BUILDING ARCHETYPES IN URBAN ENERGY MODELS. A COMPARATIVE CASE STUDY OF DETERMINISTIC AND STATISTICAL METHODS IN ANDORRA.

M.P. Guillaumet[1], P. Borges[1 2], M. Rosas-Casals[1] and O. Travesset-Baro[1 2]
[1] Research Institute for Sustainability Science and Technology – IS.UPC
Universitat Politècnica de Catalunya (UPC), Campus Nord - K2M Building, 08034 Barcelona, Spain
[2] Observatori de la Sostenibilitat d'Andorra (OBSA), Av. Rocafort 21-23, AD600 Sant Julià de Lòria, Andorra

ABSTRACT

In the field of urban building energy models (UBEM), numerous efforts have been made to establish a sound methodology to approach the urban reality of each city. Within this framework, this paper presents a comparative study of two methodologies to determine representative archetypes. We assess these methodologies through a case study in Escaldes-Engordany (the Principality of Andorra), which is representative of medium-sized urban areas in the Pyrenees.

We present a workflow for classifying residential buildings using a statistical approach to the available data. We describe the steps followed to construct the archetypes in order to set the bases for a new methodology that can be replicated in other urban contexts.

We compare our statistical methodology with a deterministic approach based on local expertise. In a deterministic method, buildings are classified according to existing bibliography and two main variables: building use and year of construction. This first classification is compared with that obtained by defining groups using the metered energy consumption and statistical variables of buildings. The statistical method draws upon the local administration's official databases and is complemented with other information sources such as population and housing censuses, regulatory and technical literature, and the monthly metered electricity consumption of all buildings. Data are processed through statistical methods to group them by similar energy consumption behaviour.

Our results show the benefits of the statistical method, as buildings can be characterised not only by their use and constructive technique, but also by their form, dimension location, and occupancy behaviour.

INTRODUCTION

According to recent studies, cities are responsible for between 40 to 70% of total anthropogenic greenhouse gas (GHG) emissions (UN-Habitat, 2011a). Buildings contribute to the vast majority of these emissions. The main measures to reverse the current rise in building energy consumption are focused principally on energy efficiency by improving the building's envelope, renewing technological devices, and by implementing user best practices (UN-Habitat, 2011b). In this context, ensuring local administrations actively participate in national (i.e. governmental) mitigation plans is key in achieving long-term objectives. In order to lead the initiative in areas where cities did not previously interfere, local actors need to develop new capacities and avail of tools to properly guide their actions. Over recent years, a multitude of energy demand models have been developed, which differ in terms of input requirements and the level of disaggregation. The improvement of these models has provided clearer and more precise results on the real building stock situation, allowing future scenarios to be drawn up to assess energy policies. Digital tools such as georeferenced cartographic representations and large existing databases provide new insights to help understand the city's complexity. Urban Building Energy Models (UBEM) apply physical models of heat and mass flows to predict the building's energy consumption (Reinhart, C.F., & Cerezo Davila, C., 2016). To simplify the enormous urban constructive diversity, UBEM use a bottom-up approach archetype to lower the simulation's time and obtain results that are more comprehensive. Current methodologies to define archetypes are generally based on deterministic methods and use building data and existing literature. However, these methods fail to include metered energy consumption and are strongly conditioned by local expertise. Some studies including metered energy data that are based on probabilistic methods have emerged in recent times (Cerezo et al., 2017).

In this paper we seek to establish the workflow necessary to generate archetypes through a hierarchical clustering method, creating groups according to similarities and internal differences. The proposed methodology addresses the shortcoming usually found in UBEMs, i.e., the non-inclusion of the building's real consumption, by including monthly electricity metered consumption in the building's parameters. These groups are later characterised to determine the particularities of each one and to build representative archetypes of a certain number of buildings. We compare the model against a deterministic method, observing the strengths and weaknesses of each method. Finally, we exemplify our methodology with a case study in Escaldes-Engordany (Principality of Andorra), an urban area with almost 15,000 inhabitants, located in a valley in the middle of the Pyrenees mountain range. The urban development is of intermediate scale, with a denser and compact centre, and isolated dwellings with a dispersed occupation of the territory in the periphery. It includes 1,300 buildings that will be catalogued. In order to contrast our methodology with other methodologies, only residential buildings will be evaluated.

METHODOLOGY

Our methodology is built by the workflow used to generate archetypes, which is later used in constructing UBEM. For this reason, we look to establish a methodology that can be easily replicated in other urban centres, giving preference to standardizing and generalizing data processing, and to avoid case-by-case analysis. Given that the available information varies according to the building stock of each municipality, we intend to establish general data-processing criteria that will allow a unified base of the building stock to be determined.

In our case study, we used data from the city of Escaldes-Engordany. It is the second urban development by size in Andorra, with 14,271 inhabitants, and represents 19% of the country's population. The main data on buildings (i.e. building footprints and locations, type of building, uses, year of construction, number of households and number of floors) was taken from local cadastral information (Comú d'Escaldes-Engordany, 2017).

A closer analysis of the city revealed different levels of occupation: Escaldes-Engordany is both a compact city, continuing the trace of Andorra la Vella, the capital of the country, and a diffuse city, expanding along the slopes of the mountains (Figure 1).

