2016

From Energy Signature To Cluster Analysis: An Integrated Approach

Lorenza Pistore
Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy, lorenza.pistore@natec.unibz.it

Giovanni Pernigotto
Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy, giovanni.pernigotto@unibz.it

Francesca Cappelletti
University Iuav of Venice, Italy, francesca.cappelletti@iuav.it

Piercarlo Romagnoni
University Iuav of Venice, Italy, piercarlo.romagnoni@iuav.it

Andrea Gasparella
Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy, andrea.gasparella@unibz.it

From energy signature to cluster analysis: an integrated approach

Lorenza PISTORE[1]*, Giovanni PERNIGOTTO[2], Francesca CAPPELLETTI[3], Piercarlo ROMAGNONI[4], Andrea GASPARELLA[5]

[1]Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy
Phone +39 0471 017700, Fax +39 0471 017009, lorenza.pistore@natec.unibz.it

[2]Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy
Phone +39 0471 017632, Fax +39 0471 017009, giovanni.pernigotto@unibz.it

[3]University IUAV of Venezia, Dept. of Design and Planning in Complex Environments, Venezia, Italy
Phone +39 041 2571295, Fax +39 041 257 1282, francesca.cappelletti@iuav.it

[4]University IUAV of Venezia, Dept. of Design and Planning in Complex Environments, Venezia, Italy
Phone +39 041 2571293, Fax +39 041 257 1282, piercarlo.romagnoni@iuav.it

[5]Free University of Bozen-Bolzano, Faculty of Science and Technology, Bolzano, Italy
Phone +39 0471017200, Fax +39 0471 017009, andrea.gasparella@unibz.it

* Corresponding Author

ABSTRACT

Energy audits of existing buildings are especially important for public buildings and, in particular, for schools, where a more efficient use of energy implies unquestionable benefits to public budgets. Schools audit can drive the public administrator to better address retrofit investments, facilitating the choices of energy efficiency measures in the renovation or operation phases. However, the energy audit of existing buildings can be onerous when the number of buildings is large and requires extensive monitoring campaigns, field surveys and energy performance calculation. A simplified method for building energy diagnosis is the "Energy Signature" (*ES*) method described in the annex B of standard EN 15603:2008. According to this approach, heating and cooling energy uses of a given building are correlated to climatic data over a suitable period. Plotting for several time periods the average heating or cooling power versus the average external temperature provides useful information on building energy performance and allows a fast detection of malfunctions or changes in the building operation/features, as well as the verification of the efficacy of any retrofit intervention. Although this method is preferably adopted in the case of constant internal temperature (e.g., fixed temperature setpoint) and when the external temperature is the most influential parameter (e.g., for buildings with stable and relatively low internal and passive solar gains), it can be applied also recording energy use for heating or cooling and accumulated temperature difference between indoor and outdoor, at average regular intervals (e.g., one hour or, for manual monitoring, a week). The *ES* is the best fitting linear regression between these two quantities and, consequently, can be characterized by means of intercept and slope.
In this paper, the building energy signature parameters have been used to analyze a large set of school buildings and to define the characteristics most influential on the energy needs. In particular, the weekly energy consumptions for heating of a set of 42 school buildings located in the province of Treviso, North East of Italy, have been considered. A cluster analysis based on multiple regression has then been used to identify the buildings' subsets homogeneous as for the features affecting the signature parameters.

1. INTRODUCTION

The Energy Efficiency Directive 2012/27/EU (European Parliament, 2012) has accelerated the spread of energy efficiency strategies by identifying existing buildings as a strategic sector for the reduction of energy consumptions. In particular, the public sector has been set within the European energy policy as the leading one for the promotion of energy efficiency measures. Public authorities represent the first example for the citizens, giving guidance and good practices to highlight the importance of energy efficiency, changing the behavior and individual choices. According to the directive, in fact, every year starting from 2014, central Government buildings are required to be renovated to meet at least the national minimum energy performance requirements set in application of the Energy Performance of Buildings Directive (European Parliament, 2010). A 3% annual renovation is the target currently prescribed. Achieving this objective requires the management of retrofitting for the whole building stock that means developing strategies suitable for the different building typologies (Geyer *et al.*, 2016). Excluding the climatic conditions, buildings' energy demand, in fact, is strictly related to the geometry shape, the envelope thermal quality, the system efficiency and building management (i.e., opening hours and windows, shading, lighting and system operation). Energy audit of existing buildings can be onerous when the number of buildings is large and it requires extensive monitoring campaigns, field surveys and energy performance calculation. For these reasons, one of the priorities, when a big stock has to be analyzed, is the classification starting from a few kind of data. In the literature, three approaches for energy audit can be distinguished: a case by case approach based on detailed on-site monitoring, the energy signature approach applied to one or more buildings and a clustering method when the number of buildings is high and there is the need of grouping buildings with similar features. In order to compare different retrofit scenarios for a primary school in Moita, Portugal, Brás *et al.* (2015) performed a standard energy audit with data collection by on-site monitoring and interviews. Dall'O and Sarto (2013) analyzed 49 Italian schools to develop three sets of retrofit measures (i.e., standard, cost-effective and high performance) to apply to the whole sample of buildings. As preliminary step, an energy audit campaign was carried out and actual energy consumption for space heating, occupants' behavior and technical characteristics of the buildings were collected. Marinosci *et al.* (2015) used the energy signature as simplified energy audit in the refurbishment of the historical building of the School of Engineering and Architecture of Bologna, Italy. As a whole, energy signature is one of the most utilized tool to investigate and assess controls and management in buildings (Lindelöf *et al.*, 2015; Belussi *et al.*, 2015; Belussi and Danza, 2012; Hitchina and Knightba, 2016), but also to detect building thermal performance information (Danov *et al.*, 2013; Nordström *et al.*, 2012). However, when the sample of buildings is large and a few representative cases have to be found, cluster analysis is the tool mostly applied. For example, Santamouris *et al.* (2007) used clustering techniques to define energy classes based on heating energy consumption of a large sample of schools in Greece. Gaitani *et al.* (2010) used principal component and cluster analysis to group school buildings with similar characteristics and to find the typical school for each level of energy class. Arambula *et al.* (2015) applied cluster analysis in order to group schools with similar characteristics and find representative architectural types and a small number of parameters for an effective description of the energy consumption for heating and hot water production.

