Prediction of monthly electric energy consumption using pattern-based fuzzy nearest neighbour regression

Paweł Pełka[1,*], and *Grzegorz* Dudek[1]

[1]Czestochowa University of Technology, Department of Electrical Engineering, Al. Armii Krajowej 17, 42-200 Czestochowa, Poland

Abstract. Electricity demand forecasting is of important role in power system planning and operation. In this work, fuzzy nearest neighbour regression has been utilised to estimate monthly electricity demands. The forecasting model was based on the pre-processed energy consumption time series, where input and output variables were defined as patterns representing unified fragments of the time series. Relationships between inputs and outputs, which were simplified due to patterns, were modelled using nonparametric regression with weighting function defined as a fuzzy membership of learning points to the neighbourhood of a query point. In an experimental part of the work the model was evaluated using real-world data. The results are encouraging and show high performances of the model and its competitiveness compared to other forecasting models.

1 Introduction

Electric energy consumption forecasting is an essential issue in power system planning and operation. Medium-term forecasting is necessary for technical and operational purposes, such as: scheduling maintenance activities, planning of production levels and fuel purchases, and planning of network investments. From an economic viewpoint, energy consumption forecasts are fundamental for negotiating contracts between energy companies and concluding contracts with customers.

Fig. 1 shows a periodical time series representing monthly energy consumption for four European countries (the data from ENTSO-E repository – www.entsoe.eu). In this figure, seasonal variations and rising tendency can be observed, caused by the influence of the economic and technological development on the electric market. Seasonal variations reflect the annual cycle and are dependent on climatic factors, which are comparable during the same month of different years. Other factors affecting directly or indirectly the level of energy consumption are political decisions and economic policy. They can disturb general rising trend and monthly fluctuations. They include: the emergence of alternative energy sources and technologies, fluctuating economic inflation, violent change in energy prices, industrial development, and global warming issues [1], [2].

The time series of monthly electric energy demand presented in Fig. 1 differ depending on the power system size and economic development of the country. Note significant share of random component in the time series and larger amplitude of annual cycles for France than for other countries.

Two approaches have been developed for medium-term electric energy consumption forecasting [3]. The first one, called conditional modelling approach, focuses on economic analysis, management and long-term planning and forecasting of energy load and energy policies. It considers socioeconomic conditions which impact energy demands, such as economic indicators and electrical infrastructure measures. These additional inputs are introduced to the model together with historical load data and weather-related variables. Such a model can be found in [4]. It includes macroeconomic indicators, such as the consumer price index, the average salary earning and the currency exchange rate.

The second approach, called autonomous modelling approach, requires a smaller set of input information to forecast future electricity demand, primarily historical loads and weather factors. Because the economic factors are not taken into consideration, this approach is more suited for stable economies.
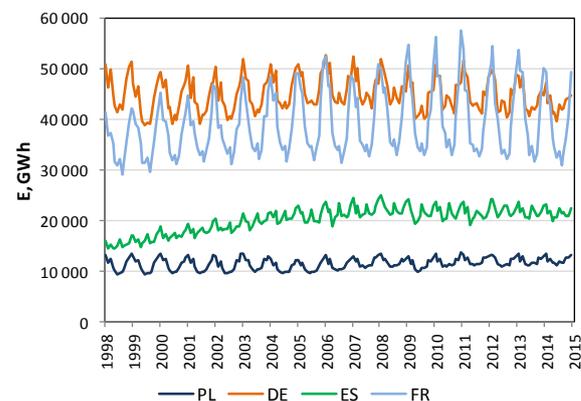
Fig. 1. Time series of monthly electric energy consumption.

---

* Corresponding author: pavelle50@gmail.com

Different forecasting models are used in this case such as classical autoregressive integrated moving average (ARIMA) and multiple linear regression [5], as well as computational intelligence methods, e.g. neural networks [6]. Examples of such models can be found in [7], where ARIMA, neural networks and neuro-fuzzy systems are employed to forecast future load demand based on various weather-related parameters and historical load profiles. Another example is a model presented in [8], where interval load forecasting is proposed using multi-output support vector regression. In addition, a memetic algorithm is used to select input variables among the variable candidates, which include time lagged loads and temperatures. In [9] neural network is used for forecasting load time series components extracted using digital filtering. Evolving fuzzy neural networks are proposed for monthly electricity demand forecasting in [4]. In this solution fuzzy neurons represent degree of importance of each input variable (loads, weather factors and daylight time). Different weights assigned to input variables lead to improved model accuracy and more precise prediction.

