

Article

Analysis and Forecasting of Monthly Electricity Demand Time Series Using Pattern-Based Statistical Methods

Paweł Pelka 

Electrical Engineering Faculty, Czestochowa University of Technology, 42-200 Czestochowa, Poland;
pawel.pelka@pcz.pl

Abstract: This article provides a solution based on statistical methods (ARIMA, ETS, and Prophet) to predict monthly power demand, which approximates the relationship between historical and future demand patterns. The energy demand time series shows seasonal fluctuation cycles, long-term trends, instability, and random noise. In order to simplify the prediction issue, the monthly load time series is represented by an annual cycle pattern, which unifies the data and filters the trends. A simulation study performed on the monthly electricity load time series for 35 European countries confirmed the high accuracy of the proposed models.

Keywords: medium-term load forecasting; pattern-based forecasting; time-series preprocessing

1. Introduction

Forecasting loads on power systems is an integral activity embedded in the long-term system operation planning processes and in the processes of the ongoing control of its operation. The system cannot function without accurate forecasts. This is because electricity cannot be stored in large quantities. The demand must be covered on an ongoing basis with production, with the limitations resulting from the flexibility of the production units and the requirements of the reliability and safety of the system operation. The accuracy of forecasts translates into the costs of production, transmission, and the degree of reliability of electricity supply to consumers. Inflated forecasts lead to the maintenance of too many generating units in order to meet the safety requirements to ensure an adequate margin of reserve capacity. Underestimated forecasts have the opposite effect—too few generating units are planned, which are not able to cover the actual demand. In such a situation, additional units with quick start-up are intervened in the traffic, generating additional operating costs [1].

Medium-term forecasting most often concerns forecasting the monthly electricity load (MEL). MEL time series contain components of nonlinear trend, annual seasonality and random disturbances. They show a significant variation in the variance and shape of the annual cycle over time. MEL is highly dependent on economic and socioeconomic as well as climatic and weather variables. The factors disturbing the MEL include unpredictable economic events, extreme weather changes and political decisions [2]. The importance of MEL forecasting in the power sector and the complexity of the problem encourage the search for forecasting models that will meet the requirements of the specificity of the task and generate accurate forecasts.

Medium-term electricity demand forecasting is a very well-researched issue. There is a great deal of solutions used to settle this issue, including classical/statistical methods [3,4], neural networks (NNs) [5–7], deep learning [8–12] or similarity-based methods [13–18]. Traditional strategies were first launched for electricity load forecasting. Linear regression, ARIMA, and exponential smoothing (ETS) methods have been widely used [19]. The restricted versatile capacities of these strategies and their linear nature have brought about an expanded interest in artificial intelligence techniques [5]. Neural networks in [5] are



Citation: Pelka, P. Analysis and Forecasting of Monthly Electricity Demand Time Series Using Pattern-Based Statistical Methods. *Energies* **2023**, *16*, 827. <https://doi.org/10.3390/en16020827>

Academic Editor: Surender Reddy Salkuti

Received: 13 December 2022

Revised: 4 January 2023

Accepted: 9 January 2023

Published: 11 January 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

utilized to predict the trend of the time series of MELs and the Fourier series are included to forecast the seasonal component. Then, at that point, the two gauges, trend and occasional changes, are totaled. In [20], a deep short-term memory network (LSTM) was carried out for the probabilistic estimating of client load profiles. The LSTM method is also used for predicting electricity prices in the article [21] with good accuracy.

The simplest classical models include naive models, which assume a selected historical load value as a forecast. In series that exhibit seasonality, such as monthly load series, this is a twelve-month shifted value. Random fluctuations and trend may negatively affect the forecast results.

Linear regression models allow for taking into account the trend (only linear), but the implementation of seasonal cycles in the model requires additional operations, e.g., decomposition of the series into individual months. An example of the application of a linear model for medium- and long-term forecasting of the electricity loads can be found in [22]. The model uses strong daily (24 h) and annual (52 weeks) correlations to forecast daily load profiles in the horizon from several weeks to several years. The forecast results are corrected for annual load increments. For forecasting with a one-year horizon, the MAPE errors were obtained with values not greater than 3.8%. In [3], the operation of the linear regression model and the ARIMA model in the task of forecasting monthly peak loads up to 12 months in advance was compared. The models are powered by the same set of inputs, including historical peak load data, weather data, and economic data. About twice the accuracy of the ARIMA model was demonstrated experimentally. For non-stationary time series with an irregular periodic trend, [23] proposed a linear regression model extended with periodic components implemented by the sine function of different frequencies.

The spatial autoregression model for medium-term load forecasting is described in [24]. The authors noted a strong correlation between the load on the system and GDP in the analyzed thirty Chinese provinces. To forecast the load in a given province, they used not only the local dependence of the load on GDP, but also the relationships identified for neighboring provinces. This allows to reduce forecast errors from 5.2–5.4% to 3.5–3.9%. Ref. [25] described ARIMA and ETS prognostic methods in combination with bootstrap aggregation. The time series were initially processed using the Box-Cox transform (a procedure often used to compensate for variances) and decomposed. The ARMA model was used for the generation of bootstrap tests, and the ARIMA and ETS models for their forecasting.

Classical methods also use Fourier series. In [5], they were used to model the seasonal component of the time series of monthly loads. Based on spectral analysis, six fundamental frequencies were identified, and a Fourier series was created for them. The forecast of the seasonal component calculated by the Fourier series was added to the trend forecast. Markov chains were used in [26] to analyze the input data and select the best forecasting model. This approach is especially useful when the upward trend of the time series is unstable. The classic models also include the model described in [27], which is based on a simple logistic function. The input variables are time, maximum atmospheric temperature and the social factor, taking into account religious holidays (the model was developed for Arab countries).

In this work, statistical methods for estimating month-to-month power demand are utilized. What differentiates these models from other traditional strategies is that they use pattern representation of seasonal cycles of the time series. The patterns allow us to unify the data and filter out the trend. The input and output variables in the pattern space are characterized by a less complex relationship contrasted with the original space. Thus, due to pattern representation, the classical forecasting method has an easier task to perform. Using time-series preprocessing by calculating yearly patterns constitutes the direct novelty of the proposed models.

Patterns represent fragments of time series and extract information about the shape of these fragments, filter out the trend and normalize the variance. By using a pattern-based

representation, the relationship between the input and output variables is simplified, and hence the predictive model is simplified. Unlike parametric systems, which are most often “black box” models, the statistical methods’ principle of operation is understandable, which is important in practical industrial applications and translates to a greater degree of confidence in the forecast.

2. MEL Time Series Analysis

A time series is a set of observations of a certain variable ordered in the time domain: $\{z_t, t = 1, 2, \dots, N\}$. Formally, the observations z_t are implementations of a sequence of random variables $\{Z_t, t = 1, 2, \dots, N\}$ having certain specified cumulative distribution. These definitions are applied to a discrete variable when the time variable t takes values at equal intervals (seconds, hours, days, months).

The purpose of the time series analysis is to detect and describe the regularities affecting the phenomenon expressed in the form of a time series. The following components of time series are distinguished, which are the effect of the influence of various factors on the studied phenomenon [28,29]:

- Development tendency (trend), which is revealed by one-way and systematic changes in the level of a given phenomenon that takes place over a long-term period;
- Periodic fluctuations, i.e., rhythmic fluctuations with a specific cycle;
- Business cycle fluctuations, i.e., systemic wave fluctuations observed in longer periods;
- Random fluctuations, i.e., irregular random disturbances.

Periodic fluctuations, also known as seasonal fluctuations [30], are characterized by a constant period, as opposed to a cyclical fluctuations in which there are no fixed period fluctuations.

A time series is called stationary if its statistical properties do not change over time. That is, a time series of $\{Z_t\}$ has the same properties as time series shifted in time $\{Z_{t+l}\}$, for each value shift l . In practice, weak stationarity is tested (so-called wide-sense stationarity), which requires immutability mean level, variance and time correlation. It can be written as follows [28]:

- $EZ_t = \text{const}$ —expected value is constant (none trend);
- $\text{Var}(Z_t) = \text{const}$ —the variance is homogeneous over time;
- $\text{Corr}(Z_r, Z_s) = \text{Corr}(Z_{r+t}, Z_{s+t}) = \rho(s - r) = \rho(l)$ —the correlation between observations is solely dependent on time shift l .

The stationarity of the time series plays a key role in a forecasting model selection and the construction of forecasts. Establishing forecasts for the stationery time series, where the basic properties do not change over time, is much easier than for non-stationary series. Therefore, in practice, it is often sought to transform the time series into a stationary form before starting the forecasting process.

A series of MELs for Poland with a box diagram, showing the distribution of monthly values in the following years, is shown in Figure 1. The variability of the median and variance is clearly visible over time. The last year of observation is characterized by almost a three times smaller interquartile range (box height) than the first one. Figure 2 shows the MEL time series for 35 European countries. The time series vary in length, from 60 to 288 months, ending with a year 2014. As we can see, these time series are non-stationary, and they show variability of mean value and variance. They are characterized by strong annual seasonality and non-linear trends.

Other forms of graphical presentation of MEL time series are shown in Figures 3 and 4. Figure 3 presents the trends set in the subsequent months of the year. Strong upward trends are observed in the summer months, which is not the case for the winter months. Figure 4) allows us to observe distinctive features of annual cycles and assess the shape similarity of these cycles over the years. Despite the reduction in the amplitude of the yearly fluctuations in subsequent years, the shapes of the annual cycles show similarities. The demand in the winter months is significantly greater than the demand in the summer

months. It is observed that there is understated demand in February with the following months (which cannot only be explained by the smaller number of days this month) and overstated demand in March and October. The shape similarities of the annual cycles are particularly important in the nonparametric regression models [31]. The seasonal graph can be compiled in polar coordinates (Figure 4)).

