ORIGINAL PAPER

Electrical energy consumption forecasting using regression method considering temperature effect for distribution network

Gülsüm Yildiriz[1] · Ali Öztürk[1]

Abstract
Load profile coefficients (LPCs) represent the pattern of electricity usage daily and yearly for electrical energy consumers. It is important to determine the LPCs accurately and reliably, in order to minimize the imbalance costs in the Electricity Energy Market. Reliable methods and sufficient measurement data are required to make accurate forecasts. The local distribution company (TLDC) already calculates the profile coefficients by taking the average of the consumptions without meteorological measurements in Turkey. TLDC determines the LPC by receiving hourly consumption data directly from the consumers. In this paper, the mathematical forecasting models (MFMs) have been produced for determining LPC Duzce in Turkey using the multiple regression analysis method for the first time. Firstly, hourly electrical energy consumption and meteorological temperatures were measured in some predetermined residential subscribers. The MFMs have been produced by using the measured data, and then, LPCs have been determined by using the MFMs. The electrical energy consumptions have been estimated using the determined LPCs, and the estimation results have been compared with the measurement data. The MFMs have been subjected to suitability tests accepted in the literature, and the performances of the models have been verified. According to the results obtained, it has been seen that the MFMs can estimate loads with an accuracy of up to 96% depending on the future changing meteorological conditions, and it has been proposed as a quick and practical method for LPCs calculation. The paper shows that the produced MFMs provide obtaining satisfactory results for energy consumption forecasting for Duzce in Turkey.

Keywords Time series method · Load forecasting · Multiple regression analysis

1 Introduction

Electrical energy cannot be stored in the generation, transmission, and distribution process. Therefore, the generation must be enough to meet the consumption. A balance must be struck between the generation and the consumption to ensure effective energy management. The difference between the generation and the consumption causes an imbalance cost. The most effective way to minimize imbalance costs is to make an accurate load forecasting. Energy Stock Exchange and the Energy Market Operating Company (EPİAŞ) engage in activities related to the imbalance costs in Turkey. The main purpose and activity of EPİAŞ are to plan, establish,

develop and operate the energy markets included in the market operating licenses in order to meet their needs in the energy market in an efficient, transparent, reliable manner. Day-ahead market, balancing power market, calculation of receivables, and debts arising from energy imbalance are the factors that constitute the imbalance cost. The accuracy of the load estimates affects the market prices and imbalance costs of distribution companies [1].

Electric load data reveal nonlinear features depending on various influences such as climatic factors, social activities, and seasonal factors [2]. There are many methods developed to determine the profile coefficients symbolizing the load estimates [3]. The changes in weather conditions, calendar effects, economic indicators, etc., are the factors affecting the power demand [4, 5]. If these factors are taken into account while estimating consumption, more accurate results will be achieved. Load forecasting is called short-term, medium-term, or long-term load forecasting based on the forecast period. The estimation models from one hour to

✉ Gülsüm Yildiriz
gulsumyildiriz@duzce.edu.tr

1 Electric-Electronic Department, Faculty of Engineering, Duzce University, Düzce, Turkey

a week are called short-term load estimation [6, 7]. There are many papers on load estimations. The most commonly used approaches in estimations are Arima models from the Box— Jeckins method [8–10], exponential smoothing models [11], autoregressive (AR) models [11, 12], and time-sensitive approaches based on a single variable. Accurate estimation of electric load is one of the most important issues for developing countries. However, univariate prediction models are insufficient in estimating the electric load. Therefore, regression models that include weather, holiday, temperature, wind conditions, humidity, and similar variables in addition to the time variable have also been proposed in the forecasting model [13]. Vrablecová et al. [14] have investigated the suitability, advantages, and disadvantages of the online support vector regression method to short-term load estimation in the publicly available Irish CER dataset. Wang et al. [15] have proposed a combined probability density model for medium-term load estimation based on quantile regression by combining a real monthly data set from the USA with three separate models consisting of random forest regression, gradient boosting decision tree and support vector regression. Liu et al. [16] have proposed a probability density estimation method based on the Copula theory to obtain the relational diagram of electric charge and real-time price with real datasets from Singapore. Zhang et al. [17] have proposed an alternative quantile determination method in addition to a parallel and improved load quantile forecasting method and solved the reliability problem of the structure of direct estimation intervals. Hu et al. [18] made short-term load estimation by using weather factors and periodicity of short-term load in generalized regression neural network method. Dordonnat et al. [19] have proposed a semi-parametric regression model for point load estimation and a multivariate time series simulation model for temperature estimation. Florian has applied a quantile regression-based forecasting method, which takes into account both weekly and annual seasonality, applied in the framework of probabilistic load prediction. He has used temperature information only to offset the long-term trend component [20]. Wu et al. [21] have estimated daily and weekly data by exponential adjustment method and regression model taking into account seasonal and trend data. Yaslan et al. [22] have made hourly electric load estimations using a mixed method created by empirical mode decomposition and support vector regression algorithms. Chen et al. [23] have proposed a new support vector regression that estimates demand using ambient temperature information two hours before the forecast time. He and Zheng have proposed a probability density estimation method based on Yeo–Johnson transformation quantile regression using Gaussian core function with 1hour load measurement data measured in August 2014 for summer months, and December 2014 for winter months in Canada [24]. In the reference [25], the researchers has been developed a forecast model

that incorporated solar capacity to predict hourly load in southern California 24-h in advance. The forecast model has based on multiple linear regression, random forest, and gradient boosting methods. A regression tree method for short-term load estimation, which is integrated with the weighted average method, has been explained in [26].

ANN methods are also frequently used for short-term load estimation. There are many studies in the literature on the ANN method. Jha et al. [27] proposed LSTM (Long short-term memory) and random forest approach ANN models that make load estimation using meteorological data, historical loads and date type. Feng et al. [28] performed the medium-term and long-term power load forecast of an economic development region in Jiangsu Province using the method of the Elman neural network. Chen et al. [29] made household load estimates using a multiple cycles self-boosted neural network (MultiCycleNet) method that includes correlation analysis of electricity consumption patterns in multiple cycles. Velasco et al. [30] performed a performance analysis of ANN models for one hour forward electric load prediction using a representative dataset of historical electric load records of a specific geographic area served by an electric utility. Aly proposed six different load estimation models based on clustering techniques for Kalman filtering (KF), wavelet and artificial neural network (WNN and ANN) schemes, using different datasets from two different locations in Egypt and Canada [31]. Piazza et al. [32], using 7-year meteorological datasets and 4.5-year hourly load power demand datasets, made hourly electrical energy demand forecasts with autoregressive network-based artificial neural network (ANN). If the mathematical model is dependent on two variables and the data are taken in hourly periods for a year, the number of available data is quite high. Although a large number of data increase the estimation performance, it will increase the computational load in complex estimation models and will require the use of additional software tools. Therefore, plain and practical methods such as MRAM can be preferred instead of more complex methods.

