

Research Article

Maksat Kalimoldayev, Aleksey Drozdenko, Igor Kopylyk, T. Marinich, Assel Abdildayeva*, and Tamara Zhukabayeva

Analysis of modern approaches for the prediction of electric energy consumption

<https://doi.org/10.1515/eng-2020-0028>

Received May 05, 2019; accepted Jan 09, 2020

Abstract: A review of modern methods of forming a mathematical model of power systems and the development of an intelligent information system for monitoring electricity consumption. The main disadvantages and advantages of the existing modeling approaches, as well as their applicability to the energy systems of Ukraine and Kazakhstan, are identified. The main factors that affect the dynamics of energy consumption are identified. A list of the main tasks that need to be implemented in order to develop algorithms for predicting electricity demand for various objects, industries and levels has been developed.

Keywords: prediction, power consumption, panel models, autoregression models, neural networks

1 Introduction

Creation of innovative intellectual systems for managing energy consumption processes is a vital task for individual objects (institutions), countries, and for the global econ-

omy as a whole. Solving such urgent problems as reducing energy consumption, ensuring energy independence, reducing greenhouse gas emissions requires identifying adequate methods for analyzing, modeling and forecasting time series of consumption and production of various types of energy, their integration with existing information systems for making management decisions across individual enterprises, cities, industries and states. The lack of development of theoretical and methodological approaches and practical aspects of the use of forecasting systems and evaluating the efficiency of electricity use in Kazakhstan and Ukraine actualize the need to create integrated automated energy management systems using modern methods of machine learning.

The purpose of this work is to compare modern methods of analysis, modeling and forecasting the consumption of electric energy at the national, sectoral and individual (by facilities) levels, as well as to study the experience of their use in various countries and industries.

In this paper we have used classical statistical “ad hoc” models, advanced ensemble methods and neural networks to predict electric power demand with a case of the wholesale energy transmission company.

The rest of the paper is organized as follows: Section 2, contain Literature review, Section 3 discuss the Comparative analysis of the methods and models, Section 4 contain Application and Results, the Conclusion of the study is provided in Section 5.

2 Literature review

The ubiquity of modern technological devices for measuring the amount of energy consumed has contributed to the development of engineering and statistical analysis methods, which make it possible to effectively plan, predict and monitor the growing load on the power grid.

Over the past decade, research has intensified in the area of forecasting electricity consumption for industrial, municipal, and energy distributing enterprises, housing complexes, business structures, and individual houses [1–

***Corresponding Author: Assel Abdildayeva:** Institute of Information and Computational Technologies of the Cabinet of Sciences of the Ministry of Education and Science of Kazakhstan, Department of Information Technologies, L.N. Gumilyov Eurasian National University, 010008, Astana, Kazakhstan; Email: abass_81@mail.ru

Maksat Kalimoldayev: Institute of Information and Computational Technologies of the Cabinet of Sciences of the Ministry of Education and Science of Kazakhstan, Kazakhstan; Email: mnk@ipic.kz

Aleksey Drozdenko: Sumy State University, Ukraine; Email: aleksey.drozdenko@gmail.com

Igor Kopylyk: Sumy State University, Ukraine; Email: igor.kopl@gmail.com

T. Marinich: Sumy State University, Ukraine; Email: email@email.edu

Tamara Zhukabayeva: Institute of Information and Computational Technologies of the Cabinet of Sciences of the Ministry of Education and Science of Kazakhstan, Department of Information Technologies, L.N. Gumilyov Eurasian National University, 010008, Astana, Kazakhstan; Email: amara_kokenovna@mail.ru

5]. This is due to the need to ensure the energy efficiency of buildings, recognized by the International Energy Agency (The International Energy Agency) as one of the five conditions that reduce the final energy consumption and associated CO₂ emissions [6]. Environmental prerequisites and economic feasibility contributed to the development of national energy-efficient design rules for various types of buildings, which gave impetus to the development of computer software for energy-efficient design of new homes, such as EnergyPlus, DOE-2, eQUEST, IES, ECOTECT, etc. [7].

Maintaining energy efficiency in buildings requires continuous monitoring of energy consumption indicators and identifying factors that affect them in real time. Most researchers identify weather conditions as the main factors determining the dynamics of demand for electricity. These include: temperature indicators (air, environment, dry lamps, dew point, wet point, room temperature); indicators of humidity, pressure, wind speed and direction, cloudiness and brightness of the sun; precipitation [8]. Among additional independent factors, the authors use variables of electrical load, heat transfer, or thermal index in models; calendar variables; size indicators and operational characteristics of buildings, urban infrastructure development; indicators of living standards and socioeconomic development [8].

For example, to predict the demand for electricity in the residential sector of Chile [4], the authors use data on average daily energy consumption in kW as the dependent variable, variables of average daily temperature in Celsius, and daily value per unit account of Chile as explanatory variables. To display calendar effects, 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 the time series used in the models is determined by the source and availability of data.

So, in the work [5] the hourly rows of electricity consumption are presented, in the study [3] - half-hour data with an annual time interval. Accordingly, the forecasts obtained on such a sample can only be short-term, for example, for a week. To obtain medium-term and long-term forecasts, models estimated using higher frequency data (for example, monthly [9]) and a longer time interval (several decades) are used. Real-time forecasting requires the acquisition of data from instrumentation by the minute or by the second.

An analysis of open statistical information on electricity consumption in Ukraine and Kazakhstan [10, 11] shows that the statistics on gross electricity consumption by all sectors of the economy are available only by year; Indi-

cators of final consumption, taking into account renewable energy sources in the context of households, industrial sectors, transport, services, agriculture, forestry and fisheries, as well as non-energy energy consumption, have been available only since 2007. You can get monthly reports from relevant ministries [12] indicators of gross energy consumption in the country, and only within the last decade.