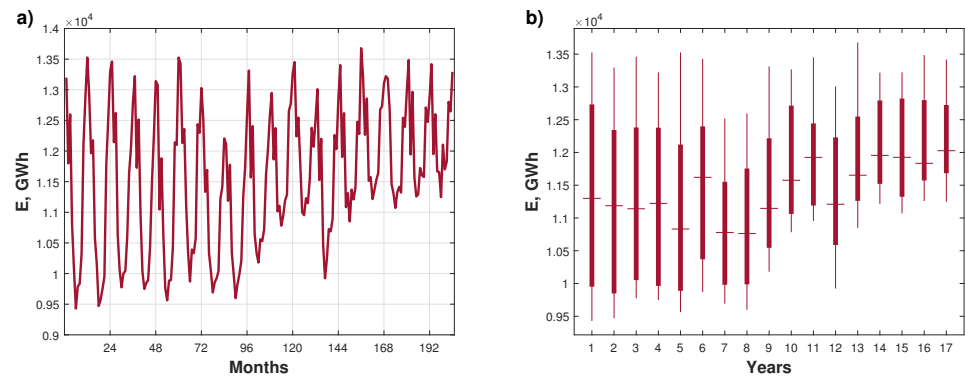


Figure 1. Graph of the monthly electricity demand time series for Poland (a) and the box diagram showing the medians and annual dispersions (b).

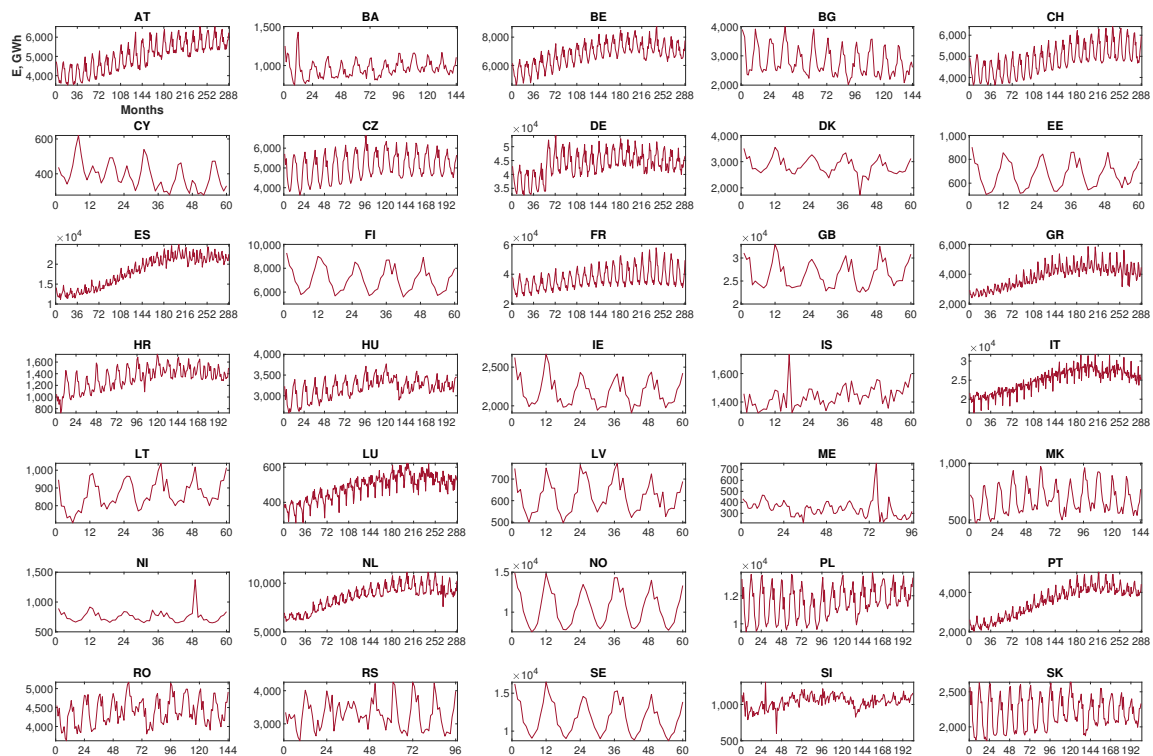


Figure 2. Monthly electricity demand time series for European countries.

Relationship between observations of the time series distant by l time units are assessed using the autocorrelation function (ACF) and the partial autocorrelation (PACF). The former has the form [28]:

$$ACF(l) = \frac{\hat{\gamma}(l)}{\hat{\gamma}(0)}, l = 0, 1, \dots, N-1, \quad (1)$$

where $\hat{\gamma}(h)$ is a sample autocovariance function:

$$\hat{\gamma}(l) = \frac{1}{N} \sum_{j=1}^{N-l} (Z_{j+l} - \bar{Z})(Z_j - \bar{Z}), \quad (2)$$

where $\bar{Z}$ is the mean value of the time series.

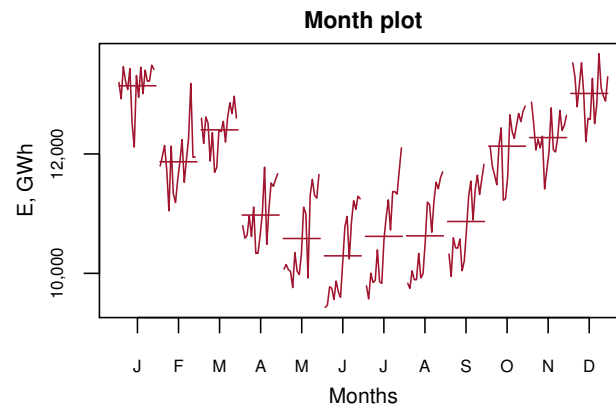


Figure 3. Month plot of energy load for Poland.

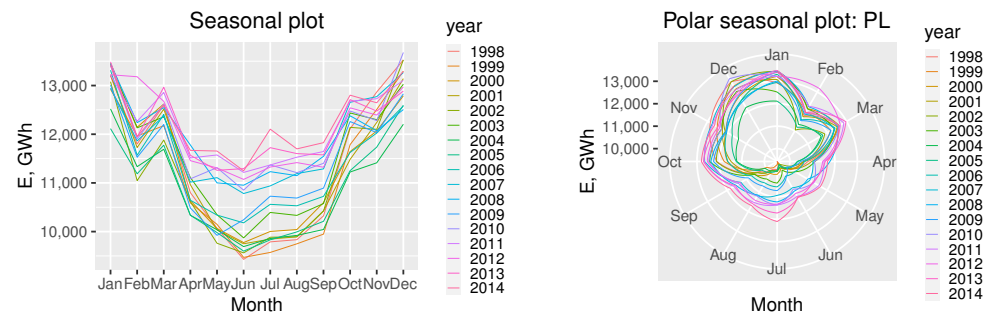


Figure 4. Seasonal MEL plots for Poland.

In the case of PACF, when determining the autocorrelation between Z_t and Z_{t+l} , the influence of intermediate observations on this is eliminated dependency, $Z_{t+1}, \dots, Z_{t+l-1}$. Sample partial autocorrelation is an estimate of a theoretical PACF function of a given pattern:

$$\alpha(l) = \begin{cases} \text{Corr}(Z_{t+1}, Z_t), & l = 1 \\ \text{Corr}(Z_{t+l} - P_{t,l}(Z_{t+l}), Z_t - P_{t,l}(Z_t)), & l > 1 \end{cases} \quad (3)$$

where $P_{t,l}(Z_t)$ is the orthogonal projection operator on linear subspace spanning over variables $Z_{t+1}, \dots, Z_{t+l-1}$.

Examples of ACF and PACF charts for the MEL time series are shown in Figure 5. The horizontal lines designate the confidence intervals, which allow a conclusion to be drawn about the statistical significance of the autocorrelation. The 95% confidence interval has the form $\left[-\frac{1.96}{\sqrt{N}}, \frac{1.96}{\sqrt{N}}\right]$. Autocorrelations outside this range are considered to be statistically significant. On the ACF chart, we can see very strong oscillations, proving a clear seasonality of the series. According to what is expected, the strongest autocorrelation occurs for lags which are multiples of twelve, which is confirmed by the annual cycle. The big PACF value for delay $l = 1$, close to 1, can signal the presence of an uptrend.

Seasonal fluctuations can also be identified using the harmonic (spectral, spectral) analysis [32]. It leads to creating a model consisting of the sum of the sine and cosine functions different frequency. A time series of length n is recorded using the Fourier series as follows [33,34]:

$$f(t) = a_0 + \sum_{i=1}^{N/2} \left[a_i \sin\left(\frac{2\pi i t}{N}\right) + b_i \cos\left(\frac{2\pi i t}{N}\right) \right], \quad (4)$$

where a_0 , a_i and b_i are the coefficients which are determined from patterns:

$$\begin{aligned}
 a_0 &= \frac{1}{N} \sum_{t=1}^N Z_t \\
 a_i &= \frac{2}{N} \sum_{t=1}^N Z_t \sin\left(\frac{2\pi it}{N}\right), i = 1, 2, \dots, \frac{N}{2} - 1 \\
 b_i &= \frac{2}{N} \sum_{t=1}^N Z_t \cos\left(\frac{2\pi it}{N}\right), i = 1, 2, \dots, \frac{N}{2} - 1, \\
 a_{N/2} &= 0 \\
 b_{N/2} &= \frac{1}{N} \sum_{t=1}^N Z_t \cos(\pi t)
 \end{aligned} \tag{5}$$

The magnitudes of the amplitudes for successive harmonics are as follows:

$$A_i = \sqrt{a_i^2 + b_i^2}. \tag{6}$$

The amplitude of i -th harmony, A_i , testifies to the participation of this harmony in explaining the variance of the considered variable. This share is expressed by the following formula [33]:

$$u_i = \frac{A_i^2}{2\text{Var}(Z_t)}, \tag{7}$$

The amplitudes of successive harmonics for the MEL time series for Poland are shown in a periodogram (Figure 6). As can be seen, the dominant period in this series is a year.

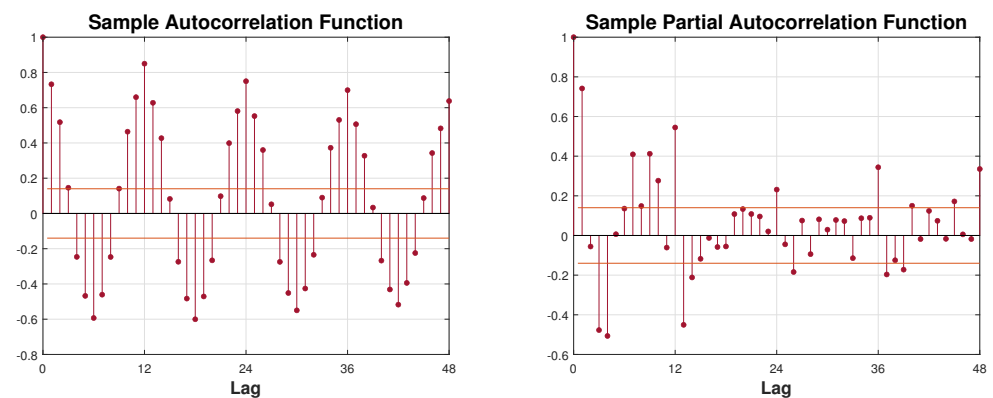


Figure 5. Graphs of autocorrelation and partial correlation of the MEL time series for Poland (lag means delay l).