In this paper, hourly load estimation was made with MRAM based on time and temperature. For the paper, MFMs that make hourly energy consumption estimations were obtained by using the temperature and residential electrical energy consumption data measured for a year in a certain region in Düzce province. The LPCs calculated with the energy consumption estimation values obtained with MFM were compared with the profile coefficients calculated by the local distribution company. The reliability of MFMs obtained with MRAM has been proven by subjecting them to R2, Adjusted R2, MAPE and RMSE tests. The back-and-forward ANN has been applied to randomly selected one-month data to measure the adequacy of the estimation performance of MFMs. Levenberg–Marquardt (LM) algorithm has been preferred because of the speed and stability,

it provides in ANN training. It has been observed that the estimation performances of the proposed simple model and the complex ANN model are similar. This result showed that the MRAM method is an adequate and reliable estimation method for the data set we have. Thus, residential electrical energy consumption estimation for Düzce was made in a plain and effective way. This article is important in terms of showing that with sufficient parameters and data sets, without the need for complex software, it is possible to reach correct predictions by only making correct mathematical models.

2 Load profile

In Turkey, profile coefficient refers to the hourly energy consumption and it is calculated using the previous year's data sets prepared by distribution companies. The calculated coefficients are used in settlement balancing in the next year. Profile coefficients are calculated separately for groups such as commercial, agricultural irrigation, industry, lighting, and housing. Profiles are classified as profiles of weekdays, weekends, and public holidays to express daily domestic consumption. Weekday profile coefficient consumption shows Monday, Tuesday, Wednesday, Thursday, and Friday. However, Monday's profile is different from other weekday profiles. The concept of weekend represents consumption on Saturday and Sunday. The holiday profile shows the consumption on official holidays, and the profile name is expressed as vacation. As a result of the classification made, the consumption data for the days in each day type are collected on the basis of the settlement period. Then, the total consumption data for each day type are calculated within each month and averaged on a daily basis. Hourly average consumption data of each day type are divided by the average of Monday type to form hourly profile coefficients [33].

Profile coefficients are announced by EPİAŞ on the official website at the end of each year [1]. Only hourly consumption is taken into account in the calculation of profile coefficients. As an example, residential load profile coefficients (RLPCs) are given in Table 1 for residential consumptions that can be used for the province of Düzce for April announced by EPİAŞ. The hourly graphical representation of these profile coefficients is shown in Fig. 1. Since the day-based curves of the profile coefficients given in Fig. 1 show similar results with each other, it does not provide the opportunity to make a detailed examination. Meteorological data are not taken into account in the calculation of RLPCs announced by EPİAŞ. These RLPCs are used after one year. The fact that this situation will negatively affect the imbalance cost calculation in the settlement is not taken into account. However, four climates can be seen in Turkey. Temperature changes during the day vary regionally. Hourly temperature changes play an effective role in electricity consumption. Correlation

coefficients have been calculated to see the relationship between hour-consumption and temperature-consumption for April. The coefficients are given in Table 2. In this paper, detailed mathematical models have been developed to observe the temperature-dependent change of consumption.

3 Mathematical models

3.1 Time series method

The regression method, one of the time series methods, proposes to plot the dependent and independent variables. If a meaningful mathematical model can be derived for the plotted graph, this model will form an estimation equation for the time series. The mathematical model of the analyzed graph can be in polynomial, sinusoidal and exponential forms [34]. The regression equation is chosen considering which mathematical model the graphical distribution of the time series resembles. If the correct mathematical model can be obtained, the model will give close results for unknown times. The closeness of the results to the measurement values is proportional to the success rate of the model. Appropriate residential subscribers, which are thought to represent the whole province in Duzce, have been determined. Measurements taken from these subscribers have been used as a data set in this paper. Temperature data have been measured synchronously in addition to hourly consumptions. The changing electrical energy consumptions depending on the time and measured temperature have been examined. The change should be examined graphically in order to obtain the right model.

In Table 3, hourly energy consumption and air temperature data for Mondays in April of Düzce province residents are given. Graphical representation of these data is shown in Fig. 2. When looking at the change in the graph given in Fig. 2, it is clearly seen that the correct mathematical model should be a polynomial model. Consequently, polynomial multiple analysis regression method (PMRAM) has been used in the paper.

The estimation equation obtained from the multiple regression model using the measurements is given in Eq. (1) [35]. Y_i, X_{1i} and X_{2i}, a_{nm} and e represent measurement data (energy consumption), the arguments (hour and temperature), the coefficients of the estimation equation, and the estimation equation error value, respectively. Increasing the performance of the estimation model is possible by minimizing the estimation error. The sum of squares of the errors is shown by E in Eg. (2). $\overline{Y}_i$ gives the forecasted value. If the derivative of Eq. (2) is taken according to each of the estimation coefficients and equalized to 0, the matrix given in Eq. (3) is obtained [36, 37]. The estimation coefficients

Table 1 RLPCs of electrical energy consumptions measured in April

Day/hour	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	23 Holiday
00	0.042865	0.042819	0.041585	0.040947	0.040809	0.042786	0.045734	0.043476
01	0.036022	0.036433	0.035282	0.034618	0.034637	0.036497	0.039191	0.03692
02	0.032093	0.032249	0.029819	0.030491	0.029604	0.032277	0.034238	0.032733
03	0.029452	0.029963	0.028957	0.029531	0.029184	0.029669	0.030613	0.029538
04	0.030021	0.030351	0.029182	0.029221	0.028843	0.029583	0.030933	0.030299
05	0.030818	0.030891	0.026362	0.02993	0.02961	0.029999	0.031413	0.030367
06	0.028771	0.029173	0.027821	0.02826	0.028322	0.027378	0.027709	0.025633
07	0.030717	0.030688	0.030379	0.029915	0.030385	0.027044	0.02621	0.027154
08	0.034687	0.034965	0.034394	0.033874	0.034472	0.031083	0.029497	0.032693
09	0.039733	0.039338	0.038007	0.038243	0.038761	0.037559	0.035688	0.037842
10	0.04339	0.042036	0.041099	0.040999	0.041866	0.043152	0.041951	0.040734
11	0.044231	0.043322	0.042617	0.042344	0.04339	0.04647	0.046592	0.042691
12	0.045734	0.043228	0.042673	0.042579	0.043099	0.047325	0.047995	0.043368
13	0.044519	0.041657	0.041534	0.041116	0.040902	0.045807	0.046334	0.042879
14	0.043035	0.040179	0.039794	0.039548	0.03992	0.04492	0.043585	0.041343
15	0.041979	0.039345	0.039259	0.038388	0.039297	0.042456	0.041684	0.03997
16	0.041541	0.039077	0.03958	0.038509	0.039147	0.04245	0.040698	0.039762
17	0.042709	0.039813	0.040517	0.039682	0.040261	0.043779	0.041736	0.040155
18	0.045745	0.042229	0.043771	0.042183	0.042594	0.046297	0.044326	0.041468
19	0.050939	0.047891	0.049372	0.048071	0.048649	0.050933	0.049639	0.045573
20	0.058671	0.057029	0.057344	0.056031	0.056568	0.057461	0.057723	0.054294
21	0.05696	0.055579	0.055649	0.054087	0.054489	0.055115	0.056433	0.052838
22	0.055096	0.054265	0.053622	0.052595	0.052509	0.054135	0.055165	0.051086
23	0.05027	0.04993	0.049602	0.048248	0.049203	0.051797	0.051015	0.047107

