

Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway

Tymofii Tereshchenko^{1*}, Dmytro Ivanko¹, Natasa Nord¹, and Igor Sartori²

¹Department of Energy and Process Technology, Norwegian University of Science and Technology (NTNU), Kolbjørn Hejes vei 1 B, Trondheim 7491, Norway

²SINTEF, Department of Building and Infrastructure, P.O. Box 124 Blindern, 0314 Oslo, Norway

Abstract. Widespread introduction of low energy buildings (LEBs), passive houses, and zero emission buildings (ZEBs) are national target in Norway. In order to achieve better energy performance in these types of buildings and successfully integrate them in energy system, reliable planning and prediction techniques for heat energy use are required. However, the issue of energy planning in LEBs currently remains challenging for district heating companies. This article proposed an improved methodology for planning and analysis of domestic hot water and heating energy use in LEBs based on energy signature method. The methodology was tested on a passive school in Oslo, Norway. In order to divide energy signature curve on temperature dependent and independent parts, it was proposed to use piecewise regression. Each of these parts were analyzed separately. The problem of dealing with outliers and selection of the factors that had impact of energy was considered. For temperature dependent part, the different methods of modelling were compared by statistical criteria. The investigation showed that linear multiple regression model resulted in better accuracy in the prediction than SVM, PLS, and LASSO models. In order to explain temperature independent part of energy signature the hourly profiles of energy use were developed.

1 Introduction

Prediction of building's heat energy use is a complex task. Particularly, this issue becomes challenging for district heating (DH) companies when heat energy planning is considered. Traditionally, the DH energy load include energy need for heating and domestic hot water (DHW).

These days low energy buildings (LEBs) such as passive houses (PHs) and zero emission buildings (ZEBs) are set as a national target for achieving energy efficiency and increase in primary energy savings. Simultaneously, characteristics of energy use in these types of building and their interactions with energy system are not fully investigated. The introduction of newly constructed LEBs and renovation of existing buildings to LEB standard brings additional volatility to heat demand in energy system.

DH production planning and operation involve decision making under uncertain conditions. Hence, accurate forecasting of daily and hourly heat loads is an important task in DH sector [1]. The need in advanced prediction technique arises, since load profiles of LEBs show variation in terms of energy use and duration of heating hours.

Energy signature diagram is a widely used instrument for analysis and prediction of energy use in the buildings. Energy signature diagram estimates energy use for DHW and DH as a function of outdoor temperature and may include other parameters. In most cases, the task of DH energy use is forecasting in the buildings that results in development of an accurate and representative energy signature diagram.

Two main approaches are available for modelling of DH energy use, which are physical modelling and data-driven approach.

Physical modelling is also called engineering approach, which employs physical principles to calculate thermal dynamics and energy behaviour on the whole building level and sub-level components [2]. A great example of physical modelling is software tools that were developed for energy use evaluation. Such simulation tools like EnergyPlus, ESP-r, IDA-ICE, BLAST, DOE eth. are well known and are mentioned in a number of research papers. The main drawback of mentioned above simulation tools that they require detailed input data for high quality modelling. To obtain these data is not always possible and economically reasonable.

On the contrary, in a data-driven approach, building energy behaviour is analysed by defining its statistical relationships with one or more different driving forces or parameters [3]. This approach got a lot of attention during recent years and is used for many applications. In particular, Machine learning techniques are widely applied for solving practical tasks in DH modelling and demonstrate high level of accuracy.

The widely used methods for energy use prediction are: Linear Regression, Support vector machine (SPV), Artificial neural networks (ANNs), Decision trees and other algorithms for development of linear and non-linear models. It should be noticed that mentioned methods are considered as advanced techniques. Some of them are quite sophisticated and may require application of special software, considerable amount of detailed input data, expert work and long computation time. In addition, these methods are not always applicable for utility companies.

* Corresponding author: tymofii.tereshchenko@ntnu.no

Nevertheless, prediction based on energy signature diagram, applying regression analysis is one of the most popular methods that is employed by a number of companies.

Regression analysis is statistical tool, which allow us to describe the variation of energy use in the building by the changes in influencing variables [2]. The goal of the regression analysis is to find an appropriate mathematical model and to determine the best fitting coefficients of the model from the given data [4].

