

Similarity analysis of the patterns of the monthly electric energy demand time series

Paweł Pelka^{1,*}

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

Abstract. The similarity analysis of the monthly electric energy demand time series sequence patterns are shown. The similarity-based forecasting models are allowed to be created because a strong relationship between input and output patterns exists. The chi-square test and the correlation tables were calculated for a few definitions of patterns.

1 Introduction

The monthly load time series for the national power systems are characterised by annual periodic variations. The annual variations are mainly related to the variable length of the day during the year and differences in air temperature depending on the season [1]. The analogies between time-series sequences with periodic variations are used successfully by the pattern similarity-based forecasting models. It is possible to deduce the course of the time series in any period based on its past behaviour [2].

The procedure for creating a forecasting model using the similarity of time-series sequences is as follows. First, the time series is divided into patterns, which usually cover a period of a few to several points. In the case studied in this paper, patterns are the points in a sequence of the monthly loads of the Polish power system (1998-2014) processed using specific functions. The pattern similarity-based forecasting methodology relies on the assumption that if the input patterns x_a and x_b are similar, then similar are also the output patterns y_a and y_b , which represent the time-series fragments following the fragments represented by the x_a and x_b patterns.

In literature medium-term load forecasting (MTLF) methods are categorized into two groups [4]. The first group consists of the conditional modelling approach and uses management, economic analysis and long term planning [5]. A MTLF model of this group is described in [6], where a few macroeconomic indicators, e.g. consumer price index are proceeded as inputs [5].

The second category consists of the autonomous modelling approach. This type requires a smaller set of inputs: firstly past loads and optionally e.g. weather variables. Models from this category are more suited for stable economies [5]. This group is represented by classical forecasting e.g. ARIMA or linear regression [7], and computational intelligence methods, like neural networks [8].

* Corresponding author: p.pelka@el.pcz.czest.pl

2 Pattern-based representation of time series

In case of forecasting m points in a series, the forecast fragment is denoted by $Y_i = \{E_{i+1} E_{i+2} \dots E_{i+m}\}$. Let us denote the preceding Y_i fragment of the series with a length of n as $X_i = \{E_{i-n+1} E_{i-n+2} \dots E_i\}$. The fragment X_i is represented by input pattern $\mathbf{x}_i = [x_{i,1} x_{i,2} \dots x_{i,n}]^T$. In this paper, the following 4 different definitions of the points in sequence X_i will apply [1]:

$$x_{i,t} = E_{i-n+t}, \quad (1)$$

$$x_{i,t} = \frac{E_{i-n+t}}{\bar{E}_i}, \quad (2)$$

$$x_{i,t} = E_{i-n+t} - \bar{E}_i, \quad (3)$$

$$x_{i,t} = \frac{E_{i-n+t} - \bar{E}_i}{D_i}, \quad (4)$$

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

In case of definition given in (1), the points in sequence X_i are copied without processing. The patterns (3) consist of differences between the values of the series and the mean value over the period covered by sequence X_i . Their quotient is defined using the equation (2). Pattern (4) is a normalised vector $[E_{i-n+1} E_{i-n+2} \dots E_i]^T$. The mean value of the components of this pattern is zero, and their variance is identical for all \mathbf{x}_i sequences.

The output pattern $\mathbf{y}_i = [y_{i,1} y_{i,2} \dots y_{i,n}]^T$ represents the fragment Y_i . The components of the output pattern are given as 4 different definitions [1]:

$$y_{i,t} = E_{i+t}, \quad (5)$$

$$y_{i,t} = \frac{E_{i+t}}{\bar{E}_i}, \quad (6)$$

$$y_{i,t} = E_{i+t} - \bar{E}_i, \quad (7)$$

$$y_{i,t} = \frac{E_{i+t} - \bar{E}_i}{D_i}. \quad (8)$$

The \bar{E}_i and D_i are determined from sequence X_i . They allow for the determination of the forecast of demand based on the forecast of \mathbf{y}_i returned by the forecasting model. For this purpose, the transformed formulas (5)-(8) are used. For example, if \mathbf{y} is defined using (8), the forecast of demand is determined using the equation [1]:

$$\hat{E}_{i+t} = \hat{y}_{i,t} D_i + \bar{E}_i. \quad (9)$$

The patterns \mathbf{x}_i and \mathbf{y}_i , which represent the preceding sequence and the forecast sequence, are combined in pairs $(\mathbf{x}_i, \mathbf{y}_i)$. A set of these pairs for a historical time series is used to create a forecasting model (for parameter estimation, teaching).