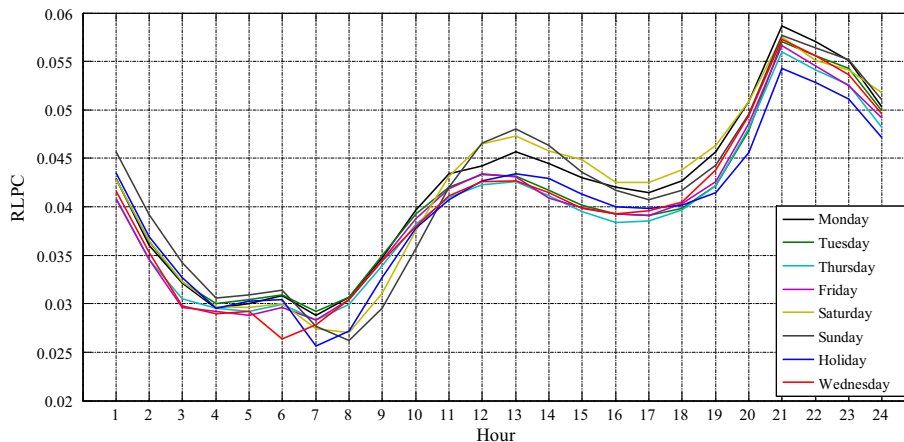**Fig. 1** Hourly change of day-type RLPCs for April

Table 2 Correlation coefficients of electrical energy consumption measured in April

Data	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Hour	0.2826	0.2447	0.3303	0.6122	0.4524	0.3292
Temperature	0.3265	0.3863	0.5416	0.4746	0.6397	0.3180

Table 3 Hourly temperature and electrical energy consumption measured on Mondays in April

| H	T (°C)	P (kWh)	H	T (°C)	P (kWh)	H	T (°C)	P (kWh)	H	T (°C)	P (kWh)
1	0.81	54.71	1	9.31	52.99	1	14.58	52.27	1	7.85	54.05
2	1.04	53.83	2	8.91	52.19	2	12.06	52.09	2	7.95	53.34
3	0.61	53.99	3	8.98	52.53	3	11.11	51.94	3	5.94	53.16
4	0.52	53.80	4	8.89	52.42	4	10.78	51.79	4	4.83	53.03
5	− 0.09	55.40	5	8.96	54.25	5	9.15	53.84	5	4.06	53.84
6	− 0.69	61.44	6	9.33	59.34	6	10.85	58.26	6	5.00	58.84
7	− 0.85	75.32	7	9.06	71.74	7	12.64	68.62	7	5.45	72.80
8	0.85	80.21	8	9.39	75.53	8	15.17	76.10	8	8.30	77.67
9	2.47	81.15	9	9.31	76.09	9	16.36	79.00	9	10.15	77.98
10	4.06	81.22	10	8.96	72.00	10	17.85	78.71	10	11.02	77.56
11	5.33	83.27	11	9.74	73.26	11	18.54	80.56	11	11.11	78.86
12	7.34	80.63	12	11.72	71.35	12	20.17	76.97	12	11.28	75.87
13	8.04	79.31	13	12.21	74.83	13	20.64	73.97	13	12.15	73.04
14	8.54	80.81	14	15.26	79.20	14	16.98	76.01	14	11.76	74.26
15	9.83	79.03	15	16.13	85.41	15	15.61	75.51	15	11.96	72.93
16	8.89	78.98	16	12.94	67.99	16	15.52	74.78	16	11.56	72.20
17	8.80	79.45	17	13.89	69.01	17	15.00	73.96	17	10.52	72.29
18	8.20	73.86	18	13.72	64.64	18	14.67	67.00	18	9.85	66.62
19	7.69	72.04	19	12.53	65.17	19	12.45	65.52	19	8.89	67.53
20	7.06	74.89	20	10.85	67.84	20	10.52	67.58	20	8.20	69.67
21	6.11	73.90	21	9.48	65.77	21	9.91	69.35	21	7.85	72.34
22	6.11	68.07	22	9.74	60.50	22	9.18	64.30	22	8.20	67.19
23	5.94	62.30	23	8.44	55.36	23	7.81	59.73	23	7.53	61.01
24	4.74	57.48	24	8.99	51.42	24	6.72	55.34	24	5.94	56.78
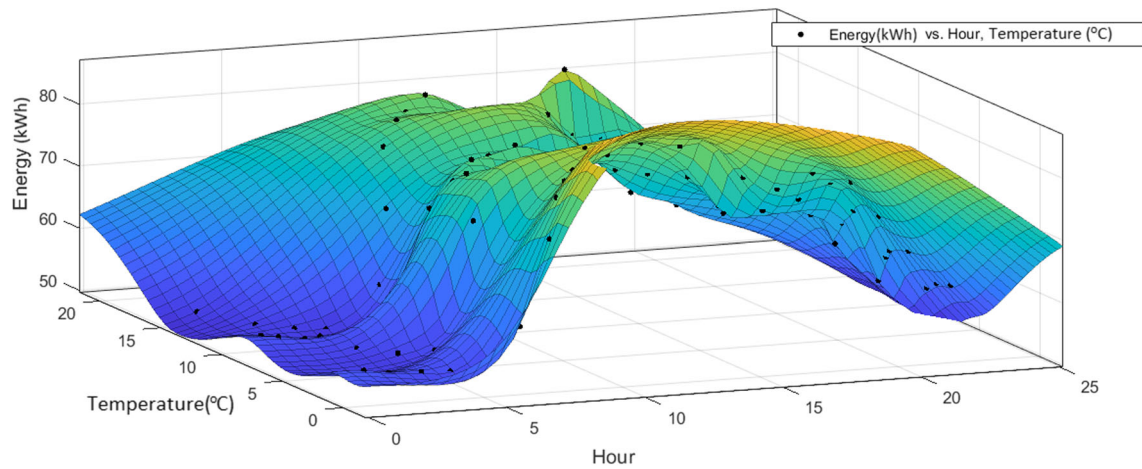
Fig. 2 3D graph of hourly temperature (°C) and consumed electrical energy (kWh) on Mondays in April

can be calculated using this matrix. In this paper, the mathematical models for detailed analysis have been solved by establishing a 5th order equation. In Eq. (4), the 5th-degree dependent variable model is given depending on two independent variables. A higher-order matrix could have been created, but it was observed that the coefficients of 5th-order matrices approached very close to 0. Therefore, it is predicted that the coefficients of higher-order matrices will be almost 0. In addition, an increase in the amount of coefficient to be calculated will increase the computational load significantly. Therefore, it was decided to create a estimation model up to the 5th-order in the paper.