Employment and comparison of algorithms based on multiple linear regression (MLR) analysis, general linear regression (GLR), ordinary least squares regression (OLS), autoregressive (AR), autoregressive integrated moving average (ARIMA), Bayesian regression, polynomial regression (poly), exponential regression, multivariate adaptive regression splines (MARS), case-based reasoning (CBR), and k-nearest neighbours (kNN) for building energy use prediction is given in [5]. Some applications of regression algorithms are described further. The study performed in [6] has proven that regression based prediction can be efficiently used as a tool for long-term energy use prediction. The modelling of monthly heating demand for residential buildings is investigated in [7]. The comparison of energy signature method and Eta method based on statistical regression model can be found in [8]. The authors found high degree of predictability for both heating and cooling loads treating them simultaneously. Aranda et. al [9] apply regression models to predict the annual energy use in the banking sector. Multiple regression model for fast prediction of heating energy demand with application on residential multifamily building is done in [4]. Prediction of annual energy use for office building from heating and cooling perspective is investigated in [10]. Hence, it can be concluded that the regression algorithm is widely used due to its simplicity and accuracy. Therefore, this paper describes improved methodology for planning and analysis of heating and DHW energy use by means of energy signature method with application of advance regression techniques.

The main objective of this study is to support heat and DHW energy planning that involves LEBs by providing rapid and simple solution of energy demand assessment with high level of accuracy.

2 Methodology

2.1 Low energy school building

The analysis performed in this work aimed to improve a degree of predictability of energy prediction tool. Due to increasing share of LEB in a building stock it is important to have a tool able to work on every building type and category. For this reason, a passive school was introduced as a source of energy use data. The school was constructed in 2010 in Oslo and has 6 454 m² heated area. Specific heat use calculated in 2011 was 31.76 kWh/m²a. The energy signature diagram is shown in Fig. 1. The characteristics of the mentioned passive school building are typical for Norwegian conditions and energy use

threshold was found in the range with other similar buildings in Norway.

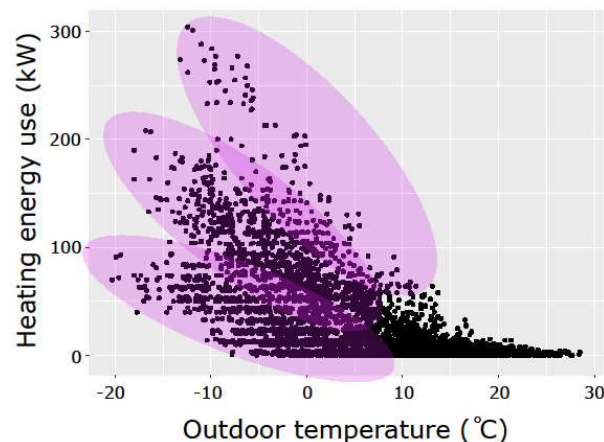


Fig. 1. Energy signature diagram of analysed building

Obtained data samples from school building were hourly based with one year duration. The building monitoring system sampled data that included various categories explained in Table 1. From Fig. 1 it can be noticed that there is no a clear pattern in energy use while the outdoor temperature is below zero. Simultaneously three-tailed pattern in the left part of Fig. 1 can be distinguished. The reason for this is unknown and may be caused by various reasons. Some of them could be due to applied control strategies or operation regimes in analysed building. Therefore, the aim of energy predictor is to capture shown volatilities in energy use data and provide reliable model that would be capable to identify them under various conditions in different types of buildings.

2.2 Regression model

In the most cases, analysis of energy signature diagram is based on using simple regression models in order to describe the behaviour of energy use in the building. Considering that energy signature diagram is dependent in terms of start and end of heating season, the division on temperature dependent and temperature independent parts that could be found in many publications is usually explained by the following equations:

$$\text{If } T_t < \text{CPT:} \\ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

$$\text{If } T_t > \text{CPT:} \\ Y = \beta_0 + \varepsilon \quad (2)$$

where, ε is random error. $\beta_0, \beta_1, \dots,$ and β_p describes the expected change in the predicted variable Y in response to a unitary change in x_i when the rest of predictors remain constant. The x_i is explanatory variables, such as wind speed, temperature, eth. CPT is change point temperature, with physical meaning of start and end of heating season. This means that temperature dependent part considers space heating and domestic hot water (SH+DHW), while temperature independent part considers DHW only.

In order to solve the introduced above equations, traditionally the least squares method is applied separately

to each segment of the model. This lead to composition of two regression lines that fit the data as closely as possible, while minimizing the sum of squares of the differences (SDD) between observed and calculated values of dependent variable. The reasonable modification of the method of energy signature modelling is proposed in the article of Lindberg et. al [11]. The authors of this work introduce a temperature moving average (TMA) of last 24 hours as additional explanatory variable in the model in order to take in account building inertia in the model. Also in the article the change point temperatures, as well as, models are identified for each hour of day.