3 Similarity analysis of the patterns

In order to confirm the validity of the assumption that if input patterns are similar, then output patterns are similar too, the analysis of the dependence between input patterns \mathbf{x} and output patterns \mathbf{y} is carried out. Similarity is measured using the Euclidean metric.

The units of the statistical population to be analysed are pairs of patterns [1]

$$((\mathbf{x}_i, \mathbf{y}_i), (\mathbf{x}_j, \mathbf{y}_j)), \tag{10}$$

where: $i = 1, 2, \dots, N; j = 1, 2, \dots, N; i \neq j; N$ – the number of pairs of patterns. The pattern lengths equal to $N=12$ were adopted.

The size of the population is equal to $M = N(N-1)$. The feature D_x which determines the distance between two patterns \mathbf{x} and the feature D_y , which means the distance between a pair of patterns \mathbf{y} are defined. The distances between all pairs of patterns are the vector of pairs of realisations of the analysed random variables [2]

$$\begin{aligned} & [(d(\mathbf{x}_i, \mathbf{x}_j), d(\mathbf{y}_i, \mathbf{y}_j))] = \\ & = [(d(\mathbf{x}_1, \mathbf{x}_2), d(\mathbf{y}_1, \mathbf{y}_2)) (d(\mathbf{x}_1, \mathbf{x}_3), d(\mathbf{y}_1, \mathbf{y}_3)) \dots \\ & \dots (d(\mathbf{x}_n, \mathbf{x}_{n-1}), d(\mathbf{y}_n, \mathbf{y}_{n-1}))], \end{aligned} \tag{11}$$

where: $d(\mathbf{x}_i, \mathbf{x}_j)$ is the distance between a pair of patterns \mathbf{x}_i and \mathbf{x}_j ; $d(\mathbf{y}_i, \mathbf{y}_j)$ is the distance between a pair of patterns \mathbf{y}_i and \mathbf{y}_j .

Table 1. Correlation table of features D_x and D_y for patterns (4)-(8).

$D_x \backslash D_y$	0-0.78	0.78-0.97	0.97-1.15	1.15-1.34	1.34-1.50	1.50-1.64	1.64-1.77	1.77-1.94	1.94-3.06	Total
0-0.68	3021	518	59	22	0	0	0	0	0	3620
0.67-0.89	599	2230	626	121	41	3	0	0	0	3620
0.89-1.07	0	871	1983	578	125	54	9	0	0	3620
1.07-1.33	0	1	952	1949	489	116	59	47	7	3620
1.33-1.50	0	0	0	881	1710	702	223	52	52	3620
1.50-1.71	0	0	0	69	1089	1306	742	296	118	3620
1.71-1.81	0	0	0	0	166	1182	1245	727	300	3620
1.81-1.89	0	0	0	0	0	246	1118	1453	803	3620
1.89-1.99	0	0	0	0	0	11	224	1045	2340	3620
Total	3620	3620	3620	3620	3620	3620	3620	3620	3620	32580

To demonstrate stochastic dependence between random variables D_x and D_y , the null hypothesis (H_0) is formulated: the existence of differences in sizes of the population units in defined categories of values of features D_x and D_y is caused by the random nature of the sample [2]. The null hypothesis can be verified using the χ^2 test. For this purpose, a correlation table (Table 1) is created, illustrating the empirical joint distribution of features D_x and D_y . The critical value for 64 degrees of freedom (9 categories for features D_x and D_y) is 83.68. The χ^2 test values are within the critical area. This entitles to reject the null hypothesis and adopt the alternative hypothesis. The number of categories for feature D_x in the table is equal to g , and for feature D_y it is h . Quantiles of the order of $0, 1/g, 2/g, \dots, 1$ were adopted as category limits for D_x and quantiles of the order of $0, 1/h, 2/h, \dots, 1$ were adopted as category limits for D_y . High values of the Cramer contingency coefficient, V , and the Pearson correlation coefficient, ρ , prove strong relationships between D_x and D_y [2]:

