

Neural Models of Demands for Electricity - Prediction and Risk Assessment

Abstract. Two neural systems for forecasting the electricity demand by the group of retail consumers are presented along with two methods for risk assessment of demand prediction models. The first forecasting system is composed of series-connected local neural predictors in the form of multi-layer perceptron (MLP) networks. The system is mainly formed on the basis of expert knowledge and statistical tests. The second forecasting system has two levels. The first contains a neural classifier and the second consists of a set of local neural predictors. The classifier is built on the basis of a self-organising neural network (SOM). MLP or radial basis function (RBF) networks are used as predictors. Finally, two methods for assessing the risk of forecasting models are proposed. These consider financial risk measures such as value at risk (VaR) and conditional value at risk (CVaR). Possible economic losses posed by the application of predictions from a forecasting model are calculated using these risk measures. The risk analysis facilitates the selection of the forecasting model that generates the smallest risk of losses when selling energy contracts. The proposed methods are tested using data from the Polish electricity market.

Streszczenie. W pracy zostały przedstawione dwa neuronowe systemy przeznaczone do prognozowania zapotrzebowania na energię elektryczną grupy konsumentów detalicznych. Ponadto zaprezentowano dwie metody oceny ryzyka modeli prognozowania. Pierwszy system prognozowania składa się z połączonych szeregowo lokalnych predyktorów neuronowych zbudowanych w oparciu o wielowarstwowe sieci perceptronowe (MLP). System powstał na podstawie wiedzy eksperckiej i testów statystycznych. Drugi system prognozowania jest dwupoziomowy. Pierwsza część zawiera klasyfikator neuronowy, a drugi składa się z zestawu lokalnych neuronowych predyktorów. Klasyfikator jest zbudowany w oparciu o samoorganizujące sieci neuronowe (SOM). Natomiast jako predyktory są wykorzystywane sieci MLP lub sieci o radialnych funkcjach bazowych (RBF). Następnie zostały zaproponowane dwie metody oceny ryzyka modeli prognostycznych. Zastosowano miary ryzyka pochodzące z rynków finansowych takie jak wartość narażona na ryzyko (VaR) i warunkowa wartość narażona na ryzyko (CVaR). Ewentualne straty ekonomiczne wynikające z zastosowania prognoz z modelu prognostycznego są obliczane przy użyciu tych miar ryzyka. Analiza ryzyka ułatwia wybór modelu prognostycznego, który generuje najmniejsze ryzyko strat dla umowy sprzedaży energii elektrycznej. Proponowane metody zostały przetestowane przy użyciu danych z polskiego rynku energii elektrycznej. (**Neuronowe modele zapotrzebowania na energię elektryczną - prognozowanie i ocena ryzyka**)

Keywords: electricity market, neural networks, electricity demand prediction, risk assessment of prediction models

Słowa kluczowe: rynek energii elektrycznej, sieci neuronowe, prognozowanie zapotrzebowania elektrycznego, ocena ryzyka modeli prognozowania

Introduction

Many countries have implemented reforms in their energy markets, thus introducing competitive forces. This has motivated the development of new and more effective methods of business process management for energy producers, consumers and distributors. These processes can be as follows: forecasting electricity prices and demand, management and portfolio optimisation, analysis and measurement and risk management. In this paper two of the above areas are considered, i.e. the short-term forecasting electricity demand by the group of retail consumers and risk assessment of prediction models.

There are many methods of short-term load prediction and electricity demand forecasting. Initially, statistical forecasting methods, such as regression models and ARIMA models, were used [10]. Currently, it is believed that these models are inadequate due to the presence of nonlinear phenomena in the stochastic process of electricity consumption, hence the widespread interest in artificial intelligence. Particularly popular are forecasting algorithms based on artificial neural networks. Also because they are flexible, easy to implement and do not have such stringent statistical assumptions as the ARIMA models have. The most often used are multi-layer perceptron (MLP) networks [8, 7, 11, 18, 21], also combined with genetic algorithms [20], and rarely used are self-organising maps (SOM) [24] and recurrent networks [23]. Other popular tools used for this purpose are fuzzy logic and neuro-fuzzy networks [5, 17, 22] or a combination of fuzzy logic and neural networks [12, 15]. Surprisingly, networks based on the radial basis function (RBF) are rather not applied in the power sector, although there are interesting applications for chaotic time series prediction [1, 13, 16, 25]. On the basis of the cited literature and own experiments, we concluded that one global model is not enough for electricity demand prediction. We propose a compound neural forecasting system which presupposes the existence of a neural classifier at the first level and a set of neural predictors at the

second level. We believe that the application of many local predictors will create a more precise forecasting model than the solution with a single predictor. It should be noted that inspiration for the forecasting system came from PREMON [14] and PREMONFS [19]. We implemented two systems, i.e. a series of MLP predictors and a system with an SOM classifier and RBF or MLP predictors.

The methods of risk measurement which are currently known have been developed for the financial sector, i.e. for banks, investment funds, insurance companies, etc. Several attempts have been made to adapt these methods for the electricity market, especially for portfolio risk assessment of electricity producers [3, 6, 4]. In these examples the portfolio risks were expressed as portfolio variance [2], value at risk (VaR) [6, 4], conditional value at risk (CVaR) [4], [6] and earning at risk (EaR) [6]. Generally, studies on risk in the energy market can be divided into the following categories [6]: general issues of risk assessment, decision analysis of energy suppliers, energy price forecasting, security techniques against risk, and market analysis. In this paper we deal with a new aspect of risk in the energy market - risk generated by forecasting models. These models generate forecasts of electricity demand that may affect the content of a company's portfolio. Yet forecasts are burdened with errors, thus generating the risk of taking wrong decisions on the market. Therefore, it is important to evaluate the forecasting models in the context of financial risk. We proposed VaR and CVaR as measures of possible economic losses posed by the application of electricity demand predictions for the construction of a company portfolio. Finally, we showed that this proposition allows to select a forecasting model which generates the smallest risk of losses on an energy selling contract.

The paper is organised as follows: in section "Prediction of electricity demands" the two forecasting systems are described in detail. Next, in section "Risk assessment of prediction models" the methods of measurement and risk assessment for the forecasting models are presented. The results

of the experiments are provided in section "experiments". The last section contains the conclusions.

