ISSN: 2219-8229
E-ISSN: 2224-0136
Founder: Academic Publishing House *Researcher*
DOI: 10.13187/issn.2219-8229
Has been issued since 2010.

European Researcher. International Multidisciplinary Journal

Technical Sciences

Технические науки

UDC 519. 711. 3

Data Set Analysis of Electric Power Consumption*

[1] Nataliia Beliaeva
[2] Anton Petrochenkov
[3] Korinna Bade

[1] Perm National Research Polytechnic University, Russian Federation
Professora Pozdeyeva 7, Perm, 614013
E-mail: beliaeva.nataliia@gmail.com
[2]Perm National Research Polytechnic University, Russian Federation
Professora Pozdeyeva 7, Perm, 614013,
E-mail: pab@msa.pstu.ac.ru
[3]Anhalt University of Applied Sciences, Germany
Postfach 1458, 06354 Koethen (Anhalt)
E-mail: k.bade@inf.hs-anhalt.de

Abstract. This paper presents the analysis of the dataset that is the consumption of electrical power in one household within practically four years in order to find out some patterns, cyclical or seasonal features or other significant information that allows us to do forecasting of the future demand with the certain degree of accuracy.
Keywords: electrical energy; analysis; forecasting; time-series; Box and Jenkins approach; ARIMA modelling.

Introduction

Nowadays practically all European countries while providing energy effective policy are concerned about reducing the total demand of energy consumption with maintaining the high level of development. Since the sector of individual consumption is one of the largest consumers of electric energy, the rational consumption of electricity at home becomes of a great importance.

The first step of rational consumption of electric energy is analyzing the level of electricity demand in order to predict future demand of electricity with a high degree of accuracy. Forecasting of the future demand can help making decisions for improving energy efficiency at home, choosing

* From 23.06.2013 on 01.07.2013 Perm National Research Polytechnic University (PNRPU) in collaboration with the University of Applied Sciences (UAS) Anhalt (Germany) held an International summer school "Information Management". During the school were conducted master classes by leading professors of UAS Anhalt, UAS Hamburg, PNRPU, Bauman Moscow State Technical University. Perm and German universities' students enrolled in the "double degree" program, presented papers on the subject of research at the partner universities. The best 3 papers presented in this journal.

the right class of the household equipment. Forecasting is also significant instrument for effective and rational planning of the budget.

Concrete task

The task to be solved in this paper is to analyze the dataset that is the consumption of electrical power in one household located in the Clamart, France (southwestern suburb of Paris) within practically four years (since December 2006 till November 2010) in order to find out some patterns, cyclical or seasonal features or other significant information that allows to do forecasting of the future demand with the certain degree of accuracy. The source of the data is the UCI Machine Learning Repository [1, 2].

Time-series data

The dataset represents the measurements of electric power consumption in one household with a one-minute sampling rate over a period of practically four years. The data presents different electrical quantities and some sub-metering values and is a typical representative of a time-series data that can be defined as a sequence of observed values. One of the most distinctive features of the time-series is that data is not generated independently; their dispersion varies in time, and often is governed by a trend and has cyclic components. An observed time series can be decomposed into three components: the trend (long term direction), the seasonal (systematic, calendar related movements) and the irregular (unsystematic, short term fluctuations).

Forecasting process

Forecasting is a process of estimating the unknown. It can be defined as the science of predicting future outcomes. Forecast should be fitted with the following characteristics: it should be timely, it should be as accurate as possible; it should be reliable; it should be in meaningful units. In order to do the forecasting process the following steps should be computed [3, 4]:

1. definition of the purpose of the forecasting;
2. data preparation;
3. preliminary analysis;
4. choosing and fitting the best model;
5. forecasting;
6. evaluation.

Data preparation

For analyzing the consumption of electrical power in the household the following attributes are required: date, time and active power as an electrical parameter that strongly depends on the electrical demand. In order to do the preliminary analysis the average daily temperature for each instance is added to the database. The source is the site with the archive of average temperature TuTiempo.

In the database there are missing values of about 1,25 % of the rows that were fulfilled with the previous values using the possibilities of R that is a free software environment for statistical computing and graphics for different platforms [5].

Preliminary analysis

The good way for understand the data is visualization in order to find out some consistent patterns or significant trend and to understand whether seasonality is important or if there is an evidence of some cycles.