$$V = \sqrt{\frac{\chi^2}{m \min(g, h)}}, \tag{12}$$

$$\rho = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n [d(\mathbf{x}_i, \mathbf{x}_j) - \bar{D}_x][d(\mathbf{y}_i, \mathbf{y}_j) - \bar{D}_y]}{(m-1)S_{D_x}S_{D_y}}, \tag{13}$$

where: \bar{D}_x, \bar{D}_y are mean values of D_x and D_y , and S_{D_x}, S_{D_y} are standard deviations of these features.

Table 2. ρ, V and χ^2 statistics values for the population determined by expression (10).

patterns	ρ	V	χ^2
(1)-(5)	0.8849	0.9096	242593.61
(2)-(6)	0.9005	0.9090	242303.21
(3)-(7)	0.8982	0.9087	242131.87
(4)-(8)	0.8543	0.9068	241104.13

Table 2 shows the χ^2, V and ρ statistics values for different pattern definitions. In the pattern similarity-based methods, the nearest neighbours of input pattern \mathbf{x} are found in the set of data and patterns \mathbf{y} paired with them are used to construct the forecast pattern. Next, it is demonstrated that there is a statistically significant relationship between the distance of pattern \mathbf{x}_i and its k^{th} nearest neighbour: $\mathbf{x}_{i^*,k}, k = 1, 2, \dots, K$ and the distance between patterns \mathbf{y}_i and $\mathbf{y}_{i^*,k}$ paired with them. In this work $K=5$ and it is the number of the last k^{th} nearest neighbour.

The units of the statistical population to be analysed are pairs of points [1]

$$((\mathbf{x}_i, \mathbf{y}_i), (\mathbf{x}_{i^*,k}, \mathbf{y}_{i^*,k})), \tag{14}$$

where: $i=1,2, \dots, n; k=1, 2, \dots, K; \mathbf{x}_{i^*,k}, \mathbf{y}_{i^*,k}$ – the k^{th} nearest neighbour of pattern \mathbf{x}_i and the pattern \mathbf{y}_i paired with it.

The population size is $n \cdot K$. The distances between all pairs of patterns determined by expression (14) are the vector of pairs of realisations of the analysed random variables [2]

$$\begin{aligned} & [(d(\mathbf{x}_i, \mathbf{x}_{i^*,k}), d(\mathbf{y}_i, \mathbf{y}_{i^*,k}))] = \\ & = [(d(\mathbf{x}_1, \mathbf{x}_{1^*,1}), d(\mathbf{y}_1, \mathbf{y}_{1^*,1})) (d(\mathbf{x}_1, \mathbf{x}_{1^*,2}), d(\mathbf{y}_1, \mathbf{y}_{1^*,2})) \dots \\ & \dots (d(\mathbf{x}_n, \mathbf{x}_{n^*,K}), d(\mathbf{y}_n, \mathbf{y}_{n^*,K}))], \end{aligned} \tag{15}$$

where: $d(\mathbf{x}_i, \mathbf{x}_{i^*,k})$ is the distance between a pair of patterns \mathbf{x}_i and its nearest k^{th} neighbour $\mathbf{x}_{i^*,k}$, and $d(\mathbf{y}_i, \mathbf{y}_{i^*,k})$ is the distance between a pair of patterns \mathbf{y}_i and $d(\mathbf{y}_i, \mathbf{y}_{i^*,k})$.

To demonstrate stochastic dependence between random variables D_x and D_y , the null hypothesis (H0), identical to that for the population determined by expression (10), is formulated.