$$Y_i = a_{00} + a_{10}X_{1i} + a_{01}X_{2i} + \cdots + e \tag{1}$$

$$E = \sum_{i=1}^{n} (Y_i - \overline{Y_i})^2 \tag{2}$$

$$\begin{bmatrix} n & \sum X_{1i} & \cdots & \sum X_{1i}X_{2i}^4 & \sum X_{2i}^5 \\ \sum X_{1i} & \sum X_{1i}^2 & \cdots & \sum X_{1i}^2 X_{2i}^4 & \sum X_{1i}X_{2i}^5 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum X_{1i}X_{2i}^4 & \sum X_{1i}^2 X_{2i}^4 & \cdots & \sum X_{1i}^2 X_{2i}^8 & \sum X_{1i}X_{2i}^9 \\ \sum X_{2i}^5 & \sum X_{1i}X_{2i}^5 & \cdots & \sum X_{1i}X_{2i}^9 & \sum X_{2i}^{10} \end{bmatrix}$$
$$\begin{bmatrix} a_{00} \\ a_{10} \\ \vdots \\ a_{14} \\ a_{05} \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{1i}Y_i \\ \vdots \\ \sum X_{1i}X_{2i}^4 Y_i \\ \sum X_{2i}^5 Y_i \end{bmatrix} \tag{3}$$

$$\begin{aligned} \overline{Y_i} = {}& a_{00} + a_{10}X_{1i} + a_{01}X_{2i} + a_{20}X_{1i}^2 + a_{11}X_{1i}X_{2i} \\ & + a_{02}X_{2i}^2 + a_{30}X_{1i}^3 + a_{21}X_{1i}^2 X_{2i} + a_{12}X_{1i}X_{2i}^2 \\ & + a_{03}X_{2i}^3 + a_{40}X_{1i}^4 + a_{31}X_{1i}^3 X_{2i} + a_{22}X_{1i}^2 X_{2i}^2 \\ & + a_{13}X_{1i}X_{2i}^3 + a_{04}X_{2i}^4 + a_{50}X_{1i}^5 + a_{41}X_{1i}^4 X_{2i} \\ & + a_{32}X_{1i}^3 X_{2i}^2 + a_{23}X_{1i}^2 X_{2i}^3 + a_{14}X_{1i}X_{2i}^4 + a_{05}X_{2i}^5 \end{aligned} \tag{4}$$

3.2 Fitness tests

The estimation methods are subjected to fitness tests to measure the accuracy and acceptability of the estimate. Thus, the estimation performance of the methods is measured. R^2 (The determination coefficient), Adj. R^2 (Adjusted R^2), RMSE (Root Mean Square Error), and MAPE (Mean Absolute Percentage Error) are commonly used in load forecasting [38–43]. R^2 is expressed by Eq. (5). Here, Y_i expresses the measurements, and these data are added for each measurement by taking the square of the difference between the measured data and the value of the estimate. Adjusted R^2 is shown in Eq. (6). The RMSE showing the square root of the mean square of the errors is given by Eq. (7). Here, n
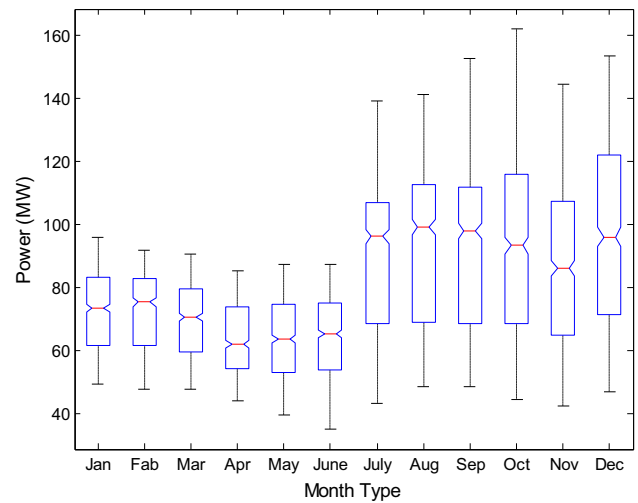

Fig. 3 Monthly box plot of data

is the total number of measurements. Another test of conformity, MAPE, gives the average absolute percentage error value and is obtained as in Eq. (8).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \overline{Y_i})^2}{\sum_{i=1}^{t} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}} \tag{5}$$

$$\text{Adj.} R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \overline{Y_i})^2}{\sum_{i=1}^{t} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}} \frac{n-1}{n-k-1} \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{\sum (Y_i - \overline{Y_i})^2}{n}} \tag{7}$$

$$\text{MAPE} = \left(\frac{100}{n} \sum_{1}^{n} \left| \frac{Y_i - \overline{Y_i}}{Y_i} \right| \right) \tag{8}$$

The determination coefficient is a value between 0 and 1. The closeness of the determination coefficient to 1 shows the forecasting performance. It is known that the smaller the RMSE test value, the better the prediction performance [44, 45]. In order to say that its performance is high in MAPE, Eq. (7) is expected to have a value below 10% [46]. The proximity of RMSE and MAPE values to 0 means an increase in forecasting performance.

4 Solution of the problem

The mathematical models developed in Chapter 3 have been solved using one of the time series methods, MRAM. The analyzed data have been divided into similar groups to increase the sensitivity of the observation. Firstly, the data have been into monthly groups. The box chart divided into months is given in Fig. 3. It is seen that the minimum,
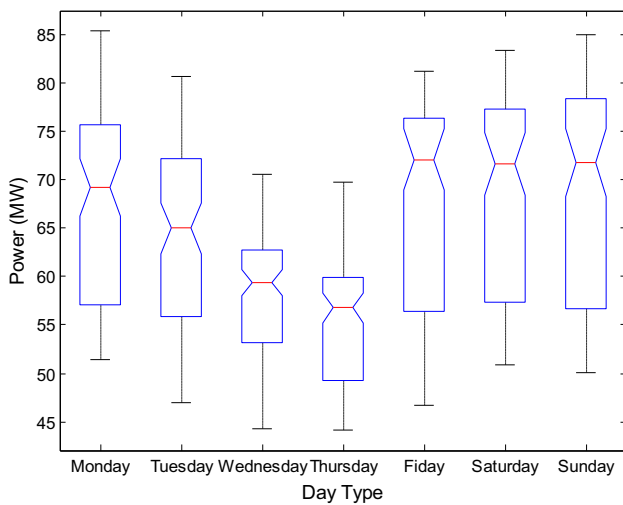
Fig. 4 Daily boxplot of April data

maximum, median, first quartile, and third quartile values of the data are different from each other in monthly periods. Figure 3 shows that there are significant differences in electricity consumption in summer and winter months as expected due to seasonal temperature differences in Düzce. Each month has been divided into daily periods in order to create more similar groups and to be able to examine them in more detail. As an example, the box graph showing the data of April divided into daily periods is shown in Fig. 4. It has been observed that the minimum, maximum, median, first quartile, and third quartile values of the weekday data are different from each other. Due to the necessities of social life, electricity consumption on weekdays and on weekends and near weekends shows different characteristics. But Saturday and Sunday (weekend) data have relatively similar characteristic values.

