consists of 174 pages and contains 25 figures and 11 tables. The list of cited literature and other sources includes 207 bibliographic names. The second volume includes 96 annexes and consists of 116 pages.

- **Novelty**

The *novelty* of the thesis is attributed to the following original results:

Based on the methods of mathematical statistics and theory of neural networks, there has been developed an advanced algorithm aimed at solving the task of short-term traffic forecasting by means of neural networks.

In comparison to most classic schemes of setting a neural network to fulfil a certain task (see, for example, [9; 53, p. 84]), the proposed algorithm focuses on the estimation of forecasting abilities rather than approximation accuracy, allows automating the process of determining an optimal solution and includes three procedures, which are:

- the use of multiple cycles of weight initialization of a neural network during a training process;
- the procedure of selecting the intermediate forecasting model, taking into account the residual autocorrelation and the estimates of information criteria;

- the procedure of selecting the final forecasting model, taking into account the accuracy of *ex ante* forecasts.

The incorporation of these procedures into a classic algorithm allows identifying neural network solutions in a more effective way and significantly improves the accuracy and reliability of produced forecasts.

Neural networks belong to *heuristic* methods. It implies that the identification of optimal parameters of architecture and learning should involve intensive test-and-trial procedures, since these parameters do not comply with strict mathematical rules. The algorithm, proposed in the thesis, is also based on the experience of the author and represents the results of years of the practical studies of real time series. However, the motivation of including these procedures into the algorithm arises from the theory of neural networks and time series forecasting.

For the first time, there has been conducted a profound statistical analysis of real network traffic forecasts, produced by non-linear neural networks, classic linear methods and "naïve" methods.

Statistically significant differences in accuracy have not been identified between the forecasts produced by neural networks and classic linear models in most cases tested by the Diebold-Mariano criterion [6]. The analyzed time series are typical for these categories of traffic loads. It means that linear statistical methods would produce reliable and accurate operative / short-term forecasts for many other time series with similar statistical properties as well.

- **Practical Significance**

The proposed algorithm and practical recommendations can be applied to produce operative and short-term forecasts of telephone network traffic as well as packet-switched Internet traffic generated at the transport and application layers. In turn, the produced forecasts can be useful in planning the capacity of transmission channels, thereby providing the required level of quality of service (QoS).

The algorithms and recommendations developed by the author can be applied to other time series with similar statistical properties – for example, in producing predictions of power consumption or road traffic.

- **Approbation**

The proposed algorithm and recommendations were successfully applied to produce the forecasts of several time series describing the traffic of real telecommunications networks. Operative and short-term traffic forecasts have been obtained for:

- the circuit switched telephone network of *Augstceltne SIA,* which specializes in maintenance of corporate customers;
- the packet switched IP network of *INBOKSS SIA,* which specializes in providing free e-mail services.

The results of the thesis were declared and discussed at the following scientific conferences:

1) The 9th International Conference *Reliability and Statistics in Transportation and Communication* (RelStat`09), Riga, Oct. 21-24, 2009. Topic of presentation – *Forecasting Network Traffic: A Comparison of Neural Networks and Linear Models.*
2) The 50th International Scientific Conference of Riga Technical University, Riga, Oct.14-16, 2009. Topic of presentation– *An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks.*

3) The 28th Annual International Symposium on Forecasting, Nice (France), June 22-25, 2008. Topic of presentation – Preprocessing *of Input Data of Neural Networks: The Case of Predicting Telecommunication Network Traffic.*

The other reports at scientific conferences associated with the subject of the thesis:

4) The 18th European Regional Conference of the International Telecommunications Society, Istanbul (Turkey), Sep. 4-6, 2007. Topic of presentation – *Perspective Evaluation of the Electronic Communications Market in Latvia*.
5) The 6th International Conference *Reliability and Statistics in Transportation and Communication* (RelStat'06), Riga (Latvia), Oct. 25-28, 2006. The theme of presentation – *Forecasting Methods and Long-term Evaluation of the Electronic Communications Market in Latvia.*
6) The 17th European Regional Conference of the International Telecommunications Society. Amsterdam (Netherlands), Aug. 22-24, 2006. The theme of presentation – *New Technologies and their Influence on the Universal Service Policy.*
7) The 16th European Regional Conference of the International Telecommunications Society (ITS Europe 2005), Porto (Portugal), Sep. 4-6, 2005. The theme of presentation – *The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.*
8) International Conference *Reliability and Statistics in Transportation and Communication* (RelStat'04), Riga (Latvia), Oct.14-15, 2004. The theme of presentation – *Financial Risk of Providing the Universal Telecommunications Service in Latvia*.

The papers in peer-reviewed scientific journals:

1) Klevecka, I. "Forecasting Traffic Loads: Neural Networks vs. Linear Models." *Computer Modelling and New Technologies* 14.2. (2010): 20–28. [ISSN 1407-5806; Thomson Reuters Researcher ID]
2) Klevecka, I. "An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks". *Scientific Journal of Riga Technical University* (*Series "Telecommunications and Electronics"*) 9 (2009): 48-55. [ISSN 1407-8880; EBSCO Host, ProQuest, VINITI]
3) Klevecka, I., and J. Lelis. "Pre-Processing of Input Data of Neural Networks: The Case of Forecasting Telecommunication Network Traffic." Spec. issue of *Telektronikk: Telecommunications Forecasting* (in co-operation with International Institute of Forecasters) 104.3/4 (2008): 168-178. [ISSN 0085-7130; Thomson Reuters Researcher ID, ACM Digital Library]
   Cited in:
   - Nikolov, V., and V. Bogdanov**. "**Integration of Neural Networks and Expert Systems for Time Series Prediction." *Proceedings of the 11th International Conference on Computer Systems and Technologies* (*CompSysTech'10*). New York: ACM Press, 2010. 534-539. [ACM Digital Library]
   - Chulaka Gunasekara, R., et. al. „Prophetia: Artificial Intelligence for TravelBox® Technology." *Advances in Computational Intelligence*. Eds. Wen Yu and Edgar N. Sanchez. Berlin: Springer-Verlag, 2009. 21-34. [Springer Link]

The abstracts in the proceedings of international scientific conferences:

4) Klevecka, I. "Forecasting Network Traffic: A Comparison of Neural Networks and Linear Models." *Abstracts of the 9th International Conference "Reliability and Statistics in Transportation and Communication"* (*RelStat'09*). Oct. 2009, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2009. 36. [ISBN 978-9984-818-22-1; Thomson Reuters Researcher ID]
5) Klevecka, I., and J. Lelis. „Preprocessing of Input Data of Neural Networks: The Case of Predicting Telecommunication Network Traffic." *Program and Abstracts of the 28th Annual*

International Symposium on Forecasting. June 2008, Nice, France. France: International Institute of Forecasters, 2008. 29. [ISSN 1997-4116; Thomson Reuters Researcher ID]

The other papers in peer-reviewed scientific journals associated with the subject of the thesis:

6) Klevecka, I., and J. Lelis. “Application of Extrapolation Methods to the Technology Diffusion Forecasting.” *Scientific Proceedings of Riga Technical University (Series "Telecommunications and Electronics")* 7 (2007): 52-59. [ISSN 1407-8880; ProQuest, VINITY]
7) Klevecka, I. and J. Lelis. “Financial Risk of Providing the Universal Telecommunications Service in Latvia.” *Transport and Telecommunication* 6.1 (2005):139-144. [ISSN 1407-6160; Thomson Reuters ResearcherID].

The papers in the proceedings of scientific conferences associated with the subject of the thesis:

8) Klevecka, I., and J. Lelis. “Perspective Evaluation of the Electronic Communications Market in Latvia.” *Proceedings of the 18th European Regional ITS Conference.* Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 1-37. CD-ROM. [Thomson Reuters ResearcherID].
9) Klevecka, I., and J. Lelis. “Forecasting Methods and Long-term Evaluation of the Electronic Communications Market in Latvia.” *Proceedings of the 6th International Conference “Reliability and Statistics in Transportation and Communication”*. Oct. 2006, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2006. 37-45. [ISBN 9984-9865-9-4; Thomson Reuters Researcher ID].
10) Klevecka, I., J. Lelis, and J. Ulmanis. “The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.” *Papers of the 16th European Regional ITS Conference*. Sep. 2005, Porto, Portugal. Berlin: International Telecommunications Society, 2005. 1-13. CD-ROM. [Thomson Reuters Researcher ID].

The other abstracts in proceedings of international scientific conferences associated with the subject of the thesis:

11) Klevecka, I., and J. Lelis. “Perspective Evaluation of the Electronic Communications Market in Latvia.” *Abstracts of the 18th European Regional ITS Conference*. Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 118-119. [Thomson Reuters Researcher ID]
12) Klevecka, I., J. Lelis, R. Bergmanis, and G. Macs. “New Technologies and their Influence on the Universal Service Policy.” *Abstract Booklet of the 17th European Regional ITS Conference*. Aug. 2006, Amsterdam, Netherlands. Berlin: International Telecommunications Society, 2006. 59-60. [ISBN 90-8559-205-4, Thomson Reuters Researcher ID]
13) Klevecka, I., J. Lelis, and J. Ulmanis. “The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.” *Abstracts of the 16th ITS European Regional Conference.* Sep. 2005, Porto, Portugal. Berlin: International Telecommunications Society, 2005. 16-17. [*Thomson* Reuters Researcher ID]
14) Klevecka, I., and J. Lelis. “Financial Risk of Providing the Universal Telecommunications Service in Latvia.” *Programme and Abstracts of the International Conference RelStat’04.”* Oct. 2004, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2004. 54. [ISBN 9984-668-76-2; Thomson Reuters Researcher ID].

SYNOPSIS OF THESIS

CHAPTER 1 MAIN ASPECTS OF NETWORK TRAFFIC FORECASTING

Chapter 1 examines the aspects of applying the models and methods of time series theory to network traffic forecasting. Main statistical properties and genesis of time series are discussed. The chapter also contains a review of scientific papers dedicated to traffic forecasting carried out by means of neural networks.

- **Concept of Time Series**

In modeling and forecasting we usually assume that network traffic is represented as time series. A *time series* is a time-ordered sequence of observation values of a physical variable, usually made at equally spaced time intervals Δt, represented as a set of discrete values $x(t_1), x(t_2), \mathbf{K}\, x(t_N)$. In statistical analysis, this sequence of *N* observations is often considered as a sample taken from a longer general sequence of random numbers. The observations or elements of time series are typically labeled in accordance with a time moment they refer to (e.g., x_1, x_2, x_3). Thus, the order of the elements of a time series is of great importance.

It is necessary to keep in mind that, unlike the observations of random variables, the elements of time series are not statistically independent [49, p.780]. Some rules and properties of the statistical analysis of random samples cannot be applied to time series, and this requires the implementattion of specific methods and approaches. On the other hand, the correlations between time series observations set up a specific base for predicting an analyzed variable, i.e. for producing the estimate $\hat{x}(N+L)$ of an unknown value $x(N+L)$ taking into account the historical values $x(t_1), x(t_2), \mathbf{K}\, x(t_N)$, where *N* is the length of an analyzed time series and *L* is a forecasting horizon.

The *genesis of observations* is the structure and classification of the main factors, under the influence of which the values of time series are formed. There are four types of such factors and components of time series [49, p.781; 52, p.242; 57, p.354]:

1) *Long-term factors* form the general dynamic tendency of an analyzed parameter $x(t)$. This tendency is usually described by a deterministic non-random function called ***a trend***.
2) *Seasonal factors* determine the tendency, which changes regularly during a certain period (a day, week, month etc.). Since this function has to be periodic (with periods, proportional to "seasons"), its analytical expression involves the use of trigonometric functions.
3) *Cyclic factors* determine the longer periods of relative rise and fall. The cyclic component may contain the cycles of economic, demographic or astrophysical nature, and varies in amplitude and length. As a rule, the length of a cyclical component exceeds one year.
4) *Random (irregular) factors* determine the stochastic nature of time series members. A random component is formed as a result of superposition of many external factors, which are not involved in the formation of a deterministic component.

The deterministic components of network traffic can be classified as follows [19, p.42; 45, p.44]:

1) *24-hour cycle*. It has been known for a long time that sigmoidal models (logistic model, Gompertz model, etc. [22; 23]) are optimal for describing the traffic dynamics within a day [58, p.202].
4) *Weekly cycle* is usually characterized by the decrease of traffic during weekends, and can be described by means of a Fourier series.
5) *Annual cycle*. It is believed that the level of network traffic is higher at the beginning of a month, after a festival season and at the beginning of each quarterly period.

6) *Linear trend*. The overall traffic increases year by year due to the influence of technical progress and socio-economic factors.

If we identify accurately a deterministic component, then the residuals of a time series will be an irregular *stochastic component*. Its behavior cannot be fully predicted in advance. In other words, every observation gives only one option among many possible. In order to describe and predict this component of a time series, the methods and concepts of probability theory are involved.

- **Aspects of Forecasting Transmission Capacity of Packet Switched Networks**

Transmission capacity is one of the most important parameters characterizing the quality of networks. For the purposes of forecasting, the measurements made at the *transport layer* are usually considered. These time series describe either the number of arriving packets or the level of traffic / transmission rate measured in bytes over discrete time intervals.

The methods of network traffic forecasting are partially determined by ITU-T Recommendations E.506 [16] and E.507 [18]. Even these recommendations have been developed for ISDN networks, some of the forecasting methods described there can be still applied to modern telecommunications networks. These methods are the autoregressive integrated moving average models (ARIMA) [3] and exponential smoothing [12].

The traffic of packet switched IP networks is characterized by such statistical effects as self-similarity and long-range dependence [20, p.150; 28; 37; 41; 56, p.53]. The stronger post-effects, the longer is a forecasting horizon, for which reliable forecasts can be produced. However, it is also a disadvantage, since the estimation and selection of an adequate model, which would take into account all significant correlations between the members of a time series, becomes labor-intensive [40].

At the same time, there has been disseminated intensively the *myth* regarding impossibility of applying traditional linear methods to predicting packet switched IP traffic and the necessity to use more complicated non-linear methods such as neural networks.

Indeed, neural networks offer some additional opportunities in modelling non-linear processes and recognizing chaotic behaviour. Owing to their great flexibility, these networks can recognize a variety of structures. However, numerous practical studies dedicated to traffic forecasting usually miss the fact that fractal properties of packet-switched traffic have a significant influence on a forecasting process only in the case of measurements *on a very large scale* – over the aggregation periods varying from milliseconds to some minutes. This fact has been confirmed by the author’s practical studies as well as a number of other research papers [36, 39, 40].

Fig. 1 is the illustration of this idea, where the real measurements of transmission rate are shown against an aggregation period. A visual analysis of the traffic aggregated over 1 and 10 seconds reveals *stochastic self-similarity* [37]. However, increasing the aggregation period up to one minute, and then – up to five minutes, we can see that a traffic trace becomes more even, its variance significantly decreases and the influence of deterministic components starts to play a leading role.

From the point of view of time series forecasting, a very fine sampling scale is unreasonable. In this case, the selection of a relevant statistical model is complicated due to the strong influence of autocorrelations between distant observations of times series as well as extraneous noises and anomalous outliers, which unavoidably accompany the large-scale measurements. Besides, an aggregation / sampling period also determines a forecasting horizon, for which reliable forecasts can be produced. In other words, a potential forecasting horizon for time series, aggregated over, for example, one-second periods is different from the one for time series aggregated over 24-hour intervals.

At present, real-time forecasting with neural networks is hard to implement in practice. Apart from the necessity to select and evaluate many parameters, often – in empirical way, some substantial time resources are necessary for training a neural network. Therefore, taking into account ITU-T Recommendation E.492 [17], it is desirable to average measurements of network traffic over 15-minute and / or one-hour read-out intervals. In doing so, the main factors determining the statistical structure of *real* network traffic are seasonal effects and monotonous trends, the main reasons of which are human behaviour and technical progress (see Fig. 2).

It has been shown in the practical part of the thesis that statistical properties of such time series become similar to statistical properties of traditional voice traffic. At an intuitive level, it gives us the opportunity to assume that the methods of modelling and forecasting of these processes would be similar as well, if the appropriate length of a read-out period is selected.

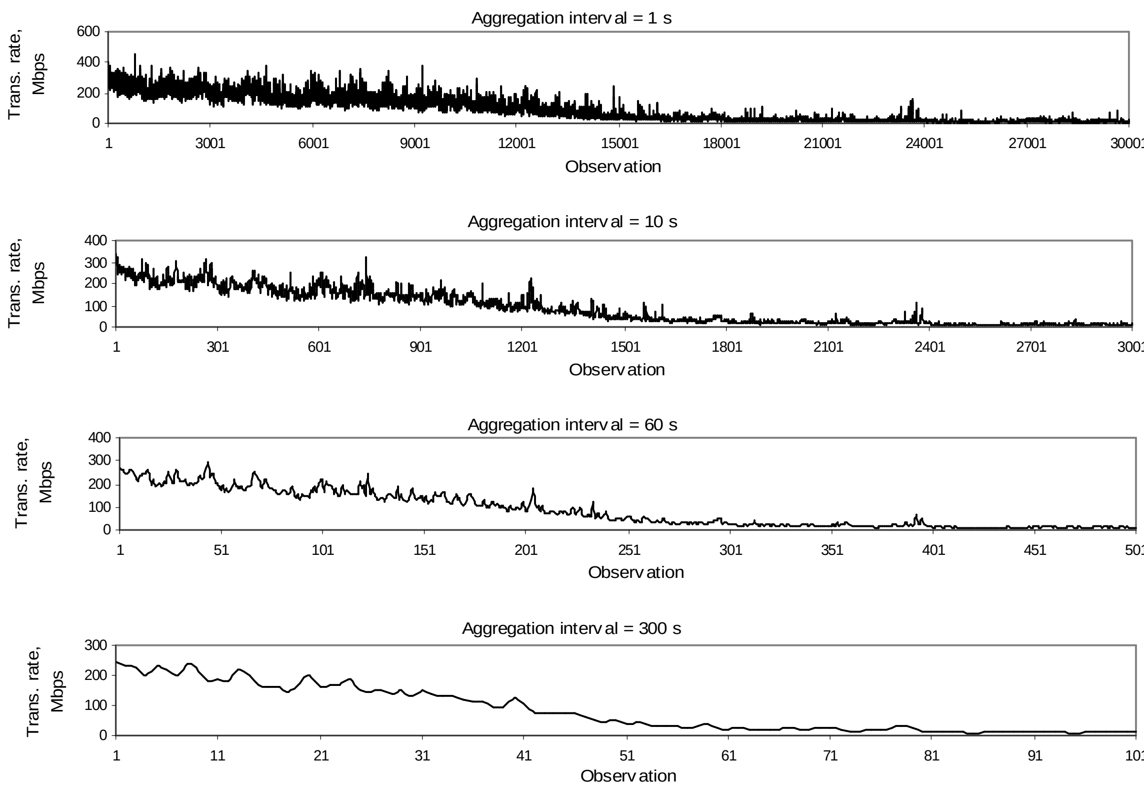


Fig. 1 Real packet-switched traffic recorded over different aggregation periods

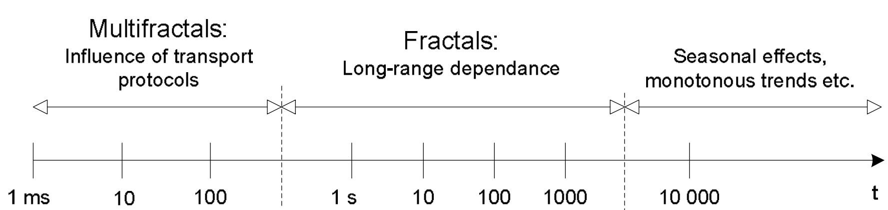


Fig. 2 Statistical effects of packet-switched traffic depending on a time scale [10, with author's amendments]

The main accent of this thesis has been put on the application of neural networks (i.e., a multilayer perceptron) to forecasting the traffic of both a traditional telephone network and a packet-switched IP network. Following the principle of Ockham's razor – *choose a parsimonious model,* – it makes sense to compare the accuracy of forecasts produced by means of non-linear models with those produced by traditional linear models. To pursue this goal, the models of ARIMA and exponential smoothing (as the methods recommended by the ITU-T) as well as "naive" methods, have been chosen. If there are no statistically significant differences between the forecasts produced by neural networks and linear methods, then the application of such a complicated and time-consuming method as neural networks becomes unnecessary.

CHAPTER 2 METHODS OF NETWORK TRAFFIC FORECASTING

Chapter 2 gives an overview of the models and methods of network traffic forecasting applied in the practical part of the thesis such as non-linear neural networks (multilayer perceptron), seasonal autoregressive integrated moving average (SARIMA), seasonal exponential smoothing and "naive" forecasting methods.

- **Neural Networks**

A *neural network* is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use [13, p.2]. The development of *artificial neural networks* started in the beginning of the 20^{th} century but only in the nineties, when some theoretical barriers were overcome and computing systems became powerful enough, neural networks have gained wide recognition. Even though neural networks can be implemented as fast hardware devices (and these realizations do exist in real life), most practical studies are performed by applying software simulations on conventional PCs. Software simulations provide low-cost flexible environment, which is sufficient for many real-life applications. Neural networks are acquiring popularity in the field of telecommunications as well, where they help solving various problems such as switching management, traffic management, routing, channel allocation in mobile transmission systems, etc. [54, p. 60].

The incorporation of time into the operation of a neural network allows to follow statistical variations in different processes described by time series, such as speech signals, radar signals, fluctuations in stock market processes, teletraffic processes and many others. The discussion of the role of time in neural processing can be found in fundamental paper [7].

The temporal structure of an analyzed sample is usually built into the operation of a neural network in *implicit* way. In this case, a *static* neural network (e.g. a multilayer perceptron) is provided with *dynamic* properties [13, p.636]. For a neural network to be dynamic, it must be given memory which may be divided into short-term and long-term memory [35]. ***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 built into the structure of a 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.

Two types of neural networks, a back-propagation network (multilayer perceptron) and a radial basis function network, are considered to be suitable for temporal processing. Due to a number of reasons, the latter has not gained acceptance[1]. At the same time, numerous practical studies have proved that a multilayer perceptron solves successfully many various tasks such as pattern recognition, regression, function approximation, time series forecasting, cluster analysis, etc. Therefore, further we will focus on this class of neural networks.

