Building portfolio analysis and benchmarking for estimating energy saving potential

Kemme, P.A.M.

Award date:
2015

Title:

Building Portfolio Analysis and Benchmarking for Estimating Energy Saving Potential
Master Thesis

Author:

P.A.M. (Paul) **Kemme** *(0587882)*

In corporation with:

Chair of Building Performance
Building Physics and Services
Architecture, Building and Planning
Eindhoven University of Technology

and

Strukton Worksphere

Supervisors:

Prof.dr.ir. J.L.M. Hensen	(Eindhoven University of Technology)
dr. M.H. Hassan Mohamed	(Eindhoven University of Technology)
B. Tuip, MSc	(Strukton Worksphere)

Abstract: In this paper, a methodology is proposed for benchmarking energy performance of buildings using a set of building data. The methodology can be used or adapted for any dataset of buildings and can be used for assessing energy performance using a set of peer buildings. In this paper, the methodology is tested using data from the United States Commercial Building Energy Consumption Survey (CBECS) 2003, because of the large available number of observations and building parameters. A case study of the Strukton Worksphere office building in Son, the Netherlands, was carried out to test practical implementation. 27 building parameters were selected based on availability and literature review, for consideration in the research. Out of these 27 parameters, 10 parameters were used for categorization of the buildings in the portfolio to define subsets of similar buildings, 14 were used as predictor parameters for the MatLab regression analysis for each subset. Resulting models were compared looking at coefficients of determination (R^2) indicating the reliability of the regression model for calculating predicted energy consumption for the subset. Residuals, defined as the difference between measured energy performance and predicted energy use, were calculated for every building in the subset building category, therewith defining the distribution of energy consumption for this type of building. Assessing a particular building in the subset, the residual of this building is calculated using the regression model determined for the subset, and compared to residual distribution. The methodology shows promising results for developing benchmarking specifically for a given building portfolio. However, more research on the choice of building parameters, improvement of the quality of data and intelligent clustering of buildings for categorization might improve the methodology. The strength of the methodology is in its flexibility to vary the predictor parameters, the possibility to create subsets of buildings of maximum similarity and the ability to compare multiple benchmarks, based on different types of categorization, which can lead to insights on the location of energy saving potential. It is expected that this methodology can also be applied when looking at monthly or daily basis, or zoom in from whole building level to floor or system level and therefore increase the accuracy of determining and locating energy saving potential. This was not tested during the course of this research due to time and data availability restrictions, but the topic is recommended for future research.

Contents

1 Introduction

The European Commission foresees an important role for Energy Service Companies, like Strukton, to achieve energy saving goals and it proposes to publish best practices or benchmarks for energy efficiency [1]. Numerous efforts have been made to benchmark energy performance of buildings for public use. However most of them are based only on building function categorization and floor area normalization [2]. It is expected that other approaches for categorization or normalization can improve the process of energy performance assessment and therefore identify buildings with high energy saving potential. In order to help Strukton improve energy efficiency in buildings in their portfolio, an efficient method to determine energy saving potential can improve internal processes. Because of the commercial nature of Strukton activities, the process of finding this saving potential should be achievable with minimum use of time and resources. The current research aims to offer Strukton or other instances a more specialized benchmarking methodology for assessing energy performance of single buildings through analysis of performance of a set of buildings or building portfolio.

In order to achieve this goal, a methodology was developed to create benchmarks using a set of buildings, a number of building parameters and measured energy consumption data of these buildings. The methodology is of an observational nature and low computational demand is a requirement because of the commercial nature of the methodology. Therefore, the choice of statistical methods is arbitrary and accuracy could be improved. MatLab R2014b and its Statistics and Machine Learning Toolbox were used to build a model for analysis of a dataset and calculation of the benchmarks. The model is a combination of built-in MatLab functions and some straightforward calculations. The aim is to be able to use any given building portfolio as an input and calculate useful benchmarks with minimal adaptations to the model. The number of buildings or observations is expected to be of no issue to the practicability of the model, however a low number of observations can lead to unreliable results.

As mentioned, the aim of the research is to develop a methodology to find benchmark values from a specific set of building data. The Strukton Worksphere building portfolio is one example of such a dataset. Unfortunately, at this moment Strukton building data is not gathered and stored systematically. However, the company is planning on improving on data collection and use this for enhancing building performance. The results of this research can help develop the data collection strategy of Strukton and propose a methodology for analysis of this data. Considering the amount of data needed to find meaningful results, it was decided to use another dataset for testing the benchmarking methodology. The Commercial Building Energy Consumption Survey *(CBECS)* 2003 a publically available dataset of over 5000 buildings (observations) and over 300 building characteristics or measured data (parameters) . After filtering, as explained in Appendix 0, close to 3692 buildings were included in the final dataset used for testing the analysis. In 2012, a new survey was conducted by the same organization [3]. The survey has a similar setup as the 2003 edition, however, at the moment of writing this paper, the data gathered in this survey is not yet released. The survey is conducted in the United States, so differences in climate characteristics, units used and similar differences must be taken into account. Therefore, data of the CBECS 2003 dataset was recalculated to unit standard units used in Europe and the Netherlands and an assumption was made on the US climate type most similar to Dutch climate. For this assumption, the Netherlands was considered one uniform climate zone [4]. Nevertheless, these differences are expected to have minor influence on the workings of the methodology in the Netherlands, although a similar dataset of Dutch commercial buildings might lead to slightly more accurate results.

2 Methodology

The introduction discussed the aim of the current research and the tools used to carry it out. The methodology section consists of a description of the steps carried out towards building the proposed benchmarking methodology, building the MatLab model used for data analysis, and the methods for calculation and interpretation of results for every step. Figure 1 shows the major methodology's stages. In detail, the methodology stages (A, B, C, D and E, Figure 1) consists of 12 steps as shown in Figure 2, along with a short description of the step and, if applicable, the MatLab function used. The subsections of the current section will describe the methods used for each step and the considerations involved. The next section is used to describe the results and findings of these steps. Steps 5 to 10 of research, as shown in Figure 2, are repeated for different subsets for different approaches for categorization or normalization and results for MRA are compared in Step 11 to decide which categorizations and normalizations are best fit to benchmark energy performance for the given dataset or building portfolio.

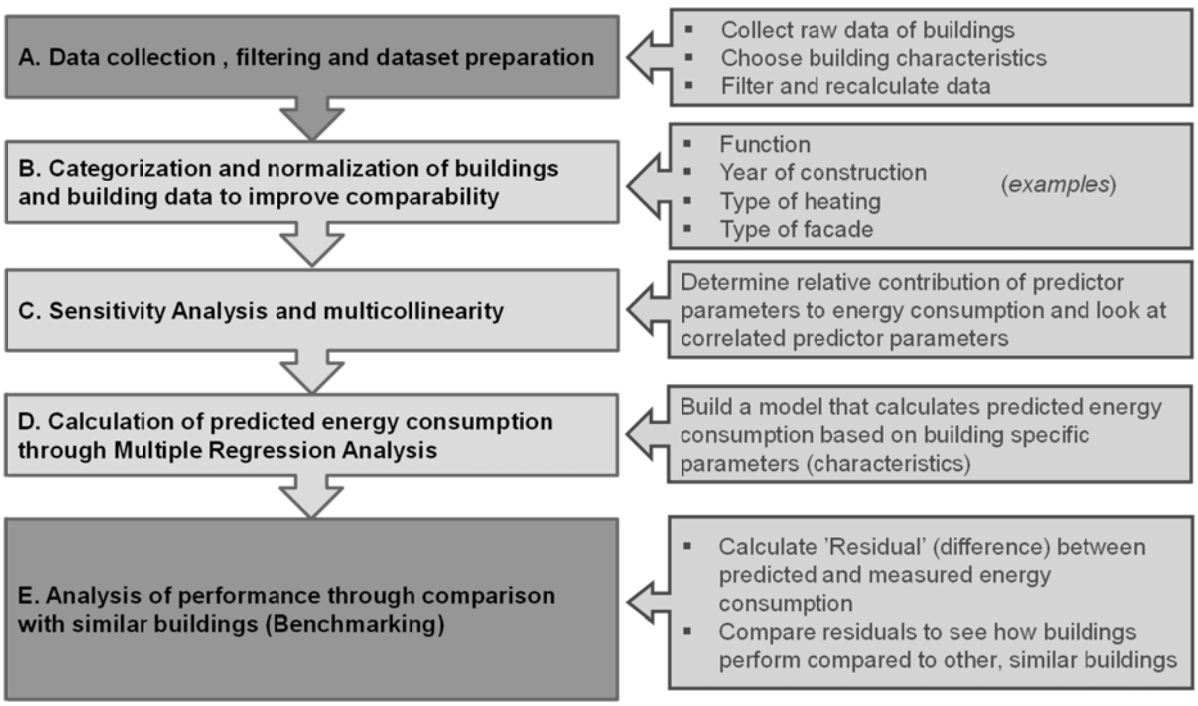

Figure 1: Simplified schematic representation of the research stages.

2.1 Data collection, filtering and dataset preparation (Stage A)

As mentioned in the introduction, the CBECS 2003 database was chosen in Step 1, Figure 2, to test the methodology. Step 2 of the research consists of selecting an (arbitrary) set of parameters, based on literature review and availability. The selected parameters can be divided in three types, as shown in Table 1 and Step 4 in Figure 2. Nominal parameters contain values divided in categories, however, these categories cannot be ranked. For example, buildings can be divided by *location*, but one cannot say one location is better than another, they cannot be ranked in this manner. These parameters will be referred to as categorization parameters. Interval parameters can be ranked and will be used as predictor parameters for the remainder of the research. In this research, *building age* is used both as categorization (10 age categories defined) and predictor parameter. The third type of parameters consists of response parameters or energy consumption parameters.

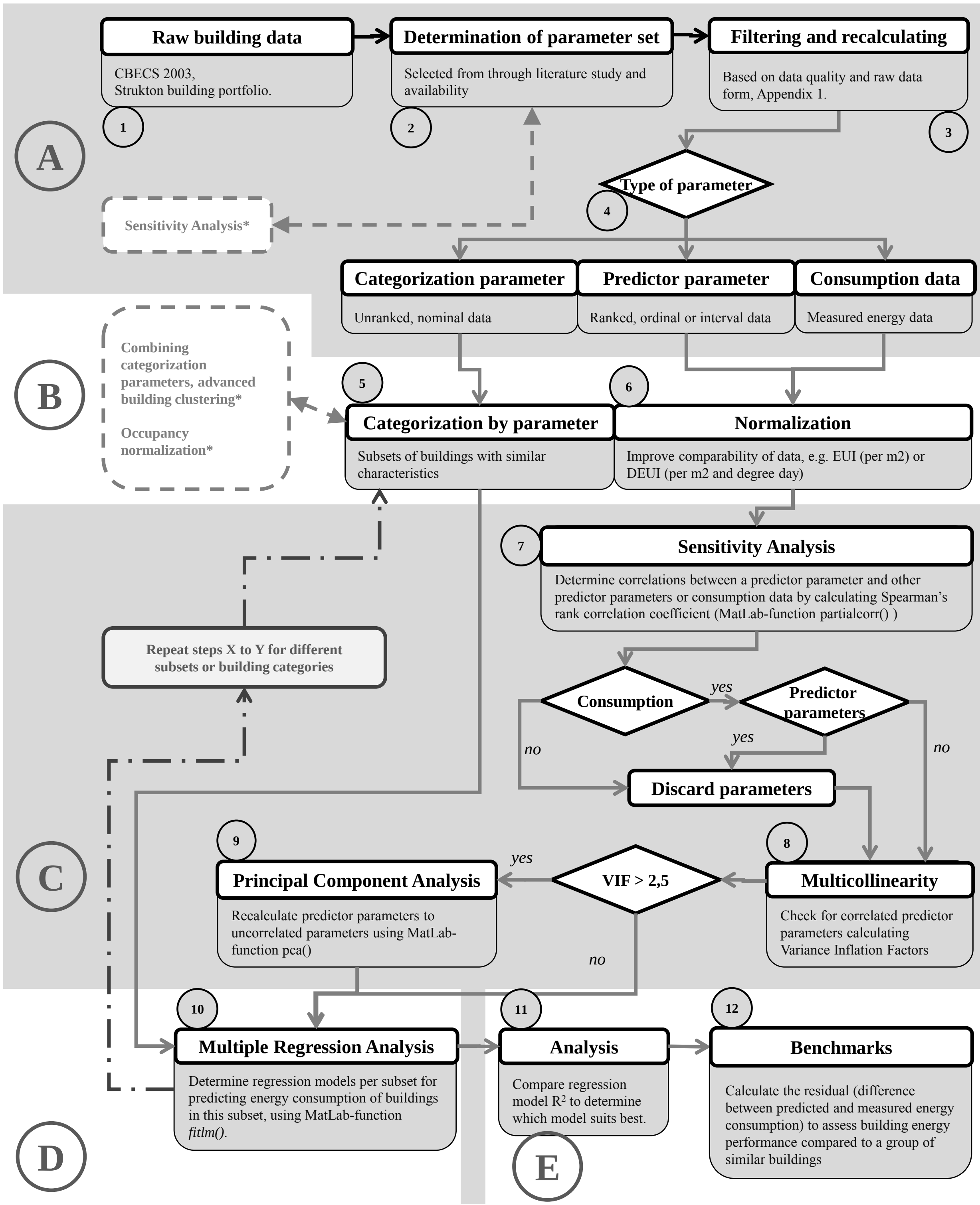

Figure 2: Detailed schematic representation of the research steps and the MatLab functions used for each step. A, B C, D and E correspond to the stages shown in Figure 1.

As mentioned before, the original CBECS 2003 dataset contained over 5000 observations, single buildings in this case. Not all observations were complete or considered usable for the remainder of the current research. Therefore, in Step 3 of the methodology shown in Figure 2, data was prepared for analysis by rewriting and recalculating some of the values, and by filtering for incomplete and unreliable data. For example, observations without consumption data of at least 12 months was left out of the dataset and also shopping malls were excluded because a lack of useful data in most cases. More detailed explanation of the filtering of data, assumptions made and recalculations performed can be found in the appendices.

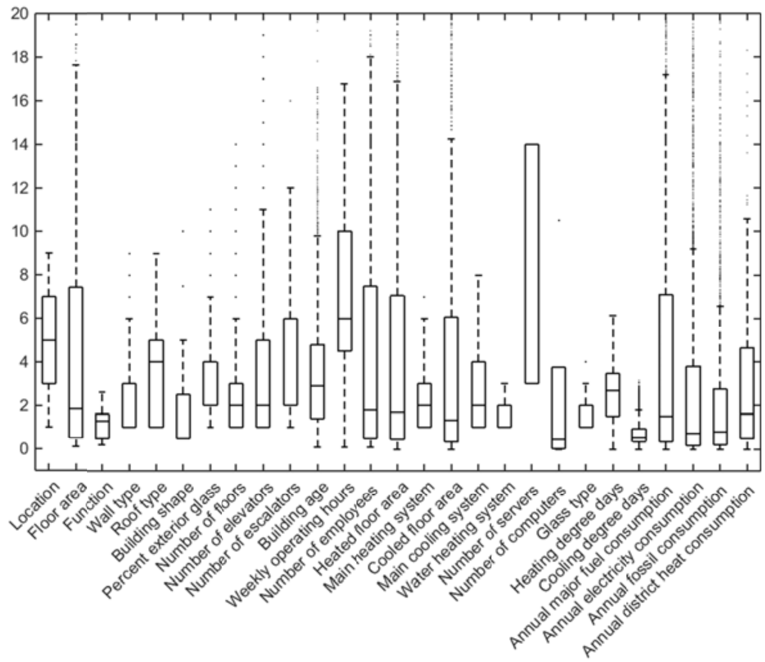

Figure 3: Output of Stage A, distribution of all values of the 27 addressed parameters for 3692 observations. Parameter values were standardized for visualization purposes, units can be found in Table 1.

