# Medium-term Electric Energy Demand Forecasting using Nadaraya-Watson Estimator

Grzegorz Dudek, Paweł Pełka
Department of Electrical Engineering
Czestochowa University of Technology
42-200 Czestochowa, Al. Armii Krajowej 17, Poland
dudek@el.pcz.czest.pl, pavelle50@gmail.com

***Abstract*—This work is focused on monthly electricity demand prediction, which is necessary for the maintenance planning in power systems as well as for negotiation forward contracts. In the proposed approach patterns of the load time series are defined, which unify input and output data and filter out the trend. Relationships between inputs and outputs simplified due to patterns are modeled using nonparametric regression: Nadaraya-Watson estimator. In the experimental part of the work the model is examined using real-world data. The results are encouraging and confirm the high accuracy of the model and its competitiveness compared to other forecasting models.**

***Keywords—Nadaraya-Watson estimator; medium-term load forecasting; pattern-based forecasting***

## I. Introduction

Medium-term electric energy demand forecasting plays a key role in power system planning and operation. There are three main factors which decide about the need of medium-term forecasting [1]:

- technical and operational - scheduling maintenance activities, planning of production levels and fuel purchases, planning of network investments,
- economical - negotiating contracts between energy companies, concluding contracts with customers,
- legal - obligation to prepare forecasts imposed by the Energy Law.

The time series of monthly electric energy demand express usually a rising tendency caused by the influence of economic and technological development on the electric market. Seasonal variations reflect the annual cycle and are caused by climatic factors, which are similar in the same month of different years. Also specific political decisions and economical factors can affect the demand and disturb general rising trend and monthly fluctuations. Examples of monthly demand time series for four European countries are shown in Fig. 1. They differ depending on the power system size and economic development of the country. Note large, regular annual cycles for France, and significant random component for other countries.

The medium-term load forecasting (MTLF) can be divided into two general categories [2]. The first category (conditional modeling approach) focuses on economic analysis, management and long term planning and forecasting of energy load and energy policies. MTLF takes into account socioeconomic conditions which impact energy demands. They are represented by some economic indicators and electrical infrastructure measures in addition to information on historical load data and weather related variables. Similar approach for long-term load forecasting can be used. An example of forecasting model of this type is presented in [3]. This model includes macroeconomic indicators, such as the consumer price index, the average salary earning and the currency exchange rate in their MTLF analysis.

The second category (autonomous modeling approach) requires a smaller set of input information, primarily historical loads and weather information to forecast future electricity demand. This approach is more suited for stable economies. It employs classical methods such as ARIMA or linear regression [4] as well as computational intelligence methods, e.g. neural networks [5]. In [6] the authors propose a MTLF model that incorporates various weather related parameters and historical load profiles. Various modeling tools, such as ARIMA, neural networks and neuro-fuzzy systems are employed to forecast future load demand. In [7] a neural network was used that includes temperature profiles to estimate medium term power demand. The authors have shown that their neural net formulation outperformed existing models that were based on linear and non-linear regressions. A

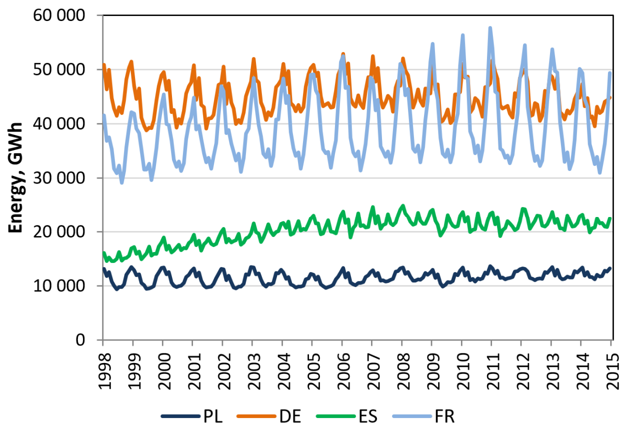


Fig. 1. Time series of monthly electric energy demand for four European countries.

