Identifying important variables of energy use in low energy office building by using multivariate analysis

Natasa Djuric[1*], Vojislav Novakovic[2]
[1]SINTEF Energy Research, Department of Energy Processes, NO-7465 Trondheim
[2]Norwegian University of Science and Technology, Department of Energy and Process Engineering, NO-7491 Trondheim, Norway
[*]E-mail address: natasa.djuric@sintef.no. Phone number: (+47) 73593338.

Abstract

The aim of the study was to indentify driving variables that contributed to energy use in a low energy office building by integrating building energy management system (BEMS) and energy use data. To take a further step towards zero emission buildings, it is necessary to identify what contributes the most to building energy use. Further, the idea was to encourage a smart use of BEMS data for energy use analysis. Principal component regression and partial least squares regression were used for the data analysis. Databases of 76 and 41 variables respectively, which included occupancy level, control signals, and water and air temperatures, were used to explain heating, electricity, and fan energy use. Variable contributions to the principal components were used to simplify the model and to find the most important variables. In this way, energy use was defined indirectly by using available variables in BEMS. The approach was tested on a low energy office building. The results showed that important variables were different for different months in the case of heating energy use. The total electricity and fan electricity use could be defined with the same variables in different months. The total electricity use could be defined by using occupancy level and input fan signals. The suggested approach could be used by building operators to identify opportunities for decreasing energy use and for energy use estimation when data are lost due to data transmission issues or other problems. A relationship between building information and energy use was established.

Key words: energy use, driving variables, principal components, low energy building, occupancy level

1. Introduction

The zero emission concept ‘‘represents a shift from the traditional industrial model in which wastes are considered the norm, to integrated technologies utilizing everything. It advocates an industrial transformation whereby businesses emulate the sustainable cycles found in nature and where society minimizes the load it imposes on the natural resource base and learns to do more with what the earth produces’’ as pointed out in [1, 2]. Further, in the work of Ulgiati et al. [1] it is emphasized that Zero Emission Strategies for the industry are significantly dependent on information for optimum use of resources, and therefore, the cost of generating, testing, disseminating, and storing information is of paramount importance for sustainability [1]. Even though use of information to achieve a zero emission building is necessary, it is still not completely utilized in everyday building design and operation practice. Therefore, the aim of the study was to encourage a smart use of building energy management system (BEMS) data by relating BEMS information to building energy use.

One of the smart grid concept’s features is increased use of digital information and control technology to improve the reliability, security, and overall efficiency of the electric system [3]. Further, the smart grid concept should provide consumers with timely information and control options.

Intelligent and advanced use of BEMS for energy efficiency and improvement in thermal comfort has been highly recognized, but still not fully utilized. For example, significant improvements in building comfort, tenant productivity, and the reduction in the labor due to implementation of information monitoring and diagnostic system are reported in [4]. An intelligent decision support model for assessing energy-saving measures by using BEMS information is reported in [5]. There are many studies and research work that are reporting advanced use of BEMS either for the entire building or HVAC components [6-9]. In general, much data are available in BEMS that should be better utilized for energy efficiency and improvement in thermal comfort.

Different methods are available to analyze and handle data organized into databases, like regression methods, probabilistic methods, neutral networks, and data mining techniques. They are used for both prediction, and fault detection and diagnosis in HVAC. For example, data mining method is used to predict total building energy demand in [10] and to identify residential load in the smart grid context in [11]. A probabilistic approach combined with the free-running concept temperature are used to describe dynamic behavior by steady-state concepts and predict building energy use in [12]. Principal component regression (PCR) and partial least squares regression (PLSR) present the application of principal component analysis (PCA) in linear models. Successful application of the PCA for fault detection and diagnosis of sensor problems in HVAC is shown in [13-15]. Further, PCA has been used for estimating trends in building heating and cooling load [16] and to analyze electricity consumption in residential dwellings [17] and office buildings [18]. Finally, PCA and PLS are used to identify significant variables of high-rise office building energy use [19] and multifamily buildings [20]. PCA and PLSR implementation for model predictive controller for condition monitoring is presented in [21]. PCA can be used to recover faulty data as shown in [22]. Since PCA has been widely applied for different applications, it was also suggested to in this study.

In this study, databases for prediction of heating energy use and electricity use were developed with BEMS data. Even though BEMS data differ in nature, temperature, control signals, pressures, etc., they can be correlated. For example, in office buildings at mid-day, outdoor temperature is usually higher, equipment is turned ON, and occupancy level is higher, while during night-time the situation is usually the opposite. In addition, BEMS data are correlated to time. Therefore, it could be useful to decouple data and establish new variables that would be uncorrelated. These new uncorrelated variables should be used to define building energy use. PCA may also be used to analyze time series, if variables of time are included as predictor variables [23].

This paper consists of four parts. The first part briefly introduces PLSR and PCR methods. Building description and predictor databases are introduced in the second part. Results including method comparison, model scaling, and defining driving variables, are given in the third part of the paper. In the fourth part of the paper, the suggested method was validated to recover energy use data that could be lost due to data transmission problems.

2. Methodology

PLSR and PCR are both methods to model a response variable when there are a large number of predictor variables, and those predictors are highly correlated or even collinear. Both methods construct new predictor variables, known as principal components (PCs), as linear combinations of the original predictor variables.

2.1. Principal components regression

In principal component regression, the predictor variables are first transformed to PCs. Further, these principal components are regressed against the original responses [23]. If X is the predictor matrix and y is the response vector, PCs of original predictors can be expressed as:

$$z = xU \tag{1}$$

where $U'U = I$. To perform PCR, transformed original predictors into PCs are used to express response as:

$$y = zb_z \tag{2}$$

where b_z denotes the regression coefficients obtained by using PCs. The regression coefficients relating the PCs to the responses will have minimum standard errors since the predictors are uncorrelated and, for the same reason, the regression coefficients will be uncorrelated [23].

In practical problems, as studied here, it can be more convenient to relate the responses to the original variables as follows:

$$y = zb_z = xUb_z \tag{3}$$

since $z = xU$, and therefore:

$$b = Ub_z \tag{4}$$

where b denotes the regression coefficients for the original variables.

2.2. Partial least squares regression

PLSR is another attempt to obtain components in a manner that directly reflects the relationship between the predictor and the response. The use of PLSR is useful in any application that has multiple predictors. PLSR takes into account the case of single-response variable and multiple responses. The PLSR technique operates in the same way as PCR in that a set of vectors are obtained from the predictor variables. PLSR is different from PCR because it immediately relates to the responses and reduces variability among the predictors. The estimation of the next vector takes that relationship into account. Simultaneously, a set of vectors for the responses is also being obtained that takes this relationship into account. The procedure for PLSR is considerably more complicated than PCR [23].