Table 1: Output of Stage A, the addressed parameters classified to three types: categorization, predictor, and response parameters. The first type is used for categorization in Stage B, the others for the regression analysis in Step C.

Categorization parameters	**Predictor parameters**	**Response parameters**
Nominal parameters	*Discrete/Continuous parameters*	*Performance Indicator*
Location	Floor area [m^2]	Annual major fuel consumption
Function	Percent exterior glass [%]	Annual electricity consumption
Wall type	Number of floors [*n*]	Annual fossil consumption
Roof type	Number of elevators [*n*]	Annual district heat consumption
Building shape	Number of escalators [*n*]	
Main heating system	Weekly operating hours [*hours*]	
Main cooling system	Number of employees [*n*]	
Water heating system	Heated floor area [m^2]	
Glass type	Cooled floor area [m^2]	
Building age	Number of servers [*n*]	
	Number of computers [*n*]	
	Heating degree days [*DD*]	
	Cooling degree days [*DD*]	
	Building age [*years*]	

2.2 Categorization and normalization of buildings and building data to improve comparability (Stage B)

To accurately assess the energy performance of buildings it is necessary to ensure the comparability of consumption data. Comparing buildings of different sizes or in climate zones can results in incorrect conclusions. In order to improve the comparability of consumption data, normalization can be carried out, Step 6 in Figure 2. The most frequently used normalized energy metric is the Energy Use Index (EUI) which divides total fuel consumption by the *floor area* of the building. Because energy consumption is highly dependent to daily weather conditions and year round climate, the Daily Energy Use Index (DEUI) is defined as energy consumption normalized by *floor area* and the number of *degree days*. In the current research the number of degree days was defined as the sum of *heating degree days* and *cooling degree days*.

In case of normalization, the values for these input parameters are divided by own values, resulting in a value of 1 which is used in further analysis. This because the effect of these parameters is included in consumption values after normalization. Also the values for *heated floor area* and *cooled floor area* are divided by the *total floor area*. Figure 4 shows both total and degree day normalized gas consumption for the Strukton case building in Son. In both cases, significantly decreasing consumption is shown and for total energy consumption these decreases are quite extreme. However, taking into account the climate conditions and heating demand, the differences are reduced. Nevertheless, the energy performance of the buildings seems to have improved due to measures taken within the building.

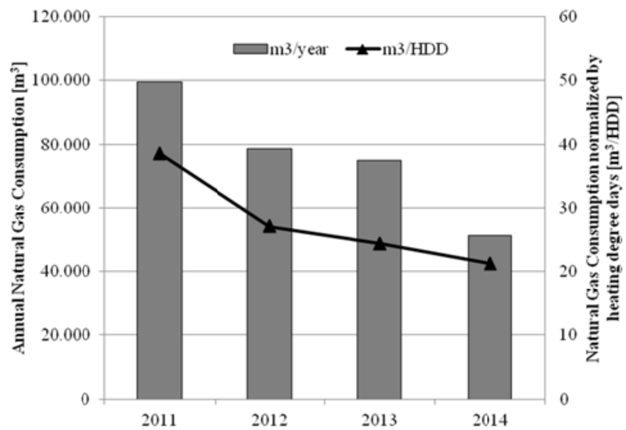

Figure 4: Annual gas consumption for the Strukton office building in Son in total and normalized by degree days. Because only one building was looked at, and its floor area did not change in time, normalization by floor area was not carried out.

Step 5, as shown Figure 2, is categorization of observations, or buildings in this case. The aim in this step is to create subsets of buildings with similar characteristics. Figure 5 shows an example of dividing a dataset or building portfolio into subsets based on categorization parameters. Buildings can be categorized by *function* or *building age*, for example. Every building can be part of only one subset for the same categorization parameter, but when looking at a number of categorization parameters, the building is part of multiple subsets. For example, it is assumed that a building can have only one (main) *function*, and belongs to only one *building age* category. However, a building can be assigned to the subset of *office* buildings and the subset of buildings built after 1999 at the same time. For the current research, single categorization parameters were used for the division of a dataset into subsets. It is possible to combine categorization in order to enhance building similarity within one subset, however, this has not been part of the current research.

As mentioned before, nominal parameters are used for categorization of buildings. Also, categories of *building age* were used introduced to be able to use it for categorization of buildings. In this case percentiles were calculated to determine the boundaries of subsets. Figure 6 shows the distribution of measured primary energy consumption for the subsets of buildings categorized by building age. Note that, for readability purposes, building age is recalculated to year of construction. Categorization parameters were not included in further analysis as predictor parameters for aforementioned reasons, building age was included in its original form.

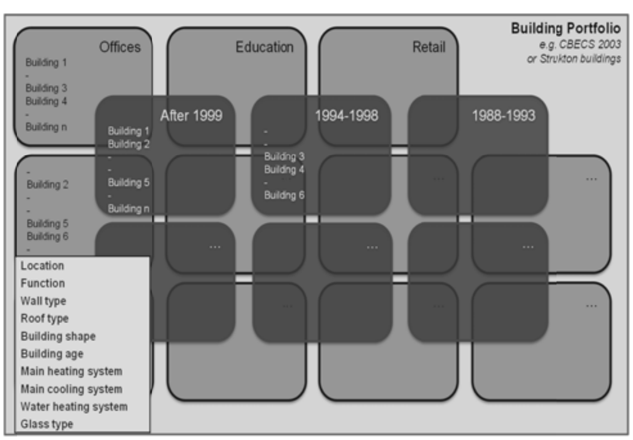

Figure 5: Schematic representation of categorization buildings as carried out in Stage B, Figure 1, in the total dataset by categorization parameter, and a list of categorization parameters used in the current research.

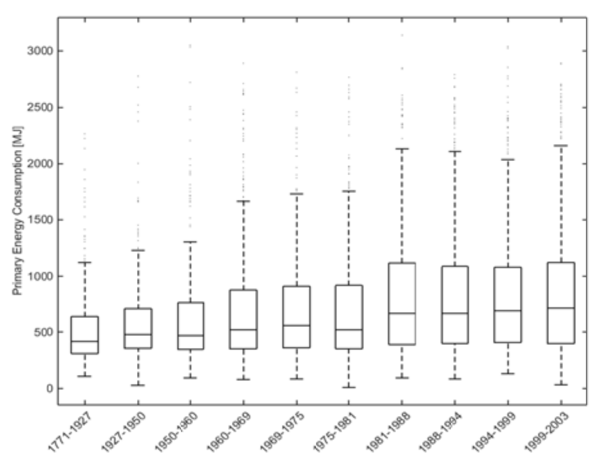

Figure 6: Total dataset categorized by year of construction category, showing the distribution of total primary energy after Step 5, Figure 2.

2.3 Sensitivity analysis and multicollinearity (Stage C)

After categorization and optional normalization of the buildings and associated predictor and response (energy consumption) parameters, Step 7 of the methodology as shown in Figure 2, consists of a sensitivity analysis. The aim of this step is to determine mutual correlation between predictor parameters and the relative contribution of the predictor parameters to energy consumption. Both correlations can be determined by calculating Spearman's partial rank correlation coefficients using built-in MatLab function *partialcorr()* [5], as was suggested by Tian et al. [6]. This function provides partial rank correlation coefficients and the p-value for the parameters. The coefficients represent the degree of correlation between two parameters, a coefficient close to 1,00 indicates a strong correlation, a coefficient close to 0,00 indicates a weak correlation. If two predictor parameters show a strong mutual correlation, the influence of one parameter on energy performance can be assumed to be explained by the other parameters to a great extent. Therefore, one of the parameters can be discarded for computational purposes. If one predictor parameter and energy consumption show a weak correlation, it can be concluded that the predictor parameter has insignificant influence on energy performance. Therefore, it is assumed that this parameter can also be discarded.

The reliability of the results of this step are dependent on the dataset worked with. The size of the dataset, distribution of parameter values and the reliability of data can influence results. Nominal data, like the categorization parameters, cannot be included in the sensitivity analysis. In order to test the statistical significance of the coefficients found, the p-value is assessed. Per assumption, a p-value smaller than 0,05 can be considered statistically significant [7], and therefore the correlation coefficient can be considered a reliable measure for strength of the correlation.

As mentioned, predictor parameters might be to some extent correlated. Moreover, in a multidimensional dataset like the one used for this research, this is likely to occur. For example, the number of degree days can be expected to have some correlation with location and the amount of heated floor area is likely to be related to total floor area. Correlated input parameters of a multidimensional dataset can cause decreased accuracy of the results of the multiple regression analysis in step 10 of the research [6, 8]. This phenomenon is referred to as the multicollinearity problem [9]. To check whether this problem is an issue to address in the current dataset, variance inflation factors (VIFs) are calculated for each input variable (Step 8, Figure 2)[7]. The equation for calculating VIFs can be found in Appendix 2. A rule of thumb introduced by Allison et al. [9] and applied by Wang et al.[8] among others, states that when the VIF of a variable exceed 2.5, this variable is significantly correlated to one or more of the other input parameters. For the current research, it is assumed that in case one or more parameters show these higher VIFs, the multicollinearity problem should be taken into account.

Two approaches are considered to address the multicollinearity problem. First, parameters could be discarded for reasons explained in the subsection on sensitivity analysis. When the multicollinearity problem remains an issue, the remaining parameters can be recalculated to uncorrelated parameters (principal components) using Principal Component Analysis (PCA) [8, 10], Step 10 in Figure 2. For this analysis, built-in MatLab function *pca();* [11] was used to create these uncorrelated parameters. More detailed information on the multicollinearity problem and PCA process can be found in the Appendix 0. Figure 7 shows the principle of PCA where a set of predictor parameters (p) are recalculated to a set of principal components (pc) which have insignificant mutual correlations. If multicollinearity is not an issue based on VIFs, this step can be skipped and predictor parameters can directly be combined into one virtual predictor parameter X, as input for the Multiple Regression Analysis (MRA). Figure 8 shows the structure of a dataset both with original predictor parameters and with these predictor parameters replaced by principal components.

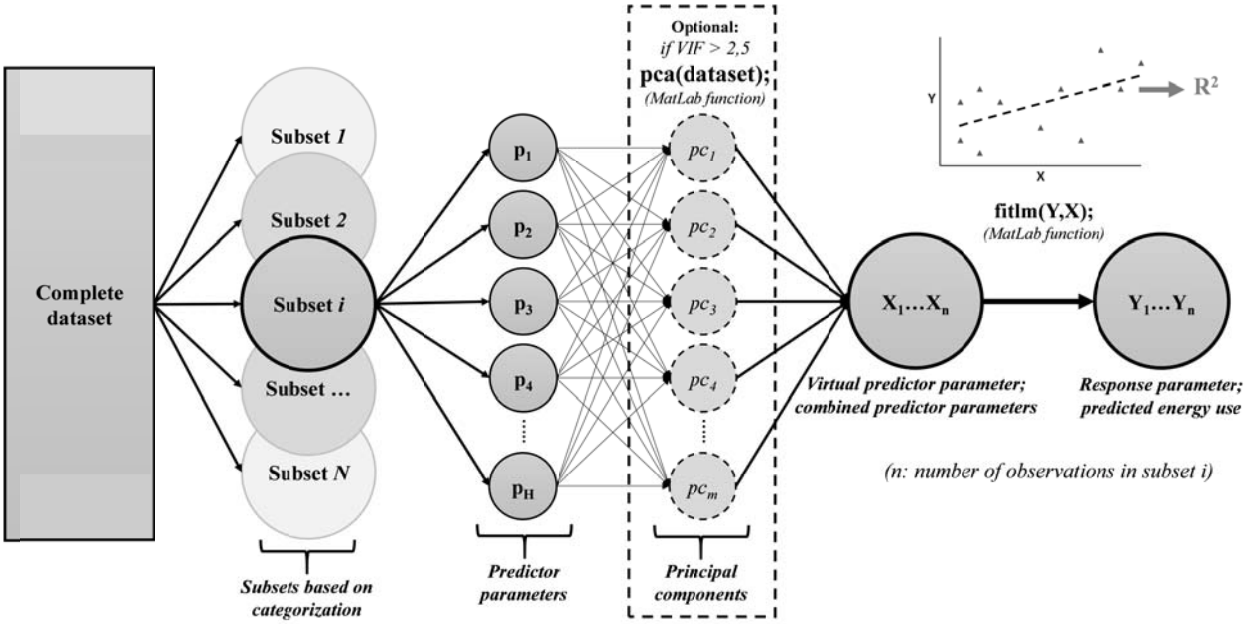

Figure 7: Schematic representation of the (optional) principal component analysis (PCA, Stage C) and Multiple Regression Analysis (MRA, Stage D). PCA recalculates the original, significantly correlated set of predictor parameters to a set of principal components with no significant correlations. This step is only carried out if the multicollinearity check proves correlations between building parameters are significant.The MRA calculates the best fit linear model for calculating energy consumption from a given set of predictor parameters for one subset. R^2 is a statistical measure to indicate to which extent the model fits the subset data. 100% means a perfict fit and every observation perfectly matches the regression model, 0% means no linear relation was found.

	Predictor 1 x_1	**Predictor 2** x_2	**…**	**Predictor h** x_h	**…**	**Predictor 10** x_H	**Combined predictor** X	**Response** Y
Observation 1 *(Building 1)*	p_1 pc_1	p_2 pc_2	… …	p_h pc_h	… …	p_h pc_h	X_1	Y_1
Observation 2 *(Building 2)*							X_2	Y_2
Observation j *(Building j)*	**Building specific predictor values**						Xj	Yj
Observation n *(Building n)*						$X_j = \sum_{h=0}^{H} (p_{j;h} * x_h)$ ***or*** $X_j = \sum_{h=0}^{H} (pc_{j;h} * x_h)$	Xn	Yn
	(j : addressed observation at subset i) *(H: number of predictor parameters)*						**Y=a*X+b;** *(MatLab: fitlm(Y,X))*	

Figure 8: Structure of the datasets used, both original dataset and subsets according to categorization. Predictor parameters (p), or calculated principal components (pc) are combined into one virtual predictor X. This virtual predictor is assumed to have a linear relation with response variable Y, in this case being energy consumption of the building. MatLab function fitlm(); determines X and finds the best fit model for the relation between X and Y based on the observations of one subset.