The forecasting model proposed in this work belongs to the latter category. It uses fuzzy nearest neighbour regression (FNNR), based on patterns of the time series fragments. An underlying assumption in this model is: 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 our earlier works, we proposed models from the same class of pattern similarity-based nonparametric regression models: the model based on *k*-nearest neighbours (*k*-NN) [11] and Nadaraya-Watson estimator [12]. The proposed FNNR allowed to consider similarity degree between shapes of the time series fragments using fuzzy set theory.

The remainder of the paper is organised as follows. In Section 2, a time series representation is described, using patterns of their fragments. In Section 3, forecasting model was defined, using fuzzy nearest neighbour regression. The model has been tested on real-world data in Section 4. Finally, the work was concluded in Section 5.

2 Patterns of time series fragments

In the first stage of the proposed approach load time series were pre-processed using methods presented in [10]. Input and output patterns were defined. The input pattern was an *n*-dimensional vector representing a time series fragment preceding the forecasted one. Let us denote the forecasted fragment by $Y_i = \{E_{i+1} \ E_{i+2} \ \ldots \ E_{i+m}\}$, and the preceding fragment by $X_i = \{E_{i-n+1} \ E_{i-n+2} \ \ldots \ E_i\}$, where E_k is the monthly energy consumption and k is the time index. An input pattern $\mathbf{x}_i = [x_{i,1} \ x_{i,2} \ \ldots \ x_{i,n}]^T$ represented the fragment X_i. Components of that vector were pre-processed points of the sequence X_i. For example [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 $t = 1, 2, \ldots, n$, $\overline{E}_i$ 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. Patterns (3) are composed of the differences between points and the mean sequence value. Pattern (4) is the normalised vector $[E_{i-n+1} \ E_{i-n+2} \ \ldots \ E_i]^T$. All patterns defined using (4) have the unity length, mean value equal to zero and the same variance.

The output pattern $\mathbf{y}_i = [y_{i,1} \ y_{i,2} \ \ldots \ y_{i,m}]^T$, representing the forecasted sequence Y_i, had the components defined similarly to the x-pattern components:

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

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

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

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

In the above formulas (5)-(8) $\overline{E}_i$ and D_i are determined from the sequence X_i, and not from the sequence Y_i. This is because the sequence Y_i is not known at the moment of forecasting. To determine the forecast of the monthly energy consumption E_{i+t} on the basis of the forecasted y-pattern generated by the forecasting model, transformed equations have been used (5)-(8). For example, in the case of (8) the forecasted energy consumption is calculated as follows:

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

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

3 Fuzzy nearest neighbour regression

The nearest neighbour estimate $m(\mathbf{x})$ is defined as the weighted average of the y-patterns in a varying neighbourhood of the query x-pattern. Typically, this neighbourhood is defined through the x-patterns which are among the *k* nearest neighbours of the query pattern

[11]. The value of k determines the number of training patterns from which the regression function is constructed and controls the degree of smoothing. The k-NN estimator gives the regression function, which is discontinuous. In the points where the set of the nearest neighbours changes, the jumps on the function graph are observed. To avoid this inconvenience, a fuzzy membership of the training points to the neighbourhood of the query point was introduced [13]. In this approach, each training point belongs to the query point neighbourhood with a degree depending on the distance between these points.

The regression function $m(\mathbf{x})$ has the nonparametric form:

$$m(\mathbf{x}) = \sum_{j=1}^{N} w(\mathbf{x}, \mathbf{x}_j) \mathbf{y}_j \qquad (10)$$

where the weighting function $w(\mathbf{x}, \mathbf{x}_j)$ is dependent on the similarity or distance between patterns $\mathbf{x}$ and $\mathbf{x}_j$. Usually it decreases monotonically with the distance. When using fuzzy approach, the weighting function has a form of the membership function, e.g. a Gaussian-type function:

$$\mu(\mathbf{x}, \mathbf{x}_j) = \exp\left(-\frac{d^2(\mathbf{x}, \mathbf{x}_j)}{\sigma^2} \right) \qquad (11)$$

where σ is a parameter controlling the width of the function, and $d(\mathbf{x}, \mathbf{x}_j)$ is a Euclidean distance between patterns $\mathbf{x}$ and $\mathbf{x}_j$.

