METHODS OF SHORT TERM ELECTRICITY DEMAND FORECASTING

G. TAGANOVA, 1 J. TUSSUPOV, 2 A. ABDILDAYEVA, 1 S. SERIKBAYEVA, 3 ZH. SADIRMEKOVA, 1 G. AZIEVA, 3 D. MAMATAYEVA, 3 M. TOLGANBAYEVA

1 L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
2 Institute of information and computational technologies CSMES RK Almaty, Kazakhstan
3 Taraz Regional University named after M.KH. Dulaty, Taraz, Kazakhstan

E-mail: 1 guldana.kileuzhanova@gmail.com, 2 tussupov@mail.ru, 3 abass_81@mail.ru, 4 information_informatik@mail.ru, 5 janna_1988@mail.ru, 6 gulmira_azieva@mail.ru, 7 dilnoza.abdrakhmanova@mail.ru, 8 tolganbaeva86@mail.ru

ABSTRACT

This article discusses the short-term forecasting of electricity consumption. Methods of smoothing the daily schedule of power consumption are considered. The possibility of using one of the methods of smoothing power consumption was analyzed in Python. The proposed method is applicable for the subjects of REM in order to approximate the retrospective data of electricity consumption. The relevance of the work is due to the demand of the subjects of the wholesale electricity and capacity market (REM) for ways to build short-term forecasts of electricity consumption in order to improve the quality and accuracy of the predictive model. From the conducted research, it was revealed that the adaptive Holt-Winters smoothing method is optimal for making short-term forecasting "for the day ahead".

Keywords: Forecasting; Power Consumption; Data Analysis, Python.

1. INTRODUCTION

Currently, the planning of electric power modes of operation of power systems (PS) is one of the main tasks that allows for the continuous maintenance of the balance of electric power in the Unified Energy System (UES) of the Republic of Kazakhstan. Improving the accuracy of the planning of operating modes ensures the reliability and safe functioning of both the power systems of the subjects of Kazakhstan and the UES as a whole. The planning of operating modes is based on the prediction of the expected power consumption [1,2].

In the last decade, research in the field of forecasting electricity consumption by industrial, municipal and energy distribution enterprises, residential complexes, business structures and individual houses has intensified [3-5]. Most researchers define weather conditions as the main factors determining the dynamics of electricity demand. These include: indicators of temperature (air, environment and indoor temperature), indicators of humidity, pressure, wind speed and direction, clouds and sun brightness; precipitation [8]. Among additional independent factors, the authors use variables in models for electrical load, heat transfer or heat index; calendar variables; indicators of the size and operational characteristics of buildings, urban infrastructure development; indicators of living standards and socio-economic development [8]. For example, the authors [4] use data on average daily energy consumption in kW as a dependent variable for forecasting electricity demand in the residential sector. To display calendar effects, the researchers include dummy variables, namely a variable for all Saturdays, a variable for all Sundays, and a variable for holidays in the study interval [4]. It should be noted that the frequency of time series used in models is determined by the source and availability of data. So, in [5], time series of electricity consumption are given; in the study [9], half-hour data with an annual time interval are given. Accordingly, the forecasts obtained from such a sample may be short-term, for example, for a week. Real-time forecasting requires data from measuring instruments every minute or every second.
To date, the development of power consumption forecasting systems is carried out by research and design and survey institutes, such as "KazNIPIEnergoprom", JSC "Kazselenergoproekt Institute", JSC "KazNII Energetiki named after academician Chokin", JSC "Kazakhstan Institute of Industry Development" [3].


Despite a large number of scientific publications in this field, there are still unresolved problems associated with improving the accuracy of forecasting power consumption models in which the accounting of influencing factors is implemented. Natural illumination, speed, air temperature and wind direction affect the increase in the accuracy of forecasting power consumption [5].

In this research, a large number of works devoted to the analysis, modeling and forecasting of time series of electricity consumption were studied. There are a large number of classification schemes of methods for predicting processes in the literature. For example, in the analytical review [1], the time series of power consumption is described by sinusoidal functions separately for night and day cycles of power consumption. The minimum average approximation error was 3.37%. In [2], the experiment showed that the average error in the annual interval was 3.72%.