In this paper, energy signature and cluster analysis approaches have been combined in order to group a set of 42 school buildings located in the North of Italy. Schools energy signature parameters, i.e., the slope and the zero of linear regression function, have been used as dependent variables to group schools in clusters. The cluster analysis has been used to identify the buildings' homogeneous subsets as for the features affecting the signature parameters.

2. METHODS

2.1 Energy Signatures Implementation

A building energy audit can be performed by the "Energy signature" method described in the annex B of the EN 15603:2008 (CEN, 2008). According to this approach, heating and cooling energy uses of a given building are correlated to climatic data over a suitable period. Plotting for several time periods the average heating or cooling power versus the average external temperature provides useful information on the building energy performance and allows a fast detection of malfunctions or changes in the building operation/features. Generally, the indoor air temperature is considered constant, and assumed to be equal to the setpoint and, for this reason, the external air temperature results the most influential parameter. Energy signature is preferably applied for buildings with stable internal gains and relatively low passive solar gains. Its application requires that energy uses for heating or cooling, as well as average external temperatures or, when possible, accumulated temperature differences are recorded or obtained at regular intervals. These intervals can be as small as one hour, but a week is often used, since this time discretization is long enough to neglect non-linear short-term behaviors due to the building thermal inertia. For this reason, in this work, weekly intervals have been adopted. The indoor air temperature has been fixed at 20 °C during the occupancy time. The average weekly power per unit volume, obtained by dividing the energy use during one week per unit volume by the amount of opening hours per week (Equation 1), has been plotted versus the weekly heating degree-hours during the opening hours (Equation 2).

Energy signatures are characterized by two main parameters: the slope of the regression function and the intersection with the x-axis, hereafter called *zero of the function*. The slope represents the energy performance of the building: the steeper the slope, the larger is the heating power needed. The zero of the function represents the minimum number of $HDH_{20,occ}$ for which the system has to be turned on: the higher is the intercept, the better is the passive building thermal performance.

$$\Phi = \frac{\sum_{i=1}^{7} EP_{h_i} / V}{\tau} \tag{1}$$

$$HDH_{20,occ} = \sum_{i=1}^{n} (20 - T_{ext})_i \quad with \quad n = opening\ hours of\ a\ week \tag{2}$$

2.2 Multiple Regression and Clustering

According to the methodology already developed and described in Arambula *et al.* (2015), the first step regards the selection of the most correlated parameters to perform the clustering. The slope and the zero of the function of each building can be correlated to some of them, such as those describing the geometry and the thermal properties of the envelope and the heating system capacity. The influence of each quantity on the slope and the zero is different and the highly correlated parameters and variables can be used to characterize effectively the sample of buildings and to develop the clustering. A list of 12 candidate descriptive quantities has been used: the area of the vertical walls exposed to the external environment, of roof, of floor and of floor in thermal contact with the ground, respectively, A_{vw}, A_r, A_f and $A_{f\text{-}g}$ [m^2]; the total area of opaque and transparent envelope, i.e., $A_{env,o}$ and $A_{env,gl}$ [m^2]; the ratios between windows' area and vertical walls' area A_{win}/A_{vw}, between windows and total floor areas A_{win}/A_f and between transparent and opaque envelope $A_{env,gl}/A_{env,op}$; the average thermal transmittance of the envelope U [W m^{-2} K^{-1}]; the shape factor of the school, expressed in terms of ratio S/V [m^{-1}] between the dissipating surface S and the conditioned volume V and the capacity of the heating system H [kW].