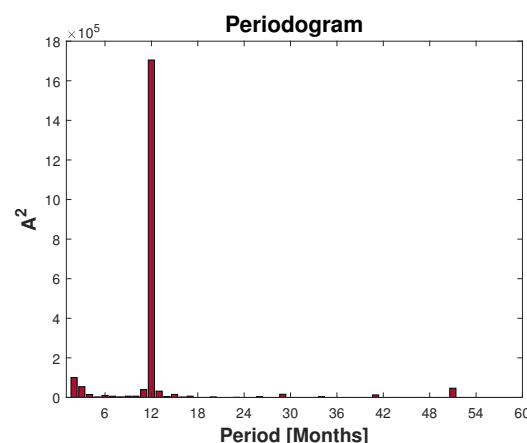


Figure 6. Periodogram of the MEL time series for Poland.

3. Forecasting Model

Let us consider the monthly electricity demand time series beginning from January and finishing off with December: $E = \{E_t : t = 1, 2, \dots, N\}$. We divide each time series into yearly fragments $E_i = \{E_t : t = 12(i-1) + 1, 12(i-1) + 2, \dots, 12(i-1) + 12\}$, $i = 1, 2, \dots, N/12$. Every fragment can be described by a vector $\mathbf{E}_i = [E_{i,1} E_{i,2} \dots E_{i,12}]^T$.

Let us formulate an x -pattern $\mathbf{x}_i = [x_{i,1}x_{i,2}\dots x_{i,12}]^T$ as a vector, which represents a yearly fragment E_i . A function, which transforms time series points into patterns, is chosen, including the character of the time series, such as trend, variance, and seasonalities. We proposed a few definitions of this function in [35]. In this study, x -patterns are defined as [8]

$$x_{i,j} = \frac{E_{i,j} - \bar{E}_i}{\sigma_i}, \quad (8)$$

where $j = 1, 2, \dots, 12$, $\bar{E}_i$ is a mean of sequence E_i , and $\sigma_i = \sqrt{\sum_{j=1}^n (E_{i,j} - \bar{E}_i)^2}$ is a measure of the sequence E_i dispersion.

The X -pattern is a normalized E_i vector. Note that yearly fragments expressed by E_i have a different mean and dispersion. After the normalization process, time-series fragments are also unified: all x -patterns have unity length, the same variance and also the mean value of these fragments equals zero. x -patterns carry information about the shapes of the yearly fragments. Then, the new time series composed of x -patterns representing successive yearly periods are created: $x = \{x_i : i = 1, 2, \dots, N/12\} = \{x_{1,1}, x_{1,2}, \dots, x_{N/12,12}\}$. Note that it is distinguished by regular character and stationarity.

The forecasting procedure using the time series composed of x -patterns requires determining the demand forecast using the x -pattern forecast. After generating the x -pattern by the forecasting model, the MELs in the forecasted yearly period are computed from the forecasted x -pattern using transformed Equation (8) (this is called decoding):

$$\hat{E}_{i,j} = \hat{x}_{i,j} * \hat{\sigma}_i + \hat{\bar{E}}_i, j = 1, 2, \dots, 12. \quad (9)$$

However, in this equation, the coding variables, $\bar{E}_i$ and σ_i , are not known because they are the mean and dispersion of the future fragment, which is forecasted. So, the coding variables must be forecasted based on their historical values. ARIMA and ETS models are used for this purpose in this work in every model and for preprocessed time-series forecasting, when we use these two models [8]. Figure 7 presents the idea of pattern-based forecasting using a block diagram of the proposed methodology presenting data structures and illustrating data flow.

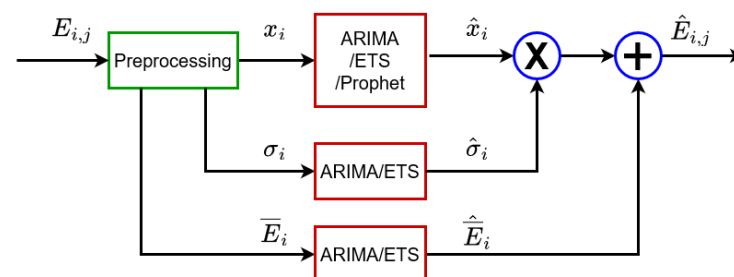


Figure 7. A block diagram of pattern-based MEL forecasting methodology.

3.1. Autoregressive Integrated Moving Average ARIMA Model

The autoregressive integrated moving average ARIMA model $ARIMA(p, d, q)(P, D, Q)_m$ [30,36] was used to model the MEL time series:

$$\Phi(B^m)\phi(B)(1-B)^D(1-B)^d z_t = c + \Theta(B^m)\theta(B)\xi_t, \quad (10)$$

where z_t are the terms of the time series (in the considered case of the series $\{E_t, t = 1, 2, \dots, N\}$), m is the length of the annual cycle ($m = 12$), B is the backward shift operator, D and d are the orders of seasonal differentiation and of the ordinary, respectively, $\phi(\cdot)$, $\Phi(\cdot)$, $\theta(\cdot)$, and $\Theta(\cdot)$ are polynomials of degree p , g , P and Q , respectively, c denotes a constant, and ξ_t is a white noise process with zero mean and the variance of σ^2 .

The ARIMA model was also used to construct predictive models for the code variables $\hat{E}_i$ and $\hat{\sigma}_i$ needed to decode the pattern x . In this case, the ARIMA model can be simplified to the non-seasonal ARIMA form (p, d, q) :

$$\phi(B)(1 - B^d)z_t = c + \theta(B)\xi_t, \quad (11)$$

where z_t represents the terms of the time series $\{\bar{E}_1, \bar{E}_2, \dots, \bar{E}_{N-1}\}$ or $\{\sigma_1, \sigma_2, \dots, \sigma_{N-1}\}$.

3.2. Exponential Smoothing ETS

The seasonal exponential smoothing method (Holt–Winters method), known since the 1960s, is, next to the ARIMA method, one of the most frequently used in practice. The idea of ETS is to assign exponentially declining weights to observations from previous periods. Thus, observations from recent periods have a greater impact on the forecast. The Holt–Winters model is based on three smoothing equations that represent the level of the forecast variable, its increment and seasonality. There are two types of Holt–Winters methods, which depend on how seasonality is modeled. The additive version of the Holt–Winters method is used when seasonal fluctuations are independent of the trend. On the other hand, the multiplicative version is used when there is a proportional relationship between seasonal fluctuations and the trend [28].

The equations for the additive version of the Holt–Winters method are [36]

$$\begin{aligned} \text{Level} : l &= \alpha(y - s_{-m}) + (1 - \alpha)(l_{-1} - b_{-1}) \\ \text{Increment} : b &= \beta(l - l_{-1}) + (1 - \beta)b_{-1} \\ \text{Seasonality} : s &= \gamma(y - l_{-1} - b_{-1}) + (1 - \gamma)s_{-m}' \\ \text{Forecast} : \hat{y}_{t+h|t} &= l + bh + s_{-m+h_m^+} \end{aligned} \quad (12)$$

where m is the period of seasonal fluctuations ($m = 12$ for the MEL series), h is the forecast horizon, $h_m^+ = [(h - 1) \bmod m] + 1$, and α, β, γ are the smoothing coefficients from the range $(0, 1)$.

Before using the model, the initial values of the states $l_0, b_0, s_{1-m}, \dots, s_0$ and the smoothing parameters α, β and γ should be given. All these values are then estimated from the observed data.

In [36], ETS models were defined by state equations and classified into 30 types. These types differ in the ways in which the model includes the trend components (the sum of the level and the increment), seasonality and error. Components can be expressed as additive or multiplicative, and the trend can be further suppressed.

The ETS model was also used to build forecasting models for the code variables $\hat{E}_i$ and $\hat{\sigma}_i$.

3.3. Prophet

Prophet is a time series forecasting method designed by Facebook for direct use in business applications [37]. It is distinguished by a completely automatic forecasting procedure with an intuitive selection of parameter values that can be adjusted without knowing the details of the base model. Prophet is immune to data deficiencies and trend changes. It usually copes well with outliers. The model implementation is publicly available in the Python and R environments.

Prophet is an additive model with three components—non-linear trend, seasonality, and a component that represents holidays:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (13)$$

where $g(t)$ is a trend function modeling non-periodic changes in the value of the time series, $s(t)$ represents seasonal changes (e.g., weekly and annual seasonality), $h(t)$ represents holiday effects, and ϵ_t is a residual component with a normal distribution.

The model specification is similar to the generalized additive model (GAM), with nonlinear smoothing and time t as the sole regressor. The trend is modeled in two ways: using a piecewise-linear model or a limited-growth model. In the latter case, the logistic curve is used:

$$g(t) = \frac{C(t)}{1 + \exp(-k(t)(t - q(t)))}, \quad (14)$$

where $C(t)$ is the carrying capacity, $k(t)$ is the increment function, and $q(t)$ is the offset function.

In (14) both the carrying capacity and the increment and offset are functions of time. They vary depending on the characteristics of the time series. This makes it possible to flexibly shape the trend function.

The seasonal component is modeled using the Fourier series:

$$s(t) = \sum_{i=1}^N \left(a_i \cos\left(\frac{2\pi it}{m}\right) + b_i \sin\left(\frac{2\pi it}{m}\right) \right), \quad (15)$$

where m is the length of the seasonal cycle, and a_i and b_i are the coefficients.

The component representing the effects of public holidays, $h(t)$, is not applicable in forecasting MEL time series.

4. Simulation Study

In this section, we analyze the monthly electricity demand time series for 35 European countries and verify our proposed forecasting model on these time series. The data were downloaded from the ENTSO-E repository, (www.entsoe.eu, accessed on 12 April 2016).