2.4 Multiple regression analysis (MRA) to calculate predicted energy consumption (Stage D)

Step 8, as shown in Figure 2, consist of multiple regression analysis (MRA) for determining what is 'normal' or predicted energy consumption for a given set of buildings (observations), each with its own set of predictor parameter values. After categorization of the buildings in order to cluster similar buildings and optional PCA to ensure predictor parameters are not significantly correlated, the subsets are considered to be suitable for multiple regression analysis (MRA). MatLab function *fitlm(X,Y);*[12] was used to create a regression model based on the given subset of observations and resulting a linear equation to calculate predicted energy consumption using a given set of predictor parameter values for one observation. In this case X is the set of predictor parameters, and Y is the response parameter, being consumption data. The MRA model gives a set of estimates (e) for coefficients for every predictor parameter (p) or principal component (pc) which can be used to calculate predicted energy consumption suing the following equation [12] :

$$Y_{pred} = e_{int} + \sum_{i=0}^{N} p_i * e_i \qquad [e_{int}: estimated\ intercept\ provided\ by\ fitlm()]$$

Along with the model for calculating predicted energy consumption, the MRA offers a number of statistics to assess the significance of the model for the given dataset. The statistics for the MRA help determine the significance and reliability of the model and its results. R^2 is the coefficient of determination of the model, which indicates to what extent the model fits the dataset used. This statistic is used to determine how well the prediction fits the data and the suitability of the model for benchmarking purposes. R^2 is shown as a percentage where high values indicate a good fit, and therefore the model is interesting for the purposes of the current research. The coefficient of determination is to some extent dependent on sensitivity to outliers and the dispersion of data. Therefore the p-value of the MRA model is evaluated. The p-value is a measure for statistical significance of the results, and therefore reliability of these results. A p-value can vary from zero to one, but again a value of more than 0,05 is assumed to indicate unreliable results [7]. Therefore, the closer the p-value is to zero, the more reliable the coefficient of determination. In the current research, Root Mean Squared Error (RMSE) is also considered a valuable statistic because it indicates the amount of dispersion of results within one subset. A high value for this RMSE might indicate a need to subdivide that particular dataset in order to improve results. All the MRA model statistics mentioned above are provided by the MatLab function *fitlm()* [12].

Figure 9 shows an example of two regression models for predicted energy consumption within a (fictional) dataset. The y-axis represents the energy consumption, x-axis a virtual variable constructed from the combination of input (predictor) parameters. The dots and circles represent building observations from the datasets used for regression analysis. The lines represents the regression models or predicted energy consumption for these dataset as a function of the combination of predictor parameter values. The relation between these predictor parameters and energy consumption are given by the equation above. Series 2 show a higher R^2 because observations from which the model is calculated are less dispersed. The regression model is more likely to predict consumption accurately.

Steps 5 to 10 of research, shown in Figure 2, are repeated for different subsets after categorization or normalization. Results of the MRA are compared in Step 11 to determine which regression model shows best fit for the subsets and therefore is most likely to produce reliable results in the upcoming step of benchmarking energy performance.

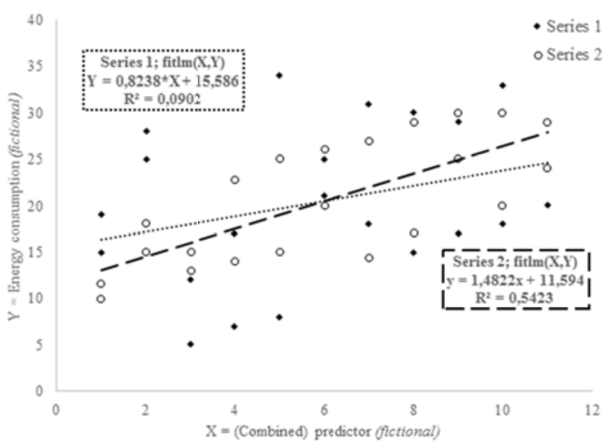

Figure 9: Examples of output of Step 10, Stage C in Figure 2. Regression models are shown with relatively high fit for Series 2 and low fit for Series 1. Energy consumption on the y-axis is a function of a virtual variable representing the combination of input factors on the x-axis. Data used in this explanatory figure is fictional.

2.5 Analysis of performance through comparison with similar buildings (Benchmarking) (Stage E)

The next step, Step 12 of the research, uses the models determined in the MRA for calculation of predicted energy consumption and compare this with measured energy consumption to assess energy performance if buildings. Residuals are calculated as the difference between measured (observed) energy consumption and calculated (predicted) energy consumption, a method also used by Wang et al. [8] among others. The residuals are calculated for all buildings in a given subset or building category, resulting in a distribution of residuals for the subset as shown in Figure 10. Given the definition of residuals, a negative residual means the building is performing better than predicted by the regression model. Therefore, a large negative residual is positive in terms of energy performance assessment.

From the calculated distribution of residuals, percentiles can be calculated to be used as boundaries for benchmark categories. Figure 10 shows arbitrary boundaries at 5%, 20%, 45%, 55%, 80% and 95%. This means buildings with a residual lower than the value belonging to the 5%-line perform better than at least 95% of the buildings in the subset and therefore can be awarded an A-label. Another possibility is to assess performance by showing the percentage of buildings in the subset showing higher residuals. As mentioned, the boundaries for labeling energy performance are arbitrary and can be adapted to the ambitions of the user of the methodology.

When assessing a building, different benchmarks can be compared for subsets of different types of categorization. For example, for one building benchmarking should be carried out for the subset of buildings with the same function as well as buildings of the same age category or heating system. A thorough analysis of multiple benchmarks could indicate where to look for energy saving potential.

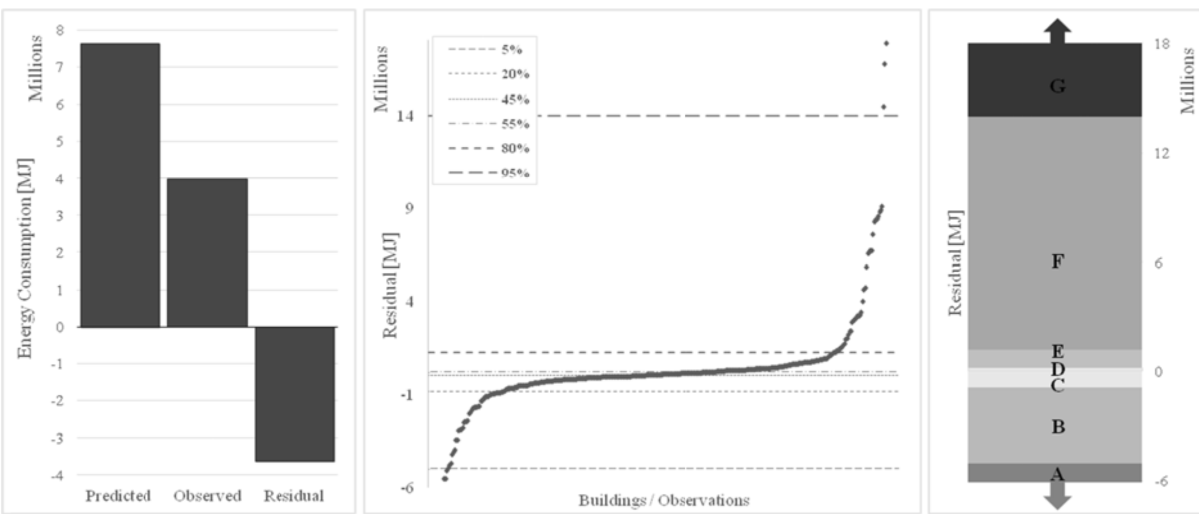

Figure 10: (a) Residual defined as the difference between measured energy consumption and consumption predicted by the regression model, (b) distribution of residuals for one subset, being buildings built after 1999 and (c) benchmark categories based on residual distribution for the subset of buildings built after 1999.

3 Results

The current research resulted in a methodology that can be used for analysis of a dataset of buildings and define custom energy performance benchmarks for this dataset. These benchmarks can help assess energy performance and indicate energy saving potential by comparing to similar buildings in subset based on a number of categorization parameters. The first part of the methodology proposed compares Multiple Regression Analysis (MRA) results for different input datasets constructed categorizing or normalizing building data, as explained in the methods section. The analysis of these MRA results is performed in order to select the most promising data subset for assessment of energy performance. After this selection, the MRA results can be used to calculate predicted energy consumption for a given set of buildings. The difference between predicted and measured energy performance, referred to as the residual, can be used to assess energy performance compared to similar buildings. The current section focusses on the results of the four crucial steps to this methodology and concludes with a case study to demonstrate the practical relevance of this methodology. Results of the benchmarking process will not be addressed separately but will be discussed in Section 3.3 on the case study.

3.1 Categorization and normalization findings

As mention in Section 2.2**,** two types of energy consumption normalization are considered in this research and compared to the results for not-normalized data. In particular, these types are Energy Use Index (EUI) and Daily Energy Use Index (DEUI). In general, the coefficients of determination (R^2) are significantly higher for energy consumption and electricity consumption when performing the MRA using the original, not-normalized dataset. When analyzing fossil fuel or district heat consumption specifically, it is advisable to use normalized data because R^2 tends to be higher for these cases. This can be explained by the fact that fossil fuels and district heat are mainly used for heating, and the size of the building and the number of degree days are known to be of significant influence on heating demand. In general, DEUI scores slightly better than EUI.

Basically, the choice of normalization type depends on which type of energy consumption is targeted for analysis. For the remainder of the current report, only primary energy consumption was taken into account, so further results were only described for total (not-normalized) energy consumption. However, detailed MRA and case study results in Appendix 4 and 5 do include the results for normalized data. In subsection 3.3 on the case study results, normalization will be addressed briefly as well.

Table 2: Coefficients of determination of the regression models for the total dataset, looking at different consumption parameters and performance indicators. P-values of all presented coefficients are significantly smaller than 0,05.

R^2	Primary	Electricity	Fossil	Heat
Total	88,94%	96,48%	51,27%	19,02%
EUI	41,14%	42,99%	35,04%	45,39%
DEUI	43,40%	43,01%	33,22%	50,41%

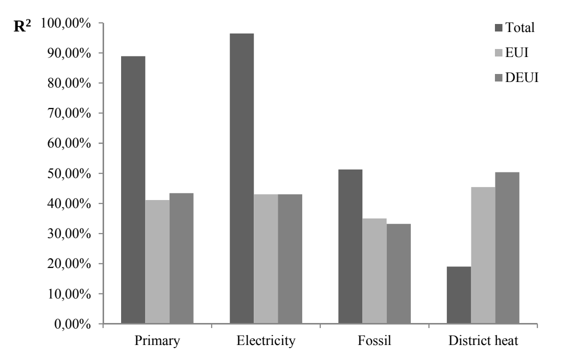

Figure 11: Graphical representation of the data presented in Table 2.

Table 2 shows that the results for normalized data were found to be significantly lower. Although normalization is considered advantageous in some cases, as discussed before, it is not recommended for primary energy consumption or electricity to use normalized data. However, using the MRA as is the case in this research, degree days and floor area are also included in calculations for predicted energy consumption, therefore making standard normalization unnecessary.

As explained in section 2.2, 10 parameters were selected for categorization of the buildings in the dataset. Every categorization parameter consists of a range of three to sixteen subsets or building categories, which can be found in the appendix. The original dataset is divided in subsets with equal values for the categorization parameters. For example, building category *office* under categorization parameter *function* is a subset of the original dataset, consisting of all buildings of the function type office. Building category *Brick, stone or stucco,* under categorization parameter *wall type* is a subset of the original dataset, consisting of all buildings with a predominant facade type of brick, stone or stucco. One building can be assigned to the building category for *offices* as well as *Brick, stone or stucco*, but can only be assigned to one subset with respect to the same categorization parameter. The original dataset is the complete dataset without categorization carried out.

Table 3: Weighted mean regression analysis results for the original dataset and subsets based on categorization parameters.

Categorization parameters	Primary			EUI			DEUI		
	R^2	p	rank	R^2	p	rank	R^2	p	rank
Original	**88,94%**	0,0000	-	**41,58%**	0,0000	-	**44,15%**	0,0000	-
Location	**88,99%**	0,0000	*1*	**25,41%**	0,0257	*11*	**23,67%**	0,0220	*11*
Function	**87,79%**	0,0000	*3*	**32,79%**	0,0000	*7*	**36,15%**	0,0000	*4*
Wall type	**85,66%**	0,0000	*4*	**34,40%**	0,0029	*5*	**41,38%**	0,0017	*1*
Roof type	**85,18%**	0,0000	*5*	**36,81%**	0,0004	*3*	**31,33%**	0,0000	*7*
Building shape	**82,19%**	0,0000	*6*	**31,40%**	0,0000	*8*	**32,68%**	0,0000	*5*
Building age	**81,48%**	0,0000	*7*	**33,06%**	0,0000	*6*	**31,30%**	0,0000	*8*
Main heating system	**80,57%**	0,0000	*8*	**37,86%**	0,0000	*2*	**38,03%**	0,0001	*2*
Main cooling system	**76,74%**	0,0000	*9*	**29,47%**	0,0138	*9*	**30,11%**	0,0008	*10*
Water heating system	**72,46%**	0,0000	*10*	**38,34%**	0,0278	*1*	**31,69%**	0,0025	*6*
Glass type	**37,04%**	0,0000	*11*	**34,99%**	0,0000	*4*	**37,25%**	0,0000	*3*

For every subset of buildings multiple regression analysis was carried out to find a model for calculating predicted energy consumption. The statistics calculated with this model, calculated by MatLab function *fitlm();* [12], can be used to assess the accuracy of predictions of this model. Table 3 shows the weighted mean R^2 and p-value for every categorization parameter. This is the mean value weighted by the number of buildings in the subset divided by the total number of buildings of all subsets combined. R^2 is a measure for the fit of the MRA-prediction to the data in the building class, p-value indicates the statistical significance of the MRA-prediction. In general, it is assumed that if the p-value remains below 0.05 [7], the results are statistically significant. As Table 3 shows, this is the case for all values of R^2.

The most widespread method of categorizing buildings is probably by function or main activity within the building. Table 3 shows this can be statistically explained by the highest R^2 of all. However, the original dataset or categorization by main heating equipment show similar results. More parameters show scores of over 80% and could be considered good options for categorization. Moreover, it can be very interesting to look at multiple categorizations for one building. A building might perform well compared to buildings of a same function, but not so well compared to buildings with the same type of heating equipment. This might indicate the heating system is not performing as well as can be expected indicating possible energy saving potential.

Table 4: Regression statistics for subsets from categorization based on building age and the weighted mean for all subsets as presented in Table 3.

Built during or after	R^2	p	Number of buildings	RMSE
1999	82,94%	1,34E-133	385	7,39E+06
1994	87,23%	1,93E-144	357	9,96E+06
1988	53,02%	2,21E-44	370	2,48E+07
1981	53,39%	5,20E-53	381	3,35E+07
1975	80,18%	1,31E-122	388	1,82E+07
1969	94,71%	3,90E-211	359	1,72E+07
1960	90,59%	1,30E-202	427	1,12E+07
1950	93,37%	1,73E-181	337	1,05E+07
1927	94,56%	5,98E-143	323	7,88E+06
1771	87,64%	1,85E-150	365	5,74E+06
Weighted mean	**81,48%**	**2,21E-45**	*(3692)*	**1,48E+07**

Details on the results per building class can be found in the appendices. As an example, Table 4 shows the statistics of the regression analysis of data categorized by building age or year of construction. As the results show, the categories for buildings built between 1981 and 1987, and between 1988 and 1993 show a relatively high RMSE indicating observations show a wide distribution of numbers for energy consumption. Notice that these categories also show a lower coefficient of determination, indication that there is a significant chance that predicted energy consumption deviates from actual, measured, energy consumption. Therefore, it might be beneficiary to subdivide these building categories to reduce the RMSE and improve R^2.