A comparative analysis of methodological approaches to the calculation of the energy security indicator revealed a number of weaknesses in the national systems for assessing energy security as part of the country's national security. In particular, the shortcomings of the approach to calculating the level of energy security of Ukraine [13] are identified. These include: the limited range of aspects of energy security for which the assessment is carried out, the lack of a base for comparison and a long series of statistical data on energy security indicators, a slow update of the threshold values of indicators embedded in the rationing algorithm. In addition to the domestic approach, an analysis was conducted of methods for assessing the energy security risk index developed by the United States Institute of Energy and the International Energy Agency [14]; a comparative analysis of these methods with the domestic approach. According to the results of the analysis, differences were found in the rationing of individual indicators, the quality characteristics of individual indicators and the method for determining the respective weights for each indicator. It was proposed to include indicators such as market volatility, energy intensity indicators, the state of global and regional fuel stocks, etc., into the list of indicators of the country's energy security. To solve the problem of modeling real statistical data represented by different frequencies, it was proposed to use mixed frequency models, Mixed-Data Sampling Models (MIDAS) [15], to determine the relationship between possible energy security factors and the energy efficiency of the national economy.

One of the solutions to the problem of a small sample of data to obtain adequate statistically significant results and qualitative forecasts can be the use of panel models that evaluate similar indicators for a group of objects, for example, all educational institutions in the region, regions of the country or countries with similar development parameters.

Thus, article [16] used a panel sample of annual data on the consumption of electricity by residential buildings in the context of Chinese cities to identify the most significant factors of the construction of "green houses". The authors of [17] examine the demand for electricity in the industrial and service sectors of Taiwan, analyzing panel

data for 23 industrial sectors and 9 service sectors for the period 1998-2015.

Article [18] assesses the efficiency of electricity consumption for an unbalanced group of 27 countries in transition and 6 OECD member countries in Europe from 1994 to 2007. Thus, it can be concluded that for countries such as Kazakhstan and Ukraine, models based on panel data are the most acceptable.

At the same time, the focus of scientific research in these countries should be biased towards modeling electricity demand by individual objects that have the appropriate equipment to measure high-frequency fixation electricity consumption, followed by extrapolation of the results to higher levels (industry, regional).

The above approach is presented in detail in the work of Canadian scientists [16], who identified two methods for modeling the demand for electricity in the residential sector: "top down" and "bottom up".

The first approach focuses on identifying key factors and forecasting electricity consumption by housing objects of different levels depending on historical housing data and top-level variables, which include macroeconomic indicators (gross domestic product, unemployment rates and inflation), prices for various types of energy, climatic factors.

The second approach is based on the use of statistical and engineering methodologies for predicting electricity consumption at the regional and national levels by extrapolating the indicators of a representative set of individual houses [19].

It should be noted that engineering models that describe final energy consumption as a natural phenomenon, based on physical laws and do not require historical data on energy consumption, are now practically not used. The rapid increase in the sources and volumes of data, their processing technologies and processing system capacities contributed to the shift of scientific interests towards statistical methods.

The variety of statistical models is due both to differences in the data structure (linear and nonlinear; discrete and continuous models), as well as the development of machine learning methods and software tools that implement them. Parametric and non-parametric methods that can be classified into regression, autoregression methods, Fourier models, neural networks, models of fuzzy logic, Wavelet analysis, Bayesian methods are widely used.

The use of parametric methods implies the availability of information on the nature of the distribution of data, which is fraught with the receipt of biased estimates of parameters and false conclusions in the case of an incorrectly chosen model. For those cases where the present distribu-

tion of data is unknown, the use of non-parametric methods is preferred. A significant drawback and limitation of non-parametric models, focused more on testing hypotheses than on estimating parameters, is the complexity of their calculations and high requirements for software and hardware [4].

3 The comparative analysis of the methods and models

Modern time series forecasting methods are based mainly on the principle of historical prediction of the future. The peculiarity of the energy consumption indicators is the presence of multidirectional trends, seasonal and cyclical fluctuations, structural breaks, makes certain requirements for the selection of appropriate methods and models. This paper represents a comparative study of approaches that can be used to make reliable predictions of energy consumption on macro, micro and sectoral levels, as well as to reveal significant predictors and causal relationships for policy conclusions. Additional interest point is model selection for energy consumption time series of different data frequency. The study focuses on classical time series techniques (autoregressions, exponential smoothing models, dynamic regressions), ensemble models and neural networks, capable to handle non-stationarity, heteroscedasticity, serial correlation of non-stable short-term data.

The methods for extrapolation past information to the future are constantly being improved in terms of complexity, interpretation and forecast accuracy. In the last 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 the Gauss least squares (OLS), non-linear least squares (NNLS) and the maximum likelihood (MLE) estimation are highly used, technologies' innovations forced active development of machine learning forecasting methods. Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Generalized Regression Neural Networks (GRNN), K-Nearest Neighbor regression (KNN), Classification And Regression Trees (CART), Support Vector Machine (SVM) demonstrated good experimental results [20]. Still, numerous studies [1, 4, 7, 20] report better model fitting but worse forecasting accuracy of these methods comparing to statistical models. The researchers [20] 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.