4.1. MEL Time Series Analysis Results

Table 1 shows the statistics and parameters describing the analyzed time series:

- Median—median as a measure of the average level of the series,
- IQR—average of the annual interquartile ranges as a measure of the annual dispersion of the series,
- $iqr_{\%}$ —mean relative annual dispersion as mean ratio of annual interquartile ranges to annual medians:

$$iqr_{\%} = \frac{100}{M} \sum_{i=1}^M \frac{IQR_i}{Median_i}, \quad (16)$$

where M is the length of the series in years, IQR_i and $Median_i$ is the interquartile range and the median of the year i ,

- $ACF(12)$ —value of the autocorrelation function for the delay $l = 12$,
- u^{12} —annual-period harmonics share in the series variance (7).

The countries classified as the largest, most developed and with the largest number of inhabitants in Europe have the greatest demand for electricity, i.e., Germany, France, Italy and Great Britain. The greatest relative annual volatility $iqr_{\%}$ is characterized in the following order: Norway, Sweden, Montenegro, Bulgaria, Estonia and France. In these cases, the interval of the median is over 25% of the median. The least annual volatility is observed for Italy and Iceland (less than 7%).

The strongest autocorrelation for the annual delay $l = 12$ ($ACF(12) \geq 0.9$), signaling the clearest annual cycles, occurs for the time series of the MEL of Switzerland, Spain, Portugal, France and Italian. The MEL time series of Montenegro, Northern Ireland and Iceland show the lowest values $ACF(12)$.

The annual-period harmonics share is the highest for Norway, Sweden, Finland and Estonia. There is a high correlation u^{12} with mean relative annual dispersion $iqr_{\%}$. The harmonic analysis in some cases gives very low values u^{12} despite the high values of the annual autocorrelation, e.g., for Italy, Greece and Spain. Periodograms for these countries

show a high bar for the year period but many lower bars for higher frequencies. This means that there are disturbances in the series of those countries with the lower value of u^{12} , masking the picture of the annual cycle.

Table 1. Statistics and describing parameters analyzed MEL time series.

Symbol	Median	IQR	iqr%
AT	4978.00 ± 641.00	720.00 ± 164.00	15.00 ± 4.36
BA	955.00 ± 49.00	141.00 ± 58.00	14.91 ± 5.66
BE	6981.00 ± 609.00	848.00 ± 96.00	12.66 ± 1.93
BG	2654.00 ± 152.00	740.00 ± 200.00	27.69 ± 6.64
CH	4885.00 ± 425.00	862.00 ± 120.00	18.15 ± 2.87
CY	375.00 ± 35.00	87.00 ± 12.00	23.33 ± 3.41
CZ	5134.00 ± 243.00	1105.00 ± 203.00	21.98 ± 4.82
DE	44,140.00 ± 2883.00	5921.00 ± 1411.00	13.79 ± 3.81
DK	2824.00 ± 86.00	431.00 ± 86.00	15.46 ± 2.95
EE	672.00 ± 16.00	181.00 ± 42.00	27.37 ± 6.66
ES	19,294.00 ± 3778.00	1466.00 ± 437.00	8.11 ± 1.49
FI	7052.00 ± 89.00	1636.00 ± 204.00	23.33 ± 2.84
FR	36,860.00 ± 2948.00	9116.00 ± 2822.00	25.01 ± 6.69
GB	25,751.00 ± 869.00	4559.00 ± 750.00	17.66 ± 2.80
GR	4094.00 ± 623.00	467.00 ± 161.00	12.03 ± 3.05
HR	1346.00 ± 158.00	196.00 ± 40.00	15.24 ± 3.86
HU	3224.00 ± 180.00	303.00 ± 92.00	9.70 ± 3.51
IE	2157.00 ± 12.00	290.00 ± 42.00	13.48 ± 1.91
IS	1413.00 ± 41.00	92.00 ± 9.00	6.48 ± 0.71
IT	25,913.00 ± 2767.00	1290.00 ± 246.00	5.15 ± 0.86
LT	864.00 ± 38.00	102.00 ± 9.00	11.99 ± 0.91
LU	509.00 ± 64.00	46.00 ± 11.00	9.53 ± 2.44
LV	622.00 ± 18.00	105.00 ± 19.00	17.01 ± 3.23
ME	345.00 ± 42.00	77.00 ± 46.00	22.25 ± 11.75
MK	641.00 ± 41.00	190.00 ± 38.00	29.60 ± 6.47
NI	742.00 ± 8.00	109.00 ± 14.00	14.79 ± 1.90
NL	8880.00 ± 1099.00	837.00 ± 243.00	9.78 ± 2.00
NO	10,339.00 ± 397.00	3827.00 ± 460.00	37.03 ± 4.90
PL	11,299.00 ± 429.00	1774.00 ± 517.00	15.65 ± 4.86
PT	3764.00 ± 716.00	292.00 ± 70.00	8.37 ± 1.20
RO	4357.00 ± 142.00	543.00 ± 104.00	12.53 ± 2.64
RS	3201.00 ± 104.00	632.00 ± 289.00	19.82 ± 9.55
SE	11,600.00 ± 299.00	3511.00 ± 462.00	30.55 ± 3.64
SI	1046.00 ± 64.00	86.00 ± 28.00	8.43 ± 3.00
SK	2165.00 ± 59.00	383.00 ± 117.00	17.70 ± 5.66

Figure 8 shows pie charts presenting the shares of each components of decomposition, i.e., trend, periodic fluctuations and random fluctuations, in the total variance of the MEL time series. The results were used as seasonal–trend decomposition using LOESS (STL). The shares were calculated from the formulas

$$u_T = \frac{\text{Var}(T_t)}{\text{Var}(E_t)}, u_S = \frac{\text{Var}(S_t)}{\text{Var}(E_t)}, u_R = \frac{\text{Var}(R_t)}{\text{Var}(E_t)}, \quad (17)$$

where T_t , S_t , R_t is the component of the trend, respectively, seasonal and random fluctuations, and E_t is the MEL time series.

By analyzing pie charts, one can divide the MEL time series from due to the dominant component. Countries whose time series have the highest content of the trend component, above 80%, are Spain, Portugal, the Netherlands and Italy. The highest share of the seasonal component, over 90%, distinguishes Norway, Finland, Estonia, Sweden and Ireland. Countries whose MEL time series show the dominance of the seasonal component comprise the most numerous group. It should be noted that the results of these analyzes are dependent on the length of the time series. The trend shines through the time series

longer ones, while in short ones, the seasonal component dominates. Montenegro is the only country whose series includes the disturbance component as dominant (53%). Other countries with a high proportion of random fluctuations, over 30% are in the following order: Northern Ireland, Serbia, Iceland and Slovenia [31].

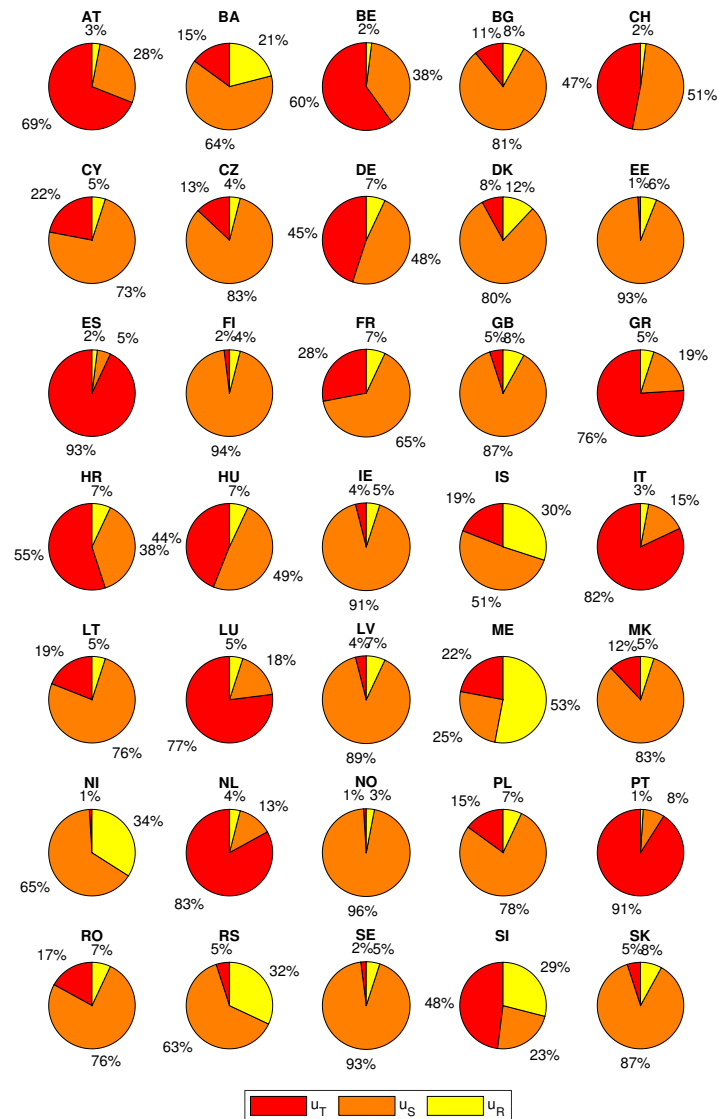


Figure 8. Shares of individual components decomposition in the total variance of the MEL time series.

4.2. MEL Time Series Forecasting Results

It follows that the time series show extensive contrasts and permit us to dependably assess predictive models. The forecasting issue is to create the multi-step-ahead forecasts for the each month of 2014 (last year of data) utilizing the information from the past period for training. For hyperparameter selection, the models were trained and validated on data up to 2013.

In this work, classical statistical models in connection with preprocessing are proposed:

- **ARIMA** — $ARIMA(p, d, q)(P, D, Q)_{12}$ model used in function `auto.arima` in R environment (package `forecast`). This function uses automatic ARIMA modeling. It combines unit root tests, maximum likelihood estimation and minimization of the Akaike information criterion (AICc) to obtain the optimal ARIMA model [38].
- **ETS** — exponential smoothing state space model [4] used in function `ets` (R package `forecast`). This implementation uses many types of ETS models depends on the consideration of the trend, seasonal and error components. It can be expressed multiplicatively or additively, and the trend could be damped or not. Similar to the

case of `auto.arima`, ETS returns the optimal model estimated by the model parameters using AICc [38].

- Prophet — modular additive regression model with nonlinear trend and seasonal components [37] implemented in function Prophet in R environment (package prophet).