All data are grouped into months and 6 days (Monday, Tuesday, Wednesday, Thursday, Friday, and Weekend) per month in order to develop a more precise forecast model in modeling as a result of the box chart observation. Thus, 72 MFMs have been created. While calculating the coefficients of mathematical models, instead of the independent variables (X_{1i} and X_{2i}) of the models, time of day information (H_i) and measured temperature (T_i) data, energy consumption (P_i) data instead of the dependent variable (Y_i) have been used. Estimation model coefficients have been calculated for each mathematical model by analyzing the obtained matrices. As a result, the general formula to which the calculated coefficients will be applied is obtained as in Eq. (9). Here, $\overline{P}_i$, H_i, T_i, and a_{mn} indicate energy consumption forecasting, hour information, temperature measurement clock and estimation coefficients, respectively. 6-day estimation coefficients for April are given as an example in Table 4. The a_{mn} values given in Table 4 show the coefficients of the estimation model calculated day by day for April. When the measured values

were examined, it has been deemed appropriate to select daily groups to produce a mathematical model. With the obtained models, the amount of electrical energy consumed depending on time and air temperature has been estimated. Then, R^2, Adj. R^2, MAPE, and RMSE values have been calculated for each of the models. So, the reliability of the mathematical models obtained has been subjected to four different suitability tests.

$$
\begin{aligned}
\overline{P}_i = {} & a_{00} + a_{10}H_i + a_{01}T_i + a_{20}H_i^2 + a_{11}H_iT_i + a_{02}H_i^2 \\
& + a_{30}H_i^3 + a_{21}H_i^2T_i + a_{12}H_iT_i^2 + a_{03}T_i^3 + a_{40}H_i^4 \\
& + a_{31}H_i^3T_i + a_{22}H_i^2T_i^2 + a_{13}H_iT_i^3 + a_{04}T_i^4 \\
& + a_{50}H_i^5 + a_{41}H_i^4T_i + a_{32}H_i^3T_i^2 + a_{23}H_i^2T_i^3 \\
& + a_{14}H_iT_i^4 + a_{05}T_i^5
\end{aligned}
\tag{9}
$$

5 Results and sensitivity analysis

The data were classified into months and 6 days each month in order to create more precise estimation models. Hence, 72 residential consumption estimation models base on time and temperature were created. Monday in April has been chosen as an example from 72 mathematical models developed in the paper. The consumption estimation diagram of the developed model according to the measured temperature and time is given in Fig. 5. Similar variations according to the measured temperature and time can be obtained for the other 71 forecasting models. Currently, local distribution companies calculate profile coefficients by taking the average of their consumption. The profile coefficients for each hour are calculated by the ratio of the total hourly energy value of Monday. Meteorological data are not taken into account in these calculations. The RLPCs calculated by the local distribution company and the RLPCs obtained by the MRAM are given in Table 5 comparatively. To support the estimation accuracy of the MRAM, estimation was made using ANN with the data and RLPCs were calculated for these estimations. Thus, it has been seen that the MRAM estimation results and the ANN estimation results are close to each other. Consumption estimates were made in cases of ± 2 °C changes in the temperature value in order to examine the effect of temperature change. Accordingly, it has been observed that even a small temperature change causes a visible change in the profile coefficients.

The performance of another forecasting model (ANN) is also examined to show the effect of the MRAM. For this, measurement data on Friday in April were taken into account. The graphical change of hourly forecasting RLPCs on this day is shown in Fig. 6. Similar to the considered Monday model, changes close to the measurement results have

Table 4 Coefficients of the mathematical day models of April

Day/coefficient	Monday	Tuesday	Wednesday	Thursday	Friday	Weekend
a00	74.21	128.1	53.99	56.79	60.03	69.22
a10	− 18.61	− 43.99	− 6.811	− 8.743	− 14.2	− 17.78
a01	1.007	− 22.69	7.721	− 0.08446	4.399	1.069
a20	4.637	8.476	2.54	2.421	5.177	5.082
a11	0.428	8.132	− 2.792	− 0.8494	− 1.537	0.6292
a02	− 0.4467	2.534	− 1.06	0.267	− 1.682	− 0.815
a30	− 0.3647	− 0.6009	− 0.3006	− 0.2569	− 0.5752	− 0.4959
a21	− 0.183	− 0.9671	0.3202	0.2262	0.3131	− 0.03035
a12	0.1248	− 0.6346	0.3338	− 0.09149	− 0.04821	− 0.06637
a03	− 0.0139	− 0.09249	− 0.03708	0.02337	0.3481	0.1111
a40	0.01236	0.01913	0.01363	0.01154	0.0232	0.02059
a31	0.009245	0.03553	− 0.008604	− 0.01044	0.001187	− 0.00067
a22	0.002335	0.05679	− 0.03259	− 0.01216	− 0.0412	0.005719
a13	− 0.01207	0.01432	− 0.005878	0.01791	0.02724	− 0.00135
a04	0.005863	0	0.009282	− 0.00713	− 0.02953	− 0.00435
a50	− 0.00016	− 0.00024	− 0.0002122	− 0.00018	− 0.0003	− 0.00031
a41	− 0.00015	− 0.00036	3.03E-05	0.000111	− 0.00044	3.73E-05
a32	− 1.65E-05	− 0.00124	0.0005975	0.000559	0.001716	− 0.00011
a23	− 0.00012	− 0.00078	0.0006744	− 0.00045	− 0.00149	− 9.44E-06
a14	0.000431	0	− 0.0002414	− 7.01E-05	0.00126	3.14E-05
a05	− 0.00025	0	− 0.0002652	0.000127	0	6.11E-05
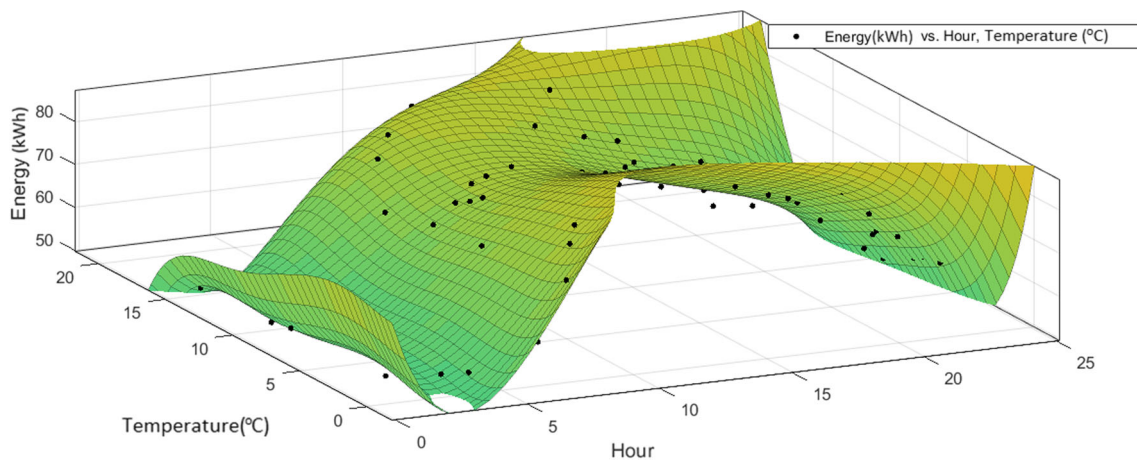
Fig. 5 3D Graph of hourly temperature (°C) and electrical energy consumption forecast (kWh) on Mondays in April

