International Information and Engineering Technology Association

Advancing the World of Information and Engineering

Development of Regression Models to Predict Energy Consumption in Industrial Sites: The Case Study of a Manufacturing Company in the Central Italy

Elisa Moretti[1*], Luca Nassuato[2], Gian Pietro Bordoni[3]

[1] Department of Engineering– University of Perugia, Via G. Duranti 93, Perugia, Italy
[2] Department of Civil and Environmental Engineering – University of Perugia, Via G. Duranti, Perugia, Italy
[3] UMBRAGROUP S.p.A., Via V. Baldaccini, 1 - Z.I. Loc. Paciana, Foligno, Italy

Corresponding Author Email: elisa.moretti@unipg.it

ABSTRACT

The paper presents the results of a preliminary electric energy consumption analysis carried out on a manufacturing company operating in the Central Italy in the aeronautics and industrial sectors. Objective of this study is the development of a Multiple Linear Regression (MLR) model to predict and analyse daily electricity consumptions. The dependent variable (electricity demand) is function of different parameters referring to outdoor temperature (which influences the energy request for cooling) and production data available in the company database. Many preliminary MLR models were developed, by considering different parameters. The outcome of the study is the creation of a 5 parameters MLR model able to simulate the electricity demand with less than 7 % error. Considering the accuracy of this model, next aim of the study is his application to the monitoring of electricity demand, aiming to detect malfunctions and inefficiencies.

1. INTRODUCTION

In the current recession scenario, the Italian energy demand of the industry has decreased during the last years. However, in Italy, the industrial sector still remains significant and it accounts for around the 40 % of total electric energy demand in 2017 [1].

Also in response to European energy targets, Italy has developed National Energy Strategies. In particular, on July 18th, 2014 was published the Legislative Decree n. 102, implementing the European Directive 27/2012 on energy efficiency. The Decree establishes a framework of measures for the improvement of energy efficiency in order to contribute to the national energy saving targets; in particular, large and energy intensive companies have to perform an energy audit every four years (starting from 5 December 2015). One of the most promising strategy of energy savings in industry is to implement an energy management system, a systematic procedure aimed at defining policies and energy targets and to identify the processes and procedures necessary to achieve them. The International Standard ISO 50001 [2] suggests the use of dedicated KPIs, called Energy Performance Indicators, to help the performance monitoring in energy management systems. Notwithstanding, this metric is not able to fully account for the process information. The energy consumption of any industrial energy use is dependent upon a large number of variables [3].

In this context, continuous energy analysis is essential and predicting energy consumption is important for detect malfunctions and inefficiencies, for tracking of industrial machines and for proposing energy saving measures. Predicting energy consumption is a complicated task, especially in industrial buildings: it depends on multiple variables such as building characteristics, weather conditions, production cycles, energy systems characteristics, control and maintenance, etc.

The analysis and prediction of energy consumptions depending on different parameters have become the focus of many recent studies [4-17]. Numerous methods and models for the energy demand forecasting have been proposed including: Fourier series models, regression models (RM) [6-15] and neural network (NN) [16-17]. In particular, Multiple regression analysis (MLR) is often utilized to investigate the impact of various design parameters such as building construction, weather data, HVAC system, lighting system, etc. on the energy performance of the buildings [6-8, 10-12, 14]. Datta et al. [16] compared NN neural network (NN) models techniques to linear regression techniques and demonstrated that nonlinear models are substantially more accurate than linear models and a significant reduction of sum squared error is possible [16-17]. However, when compared to neural networks, multiple regression analysis could be an easier and more practical solution in many situations.

The paper deals with the development of a Multiple Linear Regression (MLR) model to predict and analyse daily electricity consumptions for an industrial site. The case study consists of a manufacturing company operating in the Central Italy in the aeronautics and industrial sectors.

2. THE CASE STUDY

The factory analyzed in this paper is the located in Foligno (Perugia), a small town in central Italy. The industrial building is the headquarter of UMBRAGROUP, a manufacturing company operating in the aeronautics and industrial sectors.

The company is a large and energy intensive company, according to the Italian Legislative Decree n. 102/14 and a complex energy metering system was recently implemented in order to have a continuous monitoring and to evaluate the impact of energy saving measures.

The building has an area of about 25.000 m^2, corresponding in a total heated and cooled volume of 187.500 m^3. The total number of employees of the factory is about 730.