A multiple linear regression has been adopted to find the sets of the candidate quantities which better define homogenous groups, to perform the following clustering. Indeed, the selected quantities can be employed both to identify the groups and to develop linear predictive models for their elements. One regression has been performed for slope and another one for the zero of the function. For each one of the 12 descriptive quantities, the highest value in the whole dataset has been identified and used to normalize the characteristics of each building. The predictors have been grouped in 4083 possible combinations starting from groups with 2 to groups with 12 predictors. For each regression, the adjusted index of determination R^2_{adj} has been calculated and monitored, as well as F-tests and the p-values to check the model's statistical significance and variance inflation factors VIF for the analysis of multi-collinearity issues. Only models with significant p-value with respect to a significance level of 10 % and, preferably, without multi-collinearity issues (i.e., $VIF < 10$) have been considered for the definition of the quantities for the clustering. The combinations of predictors with the highest R^2_{adj} have been selected as set of coordinates to define the "position" of each element in the sample of schools. The next two steps involve the clustering and its validation. K-means approach is one of the most popular techniques in clustering and data mining and it is based on a simple partitional algorithm that tries to find K non-overlapping clusters (Lloyd, 1957; Wu, 2012). By this method, once defined the desired number of clusters K, an equivalent number of centroids is selected and data points are assigned to the closest centroid according to the squared Euclidean distances calculated from the closer centroid. After the definition of the clusters, it is possible to validate them by checking if the combination of predictors with the highest R^2_{adj} with respect to the whole dataset is the best for the cluster as well. If it is not, the combination of predictors with the highest R^2_{adj} is found and used as new coordinate system. If the cluster has an improved but still low R^2_{adj}, and if enough buildings are present to consider sub-clustering (i.e., with at least 25 elements), it is possible to run the algorithm again, but using, this time, the set of parameters with the highest R^2_{adj} for the cluster to split.

Since the whole dataset for the clustering includes 42 elements, and since the objective is to find a matrix of clusters (a slope-zero matrix), we imposed $K^I = K^{II} = 2$ for both clustering and sub-clustering. As it is commonly done in K-means approaches, the initial virtual centroids are randomly generated within the domain of the dataset. After the creation of the initial clusters, the centroids $C_{1,k}$ are calculated and the K-means approach is iterated. The heuristic procedure continues until the determination of the i^{th} combination of centroids $C_{i,k}$ minimizing the global squared Euclidean distances. Since this method is sensitive to the choice of the initial centroids $C_{0,k}$, each clustering has been repeated several times with different random centroid and the solution giving the best improvement with respect to the statistical models has been selected. Once all possible clusters have been defined, the results have been analyzed and the models optimized. The number of predictors, their combination and new regression models have been re-calculated with the elements of each of the cluster in order to find possible improvements to the adjusted index of determination. Also in this case, *F*-values, *p*-values and *VIF* have been analyzed to determine if the models can only describe the data in the clusters or be used also for further extrapolations and predictions. As a final step, the schools closest to the centroids are determined.

3. RESULTS

3.1 Energy Signatures

As specified in the method, the first step has been to elaborate the schools' energy signatures. The results of this approach are shown in Figure 1 where the regression functions of each school energy signature have been plot on a two-axis graph. The slope of the curves represents the normalized heating power per heating degree hour and as it can be seen in Figure 1, almost all the functions have a slope included between 0.005 and 0.075 W m^{-3} K^{-1} h^{-1}. Only school TV032_01 presents an abnormal slope of 0.286, thus revealing a great uncertainty in the reliability of the available data. Concerning the zero of the function, almost all of the curves presents a positive value included between 21.1 and 277 K h, thus revealing a great variability in the schools' thermal inertia towards the outdoor climatic variation. An extreme case is represented by school CN028_09, with an almost null zero of the function value.

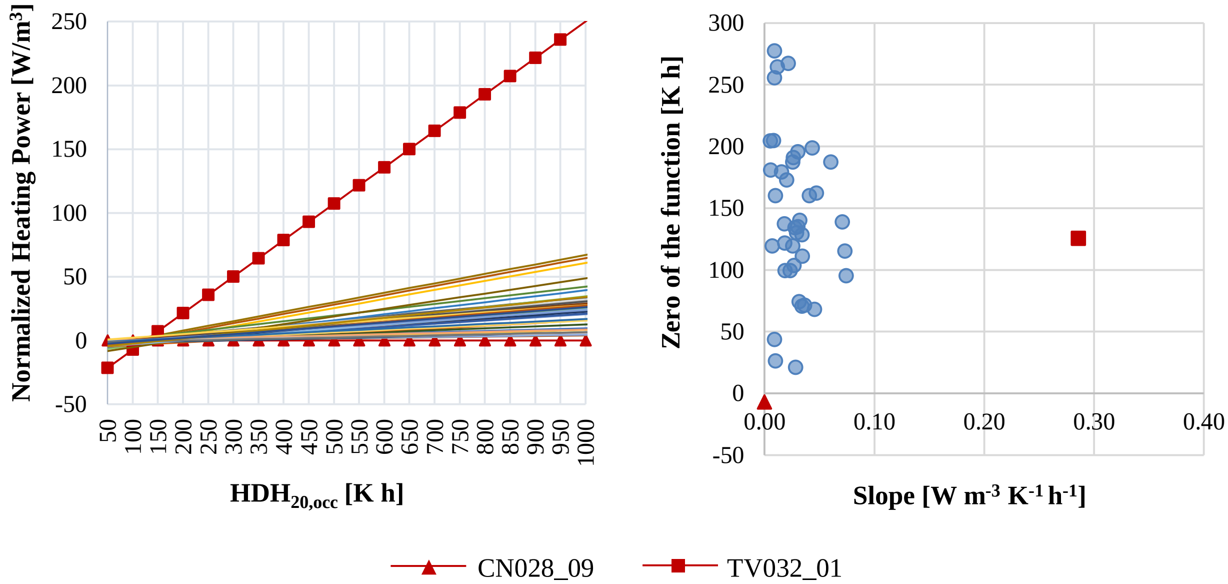

Figure 1: Schools energy signatures regression functions (left) and zero and slope of functions (right)