If X is the predictor matrix of size $n \times p$ and Y is the response matrix of size $n \times q$. The principal feature of the PLSR technique is that *two* operations will be carried out together:

$$X = TP + E \quad (T, E \; are \; n \times k, P \; is \; k \times p) \tag{5}$$

$$Y = UQ + F^{*} \quad (U, F^{*} \; are \; n \times k, Q \; is \; k \times q) \tag{6}$$

$k \leq p$ is the number of vectors associated with X. E is the matrix of residuals of X at the *k*th stage. F^{*} is an intermediate step in obtaining the residuals for Y at *k*th stage. The matrices P and Q are the characteristic vectors. The matrices T and U are referred as "X-scores" and "Y-scores", respectively. In PLSR, a prediction equation is formed by replacing U by TB (B is $k \times k$), thus producing:

$$Y = TBQ + F \tag{7}$$

In this study, PCR and PLSR equations, as given in Eq. (4) and (7) respectively, were used to relate BEMS data to building energy use. BEMS data were used as original predictor variables, while energy use data were response or target variables. In addition to BEMS data, date and time were introduced into the predictor databases.

To decrease the number of variables and find the most influencing, the PLS weights and PC loadings were used. The PLS weights are the linear combinations of the original variables that define the PCs in the PLSR. Actually, they describe how strongly each component in the PLSR depends on the original variables. Similarly, the PC loadings describe how strongly each component in the PCR depends on the original variables.

3. Database and building description

The suggested approach was tested on a low energy office building in Trondheim, Norway, with a heated area of 16 200 m^2. The ventilation system consisted of eight variable air volume systems, with a maximum air volume from 12 500 m^3/h to 22 000 m^3/h. Heating was provided by radiators, while cooling of IT rooms was provided by fan-coils. Heating energy for ventilation, space heating, and domestic hot water was supplied by district heating and heat pumps. There were two heat pumps installed. One of the heat pumps was providing part of the heating energy for ventilation in the winter period, while in the summer period the evaporator of the heat pump provided cooling for ventilation. The second heat pump was a cooling plant which provided cooling for IT rooms, while the condenser heat was utilized to support heating.

Available data from BEMS were used as predictor variables, while the heating energy use, the total electricity use, and the electricity for fans were used as target variables. BEMS data that were used for the predictor variable databases are given in Table 1. Only variables for the first air handling unit (AHU) are presented in Table 1, since the variables for the seven remaining AHUs were the same, although with different values.

Table 1. Database description predictor variables

Variable name	**Description**	**Value range**	**Application**
Day	Day of week	1 for working, 2 for nonworking	H*,El*,Fan*
Hour	Hour	0 – 1, $\sin(\pi \cdot \mathrm{Hour}/24)$	H*,El*,Fan*
Tout	Outdoor temperature	-20 – 30 ^{o}C	H*,El*,Fan*
Tin_R4031	Indoor temperature in the 4th floor office	18 – 23 ^{o}C	H*,El*,Fan*
Tin_R4010	Indoor temperature in the 4th floor office		H*,El*,Fan*
Tin_R4099	Indoor temperature in the 4th floor office		H*,El*,Fan*
OCC_R4031	Occupancy level in the 4th floor office	0.5 (not occupied), 1 – 1.5 (bypass), 3 (occupied)	H*,El*,Fan*
OCC_R4010	Occupancy level in the 4th floor office		H*,El*,Fan*
OCC_R4099	Occupancy level in the 4th floor office		H*,El*,Fan*
320.SB40	Valve position in the main branch	0 – 100 %	H*
320.RT40	Supply temperature in the main branch	30 – 70 ^{o}C	H*
320.RT50	Return temperature in the main branch	30 – 60 ^{o}C	H*
320.02.RT40	Supply temperature in floor heating	20 – 35 ^{o}C	H*
320.02.RT50	Return temperature in floor heating	20 – 30 ^{o}C	H*
320.03.SB40	Valve position for snow melting	0 – 100 %	H*
320.03.RT40	Supply temperature for snow melting	20 – 35 ^{o}C	H*
320.03.RT50	Return temperature for snow melting	15 – 25 ^{o}C	H*
320.04.SB40	Valve position in the radiator branch	0 – 100 %	H*
320.04.RT40	Supply temperature in the radiator branch	30 – 70 ^{o}C	H*
320.04.RT50	Return temperature in the radiator branch	25 – 55 ^{o}C	H*
36.01.LK	Valve position at heating/cooling coil	0 – 100 %	H*
36.01.LV	Valve position at heating coil	0 – 100 %	H*
36.01.RT55	Return temp. after LV AHU	20 – 50 ^{o}C	H*
36.01.LX01	Input signal for recovery wheel	0 – 100 %	H*,El*,Fan*
36.01.JV40	Input signal for supply fan	0 – 100 %	H*,El*,Fan*
36.01.JV50	Input signal for exhaust fan	0 – 100 %	H*,El*,Fan*
36.01.RT40	Supply air temperature	16 – 24 ^{o}C	H*,El*,Fan*

H – heating energy use, El – electricity use, Fan – fan electricity use

The BEMS of the analyzed building had only a history on control and measurements (temperature, pressure, el. signals, etc.), while energy monitoring was transferred to an energy service company. Monitoring of the energy consumption in the energy service database was done on an hourly basis. Therefore, the data in Table 1 were calculated as hourly mean values. The target variables, the heating energy use and the total electricity use, were measured by the energy provider's equipment and transferred to the energy service company's database. The target variable named "fan electricity use" is the total electricity use in the eight AHUs. This fan electricity use included electricity for two fans (supply and exhaust), heat recovery wheel, and auxiliary devices in each AHU. Electricity use of the entire AHU was measured on an hourly basis and transferred to the energy service company. There was no possibility to separate fan electricity use. Therefore, available BEMS signals were related to the AHU electricity use to find driving variables.

4. Results

Two introduced methods, PLSR and PCR, for modeling of the target variables were compared. Beside method effectiveness comparison, models based on the entire database and 10 variables were compared. Model scaling from a model based on the database to a model based on 10 variables was performed based on the predictor variable contribution to PCs. The model scaling was used to indentify driving variables. Finally, regression models and model coefficients are presented. The energy use data for three months, March, July, and November were analyzed, since for these three months the energy use data and databases were obtained.

4.1. PLSR versus PCR regression models

The database of the predictor variables for heating energy use consisted of 76 variables as presented in Table 1. The database of the predictor variables for electricity and fan electricity use consisted of 41 variables. The same database for electricity and fan electricity use was implemented. Use of the entire databases to calculate target variables could be demanding and requires specific computer programs to perform the calculation. A smaller database of predictor multi-variables could be simple for practical use and presentation of influential parameters on energy use. To introduce the approach for decreasing the number of variables gradually, method effectiveness are presented first. The method effectiveness was estimated by using the model accuracy. Similar method for determining the number of PCs and simplifying the model by using variance of reconstruction error is proposed in [24]. Accuracies for the heating energy use model for both regression methods and different amount of data are presented in Figure 1. In Figure 1, accuracy is presented by coefficient of variation of the root mean squared error (CV(RMSE)), which was estimated by using 10-fold cross validation.