An estimator (10) is a linear combination of vectors $\mathbf{y}_j$ weighted by the membership degree (11) which nonlinearly maps the distance $d(\mathbf{x}, \mathbf{x}_j)$. The greater the distance, the lower the weight. The width parameter σ decides about the bias-variance trade-off of the estimator. Too small σ value results in undersmoothing, whereas too large value results in over-smoothing. Thus, the selection of the width parameter is a key problem. In a training procedure the optimal value of σ is selected, as well as the optimal length of the input pattern n. These parameters are being searched using grid search method.

The training set contains pairs of patterns $(\mathbf{x}_i, \mathbf{y}_i)$, 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 generate the forecasts for months $i^*+1, i^*+2, ..., i^*+m$.

The forecasting procedure consists of four steps:
1. Pre-processing of load time series into x- and y-patterns.
2. Calculating the weights for the training x-patterns using membership function (11).
3. Calculating the forecasted y-pattern from (10).
4. Decoding the forecasted y-pattern using transformed equations (5)-(8) to get the monthly electricity demand for consecutive months: i*+1, i*+2, ..., i*+m.

4 Experimental study

In this section, the proposed FNNR method was applied to model and monthly electricity load demand was forecasted. Then results were compared with results of several reported statistical and machine learning methods for load demand forecasting. Data used in this research were taken from the publicly available ENTSO-E repository (www.entsoe.eu). They included monthly electricity demand for four European countries: Poland (PL), Germany (DE), Spain (ES) and France (FR). The time range of data was 1998-2015 for PL, and 1991-2015 for other countries. We constructed the forecasting models for 2015, using data from previous years to model learning. Two variants of forecasting were considered:
• Variant A – a model generated forecasts for all 12 months of 2015 (i^* was an index of December 2014, $m = 12$),
• Variant B – for each month of 2015 a separate model was created which generated one step ahead forecast (12 models created for i^* corresponding to: December 2014, January 2015, ..., and November 2015, $m = 1$).

The model parameters, σ and n, were selected using grid search in leave-one-out cross-validation procedure. The searched values of n were 3, 4, ..., 24 and $\sigma = a \cdot d_{med}$, where d_{med} was a median of distances between x-patterns in the training set, and $a = 0.02, 0.04, ..., 1$.

Tables 1-8 present optimal values of parameters and Mean Absolute Percentage Errors (MAPE) obtained with these parameter values: validation errors ($MAPE_{val}$) and test errors ($MAPE_{tst}$ for 2015). Accordingly to the tables, the selection of the best way of pattern definition seems to be difficult. Results depend on the time series features, such as a trend and level of random, irregular influences. The optimal x-pattern lengths vary between 8 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.

Fig. 2 demonstrates test errors for individual months in both variants, A and B. Note that variant B, which generates one step ahead forecasts, does not always provide better results than variant A, in which the forecast horizon is 12 months. Errors for successive months are very varied. This is caused by the significant contribution of the random component in data.

Examples of the forecasted y-pattern construction are presented in Fig. 3. Grey lines in these figures are the x- and y-patterns from the training set. A darker shade of grey indicates x-patterns which are closest to the query pattern and y-patterns paired with them. These patterns have higher value of the membership function (11), and consequently greater impact on the forecast. The query pattern and the true y-pattern paired with it are drawn with thick solid lines. The forecasted y-pattern is drawn with dotted line. Moreover, the optimal input pattern lengths are different for different pattern definitions (see Tables 1-8).

In Tables 9 and 10 results of comparative models are shown: ARIMA, exponential smoothing (ES) and Nadaraya-Watson estimator (N-WE) [12]. The proposed

Table 1. Results for PL, variant A.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	19	0.20	2.95	3.81
(2)-(6)	10	0.24	3.02	2.68
(3)-(7)	10	0.24	3.00	2.55
(4)-(8)	17	0.16	2.83	1.94

Table 2. Results for DE, variant A.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	8	0.16	3.24	3.39
(2)-(6)	12	0.26	3.09	2.18
(3)-(7)	12	0.26	3.11	2.30
(4)-(8)	10	0.24	3.16	3.07