High weighted mean coefficients of determination as shown in Table 3 can indicate the chosen building parameter is suitable for meaningful categorization of buildings for benchmarking. However, the weighted mean coefficient of determination can be significantly influenced by low coefficients for a small number of subsets. Therefore, this type of categorization should not be discarded be because the single subset can be very interesting for benchmarking particular buildings. For example, when benchmarking energy consumption of a building built between 1969 and 1974, Table 4 shows a R2 of 96%, which shows this MRA model is very interesting for this particular building although the overall R2 for categorization by age is only 82%. Note that the p-value of all subsets is significantly smaller than 0,05, so results can be assumed statistically reliable. In conclusion, in case of the assessment of a single building, it is advisable to analyse MRA statistics for specific subsets the building is part of.

In general, results of the MRA show that for every building different building parameters can be best suitable for categorization. Moreover, where one type of categorization, e.g. building function, might be more interesting for benchmarking one building than another type, e.g. heating system, this can be the other way around for another building. It is important to keep in mind that another original dataset might lead to different results and conclusions based on this same methodology. Thorough application of the methodology with respect to data collection and defining category boundaries where possible, as well as careful analysis of the results for both building benchmarks as well as choosing benchmarks for building energy performance assessment can enhance the ability to actually assess performance.

3.2 Sensitivity analysis using Spearman's partial rank correlation coefficient

After normalization and categorization of the buildings partial correlation coefficients are calculated to determine mutual correlation between two building parameters, adjusted for the other parameters. Figure 12 shows the partial correlation coefficients between all 14 predictor variables and total primary energy consumption for one subset of the building age categorization parameter, namely buildings built between 1999 and 2004. For this particular subset of buildings, numbers of escalators, elevators computers show to have a relatively small correlation to primary energy consumption. This could indicate it is less crucial to include these parameters in the following steps, and these could possibly be discarded. However, the relative importance of parameters can be different for different subsets or different categorization parameters.

As determined through the calculation of the VIFs in the example in Appendix 0, four predictor parameters showed strong correlations to one or more other predictor variables. To take a closer look at this result, partial correlation coefficients for these four parameters were calculated. Figure 13 shows the results of these calculations carried out using MatLab-function *partialcorr()* [5]. Heating and cooling degree days show strong (negative) mutual correlations, meaning that if parameters shows high values the other one most of the time shows low values. Total, heated and floor area also show strong (positive) correlations. It means large values for one often means large values for the others. In case of these high correlations it could be concluded that some of the parameters can be discarded because their effects on energy performance are to a great extent explained by the other parameters. The conclusions of both previous paragraphs in this section are based on the assumption that the results of the sensitivity analysis are statistically significant. In order to test this significance, the MatLab function used, also provides the p-values that are associated with every single correlation coefficient. For this particular subset, and for all subsets considered in the current research, the p-values tended to be larger than the allowed 0,05[7]. Therefore, at this point, the results of the sensitivity analysis were found to be of little significance and reliability. This resulted in the choice to not discard any of the chosen predictor parameters at this point. Future research on the subject of sensitivity analysis and relative significance of predictor parameters to energy performance of buildings are expected to improve the results of the current research. More on this topic is discussed in Section 4.4.

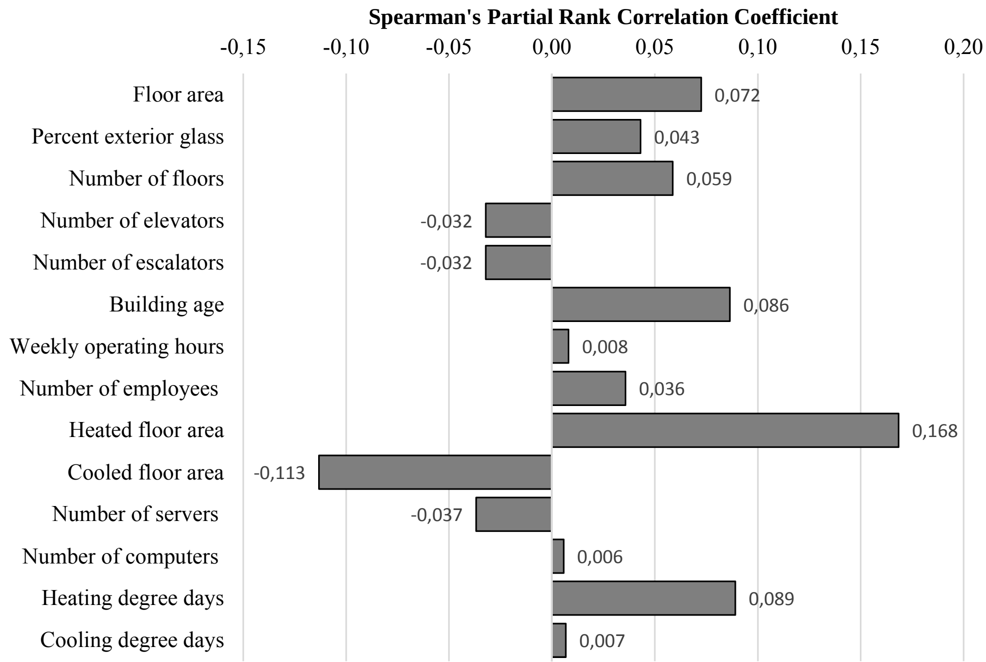

Figure 12: Spearman's partial correlation coefficients indicating the strength of the partial correlation between predictor variables and total primary energy consumption.

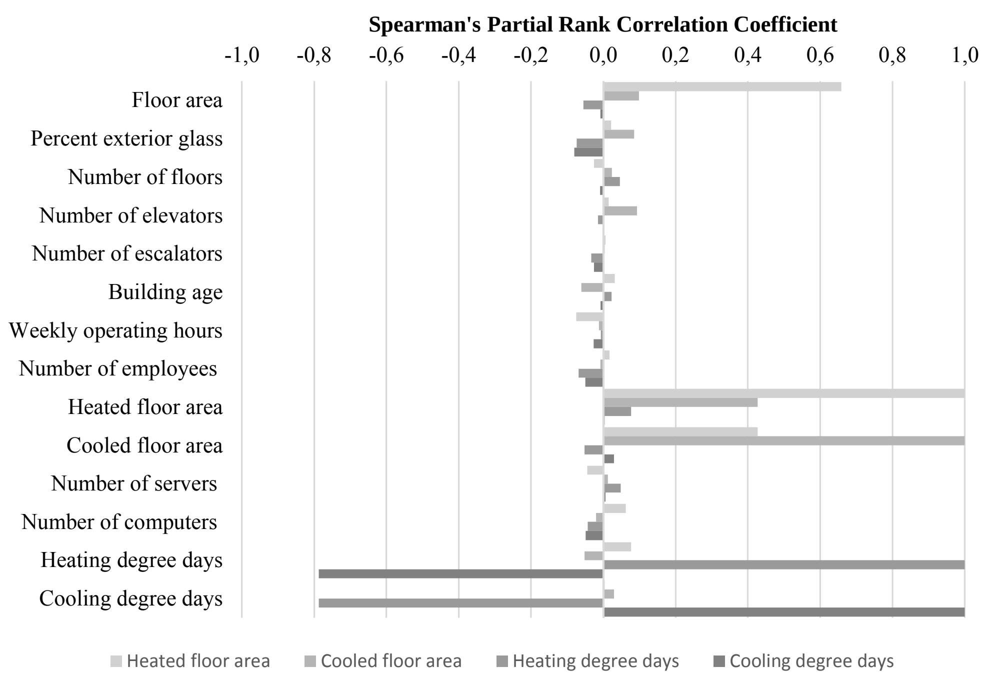

Figure 13: Spearman's rank correlation coefficient for the parial correlation found between four selected parameters and the other predictor parameters for the subset of buildings built after 1999. Heated and cooled floor area, as well as heating and cooling degree days were chosen because these parameters showed high VIF's, as show Figure 17 and Figure 16 in Appendix 0, which also discusses multicollienearity and the calculation of VIFs.

3.3 Case: Strukton office building in Son (NL)

Strukton Worksphere is a company that specializes in maintenance and operations of buildings and installations. Recently, the company also started activities in the field of ESCO's (Energy Service Companies). Also, an online platform for remote operations of buildings is under development. All this makes Strukton a partner than can use the findings of the current research to improve activities. For such companies it is important to be able to reliably determine which buildings show potential for energy savings. Because the number of buildings in their portfolio and the similarity of buildings in this portfolio, specific benchmarks can help achieve this goal.

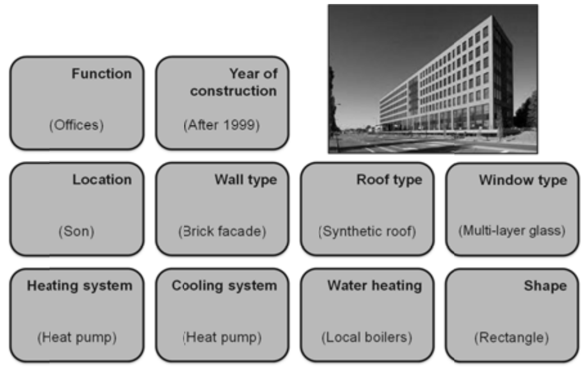

Figure 14: Case building characteristics, determining the assignment assigning it to exactly one subset (or building category) for every categorization parameter.

As mentioned, the data collection within the company is not yet on a level making it possible to determine benchmarks from Strukton data. However, one of their main office buildings is used to test the methodology using benchmarks created from the CBECS 2003 dataset. Internal monitoring of the building's energy performance showed improvements between 2011 and 2014, so data available from both years was used to check if this improvement would also show from the results yielded from the methodology. Appendix 0 shows detailed values for the parameters used as input. Note that only degree days and measured energy consumption is different between the two years. This confirms the suspicion that the current set of parameters is not optimal, as will be further discussed in the next section. Figure 14 shows a graphical representation of the building, the values assigned to the characterization parameters, and the subsets the building is part of based on these values. These subsets were used for calculating predicted energy performance using the subset specific regression model.

Table 5 shows the results for the benchmarking process of the case study for 2011 and 2012. Residuals calculated from the measured primary energy consumption and the earlier mentioned predicted energy consumption assign the building to a performance label for each subset it is part of. R^2 shows the reliability accuracy of the results and the p-value the statistical significance. After the first line, where the complete CBECS 2003 dataset was used for benchmarking, the categorization parameters are ordered by R^2.

Overall, Table 5 shows good results for energy performance of the buildings. Most categories show a B-label or higher. This is as expected, because this building was put into operation in 2010 and was designed with ambitious requirements towards energy performance. The CBECS data is from 2003 one should expect overall improved energy performance over the 7 years after the survey was carried out. Nevertheless, for the subset of 'newest' buildings the case building was assigned to, being the subset of buildings built after 1999, the case building only scores a B-label. Possible explanations for this unexpected low score are a relatively high electricity consumption due to the function of the building, the data centers present in the building and the

Table 5: Benchmarking results for the Strukton office building, located in Son, based on data from 2011 and 2014.

	Strukton Son 2011				**Strukton Son 2014**			
	R^2	**p-value**	**Residual**	**Label**	**R^2**	**p-value**	**Residual**	**Label**
Original	**88,94%**	*0,00E+00*	**-1,79E+07**	**A**	**88,94%**	0,00E+00	**-1,96E+07**	**A**
Main heating system	**96,06%**	*3,49E-172*	**-5,62E+06**	**A**	**96,06%**	3,49E-172	**-7,26E+06**	**A**
Main cooling system	**91,52%**	*1,40E-144*	**5,81E+06**	**F**	**91,52%**	1,40E-144	**4,21E+06**	**F**
Function	**89,45%**	*0,00E+00*	**-2,79E+07**	**A**	**89,45%**	0,00E+00	**-2,96E+07**	**A**
Wall type	**87,43%**	*0,00E+00*	**-1,76E+07**	**A**	**87,43%**	0,00E+00	**-1,93E+07**	**A**
Water heating system	**86,94%**	*9,56E-265*	**-8,83E+06**	**A**	**86,94%**	9,56E-265	**-1,05E+07**	**A**
Glass type	**86,91%**	*0,00E+00*	**-2,05E+07**	**A**	**86,91%**	0,00E+00	**-2,23E+07**	**A**
Building age	**82,94%**	*1,34E-133*	**-1,99E+06**	**B**	**82,94%**	1,34E-133	**-3,64E+06**	**B**
Building shape	**79,66%**	*0,00E+00*	**-1,52E+07**	**A**	**79,66%**	0,00E+00	**-1,68E+07**	**A**
Roof type	**66,06%**	*1,04E-167*	**-3,59E+07**	**A**	**66,06%**	1,04E-167	**-3,76E+07**	**A**
Location	**41,79%**	*2,46E-44*	**-1,12E+07**	**A**	**41,79%**	2,46E-44	**-1,29E+07**	**A**

amount of other technology and installations and the comfort level of this building. Take into account that these results are based on total primary energy consumption. Detailed case study results, including the results for normalized data, can be found Appendix 5. These show lower benchmarking scores for normalized data. However, R^2 for these prediction models are significantly lower than in the results shown in Table 5. These findings strengthen the presumption that working with an outdated dataset as the CBECS 2003 dataset implies shortcomings to the results. An updated dataset would imply more comparable buildings in the building age categorization.

Another noticeable result is the F-label for buildings with heat pumps for cooling. This label is really an outlier compared to the others subsets. This might be explained by the fact that the buildings uses a relatively new gas heat pump for cooling which might not be really comparable to heat pumps used for cooling that are included in the CBECS 2003 dataset. Another possible explanation is the fact that for buildings with this type of cooling, the case building showed really low numbers of cooling degree days compared to the CBECS buildings. Therefore, the MRA results for predicted energy consumption also were very low and thus the residual was high. This type of cooling might be not often used in similar climates in the CBECS dataset and therefore the benchmark might be less useful in this case. Combination of categorization parameters could be tested in order to try to solve this issue. However, two valuable conclusions yield from this result. First of all the redefining of parameter values should be revised and can lead to significant improvements, further discussed in Section 4.3. The other conclusion is that the benchmarking per categorization seems to be able to locate energy saving potential to some extent. In this case, it is most likely that further assessment of the cooling system might lead to new insight. In case of a system working suboptimal, this should be possible to detect in a robust version of this methodology. Further research on de choice of parameters and the quality of data can enhance the robustness of the methodology.

Comparing the results for the same building, the Strukton office building in Son, for two different years, the expected improved performance shows in the results. Figure 15 shows the residuals for all subsets the case building is part of, together with the relevant boundaries for the benchmark labels. On top the R^2-values for the models were used and the subsets were ordered by this value again. Although Table 5 shows no differences in labels between the two years. These differences do show in the figure, for every type of categorization. In Section 2.2, the effect of degree days on energy performance and the benefit of normalization in this case. Because of the fact that all categorization parameters for the two years of the same case building are identical, the models used for calculating predicted energy consumption are identical as well. Nevertheless a detailed look at the residual calculations, Appendix 5, show a slight difference in predicted energy performance between the two years. This can be attributed to the fact that degree days are used as predictor variables for the regression model and calculation of predicted energy consumption. Hence the influence of weather conditions is taken into account and the differences in residuals between 2011 and 2014 can be assumed to be caused, to a great extent, by improved performance.

