View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UPCommons. Portal del coneixement obert de la UPC

Energy Efficiency

Influencing factors in energy use of housing blocks: A new methodology, based on clustering and energy simulations, for decision making in energy refurbishment projects

--Manuscript Draft--

Manuscript Number:	
Full Title:	Influencing factors in energy use of housing blocks: A new methodology, based on clustering and energy simulations, for decision making in energy refurbishment projects
Article Type:	Original Research
Keywords:	Building energy use; Energy building simulation; Clustering analysis; Urban energy refurbishment
Corresponding Author:	Xavier Cipriano, M.D Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE) Terrassa, Barcelona SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE)
Corresponding Author's Secondary Institution:	
First Author:	Xavier Cipriano, M.D
First Author Secondary Information:	
Order of Authors:	Xavier Cipriano, M.D
	Alfredo Vellido, Ph.D
	Jordi Cipriano, Ph.D
	Jaime Martí-Herrero, Ph.D
	Stoyan Danov, Ph.D
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	In recent years big efforts have been dedicated to identify which are the factors with highest influence in the energy consumption of residential buildings. These factors include aspects such as weather dependence, user behaviour, socio-economic situation, type of the energy installations, and typology of buildings. The high number of factors increases the complexity of analysis and leads to a lack of confidence in the results of the energy simulation analysis. This fact grows when we move one step up and perform global analysis of blocks of buildings. The aim of this study is to report a new methodology for the assessment of the energy performance of large groups of buildings when considering the real use of energy. We combine two clustering methods; Generative Topographic Mapping and k-Means, to obtain reference dwellings that can be considered as representative of the different energy patterns, and energy systems of the neighbourhood. Then, simulation of energy demand and indoor temperature against the monitored comfort conditions in a short period is performed to obtain end-use loads disaggregation. This methodology was applied in a district at Terrassa city (Spain), and six reference dwellings were selected. Results show that the method was able to identify the main patterns, and provide occupants with feasible recommendations so that they can make required decisions at neighbourhood level. Moreover, given that the proposed method is based on the comparison with similar buildings, it could motivate building occupants to implement community improvement actions, as well as to modify their behaviour.
Suggested Reviewers:	Tianzhen Hong, Ph.D Leader Group, LBNL

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	thong@lbl.gov For his expertise in energy modelling of buildings, as one of the leaders of the Energy Plus development
	Stefano Paolo Corgnati, Ph.D Associate professor, Politecnico di Torino stefano.corgnati@polito.it For his work in IEA Annex 53, and other publications related to energy urban planning and energy efficiency in large stock of buildings
	Paul Fleming, Ph.D Director, De Montfort University PFleming@dmu.ac.uk His large expertise in energy data series modelling, as well as in energy use in buildings

The final publication is available at Springer via
https://link.springer.com/article/10.1007%2Fs12053-016-9460-9

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Title Page

Influencing factors in energy use of housing blocks: A new methodology, based on clustering and energy simulation, for decision making in energy refurbishment projects

Corresponding autor: Xavier Cipriano

X. Cipriano[*], A. Vellido[a], J. Cipriano[b], J. Martí-Herrero[c,a] S. Danov[e],

* Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE),
Building Energy and Environment Group,
Edifici GAIA (TR14), C/Rambla Sant Nebridi 22,
08222, Terrassa, Barcelona, Spain
e-mail: xciprian@cimne.upc.edu
Tel: +34937398575/+34647135392

[a] Computer Science,Universitat Politècnica de Catalunya (UPC BarcelonaTech),
Campus Nord UPC
08034, Barcelona, Spain
e-mail: avellido@lsi.upc.edu
Tel: +34 93 4137796

[b] Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE),
Building Energy and Environment Group,
CIMNE-UdL Classroom. Pere de Cabrera s/n. Building
CREA. O-ce 1.15.
25001 Lleida, Spain
e-mail: cipriano@cimne.upc.edu
Tel: +34 973003574

[c] PROMETEO Researcher. Instituto Nacional de Eficiencia Energética y Energías Renovables (INER).
6 de Diciembre N33-32,
Quito, Ecuador.
e-mail: jaimemarti@cimne.upc.edu

[d] Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE),
Building Energy and Environment Group,
Edifici GAIA (TR14), C/Rambla Sant Nebridi 22,
08222, Terrassa, Barcelona, Spain
e-mail: sdanov@cimne.upc.edu
Tel: +34937398575

Corresponding Author: Xavier Cipriano
Contact: xciprian@cimne.upc.edu. Tel: +34937398575 / +34647135392

Acknowledgements

Xavier Cipriano is thankful to the Catalan Housing Agency of the Catalan Government, for funding the main en research, as well as to María Heras, Nuria Cardellach, and Eduard Calderón for their participation in the energy audits and surveys to tenants. We also are thankful to Dr. Nicolas Kosoy for his suggestions in the socio-economic analysis of surveys, as well as the Can Jofresa's neighbours association for their answers and information about their homes. Jaime Martí-Herrero is thankful to the Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation of the Republic of Ecuador that funded part of his work.

Blinded Manuscript
Click here to view linked References

Influencing factors in energy use of housing blocks: A new methodology, based on clustering and energy simulations, for decision making in energy refurbishment projects

ABSTRACT

In recent years big efforts have been dedicated to identify which are the factors with highest influence in the energy consumption of residential buildings. These factors include aspects such as weather dependence, user behaviour, socio-economic situation, type of the energy installations, and typology of buildings. The high number of factors increases the complexity of analysis and leads to a lack of confidence in the results of the energy simulation analysis. This fact grows when we move one step up and perform global analysis of blocks of buildings. The aim of this study is to report a new methodology for the assessment of the energy performance of large groups of buildings when considering the real use of energy. We combine two clustering methods; Generative Topographic Mapping and k-Means, to obtain reference dwellings that can be considered as representative of the different energy patterns, and energy systems of the neighbourhood. Then, simulation of energy demand and indoor temperature against the monitored comfort conditions in a short period is performed to obtain end-use loads disaggregation. This methodology was applied in a district at Terrassa city (Spain), and six reference dwellings were selected. Results show that the method was able to identify the main patterns, and provide occupants with feasible recommendations so that they can make required decisions at neighbourhood level. Moreover, given that the proposed method is based on the comparison with similar buildings, it could motivate building occupants to implement community improvement actions, as well as to modify their behaviour.

Keywords: Building energy use; Energy building simulation; Clustering analysis; Urban energy refurbishment

Highlights:

- Energy audits, tenant's surveys, and empirical tests in households are performed
- Normalization and selection of relevant variables with respect to energy use are obtained
- Clustering of variables is carried out to characterize the different groups of dwellings
- Refinement of energy simulation of representative dwellings with monitoring data is presented
- Energy disaggregation and stock aggregation to the whole district is calculated
- Results serve to evaluate the energy current situation and related socio-economical impacts
- Potential impact of energy saving measures are finally presented

Abbreviations: IEA-EBC: International Energy Agency-Energy in Building and Communities. EPBD: European Union Energy Performance of Buildings Directive. EUI: Energy Use Intensity.

1. INTRODUCTION

In recent years, there has been a growing interest in understanding and analysing the real energy performance of buildings. This interest has been mainly driven by the evidence of the high variability in the energy consumption of buildings with very similar characteristics (IEA-CBCS Annex 33 2010; IEA-CBCS Annex 53 2013). This high variability, along with the lack of confidence in the estimation of real energy use is at its most relevant, and sometimes a critical factor, in projects oriented at the neighbourhood level (IEA-CBCS Annex 33). At such level, the improvement in energy efficiency is not only determined by urban and architectural aspects, but also by the upgrading of installations and/or by changes in users' behaviour. This variety of factors increases the complexity of analysis of the real energy use of buildings. This complexity can be partly explained by defining the influencing factors in energy use of buildings. In the IEA-EBC Annex 53 the main influencing factors of building energy consumption were proposed to fall into six categories: 1) climate, 2) building envelope, 3) building services and energy systems, 4) building operation and maintenance, 5) occupant activities and behaviour, and 6) indoor environmental quality. The three first categories are related to variables influencing building energy performance and, as defined by the European Union Energy Performance of Buildings Directive (EPBD) (Directive 2010/31/EU 2010) they are usually calculated by fixing standard conditions for the other three categories, which are specifically related to actual building functions. As a consequence, the building energy performance is calculated assuming that all of the analysed buildings operate under the same standardized functioning conditions, as outlined in (IEA-EBC Annex 33 2010). This approach allows a coherent comparison of the calculated building energy performance, but this calculation is not strictly related to the real energy consumption (IEA-CBCS Annex 53 2013). This type of calculation allows obtaining the Asset Ratings energy performance indicators, in contrast to the Operational Ratings which are based on measured energy use, often normalized for relevant variables like climate and level of energy service (Goldstein 2014). When the focus moves to the real use of buildings, all six categories of influencing factors must be taken into account. The influencing factors could be seen as driving forces for changing energy use and are of great relevance if we extend the analysis from an isolated building to a group of buildings, or to a neighbourhood level.

Several models for the integration of the influence of occupants' behaviour into building energy performance calculations have recently been proposed. A number of studies (Nakagami 1996; Lopes et al 2005; Yu et al 2011) suggest that an optimal approach to the quantification of the global effect of occupants' behaviour should be based on knowledge extraction from monitored data and from occupants' surveys rather than on improving theoretical building energy simulation models. Moreover, neighbourhoods or large group of buildings often yield less data (and are less frequently surveyed) than individual buildings. This fact increases the uncertainty of simulations, due to the broad assumptions about input data that must be often relied upon. The results of the reviewed studies (IEA-CBCS Annex 33 2010; Yu et al 2010; Yu et al 2011) show that a combination of statistical analysis with prediction models (both heuristic simulation and inverse models), complemented in some cases with monitoring data analysis, can be a powerful tool for the development of energy urban actions aiming at reducing the energy consumption in existing buildings. According to this approach, descriptive statistics have been used to identify the most important factors and reference members of the set, by grouping the buildings/houses according to them (Räsänen et al 2008; Loughlin et al 2012). The identification of factors may help the better implementation of subsequent steps of the simulation of the current situation and of energy improvement scenarios (Yu et al 2011; Ueno 2006). In the same way, results of a framework to model personalized occupancy profiles for representing occupants' long-term presence patterns presented in (Yang 2014), shows that the personalized occupancy profiles acquired through time-series modelling, pattern recognition modelling and stochastic process modelling, outperform the fixed design profiles currently used in building energy simulations. A brief description of common bottom-up modelling techniques (statistical and building physics-based) can be found in (Kavgic et al 2010; Murray et al 2014). An example of statistical modelling is also described in (Yu et al 2010), where a decision tree method for building energy demand characterization was proposed and applied to historical data from a sample of Japanese residential buildings. Taking the same statistical approach, some studies about classification of buildings according to the relevant factors and the different hourly profiles of users have been carried out. The user behaviour in these studies are usually represented as time-based profiles or patterns. As a general rule in this approach, clustering is used to group energy consumers of similar characteristics (Chicco et al 2003; Chicco 2012), to predict future energy demand, or to detect atypical, usually undesired, behaviours (Räsänen et al 2008, Tsekouras et al 2008; yang et al 2014; Li et al 2010).