Prediction of electricity demand

Problem formulation. Let us consider an electricity distribution company which delivers electricity to retail customers located in a particular area. The company must have some information about the customers' whole energy consumption in the future in order to properly build a portfolio of contracts (namely the purchase of electricity from the wholesale market to meet the customers' needs). Let us assume that all of the participants in the wholesale energy market must submit their contracts of purchase and sale of electricity to the national transmission system operator. This should be done before the hour 11:00 on day N-1 if energy delivery in an agreement is planned for day N. Therefore, the company must have forecasts for the whole of electricity consumption by retail consumers at least two days ahead, but longer-term forecasts are also welcome. In conclusion, we can assume that the task is to forecast the electricity demand of the whole group of retail consumers. Forecasts should be performed every hour for a minimum of two or a maximum of seven days in advance. The values of electricity consumption (note that it is not load) by the group of retail consumers are known for every hour (k) for a certain period in the past $e = [e_1, e_2, \dots, e_K]^T$ [MWh]. The following meteorological factors are also known for the same period: temperature $t = [t_1, t_2, \dots, t_K]^T$ [$^{\circ}C$]; insolation $i = [i_1, i_2, \dots, i_K]^T$ [Wh/m²]; and humidity $h = [h_1, h_2, \dots, h_K]^T$ [0 – 100%]. Note that these assumptions can be easily adapted to any particular energy market. However, the experiments presented in this paper were conducted using data for the years 2002-2004 which came from a Polish power distribution company.

The overall concept of the forecasting system. A prerequisite for the selection of an appropriate forecasting model is the stationarity of a stochastic process. The stationarity analysis of the energy consumption process e was carried out by analysis of its autocorrelation function. Studies have shown the occurrence of cyclical components (daily and weekly) and an insignificant seasonal component in the process. Removing these components causes the process to become stationary. Then, studies of the impact of weather factors (temperature t , insolation i , humidity h) on energy consumption (e) were carried out in order to determine which factors should be used in the forecasting model. Pearson correlation coefficients were calculated separately for each month of 2002-2004 (Tab. 1).

Table 1. Linear correlation: (+) positive, (-) negative.

energy (e)	temperature (t)	insolation (i)	humidity (h)
spring	poor (+/-)	poor (+)	poor (-)
summer	moderate (+)	moderate (+)	moderate (-)
autumn	poor (+/-)	lack	poor (-)
winter	lack	lack	lack

Next, additional tests of the correlation coefficient rank $\rho - Spearman$ were conducted. For each month of 2002-2004 two hypotheses were tested: (H) - features are independent; (A) - features are dependent. Based on the tests, hypothesis H was rejected in the pairs: e and i , e and h , e and t respectively for 100%, 94.5% and 72% of the months. In conclusion, it can be assumed that process electricity consumption contains daily, weekly and seasonal components.

There is a linear and rank correlation between electricity consumption and temperature and insolation and humidity, respectively.

On the basis of the analyses, the autocorrelation function of electricity consumption and reports in literature [7, 11] we can state that electricity consumption varies depending on the hour of the day, the type of day and the season of the year. It can be modeled by a single forecasting model where the changes in month, in day, in holiday are represented by 1-of-N binary coding in the input layer. But we believe that the application of many local predictors will create a more precise forecasting model than the use of a single forecasting model which describes the whole electricity consumption. The overall concept of the forecasting system presupposes the existence of a classifier at the first level and a set of predictors at the second level (Fig. 1a).

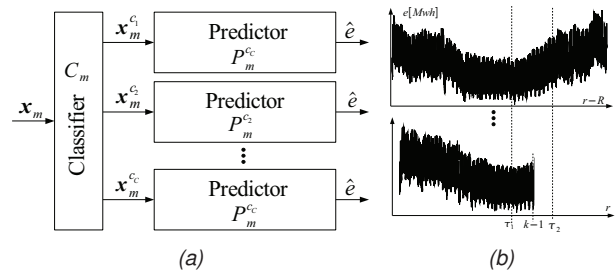


Figure 1. (a) The overall concept of the forecasting system and (b) an example of a measuring window.

The classifier (C_m) is responsible for grouping data (vectors x_m) into sets c_1, \dots, c_c . A local predictor P_m^c is trained on data set c . Therefore, the predictor P_m^c describes an electricity customer's behaviour in sections of electricity consumption time that are similar to each other. The classifier and predictors are in the scope of a forecasting model. The content of input vector x_m is determined by the forecasting model m . The classifier will be discussed later in this article while the general form of the local predictor is determined as follows:

$$(1) \quad \hat{e}_k = f(W, x_{k,m}^c)$$

$$(2) \quad x_{k,m}^c = [e_{k-1}^c, \dots, e_{k-k_e}^c, t_k, \dots, t_{k-k_t}, i_k, \dots, i_{k-k_i}, h_k, \dots, h_{k-k_h}]$$

where: c is the cluster of data; W are properties and the internal structure of the predictor; $f(*)$ is a function which represents the dynamic properties of the predictor; k_e, k_t, k_i, k_h are the maximum delays of time series: consumption of energy, temperature, insolation and humidity, respectively; k is the time index for which the value has to be predicted.

Due to the existence of a daily cycle in the process of electricity consumption, it can be assumed that demand for electricity in hour k depends on each hour from the previous day to some extent, so $k_e = 24$ (eq.2). In view of the fact that correlations were found between energy consumption and temperature, insolation and humidity, we decided to apply these factors in the forecasting models. It was assumed that they are taken into account for the time of the forecast, so $k_t = k_i = k_h = 0$ (eq.2). Partial correlation analyses to investigate the impact of individual model inputs on the significance of others were not carried out. Instead, we decided that tests for different combinations of inputs would be carried out. We propose the following forecasting models, assuming

that $e_k = \{e_{k-1}, \dots, e_{k-24}\}$:

$$\begin{aligned} m_1 : \quad \mathbf{x}_{k,m}^c &= [e_k]^T & m_5 : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k]^T \\ m_2 : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k]^T & m_6 : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, h_k]^T \\ m_3 : \quad \mathbf{x}_{k,m}^c &= [e_k, i_k]^T & m_7 : \quad \mathbf{x}_{k,m}^c &= [e_k, i_k, h_k]^T \\ m_4 : \quad \mathbf{x}_{k,m}^c &= [e_k, h_k]^T & m_8 : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k, h_k]^T \end{aligned}$$