A multilayer perceptron usually consists of multiple sensor elements (i.e., input nodes) forming an input layer, one or several hidden layers containing computational nodes, and one output layer. It is often trained according to the *error back propagation algorithm*. It is a supervised training algorithm, which is based on the *error correction training rule* and requires two computational flows – a direct one and a backward one – through all the layers of a network.

Temporal pattern recognition demands 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

[1] Both a *radial-basis function* (RBF) network and a *multilayer perceptron* (MPP) belong to the class of universal approximations. Due to that, there always exists an RBF network capable of accurately mimicking a specified MLP, and vice versa. An MLP develops global approximations to nonlinear input-output mapping. In turn, an RBF network constructs local approximations using exponentially decaying localized nonlinearities (e.g., Gaussian functions). The latter is the reason of the popularity of MLPs – in order to approximate a nonlinear input-output mapping at the same degree of accuracy, an MLP requires the smaller number of parameters to determine and consequently, less time for completing a training cycle, in comparison with an RBF network [13, p. 293].

its past values. Fig. 3 shows the diagram of a nonlinear filter built on a static neural network. Given a specific input signal consisting of the current time series value $x(t)$ and the Ω past values $\{x(t-1), x(t-2)\,\mathbf{K}, x(t-\Omega)\}$ stored in a delay line memory of order Ω, the free parameters are adjusted to minimize the training error between the output of the network, $y(t)$, and the desired response, $d(t)$ [13, p.645]. The structure shown in Fig. 3 can be implemented at the level of a single neuron or a network of neurons.

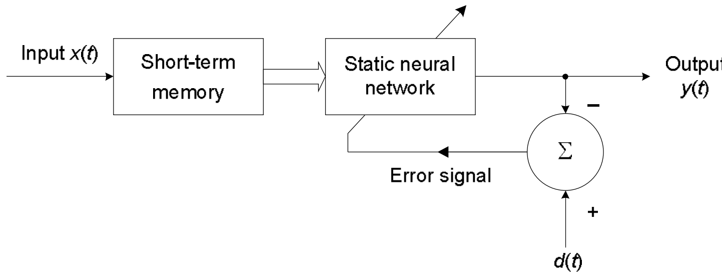


Fig. 3 Temporal processing – nonlinear filter built on a static neural network [13, p.643]

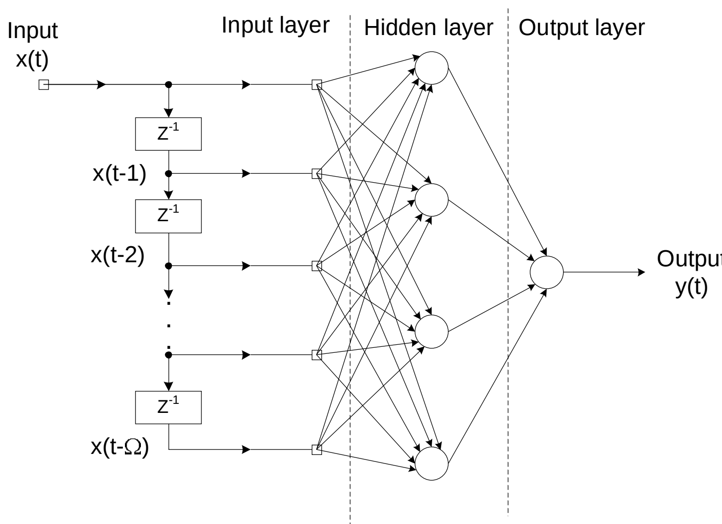


Fig. 4 Time lagged feed-forward network[2] [13, p.644; 35]

The diagram of a *time lagged feed-forward network* is shown in Fig. 4. It is a powerful nonlinear filter consisting of a tapped delay memory of order Ω and a multilayer perceptron. The standard back propagation algorithm can be used to train this type of neural networks. At time *t*, the temporal pattern applied to the input layer of the network is the signal vector:

$$x(t) = \{x(t), x(t-1), x(t-2)\,\mathbf{K}, x(t-\Omega)\}^{N^{TR}}, \tag{1}$$

where
N^{TR} – the length of a time series or a training subset.

Eq. (1) describes the state of the nonlinear filter at time *t*. One training *epoch* consists of a sequence of patterns (states), the number of which is determined by the memory order Ω and the size of a training sample N^{TR}. The output of a nonlinear filter, assuming that a multilayer perceptron has a single hidden layer and one output neuron, is computed from:

[2] The bias levels are omitted for convenience of representation.

$$y(t) = j_2\left(b_j + \sum_{j=1}^{m_1} w_j j_1\left(\sum_{z=0}^{\Omega} w_{j,z}\, x_{t-z} + b_{j,z}\right)\right), \tag{2}$$

where

j_1 – activation function of a hidden layer;

j_2 – activation function of an output layer;

$w_{j,z}$ – synaptic weight of input synapse ζ of hidden neuron j;

w_j – synaptic weight of input synapse j of an output neuron;

$b_{j,z}$ and b_j – biases;

m_1 – number of hidden neurons;

Ω – order of linear delay memory.

- **Autoregressive Integrated Moving Average Models**

One of the most popular class of *linear time series models* refers to *autoregressive moving average models* (ARMA), including purely autoregressive (AR) and purely moving-average (MA) models as special cases. These models were initially introduced in the twenties of the last century but have been using actively only since 1970, when the fundamental book of Box and Jenkins [3] was published.

In real-life network traffic modelling and forecasting, it is necessary to put an accent on the *seasonal modifications* of linear models, which are able to model and forecast the periodic time series. The application of non-seasonal models to seasonal time series can lead to the erroneous conclusions that linear models are not capable to model and make reliable forecasts of network traffic dynamics.

The *seasonal autoregressive integrated moving average model*, denoted as $SARIMA(p, d, q)(P, D, Q)_s$, is given by [3, p.305]:

$$f_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D x_t = q_q(B)\Theta_Q(B^s)e_t, \tag{3}$$

where

s – period of the seasonal component;

p – order of the non-seasonal autoregressive operator;

q –order of the non-seasonal moving average operator;

d –order of the non-seasonal differencing operator;

P – order of the seasonal autoregressive operator;

Q – order of the seasonal moving average operator;

D – order of the seasonal differencing operator;

$\nabla = \nabla_1 = 1 - B$ – non-seasonal differencing operator;

$\nabla_s = 1 - B^s$ – seasonal differencing operator;

$f(B)$ и $q(B)$ – polynomials in B of order p и q, respectively, which satisfy the conditions of stationarity and invertibility;

$\Phi(B^s), \Theta(B^s)$ – polynomials in B^s of order P и Q, respectively, which satisfy the conditions of stationarity and invertibility;

$e_t \sim WN(0, s^2)$– a white noise process.

Let us assume that all the values of a time series $x_t, x_{t-1}, \mathbf{K}$ are known until time moment t. Then, the minimal mean squared error forecast $\hat{x}_t(L), L \geq 1$ at lead time L and origin t is the *conditional expectation* of x_{t+L} [3, p.306]:

$$\hat{x}_t(L) = [x_{t+L}] = E[x_{t+L} \mid q, \Theta, x_t, x_{t-1}, \mathbf{K}] \tag{4}$$

Box and Jenkins proved that a forecast of the minimal mean squared error can be calculated directly from the model represented as a *difference equation*. For example, for the seasonal process of period $s = 12$, the forecast is given by [3, p.306]

$$\hat{x}(L) = [x_{t+L}] = x_{t+L-1} + x_{t+L-12} - x_{t+L-13} + e_{t+1} - \theta\, e_{t+L-1} - \Theta e_{t+L-12} + \theta\, \Theta e_{t+L-13} \tag{5}$$

After inserting the values of parameters θ и Θ into Eq. (5), we immediately obtain a minimal mean squared error forecast at lead time *L* calculated at origin *t* [3, p. 307]. The parameters of Eq. (5) are assumed to be known precisely, and a time series $x_t, x_{t-1}, \ldots$ is assumed to extend into the remote past.

- **Exponential Smoothing**

The method of exponential smoothing allow to produce the description of a process, according to which the last historical observations have the larger weights as compared to the earlier ones, and the weights decrease exponentially.

The simple exponential smoothing model is defined as [12]:

$$S_t = S_{t-1} + \alpha\, e_t \tag{6}$$

where

S_t – smoothed level of the series computed after x_t is observed;

α – smoothing parameter for the level of the time series;

e_t – one-step-ahead forecast error; $e_t = x_t - \hat{x}_{t-1}(1)$;

x_t – observed value of the time series at moment *t*;

$\hat{x}_{t-1}(1)$ - one-step-ahead forecast from origin $t-1$;

There exist different modifications of exponential smoothing aimed at the analysis of non-stationary time series with linear and nonlinear trend components, as well as seasonal time series with multiplicative and additive seasonality. Only the models with additive seasonality and / or a linear trend will be considered further in this section as they are the most suitable for the analysis of teletraffic processes. In exponential smoothing models, the additive seasonal component and linear trend of a time series are calculated from [12]:

$$I_t = I_{t-s} + \delta(1-\alpha)e_t \tag{7}$$

$$T_t = T_{t-1} + \alpha\,\gamma\, e_t \tag{8}$$

where

I_t – smoothed value of the seasonal component at the end of period *t*;

T_t – smoothed value of the trend component at the end of period *t*;

δ – smoothing parameter for the seasonal component;

γ – smoothing parameter for the trend component.

The estimation of a time series incorporating seasonal and / or trend components, is based on *decomposition*. The seasonal and trend components are estimated at each time moment independently by using a simple exponential smoothing model with parameters δ and γ. This method is known as the *Holt-Winters' exponential smoothing* and is based on three smoothing equations – one for the level, one for trend and one for seasonality.

For the time series with additive seasonality and a linear trend, the forecast for horizon *L* is given by [12]:

$$\hat{x}_t(L) = S_t + I_{t-s+L} + L \cdot T_t \tag{9}$$

where

$\hat{x}_t(L)$ – forecast produced for horizon L from origin t.

The complete classification as well as the description of the methods of determining the optimal parameters of exponential smoothing models is given in [12].

- **Naïve Forecasting**

It is desirable to compare the results, produced by various forecasting models, with so called *naïve fore*casts. In practical analysis of time series, a naïve forecast serves as the simplest *benchmark* forecast and can be produced in several different ways, of which the following two were applied in the thesis:

1) The naïve forecast is the sample mean of an examined time series [33]:

$$\hat{x}_t^{\,a}(L) = E[x_{t+L} \mid x_t, x_{t-1}, \mathbf{K}, x_1] = \hat{m} \text{ for } L \geq 1 \tag{10}$$

2) The seasonal naïve forecast can be used with seasonal data and postulates that the forecast for one period ahead is equal to the same value of the last historical period of a time series [44]:

$$\hat{x}_t^{\,b}(L) = E[x_{t+L} \mid x_t, x_{t-s}, \mathbf{K}, x_1] = x_{t+L-s} \text{ for } L \geq 1 \tag{11}$$

If the comparison of forecasts produced by neural networks or statistical models with those produced by naïve methods does not reveal any statistically significant differences, then, perhaps, the use of the models in further forecasting of this particular time series is not required.

CHAPTER 3 ALGORITHM FOR SOLVING A FORECASTING TASK WITH NEURAL NETWORKS

Chapter 3 describes the author's developed algorithm aimed at solving a time series forecasting task by means of neural networks. The innovative aspects of the algorithm are considered in detail. The chapter examines the methods of determining the parameters of multilayer perceptrons and provides some recommendations on the practical use of the algorithm, e.g., the preparation and pre-processing of input data, development of forecasts, estimation of the accuracy of forecasts, etc.

As mentioned above, unlike classic linear methods, the method of neural networks was not initially aimed at modelling and forecasting time series. When applied to time series forecasting, neural networks are often criticized for the necessity to set many different parameters through test-and-trial procedures, complications with producing and replicating a stable solution, high probability of over-learning, high demands for time resources and computational capacities. Besides, it is necessary to keep in mind that neural networks are sensitive to the quality of input data [24, 27, 55].

In order to facilitate and automate the process of time series modelling and forecasting, and compensate for the problems associated with instability of a produced solution, an advanced algorithm has been developed in the thesis.

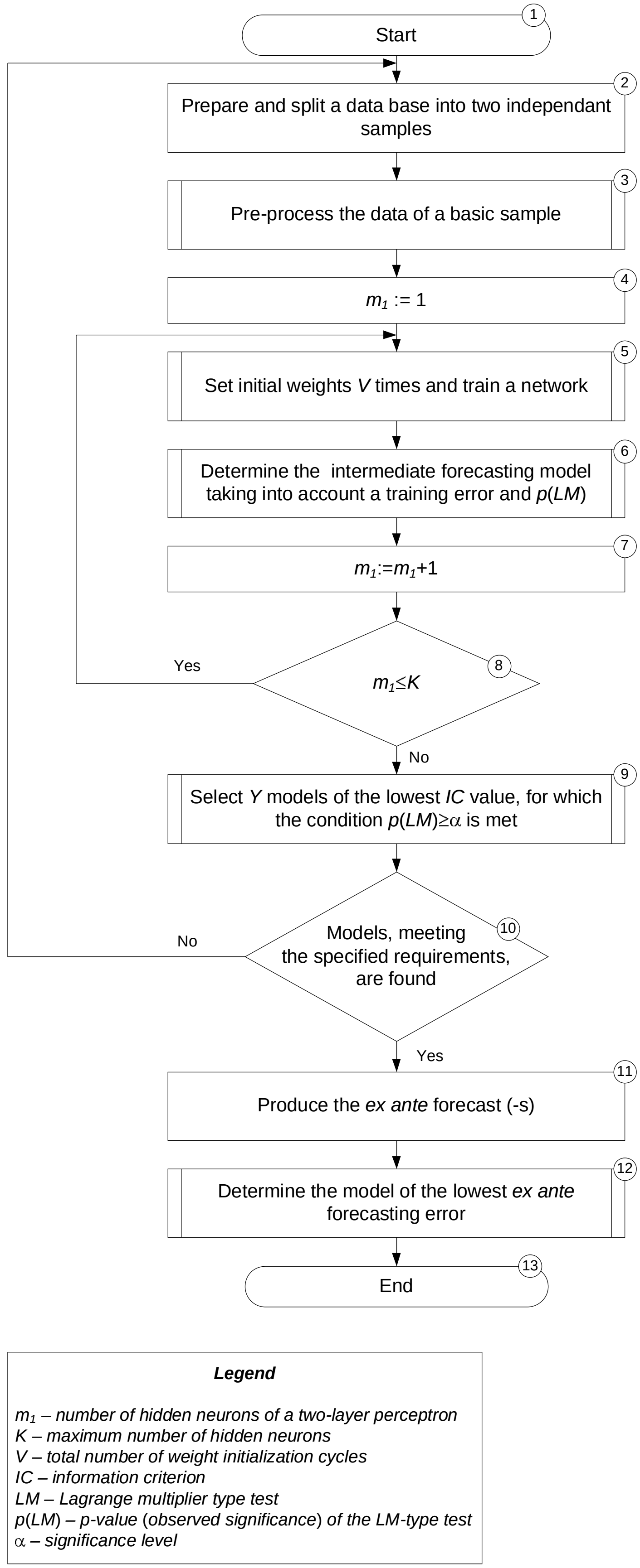


Fig. 5 Advanced algorithm aimed at solving a forecasting task by means of neural networks

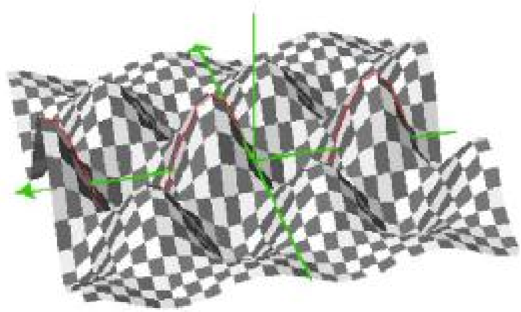
The algorithm, the main block diagram of which is shown in Fig. 5, produces the operative / short-term forecasts of traffic loads represented as univariate / multivariate time series. In contrast to some classic algorithms for accomplishing a pre-defined task (see, for example, [9; 53, p.84])), the proposed algorithm incorporates three procedures:

- implementation of multiple cycles of weight initialization of a neural network during a training process;
- method of selecting the intermediate forecasting models, taking into account the level of residual autocorrelation and the estimates of information criteria;
- method of selecting the final forecasting model, taking into account the accuracy of *ex ante* forecasts.

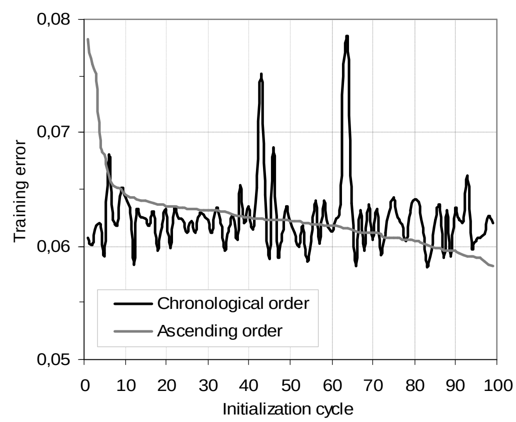
Let us consider the theoretical arguments in defence of the necessity of implementing these procedures to identify a relevant forecasting model.

- **Implementation of multiple cycles of weight initialization of a neural network**

It is required to set some initial values to all the weights and biases of a neural network before a training process starts. The aim of initialization is, probably, to find the best approximation to an optimal solution and, in this way, to decrease training time and facilitate the convergence of a training algorithm.



(a)



(b)

Fig. 6 Illustration of the necessity of implementing multiple initialization cycles: (a) a simplified example of two-dimensional error surface where a vertical axis represents an error. This demonstrates a key problem – several local minima can exist on the training error surface; (b) the training errors of a neural network (without applying cross-validation) produced as a result of a hundred cycles of weight initialization[3]

If the initial weights and biases are set to large values, then neurons usually approach the saturation level very quickly. In this case, the local gradients, calculated according to the back-propagation algorithm, take small values, which, in turn, would significantly increase training time. If the initial values are set to small values, then the algorithm works very slowly about the origin of the error surface. This is specifically true in the case of anti-symmetric activation functions such as hyperbolic tangent.

During the last two decades many heuristic methods of weight initialization have been proposed, some of which are described in Chapter 3 of the thesis. Despite this, the optimal solution of

[3] The values of training errors are displayed in chronological order of their evaluation at the end of each training epoch as well as sorted in descending order to facilitate their comparison.

this issue has not still been found. Due to its simplicity, the most common method is the random initialization of weights and biases from a uniformly or normally distributed narrow range of small values.

Regardless of the initialization method, starting values of weight coefficients influence the final result of training. This is due to the properties of a training algorithm as well as the fact that several local minima can exist on the error surface (see Fig. 6-a). Therefore, in order to find an actual global minimum, it is necessary to train one and the same neural network multiple times under the same conditions, changing only the initial values of weights and biases.

In spite of these problems, most researchers still do not pay a proper attention to this aspect. It is possible to find references to 5 [21], 10 [42], 25 [25], 50 [8; 46] and 100 [25] cycles of initialization. However, most practical studies restrict the number of initialization cycles to one, and this can mislead a researcher regarding the adequacy of a produced solution.

The example shown in Fig. 6-b illustrates the uncertainty in producing a final solution and its dependence on the starting values of weights and biases. The training errors, shown here, are produced as a result of training a neural network of the same architecture and applying the same training parameters but changing the initial values of weights and biases. It is easy to notice that the difference between the largest and smallest error comprises more than 25 per cent, which can significantly influence the identification of a relevant forecasting model and the accuracy of produced forecasts.

The number of initialization cycles is usually chosen mandatory, and depends on the complexity of a task as well as on time resources a researcher has on his / her disposal.

- **Selection of the intermediate forecasting models, taking into account the residual autocorrelation and estimates of information criteria**

According to the developed algorithm, the selection of the intermediate forecasting model / models among the trained networks, with a certain number of hidden neurons m_1, is carried out taking into account the level of *residual autocorrelation* and the value of an *information criterion*.

Some standard parameters, such as the correlation coefficient (R), mean squared error (MSE), mean absolute error (MAE), are traditionally used to evaluate a general forecasting ability of a statistical model. However, these parameters provide little information about the accuracy of a fitted model, and are also useless in identifying the statically significant differences between the forecasts produced by various methods [4, 6, 26].

The most accurate indicator of the adequacy of a forecasting model can be the absence of *autocorrelation in residuals*. The residuals of a fitted model are defined as n differences given by $e_t = x_t - \hat{x}_t, t = 1, 2, \ldots, n$, where x_t is the observed value and $\hat{x}_t$ is a corresponding predicted value produced by means of a fitted statistical model [32, p. 94]. These differences cannot be explained by a forecasting model. Therefore, we can consider residuals e_t to be observed errors.

The acceptance of the hypothesis of no autocorrelation in residuals at a pre-defined significance level means that the residuals are similar to *white noise* and further analysis will not discover any statistically significant dependencies. In classic regression analysis, the *Durbin-Watson* criterion [50, p.245] is traditionally used for testing the autocorrelation of residuals. However, this test is not suitable if the regressor is a lagged explanatory variable [51, p.256]. For the same reason, the *Box-Pierce* and *Ljung-Box* [30] criteria cannot be applied to neural networks, although the last one is widely used and, despite its theoretical inconsistence, is still included in most statistical and econometric software packages.

At present, the most relevant estimate of the residual autocorrelation of neural networks (as well as ARIMA models) is a powerful *Lagrange Multiplier* (*LM*) *type test* [34], which is also known as *Breusch–Godfrey test.* The LM-type test belongs to classic asymptotic tests and is capable to identify the autocorrelation of any order.

In turn, the use of *information criteria* is based on one of the main idea of time series forecasting – "chose a parsimonious model" (known as the *Ockham's razor principle*). It means that, all other things being equal, one should prefer the model with the fewest free parameters.