3.1 Autoregressive approach

One of the most widely used classical “ad hoc” time series techniques is Autoregressive moving average (ARMA) or Autoregressive integrated moving average (ARIMA) models that apply the Box-Jenkins methodology [21]. 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:

$$\Delta^d y_t = c + \varphi_1 \Delta^d y_{t-1} + \dots + \varphi_p \Delta^d y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

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; $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ are the parameters of the model; ε_t is a error term that is assumed to be a stationary Gaussian white-noise process with mean zero and constant variance σ^2 [21]. Model (1) can be rewritten using backshift lag operator (L) notation as:

$$\phi(L)(1-L)^D y_t = c + \theta(L)\varepsilon_t \quad (2)$$

A special case of model (1) is the Seasonal autoregressive integrated moving average model SARIMA (p, d, q) \times (P, D, Q) s [21]:

$$\Phi(L^s)\phi(L)\Delta^d \Delta_s^D y_t = \theta_0 + \theta(L^s)\theta(L)\varepsilon_t, \quad (3)$$

where s is the seasonal length – the number of periods in a season ($s=12$ for monthly series); L is the lag operator; Δ_s^D 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 [22].

A common technique to assess the stationarity of the series is the Augmented Dickey-Fuller (ADF) test. It estimates the model (4) to test the null hypothesis of a unit root against the alternative of stationarity [23]:

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \quad (4)$$

where α is a constant, β the coefficient of a simple time trend, ρ is the parameter of interest, Δ is the first difference operator, δ_i are parameters and p the lag order of the autoregressive process. 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 [21]. Selection the optimal model parameters (p, d, q), (P, D, Q) is justified by minimization of the information criteria (see Appendix 1). The Hyndman-Khandakar algorithm automates this procedure with the function *auto.arima* of the “forecast” R package [22].

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 [24].

In the presence of the consistent change in the variance over time, the Autoregressive model of conditional heteroscedasticity (ARCH) [25] or Generalized uutoregressive conditional heteroscedasticity model of (GARCH) are appropriate [26]. The models predict the future conditional and unconditional variance presuming the stationarity of the series (no trend or seasonal component) [26]:

$$\varepsilon_t = \sigma_t z_t \quad (5)$$

Here the error term ε_t accounts for a stochastic white-noise process z_t , and a time-dependent standard deviation, σ_t .

For ARCH(q) the squared innovations σ_t^2 are modeled as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2, \quad (6)$$

where $\alpha_0 > 0$ and $\alpha_i \geq 0, i > 0$ for all t .

For GARCH(p, q) the series σ_t^2 is modeled as:

$$\sigma_t^2 = k + \gamma_1 \sigma_{t-1}^2 + \dots + \gamma_p \sigma_{t-p}^2 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2, \quad (7)$$

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.

The researchers emphasize the improved forecast accuracy of the SARIMAX models [1, 4, 22], which assesses not only historical energy consumption data, but additional exogenous variables as well. Thus, considering holidays’ and weather effects, changes in the law, market situation and demographics, may explain significant data

variation giving more reliable predictions. Dynamic regression models that include external variables and allow the model errors to contain autocorrelation describing them as ARIMA process showed good results as well [22, 27].

A common way to account for causality of nonstationary series, make structural inference and policy conclusions is to use vector autoregressive (VAR) models and structural vector autoregressive (SVAR) models. They treat simultaneous sets of variables equally, regressing each endogenous variable on its own lags and the lags of all other variables in a finite-order system [28]. The basic p -lag VAR has the form:

$$Y_t = c + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \varepsilon_t, \quad t = 1, \dots, T, \quad (8)$$

where $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ is an $(n \times 1)$ vector of time series variables; Π_i are $(n \times n)$ coefficient matrices; ε_t is an $(n \times 1)$ unobservable zero mean white noise independent vector process with time invariant covariance matrix Σ .

The simplicity of estimation and interpretation of VAR/SVAR models with impulse response functions and forecast error vector decompositions made them a good alternative to structural models. The authors [29] used VAR approach to empirically prove the existence of bi-directional causality between electricity consumption and GDP in Russia.

The drawbacks of VAR approach in terms of explaining the long-term dynamics of the series is successfully realized by the Vector error correction models, VECM, used to describe the cointegration relationships between the variables. The basic VECM form a relationship [30]:

$$\begin{aligned} \Delta Y_t &= \Phi D_t + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t \quad (9) \\ \Pi &= \Pi_1 + \dots + \Pi_p - I_n \\ \Gamma_k &= - \sum_{j=k+1}^p \Pi_j, \quad k = 1, \dots, p-1, \end{aligned}$$

where ΔY_t and its lags are differenced $I(0)$ series; D_t is a deterministic term; ΠY_{t-1} contains the cointegrating relations.

Authors [31] employed Johansen cointegration to determine the long run relationship between energy consumption and its determinants for different sectors and to forecast future energy demand using scenario analysis.

Taking into consideration deep theoretical development, outstanding empirical results, simplicity and feasibility of justification and deployment, autoregressive models are highly recommended for use in experimental studies. It is important to mention that vector autoregressive and cointegration models are suitable mostly for macroeconomic analysis of energy consumption by sectors, regions and sources. ARIMA/SARIMAX models

3.2 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 [32] this approach has motivated some of the most successful forecasting methods.

The taxonomy of exponential smoothing models differs depending on the trend and seasonality 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 [22]:

$$\hat{y}_{t+1|t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots, \quad (10)$$

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 [22]. 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 [22]. 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 [33]. 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 (11) includes level (12), trend (13), seasonal (14) and ARMA error term (15) equations:

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_t^{(i)} + d_t \quad (11)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (12)$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t \quad (13)$$

$$s_t^{(i)} = s_{j, t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j, t-1}^{(i)} \sin \lambda_j^{(i)} + \gamma_i d_t \quad (14)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (15)$$

Here $y_t^{(\omega)}$ is Box-Cox transformed observations at time t with the parameter ω ; l_t is the local level at time t ; b is the long-run trend; b_t is the short-term trend at time t ; seasonal periods; $s_t^{(i)}$ is i^{th} seasonal component of the series at time t ; d_t is an ARMA(p, q) error process; ε_t is the Gaussian white-noise process with zero mean and constant variance; α, β, γ_i are smoothing parameters; ϕ is damped parameter; $s_{j, t-1}^{(i)}$ is the stochastic level; k_i is the number of harmonics for the i^{th} seasonal component, $\lambda_j^{(i)} = 2\pi j/m_i$ where m_i is period of the i^{th} seasonal cycles [33].