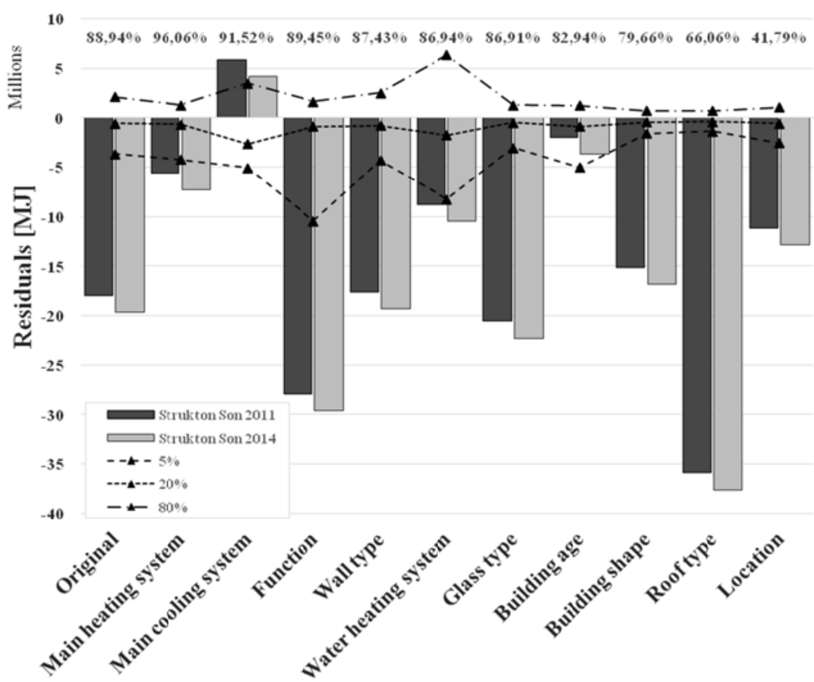

Figure 15: Case results for the Strukton office building showing the calculated residuals for the building and the percentile lines used for defining performance labels. On top, R^2 of the models used are shown indicating significance of the results.

4 Conclusions and discussion

This section rounds up the conclusions that yielded from the current research, discusses strengths and limitations of the current state of the methodology and offers recommendations for further development of the methodology.

4.1 The proposed benchmarking methodology

The research presented in this paper resulted in a proposal for a methodology to develop benchmarks based on data analysis of a given dataset. This dataset consists of set of buildings, technical building characteristics and measured data. Therefore the methodology is considered to be advantageous for companies, institutions or collectives targeting energy efficiency of buildings in their portfolio. The only prerequisite is availability of a significantly large set of buildings and associated building data. Significance of the size of the dataset depends on the diversity of buildings within the set. The methodology is advantageous in comparison to most public benchmarks because more types of building categorization can be considered and multiple benchmarks can be used for a thorough assessment. The research showed interesting alternatives for categorization by building function, showed the potential of benchmarking using subsets of similar buildings and can be used to assess historic performance. Also, to some extent, energy saving potential could be located within the building, through analysis of multiple benchmarks for one building. If the benchmarking results of one type of categorization are significantly lower than others, the defining categorization parameter is assumed to be a good place to start looking for possibilities to enhance performance. Also, because of the application of regression analysis, the normalization process seems to be superfluous. Floor area and degree days, for example, are implemented in the regression model used to calculate predicted energy consumption. This finding was supported by the coefficients of determination (R^2) for regression models found for subsets with normalized data (Table 3, Section 3.1, and Appendix 4). The methodology is flexible to numerous approaches for categorizing buildings and choosing predictor parameters. Therefore it is possible to start with a limited set of buildings and predictor parameters and get initial results. Increasing the number of buildings or predictor parameters can be carried out over time to increase accuracy of the results. In the same manner, submetering and time resolution can be implemented to amplify the possibilities of using results for assessing building energy performance. The current state of the methodology shows interesting results, however, robustness of the methodology can be significantly increased and the methodology could be expanded. The conclusions on these possible improvements are discussed in the next subsections. Finally, one of the major conclusions of the research is that the assessment of multiple benchmarks, based on multiple (single or combinations of) categorization parameters, can help enhance the understanding of building performance and therefore help improve it.

4.2 Combining categorization parameters and clustering buildings

The results presented in this paper indicate that further research on the categorization of buildings can lead to improved results of the method. The aim of categorization is to define subsets of buildings with similar characteristics. This can be done based on one categorization of a combination of parameters. Based on the final conclusion of the previous subsection, it must be noted that it is not the goal of the methodology to determine one ultimate type of categorization to benchmark performance. Multiple types can show sufficient reliability and the comparison of those can lead to new insight on energy performance.

Looking back at the building age categorization, the low scoring categories show high root mean squared error (RMSE) indicating a wide range of energy consumption within this category. Dividing these categories in multiple subcategories might improve the results. Furthermore, the focus of this research has been on categorization based on one single building parameter. However, it is expected that combining building parameters for categorization could improve the ability to assess energy performance of buildings. For example, looking at all office buildings built after 2000 or buildings with a façade of mainly glass and a heat pump cooler. The combination of building parameters can be made as complex as desired, increasing the similarity of buildings within the building category. Future research could investigate the added value of applying state of the art clustering techniques for building categorization.

4.3 Challenges in data availability

Building categories can contain as many buildings as available, but a larger set of buildings is likely to generate more reliable results. In general, smaller sets show lower R^2 except when the RMSE or dispersion of energy consumption is relatively low. To test the minimum amount of buildings the p-value can be used, a p-value exceeding 0,05 is considered as an indication for unreliable results [7]. In the current research it was found that a minimum of 35 buildings was required for the regression model to be usable for predicting energy consumption. This number should be reconsidered every time the methodology is applied.

The research was carried out using the CBECS 2003 data because of the large number of buildings and building parameters. Although the data is considered useful for testing the methodology, a number of shortcomings have emerged as well. The data from the survey was collected more than a decade before the research was carried out. In this decade, a lot of progress was made concerning energy efficiency in both existing and newly constructed buildings. For example, on-site generation of electricity and heat using solar or wind energy was not included in the dataset. Also, overall energy consumption patterns most likely have changed quite a lot due to developments in appliances and installations. A similar survey was carried out in 2012, unfortunately the complete dataset is only scheduled to be released end of 2015. A more recent dataset might be useful for further testing and developing of the methodology.

The dataset contained a lot of categorical data where interval data might be expected or more useful. For example, wall type is only included in the research using categories like *brick wall* or *glass facades*. Window glass types in the survey are divided in *single glass*, *multi-layer glass* or a *combination*. A transition towards working with insulation values, air tightness, etc., is expected to have a significant positive effect on the ability to assess energy performance. The same can be said for *heating* and *cooling systems*, *servers* and *computers*. Capacities, efficiencies and running time might be interesting data to take into account. General tendency is to install increasing amounts of measurement instrument, this data can be used for energy performance assessment using the proposed methodology. Also, a number of parameters were not found in the survey data which might have significant influence, like façade area and air tightness of the building.

4.4 Choice of parameters

The set of building parameters selected from the survey data was based on literature study and availability. This makes this set arbitrary and leaves it up for discussion. Improvement of the availability of good quality data as mentioned above and a more thorough sensitivity analysis of building parameters might improve this set of building parameters. For example, the parameter wall type is a nominal value which means it cannot be ranked, so one type cannot be considered better or worse than another type. This makes it very hard to include this parameter in a sensitivity analysis or regression analysis. Therefore, this parameter is only used as a categorization parameter while it is expected to be significant to predicting energy consumption when expressed as insulation values. A simple sensitivity analysis was carried out for the selected set of parameters, but results were not very conclusive so for this research the set of parameters presented in Table 1, no parameters were discarded for the next steps of the research.

4.5 Resolution of the dataset and methodology

The methodology can be adapted from a low dimensional dataset, like annual whole building energy consumption and a limited number of parameters, to a higher dimension for time, building level and level of detail of building parameters. The current research focuses on annual whole building energy consumption. However, the methodology is expected to be suitable for assessing on different levels as well. Time resolution might be changed to a monthly, daily or even hourly scale to look at the difference between seasons, weekends and weekdays or day and night. Also zooming in from whole building level to floor, room, or system level might help locate inefficient energy performance and enable significant savings. Future research is needed to further develop the methodology for this purpose, especially concerning interpreting results of an assessment.

4.6 Recommendations for future research

From the findings of this results and the observed shortcomings and opportunities to improve were identified. The methodology proposed in shows promising results, but is not yet robust enough. For example, the case study showed deviant results for cooling, for which possible explanation were suggested. Uncertainties in data collection and parameter selection make it difficult to draw conclusions in such a case. Further research on the following topics can probably significantly improve the robustness of the methodology. Update the database used for testing the methodology is not one of the steps mentioned, but is expected to be useful for any kind of future research.

1. Reconsider the parameter set from Table 1 to develop a minimal dataset for robust performance assessment of buildings, including the possible values of categorical data.
2. Test the methodology for building categories defined by a combination of categorization parameters or clustering using state-of-the-art techniques to maximize similarity of buildings within the subsets.
3. Adapt the methodology from annual whole building, to monthly or smaller time resolution and to submetering on floor, system or component level to more precisely locate energy saving potential.
4. Further investigate the possibilities to locate energy saving potential using the results of multiple benchmarks, based on multiple categorization approaches.

5 References

1. European Commission. The Commission's new Energy Efficiency Plan. (2011).
2. J. Field, Cibse TM46: Energy Benchmarks, Chartered Institute for Building Services Engineering, London, 2008.
3. http://www.eia.gov/consumption/commercial/data/2012/, author unknown, last visited: 24/8/2015, *Commercial Buildings Energy Consumption Survey (CBECS): 2012 CBECS Survey Data*, U.S. Energy Information Administration.
4. Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated. Meteorol. Z., 15, 259-263. DOI: 10.1127/0941-2948/2006/0130.
5. http://nl.mathworks.com/help/stats/partialcorr.html, author unknown, last visited: 24/8/2015, *MatLab R2015a Documentation: partialcorr*, Mathworks.
6. Tian, W. A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews* **20,** 411–419 (2013).
7. De Sá, J. Applied Statistics Using SPSS, STATISTICA, MATLAB and R. (2007). doi:10.1007/978-3-540-71972-4
8. Wang, E., Shen, Z. & Grosskopf, K. Benchmarking energy performance of building envelopes through a selective residual-clustering approach using high dimensional dataset. *Energy and Buildings* **75,** 10–22 (2014).
9. Allison, P. D. *Multiple Regression: A Primer*. (Pine Forge Press, 1999).
10. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and intelligent laboratory systems* **2,** 37–52 (1987).
11. http://nl.mathworks.com/help/stats/pca.html, author unknown, last visited: 24/8/2015, *MatLab R2015a Documentation: pca*, Mathworks.
12. http://nl.mathworks.com/help/stats/fitlm.html, author unknown, last visited: 24/8/2015, *MatLab R2015a Documentation: fitlm*, Mathworks.
13. http://www.eia.gov/consumption/commercial/data/2003, author unknown, last visited: 24/8/2015, *Commercial Buildings Energy Consumption Survey (CBECS): 2003 CBECS Survey Data*, U.S. Energy Information Administration.
14. http://www.rapidtables.com/, author unknown, last visited: 24/8/2015, *Rapidtables online reference & tools: Online Scientific and Engineering Resource*, Rapidtables.com.
15. Neumann, C. & Jacob, D. *Building EQ: Guidelines for the evaluation of Building Performance*. (Fraunhofer institute for Solar Energy Systems, 2008).
16. Heiselberg, P. K. & Brohus, H. Sensitivity Analysis Applied in Design of Low Energy Office Building . in **2,** 33–40 (AIVC, 14AD).
17. Lomas, K. & Eppel, H. Sensitivity analysis techniques for building thermal simulation programs. *Energy and Buildings* **19,** 21–44 (1992).
18. Heller, J., Heater, M. & Frankel, M. *Sensitivity Analysis: Comparing the Impact of Design, Operation, and Tenant Behavior on Building Energy Performance*. (New Buildings Institute, 2011).
19. Olofsson, T., Andersson, S. & Sjögren, J.-U. Building energy parameter investigations based on multivariate analysis. *Energy and Buildings* **41,** 71–80 (2009).
20. http://nl.mathworks.com/help/econ/examples/time-series-regression-ii-collinearity-and-estimator-variance.html, author unknown, last visited: 24/8/2015, *MatLab R2015a Documentation: Time Series Regression II: Collinearity and Estimator Variance*, Mathworks.
21. Lam, J., Wan, K., Cheung, K. & Yang, L. Principal component analysis of electricity use in office buildings. *Energy and Buildings* **40,** 828–836 (2008).

Appendices

1. Data preparation, filtering and assumptions

As mentioned, in order to test the methodology the CBECS 2003 dataset was used as can be found on the website of U.S. Energy Information Administration [13]. The total set of micro data contains of over 5000 buildings and per building hundreds of possible building parameters in combination with energy consumption data. Therefore it was considered a suitable source of data for the current research. The actual data showed to be usable but not perfect for the methodology during the course of this research. Some weaknesses of the data could be eliminated by recalculating and filtering the data. Others have to be considered while interpreting the results and efforts should be made to improve data collection quality in future work, as discussed in Section 4.

First of all, the dataset was recalculated to European standard units using the conversion factors in Table 6. The research focusses on energy consumption so observations without data for primary consumption were removed from the dataset. Primary consumption is defined as the total of energy consumption from all individual energy sources. Primary energy is measured in British Thermal Units (Btu) in the U.S. system and in mega joule (MJ) in the European system. Observations with less than 12 months of energy consumption data were also removed from the dataset. Also, malls, vacant buildings, and buildings with function description ‘other function’ were removed because of a lack of data.

Table 6: Conversion factors [14] *for recalculating CBECS 2003 (US) data to European standards.*

Units	*US*	*EU*	Factor
Floor area	ft^2	m^2	0,09290
Energy	*thous Btu*	*MJ*	1,05597
Degree day	$DD_{F;65}$	$DD_{C;18}$	5/9

As mentioned before, a selection was made concerning the building parameters used, as presented in Table 1. One group of parameters was used for categorizing buildings into smaller subsets of higher similarity. Another group of parameters was used as input for regression analysis and the benchmarking process. The selection of parameters was made based on availability of data, the type of parameter values and literature study [15-19]. Consequently some of the parameters are not used as might be expected, because the type of data was not as desired. For example, *wall type* is incorporated in the CBECS 2003 dataset as a nominal parameter while insulation values might me more interesting. More on this subject can be found in Section 4.