The production department's working time includes 128 hours per week from 06:00 on Monday to 14:00 on Saturday, while office hours include 40 hours per week from Monday to Friday. On Sunday the company is closed.

The building is an industrial building with prefabricated prestressed concrete panels.

The production process consists of two main lines, the one for the ball bearings and the one for the production of recirculating ball screws (Figure 1a and Figure 1b). There is also EMA (electro-mechanical actuators) production line (Figure 1.c), but we don't describe it because the electrical energy consumption of this line is not significant when compared to that of the other 2 lines.
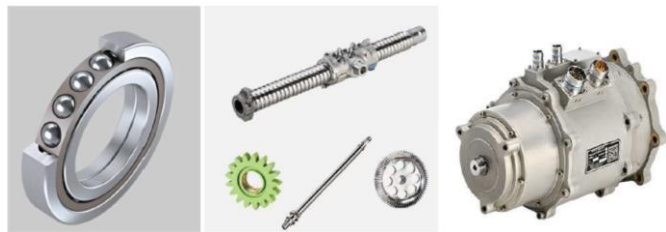

Figure 1. Main products: a) Bearings b) Ball screws, Linear Bearings, Gears c) Electro-mechanical actuators

The production cycle of ball bearings (Figure 2a) begins with the arrival of semi-finished products and as cages, spheres, rings. The rings and the spheres are subjected to heat treatments, in the TTC department (the most energy-intensive department, figure 3a). In the technological laboratory the physical-chemical properties of heat-treated materials are verified. Subsequently the components are subjected to grinding and lapping. A dimensional check of the bearing that is performed by the machine operators. Other checks are carried out in the metrological laboratory to check surface roughness and other dimensional characteristics. Finally, after a complete washing, the ball bearing is assembled and the marking is carried out and then the shipment.

The cycle of recirculating ball screws (Figure 2b) begins with the entry of the raw materials (steel bars) and the semi-finished products necessary for their realization. The bars are cut according to production requirements and then sent to the turning department where they are subjected to drilling, milling and threading. Subsequently they are subjected to heat treatments, (the third energy-intensive department Figure 3b), and galvanic treatments. At the end of the treatments the components are sent to the grinding department, (the second most energy-intensive department). Once the dimensional and technological compliance of the components has been ascertained, assembly is carried out. Finally, the screws are tested through performance checks carried out on special test benches. The last work phase consists of packaging and shipping.
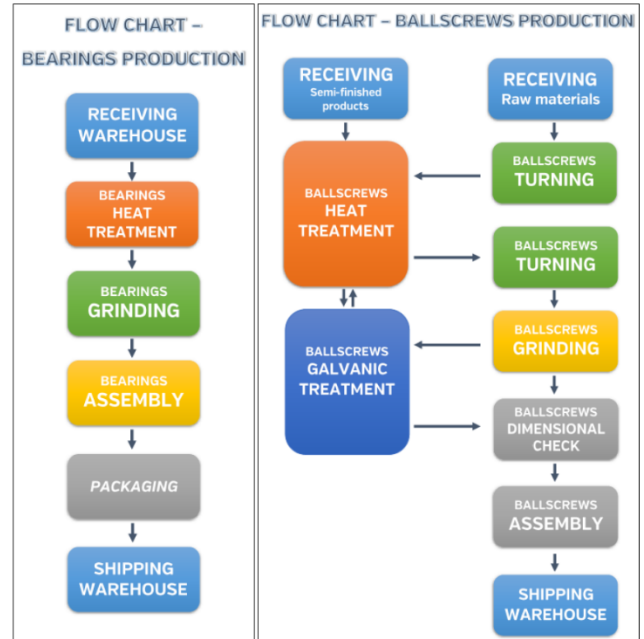

Figure 2. Flow chart of bearing and ballscrews production lines

Figure 3. a) Heat treatment bearings; b) Heat Treatment screws

To monitor electricity consumption in real time, a Schneider monitoring system was installed consisting of 76 multimeters which allow total control of electricity consumption. The plant is powered by a medium voltage cabin (20,000 volts) which powers 2 power centers with a total of 10 transformers that transform into low voltage. The monitoring system performs measurements on arrival in Medium voltage (using 2 meters), which allow you to monitor the energy purchased by the medium voltage supplier. The other 74 multimeters are installed in low voltage switchboards. The meters are connected by a mod-bus network, and by 2 mod-bus / TCPIP gateways, the data is sent to a dedicated server, where the Power Monitoring Expert 8.2 (program of Schneider) processes them. The data collected is displayed in multiple ways through reports or diagrams customized by the user. Figure 4 shows two PME dashboards.

a)

b)

Figure 4. The energy metering system

Referring to 2018, the natural gas consumption for heating and hot water was quite limited (about 565.000 Sm^3), whereas the electricity consumption was 20.98 GWh, therefore the work focussed on electric energy demand.