Papers [7, 34, 35] verified excellent forecast accuracy characteristics and opportunities for long-term forecast of electric energy demand using TBATS and based on it hybrid models.

Other useful models based on time series smoothing and decomposition are Seasonal and Trend decomposition using Loess (STL) and Multiple Seasonal Decomposition (MSTL). They use local regression nonlinear smoothing algorithm (Loess) for parameter estimation [22].

3.3 Machine learning methods

Artificial intelligence methods are becoming increasingly popular in the scientific and business environment [20]. There are numerous applications of machine learning methods in forecasting energy consumption and demand [5, 7, 8, 17, 19, 35].

Deep learning with artificial neural networks (ANN) are widely used and discussed nowadays. A significant advantage of ANN models is their ability to model non-linear relationships not restricting on the stationarity of the parameters. Its shortcoming refers to the requirement of big data sample for training and complexities with interpretation of the "black box" output.

Neural network is organized in form of layers having the predictors' or inputs' bottom layers, the forecasts' or outputs' top layer and intermediate layers containing "hidden neurons" [22]. Frequently used nonlinear autoregressive neural networks model NNAR(p, P, k) $_m$ [36] can be described by the following equation:

$$y'_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-m}, y_{t-2m}, y_{t-3m} + \varepsilon_t, \quad (16)$$

where p, P represent lagged autoregressive and seasonal inputs respectively, k – nodes in the hidden layer, m – the number of seasonal periods' inputs

Another ANN model that showed outstanding forecasting abilities is Multi-Layer Perceptron (MLP), where each layer of nodes receives inputs from the previous lay-

ers. The matrix notation of the MLP model is:

$$f(x) = G(b^{(2)} + W^{(2)}(s(b^{(1)} + W^{(1)}x))) \quad (17)$$

$$h(x) = \Phi(x) = s(b^{(1)} + W^{(1)}x) \quad (18)$$

$$o(x) = G(b^{(2)} + W^{(2)}h(x)), \quad (19)$$

Here $b^{(1)}, b^{(2)}$ are the bias vectors; $W^{(1)}, W^{(2)}$ are weight matrices connecting the input vector to the hidden layer; G, s – activation functions; $h(x)$ forms the hidden layer; $o(x)$ is the output vector.

The proposed MLP approach [37] was used to classify residential buildings according to their energy consumption and make corresponding hourly predictions for high and low power consumption buildings.

To sum up, neural networks models often provide an ideal approximation of actual and predicted data within a training sample, but in the case of insufficient training data, give large forecast errors. A variety of methods are used to improve predictive qualities of ANNs, including cross-validation, noise reduction, error regularization, error-reversal method, optimized approximation, SVM algorithms [22].

Currently, scientists are offering a range of hybrid models that are based on two or more traditional machine learning techniques or artificial intelligence methods [7, 19, 35]. Traditional methods for predicting time series, such as ANN and ARIMA, are complemented by optimization methods – Particle Swarm Optimization Algorithm (PSO), genetic algorithm, ant colony genetic algorithm etc. For instance, in paper [8] the authors introduced the hybrid model that combines the ARIMA model to identify periodicity, seasonality and linearity with an evolutionary algorithm (EA) for efficiently determining and optimizing residuals. Researchers [35] developed a hybrid model based on the TBATS and neural networks algorithms to forecast the electricity load demand.

Ensemble methods build a model by training several relatively simple base models (also known as weak learners) and then combine them to create a more predictive model. The most well known ensemble learning algorithms use bootstrap aggregation, also known as bagging (Breiman 1996); random forests (Breiman 2001); extremely randomized trees, also called extra trees (Geurts et al. 2006); and boosting (Schapire 1990). The bagging, extra trees, and random forests are based on a simple averaging of the base learner, while the boosting algorithms are built upon a constructive iterative strategy.

The recent advances in machine learning methods refer to ensemble methods that combine several low accuracy base models ("weak learners") are used to create a higher quality predictive model ("strong learner"). The

most popular ensemble learning algorithms are bootstrap aggregation (also known as bagging); random forests; extremely randomized trees (also known as extra trees), and boosting [38]. The first three methods are based on a simple averaging of the base models, while boosting methods apply iterative optimization algorithms based on decision trees and minimization of the loss function [39]. Boosting algorithms like Gradient Boosting [39], XGBoost, AdaBoost, Gentle Boost are frequently demonstrate state-of-the-art results at Kaggle and other machine learning competitions [40]. The improvement of prediction accuracy by the gradient boosting machine model comparing to piecewise linear regression and to a random forest algorithm have been proved in [38] on the example of energy consumption of commercial buildings.

4 Application and Results

In this section we use the hourly energy consumption data (2012 – 2017) of the US wholesale transmission organization [41] to test the prediction accuracy and deployment features of the statistical and machine learning methods described in Section 3. Visual inspection of the electricity consumption time series point on the possible sources of data variation, including weather, holidays, daily, weekly and monthly periodicity (Figure 1).

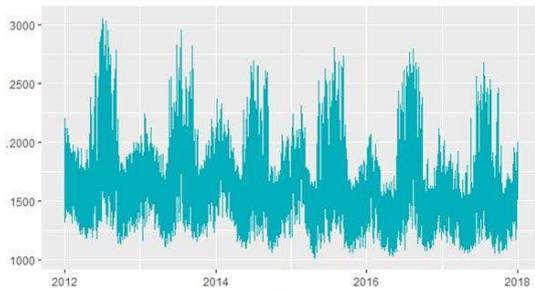


Figure 1: Hourly electricity consumption in MW, 2012-2018.