Within the framework of the structural approach, since the 1980s, dynamic linear and nonlinear Baess models using the Kalman filter have successfully proved themselves [9-11]. The substantiation of the optimal model in terms of statistical characteristics and predictive qualities requires the determination of the main components of the series, the nature of its stationarity, specification, parameterization, verification and approbation of models, testing of fictitious variables in order to improve their qualities.

2. MATERIALS AND METHODS

Modern time series forecasting methods are built mainly on the principle of historical prediction of the future. The peculiarity of the energy consumption indicators is the presence of multidirectional trends, structural breaks, cyclical and seasonal fluctuations, makes certain requirements for the selection of appropriate models and methods. The study focuses on classical time series techniques: autoregressions and exponential smoothing models [19].

The extrapolation methods past information to the future are constantly being improved in terms of interpretation, complexity and forecast accuracy. In the recent decades scholars’ attention has shifted from structural models based on the system of equations and restrictions on parameters to special “ad hoc” models that are not theoretically justified. Although statistical techniques based on non-linear least squares (NLNS), Gauss least squares (OLS) and the maximum likelihood (MLE) estimation are highly used, technologies’ innovations forced active development of machine learning forecasting methods. K-Nearest Neighbor regression (KNN), Classification And Regression Trees (CART), Bayesian Neural Network (BNN), Generalized Regression Neural Networks (GRNN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM) demonstrated good experimental results. There are studies report better model fitting but worse forecasting accuracy of these methods comparing to
statistical models. The researchers state the need for improvement and further development of machine learning models in terms of their better interpretability and specification of the uncertainty around the point forecasts [23].

2.1 Autoregressive Approach

Autoregressive moving average (ARMA) or Autoregressive integrated moving average (ARIMA) models - one of the most widely used classical time series techniques, that apply the Box-Jenkins methodology. These models predict time series’ future values based on a linear combination of its previous values and disturbances. The ARIMA model with parameters \( p \) (the autoregressive order or the lag of the model), \( d \) (the integration or differencing order), \( q \) (the moving average order) fit an equation [20]:

\[
\Delta^{d}y_{t} = c + \varphi_{1}\Delta^{d}y_{t-1} + \ldots + \varphi_{p}\Delta^{d}y_{t-p} + \theta_{1}\epsilon_{t-1} + \ldots + \theta_{q}\epsilon_{t-q} + \epsilon_{t} \tag{1}
\]

Here, \( y_{t} \), represent the actual time series values in time period \( t \); \( \Delta^{d} = (y_{t-1} - y_{t})^{d} \) is the difference operator of the \( d \)th order, applied to remove a stochastic trend; \( \varphi_{1},\ldots,p \)

\( \theta_{1},\ldots,q \) are the parameters of the model; \( \epsilon_{t} \) is an error term that is assumed to be a stationary Gaussian white-noise process with mean zero and constant variance \( \sigma^{2} \) [20]. Model (1) can be rewritten using backshift lag operator (L) notation as:

\[
\phi(L)(1-L)^{d}y_{t} = c + \theta(L)\epsilon_{t} \tag{2}
\]

A special case of model (1) is the Seasonal autoregressive integrated moving average model SARIMA \((p, d, q)x(P, D, Q)s\) [20]:

\[
\Phi(Ls)\phi(L)\Delta^{d}s\Delta y_{t} = \Theta_{0} + \Theta(Ls)\theta(L)\epsilon_{t}, \tag{3}
\]

where \( s \) is the seasonal length - the number of periods in a season \((s=12\) for monthly series); \( L \) is the lag operator; \( \Delta^{d}_{s} \) is the seasonal difference operator.

An iterative modeling approach implies assessing stationarity and seasonality patterns; identification of the model parameters and their estimation with maximum likelihood or non-linear least squares methods; checking adequacy and prediction accuracy of the model [20].

Specification of the ARMA/ARIMA/SARIMA models is commonly facilitated by the graphical analysis of the correlograms (the autocorrelation function, ACF, and partial autocorrelation function, PACF) of the original and differenced series. Selection the optimal model parameters \((p, d, q), (P, D, Q)\) is justified by minimization of the information criteria. The Hyndman-Khandakar algorithm automates this procedure with the function auto.arima of the “forecast” R package [20].