The proposed models were here compared with other computational intelligence models as well as classical statistical models. They include the following:

- k -NN— k nearest neighbor is an ML-type model designed for the medium-term load forecasting (MTLF) [13]. It learns from patterns defined by (8).
- N-WE—Nadaraya–Watson estimator [39] for MTLF using pattern representation (8).
- LSTM—Long short-term memory (LSTM) network developed for MEL forecasting in [8]. It works on patterns (8).
- MLP—Multilayer perceptron proposed in [6]. This model is developed for MEL forecasting. It learns from patterns defined by (8).
- ANFIS—adaptive neuro-fuzzy inference system proposed for MTLF in [40]. It works on patterns (8).
- SVM—Support vector machine proposed for MTLF in [41]. It works on patterns (8).
- ES-RNN—A hybrid method of exponential smoothing and recurrent neural networks [42–44].

The length of the x -patterns is the one of the main hyperparameters for k -NN, N-WE, LSTM, MLP, ANFIS, and SVM models. Despite the natural choice for this hyperparameter, which is equal to the seasonal cycle length, i.e., 12 for the monthly type of time series, the optimal value of n , in a range from 3 to 24, for each model and each time series was selected in leave-one-out procedure using historical data.

The parameters of the ARIMA, ETS, and Prophet models were selected in the optimization procedures implemented in the `auto.arima`, `ets`, and `prophet` functions, respectively. These functions ensure a fully automatic selection of model structure and parameters for each time series individually. The MLP model learned from x patterns. A separate MLP network was trained for each time series and each month of the forecast period (2014). A single-hidden layer network with sigmoid activation functions was used. The networks were trained using the Levenberg–Marquardt method with Bayesian regularization, which helped prevent overfitting. The number of hidden nodes was selected from 1 to 10, individually for each time series and each forecasted month. The adaptive-network-based fuzzy inference system, ANFIS, like MLP, learned from x patterns. A separate ANFIS model was trained for each time series and month of the forecast period. The initial parameters of the Gaussian membership functions in the premise parts of the rules were selected using fuzzy c-means clustering. ANFIS was trained with a hybrid method that uses a combination of the least squares method to estimate the consequent parameters and the backpropagation gradient descent method to select the premise parameters. The number of fuzzy rules M was selected from 2 to 13. SVM, like MLP and ANFIS, is learned from x patterns. For each time series and each month of the forecast period, a separate SVM model was trained with kernels in the form of a dot product $K(x_i, x_j) = x_i^T x_j$. The length of the input pattern was selected for each series. The remaining hyperparameters (BoxConstraint, KernelScale, and Epsilon) were selected in the automatic optimization procedure implemented in the `fitrsvm` function from the Statistics and Machine Learning Toolbox in the Matlab environment. LSTM networks in all variants were trained using the Adam optimization algorithm (adaptive moment estimation). The number of hidden units was selected for each time series individually from the set $\{1, 2, \dots, 10, 15, \dots, 50, 60, \dots, 200\}$. The remaining hyperparameters assumed default values: number of epochs, 250; initial value of the learning rate, 0.005; and a threshold value of the gradient (to prevent gradient explosion), 1. In the middle of the learning process, the learning rate was reduced to 0.001. The ETS+RD-LSTM model learned on all-time series simultaneously (cross-learning). The optimal values of the hyperparameters of this model for the set of 35-time series were as follows: number of epochs, 16; learning rate, 0.001; length of the cell and hidden state, 40; pinball loss, 0.4; regularization parameter, 50; and ensembling parameters, $L = 5$, $K = 3$, and $R = 3$.

In the case of the k -NN, MLP, ANFIS, SVM, and LSTM models, the optimal values of their hyperparameters were selected individually for each of the 35 time series. The selection of parameters was carried out separately for each of the variants of the models, V1, or (V2, and V3). The selection of hyperparameters was performed on the training set in the grid search procedure using cross-validation—the leave-one-out cross-validation method or by treating the last year of training data (2013) as validation data (SVM and LSTM models). The training set contained all terms of a given series from the historical period, up to and including 2013. The model with the optimal set of hyperparameter values was used to forecast the time series in the test period, which covered 12 months in 2014. Optimization procedures for ARIMA, ETS, and Prophet models are fully automatic, built into the functions that implement these models (`auto.arima`, `ets`, `prophet`). The ETS+RD-LSTM model was optimized on training data, treating the last year of training data (2013) as validation data.

The k -NN, N-WE, ARIMA and ETS are deterministic models, and they return the same results for the same data. NN-based models, for example, MLP, ANFIS, LSTM, and SVM, return various outcomes for the same data because of the stochastic type of the learning system. In this study, these models were run 100 times, and the final errors were averaged from 100 independent trials.

Taking into account the x -pattern encoding variants described in Section 3, three variants of each models (which also work with pattern representation usage) are considered:

- V1. The basic variant, where the coding variables for x -patterns are the mean and dispersion of previous sequence X_{i-1} for k -NN, N-WE, LSTM, MLP, ANFIS, and SVM. No patterns and coding variables are used for ES-RNN, ARIMA, ETS, LSTM, and Prophet in this case. We can use this variant to forecast the MEL from (8) without additional forecasting for coding variables.
- V2. The variant, for which the mean and dispersion of sequence X_i serve as the coding variables. Using ARIMA model, they are both independently forecasted for the query pattern based on their previous values. This variation broadens the denotations by “+AR”, e.g., “ k -NN + AR”, and “ANFIS + AR”.
- V3. The mean and dispersion of sequence X_i serve as the coding variables, as in variant V2. However, in this instance, they are predicted using ETS for the query pattern. “+ETS” is used to extend the denotations in this variant, such as “ k -NN + ETS”, and “ANFIS + ETS”.

The following measures are used to assess the quality of forecasts and forecasting models:

- Percentage error (PE):

$$PE = \frac{E - \hat{E}}{E} \cdot 100, \quad (18)$$

where E is the actual value and $\hat{E}$ is the forecasted value.

- Mean percentage error (MPE):

$$MPE = \frac{1}{N} \sum_{i=1}^N PE_i. \quad (19)$$

- Absolute percentage error (APE) :

$$APE = |PE|. \quad (20)$$

- Mean absolute percentage error (MAPE):

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{E_i - \hat{E}_i}{E_i} \right|. \quad (21)$$

- Interquartile range of absolute percentage error (IQR):

$$IQR(APE) = Q3(APE) - Q1(APE), \quad (22)$$

where $Q1(APE)$, $Q3(APE)$ are the lower and upper quartiles, respectively.

The quarter range allows you to assess the variability of the APE error. It includes 50% of all observations located centrally in the distribution.

- Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - \hat{E}_i)^2}. \quad (23)$$

- Standard deviation of percentage errors ($stdPE$):

$$Std(PE) = \sqrt{\frac{1}{N} \sum_{i=1}^N (PE_i - MPE)^2}. \quad (24)$$

- Coefficient of asymmetry (skewness) of the distribution of percentage errors ($skewPE$):

$$Skew(PE) = \frac{\frac{1}{N} \sum_{i=1}^N (PE_i - MPE)^3}{Std(PE)^3}, \quad (25)$$

The coefficient of asymmetry is zero for symmetric distributions, negative values for left asymmetric distributions (most of the population is below average) and positive for right asymmetric distributions (most of the population is above average).

- Kurtosis of percentage error distribution ($kuPE$):

$$ku(PE) = \frac{\frac{1}{N} \sum_{i=1}^N (PE_i - MPE)^4}{Std(PE)^4} - 3, \quad (26)$$

$kuPE$ is a measure of the clustering of PE errors around the mean value of MPE. The higher the kurtosis value, the more slender the distribution of errors and the greater the concentration of their values around the mean.

The median absolute percentage error (MdAPE), mean absolute percentage error (MAPE), interquartile range of absolute percentage error (APE) as a measure of the forecast dispersion, and root mean square error (RMSE) are all shown in Table 2. The RMSE error measure, as can be seen from this table, indicates that Prophet + ETS is the most accurate model when compared to its rivals. In cases of MdAPE and MAPE metrics, Prophet + ETS results are very close to the state-of-the-art model ETS-RNN + ETS, but the predictions are fully interpretable, and the model has a small number of parameters to estimate.

Table 2. Results comparison among proposed and comparative models.

Model	MdAPE	MAPE	IQR	RMSE
k-NN	3.11	5.19	4.17	385.68
k-NN + AR	2.88	4.71	4.25	352.42
k-NN + ETS	2.72	4.58	3.58	333.27
N-WE	2.84	5.00	3.97	352.01
N-WE + AR	2.85	4.59	3.95	340.26
N-WE + ETS	2.68	4.37	3.36	320.51
LSTM	3.73	6.11	4.50	431.83
LSTM + AR	3.43	5.28	4.79	392.47
LSTM + ETS	3.08	5.19	4.54	366.45
MLP	2.97	5.27	3.84	378.81
MLP + AR	3.12	4.83	4.26	362.03
MLP + ETS	3.11	4.80	4.12	358.07

Table 2. Cont.

Model	MdAPE	MAPE	IQR	RMSE
ANFIS	3.56	6.18	4.87	488.75
ANFIS + AR	3.66	6.05	5.07	473.80
ANFIS + ETS	3.54	6.32	4.26	464.29
SVM	2.80	5.41	3.97	382.60
SVM + AR	3.14	4.91	4.09	348.52
SVM + ETS	2.85	4.74	3.60	330.94
ETS-RNN	2.74	4.48	3.55	347.24
ETS-RNN + AR	2.58	4.23	3.47	332.74
ETS-RNN + ETS	2.64	4.09	3.13	314.01
ARIMA	3.32	5.65	5.24	463.07
ARIMA + AR	2.99	4.64	3.95	357.84
ARIMA + ETS	2.85	4.52	3.61	339.49
ETS	3.50	5.05	4.80	374.52
ETS + AR	2.94	4.50	3.68	345.37
ETS + ETS	2.76	4.30	3.19	326.94
Prophet	3.08	4.72	4.37	349.01
Prophet + AR	2.99	4.39	3.65	334.46
Prophet + ETS	2.68	4.15	3.44	311.56

Figure 9 provides more specific results, such as MAPE for each nation. It is important to note that Prophet + ETS is frequently one of the most precise models. The model rankings, based on MAPE and RMSE, are shown in Figure 10. They display the models' average positions in the country-by-country rankings. Take note of Prophet + ETS's top spot in both rankings.

