Engineering Sustainability

Data mining for energy analysis of a large data set of flats
Capozzoli, Serale, Piscitelli and Grassi

ice | proceedings

Proceedings of the Institution of Civil Engineers
Engineering Sustainability
http://dx.doi.org/10.1680/jensu.15.00051
Paper 1500051
Received 30/09/2015 Accepted 31/03/2016
Keywords: energy/statistical analysis/sustainability

ICE Publishing: All rights reserved

ice
Institution of Civil Engineers
publishing

Data mining for energy analysis of a large data set of flats

1 **Alfonso Capozzoli** PhD
Assistant Professor, Technology Energy Building Environment (TEBE) Research Group, Department of Energy, Politecnico di Torino, Turin, Italy (corresponding author: alfonso.capozzoli@polito.it)

2 **Gianluca Serale**
Engineer/PhD Student, Technology Energy Building Environment (TEBE) Research Group, Department of Energy, Politecnico di Torino, Turin, Italy

3 **Marco Savino Piscitelli**
Engineer/Grant Researcher, Technology Energy Building Environment (TEBE) Research Group, Department of Energy, Politecnico di Torino, Turin, Italy

4 **Daniele Grassi**
Engineer/Grant Researcher, Technology Energy Building Environment (TEBE) Research Group, Department of Energy, Politecnico di Torino, Turin, Italy

1

2

3

4

To improve the energy efficiency of a large building stock, authority planners and designers need to identify which buildings consume most energy and why. For this purpose, this paper provides a data mining-based methodology for setting decision-making rules to identify patterns of energy consumption for a large data set of flats and evaluate the potential effects achievable by retrofitting actions. The calculated normalised primary energy demand (E_{PDn}) and the geometrical, thermo-physical and heating system attributes of 92 906 flats are analysed. Firstly, an accurate statistical description of the building stock and its main technological features is provided. Secondly, a supervised classification algorithm to rank flats as 'low', 'medium' or 'high' E_{PDn} is developed based on the flats' attributes. To classify E_{PDn}, reference threshold values are set between the attributes. These values will benefit authority planners and designers when setting performance objectives. Finally, the high-E_{PDn} flats are analysed in depth through an unsupervised classification algorithm. Thus, intrinsic properties and hidden dependencies are discovered. Moreover, a manageable number of real reference flats representative of the entire high-consumption class are identified. These real reference flats can be used to study the causes of high-E_{PDn} and propose different energy retrofit actions.

Notation

DD	degree day
E_{PD}	primary energy demand
E_{PDn}	normalised primary energy demand
E_{PDnDD}	normalised primary energy demand on degree day
rF_i	real reference flat of *i*-cluster
S/V	aspect ratio (ratio of heat transfer surface on heated volume)
U_{env}	average *U*-value of the vertical opaque envelope
U_{w}	average *U*-value of the windows
$z(c)$	*z*-score centroid for an attribute in a specific cluster
$z(x)_n$	*z*-score of the n attribute
η	average global efficiency of the system for space heating and domestic hot water

1. Introduction

In recent years, the application of energy efficiency and sustainable green design measures in new and existing buildings has become a crucial issue for building owners, designers, contractors and facility managers (Kim *et al.*, 2011; Xiao and Fan, 2014). Moreover, the amount of data generated by energy simulations, surveys and building management systems has increased dramatically. In the study of Swan and Cantab (2015), different UK practitioners were interviewed, highlighting the need for standardised and structured analysis methods to extract and transfer knowledge from these huge amounts of data.

In particular, the application of intelligent analysis methods to large data sets would benefit designers and authority planners who need to

- identify the major causes of high energy consumption and suggest rules for incentivising energy retrofit actions (Fracastoro and Serraino, 2011)
- evaluate benchmark values to drive policies for building sustainability design approaches (Capozzoli *et al.*, 2003; Elghali *et al.*, 2008; Parkin *et al.*, 2003)

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

- have a framework of building stocks (Aksoezen *et al*., 2015; Capozzoli *et al*., 2015a) and evaluate a manageable number of reference buildings representative of the entire data set (Filogamo *et al*., 2014)
- provide simple tools for a fast estimation of energy consumption classes (Motawa, 2015).

In the past decade the use of data mining in the building energy sector has increased considerably in different applications (Capozzoli *et al*., 2015c; Fan *et al*., 2015; Khan *et al*., 2013; Kumar, 2011; Yu *et al*., 2013). In this paper, some of these techniques are proposed to analyse a data set of 92 906 energy certificates related to residential flats. The data set contains information on envelope and technical plant features and on primary energy demand (E_{PD}) for space heating and domestic hot water (DHW) for each flat, calculated in 'standard rating' conditions, according to the methodology proposed in EN ISO 13790 (ISO, 2008), UNI TS 11300-1 (UNI, 2008a) and UNI TS 11300-2 (UNI, 2008b).

This paper aims to cover some crucial aspects of practical relevance for both authority planners and building energy experts and designers. Section 2 provides an overview of the applied methodology, and Section 3 offers briefly the theoretical basis for the data mining techniques adopted in the present work. In Section 4 an accurate description of the main attributes and construction typologies of the flats composing the data set was carried out. Section 5 describes the results obtained by a classification process of the data set according to the work carried out in Capozzoli *et al*. (2015b), while Section 6 investigates the intrinsic properties and hidden dependencies of high-consumption flats and proposes specific tailored retrofit actions. In particular, the present work on the basis of a classification process (Capozzoli *et al*., 2015b) does the following.

- It offers authority planners a simple method to set reference threshold values (to respect or to create incentives) for some thermo-physical attributes that drive the classification of energy consumption (Section 5). Moreover, it provides a method to evaluate a manageable number of real reference flats representative of the entire high-consumption class (Section 6).
- It provides building energy experts and designers with a set of decision-making rules, based on a small number of attributes that can drive different patterns of normalised primary energy demand (Section 5). The intrinsic properties and hidden dependencies of the high-consumption flats are identified with the aim of finding specific possible retrofit actions on the basis of the small number of variables available (Section 6).

2. Methodology

Figure 1 highlights the main steps that were carried out in this paper. A pre-processing analysis (data preparation) in the first part of the work was helpful to clean the data set by removing outliers. Afterwards, a data transformation analysis was performed introducing criteria for labelling each building as having a 'high', 'medium' or 'low' normalised primary energy demand (E_{PDn}). The classification and regression tree (CART) algorithm, which consists of a supervised multistage decision-making process to classify the observations in a finite number of classes, was implemented. The output of the model is a flow chart subdividing the observations into homogeneous subsets (Yu *et al*., 2010) according to respect response, represented in the model by categorical variables related to primary space heating and DHW

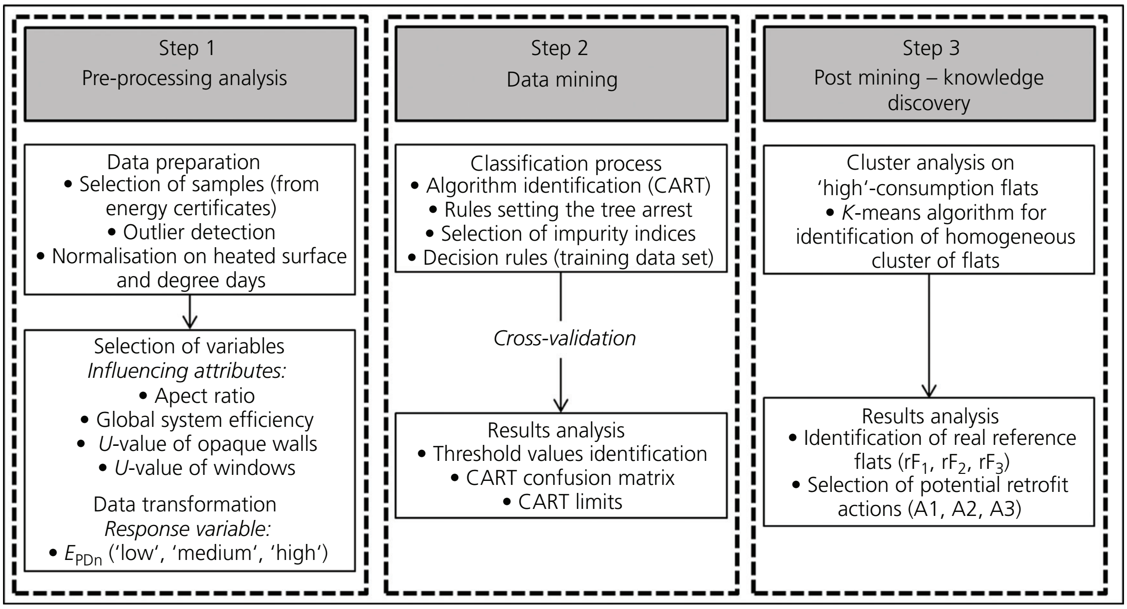

Step 1
Pre-processing analysis
Data preparation
• Selection of samples (from energy certificates)
• Outlier detection
• Normalisation on heated surface and degree days
Selection of variables
Influencing attributes:
• Apect ratio
• Global system efficiency
• U-value of opaque walls
• U-value of windows
Data transformation
Response variable:
• E_{PDn} ('low', 'medium', 'high')
Step 2
Data mining
Classification process
• Algorithm identification (CART)
• Rules setting the tree arrest
• Selection of impurity indices
• Decision rules (training data set)
Cross-validation
Results analysis
• Threshold values identification
• CART confusion matrix
• CART limits
Step 3
Post mining – knowledge discovery
Cluster analysis on 'high'-consumption flats
• K-means algorithm for identification of homogeneous cluster of flats
Results analysis
• Identification of real reference flats (rF_1, rF_2, rF_3)
• Selection of potential retrofit actions (A1, A2, A3)

Figure 1. Framework of the paper

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

energy demand. The classification process made it possible to introduce a set of decision rules capable of outlining the splitting criteria. The outcome of this process consists of useful information that helps to recognise the patterns which drive the evaluation of the energy performance of buildings. Furthermore, a detailed analysis on the high-consumption flats was performed using *K*-means algorithm. This kind of analysis made it possible to group the high-consumption samples into similar clusters and to find for each of them a real reference flat. Some useful information was retrieved regarding the attributes that need to be considered in potential retrofitting measures.