978-1-5090-6406-9/17/$31.00 ©2017 IEEE

knowledge based expert system was proposed in [8], that uses a set of decision rules for forecasting demand of fast growing utilities. It can identify key variables as well as the most suitable load-forecasting model.

A method which is proposed in this work can be classified into autonomous modeling approaches and similarity-based methods which are a basis of many algorithms of machine learning, pattern recognition and data mining. They are very useful for problems of data classification and clustering, regression, filling missing values and create associative memory [9]. An underlying assumption in similarity-based models is: unknown features of the object can be predicted on the basis of a set of similar objects for which these features are known. Objects are seen through their patterns, i.e. sets of measurable parameters, attributes, properties. Definition of patterns and data representation are the key issue here. Another key issue is definition of similarity measure between patterns.

In the case of time series the similarity-based methods use a similarity measure between shapes of the time series fragments. It is assumed that if two fragments of the time series are similar in shape, then the fragments following them are also similar in shape [10]. This approach is especially attractive when the time series expresses seasonal pattern. In [11] we used $k$ nearest neighbor as a similarity-based forecasting model. In this work we use Nadaraya-Watson estimator (N-WE), which allows us to take into account similarity degree between shapes of the time series fragments.

The remainder of the paper is organized as follows. In Section II, input and output patterns of the time series fragments are defined. Section III describes N-WE and the method of its parameters estimation. In Section IV, the proposed model is tested on real-world data. Finally, Section V concludes this paper.

## II. Patterns of Time Series Fragments

Fragments of time series are represented by patterns, which are vectors of $n$ or $m$ components. Each component is a preprocessed time series point. Input and output patterns are defined: $\mathbf{x} = [x_1\ x_2\ \dots\ x_n]^T$ and $\mathbf{y} = [y_1\ y_2\ \dots\ y_m]^T$, respectively. The patterns are paired ($\mathbf{x}_i$, $\mathbf{y}_i$), where $\mathbf{y}_i$ is a pattern of the time series fragment succeeding the fragment represented by pattern $\mathbf{x}_i$. A distance in time between these fragments is equal to the forecast horizon $\tau$.

Pattern similarity-based forecasting methods are based on the following assumption [10]: If the input patterns $\mathbf{x}_a$ and $\mathbf{x}_b$ are similar to each other, then patterns $\mathbf{y}_a$ and $\mathbf{y}_b$ paired with them are similar to each other as well. The above assumption allows us to forecast the pattern $\mathbf{y}_a$ on the basis of known patterns $\mathbf{x}_a$, $\mathbf{x}_b$ and $\mathbf{y}_b$, which are determined from the history. In N-WE we calculate similarity between query pattern $\mathbf{x}_a$ and patterns $\mathbf{x}$ from the training set. Then to get the forecasted pattern $\mathbf{y}_a$ we aggregate training patterns $\mathbf{y}$ using weights dependent on the similarity between x-patterns. The method of preprocessing the time series fragments to get their patterns are dependent on the time series specificity (seasonal variations, trend), forecast period and horizon.

Let us consider the task of prediction of $m$ consecutive points of the monthly electricity demand time series. Let us denote the forecasted fragment by $Y_i = \{E_{i+1}\ E_{i+2}\ \dots\ E_{i+m}\}$, and the preceding fragment of $n$ points by $X_i = \{E_{i-n+1}\ E_{i-n+2}\ \dots\ E_i\}$. An input pattern $\mathbf{x}_i = [x_{i,1}\ x_{i,2}\ \dots\ x_{i,n}]^T$ represents the fragment $X_i$. The components of this vector are preprocessed points of the sequence $X_i$. Some examples of preprocessing are [11]:

$$x_{i,t} = E_{i-n+t} \tag{1}$$

$$x_{i,t} = \frac{E_{i-n+t}}{\overline{E}_i} \tag{2}$$

$$x_{i,t} = E_{i-n+t} - \overline{E}_i \tag{3}$$

$$x_{i,t} = \frac{E_{i-n+t} - \overline{E}_i}{D_i} \tag{4}$$

where $t = 1, 2, \dots, n$, $\overline{E}_i$ is the mean value of the points in sequence $X_i$, and $D_i = \sqrt{\sum_{j=1}^{n}(E_{i-n+j} - \overline{E}_i)^2}$ is a measure of their dispersion.