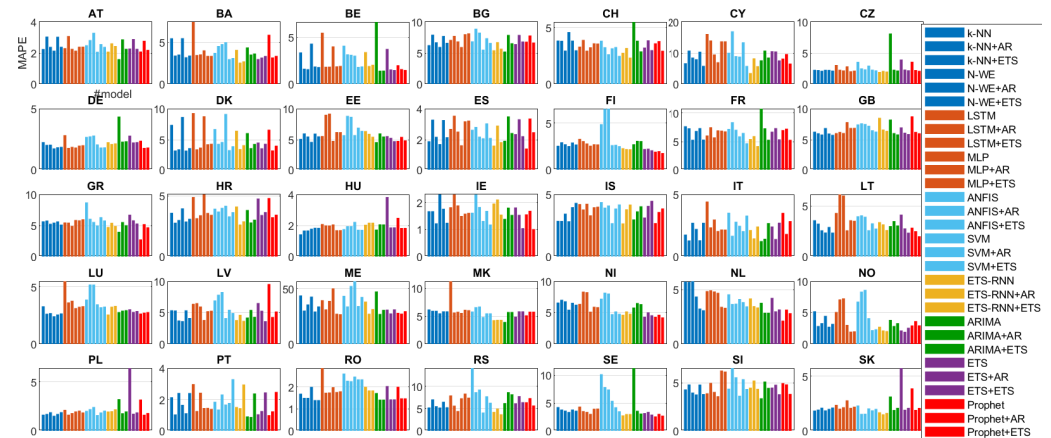


Figure 9. MAPE for each country.

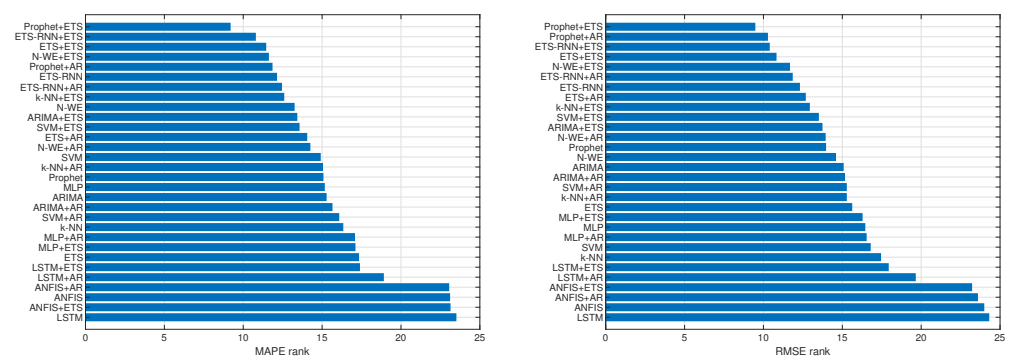


Figure 10. Rankings of the models.

In Table 3, the forecast errors and the descriptive statistics for the percentage errors are shown. The bias of the forecasts can be estimated from the average value of the PE error. For all models, the mPE is negative, which means that the forecast is overstated. The least loaded forecasts are made by the Prophet model (mPE = -0.70), and the most by the LSTM model (mPE = -3.12). The distribution of PE errors is described by the statistics: medPE, stdPE, skewPE, and kurtPE. The distribution of Prophet + ETS errors (medPE = 0.00) is the closest to zero, and the most distant LSTM (medPE = -1.81). The PE errors for the SVM model show the greatest dispersion around the mean (stdPE = 16.76), and the Prophet + AR errors the smallest (stdPE = 7.02). The most slender and centered around the mean value are the error distributions MLP + AR, MLP + ETS, ANFIS + AR, ARIMA + AR, ETS, ETS + AR, Prophet, Prophet + AR (kurtPE > 15), and the most flattened is error distribution MLP + ETS (kurtPE = 11.83). The negative skewness values, skewPE, which characterize the PE distributions of all models, indicate the left-hand skewness (the greater part of the population has values below the average). The most flattened distribution shows the greatest symmetry, obtained for ANFIS + ETS (skewPE = 0.96), and the smallest, the most slender distribution, obtained for SVM (skew (PE) < -13.40).

Table 3. Descriptive statistics of percentage errors among proposed and comparative models.

Model	mPE	medPE	stdPE	skewPE	kuPE
k-NN	−1.96	−1.27	10.83	−4.88	49.39
k-NN + AR	−1.76	−0.75	8.10	−2.66	20.96
k-NN + ETS	−1.26	−0.20	9.11	−4.47	38.22
N-WE	−1.91	−1.18	10.82	−5.41	48.94
N-WE + AR	−1.75	−0.85	7.82	−2.68	21.38
N-WE + ETS	−1.26	−0.17	8.68	−4.63	40.75
LSTM	−3.12	−1.81	9.49	−2.86	22.21
LSTM + AR	−1.86	−0.78	8.66	−2.75	21.20
LSTM + ETS	−1.41	−0.55	10.15	−5.35	50.04
MLP	−1.37	−0.68	11.88	−7.52	109.64
MLP + AR	−1.64	−0.92	7.45	−1.64	12.16
MLP + ETS	−1.71	−1.03	7.32	−1.55	11.83
ANFIS	−2.51	−1.43	11.37	−4.35	34.93
ANFIS + AR	−1.94	−0.65	9.63	−1.67	13.29
ANFIS + ETS	−1.30	−0.40	12.65	−0.96	39.37
SVM	−2.22	−0.91	16.76	−13.40	229.68
SVM + AR	−1.66	−0.61	8.58	−3.44	28.95
SVM + ETS	−1.28	−0.09	11.06	−7.84	95.91
ETS-RNN	−1.11	−0.27	10.07	−6.37	63.61
ETS-RNN + AR	−0.86	−0.20	7.30	−2.75	24.75
ETS-RNN + ETS	−0.32	0.47	8.48	−5.36	52.51
ARIMA	−2.35	−1.03	13.62	−9.01	119.20
ARIMA + AR	−1.69	−0.77	7.43	−1.65	12.67
ARIMA + ETS	−1.15	−0.26	8.40	−3.76	30.82
ETS	−1.04	−0.31	7.97	−1.89	13.52
ETS + AR	−1.71	−0.90	7.20	−1.73	13.02
ETS + ETS	−1.17	−0.28	8.20	−4.06	33.64
Prophet	−0.70	−0.18	7.47	−1.37	12.27
Prophet + AR	−1.30	−0.55	7.02	−1.66	12.84
Prophet + ETS	−0.76	0.00	7.86	−3.79	30.82

Figure 11 shows the MAPE errors broken down into individual months of the forecast period (2014). Attention should be paid to errors that are lower for months 8–10 and higher for months 1–4 and 12 [18]. The lowest errors were most often achieved by models with the forecast of code variables using ETS. Examples of forecasts generated by the models with the + ETS representation for several European countries are shown in Figure 12. The

forecast errors of the PL series, except for the ETS model, do not exceed 2%, which should be considered a very good result. DE, ES, and IT ranks are forecast with a slightly greater error. In the case of DE, there are greater deviations in the forecasts set by ANFIS + ETS for September. The GB series is projected well below the actual mileage. This is due to the unexpected increase in demand in the UK system in 2014, despite the downward trend observed in 2010–2013 (see Figure 5). The opposite situation for FR, ES, and IT resulted in a slight overestimation of forecasts.

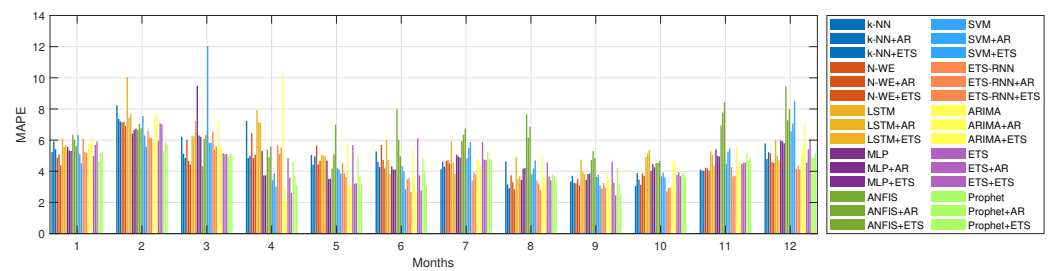


Figure 11. MAPE for each month.

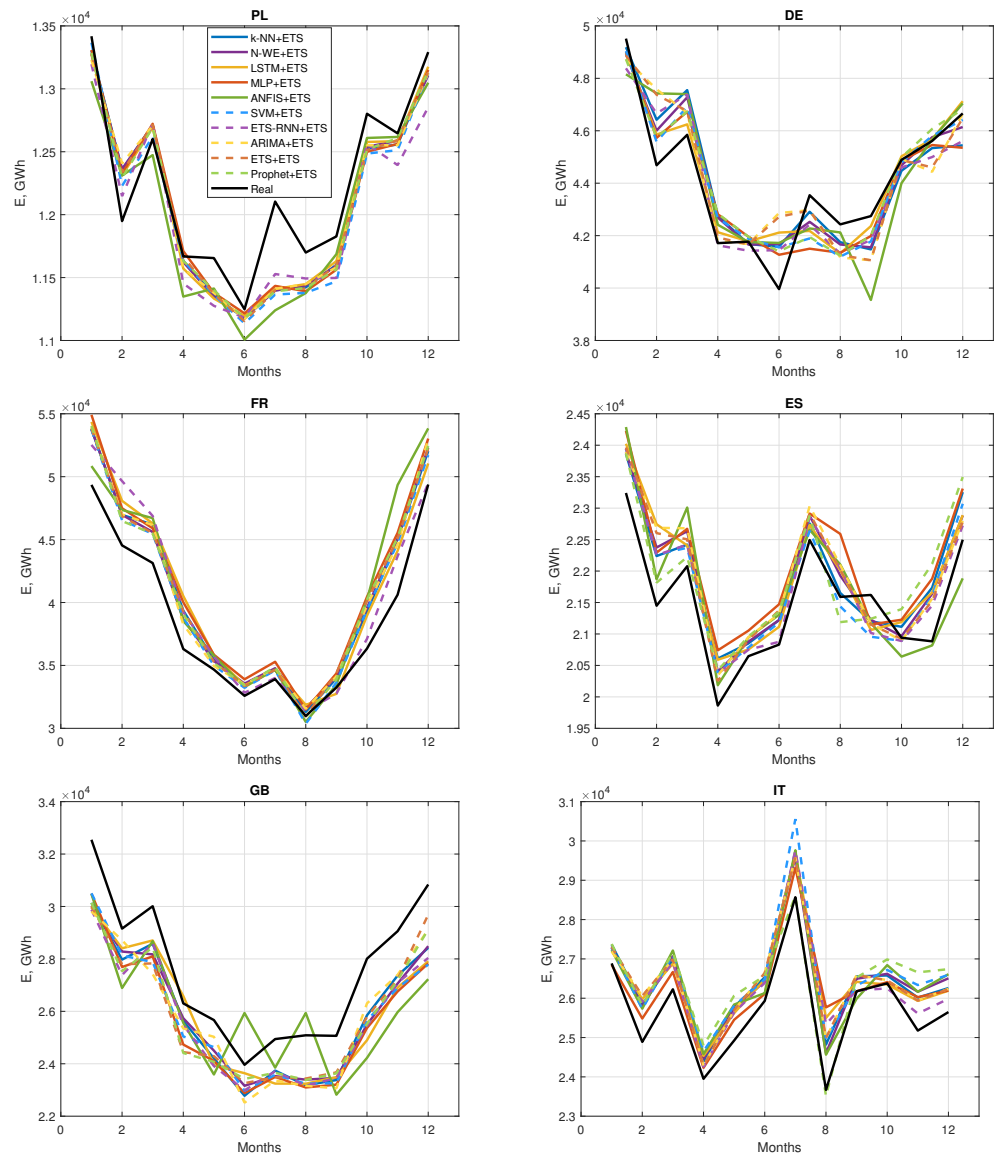


Figure 12. Examples of forecasts of MOD.