been observed in the Friday model. It has been determined that there is a noticeable change in the consumption estimation results when the temperature increases by + 2 °C and decreases by − 2 °C. If the temperature effect is not taken into account, it is clearly seen that there will be deviations in the estimates used in settlement, and thus, the imbalance cost will increase. This shows that temperature has an effect on consumption.

It is of great importance that the created MFMs pass the reliability test successfully. The fitness tests, which are widely used in the literature, were applied to the estimation results and the results are given in Table 6. On Friday, April, it was the best performing model among the mathematical models. According to the R2 test results of this mathematical model, it has been seen that the developed mathematical model can estimate consumption with an accuracy of up to 96%. In the RMSE test of the model, the RMSE value has

Table 5 Comparison of RLPCs calculated with different forecasting methods for Monday in April

Hour	RLPCs by the present method	RLPCs by the MRAM (proposed)	RLPCs by the ANN method	RLPCs temperature change + 2 °C	RLPCs temperature change − 2 °C
00	0.033018789	0.033879691	0.032034161	0.033039324	0.0333821
01	0.032622083	0.031743599	0.032007933	0.033027585	0.030612259
02	0.032647035	0.03154554	0.032113324	0.033113597	0.030366568
03	0.032557829	0.032961161	0.032418103	0.034589177	0.031852455
04	0.033527894	0.035269613	0.033614027	0.036610799	0.034765062
05	0.036697834	0.038690451	0.037621672	0.040036536	0.038669225
06	0.044504249	0.042473466	0.043946515	0.043494216	0.042919356
07	0.047749236	0.045102456	0.047838671	0.04596854	0.044758493
08	0.048477908	0.047023179	0.048710387	0.047476403	0.046449296
09	0.047744497	0.048222611	0.048806958	0.047370294	0.04799879
10	0.048741663	0.048723544	0.048266574	0.0473686	0.048743321
11	0.047024873	0.047513265	0.046833757	0.044194079	0.048387682
12	0.046460869	0.047231648	0.045966178	0.043881051	0.048116824
13	0.047868409	0.048116634	0.047297366	0.049786457	0.046868593
14	0.048270281	0.046716205	0.046566822	0.048863482	0.045552833
15	0.045349091	0.045493992	0.045441384	0.046888787	0.045325
16	0.045465839	0.044574287	0.044964908	0.045778457	0.044589619
17	0.041979917	0.043730468	0.044133084	0.044549685	0.044008106
18	0.041695086	0.043087694	0.043287118	0.042701341	0.044078891
19	0.043193954	0.043021938	0.043283366	0.041793216	0.044183157
20	0.043406682	0.042295158	0.042309853	0.040901248	0.043318901
21	0.040120594	0.040089813	0.039576654	0.03857152	0.041013453
22	0.036778148	0.038066037	0.037287244	0.036709828	0.038614764
23	0.03409724	0.03442755	0.035084158	0.033285779	0.035425252

Fig. 6 Hourly forecasted RLPCs for Friday, April
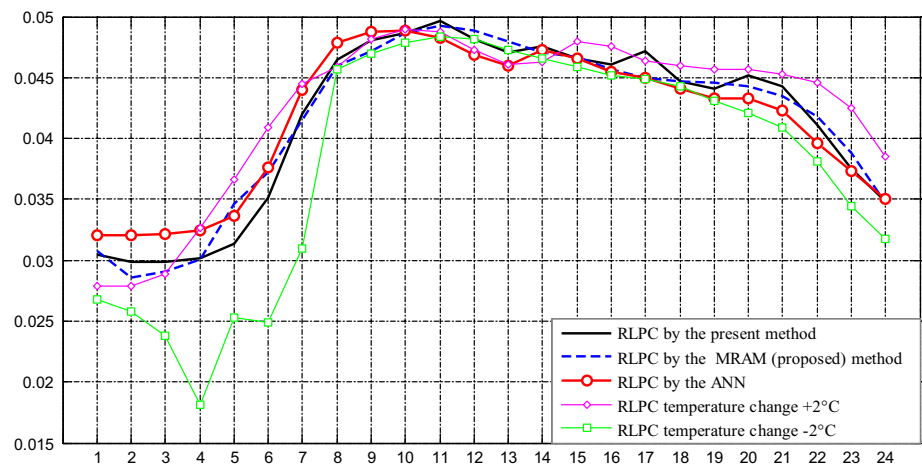
decreased to 2.34. According to the MAPE test results, the MAPE value drops to 2.67%. Similar test performances were observed for the other mathematical models. Consequently, the reliability of the proposed models is acceptable for all fitness tests. The change of temperature affects the electrical energy consumption forecasting, as shown in Table 5

and Fig. 6. So, meteorological data should also be taken into account in order to increase the consumption forecasting accuracy and reduce the imbalance cost. RLPCs can be obtained more accurately and reliably with the mathematical models created using hourly temperature data. Thus, an

Table 6 Estimation performances of the developed mathematical models according to months and days