A pattern defined using (1) is a copy of the sequence $X_i$ without processing. Pattern components defined using (2) are the points of the sequence $X_i$ divided by the mean value of this sequence. The differences between points and the mean sequence value are pattern components (3). The pattern defined using (4) is a normalized vector $[E_{i-n+1}\ E_{i-n+2}\ \dots\ E_i]^T$. It has the unity length and the mean value equal to zero. Moreover, all x-patterns have the same variance.

The output pattern $\mathbf{y}_i = [y_{i,1}\ y_{i,2}\ \dots\ y_{i,m}]^T$ represents the forecasted sequence $Y_i$. The components of y-pattern can be defined similarly to the components of x-patterns:

$$y_{i,t} = E_{i+t} \tag{5}$$

$$y_{i,t} = \frac{E_{i+t}}{\overline{E}_i} \tag{6}$$

$$y_{i,t} = E_{i+t} - \overline{E}_i \tag{7}$$

$$y_{i,t} = \frac{E_{i+t} - \overline{E}_i}{D_i} \tag{8}$$

In formulas (5)-(8) $\overline{E}_i$ and $D_i$ are determined from the sequence $X_i$, and not, as one might expect $Y_i$. These values at the moment of forecasting are known and enable us to calculate the forecast of demand based on the forecast of y-pattern returned by the forecasting model. We use for this the transformed equations (5)-(8). For example, in the case of (8) the forecast of demand is calculated as follows:

$$\widehat{E}_{i+t} = \widehat{y}_{i,t} D_i + \overline{E}_i \tag{9}$$

Patterns $\mathbf{x}_i$ (representing the sequence preceding the forecasted one) and $\mathbf{y}_i$ (representing the forecasted sequence) are paired ($\mathbf{x}_i$, $\mathbf{y}_i$). The set of these pairs determined from history is used for learning the forecasting model.

## III. Nadaraya-Watson Estimator

N-WE is an example of nonparametric kernel regression for estimation the conditional expectation of a random variable. A nonlinear relationship between a pair of random variables **x** and **y** is estimated as a locally weighted average, using a kernel as a weighting function. In [12] N-WE as a forecasting tool was derived from the conditional density estimator. For vector-input, vector-output case the kernels are expressed using a multidimensional product kernel function and the N-WE is defined as follows [13]:

$$m(\mathbf{x}) = \frac{\sum_{j=1}^{N} \prod_{t=1}^{n} K\left(\frac{x_t - x_{j,t}}{h_t}\right) \mathbf{y}_j}{\sum_{j=1}^{N} \prod_{t=1}^{n} K\left(\frac{x_t - x_{j,t}}{h_t}\right)} \tag{10}$$

where $N$ is a number of training points, $K(.)$ is a kernel function and $h$ is its bandwidth.