With the help of Tableau 8.0 that is a powerful statistic tool for exploration and visualization of the datasets the graphics for the different time periods are constructed. To aggregate data values to the required time period the median of the active power is used. For example, figure 1 shows month graphic of the median of active power in comparison with average temperature.

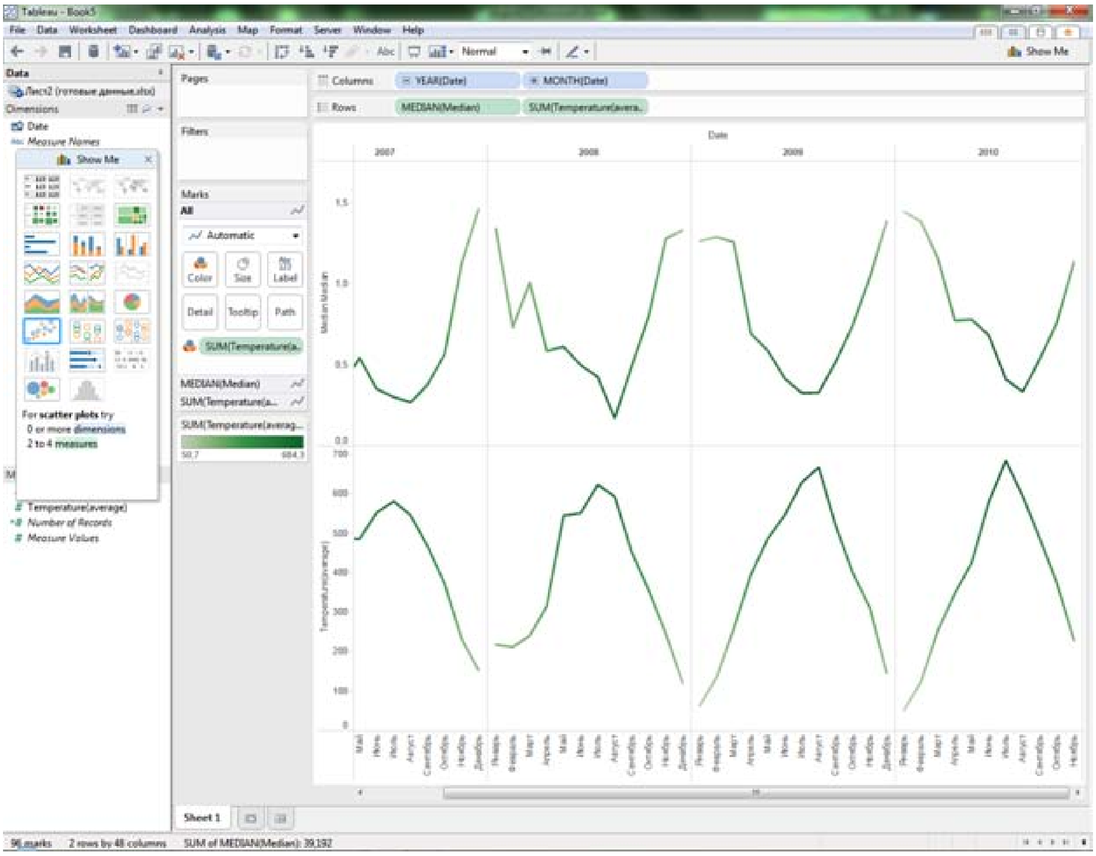

Fig. 1. Month graphic of the median of active power with average temperature

The most interesting from the practical point of view is probably the analysis of the day and week consumption that can be used in the future forecasting. Additionally there is enough data for doing the forecasting of the future demand; all in all there are 207 weeks and 1442 days in the observed period.

The results of the preliminary analysis show that active power of electrical demand is strongly seasonal dependent, some trend can be definitely observed there and also some random factor influences data distribution.

Decomposition of the data is frequently used for analysing the time-series data that contains trend, seasonal and the irregular components. Using Tableau 8.0 we receive several smaller datasets for analysing day and week consumption that have an appropriate form for doing decomposition. With the help of R the time-series objects with the required length and frequency are created and decomposed.

The results of decomposition with respect to days and weeks periods are presented in the figure 2 and 3.