To eliminate the problem of unreliable MLE parameter estimation and to reveal unobservable state of the series frequently the Kalman filter algorithm is used for ARIMA state-space models [20].

In the presence of the consistent change in the variance over time, the Autoregressive model of conditional heteroscedasticity (ARCH) or Generalized autoregressive conditional heteroscedasticity model of (GARCH) are appropriate. The models predict the future conditional and unconditional variance presuming the stationarity of the series (no trend or seasonal component) [20]:

\[
\epsilon_{t} = \sigma_{t}\epsilon_{t} \tag{4}
\]

Here the error term \( \epsilon_{t} \) accounts for a stochastic white-noise process \( Z_{t} \), and a time-dependent standard deviation, \( \sigma_{t} \).

For ARCH(q) the squared innovations \( \sigma_{t}^{2} \) are modeled as:

\[
\sigma_{t}^{2} = \alpha_{0} + \alpha_{1}\epsilon_{t-1}^{2} + \ldots + \alpha_{q}\epsilon_{t-q}^{2} \tag{5}
\]

where \( \alpha_{0} > 0 \) and \( \alpha_{i} \geq 0, i > 0 \) for all \( t \).

For GARCH(p, q) the series \( \sigma_{t}^{2} \) is modeled as:

\[
\sigma_{t}^{2} = k + \gamma_{1}\sigma_{t-1}^{2} + \ldots + \gamma_{p}\sigma_{t-p}^{2} + \alpha_{1}\epsilon_{t-1}^{2} + \ldots + \alpha_{q}\epsilon_{t-q}^{2} \tag{6}
\]

Here \( p \) and \( q \) are nonnegative integers, representing the number of lagged conditional variances and the lagged squared innovations, respectively.

GARCH models have numerous applications in financial time series analysis. The ARIMA/SARIMAX models fit energy consumption series better due to relatively stable dynamics and seasonal characteristics.

Despite the active development of machine learning models, autoregressive methods (ARMA/ARIMA/SARIMA, dynamic regression models, vector autoregressions, VAR, and
cointegration models, VEC) are still widely used to predict the electric energy consumption.

2.2 Exponential Smoothing Approach

Exponential smoothing is a powerful time series forecasting method for univariate data, frequently used as an alternative to autoregressive approach. This framework has multiple applications in different fields of studies due to its flexibility, reliability of the forecasts and low expenses. Proposed in the late 1950s [25] this approach has motivated some of the most successful forecasting methods.

The taxonomy of seasonality exponential smoothing models differs depending on the trend and nature. The simple exponential smoothing model applicable for data with no clear trend or seasonality produces forecasts as weighted averages of past observations, decaying exponentially depending on the timing of observations [20]:

\[
y_{t-1} | t = \alpha y_t + (1-\alpha) y_{t-1} + (1-\alpha)^2 y_{t-2} + \ldots, \quad (7)
\]

where \(0 \leq \alpha \leq 1\) is the smoothing parameter.

Holt-Winters additive and multiplicative models suggested improvement of the model (16) to account for trend and seasonal patterns [20]. The more advanced state space exponential smoothing models with additive or multiplicative errors contain a measurement equation that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonal) change over time [21]. One of the most successful recent advancement in exponential smoothing state space models refers to TBATS model with Box-Cox transformation, ARMA error, trend and representation of seasonal components by Fourier series. This approach produces high accuracy forecasts handling multiple nested and non-nested seasonality. Although it requires extra calculation time, especially for big time series data.