3. METHODOLOGY

In the present study, multi-linear regression analysis was carried out in order to predict and analyze daily electricity demand in the investigated factory building.

Multiple linear regression models the relationship between two or more variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. The following form of the regression equation was used to predict the daily electricity consumption:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \varepsilon \tag{1}$$

where:

y is the predicting daily electricity (kWh);

x_i is the value of the chosen parameter;

β_i is the corresponding regression coefficient;

ε is the statistical error.

The regression model coefficients are estimated by using the ordinary least square or linear least square method. This method tries to minimise the sum of the squares of the error terms.

Accuracy of the models was assessed by the coefficient of determination (R^2): the adjusted coefficients of determination (R^2adj) is a statistical index that provides information about the goodness of fit of a model. The significance of the MLR model was tested by using the Fisher–Snedecor test (F) [18].

The residuals analysis and the coefficient of determination are not enough to evaluate the performance of the estimation models. Therefore, the Root Mean Square Error (RMSE) and the Mean Absolute Percentage error (MAPE) were calculated to test the performance of the models [19].

The model validation is an important step in developing a model, especially when dealing with multiple parameters and the above mentioned parameters are not sufficient in order to determine the quality of the model. Therefore, the validation of the model was carried out by dividing the dataset in two groups: yearly data (2018) were used in creating the regression model, whereas the data of the first three months of the 2019 were used for testing the model. The validation was necessary to demonstrate the precision and the feasibility of the developed model.

4. RESULTS AND DISCUSSION

4.1 Data analysis

The data required to develop model were collected every day over a period of 12 months from January 2018 to December 2018. The database contains information for 365 days, but the sample was reduced to 357 data by excluding some days when maintenance work on the production lines were carried out.

The monthly electricity demand (Figure 5) varies significantly depending on the season: the mean daily electricity is in the 61583- 70746 kWh range when the cooling system is on (May, 1st- October, 19th), whereas it decreases up to 47394 – 57369 kWh in the other periods.

The electric energy consumptions are mainly due to production lines (79 %), whereas the contribution of the general services (HVAC system, offices, etc.) is 14 %. The compressed air accounts for about 7 % (Figure 6). The energy metering system allowed to divide the electricity demand between the different production lines (Figure 7): the EMA line is not significant (3 %) when compared to that of the other 2 lines; the majority of electric energy consumption is due to screws production (57 %).
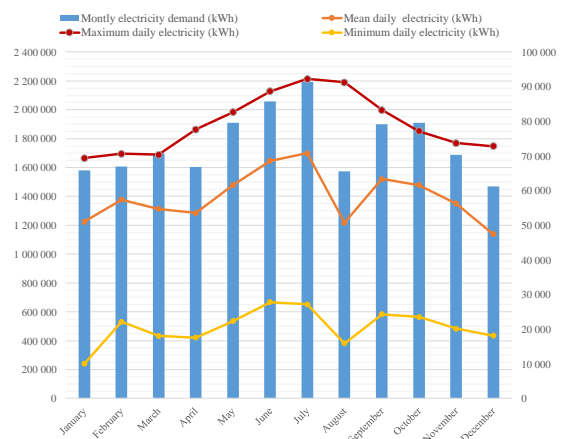

Figure 5. Monthly Electricity demand and daily variations (mean, maximum and minimum values, 2018)