Moreover, we propose forecasting models containing additional information about the type of working day (d_k) and holidays (s_k) as well as daily (\bar{e}_k^{24}) and weekly (\bar{e}_k^{168}) moving averages defined as follows $\bar{e}_k^n = \frac{1}{n} \sum_{j=0}^{n-1} e_{k-j}$. Models having various combinations of these factors are presented below:

$$\begin{aligned} m_9 : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k, h_k, d_k, s_k]^T \\ m_{10} : \quad \mathbf{x}_{k,m}^c &= [e_k, d_k, s_k, \bar{e}_k^{24}, \bar{e}_k^{168}]^T \\ m_{11} : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, d_k, s_k, \bar{e}_k^{24}, \bar{e}_k^{168}]^T \\ m_{12} : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k, d_k, s_k, \bar{e}_k^{24}, \bar{e}_k^{168}]^T \\ m_{13} : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k, h_k, d_k, s_k, \bar{e}_k^{24}, \bar{e}_k^{168}]^T \\ m_{14} : \quad \mathbf{x}_{k,m}^c &= [e_k, t_k, i_k, h_k, \bar{e}_k^{24}, \bar{e}_k^{168}]^T \end{aligned}$$

The series of predictors. The overall idea of the forecasting system (Fig. 1a) is that there is a classifier and a set of local predictors. The clustering of electricity consumption data may be based on a calendar and on expert knowledge, because the profile of energy or power in a day depends on the calendar date, which is derived from the activities of energy consumers in a day, week or month. The implementation of the main idea (Fig. 1a) is the use of expert classification and the serial connection of predictors (Fig. 2).

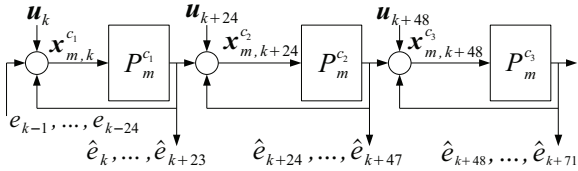


Figure 2. Example of the series of MLP predictors.

Here the classifier is not a real object, but only a set of rules describing how many classes of predictors are needed and how the predictors should be connected. Note that the classification does not depend on weather factors. A vector in the cluster of data includes the whole daily profile of energy consumption (24 hours). Each predictor corresponds to one cluster of data. The predictor describes one type of day, but it has only one output, so a forecast for the whole day is performed recursively in 24 steps, so that its forecast is turned back to its input at each step of forecasting. It is easy to notice (Fig. 2) that the input data for the first predictor in the series at the first step of the forecast are the following: 24 known values of energy consumption $\{e_{k-1}, \dots, e_{k-24}\}$ and a regression input vector (u_k). The vector u_k contains additional inputs (weather factors) selected in pursuance of the prediction model m (for example, if $m = 8$ then $u_k = [t_k, i_k, h_k]^T$). The combination of these two vectors gives the input vector $x_{m,k}^c$. However, at the last step of the forecast, the input vector $x_{m,k}^c$ to the predictor (the first in the series) contains only one real value of energy consumption e_{k-24} and 23 last forecasts $\{\hat{e}_k, \dots, \hat{e}_{k+22}\}$ and weather factors u_k . Inputs for the next predictor in the series at its first forecasting step contain 24 predictions from the previous predictor and weather factors. This predictor also performs 24 forecasting steps, just

as the previous one. The number of predictors in the series is not limited.

A training set for each of the neural predictors is determined by a special procedure which we called the measuring window. The method uses the fact that there is an annual cyclical component in the process of energy consumption and the assumption that this process is stationary in the neighbourhood of a forecast starting point (k). The measuring window is defined by R - the number of years back and variables τ_1 and τ_2 specifying, respectively, the number of days left and right from $k - 1$. The following conditions $\tau_1 < k - 1 < \tau_2$ must be met, where $k - 1$ is the last known observation. The training set for the predictor P_m^c is chosen from data restricted by the window and belonging only to class c (Fig.1b).

Finally, the complete algorithm is as follows. Suppose that:

- From $\{m_f\}_{f=1}^M$ forecasting models, m_f is chosen.
- There is a set of data \mathbb{Z} containing the time series: electricity consumption $\{e_k\}_{k=1}^K$, temperature $\{t_k\}_{k=1}^K$, insulation $\{i_k\}_{k=1}^K$, humidity $\{h_k\}_{k=1}^K$.
- K is the start point of forecasting; a is its max advance.

Data classification:

- Suppose that there are $\{c_i\}_{i=1}^C$ classes of daily electricity consumption (24 hour profiles). The number and range of these clusters were set by expert classification.
- Each day of a calendar year is assigned to an appropriate class. Therefore, each element e_k^c has information about membership to the appropriate class c .

Creation and preliminary training of predictors:

- A predictor is an MLP neural network. There are as many neural networks created as there are classes of data.
- Each predictor P_m^c is trained on an appropriate set $Z_{m,p}^c$ selected from the whole set \mathbb{Z} . The set $Z_{m,p}^c$ consists of pairs: an expected output e_k^c and input vector $x_{m,k}^c$. Each value e_k^c must belong to class c and each vector $x_{m,k}^c$ must satisfy the criteria of the model m .
- Finally, each predictor P_m^c has weights $W_{m,p}^c$ as a result of preliminary training. They are stored in a database.

Forecast execution and additional training (Fig. 2):

- The period $\{K + 1, a\}$, which will be predicted, is divided into days. Each of these days is assigned to the appropriate class c . Thus, corresponding predictors P_m^c are selected from the database and connected in a series.
- Each predictor P_m^c in the series is additionally trained using the set $Z_{m,p}^c$. Training begins from its last weights $W_{m,p}^c$. The set $Z_{m,p}^c$ contains pairs: $e_k^c, x_{m,k}^c$. Each value e_k^c must belong to cluster c and must be located within the measuring window defined by $O_k = \{R, \tau_1, \tau_2\}$ (Fig.1b). The vector $x_{m,k}^c$ must satisfy the criteria of the model m .
- Each predictor in the series performs forecasts for 24 steps ahead. The forecasted vector of energy consumption carried out by one predictor is supplemented by the weather factors and becomes an input for the next predictor in the series.