3. Methods

In recent years, the techniques of machine learning, data mining and knowledge discovery in data set were successfully applied for energy saving purposes (Yu *et al*., 2013). In this scope, pattern recognition is a subarea of data mining and consists of the analysis of patterns within the data in order to identify a correct classification. The aim of pattern recognition is to learn classifier data (patterns) based on prior knowledge or statistical information extracted from the pattern. In general, these classification algorithms treat groups of measurements or observations, defining points in an appropriate multidimensional space. In this study, a supervised classification algorithm (CART) was developed. This technique produces only a binary split (considering all $2k - 1$ ways of creating a binary partition of k attribute values) beginning with the root node, which contains the whole learning sample, and splitting each subsequent parent node into two child nodes. The split is an iterative process that splits the data set into subclasses. The best way to divide the record depends on the type of measure chosen. This measure is defined in terms of the record's class distribution before and after splitting. In this work, the Gini index was used as a degree of impurity of each node. The statistical performance of each classification algorithm has to be evaluated in order to apply it into a new data set. The *k*-fold cross-validation is the method used in this paper to evaluate the accuracy of the classification tree.

According to D'Oca and Hong (2015), in a classification process, a minimum confidence of 50% ensures the reliability of each leaf node. In the studies of Gao *et al*. (2010) and Yu *et al*. (2010), the accuracy of the whole classification process is considered acceptable where the uncertainty is lower than 20–30%.

For a further investigation of a determined group of samples, an unsupervised classification algorithm (*K*-means) was performed (Wu, 2012) on the most energy-consuming samples. This is an algorithm that allows for objects with similar characteristics to be grouped together into clusters. In particular, each cluster captures the natural structure of the data. Since the data are located in an *n*-dimensional space, the similarities according to distance-based metrics were evaluated. In this study, the Euclidian distance was used in order to apply the *K*-means algorithm correctly. This process requires as an input parameter the number *k* of partitions. The optimal number of partitions (*k*) was valued using the minimisation of the Davies-Bouldin index as the internal validation method.

4. The adopted data set

4.1 Construction of the data set

The value of E_{PD} was calculated using the standard rating methodology suggested in the aforementioned technical standard and considering energy needs for DHW production and space heating. The DHW energy demand was calculated by considering standard values referring to floor area, while the space heating energy demand was evaluated by considering building energy balance. The modelling of the building geometry considers real shapes and self-shading or overshading of other buildings. The quasi-steady-state calculation method is based on the monthly balance of heat losses (transmission and ventilation) and heat

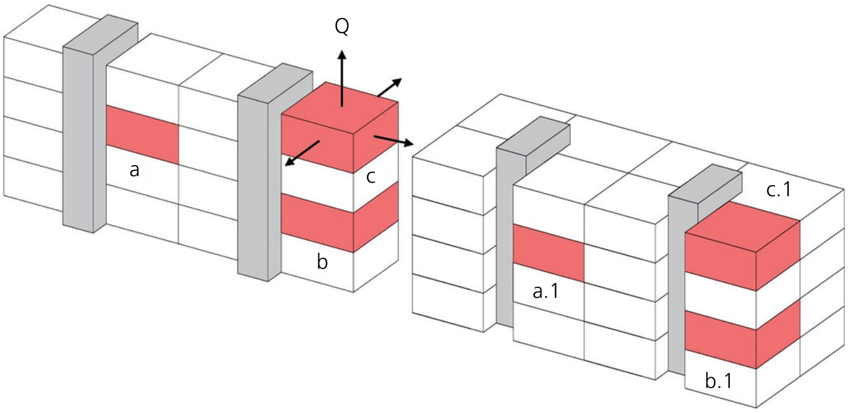
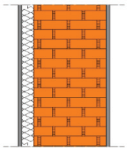


Figure 2. Examples of possible construction typologies with different positions of the flats in multifamily buildings. Flats indicated as 'a' are characterised by lower aspect ratios, while 'c' flats have higher aspect ratios

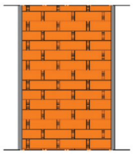
Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

Typology 1: U = 0·338 W/m^2 K, s = 0·615 m



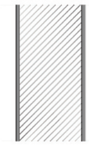
Material	s: m	λ: W/m K
Plaster	0·020	0·800
Polystyrene	0·080	0·041
Brick	0·500	0·676
Plaster	0·015	0·800

Typology 2: U = 0·984 W/m^2 K, s = 0·640 m



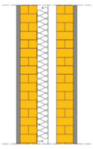
Material	s: m	λ: W/m K
Plaster	0·020	0·800
Brick	0·600	0·800
Plaster	0·020	0·800

Typology 3: U = 2·167 W/m^2 K, s = 0·490 m



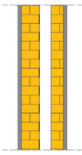
Material	s: m	λ: W/m K
Plaster	0·020	0·800
Stonewall	0·450	2·300
Plaster	0·020	0·800

Typology 4: U = 0·638 W/m^2 K, s = 0·355 m



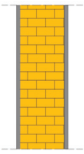
Material	s: m	λ: W/m K
Plaster	0·200	0·800
Hollow brick	0·120	0·387
Polystyrene	0·300	0·059
Air cavity	0·500	0·278
Hollow brick	0·120	0·387
Plaster	0·015	0·800

Typology 5: U = 1·053 W/m^2 K, s = 0·315 m



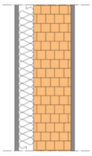
Material	s: m	λ: W/m K
Plaster	0·200	0·800
Hollow brick	0·120	0·387
Air cavity	0·080	0·444
Hollow brick	0·080	0·400
Plaster	0·015	0·800

Typology 6: U = 1·285 W/m^2 K, s = 0·235 m



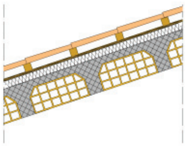
Material	s: m	λ: W/m K
Plaster	0·020	0·800
Hollow brick	0·200	0·387
Plaster	0·015	0·800

Typology 7: U = 0·246 W/m^2 K, s = 0·385 m



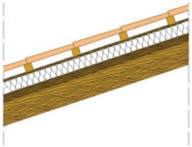
Material	s: m	λ: W/m K
Plaster	0·020	0·800
Polystyrene	0·100	0·041
Alveolar brick	0·250	0·183
Plaster	0·015	0·800

Typology 8: U = 0·313 W/m^2 K, s = 0·380 m



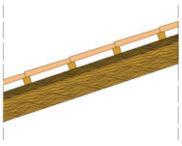
Material	s: m	λ: W/m K
Roof tile	0·030	1·000
Wood deck	0·030	0·120
Polystyrene	0·100	0·041
Concrete	0·200	0·720
Plaster	0·020	0·800

Typology 9: U = 0·901 W/m^2 K, s = 0·095 m



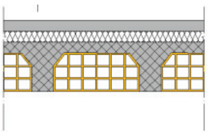
Material	s: m	λ: W/m K
Roof tile	0·030	1·000
Wood deck	0·020	0·180
Fibreglass	0·030	0·043
Wood deck	0·015	0·180

Typology 10: U = 2·019 W/m^2 K, s = 0·065 m



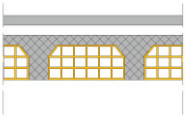
Material	s: m	λ: W/m K
Roof tile	0·030	1·000
Air cavity	0·030	—
Wood deck	0·050	0·180

Typology 11: U = 0·288 W/m^2 K, s = 0·495 m



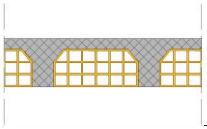
Material	s: m	λ: W/m K
Underlayer	0·005	0·260
Concrete	0·050	1·490
Fibreglass	0·120	0·043
Concrete	0·300	0·720
Plaster	0·020	0·800

Typology 12: U = 1·190 W/m^2 K, s = 0·425 m



Material	s: m	λ: W/m K
Underlayer	0·005	0·260
Concrete	0·050	1·490
Air cavity	0·050	0·313
Concrete	0·300	0·720
Plaster	0·020	0·800

Typology 13: U = 1·546 W/m^2 K, s = 0·325 m



Material	s: m	λ: W/m K
Underlayer	0·005	0·260
Concrete	0·300	0·720
Plaster	0·020	0·800

Materials are listed from outdoor to indoor

Table 1. An overview of possible reference construction technologies in the Piedmont region for vertical and horizontal opaque envelopes

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

gains (solar and internal) evaluated in monthly average conditions. Transmission heat losses were estimated by taking into consideration opaque and transparent surfaces and the thermal bridging effect. In standard rating, parametrical values depending on floor area or heated net volume are taken into consideration when evaluating the ventilation rate and internal heat gains. The dynamic effects on the net space heating energy demand are considered by introducing the dynamic parameters utilisation factor. These parameters depend on the thermal inertia of the building, on the ratio of heat gains to heat losses and on the occupancy/system management schedules. The annual E_{PD} is calculated from the net energy demand through different system efficiencies, which take into account the thermal losses in the various subsystems related to both space heating and DHW. For the heating season, the average global system efficiency represents the ratio between the annual building net energy need and the annual E_{PD} for space heating and DHW.