On the other hand, and considering only the building physics models at district level, the reported approaches generally include the energy calculation of a sample of houses considered to be representative of the neighbourhood/district/nation stock, as described in (Swan et al 2009). In some cases, simulation methods were used to conduct building energy consumption calculations, in order to identify the correlation between building energy consumption and different influencing factors (e.g., building relative compactness, building control strategies) (Ourghi et al 2007).. However, simulation methods do not perform so well in simulating energy performance for occupied buildings as compared to non-occupied buildings, due to a lack of sufficient knowledge about occupant behaviour patterns, which are normally very difficult to parameterize. Moreover, the calibration of building simulation programs against real conditions is a normally complex undertaking and the learning process is time-consuming (Yu et al 2011).

The main goal of the current study is the assessment of the energy performance of medium and large groups of buildings when aiming at eliciting common characteristics of building/dwelling typologies and the main factors influencing in their energy consumption. Improving the understanding of these influencing factors will allow us not only to improve the accuracy of prediction or classification methods, but also to incorporate the socio-economic impact in decision making for urban refurbishment projects. In order to achieve this objective, the paper is structured as follows: First, the methodology is described as a combination of innovative and standard statistics methods for clustering (Generative Topographic Mapping and K-means), together with simulation tools that are employed to obtain realistic assumptions about user behaviour in the main representative groups of dwellings of a neighbourhood. These assumptions enable the estimation of present energy consumption at the level of individual properties as well as at the neighbourhood level. Then, this approach is implemented in a case study involving a district of the city of Terrassa (Spain), including an estimation of the potential impact of improvement measures. Finally, a discussion of the appropriateness of the approach is also provided.

2. METHODOLOGY

The analytical framework involves both quantitative and qualitative household information (inputs), the steps of the working process (process), and the results obtained at each step (outputs). The developed methodology is presented in Figure 2. The quantitative dataset comprises electricity and gas bills, complemented with electricity consumption and indoor temperature measurements over 15-minute periods, as well as results of blowing door tests. The qualitative dataset includes the household occupants' responses to surveys and interviews carried by the researcher team.

2.1 Data acquisition and treatment

A study entitled "Diagnostic and analysis of energy improvements in low income districts in Catalonia region: case study in Can Jofresa's neighbourhood (Terrassa)" was carried out by CIMNE from May 2008 to May 2010. The investigated neighbourhood is located in the city of Terrassa (Barcelona, population 215,517, as of 2014), in NE Spain, and consists of twelve H-shaped fifteen-story tower blocks (60 dwellings per tower, 720 households in total, see Figure **1**

Fig. 1 General site view (left) and pilot tower detail (right)

For this study, field surveys of energy-related data and other relevant information were carried out in 166 of the 720 residential dwellings. Table 1 shows the surveyed items and the corresponding extracted variables. A blowing door test was also carried out in four dwellings representing the four different types of existing windows, in order to determine the most common infiltration rates. Real samples, together with thermography of façades, were also performed to estimate the average U-value of the external walls and to detect the main existing thermal bridges.
Data reduction and aggregation was then performed to obtain a more parsimonious representation of the original data. Normalization of the yearly energy consumption per unit of surface ($kWh/m^2 \cdot yr$) was applied (called Energy Use Intensity, EUI). An aggregation was carried out in some of the surveyed items for a more clear understanding of the variables under analysis. For instance, questions related to the type of window frames, type of glass, and degree of windows tightness were grouped in a categorization of the quality of windows (1=very poor, 2=poor, 3=good, 4=very good). This process of related-questions grouping was also carried out for the categorical answers, in order to have a more understandable classification.

Subsequently, a data transformation was applied to variables showed in Table 1 to deal with the differences in scale and in categories of the obtained dataset. Specifically, *Min–Max* normalization was performed to scale the values so that they fell within a predetermined range. This technique of linear normalization has the advantage of preserving the relationships between the original data. In this study, the new range is defined as (0, 1).

Fig. 2. Global view of the working process

2.2 Selection of relevant variables

The first approach to obtain the relevance of the attributes with respect to the EUI entails correlation analysis. This relevance corresponds to a weighting scheme that returns the squared value of the correlation as the attribute weight. Only those variables with a correlation weight over 0.35 were selected as relevant (typical). In a subsequent step, the covariance matrix is calculated in order to quantify the rank correlations between variables. A threshold value of 0.8 was set as a minimum criterion to consider two variables as highly correlated. These two processes yield a selection of the most relevant (typical) variables.

Table 1 List of variables selected in this study to analyse their influence in energy consumption.

Code	Name	Acronym	Range value	Value
X1	Space heating	Heating	yes/no	(1,0)
X2	Type of space heating	TypeHeating	elec.stove/gas stove/gas boiler/heat pump	(1 to 4)
X3	Type of windows	Typewindows	very poor/poor/ normal/good	(1 to 4)
X4	Number of months heating	NumMonthHeating	Number (month)	(0 to 6)
X5	Heating schedules	HeatingSchedule	little/morning-afternoon/ lunch-dinner/afternoon/night/all day	(1 to 6)
X6	Degree of comfort	Comfort	very low, low, medium, high	(1 to 4)
X7	Number of rooms unheated	NumRooms Unheat	all/all bedrooms/one bedroom/only dining room/only kitchen/none	(1 to 6)
X8	Air conditioning (AC)	AC	yes/no	(1,0)
X9	Use of AC	UseAC	never/occasionally/few/noon-night/always	(1 to 4)
X10	Number of adults and children	NumPeople	number (person)	(1 to 6)
X11	Total monthly income	TotalMonthIncome	number (€/month)	1160-6000
X12	Degree of good practices in heating	BPSwitch Heating	little awareness/normal awareness/high awareness	(1 to 3)
X13	Use of awnings	AwningsUse	much use/none	(1,0)
X14	Type of cooking facilities	TypeKitchen	gas/ceramic hob	(1,0)
X15	Type of fridge	Fridge	large/medium	(1,0)
X16	Use of washer	UseWasher	very inefficient/inefficient/normal/efficient/ very efficient	(1 to 5)
X17	Number of appliances	Nuppliances	number (units)	(0 to 27)
X18	Switch off appliances by night	StndbyOffSleep	yes/no	(1,0)
X19	Number of energy saving lamps	NumEfficLamps	number (units)	(0 to 20)
X20	Number of fluorescent tubes	NumFluoresc	number (units)	(0 to 5)

2.3 Data clustering with the GTM and k-Means algorithms

Clustering is a process in which we aim to infer data grouping structure that is unknown beforehand. It is often used as an exploratory strategy that attempts to partition the data into groups that are internally homogeneous and different enough from other groups. Unlike in classification, no groups are predefined and there is no explicit modelling of the relationship between data and class labels. In this study, we combine two clustering methods, the namely Generative Topographic Mapping (GTM) (Bishop et al 1998) and k-Means (Jain 2010)], in the exploratory process of grouping the parameterized data. GTM is a probabilistic alternative to the well-known Self-Organizing Maps (SOM) (Kohonen 2001), which has successfully been applied to energy use profiling. In both methods, data clustering becomes secondary to exploratory data visualization in a low-dimensional space (usually, in 2-D), as outlined in (Vellido et al 2011). GTM is preferred to the more standard SOM in this study because its probabilistic definition ensures the convergence towards a minimum of a properly defined error function, as well as the adaptive estimation of the optimum values of some of its variables. Formally, GTM is a non-linear latent variable model (Bishop 1998) of the manifold learning family and, as such, data are modelled through a low-dimensional manifold embedded in the data space. Such manifold is defined as a mesh whose knots are the centers of probability distributions (usually Gaussians) that become prototype representatives of groups of data. These prototypes are cluster centers and also the building elements of a mixture of distributions. In different variants, GTM has been used for missing data imputation (Vellido et al 2010) , outlier detection (Tosi and Vellido 2013), or time series analysis (Tosi et al 2014), as well as applied in areas such as medicine (Cruz and Vellido 2011)or e-learning (Etchells et al 2006), amongst others.

The GTM is mathematically defined as a nonlinear mapping from points $\mathbf{u}$ in a low dimensional latent space onto the prototypes $\mathbf{y}$ residing in data space, with a functional form described as: $y = \Phi(u)W$, where Φ is a set of $\mathbf{M}$ basis functions $\Phi(u) = \{ {}_1(u), \ldots, {}_M(u)\}$ and $\mathbf{W}$ is a matrix of adaptive weights that defines an specific mapping. The probability distribution for data point $\mathbf{x}$ in a data space $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ with $\mathbf{x} \in \Re^D$, being generated by a latent point $\mathbf{u}$, is defined as an isotropic Gaussian noise distribution with common inverse variance β, from which the likelihood of the model can be defined. The adaptive parameters of the model ($\mathbf{W}, \beta$) can then be estimated through maximum likelihood using, for instance, the Expectation-Maximization (EM) algorithm (Dempster et al 1977). In order to use GTM for visualization, the relation between each data point $\mathbf{x}$ and each latent space point $\mathbf{u}$ is quantified as a conditional probability $\mathbf{p}(\mathbf{u_k} \mid \mathbf{x_n})$ and its calculation is a by-product of the maximization step of EM. This probability is known as the *responsibility* $\mathbf{r_{kn}}$ of each latent point $\mathbf{u_k}$ for the generation of each data point $\mathbf{x_n}$. Each data point $\mathbf{x_n}$ can therefore be visualized by its assignment to the location in the latent space (to the cluster) where the mode of the corresponding conditional probability is highest, that is: $\hat{\mathbf{u}}_{\text{mode}} = \arg_{\mathbf{u_k}} \mathbf{r_{kn}}$

This type of visualization, known as *mode projection*, was used in the experiments reported in Section 3. The fact that clustering is somehow subordinated to visualization in GTM means that the resulting clustering solution is often too detailed for practical purposes. To overcome this limitation, the well-trodden k-Means algorithm, which, as SOM, has been used for energy use profiling, was used to cluster the prototypes resulting from GTM. This becomes, *de facto*, a two-stage clustering procedure that yields a parsimonious final cluster partition that can be interpreted in terms of the original data variables with the assistance of the GTM visualization maps.