The neural system with a classifier and predictors.

Grouping data and modelling based on a calendar turned out to be an effective tool, but it requires a considerable amount of knowledge about the forecasted process. Instead of the expert classification we proposed a system consisting of a

classifier and a set of predictors. The classifier was built on the basis of a self-organisation network with the Kohonen algorithm, while the predictors were built based on a multi-layer perceptron network or a radial basis functions network. Profiles containing energy consumption and other additional factors are grouped by the classifier. Each predictor corresponds to the group of data specified by the classifier. It was assumed that each predictor performs a forecast only for one hour ahead. The proposed system operates in two modes - training and forecasting. The forecasting algorithm is as follows: making the same assumptions about the available time series and forecasting models and advance predictions as in the previous algorithm.

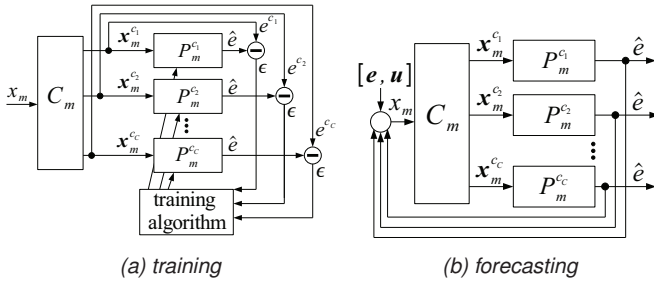


Figure 3. The neural system with a classifier and predictors.

The preliminary training stage (Fig. 3a):

- A training set \mathbb{Z}_T of pairs $\{\mathbf{x}_{k,m}, e_k\}_{k=1}^K$ is generated from the data set \mathbb{Z} , where $\mathbf{x}_{k,m}$ is an input vector dependent on the chosen forecasting model (m) and e_k is an expected output of this model.
- The classifier C_m performs a grouping of data in a self-organising process. Vectors $\mathbf{x}_{k,m}$ are the input data for the classifier which divides them into $c = \{1, \dots, C\}$ clusters. As the classification result for each vector $\mathbf{x}_{k,m}$ is assigned a corresponding class c , thus the following sets are obtained $\mathbb{Z}_m^c = \{\mathbf{x}_{k,m}^c, e_k^c\}_{k=1}^K$.
- For each set \mathbb{Z}_m^c a suitable predictor P_m^c is created, which means that a neural network is trained on the set \mathbb{Z}_m^c .
- The classifier and predictors are stored in a database.

The forecasting stage (Fig. 3b):

- The classifier C_m is read from the database.
- Step k is executed. The vector $\mathbf{x}_{k,m}$ is formed from the real values of energy consumption e and additional factors \mathbf{u}_k depending on the forecasting model m .
- (L) The classifier determines c class of $\mathbf{x}_{k,m}^c$ vector.
- The predictor P_m^c of class c is read from the database. The predictor is additionally trained on \mathbb{Z}_m^c set (training begins from recently obtained weights).
- The forecast of value \hat{e}_k is performed and the weights of the predictor P_m^c are re-stored in the database.
- If $k = a$ then the algorithm stops, otherwise $k = k + 1$, the vector $\mathbf{x}_{k,m}$ is rebuilt in such a way that the last predicted value is added to $\mathbf{x}_{k,m}$ and the oldest value of energy consumption (real or predicted) is removed from $\mathbf{x}_{k,m}$. Finally, additional inputs \mathbf{u}_k are upgraded. The next step is executed from (L) point.

Risk assessment of prediction models

Problem formulation. It was assumed that there is an electric energy distribution company which supports a group of retail customers. The company purchases energy on the wholesale market in order to ensure energy supplies to its customers. The first contract in the company's portfolio will be a retail electricity contract S_r , which is the sum of all con-

tracts for providing energy to retail customers. Retail agreements do not indicate how much electricity will be supplied at every hour, it is only an obligation to maintain an energy link with a certain maximum power. Therefore, the company, in order to know the amount of energy in the S_r contract, must perform the forecast demand $\hat{e}_{m,k}$ for this group of energy consumers (prediction algorithms have been discussed in the previous chapters). The value of the sales contract for one day is the following:

$$(3) \quad S_r = \sum_{k=1}^{24} (\hat{e}_{m,k} \cdot p_r)$$

where p_r is the retail price for $1MWh$. Energy retail prices are determined according to the tariff of the Energy Regulatory Office, so they are permanent and are not a source of risk, while the source is the inaccuracy of energy demand forecasts. Measuring risk is a representation of risk through appropriate measures. Thus, the obtained values express an amount of probable loss or likelihood of risky events. Moreover, the measurement of risk allows us to take appropriate risk management strategies for instance in order to its minimise. Energy prices varies during the day, season, therefore the prediction errors are more or less valuable. In this way, it is possible to attach a weight to the prediction errors and prepare a prediction model to minimize a function of weighted least squares error. This is a possible way for quantifying risk but we propose, as we believe, more sophisticated methods. Statistical risk measures, such as VaR and CVaR, can be used to assess the risk of electricity demand forecasting models. If we use these metrics to determine which prediction models are less risky we can minimise the risk of the energy sales contract (3).

Value-at-Risk (VaR). VaR is defined as the maximum value of loss that can occur on a single financial instrument or the whole of a portfolio for a given time horizon (T) and a confidence interval (r). A source of risk in financial markets is price volatility, so this is a measure of the market risk of price. If it is assumed that prices are normally distributed (σ is the price volatility) and S is the value of a financial instrument, VaR is calculated as follows:

$$(4) \quad VaR = \sqrt{T} \cdot r \cdot \sigma \cdot S$$

The above equation is true for financial markets, but for the electricity market it requires some modifications. It has been noted that the source of risk in the retail electricity contract is the inaccuracy of the forecasting models. If we consider only one day $T = 1$ and the confidence interval r equal to 95% the risk of the sales contract S_r (3) can be determined by the relationship:

$$(5) \quad VaR_{S_r}^{95} = 1.65 \cdot \sum_{k=1}^{24} (\sigma_{m,k} \cdot \hat{e}_{m,k} \cdot p_r)$$

where $\sigma_{m,k}$ is forecast error variability and $\hat{e}_{m,k}$ is the value of forecast in k hour generated by the model m . The value of 1.65 means that the percentage change in errors of the forecast ($\sigma_{m,k}$) corresponds to a distance of no more than 1.65 standard deviations σ . This means that 90% of all error changes of the forecast do not exceed the 1.65 standard deviation. Thus, only 5% of changes up and 5% of changes down of the forecast errors will exceed the limit of $1.65 \cdot \sigma$. Finally, we can say that the loss on the sales S_r will not exceed the value of $VaR_{S_r}^{95}$ with 95% certainty.