The mean squared residuals usually decrease once the model becomes more “complicated” with addition of new free parameters. Increasing the number of free parameters of a neural network (which is associated with the addition of new neurons / layers of neurons), one can fit a model to historical data with infinite accuracy. The *universal approximation theorem* [5; 15] explains this property of neural networks. However, such a neural network may have a poor ability to make generalizations due to *over-training* [13, p.206]. Besides, once a certain limit is reached, the gain in accuracy of fitting with addition of new parameters tends to be insignificant. On the other hand, time required for selecting the optimal values of free parameters can lead to a sharp decrease in the performance of a network. Therefore, it is very important to look for the balance between the preciseness of approximation and the complexity of a statistical model.

Information criteria have been found to be quite useful in solving this problem. The estimate of the criterion consists from the penalty for poor fitting and the penalty for over-parameterization. The most popular criteria of this type, applied in the practical part of the thesis, are the *Akaike's information criterion* (AIC) and the *Bayesian information criterion* (BIC) given by [11, p.38; 38, p.373]:

$$AIC(l) = N_{ef} \ln \hat{s}_e^2 + 2l \tag{12}$$

$$BIC(l) = N_{ef} \ln \hat{s}_e^2 + 2l + l \ln(N_{ef}) \tag{13}$$

where

N_{ef} – number of effective observations, to which the model is fitted;

l – number of adjusted parameters;

$\hat{s}_e^2$ – estimate of the residual variance, $\hat{s}_e^2 = \dfrac{\sum_{t=1}^{Nef} e_t^2}{N_{ef}}$.

Information criteria are evaluated separately for each analyzed specification (architecture) of neural networks. The models that possess the lowest value of the criterion should be selected for further analysis. It has been also noticed that, in practice, the *BIC* “selects” very parsimonious models with only few parameters. Therefore, this criterion is often used for non-linear models, where insignificant gain in fitting quality is directly related to the necessity of calculating a large number of additional parameters.

In the practical part of the thesis the criteria given by Eqs. (12) and (13) were applied. However, some other modifications of information criteria have been proposed as well. The *Schwarz's Bayesian criterion* (SBC) [38, p.376] and the *Hannan-Quinn* criterion [32, p. 86] are among them.

- **Selection of the final forecasting model, taking into account the accuracy of *ex ante* forecasts**

If the models, meeting the above specified conditions, are found, it is required to test their generalization ability (i.e., the ability to produce reliable forecasts) for an independent test set, which is not involved in training. The forecast developed for an independent test set we will call an *ex ante forecast* or *a pseudo-forecast*. The necessity *of ex ante forecasting* hinges upon the fact that even if a neural network provides a high accuracy of approximation and uncorrelated residuals, it would be still over-trained on historical data.

The approach of splitting a time series into two independent subsets has gained wide acceptance in the practical studies dedicated to time series forecasting (see, e.g., [31; 43]) but it is still rarely used in the case of neural networks. The first, largest data subset, called the *basic* or *retrospective sample*, is used to select and verify a statistical model. The second, *ex ante forecasting sample* is used to examine the quality of *ex ante* forecasts, comparing them against historical data. It provides the opportunity to evaluate independently the forecasting ability of the model fitted to the basic sample.

The last historical values of an analyzed time series are traditionally used for developing the *ex ante* forecasting sample. However, it is necessary to keep in mind, that these observations influence the direction of the actual *real-life* forecast much more than the earlier ones. Therefore, the last historical data are the most valuable for the process of selecting an appropriate forecasting model, and using them as a testing sample is not always reasonable.

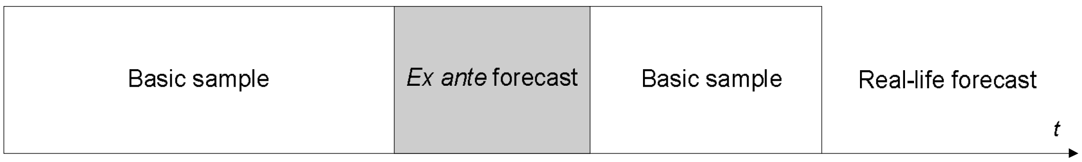


Fig. 7 Chronological division of a time series into the basic and *ex ante* forecasting samples

According the proposed algorithm, it is recommended to divide a time series into the basic and ex ante samples in the way shown in Fig. 7. This original approach allows increasing the quality of real, *ex post* forecasts as the last historical observations are used for fitting a statistical model rather than testing.

The accuracy of *ex ante* forecasts is evaluated by one of the standard error parameters. In the practical studies of the thesis, the mean absolute percentage error (*MAPE*) was applied. It is calculated from [1, p. 347]:

$$MAPE = \frac{\sum_{t=1}^{L} \left| \frac{x_t - \hat{x}_t}{x_t} \right|}{L} \cdot 100\% \tag{14}$$

where L – the size of an *ex ante forecasting* sample (i.e., forecasting horizon).

The *MAPE* is a relative, dimensionless measure of the accuracy of an approximation curve or a forecast. It is helpful in comparing forecast performance across different data sets, or comparing the performance of different statistical methods.

The model of the lowest MAPE is the final model assigned to further *ex post* forecasting. The interpretation of *MAPE* introduced in [29] allows judging about the accuracy of a forecast: less than 10 per cent is a highly accurate forecast, 11 to 20 per cent is a good forecast, 21 to 50 per cent is a reasonable forecast, and 51 per cent or more is an inaccurate forecast.

Thus, the choice of a final forecasting model is based on the results of multiple sequential procedures and tests. The final model is characterized by the lowest value of the information criterion, uncorrelated residuals and the lowest error of an *ex ante* forecast.

CHAPTER 4 PRACTICAL STUDIES

Chapter 4 contains the description of practical research studies and the analysis of produced results.

The effectiveness of the developed algorithm and the ability of different methods to accomplish a traffic forecasting task were examined on real data sets represented as time series of different lengths and aggregation rates. Two data samples, characterizing the intensity of total carried traffic of a conventional telephone network and the transmission rate of outgoing international traffic, were considered in the thesis. Following ITU-T Recommendation E.492 [17], the initial traffic measurements were averaged over the periods equal to 15 minutes and one hour.

Table 1 General description of examined time series

Network Type	Traffic Type	Read-out period	Label	Basic sample	Size of a basic sample	Size of an *ex ante* forecasting 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 h	C	May 12 – Jul. 13, 2008	9 weeks (1512 obs.)	
			D	May 12 – Aug. 3, 2008	12 weeks (2016 obs.)	
Telephone network (POTS)	Total carried traffic	15 min	E	Jan. 8 – Mar. 11, 2007	9 weeks (6048 obs.)	1- 14 days
			F	Jan. 8 – Apr. 1, 2007	12 weeks (8064 obs.)	
			G	Jan. 8 – May 13, 2007	18 weeks (12096 obs.)	
		1 h	H	Jan. 8 – Mar. 11, 2007	9 weeks (1512 obs.)	
			I	Jan. 8 – Apr. 1, 2007	12 weeks (2016 obs.)	
			J	Jan. 8 – May 13, 2007	18 weeks (3024 obs.)	

A secondary goal of the practical studies was to examine how both the size of a basic data sample and the rate of aggregation influenced the accuracy of *ex ante* forecasts. The size of a basic sample was equal to 9 and 12 weeks for the first analyzed variable, and to 9, 12 and 18 weeks – for the second variable. Thus, the total number of time series considered was equal to ten.

The size of an *ex ante* forecasting sample, which determined a total forecasting horizon, varied for each time series from one to 14 days, with the sampling step of one day.

The general description of the considered time series is given in Table 1. The fragments[4] of the time series are displayed in Fig. 8.

Prior to determining an appropriate forecasting model and developing *ex ante* forecasts, the main statistical parameters and properties of each time series were estimated. The corresponding procedures included:

- assessment of the main sample parameters (mean, variance, median, etc.);
- testing for stationarity by means of the runs test and reverse arrangement test [24; 49, p.767];
- evaluation of the autocorrelation function;
- evaluation of the Hurst coefficient;
- testing for periodicity.

In the case of time series (E)-(J), characterizing telephone network traffic, the reverse arrangement test accepted the null hypothesis of the stationarity of both the mean and the variance at significance level $\alpha = 0.05$. For time series (A)-(D), characterizing Internet network traffic, the reverse arrangement test revealed the instability of the variance at significance level $\alpha = 0.05$. Nevertheless, the deviation of the number of reversals from the critical limits was slight. Already at significance level $\alpha = 0.02$, the hypothesis about the variance stationarity was accepted in most cases considered. Therefore, it was decided not to apply further measures to stabilize the variance.

The analysis of the autocorrelation function indicated the presence of periodic components. The influence of strong autocorrelation dependencies was observed not only between the adjacent members but also between the quite remote ones. It points at the "long history" of an underlying process, which provides the opportunity to produce reliable forecasts into a rather distant future. The persistency of the analyzed time series was confirmed by the Hurst coefficient as well, which exceeded 0.5 for all the time series analyzed. However, this estimate is often criticized and purely optional due to its inaccuracy. Besides, it is worth noting that the value of the Hurst coefficient cannot be directly incorporated into a forecasting model.

[4] Each fragment displays the observations over the first two weeks of a considered data sample

The spectral analysis of the IP network traffic pointed at the periodical components of the periods equal to 24 hours and 7 days. In the case of telephone traffic, the largest periods of the seasonal component comprised 12 hours, 24 hours and 7 days.

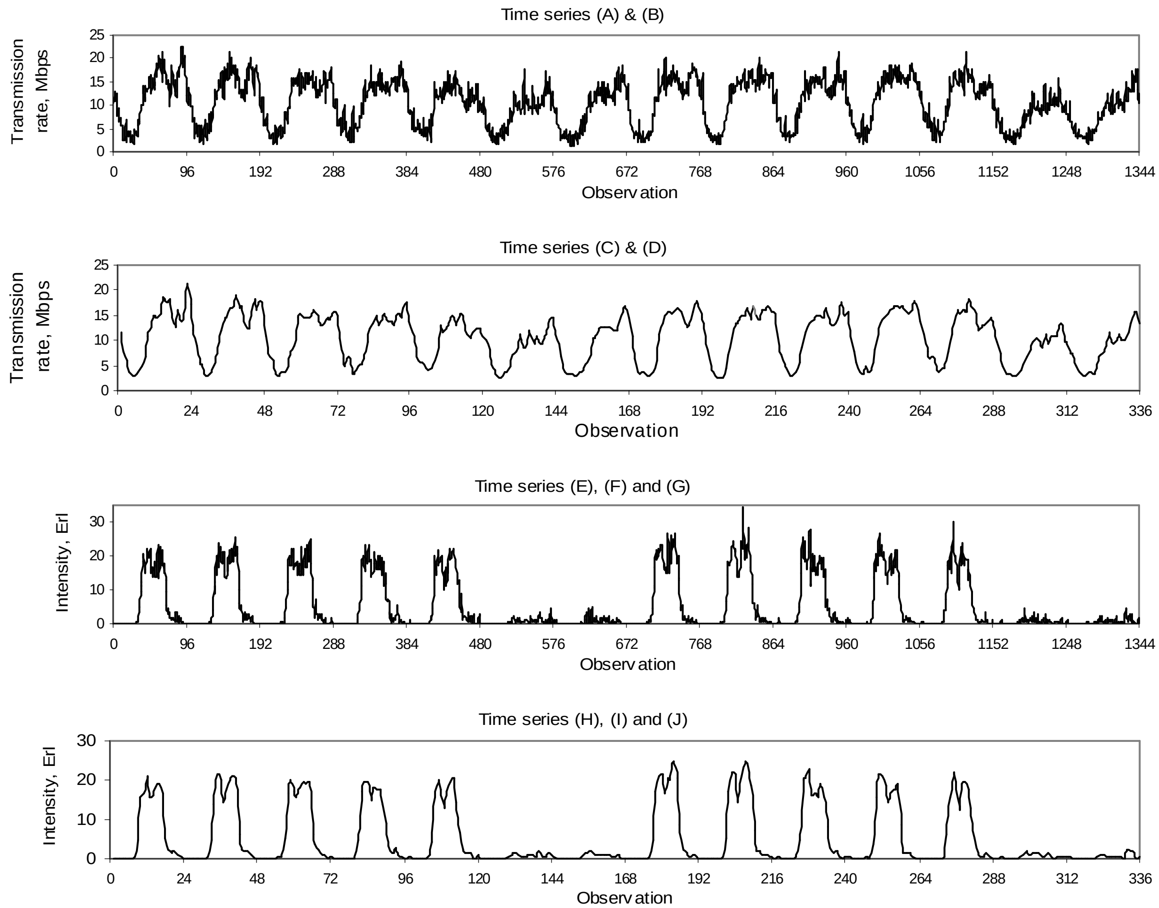


Fig. 8 Fragments of examined time series

- **Description of Forecasting Technique**

The main goal of the practical studies was to examine the statistical properties of certain data sets and to develop such a neural network, which was capable of modelling the underlying processes and producing the reliable low-error forecasts for a pre-defined forecasting horizon.

The selection of appropriate neural network models and the development of *ex ante* forecasts were fulfilled according to the algorithm shown in Fig. 5. The diagram of the neural network, applied in the empirical studies, is shown in Fig. 9 . The main parameters of the neural network, which stayed unchanged for all the models during a training process, are summarized in Table 2.

The appropriate architecture of a neural network was determined as follows. According to the universal approximation theorem [5, 15] the number of hidden layers in all the examined neural networks was equal to one. The size of an input window was set according 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 one-step-ahead forecasting. The number of hidden neurons varied from one to ten. The adaptive methods of network pruning [2, p.359] or growing [2, p. 357; 13, p. 250] were not implemented. The procedures of verification and residual testing were applied to each of these models. Although the process of verifying all the possible architectures is time-consuming, it provides an opportunity to preserve the purity of experiments.

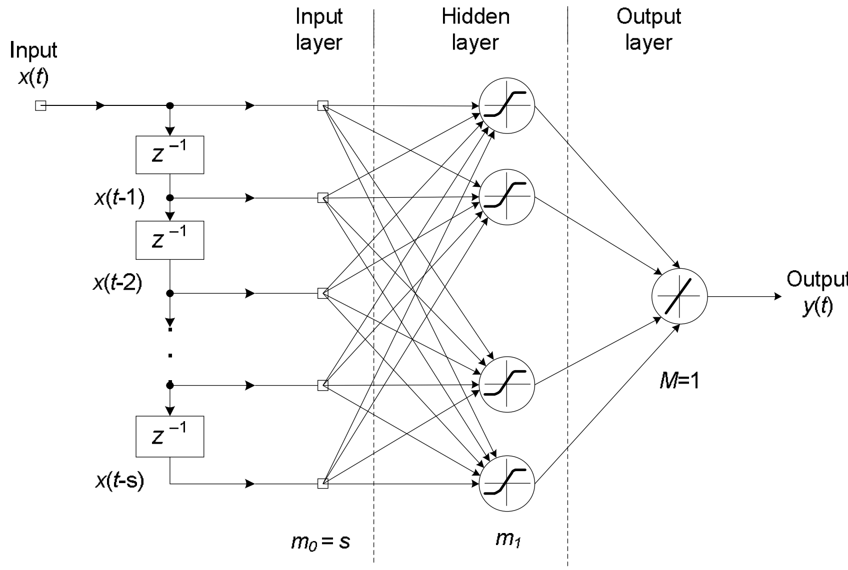


Fig. 9 Diagram of the neural network (multilayer perceptron) applied in the empirical studies

The initial weights were randomly drawn from the diapason of uniformly distributed small values. All the network architectures were reinitialized and retrained a hundred times.

Table 2 General specification of the developed neural network

Stage	Parameter / Procedure	Parameter Value / Procedure Description
Selection of network type and topology	Type of a network	Fully connected time-lagged feed forward network
	Number of hidden layers	1
	Number of output neurons	1
	Activation function	Hidden layer – hyperbolic tangent; output layer – linear function
Selection of training parameters	Number of training epochs	600
	Training algorithm	Back propagation & conjugate gradient descent
	Error function	Mean squared error
	Learning rate	0.1
	Momentum term	0.3
	Method of weight initialization	Randomized values from a uniform distribution
	Number of times to randomize weights	100
Training optimization	Methods to prevent over-learning	Cross-validation [13, p.218] , weight regularization [47]
	Size of training, validation and test subsets	At a ratio of 3:1:1
	Stopping criterion	Invariable or increasing training error during 50 epochs
In-sample and out-of-sample evaluation	Parameters of in-sample evaluation	R, MAE, RMSE, MAPE, AIC, BIC
	Diagnostic testing of residuals	LM- type test , χ^2 - test
	Parameters of out-of-sample evaluation	RMSE, MAE, MAPE, Diebold-Mariano criterion [6]

In order to avoid the effect of over-training, the ***cross-validation*** technique [13, p.218] was implemented. The basic sample was divided into training, validation and test subsets at a ratio 3:1:1. A splitting scheme was random and changed for each training cycle. This approach does not allow “getting stuck” in local minima and increases the stability of a system, since the process of searching a global minimum is carried out in different directions and do not rely on a particular set of time series observations. Another measure to avoid over-training was *weight regularization* [47] applied without subsequent deletion of synaptic connections and neurons.

A training process was realized by means of the software package StatSoft STATISTICA© 7.0. A two-stage training process was implemented. During the first stage a multilayer perceptron was trained by 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 (500 epochs) was used, with a stopping window of 50, to terminate training once convergence stopped or over-learning occurred. Once the algorithm stopped, the best network from the training run was restored.

The input data of a neural network were corrected for obvious anomalous outliers, the reason of which was temporal malfunction of network equipment, and for anomalous patterns, which took place as a result of public holidays falling on the days of a workweek. The input data sets were also transformed to the range $[-1, 1]$ by means of the linear transformation.

The final forecasts produced by neural networks were compared to those produced by the models of seasonal ARIMA, seasonal exponential smoothing as well as "naïve" methods. In order to evaluate statistically significant differences between the forecasts developed for various forecasting horizons, the Diebold-Mariano [6] criterion was implemented. It is non-parametric and tolerant to different deviations from the classic assumptions about the properties of forecast errors. In particular, it can be applied even if forecast errors are non-Gaussian, serially correlated, contemporaneously correlated and have a non-zero mean.

- **Estimation of Practical Results**

The results of fitting and verifying the statistical models and neural networks, the estimates of their in-sample and *ex ante* accuracy are summarized in Volume 2 of the thesis. The following main operations were conducted for each time series:

- appropriate models of multilayer perceptrons, SARIMA models and exponential smoothing were identified and varified;
- *ex ante* forecasts were produced by means of different models and evaluated for the accuracy;
- statistically significant differences between the final *ex ante* forecasts developed for various forecasting horizons were identified by means of the Diebold-Mariano test.

The accuracy of final forecasts was evaluated by means of such standard parameters as MAE (mean absolute error), RMSE (root mean squared error) and MAPE (mean absolute percentage error), the latter of which raises the greatest interest (see Fig. 10).

Accuracy of neural network forecasts evaluated by the mean absolute percentage error

For time series (A)-(D) describing IP network traffic, the MAPE estimates do not practically change or slowly grow with the increase of a forecasting horizon from 24 hours to 14 days. It points at the opportunity to increase a lead time, for which reliable forecasts can be produced.

For time series (E)-(G) characterizing telephone network traffic, the MAPE estimates grow fast with a forecasting horizon, and already for the forecasts obtained two weeks ahead, exceed 30 per cent. It means that a maximum forecasting horizon is achieved. In turn, for time series (H)-(J) aggregated over one-hour intervals, the MAPE estimates do not practically change or slightly increase with a forecasting horizon. It would allow to extend a forecasting horizon further.

For time series (A) and (B), the MAPE estimates of neural networks and statistical models comprise 20-25 per cent. This reveals a satisfactory accuracy of the produced forecasts. For time series (C) and (D) the MAPE of statistical models and neural networks is around 10-15 per cent, which demonstrates a good accuracy of produced forecasts.

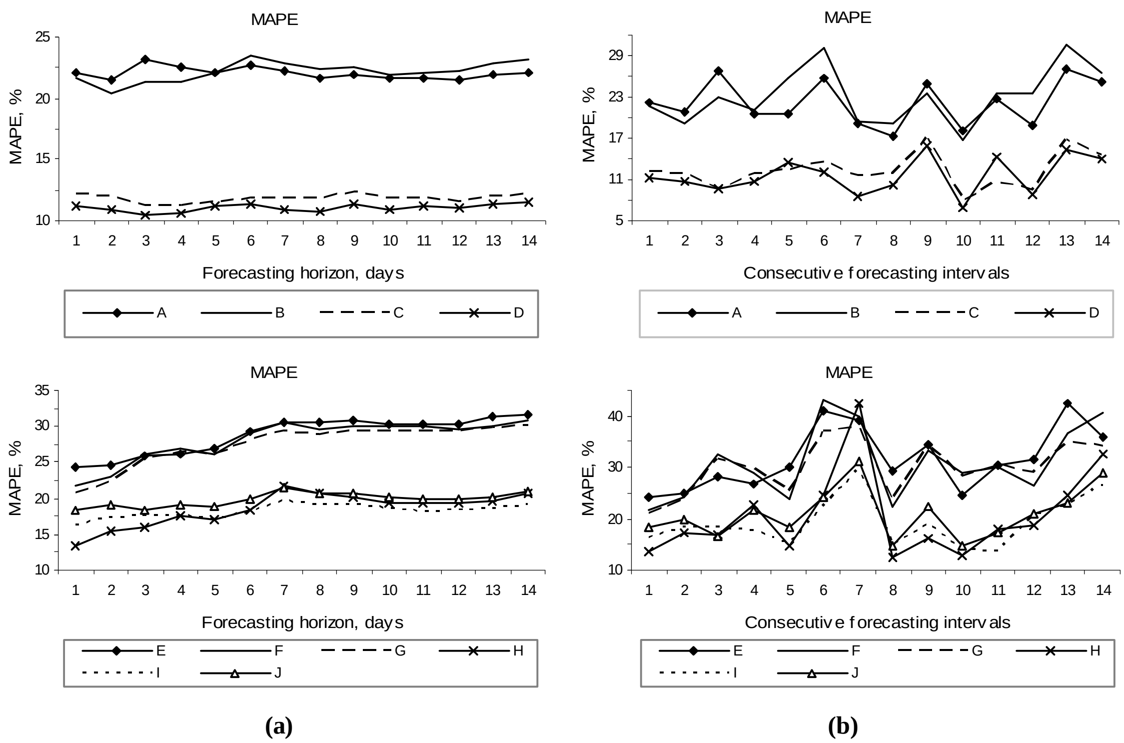


Fig. 10 Estimates of the accuracy of the final *ex ante* forecasts produced by neural networks (a) against the length of a forecasting horizon; (b) at consecutive one-day sampling intervals[5]

For time series (E)-(G) the MAPE estimates of neural networks comprise 21-32 per cent, which is the evidence of a satisfactory accuracy of the produced forecasts. For time series (H)-(J), the MAPE of neural networks is around 14-21 per cent. It points at a good accuracy of the forecasts.