Despite the relative simplicity, there are many drawbacks and unsolved issues in considered methods. For example, the question of identifying exact value of CPT is usually not considered. Moreover, in a number of publications, selection of CPT is relies only on intuition and experience of researches. In addition, the possibility to increase the accuracy of the energy signature model by using advanced modelling techniques has not been studied sufficiently. Therefore, the improved approach of energy signature analysis is proposed in this article.

The analysis of energy signature proposed to conduct in the following way. First, the available statistical data should be pre-processed by division in two samples related to (SH+DHW) needs and related to DHW needs. This was done by a piecewise regression method combined with additional conditions related to month of the year. The piecewise regression allows us automatically to figure out the exact value of CPT in energy signature diagram. In such way, two regression lines are used to fit the data set as closely as possible. This minimizes the sum of squares of the differences (SSD) between the observed and the calculated values of energy use in different segments of energy use diagram.

When there is only one breakpoint, at $x = CPT$, the model can be written as follows [12]:

$$\begin{aligned} y &= a_1 + \beta_1 x & \text{for } x \leq CPT \\ y &= a_2 + \beta_2 x & \text{for } x \geq CPT \end{aligned} \quad (3)$$

In order for the regression function to be continues at breaking point, the two equations for y need to be equal at breakpoint (when $x = CPT$):

$$a_1 + \beta_1 CPT = a_2 + \beta_2 CPT \quad (4)$$

Solving for one of the parameters in terms of the others by rearranging the equation above:

$$a_2 = a_1 + CPT(\beta_1 - \beta_2) \quad (5)$$

Then by replacing a_2 with the equation above, the result is a piecewise regression model that is continuous at $x = CPT$:

$$\begin{aligned} y &= a_1 + \beta_1 x & \text{for } x \leq CPT \\ y &= a_1 + CPT(\beta_1 - \beta_2) + \beta_2 x & \text{for } x \geq CPT \end{aligned} \quad (6)$$

where: β_1 and β_2 are regression coefficients, which indicating the slope of the line segments; a_1, a_2 are regression constants, which indicates the intercept at the Y-axis.

In this study the CPT was defined by computational means applying Python software. Temperature dependent heating energy use was explained via multivariable regression models. In order to obtain reliable models that considers the variety of data points, energy use data were separated in four sample groups: 1) weekends, 2) working days – working hours 3) working days – non-working hours 4) non-typical energy use. Unusual energy use data were investigated based on confidence intervals to regression models. For each group of samples, separate models were obtained. In order to choose the best model that explains energy signature diagram, several advanced models like: support vector machines (SPV), partial least square regression (PLS), least absolute shrinkage selector operator (LASSO) were compared applying statistical criteria. The final model is a combination of four sub-models separated by four sample groups and used to generate prediction output. The testing and training sets were defined and model was tested.

Instead of considering values of temperature independent energy use as an intercept shown by Eq.(2), it is suggested to present them by means of energy profiles with division on months, working and non-working days.

The statistical analysis and model development of energy signature diagram for LEBs was implemented with software tools like R and Python.

3 Results

The results given in this section show workflow how the analysis on model was done and improvements were introduced. The results are divided in several sections with specific tasks to analyse.

3.1 Correlation analysis

One of the most important tasks in prediction of building energy use is selection of input variables. A number of studies dealing with parameter evaluation could be found in literature. Several studies show that solar irradiation has impact, others that wind speed effects energy use [13, 14]. In addition, the day of the week or working hours correlate a lot. In order to figure out which variables have the highest impact, a correlation analysis was introduced in this study. A correlation analysis is a simple way to select the input variables and see the degree of linear relationship between them. Quite often, it is not always possible to collect all the variables when it comes to real operation, therefore, correlation analysis aimed to identify factors that have the highest impact on analysed parameter and to reduce the total number of components. In such way, the most insignificant parameters are eliminated. The building monitoring system sampled various data categories shown below, together with energy signature diagram shown in Fig. 1 the correlation analysis was conducted. The correlation analysis aimed to find out how various factors effect heating load when system is

operated under temperature dependent (SH + DHW) and temperature independent (DHW) modes. A correlation matrix was created and the results are shown in Table 1.