The standard rating approach could produce results for E_{PD} also far from actual energy requests, because standard assumptions for occupant behaviour, climatic conditions and ventilation are taken into consideration (Summerfield *et al*., 2011). However, since a large data set was analysed in this paper, the potential information that can be extracted in relation to the main patterns driving the E_{PD} can be considered consistent.

4.2 Description of the data set

The samples analysed in the present work was retrieved from a data set of energy certificates until 2014 for several buildings and single habitation units sited in Piedmont region (Northern Italy). In Piedmont, all energy certificates were collected on a Web platform developed by CSI Piemonte (Consorzio per il Sistema Informativo) and are regulated by the authority Piedmont region (Settore Sviluppo Energetico Sostenibile). Designer and energy labellers upload the data directly onto this platform by using a

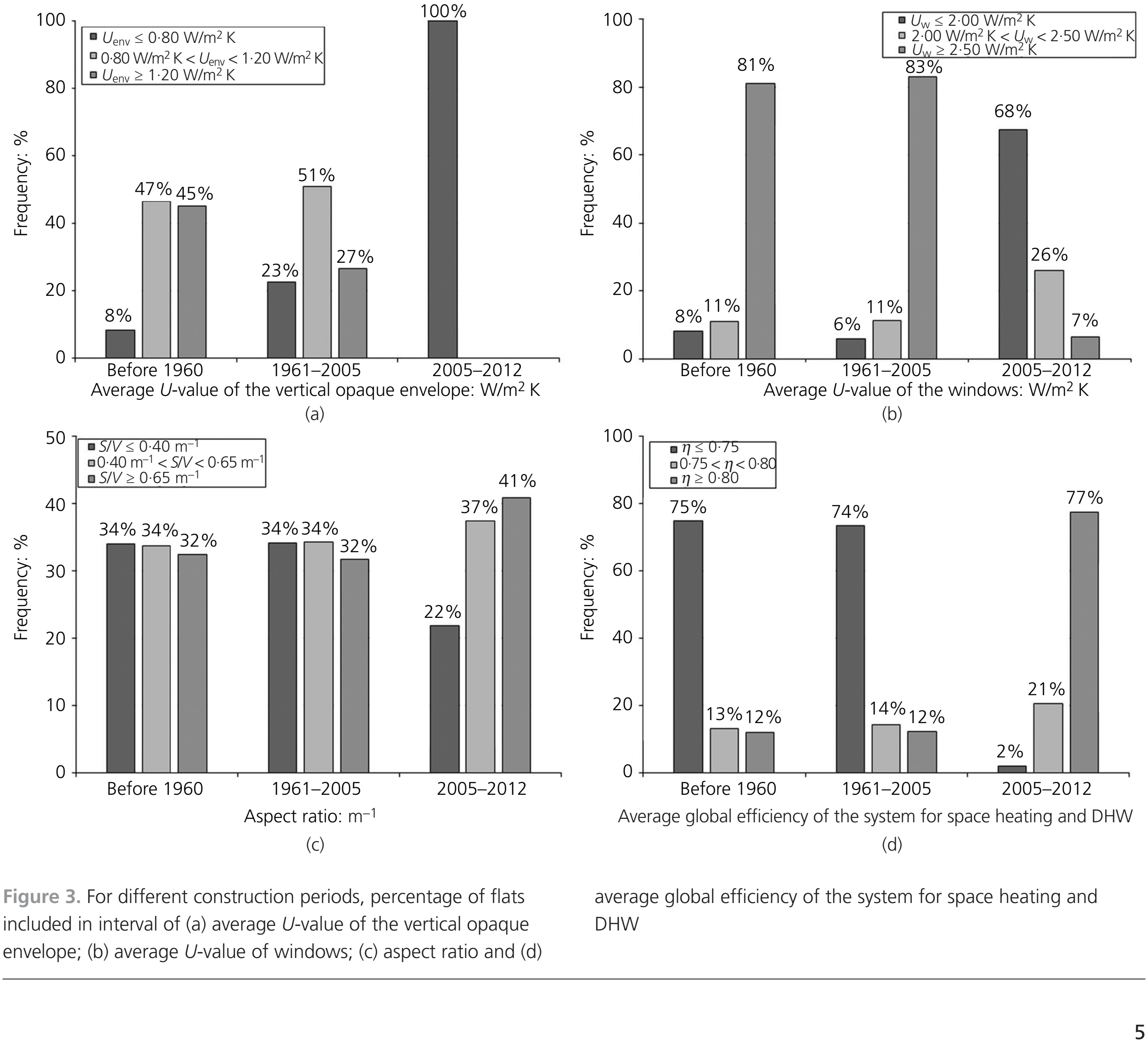


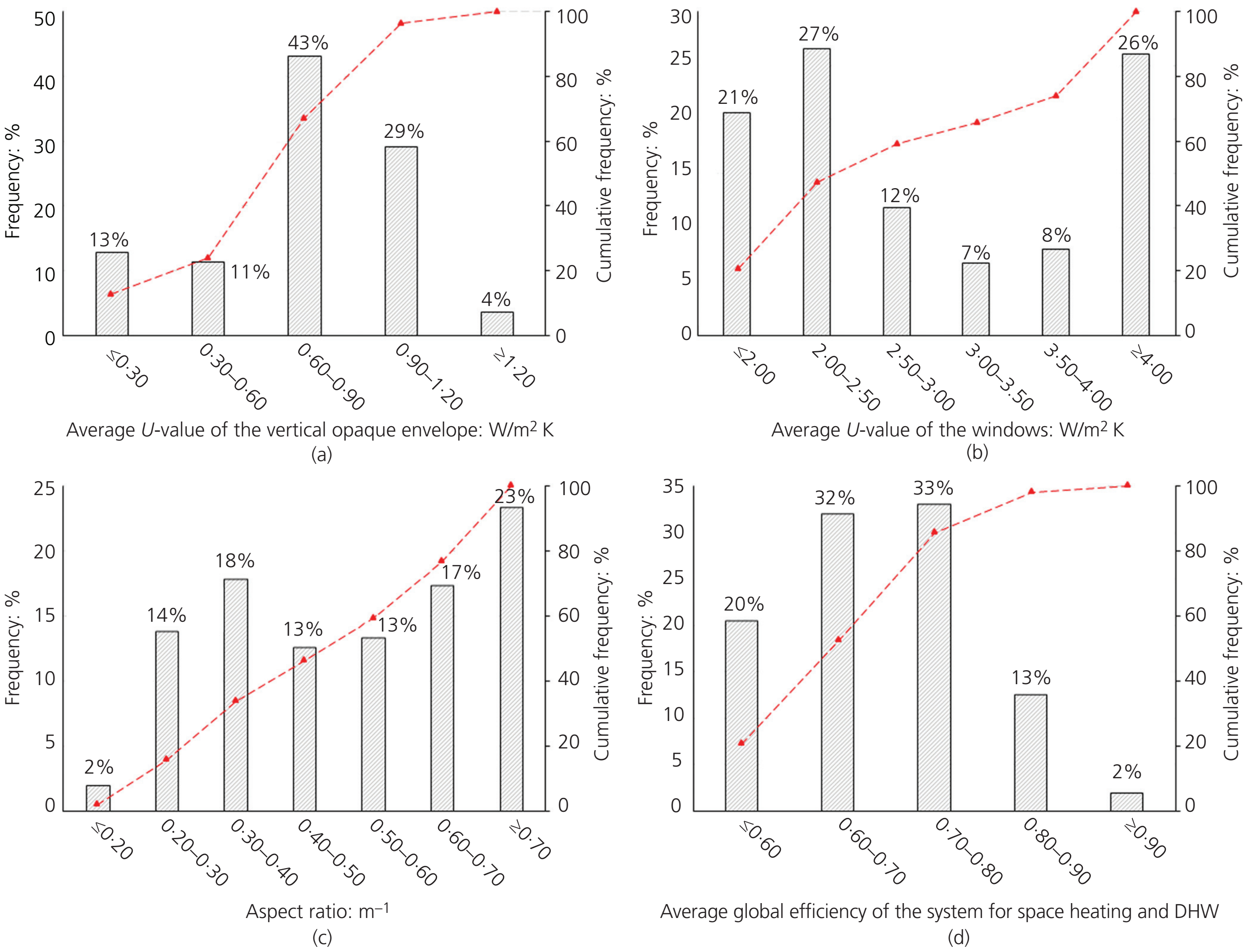
Figure 3. For different construction periods, percentage of flats included in interval of (a) average *U*-value of the vertical opaque envelope; (b) average *U*-value of windows; (c) aspect ratio and (d) average global efficiency of the system for space heating and DHW

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

specific application. In recent years, the EU projects Tabula and Episcope (Ballarini *et al*., 2011, 2014) individuated in this data set a precious source of information to have a framework of the building stock in terms of energy performance.