2.4 Definition of reference dwellings

The definition of reference dwellings is then carried out by identifying those which meet two main criteria: first, having the values of relevant variables closest to the values of the centroid of each cluster in terms of Euclidean distance; second, having the monthly EUI (kWh/m^2·month) of gas and electricity closest to the median monthly value of each cluster, also in terms of Euclidean distance.

2.5 Thermal simulation and refinement of the reference dwellings

In order to check the real indoor conditions in the selected reference dwellings and to calculate their related heating and cooling demands, a thermal simulation was performed with Energy Plus software [34]. The adjustment parameters for simulation acceptance were *natural ventilation rate, use of shading devices*, *indoor set point temperature*, *hours/day being at home*, *internal energy demand* (number of electrical appliances, nominal power and artificial lighting) and *use of AC* (in dwellings with AC system).The monitoring indoor temperature, electricity hourly consumption, and outdoor temperature in the representative dwellings (see section 3.8) were chosen as reference for adjusting the energy demand simulation.

2.6. Extrapolation of results for the entire neighbourhood

Extrapolation of results and calculation of disaggregated energy consumption for the whole district were carried out considering a tower of 60 dwellings with the same distribution of types of households as those obtained from the clustering procedure as a pilot. This pilot tower was considered to be an appropriate representative of the 12 towers of the district. A proportional aggregation according to the surface area, of dwellings in each cluster was implemented to estimate the total energy demand of the neighbourhood. We used the so-called weighting coefficient which is the number of buildings of the stock which are represented by each archetype building, as presented in (Mata et al 2014)

3. EXPERIMENTAL RESULTS

3.1. Data collection and pre-processing

A close scrutiny of the data from the 166 surveys, for which 59 have a yearly period of monthly energy data, revealed that only 146 sets of socio-economic data and 51 sets of energy monthly data were complete. As previously explained, aggregation of related questions from surveys was carried out resulting on a dataset of 26 variables for the 146 dwellings. Table 1 shows the complete list of variables extracted from the questionnaires. Finally, and also as previously described, data transformation to a range of [0, 1] was applied to deal with the differences in scale and in categories.

3.2 Selection of influencing variables

Correlation between variables and EUI was calculated for the 51 samples with monthly bills, under the assumption that this result will apply to the rest of samples. The covariance matrix was then calculated over the remaining variables of the 146 dwellings. This resulted in a selection of 13 variables for clustering. From those variables (see Table 1 for coding), five are related to heating use and comfort (X1, X4, X5, X6, and X7), two are related to electricity consumption (X15, X17), four to energy behaviour and awareness (X12, X13, X18, and X24), one to the economic situation (X11), and one to hot water consumption (X22). Note that none of the variables are related to summer comfort, air conditioning, or the kitchen. The reason for this is that only a few dwellings have air conditioning systems (around 25% of the total) and their use is low (as will be confirmed in the next section).

3.3 Two-stage clustering results

The selected data (13 relevant variables from 146 dwellings) underwent a two-stage, fully unsupervised, clustering process in which GTM was first used to obtain a loose data partition into natural groups with a focus on exploratory visualization. For the experiments reported in this study, the visualization grid of GTM latent nodes was fixed to square layouts of 10×10 nodes (i.e., 100 constrained mixture components). Figure 4 (right) shows the 2-D representation of the 146 13-D points on the GTM 2-D visualization space, according to their *mode projection* as described in section 2.3. Each square corresponds to one of the latent points in the 10x10 grid and its relative size corresponds to the ratio of cases (dwellings) assigned to that point. The different square sizes and the empty spaces in some areas are a clear indication that the analysed data have some intrinsic cluster structure. The k-Means algorithm was then applied to the obtained prototypes (the functional images of the latent points) in order to further group the different visualization regions into a specific number of clusters.
The adequate number of clusters must be estimated according to some criterion. In our experiments, we used the *silhouette* index, which provides a succinct graphical representation of how well each data item (a dwelling in our experiments) lies within its cluster. It was first described by (Rousseeuw 1987) and a value near 1 indicates that the data item are assigned to the right cluster whereas a value near -1 indicates that the items should have been assigned to a different one. This index suggested a number of six clusters as the optimal choice, yielding a maximum *silhouette* value of 0.3, which is acceptable taking into account the small dimension of the data matrix (146×13). The distribution of colours in the map of Figure 3(right) reflects the results of this 2nd stage of the clustering process. Importantly in terms of usability and data consistency assessment, the 6 cluster solution is shown to partition the data in mostly self-contained independent map areas with minimal cluster overlapping and very little cluster discontinuity.

3.4 Reference maps and feature-Based Interpretation

In order to make practical sense of these 6 clusters, an interpretation on the basis of the original 13 variables is needed. GTM provides such interpretation through the so-called *reference maps*, shown in Figure 3(left). Each reference map displays a variable's relative contribution over the representation map (and thus how it contributes to the clustering solution as a whole). Given that features, whose reference maps that exhibit a "regular" distribution over the map are likely to induce a similarly regular structure on the GTM visualization map, this property could be used to select a subset of data features that are the most relevant at determining the shape of the map and consequently the clustering of groups of households.

The reference maps are coded in grey-scale, from black (lowest values) to white (highest values), allowing a straightforward interpretation. It can be seen, for instance, that the reference map relative to variable X1 (space heating), which reflects whether the household has central space heating system (yes) or stoves (no), is neatly partitioned vertically according to low/high values. Its correspondence with the GTM map in Figure 3(right) reveals that almost all households with only stoves (electric or gas) are located on the left-hand side of the map, which corresponds strictly to cluster 3 (in yellow). This cluster also seems to be neatly characterized by low values (black colour on the left) of X4 and X5 (*months of heating* and *heating schedule*) and, therefore, variable X1dominates the first level of the partition. Similar exploratory interpretations can be carried for other clusters using the reference maps

Fig. 3 a)(right): GTM 10×10 cluster map: blue)cluster 0, light blue)cluster 1, green) cluster 2, yellow) cluster 3, red)cluster 4, brown)cluster 5.**b)** (left) reference maps of single 13 variables.

Through visual interpretation, we can say that only some variables seem to significantly contribute to explaining the final cluster partition in Figure 3 (right). The selected 9 variables are: *X1=Space Heating*, *X4=Num. month heating, X5= Heating Schedule, X7 = Num. rooms unheated, X11= Total month Income, X13=Use of awnings in summer, X17= Number appliances*, *X18=Switch off appliances by night,* and *X22= Time shower*.

3.5 Characterization of groups of households

Beyond visual exploration, we would like to quantify the specificity of the clusters according to actual characteristics. For this, we could display, in the categorical variables, the percentage of households that exhibit a specific categorical value, while, for real-valued variables, we could display the quartile values of their distribution in each of the identified clusters identified (see Figure 6).

Beyond the subset of relevant variables considered in the previous section, there are others that are unevenly distributed over the different clusters. Due to this different distribution over groups, they could also be considered when trying to understand the characteristics of groups. This second set of variables could be denoted as complementary features of the different groups of households. It includes *X3 = Quality of windows, X8= AC system, X9= Use of AC*, *X26= Time at home* (see Figure 7 top). Additionally, the distribution of the yearly aggregated monthly bills (gas and electricity collected in some of the households), can be obtained for each cluster (Figure 6 bottom). A description that summarises the mean values for the predicted variables and the energy consumption in each cluster is shown, in Table 2, as a characterization of the groups of households. In this table, we can see that 12 variables have finally been selected (both representative and complementary) as the main representative to characterize the different clusters. It should be noted that, rather than selecting X1, which can only discriminate those households without centralized space heating (see Figure 3 left), we selected X2 as representative variable (it is in fact highly correlated to X1) because households without centralized space heating will be determinant in making decisions about energy improvements at building level due to their low comfort, consumption and low incomes (they represent 10% of the whole analysed sample). In the practical description of groups we have also included the number of people living at home, but only to complement the selected information.

Some relevant conclusions can also be extracted from the rejected variables. For instance, the number of efficient lamps and the number of people are not directly related to the electricity or the gas consumption, as the variation in these variables is not coherent with their equivalent consumption. (Most clusters have the same median and range of number of people at home (see Figure 7 bottom). As outlined in Table 2, the average income per family is split into two main groups, with incomes around 2,300€ and 3.000€ per month (in year 2008). Furthermore, when asked about the possibility of engaging in energy efficiency projects that entailed the introduction of a fee, the vast majority of households (97%) rejected the idea. They are, however, prone to accept the measures if no extra payment is required. The majority of investments carried out in the past in the analysed area focused on improving the quality of windows (around of 40% of windows are double glazed, with the exception of Cluster 3) and installing natural gas heating systems (around 85%, again with the exception of Cluster 3).

In relation to energy consumption, we can affirm that households of people spending long periods at home (Cluster 5) and small families spending very limited time at home (Cluster 3) are representative of the group with lowest energy consumption rates (both gas and electricity). In contrast, families with high comfort and medium time spent at home (Cluster 4) represent the highest energy consumption, followed by Cluster 2, Cluster 1, and Cluster 0. These last three clusters show very similar gas and electricity consumption rates, due to the similarities in the time spent at home, income and comfort. The small increase in energy consumption observed for Cluster 2 over clusters 0 and 1 is due to small variations in combinations of these variables related to energy use (mainly use of heating, income, and comfort). The big differences observed between Clusters 3 and 5, on one hand, and Cluster 4, on the other, are mainly due to: Firstly, the different type and use of their heating energy systems as well as their thermal comfort; secondly, their different monthly income and number of appliances; and thirdly, their level of energy awareness related to the use of AC, awnings and appliances by night. In contrast, these high differences in energy consumption are not reflected in same differences in energy cost, due to the structure of energy tariffs in Spain, where the fixed terms are very high (especially in electricity). All groups spent around 3% of their monthly income in energy consumption. However, this percentage may increase up to 11% for Cluster 3 and Cluster 5 in months of high energy consumption (winter).