Statistically significant differences between the forecasts produced by different methods

The final forecasts produced by various methods (neural networks, SARIMA and seasonal exponential smoothing) look very similar. Besides, it is not easy to select a forecasting model, taking into account only the standard accuracy parameters. Therefore, the identification of statistically significant differences in accuracy of the forecasts, developed by different methods, raises a special interest.

For these purposes, the Diebold-Mariano test was applied to the forecasts produced 24 hours, 7 days and 14 days ahead. The first group of forecasts can be considered as operative forecasts, while the second and the third ones – as short-term forecasts. The results of testing a null hypothesis of the absence of statistically significant differences between the forecasts, at significance level $\alpha = 0.05$, are summarized in Table 3.

For time series (A)-(D) characterizing IP network traffic, the forecasts produced by one or another linear method (SARIMA or exponential smoothing) do not lose in accuracy to the forecasts of neural networks, in all 12 analyzed cases. In two out of 12 cases, the forecasts produced by neural networks are statistically equivalent to the seasonal naïve forecasts as well.

For time series (E)-(J) characterizing telephone network traffic, the forecasts produced by neural networks are statistically equivalent to the forecasts produced by one or another linear method in 14 out of 18 analyzed cases. In the other four cases, a neural network outperforms in forecasting

[5] The MAPE estimates are calculated for the data moved along the y-axis by the distance of 0.5 Erl. It is required as the telephone traffic contains zero values. Even if this procedure distorts the actual values of absolute percentage errors of the observations which are not zero, it gives the opportunity to evaluate the dynamic changes of MAPE over different forecasting horizons.

accuracy both SARIMA and seasonal exponential smoothing. In 4 out of 18 analyzed cases, statistically significant differences were not identified between the forecasts produced by neural networks and seasonal naïve methods.

Table 3 Final neural network forecasts in comparison with the forecasts produced by other methods[6]

Forecasting method \ Time series/ Forecasting horizon	A			B			C			D		
	24 h	7 d	14 d	24 h	7 d.	14 d.	24 h	7 d	14 d	24 h	7 d	14 d
SARIMA	↑	=	=	↑	=	=	↑	=	↑	=	=	=
Seasonal exponential smoothing	=	↓	↓	=	↓	↓	=	↓	↓	↓	↓	↓
"Naïve" forecast	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Seasonal "naïve" forecast	↓	↓	↓	↓	↓	↓	=	↓	=	↓	↓	↓

Forecasting method \ Time series/ Forecasting horizon	E			F			G			H			I			J		
	24 h	7 d	14 d	24 h	7 d	14 d	24 h	7 d	14 d	24 h	7 d	14 d	24 h	7 d	14 d.	24 h	7 d	14 d
SARIMA	=	=	=	=	↓	↓	=	↓	↓	=	=	=	=	=	=	=	=	=
Seasonal exponential smoothing	=	=	=	=	↓	↓	=	↓	↓	=	=	=	=	=	=	=	=	=
"Naïve" forecast	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Seasonal "naïve" forecast	=	↓	↓	↓	↓	↓	↓	↓	↓	=	↓	↓	=	↓	↓	=	↓	↓

Notes:
= *the forecast, produced by the specified method, is statistically equivalent to the neural network forecast;*
↑ *the forecast, produced by the specified method, outperforms in accuracy the neural network forecast;*
↓ *the forecast, produced by the specified method, loses in accuracy to the neural network forecast.*

Impact of the length of a read-out period on forecasting accuracy

For all the analyzed time series, the increase of a read-out period allowed increasing the accuracy of the produced *ex ante* forecasts. On average, the increase of a read-out period from 15 minutes to one hour decreased the MAPE values of a neural network for 10 per cent. This is primarily due to the reduction of time series variance, which simplified the selection of an appropriate statistical model as well.

Impact of the size of a basic sample on forecasting accuracy

For time series describing the traffic of both an IP network and a conventional telephone network, the increase of the size of a basic fit sample (i.e., the increase of the number of training patterns) did not lead to a substantial increase in accuracy of neural network forecasts.

In the case of IP network traffic, the MAPE estimate is slightly lower for time series (D) than for time series (C), although these differences are insignificant. For telephone network traffic, the MAPE estimates are a bit lower for time series (G) than for time series (E) and (F) but these differences are also insignificant.

In the case of telephone traffic aggregated over one-hour intervals, the increase of a basic sample from 12 to 18 weeks, in contrast, resulted in a slight increase in the level of forecasting errors.

[6] The identification of statistically significant differences in forecasting accuracy was conducted by the Diebold-Mariano test at significance level α=0.05.

MAIN CONCLUSIONS AND RECOMMENDATIONS

The *main aim* of the research – to address and solve the problems, raised by the production of short-term traffic forecasts by means of neural networks, has been achieved. Taking into account the results of practical and theoretical studies conducted in the thesis, the following main conclusions and recommendations, regarding the application of forecasting models in network traffic forecasting, have been specified.

The proposed algorithm, aimed at short-term traffic forecasting by means of neural networks, allow automating the identification of a neural network solution with minimal involvement and influence of expert assessment and human factors.

Neural networks traditionally involve expert assessment at all the stages of application. The algorithm, developed in the thesis, allows minimizing the influence of a human factor and helps finding the solutions resulting in reliable forecasts with a minimum level of errors. The *MAPE* estimates of examined *ex ante* forecasts vary from 10 per cent (in the case of averaging over one-hour intervals) to 30 per cent (in the case of averaging over 15-minute intervals). This confirms the possibility of applying these models in real-life conditions. The criteria for selecting a final forecasting model proposed in the thesis – the lowest estimate of the information criterion and statistically insignificant residual autocorrelation can be successfully applied to linear statistical methods as well.

The properties of self-similarity and long-range dependence of packet-switched network traffic are only observable in the case of aggregation in a very large scale – usually, over the intervals from a few milliseconds to a few minutes.

From the point of view of analysis and forecasting, an excessively large scale of time series is not useful. The process of fitting a forecasting model to an examined time series will be complicated due to correlations between remote observations as well as strong influence of extraneous noises and anomalous outliers, which inevitably accompany the large-scale measurements. It is also necessary to understand that the longer the period of sampling / aggregation, the longer is the horizon, for which reliable forecasts can be produced. Therefore, taking into account ITU-T Recommendation E.492 [17], it is advised to average the initial measurements of network traffic over 15-minute and / or one-hour intervals. In this case, the factors determining the statistical structure of a *real* traffic process refer to seasonal effects and monotonous trends, which are primarily associated with the behaviour of subscribers / users and the influence of technological progress.

Reliable operative and short-term forecasts of traffic dynamics can be produced by means of linear statistical models, if the aggregation / sampling period of time series is set in compliance with ITU-T Recommendation E.492.

The real time series analyzed in the thesis are *typical* for these types of loads and incorporate both daily and weekly cycles. It means that for many other time series with similar read-out periods, statistical properties and autocorrelation structure, the production of reliable operative and short-term forecasts can be conducted by applying linear statistical models and methods. The process of forecasting by means of neural networks requires substantial time resources for training, apart from an intensive test-and-trial estimation of many parameters. The time required for series modelling and forecasting by means of neural networks exceeds the time required for producing a forecast by means of linear methods by several orders of magnitude. Therefore, this conclusion is important for successful forecasting in real operating conditions.

A neural network is capable of modelling and predicting seasonal time series in a direct way, without prior deseasonalization.

In developing a seasonal neural network, the most important aspect to consider is the appropriate size of an input window, which has to be set according to the largest period of a seasonal component.

In order to produce reliable forecasts of network traffic represented as periodic time series, it is necessary to focus on the seasonal modifications of such linear models as ARIMA and exponential smoothing.

In contrast to the series, synthesized by such formal models as *fractal Brownian motion*, real time series often incorporate seasonal and / or cyclic components and require the application of seasonal modifications of classic linear models. Just as in the case of neural networks, correct identification of the periods of a seasonal component is of great importance.

- **Potential Directions for Further Research**

The method of neural networks is, certainly, one of the most perspective tools in traffic forecasting. Further study directions has to be aimed at developing the algorithms and methods of *real-time forecasting* that currently raise some difficulties due to insufficient capacity of computing equipment and intensive involvement of human expertise.

Another interesting research direction is to develop fast and reliable methods of constructing the *confidence intervals* for the forecasts produced by neural networks. This problem has not been solved yet due to the necessity of estimating a very large number of free parameters, each of which contributes a share of uncertainty.

BIBLIOGRAPHY

1. Armstrong, J.S. *Long-Range Forecasting*. New York: John Willey & Sons, 1985.
2. Bishop, C. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
3. Box, G.E.P., and G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, Inc., 1970.
4. Clements, M.P., and D.F. Hendry. "On the Limitations of Comparing Mean Square Forecast Errors". *Journal of Forecasting* 12 (1993): 617-637.
5. Cybenko, G. "Approximation by Superpositions of Sigmoidal Function." *Mathematics of Control, Signals, and Systems* 2 (1989): 303-314.
6. Diebold, F.C., and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13.3 (July 1995): 253-263.
7. Elman, J.L. "Finding Structure in Time." *Cognitive Science* 14 (1990): 179-211.
8. Faraway, J., and C. Chatfield. "Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data." *Applied Statistics* 47.2 (1998): 231-250.
9. Fildes, R., and K.-P. Liao. "The Accuracy of Procedural Approach to Specifying Feedforward Neural Networks for Forecasting." *Computers & Operations Research* 32.8 (August 2005): 2151-2169.
10. Fowler, T.B. "A Short Tutorial on Fractals and Internet Traffic." *The Telecommunications Review* 10 (1999): 1-15.
11. Franses, H.P., and D. van Dijk. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, 2000.
12. Gardner, E.S.Jr. "Exponential Smoothing: The State of Art." *Journal of Forecasting* 4 (1985): 1-28.
13. Haykin, S. Neural Networks: A Comprehensive Foundation. 2nd Ed. Prentice Hall, 1999.
14. Hill, T., M. O'Connor, and W. Remus. "Neural Network Models for Time Series Forecasts." *Management Science* 42.7 (Jul. 1996): 1082-1092.
15. Hornik, K. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* 2 (1989): 359-366.
16. ITU-T (CCITT) Recommendation E.506 (rev.1). Forecasting International Traffic. Geneva, 1992.
17. ITU-T Recommendation E.492. Traffic Reference Period (Telephone Network and ISDN. Quality of Service, Network Management and Traffic Engineering). ITU, 02/1996.
18. ITU-T Recommendation E.507. Models for Forecasting International Traffic. ITU, 1993 (1988).
19. Iversen, V. B. *Teletraffic Engineering and Networks Planning*. Technical University of Denmark, 2008.
20. Janevski, T. *Traffic Analysis and Design of Wireless IP Networks*. ARTECH HOUSE, INC., 2003.
21. Kaastra, I., and M. Boyd. "Designing a Neural Network for Forecasting Financial and Economic Time Series." *Neurocomputing* 10 (1996): 215-236.
22. Klevecka, I., and J. Lelis. "Application of Extrapolation Methods to the Technology Diffusion Forecasting." *Scientific Proceedings of Riga Technical University* (*Series "Telecommunications and Electronics "*) 7 (2007: 52-59.
23. Klevecka, I., and J. Lelis. "Perspective Evaluation of the Electronic Communications Market in Latvia." *Proceedings of the 18th European Regional ITS Conference*. Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 1-37. CD-ROM.
24. Klevecka, I., and J. Lelis. "Pre-Processing of Input Data of Neural Networks: The Case of Forecasting Telecommunication Network Traffic." Spec. issue of *Telektronikk* 104.3/4 (2008): 168-178.
25. Kolen, P., and J.B. Pollack. "Back Propagation in Sensitive to Initial Conditions." *Proceedings of the Conference on Advances in Neural Information Processing Systems*. Nov. 1989, Denver, United States. San Francisco: The MIT Press, 1990. 860-867.
26. Kunst, R.M. *Testing for Relative Predictive Accuracy: A Critical Viewpoint*. Economics Series 130. Vienna: Institute for Advanced Studies (IHS), 2003.
27. LeCun, Y., L. Bottou, G.B. Orr, and K.-R. Müller. "Efficient BackProp." *Neural Networks: Tricks of the Trade*. Eds. G.B. Orr and K.-R. Müller. Berlin: Springer, 1998. 9-50.
28. Leland, W.E., M.S. Taqqu, W. Willinger, and D.V. Wilson. "On the Self-Similar Nature of Ethernet Traffic (Extended Version)." *IEEE/ACM Transactions on Networking* 2.1 (February 1994): 1-15.
29. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. London: Butterworth Scientific, 1982.
30. Ljung, G.M., and G.E. Box. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (1978): 297–303.
31. Makridakis, S., and R.L. Winkler. "Sampling Distributions of Post-sample Forecasting Errors." *Applied Statistics* 38.2 (1989): 331-342.
32. McNelis, P.D. *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Elsevier Academic Press, 2005.
33. Meade, N. "Evidence for the Selection of Forecasting Methods." *Journal of Forecasting* 19 (2000): 515-535.
34. Medeiros, M.C., T. Teräsvirta, and G. Rech "Building Neural Network Models for Time Series: A Statistical Approach." *Journal of Forecasting* 25 (2006): 49-75.

35. Mozer, M. C. “Neural Network Architectures for Temporal Pattern Processing.” *Time Series Prediction: Forecasting the Future and Understanding the Past.* Eds. A.S. Weigend and N. A. Gershenfeld. Perseus Book Publishing, 1993. 243-264.
36. Park, C., F. Hernandez-Campos, J.S. Marron, and F.D. Smith. “Long-Range Dependence in a Changing Traffic Mix.” *Computer Networks* 48 (2005): 401-422.
37. Park, K., and W. Willinger. “Self-Similar Network Traffic: An Overview.” *Self-Similar Network Traffic and Performance Evaluation.* Eds. K. Park, and W. Willinger. John Wiley & Sons, 2000. III-XLIX.
38. Priestley, M.B. *Spectral Analysis and Time Series.* London: Academic Press, 1981.
39. Ryu, B.K., and A. Elwalid. “The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities.” *ACM SIGCOMM Computer Communication Review* 26.4 (1996): 3-14.
40. Sang, A., and S.-q. Li. “A Predictability Analysis of Network Traffic.” *Computer Networks* 39 (2002): 329-345.
41. Sheluhin, O.I., S.M. Smolskiy, and A.V. Osin. *Self-Similar Processes in Telecommunications.* John Willey & Sons, 2007.
42. Tang, Z., and P.A. Fishwick. „Feedforward Neural Nets as Models for Time Series Forecasting” *ORSA Journal on Computing* 5.4 (Fall 1993): 374-385.
43. Tashman, L.J. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting* 16.4 (October 2000): 437-450.
44. Taylor, J.W., L.M. de Menezes, and P.E. McSharry. „A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead.” *International Journal of Forecasting* 22 (2006): 1-16.
45. *Teletraffic Engineering Handbook.* ITU-D (Study Group 2, Question 16/2), June 2006.
46. Thimm, G., and E. Fiesler. *Optimal Setting of Weights, Learning Rate, and Gain.* IDIAP Research Report 97-04. Switzerland: IDIAP, 1997.
47. Weigend, A. S., D. E. Rumelhart, and B. A. Huberman. “Generalization by Weight-Elimination with Application to Forecasting.” *Advances in Neural Information Processing Systems 3.* Eds. R.P.Lippmann, J. Moody, and D. S. Touretzky. San Mateo, CA: Morgan Kaufmann, 1991. 875-882.
48. Yaffe, R.A, and M. McGee. *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS.* Academic Press, 2000.
49. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Издательское объединение «ЮНИТИ», 1998. – 1005с.
50. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2007. – 912с.
51. Канторович Г.Г. Анализ временных рядов // Экономический журнал ВШЭ. №1 2002, с.85-116.
52. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. – М.: ИНФРА-М, 1997.- 302с.
53. Комарцова Л., Максимов А. Нейрокомпьютеры: Учеб. пособие для вузов. - М.: Издательство МГТУ имени Н.Э. Баумана, 2002. – 320с.
54. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи. – М.: Горячая линия-Телеком, 2003. – 94с.
55. Крисилов В., Чумичкин К., Конратюк А. Представление исходных данных в задачах нейросетевого программирования// V Всероссийская научно-техническая конференция "Нейроинформатика-2003": Сборник научных трудов. Ч. 1. - М: МИФИ, 2003, с 184-191.
56. Крылов В.В., Самохвалова С.С. Теория телетрафика и ее приложения. – Спб.: БХВ-Петербург, 2005. – 288с.
57. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. – М.: ИНФРА, 1998. – 528с.
58. Шмертер Х., Белендорф Э. и др. Теория телетрафика. – М.: Связь, 1971. – 319с.

RIGA TECHNICAL UNIVERSITY
Faculty of Electronics and Telecommunications
Institute of Telecommunications

Irina KLEVECKA

Doctoral student of the study program "Telecommunications and Computer Networks"

NEURAL NETWORKS
FOR SHORT-TERM FORECASTING OF NETWORK TRAFFIC

Summary of Promotion Thesis

Scientific Supervisor
Dr.Sc.Ing. J. LELIS

Riga 2011

Klevecka I. Neural Networks for Short-Term
Forecasting of Network Traffic. Summary of Promotion
Thesis.-R.:RTU, 2011.-32 p.

Printed in accordance with resolution of Promotion
Council RTU-P-08 of Riga Technical University,
Minute No.1 dated February 22, 2011.

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».

PROMOTION THESIS
IS SUBMITTED TO RIGA TECHNICAL UNIVERSITY IN FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF DOCTOR OF ENGINEERING SCIENCES

The public defense of the promotion thesis, submitted for the degree of Doctor of Engineering Sciences, takes place at the Faculty of Electronics and Telecommunications of Riga Technical University, 12 Azenes St., Room 210, Riga, Latvia, on the 9th June of 2011 at 15:00.

OFFICIAL REVIEWERS

Dr.habil.sc.ing. V. Štrauss, Senior Scientist
Institute of Polymer Mechanics of University of Latvia, Riga, Latvia

Dr.habil.sc.ing. M.L. Šneps-Šnepe, Professor
Ventspils University College, Ventspils, Latvia

Dr.sc.ing. I. Jackiva, Professor
Transport and Telecommunication Institute, Riga, Latvia

CONFIRMATION

I confirm that the promotion thesis, submitted to Riga Technical University in fulfillment of the requirement for the degree of Doctor of Engineering Sciences, is my own work. The promotion thesis has not been submitted for a scientific degree to any other university.

Irina Klevecka…………………………….

Date: ……………………………………….

The promotion thesis is written in English and consists of two volumes. The first volume contains introduction, four chapters, main conclusions and recommendations and the list of bibliography, 174 pages in total. The second volume contains 96 appendices, 116 pages in total. The bibliography includes 207 sources.

GENERAL DESCRIPTION OF THESIS

- **Topicality of Subject Matter**

The *subject* of the promotion thesis is short-term network traffic forecasting by means of neural networks. *Forecasting* is a special research study aimed at the evaluation of future development of objects or phenomena. Forecasts have to precede strategic planning, assess the potential directions of development, and take into account the consequences of fulfilment or failure of plans.

There are three main parameters, the prediction of which plays an important role in design, optimization and performance of modern telecommunications networks. They are:

1) Traffic loads produced by users or subscribers,
2) Number of users / subscribers or telephone lines, and
3) Demand of inhabitants for telecommunications services.

These parameters are closely interrelated and their reliable forests are determined not only by accurate computing methods but also by financial capabilities of an operator. However, from both a theoretical and practical point of view, forecasts of traffic dynamics raise the most interest. It is also one of the strategic engineering tasks specified by ITU-T.

The topicality of the problem of forecasting lies in the fact that knowledge of network performance facilitates network management, in particular – helps to develop the algorithm of preventing an overload of transmission channels. Accurate and reliable forecasts allow planning the capacity of a network on time and sustain the required level of quality of service. Besides, the properties of network traffic directly influence both capital costs of equipment and expected income of an operator.

The emergence of packet-switched Internet networks as well as transformation of traditional telephone networks into multi-service systems provides new opportunities to a user in the sphere of his/ her activities. It has changed not only the architecture of a network but also statistical nature of teletraffic, which is now characterized by the effects of self-similarity and strong long-range dependence. Therefore, new approaches to the analysis and forecasting of states and parameters of packet-switched networks are strongly required. A *non-linear neural network* is one of these methods, which is rapidly gaining recognition in time series forecasting.

We can often hear that *neural networks are more art than science*. This is primarily due to the lack of a functional algorithm for applying neural networks to time series forecasting. Because of that, the active expert assessment is still necessary at all the stages of implementation, which prevents the automation of a forecasting process. It is also important to understand that, in contrast to some linear time series methods, neural networks have not originally been developed to meet the challenges of forecasting.

The attempts to predict the traffic loads of both a conventional telephone network and a packet-switched Internet network by means of neural networks have been made many times in the past. However, most of these research papers solve a trivial task of time series approximation, with more or less success, without taking into account a general theory of neural networks and time series forecasting. Thus, the *main problems* of applying neural networks to network traffic forecasting are:

- the absence of a consistent and efficient algorithm for applying the method of neural networks in time series forecasting;
- the absence of efficient and comprehensive criteria for selecting the final forecasting model and evaluating its quality.

- **Objects of Research**

The objects of this research are the time series of different lengths and aggregation rates, which describe:

- traffic of a conventional circuit-switched telephone networks (i.e. POTS);
- traffic of a packet-switched IP network.