Table 3. ρ, V and χ^2 statistics values for the population determined by expression (14), $K = 5$.

patterns	ρ	V	χ^2
(1)-(5)	0.2061	0.8908	6428.24
(2)-(6)	0.0401	0.8902	6418.72
(3)-(7)	0.0193	0.8901	6417.74
(4)-(8)	0.0468	0.8904	6421.60

Table 3 shows the χ^2 , V and ρ statistics values for different pattern definitions and $K=5$. Low values of ρ arise from a large dispersion of values among the intervals in the correlation table. They are not concentrated around diagonals. This results from the fact that all input patterns are collected and there are only 5 output patterns for each input pattern. Therefore, the maximum distance D_y is shorter and the dispersion of patterns y within these intervals is higher.

Associated with the nearest neighbour of input pattern x_i , the forecast pattern $y_{i^*, l}$ may not be the nearest neighbour of the forecast pattern y_i . The ratio of the distance between patterns y_i and $y_{i^*, k}$ (pattern associated with the k^{th} successive neighbour of pattern x_i) and the distance between patterns y_i and $y_{i', k}$ (pattern associated with the k^{th} successive neighbour of pattern y_i) is calculated [2]

$$w_d = \frac{d(y_i, y_{i^*, k})}{d(y_i, y_{i', k})} - 1. \quad (16)$$

The best accuracy of the forecasting model is achieved when patterns $y_{i', k}$, $y_{i^*, k}$ are identical, i.e. w_d takes the value 0. If these patterns are very similar, it means a strong link between data and good quality of the model. However, if they vary considerably, it means a weak link between random variables and poor quality of the model.

In Figure 1 relationship between distances D_x and D_y are shown. It is drawn in two variants: depending on y^* (red colour) and y' (blue colour). As can be seen in the Figure 1, the distances D_x and D_y are very similar.

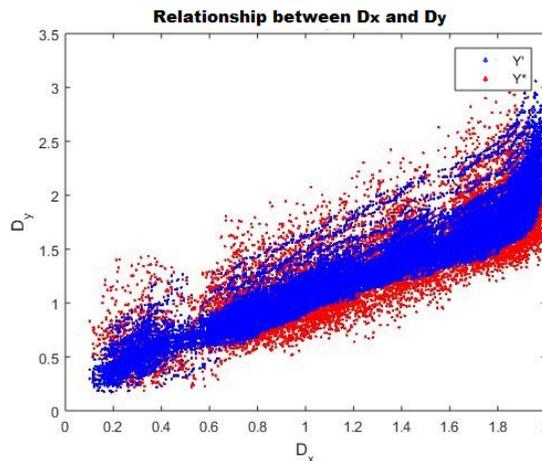


Fig. 1. Relationship between D_x and D_y for patterns determined by expressions (4)-(8).

In Figure 2, the monthly loads of the Polish power system (1998-2014) are visualised. For the analysed time series, these patterns are characterised by a high degree of similarity. Figure 3 shows mean values of w_d from successive neighbours of each pattern. It can be observed that these values are highest for the first neighbours. The values are lower than 0 in the vicinity of the 15th neighbour. It means that these are the nearer neighbours of pattern y . This suggests the use of models that consider several neighbours of pattern x to improve the accuracy of forecasting.

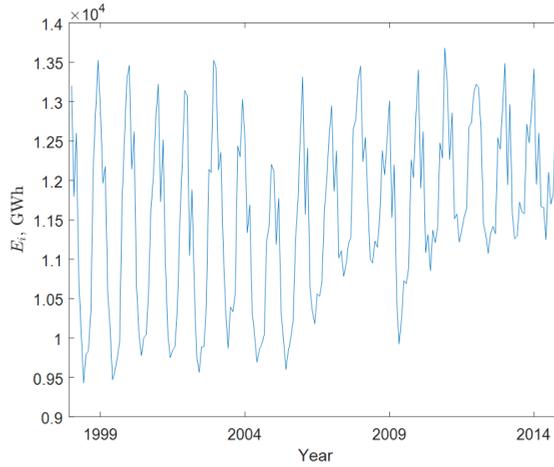


Fig. 2. Monthly loads of the Polish power system.