3.6 Differences of energy consumption within clusters

In order to examine the variability in annual gas and electricity consumption within each cluster, the yearly EUI of gas and electricity was normalized and plotted (see Figure 4). The normalization is based on dividing all the EUI of dwellings by the median value in each cluster, thus highlighting the variability and allowing the EUI to be plotted together on the same scale. We consider large variations those with values over 1.5 or below 0. 5. Regarding the figures, all clusters have their values corresponding to the first and third quartile included within this variation in the case of gas, with the exception of Cluster 3 and Cluster 4, for which higher ranges of variation in electricity consumption were observed. This means that the clustering procedure has predicted a very good similarity of gas consumption within clusters, and an acceptable one in the case of electricity consumption. This can be explained because it is easier to define variables that explain the use of heating than defining variables that explain the use of electrical appliances adequately.

3.7 Selection of reference dwellings

The next step of the analysis consists on the characterization of the reference dwellings that can represent each cluster. This task was carried out by selecting those dwellings that complied with the two criteria previously defined in section 2.6 (minimum Euclidean distance of all relevant variables and monthly energy consumption to the centroid of each cluster). An example for Cluster 0 of visualization of the monthly EUI values of gas and electricity, with their corresponding median value is shown in Figure 5. The dashed grey line corresponds to the median value, while the red line corresponds to the reference building. It can be seen that the monthly distance between the median and the selected reference dwelling is acceptable, as significant differences are identified only in two months. According to this procedure, the ID of the six selected reference dwellings, together with their related energy consumption, were obtained (see Table 2). The results of corresponding variables of each reference dwelling are summarized in Table 3.

Fig. 4 Box plots of the normalized EUI of gas (left) and electricity (right) in each cluster.

Fig. 5 Monthly gas and electricity consumption of households in Cluster 0.

Table 2 Yearly EUI of gas and electricity, and energy thermal demand for cooling and heating of reference buildings

GROUP	**% Dwellings**	**Surface (m^2)**	**ID Reference Dwelling**	**Gas ($kWh/yr\cdot m^2$)**	**Electricity ($kWh/yr\cdot m^2$)**	**Heating Demand ($kWh/m^2\cdot yr$)4***	**Cooling Demand ($kWh/m^2\cdot yr$)5***
C0	21%	89	38	57	27	38	0,0
C1	19%	89	44	53	46	36,5	-16,0
C2	21%	65,5	139	64	42	44,3	-10,5
C3	10%	65,5	123	32	23	21,23	0,0
C4	11%	89	32	93	68	63,20	-17,6

4* heating demand is obtained by simulating with a ideal heating system that covers the defined equivalent winter T set point
5* cooling demand is obtained by simulating with an ideal AC system that covers the set point T when the household has air conditioned system

3.8 Energy simulation of the reference buildings of each cluster

We then proceeded to the estimation of the parameters for the calculation of the hourly thermal demand of the reference dwellings. A summary of those simulation parameters is shown in Table 4. Some of these parameters, like *U values*, *infiltration rate* (ACH), and *type of windows* were obtained from *ad-hoc* tests and measures (a hole in walls of two unoccupied dwellings; visits to check the quality of windows; and blowing door tests were conducted). Other parameters such as number of people, internal gains, use of night ventilation, comfort temperature and schedule time at home were estimated according to relevant variables of each cluster. These last group of parameters, in the case of our study, were refined with data obtained from a monitoring period (summer) of indoor temperature and hourly electricity consumption in the reference dwellings (see next section), in order to check the reliability of our parameters. In the energy model created with Energy Plus simulation software (Energyplus Software 2009), each zone represents one room of the dwelling. The measurements of *indoor T* are also carried out in one bedroom and in the dining room. The rest of the dwellings in the building were used only for shadowing simulation purposes, as well as the rest of the buildings in the district. Figure 9 displays views of the computer modelled buildings.

Figure 6.a) top) Distribution of relevant categorical variables for each cluster; **b)**bottom) max., min., median, 25% and 75% quartiles of continuous variables within each cluster.

Fig. 7a). Top) Distribution of predicted categorical variables within each cluster; **b)**Bottom) max., min., median, 25% and 75% quartiles of EUI of gas and electricity.

Table 3 Relevant variables which define the different features of the groups of households

GROUP	% dwellings	Description	Total income (€/month)	Gas (kWh/yr·m²)	Electricity (kWh/yr·m²)	Average Energy costs (€/Month)*	Stand by Appliances off by night	Use of Awnings
		X26, X7, X2, X4, X5, X17, X8, X9	X11				X18	X13
C0	21%	Family with around 3-4 members (middle age adults) . Medium-High time at home and thermal comfort. Centralized heating. Medium heating period with medium use. High number appliances. Few of them with AC and low use	3190±1580	57±32	34±13	92,52	MOSTLY YES (70%)	MOST OF THEM ALWAYS (70% always)
C1	19%	Family with 2-3 members. Medium time at home and thermal comfort. Centralized heating. Highest heating period with high use. High number appliances. Some of them with AC with low use	3000±1327	60±39	32±20	91,55	MIXED (60% no)	ALWAYS (90%)
C2	21%	family with 3-4 members (middle age adults). Little time at home. Medium-high thermal comfort and appliances. Central heating. High period of heating with medium use. No AC	2900±1814	64±33	34±18	94,96	MOSTLY YES (70%)	NEVER (100%)
C3	10%	Young couple or little family (2-3 members), little time at home. Heating with gas stoves. Low thermal comfort. No AC. Low number of appliances. Short period heating	2300±1100	34±12	21±20	71,41	MOSTLY YES (65%)	MIXED (50% never, 50% always)
C4	11%	Family with around 3 members. Medium time at home. High thermal comfort and appliances. Central Heating. Short period of heating, but intensive use. With AC and some use.	3150±950	93±11	29±22	100,03	NO	NEVER (100%)
C5	19%	Family with 2-3 members (mainly Elderly, or family with elderly). Long time at home. Central heating. Medium-low thermal comfort. Low period of heating with medium use. Low-medium number of appliances. Some of them with AC but low use.	2400±1100	37±11	25±17	76,48	YES (90%)	ALWAYS (90%)

3.8.1 Refinement of thermal simulations

As outlined above, the adjustment of an overall infiltration rate for each dwelling was defined according to blowing door tests (see Table 4: Infiltration rate column). For *night ventilation*, *time at home*, *internal gains* and *heating period* definition, the indoor temperature was simulated (without HVAC system) against the real *indoor T* during the monitoring period. Within this refinement process, the outdoor temperature for the simulation was taken from an automated meteorological station installed on the roof, and the indoor real temperature corresponds to the measured temperature in the dining room. Results for both simulated and real indoor temperature (dining room) for the reference dwelling of Cluster 3 (low consumption) and Cluster 4 (high consumption) in August are shown in Figure 10. These results show that the relative error was above 10% only the 0.94% of the total monitoring time, which reaches the acceptable accuracy threshold in adjustment of an energy simulation model defined by (AHSRAE 1999). To reach this accuracy, the infiltration rate in Cluster 3 reference dwelling was defined as 1,5 ACH for June and September (from 21h to 10h), 3ACH for July (21h to 10h), 1,5ACH for August half time unoccupied. This adjustment procedure was applied also in Cluster 4 reference and results also showed an error above 10% in only 1,94% of total hours. Small variations of these conditions were selected for the rest of reference dwellings according to their related influencing variables as shown in Table 4. An extended simulation based on results obtained from the monitoring period was performed in order to obtain the energy demands for the whole year (also in winter).

Figure 10 shows results of typical Cluster 3 non-air conditioned dwelling, where the hourly indoor temperature in July is around 26ºC during the day and around 24ºC at night (average in summer is 25.5 ºC during the day and 23 ºC during the night). In winter, the average of measured temperature is 18ºC for the complete day and 14ºC over night (these types of dwellings have butane gas or electric stoves as heating systems). Thermal energy demands for dwellings considering these comfort conditions are 21,23kWh/m^2·yr for heating (see Table 3). Small variations when simulating the Cluster 5 reference dwelling were obtained. In the case of Cluster 0, the higher heating demand is due to the use of a centralized heating system that allows the inhabitants to get better thermal set points and comfort. Thermal energy demands for Cluster 4 representative dwelling are 77,4kWh/m^2·yr for heating and -17,58Kwh/m^2·yr for cooling, as shown in Table 3. Conditions of less comfort, especially in winter, were assumed for Clusters 2 and 1.

3.9 Disaggregation of consumption and stock aggregation for the whole district

To obtain energy consumption and demand for the whole district we assume that results obtained in the pilot tower may be extrapolated, after refinement, to the whole district. Taking the percentage of dwellings in each cluster together with their respective energy demands, the total energy demand of the tower was calculated. Then, some assumptions were made in order to disaggregate the gas and electricity consumption: for the gas stoves and gas heater (centralized), performance were assumed to be, in turn, 75%, and 79%, and for split units, they were assumed to be 111,6%, according to the official annex document of the national energy certification (Salmerón et al 2009; CTE 1999). Results are shown in Table 5.

.

Fig. 8. Computer Model. Top left) complete district, top right) complete building (15 stories, 60 dwellings) reference dwellings of each cluster in grey shades, bottom left) WO dwelling. Bottom right) East-oriented dwelling.

Fig. 9 Simulation results for indoor T for the C3 reference non Air-Conditioned Dwelling's Calibrated Model in August (top), and the Cluster 4 reference air conditioned dwelling, also in August (bottom).

4. CONCLUSIONS

4.1. Analysis of the current situation

In summary, we can affirm that in dwellings with only gas stoves or electric heaters for space heating, thermal conditions in summer are not as bad as expected prior to analysis. In visits, even during some especially hot summer days, indoor conditions were found to be comfortable. Even though the building's envelope characteristics were in general rather bad (low insulation level), it was found that the building's cooling loads in the dwelling are extremely low, and heating loads are not very high, as a consequence of climate's characteristics in the area, natural ventilation potential, passive techniques (such as orientation and canopies), and low internal gains. Instead, conditions of discomfort were found in winter. This is a rather interesting finding, as there are a lot of similar social-housing districts in the country, all with similar shapes, surroundings (in the outskirts of cities, free of the obstructions and, as a result, exposed to the wind) and envelope's characteristics.