Conditional Value-at-Risk (CVaR). VaR says how much you can lose in 95% of cases, but it says nothing about the remaining 5% of cases corresponding to extreme events. The answer to this problem is conditional value at risk, otherwise known as a measure of the average shortfall. It measures the average of potential losses that exceed the value of VaR. In practice, on financial markets CVaR is calculated using the historical simulation procedure. This procedure, after some modifications, can be used for the electricity market. CVaR of the contract S_r (3) is calculated as follows:

- Suppose that there are predictions generated by $\{m_f\}_{f=1}^M$ forecasting models, so M time series with forecast errors can be determined. These errors are calculated as follows:

$$(6) \quad APE_{m,k} = \left| \frac{e_{m,k} - \hat{e}_{m,k}}{e_{m,k}} \right| \cdot 100\%$$

- Time series containing forecast error changes ΔAPE (7) are created. Forecasts from each forecasting model (m) and for each hour ($k = 1, \dots, 24$) are considered as a separate time series, so $24 \cdot M$ series are obtained.

$$(7) \quad \Delta APE_{m,k}^d = (APE_{m,k}^{d-1} - APE_{m,k}^d) / 100$$

d is the day of past data and it is no greater than d_{max} .

- Next, $24 \cdot M$ time series $\Delta S_{m,k}^d$ containing changes in the values of the contract S_r (3) are calculated as follows

$$(8) \quad \Delta S_{k,m}^d = \Delta APE_{k,m}^d \cdot \hat{e}_{k,m} \cdot p_r$$

- Finally, each time series ($\Delta S_{m,k}$) is ordered ascending. Then, the level 100% of confidence is assigned to the first item in the time series and 1% to the last one - this way empirical inverse distributions are designated. $VaR_{S_r}^{95}$ is the value in the series which corresponds to a 6% level of confidence (if more than one value corresponds to this level then its average should be counted). The $CVaR_{S_r}$ is calculated as the average of values in the series which correspond to confidence levels of 5% to 1%.

Experiments

Forecasting by the series of MLP predictors. The first experiment consisted of forecasting electricity demand for seven days ahead by using a series of predictors. MLP networks were used as predictors so that the forecasting algorithm was called a series of MLP predictors.

Twenty types of daily profiles of energy consumption have been defined based on expert knowledge, the calendar and data analysis. Among them, seven classes stand out which correspond to the days of the week, and thirteen classes which correspond to the holidays and significant days in Poland. This means that we have twenty types of predictors $\{P^c\}_{c=1}^{20}$, namely neural networks, each with its own set of weights W_p .

Many experiments were performed to determine the parameters of the predictors. They were carried out by the 10-fold cross-validation procedure, using data selected from 2004, and using the forecasting model m_2 . It was found that the optimal network is the one with one hidden layer which contains ten hidden neurons. The number of inputs and outputs of the network depends on the selected model m . All neurons in the network have unipolar sigmoidal activation functions. The derivatives of these functions are modified according to Fahlman's proposition [9]. It was found that

the application of a two-phase neural network training would be optimal. At the first stage, the *QuickProp* algorithm [9] tries to find the global minimum of the objective function. At the second stage of training, the *backpropagation* algorithm without momentum but with a low value of learning ratio tries to precisely determine the minimum of the objective function. The stop condition for both training algorithms is achieving the maximum number of iterations $it = 10^3$ or the $MAPE < 1\%$ (9).

$$(9) \quad MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i - \hat{e}_i}{e_i} \right| \cdot 100\%$$

Data were available for the years 2002-2004. Forecasts were made for the period from 7 January 2004 to 24 December 2004 in such a way that for each day of the period they were performed 7 days ahead. We omitted one week at the beginning and one week at the end of 2004 because the set of training data was not adequate for these days. The following dimensions of the measuring window O_p were chosen: $R = 2$ years, $\tau_1 = \tau_2 = 15$ days. Predictors were trained using data lying between 1 January 2002 and the starting point of forecast, and which belonged to the appropriate class c inside the window O_p . We tested eight forecasting models $m_1 - m_8$.

Table 2. $MAPE$ of forecasts in 2004 and in September 2004 using forecasting models $m_1 - m_8$ (the best results are in bold).

ahead	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
annual average of 2004								
1	1,81	1,82	1,76	1,89	1,76	1,88	1,86	1,79
2	2,53	2,45	2,39	2,65	2,33	2,55	2,54	2,40
3	3,05	2,93	2,89	3,21	2,75	3,05	3,04	2,81
4	3,56	3,34	3,32	3,71	3,06	3,54	3,46	3,20
5	4,16	3,84	3,75	4,23	3,38	3,99	3,84	3,62
6	4,79	4,31	4,07	4,69	3,65	4,41	4,16	3,93
7	5,44	4,74	4,39	5,18	3,91	4,82	4,51	4,31
Avg	3,62	3,35	3,22	3,65	2,98	3,46	3,44	3,15
September 2004 (the best case)								
1	1,23	1,26	1,27	1,26	1,20	1,25	1,29	1,28
2	1,64	1,74	1,64	1,70	1,59	1,58	1,70	1,64
3	1,77	1,93	1,82	1,84	1,78	1,70	1,85	1,80
4	2,00	2,13	1,96	1,99	1,82	2,01	1,99	1,86
5	2,19	2,24	2,01	2,06	1,83	2,05	2,00	2,00
6	2,54	2,38	2,13	2,13	1,90	2,07	2,04	2,15
7	2,95	2,66	2,16	2,23	1,97	2,25	2,18	2,36

An overview of the forecast errors is in table 2 (annual average and the best month). An adverse phenomenon of a rising forecast error with forecast advancement was observed since the errors propagate from one model to the next in the series. The best results are achieved with models (m_5, m_8) which contain the following inputs: energy, temperature and insolation. The supposition that humidity will affect the amount of energy used was not confirmed because models which take this parameter into account do not give less false predictions than models that do not contain it. However, an improvement of forecast accuracy was observed for certain months of winter and spring when this parameter was taken into account. Hence, it can be concluded that, depending on the time of year, models with a different set of weather factors should be used. The forecasting algorithm, constructed as a series of MLP predictors where each model represents an appropriate class of the day, gives a satisfactory estimate of

energy demand in the short-term horizon and has been confirmed experimentally. It clearly indicates that the proposed methods of forecasting can be useful in the everyday work of an electricity trading company.