The popular choice for the kernel function is a normal kernel. In this case N-WE is of the form:

$$m(\mathbf{x}) = \frac{\sum_{j=1}^{N} \exp\left(-\sum_{t=1}^{n} \frac{(x_t - x_{j,t})^2}{2h_t^2}\right) \mathbf{y}_j}{\sum_{j=1}^{N} \exp\left(-\sum_{t=1}^{n} \frac{(x_t - x_{j,t})^2}{2h_t^2}\right)} \tag{11}$$

As we can see from (10) and (11) N-WE is a linear combination of vectors $\mathbf{y}_j$ weighted by the normalized kernels, which nonlinearly map the distance between patterns **x** and $\mathbf{x}_j$. The distance is parameterized by the bandwidths. The bandwidth $h_t$ controls the share of the $t$-th component of **x** in the distance. The bandwidth values decide about the bias-variance tradeoff of the estimator. When they are too small the estimator tends to undersmoothing, and when they are too large it tends to oversmoothing. Thus the selection of the bandwidth values is a crucial issue. The simplest way is to adopt them from the formula proposed by Scott [14] for the normal product density estimators:

$$h_t^S = \hat{\sigma}_t N^{-\frac{1}{n+4}} \tag{12}$$

where $\hat{\sigma}_t$ is the standard deviation of the $t$-th component of **x** estimated from the learning sample, $N$ is a size of learning sample and $n$ is a number of components in **x**.

The bandwidths can be tuned by searching the neighborhood of the point $\mathbf{h}^S = [h_1^S \; h_2^S \; \ldots \; h_n^S]$. This is performed using iteration process where vectors **h** are generated one by one according to:

$$\mathbf{h}_l = a_l \mathbf{h}^S \tag{13}$$

where $a_l = a_{\min} + \Delta l$, $l = 1, 2, ..., L$.

Note, that the multi-dimensional optimization problem (searching of $h_1$, $h_2$, ..., $h_n$) is replaced with a simple one-dimensional optimization problem (searching of $a$ value). In the training process the optimal value of $a$ is selected as well as the optimal length of the input pattern $n$. These both parameters are searched using grid search method. More sophisticated methods for searching the bandwidth space using evolutionary algorithms and tournament searching in [13] are