Month	Fitness test	Monday	Tuesday	Wednesday	Thursday	Friday	Weekend
January	R^2	0.89	0.85	0.90	0.87	0.86	0.90
	Adj. R^2	0.89	0.85	0.90	0.87	0.86	0.90
	RMSE	3.87	3.86	4.62	3.4	5.04	3.77
	MAPE	4.3	4.55	5.52	3.69	5.03	3.81
February	R^2	0.92	0.93	0.90	0.87	0.91	0.93
	Adj. R^2	0.92	0.93	0.90	0.86	0.91	0.93
	RMSE	3.14	2.8	2.49	3.17	3.59	4.25
	MAPE	3.31	3.09	2.94	3.41	3.86	3.43
March	R^2	0.91	0.91	0.82	0.89	0.95	0.92
	Adj. R^2	0.91	0.91	0.81	0.89	0.95	0.92
	RMSE	3.26	3.14	3.02	2.73	2.51	3.17
	MAPE	3.37	3.36	3.87	3.21	2.85	3.33
April	R^2	0.92	0.91	0.83	0.85	0.96	0.94
	Adj. R^2	0.92	0.91	0.82	0.84	0.95	0.94
	RMSE	2.87	2.86	2.58	2.34	2.37	2.63
	MAPE	3.12	3.32	3.68	3.2	2.67	3.18
May	R^2	0.94	0.94	0.89	0.87	0.95	0.89
	Adj. R^2	0.94	0.93	0.88	0.86	0.95	0.89
	RMSE	2.55	2.58	2.13	2.69	2.84	4.07
	MAPE	3.02	3.07	2.76	4.19	3.96	4.73
June	R^2	0.90	0.94	0.93	0.90	0.96	0.94
	Adj.R^2	0.90	0.94	0.93	0.89	0.96	0.94
	RMSE	3.7	2.7	1.81	2.66	2.45	2.88
	MAPE	4.23	3.12	2.41	3.7	3.11	3.42
July	R^2	0.90	0.94	0.89	0.89	0.68	0.83
	Adj. R^2	0.90	0.94	0.89	0.89	0.68	0.83
	RMSE	7.19	5.53	7.98	8.02	12.61	9.48
	MAPE	5.51	5.2	6.58	7.45	9.71	8.19
Agust	R^2	0.90	0.91	0.94	0.94	0.90	0.91
	Adj. R^2	0.90	0.90	0.94	0.93	0.90	0.91
	RMSE	7.73	7.19	5.75	6.49	7.95	6.9
	MAPE	5.77	5.07	4.56	4.98	6.17	5.75
September	R^2	0.91	0.94	0.88	0.82	0.91	0.93
	Adj. R^2	0.91	0.94	0.88	0.81	0.91	0.93
	RMSE	7.06	5.76	9.03	10.69	6.81	6.99
	MAPE	5.94	5.11	7.47	7.6	5.66	4.83
October	R^2	0.91	0.93	0.94	0.91	0.88	0.88
	Adj. R^2	0.91	0.93	0.93	0.91	0.88	0.88
	RMSE	7.69	7.03	6.91	8.43	10.64	11.19
	MAPE	7.6	5.84	6.4	7.28	6.9	7.53
November	R^2	0.90	0.92	0.93	0.93	0.93	0.87
	Adj. R^2	0.89	0.92	0.93	0.93	0.92	0.87
	RMSE	8.06	6.61	5.71	6.22	6.4	10.34

Table 6 (continued)

Month	Fitness test	Monday	Tuesday	Wednesday	Thursday	Friday	Weekend
	MAPE	6.94	6.3	5.25	5.99	6.6	7.66
December	R^2	0.91	0.88	0.90	0.92	0.94	0.92
	Adj. R^2	0.90	0.88	0.89	0.92	0.94	0.92
	RMSE	7.82	9.94	7.76	7.12	5.41	6.61
	MAPE	7.11	6.92	7.37	6.66	5.47	6.90

important solution proposal is presented to reduce the imbalance cost in the paper.

6 Conclusions

The greater the error to be made in the consumption estimation, the imbalance costs will increase in the same proportion. In order to minimize imbalance costs, estimation methods with high accuracy are needed. In this paper, by using PMRAM, one of the time series methods, short-term load estimation mathematical models were developed using hourly air temperature and residential consumption data measured in a certain region in Duzce for one year. Similar mathematical models can be made for different geographic regions using NWPs instead of temperature measurement.

Electrical energy consumption forecastings have been made based on the mathematical models developed in this paper. In order to examine the estimation performance of the proposed models, the months, days, and hours with high seasonal temperature changes can be taken into account. Accordingly, an increase of 3.68 °C in ambient temperature was measured between 08:00 and 09:00 on Friday in April. While the consumption measured between these hours is 76.77 kWh, the electrical energy consumption has been estimated at 76.24 kWh with the proposed model, and 76.32 kWh with the ANN. There was a 5.52 °C change in the ambient temperature between 00:00 and 01:00 on the same day. Between these hours, the energy consumption was measured as 56.65 kWh. The energy consumption has been estimated at 56.95 kWh with the developed model, and 56.05 kWh with the ANN. According to these results, the accuracy of the proposed models is also supported by a more complex method, ANN.

It has been determined that the increase or decrease in the temperature values for the measured energy consumption region (Duzce) plays a decisive role in residential consumption. Therefore, it has been observed that taking into account the temperature forecasts while estimating the electrical energy consumption increases the forecast performance. This paper shows that the efficient use of MFM will improve the performance of energy consumption estimation and thus reduce imbalance costs.

Declarations

Conflict of interest Conflict of interest the authors declare that they have no conflict of interest.

References

1. https://www.epias.com.tr Accessed 19 Oct 2020
2. Hong WC (2009) Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model. Energy Convers Manag 50(1):105–117. https://doi.org/10.1016/j.enconman.2008.08.031
3. Russell FP, Wilkinson KA, Kelly PH et al (2015) Optimised three-dimensional Fourier interpolation: an analysis of techniques and application to a linear-scaling density functional theory code. Comput Phys Commun 187:8–19. https://doi.org/10.1016/j.cpc.2014.09.019
4. Yang Y, Li S, Li W et al (2018) Power load probability density forecasting using Gaussian process quantile regression. Appl Energy 213:499–509. https://doi.org/10.1016/j.apenergy.2017.11.035
5. Sheikh SK, Unde MG (2012) Short term load forecasting using ann technique. Int J Eng Sci Emerg Technol 1(2):97–107. https://doi.org/10.7323/ijeset/v1_i2_12
6. Panapakidis IP (2016) Application of hybrid computational intelligence models in short-term bus load forecasting. Expert Syst Appl 54:105–120. https://doi.org/10.1016/j.eswa.2016.01.034
7. Öztürk A, Taşpinar F (2019) Short term load forecasting for Turkey energy distribution system with artificial neural networks. Tehnički Vjesnik 26(6):1545–1553
8. Box GE, Jenkins GM, Reinsel GC et al (2015) Time series analysis: forecasting and control. John Wiley & Sons, New Jersey
9. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175. https://doi.org/10.1016/S0925-2312(01)00702-0
10. Tseng FM, Yu HC, Tzeng GH (2002) Combining neural network model with seasonal time series ARIMA model. Technol Forecast Soc Chang 69(1):71–87
11. Park JH, Park YM, Lee KY (1991) Composite modeling for adaptive short-term load forecasting. IEEE Trans Power Syst 6(2):450–457. https://doi.org/10.1109/59.76686
12. Marcellino M, Stock JH, Watson MW (2006) A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. J Econom 135(1–2):499–526. https://doi.org/10.1016/j.jeconom.2005.07.020
13. Papalexopoulos A, Hesterberg TC (1990) A regression-based approach to short-term system load forecasting. IEEE Trans Power Syst 5(4):1535–1547. https://doi.org/10.1109/59.99410
14. Vrablecová P, Ezzeddine AB, Rozinajová V et al (2018) Smart grid load forecasting using online support vector regression. Comput Electr Eng 65:102–117. https://doi.org/10.1016/j.compeleceng.2017.07.006