Table 4 Simulation parameters that define the different reference buildings of each group of households

	T comfort cooling[6]*	**T comfort heating[6]***	**Windows quality. Equivalent U $(W/m^2 \cdot K)$**	**Use of AC**	**Natural ventilation**	**Infiltration rate ACH*** (1/h)**	**Time at home**	**N**	**Internal gains (W/m^2)**	**Lighting internal gains (W/m^2)**	**Heating period**
C0	23º by night, 26ºC by day	18ºC by night, 20ºC by day	50% double glass (U=3,24), 50% single glass (U=6,14)	no	Winter: 0.4ACH from 0h to 24h. Spring and Autumn: 1.5ACH from 21h to 10h. July: 3ACH from 21h to 10h. August: 1.5ACH from 0h to 24h.	0,4	11 months a year: 0.015 people/m2 from 7:00 to 10:00, and 16:00 to 24:00	3	55	Winter: 1.5 from 8:30 to 9:30 and from 17:30 to 24:00. Summer: 1.2 from 20:00 to 24:00.	evening, mid-lunch and breakfast
C1	23º by night, 24ºC by day	17ºC by night, 19ºC by day	55% double glass (U=3,24), 45% single glass (U=6,14)	medium	Winter: 0.4 ACH from 0h to 24h. Spring and Autumn: 0.9ACH from 18:00 to 24:00. Summer: 2ACH from 01:00 to 6:00.	0,4	January to December: 0.02 people/m2 from 8:30 to 9:30 and from 15:00 to 24:00	4	55	Winter: 1.5 from 8:30 to 9:30 and from 17:30 to 24:00. Summer: 1.2 from 20:00 to 24:00.	Mid-lunch and evening
C2	23º by night, 24,5ºC by day	18ºC by night, 21ºC by day	50% double glass (U=3,24), 50% single glass (U=6,14)	low	Winter: 0.4 ACH from 0h to 24h. Spring and Autumn: 0.9ACH from 18:00 to 24:00. Summer: 2ACH from 01:00 to 6:00.	0,4	11 months a year: 0.015 people/m2 from 7:00 to 10:00, and 20:00 to 24:00	3	60	Winter: 1.5 from 8:30 to 9:30 and from 17:30 to 24:00. Summer: 1.2 from 20:00 to 24:00.	only in the evening (from 20h-24h in week days, and from 18h-24h week end)
C3	24º by night, 26ºC by day	15ºC by night, 18ºC by day	15% double glass (U=3,24), 85% single glass (U=6,14)	no	Winter: 0.6ACH from 0h to 24h. Spring and Autumn: 1.5ACH from 21h to 10h.July: 3ACH from 21h to 10h. August: 1.5ACH from 0h to 24h.	0,6	11 months a year: 0.015 people/m2 from 7:00 to 10:00, and 20:00 to 24:00	2	45	Winter: 2.4 from 7:00-9:00 and 4 from 17:30-24:00. Spring, Summer and Autumn (except August): 2.4 from 7:00 – 8:00, and 3.2 from 20:00-24:00.	only in the evening (week days, and from 18h-24h week end)
C4	23º by night, 24ºC by day	18ºC by night, 23.5ºC by day	45% double glass (U=3,24), 55% single glass (U=6,14)	medium	Winter: 0.5ACH from 0h to 24h. Spring and Autumn: 0.9ACH from 18:00 to 24:00. Summer: 2ACH from 01:00 to 6:00.	0,5	January to December: 0.02 people/m2 from 8:30 to 9:30 and from 15:00 to 24:00	3,5	65	Winter: 2.4 from 7:00-9:00 and 4 from 17:30-24:00. Spring, Summer and Autumn (except August): 2.4 from 7:00 – 8:00, and 3.2 from 20:00-24:00.	Mid-lunch and evening
C5	24º by night, 26ºC by day	15ºC by night, 19ºC by day	50% double glass , 50% single glass (U=6,14)	no	Winter: 0.4ACH from 0h to 24h. Spring and Autumn: 1.5ACH from 21h to 10h.July: 3ACH from 21h to 10h. August: 1.5ACH from 0h to 24h.	0,4	September to July and September to December: 0.01 people/m2 from 7:00 to 23:00.	2,5	45	Winter: 2.4 from 7:00-9:00 and 4 from 17:30-24:00. Rest of year(except August): 2.4 from 7:00 – 8:00, and 3.2 from 20:00-24:00.	All day (from 8h to 24h)

Table 5 aggregated results for the whole pilot tower and district.

	number dwellings	Yearly heating consumption (Mwh/yr)	HW + kitchen (Mwh/yr)	Cooling consumption (Mwh/yr)	Appliances + light (MWh/yr)
C0	13	53,7	10,2	0,0	30,3
C1	11	46,7	7,1	14,7	32,0
C2	12	43,9	6,4	7,5	25,5
C3	6	11,1	1,5	0,0	9,0
C4	7	46,8	7,9	9,3	30,6
C5	11	22,6	5,1	0,0	20,2
Tower	60	225	38,2	31	148
Whole district	720	2696	457,9	378	1771
%		51%	9%	7%	33%

As consequence, results of energy disaggregation are quite different than expected: 29% of the residences consumed in similar patterns to the two non AC dwellings, which was originally supposed to be one of the minor energy consumers in the district. These dwellings are often occupied by one or two elderly people (high time at home), little families, or couples spending little time at home, with low incomes and sacrificing their thermal comfort in winter to not to spend too much money in heating their homes.

On the other hand, dwellings with AC and central heating radiators, with higher incomes and with a high number of appliances, represent only 11% of the district. Dwellings with centralized space heating but normal conditions of comfort (around 55-60% of hours per year with comfort), some of them with AC but low use in summer, and with middle-range time spent at home and middle-range incomes, are the main group of the district (Cluster1, Cluster2 and Cluster0 represent 61% of the total). The reference consumption of gas for heating in Barcelona Metropolitan Region is 62.3 kWh /m^2·yr (according to the national energy code [37]. The consumption in 89% of the analysed dwellings is lower than this reference value. However, in the minority of households that consume more, this difference is considerable: up to 35%. In these dwellings, the greatest consumption is due to heating, followed by electricity consumption due to light and appliances, which represents 29% to 41% of total energy consumption. These results also strongly correlate with the level of income of the families, since their 36% of difference in monthly income (between Clusters 3 and 4) is enough to be the cause of big changes in the type of space heating systems and in number and use of appliances.

4.2 Decision making for the implementation of energy improvement actions

It can be concluded that the methodology used in this study is relatively simple and reliable, as intended. The clustering of data obtained from surveys, in combination with refined simulation models, allowed the evaluation of the current situation and of the impact of tenants' behaviour in a realistic way. The increase in time effort, as compared to simulations based on standard characterization of dwellings, is compensated by the higher quality of results in to the objective of understanding the real situation. The proposed methodology is useful in evaluating the possibilities of implementing a real project of refurbishment in existing districts, where the different impacts in the different groups of tenants could be a key factor in decision making, and where the big differences (almost a factor of 2.5) in energy consumption between dwellings due to socio-economical aspects is also a relevant aspect to take into account.

Is there any room for action in these types of neighbourhoods?. Yes, provide any improvement that implies a certain level of economic investment should consider finance mechanisms that help make costs affordable, offering subsidies to the poorest families. Innovative financial measures based on sharing of investments costs according to the potential of economic savings, or on the level of energy consumption should be considered in designing of theses financial schemes. In future research, other actions beyond traditional isolation systems, such as energy management control systems, boiler replacement, micro-generation systems, solar heat water systems, or freezer replacement, will be considered to obtain improvements with low or even none economical cost. It should also be considered that improvements in heating systems and in thermal comfort might lead to an increase of the energy consumption in households with less comfort, so any technology to be implemented should offset this increase through a corresponding increase in energy efficiency or renewable energy contribution. In this sense the effect of the phenomenon of "heat theft" should be also considered, as it is showed in (Dall'O 2014) where it has been proven that is an important issue which can provide the possibility of reducing to zero the quota of expenses from consumption whilst still benefiting from comfortable temperatures

References

American Society of Heating, Refrigerating and Air-Conditioning Engineers. (ASHRAE). (1999). Handbook of HVAC Applications. Chapter 38: 8-9.

Bishop, C, Svensén,M, and.Williams, C. K. I, (1998). GTM: The Generative Topographic Mapping. Neural Computing, 10(1):215–234.

Bishop, C.M., (1998). Latent variable models. In Learning in Graphical Models. Springer Netherlands, 371-403.

Chicco, G., 2012. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 42, 68–80. doi:10.1016/j.energy.2011.12.031

Chicco, G., Napoli, R., Postolache, P., Scutariu, M., Toader, C., (2003). Customer characterization options for improving the tariff offer. IEEE Transactions on Power Systems 18, 381–387. doi:10.1109/TPWRS.2002.807085

Código Técnico de la Edificación (CTE), (1999). Ley 38/1999 de 5 de Noviembre, de Ordenación de Ordenación de la Edificación (LOE). Ministerio de Indústria. Spain.

Cruz, R., Vellido, A., (2011). Semi-supervised analysis of human brain tumours from partially labelled MRS information, using manifold learning models. International Journal of Neural Systems 21, 17-29.

Dall'O', G., Sarto, L., (2014). Individual metering of energy in existing buildings: potential and critical aspects. Energy Efficiency 7, 467–476. doi:10.1007/s12053-013-9235-5

Dempster A.P., Laird N.M., Rubin D.B., (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal Royal Statistical Society B 39, 1–38.

Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings. European Commission, (2010).

EnergyPlus Energy Simulation Software (v6), US Department of Energy, (2009). http://apps1.eere.energy.gov/buildings/energyplus/energyplus_about.cfm

Etchells,T.A., Nebot, A., Vellido, A., Lisboa, P.J.G., and Mugica, F., (2006). Learning what is important: feature selection and rule extraction in a virtual course. Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN 2006), Bruges, Belgium, 401-406.

Goldstein D.B. Eley. (2014). A classification of building energy performance indices. Energy Efficiency 7:353–375. 10.1007/s12053-013-9248-0.