3.2 Multiple Linear Regression and Clustering

The energy signatures of the set of schools have been used to implement a multiple linear regression taking into account the 12 predictors used to describe the entire stock of buildings as independent variables and the zero and the slopes in turn as dependent variables of a linear function. The linear models have been elaborated starting from the smallest groups (i.e., 2 predictors) and evaluated in terms of R^2_{adj}, *F*-value and *p*-value, taking into account the top-ten configurations output for each set of variables. Results show that the most descriptive configurations have been obtained with a maximum of 6 variables, since the R^2_{adj} tend to increase only from 2 to 6 predictors.

The top-ten configurations among the 924 combinations obtained with 6 predictors, each one identified by an ID number, have been selected for both zero and slopes of the functions (Tables 1 and 2). As it can be seen, the statistical indexes are similar for all of them.

As regards the involved predictors, it can be noticed that floor area, opaque and transparent envelope area and ratio between windows and floor area are the most common variables for the zero of the function's top-10 combinations, whereas for the slope's combinations the most recurring ones are the ratio between vertical walls and windows, the average between the opaque and the transparent envelope, the average thermal transmittance of the envelope, and the compactness ratio *S/V*. In addition to these common predictors, some others descriptive variables differentiate each model.

All the top ten combinations have been used to perform the first clustering and the results are shown in Table 3 and 4. In order to choose the best ID to perform the further steps, a preliminary analysis has been managed according to the statistics and to the number of elements included in each resulted cluster. Looking at their R^2_{adj}, *F*-value, *p*-value and their numeric quantity, the configurations chosen for the further steps are ID805 and ID825 respectively for the zero of the function and the slope.

3.2.1 Zero of the function clustering. The results for the zero of the function clustering are shown in Table 5. ID805 have been grouped in two clusters (i.e., CL_{z1} and CL_{z2}) of 15 and 27 elements. The multiple regression of each cluster has been performed in order to find the best descriptors combination. After this regression, ID819 is the best combination for CL_{z1}, having a R^2_{adj} equal to 0.85, and ID755 the best for CL_{z2}, even with a still very low R^2_{adj}. So, a second clustering has been implemented for CL_{z2}. Combination ID755 has been used and $CL_{z2.1}$ and $CL_{z2.2}$, respectively made of 14 and 13 elements, have been formed. The results have been further optimized with a second regression, performed by ID427 for $CL_{z2.1}$ and ID148 for $CL_{z2.2}$. In this way, three final clusters have been obtained: CL_{z1} with a R^2_{adj} of 0.85, and $CL_{z2.1}$ and $CL_{z2.2}$ with a R^2_{adj} respectively of 0.82 and 0.94, with a considerable improvement of the initial R^2_{adj} values.

Table 1: Zero of the function. Top-10 combinations selected for the first clustering and their descriptors.

Zero of the function										
ID	**805**	**286**	**861**	**346**	**356**	**614**	**871**	**851**	**872**	**362**
Predictors										
A_{vw}		**x**		**x**	**x**					**X**
A_r						**x**				
A_f	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**X**
$A_{f\text{-}g}$	**x**	**x**								
$A_{env,op}$			**x**	**x**	**x**	**x**	**x**	**x**	**x**	**X**
$A_{env,gl}$	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**X**
A_{win}/A_{vw}	**x**		**x**			**x**	**x**	**x**	**x**	**X**
A_{win}/A_f	**x**	**x**	**x**		**x**	**x**	**x**	**x**	**x**	**X**
$A_{env,gl}/A_{env,op}$									**x**	
U				**x**				**x**		
S/V	**x**	**x**	**x**	**x**	**x**					
H							**x**			
$\boldsymbol{R^2_{adj}}$	0.16	0.15	0.15	0.15	0.15	0.14	0.14	0.14	0.14	0.14
F value	2.27	2.24	2.23	2.17	2.16	2.16	2.15	2.15	2.11	2.11
p-value	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.08	0.08

Table 2: Slopes. Top-10 combinations selected for the first clustering and their descriptors.