In this study, homogenous end use and construction typologies were carefully chosen to allow the comparison between the samples. Indeed, among the 269 544 samples classified as ‘residential dwelling with continuous utilisation’, only the 92 906 ‘single flats’ included in multifamily houses and blocks of flats were selected (excluding villas, single houses, co-housing, etc.). The data set collects information related to E_{PD} for space heating and DHW, year of construction and last refurbishment, floor area, heated volume, heat transfer surface, aspect ratio, average *U*-values of the opaque and transparent envelope, subsystem efficiencies of the heating plant (emission, distribution, control and generation subsystem), average global efficiency for space heating and DHW systems, boiler size and Italian energy label (according to Italian legislation updated to 2014).

A frequency distribution analysis of the geometrical features of the samples reveals that 44% of the data set is composed of flats with a floor area ranging between 60 and 90 m^2, 37% ranging between 30 and 60 m^2, 15% ranging between 90 and 120 m^2 and the remaining 4% with other dimensions. Since the data set is very large, the previous analysis could be representative of the typical dimensions of single flats in Italy. Considering the construction periods, three different clusters were highlighted. The first one includes 38% of the data set, and it is composed of flats built before 1960. In general, their thermo-physical characteristics are very poor and an energy refurbishment should be implemented. The second set considers the samples built between 1960 and



(a)

(b)

(c)

(d)

Figure 4. Frequency distribution and cumulative frequency distribution of (a) average *U*-value of the opaque envelope; (b) average *U*-value of windows; (c) aspect ratio and (d) average global efficiency of the system for space heating and DHW

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

2005, while new flats built within the last decade are included in the third cluster construction period. The second and the third subset refer to 58% and 4% of the data set, respectively.

Through a sensitivity analysis, the attributes selected as the most important to consider in the data mining analysis are listed below

- S/V – aspect ratio (ratio of heat transfer surface on heated volume): m^{-1}
- U_{env} – average U-value of the vertical opaque envelope: $W/m^2 K$
- U_w – average U-value of the windows: $W/m^2 K$
- η – average global efficiency of the systems for space heating and DHW.

Figure 2 provides a schematic representation of possible shapes and construction typologies of the single flats in multifamily houses and their relative aspect ratio. The aspect ratio determines how large the surface exposed to the external environment is, and consequently it provides information on the heat gain and loss through the building envelope.

Table 1 shows some possible reference technological constructive typologies that were individuated for the vertical and horizontal opaque envelopes. These technologies were inferred among the ones indicated by the Tabula project (Ballarini *et al.*, 2011, 2014) as the most diffused in the Piedmont region in different periods. Figure 3 illustrates the frequency distribution of these attributes according to the different building construction periods. Figure 4 shows an overall frequency distribution. The average U-value of the vertical opaque envelope and the windows influences heat losses by transmission, while the average global efficiency provides information on the quality of the space heating and DHW system. The technological improvements changing the performance of buildings in the past decade can be deduced from Figures 3 and 4. In particular, the building stock in the last decade has reached U-values that are significantly below 0·80 $W/m^2 K$ and the majority of them present an average global efficiency of the system for space heating and DHW of over 80·0%. Meanwhile, for older buildings the performance is poor.

4.3 Pre-processing analysis

A preliminary analysis was conducted first normalising the primary energy demand on the floor area of each flat. In this way a normalised primary energy demand (E_{PDn}) was obtained for each flat. The average E_{PDn} of the data set is 214·22 kWh/m^2, while the median value is 205·54 kWh/m^2. Figure 5 reports the frequency distribution of E_{PDn}.

A data transformation analysis was performed introducing criteria for labelling each flat as ‘high’, ‘medium’ or ‘low’ E_{PDn}. This data transformation is necessary for the construction of the classification tree, which is based on a categorical response variable. The selection of threshold values between consumption classes must be accurate to obtain reliable information from the data set (Fracastoro and Serraino, 2011). In the Piedmont region, residential flats with an energy demand lower than 82·00 kWh/m^2 are considered low-consumption buildings (energy class labels

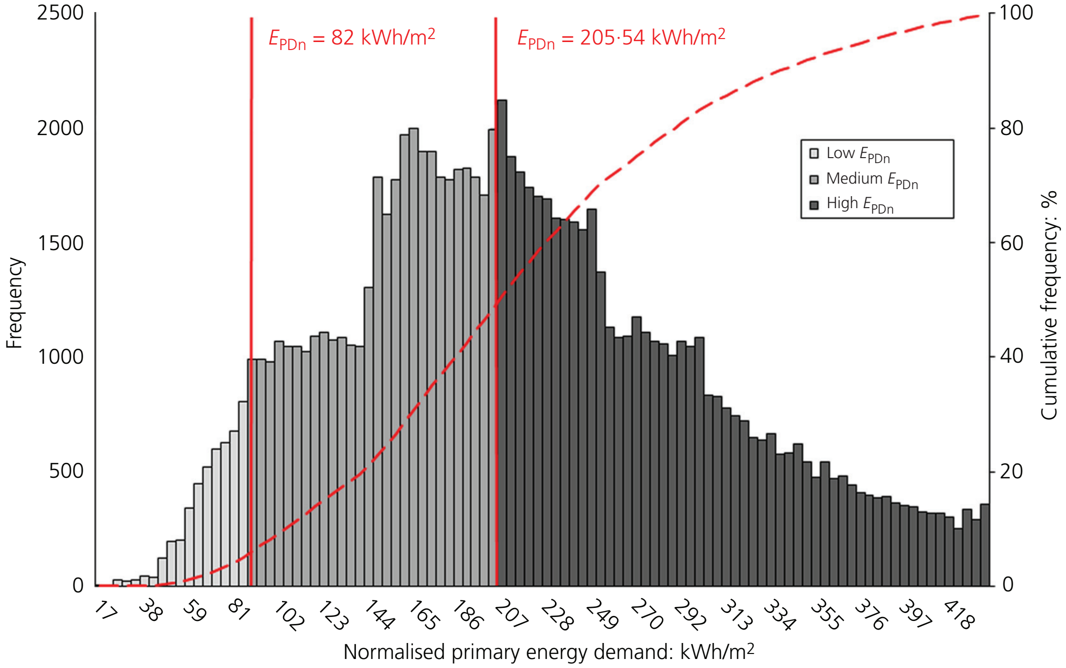


Figure 5. E_{PDn} frequency distribution (classes are highlighted with different shades of grey)

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

Consumption class	E_{PDn}: kWh/m^2	E_{PDnDD}: kWh/DD m^2	Percentile
Low	$0 \leq E_{\mathrm{PDn}} \leq 82$	$0 \leq E_{\mathrm{PDnDD}} \leq 3{\cdot}13 \times 10^{-2}$	1–5
Medium	$82 \leq E_{\mathrm{PDn}} \leq 205{\cdot}54$	$3{\cdot}13 \times 10^{-2} \leq E_{\mathrm{PDnDD}} \leq 7{\cdot}85 \times 10^{-2}$	6–50
High	$E_{\mathrm{PDn}} \geq 205{\cdot}54$	$E_{\mathrm{PDnDD}} \geq 7{\cdot}85 \times 10^{-2}$	51–100

DD, degree day

Table 2. E_{PDn} classes

A+, A and B). In this paper, the same criterion was adopted and these samples were labelled as 'low consuming'. Furthermore, the authors noticed that this cluster represents the 5th percentile of flats with the better E_{PDn}. This is a value that could be used for a further generalisation: in a generic data set samples labelled 'low' are the 5th percentile of E_{PDn}. Afterwards, the median was used as a threshold value for splitting medium-E_{PDn} flats from high-E_{PDn} flats. For this reason, flats with an E_{PDn} higher than 205·54 kWh/m^2 were classified as 'high'. Table 2 summarises the selected threshold values between classes. Moreover, in Table 2 a further normalisation of E_{PDn} considering Turin degree day (DD) was performed (E_{PDnDD}).

5. Supervised classification process

5.1 Results of classification tree (CART)

A classification tree was built (Capozzoli *et al.*, 2015b) based on the most important attributes influencing the E_{PDn} (aspect ratio, opaque and transparent envelope average U-value and average global efficiency of the system for space heating and DHW). The classification process involved the introduction of a set of decision rules for the characterisation of the splitting criteria.

By considering the four input attributes, a classification tree model was developed to predict three categorical variables of E_{PDn}: low, medium and high. The classification tree was initially developed to its maximum size by setting the minimum number of cases in the parent and child nodes (1000 and 800 cases, respectively) and the maximum decrease in the impurities of each split (impurity$_{\mathrm{SPLIT}}$ = 0·001). Subsequently, a pruning analysis was carried out to remove the leaf nodes, which did not improve the classification process. Thus, each leaf node with an error rate higher than 25% was removed. Each leaf node in the final tree contains at least 1% sample of the total and has a minimum accuracy of 75%. To evaluate the performance of the learning process, the number of validation k was set equal to 15 (cross-validation). In Table 3 the confusion matrix is reported, illustrating for each class how instances from a specific class received various classifications. The rows show the real categorical label attributes, whereas the columns illustrate the label attributes given by the classification process. The numbers of correctly classified cases appear as bold values in Table 3. The last row shows that 83·70% of all training records are correctly classified as low, medium and high E_{PDn}.