The real measurements of IP network traffic were taken at the transport layer. All the measurements were brought to a form suitable for further statistical analysis and forecasting.

- **Main Goal and Tasks**

The main goal of the promotion thesis is to address and solve the problems related to the use of neural networks in time series forecasting, and, after *real data* testing and comparing the produced results, give practical recommendations on the effective application of neural networks and statistical models to network traffic forecasting.

The *main tasks* of the thesis are formulated as follows:

1) *Develop a functional algorithm* for implementing neural networks to solve a short-term forecasting task, which would provide a maximum *automation* of the process of selecting an optimal model and guarantees an appropriate quality of forecasts.
2) *Give practical recommendations* on
- selecting the models and methods to produce the operative forecasts (for 24 hours ahead) and short-term forecasts (for up to two weeks ahead);
- verifying a forecasting model;
- producing operative and short-term forecasts (or *ex ante* forecasts);
- assessing the accuracy of forecasts (or *ex ante* forecasts).
3) *Produce the empirical operative and short-term forecasts* of traffic of conventional telephone networks and packet switched IP networks by applying the method of neural networks, evaluate the accuracy of forecasts and compare them with the forecasts produced by traditional linear models and "naïve" methods. The topicality of this problem is mainly related to the fact that complexity of neural networks has provoked strong opinion about their advantages over simpler linear methods in solving a forecasting task. However, none of the scientific papers or books, known to the author, has accomplished a comprehensive comparison of the results produced by non-linear neural networks and traditional linear methods.

- **Hypothesis to Defend**

The author advances the hypothesis that:

- the proposed *algorithm of solving a forecasting task with neural networks* allows automating the identification of neural network solutions, which are capable of producing reliable forecasts;
- reliable operative and short-term forecasts can be produced by using not only non-linear neural networks, but also simpler linear models such as ARIMA and exponential smoothing, if an aggregation / sampling period of packet-switched network traffic is properly selected (usually over some minutes);
- neural networks outperform linear methods in the case of predicting traffic of real telephone networks, if the observations are taken over relatively small read-out periods (e.g., over 15-minute periods following ITU-T Recommendation E.492).

- **Main Methods of Research**

In order to fulfil the indicated tasks, the following main methods were applied:

1) *Neural networks* – the method of artificial intelligence and universal approximator, which allows revealing non-linear dependencies of stochastic processes.
2) *Autoregressive integrated moving average models* (ARIMA) and *exponential smoothing* – the classic linear methods of time series forecasting, which are useful for modelling and forecasting short-range dependent processes.
3) *Spectral analysis* helps to identify the periodic and quasi-periodic components of time series. It is also useful in evaluating the Hurst exponent.
4) *Correlation analysis* allows identifying the statistical dependencies between members of a time series taken with a time shift (autocorrelation) or statistical bonds between two processes (cross-correlation).
5) *Regression analysis* is helpful in identifying the reliable model for a trend component, if any.
6) *The methods of non-parametric statistics* are useful in testing the hypothesis of stationarity or normality of time series in the absence of clear knowledge about the probability distribution of a process.

- **Structure**

The thesis consists of two volumes. The first volume contains four main chapters:

- *Chapter 1* examines the main aspects and prerequisites of network traffic forecasting.
- *Chapter 2* describes in detail the methods applied in the practical studies, such as non-linear neural networks, ARIMA models, exponential smoothing models and “naïve” methods.
- *Chapter 3* highlights the main contribution of the author to the theory of time series forecasting. The description of the proposed advanced algorithm and the main aspects of its practical realization are given in detail.
- *Chapter 4* analyzes the results of practical studies.

The first volume consists of 174 pages and contains 25 figures and 11 tables. The list of cited literature and other sources includes 207 bibliographic names. The second volume includes 96 annexes and consists of 116 pages.

- **Novelty**

The *novelty* of the thesis is attributed to the following original results:

Based on the methods of mathematical statistics and theory of neural networks, there has been developed an advanced algorithm aimed at solving the task of short-term traffic forecasting by means of neural networks.

In comparison to most classic schemes of setting a neural network to fulfil a certain task (see, for example, [9; 53, p. 84]), the proposed algorithm focuses on the estimation of forecasting abilities rather than approximation accuracy, allows automating the process of determining an optimal solution and includes three procedures, which are:

- the use of multiple cycles of weight initialization of a neural network during a training process;
- the procedure of selecting the intermediate forecasting model, taking into account the residual autocorrelation and the estimates of information criteria;

– the procedure of selecting the final forecasting model, taking into account the accuracy of *ex ante* forecasts.

The incorporation of these procedures into a classic algorithm allows identifying neural network solutions in a more effective way and significantly improves the accuracy and reliability of produced forecasts.

Neural networks belong to *heuristic* methods. It implies that the identification of optimal parameters of architecture and learning should involve intensive test-and-trial procedures, since these parameters do not comply with strict mathematical rules. The algorithm, proposed in the thesis, is also based on the experience of the author and represents the results of years of the practical studies of real time series. However, the motivation of including these procedures into the algorithm arises from the theory of neural networks and time series forecasting.

For the first time, there has been conducted a profound statistical analysis of real network traffic forecasts, produced by non-linear neural networks, classic linear methods and "naïve" methods.

Statistically significant differences in accuracy have not been identified between the forecasts produced by neural networks and classic linear models in most cases tested by the Diebold-Mariano criterion [6]. The analyzed time series are typical for these categories of traffic loads. It means that linear statistical methods would produce reliable and accurate operative / short-term forecasts for many other time series with similar statistical properties as well.

- **Practical Significance**

The proposed algorithm and practical recommendations can be applied to produce operative and short-term forecasts of telephone network traffic as well as packet-switched Internet traffic generated at the transport and application layers. In turn, the produced forecasts can be useful in planning the capacity of transmission channels, thereby providing the required level of quality of service (QoS).

The algorithms and recommendations developed by the author can be applied to other time series with similar statistical properties – for example, in producing predictions of power consumption or road traffic.

- **Approbation**

The proposed algorithm and recommendations were successfully applied to produce the forecasts of several time series describing the traffic of real telecommunications networks. Operative and short-term traffic forecasts have been obtained for:

– the circuit switched telephone network of *Augstceltne SIA,* which specializes in maintenance of corporate customers;
– the packet switched IP network of *INBOKSS SIA,* which specializes in providing free e-mail services.

The results of the thesis were declared and discussed at the following scientific conferences:

1) The 9th International Conferenc*e Reliability and Statistics in Transportation and Communication* (RelStat`09), Riga, Oct. 21-24, 2009. Topic of presentation – *Forecasting Network Traffic: A Comparison of Neural Networks and Linear Models.*
2) The 50th International Scientific Conference of Riga Technical University, Riga, Oct.14-16, 2009. Topic of presentation– *An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks.*

3) The 28th Annual International Symposium on Forecasting, Nice (France), June 22-25, 2008. Topic of presentation – Preprocessing *of Input Data of Neural Networks: The Case of Predicting Telecommunication Network Traffic.*

The other reports at scientific conferences associated with the subject of the thesis:

4) The 18th European Regional Conference of the International Telecommunications Society, Istanbul (Turkey), Sep. 4-6, 2007. Topic of presentation – *Perspective Evaluation of the Electronic Communications Market in Latvia*.
5) The 6th International Conference *Reliability and Statistics in Transportation and Communication* (RelStat'06), Riga (Latvia), Oct. 25-28, 2006. The theme of presentation – *Forecasting Methods and Long-term Evaluation of the Electronic Communications Market in Latvia.*
6) The 17th European Regional Conference of the International Telecommunications Society. Amsterdam (Netherlands), Aug. 22-24, 2006. The theme of presentation – *New Technologies and their Influence on the Universal Service Policy.*
7) The 16th European Regional Conference of the International Telecommunications Society (ITS Europe 2005), Porto (Portugal), Sep. 4-6, 2005. The theme of presentation – *The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.*
8) International Conference *Reliability and Statistics in Transportation and Communication* (RelStat'04), Riga (Latvia), Oct.14-15, 2004. The theme of presentation – *Financial Risk of Providing the Universal Telecommunications Service in Latvia*.

The papers in peer-reviewed scientific journals:

1) Klevecka, I. "Forecasting Traffic Loads: Neural Networks vs. Linear Models." *Computer Modelling and New Technologies* 14.2. (2010): 20–28. [ISSN 1407-5806; Thomson Reuters Researcher ID]
2) Klevecka, I. "An Advanced Algorithm for Forecasting Traffic Loads by Neural Networks". *Scientific Journal of Riga Technical University* (*Series "Telecommunications and Electronics"*) 9 (2009): 48-55. [ISSN 1407-8880; EBSCO Host, ProQuest, VINITI]
3) Klevecka, I., and J. Lelis. "Pre-Processing of Input Data of Neural Networks: The Case of Forecasting Telecommunication Network Traffic." Spec. issue of *Telektronikk: Telecommunications Forecasting* (in co-operation with International Institute of Forecasters) 104.3/4 (2008): 168-178. [ISSN 0085-7130; Thomson Reuters Researcher ID, ACM Digital Library]
   Cited in:
   – Nikolov, V., and V. Bogdanov**. "**Integration of Neural Networks and Expert Systems for Time Series Prediction." *Proceedings of the 11th International Conference on Computer Systems and Technologies* (*CompSysTech'10*). New York: ACM Press, 2010. 534-539. [ACM Digital Library]
   – Chulaka Gunasekara, R., et. al. „Prophetia: Artificial Intelligence for TravelBox® Technology." *Advances in Computational Intelligence*. Eds. Wen Yu and Edgar N. Sanchez. Berlin: Springer-Verlag, 2009. 21-34. [Springer Link]

The abstracts in the proceedings of international scientific conferences:

4) Klevecka, I. "Forecasting Network Traffic: A Comparison of Neural Networks and Linear Models." *Abstracts of the 9th International Conference "Reliability and Statistics in Transportation and Communication"* (*RelStat'09*). Oct. 2009, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2009. 36. [ISBN 978-9984-818-22-1; Thomson Reuters Researcher ID]
5) Klevecka, I., and J. Lelis. „Preprocessing of Input Data of Neural Networks: The Case of Predicting Telecommunication Network Traffic." *Program and Abstracts of the 28th Annual*

International Symposium on Forecasting. June 2008, Nice, France. France: International Institute of Forecasters, 2008. 29. [ISSN 1997-4116; Thomson Reuters Researcher ID]

The other papers in peer-reviewed scientific journals associated with the subject of the thesis:

6) Klevecka, I., and J. Lelis. “Application of Extrapolation Methods to the Technology Diffusion Forecasting.” *Scientific Proceedings of Riga Technical University (Series "Telecommunications and Electronics")* 7 (2007): 52-59. [ISSN 1407-8880; ProQuest, VINITY]
7) Klevecka, I. and J. Lelis. “Financial Risk of Providing the Universal Telecommunications Service in Latvia.” *Transport and Telecommunication* 6.1 (2005):139-144. [ISSN 1407-6160; Thomson Reuters ResearcherID].

The papers in the proceedings of scientific conferences associated with the subject of the thesis:

8) Klevecka, I., and J. Lelis. “Perspective Evaluation of the Electronic Communications Market in Latvia.” *Proceedings of the 18^{th} European Regional ITS Conference.* Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 1-37. CD-ROM. [Thomson Reuters ResearcherID].
9) Klevecka, I., and J. Lelis. “Forecasting Methods and Long-term Evaluation of the Electronic Communications Market in Latvia.” *Proceedings of the 6^{th} International Conference “Reliability and Statistics in Transportation and Communication”*. Oct. 2006, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2006. 37-45. [ISBN 9984-9865-9-4; Thomson Reuters Researcher ID].
10) Klevecka, I., J. Lelis, and J. Ulmanis. “The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.” *Papers of the 16^{th} European Regional ITS Conference*. Sep. 2005, Porto, Portugal. Berlin: International Telecommunications Society, 2005. 1-13. CD-ROM. [Thomson Reuters Researcher ID].

The other abstracts in proceedings of international scientific conferences associated with the subject of the thesis:

11) Klevecka, I., and J. Lelis. “Perspective Evaluation of the Electronic Communications Market in Latvia.” *Abstracts of the 18^{th} European Regional ITS Conference*. Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 118-119. [Thomson Reuters Researcher ID]
12) Klevecka, I., J. Lelis, R. Bergmanis, and G. Macs. “New Technologies and their Influence on the Universal Service Policy.” *Abstract Booklet of the 17^{th} European Regional ITS Conference*. Aug. 2006, Amsterdam, Netherlands. Berlin: International Telecommunications Society, 2006. 59-60. [ISBN 90-8559-205-4, Thomson Reuters Researcher ID]
13) Klevecka, I., J. Lelis, and J. Ulmanis. “The Necessity of Including Mobile Telephony in a Minimum Set of Universal Service.” *Abstracts of the 16^{th} ITS European Regional Conference.* Sep. 2005, Porto, Portugal. Berlin: International Telecommunications Society, 2005. 16-17. [*Thomson* Reuters Researcher ID]
14) Klevecka, I., and J. Lelis. “Financial Risk of Providing the Universal Telecommunications Service in Latvia.” *Programme and Abstracts of the International Conference RelStat’04.”* Oct. 2004, Riga, Latvia. Riga: Transport and Telecommunication Institute, 2004. 54. [ISBN 9984-668-76-2; Thomson Reuters Researcher ID].

SYNOPSIS OF THESIS

CHAPTER 1 MAIN ASPECTS OF NETWORK TRAFFIC FORECASTING

Chapter 1 examines the aspects of applying the models and methods of time series theory to network traffic forecasting. Main statistical properties and genesis of time series are discussed. The chapter also contains a review of scientific papers dedicated to traffic forecasting carried out by means of neural networks.

- **Concept of Time Series**

In modeling and forecasting we usually assume that network traffic is represented as time series. A *time series* is a time-ordered sequence of observation values of a physical variable, usually made at equally spaced time intervals Δt, represented as a set of discrete values $x(t_1), x(t_2), \ldots x(t_N)$. In statistical analysis, this sequence of N observations is often considered as a sample taken from a longer general sequence of random numbers. The observations or elements of time series are typically labeled in accordance with a time moment they refer to (e.g., x_1, x_2, x_3). Thus, the order of the elements of a time series is of great importance.

It is necessary to keep in mind that, unlike the observations of random variables, the elements of time series are not statistically independent [49, p.780]. Some rules and properties of the statistical analysis of random samples cannot be applied to time series, and this requires the implementattion of specific methods and approaches. On the other hand, the correlations between time series observations set up a specific base for predicting an analyzed variable, i.e. for producing the estimate $\hat{x}(N+L)$ of an unknown value $x(N+L)$ taking into account the historical values $x(t_1), x(t_2), \ldots x(t_N)$, where N is the length of an analyzed time series and L is a forecasting horizon.

The *genesis of observations* is the structure and classification of the main factors, under the influence of which the values of time series are formed. There are four types of such factors and components of time series [49, p.781; 52, p.242; 57, p.354]:

1) *Long-term factors* form the general dynamic tendency of an analyzed parameter $x(t)$. This tendency is usually described by a deterministic non-random function called ***a trend***.
2) *Seasonal factors* determine the tendency, which changes regularly during a certain period (a day, week, month etc.). Since this function has to be periodic (with periods, proportional to “seasons”), its analytical expression involves the use of trigonometric functions.
3) *Cyclic factors* determine the longer periods of relative rise and fall. The cyclic component may contain the cycles of economic, demographic or astrophysical nature, and varies in amplitude and length. As a rule, the length of a cyclical component exceeds one year.
4) *Random (irregular) factors* determine the stochastic nature of time series members. A random component is formed as a result of superposition of many external factors, which are not involved in the formation of a deterministic component.

The deterministic components of network traffic can be classified as follows [19, p.42; 45, p.44]:

1) *24-hour cycle*. It has been known for a long time that sigmoidal models (logistic model, Gompertz model, etc. [22; 23]) are optimal for describing the traffic dynamics within a day [58, p.202].
4) *Weekly cycle* is usually characterized by the decrease of traffic during weekends, and can be described by means of a Fourier series.
5) *Annual cycle*. It is believed that the level of network traffic is higher at the beginning of a month, after a festival season and at the beginning of each quarterly period.

6) *Linear trend*. The overall traffic increases year by year due to the influence of technical progress and socio-economic factors.

If we identify accurately a deterministic component, then the residuals of a time series will be an irregular *stochastic component*. Its behavior cannot be fully predicted in advance. In other words, every observation gives only one option among many possible. In order to describe and predict this component of a time series, the methods and concepts of probability theory are involved.

- **Aspects of Forecasting Transmission Capacity of Packet Switched Networks**

Transmission capacity is one of the most important parameters characterizing the quality of networks. For the purposes of forecasting, the measurements made at the *transport layer* are usually considered. These time series describe either the number of arriving packets or the level of traffic / transmission rate measured in bytes over discrete time intervals.

The methods of network traffic forecasting are partially determined by ITU-T Recommendations E.506 [16] and E.507 [18]. Even these recommendations have been developed for ISDN networks, some of the forecasting methods described there can be still applied to modern telecommunications networks. These methods are the autoregressive integrated moving average models (ARIMA) [3] and exponential smoothing [12].

The traffic of packet switched IP networks is characterized by such statistical effects as self-similarity and long-range dependence [20, p.150; 28; 37; 41; 56, p.53]. The stronger post-effects, the longer is a forecasting horizon, for which reliable forecasts can be produced. However, it is also a disadvantage, since the estimation and selection of an adequate model, which would take into account all significant correlations between the members of a time series, becomes labor-intensive [40].

At the same time, there has been disseminated intensively the *myth* regarding impossibility of applying traditional linear methods to predicting packet switched IP traffic and the necessity to use more complicated non-linear methods such as neural networks.

Indeed, neural networks offer some additional opportunities in modelling non-linear processes and recognizing chaotic behaviour. Owing to their great flexibility, these networks can recognize a variety of structures. However, numerous practical studies dedicated to traffic forecasting usually miss the fact that fractal properties of packet-switched traffic have a significant influence on a forecasting process only in the case of measurements *on a very large scale* – over the aggregation periods varying from milliseconds to some minutes. This fact has been confirmed by the author’s practical studies as well as a number of other research papers [36, 39, 40].

Fig. 1 is the illustration of this idea, where the real measurements of transmission rate are shown against an aggregation period. A visual analysis of the traffic aggregated over 1 and 10 seconds reveals *stochastic self-similarity* [37]. However, increasing the aggregation period up to one minute, and then – up to five minutes, we can see that a traffic trace becomes more even, its variance significantly decreases and the influence of deterministic components starts to play a leading role.

From the point of view of time series forecasting, a very fine sampling scale is unreasonable. In this case, the selection of a relevant statistical model is complicated due to the strong influence of autocorrelations between distant observations of times series as well as extraneous noises and anomalous outliers, which unavoidably accompany the large-scale measurements. Besides, an aggregation / sampling period also determines a forecasting horizon, for which reliable forecasts can be produced. In other words, a potential forecasting horizon for time series, aggregated over, for example, one-second periods is different from the one for time series aggregated over 24-hour intervals.

At present, real-time forecasting with neural networks is hard to implement in practice. Apart from the necessity to select and evaluate many parameters, often – in empirical way, some substantial time resources are necessary for training a neural network. Therefore, taking into account ITU-T Recommendation E.492 [17], it is desirable to average measurements of network traffic over 15-minute and / or one-hour read-out intervals. In doing so, the main factors determining the statistical structure of ***real*** network traffic are seasonal effects and monotonous trends, the main reasons of which are human behaviour and technical progress (see Fig. 2).

It has been shown in the practical part of the thesis that statistical properties of such time series become similar to statistical properties of traditional voice traffic. At an intuitive level, it gives us the opportunity to assume that the methods of modelling and forecasting of these processes would be similar as well, if the appropriate length of a read-out period is selected.

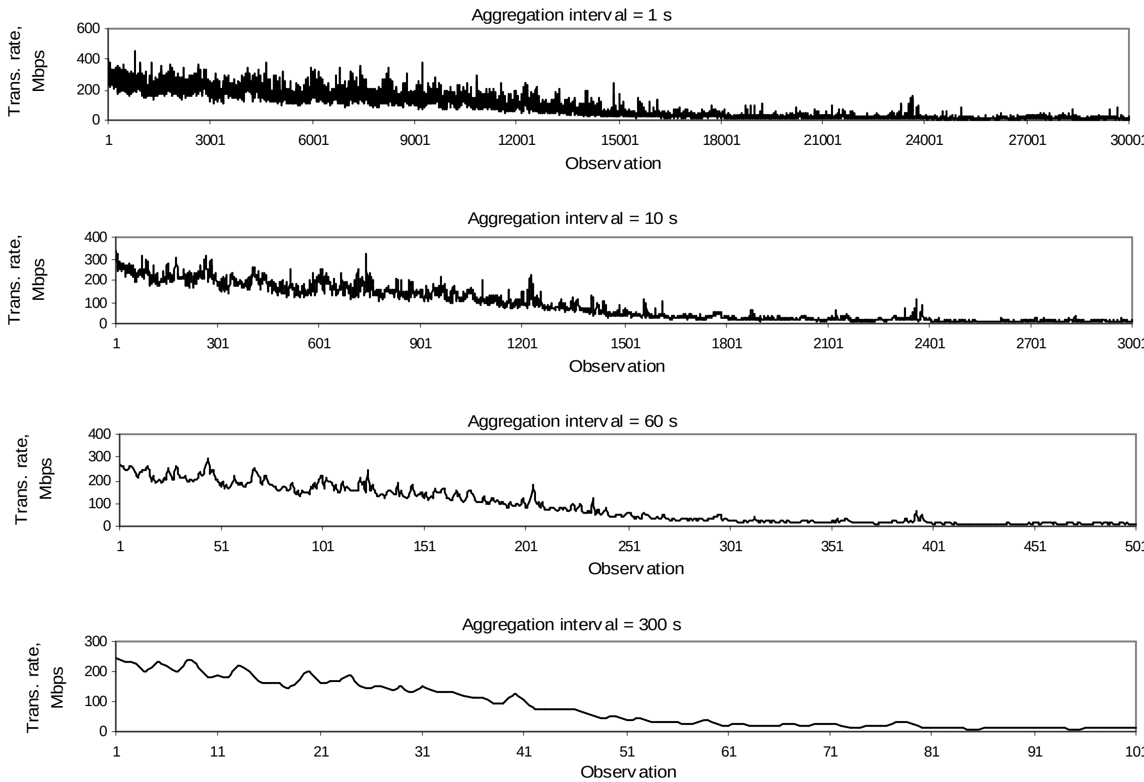


Fig. 1 Real packet-switched traffic recorded over different aggregation periods

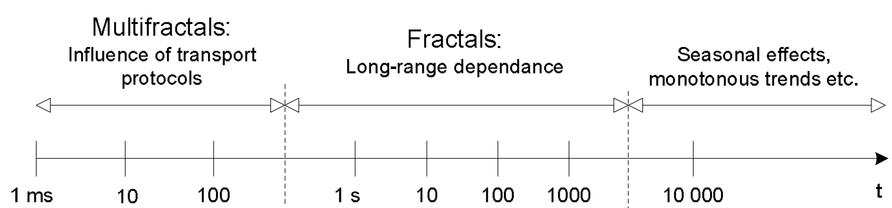


Fig. 2 Statistical effects of packet-switched traffic depending on a time scale [10, with author's amendments]

The main accent of this thesis has been put on the application of neural networks (i.e., a multilayer perceptron) to forecasting the traffic of both a traditional telephone network and a packet-switched IP network. Following the principle of Ockham's razor – *choose a parsimonious model,* – it makes sense to compare the accuracy of forecasts produced by means of non-linear models with those produced by traditional linear models. To pursue this goal, the models of ARIMA and exponential smoothing (as the methods recommended by the ITU-T) as well as "naive" methods, have been chosen. If there are no statistically significant differences between the forecasts produced by neural networks and linear methods, then the application of such a complicated and time-consuming method as neural networks becomes unnecessary.