Slopes										
ID	**902**	**557**	**662**	**761**	**661**	**822**	**627**	**881**	**825**	**762**
Predictors										
A_{vw}										
A_r		**x**	**x**		**x**		**x**			
A_f										
$A_{f\text{-}g}$		**x**		**x**		**x**			**x**	**X**
$A_{env,op}$				**x**			**x**	**x**		**X**
$A_{env,gl}$	**x**		**x**		**x**					
A_{win}/A_{vw}	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**X**
A_{win}/A_f				**x**		**x**				**X**
$A_{env,gl}/A_{env,op}$	**x**	**x**	**x**			**x**	**x**	**x**	**x**	**X**
U	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**x**	**X**
S/V	**x**	**x**	**x**		**x**	**x**	**x**	**x**	**x**	
H	**x**			**x**	**x**			**x**	**x**	
$\boldsymbol{R^2_{adj}}$	0.17	0.16	0.16	0.15	0.15	0.15	0.15	0.15	0.15	0.15
F value	2.38	2.34	2.28	2.23	2.22	2.21	2.20	2.20	2.19	2.18
p-value	0.05	0.05	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07

3.2.2 Slope clustering. The same process has been implemented with the slope combinations (Table 6). For the first clustering, combination ID825 has been used and grouped in 2 clusters (i.e., CL_{s1} and CL_{s2}) of 18 and 14 elements. After the regression for the selection of better predictors, ID442 and ID818 are found optimizing, respectively, CL_{s1} with R^2_{adj} equal to 0.41 and CL_{s2} with a R^2_{adj} of 0.73. While the latter can be considered satisfying, the first cluster has been divided again into 2 clusters, using ID442 which is the best ID that has optimized the results. $CL_{s1.1}$ and $CL_{s1.2}$ are composed of respectively 14 and 13 buildings. The results have been further optimized with a second regression, performed with ID222 for $CL_{s1.1}$ and ID769 for $CL_{s1.2}$. In this way, three final clusters have been obtained: CL_{s2} with a R^2_{adj} of 0.73, and $CL_{s1.1}$ and $CL_{s1.2}$ with R^2_{adj} respectively of 0.60 and 0.99, with a good increasing of the CL_{s1} and CL_{s2} R^2_{adj} values output from the first clustering. In Table 7, the 6 predictors of the best IDs configurations can be observed.

Table 3: Results of the 1st clustering for each ID configuration (zero of the function as dependent variable).

Zero of the function											
ID		**805**	**286**	**861**	**346**	**356**	**614**	**871**	**851**	**872**	**362**
CL_{z1}	R^2_{adj}	0.516	0.322	0.121	0.558	0.100	0.225	-0.282	-0.011	-0.006	0.004
	F value	3.486	1.870	1.734	3.107	1.558	1.339	0.707	0.941	0.966	1.020
	p-value	0.053	0.254	0.153	0.146	0.203	0.579	0.686	0.483	0.466	0.434
	N	15	12	33	11	31	8	9	34	34	34
CL_{z2}	R^2_{adj}	-0.086	0.102	0.325	0.066	0.286	0.006	0.012	0.597	0.836	0.255
	F value	0.657	1.550	1.641	1.355	1.666	1.034	1.067	2.730	6.937	1.399
	p-value	0.684	0.207	0.426	0.273	0.323	0.425	0.407	0.433	0.283	0.570
	N	27	30	9	31	11	34	33	8	8	8

Table 4: Results of the 1st clustering for each ID configuration (slope as dependent variable).

Slopes											
ID		**902**	**557**	**662**	**761**	**661**	**822**	**627**	**881**	**825**	**762**
CL_{s1}	R^2_{adj}	0.109	0.200	0.091	0.249	1.00	0.020	0.117	0.141	0.226	0.522
	F value	1.388	1.960	1.469	2.772	-	1.078	1.732	1.929	2.315	2.640
	p-value	0.290	0.129	0.234	0.032	-	0.413	0.152	0.111	0.071	0.228
	N	20	24	29	33	1	24	34	35	28	10
CL_{s1}	R^2_{adj}	0.489	0.566	-0.060	0.896	0.148	0.014	0.763	-	0.629	0.149
	F value	4.349	4.701	0.886	12.484	2.155	1.040	4.753	-	4.676	1.906
	p-value	0.010	0.013	0.556	0.076	0.072	0.450	0.337	-	0.031	0.119
	N	22	18	13	9	41	18	8	7	14	32

Table 5: Results of the clustering and optimization basing on the zero of the function as dependent variable.

Zero of the function	**1st clustering**	**1st regression**	**755- subclustering**	**2nd regression**
ID	**805**	**819**	**755**	**427**
	CL_{z1}	**CL_{z1}**	**$CL_{z2.1}$**	**$CL_{z2.1}$**
R^2_{adj}	0.516	**0.852**	0.285	**0.823**
F value	3.486	14.479	1.862	11.101
p-value	0.053	0.001	0.217	0.003
N	15	15	14	14
ID		**755**		**148**
	CL_{z2}	**CL_{z2}**	**$CL_{z2.2}$**	**$CL_{z2.2}$**
R^2_{adj}	-0.086	0.033	0.906	**0.945**
F value	0.657	1.148	20.211	35.493
p-value	0.684	0.372	0.001	0.000
N	27	27	13	13

Table 6: Results of the clustering and optimization basing on the slope as dependent variable.

Slope	**1st clustering**	**1st regression**	**442- subclustering**	**2nd regression**
ID	**825**	**442**	**442**	**222**
	CL_{s1}	**CL_{s1}**	**$CL_{s1.1}$**	**$CL_{s1.1}$**
R^2_{adj}	0.226	0.302	0.563	**0.606**
F value	2.315	2.947	4.865	5.614
p-value	0.071	0.030	0.010	0.006
N	28	28	19	19
ID		**818**		**769**
	CL_{s2}	**CL_{s2}**	**$CL_{s1.2}$**	**$CL_{s1.2}$**
R^2_{adj}	0.629	**0.731**	0.850	**0.997**
F value	4.676	6.882	8.548	389.260
p-value	0.031	0.011	0.108	0.003
N	14	14	9	9

Table 7: The 6 predictors resulted from each best ID configuration.