The boxes in Figure 6 represent the different nodes of the classification tree. The first node is the root node, which considers the whole data set of 92 906 flats. The leaf nodes report the final class of E_{PDn} in which the samples are classified. Furthermore, in each node the number of split samples and their percentage of the total are also reported. When the node is not a leaf node, the logic condition for the following split is marked in the third row. In this case, if the logic condition is fulfilled, branch Y (yes) has to be followed; otherwise, branch N (no) has to be followed.

The algorithm can be translated into a set of decision rules, which have the following form: if antecedent conditions, then consequent conditions. In Table 4 the results of the classification tree are presented in terms of decision rules, starting from the root node and following all the possible ways of reaching each leaf node. The first column titled 'E_{PDn} class' shows the final nodes of the tree, which classify the E_{PDn}. The second column shows the rules that have to be respected in order to classify flats in categorical energy classes, considering the conditions in different rows. The third column indicates the amount of samples included in a final node and their percentage on the total data set.

5.2 Critical analysis of the classification tree split variables

Useful information and benefits can be inferred from a classification tree. Therefore, by examining the decision rules, the significant factors influencing E_{PDn} profiles can be identified and threshold values of the influencing attributes can be derived (Mikučionienė *et al.*, 2014). The first split is driven by the

		Classified			
		Low	Medium	High	Correct: %
Real	Low	**3188**	1327	0	70·6
	Medium	232	**33 151**	8440	79·3
	High	0	5131	**41 437**	89·0
	Accuracy				83·7

Bold numbers are the numbers of correctly classified cases

Table 3. Confusion matrix

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

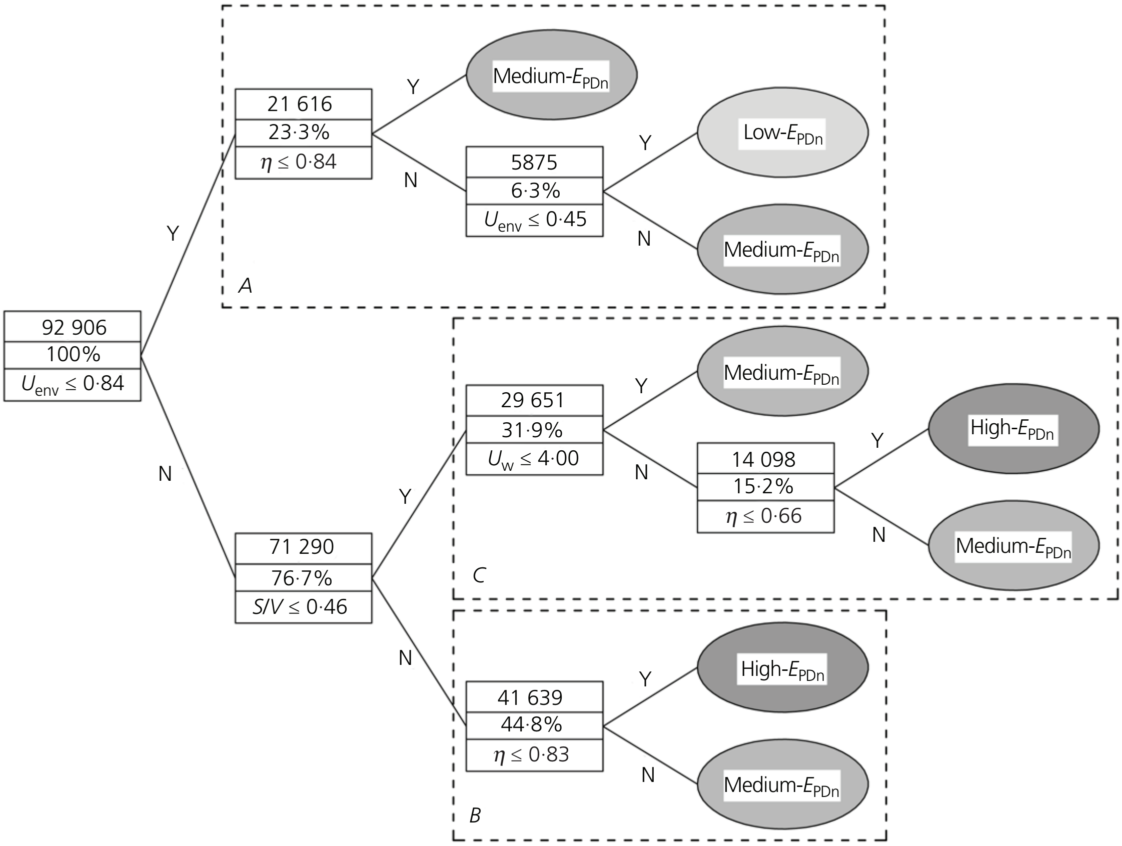


Figure 6. Classification tree obtained using the CART algorithm

attribute that most influences the E_{PDn}. As shown in Figure 6 the average U-value of the opaque envelope is the first split variable of the classification model. In particular, with U-value lower than 0·84 W/m^2 K, all the flats are classified as 'medium' or 'low'. This part of the tree is highlighted in Figure 6 with the area marked as A. In this area, the flats classified as low-E_{PDn} (≤82·00 kWh/m^2) can be divided from flats with medium-E_{PDn}. In comparing the threshold U-value with the ones reported in Figure 3, it is clear that each flat built after 2005 is included in area A. In particular, following the rules listed in Table 4, the 'low' samples are characterised by average U-value of the opaque envelope lower than 0·45 W/m^2 K and an average global efficiency of the system higher than 0·84.

If the average U-value of the opaque envelope is higher than 0·84 W/m^2 K, branch N should be followed after the first split.

E_{PDn} class	Attributes				Amount	
Low	$U_{env} \leq 0{\cdot}45$	—	$\eta > 0{\cdot}84$	—	4275	4·6%
Medium	$0{\cdot}45 \leq U_{env} \leq 0{\cdot}84$	—	$\eta > 0{\cdot}84$	—	1600	1·7%
	$U_{env} \leq 0{\cdot}84$	—	$\eta \leq 0{\cdot}84$	—	15 741	16·9%
	$U_{env} > 0{\cdot}84$	—	$\eta > 0{\cdot}83$	$S/V > 0{\cdot}46$	2171	2·3%
	$U_{env} > 0{\cdot}84$	$U_w \leq 4{\cdot}00$	—	$S/V \leq 0{\cdot}46$	15 553	16·7%
	$U_{env} > 0{\cdot}84$	$U_w > 4{\cdot}00$	$\eta > 0{\cdot}66$	$S/V \leq 0{\cdot}46$	5416	5·8%
High	$U_{env} > 0{\cdot}84$	—	$\eta \leq 0{\cdot}83$	$S/V > 0{\cdot}46$	39 468	42·5%
	$U_{env} > 0{\cdot}84$	$U_w > 4{\cdot}00$	$\eta \leq 0{\cdot}66$	$S/V \leq 0{\cdot}46$	8682	9·3%

Table 4. E_{PDn} classes and classification criteria

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

Real reference flat	S/V: m^{-1}	U_{env}: W/m^2 K	U_{w}: W/m^2 K	η	E_{PDn}: kWh/m^2
rF$_1$ (cluster 1)	0·65	1·17	3·74	0·74	259
rF$_2$ (cluster 2)	0·44	1·37	4·02	0·60	263
rF$_3$ (cluster 3)	0·71	1·55	3·92	0·65	359

Table 5. Real reference flat attributes for each cluster

The following child node takes into account the flat aspect ratio. This second split highlights that the aspect ratio is the principal attribute affecting the E_{PDn} of flats with a higher average U-value for the opaque envelope. If the aspect ratio is higher than 0·46, the B area is defined. In general, the samples in this area are mainly classified in the high energy demand class. Only a small percentage with an average global efficiency of the system for space heating and DHW higher than 0·83 belongs to the medium-E_{PDn} class.

Finally, an aspect ratio lower than 0·46 leads to the C area. Once again, the flats included in this area belong to medium and