The weekly electricity demand is shown in Figure 8: the data are quite similar for typical working days (Tuesday-Friday), when the working time is 0 a.m.-12 p.m and the differences are due to the production volume; the first day of

the week the demand is lower, due to the reduced working time (6 a.m-12 p.m). Likewise, on Saturday the electric energy demand is lower than typical working days (working time: 0 a.m- 2 p.m). Finally, the minimum request was achieved on Sunday when the company is closed. When the cooling system is on an increase for all days is observed.
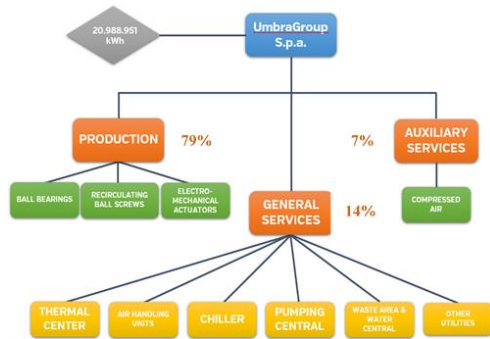

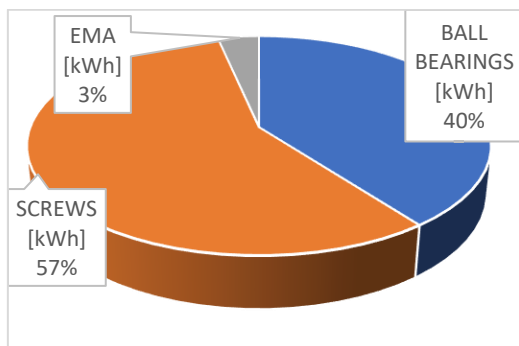**Figure 6.** Breakdown of electricity consumption

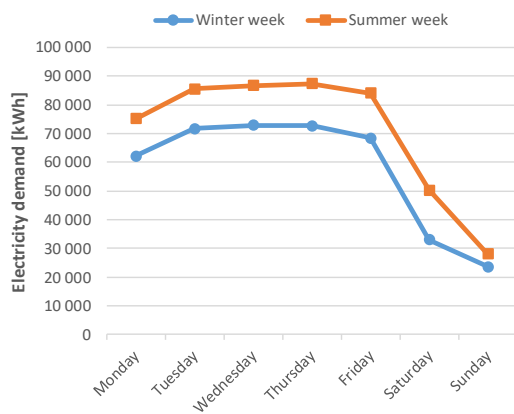**Figure 7.** Breakdown of production lines percentage

Figure 8. Weekly electricity demand

4.2 Model validation

Multiple linear regressions were used to model the relationship between the explanatory variables and the daily electricity consumption which is the response variable by fitting a linear regression.

The available daily data collected to characterize the electricity demand (kWh) were:

(1). Number of Screws (-);
(2). Economic production value for Screws (€);
(3). Number of ball bearings (-);

(4). Economic production value for ball bearings (€);
(5). Number of EMA (-);
(6). Economic production value for EMA (€);

Moreover, in order to take into account the increase due the cooling system (Figures 5 and 8) the mean outdoor temperature (°C) was included in the database when the cooling system is on.

Several models were tested to find the best fit between the simulated data and the model results and it was found that a five input model is the most appropriate solution. The different working days were taken into account by including a variable for the day type, according to figure 8. The input variables for the final model are shown in Table 1.

Table 1. Input variables

Variable	Values range
x_1= Mean outdoor temperature, when the cooling system is on (°C)	12-29
x_2= Economic production value for Screws (€)	0-334'216
x_3= Economic production value for ball bearings (€)	0-1'035'099
x_4= Day type	0: Sunday or holiday (the company is closed); 1: Saturday (working time: 0 a.m- 2 p.m.) 2: Special saturday (working time: 6 a.m- 10 p.m.) 3: Monday (working time: 6 a.m-12 p.m.) 4: Typical working day (Tuesday-Friday: working time: 0 a.m-12 p.m.)
x_5= Cooling system operation	0: Cooling system is off 1: Cooling system is on

Choosing of the five input model is an optimal solution because a higher number of inputs would have made the model too complicated and difficult to be used, while with less inputs the errors would have been higher.

The output of MLR model is an equation which accounts for all of the major variables affecting electric energy consumption. The analysis highlighted that the electricity demand was mostly influenced by the day type (strictly correlated to working time), the cooling system switch-on, the outdoor air temperature, and the Economic production value for Screws and ball bearings. On the contrary, the number of components (Screws and ball bearings) per day have instead a negligible significance in the model.

Figure. 9 compares the measured daily electricity consumption and regression model for the whole year. A Mean Absolute Percentage error less than 7 % was observed (Table 2). However, the mean percentage error for typical working days is quite limited (3 %), whereas the model performs worse for Saturdays (12 %) due to a high variability on working time and production typology for these days.