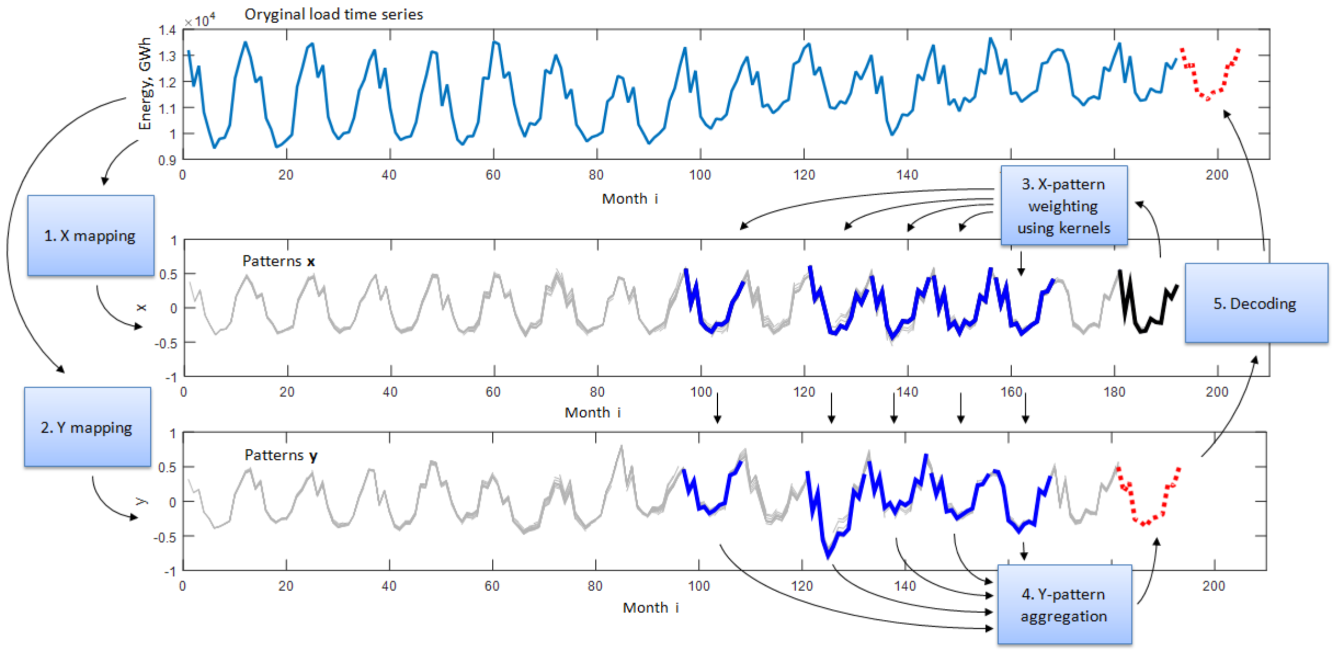


Fig. 2. Idea of the proposed forecasting model based on Nadaraya-Watson estimator.

described.

The training set contains pairs ($\mathbf{x}_i$, $\mathbf{y}_i$), which are historical for the forecasted sequence, i.e. these ones for which $i = n, n+1, ..., i^*-m$, where $i^*$ is an index of the last month before the forecasted sequence. The forecasting task is to prepare the forecasts for months $i^*+1, i^*+2, ..., i^*+m$.

The idea of the forecasting model based on N-WE in Fig. 2 is shown. In the first steps the load time series is preprocessed into patterns (steps 1 and 2). Having the query pattern $\mathbf{x}_{i^*}$ the weights for each x-pattern from history are calculated using kernels (step 3). The y-patterns are weighted and aggregated to get the forecasted pattern $\hat{\mathbf{y}}_{i^*}$ (step 4). This pattern is decoded using one of the transformed equation (5)-(8) (step 5). As a result we get the monthly electricity demand for consecutive months: $i^*+1, i^*+2, ..., i^*+m$.

## IV. Experimental Study

The proposed forecasting model was used to forecast monthly electricity demand for four European countries: Poland (PL), Germany (DE), Spain (ES) and France (FR). The data were obtained from the ENTSO-E repository (www.entsoe.eu). Data for PL cover the period 1998-2014 and data for the other countries cover the period 1991-2014. The forecasts are made for data from 2014, using data from previous years to model learning. The forecasts were prepared in two variants:

- Variant A - for all 12 months 2014 simultaneously ($i^*$ is an index of December 2013, $m = 12$),
- Variant B - individually for 12 consecutive months of 2014 (12 models created for $i^*$ corresponding to: December 2013, January 2014, ..., and November 2014, $m = 1$).

The model parameters were selected using grid search ($a_{\min} = 0.1$, $\Delta = 0.05$, $l = 1, 2, ..., 38$, and $n = 3, 4, ..., 24$) in leave-one-out cross-validation procedure. Tables I-VIII show the results: optimal values of parameters, validation errors ($MAPE_{val}$) and test errors ($MAPE_{tst}$ for 2014). In most cases the lowest errors were achieved for patterns (4)-(8). The optimal x-pattern lengths vary between 9 and 24 depending on time series and pattern definition. Note that the optimal lengths are rarely equal to the annual cycle length, which is characteristic for these time series.

The test errors for individual months in variants A and B in Fig. 3 are shown. As can be seen from this figure, variant B which generates one step ahead forecasts, not always provides better results than variant A, in which the forecast horizon is 12 months. Due to significant contribution of the random component in the time series the errors for successive months are very varied. Different level of heteroscedasticity in time series for different countries and also occurrence of nonlinear trend in some of them decrease the accuracy of the model.