It is expected that the electricity consumption should be lower during weekends and nights, and may be higher at holidays and in summer and winter months. To account for possible multiple seasonality different exogenous variables are considered: outside air temperature, time of the week, hour of the day. The models are implemented in R programming language using ‘Forecast’ [42], ‘Segmented’ [43], ‘XGBoost’ [44], ‘rnn’ [45] R packages. To check the

prediction accuracy of the models the dataset is split into training (49660 observations) and test (2944 observations) data sets.

The study compares results for univariate and multivariate models. Predictions, based on the univariate series of electricity consumption, include:

1. Autoregressive Integrated Moving Average models (ARIMA);
2. Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components (TBATS);
3. Multiple Seasonal Decomposition model (MSTL);
4. Dynamic Harmonic Regression with ARMA error term.

For multivariate time series analysis we estimate ARIMA, piecewise linear regression that include additional input variables. To improve the prediction accuracy of the energy consumption modeling we used the gradient boosting [44] and neural network [45] machine learning algorithms.

Tables 1-2 demonstrate the model and forecast accuracy for training and test sets.

The energy consumption models are ranked by the forecast accuracy in this order:

1. Recurrent multilayer perceptron network (RMLP) with four selected input features.
2. Gradient boosting tree model with input variables (date, hour, day of week, month, quarter, year, day of week, day of month, week of year).
3. Piecewise linear regression with exogenous variables (air temperature – used for segmentation, day of the week, hour of the day).
4. Nonparametric multiple seasonal decomposition model (MSTL) with daily, weekly and yearly seasonality.
5. Autoregressive integrated moving average models (ARIMA) with exogenous variables (air temperature, day of the week, hour of the day).

The ARIMA model fits the following equation (with standard error given in brackets):

$$\begin{aligned}
 y_t = & \frac{1.802y_{t-1}}{(0.003)} - \frac{0.879y_{t-2}}{(0.003)} - \frac{1.168\varepsilon_{t-1}}{(0.007)} \\
 & + \frac{0.213\varepsilon_{t-2}}{(0.007)} \\
 & + \frac{0.611 \cdot \text{Temperature}}{(0.148)} + \frac{0.188 \cdot \text{day_of_week}}{(0.285)} \\
 & + \frac{0.762 \cdot \text{hour_of_day}}{(0.031)} + \varepsilon_t
 \end{aligned} \quad (20)$$

Table 1: Model accuracy for training data sets.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	AIC
Baseline μ_y mean value	-8.5e-14	297.20	232.17	-3.18	14.46	4.62	0.97	NA
TBATS	6.85e-03	41.03	25.31	-0.04	1.58	0.50	0.043	147954.9
MSTL	-0.0036	28.67	18.34	-0.02	1.16	0.36	0.086	NA
Dynamic harmonic regression	0.32	43.34	28.11	-0.039	1.75	0.56	0.03	NA
ARIMA (2,1,2)	1.97e-04	35.76	26.49	-0.028	1.62	0.55	0.018	430150.7
ARIMA (2,1,2) with Xreg	3.65e-04	35.49	26.22	-0.026	1.61	0.54	0.021	429599.4
Piecewise linear regression	-1.62e-7	101.48	62.6	-0.10	3.91	1.23	0.11	553966.8

Table 2: Forecast accuracy for test data sets

	ME	RMSE	MAE	MPE	MAPE	MASE
Baseline mean value μ_y	-8.9e+01	323.14	268.81	-9.77	18.45	5.35
TBATS	4.3e+02	515.81	440.46	28.16	28.97	8.76
Dynamic harmonic regression	236.13	361.49	274.90	12.99	16.05	5.47
ARIMA (2,1,2) with Xreg	-1.1e+02	252.90	199.98	-10.27	14.57	4.15
MSTL	96.88	278.27	206.52	3.91	12.38	4.11
Piecewise linear regression	-112.31	164.84	135.55	-7.92	9.31	3.09
Gradient Boosting	70.04	158.46	130.31	-7.61	8.95	2.97
Neural network	27.38	61.97	50.96	-2.97	3.5	0.99

In model (20) the exogenous variable *day_of_week* turns to be insignificant. According to the Hyndman-Khandakar algorithm [22] the optimal ARIMA model doesn't include any seasonal parameters. At the same time variables *Temperature* and *hour_of_day* significantly influence the hourly energy consumption.

Dependency of the energy consumption on the year seasons (rise in summer and winter months and slowdown in the rest of the year) explain the choice of the piecewise linear regression. Effect of temperature on electricity consumption is presented in Figure 2.

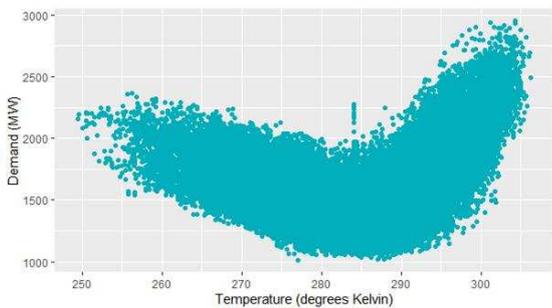


Figure 2: Temperature effect on the energy consumption

The estimated temperature breakpoint by segmentation algorithm [43] is 287.74 degrees of Kelvin. The piecewise model explains 74.42% of electricity consumption variation. The *Temperature* coefficients of the piecewise linear model are given in (21), the full model is given in Appendix 2.

$$y = \begin{cases} 6289.88 - 17.45Temperature \\ (36.08) (0.129) \\ 6289.88 - 71.60 \times 287.74 \\ +(287.74 - 17.45)Temperature \\ (36.08) (0.32) \end{cases} \quad (21)$$

for $x \leq 287.74$
for $x > 287.74$

Gradient boosting [44] and neural network machine learning algorithms used in the study are based on the extraction of the influential data features for models' training. The main features of the energy consumption time series are hour, day of week, month, quarter, year, day of week, day of month, week of year. Feature importance according to the gradient boosting tree algorithm is presented in Figure 3.