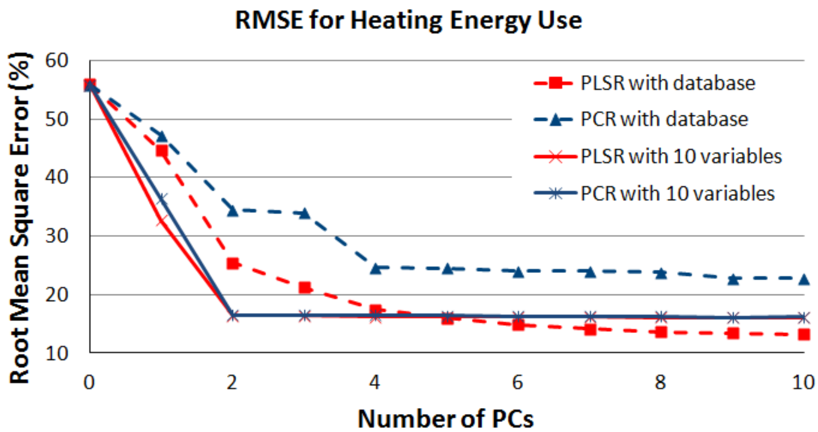

Figure 1. Accuracy of the heating energy use model for different amount of data and different methods

In Figure 1, the model accuracies are presented as a function of the number of used PCs to model the target variable. Models are declared to be calibrated if they produce CV(RMSE) within ±30% when using hourly data [25]. This means, that the models developed by using the entire database in Figure 1 have acceptable accuracy when four PCs were used, while the models described with 10 variables produced acceptable accuracy already with two PCs. In Figure 1, it is also possible to notice that the 10 variable models had faster improvement than the models based on the database. This faster improvement of the simpler models indicated that there were redundancy and mutual correlation among the variables in the database. In Figure 1, when comparing the regression methods, it is possible to notice that PLSR has a faster improvement than PCR, either by using the entire database or 10 variables. This result was expected, since in the PLSR method PCs are obtained by directly reflecting the relationship between the predictor and the response [23]. In the PCR method, the PCs explain only variation in the predictor variables, with no regard to the target variables. Since the PLSR method gave a faster model improvement with a better accuracy, it will be used further in the study. Model accuracies for the electricity and the fan electricity use are displayed in Figures 2 and 3, respectively.

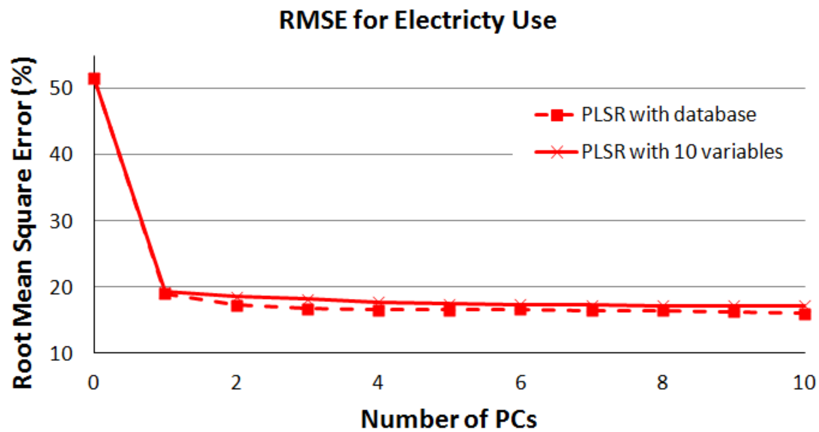

Figure 2. Accuracy of the electricity use model for different amount of data obtained using PLSR

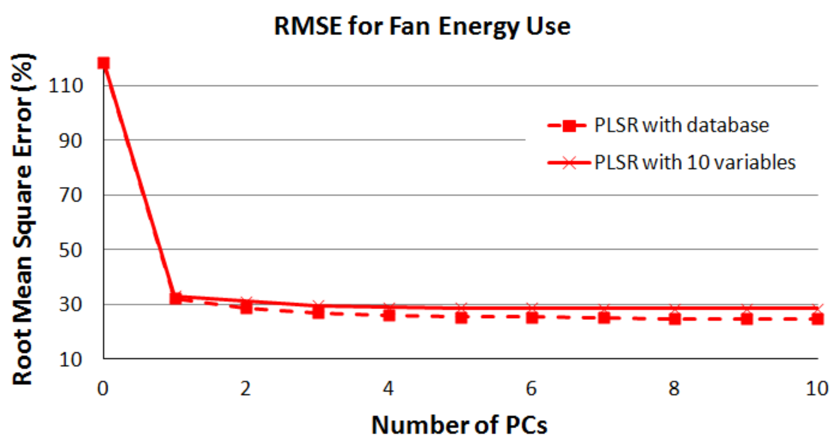

Figure 3. Accuracy of the fan electricity use model for different amount of data obtained using PLSR

Results in Figures 2 and 3 show that acceptable model accuracies were achieved when using both the entire database and the 10 variables. The high accuracy of the total electricity use model in Figure 2 indicates that the 41 variables in Table 1 could be highly related to the electricity use. Among the 41 variables in Table 1, 24 variables were the AHU electrical input signals. Even though the database in Table 1 provided more data related to the building ventilation system, the achieved accuracy of the total electricity use model in Figure 2 was higher than the accuracy of the fan electricity use model in Figure 3. There were two possible reasons for this issue. The first one might be due to different type of energy measurements for the total electricity use and the electricity use for AHUs (in the text named fan electricity use as explained in Section 3). The total electricity use was measured by the energy provider, while the electricity for AHUs was measured by the energy service company. Further, even thought AHU measurement was labeled to measure electricity for AHUs, it might happen that the measurement included additional users, like light in technical room, electricity for control devices, etc. However, the variables in Table 1 did not explain these additional users and auxiliary devices.

In Figures 1, 2, and 3 it was shown that the models with 10 variables could achieve acceptable accuracy and even faster improvement due to decreased redundancy. These 10 variables were chosen based on partial least squares (PLS) weights. The original variable relations to PLS weights are explained in the next section.

4.2. Model scaling and defining driving variables

To simplify the models based on the databases and to find the most influencing variables, values of PLS weights for the first four PCs were used. 97 % of the model variance was explained in the first four PCs for the heating energy use model. 99 % of the model variance was explained in the first four PCs for the electricity use model. Therefore, the first four PCs were assumed to be sufficient for this analysis. The procedure of defining the driving variables consists of two parts. First, matrices of the original variables defined in Table 1 were normalized. Afterwards, the first 10 variables that had highest contribution to the first four PCs were chosen as the model important or driving variables. The same procedure was repeated for each target variable. To prove that the first 10 variables chosen in this way were important variables, contribution percentage to the model was also calculated.