International Energy Agency Energy Conservation in Buildings and Community Systems Programme, (IEA-CBCS) (2010) Annex 33: Energy Related Environmental Impact in Buildings.

International Energy Agency Energy Conservation in Buildings and Community Systems Programme, , (IEA-CBCS) Annex 53, (2013). Total Energy Use in Buildings: Analysis & Evaluation Methods. Final report of ST-C - Statistical analysis and prediction methods, Chapter 10 - statistical analysis of total energy use.

Jain, A. K., (2010). Data clustering: 50 years beyond k-means. Pattern Recognition Letters, 31,651–666.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., Djurovic-Petrovic, M., (2010). A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment 45, 1683–1697. doi:10.1016/j.buildenv.2010.01.021

Kohonen, T., (2001). Self-Organizing Maps. Springer-Verlag, Berlin, 3rd edition.

Li, X., Bowers, C.P., Schnier, T., (2010). Classification of Energy Consumption in Buildings With Outlier Detection. IEEE Transactions on Industrial Electronics 57, 3639–3644. doi:10.1109/TIE.2009.2027926

Lopes, L., Hokoi, S., Miura, H., Shuhei, K., (2005). Energy efficiency and energy savings in Japanese residential buildings—research methodology and surveyed results. Energy and Buildings 37, 698–706. doi:10.1016/j.enbuild.2004.09.019

Loughlin, F.M., Duffy, A., Conlon, M., (2012). Analysing domestic electricity smart metering data using self organising maps, in: Integration of Renewables into the Distribution Grid, CIRED 2012 Workshop. IET, pp. 1–4.

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

Murray, S.N., Walsh, B.P., Kelliher, D., O'Sullivan, D.T.J., (2014). Multi-variable optimization of thermal energy efficiency retrofitting of buildings using static modelling and genetic algorithms – A case study. Building and Environment 75, 98–107. doi:10.1016/j.buildenv.2014.01.011

Nakagami, H., (1996). Lifestyle change and energy use in Japan: household equipment and energy consumption. Energy 21, 1157–1167.

Ourghi, R., Al-Anzi, A., Krarti, M., (2007). A simplified analysis method to predict the impact of shape on annual energy use for office buildings. Energy Conversion and Management 48, 300–305. doi:10.1016/j.enconman.2006.04.011.

Räsänen, T., Ruskanen, J., Kolehmainen, M., (2008). Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. Applied Energy 85, 830–840. doi:10.1016/j.apenergy.2007.10.012

Rousseeuw, Peter J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Computational and Applied Mathematics 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Salmerón, J.M., Cerezuela, A, and Salmerón, R., (2009). Escala de calificación energética para edificios existentes. Cuadernos de Eficiencia Energética. Publicaciones IDAE, Madrid, Spain.

Swan, L., Ugursal, V.I., Beausoleil-Morrison, I., (2009). Implementation of a Canadian residential energy end-use model for assessing new technology impacts. Proceedings of Building Simulation, Glasgow 1429–1436.

Tosi, A., Olier, I., Vellido, A., (2014). Probability ridges and distortion flows: Visualizing multivariate time series using a variational Bayesian manifold learning method. 10th Workshop on Self-Organizing Maps (WSOM 2014). Advances in Intelligent Systems and Computing, Vol.295, 55-64.

Tosi, A., Vellido, A., (2013). Robust cartogram visualization of outliers in manifold learning. Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), Bruges, Belgium, pp.555-560

Tsekouras, G.J., Kotoulas, P.B., Tsirekis, C.D., Dialynas, E.N., Hatziargyriou, N.D., (2008). A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. Electric Power Systems Research 78, 1494–1510. doi:10.1016/j.epsr.2008.01.010

Ueno, T., Sano, F., Saeki, O., Tsuji, K., (2006). Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data. Applied Energy 83, 166–183. doi:10.1016/j.apenergy.2005.02.002

Vellido, A., Martí, E., Comas, J., Rodríguez-Roda, I., and Sabater, F (2010). Exploring the ecological status of human altered streams through Generative Topographic Mapping. Environmental Modelling & Software 22, 1053-1065.

Vellido, A., Martín, J.D., Rossi, F., and Lisboa, P.J.G., (2011). Seeing is believing: The importance of visualization in real-world machine learning applications. Procs. of the 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2011), 219-226.

Yang, Z., Becerik-Gerber, B., (2014). Modeling personalized occupancy profiles for representing long term patterns by using ambient context. Building and Environment 78, 23–35. doi:10.1016/j.buildenv.2014.04.003

Yu, Z. (Jerry), Haghighat, F., Fung, B.C.M., Morofsky, E., Yoshino, H., (2011). A methodology for identifying and improving occupant behaviour in residential buildings. Energy 36, 6596–6608. doi:10.1016/j.energy.2011.09.002.

Yu, Z., Fung, B.C.M., Haghighat, F., Yoshino, H., Morofsky, E., (2011). A systematic procedure to study the influence of occupant behaviour on building energy consumption. Energy and Buildings 43, 1409–1417. doi:10.1016/j.enbuild.2011.02.002

Yu, Z., Haghighat, F., Fung, B.C.M., Yoshino, H., (2010). A decision tree method for building energy demand modelling. Energy and Buildings 42, 1637–1646. doi:10.1016/j.enbuild.2010.04.006