Detailed analysis of the errors of the gradient boosting model revealed the worst prediction accuracy for holidays. Inclusion of holidays' dummy variable (takes 1 if the day

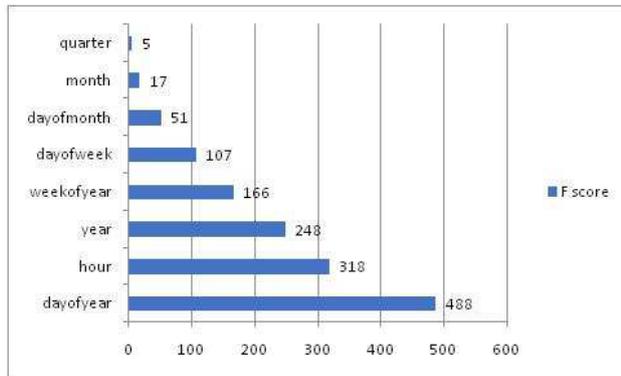


Figure 3: Gradient boosting: feature importance.

is a holiday and 0 otherwise) helped to improve the forecast accuracy MAPE to 5.45%. Estimated neural networks model showed the best forecast accuracy using two layers and four input variables that form the most important features of the energy consumption series. At the same time empirical analysis revealed deterioration of the forecast accuracy of the estimated machine learning methods in favor of TBATS and ARIMA model with exogenous variables.

5 Conclusions

The paper contains analytical review of theoretical and practical issues of effective energy management system based on the analysis of internal (technical, economic, structural, regime) and external (meteorological, environmental, energy, macroeconomic) factors. A comparative assessment of modeling techniques used to forecast electricity demand is considered. Two areas of research have been identified: forecasting electricity consumption based on panel data (by countries; regions; sectors, industries) and by individual objects that have the appropriate equipment to measure high-frequency consumption.

The findings point on the evolving shift from classical regression models to machine learning algorithms. Classical statistical techniques are still used but mostly in terms of hybrid models designed to reduce the model error or eliminate the existing assumptions for parameter estimation. In this respect, exponential smoothing model TBATS, seasonal trend decomposition model STL and seasonal autoregressive model SARIMAX form the top list of statistical techniques according to publications' review and empirical assessments.

The empirical analysis proves the extreme importance of clean high-frequency long statistics for high accuracy forecasts of energy consumption. Verification of signifi-

cant independent variables that explain variation of energy consumption is found to be another factor that improves the quality of predictions, especially for short data samples.

The increasing popularity of machine learning methods, and gradient boosting and neural networks in particular, is their ability to extract features from the series and include them in the models without specifying the parameters, as is the case with standard statistical algorithms. The empirical study proved their superiority in terms of forecast accuracy, especially for long samples. Besides these models are less prone to overfitting and let the user to include non-significant variables and parameters without the loss in the predictability of the model [38]. The empirical model evaluation in RStudio Integrated Development Software revealed problems associated with huge computation time undertaken for neural networks model. The XGBoost gradient boosting algorithm realized in [44] sufficiently decreases this time applying paralleling technique. Still much effort should be taken to help the final user to interpret these models not only by the accuracy metrics, but also by the "black box" investigation. Real time analytical solutions enabling in-time detection of the energy demand and its high and low picks, require further research considerations.

References

- [1] Sen P., Roy M., Pal P., Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization, *Energy*, 2016, 116, 1031–1038 <http://dx.doi.org/10.1016/j.energy.2016.10.068>
- [2] Kalinchyk V. P. Methodology of operative management to the electric power consumption, *Energetika*, 2013, 1, 49–53 http://nbuv.gov.ua/UJRN/eete_2013_1_10
- [3] Zhang F., Deb C., Lee S. and oth., Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique, *Energy and Buildings*, 2016, 126, 94–103 <http://dx.doi.org/10.1016/j.enbuild.2016.05.028>
- [4] Verdejo H., Awerkin A., Becker C., Olguin G., Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile, *Renewable and Sustainable Energy Reviews*, 2017, 74, 512–521 <http://dx.doi.org/10.1016/j.rser.2017.01.110>
- [5] Rahman A., Srikumar V., Smith A., Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks, *Applied Energy*, 2018, 212, 372–385 <https://doi.org/10.1016/j.apenergy.2017.12.051>
- [6] OECD/IEA 2015 Energy and Climate Change, World Energy Outlook Special Report, IEA Publishing <https://www.iea.org/publications/freepublications/publication/WEO2015SpecialReportonEnergyandClimateChange.pdf>