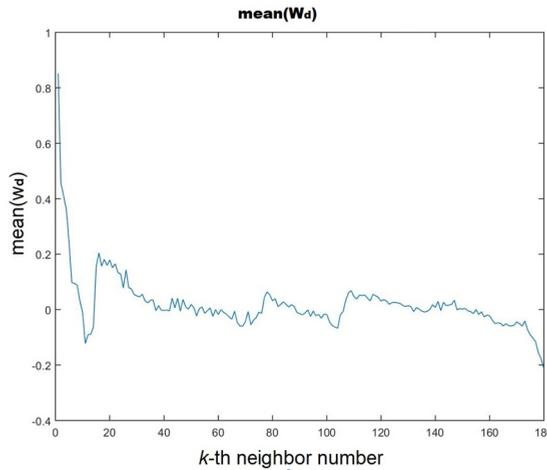


Fig. 3. Mean values of w_d depending on indices of the nearest neighbours of pattern for patterns determined by expression (4)-(8).

4 Decomposition of series

Decomposition of a time series can be described using the formula [3]:

$$X_t = f(s_t, m_t, z_t), \quad (17)$$

where $f()$ represents the function, s_t is a seasonal component, m_t is a long-term trend and z_t is a random disturbance.

There are two most popular decomposition models [3]:
 additive decomposition

$$X_t = s_t + m_t + z_t, \quad (18)$$

multiplicative decomposition

$$X_t = s_t \cdot m_t \cdot z_t. \quad (19)$$

If the variance around the trend or the size of seasonal variations does not change with the series level, an additive decomposition should be used.

In the event when the variance or the amplitude of seasonal variations is proportional to the series level, a multiplicative decomposition will be more suitable.

To isolate the components of a series (trend, seasonal and random variations), different methods are used. Two groups of methods can be distinguished:

- parametric methods, in which the model for regular components within a series is adopted (e.g. linear or square model). In this case, the estimators (estimates) need to be found for unknown parameters,

- non-parametric methods, in which the trend is not determined in the form of an analytical formula. The advantage of this group is a greater flexibility.

A popular trend determination method is the use of the moving average, which is defined as follows [3]:

$$m(t) = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j}, \quad (20)$$

where q is the parameter (an order of the moving average) used to control the data smoothness degree [3].

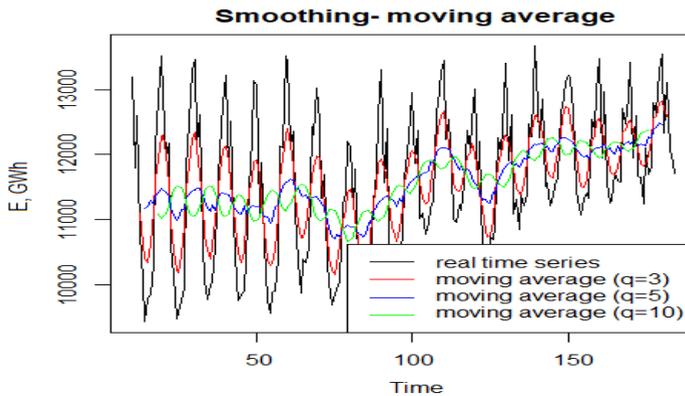


Fig. 4. Smoothing using the moving average for the Polish power system (PL) time series.

The graph illustrating smoothing using the moving average for the analysed time series is shown in Figure 4. Three types of moving average was shown. As can be seen, higher values of the q parameter give more smoothness degree of the moving average function.

If we want to assign different weights to the points of the series depending on the time distance from the moment t , we use the weighted moving average, which takes the following form [3]:

$$m(t) = \sum_{j=-q}^q w_j \cdot X_{t-j}, \quad (21)$$

where $w_{-q} + w_{-q+1} + \dots + w_{q-1} + w_q = 1$ and $w_j = w_{-j}$ (for symmetric moving average).