	Zero of the function			**Slope**		
ID	**819**	**427**	**148**	**818**	**222**	**769**
Predictors						
A_{vw}		**x**	**x**		**x**	
A_r			**x**	**x**	**x**	**x**
A_f	**x**	**x**		**x**		**x**
$A_{f\text{-}g}$	**x**					
$A_{env,op}$				**x**	**x**	**x**
$A_{env,gl}$	**x**	**x**	**x**	**x**	**x**	
A_{win}/A_{vw}		**x**	**x**	**x**	**x**	
A_{win}/A_f	**x**	**x**			**x**	
$A_{env,gl}/A_{env,op}$	**x**			**x**		**x**
U			**x**			**x**
S/V		**x**				
H	**x**		**x**	**x**		**x**

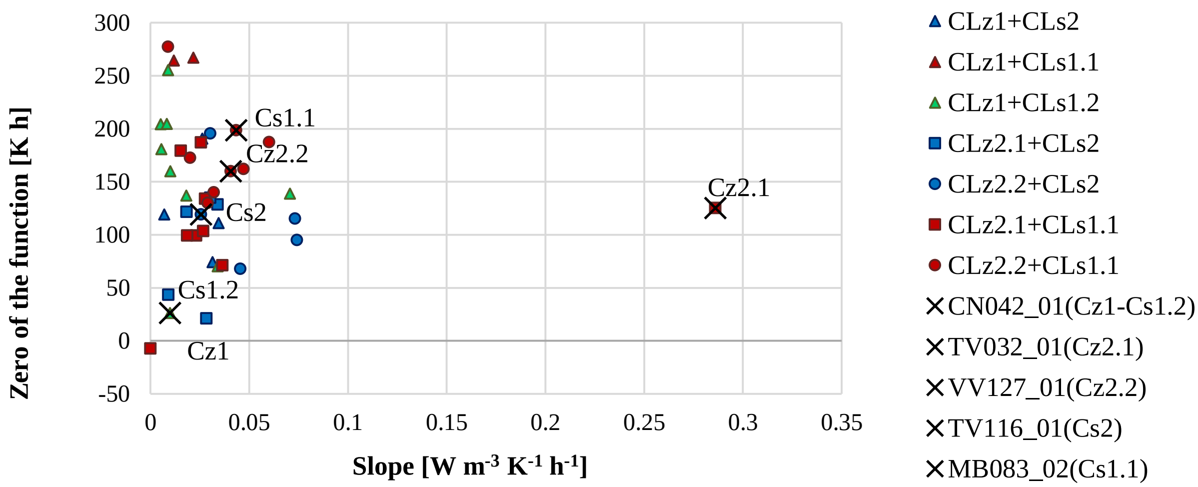

Figure 2: Schools energy signatures slope and zero of function, grouped by clusters

3.2.3 Comprehensive results. In Figure 2, the entire sample of schools grouped by the cluster analysis is represented. Each school position is given according to its belonging to a cluster for both the zero of the function and the slope of its energy signature. The clusters combinations, expressed in the legend, are elaborated using a matrix in which the zero clusters are crossed with the slope ones. The matrix consists of 3 rows and 3 columns, that combined together give a total of 9 possible combinations.