PLS weights on the first four PCs for the heating energy use, the total electricity use, and the fan electricity use are displayed in Figures 4, 5 and 6, respectively. If the PLS weights of an observed variable have higher values, then that variable has high a contribution to PCs. Consequently, it can be concluded that the observed variable contributes more to the target variable. In Figure 4, 5, and 6 the driving variables for the energy use in November are shown.

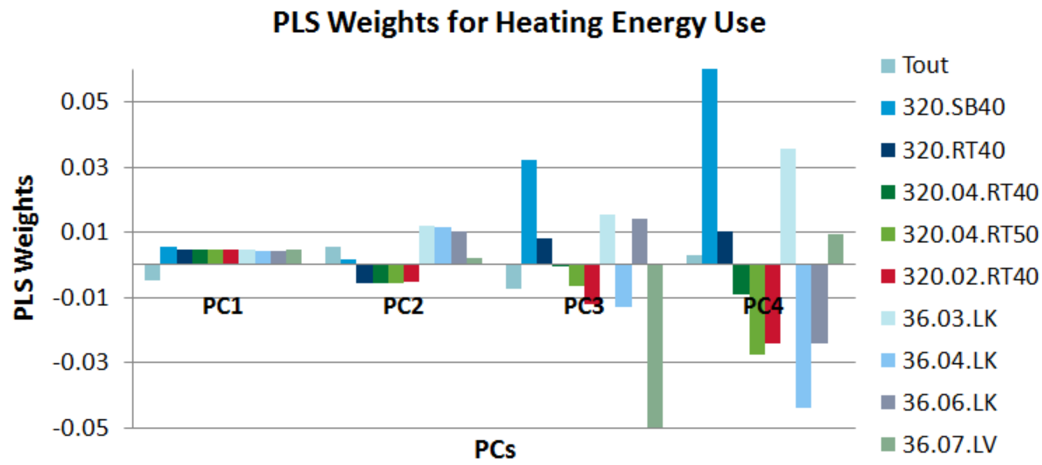

Figure 4. PLS weights of 10 important variables for heating energy use model

By using the procedure for model scaling and finding driving variables based on PLS weights, it was found that the most important variables of the heating energy use are outdoor temperature, control parameters and temperatures in the substation, and some of the ventilation parameters. These ventilation parameters were related to the AHUs that were mostly in use. In Figure 4, PLS weights of the different variables on the first and second PC had quite similar values, while on the third and fourth PLS weights were quite different. Therefore, the values of PLS weights on the third and fourth PC could explain the variable importance. Based on that, it is possible to conclude that the heating energy use was influenced by the operation parameters rather than by the outdoor temperature.

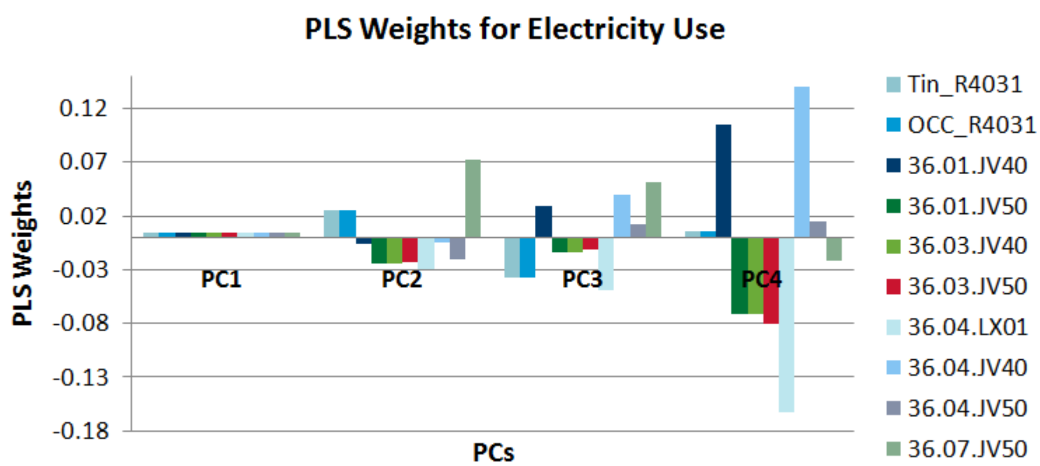

Figure 5. PLS weights of 10 important variables for electricity use model

Results in Figure 5 show variables that could explain the total electricity use in November. Among the 10 variables, occupancy level, indoor temperature, and parameters of the fourth AHU were variables that could explain the electricity use well. The fourth AHU was supplying the most typically occupied part of the building. Results in Figure 5 shows that the occupancy level could be included in the electricity use model. In this model, occupancy level and indoor temperature had significant contribution already on the second and third PC. Negative values of PLS weights of the occupancy level and indoor temperature should not imply directly negative influence on the target variable, because the original variable matrix was normalized. In the case of the normalized original variable matrix, values of PLS weights should indicate variable importance to the model.

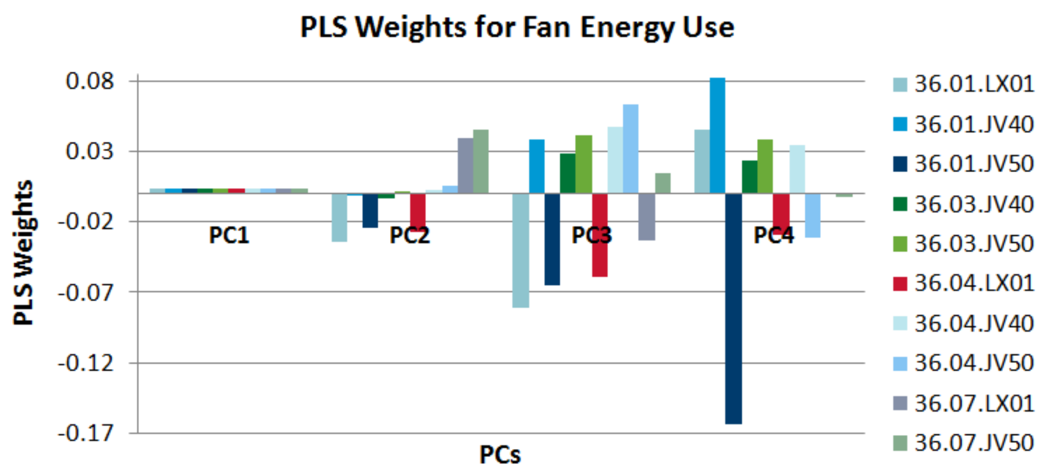

Figure 6. PLS weights of 10 important variables for fan electricity use model

Results in Figure 6 show that, except for the input signal to the recovery wheel of the first AHU at the third PC, the most contributing variables to the AHU electricity use were the input signals to the supply and exhaust fans. This indicated that the electricity use of AHUs could be explained by the fan use. Consequently, it can be concluded that the electricity use for the AHUs could be decreased by influencing the fans.