CHAPTER 2 METHODS OF NETWORK TRAFFIC FORECASTING

Chapter 2 gives an overview of the models and methods of network traffic forecasting applied in the practical part of the thesis such as non-linear neural networks (multilayer perceptron), seasonal autoregressive integrated moving average (SARIMA), seasonal exponential smoothing and “naive” forecasting methods.

- **Neural Networks**

A *neural network* is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use [13, p.2]. The development of *artificial neural networks* started in the beginning of the 20^{th} century but only in the nineties, when some theoretical barriers were overcome and computing systems became powerful enough, neural networks have gained wide recognition. Even though neural networks can be implemented as fast hardware devices (and these realizations do exist in real life), most practical studies are performed by applying software simulations on conventional PCs. Software simulations provide low-cost flexible environment, which is sufficient for many real-life applications. Neural networks are acquiring popularity in the field of telecommunications as well, where they help solving various problems such as switching management, traffic management, routing, channel allocation in mobile transmission systems, etc. [54, p. 60].

The incorporation of time into the operation of a neural network allows to follow statistical variations in different processes described by time series, such as speech signals, radar signals, fluctuations in stock market processes, teletraffic processes and many others. The discussion of the role of time in neural processing can be found in fundamental paper [7].

The temporal structure of an analyzed sample is usually built into the operation of a neural network in *implicit* way. In this case, a *static* neural network (e.g. a multilayer perceptron) is provided with *dynamic* properties [13, p.636]. For a neural network to be dynamic, it must be given memory which may be divided into short-term and long-term memory [35]. *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 built into the structure of a 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.

Two types of neural networks, a back-propagation network (multilayer perceptron) and a radial basis function network, are considered to be suitable for temporal processing. Due to a number of reasons, the latter has not gained acceptance[1]. At the same time, numerous practical studies have proved that a multilayer perceptron solves successfully many various tasks such as pattern recognition, regression, function approximation, time series forecasting, cluster analysis, etc. Therefore, further we will focus on this class of neural networks.

A multilayer perceptron usually consists of multiple sensor elements (i.e., input nodes) forming an input layer, one or several hidden layers containing computational nodes, and one output layer. It is often trained according to the *error back propagation algorithm*. It is a supervised training algorithm, which is based on the *error correction training rule* and requires two computational flows – a direct one and a backward one – through all the layers of a network.

Temporal pattern recognition demands 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

---

[1] Both a *radial-basis function* (RBF) network and a *multilayer perceptron* (MPP) belong to the class of universal approximations. Due to that, there always exists an RBF network capable of accurately mimicking a specified MLP, and vice versa. An MLP develops global approximations to nonlinear input-output mapping. In turn, an RBF network constructs local approximations using exponentially decaying localized nonlinearities (e.g., Gaussian functions). The latter is the reason of the popularity of MLPs – in order to approximate a nonlinear input-output mapping at the same degree of accuracy, an MLP requires the smaller number of parameters to determine and consequently, less time for completing a training cycle, in comparison with an RBF network [13, p. 293].

its past values. Fig. 3 shows the diagram of a nonlinear filter built on a static neural network. Given a specific input signal consisting of the current time series value $x(t)$ and the Ω past values $\{x(t-1), x(t-2) \mathbf{K}, x(t-\Omega)\}$ stored in a delay line memory of order Ω, the free parameters are adjusted to minimize the training error between the output of the network, $y(t)$, and the desired response, $d(t)$ [13, p.645]. The structure shown in Fig. 3 can be implemented at the level of a single neuron or a network of neurons.

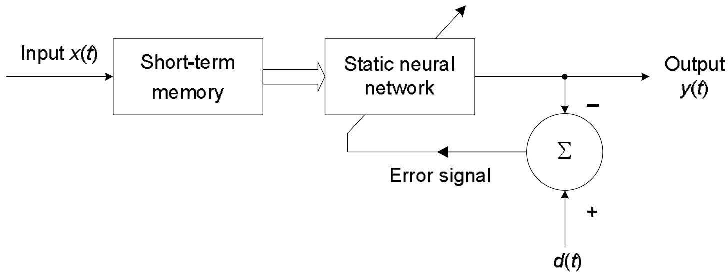


Fig. 3 Temporal processing – nonlinear filter built on a static neural network [13, p.643]

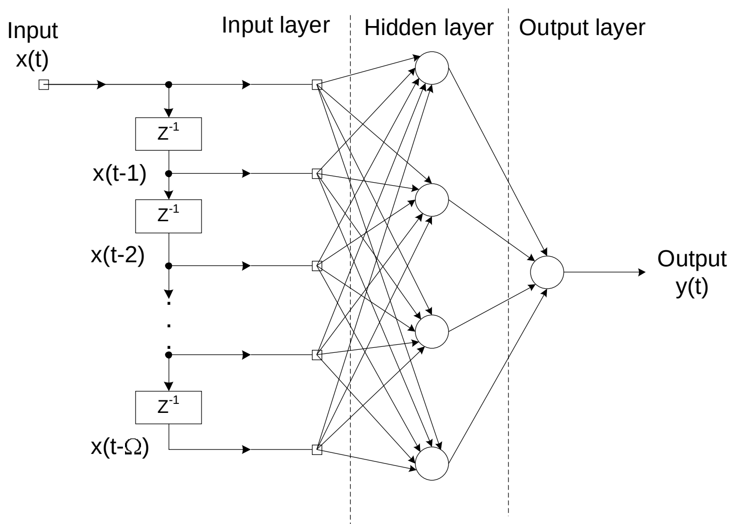


Fig. 4 Time lagged feed-forward network[2] [13, p.644; 35]

The diagram of a *time lagged feed-forward network* is shown in Fig. 4. It is a powerful nonlinear filter consisting of a tapped delay memory of order Ω and a multilayer perceptron. The standard back propagation algorithm can be used to train this type of neural networks. At time *t*, the temporal pattern applied to the input layer of the network is the signal vector:

$$x(t) = \{x(t), x(t-1), x(t-2) \mathbf{K}, x(t-\Omega)\}^{N^{TR}}, \tag{1}$$

where
N^{TR} – the length of a time series or a training subset.

Eq. (1) describes the state of the nonlinear filter at time *t*. One training *epoch* consists of a sequence of patterns (states), the number of which is determined by the memory order Ω and the size of a training sample N^{TR}. The output of a nonlinear filter, assuming that a multilayer perceptron has a single hidden layer and one output neuron, is computed from:

[2] The bias levels are omitted for convenience of representation.

$$y(t) = j_2\left(b_j + \sum_{j=1}^{m_1} w_j j_1\left(\sum_{z=0}^{\Omega} w_{j,z}\, x_{t-z} + b_{j,z}\right)\right), \tag{2}$$

where

j_1 – activation function of a hidden layer;

j_2 – activation function of an output layer;

$w_{j,z}$ – synaptic weight of input synapse ζ of hidden neuron j;

w_j – synaptic weight of input synapse j of an output neuron;

$b_{j,z}$ and b_j – biases;

m_1 – number of hidden neurons;

Ω – order of linear delay memory.

- **Autoregressive Integrated Moving Average Models**

One of the most popular class of *linear time series models* refers to *autoregressive moving average models* (ARMA), including purely autoregressive (AR) and purely moving-average (MA) models as special cases. These models were initially introduced in the twenties of the last century but have been using actively only since 1970, when the fundamental book of Box and Jenkins [3] was published.

In real-life network traffic modelling and forecasting, it is necessary to put an accent on the *seasonal modifications* of linear models, which are able to model and forecast the periodic time series. The application of non-seasonal models to seasonal time series can lead to the erroneous conclusions that linear models are not capable to model and make reliable forecasts of network traffic dynamics.

The *seasonal autoregressive integrated moving average model*, denoted as $SARIMA(p, d, q)(P, D, Q)_s$, is given by [3, p.305]:

$$f_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D x_t = q_q(B)\Theta_Q(B^s)e_t, \tag{3}$$

where

s – period of the seasonal component;

p – order of the non-seasonal autoregressive operator;

q –order of the non-seasonal moving average operator;

d –order of the non-seasonal differencing operator;

P – order of the seasonal autoregressive operator;

Q – order of the seasonal moving average operator;

D – order of the seasonal differencing operator;

$\nabla = \nabla_1 = 1 - B$ – non-seasonal differencing operator;

$\nabla_s = 1 - B^s$ – seasonal differencing operator;

$f(B)$ и $q(B)$ – polynomials in B of order p и q, respectively, which satisfy the conditions of stationarity and invertibility;

$\Phi(B^s), \Theta(B^s)$ – polynomials in B^s of order P и Q, respectively, which satisfy the conditions of stationarity and invertibility;

$e_t \sim WN(0, s^2)$– a white noise process.

Let us assume that all the values of a time series $x_t, x_{t-1}, \mathbf{K}$ are known until time moment t. Then, the minimal mean squared error forecast $\hat{x}_t(L), L \geq 1$ at lead time L and origin t is the *conditional expectation* of x_{t+L} [3, p.306]:

$$\hat{x}_t(L) = [x_{t+L}] = E[x_{t+L} \mid q, \Theta, x_t, x_{t-1}, \mathbf{K}] \tag{4}$$

Box and Jenkins proved that a forecast of the minimal mean squared error can be calculated directly from the model represented as a *difference equation*. For example, for the seasonal process of period $s = 12$, the forecast is given by [3, p.306]

$$\hat{x}(L) = [x_{t+L}] = x_{t+L-1} + x_{t+L-12} - x_{t+L-13} + e_{t+1} - \theta\, e_{t+L-1} - \Theta e_{t+L-12} + \theta\, \Theta e_{t+L-13} \qquad (5)$$

After inserting the values of parameters θ и Θ into Eq. (5), we immediately obtain a minimal mean squared error forecast at lead time *L* calculated at origin *t* [3, p. 307]. The parameters of Eq. (5) are assumed to be known precisely, and a time series $x_t, x_{t-1}, \ldots$ is assumed to extend into the remote past.

- **Exponential Smoothing**

The method of exponential smoothing allow to produce the description of a process, according to which the last historical observations have the larger weights as compared to the earlier ones, and the weights decrease exponentially.

The simple exponential smoothing model is defined as [12]:

$$S_t = S_{t-1} + \alpha\, e_t \qquad (6)$$

where

S_t – smoothed level of the series computed after x_t is observed;

α – smoothing parameter for the level of the time series;

e_t – one-step-ahead forecast error; $e_t = x_t - \hat{x}_{t-1}(1)$;

x_t – observed value of the time series at moment *t*;

$\hat{x}_{t-1}(1)$ - one-step-ahead forecast from origin $t-1$;

There exist different modifications of exponential smoothing aimed at the analysis of non-stationary time series with linear and nonlinear trend components, as well as seasonal time series with multiplicative and additive seasonality. Only the models with additive seasonality and / or a linear trend will be considered further in this section as they are the most suitable for the analysis of teletraffic processes. In exponential smoothing models, the additive seasonal component and linear trend of a time series are calculated from [12]:

$$I_t = I_{t-s} + \delta(1-\alpha)e_t \qquad (7)$$

$$T_t = T_{t-1} + \alpha\, \gamma\, e_t \qquad (8)$$

where

I_t – smoothed value of the seasonal component at the end of period *t*;

T_t – smoothed value of the trend component at the end of period *t*;

δ – smoothing parameter for the seasonal component;

γ – smoothing parameter for the trend component.

The estimation of a time series incorporating seasonal and / or trend components, is based on *decomposition*. The seasonal and trend components are estimated at each time moment independently by using a simple exponential smoothing model with parameters δ and γ. This method is known as the *Holt-Winters' exponential smoothing* and is based on three smoothing equations – one for the level, one for trend and one for seasonality.

For the time series with additive seasonality and a linear trend, the forecast for horizon *L* is given by [12]:

$$\hat{x}_t(L) = S_t + I_{t-s+L} + L \cdot T_t \tag{9}$$

where
$\hat{x}_t(L)$ – forecast produced for horizon L from origin t.

The complete classification as well as the description of the methods of determining the optimal parameters of exponential smoothing models is given in [12].

- **Naïve Forecasting**

It is desirable to compare the results, produced by various forecasting models, with so called *naïve fore*casts. In practical analysis of time series, a naïve forecast serves as the simplest *benchmark* forecast and can be produced in several different ways, of which the following two were applied in the thesis:

1) The naïve forecast is the sample mean of an examined time series [33]:

$$\hat{x}_t^{\,a}(L) = E[x_{t+L} \,|x_t, x_{t-1}, \mathbf{K}, x_1] = \hat{m} \text{ for } L \geq 1 \tag{10}$$

2) The seasonal naïve forecast can be used with seasonal data and postulates that the forecast for one period ahead is equal to the same value of the last historical period of a time series [44]:

$$\hat{x}_t^{\,b}(L) = E[x_{t+L} \,|x_t, x_{t-s}, \mathbf{K}, x_1] = x_{t+L-s} \text{ for } L \geq 1 \tag{11}$$

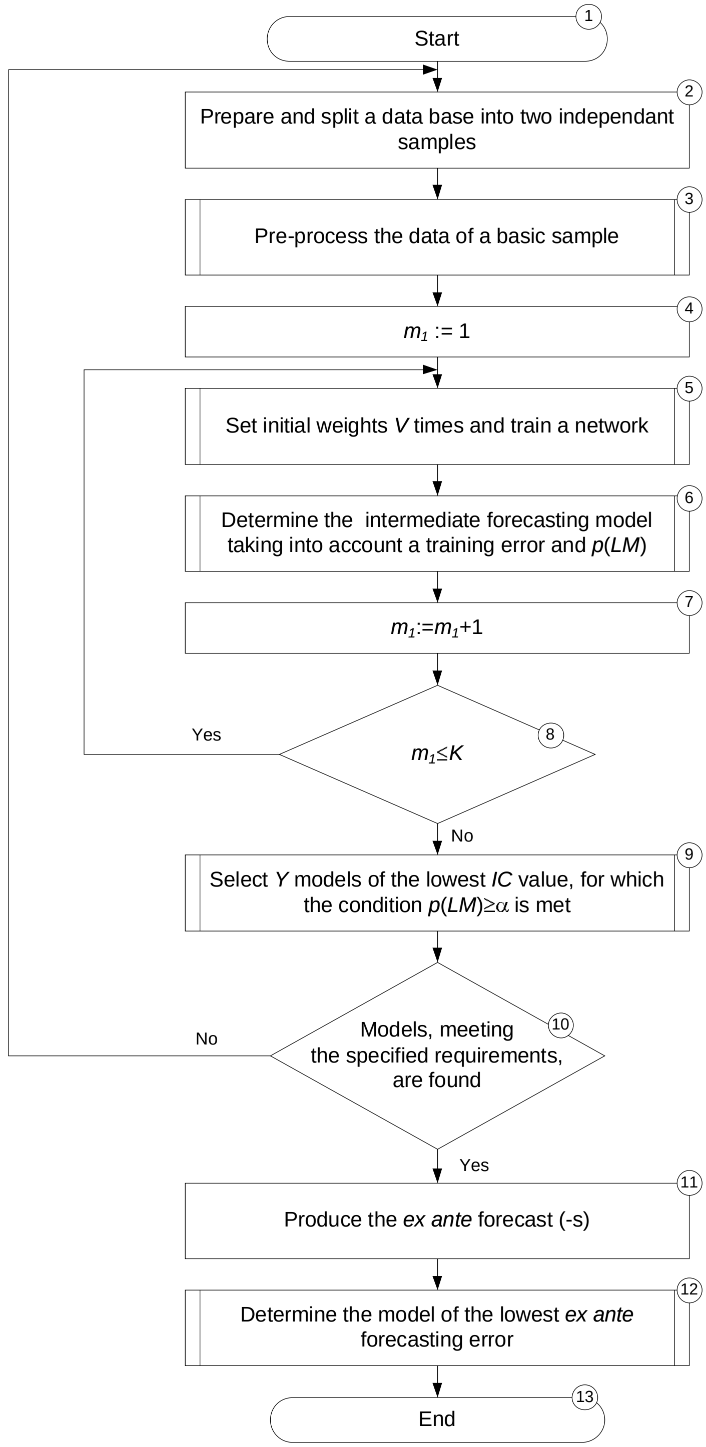
If the comparison of forecasts produced by neural networks or statistical models with those produced by naïve methods does not reveal any statistically significant differences, then, perhaps, the use of the models in further forecasting of this particular time series is not required.

CHAPTER 3 ALGORITHM FOR SOLVING A FORECASTING TASK WITH NEURAL NETWORKS

Chapter 3 describes the author's developed algorithm aimed at solving a time series forecasting task by means of neural networks. The innovative aspects of the algorithm are considered in detail. The chapter examines the methods of determining the parameters of multilayer perceptrons and provides some recommendations on the practical use of the algorithm, e.g., the preparation and pre-processing of input data, development of forecasts, estimation of the accuracy of forecasts, etc.

As mentioned above, unlike classic linear methods, the method of neural networks was not initially aimed at modelling and forecasting time series. When applied to time series forecasting, neural networks are often criticized for the necessity to set many different parameters through test-and-trial procedures, complications with producing and replicating a stable solution, high probability of over-learning, high demands for time resources and computational capacities. Besides, it is necessary to keep in mind that neural networks are sensitive to the quality of input data [24, 27, 55].

In order to facilitate and automate the process of time series modelling and forecasting, and compensate for the problems associated with instability of a produced solution, an advanced algorithm has been developed in the thesis.



Legend

m_1 – number of hidden neurons of a two-layer perceptron
K – maximum number of hidden neurons
V – total number of weight initialization cycles
IC – information criterion
LM – Lagrange multiplier type test
p(*LM*) *– p-value* (*observed significance*) *of the LM-type test*
α – significance level

Fig. 5 Advanced algorithm aimed at solving a forecasting task by means of neural networks

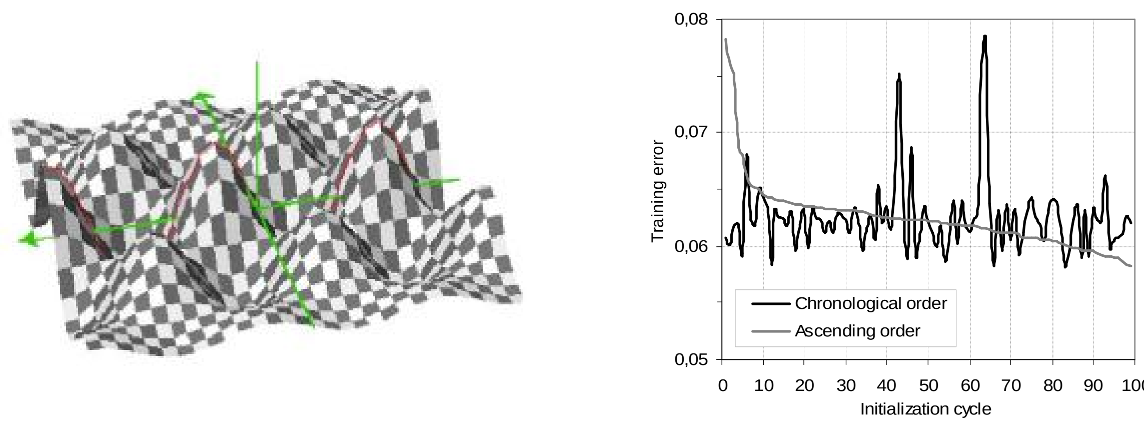
The algorithm, the main block diagram of which is shown in Fig. 5, produces the operative / short-term forecasts of traffic loads represented as univariate / multivariate time series. In contrast to some classic algorithms for accomplishing a pre-defined task (see, for example, [9; 53, p.84])), the proposed algorithm incorporates three procedures:

– implementation of multiple cycles of weight initialization of a neural network during a training process;
– method of selecting the intermediate forecasting models, taking into account the level of residual autocorrelation and the estimates of information criteria;
– method of selecting the final forecasting model, taking into account the accuracy of *ex ante* forecasts.

Let us consider the theoretical arguments in defence of the necessity of implementing these procedures to identify a relevant forecasting model.

- **Implementation of multiple cycles of weight initialization of a neural network**

It is required to set some initial values to all the weights and biases of a neural network before a training process starts. The aim of initialization is, probably, to find the best approximation to an optimal solution and, in this way, to decrease training time and facilitate the convergence of a training algorithm.