The obtained results demonstrate that the energy prediction models could be applied with success and with a sufficient precision on the results.
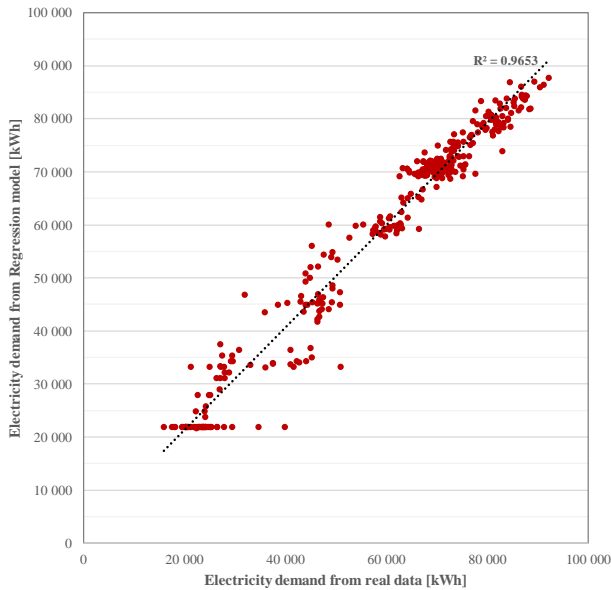
Figure 9. Regression model validation

Table 2. Model error statistics

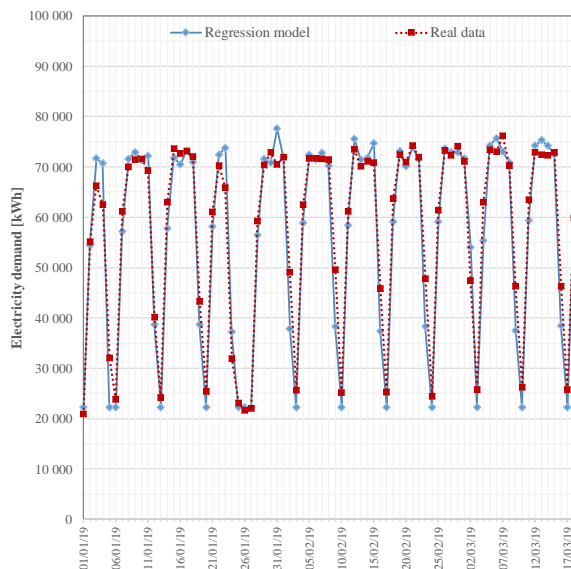| Parameter	Value
$R^2_{\,adj}$ (-)	0.96
RMSE (kWh)	3962
MAPE (%)	6.5
F-value (-)	1951

4.3 Electricity demand forecast

Figure 10. Comparison between simulated data and real data in the period 2019, January, 1st- March, 18th

The application of the regression model to the first three months of 2019 is plotted in Figure. 10. It can be seen there are some data with great differences between the predicted value and the real value. In particular, the model underestimates the daily electricity demand for all Saturdays and Sundays. The model performs well for typical working days instead. The deviation between real and predicted data was mainly caused by an increase in production during the weekend: the working hours on Saturdays were increased for

some production departments and some operations (Heat treatments) finish on Sundays, allowing higher consumptions without producing screws or bearings. Therefore, the developed model could not adequately represent the new management of the production process during the weekends and it is highly recommended to further investigate this issue.

5. CONCLUSIONS AND FUTURE WORK

In this study a total of 5 variables were identified and considered as inputs in the regression models. The coefficient of determination R^2 is about 0.96 indicating that 96 % of variation in daily electricity demand for the factory building can be explained by change in 5 parameters. The electricity demand is affected by the day type, the cooling system switch-on, the outdoor air temperature (when the cooling system is on), and the Economic production value for Screws and ball bearings.

The developed model helps the company energy coordinator and the decision makers to predict the electricity demand without extensive analysis. The regression model will act as pre-diagnostic tool to estimate the electricity request in the factory and to evaluate the impact of energy saving measures by the comparison between the data estimated from the validated regression model and the real data measured after the intervention.