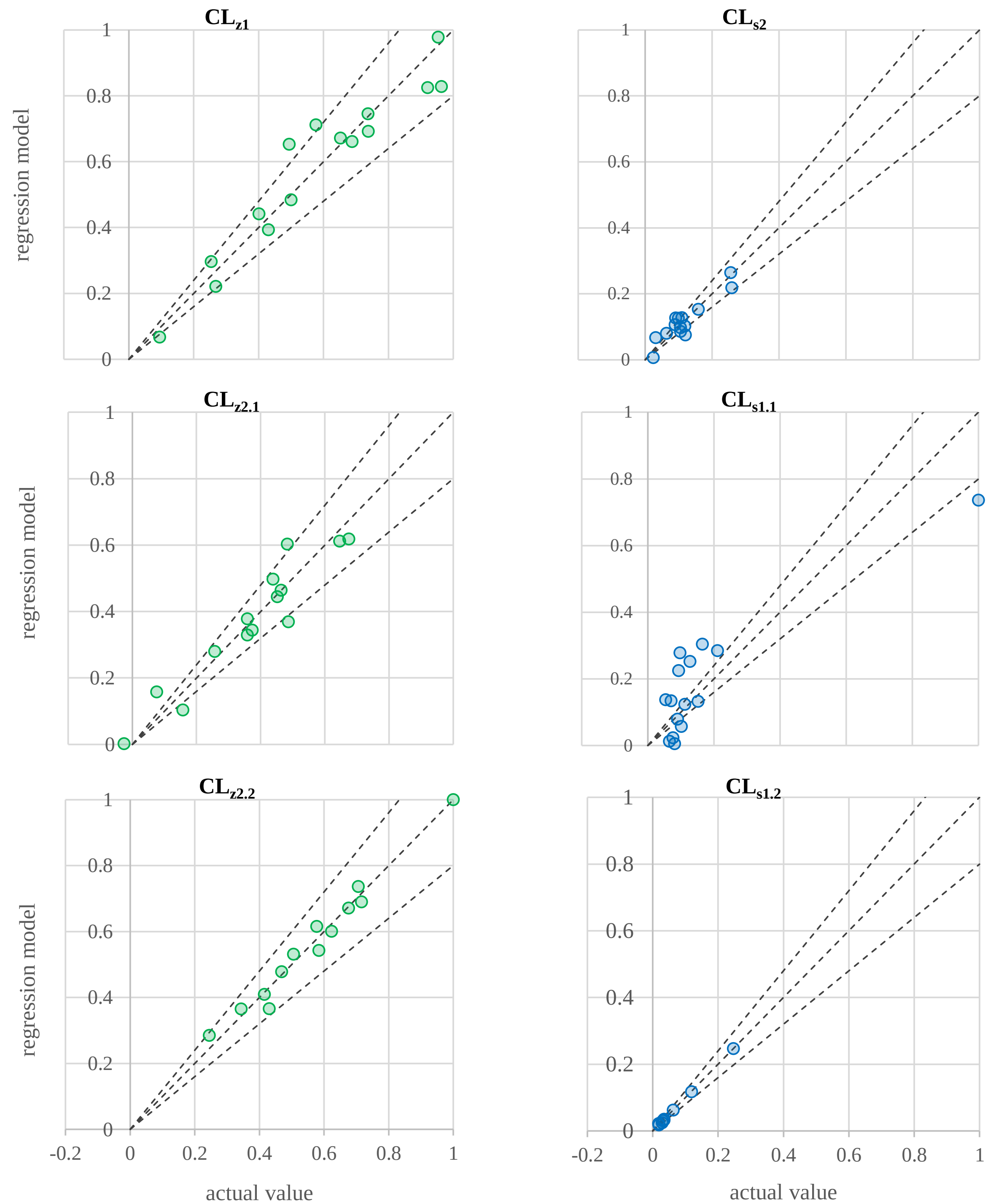

Figure 3: Actual vs. Estimated zero of the function of schools in clusters C_{z1}, $C_{z2.1}$, and $C_{z2.2}$ in green color, and clusters C_{s2}, $CL_{s1.1}$ and $CL_{s1.2}$ in blue color. The dashed lines indicate a deviation of ±20 %.

In the legend, only 7 of these are shown, since no school is included in the $CLs_{1.2}$+$CLz_{2.1}$ and in the $CLs_{1.2}$+$CLz_{2.2}$ cases. Each X indicates the school nearest to the centroid of each cluster, with the exception of school CN042_01 which refers to two clusters, one for the zero and one for the slope.

Finally, Figure3 shows the comparison of the linear regression models to the actual data, for schools included in each cluster, the normalized values of the outputs (i.e., the zero of the function in the left side, the slopes in the right side). In the x-axis there are the measured data, in the y-axis the results of the regression.

As it can be observed, the models fit well for almost all the clusters, since almost all the elements in the stock are within the error band of ±20%. The only exception is $CL_{s1.1}$, in which some points exceed the deviation trend line. This confirms the results of Table 5 and Table 6, where it can be seen that ID222 model regression obtained the lowest value of R^2_{adj}.

4. CONCLUSIONS

In this work the problem of energy auditing a large stock of buildings is discussed. The sample study includes 42 schools in the Province of Treviso, in the North-East of Italy. In order to achieve the purpose, firstly the schools' energy consumptions have been analyzed with the energy signature method, then the K-means cluster analysis has been used for their classification. In addition to this, the multiple linear regression approach has been implemented to validate and optimize the cluster analysis results. The main steps and final goals of the method have been the following ones:

- Starting from relatively few data from a large building stock, using the energy signature method it has been possible to analyze and compare buildings energy performance without onerous and long-term monitoring campaigns.
- Starting from a set of 12 descriptive variables (i.e., predictors), correlated with the normalized zero of the functions and the normalized slopes of the outcome energy signatures, through a first multiple linear regression these predictors have been reduced from 12 to 6.
- After the identification of the best group of predictors, a cluster analysis has been performed and validated through some statistical indices, the adjusted index of determination, *p*-values and *F*-values. This step has been iterated until the clustering output was no more meaningful nor improvable.
- Data in the clusters have been studied and optimized with regressions. For the zero of the function, CL_{z1} has a R^2_{adj} of 0.85, and $CL_{z2.1}$ and $CL_{z2.2}$ with a R^2_{adj} respectively of 0.82 and 0.94, with a considerable improvement from the CL_{z1} and CL_{s2} first R^2_{adj} values. For the slopes, it was obtained a CL_{s2} with a R^2_{adj} of 0.73, $CL_{s1.1}$ and $CL_{s1.2}$ with a R^2_{adj} respectively of 0.60 and 0.99, with also in this case a good increasing of the CL_{z1} and CL_{s2} R^2_{adj} values output from the first clustering.
- For each best ID configuration, it has been possible to select the 6 predictors that describe the set of school included in the clusters, which can be, in additional phase, the areas of intervention to act on.