Table 1. Correlation matrix results

Component	SH+DHW	DHW
Month	-0.370	-0.070
Weekend day	-0.235	-0.167
Working day	-0.362	-0.058
Day of the week	-0.202	-0.127
Hour	0.021	0.012
Outdoor temperature	-0.516	-0.237
Medium outdoor temperature during 24 hours	-0.554	-0.426
Wind	-0.006	0.270
Medium wind during 24 hours	-0.160	0.292
Season	-0.364	-0.455
Electricity use	0.595	0.431

To recall, the correlation coefficient measures strength and direction of a linear relationship between the variables. The week positive or negative correlation starts from value of ± 0.30 .

From Table 1, it can be noticed that SH+DHW heating load shows low correlation with wind, medium wind, and hour of the day. Hour of the day and wind are below significance level, while medium wind has low correlation with heating load. The weak correlation was observed for parameters like weekend/weekday, month and season. The correlation is negative for all parameters and this can be explained as all these components are similar in terms of time factor. The highest negative correlation was found for outdoor temperature and medium outdoor temperature. This indicates that heating load increases while temperature decreases. The electricity use has moderate positive correlation with heating. It might be that some school areas are heated up with electricity panels and this is the reason of positive correlation. The analysis of temperature independent part (DHW) shows moderate correlation with medium outdoor temperature, season and electricity use. This is reasonable, because energy use in DHW part would be different due to seasonality, e.g. summer vacation, Easter holidays or beginning of school season.

3.2 Building's energy use profile

The analysis of energy signature diagram is key to understanding of future energy use for a particular building type and a building category. Therefore, it is important to figure out the reasons for typical and untypical energy use patterns that were found in Fig. 1. The tailed data shown in left side of Fig. 1 were investigated by separation of existing data points on hourly basis intervals. The idea behind this was to find cluster formations that could explain tails.

Unfortunately, hourly data distribution could not provide clear explanation about energy use extremes. In

order to see more clearly energy use pattern, the boxplot was established and is shown in Fig. 2. The spikes in data were observed practically during each hour of the day. However, it can be noticed that energy use increases drastically starting from 7 AM and decreasing by 5 PM (17 o'clock in Fig. 3). Before that time energy use showed maximum at 150 kW, but later increased up 250-300 kW.

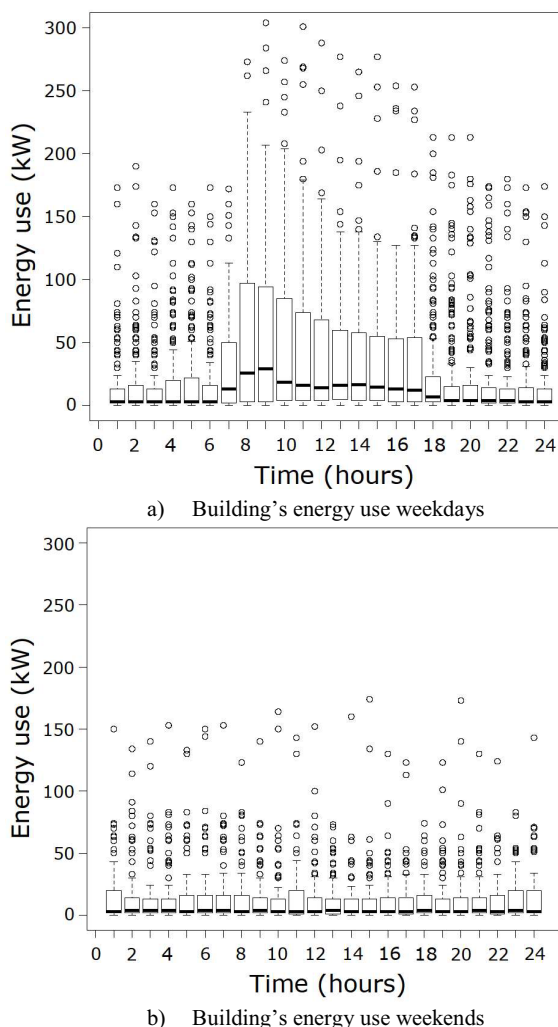


Fig. 2. Hourly energy use

The boxplot depicts data for weekdays and weekend days. The maximum, median and minimum energy use and density of data points are shown. In addition, outliers and suspected outliers that are not typical to the analysed data range are shown. The suspected outliers are shown as dots with higher density, while outliers are randomly distributed dots far beyond of suspected outliers. Fig. 2 shows that the building energy use has clear visible pattern, showing increase during working hours in the period between 7–17 o'clock. The weekend profile shows pretty smooth energy use pattern without sudden peaks and drops during the day.