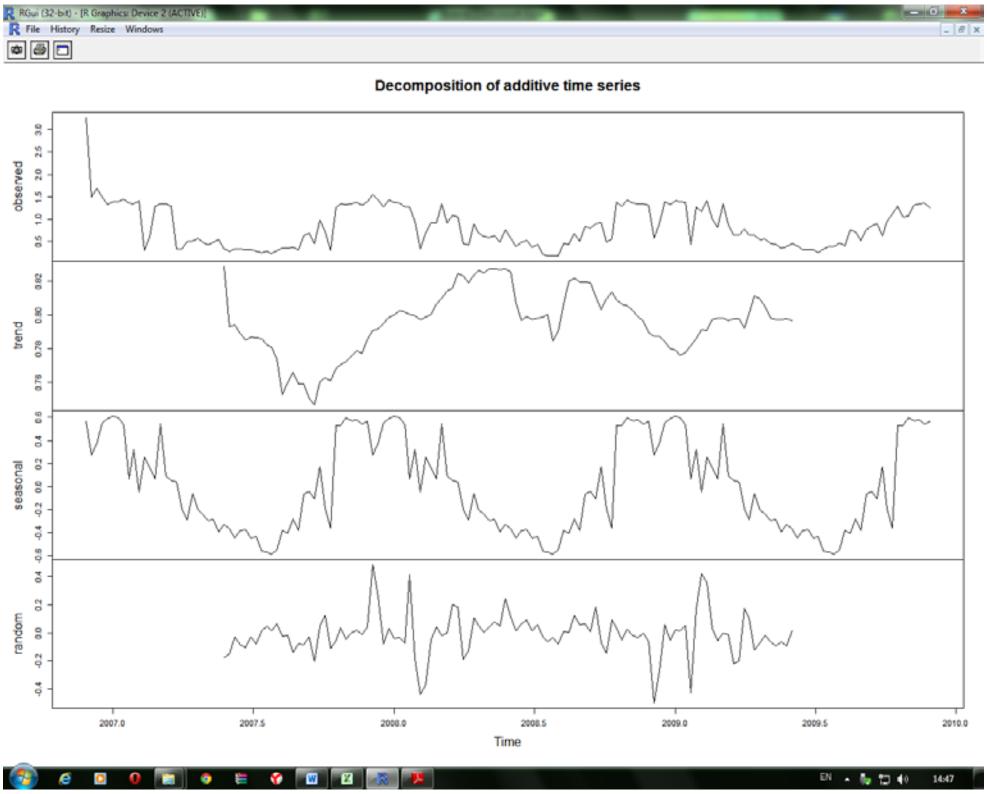

Fig. 2. Visualization of the decomposition of the additive time-series object with respect to the week period