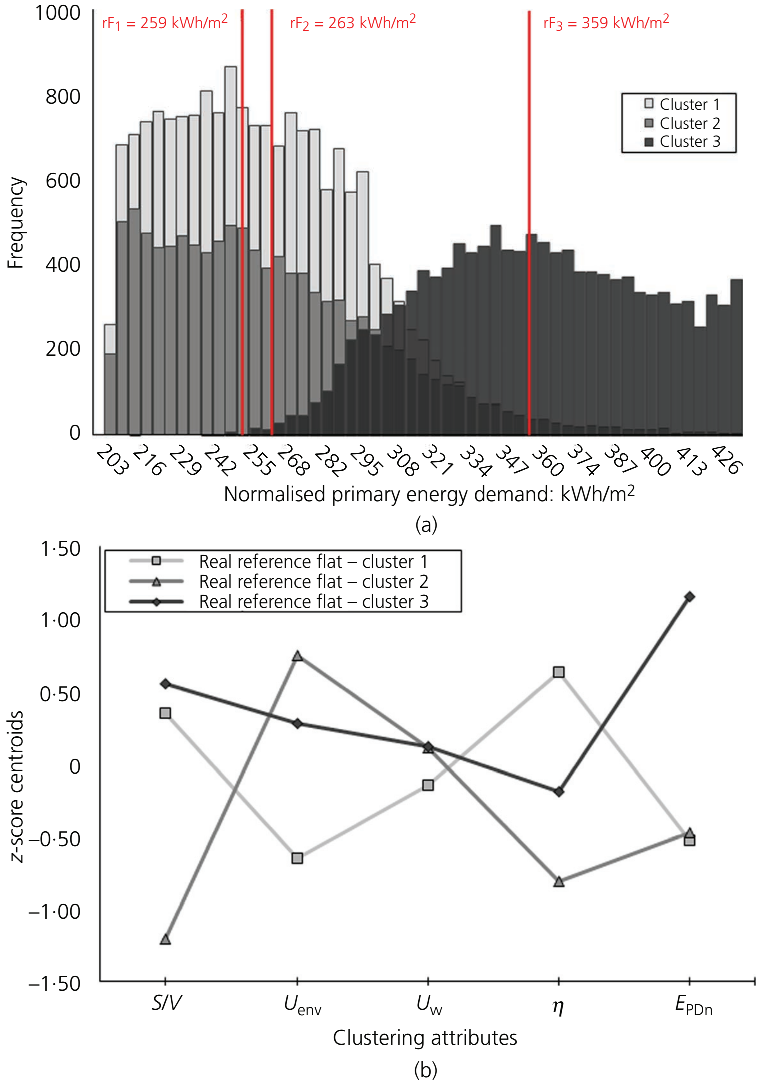


Figure 7. (a) E_{PDn} for high-consumption flats, highlighted with different shades of grey for each cluster; (b) vector components of cluster centroids

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

high-consumption classes. The parent node of area C splits the data based on the average *U*-value of the windows. If it is lower than 4·00 W/m^2 K, the energy demand is classified into a leaf node belonging to the medium-consumption class. Additionally, 21·50% of the flats (3351) grouped in this leaf node were built before 1960 and the average *U*-value of window of these sample is lower than 2·50 W/m^2 K. Therefore, it can be deduced that these windows were subject to a refurbishment.

5.3 Classification accuracy

In the developed classification tree, 83·70% of the data set was correctly classified, demonstrating a reliable accuracy. The best classified class includes the high-consumption flats. On the contrary, the worst classified samples belong to the low-consumption class. This result was predictable mainly because of the intrinsic definition of this class. Indeed, having to include the best 5% performing samples, the dimension of this cluster is significantly lower than the others and some affecting attributes could be neglected. Nevertheless, 70·60% of samples in this class are correctly classified and this accuracy is still acceptable. Moreover, misclassifications between extreme classes (‘low’ to ‘high’ and vice versa) were not present. Lastly, the remaining 16·30% of inaccuracy of the model can be amply explained.

Furthermore, some of the misclassification drawbacks are due to the restricted number of attributes considered in the classification tree. Indeed, on the one hand, the lower the number of attributes, the simpler the usability of the classification model. On the other hand, a low number of attributes might cause the neglecting of some physical processes. Considering low-E_{PDn} samples misclassified as ‘medium’, some of these drawbacks can be attributed to the neglecting of the data regarding ventilation need. In fact, to split the samples, the classification tree does not use any variables related to the efficiency of a potential mechanical ventilation heat-recovery system installed. It is commonly known that for low-consumption buildings, ventilation represents an important voice to be considered for the evaluation of E_{PDn}.

6. Descriptive learning from high-consumption flats

6.1 Unsupervised classification and real reference flat selection

The 46 568 flats labelled as high-E_{PDn} were further investigated since in this class high energy saving opportunities exist. Other authors have demonstrated that a large building stock can be efficiently simulated by using a small number of reference buildings (Filogamo *et al.*, 2014; Mata *et al.*, 2014; Petcharat *et al.*, 2012). Furthermore, an important step to promote the efficient use of energy is to establish benchmark values and to identify the flats that most need energy improvements (Mikučionienė *et al.*, 2014).

In this research, a *K*-means algorithm (Wu, 2012) was adopted to find clusters of high-consumption flats with common features. Before performing the cluster analysis, each attribute was standardised by the *z*-score method to compare attributes between them by assuming the same distribution ($\mu = 0$; $\sigma = 1$). The same input attributes used in the classification tree were selected. The evaluation of the Davies-Bouldin index (Wu, 2012) showed that the *K*-means algorithm with three clusters produced the best clustering output. In particular, the samples are evenly distributed in the three clusters (13 970 samples in cluster 1; 14 436 samples in cluster 2; 18 162 in cluster 3), allowing to have a balanced and representative segmentation.

For each cluster the real reference flat (rF$_i$) closest to the centroid was selected as the most representative. The reference flats characterised by minimum distance d$_i$ were found using the least squares method (Equation 1). In particular, Table 5 shows

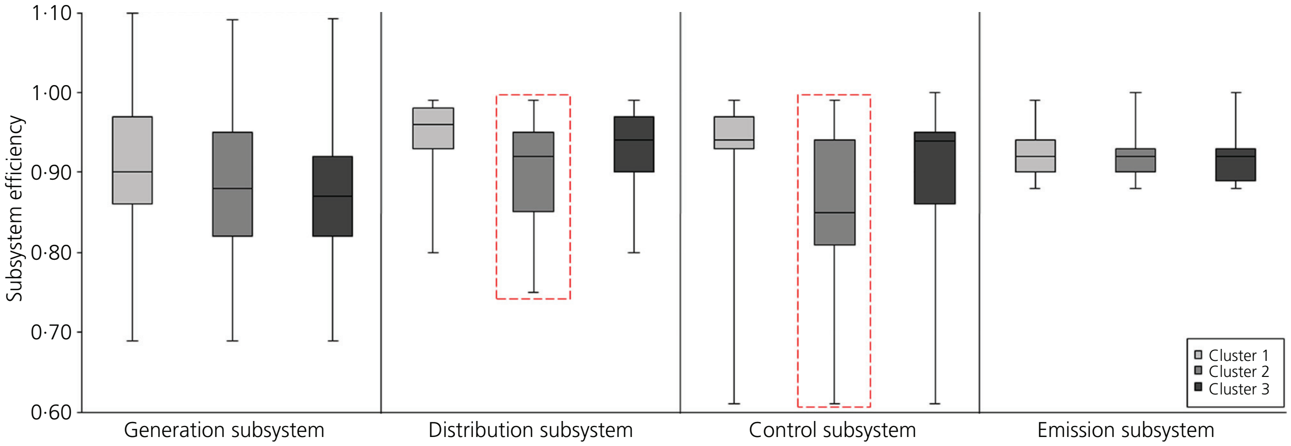


Figure 8. Box plot analysis of the heating subsystems efficiencies for each high-E_{PDn} cluster

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

the attributes of the reference flats (rF_1, rF_2 and rF_3) for each cluster.

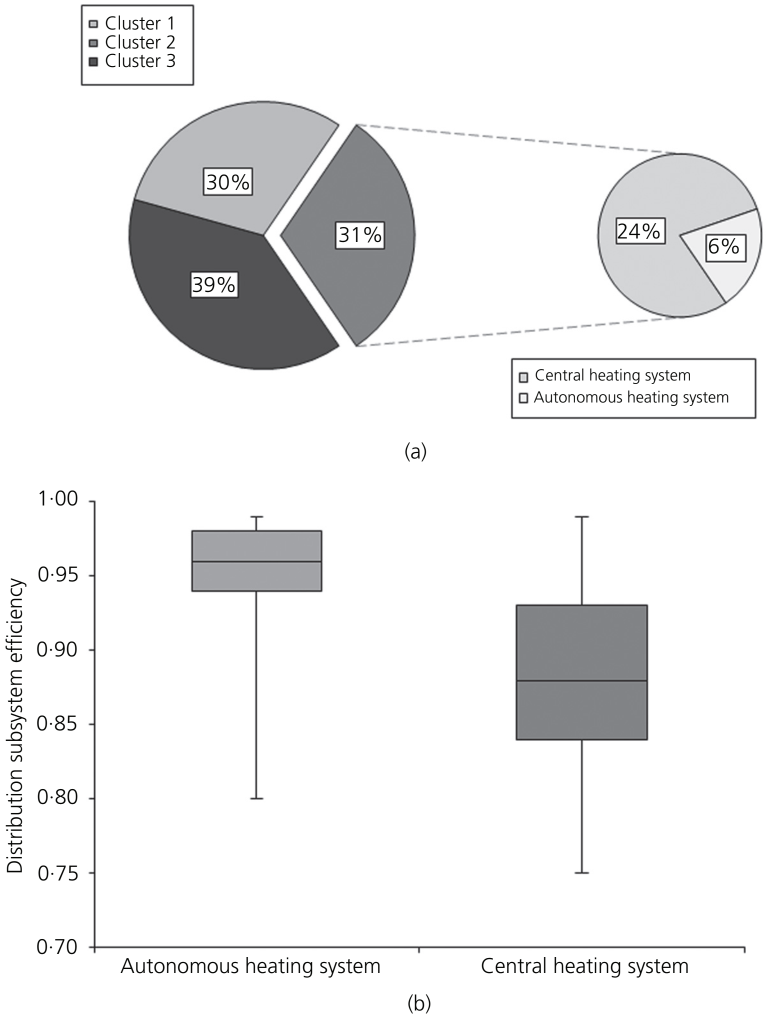
$$\text{d}_i = \min\left\{ \left[\left\{ \sum_{j=1}^{n} \left[z(x)_j - z(c)_j \right]^2 \right\}^{1/2} \right]_{\text{f}} \right\} \tag{1}$$

where i = cluster, n = attribute and f = flats of i cluster.