To prove importance of the 10 variables chosen as driving variables, the contribution percentage to the target variable was calculated. In Figure 7, the contribution percentage for the total electricity use in November is displayed.

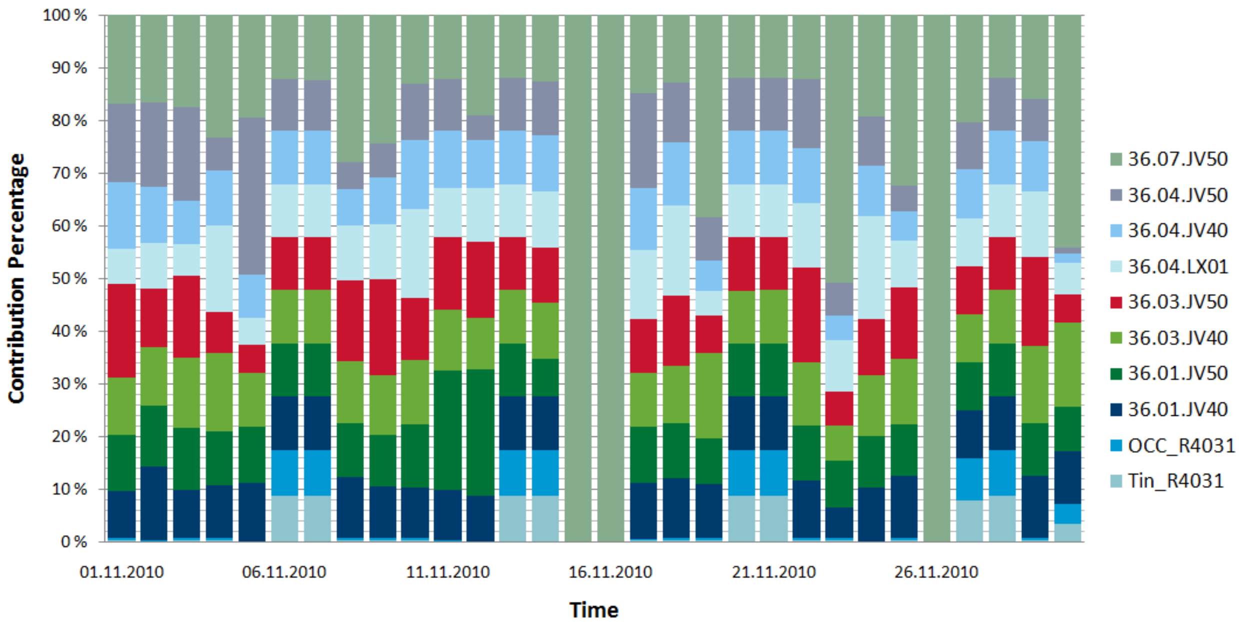

Figure 7. Contribution percentage of the 10 most important variables to the electricity use

If arbitrarily predictor variables would be chosen to explain target variable, their contribution percentage would not be uniform over time as in Figure 7, even though the model could achieve acceptable accuracy. Uniformity of the contribution percentage means that contribution percentage in Figure 7 had similar values for working days and different but similar values for weekends. During three days in November, 15th, 16th, and 26th, BEMS historical data were lost except for the seventh AHU. However, variance in the input signal of the exhaust fan in the seventh AHU was enough to explain the total electricity use in November. Results in Figure 7 show that BEMS data could be used for better understanding and explaining building energy use.

4.3. Regression models

Regression models and model coefficients of the target variables are presented in this section. Firstly, model comparison based on the database and 10 variables is presented. Afterwards, model coefficients for the regression models with 10 variables are presented. The model coefficients are given for March, July, and November. Comparisons of the heating energy models are displayed in Figures 8 and 9, while comparison of the electricity use model in November is displayed in Figure 10.