For a significant number of observations, one or more parameter values were missing. Although the used MatLab functions mostly use built in strategies to deal with missing values, in some cases it was considered to be desirable to make assumptions on parameter values for missing data. For number of elevators and escalators, as well as number of servers and missing energy consumption data, is was assumed that missing values indicate the value is 0. The percentage of exterior glass was assumed to be 25% when in case of missing data and the number of computers per person was assumed to be the average number over the filtered dataset, being approximately 0.86 computers per person. Furthermore, *server* and *computer categories* were replaced by the median value of the category interval.

Finally, some parameters were recalculated. *Natural gas* and *fuel oil* were combined and renamed *fossil fuels*. Years of construction were recalculated to building age with 2004 as base value. In other words, buildings age is 2004 minus the year of construction. Percentages for *heated* and *cooled floor area* were recalculated to actual areas by multiplying the percentages with the value of *total floor area*.

During the course of the research, it was concluded that a subset of building data should contain at least 35 buildings to in order to provide useful results. This number was found using the current dataset and might be different when analyzing another building portfolio. The current research, testing of the methodology, analyses and benchmarking process were carried out using a dataset from the United States. By carrying a case study using Strukton buildings the results were tested for Dutch buildings. Because of the large variety of climatic conditions within the U.S. one climate zone was assumed to be most similar to Dutch climate. This was assumed to Seattle, located in Census Region 9, Pacific.

2. Multicollinearity and Principal Component Analysis

The danger of a high dimensional dataset is the existence of correlations between parameters representing building characteristics. In case of Multiple Regression Analysis this can lead to unreliable results 9and therefore the presence of multicollinearity has to be checked. Variance Inflation Factors (VIFs) are calculated to this extent [8,20].

$$VIF_p = \left|\frac{1}{1-R_p^2}\right| \tag{1}$$

```
MatLab code:      VIF=abs (diag(inv(rho))); [20]
```

With R_p^2(=ρ) the coefficients of multiple determination or Spearman's ρ (Partial Rank Correlation Coefficients) of the predictor parameters found by MatLab-function *partialcorr();* [5] and VIF an Nx1 matrix if VIFs for every predictor parameter. N is the number of predictor parameters.

Depending on the results and rule of thumb proposed by Allison et al. [9], a ruling on the presence of multicollinearity is made. It is expected that multicollinearity will be found. Correlated parameters might be discarded and benchmarks can be developed for the remaining predictor parameters. Another approach to the multicollinearity issue is Principal Component Analysis (PCA) [8, 10, 21]. This technique converts the predictor parameters into a number of uncorrelated components by combining correlated parameters. In order to create uncorrelated input for the regression analysis, built-in MatLab function *pca();* [11] was used. This function determines principal component coefficients which can be used to recalculate parameter values of every observation to principal components in such a manner that principal component are not significantly correlated and therefore are suitable as predictor parameters for multiple regression analysis.

Every principal component explains a certain amount of variance in the data. The number of principal components is maximum the number of original parameters, however, it is possible that 100% variance explained is reached using less principal components. Moreover, in literature it is suggested to only keep principal components up to a number of approximately 75% of variance explained [21] in order to simplify further steps in research. Discarded principal components represent such a small significance in influence on energy consumption that they can be ignored. The percentage of variance explained is provided by the MatLab function as explained. The function also provides in scores, being representations of the observations using principal components instead of the original predictor parameters. These scores result in a new dataset for which parameters, principal components in this case, are supposed to be not significantly correlated. This can be checked by recalculating VIFs for the principal components.

Figure 17 shows the VIFs calculated for the predictor parameters using Equation 1, for the subset of buildings built after 1999, indicating multicollinearity may cause problems due to correlated predictor parameters. Therefore, PCA was performed for the predictor parameters of this dataset resulting in principal component coefficients and a recalculated dataset for this subset. Again, VIFs were calculated and results are shown in Figure 16. VIFs show to be significantly lower than 2.5 for the principal components, so multicollinearity should not cause problems in the regression analysis.

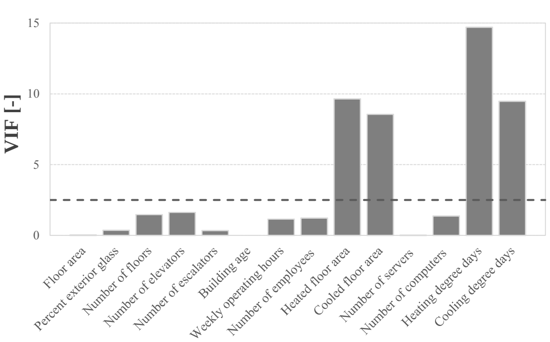

Figure 17: Variance inflation factors for the predictor parameters of the subset of buildings categorized by building age. The category used as an example is buildings built after 1999.

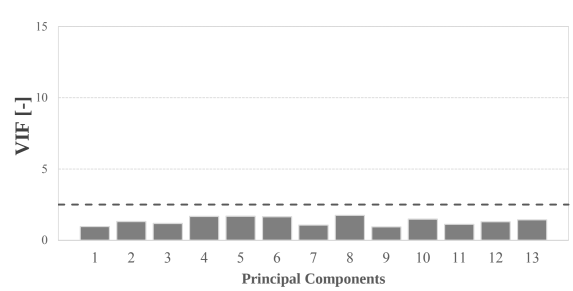

Figure 16: Variance inflation factors for principal components calculated from the predictor parameters of the subset of buildings built after 1999.

3. Subsets or building categories defined for categorization of buildings

Building function	
(predominant activity in the building)	
2	Office
4	Laboratory
5	Non-refrigerated warehouse
6	Food sales
7	Public order and safety
8	Outpatient health care
11	Refrigerated warehouse
12	Religious worship
13	Public assembly
14	Education
15	Food service
16	Inpatient health care
17	Nursing
18	Lodging
25	Retail other than mall
26	Service

Year of construction*	
	Range
1	1999-2003
2	1994-1998
3	1988-1993
4	1981-1987
5	1975-1980
6	1969-1974
7	1960-1968
8	1950-1959
9	1927-1949
10	1771-1926

* For categorization and further analysis building age is used by subtracting the year of construction from base year 2004. Case buildings from after 2004 are categorized in the first building class (1999-2003)

Location	
*US Census Divisions**	
1	New England
2	Middle Atlantic
3	East North Central
4	West North Central
5	South Atlantic
6	East South Central
7	West South Central
8	Mountain
9	Pacific

* Dutch climate was assumed to be most similar to the Pacific division. Census Divisions can be found following this link:

www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Heating equipment type	
1	Furnaces that heat air directly
2	Boilers inside the building
3	Packaged heating units
4	Individual space heaters
5	Heat pumps for heating
6	District steam or hot water
7	Other heating equipment

Cooling equipment type	
1	Packaged A/C units
2	Residential-type central A/C
3	Individual room A/C
4	Heat pumps for cooling
5	District chilled water piped in
6	Central chillers inside the building
7	"Swamp"-coolers or evaporative coolers

Building Shape	
1	Square
2	Wide rectangle
3	Narrow rectangle
4	Rectangle/square with courtyard
5	"H"-shaped
6	"U"-shaped
7	"E"-shaped
8	"T"-shaped
9	"L"-shaped
10	"+"-or cross shaped
11	Other shape

Wall type	
1	Brick, stone, or stucco
2	Pre-cast concrete panels
3	Concrete block or poured concrete
4	Siding, shingles, tiles, or shakes
5	Sheet metal panels
6	Window or vision glass

Window glass type	
1	Single layer glass
2	Multi-layer glass
3	Combination of both
4	No windows

Roof type	
1	Built-up
2	Slate or tile shingles
3	Wood shingles/shakes/other wood
4	Asphalt/fiberglass/other shingles
5	Metal surfacing
6	Plastic/rubber/synthetic sheeting
7	Concrete

Water heating type	
1	One or more centralized
2	One or more "point-of-use"-
3	Both types of water heaters

4. Results from Multiple Regression Analysis (MRA) of individual building categories

Building Age

[m²]	R^2	*p*	number of buildings	RMSE
after 1999	**82,94%**	*1,34E-133*	385	7,39E+06
1994-1998	**87,23%**	*1,93E-144*	357	9,96E+06
1988-1993	**53,02%**	*2,21E-44*	370	2,48E+07
1981-1987	**53,39%**	*5,20E-53*	381	3,35E+07
1975-1980	**80,18%**	*1,31E-122*	388	1,82E+07
1969-1974	**94,71%**	*3,90E-211*	359	1,72E+07
1960-1968	**90,59%**	*1,30E-202*	427	1,12E+07
1950-1959	**93,37%**	*1,73E-181*	337	1,05E+07
1927-1949	**94,56%**	*5,98E-143*	323	7,88E+06
1771-1926	**87,64%**	*1,85E-150*	365	5,74E+06
	81,48%	***2,21E-45***	**3692**	1,48E+07
after 1999	**27,07%**	*1,01E-19*	385	901
1994-1998	**66,42%**	*5,48E-74*	357	996
1988-1993	**31,82%**	*1,16E-23*	370	904
1981-1987	**28,13%**	*1,44E-20*	381	907
1975-1980	**30,36%**	*1,99E-23*	388	999
1969-1974	**29,03%**	*5,37E-20*	359	1036
1960-1968	**30,25%**	*3,77E-26*	427	878
1950-1959	**42,53%**	*1,58E-32*	337	729
1927-1949	**29,81%**	*2,19E-13*	323	907
1771-1926	**17,39%**	*6,95E-10*	365	915
	33,06%	***6,88E-11***	**3692**	918
after 1999	**29,02%**	*5,86E-23*	385	0,29
1994-1998	**50,86%**	*1,14E-47*	357	0,41
1988-1993	**32,91%**	*4,35E-26*	370	0,30
1981-1987	**28,76%**	*2,12E-22*	381	0,32
1975-1980	**30,03%**	*2,74E-24*	388	0,33
1969-1974	**30,49%**	*1,21E-22*	359	0,33
1960-1968	**30,81%**	*2,00E-23*	427	0,27
1950-1959	**34,78%**	*2,63E-25*	337	0,23
1927-1949	**25,12%**	*2,85E-15*	323	0,30
1771-1926	**20,54%**	*1,79E-13*	365	0,26
	31,30%	***1,79E-14***	**3692**	0,30

Location

[m²]	R^2	*p*	number of buildings	RMSE
1=New England	**93,91%**	*3,00E-75*	151	9,35E+06
2=Middle Atlantic	**89,89%**	*7,54E-231*	500	1,98E+07
3=East North Central	**78,44%**	*5,33E-207*	669	2,22E+07
4=West North Central	**92,75%**	*1,24E-175*	340	5,28E+06
5=South Atlantic	**89,30%**	*0,00E+00*	694	1,23E+07
6=East South Central	**97,15%**	*1,04E-130*	211	1,02E+07
7=West South Central	**86,44%**	*1,18E-172*	436	9,47E+06
8=Mountain	**91,23%**	*3,85E-104*	228	9,29E+06
9=Pacific	**41,79%**	*2,46E-44*	463	2,21E+07
	82,19%	***3,09E-45***	**3692**	1,49E+07
1=New England	**62,36%**	*1,10E-14*	151	1082
2=Middle Atlantic	**28,63%**	*1,39E-28*	500	1080
3=East North Central	**26,33%**	*5,90E-36*	669	901
4=West North Central	**47,59%**	*1,82E-38*	340	679
5=South Atlantic	**31,34%**	*1,61E-47*	694	804
6=East South Central	**24,33%**	*9,52E-08*	211	1235
7=West South Central	**30,87%**	*1,24E-23*	436	926
8=Mountain	**28,91%**	*1,02E-10*	228	974
9=Pacific	**24,76%**	*2,90E-21*	463	747
	31,40%	***5,44E-09***	**3692**	901
1=New England	**32,88%**	*3,15E-08*	151	0,23
2=Middle Atlantic	**29,41%**	*5,77E-31*	500	0,28
3=East North Central	**27,31%**	*3,88E-39*	669	0,23
4=West North Central	**46,91%**	*5,14E-39*	340	0,16
5=South Atlantic	**34,28%**	*2,79E-55*	694	0,28
6=East South Central	**34,67%**	*2,32E-14*	211	0,42
7=West South Central	**35,30%**	*4,66E-34*	436	0,30
8=Mountain	**29,07%**	*1,00E-11*	228	0,27
9=Pacific	**29,42%**	*2,56E-28*	463	0,45
	32,68%	***1,29E-09***	**3692**	0,29

Function

[m²]	R^2	*p*	number of buildings	RMSE
02=Office	**89,45%**	*0,00E+00*	783	1,04E+07
04=Laboratory	**94,21%**	*3,91E-11*	37	1,91E+07
05=Nonrefrigerated warehouse	**84,02%**	*1,01E-124*	349	3,72E+06
06=Food sales	**95,30%**	*3,17E-51*	100	1,05E+06
07=Public order and safety	**98,10%**	*8,06E-43*	71	3,03E+06
08=Outpatient health care	**74,19%**	*8,02E-26*	124	5,90E+06
12=Religious worship	**95,25%**	*4,39E-116*	247	4,49E+05
13=Public assembly	**96,17%**	*3,79E-113*	221	1,03E+07
14=Education	**83,26%**	*7,50E-190*	531	4,02E+06
15=Food service	**88,42%**	*4,71E-79*	201	9,69E+05
16=Inpatient health care	**62,53%**	*1,96E-28*	177	9,21E+07
17=Nursing	**96,45%**	*3,64E-29*	59	5,81E+06
18=Lodging	**97,44%**	*3,98E-153*	218	3,94E+06
25=Retail other than mall	**92,79%**	*5,28E-149*	292	1,43E+06
26=Service	**99,76%**	*8,79E-223*	282	1,18E+06
	88,99%	***3,91E-13***	**3692**	9,26E+06
02=Office	**19,73%**	*1,66E-29*	783	565
04=Laboratory	**35,87%**	*4,25E-01*	37	1734
05=Nonrefrigerated warehouse	**27,28%**	*8,58E-12*	349	371
06=Food sales	**15,78%**	*2,02E-01*	100	1405
07=Public order and safety	**25,30%**	*1,52E-01*	71	699
08=Outpatient health care	**57,75%**	*1,14E-14*	124	676
12=Religious worship	**23,31%**	*8,61E-07*	247	308
13=Public assembly	**41,96%**	*1,58E-18*	221	746
14=Education	**23,49%**	*2,23E-23*	531	586
15=Food service	**33,37%**	*2,55E-11*	201	2571
16=Inpatient health care	**8,87%**	*2,71E-01*	177	1330
17=Nursing	**40,10%**	*5,96E-03*	59	713
18=Lodging	**30,78%**	*4,40E-09*	218	647
25=Retail other than mall	**25,24%**	*3,84E-12*	292	666
26=Service	**16,89%**	*2,37E-06*	282	1136
	25,41%	***2,57E-02***	**3692**	789
02=Office	**17,97%**	*2,72E-27*	783	0,19
04=Laboratory	**20,93%**	*7,48E-01*	37	0,73
05=Nonrefrigerated warehouse	**36,85%**	*1,28E-28*	349	0,09
06=Food sales	**15,55%**	*2,24E-01*	100	0,51
07=Public order and safety	**23,67%**	*1,73E-01*	71	0,23
08=Outpatient health care	**45,00%**	*1,52E-10*	124	0,22
11=Refrigerated warehouse	**0,00%**	*0,00E+00*	0	0,00
12=Religious worship	**21,33%**	*1,15E-06*	247	0,09
13=Public assembly	**41,08%**	*5,02E-19*	221	0,22
14=Education	**18,15%**	*1,95E-17*	531	0,20
15=Food service	**29,80%**	*2,46E-10*	201	0,97
16=Inpatient health care	**21,91%**	*7,45E-06*	177	0,45
17=Nursing	**18,36%**	*3,17E-01*	59	0,28
18=Lodging	**21,45%**	*8,74E-06*	218	0,22
25=Retail other than mall	**27,46%**	*5,33E-13*	292	0,21
26=Service	**11,54%**	*4,13E-04*	282	0,47
	23,67%	***2,20E-02***	**3692**	0,27