3.3 Analysis of temperature lag

As it was mentioned before, TMA of last 24 hours has to be introduced in order to consider building inertia in the

model, see [11]. The analysis on TMA aimed to figure out how lagged outdoor temperature in terms of hour of the day effects on building energy. It is known that building it a subject to thermal inertia. Depending on building constructions, some buildings can accumulate more heat and use it afterwards to improve indoor thermal comfort. In order to figure out to which extent temperature lag has effect on energy use, the outdoor temperature was shifted by each hour for 48 hours see correlation between mentioned parameters. Fig. 3 shows correlation between TMA and heating energy use.

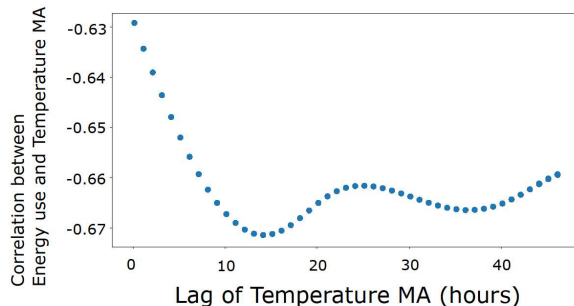


Fig. 3. Effect of temperature lag on energy use of building

Fig. 3 shows that if the correlation factor has a bigger absolute value, this indicates better fit between TMA and heating energy use. This is a good indication that thermal inertia in the building takes place, because better correlation was achieved when the TMA was shifted. The largest correlation for TMA was found for 14 hours. This shows that lag of 24 or 48 hours that are commonly used for model definition and description of building inertia is misleading. The found results showed that value of this parameter is dependent on building's construction type and time constant parameter.

3.4 Model formulation

Next step of the analysis aimed to check if cluster formations could be distinguished among available data. Therefore, several clustering algorithms have been tested. The main focus here was on temperature dependent part (SH+DHW), due to high variety in data points. The tests were conducted with the following algorithms: k-means clustering, hierarchical clustering, density based clustering, and model based clustering. In spite of different approach used in each method, most of the them did not show a good degree of clusterization. The cluster formations did not follow specific pattern that would explain tailed data. For this reason, the decision was made to apply techniques that would allow to separate untypical data points in existing data range. Hence, confidence interval (CI) was applied to analysed data. The CI was calculated by next equation [15]:

$$C.I = \hat{Y}_i \pm S_e \times t\left(1 - \frac{\alpha}{2}, f_e\right) \times \sqrt{1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)S_x^2}} \quad (7)$$

where, \hat{Y}_i is predicted value of energy use; $t(1 - \alpha/2, f_e)$ is Student's criteria, which depends on probability α and

f_e degrees of freedom; n is the sample size; S_e is the residual standard deviation of actual energy about the regression line; \bar{X} is the mean value of independent variable; X_i actual value of independent variable; S_x is the standard deviations of the of independent variable.

Fig. 4 show regression model with upper and lower bounds of the confidence interval. This step aimed to separate data that did not fit in typical model population. The shape of many sample distributions can be approximated by a normal distribution. A convenient aspect of normal population distribution is that we can apply 95% confidence interval to describe desired population range. The confidence interval was created for each group of dataset described in Section 2.1, weekends, working days – working hours and working days – non-working hours. The results are given in Fig. 4.