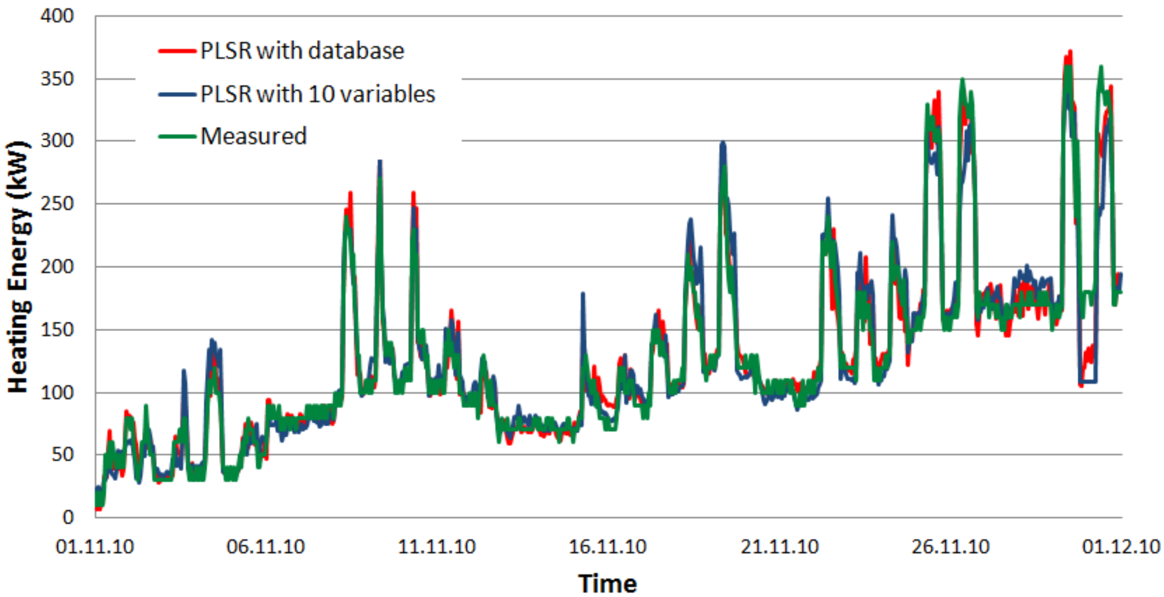

Figure 8. PLSR models of the heating energy use

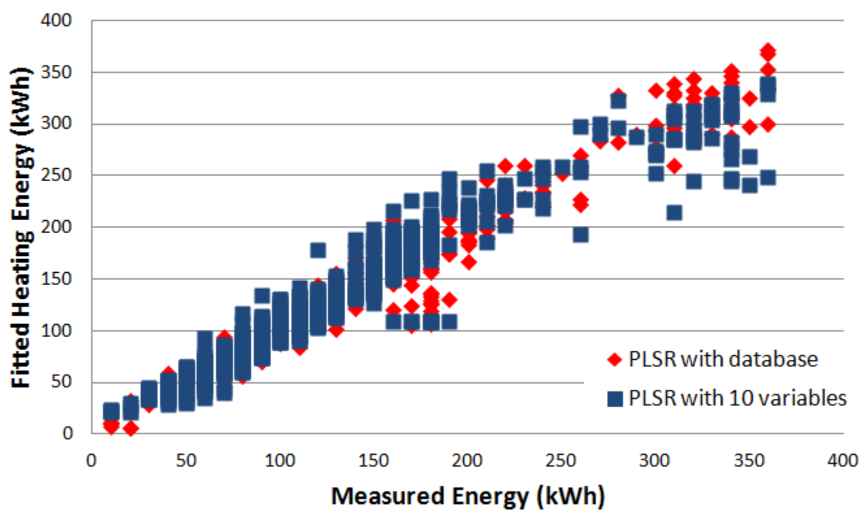

Figure 9. Comparison of the PLSR models based on database and 10 variables

Results in Figures 8 and 9 compared both models for the heating energy use, with the entire database and with 10 variables, with the measured data. Both models with the entire database and with 10 variables were developed by using 10 PCs. In Figures 8 and 9, it is possible to notice a small difference between the models and the measurements. As shown in Figure 1, the accuracy of the heating energy use model with the database was higher when modeled with 10 PCs. Specifically, CV(RMSE) was 13.3 % for the model based on the database, and

CV(RMSE) was 16.1 % for the model based on 10 variables. The small difference between models and acceptable accuracy of the model with 10 variables indicated that use of the model with 10 variables might be acceptable in practice and for indentifying driving variables.

Figure 10. PLSR models of the electricity use

In Figure 10, both models with the entire database and with 10 variables were developed by using 10 PCs. Results in Figure 10 confirmed results on the accuracy of the electricity use model from Figure 2, where it was shown that CV(RMSE) was 16 % for the model based on the database, and CV(RMSE) was 17.2 % for the model based on 10 variables. The acceptable model accuracy indicated that these driving variables from BEMS could be used to explain building energy use and as an additional energy estimation tool.

To directly utilize BEMS data for building energy use, as explained in the previous text, it is necessary to have available model coefficients. The model coefficients were estimated directly from BEMS training data, with no data normalization. After several tests, it was found that the heating energy use model with 10 variables required different driving variables in different months. The electricity use and the fan electricity use model could be explained with the same variables in different months. The model coefficients for the heating energy use model are given in Table 2, while the model coefficients for the electricity use model and the fan electricity use model are displayed in Figures 11 and 12.

Table 2. PLS model coefficients for the heating energy use model with 10 variables

Month					
March		**July**		**November**	
Variables	Value	Variables	Value	Variables	Value
	-15.1207		0.1956		-149.9317
T_{out}	-0.2682	Day	0.0186	T_{out}	0.0264
320.SB40	2.1546	Hour	0.0165	320.SB40	1.8980
320.04.SB40	0.1079	T_{out}	-0.2106	320.RT40	3.6985
320.04.RT40	5.3074	T_{in}_R4031	-0.0050	320.04.RT40	0.5511
320.04.RT50	-5.1055	T_{in}_4010	0.0320	320.04.RT50	-1.8820
320.03.SB40	3.6336	T_{in}_4099	0.0288	320.02.RT40	3.6833
320.03.RT40	0.8551	OCC_R4031	-0.0050	36.03.LK	0.8813
320.03.RT50	-0.4869	OCC_R4010	0.0320	36.04.LK	-0.0032
36.02.LK	0.5160	OCC_R4099	0.0288	36.06.LK	0.6239
36.02.LV	0.5160	36.01.LX01	-0.0004	36.07.LV	0.5708
CV(RSME) (%)	15.84	CV(RSME) (%)	154.94	CV(RSME) (%)	16.14

In Table 2, the original variables with corresponding coefficients for the heating energy use model for three months are given. Also, the model accuracies are presented in Table 2. As mentioned before, models are declared to be calibrated if they produces CV(RMSE) within ±30% when using hourly data [25]. In Table 2, the PLSR model for the heating energy use in July had five times worse accuracy than allowed. By using the entire database, better model accuracy was not possible to achieve. Therefore, it can be concluded that the available 76 variables were not capable to describe the heating energy in July. Unfortunately, hot tap water data were not available in the BEMS history. Use of these data might improve the model in July.

Unlike the heating energy use model, both the models for electricity use and fan electricity use could be explained with the same variables in different months, as shown in Figures 11 and 12. Model accuracy for both models in each month was acceptable. Specifically, for the electricity use model CV(RSME) was 23.2% for March, 11.3% for July, and 17.2% for November. For the fan electricity use, CV(RSME) was 26% for March, 22.9% for July, and 28.5% for November.