(a) **(b)**

Fig. 6 Illustration of the necessity of implementing multiple initialization cycles: (a) a simplified example of two-dimensional error surface where a vertical axis represents an error. This demonstrates a key problem – several local minima can exist on the training error surface; (b) the training errors of a neural network (without applying cross-validation) produced as a result of a hundred cycles of weight initialization[3]

If the initial weights and biases are set to large values, then neurons usually approach the saturation level very quickly. In this case, the local gradients, calculated according to the back-propagation algorithm, take small values, which, in turn, would significantly increase training time. If the initial values are set to small values, then the algorithm works very slowly about the origin of the error surface. This is specifically true in the case of anti-symmetric activation functions such as hyperbolic tangent.

During the last two decades many heuristic methods of weight initialization have been proposed, some of which are described in Chapter 3 of the thesis. Despite this, the optimal solution of

[3] The values of training errors are displayed in chronological order of their evaluation at the end of each training epoch as well as sorted in descending order to facilitate their comparison.

this issue has not still been found. Due to its simplicity, the most common method is the random initialization of weights and biases from a uniformly or normally distributed narrow range of small values.

Regardless of the initialization method, starting values of weight coefficients influence the final result of training. This is due to the properties of a training algorithm as well as the fact that several local minima can exist on the error surface (see Fig. 6-a). Therefore, in order to find an actual global minimum, it is necessary to train one and the same neural network multiple times under the same conditions, changing only the initial values of weights and biases.

In spite of these problems, most researchers still do not pay a proper attention to this aspect. It is possible to find references to 5 [21], 10 [42], 25 [25], 50 [8; 46] and 100 [25] cycles of initialization. However, most practical studies restrict the number of initialization cycles to one, and this can mislead a researcher regarding the adequacy of a produced solution.

The example shown in Fig. 6-b illustrates the uncertainty in producing a final solution and its dependence on the starting values of weights and biases. The training errors, shown here, are produced as a result of training a neural network of the same architecture and applying the same training parameters but changing the initial values of weights and biases. It is easy to notice that the difference between the largest and smallest error comprises more than 25 per cent, which can significantly influence the identification of a relevant forecasting model and the accuracy of produced forecasts.

The number of initialization cycles is usually chosen mandatory, and depends on the complexity of a task as well as on time resources a researcher has on his / her disposal.

- **Selection of the intermediate forecasting models, taking into account the residual autocorrelation and estimates of information criteria**

According to the developed algorithm, the selection of the intermediate forecasting model / models among the trained networks, with a certain number of hidden neurons m_1, is carried out taking into account the level of *residual autocorrelation* and the value of an *information criterion*.

Some standard parameters, such as the correlation coefficient (R), mean squared error (MSE), mean absolute error (MAE), are traditionally used to evaluate a general forecasting ability of a statistical model. However, these parameters provide little information about the accuracy of a fitted model, and are also useless in identifying the statically significant differences between the forecasts produced by various methods [4, 6, 26].

The most accurate indicator of the adequacy of a forecasting model can be the absence of *autocorrelation in residuals*. The residuals of a fitted model are defined as *n* differences given by $e_t = x_t - \hat{x}_t, t = 1, 2, \mathbf{K}, n$, where x_t is the observed value and $\hat{x}_t$ is a corresponding predicted value produced by means of a fitted statistical model [32, p. 94]. These differences cannot be explained by a forecasting model. Therefore, we can consider residuals e_t to be observed errors.

The acceptance of the hypothesis of no autocorrelation in residuals at a pre-defined significance level means that the residuals are similar to *white noise* and further analysis will not discover any statistically significant dependencies. In classic regression analysis, the *Durbin-Watson* criterion [50, p.245] is traditionally used for testing the autocorrelation of residuals. However, this test is not suitable if the regressor is a lagged explanatory variable [51, p.256]. For the same reason, the *Box-Pierce* and *Ljung-Box* [30] criteria cannot be applied to neural networks, although the last one is widely used and, despite its theoretical inconsistence, is still included in most statistical and econometric software packages.

At present, the most relevant estimate of the residual autocorrelation of neural networks (as well as ARIMA models) is a powerful *Lagrange Multiplier* (*LM*) *type test* [34], which is also known as *Breusch–Godfrey test*. The LM-type test belongs to classic asymptotic tests and is capable to identify the autocorrelation of any order.

In turn, the use of *information criteria* is based on one of the main idea of time series forecasting – “chose a parsimonious model” (known as the *Ockham’s razor principle*). It means that, all other things being equal, one should prefer the model with the fewest free parameters.

The mean squared residuals usually decrease once the model becomes more "complicated" with addition of new free parameters. Increasing the number of free parameters of a neural network (which is associated with the addition of new neurons / layers of neurons), one can fit a model to historical data with infinite accuracy. The ***universal approximation theorem*** [5; 15] explains this property of neural networks. However, such a neural network may have a poor ability to make generalizations due to ***over-training*** [13, p.206]. Besides, once a certain limit is reached, the gain in accuracy of fitting with addition of new parameters tends to be insignificant. On the other hand, time required for selecting the optimal values of free parameters can lead to a sharp decrease in the performance of a network. Therefore, it is very important to look for the balance between the preciseness of approximation and the complexity of a statistical model.

Information criteria have been found to be quite useful in solving this problem. The estimate of the criterion consists from the penalty for poor fitting and the penalty for over-parameterization. The most popular criteria of this type, applied in the practical part of the thesis, are the *Akaike's information criterion* (AIC) and the *Bayesian information criterion* (BIC) given by [11, p.38; 38, p.373]:

$$AIC(l) = N_{ef} \ln \hat{s}_e^2 + 2l \tag{12}$$

$$BIC(l) = N_{ef} \ln \hat{s}_e^2 + 2l + l \ln(N_{ef}) \tag{13}$$

where

N_{ef} – number of effective observations, to which the model is fitted;

l – number of adjusted parameters;

$\hat{s}_e^2$ – estimate of the residual variance, $\hat{s}_e^2 = \dfrac{\sum_{t=1}^{Nef} e_t^2}{N_{ef}}$.

Information criteria are evaluated separately for each analyzed specification (architecture) of neural networks. The models that possess the lowest value of the criterion should be selected for further analysis. It has been also noticed that, in practice, the *BIC* "selects" very parsimonious models with only few parameters. Therefore, this criterion is often used for non-linear models, where insignificant gain in fitting quality is directly related to the necessity of calculating a large number of additional parameters.

In the practical part of the thesis the criteria given by Eqs. (12) and (13) were applied. However, some other modifications of information criteria have been proposed as well. The *Schwarz's Bayesian criterion* (SBC) [38, p.376] and the *Hannan-Quinn* criterion [32, p. 86] are among them.

- **Selection of the final forecasting model, taking into account the accuracy of *ex ante* forecasts**

If the models, meeting the above specified conditions, are found, it is required to test their generalization ability (i.e., the ability to produce reliable forecasts) for an independent test set, which is not involved in training. The forecast developed for an independent test set we will call an *ex ante forecast* or *a pseudo-forecast*. The necessity *of ex ante forecasting* hinges upon the fact that even if a neural network provides a high accuracy of approximation and uncorrelated residuals, it would be still over-trained on historical data.

The approach of splitting a time series into two independent subsets has gained wide acceptance in the practical studies dedicated to time series forecasting (see, e.g., [31; 43]) but it is still rarely used in the case of neural networks. The first, largest data subset, called the *basic* or *retrospective sample*, is used to select and verify a statistical model. The second, *ex ante forecasting sample* is used to examine the quality of *ex ante* forecasts, comparing them against historical data. It provides the opportunity to evaluate independently the forecasting ability of the model fitted to the basic sample.

The last historical values of an analyzed time series are traditionally used for developing the *ex ante* forecasting sample. However, it is necessary to keep in mind, that these observations influence the direction of the actual *real-life* forecast much more than the earlier ones. Therefore, the last historical data are the most valuable for the process of selecting an appropriate forecasting model, and using them as a testing sample is not always reasonable.

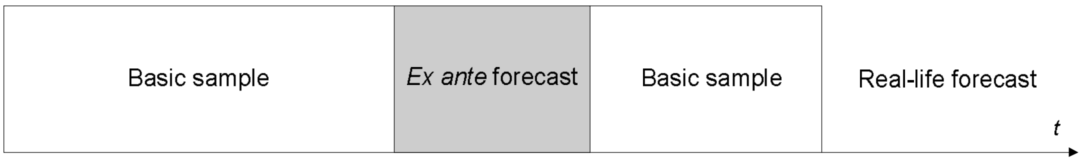


Fig. 7 Chronological division of a time series into the basic and *ex ante* forecasting samples

According the proposed algorithm, it is recommended to divide a time series into the basic and ex ante samples in the way shown in Fig. 7. This original approach allows increasing the quality of real, *ex post* forecasts as the last historical observations are used for fitting a statistical model rather than testing.

The accuracy of *ex ante* forecasts is evaluated by one of the standard error parameters. In the practical studies of the thesis, the mean absolute percentage error (*MAPE*) was applied. It is calculated from [1, p. 347]:

$$MAPE = \frac{\sum_{t=1}^{L} \left| \frac{x_t - \hat{x}_t}{x_t} \right|}{L} \cdot 100\% \qquad (14)$$

where L – the size of an *ex ante forecasting* sample (i.e., forecasting horizon).

The *MAPE* is a relative, dimensionless measure of the accuracy of an approximation curve or a forecast. It is helpful in comparing forecast performance across different data sets, or comparing the performance of different statistical methods.

The model of the lowest MAPE is the final model assigned to further *ex post* forecasting. The interpretation of *MAPE* introduced in [29] allows judging about the accuracy of a forecast: less than 10 per cent is a highly accurate forecast, 11 to 20 per cent is a good forecast, 21 to 50 per cent is a reasonable forecast, and 51 per cent or more is an inaccurate forecast.

Thus, the choice of a final forecasting model is based on the results of multiple sequential procedures and tests. The final model is characterized by the lowest value of the information criterion, uncorrelated residuals and the lowest error of an *ex ante* forecast.

CHAPTER 4 PRACTICAL STUDIES

Chapter 4 contains the description of practical research studies and the analysis of produced results.

The effectiveness of the developed algorithm and the ability of different methods to accomplish a traffic forecasting task were examined on real data sets represented as time series of different lengths and aggregation rates. Two data samples, characterizing the intensity of total carried traffic of a conventional telephone network and the transmission rate of outgoing international traffic, were considered in the thesis. Following ITU-T Recommendation E.492 [17], the initial traffic measurements were averaged over the periods equal to 15 minutes and one hour.

Table 1 General description of examined time series

Network Type	Traffic Type	Read-out period	Label	Basic sample	Size of a basic sample	Size of an *ex ante* forecasting 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 h	C	May 12 – Jul. 13, 2008	9 weeks (1512 obs.)	
			D	May 12 – Aug. 3, 2008	12 weeks (2016 obs.)	
Telephone network (POTS)	Total carried traffic	15 min	E	Jan. 8 – Mar. 11, 2007	9 weeks (6048 obs.)	1- 14 days
			F	Jan. 8 – Apr. 1, 2007	12 weeks (8064 obs.)	
			G	Jan. 8 – May 13, 2007	18 weeks (12096 obs.)	
		1 h	H	Jan. 8 – Mar. 11, 2007	9 weeks (1512 obs.)	
			I	Jan. 8 – Apr. 1, 2007	12 weeks (2016 obs.)	
			J	Jan. 8 – May 13, 2007	18 weeks (3024 obs.)	

A secondary goal of the practical studies was to examine how both the size of a basic data sample and the rate of aggregation influenced the accuracy of *ex ante* forecasts. The size of a basic sample was equal to 9 and 12 weeks for the first analyzed variable, and to 9, 12 and 18 weeks – for the second variable. Thus, the total number of time series considered was equal to ten.

The size of an *ex ante* forecasting sample, which determined a total forecasting horizon, varied for each time series from one to 14 days, with the sampling step of one day.

The general description of the considered time series is given in Table 1. The fragments[4] of the time series are displayed in Fig. 8.

Prior to determining an appropriate forecasting model and developing *ex ante* forecasts, the main statistical parameters and properties of each time series were estimated. The corresponding procedures included:

- assessment of the main sample parameters (mean, variance, median, etc.);
- testing for stationarity by means of the runs test and reverse arrangement test [24; 49, p.767];
- evaluation of the autocorrelation function;
- evaluation of the Hurst coefficient;
- testing for periodicity.

In the case of time series (E)-(J), characterizing telephone network traffic, the reverse arrangement test accepted the null hypothesis of the stationarity of both the mean and the variance at significance level $\alpha = 0.05$. For time series (A)-(D), characterizing Internet network traffic, the reverse arrangement test revealed the instability of the variance at significance level $\alpha = 0.05$. Nevertheless, the deviation of the number of reversals from the critical limits was slight. Already at significance level $\alpha = 0.02$, the hypothesis about the variance stationarity was accepted in most cases considered. Therefore, it was decided not to apply further measures to stabilize the variance.

The analysis of the autocorrelation function indicated the presence of periodic components. The influence of strong autocorrelation dependencies was observed not only between the adjacent members but also between the quite remote ones. It points at the "long history" of an underlying process, which provides the opportunity to produce reliable forecasts into a rather distant future. The persistency of the analyzed time series was confirmed by the Hurst coefficient as well, which exceeded 0.5 for all the time series analyzed. However, this estimate is often criticized and purely optional due to its inaccuracy. Besides, it is worth noting that the value of the Hurst coefficient cannot be directly incorporated into a forecasting model.

[4] Each fragment displays the observations over the first two weeks of a considered data sample

The spectral analysis of the IP network traffic pointed at the periodical components of the periods equal to 24 hours and 7 days. In the case of telephone traffic, the largest periods of the seasonal component comprised 12 hours, 24 hours and 7 days.

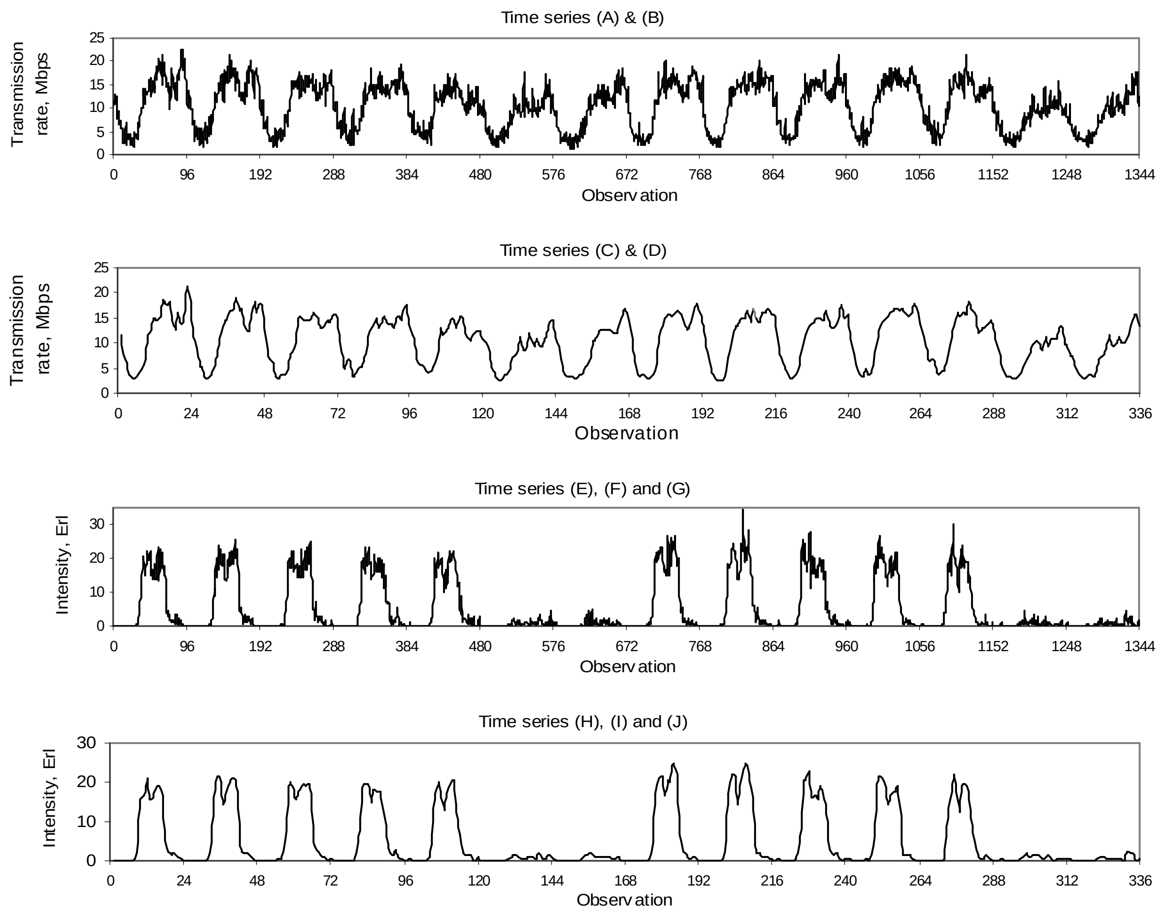


Fig. 8 Fragments of examined time series

- **Description of Forecasting Technique**

The main goal of the practical studies was to examine the statistical properties of certain data sets and to develop such a neural network, which was capable of modelling the underlying processes and producing the reliable low-error forecasts for a pre-defined forecasting horizon.

The selection of appropriate neural network models and the development of *ex ante* forecasts were fulfilled according to the algorithm shown in Fig. 5. The diagram of the neural network, applied in the empirical studies, is shown in Fig. 9 . The main parameters of the neural network, which stayed unchanged for all the models during a training process, are summarized in Table 2.

The appropriate architecture of a neural network was determined as follows. According to the universal approximation theorem [5, 15] the number of hidden layers in all the examined neural networks was equal to one. The size of an input window was set according 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 one-step-ahead forecasting. The number of hidden neurons varied from one to ten. The adaptive methods of network pruning [2, p.359] or growing [2, p. 357; 13, p. 250] were not implemented. The procedures of verification and residual testing were applied to each of these models. Although the process of verifying all the possible architectures is time-consuming, it provides an opportunity to preserve the purity of experiments.

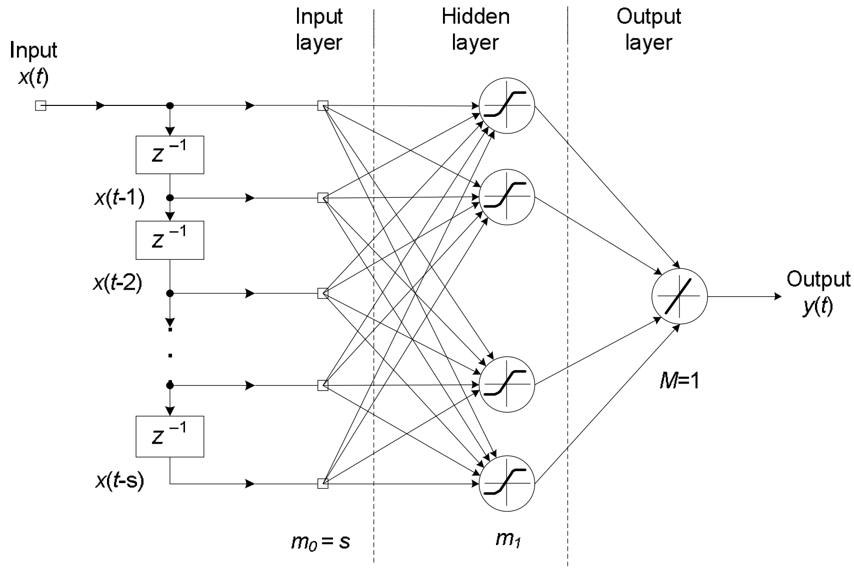


Fig. 9 Diagram of the neural network (multilayer perceptron) applied in the empirical studies

The initial weights were randomly drawn from the diapason of uniformly distributed small values. All the network architectures were reinitialized and retrained a hundred times.

Table 2 General specification of the developed neural network

Stage	Parameter / Procedure	Parameter Value / Procedure Description
Selection of network type and topology	Type of a network	Fully connected time-lagged feed forward network
	Number of hidden layers	1
	Number of output neurons	1
	Activation function	Hidden layer – hyperbolic tangent; output layer – linear function
Selection of training parameters	Number of training epochs	600
	Training algorithm	Back propagation & conjugate gradient descent
	Error function	Mean squared error
	Learning rate	0.1
	Momentum term	0.3
	Method of weight initialization	Randomized values from a uniform distribution
	Number of times to randomize weights	100
Training optimization	Methods to prevent over-learning	Cross-validation [13, p.218] , weight regularization [47]
	Size of training, validation and test subsets	At a ratio of 3:1:1
	Stopping criterion	Invariable or increasing training error during 50 epochs
In-sample and out-of-sample evaluation	Parameters of in-sample evaluation	R, MAE, RMSE, MAPE, AIC, BIC
	Diagnostic testing of residuals	LM- type test , χ^2 - test
	Parameters of out-of-sample evaluation	RMSE, MAE, MAPE, Diebold-Mariano criterion [6]

In order to avoid the effect of over-training, the *cross-validation* technique [13, p.218] was implemented. The basic sample was divided into training, validation and test subsets at a ratio 3:1:1. A splitting scheme was random and changed for each training cycle. This approach does not allow "getting stuck" in local minima and increases the stability of a system, since the process of searching a global minimum is carried out in different directions and do not rely on a particular set of time series observations. Another measure to avoid over-training was *weight regularization* [47] applied without subsequent deletion of synaptic connections and neurons.

A training process was realized by means of the software package StatSoft STATISTICA© 7.0. A two-stage training process was implemented. During the first stage a multilayer perceptron was trained by 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 (500 epochs) was used, with a stopping window of 50, to terminate training once convergence stopped or over-learning occurred. Once the algorithm stopped, the best network from the training run was restored.

The input data of a neural network were corrected for obvious anomalous outliers, the reason of which was temporal malfunction of network equipment, and for anomalous patterns, which took place as a result of public holidays falling on the days of a workweek. The input data sets were also transformed to the range $[-1, 1]$ by means of the linear transformation.