MAPE and RMSE results are presented in Figure 13. The time series of code variables and their forecasts for six sample countries are shown in Figure 14). Figure 15 shows Wilcoxon tests and the distribution of APE errors. The models were ranked from the lowest to the highest median (APE) value. The first two positions are occupied by the ETS-RNN models using the forecast of code variables. It should be noted that the proposed statistical models achieve higher positions in the case of using ETS to forecast code variables. ANFIS models take final positions to confirm the statistical significance of the differences in APE errors. Wilcoxon tests were performed for each pair of models. A white diagram element means that the models intersecting this element do not differ statistically in terms of APE. A yellow element means that the model pointed to by this element on the OY axis has reached a smaller error than the model indicated on the OX axis. A red element means that the model pointed to by this element on the OY axis has reached a greater error than the model pointed to on the OX axis. Figure 15 shows that models without prediction of code variables in many cases bring greater errors than hybrid models that contain predictions of these variables. The results of the Wilcoxon test for the ETS-RNN + ETS model, which is characterized by the lowest MAPE error (see Table 3), is similar to the results of the Prophet + ETS model. Both of these models show advantages over at least twenty-three other models (statistically significant difference in errors) and are as accurate as the other models. The ETS-RNN + ETS model shows the greatest advantage: in terms of error, the APE is more accurate than twenty-five other models.

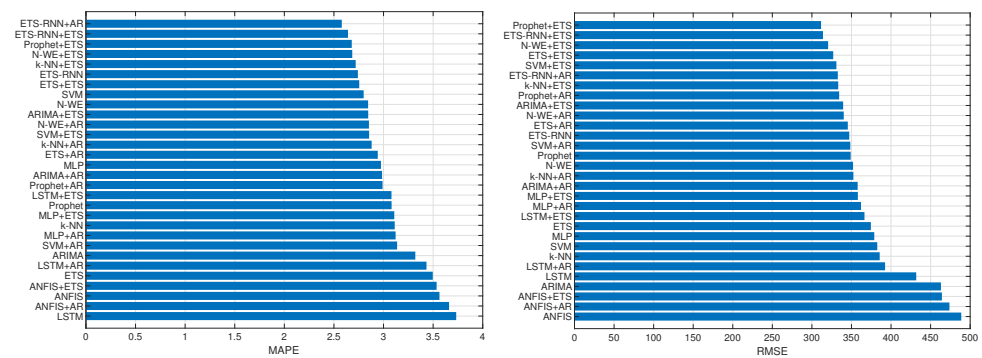


Figure 13. Errors of the models.

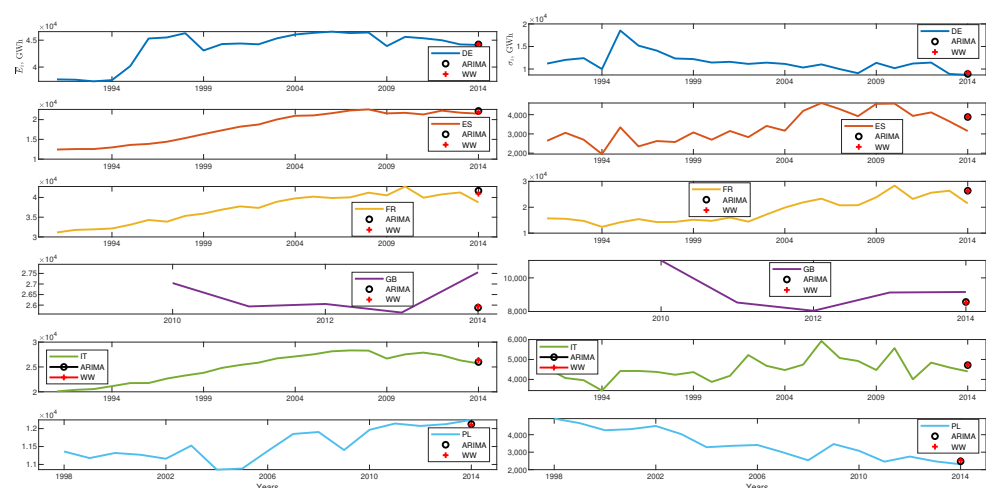


Figure 14. Forecasts of coding variables.

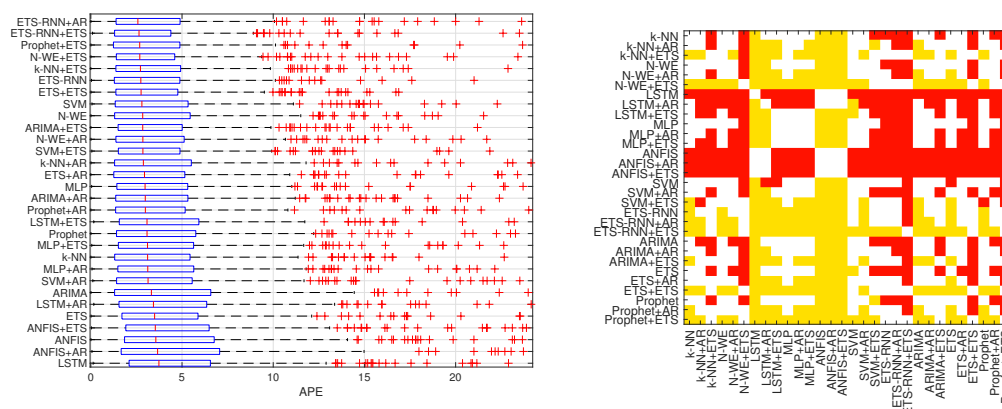


Figure 15. APE and Willcoxon test results for each model.

Summarizing the preliminary results, it should be noted that the performance of the prediction model is completely dependent on proper TS preprocessing. The introduction of initial normalization and predictive coding variables through ETS in the classic models significantly improves performance.

5. Discussion

In the ARIMA and ETS models, the optimization is global, i.e., their parameters are adjusted to ensure the lowest error for all terms of the time series (however, the length of the series can be restricted to the last terms so that the model takes into account only the specificity of the last historical period). The memory of the ARIMA model is limited to the last values of the time series (the size of the memory is determined by the rows of the AR and MA processes). Thus, the predicted value is “constructed” from the last terms. The ETS model is based on all values, but their influence on the predicted value decreases exponentially with time, i.e., more distant points have lower weights. Unlike these models, for example, k-NN constructs a local model individually for each query pattern. For the construction of the forecast, it uses all terms of the time series, not limited to the last period, as in ARIMA, and without introducing weights depending on the time distance, as in ETS. To construct a local regression function, k-NN considers values that may be distant in time from the query pattern if such distant cases are similar in shape to the query pattern. It is worth noting, however, that the time distance information can be easily introduced into, for example, N-WE by additional weights (decreasing in time) assigned to the x output patterns in the model. We can also enter weights based on seasonality. Outliers in the time series interfere with the selection of ARIMA and ETS parameters, leading to suboptimal models. In, for example, k-NN, as previously stated, outliers have a reduced impact in part. An additional distinction between statistical models and, for example, N-WE, is that the former generates forecasts one step ahead. Forecasts with longer horizons are achieved recursively, taking the the prediction for the preceding time step as an input for the prediction of the following time step. In contrast, for example, k-NN predicts the x -pattern, representing the entire predicted sequence, in one step.

6. Conclusions

In this paper, statistical methods for mid-term load forecasting were proposed. The input data for the models represent the normalized annual seasonal cycle of a load time series with filtered trend and unified variance. They express shapes of the yearly cycles. The proposed approach uses statistical methods for forecasting annual patterns and also for forecasting coding variables for decoding these patterns. Due to the pattern representation of the time series, forecasting models do not need to grasp the essence of the complex time series, which simplifies the predictive problem.

In the experimental part of the work, classical models with pattern representation were tested on MEL predicting issues for 35 European countries. The outcomes showed the advantages of the proposed approach over an alternative approach without forecasting coding variables. For statistical models and also for many comparative models, the proposed approach improved the accuracy. For example, in the cases of Prophet + ETS, ETS + ETS, and ARIMA + ETS usage, it leads to outperform predecessors Prophet, ETS, and ARIMA about 13.7%, 17.4%, and 25% in case of MAPE error. The proposed models can be further implemented for short-term load forecasting.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We use real-world data collected from www.entsoe.eu (accessed on 12 April 2016).