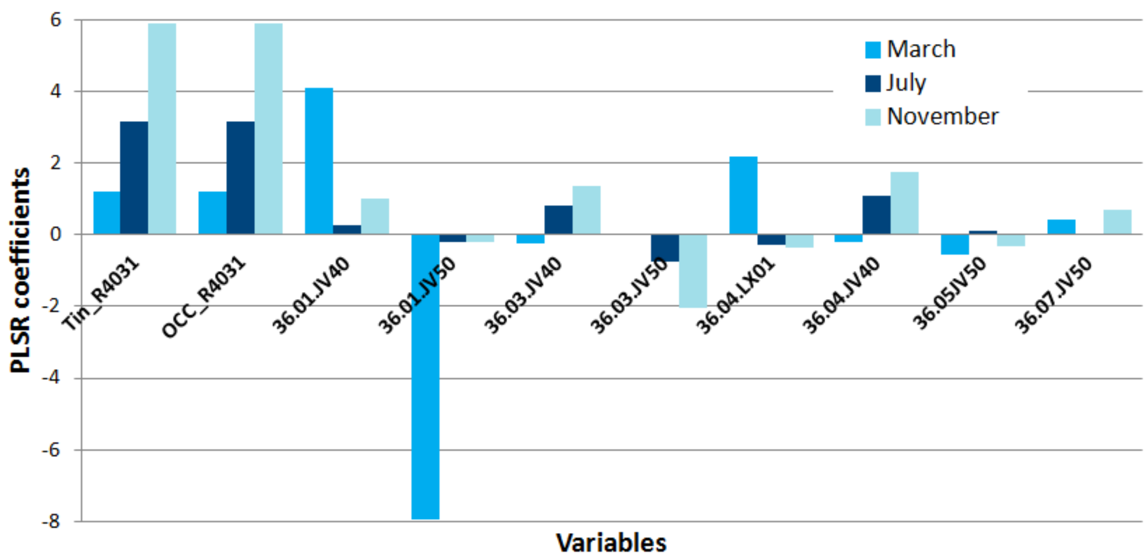

Figure 11. PLSR model coefficients for electricity use model with 10 variables

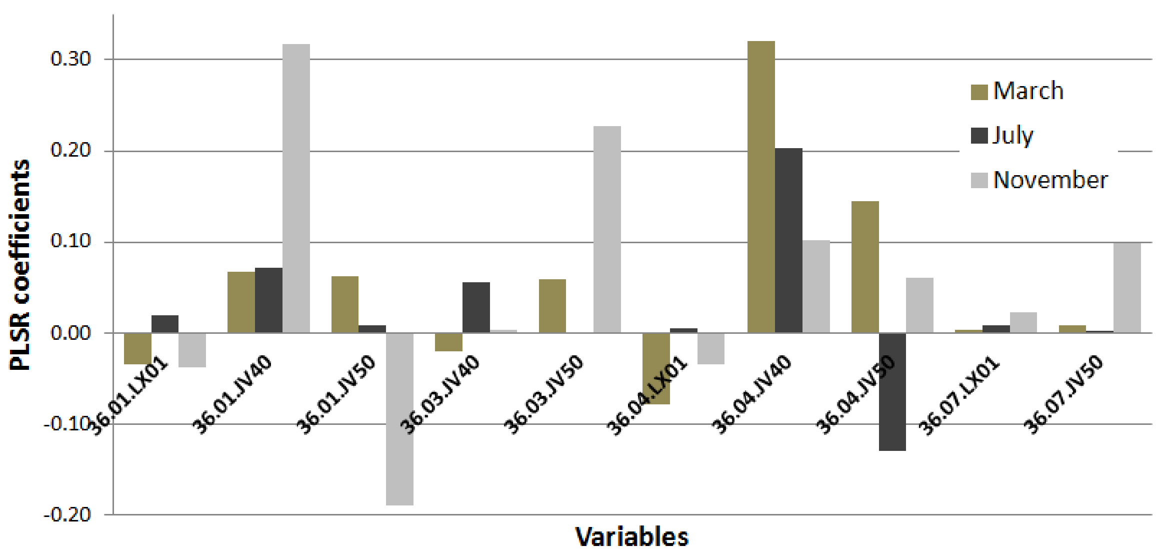

Figure 12. PLSR model coefficients for fan electricity use model with 10 variables

One of the reasons for the different coefficients for the different months could be explained by a different building use during the year. The analyzed building was in use since September 2009, and analyzed data were from 2010. The number of occupants from March to November 2010 increased about 60%. Another reason could be operation adjustment, since the building was still new. However, these different regression coefficients for the different months indicated that the regression models should be updated on a monthly basis.

5. Verification and data recovery

The developed approach for using BEMS data to explain the building energy use was verified on a data recovery problem. In the work of Hao et al. [22] with data recovery is meant replacing faulty data using PCA. In this study, the suggested approach implied recovering the energy use data that were lost due to data transmission problems. In the work of Sjorgen et al. [26], the problem of missing data on the monthly building total energy use is solved by using consumption profiles. Energy use data from BEMS or energy provider are used by energy service companies for energy efficiency improvements. Such data can be corrupted or partially lost after data transmission. However, energy service companies usually do not implement any robust method to correct data, yet use simple data interpolation or do not correct data. The developed approach in this study could be used for the purpose of data recovery. To validate the suggested approach, three methods for data recovering were compared: data recovering based on PCR, data recovering based on PLSR, and simple data recovering. Simple data recovering means that lost data are replaced by the column's mean, if the energy use data on an hourly basis are organized like: columns represent hours during the day (from 1 to 24) and rows are day of month.

For the validation purpose, hourly data for November were separated into two groups: training and recovering data. The training data were used to develop model coefficients that were used afterwards on the recovering data. Regarding the target variable or the energy use data, one part of the data was used for training, while the other was treated as lost data due to transmission problem. Robustness of the suggested approach was proven by increasing the number of lost data. In Figures 13 and 14 are displayed estimation error of the heating energy use and the electricity use, respectively. Error in the energy use estimation was calculated by comparing real energy use with the recovered energy use for November, when the total heating energy use was 90 980 kWh and the total electricity use was 78 725 kWh.

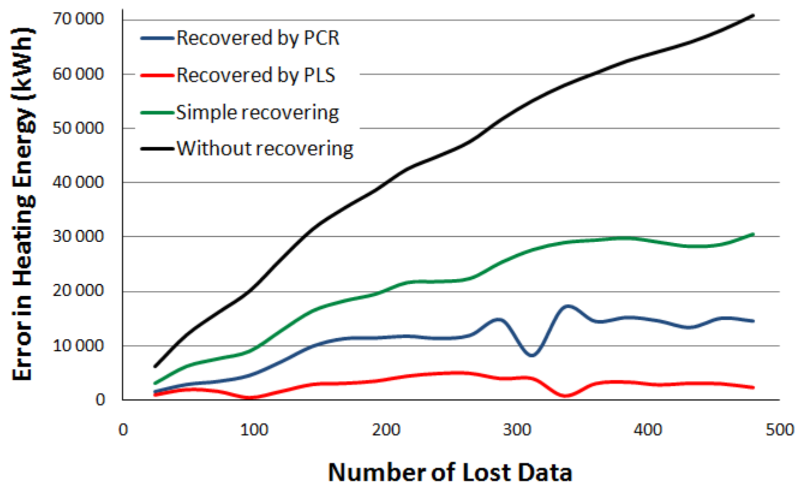

Figure 13. Error in the heating energy use estimation due to implementation of different data recovering methods

Results in Figure 13 show that the PLSR method gave the best results in data recovering. For example, in the case when 120 hours of data (17 %) were lost, and if the data could not be recovered, the error in the total heating energy use estimation would be about 28 %. If the simple recovering would be implemented, the error in the heating energy estimation would be about 14 %, while use of the PLSR and BEMS data would produce an error of only 2 %. In Figure 13, it is also possible to notice that the error in the heating energy estimation based on the PLSR was not increased when the amount of lost data was increased, which proved robustness of the suggested approach.

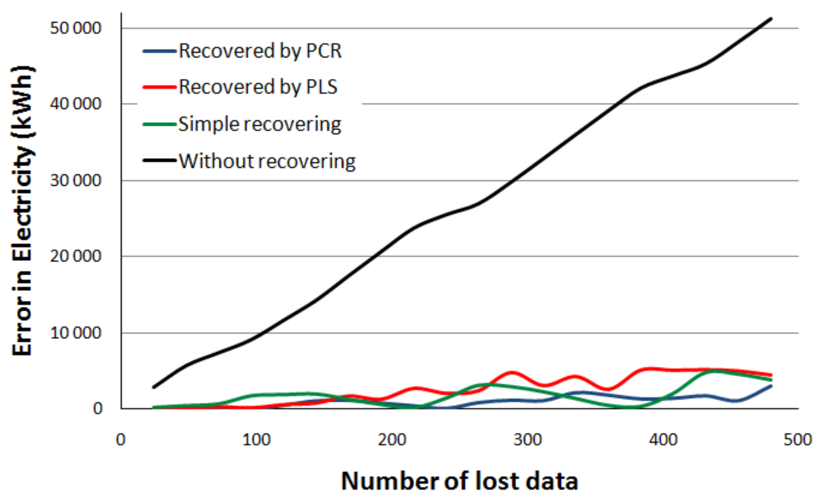

Figure 14. Error in the electricity use estimation due to implementation of different data recovering methods

Results in Figure 14 show that all three methods were good in data recovering, while recovering based on the PCR method gave the best results. The reason that even simple data recovering based on mean values gave good results could be explained by the fact that the total electricity use had quite periodical values as shown in Figure 10. Results in Figures 13 and 14 show the importance of energy use and information on energy use.