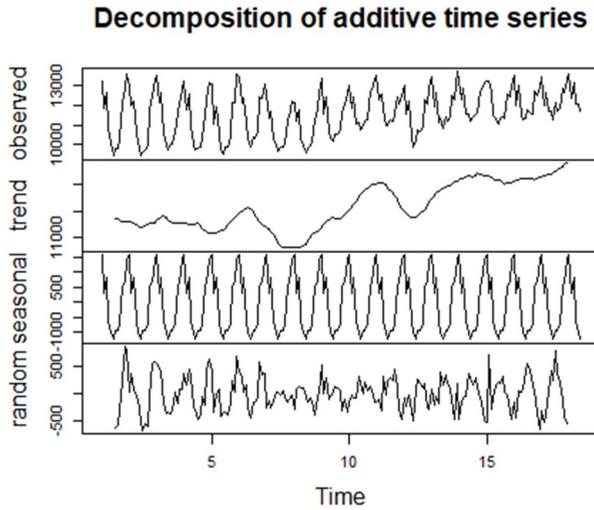


Fig. 5. Additive decomposition of time series PL.

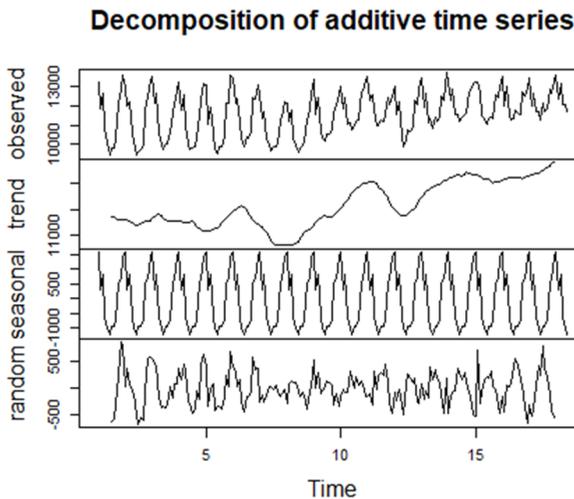


Fig. 6. Multiplicative decomposition of time series PL.

As a result of the use of one of the decomposition methods, the estimates are obtained for individual components of the series, i.e. trend and seasonality (\hat{s}_t and \hat{m}_t). It is possible to remove them and determine a random series of residuals (in case of additive decomposition)[3]

$$\hat{Z}_t = X_t - \hat{m}_t - \hat{s}_t. \quad (22)$$

If multiplicative relationships can be observed in the analysed series (the amplitude of seasonal variations or variance is proportional to the level of data), the conversion of (19) is used instead of (22)

$$\hat{Z}_t = X_t / (\hat{m}_t \cdot \hat{s}_t). \quad (23)$$

It is necessary to eliminate the trend and seasonality to match the stationary model [3].

The examples of graphs illustrating the effects of decomposition for the analysed series are shown in Figures 5 and 6.

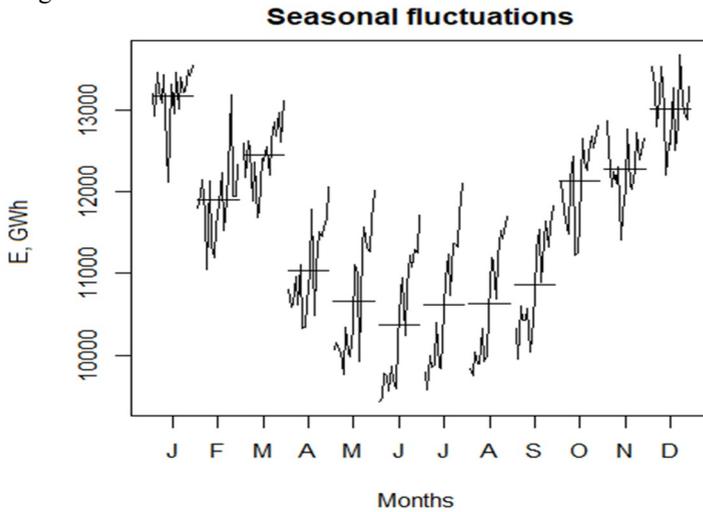


Fig. 7. Seasonal variations of the time series PL.

Figure 7 shows the monthly variations of values within the series in particular years, while Figure 8 shows the annual variations within the series. A considerable similarity of the values in individual months over the subsequent years can be observed in the graphs, taking into account the rising trend, which shows a strong upward tendency in the summer months. Strong annual cycles with a higher demand in the winter months are observed. It is worth emphasising the reduction in variance of the series in the final period where the annual cycles have a lower amplitude.