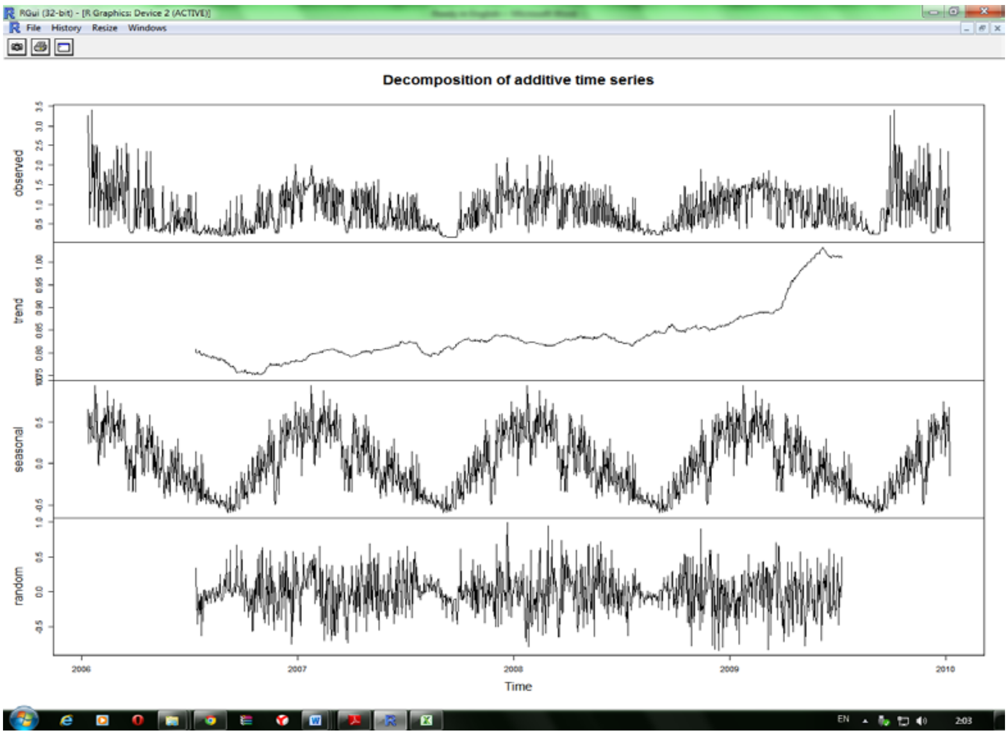

Fig. 3. Visualization of the decomposition of the additive time-series object with respect to the day period

The result of the dividing the data into components shows that time-series object is composed of trend, seasonal and irregular component. The graphics clearly shows that the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls. It means our data is additive and the observed time series (O_t) is the sum of three independent components: the seasonal (S_t), the trend (T_t) and the irregular (I_t): $O_t = S_t + T_t + I_t$.

Choosing and fitting the model

Next step is to determine the appropriate model that fits the data. For that purpose we use Box and Jenkins approach [6] that allows selecting from a group of forecasting models the one that is the best to fit the time series data. The ARIMA (autoregressive integrated moving average) modeling can be applied to the most types of time series data. The forecasting accuracy of ARIMA model is considered by scientist to be of a high degree.

In R environment there is a function that allows automatically detect the best fitted model to the given time-series with the smallest values of the Akaike's information criteria (AIC). The output of (auto.arima) for week`s period is ARIMA(2,0,1)(1,0,1), for day`s demand is ARIMA(2,1,1).

Forecasting

The forecasting is done for the autumn months of the year 2010 to understend the degree of accuracy that gives us the fitted ARIMAmodels for weeks and days periods.

The results of the prediction with respect to week`s period are presented at the figure 4.