Forecasting by the system SOM-MLP. The analysis results of electricity demand forecasts which had been made with the series of MLP predictors showed that by far the best results were achieved for September 2004, irrespective of the chosen forecasting model. Therefore, we decided to perform a second experiment in order to verify whether it is possible to obtain more accurate forecasts for this month. The experiment consisted of forecasting electricity demand in September 2004 for two days ahead by using the system with a SOM classifier and MLP predictors.

The configuration of MLP predictors was the same as in the previous experiment (section). The classifier training algorithm is the winner takes most. The algorithm parameters were determined experimentally. The Gaussian neighbourhood and the Euclidean distance measure were used. The neighbourhood and the learning coefficient were reduced exponentially in each epoch of the algorithm. Self-organisation of the network was stopped when 500 epochs of training were carried out or the aggregate change in weights was less than 10^{-3} .

It was assumed that the classifier (SOM network) was taught on data for the period from 1 January 2004 to 30 August 2004 and it was not additionally trained. Forecasts for two days ahead (48 hours) were made every day for the period from 1 September 2004 to 30 September 2004. Firstly, the forecasting model m_8 was tested. Surprisingly, very inaccurate forecasts were obtained for the non-typical days. This was due to the fact that the classifier, when deciding which predictor should make a forecast for the next hour, had no adequate information in the input vector that the hour belonged to one of the significant days. Hence the classifier could not select the appropriate predictor. Therefore, it was decided to use models $m_9 - m_{11}$, which contain additional data, such as the type of week which is represented by numerical values: $D = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$; the type of Polish holiday or other significant days (13 days) which are represented by numerical values: $S = \{0.05, 0.1, \dots, 0.65\}$; the daily \bar{e}_{24} and weekly \bar{e}_{168} moving averages which vary across seasons. It was found that the optimal size of the SOM matrix is 5×5 , so no more than 25 clusters could be obtained. This gives an average of 930 vectors per training set.

Table 3. MAPE for forecasts made by the systems SOM-MLP and SOM-RBF (models $m_{10} - m_{14}$, SOM matrix 5×5).

the system	ahead	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}
SOM-MLP	1	2,02	4,62	1,98	1,86	2,12	2,04
SOM-MLP	2	2,59	7,66	2,54	2,49	2,87	3,02
SOM-RBF	1	—	3.40	3.31	3.06	3.44	5.04
SOM-RBF	2	—	4.47	4.53	3.46	3.88	6.33

The results are in table 3. It was found that adding information about the type of working days and holidays to the input vector made the classifier correctly select an adequate predictor. A comparison of errors proves that the moving average as input to the forecasting model improves the quality of predictions. The relatively large error of the model m_{10} indicates that weather factors are essential for the proper operation of the forecasting model. If one compares the forecast errors for September 2004 for the second day ahead which are made by the system 'series of MLP predictors' and by

the system 'classifier SOM-MLP predictors', it turns out that the error of the first system for any forecasting models does not exceed 1.75%, while the smallest error from the second system is 2.49%.

Forecasting by the system SOM-RBF. In previous experiments MLP networks were used as predictors. The obtained forecasts were satisfactory but this does not exempt us from seeking other, better solutions. For this reason the possibility of applying RBF networks as predictors was investigated. The third experiment consisted of forecasting electricity demand in September 2004 for two days ahead by using the system with the SOM classifier and RBF predictors.

The classifier configuration parameters were the same as in the previous experiment with the system SOM-MLP, whereas the parameters of the RBF network were determined experimentally. According to the theory, the RBF network always has three layers: inputs, rbf layer and outputs. The number of inputs depends on the forecasting model used. The output layer is a linear adder with a single output since the network is designed to perform a single prediction in accordance with the adopted forecasting model (1). The widths of all functions were constant and chosen experimentally. The centres of the functions were selected randomly from a training set, so its quantity depended on the size of the set. It has been experimentally established that it amounts to 1% of this set. The weights of the linear combiner were calculated by the pseudo-inversion procedure. In cases where there was an irreversible singular matrix, the centres c were drawn again and weights w were again calculated.

Experiments were performed on the same conditions as in the previous experiment (system SOM-MLP): the classifier (SOM network) was taught on data for the period from 1 January 2004 to 30 August 2004; forecasts were performed for two days ahead (48 hours) for each day of the period from 1 September 2004 to 30 September 2004; five forecasting models $m_{10} - m_{14}$ were used. The results are in table 3. Unfortunately, we did not achieve better predictions than those achieved from the system SOM - MLP. RBF networks have a lower ability of generalising a problem than MLP networks do and, therefore, incorrect results were obtained in cases where the forecast went beyond the training data.

The best forecasting model according to VaR. Many energy consumption forecasts for the same period have been obtained in previous experiments. Now we will try to evaluate which of the forecasts is the most appropriate in order to use it to estimate aggregate value of a sales contract (3) in the retail market. Three evaluation criteria of the forecasting models are applied.

Criterion 1. The forecasts are selected from only one forecasting model (m) which generates the smallest $MAPE_m$ (9) in the given predictions period

$$(10) \quad \min \{MAPE_m\}$$

Criterion 2. The forecasts are selected independently for each hour $k = 1 \dots 24$ from this forecasting model (m) which gives at hour k the smallest

$$(11) \quad \min \{MAPE_{k,m}\}$$

$$(12) \quad MAPE_{k,m} = \frac{\sum_{i=1}^z APE_{k,m}(i)}{z}$$

where $APe_{k,m}$ (6) is calculated only in hours (k) in the given predictions period ($i = 1..z$).