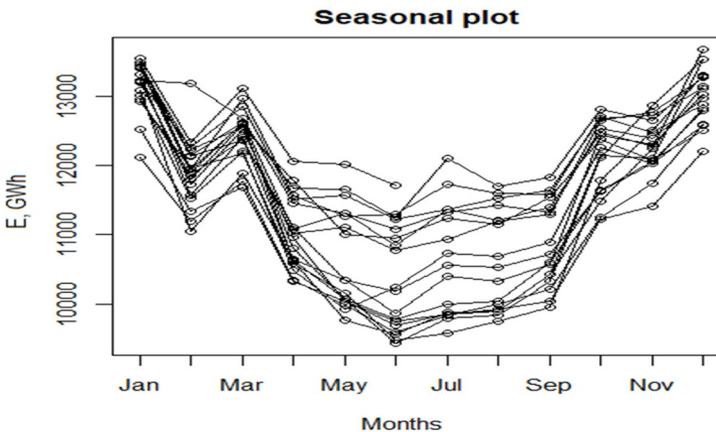


Fig. 8. Seasonal graph of the time series PL.

Summary

The analysis carried out in this paper for the Polish power system monthly load time series confirms that there is a strong relationship between the similarity of patterns x and the similarity of patterns y paired with them and that it does not result from the random nature of the sample. Such a conclusion for the specific time series justifies the sense of

constructing and applying the forecasting models using the similarity of patterns of seasonal cycles. In the further work, it is planned to examine the similarity of patterns for the monthly load time series for power systems in other European countries.

In a few articles [5, 9-12] authors demonstrate the effectiveness of a similarity-based approach on real-world data. Comparing with commonly used methods, e.g. ARIMA and exponential smoothing, the similarity-based models results achieve close errors on average [10]. For more regular time series with lower noise component and stable relationship between input and output patterns better performance of this kind of models is observed. The factors which decrease this stability are heteroscedasticity of time series and the nonlinear trend [10].

References

1. G. Dudek, *Systemy uczące się oparte na podobieństwie obrazów do prognozowania szeregów czasowych obciążeń elektroenergetycznych*, Akademicka Oficyna Wydawnicza EXIT, Warszawa (2012)
2. G. Dudek, *Przegląd Elektrotechniczny*, 85(3), 149-152 (2009)
3. A. Zagdański, A. Suchwałko: *Analiza i prognozowanie szeregów czasowych. Praktyczne wprowadzenie na podstawie środowiska R*, Wydawnictwo naukowe PWN SA, Warszawa (2016)
4. M. Ghiassi, D.K., Zimbra, H. Saidane, *Electric Power Systems Research*, 76, 302-316 (2006)
5. P. Pełka, G. Dudek, In: Borzemski L., Świątek J., Wilimowska Z. (Eds), *Advances in Intelligent Systems and Computing*, 853, 218-227 (2019)
6. M. Gavrilas, I. Ciutea, C. Tanasa, In: 16th International Conference and Exhibition on Electricity Distribution, IET, Amsterdam (2001).
7. C.L. Hor, S. Watson, S. Majithia, *IEEE Transactions On Power Systems*, 20, 2078-2085 (2005)
8. E. González-Romera, M. Jaramillo-Morán, D. Carmona-Fernández, *IEEE Transactions On Power Systems*, 21, 1946-1953 (2006).
9. G. Dudek, P. Pełka, *Przegląd Elektrotechniczny*, 93(4), 62-65 (2017)
10. P. Pełka, G. Dudek: *Proc. 2nd Int. Conf. Computational Methods in Engineering Science (CMES'17)*, ITM Web Conf. 15, 1-5, (2017)
11. G. Dudek, P. Pełka, *Proc. 18th Int. Scientific Conf. on Electric Power Engineering (EPE'17)*, 1-6 (2017)
12. P. Pełka, G. Dudek: In: Borzemski L., Świątek J., Wilimowska Z. (Eds), *Advances in Intelligent Systems and Computing*, 655, 38-47 (2018)