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning	ES	Spain
MEL	Monthly Energy Load	FI	Finland
MTLF	Mid-term Load Forecasting	FR	France
ARIMA	Autoregressive Integrated Moving Average	GB	Great Britain
ETS	Exponential Smoothing	GR	Greece
NN	Neural Network	HR	Croatia
ACF	Autocorrelation Function	HU	Hungary
PACF	Partial Autocorrelation Function	IE	Ireland
k-NN	k Nearest Neighbor	IS	Iceland
N-WE	Nadaraya–Watson estimator	IT	Italy
LSTM	Long Short-Term Memory	LT	Lithuania
MLP	Multilayer Perceptron	LU	Luxembourg
ANFIS	Adaptive Neuro-Fuzzy Inference System	LV	Latvia
SVM	Support Vector Machine	ME	Montenegro
ETS-RNN	Exponential Smoothing and Recurrent Neural Networks	MK	Macedonia
AT	Austria	NI	Northern Ireland
BA	Bosnia and Herzegovina	NL	Netherlands
BE	Belgium	NO	Norway
BG	Bulgaria	PL	Poland
CH	Switzerland	PT	Portugal
CY	Cyprus	RO	Romania
CZ	Czech Republic	RS	Serbia
DE	Germany	SE	Sweden
DK	Denmark	SI	Slovenia
EE	Estonia	SK	Slovakia

References

1. Pełka, P. Przegląd Metod Wykorzystywanych Do Średnioterminowego Prognozowania Obciążeń Elektroenergetycznych. *Przegląd Elektrotechniczny* **2019**, *95*, 155–159. [\[CrossRef\]](#)
2. Dogan, E. Are Shocks to Electricity Consumption Transitory or Permanent? Sub-National Evidence from Turkey. *Utilities Policy* **2016**, *41*, 77–84. [\[CrossRef\]](#)
3. Bunnoon, P.; Chalermyanont, K.; Limsakul, C. Mid Term Load Forecasting of the Country Using Statistical Methodology: Case Study in Thailand. In Proceedings of the 2009 International Conference on Signal Processing Systems, Singapore, 15–17 May 2009; pp. 924–928.
4. Hyndman, R.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.

5. González-Romera, E.; Jaramillo-Morán, M.A.; Carmona-Fernández, D. Monthly Electric Energy Demand Forecasting with Neural Networks and Fourier Series. *Energy Convers. Manag.* **2008**, *49*, 3135–3142. [\[CrossRef\]](#)
6. Pełka, P.; Dudek, G. Pattern-Based Forecasting Monthly Electricity Demand Using Multilayer Perceptron. In *Artificial Intelligence and Soft Computing*; Springer: Cham, Switzerland, 2019; pp. 663–672.
7. Gao, T.; Niu, D.; Ji, Z.; Sun, L. Mid-Term Electricity Demand Forecasting Using Improved Variational Mode Decomposition and Extreme Learning Machine Optimized by Sparrow Search Algorithm. *Energy* **2022**, *E261*, 125328. [\[CrossRef\]](#)
8. Pełka, P.; Dudek, G. Pattern-Based Long Short-Term Memory for Mid-Term Electrical Load Forecasting. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
9. Peng, L.; Wang, L.; Xia, D.; Gao, Q. Effective Energy Consumption Forecasting Using Empirical Wavelet Transform and Long Short-Term Memory. *Energy* **2022**, *238*, 121756. [\[CrossRef\]](#)
10. Sharma, K.; Dwivedi, Y.K.; Metri, B. Incorporating Causality in Energy Consumption Forecasting Using Deep Neural Networks. *Ann. Oper.* **2022**, 1–36. [\[CrossRef\]](#)
11. Li, J.; Wei, S.; Dai, W. Combination of Manifold Learning and Deep Learning Algorithms for Mid-Term Electrical Load Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [\[CrossRef\]](#)
12. Xu, A.; Tian, M.-W.; Firouzi, B.; Alattas, K.A.; Mohammadzadeh, A.; Ghaderpour, E. A New Deep Learning Restricted Boltzmann Machine for Energy Consumption Forecasting. *Sustainability* **2022**, *14*, 10081. [\[CrossRef\]](#)
13. Dudek, G.; Pełka, P. Prognozowanie Miesięcznego Zapotrzebowania Na Energię Elektryczną Metodą k Najbliższych Sąsiadów. *Przegląd Elektrotechniczny* **2017**, *1*, 64–67. [\[CrossRef\]](#)
14. Pełka, P.; Dudek, G. Prediction of Monthly Electric Energy Consumption Using Pattern-Based Fuzzy Nearest Neighbour Regression. *ITM Web Conf.* **2017**, *15*, 02005. [\[CrossRef\]](#)
15. Pełka, P.; Dudek, G. Medium-Term Electric Energy Demand Forecasting Using Generalized Regression Neural Network. In *International Conference on Information Systems Architecture and Technology*; Springer: Cham, Switzerland, 2018; pp. 218–227.
16. Pełka, P.; Dudek, G. Ensemble Forecasting of Monthly Electricity Demand Using Pattern Similarity-Based Methods. In *International Conference on Artificial Intelligence and Soft Computing*; Springer: Cham, Switzerland, 2020; pp. 712–723.
17. Castán-Lascorz, M.A.; Jiménez-Herrera, P.; Troncoso, A.; Asencio-Cortés, G. A New Hybrid Method for Predicting Univariate and Multivariate Time Series Based on Pattern Forecasting. *Inf. Sci.* **2022**, *586*, 611–627. [\[CrossRef\]](#)
18. Dudek, G.; Pełka, P. Pattern Similarity-Based Machine Learning Methods for Mid-Term Load Forecasting: A Comparative Study. *Appl. Soft Comput.* **2021**, *104*, 107223. [\[CrossRef\]](#)
19. Papalexopoulos, A.D.; Hesterberg, T.C.: A regression-based approach to short-term system load forecasting. *IEEE Trans. Power Syst.* **1990**, *5*, 1535–1547. [\[CrossRef\]](#)
20. Wang, Y.; Gan, D.; Sun, M.; Zhang, N.; Lu, Z.; Kang, C. Probabilistic Individual Load Forecasting Using Pinball Loss Guided LSTM. *Appl. Energy* **2019**, *235*, 10–20. [\[CrossRef\]](#)
21. Miseta, T.; Fodor, A.; Vathy-Fogarassy, Á. Energy Trading Strategy for Storage-Based Renewable Power Plants. *Energy* **2022**, *250*, 123788. [\[CrossRef\]](#)
22. Al-Hamadi, H.M.; Soliman, S.A. Long-Term/Mid-Term Electric Load Forecasting Based on Short-Term Correlation and Annual Growth. *Electr. Power Syst. Res.* **2005**, *74*, 353–361. [\[CrossRef\]](#)
23. Barakat, E.H. Modeling of Nonstationary Time-Series Data. Part II. Dynamic Periodic Trends. *Int. J. Electr. Power Energy Syst.* **2001**, *23*, 63–68. [\[CrossRef\]](#)
24. Cai, G.; Yang, D.; Jiao, Y.; Pan, C. The Characteristic Analysis and Forecasting of Mid-Long Term Load Based on Spatial Autoregressive Model. In Proceedings of the 2009 International Conference on Sustainable Power Generation and Supply, Nanjing, China, 6–7 April 2009; pp. 1–5.
25. de Oliveira, E.M.; Oliveira, F.L.C. Forecasting Mid-Long Term Electric Energy Consumption through Bagging ARIMA and Exponential Smoothing Methods. *Energy* **2018**, *144*, 776–788. [\[CrossRef\]](#)
26. Zhang, D.-L.; Yan, J.; Wang, W.-H.; Yang, X.-L. Mid-Long Term Load Forecasting of the Unstable Growth Sequence Based on Markov Chains Screening Combination Forecasting Models. In Proceedings of the 2016 China International Conference on Electricity Distribution (CICED), Xi'an, China, 10–13 August 2016; pp. 1–5.
27. Barakat, E.E.; Al-Rashed, S.A. Long Range Peak Demand Forecasting under Conditions of High Growth. *IEEE Trans. Power Syst.* **1992**, *7*, 1483–1486. [\[CrossRef\]](#)
28. Zagdański, A.; Suchwałko, A. *Analiza i Prognozowanie Szeregów Czasowych: Praktyczne Wprowadzenie Na Podstawie Środowiska R*; Wydawnictwo Naukowe PWN: Warsaw, Poland, 2016.
29. Józwiak, J.; Podgórski, J. *Statystyka Od Podstaw*; Wyd. PWE: Warszawa, Poland, 2001.
30. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2013; ISBN 978-0-9875071-0-5.
31. Pełka, P. Nonparametric Machine Learning Methods for Mid-Term Forecasting of Power System Loads. Ph.D. Thesis, Czestochowa University of Technology, Czestochowa, Poland, 2021.
32. Chatfield, C. *The Analysis of Time Series: An Introduction*, 6th ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2003; ISBN 978-1-58488-317-3.
33. Dittmann, P. *Prognozowanie w Przedsiębiorstwie*; Oficyna Ekonomiczna: Kraków, Poland, 2004.
34. Dudek, G. Wybrane Metody Analizy Szeregów Czasowych Obciążeń Elektroenergetycznych. In Proceedings of the Materiały VII Konferencji Naukowej Prognozowanie w elektroenergetyce, Czestochowa, Poland, 26–27 September 2004; pp. 116–125.

35. Dudek, G. Pattern Similarity-Based Methods for Short-Term Load Forecasting—Part 1: Principles. *Appl. Soft Comput.* **2015**, *37*, 277–287. [[CrossRef](#)]
36. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The Forecast Package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [[CrossRef](#)]
37. Taylor, S.J.; Letham, B. Forecasting at Scale. *Am. Stat.* **2018**, *72*, 37–45. [[CrossRef](#)]
38. Hyndman, R.J.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O'Hara-Wild, M.; Petropoulos, F.; Razbash, S.; Wang, E.; Yasmeen, F. *Forecast: Forecasting Functions for Time Series and Linear Models*; R Package: Melbourne, Australia, 2018.
39. Dudek, G.; Pełka, P. Medium-Term Electric Energy Demand Forecasting Using Nadaraya-Watson Estimator. In Proceedings of the 2017 18th International Scientific Conference on Electric Power Engineering (EPE), Kouty nad Desnou, Czech Republic, 17–19 May 2017; pp. 1–6.
40. Pełka, P.; Dudek, G. Neuro-Fuzzy System for Medium-Term Electric Energy Demand Forecasting. In *International Conference on Information Systems Architecture and Technology*; Springer: Cham, Switzerland, 2017; pp. 38–47.
41. Pełka, P. Pattern-Based Forecasting of Monthly Electricity Demand Using Support Vector Machine. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021.
42. Smyl, S. A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [[CrossRef](#)]
43. Dudek, G.; Pełka, P.; Smyl, S.: A Hybrid Residual Dilated LSTM and Exponential Smoothing Model for Midterm Electric Load Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 2879–2891. [[CrossRef](#)] [[PubMed](#)]
44. Dudek, G.; Pełka, P.; Smyl, S. 3ETS+ RD-LSTM: A New Hybrid Model for Electrical Energy Consumption Forecasting. In Proceedings of the International Conference on Neural Information Processing, Bangkok, Thailand, 23–27 November 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 519–531.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.