In Fig. 4 the construction of the forecasted y-pattern is shown. Gray lines in this figures are the training x- and y-patterns. A darker shade of gray indicates closer patterns to the query one $\mathbf{x}_{i^*}$. These patterns have higher weights, and consequently greater impact on the forecast. Patterns $\mathbf{x}_{i^*}$ and

TABLE I. Results for PL, Variant A

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 18 | 1.10 | 3.03 | 3.28 |
| (2)-(6) | 9 | 0.95 | 3.11 | 1.70 |
| (3)-(7) | 9 | 0.95 | 3.08 | 1.60 |
| (4)-(8) | 9 | 0.90 | 2.93 | 1.54 |

TABLE II. Results for DE, Variant A

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 21 | 0.40 | 3.29 | 3.64 |
| (2)-(6) | 10 | 1.15 | 3.18 | 3.68 |
| (3)-(7) | 10 | 1.15 | 3.18 | 3.93 |
| (4)-(8) | 10 | 1.20 | 3.24 | 3.11 |

TABLE III. Results for ES, Variant A

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 24 | 0.65 | 2.91 | 1.88 |
| (2)-(6) | 22 | 1.25 | 2.83 | 2.03 |
| (3)-(7) | 9 | 1.20 | 2.87 | 1.49 |
| (4)-(8) | 22 | 1.30 | 2.90 | 1.71 |

TABLE IV. Results for FR, Variant A

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 12 | 0.80 | 3.22 | 4.71 |
| (2)-(6) | 12 | 0.75 | 3.29 | 6.86 |
| (3)-(7) | 12 | 0.70 | 3.34 | 7.06 |
| (4)-(8) | 13 | 0.90 | 3.36 | 6.89 |

TABLE V. Results for PL, Variant B

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 10 | 0.95 | 2.40 | 2.12 |
| (2)-(6) | 10 | 0.85 | 2.24 | 1.94 |
| (3)-(7) | 10 | 0.85 | 2.22 | 1.93 |
| (4)-(8) | 10 | 0.70 | 2.00 | 1.30 |

TABLE VI. Results for DE, Variant B

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 9 | 0.60 | 2.48 | 5.87 |
| (2)-(6) | 10 | 0.80 | 2.54 | 3.20 |
| (3)-(7) | 12 | 0.90 | 2.53 | 3.39 |
| (4)-(8) | 9 | 0.80 | 2.51 | 2.78 |

TABLE VII. Results for ES, Variant B

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 24 | 0.60 | 2.86 | 1.40 |
| (2)-(6) | 14 | 1.40 | 2.32 | 2.08 |
| (3)-(7) | 13 | 1.30 | 2.43 | 1.72 |
| (4)-(8) | 13 | 1.20 | 2.41 | 1.16 |

TABLE VIII. Results for FR, Variant B

| Patterns | $n$ | $a_{opt}$ | $MAPE_{val}$ % | $MAPE_{tst}$ % |
|---|---|---|---|---|
| (1)-(5) | 22 | 1.15 | 3.07 | 4.29 |
| (2)-(6) | 14 | 0.75 | 2.93 | 2.96 |
| (3)-(7) | 14 | 0.85 | 2.97 | 2.84 |
| (4)-(8) | 13 | 0.85 | 2.86 | 4.11 |