Table 3. Results for ES, variant A.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	24	0.14	2.82	1.99
(2)-(6)	22	0.22	2.75	1.86
(3)-(7)	9	0.26	2.81	2.71
(4)-(8)	21	0.22	2.83	2.08

Table 4. Results for FR, variant A.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	13	0.16	3.21	3.95
(2)-(6)	12	0.14	3.28	3.56
(3)-(7)	12	0.14	3.31	3.65
(4)-(8)	12	0.16	3.36	3.53

Table 5. Results for PL, variant B.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	10	0.20	2.37	3.57
(2)-(6)	10	0.20	2.20	1.90
(3)-(7)	10	0.18	2.18	1.71
(4)-(8)	10	0.14	1.95	1.38

Table 6. Results for DE, variant B.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	8	0.14	2.46	2.43
(2)-(6)	8	0.18	2.51	2.46
(3)-(7)	8	0.18	2.50	2.40
(4)-(8)	10	0.16	2.49	2.68

Table 7. Results for ES, variant B.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	24	0.14	2.77	2.42
(2)-(6)	15	0.24	2.25	2.89
(3)-(7)	15	0.24	2.33	2.97
(4)-(8)	13	0.20	2.27	3.44

Table 8. Results for FR, variant B.

Patterns	*n*	*a*	*MAPE$_{val}$*	*MAPE$_{tst}$*
(1)-(5)	22	0.18	3.08	3.65
(2)-(6)	14	0.14	2.90	3.70
(3)-(7)	14	0.16	2.93	3.76
(4)-(8)	14	0.14	2.89	3.91

FNNR model belongs to the same group of nonparametric regression methods as N-WE; thus, results of both models are similar. When comparing errors of all models, it can be concluded that FNNR is competitive with other models, but it should be noted also that the classical ES model outperformed all other models in six of eight cases.
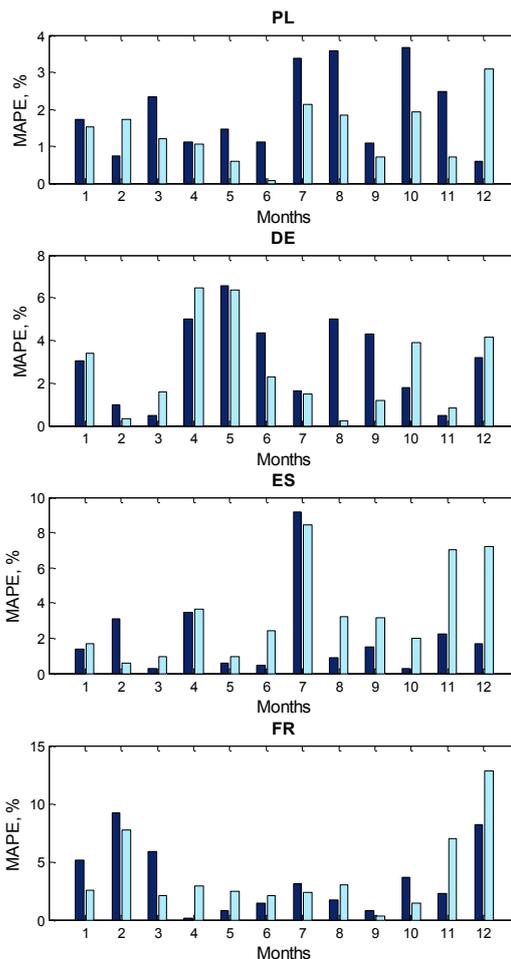
Fig. 2. Forecast errors for consecutive months of 2015. Variant A - dark bars, variant B - light bars.

Table 9. MAPE of the forecasting models, variant A.

Model	PL	DE	ES	FR
FNNR (4)-(8)	1.94	3.07	2.08	3.53
ARIMA	2.08	2.54	2.67	4.02
ES	1.92	2.32	2.17	3.02
N-WE	2.03	3.12	2.08	3.56

Table 10. MAPE of the forecasting models, variant B.