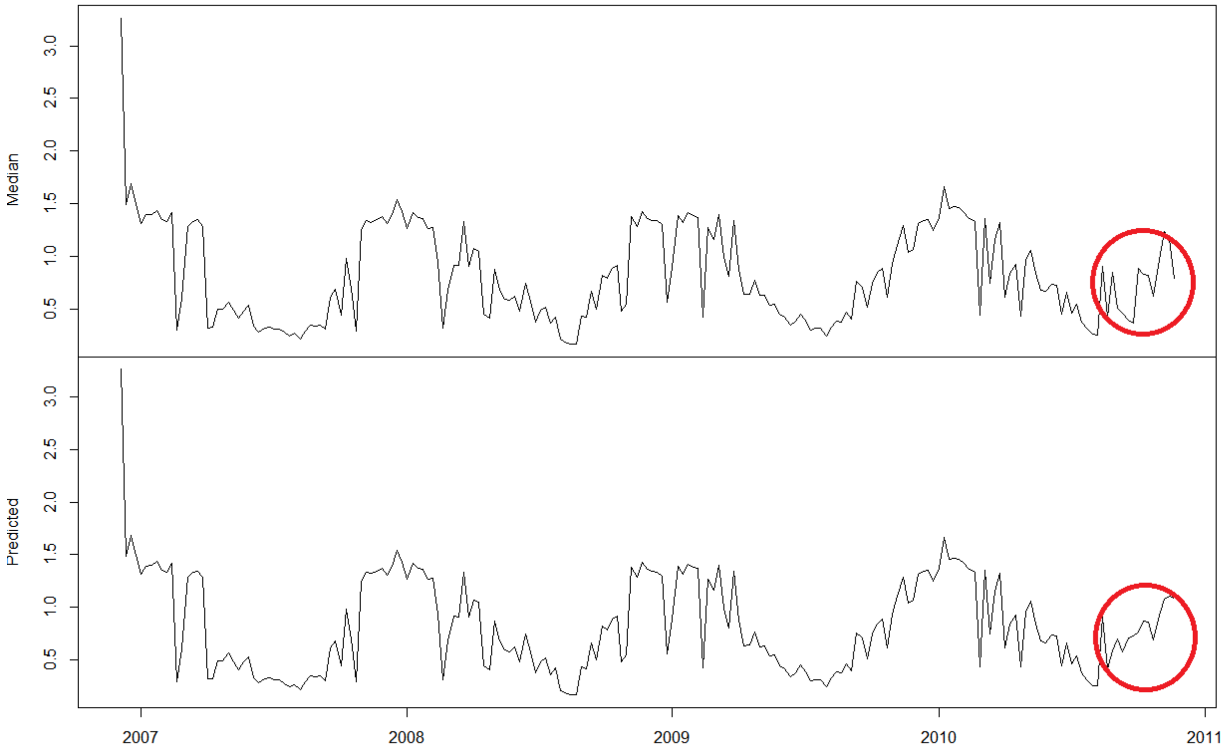

Fig. 4. Visualization of the forecasting of time-series data for autumn period 2010 with respect to weeks demand of electrical power consumption

Evaluation

Evaluation was made using the mean squared error (MSE) by calculating the difference between the forecasted values and the true values of the parameter with the following formula:

$$MSE\ (\bar{y}_i y_i) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \bar{y}_i)^2$$

The accuracy of the forecast with respect to week period is 0,039 and 0,201 for the day period. These results represent the high accuracy of the forecast. This proves the assurance of scientists that forecasting accuracy of ARIMA model is normally of a high degree.

Conclusion

The result of the research is the received forecast of the time-series data that is the day and week values of active power within individual household of a high degree of accuracy.

The results of the research can be applied in different fields. For example, knowing the amounts of energy consumption is of great importance for several reasons. First of all for consumers of electrical energy knowledge about the electric load and the targeted is important for understanding their bills and better controlling their consumption [7, 8]. For organizations it is also useful to know periods of minimum and maximum of the consumption for planning the technological cycles, for planning budget costs.

Secondly analyzing the data of energy consumption is useful for in energy sales companies to predict probable future consumption and applying the costs for electrical units on the opt market of electrical energy [9].

Thirdly it is useful for power grid companies to regulate and determine the optimal loading of transformer substations [10]. And at last analyzing the consumption of electrical consumption can be applied in governmental sector for calculating the optimal tariff schemes for different groups of consumers.

References:

1. UCI Machine Learning Repository http: //archive.ics.uci.edu /ml/datasets/Individual+household+electric+power+consumption;
2. Site TuTiempo: http://www.tutiempo.net/en/Climate/France/FR_3.html;
3. Bryan Nolan "Data Analysis: An Introduction", Wiley, 1994;
4. J.S. Armstrong "Principles of Forecasting: A Handbook for Researchers and Practitioners", Springer, 2001;
5. For R a-little-book-of-r-for-time-series: https://media.readthedocs.org /pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf;
6. Rob J Hyndman, article "Box-Jenkins modeling", 2001: http://robjhyndman. com/ papers/BoxJenkins.pdf.
7. Beliaeva N., Beliaev A., Petrochenkov A. Design and rationale of the algorithm of automated management of selected energy saving equipment for "Perm wooden house" // Solutions of applied problems in control, data processing and data analysis: 2nd International Conference of young scientists (4 April, 2011 Koethen, Germany). – Perm, Perm State Technical University Publishing house, 2011. P.3-13.
8. Beliaeva N., Beliaev A., Petrochenkov A. Energy saving controlling system for «PERM WOODEN HOUSE» // Innovation Technologies: Theory, Instruments, Practice: Proceedings oft he 3rd International Conference of young scientists and students. – Perm, 2012. – pp.329-338. – URL: http://conference.msa.pstu.ru/arhiv/skachat-sbornik-innotech-2011/view.
9. Kazantsev V.P., Petrochenkov A.B., Romodin A.V., Khoroshev N.I. Some aspects of Technology of Use of Electrical Objects on the Basis of Methods of Short Term Forecasting of Technical Condition. Russian Electrical Engineering, 2011, Vol. 82, No.11., pp.600-606, available at: http://link.springer.com/content/pdf/10.3103%2FS1068371211110058 (**DOI: 10.3103 /S1068371211110058**).
10. Petrochenkov A.B., Bochkarev S.V., Romodin A.V., Eltyshev D.K. The Planning Operation Process of Electrotechnical Equipment Using the Markov Process. Russian Electrical Engineering, 2011, Vol. 82, No.11., pp.592-595, available at: http://link.springer.com/content/pdf/10.3103%2FS1068371211110113 (**DOI: 10.3103/S1068371211110113**).