Frequency distribution of E_{PDn} reported in Figure 7(a) for high-consumption flats shows the location of the objects in each cluster. However, the causes of high consumption are more evident if the attributes standardised with z-scores are analysed for each reference flat (Figure 7(b)).

From Figure 7 it is possible to see that the three reference flats are characterised by different values of geometrical, constructive and system variables. This allows the data set to be further investigated due to the cluster analysis ability to emphasise the inter-cluster similarities and intra-cluster dissimilarities at the same time. In the first instance, the aspect ratio factor has been considered due to its direct effect on the energy losses through the building envelope.

In particular rF_1 and rF_3 are characterised by a higher aspect ratio than rF_2. For this reason the E_{PDn} of rF_1 and rF_2 are much more influenced by the thermo-physical performance of the opaque and transparent envelope. Indeed, the E_{PDn} of rF_3 is significantly higher than the E_{PDn} of rF_1 due to the worst combination of the constructive and system attributes. A different reasoning can be applied to flats grouped in cluster 2. Indeed, for this cluster the low value of the E_{PDn} is mostly due to the low aspect ratio. In this case, the geometrical shape of the flats belonging in this cluster compensates for the low efficiency of the building envelope and system.



(a)

(b)

Figure 9. (a) Pie chart highlighting the distribution of the flats among the different clusters and pie chart highlighting system typologies (central or autonomous) in cluster 2; (b) box plot analysis on the distribution subsystem efficiency for the two different system typologies (central or autonomous) in cluster 2

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

A1								A2	A3	
Wall insulation polystyrene ($\lambda = 0{\cdot}040$ W/m K)		Wall insulation cork ($\lambda = 0{\cdot}041$ W/m K)		Air cavity cork insufflation ($\lambda = 0{\cdot}037$ W/m K)		Roof insulation fibreglass ($\lambda = 0{\cdot}040$ W/m K)		Window substitution	Boiler substitution	Thermostatic valve installation
s: cm	Cost: €/m^2	*s*: cm	Cost: €/m^2	*s*: cm	Cost: €/m^2	*s*: cm	Cost: €/m^2	Cost: €/m^2	Cost: €	Cost: €
8–12	60–65	8–12	75–80	6–10	25–30	8–12	35–40	350–400	1700–1900	80–100

Reference prices according to Piedmont Region guidelines

Table 6. Possible cost related to retrofit actions (A1, A2 and A3) individuated as the most suitable and commonly used for the investigated flat typologies

Additional information for high-consumption flats can be extracted by means of an expert analysis supported by visualisation tools, also using attributes previously not considered for the classification tree and clustering analysis. For example Figure 8 shows the box plots of the efficiencies of each heating subsystem. It is clear how flats included in cluster 2 present efficiencies of control and distribution subsystems with the lowest median value and the highest inter-quartile range. As shown in Figure 9, 80% of the flats grouped in cluster 2 have a centralised heating system. In general, these flats are characterised by higher heat losses in the distribution subsystem, especially in old systems where pipes are not well insulated. Moreover, the control subsystems have low performance because they are generally ‘climatic’ and based and not specific for each thermal zone/ambient. The combination of these two aspects suggests the assumption that high energy savings are achievable by retrofitting the control and distribution subsystem of flats equipped with old centralised heating systems. It is important to have dominant features in a cluster in order to extract useful information to formulate strategies for energy saving.

6.2 Analysis of possible retrofit actions

Benefits achievable from some possible common retrofit actions was analysed for each reference flat. The aspect ratio is an intrinsic feature of each flat; thus, it cannot be improved through refurbishment actions. However, energy retrofitting designers can improve the other three construction attributes with different retrofit actions, called A1, A2 and A3. Action A1 is related to the increasing of the insulation of the vertical opaque envelope and A2 to the substitution of the existing windows with new high-performing ones. Thus, *U*-value of the vertical opaque envelope and *U*-values of the windows become lower than 0·30, and 1·90 W/m^2 K, respectively, which are the Italian legislation limits for the Turin climatic zone in a refurbishment process (Ministero dello Sviluppo Economico, 2015). Moreover, action A3 consists of the refurbishment of the heating and DHW system

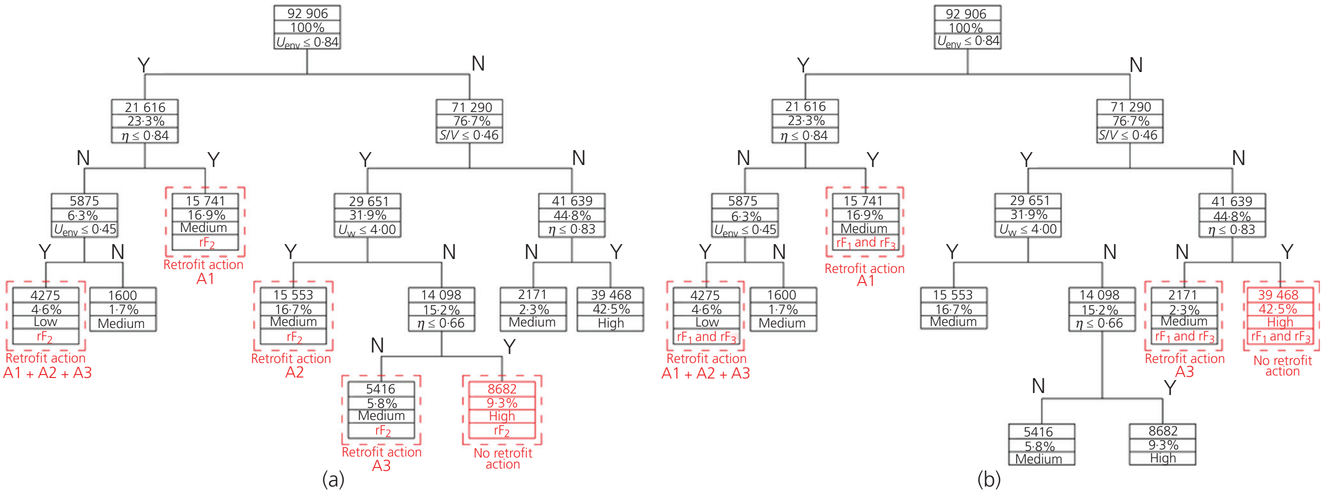


Figure 10. (a) Classification tree after the application of the retrofit actions A1, A2 and A3 on the real reference flat rF$_2$; (b) classification tree after the application of the retrofit actions A1, A2 and A3 on the real reference flats rF$_1$ and rF$_3$

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

(distribution subsystem refurbishment, boiler-controller-terminal substitution). An $\eta > 0{\cdot}85$ can be considered a target threshold in a refurbishment process. In particular, for the flats equipped with autonomous boilers, action A3 includes boiler substitution and installation of zone thermostatic valves. For flats equipped with centralised boilers (mainly similar to rF_2), action A3 should include also additional refurbishment of the distribution subsystem at building level. Table 6 provides an evaluation of unitary reference prices (according to Piedmont region guidelines) required for each of the actions A1, A2 and A3 (only for autonomous heating system) to respect the limit values in the retrofitting process. The cost of roof insulation is reported for the sake of completeness.

The classification tree previously trained was tested to classify rF_1, rF_2 and rF_3 considering the three retrofit actions and the limit values for the attributes suggested by the Italian energy legislation. In this way it is possible to evaluate roughly the effect of a single action or of a combination of actions. The classification testing is shown in Figure 10 for the three reference flats. Table 7 summarises the results of this analysis. On the one hand, the application of one single refurbishment action can provide indications on the most suitable energy efficiency measure for each cluster. On the other hand, the application of all the refurbishment actions at the same time gives an indication of the best class in which a reference flats can be classified.

As it can be seen from Table 7, in seven over nine cases the retrofit of a single attribute allows the flats to switch from the high to the medium-consumption class. In the remaining two over nine cases, the consumption remained high. These cases coincide with the substitution of windows (A2) in rF_1 and rF_3. This fact highlights that this energy efficiency measure is not very effective on flats with a high value of aspect ratio. Furthermore, it is clear that the adoption of all three measures at the same time allows the all reference flats to switch from the high to the low-consumption class. This means that the energy consumption after the refurbishment could reach potentially a value lower of 82 kWh/m^2 K. The average achievable energy saving for each high-E_{PDn} flat would be higher than 208 kWh/m^2.

Real reference dwelling	Energy retrofit action	E_{PDn} class
rF_1 (cluster 1)	A1	Medium
rF_1 (cluster 1)	A2	High
rF_1 (cluster 1)	A3	Medium
rF_1 (cluster 1)	A1 + A2 + A3	Low
rF_2 (cluster 2)	A1	Medium
rF_2 (cluster 2)	A2	Medium
rF_2 (cluster 2)	A3	Medium
rF_2 (cluster 2)	A1 + A2 + A3	Low
rF_3 (cluster 3)	A1	Medium
rF_3 (cluster 3)	A2	High
rF_3 (cluster 3)	A3	Medium
rF_3 (cluster 3)	A1 + A2 + A3	Low

Table 7. E_{PDn} classification after the application of the three retrofit actions: A1, improvement of opaque envelope *U*-value; A2, substitution of windows with more efficient ones; A3, refurbishment of the space heating and DHW system

7. Conclusions

In this paper, a classification process involving 92 906 flats was conducted. Due to the large dimensions of the adopted data set, the information provided can be considered representative of the Piedmont region residential flat stock. The method is easily adaptable to different data sets and attributes, since the classification criteria are based on statistical variables.