Further outlooks of this research project are to identify the representative schools for each cluster. Then, taking into account the set of variables most influencing each stock of schools, to compile a list of possible interventions and classify the schools according to a priority of refurbishment. In addition to this, applying a cost-optimal approach, the most convenient solutions can be pointed out. Finally, conducting this approach also in the further years, there is the possibility to check the potentialities of the method described in this paper to highlight the effects of performed energy efficiency measures.

NOMENCLATURE

Symbol		*Subscripts*	
A	area (m²)	*0*	initial
CL	referred to cluster	*20*	referred to an indoor set temperature of 20°C
C	referred to centroid		
EP	Primary Energy	*adj*	adjusted
F	F-test statistic (-)	*env,gl*	referred to transparent envelope
H	capacity of the heating system (kW)	*env,o*	referred to opaque envelope
h	hour (h)	*ext*	external
HDH	heating degree hours (K h)	*f*	referred to floor
K	number of partitions for *K-means* algorithm	*f-g*	floor in thermal contact with the ground

N	number of elements	*h*	referred to heating
R^2	index of determination (-)	*occ*	occupancy time
S	dissipating surface (m²)	*r*	referred to the roof
T	temperature (°C)	*s*	referred to the slope
τ	time (h)	*vw*	vertical walls exposed to the external environment
U	thermal transmittance (W m^{-2} K^{-1})		
V	conditioned volume (m³)	*win*	referred to the windows
VIF	variance inflation factor (-)	*z*	referred to the zero of the function

REFERENCES

Arambula Lara, R. Pernigotto, G., Cappelletti, F., Romagnoni, P., Gasparella, A. (2015). Energy audit of schools by means of cluster analysis, *Energy Build. 95*, 160-171.

Belussi, L., Danza, L., Meroni, I., Salamone, F. (2015). Energy performance assessment with empirical methods: application of energy signature, *Opto−Electronics Review 23*(1), 83–87.

Belussi L. Danza L., (2012). Method for the prediction of malfunctions of consumption analysis: Holistic and multidisciplinary approach of Energybuildings through real energy Signature, *Energy Build. 55*, 715–720.

Brás A., Rocha A., Faustino P., (2105). Integrated approach for school buildings rehabilitation in a Portuguese city andanalysis of suitable third party financing solutions in EU, *Journal of Building Engineering 3*, 79–93.

CEN, (2008). EN 15603:2008 Energy performance of buildings. Overall energy use and definition of energy ratings, Brussels, Belgium.

Dall'O G. and Sarto L., (2013). Potential and limits to improve energy efficiency in space heating in existing school buildings in northern Italy, *Energy and Buildings 67*, 298–308.

Danov S., Carbonell J., Cipriano J., Martí-Herrero J., (2013). Approaches to evaluate building energy data considering dynamic and solar gain effects performance from daily consumption, *Energy Build. 57*, 110–118.

European Parliament, (2012), Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, in Official Journal of the European Union, L315/1.

European Parliament, (2010), Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), in Official Journal of the European Union, L153/13.

N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, P. Patargias, (2010). Using principal component and cluster analysis in the heating evaluation of the school building sector, *Appl. Energy 87*(6), 2079–2086.

Geyer P., Schlüter A., Cisar S., (2016). Application of clustering for the development of retrofit strategies for large building stocks, *Advanced Engineering Informatics*. In press

Hitchina R. and Knightba I., (2016). Daily energy consumption signatures and control charts for air-conditioned buildings, *Energy Build. 112*, 101–109.

Lindelöf D., Afshari H., Alisafaee M., Biswas J., Miroslav J., Mocellin X., Viaene J., (2015). Field tests of an adaptive, model-predictive heating controller for residential buildings. *Energy Build. 99*, 292–302.

Lloyd S.P., 1957, Least squares quantization in PCM. Bell Telephone Labs Memorandum, Murray Hill, NJ. Reprinted in: *IEEE Trans. Information Theory IT-28* (1982), *vol. 2*, 129–137.

Marinosci C., Morini G.L., Semprini G., Garai M., (2015). Preliminary energy audit of the historical building of the School of Engineering and Architecture of Bologna, *Energy Procedia 81*, 64 – 73.

Nordström G., Johnsson H., Lidelöw S., (2012). Using the Energy Signature Method to Estimate the Effective U-Value of Buildings, in Sustainability in Energy and Buildings, *Proceedings of the 4th International Conference in Sustainability in Energy and Buildings (SEB´12), Volume 22 of the series Smart Innovation, Systems and Technologies, Chapter 4*, pp 35-44.

Santamouris M., Mihalakakou G., Patargias P., Gaitani N., Sfakianaki K., Papaglastra M., Pavlou C., Doukas P., Primikiri E., Geros V., Assi-makopoulos M. N., Mitoula R., Zerefos S., (2007). Using intelligent clustering techniques toclassify the energy performance of school buildings, *Energy Build. 39*, 45–51.

Wu Junjie,(2012). Advances in K-means clustering. A data mining thinking, Springer Thesis, Springer.

ACKNOWLEDGEMENT

The authors would like to thank the Province of Treviso (Provincia di Treviso) for making the schools database available for this research.