- [7] Deb C., Zhang F., Yang J. and oth., A review on time series forecasting techniques for building energy consumption, *Renewable and Sustainable Energy Reviews*, 2017, 74, 902–924 <http://dx.doi.org/10.1016/j.rser.2017.02.085>
- [8] Daut M., Hassan M., Abdullah H. and oth., Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review, *Renewable and Sustainable Energy Reviews*, 2016 <http://dx.doi.org/10.1016/j.rser.2016.12.015>
- [9] Son H., Kim C., Short-term forecasting of electricity demand for the residential sector using weather and social variables, *Resources, Conservation and Recycling*, 2017, 123, 200–207 <https://doi.org/10.1016/j.resconrec.2016.01.016>
- [10] Statistical information, State Statistics Service of Ukraine, <http://www.ukrstat.gov.ua>
- [11] Official statistics, Statistics Committee, Ministry of National Economy of the Republic of Kazakhstan <http://stat.gov.kz>
- [12] Statistical information, Ministry of Energy and Coal Industry of Ukraine <http://mpe.kmu.gov.ua>
- [13] On approval of the Methodological recommendations on the calculation of the level of economic security of Ukraine: Order of the Ministry of Economic Development and Trade of Ukraine of October 29, 2013, No. 1277 <https://zakon.rada.gov.ua/go/v1277731-13>
- [14] Energy Security Risk Index Report, U.S. Chamber of Commerce, www.energyxxi.org/energy-security-risk-index
- [15] Ghysels E., Kvedaras V., Zemlyš V., Mixed Frequency Data Sampling Regression Models: The R Package midasr, *Journal of Statistical Software*, 2016, 72(4), 1–35. <http://dx.doi.org/10.18637/jss.v072.i04>.
- [16] Zhang L., Wu J., Liu H., Policies to enhance the drivers of green housing development in China, *Energy Policy*, 2018, 121, 225–235 <https://doi.org/10.1016/j.enpol.2018.06.029>
- [17] Su Y.-W., Electricity demand in industrial and service sectors in Taiwan, *Energy Efficiency*, 2018, 11(6), 1541–1557 <https://doi.org/10.1007/s12053-018-9615-y>
- [18] Carvalho A., Energy efficiency in transition economies: A stochastic frontier approach, *Economics of Transition*, 2018, 26(3), 553–578 <https://doi.org/10.1111/ecot.12152>
- [19] Swan L., Ugursal V., Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, *Renewable and Sustainable Energy Reviews*, 2009, 13(8), 1819–1835 <https://doi.org/10.1016/j.rser.2008.09.033>
- [20] Makridakis S., Spiliotis E., Assimakopoulos V., Statistical and Machine Learning forecasting methods: Concerns and ways forward, *PLoS ONE*, 2018, 13(3) <https://doi.org/10.1371/journal.pone.0194889>
- [21] Box G., Jenkins G., Reinsel G., Ljung G., *Time Series Analysis: Forecasting and Control*, John Wiley & Sons, Fifth edition, New Jersey, 2016, 88–339.
- [22] Hyndman R., Athanasopoulos G., *Forecasting: principles and practice*, OTexts, 2nd edition, 2018 <https://www.otexts.org/fpp>
- [23] Banerjee A., Dolado J., Galbraith J., Hendry D., *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford, 1993.
- [24] Kalman R., A new approach to linear filtering and prediction problems, *Transactions of the ASME—Journal of Basic Engineering*, 1960, 82(Series D), 35–45.
- [25] Bollerslev T., Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 1986, 31, 307–327.
- [26] Engle R., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 1982, 50(4), 987–1007.
- [27] Hyndman R., Fan S., *Monash Electricity Forecasting Model*, Monash University, 2015 <https://robjhyndman.com/papers/MEFMR1.pdf>
- [28] Sims C., *Macroeconomics and Reality*, *Econometrica*, 1980, 48, 1–48.
- [29] Faisal, Tursoy T., Resatoglu N. Energy Consumption, Electricity, and GDP Causality; The Case of Russia, 1990-2011, *Procedia Economics and Finance*, 2016, 39, 653-659 [https://doi.org/10.1016/S2212-5671\(16\)30312-4](https://doi.org/10.1016/S2212-5671(16)30312-4)
- [30] Johansen S., *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press, 1995.
- [31] Faisal M., Nishat F., Kafait U. Impact of China-Pakistan economic corridor on Pakistan’s future energy consumption and energy saving potential: Evidence from sectoral time series analysis, *Energy Strategy Reviews*, 2019, 25, 34–46 <https://doi.org/10.1016/j.esr.2019.04.015>
- [32] Holt C., *Forecasting seasonals and trends by exponentially weighted averages* (O.N.R. Memorandum No. 52), Carnegie Institute of Technology, Pittsburgh USA, 1957 <https://doi.org/10.1016/j.ijforecast.2003.09.015>
- [33] De Livera A., Hyndman R., and Snyder R., Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 2011, 106(496), 1513–1527 <https://doi.org/10.1198/jasa.2011.tm09771>
- [34] Brozyna J., Mentel G., Szetela B., Strielkowski W., Multi-Seasonality in the TBATS Model Using Demand for Electric Energy as a Case Study, *Economic computation and economic cybernetics studies and research*, Academy of Economic Studies, 2018, 52(1), 229-246 <https://doi.org/10.24818/18423264/52.1.18.14>
- [35] Sulandari W., Subanar, Suhartono, Utami H., Forecasting electricity load demand using hybrid exponential smoothing-artificial neural network model, *International Journal of Advances in Intelligent Informatics*, 2016, 2(3), 131-139 <https://doi.org/10.26555/ijain.v2i3.69>
- [36] Energy Consumption Forecasting Using ARIMA and Neural Network Models / C. Nichiforov, I. Stamatescu, I. Fagarasan, G. Stamatescu // 5th International Symposium on Electrical and Electronics Engineering (ISEEE). – 2017. <https://doi.org/10.1109/ISEEE.2017.8170657>
- [37] Wahid F., Ghazali R., Shah A., Fayaz M., Prediction of Energy Consumption in the Buildings Using MultiLayer Perceptron and Random Forest, *International Journal of Advanced Science and Technology*, 2017, 101, 13–22 <http://dx.doi.org/10.14257/ijast.2017.101.02>
- [38] Touzani S., Granderson J., Fernandes S., Gradient boosting machine for modeling the energy consumption of commercial buildings, *Energy and Buildings*, 2018, 158, 1533-1543 <https://doi.org/10.1016/j.enbuild.2017.11.039>
- [39] Friedman J., Greedy function approximation: A gradient boosting machine, *Annals of statistics*, 2001, 1189–1232.
- [40] Kaggle Data Science Platform <https://www.kaggle.com/>
- [41] PJM Hourly Energy Consumption Data, PJM Interconnection LLC <https://www.pjm.com/>
- [42] Hyndman R., R package ‘forecast’: Forecasting Functions for Time Series and Linear Models, 2019 <https://cran.r-project.org/web/packages/forecast/forecast.pdf>