Criterion 3. The forecasts are selected independently for each hour $k = 1..24$ from this model (m) which gives the smallest standard deviation of prediction error

$$(13) \quad \min \{ \sigma_{k,m} \}$$

$$(14) \quad \sigma_{k,m} = \sqrt{\frac{\sum_{i=1}^z (PE_{k,m}(i) - \overline{PE}_{k,m})^2}{(z-1)}}$$

where $PE_{k,m} = \left| \frac{e_{k,m} - \hat{e}_{k,m}}{e_{k,m}} \right| \cdot 100\%$ is calculated only in hours (k) in the given predictions period ($i = 1..z$), and $\overline{PE}_{k,m}$ is the average of $PE_{k,m}$ in this period.

The evaluation of predictions was carried out as follows:

- Only predictions generated by the forecasting models $m_1 - m_8$ were taken into account.
- Calculations of errors ($MAPE_m, MAPE_{k,m}, \sigma_{k,m}$) were made for the period from 7 January 2004 to 14 September 2004, but only forecasts for the second day ahead were taken into account.
- Using the above criteria, three of the forecasting models were selected at each hour $k = 1..24$ of 15 September 2004 (example results are in Tab. 4). Three sets of predictions were obtained as the results.
- By using these sets the values S_r (3) and VaR of the contract (5) were calculated separately for every hour of 15 September 2004.
- Finally, contract values and contract risks were summed up for the day (these values are given in Tab. 5).

As can be noted, the selection of forecasts based on the risk assessment of the forecasting model (criterion 3) leads to contract risk minimisation. The values given in the row of the table for criterion 3 are to be understood as follows: it is forecast that revenue from the sales contract will be 3.583.925, 80 PLN, but it is expected, with 95% confidence, that revenue from the contract will not fall by more than 175.963, 85 PLN. It should be noted that the value of the contract does not affect the value of risk of the forecasting model. For criterion 3 we see that the value of the contract is the largest, but the risk is lower than for criteria 1 and 2. Finally, we can say that the choice of the forecasting model according to the criterion of risk (expressed as standard deviation error forecasts of electricity demand, i.e. the VaR) carried out separately for each hour of the day is well-founded as it leads to the choice of the least risky forecasting model, which translates into lower economic risks.

Table 4. Examples of forecasting model selection.

Criterion	1	2	3	...	22	23	24
1. (eq. 10)	m_5	m_5	m_5	...	m_5	m_5	m_5
2. (eq. 11)	m_8	m_8	m_8	...	m_5	m_5	m_1
3. (eq. 13)	m_8	m_8	m_8	...	m_3	m_2	m_2
4. (eq. 15)	m_3	m_8	m_3	...	m_4	m_7	m_8
5. (eq. 16)	m_8	m_8	m_8	...	m_3	m_3	m_1

The best forecasting model according to CVaR. The conditional value at risk provides information about potential losses generated by the prediction models that go beyond the confidence level of VaR. In this experiment we will try to evaluate CVAR (and also VaR) of the forecasting models $m_1 - m_8$. Two evaluation criteria of the forecasting models are applied:

Criterion 4. The forecasts are selected independently for each hour $k = 1..24$ from this model (m) which gives the smallest VaR loss

$$(15) \quad \min \{ VaR_{k,m} \}$$

Criterion 5. The forecasts are selected independently for each hour $k = 1..24$ from this model (m) which gives the smallest CVaR loss

$$(16) \quad \min \{ CVaR_{k,m} \}$$

Calculations were carried out using the historical simulation procedure. It involves the following steps:

- Only predictions generated by the forecasting models $m_1 - m_8$ were taken into account and
- Firstly, the series $\Delta APE_{m,k}$ was created for $m_f, f = \{1, \dots, 8\}$ models and for each hour of $h = \{1..24\}$ separately. For this purpose only data for the last 100 days, i.e. the period from 6 June 2004 to 14 September 2004 was used. Thus, $8 \times 24 = 192$ series were obtained, each the size of 100 samples.
- Next, 192 of series $\Delta S_{k,m}$ were calculated (eq. 8) using the demand forecasts of energy ($\hat{e}_{k,m}$) per day 15.9.2004 and assuming that the price of energy was $c_r = 250 PLN$.
- Each series $\Delta S_{k,m}$ was sorted ascending. Then, for each item of the series the confidence level range from 100% to 1% was assigned. In this way 192 inverse empirical distributions were calculated.
- Then VaR was read as the item with a confidence level equal to 6% and CVaR was read as the average value of items with a confidence level of 5% to 1%. Values of VaR and CVaR for every 8 models and for each of the 24 hours were received as the results.
- Then the forecasting models for each hour of the day of 15 September 2004 were selected on the basis of criterion 4 - minimum VaR or criterion 5 - minimum CVaR (example results are in table 4).
- Finally, the values VaR and CVaR and values of the contract were respectively summed up for the entire day (the results of this operation are shown in table 5).

Table 5. Summary risk estimates (VaR, CVaR) and contract values (S).

Criterion	S	VaR	CVaR
1. (eq. 10)	3 581 295,02	179 859,31	—
2. (eq. 11)	3 580 317,47	178 563,85	—
3. (eq. 13)	3 583 925,80	175 963,85	—
4. (eq. 15)	3 581 092,95	100 499,00	172 057,72
5. (eq. 16)	3 587 083,89	105 327,03	141 387,35

The values VaR and CVaR in table 5 should be understood as follows: It is forecast that the income from the sales contract will be S , but it is to be expected, with 95% certainty, that the revenue from the contract will not fall by more than VaR, and it is also to be expected that the average decline in the value of the contract will not be greater than CVaR if events with 5% certainty occur. Provided that in the first case (criterion 4) it is possible to find such parameters of the contract which will give a lower risk of VaR loss and in the second case (criterion 5) it is possible to find such contract parameters which will give a lower risk of CVaR loss. It should be noted that the VaR which was calculated by the historical simulation method is less than the VaR which was calculated by the analytical method. Therefore, it was concluded that the

choice of methods of calculation risk and data used for the simulation have an influence on the results.