Figure 1
Click here to download Figure: Fig 1.tif

Figure 2
Click here to download Figure: Fig 2.tif

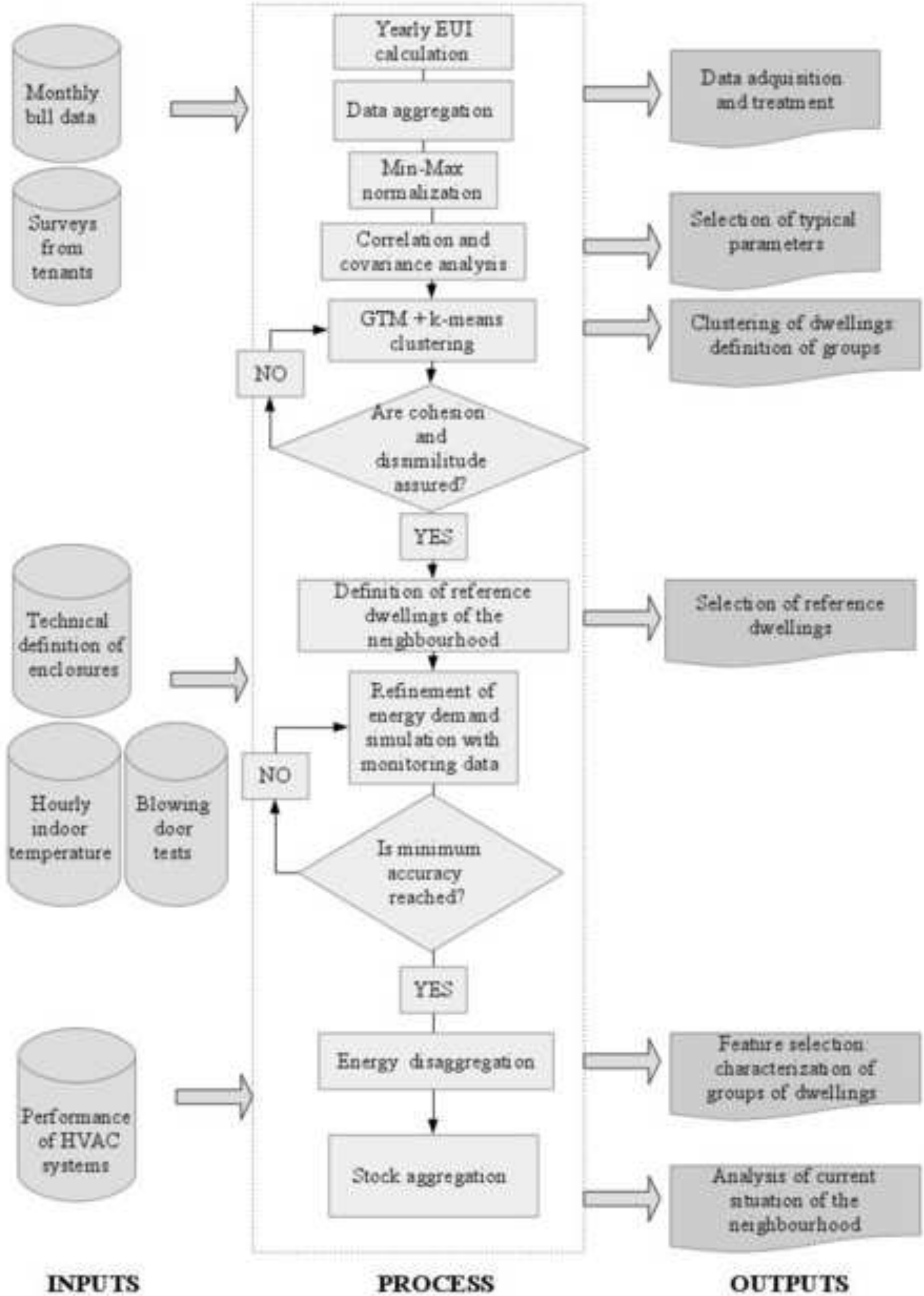


Figure 3
Click here to download Figure: Fig 3.tif

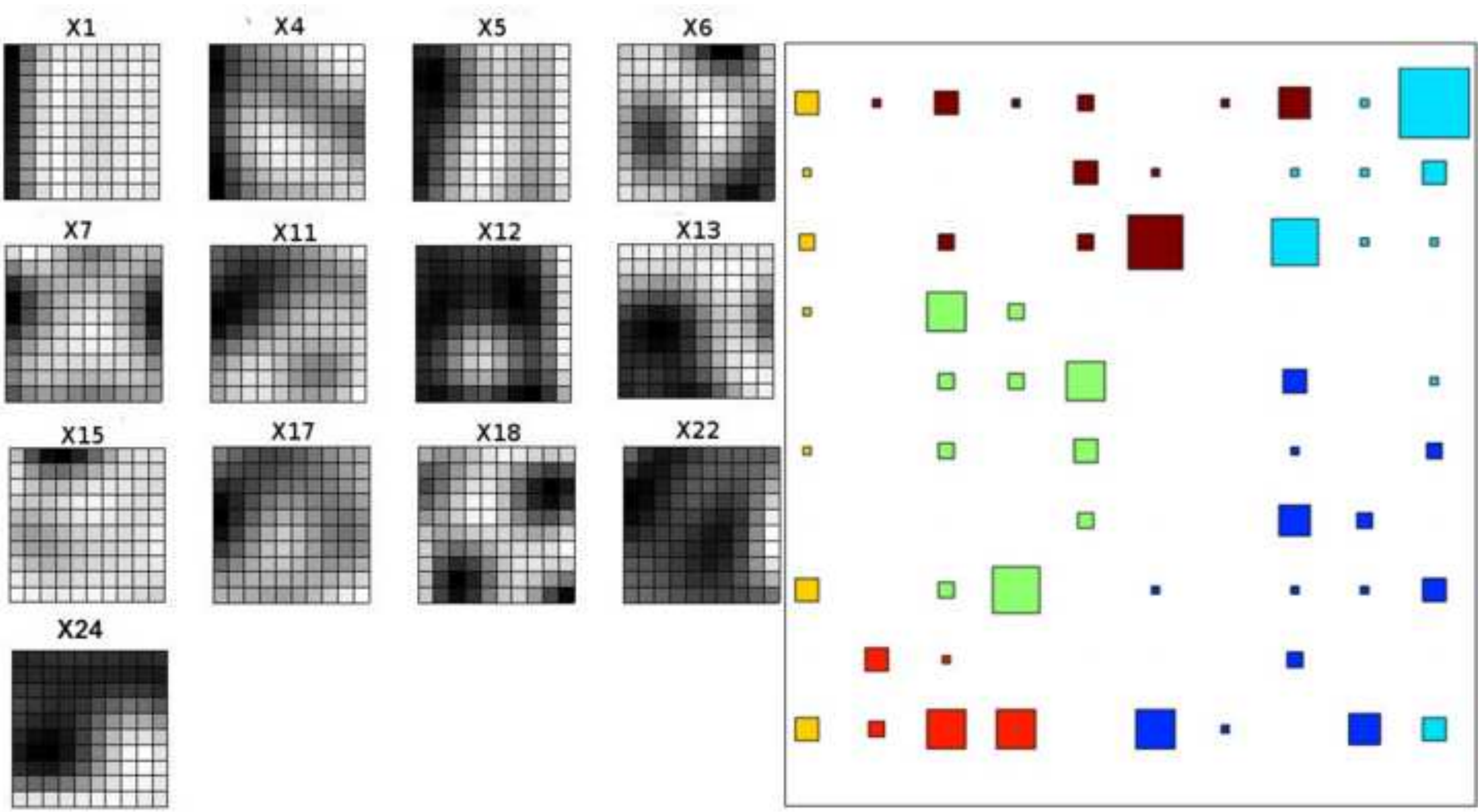
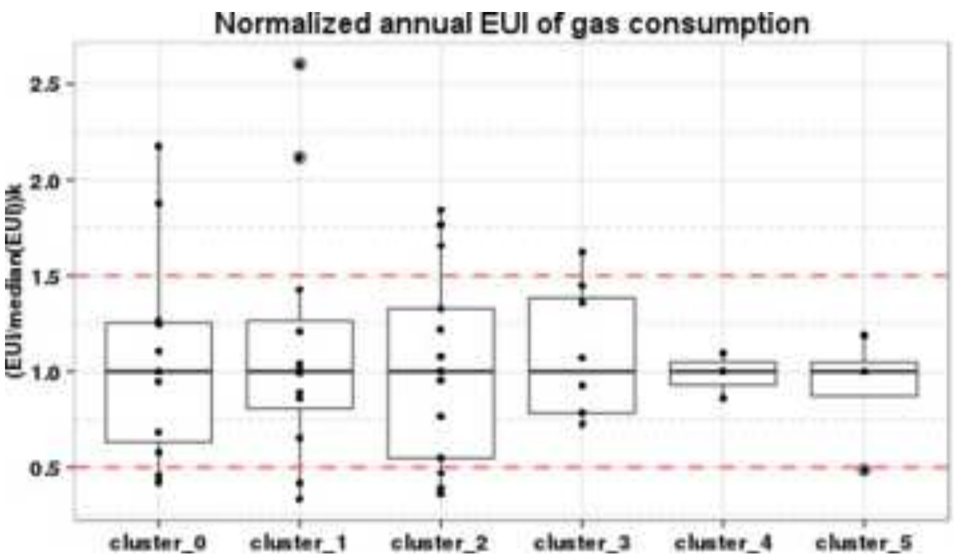