15. Wang S, Wang S, Wang D (2019) Combined probability density model for medium term load forecasting based on quantile regression and kernel density estimation. Energy Procedia 158:6446–6451. https://doi.org/10.1016/j.egypro.2019.01.169

16. He Y, Liu R, Li H et al (2017) Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. Appl Energy 185:254–266. https://doi.org/10.1016/j.apenergy.2016.10.079

17. Zhang W, Quan H, Srinivasan D (2018) Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. Energy 160:810–819. https://doi.org/10.1016/j.energy.2018.07.019

18. Hu R, Wen S, Zeng Z et al (2017) A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. Neurocomputing 221:24–31. https://doi.org/10.1016/j.neucom.2016.09.027

19. Dordonnat V, Pichavant A, Pierrot A (2016) GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models. Int J Forecast 32(3):1005–1011. https://doi.org/10.1016/j.ijforecast.2015.11.010

20. Ziel F (2019) Quantile regression for the qualifying match of GEFCom2017 probabilistic load forecasting. Int J Forecast 35(4):1400–1408. https://doi.org/10.1016/j.ijforecast.2018.07.004

21. Wu J, Wang J, Lu H et al (2013) Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. Energy Convers Manag 70:1–9. https://doi.org/10.1016/j.enconman.2013.02.010

22. Yaslan Y, Bican, (2017) Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting. Measurement 103:52–61. https://doi.org/10.1016/j.measurement.2017.02.007

23. Chen Y, Xu P, Chu Y et al (2017) Short-term electrical load forecasting using the support vector regression (svr) model to calculate the demand response baseline for office buildings. Appl Energy 195:659–670. https://doi.org/10.1016/j.apenergy.2017.03.034

24. He Y, Zheng Y (2018) Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function. Energy 154:143–156. https://doi.org/10.1016/j.energy.2018.04.072

25. Zhang N, Li Z, Zou X et al (2019) Comparison of three short-term load forecast models in Southern California. Energy 189:116358. https://doi.org/10.1016/j.energy.2019.116358

26. Yang J, Stenzel J (2006) Short-term load forecasting with increment regression tree. Electr Power Sys Res 76(9–10):880–888. https://doi.org/10.1016/j.epsr.2005.11.007

27. Jha N, Prashar D, Rashid M, Gupta SK, Saket RK (2021) Electricity load forecasting and feature extraction in smart grid using neural networks. Comput Electr Eng 96:107479. https://doi.org/10.1016/j.compeleceng.2021.107479

28. Feng Q, Qian S (2021) Research on power load forecasting model of economic development zone based on neural network. Energy Rep 7:1447–1452. https://doi.org/10.1016/j.egyr.2021.09.098

29. Chen R, Lai CS, Zhong C, Pan K, Ng WW, Li Z, Lai LL (2022) MultiCycleNet: multiple cycles self-boosted neural network for short-term electric household load forecasting. Sustain Cities Soc 76:103484. https://doi.org/10.1016/j.scs.2021.103484

30. Velasco LCP, Arnejo KAS, Macarat JSS (2022) Performance analysis of artificial neural network models for hour-ahead electric load forecasting. Procedia Comput Sci 197:16–24. https://doi.org/10.1016/j.procs.2021.12.113

31. Aly HH (2020) A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid. Electr Power Sys Res 182:106191. https://doi.org/10.1016/j.epsr.2019.106191

32. Di Piazza A, Di Piazza MC, La Tona G, Luna M (2021) An artificial neural network-based forecasting model of energy-related time series for electrical grid management. Math Comput Simul 184:294–305. https://doi.org/10.1016/j.matcom.2020.05.010

33. https://www.epdk.gov.tr/Detay/DownloadDocument?id=KKs7oAf43Ys=. Accessed 19 Oct 2020

34. Aslan Y, Yavasca S, Yasar C (2011) Long term electric peak load forecasting of Kutahya using different approaches. Int J Tech Phys Probl Eng 3(2):87–91

35. Chapra SC, Canale RP (2003) Numerical methods for engineers. Mc Graw Hill, USA

36. Heizer J, Render B, Munson C (2014) Operations management-sustainability and supply chain management. Pearson, Essex

37. Krajewski LJ, Ritzman LP, Malhotra MK (2013) Operations management: processes and supply chains. Pearson, New Jersey

38. Saha P, Debnath P, Thomas P (2019) Prediction of fresh and hardened properties of self-compacting concrete using support vector regression approach. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04267-w

39. Liu Z, Jiang P, Zhang L et al (2020) A combined forecasting model for time series: application to short-term wind speed forecasting. Appl Energy 259:114137. https://doi.org/10.1016/j.apenergy.2019.114137

40. Li HZ, Guo S, Li CJ et al (2013) A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. Knowl-Based Syst 37:378–387. https://doi.org/10.1016/j.knosys.2012.08.015

41. Kazemzadeh MR, Amjadian A, Amraee T (2020) A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. Energy. https://doi.org/10.1016/j.energy.2020.117948

42. Yildiz B, Bilbao JI, Sproul AB (2017) A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 73:1104–1122. https://doi.org/10.1016/j.rser.2017.02.023

43. Leach LF, Henson RK (2007) The use and impact of adjusted R^2 effects in published regression research. Mult Linear Regres Viewp 33(1):1–11

44. Alkan Ö, Öztürk A, Tosun S (2018) Rüzgar Ve Güneş Santrallerinde Kısa Dönem Enerji Üretim Tahmini İçin Matematiksel Modellerin Oluşturulması. Düzce Üniversitesi Bilim Ve Teknoloji Dergisi 6(1):188–195

45. Crawford GW, Fratantoni MC (2003) Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. Real Estate Econ 31(2):223–243

46. Chen RJ, Bloomfield P, Fu JS (2003) An evaluation of alternative forecasting methods to recreation visitation. J Leis Res 35(4):441–454. https://doi.org/10.1080/00222216.2003.11950005