The general representation of TBATS model (8) includes level (9), trend (10), seasonal (11) and ARMA error term (12) equations:

\[
y(w)_t = l_t + \phi b_t + \sum_{i=1}^{T} s(i)_t + d_t \quad (8)
\]

\[
l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (9)
\]

\[
b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t \quad (10)
\]

\[
s(i)_t = s((i), t-1) \cos(\lambda(i)j + s(i), t-1) \sin(\lambda(i)j + \gamma d_{t}) \quad (11)
\]

\[
d_t = \sum_{i=1}^{P} \phi_{i} d_{i,t} + \sum_{i=1}^{Q} \theta_i e_{i,t} + \epsilon_t \quad (12)
\]

Here \(y(w)_t\) is Box-Cox transformed observations at time \(t\) with the parameter \(\omega\); \(l_t\) is the local level at time \(t\); \(b_t\) is the long-run trend; \(b_t\) is the short-term trend at time \(t\); \(s(i)_t\) is \(i\) th seasonal component of the series at time \(t\); \(d_t\) is an ARMA(p, q) error process; \(\epsilon_t\) is the Gaussian white-noise process with zero mean and constant variance; \(\alpha, \beta, \gamma\) are smoothing parameters; \(\phi\) is the damped parameter; \(s((i), t-1)\) is the stochastic level; \(k_j\) is the number of harmonics for the \(i\) th seasonal component, \(\lambda(i)j = 2\pi j/m_i\) where \(m_i\) is period of the \(i\) th seasonal cycles [22-25].

3. RESULTS

The data (power consumption of Nur-Sultan, Republic of Kazakhstan) used in this study are daily based and they refer to the period from 1 October 2019 to 31 December 2019. Basic statistics about power consumption are listed in Table 1. According to the figure 1, it is possible to notice the daily trend, which presents higher values in working days of the week.

Table 1. Main power consumption statistics

<table>
<thead>
<tr>
<th>Day</th>
<th>Power consumption kW</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.10.2019</td>
<td>17 347 706</td>
</tr>
<tr>
<td>02.10.2019</td>
<td>17 193 388</td>
</tr>
<tr>
<td>03.10.2019</td>
<td>18 554 473</td>
</tr>
<tr>
<td>04.10.2019</td>
<td>18 418 246</td>
</tr>
<tr>
<td>05.10.2019</td>
<td>18 469 078</td>
</tr>
</tbody>
</table>

Considering the graph of power consumption shown in Figure 1, the authors decided first of all to determine the moving average (13). In the moving average, the future value of the variable depends on the average of its previous values. A forecast based on the last observed day (24 hours).
It is easy to make a forecast for a stationary series, since its future statistical characteristics will not differ from the observed current ones. Most time series forecasting solutions manage to successfully model and predict these characteristics (for example, the expectation or variance), so in the case of non-stationarity of the original series, the predictions will turn out to be incorrect. A stationary time series can be transformed from most time series through white noise (Figure 3).

The process generated by the standard normal distribution, stationary, oscillates around zero with a deviation of 1. Now, based on this observation, we can generate a new process in which each subsequent value will depend on the previous one:

$$x_t = px_{t-1} + e_t$$  \hspace{1cm} (14)

[1] Another method for checking the stationarity of data is V. Fuller and D. Dickey and its improved form of the extended Dickey-Fuller test (ADF). This technique consists in testing the statistical hypothesis about the presence of a single root (with an alternative hypothesis about the presence of a root less than one). If the t-statistic is less than the critical values of the ADF statistic, then the null hypothesis is rejected, which indicates the stationarity of the series. In the case of the presence of unit roots, the series is considered to be integrated k-th order I (k) and requires differentiation to bring it to stationarity [24].

The figure 4 shows exactly the same stationary white noise that was built earlier. At next stage in the figure 5, the p value increased to 0.6, as a result of which wider cycles began to appear on the graph, but in general, it has not yet ceased to be stationary.

Fig. 2. Moving Average Value

Fig. 3. White Noise

Fig. 4. Dickey-Fuller, Rh=0

Fig. 5. Dickey-Fuller, Rh=0.6

Fig. 6. Dickey-Fuller, Rh=0.9

Fig. 7. Dickey-Fuller, Rh=1
The figure 6 deviates more and more from the zero average value, but still fluctuates around it. Finally, in the figure 7 a value of \( p \) equal to one gave a random walk process — the series is not stationary\[14-16\].