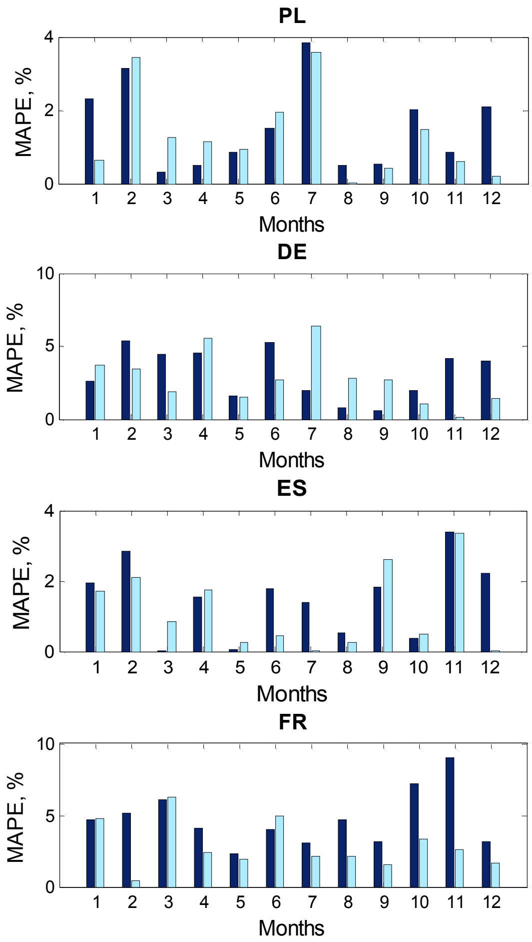


Fig. 3. Forecast errors for consecutive months of 2014. Variant A - dark bars, variant B - light bars.

$\mathbf{y}_{i*}$ are drawn with thick solid lines. The forecasted y-pattern is drawn with dotted line. Note, that the optimal input pattern lengths are different for different pattern definitions.

The proposed forecasting model is competitive with the models described in the literature. In [15] to forecast monthly electricity demand in Spain the neural network was applied. The forecasting problem was decomposed by splitting the time series into two new series: one representing its trend and the other describing a fluctuation around that trend. Six components of fluctuation was extracted by digital filtering. All components were forecasted using neural network and then were added to obtain the monthly demand forecast. The error obtained by this model was 3.15% and the error obtained by ARIMA which was used as a reference model was even larger. The N-WE model described in this work generates errors from 1.49 to 2.03 for ES data. For one step ahead forecasting the authors of [5] have obtained for Spain *MAPE* = 1.89% using neural networks and 4.19% using ARIMA. In our case when using N-WE the lower error was achieved: 1.16%.

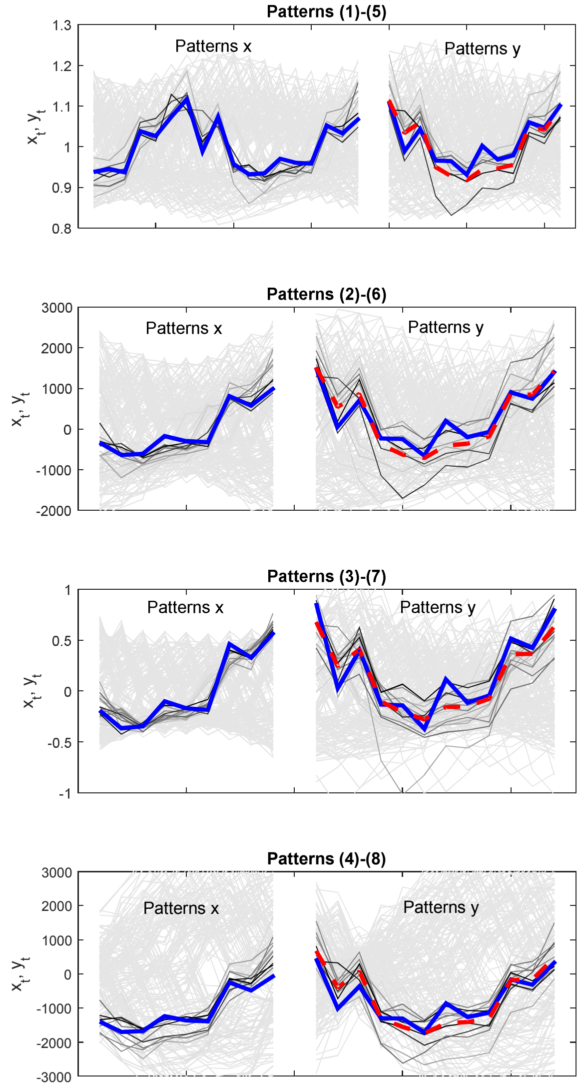


Fig. 4. Construction of the forecast pattern when using different pattern definitions for PL data, variant A. Patterns $\mathbf{x}_{i*}$ and $\mathbf{y}_{i*}$ are drawn with thick lines. The forecasted y-pattern is drawn with dotted line. Training patterns are drawn with thin gray lines.