A possible development of the study is the development of a different regression model addressed to the daily monitoring of the consumptions for each production line of the factory. This possibility is given by the use of a different database including data regarding the effective use of machinery: "Hours" (the number of hours of effective use of a machinery in a day) and "Hour cost" (the cost of an hourly operation of a machinery). The product of these two parameters provides the quantity "Amount": the total operation cost of a machinery. The resulting independent variables are the amounts of all the treatments needed for the production with daily frequencies. This model can provide more accurate results than the already developed model but it can be applied only after the production has occurred. However, it is expected to be automated in order to provide a tool for daily comparison between simulated and measured values for electricity demand in order to highlight possible inefficiencies.

ACKNOWLEDGMENT

REFERENCES

[1] Arera (Autorità di regolazione per Energia Reti e Ambiente), Bilancio energetico nazionale nel 2017. Available: https://www.arera.it/it/dati/bilancio_en.htm

[2] Sistemi di gestione dell'energia. (2018). Requisiti e linee guida per l'uso, UNI CEI EN ISO 50001.

[3] Schulze M, Nehler H, Ottosson M, Thollander P. (2016). Energy management in industry - A systematic review of previous findings and an integrative conceptual framework. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2015.06.060

[4] Asadi S, Amiri SS, Mottahedi M. (2014). On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. Energy Build 85: 246-255. https://doi.org/10.1016/j.enbuild.2014.07.096

[5] Guo Y, Wang J, Chen H, Li G, Liu J, Xu C, Huang R, Huang Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Appl. Energy 222(1): 16-27. https://doi.org/10.1016/j.apenergy.2018.03.125

[6] Kipping A, Trømborg E. (2017). Modeling hourly consumption of electricity and district heat in non-residential buildings. Energy 123. https://doi.org/10.1016/j.energy.2017.01.108

[7] McLoughlin F, Duffy A, Conlon M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. Energy Build 48(5): 240-248. https://doi.org/10.1016/j.enbuild.2012.01.037

[8] Mottahedi M, Mohammadpour A, Amiri SS, Riley D, Asadi S. (2015). Multi-linear Regression models to predict the annual energy consumption of an office building with different shapes. In: Procedia Eng 118: 622-629. https://doi.org/10.1016/j.proeng.2015.08.495

[9] Li Q, Gu L, Augenbroe G, Jeff Wu CF, Brown J. (2015). Calibration of dynamic building energy models with multiple responses using Bayesian inference and linear regression models. In: Energy Procedia 78: 979-984. https://doi.org/10.1016/j.egypro.2015.11.037

[10] Catalina T, Virgone J, Blanco E. (2008). Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build 40(10): 1825-1832. https://doi.org/10.1016/j.enbuild.2008.04.001

[11] Aranda A, Ferreira G, Mainar-Toledo MD, Scarpellini S, Sastresa EL. (2012). Multiple regression models to predict the annual energy consumption in the Spanish banking sector. Energy Build 49: 380-387. https://doi.org/10.1016/j.enbuild.2012.02.040

[12] Catalina T, Iordache V, Caracaleanu B. (2013). Multiple regression model for fast prediction of the heating energy demand. Energy Build 57(2): 302-312. https://doi.org/10.1016/j.enbuild.2012.11.010

[13] Fang T, Lahdelma R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Appl. Energy 177(6): 326-333. https://doi.org/10.1016/j.apenergy.2016.06.133

[14] Asadi S, Hassan M, Beheshti A. (2012). Development and validation of a simple estimating tool to predict heating and cooling energy demand for attics of residential buildings. Energy Build 54(11): 12-21. https://doi.org/10.1016/j.enbuild.2012.07.037

[15] Behl M, Smarra F, Mangharam R. (2016). DR-Advisor: A data-driven demand response recommender system. Appl. Energy 170(15): 30-46. https://doi.org/10.1016/j.apenergy.2016.02.090

[16] Datta D, Tassou SA, Marriott D. (1997). Application on neural networks for the prediction of the energy consumption in a supermarket. Belgium. Proceedings of the Clima 2000 Conference, Brussels, August 30th to September 2nd.

[17] Dodier RH, Henze GP. (2004). Statistical analysis of neural networks as applied to building energy prediction. Transactions of ASME 126: 592–600. https://doi.org/10.1115/1.1637640

[18] Kilic S. (2013). Linear regression analysis. J. Mood Disord. https://doi.org/10.5455/jmood.20130624120840

[19] Capozzoli A, Grassi D, Causone F. (2015). Estimation models of heating energy consumption in schools for local authorities planning. Energy Build 105(15): 302-313. https://doi.org/10.1016/j.enbuild.2015.07.024