Water heating system

[m²]	R^2	*p*	number of buildings	RMSE
1=One or more centralized	**93,59%**	*0,00E+00*	2194	7,21E+06
2=One or more "point-of-use"	**86,94%**	*9,56E-265*	639	6,20E+06
3=Both types of water heaters	**62,55%**	*3,77E-90*	482	4,35E+07
	87,79%	***5,49E-91***	**3315**	1,23E+07
1=One or more centralized	**35,07%**	*2,23E-193*	2194	750
2=One or more "point-of-use"	**25,91%**	*2,43E-33*	639	780
3=Both types of water heaters	**31,57%**	*2,12E-31*	482	855
	32,79%	***3,14E-32***	**3315**	771
1=One or more centralized	**39,27%**	***1,07E-201***	**2194**	0,23
2=One or more "point-of-use"	**30,07%**	*3,09E-42*	639	0,24
3=Both types of water heaters	**29,97%**	*1,93E-30*	482	0,28
	36,15%	***2,81E-31***	**3315**	0,24

Wall type

[m²]	R^2	*p*	number of buildings	RMSE
1=Brick, stone, or stucco	**87,43%**	*0,00E+00*	1934	9,67E+06
2=Pre-cast concrete panels	**79,84%**	*1,67E-82*	277	2,30E+07
3=Concrete block or poured concrete	**92,66%**	*0,00E+00*	640	9,59E+06
4=Siding, shingles, tiles, or shakes	**91,72%**	*1,94E-173*	351	9,88E+05
5=Sheet metal panels	**62,66%**	*2,35E-66*	366	1,18E+07
6=Window or vision glass	**85,71%**	*1,09E-08*	53	1,56E+07
	85,66%	***1,60E-10***	**3621**	1,01E+07
1=Brick, stone, or stucco	**38,79%**	*1,07E-193*	1934	755
2=Pre-cast concrete panels	**29,13%**	*4,87E-14*	277	993
3=Concrete block or poured concrete	**29,72%**	*2,98E-40*	640	897
4=Siding, shingles, tiles, or shakes	**21,79%**	*6,29E-13*	351	865
5=Sheet metal panels	**35,82%**	*4,07E-27*	366	548
6=Window or vision glass	**31,93%**	*2,00E-01*	53	722
	34,40%	***2,93E-03***	**3621**	787
1=Brick, stone, or stucco	**39,94%**	*1,30E-203*	1934	0,24
2=Pre-cast concrete panels	**27,73%**	*5,21E-14*	277	0,32
3=Concrete block or poured concrete	**30,00%**	*3,67E-42*	640	0,29
4=Siding, shingles, tiles, or shakes	**22,34%**	*2,81E-11*	351	0,30
5=Sheet metal panels	**98,97%**	*0,00E+00*	366	0,17
6=Window or vision glass	**31,02%**	*1,14E-01*	53	0,29
	41,38%	***1,67E-03***	**3621**	0,26

Building Shape

[m²]	R^2	*p*	number of buildings	RMSE
1=Square	**82,69%**	*1,91E-110*	329	1,02E+07
2=Wide rectangle	**79,66%**	*0,00E+00*	1898	7,43E+06
3=Narrow rectangle	**76,03%**	*7,91E-41*	215	1,07E+07
4=Rectangle/square with courtyard	**57,45%**	*1,68E-13*	117	5,98E+07
5="H" shaped	**88,90%**	*3,45E-21*	79	1,73E+07
6="U" shaped	**84,45%**	*1,23E-27*	98	1,16E+07
7="E" shaped	**86,34%**	*3,06E-08*	42	3,77E+07
8="T" shaped	**92,32%**	*2,22E-35*	97	1,35E+07
9="L" shaped	**75,63%**	*1,30E-45*	216	4,59E+06
10="+" or cross shaped	**96,18%**	*2,05E-30*	70	1,63E+07
11=Other shape	**93,90%**	*2,01E-86*	170	2,04E+07
	80,57%	***3,85E-10***	**3331**	1,13E+07
1=Square	**17,86%**	*1,51E-08*	329	1354
2=Wide rectangle	**38,71%**	*2,19E-189*	1898	656
3=Narrow rectangle	**43,88%**	*2,31E-14*	215	1010
4=Rectangle/square with courtyard	**32,72%**	*4,65E-05*	117	868
5="H" shaped	**38,78%**	*1,03E-03*	79	798
6="U" shaped	**33,77%**	*4,51E-04*	98	800
7="E" shaped	**71,59%**	*2,27E-04*	42	1250
8="T" shaped	**82,77%**	*2,06E-23*	97	653
9="L" shaped	**24,97%**	*1,63E-06*	216	670
10="+" or cross shaped	**53,47%**	*1,09E-05*	70	793
11=Other shape	**40,87%**	*3,46E-11*	170	811
	37,86%	***4,26E-05***	**3331**	782
1=Square	**16,96%**	*6,56E-07*	329	0,46
2=Wide rectangle	**42,03%**	*4,93E-214*	1898	0,21
3=Narrow rectangle	**33,75%**	*2,85E-10*	215	0,32
4=Rectangle/square with courtyard	**25,13%**	*8,70E-04*	117	0,31
5="H" shaped	**32,82%**	*2,85E-03*	79	0,30
6="U" shaped	**32,57%**	*1,98E-04*	98	0,29
7="E" shaped	**67,66%**	*6,92E-05*	42	0,42
8="T" shaped	**69,77%**	*9,26E-18*	97	0,24
9="L" shaped	**26,47%**	*1,61E-09*	216	0,20
10="+" or cross shaped	**57,32%**	*1,11E-06*	70	0,27
11=Other shape	**35,27%**	*1,33E-09*	170	0,29
	38,03%	***1,05E-04***	**3331**	0,26

Roof type

[m²]	R^2	*p*	number of buildings	RMSE
1=Built-up	**83,70%**	*0,00E+00*	1147	1,37E+07
2=Slate or tile shingles	**96,02%**	*5,68E-109*	203	3,97E+06
3=Wood shingles/shakes/other wood	**93,74%**	*1,11E-27*	68	1,63E+06
4=Asphalt/fiberglass/other shingles	**30,52%**	*2,22E-48*	737	1,39E+07
5=Metal surfacing	**95,96%**	*0,00E+00*	638	1,60E+06
6=Plastic/rubber/synthetic sheeting	**66,06%**	*1,04E-167*	776	2,21E+07
7=Concrete	**95,05%**	*1,34E-21*	78	9,00E+06
	72,46%	***2,87E-23***	**3647**	1,25E+07
1=Built-up	**34,04%**	*5,58E-93*	1147	883
2=Slate or tile shingles	**19,29%**	*7,48E-05*	203	1023
3=Wood shingles/shakes/other wood	**20,67%**	*5,43E-01*	68	994
4=Asphalt/fiberglass/other shingles	**30,02%**	*5,28E-48*	737	715
5=Metal surfacing	**73,07%**	*6,20E-168*	638	516
6=Plastic/rubber/synthetic sheeting	**32,78%**	*2,43E-57*	776	907
7=Concrete	**16,41%**	*8,24E-01*	78	1256
	38,34%	***2,78E-02***	**3647**	808
1=Built-up	**32,85%**	*2,22E-90*	1147	0,30
2=Slate or tile shingles	**15,20%**	*7,05E-04*	203	0,46
3=Wood shingles/shakes/other wood	**39,68%**	*7,24E-03*	68	0,33
4=Asphalt/fiberglass/other shingles	**29,77%**	*6,60E-49*	737	0,23
5=Metal surfacing	**35,50%**	*8,15E-53*	638	0,16
6=Plastic/rubber/synthetic sheeting	**33,27%**	*4,50E-60*	776	0,28
7=Concrete	**21,51%**	*1,07E-01*	78	0,43
	31,69%	***2,46E-03***	**3647**	0,27

Glass type

[m²]	R^2	*p*	number of	RMSE
1=Single layer glass	**91,00%**	*0,00E+00*	1392	5,71E+06
2=Multi-layer glass	**86,91%**	*0,00E+00*	1538	1,23E+07
3=Combination of both	**67,02%**	*7,14E-146*	664	2,83E+07
4=No windows	**98,17%**	*8,41E-66*	98	1,57E+06
	85,18%	***2,23E-67***	**3692**	1,24E+07
1=Single layer glass	**31,33%**	*8,40E-103*	1392	691
2=Multi-layer glass	**36,28%**	*6,84E-139*	1538	772
3=Combination of both	**47,70%**	*1,38E-82*	664	840
4=No windows	**49,19%**	*1,60E-02*	98	509
	36,81%	***4,25E-04***	**3692**	747
1=Single layer glass	**34,10%**	*9,52E-117*	1392	0,25
2=Multi-layer glass	**38,74%**	*1,06E-153*	1538	0,23
3=Combination of both	**50,92%**	*3,38E-93*	664	0,24
4=No windows	**63,61%**	*1,03E-14*	98	0,11
	39,84%	***2,75E-16***	**3692**	0,24

Heating system

[m²]	R^2	*p*	number of buildings	RMSE
1=Furnaces that heat air directly	**91,93%**	*0,00E+00*	1015	1,42E+06
2=Boilers inside the building	**87,22%**	*0,00E+00*	912	1,61E+07
3=Packaged heating units	**86,62%**	*2,85E-308*	747	3,33E+06
4=Individual space heaters	**96,82%**	*1,77E-149*	226	2,61E+06
5=Heat pumps for heating	**96,06%**	*3,49E-172*	273	2,02E+06
6=District steam or hot water	**64,99%**	*1,65E-42*	235	7,06E+07
7=Other heating equipment	**98,32%**	*9,43E-62*	109	2,86E+06
	88,62%	***1,10E-43***	**3517**	1,04E+07
1=Furnaces that heat air directly	**27,79%**	*3,75E-62*	1015	672,31
2=Boilers inside the building	**32,58%**	*4,10E-68*	912	778,49
3=Packaged heating units	**32,47%**	*3,22E-54*	747	920,18
4=Individual space heaters	**24,92%**	*2,05E-08*	226	463,02
5=Heat pumps for heating	**14,90%**	*5,11E-05*	273	891,84
6=District steam or hot water	**19,83%**	*2,93E-06*	235	1315,96
7=Other heating equipment	**28,83%**	*6,10E-04*	109	523,02
	28,34%	***2,31E-05***	**3517**	794,46
1=Furnaces that heat air directly	**29,89%**	*1,45E-61*	1015	0,20
2=Boilers inside the building	**35,62%**	*1,46E-78*	912	0,25
3=Packaged heating units	**31,61%**	*9,36E-54*	747	0,31
4=Individual space heaters	**20,49%**	*6,09E-07*	226	0,18
5=Heat pumps for heating	**17,21%**	*5,94E-07*	273	0,29
6=District steam or hot water	**17,76%**	*6,00E-06*	235	0,44
7=Other heating equipment	**66,23%**	*6,24E-15*	109	0,17
	30,47%	***4,86E-07***	**3517**	0,26

Cooling system

[m²]	R^2	*p*	number of buildings	RMSE
1=Packaged A/C units	**74,02%**	*0,00E+00*	1424	5,37E+06
2=Residential-type central A/C	**72,62%**	*3,31E-134*	528	4,30E+06
3=Individual room A/C	**93,82%**	*7,81E-196*	355	2,07E+06
4=Heat pumps for cooling	**91,52%**	*1,40E-144*	313	2,36E+06
5=District chilled water piped in	**84,85%**	*4,38E-45*	142	4,69E+07
6=Central chillers inside the building	**63,43%**	*1,93E-97*	505	4,21E+07
7="Swamp" coolers or evaporative coolers	**95,66%**	*1,64E-20*	50	1,22E+06
	76,74%	***2,48E-22***	**3317**	1,19E+07
1=Packaged A/C units	**34,92%**	*1,53E-121*	1424	713
2=Residential-type central A/C	**29,42%**	***1,08E-31***	**528**	841
3=Individual room A/C	**19,03%**	*2,12E-10*	355	674
4=Heat pumps for cooling	**15,40%**	*4,31E-06*	313	864
5=District chilled water piped in	**25,71%**	*1,59E-04*	142	1447
6=Central chillers inside the building	**32,85%**	*4,13E-35*	505	878
7="Swamp" coolers or evaporative coolers	**13,40%**	*9,15E-01*	50	1693
	29,47%	***1,38E-02***	**3317**	815
1=Packaged A/C units	**36,62%**	*2,12E-131*	1424	0,23
2=Residential-type central A/C	**30,52%**	*5,14E-33*	528	0,27
3=Individual room A/C	**21,96%**	*9,47E-14*	355	0,19
4=Heat pumps for cooling	**17,39%**	*1,89E-08*	313	0,28
5=District chilled water piped in	**17,65%**	*1,21E-02*	142	0,54
6=Central chillers inside the building	**27,30%**	*2,64E-28*	505	0,31
7="Swamp" coolers or evaporative coolers	**41,57%**	*1,88E-02*	50	0,20
	30,11%	***8,01E-04***	**3317**	0,26

5. Detailed results case buildings

Table 7: Input data for 2 years of the same case study building. For the Strukton office building in Son the results for 2011 and 2014 are compared.