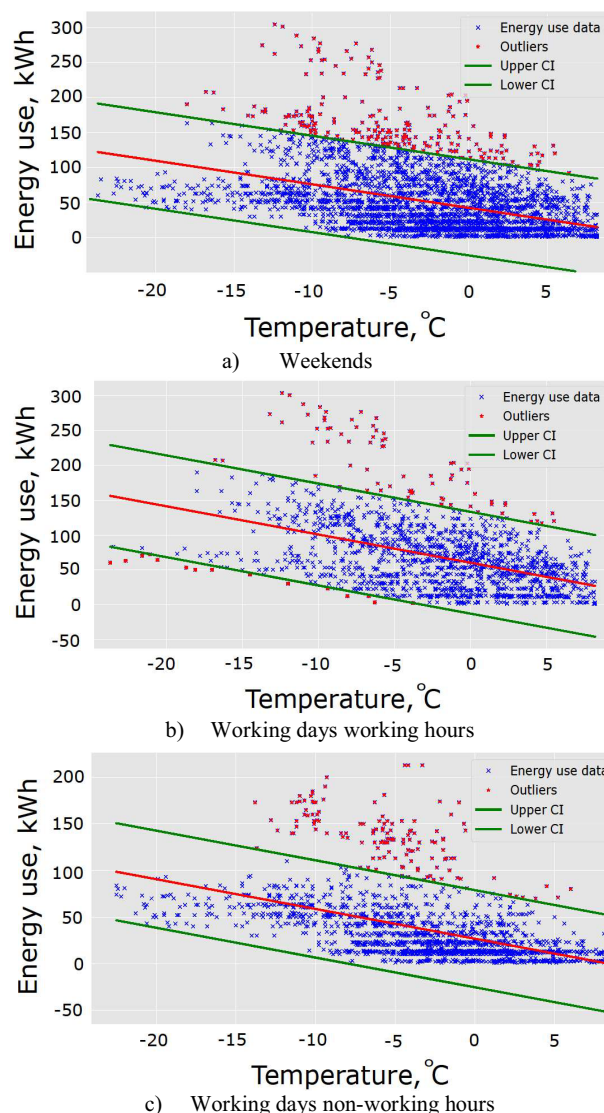


Fig. 4. Regression models with confidence interval

The accuracy of all models was evaluated by few statistical criteria: multiple determination coefficient (R^2), mean absolute error (MAE), and mean square error (MSE). The mathematical formulation of statistical criteria is shown below.

$$R^2 = \frac{Var(y'_i)}{Var(y_i)} = 1 - \frac{Var(e)}{Var(y_i)} \quad (8)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - y'_i|}{n} \quad (9)$$

$$MSE = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (10)$$

where, y_i is the predicted output variable and y'_i is the actual output variable for i^{th} entry in the analyzed database, and n is the number of samples in training subset. e is error term. The results are summarized in Table 2 and Table 3.

Table 2 and Table 3 shows prediction results for testing and training sets for different predictive algorithms under various statistical criteria. To recall, R^2 criteria means the better model when it is closer to 1, for MAE and MSE, the lower value, the better model.

Table 2. The accuracy of main models

№	Model type	R^2		MAE		MSE	
		Train ing set	Testi ng set	Traini ng set	Testi ng set	Trainin g set	Testin g set
1	Linear regression	0.834	0.832	7.71	7.84	202.18	208.09
2	SVM	0.828	0.829	7.26	7.60	208.41	209.71
3	PLS	0.833	0.829	7.59	7.80	200.92	206.80
4	LASS O	0.834	0.831	7.29	7.49	185.41	189.04

Table 3. The accuracy of outliers' model

№	Model type	R2		MAE		MSE	
		Traini ng set	Testin g set	Traini ng set	Testin g set	Traini ng set	Testin g set
1	Linear regression	0.814	0.725	14.94	17.14	485.48	661.21
2	SVM	0.801	0.709	14.26	17.50	514.37	706.05
3	PLS	0.773	0.657	15.35	18.36	461.16	644.74
4	LASS O	0.717	0.554	13.20	18.89	351.20	497.96

From Table 2, it can be noticed that obtained regression model shows good prediction ability to explain heating energy use. Both training and testing results scores were found in the same range for all statistical criteria. It can also be noticed that more advanced algorithms resulted in close values to regression model. This shows that improvements introduced to regular regression model resulted in good explanation degree of analysed energy use data of school building. The results for outliers' model that are shown in Table 3 were found as less accurate. This can be explained by sparse data and occasion occurrence of it. In general, it can be concluded that improvements made to regression model led to better prediction capability. This is valuable information, since

in such way the prediction done by utility companies become more reliable and security of supply increases.

3.5 Analysis of outliers

The analysis of data points separated as outliers from Fig. 4 was investigated. The total number of identified points was 2.6% of total annual data points, which corresponds to 17% of total heating energy use. The analysis showed that the occurrence of outliers mainly appeared in two consecutive months such as January and December. The distribution of these points showed random pattern without clear cluster formations except Thursday. During that day energy use cluster was identified between 18–22 o'clock (6-10 PM). This is particularly relevant for January. The reason for this could be that the building was used for purposes other than education. It is quite common that in Norway schools are booked for Christmas celebration by local companies. Other reasons could be particularities in the in operation of the heating and ventilation system in the building. Unfortunately, analysed data were received without any explanation about equipment installed inside the building and therefore, it was hard to conclude something about its operation. During other months the number of outliers was negligible and this information was considered as insignificant.

3.6 Analyses of temperature independent part of energy signature

As has been mentioned above, energy signature consists of two different parts – temperature dependent (SH+DHW) and temperature independent (DHW) energy use. These parts are separated by CPT and the additional condition related to the month of the year. The analysis of temperature independent part showed that statistical data covered months from April to October. It was found that occasionally temperatures lower than CPT were observed during this period. Nevertheless, analysis shows that these temperatures do not last for a long time and the need in SH does not occur. Therefore, it can be concluded that temperature independent energy use can only be observed within considered months.