- [43] Muggeo R., R package ‘segmented’: Regression Models with Break-Points / Change Points Estimation, 2019 <https://cran.r-project.org/web/packages/segmented/segmented.pdf>
- [44] Chen T., Guestrin C., R package Xgboost: A scalable tree boosting system, 2016 <https://arxiv.org/abs/1603.02754>
- [45] Quast B., Fichou D., R package ‘rnn’: Recurrent Neural Network, 2019 <https://cran.r-project.org/web/packages/rnn/rnn.pdf>

Appendix 1

Statistical Metrics to Evaluate Model Adequacy and Forecast Accuracy

To check the model fit we need to check significance of the coefficients, overall model adequacy and stability, correspondence to the model assumptions: no serial correlation, homoscedasticity and normal distribution of residuals.

The model adequacy is estimated on the basis of residual standard error (σ^2), coefficient of determination (R^2) that refers to the percentage of energy consumption variance explained by the model:

$$R^2 = \left[1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_t - \hat{y}_t)^2}{\text{var}(y)} \right] \times 100 \quad (22)$$

The model selection among alternatives is based on the information criteria usage. Akaike (AIC) and Schwarz Bayesian (BIC) criteria choose the most parsimonious model from the degrees of freedom point of view [23].

$$AIC = \ln \left(\frac{\varepsilon' \varepsilon}{T} \right) + \frac{2(p+q)}{T} \quad (23)$$

$$BIC = \ln \left(\frac{\varepsilon' \varepsilon}{T} \right) + \frac{(p+q)}{T} \ln T \quad (24)$$

The accuracy of the forecasts is verified on the basis of the following error measurements [22]:

1. ME (Mean Error):

$$ME = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t) \quad (25)$$

2. RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (26)$$

3. MAE (Mean Absolute Error):

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (27)$$

4. MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (28)$$

5. MASE (Mean Absolute Scaled Error):

$$MASE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|} \right| \quad (29)$$

Here m is the number of seasonal periods, for non-seasonal time series $m=1$.

Serial correlation can be assessed with autocorrelation function (ACF). ACF at lag=1 can be expressed as [22]:

$$ACF1 = r_1 = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}, \quad (30)$$

where y_t is the actual value of the series; \bar{y} is the mean value of the series; T is the number of time periods.

Appendix 2

Regression Model with Segmented Relationship(s)
 segmented.lm(obj = lm(DUQ_MW ~ Pittsburgh + hour_of_day + day_of_week,temp_power_train), seg.Z = ~Pittsburgh)

Estimated Break-Point(s):

Est. St.Err

psi1.Pittsburgh 287.742 0.039

Meaningful coefficients of the linear terms:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6289.8786 36.0835 174.315 < 2e-16 ***

Pittsburgh -17.4485 0.1295 -134.783 < 2e-16 ***

hour_of_day01 -57.8623 5.0219 -11.522 < 2e-16 ***

hour_of_day02 -89.0548 5.0285 -17.710 < 2e-16 ***

hour_of_day03 -106.9933 5.0321 -21.262 < 2e-16 ***

hour_of_day04 -111.4492 5.0311 -22.152 < 2e-16 ***

hour_of_day05 -94.6977 5.0331 -18.815 < 2e-16 ***

hour_of_day06 -37.7681 5.0343 -7.502 6.4e-14 ***

hour_of_day07 53.9267 5.0357 10.709 < 2e-16 ***

hour_of_day08 127.5685 5.0367 25.328 < 2e-16 ***

hour_of_day09 183.5881 5.0373 36.446 < 2e-16 ***

hour_of_day10 235.9642 5.0384 46.833 < 2e-16 ***

hour_of_day11 275.7144 5.0376 54.732 < 2e-16 ***

hour_of_day12 297.0947 5.0323 59.038 < 2e-16 ***

hour_of_day13 297.5840 5.0283 59.182 < 2e-16 ***

hour_of_day14 291.9243 5.0226 58.122 < 2e-16 ***

hour_of_day15 274.3167 5.0199 54.646 < 2e-16 ***

hour_of_day16 251.7429 5.0204 50.144 < 2e-16 ***

hour_of_day17 246.0881 5.0231 48.991 < 2e-16 ***

hour_of_day18 255.1194 5.0254 50.766 < 2e-16 ***
hour_of_day19 239.2830 5.0284 47.586 < 2e-16 ***
hour_of_day20 214.1266 5.0295 42.574 < 2e-16 ***
hour_of_day21 200.5399 5.0269 39.893 < 2e-16 ***
hour_of_day22 160.0846 5.0251 31.857 < 2e-16 ***
hour_of_day23 79.5914 5.0213 15.851 < 2e-16 ***
day_of_week2 23.7112 2.7093 8.752 < 2e-16 ***
day_of_week3 34.0968 2.7093 12.585 < 2e-16 ***
day_of_week4 29.0373 2.7121 10.706 < 2e-16 ***
day_of_week5 2.6755 2.7123 0.986 0.324
day_of_week6 -119.0771 2.7121 -43.906 < 2e-16 ***
day_of_week7 -148.5501 2.7129 -54.756 < 2e-16 ***
U1.Pittsburgh 71.5975 0.3233 221.449 NA
Residual standard error: 150.4 on 43027 degrees of freedom
Multiple R-Squared: 0.7442, Adjusted R-squared: 0.744