Input data Cases		Strukton Building Son 2011		Strukton Building Son 2014	
Location	*[-]*	Pacific	*9*	Pacific	*9*
Floor area	*[m^2]*	7550		7550	
Function	*[-]*	Office	*2*	Office	*2*
Wall type	*[-]*	Brick, stone or stucco	*1*	Brick, stone or stucco	*1*
Roof type	*[-]*	Plastic/rubber/synthetic	*6*	Plastic/rubber/synthetic	*6*
Building shape	*[%]*	25		25	
Percent exterior glass	*[-]*	Wide Rectangle	*2*	Wide Rectangle	*2*
Number of floors	*[-]*	5		5	
Number of elevators	*[-]*	2		2	
Number of escalators	*[-]*	0		0	
Building age	*[years]*	1	*1*	1	*1*
Weekly operating hours	*[hours]*	60		60	
Number of employees	*[persons]*	350		350	
Heated floor area	*[m^2]*	7550		7550	
Main heating system	*[-]*	Heat pump for heating	*5*	Heat pump for heating	*5*
Cooled floor area	*[m^2]*	7000		7000	
Main cooling system	*[-]*	Heat pump for cooling	*4*	Heat pump for cooling	*4*
Water heating system	*[-]*	Point of use	*2*	Point of use heating	*2*
Number of servers	*[-]*	5		5	
Number of computers	*[-]*	350		350	
Glass type	*[-]*	Multi-layer	*2*	Multi-layer	*2*
Heating degree days	*[$DD_{C;18}$]*	2587		2406	
Cooling degree days	*[$DD_{C;18}$]*	79		109	
Annual major fuel consumption	*[MJ]*	5.579.379		3.977.645	
Annual electricity consumption	*[kWh]*	673.149		654.733	
Annual fossil consumption	*[MJ]*	3.156.043		1.620.607	
Annual district heat consumption	*[MJ]*	0		0	

Benchmarking using residuals based on total energy comsumption *[TJ]*

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2011		**Original**	**88,94%**	*0,00E+00*	-109,460	-3,670	-0,576	0,000	0,153	2,142	34,520	1264,723	**-17,936**	**A**
	10	**Location**	**41,79%**	*2,46E-44*	-56,678	-4,241	-0,665	-0,056	0,094	1,263	20,506	1008,250	**-11,218**	**A**
	3	**Function**	**89,45%**	*0,00E+00*	-75,291	-10,468	-0,880	-0,017	0,132	1,626	19,589	274,247	**-27,913**	**A**
	4	**Wall type**	**87,43%**	*0,00E+00*	-121,264	-4,342	-0,820	0,032	0,230	2,521	40,998	978,928	**-17,597**	**A**
	9	**Roof type**	**66,06%**	*1,04E-167*	-73,165	-8,260	-1,752	-0,011	0,420	6,324	71,054	981,216	**-35,899**	**A**
	8	**Building shape**	**79,66%**	*0,00E+00*	-54,024	-3,055	-0,488	-0,035	0,084	1,286	16,705	789,869	**-15,176**	**A**
	7	**Building age**	**82,94%**	*1,34E-133*	-32,833	-5,044	-0,902	-0,033	0,145	1,205	13,989	601,473	**-1,992**	**B**
	1	**Main heating system**	**96,06%**	*3,49E-172*	-9,256	-1,608	-0,458	0,014	0,097	0,680	5,959	23,371	**-5,617**	**A**
	2	**Main cooling system**	**91,52%**	*1,40E-144*	-11,321	-1,381	-0,393	0,005	0,077	0,686	6,067	44,327	**5,808**	**F**
	5	**Water heating system**	**86,94%**	*9,56E-265*	-112,762	-2,544	-0,562	-0,013	0,137	1,038	12,128	207,549	**-8,832**	**A**
	6	**Glass type**	**86,91%**	*0,00E+00*	-150,496	-6,932	-1,018	-0,028	0,213	2,407	37,866	969,665	**-20,542**	**A**

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2014		**Original**	**88,94%**	*0,00E+00*	-109,46	-3,67	-0,58	0,00	0,15	2,14	34,52	1264,72	**-19,63**	**A**
	10	**Location**	**41,79%**	*2,46E-44*	-56,68	-4,24	-0,66	-0,06	0,09	1,26	20,51	1008,25	**-12,89**	**A**
	3	**Function**	**89,45%**	*0,00E+00*	-75,29	-10,47	-0,88	-0,02	0,13	1,63	19,59	274,25	**-29,63**	**A**
	4	**Wall type**	**87,43%**	*0,00E+00*	-121,26	-4,34	-0,82	0,03	0,23	2,52	41,00	978,93	**-19,30**	**A**
	9	**Roof type**	**66,06%**	*1,04E-167*	-73,16	-8,26	-1,75	-0,01	0,42	6,32	71,05	981,22	**-37,62**	**A**
	8	**Building shape**	**79,66%**	*0,00E+00*	-54,02	-3,06	-0,49	-0,03	0,08	1,29	16,71	789,87	**-16,85**	**A**
	7	**Building age**	**82,94%**	*1,34E-133*	-32,83	-5,04	-0,90	-0,03	0,14	1,21	13,99	601,47	**-3,64**	**B**
	1	**Main heating system**	**96,06%**	*3,49E-172*	-9,26	-1,61	-0,46	0,01	0,10	0,68	5,96	23,37	**-7,26**	**A**
	2	**Main cooling system**	**91,52%**	*1,40E-144*	-11,32	-1,38	-0,39	0,00	0,08	0,69	6,07	44,33	**4,21**	**F**
	5	**Water heating system**	**86,94%**	*9,56E-265*	-112,76	-2,54	-0,56	-0,01	0,14	1,04	12,13	207,55	**-10,50**	**A**
	6	**Glass type**	**86,91%**	*0,00E+00*	-150,50	-6,93	-1,02	-0,03	0,21	2,41	37,87	969,67	**-22,27**	**A**

Benchmarking using residuals based on floor area normalized energy comsumption *[MJ]*

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2011		**Original**	**41,14%**	*0,00E+00*	-1886	-696	-375	-86	47	720	2494	16983	**-373**	**C**
	7	**Location**	**24,76%**	*2,90E-21*	-838	-505	-298	-48	46	508	2103	12685	**47**	**E**
	9	**Function**	**19,73%**	*1,66E-29*	-1071	-590	-318	-80	28	439	1354	8879	**-402**	**B**
	1	**Wall type**	**38,79%**	*1,07E-193*	-2717	-752	-408	-98	59	764	2687	16989	**-673**	**B**
	4	**Roof type**	**32,78%**	*2,43E-57*	-2537	-834	-450	-105	63	762	2729	12427	**-996**	**A**
	2	**Building shape**	**38,71%**	*2,19E-189*	-1287	-635	-334	-71	44	657	2479	13176	**-287**	**C**
	5	**Building age**	**27,07%**	*1,01E-19*	-875	-589	-297	-63	30	496	1818	17091	**-132**	**C**
	11	**Main heating system**	**14,90%**	*5,11E-05*	-748	-451	-244	-41	37	428	1583	16212	**-189**	**C**
	10	**Main cooling system**	**15,40%**	*4,31E-06*	-896	-465	-245	-37	45	409	1627	16225	**1139**	**F**
	6	**Water heating system**	**25,91%**	*2,43E-33*	-1315	-544	-323	-75	52	596	2505	13276	**-256**	**C**
	3	**Glass type**	**36,28%**	*6,84E-139*	-1784	-731	-396	-77	60	802	2818	16982	**-436**	**B**

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2014		**Original**	**41,14%**	*0,00E+00*	-1886	-696	-375	-86	47	720	2494	16983	**-580**	**B**
	7	**Location**	**24,76%**	*2,90E-21*	-838	-505	-298	-48	46	508	2103	12685	**-160**	**C**
	9	**Function**	**19,73%**	*1,66E-29*	-1071	-590	-318	-80	28	439	1354	8879	**-611**	**A**
	1	**Wall type**	**38,79%**	*1,07E-193*	-2717	-752	-408	-98	59	764	2687	16989	**-883**	**A**
	4	**Roof type**	**32,78%**	*2,43E-57*	-2537	-834	-450	-105	63	762	2729	12427	**-1208**	**A**
	2	**Building shape**	**38,71%**	*2,19E-189*	-1287	-635	-334	-71	44	657	2479	13176	**-495**	**B**
	5	**Building age**	**27,07%**	*1,01E-19*	-875	-589	-297	-63	30	496	1818	17091	**-346**	**B**
	11	**Main heating system**	**14,90%**	*5,11E-05*	-748	-451	-244	-41	37	428	1583	16212	**-406**	**B**
	10	**Main cooling system**	**15,40%**	*4,31E-06*	-896	-465	-245	-37	45	409	1627	16225	**929**	**F**
	6	**Water heating system**	**25,91%**	*2,43E-33*	-1315	-544	-323	-75	52	596	2505	13276	**-467**	**B**
	3	**Glass type**	**36,28%**	*6,84E-139*	-1784	-731	-396	-77	60	802	2818	16982	**-645**	**B**

Benchmarking using residuals based on floor area and degree day normalized energy comsumption *[MJ]*

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2011		**Original**	**43,40%**	*0,00E+00*	-0,6401	-0,2140	-0,1196	-0,0270	0,0167	0,2268	0,9031	11,1490	**0,0375**	**E**
	7	**Location**	**29,42%**	*2,56E-28*	-0,9573	-0,2619	-0,1599	-0,0408	0,0215	0,3213	1,3181	11,1032	**0,2514**	**E**
	9	**Function**	**17,97%**	*2,72E-27*	-0,3203	-0,1992	-0,1118	-0,0287	0,0120	0,1625	0,5180	3,4649	**0,0061**	**D**
	5	**Wall type**	**31,02%**	*1,14E-01*	-0,3933	-0,3284	-0,2115	-0,0329	0,0156	0,1763	0,4758	0,7576	**-0,1787**	**C**
	4	**Roof type**	**33,27%**	*4,50E-60*	-0,8087	-0,2360	-0,1357	-0,0302	0,0235	0,2227	0,8912	5,0375	**0,0259**	**E**
	1	**Building shape**	**42,03%**	*4,93E-214*	-0,3682	-0,1902	-0,1026	-0,0192	0,0189	0,2201	0,8916	5,8345	**0,0619**	**E**
	8	**Building age**	**29,02%**	*5,86E-23*	-0,2646	-0,1740	-0,0891	-0,0156	0,0183	0,1883	0,7485	6,7958	**0,0133**	**E**
	11	**Main heating system**	**17,21%**	*5,94E-07*	-0,2487	-0,1657	-0,0937	-0,0155	0,0255	0,1588	0,6021	5,2240	**0,0352**	**E**
	10	**Main cooling system**	**17,39%**	*1,89E-08*	-0,3089	-0,1652	-0,0931	-0,0186	0,0211	0,1544	0,5856	5,2272	**0,0353**	**E**
	6	**Water heating system**	**30,07%**	*3,09E-42*	-0,6129	-0,1735	-0,1056	-0,0215	0,0139	0,2120	0,8788	4,1254	**0,0138**	**E**
	2	**Glass type**	**38,74%**	*1,06E-153*	-0,7393	-0,2112	-0,1149	-0,0242	0,0184	0,2357	0,9276	6,7564	**0,0282**	**E**

	rank		R^2	*p-value*	0%	5%	20%	45%	55%	80%	95%	100%	Case building	Label
Strukton Son 2014		**Original**	**43,40%**	*0,00E+00*	-0,6401	-0,2140	-0,1196	-0,0270	0,0167	0,2268	0,9031	11,1490	**-0,0228**	**D**
	6	**Location**	**29,42%**	*2,56E-28*	-0,9573	-0,2619	-0,1599	-0,0408	0,0215	0,3213	1,3181	11,1032	**0,1955**	**E**
	9	**Function**	**17,97%**	*2,72E-27*	-0,3203	-0,1992	-0,1118	-0,0287	0,0120	0,1625	0,5180	3,4649	**-0,0530**	**C**
	2	**Wall type**	**39,94%**	*1,30E-203*	-0,9509	-0,2212	-0,1289	-0,0289	0,0196	0,2381	0,9864	11,1432	**-0,0771**	**C**
	4	**Roof type**	**33,27%**	*4,50E-60*	-0,8087	-0,2360	-0,1357	-0,0302	0,0235	0,2227	0,8912	5,0375	**-0,0366**	**C**
	1	**Building shape**	**42,03%**	*4,93E-214*	-0,3682	-0,1902	-0,1026	-0,0192	0,0189	0,2201	0,8916	5,8345	**0,0058**	**D**
	7	**Building age**	**29,02%**	*5,86E-23*	-0,2646	-0,1740	-0,0891	-0,0156	0,0183	0,1883	0,7485	6,7958	**-0,0578**	**C**
	11	**Main heating system**	**17,21%**	*5,94E-07*	-0,2487	-0,1657	-0,0937	-0,0155	0,0255	0,1588	0,6021	5,2240	**-0,0273**	**C**
	10	**Main cooling system**	**17,39%**	*1,89E-08*	-0,3089	-0,1652	-0,0931	-0,0186	0,0211	0,1544	0,5856	5,2272	**-0,0316**	**C**
	5	**Water heating system**	**30,07%**	*3,09E-42*	-0,6129	-0,1735	-0,1056	-0,0215	0,0139	0,2120	0,8788	4,1254	**-0,0380**	**C**
	3	**Glass type**	**38,74%**	*1,06E-153*	-0,7393	-0,2112	-0,1149	-0,0242	0,0184	0,2357	0,9276	6,7564	**-0,0358**	**C**

Table 8: Detailed benchmarking results for the 2 years analyzed for the case building.

		Strukton Son 2011			Strukton Son 2014		
		Predicted	**Observed**	**Residual**	**Predicted**	**Observed**	**Residual**
		[TJ]	*[TJ]*	*[TJ]*	*[TJ]*	*[TJ]*	*[TJ]*
Total consumption	**Original**	23,515	5,579	-17,936	23,610	3,978	-19,633
	Location	16,798	5,579	-11,218	16,872	3,978	-12,895
	Function	33,493	5,579	-27,913	33,612	3,978	-29,635
	Wall type	23,176	5,579	-17,597	23,274	3,978	-19,297
	Roof type	41,478	5,579	-35,899	41,593	3,978	-37,615
	Building shape	20,756	5,579	-15,176	20,824	3,978	-16,846
	Building age	7,571	5,579	-1,992	7,618	3,978	-3,640
	Main heating system	11,197	5,579	-5,617	11,237	3,978	-7,259
	Main cooling system	-0,229	5,579	5,808	-0,230	3,978	4,208
	Water heating system	14,411	5,579	-8,832	14,474	3,978	-10,497
	Glass type	26,121	5,579	-20,542	26,246	3,978	-22,268

		Predicted	**Observed**	**Residual**	**Predicted**	**Observed**	**Residual**
		[MJ]	*[MJ]*	*[MJ]*	*[MJ]*	*[MJ]*	*[MJ]*
EUI	**Original**	1112	739	-373	1107	527	-580
	Location	692	739	47	687	527	-160
	Function	1141	739	-402	1138	527	-611
	Wall type	1412	739	-673	1410	527	-883
	Roof type	1735	739	-996	1735	527	-1208
	Building shape	1026	739	-287	1022	527	-495
	Building age	871	739	-132	873	527	-346
	Main heating system	928	739	-189	932	527	-406
	Main cooling system	-400	739	1139	-402	527	929
	Water heating system	995	739	-256	994	527	-467
	Glass type	1175	739	-436	1171	527	-645

		Predicted	**Observed**	**Residual**	**Predicted**	**Observed**	**Residual**
		[MJ]	*[MJ]*	*[MJ]*	*[MJ]*	*[MJ]*	*[MJ]*
DEUI	**Original**	0,232	0,277	0,045	0,232	0,209	-0,023
	Location	0,014	0,277	0,263	0,014	0,209	0,196
	Function	0,262	0,277	0,015	0,262	0,209	-0,053
	Wall type	0,287	0,277	-0,009	0,287	0,209	-0,077
	Roof type	0,246	0,277	0,031	0,246	0,209	-0,037
	Building shape	0,204	0,277	0,073	0,204	0,209	0,006
	Building age	0,267	0,277	0,010	0,267	0,209	-0,058
	Main heating system	0,237	0,277	0,040	0,237	0,209	-0,027
	Main cooling system	0,242	0,277	0,036	0,241	0,209	-0,032
	Water heating system	0,248	0,277	0,030	0,247	0,209	-0,038
	Glass type	0,245	0,277	0,032	0,245	0,209	-0,036

Table 9: Residual calculations to assess energy performance for the 2 years analyzed for the case building.