Conclusions

The ideas presented in this paper combine aspects of artificial intelligence, forecasting, risk analysis and risk management, and present them on the example of the electricity market in Poland. Statistical autocorrelation tests of electricity consumption and analyses of correlation between electricity consumption and climatic factors (humidity, insolation and temperature) were performed. On this basis, and based on the literature, algorithms for forecasting the demands for electricity were developed. The first algorithm of 'the series of MLP predictors' was successfully tested in a real business environment. Forecasts of energy consumption of a group of consumers for two days ahead, obtained using this algorithm, were the basis for planning the purchase of energy on the wholesale market. The second proposition was the two-tier algorithm containing the classifier (SOM) and predictors (MLP or RBF). We believe that this algorithm is suitable for forecasting little-known phenomena (time series), as opposed to the first algorithm, which needs large expert knowledge about the modelled process. Finally, we presented methods of risk measurement for model forecasting based on VaR and CVaR. It has been shown, through computational examples, that by choosing energy forecasts which are generated by the forecasting model with less risk the company is less vulnerable to economic losses than by choosing projections based on historical prediction errors of the model. In other words, the selection criterion of the forecasting model is the risk rather than the accuracy of the model. We believe that the presented method for assessing forecasting models may be useful in decision-making in the electricity market.

REFERENCES

- [1] S. A. Billings and X. Hong. Dual-orthogonal radial basis function networks for nonlinear time series prediction. *Neural Networks*, 11:479–493, 1998.
- [2] R. Bjorgan, Chen-Ching Liu, and J. Lawarree. Financial risk management in a competitive electricity market. *IEEE Transactions on Power Systems*, 14(4):1285–1291, 1999.
- [3] H. Choungfu. Fuzzy risk vs. probability risk analysis. In *Soft Computing for Risk Evaluation and Management. Applications in technology, Environment and Finance*. Physica-Verlag, Heidelberg, 2001.
- [4] R. Dahlgren, Chen-Ching Liu, and J. Lawarree. Risk assessment in energy trading. *IEEE Transactions on Power Systems*, 18(2):503–511, 2003.
- [5] P. K. Dash, A. C. Liew, S. Rahman, and S. Dash. Fuzzy and neuro-fuzzy computing models for electric load forecasting. *Engineering Applications of Artificial Intelligence*, 8(4):423–433, 1995.
- [6] M. Denton, A. Palmer, R. Masiello, and P. Skantze. Managing market risk in energy. *IEEE Transactions on Power Systems*, 18(2):494–502, 2003.
- [7] I. Drezga and S. Rahman. Input variable selection for ann-based short-term load forecasting. *IEEE Transaction on Power System*, 13(4):1238–1244, 1998.
- [8] I. Drezga and S. Rahman. Short-term load forecasting with local ann predictors. *IEEE Transaction on Power System*, 14(3):844–850, 1999.
- [9] Scott E. Fahlman. An empirical study of learning in back-propagation networks. *Computer Science Technical Report CMU-CS-88-162*, 1998.
- [10] Gregory C. Reinsel George E. P. Box, Gwilym M. Jenkins. *Time series analysis : forecasting and control*. Prentice-Hall, Englewood Cliffs, 1994.
- [11] H. S. Hippert, C. E. Pedreira, and R. C. Souza. Neural network for short-term load forecasting: A review and evaluation. *IEEE Transaction on Power System*, 16(1):44–55, 2001.
- [12] H. R. Kassaei, A. Keyhani, T. Woung, and M. Rahman. A hybrid fuzzy, neural network bus load modeling and prediction. *IEEE Transaction on Power System*, 14(2):718–724, 1999.
- [13] R. Katayama, K. Kuwata, Y. Kajiytani, and M. Watanabe. Embedding dimension estimation of chaotic time series using self-generating radial basis function network. *Fuzzy sets and systems*, 71:311–327, 1995.
- [14] Ath. Kehagias and Vas. Petridis. Predictive modular neural networks for time series classification. *Neural Networks*, 10(1):31–49, 1997.
- [15] Kwang-Ho Kim, Hyoung-Sum Youn, and Young-Cheol Kang. Short-term load forecasting for special days in anomalous load conditions using neural networks and fuzzy inference method. *IEEE Transaction on Power System*, 15(2):559–565, 2000.
- [16] H. Leung, T. Lo, and S. Wang. Prediction of noisy chaotic time series using an optimal radial basic function neural network. *IEEE Transaction on Neural Networks*, 12(5):1163–1172, 2001.
- [17] P. A. Mastorocostas, J. B. Theocharis, and A. G. Bakirtzis. Optimal fuzzy inference for short-term load forecasting. *IEEE Transaction on Power System*, 14(1):29–36, 1999.
- [18] H. Mori and A. Yuihara. Deterministic annealing clustering for ann-based short-term load forecasting. *IEEE Transaction on Power System*, 16(3):545–551, 2001.
- [19] V. Petridis and A. Kehagias. Predictive modular fuzzy systems for time-series classification. *IEEE Transaction on Fuzzy System*, 5(3):381–397, 1997.
- [20] P. Piotrowski. Neural network with genetic algorithms for monthly electric energy consumption and peak power middle - term forecasting. *Journal of Applied Computer Science*, 10(1):105–115, 2002.
- [21] T. Senjyu, H. Takara, K. Katsumi, and T. Funabashi. One-hour-ahead load forecasting using neural network. *IEEE Transaction on Power System*, 17(1):113–118, 2002.
- [22] D. Srinivasan, C. S. Chang, and A. C. Liew. Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting. *IEEE Transaction on Power System*, 10(4):1897–1903, 1995.
- [23] J. Vermaak and E. C. Botha. Recurrent neural network for short-term load forecasting. *IEEE Transaction on Power System*, 13(1):126–132, 1998.
- [24] H. Yoo and R. L. Pimmel. Short term load forecasting using a self-supervised adaptive neural network. *IEEE Transaction on Power System*, 14(2):779–784, 1999.
- [25] R. Zemouri, D. Racoceanu, and N. Zerhouni. Recurrent radial basis function network for time-series prediction. *Engineering Applications of Artificial Intelligence*, 16:453–463, 2003.

Authors:

Jarosław Protasiewicz (PhD); Information Processing Institute, Laboratory of Intelligent Information Systems; al. Niepodległości 188 b, 00-608 Warsaw, Poland; e-mail: jaroslaw.protasiewicz@opi.org.pl.

Piotr Stanisław Szczepaniak (Full Professor); Technical University of Łódź; Faculty of Technical Physics, Information Technology and Applied Mathematics; Institute of Information Technology; ul. Wólczańska 215, 90-924 Łódź, Poland; e-mail: office@ics.p.lodz.pl