DHW energy use profiles is the primary instrument for understanding people behaviour and their effect on DHW use in buildings. Analysis of DHW profiles showed changes of energy use under different time intervals. Primarily, DHW energy use depends on a number of people who are present in a building. However, information about people presence is usually not available. The month of the year and the day of the week are factors that have direct influence on building attendance and, consequently, DHW energy use. Statistical analysis showed that unlike SH, DHW energy use has no other important explanatory variables. Therefore, an approach that differs from regression analysis should be used to explain temperature independent part of energy signature. The profiles found for working days and weekends are shown in Fig. 5 and represent average values in selected period.

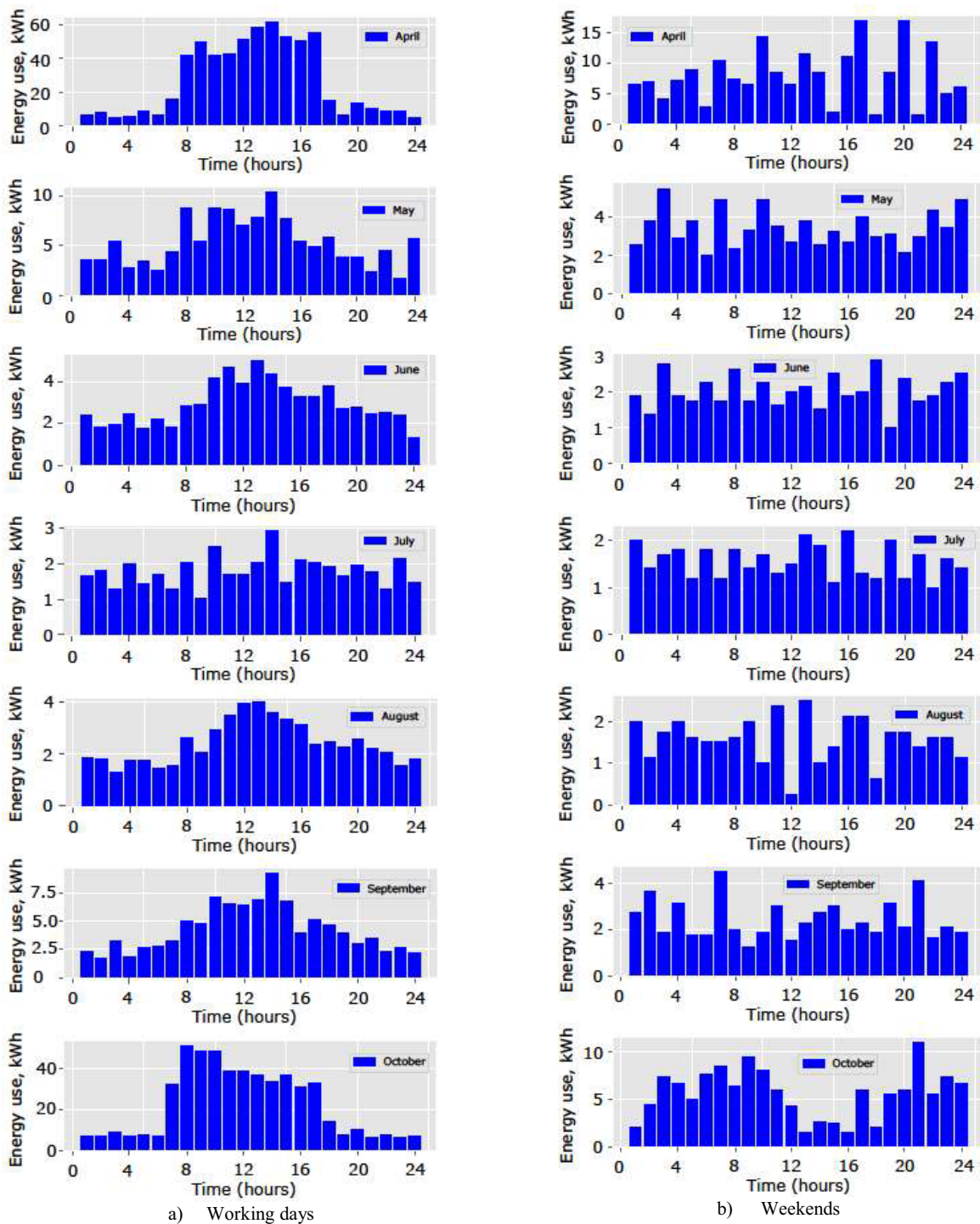


Fig. 5 Profiles of temperature independent part of energy signature

The value of R^2 showed that the proportion of total variation of outcomes explained by the profiles was equal to 0.71. This value of R^2 justified the expediency of using profiles. Moreover, the profiles obtained in this way were quite informative and allow us retrieve additional information about DHW energy use in buildings.