This is due to the fact that when the critical unit is reached, the series \( x_t = px_{t-1} \) stops returning to its average value. If we subtract \( x_t, x_{t-1} = (p - 1)x_{t-1} + e_t \), where the expression on the left is the first differences. If \( p = 1 \), then the first differences will give a stationary white noise \( e_t \). This fact formed the basis of the Dickey-Fuller test for the stationarity of a series (the presence of a single root). If it is possible to obtain a stationary one from a non-stationary series by the first differences, then it is called an integrated one of the first order. The null hypothesis of the test—the series is not stationary, was rejected on the first three graphs, and started on the last one. It is worth saying that the first differences are not always enough to obtain a stationary series, since the process can be integrated with a higher order (have several unit roots), to check such cases, an extended Dickey-Fuller test is used, which checks several lags at once.

If seasonality is detected, it is necessary to eliminate it or include fictitious seasonality variables in order to assess the stationarity of the data and further correct construction of the model. Common smoothing methods are: exponential and adaptive smoothing methods, additive and multiplicative Holt-Winters models \[17-19\].

When using the exponential smoothing method, a new representation of a data series is performed according to the rule:

\[
S_t = \gamma_t, S_t = \alpha \gamma_t + (1 - \alpha) S_{t-1}, t = 1, T
\]  

(14)

where \( S_t \) is the new value of the row level; \( \gamma_t \) is the original value of the row level; \( \alpha \) is the smoothing constant. This method is advisable to use in the case when the data has a slowly growing or horizontal trend.

The adaptive smoothing method allows you to change the smoothing constant during the calculation, for which the scheme is used:

\[
S_{t+1} = \alpha \gamma_t + (1 - \alpha) S_t
\]  

(15)

Here \( \alpha \) changes in time according to the rule:

\[
\alpha_t = \frac{E_t}{M_t}, E_t = \beta (y_t - \hat{y}_t) + (1 - \beta) E_{t-1},
M_t = \beta |y_t - \hat{y}_t| + (1 - \beta) M_{t-1}, \beta \in (0; 1).
\]  

(16)

A more advanced modification of exponential smoothing is the additive Holt-Winters model, based on the use of smoothed data, a trend component and a seasonality index. Smoothing in this case occurs according to the scheme:

\[
S_{t+p} = \alpha_t + b_t p + c_{t+p}
\]  

(17)

where \( b_t \) is the trend parameter; \( p = 1, 2 \ldots \) - the number of forecast periods, \( c_t \) the seasonality.
parameter. The components $\alpha$, $b$, $c$ are calculated using the formulas:

\[
\alpha_i = \alpha(y_i - c_{i-s}) + (1 - \alpha)(\alpha_{i-1} + b_{i-1}),\\
b_i = \beta(\alpha_i - \alpha_{i-1}) + (1 - \beta)b_{i-1},\\
c_i = \gamma(y_i - \alpha_i) + (1 - \gamma)c_{i-s},\quad 0 \leq \alpha, \beta, \gamma \leq 1.
\] (18)

Fig. 10. Additive Holt-Winters Model

Here $s$ - is the number of seasonality cycles; $\alpha$, $\beta$, $\gamma$ - are the smoothing parameters, respectively, for the level of the series, trend and seasonality.

The idea of this method is to add one more, third, component-seasonality. Accordingly, the method is applicable only if the series is not deprived of this seasonality, which is true in our case. The seasonal component in the model will explain the repeated fluctuations around the level and trend, and it will be characterized by the length of the season — the period after which the repeated fluctuations begin. For each observation in the season, its own component is formed, for example, if the length of the season is 7 (for example, weekly seasonality), then we get 7 seasonal components, one for each of the days of the week.

The level now depends on the current value of the series minus the corresponding seasonal component, the trend remains unchanged, and the seasonal component depends on the current value of the series minus the level and on the previous value of the component. At the same time, the components are smoothed through all available seasons, for example, if this is the component responsible for Monday, then it will be averaged only with other Mondays.

### 4. CONCLUSIONS

The article contains an analytical review of the autoregressive and Exponential smoothing approach. A comparative assessment of the modeling methods used to predict the demand for electricity is considered. It follows from the study that it is advisable to use the adaptive Holt-Winters smoothing method, where alpha and beta take the value of 0.9. This method of a time series of power consumption can be used when making short-term forecasting "for the day ahead".

### REFERENCES


[16] Kostin N. S.; Gritsay A. S. Expansion of the Rapidminer analytical platform for building predictive models of OREM.