The influence on the normalised primary energy demand (E_{PDn}) of four influencing attributes (aspect ratio, *U*-value of vertical opaque envelope and windows and average global efficiency of the system for space heating and DHW) was analysed through a classification tree. Further analyses on flats classified as high-consumption were carried out. Three different clusters with similar feature patterns were identified through a cluster analysis. For each of them a reference flat was located and the effect of different retrofit actions was investigated.

Future works will firstly investigate additional data sets, to lower the error rate limit of the classification tree and to increase further the reliability of the proposed methodology. Secondly, the real reference flats for the high-E_{PDn} class can be used as reference buildings for more accurate energy simulations (Filogamo *et al*., 2014). Finally, the influence of building owners decision on the application of the proposed retrofit actions (Galiotto *et al*., 2015) and a cost optimal analysis (Ferrara *et al*., 2014) will be investigated.

Acknowledgements

The authors express their gratitude to Giovanni Nuvoli (Settore Sviluppo Energetico Sostenibile – Regione Piemonte) and to CSI Piemonte.

REFERENCES

Aksoezen M, Daniel M, Hassler U and Kohler N (2015) Building age as an indicator for energy consumption. *Energy and Buildings* **87**: 74–86, http://dx.doi.org/10.1016/j.enbuild.2014.10.074.

Ballarini I, Corgnati SP, Corrado V and Talà N (2011) Definition of building typologies for energy investigations on residential sector by Tabula IEE project: application to Italian case studies. *Roomvent, Trondheim, Germany*, pp. 19–22.

Ballarini I, Corgnati SP and Corrado V (2014) Use of reference building to asses the energy saving potentials of the residential building stock: the experience of TABULA project. *Energy Policy* **68**: 273–284, http://dx.doi.org/10.1016/j.enpol.2014.01.027.

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

Capozzoli A, Grassi D and Causone F (2015a) Estimation models of heating energy consumption in schools for local authorities planning. *Energy and Buildings* **105**: 302–313, http://dx.doi.org/10.1016/j.enbuild.2015.07.024.

Capozzoli A, Grassi D, Piscitelli MS and Serale G (2015b) Discovering knowledge from a residential building stock through data mining analysis for engineering sustainability. *Energy Procedia* **83**: 370–379, http://dx.doi.org/10.1016/j.egypro.2015.12.212.

Capozzoli A, Lauro F and Khan I (2015c) Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications* **42(9)**: 4324–4338, http://dx.doi.org/10.1016/j.eswa.2015.01.010.

Capozzoli A, Piscitelli MS, Neri F, Grassi D and Serale G (2016) A novel methodology for energy performance benchmarking of buildings by means of linear mixed effect model: the case of space and DHW heating of out-patient healthcare centres. *Applied Energy* **171**: 592–607, http://dx.doi.org/10.1016/j.apenergy.2016.03.083.

D'Oca S and Hong T (2015) Occupancy schedules learning process through a data mining framework. *Energy and Buildings* **88**: 395–408, http://dx.doi.org/10.1016/j.enbuild.2014.11.065.

Elghali L, Clift R, Begg KG and McLaren S (2008) Decision support methodology for complex contexts. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **161(1)**: 7–22, http://dx.doi.org/10.1680/ensu.2008.161.1.7.

Fan C, Xiao F and Yan C (2015) A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Automation in Construction* **50**: 81–90, http://dx.doi.org/10.1016/j.autcon.2014.12.006.

Ferrara M, Fabrizio E and Virgone J (2015) Appraising the effect of the primary systems on the cost optimal design of nZEB: a case study in two different climates. *Energy Procedia* **78**: 2028–2033, http://dx.doi.org/10.1016/j.egypro.2015.11.200.

Filogamo L, Peri G, Rizzo G and Giaccone A (2014) On the classification of large residential buildings stocks by sample typologies for energy planning purposes. *Applied Energy* **135**: 825–835, http://dx.doi.org/10.1016/j.apenergy.2014.04.002.

Fracastoro GV and Serraino M (2011) A methodology for assessing the energy performance of large scale building stocks and possible applications. *Energy and Buildings* **43(4)**: 844–852, http://dx.doi.org/10.1016/j.enbuild.2010.12.004.

Galiotto N, Heiselberg P and Knudstrup MA (2015) The Integrated Renovation Process: application to family homes. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **168(6)**: 245–257, http://dx.doi.org/10.1680/ensu.14.00020.

Gao Y, Tumwesigye E, Cahill B and Menzel K (2010) Using data mining in optimisation of building energy consumption and thermal comfort management. In *2nd International Conference on Software Engineering and Data Mining (SEDM)* (IEEE (ed.)). IEEE, Piscataway, NJ, USA, pp. 434–439.

ISO (International Organization for Standardization) (2008) EN ISO 13790: 2008: Energy performance of buildings – calculation of energy use for space heating and cooling. EU, Brussels, Belgium.

Khan I, Capozzoli A, Corgnati SP and Cerquitelli T (2013) Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia* **42**: 557–566, http://dx.doi.org/10.1016/j.egypro.2013.11.057.

Kim H, Stumpf A and Kim W (2011) Analysis of an energy efficient building design through data mining approach. *Automation in Construction* **20(1)**: 37–43, http://dx.doi.org/10.1016/j.autcon.2010.07.006.

Kumar B (2011) Data mining approach for friction factor in mobile bed channel. *Proceedings of the Institution of Civil Engineers – Water Management* **164(1)**: 15–25, http://dx.doi.org/10.1680/wama.1000031.

Mata É, Sasic Kalagasidis A and Johnsson F (2014) Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment* **81**: 270–282, http://dx.doi.org/10.1016/j.buildenv.2014.06.013.

Mikučionienė R, Martinaitis V and Keras E (2014) Evaluation of energy efficiency measures sustainability by decision tree method. *Energy and Buildings* **76**: 64–71, http://dx.doi.org/10.1016/j.enbuild.2014.02.048.

Ministero dello Sviluppo Economico (2015) DM 26/06/2015: Decreto interministeriale 26 giugno 2015 – applicazione delle metodologie di calcolo delle prestazioni energetiche e definizione delle prescrizioni e dei requisiti minimi degli edifici. Ministro dello Sviluppo Economico, Rome, Italy.

Motawa I (2015) Dynamic modelling for sustainable dwellings. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **168(4)**: 182–190, http://dx.doi.org/10.1680/ensu.14.00051.

Parkin S, Sommer F and Uren S (2003) Sustainable development : understanding the concept and practical challenge. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **156(1)**: 19–26, http://dx.doi.org/10.1680/ensu.156.1.19.37055.

Petcharat S, Chungpaibulpatana S and Rakkwamsuk P (2012) Assessment of potential energy saving using cluster analysis: a case study of lighting systems in buildings. *Energy and Buildings* **52**: 145–152, http://dx.doi.org/10.1016/j.enbuild.2012.06.006.

Summerfield AJ, Raslan R, Lowe RJ and Oreszczyn T (2011) How useful are building energy models for policy? A UK perspective. *12th Conference of the International Building Performance Simulation, Sydney, Australia*, pp. 14–16.

Swan W and Cantab MA (2015) A UK practitioner view of domestic energy performance measurement. *Proceedings of the Institution of Civil Engineers – Engineering Sustainability* **168(3)**: 140–147, http://dx.doi.org/10.1680/ensu.14.00056.

UNI (Ente Nazionale Italiano di Unificazione) (2008a) UNI/TS 11300-1: 2008: Energy performance of buildings – Part 1: evaluation of energy need for space heating and cooling. Milan, Italy (in Italian).

UNI (2008b) UNI/TS 11300-2: 2008: Energy performance of buildings – Part 2: evaluation of primary energy and system

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.

efficiencies for space heating and domestic hot water production. Milan, Italy (in Italian).

Wu J (2012) *Advances in K-means Clustering: a Data Mining Thinking*, 1st edn. New York, NY, USA.

Xiao F and Fan C (2014) Data mining in building automation system for improving building operational performance. *Energy and Buildings* **75**: 109–118, http://dx.doi.org/10.1016/j.enbuild.2014.02.005.

Yu Z, Fung BCM and Haghighat F (2013) Extracting knowledge from building-related data – a data mining framework. *Building Simulation* **6(2)**: 207–222, http://dx.doi.org10.1007/s12273-013-0117-8.

Yu Z, Haghighat F, Fung BCM and Yoshino H (2010) A decision tree method for building energy demand modeling. *Energy and Buildings* **42(10)**: 1637–1646, http://dx.doi.org10.1016/j.enbuild.2010.04.006.

WHAT DO YOU THINK?

To discuss this paper, please submit up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial panel, will be published as a discussion in a future issue of the journal.

Proceedings journals rely entirely on contributions sent in by civil engineering professionals, academics and students. Papers should be 2000–5000 words long (briefing papers should be 1000–2000 words long), with adequate illustrations and references. You can submit your paper online via www.icevirtuallibrary.com/content/journals, where you will also find detailed author guidelines.

Downloaded by [POLITECNICO DI TORINO] on [27/06/16]. Copyright © ICE Publishing, all rights reserved.