As we can see from Fig. 5 the energy use in April and October are higher than in other months. This is due to the fact that in these months the school building was fully occupied by students. In addition, the outdoor temperature was lower and could induce to extra energy use. The smallest DHW energy use was in June, July, and August when there was no classes in the school and most of employees were on vacation. Further, DHW energy use is oscillating along zero line during the summer time, the reason for this could be that hot water circulation (bypass) took place to keep the system in operation. The energy use

in working and non-working days was different. In working days the maximum value was higher, as well as variation of energy use in general. Finally, it can be concluded that information retrieved from profiles is very useful to understand occupant behavior and will be used in further research.

4 Conclusions

This paper aimed to improve prediction of heating energy use by introducing changes to general regression model algorithm. The analysis was done on the passive school on hourly data resolution. The model was divided into the sub-models that helped to separate untypical energy use data from typical energy data range points. The correlation analysis was performed and most influential variables were selected for model formulation. The results showed that introduced improvements resulted in high accuracy in comparison with more sophisticated algorithms like SVM, PLS, and LASSO. This is a good observation, because regression algorithm does not require sophisticated knowledge, high computational time, or expert work for its implementation. Further, the analysis of the temperature lag showed that it is misleading to introduce lag of 24 hours and 48 hours that could often be found in the literature. The reason for this is due to differences in thermal inertia of building types. The analysis of outliers showed some degree of clusterization during January and this could be explained by non-educational activities in the analysed building and operation particularities. The temperature independent part of energy use was analysed and hourly profiles were developed. In general, it can be concluded that improvements made to regression model led to better prediction capability. This is valuable information, since in such way the prediction done by utility companies become more reliable and security of supply increases.

The authors gratefully acknowledge the support from the Research Council of Norway through the research projects: the Research Centre on Zero Emission Neighbourhoods in Smart Cities (FME ZEN) and Energy for domestic hot water in the Norwegian low emission society under VarmtVann2030 within EnergiX program.

References

- [1] Dahl M, Brun A, Andresen GB. Using ensemble weather predictions in district heating operation and load forecasting. *Applied Energy*. 2017;193:455-65.
- [2] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*. 2012;16(6):3586-92.
- [3] Jovanović RŽ, Sretenović AA, Živković BD. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*. 2015;94:189-99.
- [4] Catalina T, Iordache V, Caracaleanu B. Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*. 2013;57:302-12.
- [5] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*. 2018;81:1192-205.
- [6] Wang Z, Srinivasan RS. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*. 2017;75:796-808.
- [7] Catalina T, Virgone J, Blanco E. Development and validation of regression models to predict monthly heating demand for residential buildings. *Energy and Buildings*. 2008;40(10):1825-32.
- [8] Bauer M, Scartezzini JL. A simplified correlation method accounting for heating and cooling loads in energy-efficient buildings. *Energy and Buildings*. 1998;27(2):147-54.
- [9] Aranda A, Ferreira G, Mainar-Toledo MD, Scarpellini S, Llera Sastresa E. Multiple regression models to predict the annual energy consumption in the Spanish banking sector. *Energy and Buildings*. 2012;49:380-7.
- [10] Farzana S, Liu M, Baldwin A, Hossain MU. Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China. *Energy and Buildings*. 2014;81:161-9.
- [11] Lindberg KB, Doorman G. Hourly load modelling of non-residential building stock. *Conference Hourly load modelling of non-residential building stock*. p. 1-6.
- [12] Ryan SE, Porth LSJGTRR-G-FC, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 41 p. A tutorial on the piecewise regression approach applied to bedload transport data. 2007;189.
- [13] Energy LPLMoBiM, Distribution Systems PDD, Dept. Process Engineering,, The Norw. Univ. of Science and Techn.(NTNU) T, 2007
- [14] Djuric N, Novakovic V. Identifying important variables of energy use in low energy office building by using multivariate analysis. *Energy and Buildings*. 2012;45:91-8.
- [15] Altman DG, Gardner MJJBmj. *Statistics in Medicine: Calculating confidence intervals for regression and correlation*. 1988;296(6631):1238.