The final forecasts produced by neural networks were compared to those produced by the models of seasonal ARIMA, seasonal exponential smoothing as well as "naïve" methods. In order to evaluate statistically significant differences between the forecasts developed for various forecasting horizons, the Diebold-Mariano [6] criterion was implemented. It is non-parametric and tolerant to different deviations from the classic assumptions about the properties of forecast errors. In particular, it can be applied even if forecast errors are non-Gaussian, serially correlated, contemporaneously correlated and have a non-zero mean.

- **Estimation of Practical Results**

The results of fitting and verifying the statistical models and neural networks, the estimates of their in-sample and *ex ante* accuracy are summarized in Volume 2 of the thesis. The following main operations were conducted for each time series:

- appropriate models of multilayer perceptrons, SARIMA models and exponential smoothing were identified and varified;
- *ex ante* forecasts were produced by means of different models and evaluated for the accuracy;
- statistically significant differences between the final *ex ante* forecasts developed for various forecasting horizons were identified by means of the Diebold-Mariano test.

The accuracy of final forecasts was evaluated by means of such standard parameters as MAE (mean absolute error), RMSE (root mean squared error) and MAPE (mean absolute percentage error), the latter of which raises the greatest interest (see Fig. 10).

Accuracy of neural network forecasts evaluated by the mean absolute percentage error

For time series (A)-(D) describing IP network traffic, the MAPE estimates do not practically change or slowly grow with the increase of a forecasting horizon from 24 hours to 14 days. It points at the opportunity to increase a lead time, for which reliable forecasts can be produced.

For time series (E)-(G) characterizing telephone network traffic, the MAPE estimates grow fast with a forecasting horizon, and already for the forecasts obtained two weeks ahead, exceed 30 per cent. It means that a maximum forecasting horizon is achieved. In turn, for time series (H)-(J) aggregated over one-hour intervals, the MAPE estimates do not practically change or slightly increase with a forecasting horizon. It would allow to extend a forecasting horizon further.

For time series (A) and (B), the MAPE estimates of neural networks and statistical models comprise 20-25 per cent. This reveals a satisfactory accuracy of the produced forecasts. For time series (C) and (D) the MAPE of statistical models and neural networks is around 10-15 per cent, which demonstrates a good accuracy of produced forecasts.

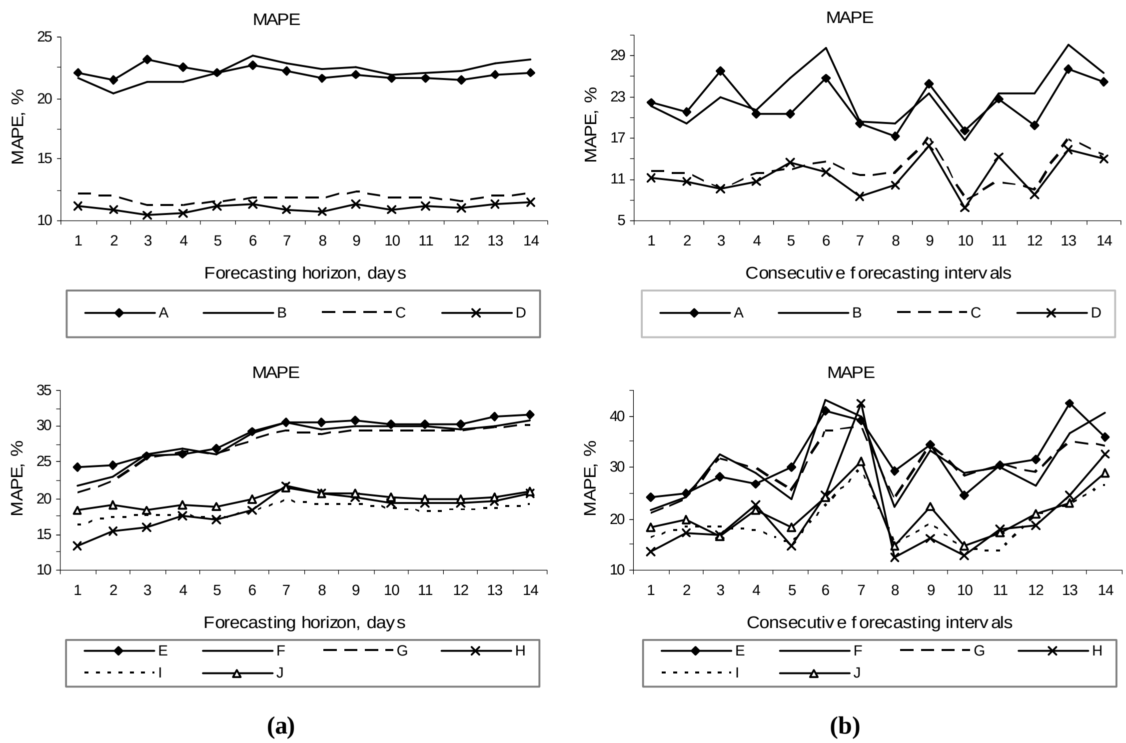


Fig. 10 Estimates of the accuracy of the final *ex ante* forecasts produced by neural networks (a) against the length of a forecasting horizon; (b) at consecutive one-day sampling intervals[5]

For time series (E)-(G) the MAPE estimates of neural networks comprise 21-32 per cent, which is the evidence of a satisfactory accuracy of the produced forecasts. For time series (H)-(J), the MAPE of neural networks is around 14-21 per cent. It points at a good accuracy of the forecasts.

Statistically significant differences between the forecasts produced by different methods

The final forecasts produced by various methods (neural networks, SARIMA and seasonal exponential smoothing) look very similar. Besides, it is not easy to select a forecasting model, taking into account only the standard accuracy parameters. Therefore, the identification of statistically significant differences in accuracy of the forecasts, developed by different methods, raises a special interest.

For these purposes, the Diebold-Mariano test was applied to the forecasts produced 24 hours, 7 days and 14 days ahead. The first group of forecasts can be considered as operative forecasts, while the second and the third ones – as short-term forecasts. The results of testing a null hypothesis of the absence of statistically significant differences between the forecasts, at significance level $\alpha = 0.05$, are summarized in Table 3.

For time series (A)-(D) characterizing IP network traffic, the forecasts produced by one or another linear method (SARIMA or exponential smoothing) do not lose in accuracy to the forecasts of neural networks, in all 12 analyzed cases. In two out of 12 cases, the forecasts produced by neural networks are statistically equivalent to the seasonal naïve forecasts as well.

For time series (E)-(J) characterizing telephone network traffic, the forecasts produced by neural networks are statistically equivalent to the forecasts produced by one or another linear method in 14 out of 18 analyzed cases. In the other four cases, a neural network outperforms in forecasting

[5] The MAPE estimates are calculated for the data moved along the y-axis by the distance of 0.5 Erl. It is required as the telephone traffic contains zero values. Even if this procedure distorts the actual values of absolute percentage errors of the observations which are not zero, it gives the opportunity to evaluate the dynamic changes of MAPE over different forecasting horizons.

accuracy both SARIMA and seasonal exponential smoothing. In 4 out of 18 analyzed cases, statistically significant differences were not identified between the forecasts produced by neural networks and seasonal naïve methods.

Table 3 Final neural network forecasts in comparison with the forecasts produced by other methods[6]

Forecasting method \ Time series/ Forecasting horizon	A 24 h	A 7 d	A 14 d	B 24 h	B 7 d.	B 14 d.	C 24 h	C 7 d	C 14 d	D 24 h	D 7 d	D 14 d
SARIMA	↑	=	=	↑	=	=	↑	=	↑	=	=	=
Seasonal exponential smoothing	=	↓	↓	=	↓	↓	=	↓	↓	↓	↓	↓
"Naïve" forecast	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Seasonal "naïve" forecast	↓	↓	↓	↓	↓	↓	=	↓	=	↓	↓	↓

Forecasting method \ Time series/ Forecasting horizon	E 24 h	E 7 d	E 14 d	F 24 h	F 7 d	F 14 d	G 24 h	G 7 d	G 14 d	H 24 h	H 7 d	H 14 d	I 24 h	I 7 d	I 14 d.	J 24 h	J 7 d	J 14 d
SARIMA	=	=	=	=	↓	↓	=	↓	↓	=	=	=	=	=	=	=	=	=
Seasonal exponential smoothing	=	=	=	=	↓	↓	=	↓	↓	=	=	=	=	=	=	=	=	=
"Naïve" forecast	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Seasonal "naïve" forecast	=	↓	↓	↓	↓	↓	↓	↓	↓	=	↓	↓	=	↓	↓	=	↓	↓

Notes:
= *the forecast, produced by the specified method, is statistically equivalent to the neural network forecast;*
↑ *the forecast, produced by the specified method, outperforms in accuracy the neural network forecast;*
↓ *the forecast, produced by the specified method, loses in accuracy to the neural network forecast.*

Impact of the length of a read-out period on forecasting accuracy

For all the analyzed time series, the increase of a read-out period allowed increasing the accuracy of the produced *ex ante* forecasts. On average, the increase of a read-out period from 15 minutes to one hour decreased the MAPE values of a neural network for 10 per cent. This is primarily due to the reduction of time series variance, which simplified the selection of an appropriate statistical model as well.

Impact of the size of a basic sample on forecasting accuracy

For time series describing the traffic of both an IP network and a conventional telephone network, the increase of the size of a basic fit sample (i.e., the increase of the number of training patterns) did not lead to a substantial increase in accuracy of neural network forecasts.

In the case of IP network traffic, the MAPE estimate is slightly lower for time series (D) than for time series (C), although these differences are insignificant. For telephone network traffic, the MAPE estimates are a bit lower for time series (G) than for time series (E) and (F) but these differences are also insignificant.

In the case of telephone traffic aggregated over one-hour intervals, the increase of a basic sample from 12 to 18 weeks, in contrast, resulted in a slight increase in the level of forecasting errors.

[6] The identification of statistically significant differences in forecasting accuracy was conducted by the Diebold-Mariano test at significance level α=0.05.

MAIN CONCLUSIONS AND RECOMMENDATIONS

The *main aim* of the research – to address and solve the problems, raised by the production of short-term traffic forecasts by means of neural networks, has been achieved. Taking into account the results of practical and theoretical studies conducted in the thesis, the following main conclusions and recommendations, regarding the application of forecasting models in network traffic forecasting, have been specified.

The proposed algorithm, aimed at short-term traffic forecasting by means of neural networks, allow automating the identification of a neural network solution with minimal involvement and influence of expert assessment and human factors.

Neural networks traditionally involve expert assessment at all the stages of application. The algorithm, developed in the thesis, allows minimizing the influence of a human factor and helps finding the solutions resulting in reliable forecasts with a minimum level of errors. The *MAPE* estimates of examined *ex ante* forecasts vary from 10 per cent (in the case of averaging over one-hour intervals) to 30 per cent (in the case of averaging over 15-minute intervals). This confirms the possibility of applying these models in real-life conditions. The criteria for selecting a final forecasting model proposed in the thesis – the lowest estimate of the information criterion and statistically insignificant residual autocorrelation can be successfully applied to linear statistical methods as well.

The properties of self-similarity and long-range dependence of packet-switched network traffic are only observable in the case of aggregation in a very large scale – usually, over the intervals from a few milliseconds to a few minutes.

From the point of view of analysis and forecasting, an excessively large scale of time series is not useful. The process of fitting a forecasting model to an examined time series will be complicated due to correlations between remote observations as well as strong influence of extraneous noises and anomalous outliers, which inevitably accompany the large-scale measurements. It is also necessary to understand that the longer the period of sampling / aggregation, the longer is the horizon, for which reliable forecasts can be produced. Therefore, taking into account ITU-T Recommendation E.492 [17], it is advised to average the initial measurements of network traffic over 15-minute and / or one-hour intervals. In this case, the factors determining the statistical structure of a *real* traffic process refer to seasonal effects and monotonous trends, which are primarily associated with the behaviour of subscribers / users and the influence of technological progress.

Reliable operative and short-term forecasts of traffic dynamics can be produced by means of linear statistical models, if the aggregation / sampling period of time series is set in compliance with ITU-T Recommendation E.492.

The real time series analyzed in the thesis are *typical* for these types of loads and incorporate both daily and weekly cycles. It means that for many other time series with similar read-out periods, statistical properties and autocorrelation structure, the production of reliable operative and short-term forecasts can be conducted by applying linear statistical models and methods. The process of forecasting by means of neural networks requires substantial time resources for training, apart from an intensive test-and-trial estimation of many parameters. The time required for series modelling and forecasting by means of neural networks exceeds the time required for producing a forecast by means of linear methods by several orders of magnitude. Therefore, this conclusion is important for successful forecasting in real operating conditions.

A neural network is capable of modelling and predicting seasonal time series in a direct way, without prior deseasonalization.

In developing a seasonal neural network, the most important aspect to consider is the appropriate size of an input window, which has to be set according to the largest period of a seasonal component.

In order to produce reliable forecasts of network traffic represented as periodic time series, it is necessary to focus on the seasonal modifications of such linear models as ARIMA and exponential smoothing.

In contrast to the series, synthesized by such formal models as *fractal Brownian motion*, real time series often incorporate seasonal and / or cyclic components and require the application of seasonal modifications of classic linear models. Just as in the case of neural networks, correct identification of the periods of a seasonal component is of great importance.

- **Potential Directions for Further Research**

The method of neural networks is, certainly, one of the most perspective tools in traffic forecasting. Further study directions has to be aimed at developing the algorithms and methods of *real-time forecasting* that currently raise some difficulties due to insufficient capacity of computing equipment and intensive involvement of human expertise.

Another interesting research direction is to develop fast and reliable methods of constructing the *confidence intervals* for the forecasts produced by neural networks. This problem has not been solved yet due to the necessity of estimating a very large number of free parameters, each of which contributes a share of uncertainty.

BIBLIOGRAPHY

1. Armstrong, J.S. *Long-Range Forecasting*. New York: John Willey & Sons, 1985.
2. Bishop, C. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
3. Box, G.E.P., and G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, Inc., 1970.
4. Clements, M.P., and D.F. Hendry. "On the Limitations of Comparing Mean Square Forecast Errors". *Journal of Forecasting* 12 (1993): 617-637.
5. Cybenko, G. "Approximation by Superpositions of Sigmoidal Function." *Mathematics of Control, Signals, and Systems* 2 (1989): 303-314.
6. Diebold, F.C., and R.S. Mariano. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13.3 (July 1995): 253-263.
7. Elman, J.L. "Finding Structure in Time." *Cognitive Science* 14 (1990): 179-211.
8. Faraway, J., and C. Chatfield. "Time Series Forecasting with Neural Networks: A Comparative Study Using the Airline Data." *Applied Statistics* 47.2 (1998): 231-250.
9. Fildes, R., and K.-P. Liao. "The Accuracy of Procedural Approach to Specifying Feedforward Neural Networks for Forecasting." *Computers & Operations Research* 32.8 (August 2005): 2151-2169.
10. Fowler, T.B. "A Short Tutorial on Fractals and Internet Traffic." *The Telecommunications Review* 10 (1999): 1-15.
11. Franses, H.P., and D. van Dijk. *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press, 2000.
12. Gardner, E.S.Jr. "Exponential Smoothing: The State of Art." *Journal of Forecasting* 4 (1985): 1-28.
13. Haykin, S. Neural Networks: A Comprehensive Foundation. 2nd Ed. Prentice Hall, 1999.
14. Hill, T., M. O'Connor, and W. Remus. "Neural Network Models for Time Series Forecasts." *Management Science* 42.7 (Jul. 1996): 1082-1092.
15. Hornik, K. "Multilayer Feedforward Networks are Universal Approximators." *Neural Networks* 2 (1989): 359-366.
16. ITU-T (CCITT) Recommendation E.506 (rev.1). Forecasting International Traffic. Geneva, 1992.
17. ITU-T Recommendation E.492. Traffic Reference Period (Telephone Network and ISDN. Quality of Service, Network Management and Traffic Engineering). ITU, 02/1996.
18. ITU-T Recommendation E.507. Models for Forecasting International Traffic. ITU, 1993 (1988).
19. Iversen, V. B. *Teletraffic Engineering and Networks Planning*. Technical University of Denmark, 2008.
20. Janevski, T. *Traffic Analysis and Design of Wireless IP Networks*. ARTECH HOUSE, INC., 2003.
21. Kaastra, I., and M. Boyd. "Designing a Neural Network for Forecasting Financial and Economic Time Series." *Neurocomputing* 10 (1996): 215-236.
22. Klevecka, I., and J. Lelis. "Application of Extrapolation Methods to the Technology Diffusion Forecasting." *Scientific Proceedings of Riga Technical University* (*Series "Telecommunications and Electronics "*) 7 (2007: 52-59.
23. Klevecka, I., and J. Lelis. "Perspective Evaluation of the Electronic Communications Market in Latvia." *Proceedings of the 18th European Regional ITS Conference*. Sep. 2007, Istanbul, Turkey. Berlin: International Telecommunications Society, 2007. 1-37. CD-ROM.
24. Klevecka, I., and J. Lelis. "Pre-Processing of Input Data of Neural Networks: The Case of Forecasting Telecommunication Network Traffic." Spec. issue of *Telektronikk* 104.3/4 (2008): 168-178.
25. Kolen, P., and J.B. Pollack. "Back Propagation in Sensitive to Initial Conditions." *Proceedings of the Conference on Advances in Neural Information Processing Systems*. Nov. 1989, Denver, United States. San Francisco: The MIT Press, 1990. 860-867.
26. Kunst, R.M. *Testing for Relative Predictive Accuracy: A Critical Viewpoint*. Economics Series 130. Vienna: Institute for Advanced Studies (IHS), 2003.
27. LeCun, Y., L. Bottou, G.B. Orr, and K.-R. Müller. "Efficient BackProp." *Neural Networks: Tricks of the Trade*. Eds. G.B. Orr and K.-R. Müller. Berlin: Springer, 1998. 9-50.
28. Leland, W.E., M.S. Taqqu, W. Willinger, and D.V. Wilson. "On the Self-Similar Nature of Ethernet Traffic (Extended Version)." *IEEE/ACM Transactions on Networking* 2.1 (February 1994): 1-15.
29. Lewis, C.D. *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. London: Butterworth Scientific, 1982.
30. Ljung, G.M., and G.E. Box. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (1978): 297–303.
31. Makridakis, S., and R.L. Winkler. "Sampling Distributions of Post-sample Forecasting Errors." *Applied Statistics* 38.2 (1989): 331-342.
32. McNelis, P.D. *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Elsevier Academic Press, 2005.
33. Meade, N. "Evidence for the Selection of Forecasting Methods." *Journal of Forecasting* 19 (2000): 515-535.
34. Medeiros, M.C., T. Teräsvirta, and G. Rech "Building Neural Network Models for Time Series: A Statistical Approach." *Journal of Forecasting* 25 (2006): 49-75.

35. Mozer, M. C. “Neural Network Architectures for Temporal Pattern Processing.” *Time Series Prediction: Forecasting the Future and Understanding the Past.* Eds. A.S. Weigend and N. A. Gershenfeld. Perseus Book Publishing, 1993. 243-264.
36. Park, C., F. Hernandez-Campos, J.S. Marron, and F.D. Smith. “Long-Range Dependence in a Changing Traffic Mix.” *Computer Networks* 48 (2005): 401-422.
37. Park, K., and W. Willinger. “Self-Similar Network Traffic: An Overview.” *Self-Similar Network Traffic and Performance Evaluation.* Eds. K. Park, and W. Willinger. John Wiley & Sons, 2000. III-XLIX.
38. Priestley, M.B. *Spectral Analysis and Time Series*. London: Academic Press, 1981.
39. Ryu, B.K., and A. Elwalid. “The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities.” *ACM SIGCOMM Computer Communication Review* 26.4 (1996): 3-14.
40. Sang, A., and S.-q. Li. “A Predictability Analysis of Network Traffic.” *Computer Networks* 39 (2002): 329-345.
41. Sheluhin, O.I., S.M. Smolskiy, and A.V. Osin. *Self-Similar Processes in Telecommunications*. John Willey & Sons, 2007.
42. Tang, Z., and P.A. Fishwick. „Feedforward Neural Nets as Models for Time Series Forecasting” *ORSA Journal on Computing* 5.4 (Fall 1993): 374-385.
43. Tashman, L.J. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting* 16.4 (October 2000): 437-450.
44. Taylor, J.W., L.M. de Menezes, and P.E. McSharry. „A Comparison of Univariate Methods for Forecasting Electricity Demand up to a Day Ahead.” *International Journal of Forecasting* 22 (2006): 1-16.
45. *Teletraffic Engineering Handbook.* ITU-D (Study Group 2, Question 16/2), June 2006.
46. Thimm, G., and E. Fiesler. *Optimal Setting of Weights, Learning Rate, and Gain.* IDIAP Research Report 97-04. Switzerland: IDIAP, 1997.
47. Weigend, A. S., D. E. Rumelhart, and B. A. Huberman. “Generalization by Weight-Elimination with Application to Forecasting.” *Advances in Neural Information Processing Systems 3*. Eds. R.P.Lippmann, J. Moody, and D. S. Touretzky. San Mateo, CA: Morgan Kaufmann, 1991. 875-882.
48. Yaffe, R.A, and M. McGee. *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. Academic Press, 2000.
49. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: Издательское объединение «ЮНИТИ», 1998. – 1005с.
50. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 3-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2007. – 912с.
51. Канторович Г.Г. Анализ временных рядов // Экономический журнал ВШЭ. №1 2002, с.85-116.
52. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. – М.: ИНФРА-М, 1997.- 302с.
53. Комарцова Л., Максимов А. Нейрокомпьютеры: Учеб. пособие для вузов. - М.: Издательство МГТУ имени Н.Э. Баумана, 2002. – 320с.
54. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи. – М.: Горячая линия-Телеком, 2003. – 94с.
55. Крисилов В., Чумичкин К., Конратюк А. Представление исходных данных в задачах нейросетевого программирования// V Всероссийская научно-техническая конференция "Нейроинформатика-2003": Сборник научных трудов. Ч. 1. - М: МИФИ, 2003, с 184-191.
56. Крылов В.В., Самохвалова С.С. Теория телетрафика и ее приложения. – Спб.: БХВ-Петербург, 2005. – 288с.
57. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. – М.: ИНФРА, 1998. – 528с.
58. Шмертер Х., Белендорф Э. и др. Теория телетрафика. – М.: Связь, 1971. – 319с.