6. Conclusions

The study presents a smart approach to utilize BEMS data for energy use estimation and identification of driving variables of the energy use. The idea was to relate building information with the building energy use. PCR and PLSR were used to relate BEMS data to the building energy use. BEMS data were used as original predictor variables, while the heating energy use, the electricity use, and the fan electricity use were target variables. To simplify models and find the most influencing variables, the values of PLS weights for the first four PCs were used. The suggested approach was tested on the low energy office building located in Trondheim.

The results showed that the heating energy use in the low energy office building was influenced by the operation parameters rather than by the outdoor temperature. The total electricity use could be explained by using occupancy level, indoor temperature, and some of the AHU electrical signals. The AHU electricity use could be explained by using the input electrical signals of supply and exhaust fans. However, results indicated that the regression models should be updated on a monthly basis. All the simplified regression models with 10 variables had acceptable accuracy. This indicated that driving variables obtained by using the suggested approach could be used to explain the building energy use. Future work should include variables that are easier to obtain or variables easier for occupant's understanding.

The suggested approach in this study could be used as the additional virtual energy measurement tool, to calibrate energy measurements, and to check the quality of energy measurements. Further, this approach could indicate possible reasons of changes in building energy use. Current results showed that the PLSR method was more accurate in recovering the heating energy use, while PCR was more accurate in recovering the electricity use. The results on the method validation showed relationships and importance of data for accurate energy use estimation. For example, 17 % of lost data could imply 28 % error in the heating energy use, while use of PLSR and BEMS data would produce an error of only 2 %.

7. Acknowledgment

This work has been supported by the Research Council of Norway and several partners through the research project ”*The Research Centre on Zero Emission Buildings*” (ZEB). ZEB is one of eleven national Centers for Environment-friendly Energy Research.

8. References

[1] S. Ulgiati, S. Bargigli, M. Raugei, An emergy evaluation of complexity, information and technology, towards maximum power and zero emissions, Journal of Cleaner Production, 15 (13-14) (2007) 1359-1372.
[2] Zero Emissions Forum, United Nations University, Available from: http://unu.edu/zef/.
[3] N. Armaroli, V. Balzani, Towards an electricity-powered world, Energy & Environmental Science, (2011)
[4] M. A. Piette, S. K. Kinney, P. Haves, Analysis of an information monitoring and diagnostic system to improve building operations, Energy and Buildings, 33 (8) (2001) 783-791.
[5]H. Doukas, C. Nychtis, J. Psarras, Assessing energy-saving measures in buildings through an intelligent decision support model, Building and Environment, 44 (2) (2009) 290-298.
[6] N. Djuric, V. Novakovic, G. Huang, Lifetime commissioning as a tool to achieve energy-efficient solutions, International Journal of Energy Research, (2011)
[7] N. Djuric, G. Huang, V. Novakovic, Data fusion heat pump performance estimation, Energy and Buildings, 43 (2-3) (2011) 621-630.
[8] S. Wang, Z. Ma, Supervisory and optimal control of building HVAC systems: A review, HVAC and R Research, 14 (1) (2008) 3-32.
[9] S. Wang, Z. Xu, J. Cao, J. Zhang, A middleware for web service-enabled integration and interoperation of intelligent building systems, Automation in Construction, 16 (1) (2007) 112-121.
[10] Z. Yu, F. Haghighat, B. C. M. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy and Buildings, 42 (10) (2010) 1637-1646.
[11] R. A. S. Fernandes, I. N. Da Silva, M. Oleskovicz. Identification of residential load profile in the smart grid context, IEEE PES General Meeting, PES 2010 , art. no. 5589653, 2010.
[12] C. Ghiaus, Equivalence between the load curve and the free-running temperature in energy estimating methods, Energy and Buildings, 38 (5) (2006) 429-435.
[13] S. Wang, F. Xiao, Sensor fault detection and diagnosis of air-handling units using a condition-based adaptive statistical method, HVAC and R Research, 12 (1) (2006) 127-150.
[14] S. Wang, F. Xiao, Detection and diagnosis of AHU sensor faults using principal component analysis method, Energy Conversion and Management, 45 (17) (2004) 2667-2686.
[15] S. Wang, F. Xiao, AHU sensor fault diagnosis using principal component analysis method, Energy and Buildings, 36 (2) (2004) 147-160.

[16] K. K. W. Wan, D. H. W. Li, D. Liu, J. C. Lam, Future trends of building heating and cooling loads and energy consumption in different climates, Building and Environment, 46 (1) (2011) 223-234.
[17] D. Ndiaye, K. Gabriel, Principal component analysis of the electricity consumption in residential dwellings, Energy and Buildings, 43 (2-3) (2011) 446-453.
[18] J. C. Lam, K. K. W. Wan, K. L. Cheung, L. Yang, Principal component analysis of electricity use in office buildings, Energy and Buildings, 40 (5) (2008) 828-836.
[19] J. C. Lam, S. C. M. Hui, A. L. S. Chan, Regression analysis of high-rise fully air-conditioned office buildings, Energy and Buildings, 26 (2) (1997) 189-197.
[20] T. Olofsson, S. Andersson, J. U. Sjögren, Building energy parameter investigations based on multivariate analysis, Energy and Buildings, 41 (1) (2009) 71-80.
[21] A. AlGhazzawi, B. Lennox, Model predictive control monitoring using multivariate statistics, Journal of Process Control, 19 (2) (2009) 314-327.
[22] X. Hao, G. Zhang, Y. Chen, Fault-tolerant control and data recovery in hvac monitoring system, Energy and Buildings, 37 (2) (2005) 175-180.
[23] J. E. Jackson, A user's guide to principal components. 2003, Hoboken, New Jersey: John Wiley. 569 s.
[24] S. J. Qin, R. Dunia, Determining the number of principal components for best reconstruction, Journal of Process Control, 10 (2) (2000) 245-250.
[25] Measurement of Energy and Demand Savings. Vol. Guideline 14 2002: ASHRAE.
[26] J. U. Sjögren, S. Andersson, T. Olofsson, An approach to evaluate the energy performance of buildings based on incomplete monthly data, Energy and Buildings, 39 (8) (2007) 945-953.