## V. Conclusion

This work presents a new forecasting model based on Nadaraya-Watson estimator for monthly electricity demand. The model uses patterns representing time series fragments which are the key element of this approach. Due to patterns the problem of the time series nonstationarity can be reduced and the trend can be filtered out. The advantages of the model are its simple and understandable principle of operation and a small number of parameters to estimate.

The best performance of the model is observed for more regular time series with lower noise component and stable relationship between input and output patterns. The factors which decrease this stability are the nonlinear trend and variable in time variance of time series.

Results of simulations demonstrate high accuracy of the proposed model. The errors around 2% should be considered as a good result comparing to results of other models proposed in the literature. In the future work, we focus on the comparison our approaches based on patterns with other models used for medium-term electric energy demand forecasting, such as ARIMA, exponential smoothing and neural networks.

## REFERENCES


[1] P. Piotrowski, "Forecasting in power engineering in different time horizons," Warsaw University of Technology Publishing House, Warsaw 2013.

[2] M. Ghiassi , D. K. Zimbra, H. Saidane, "Medium term system load forecasting with a dynamic artificial neural network model," Electric Power Systems Research, vol. 76, pp. 302-316, 2006.

[3] M. Gavrilas, I. Ciutea, C. Tanasa, "Medium-term load forecasting with artificial neural network models," IEEE Conf. Elec. Disl. Pub. 482 (2001) 383.

[4] C.L. Hor, S.Watson, S. Majithia, "Analyzing the impact of weather variables on monthly rlectricity demand," IEEE Transactions On Power Systems, vol. 20 , pp. 2078-2085, 2005.

[5] E. González-Romera, M. Jaramillo-Morán, D. Carmona-Fernández, "Monthly electric energy demand forecasting based on trend extraction," IEEE Transactions On Power Systems, vol. 21, pp. 1946-1953, 2006.

[6] M.M. Elkateb. K. Solaiman. Y. AI-Turki, "A comparative study of medium-weather-dependent load forecasting using enhanced artificial fuzzy neural network and statistical techniques," Neurocomputing, vol. 23, pp. 3-13, 1998.

[7] E. Doveh, P. Feigin, L. Hyams, "Experience with FNN models for medium term power demand predictions," IEEE Transactions On Power Systems, vol. 14, no. 2, pp. 538-546, 1999.

[8] M.S. Kandil, S.M. EI-Debeiky, N.E. Hasanien, "Long-term load forecasting for fast developing utility using a knowledge-based expert system," IEEE Transactions On Power Systems, vol. 17, no. 2, pp. 491-496, 2002.

[9] W. Duch, "Similarity-based methods: A general framework for classification, approximation and association," Control and Cybernetics, vol. 29, pp. 937-968, 2000.

[10] G. Dudek, "Pattern similarity-based methods for short-term load forecasting – Part 1: Principles," Applied Soft Computing, vol. 37, pp. 277-287, 2015.

[11] G. Dudek, "Forecasting monthly electricity demand using k nearest neighbor method," Przegląd Elektrotechniczny (Electrical Review; in Polish; in print).

[12] G. Dudek, "Short-term load forecasting based on kernel conditional density estimation," Przegląd Elektrotechniczny (Electrical Review), vol. 86, no. 8, pp. 164-167, 2010.

[13] G. Dudek, "Pattern similarity-based methods for short-term load forecasting – Part 2: Models," Applied Soft Computing, vol. 36, pp. 422-441, 2015.

[14] D.W. Scott, "Multivariate density estimation: theory, practice, and visualization," Wiley, 1992.

[15] M. Jaramillo-Morán, E. González-Romera, D. Carmona-Fernández, "Monthly electric demand forecasting with neural filters," Electrical Power and Energy Systems, vol. 49, pp. 253-263, 2013.