Figure 4

Click here to download Figure: Fig 4.tif



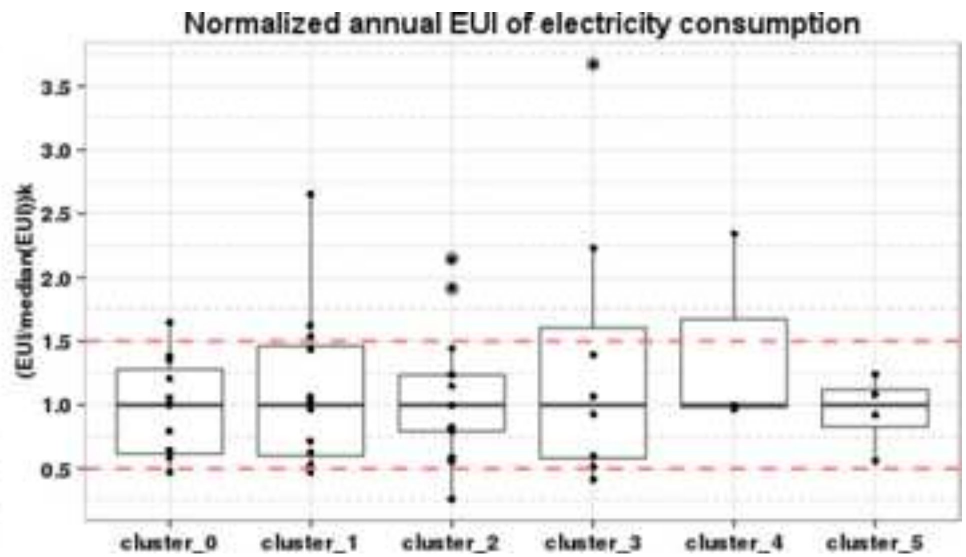


Figure 5
Click here to download Figure: Fig 5.tif

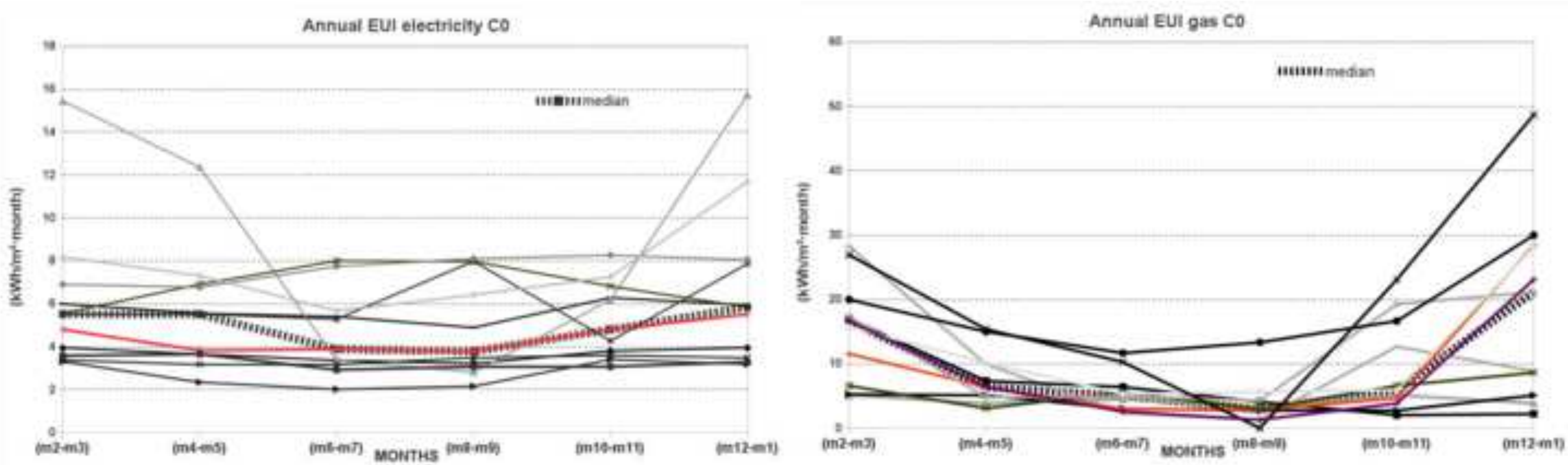


Figure 6a

Click here to download Figure: Fig 6 a.tif

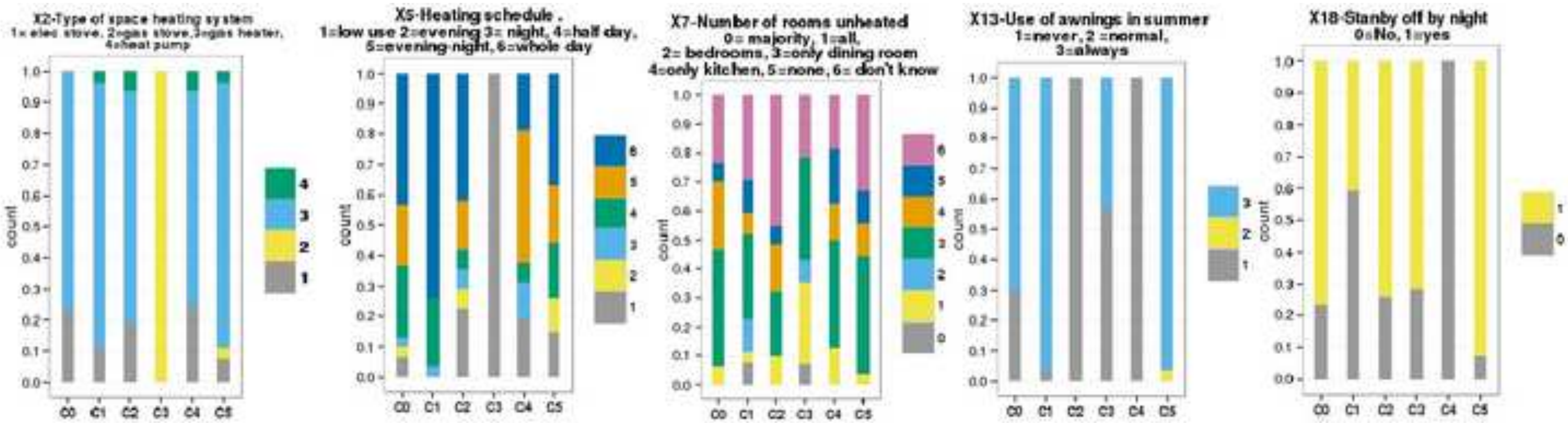


Figure 6b

Click here to download Figure: Fig 6 b.tif

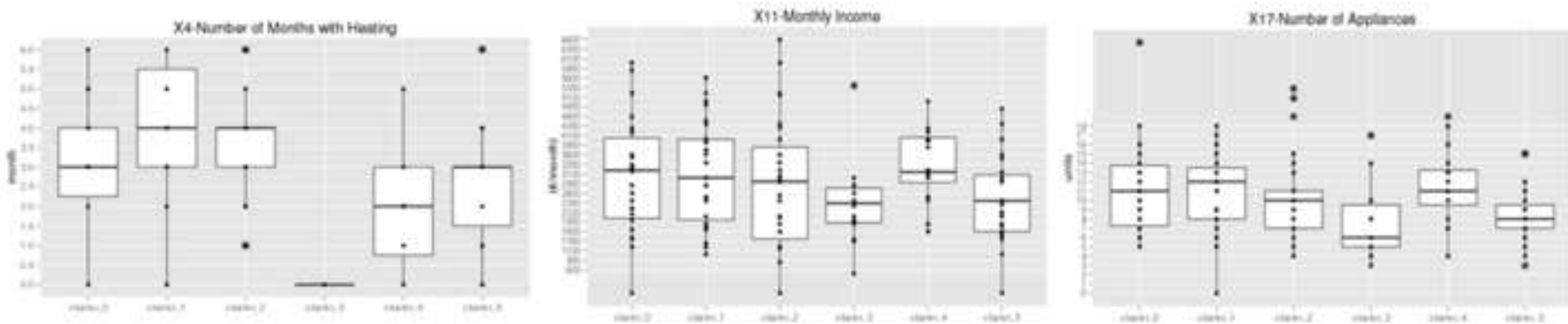


Figure 7
Click here to download Figure: Fig 7.tif

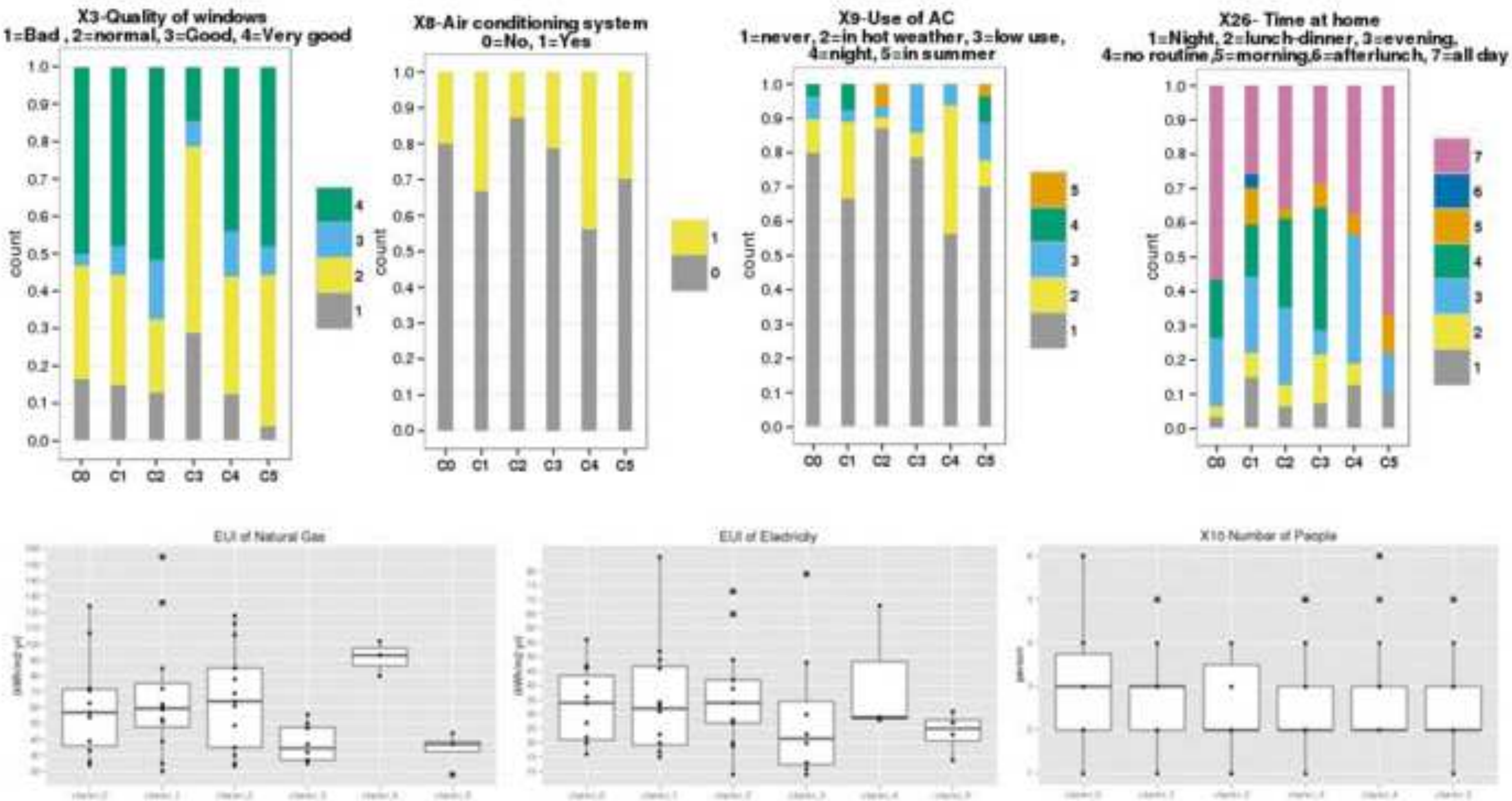


Figure 8
Click here to download Figure: Fig 8.tif

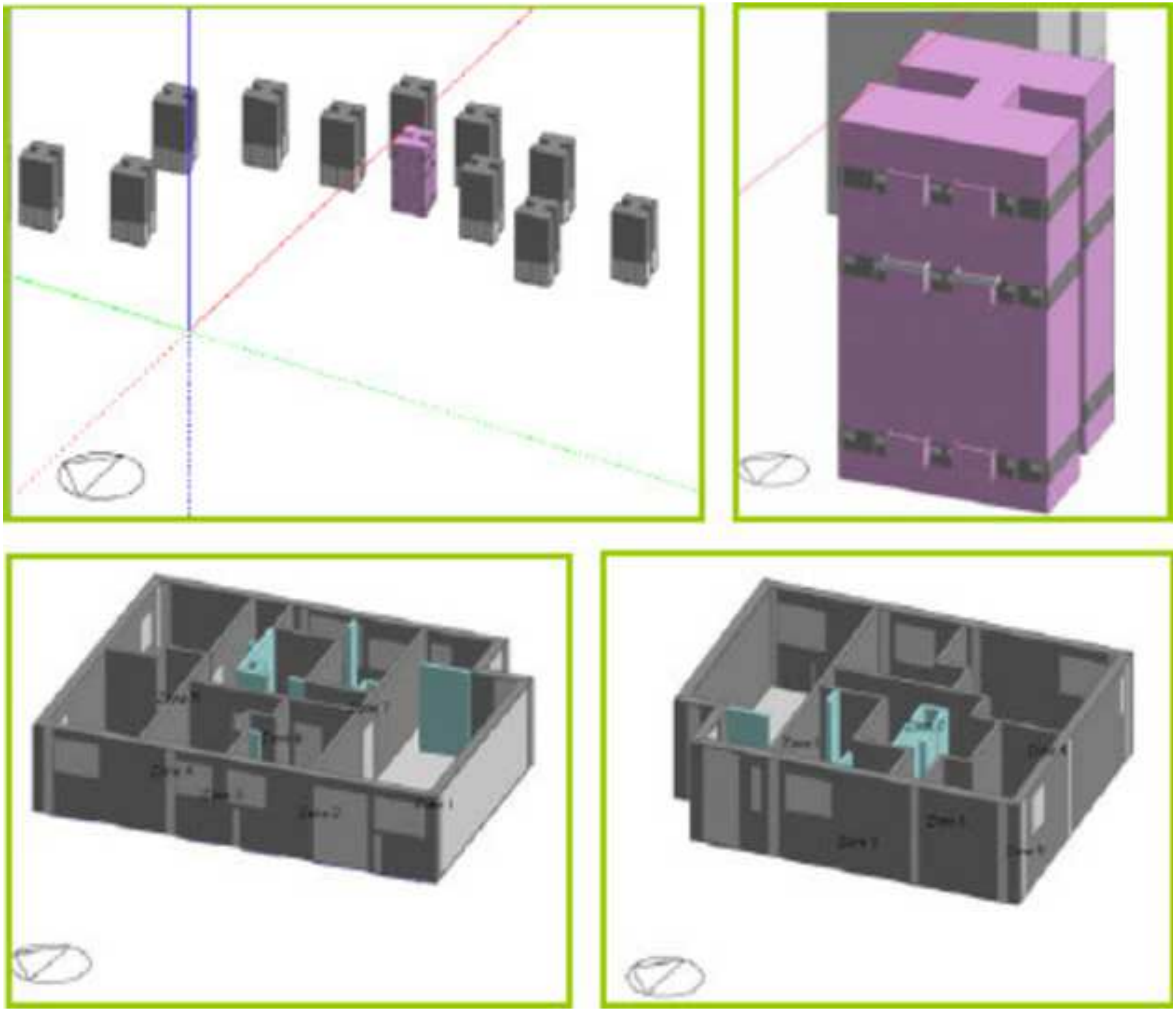


Figure 9
Click here to download Figure: Fig 9.tif