Model	PL	DE	ES	FR
FNNR (4)-(8)	1.38	2.68	3.44	3.91
ARIMA	2.02	2.56	2.18	3.91
ES	1.92	2.32	2.16	2.98
N-WE	1.35	2.72	3.42	3.99

5 Conclusion

This work proposes a practical methodology to forecast the monthly electric energy consumption using fuzzy nearest neighbour regression. This model is based on the assumption that the similarity of the input patterns implies the similarity of the output patterns paired with them. The patterns representing time series fragments are the key element of this approach. They unify data, reduce nonstationarity and filter out the trend.
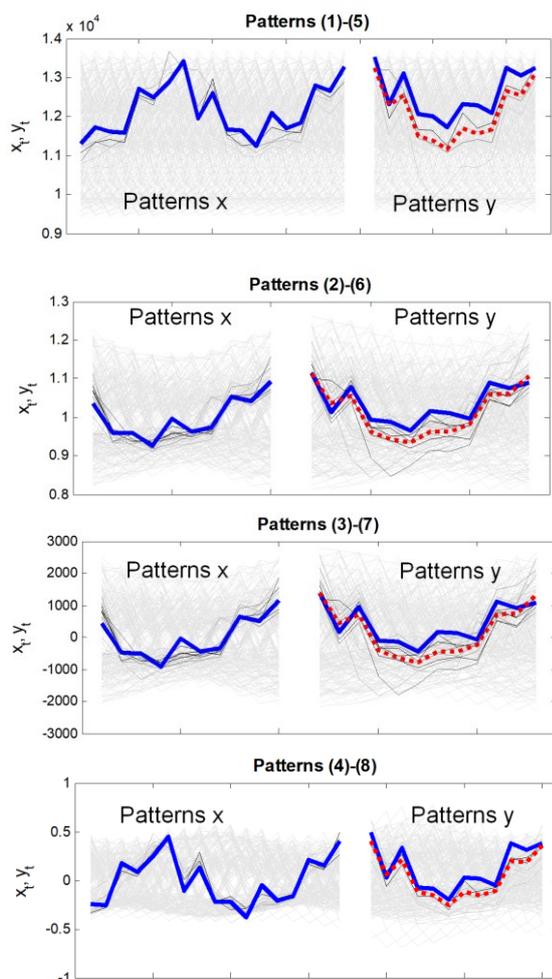
Fig. 3. Construction of the forecast pattern when using different pattern definitions for PL data, variant A.

The main advantages of the model are the simple and understandable principle of operation and only two parameters to estimate: the length of the input pattern and the width of the membership function. Models with fewer parameters have better generalisability and do not require complex learning procedures.

We demonstrate the effectiveness of our approach on real-world data. Comparing with commonly used methods, such as ARIMA and exponential smoothing, the proposed model results in similar errors on average. Better 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 heteroscedasticity of time series.

References

1. J.-F. Chen, S.-K. Lo, Q.-H. Do, Information, vol. 8, 31, (2017)

2. P.C. Chang, C.Y. Fan, J.J. Lin, Int. J. Electr. Power Energy Syst., vol. 33, pp. 17–27, (2011)

3. M. Ghiassi, D.K. Zimbra, H. Saidane, Electric Power Systems Research, vol. 76, pp. 302-316, (2006)

4. M. Gavrilas, I. Ciutea, C. Tanasa, IEEE Conf. Electricity Distribution (2001)

5. C.L. Hor, S.Watson, S. Majithia, IEEE Transactions On Power Systems, vol. 20 , pp. 2078-2085, (2005)

6. E. González-Romera, M. Jaramillo-Morán, D. Carmona-Fernández, IEEE Transactions On Power Systems, vol. 21, pp. 1946-1953, (2006)

7. M.M. Elkateb, K. Solaiman, Y. AI-Turki, Neurocomputing, vol. 23, pp. 3-13, (1998)

8. Z. Hu, Y. Bao, R. Chiong, T. Xiong, Energy 84, pp. 419-431 (2015)

9. M. Jaramillo-Morán, E. González-Romera, D. Carmona-Fernández, Electrical Power and Energy Systems, vol. 49, pp. 253-263 (2013).

10. G. Dudek, Applied Soft Computing, vol. 37, pp. 277-287, (2015)

11. G. Dudek, P. Pełka, Przegląd Elektrotechniczny (Electrical Review), r. 93(4), pp. 62-65, (2017) (in Polish)

12. G. Dudek, P. Pełka, Proc. 18th Conf. Electric Power Engineering (EPE'2017), pp. 1-6, (2017).

13. G. Dudek, Applied Soft Computing, vol. 36, pp. 422-441, (2015)