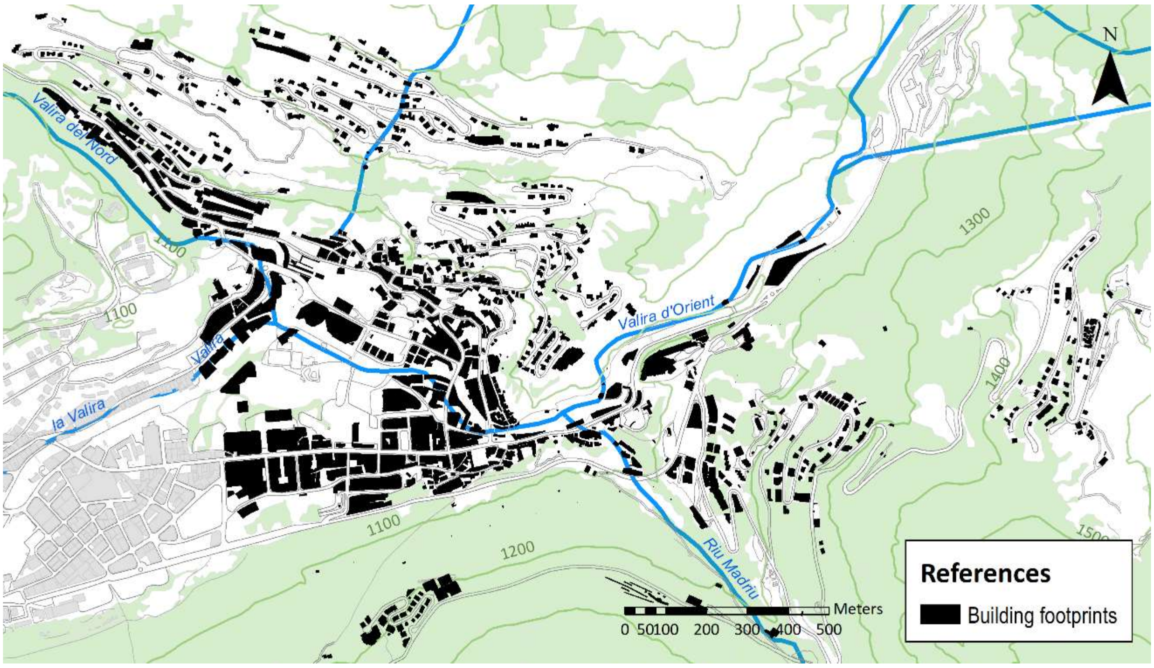

Figure 1. Building footprints of Escaldes-Engordany.

The city has 1,277 registered buildings according to local cadastral information. The constructive footprint is around 350,000 m^2, with a total useful surface of 2.2 km^2. The distribution of the building stock by quantity and occupied square metres shows that 65% of buildings and 62% of the surface are catalogued as residential (Figure 2).

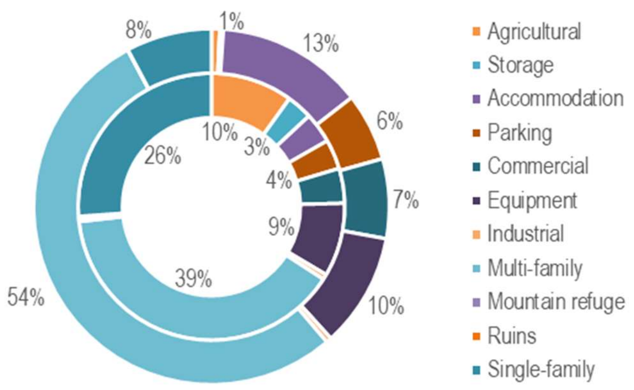

Figure 2. Distribution of building stock by cadastral use.
Percentage of number of buildings (inner ring) and m² (outer ring).

The construction of the model requires the use of aggregated databases in order to concentrate all available information. In our case, official data from the public administration, private companies, and existing publications were included in a database. The main sources of information are the Escaldes-Engordany municipality[1], the local electricity company (FEDA)[1], the Andorran College of Architects (COAA, 2012), and existing academic publications (Borges, 2016; Garrido, 2010). The data extracted from each source, the type of variable, and the year of update are specified in Table 1.

NAME	DATA SOURCE	YEAR OF UPDATE	TYPE OF VARIABLE	DESCRIPTION
ID	Comú d'Escaldes-Engordany	2017	Identifier	Cadastral identification code.
USO_UC	Comú d'Escaldes-Engordany	2017	Qualitative	Cadastral use.
ANO	Comú d'Escaldes-Engordany Borges Martins, 2016	2017 2016	Quantitative	Year of construction.
PLANTA	Comú d'Escaldes-Engordany	2017	Quantitative	Levels of the building.
SUP_TERR	Comú d'Escaldes-Engordany	2017	Quantitative	Building footprint area (m²).
AGRU	Comú d'Escaldes-Engordany	2017	Qualitative	Existence of buildings sharing dividing walls or isolated buildings.
SUP_RES	Comú d'Escaldes-Engordany	2017	Quantitative	Residential area (m²).
CANT_VIV	Comú d'Escaldes-Engordany	2017	Quantitative	Quantity of households per building.
SUP_VIV	Comú d'Escaldes-Engordany	2017	Quantitative	Household area (m²).
SUP_ALOJ	Comú d'Escaldes-Engordany	2017	Quantitative	Accommodation area (m²).
SUP_COM	Comú d'Escaldes-Engordany	2017	Quantitative	Commercial area (m²).
TIPO_CONS	COAA (COAA, 2012)	2012	Qualitative	Construction type that indicates the technic and materials used in the building.
X201X_XX	FEDA	2018	Quantitative	Monthly electricity consumption (2014-2017) (kWh)

Table 1. Sources of information and inputs.

Data are pre-processed to compile existing data in a unified database, with each row corresponding to a building and its characteristics. This step is usually the most time-intensive, as often the data comes from sources that use different systems and names to identify each building. The result is a database and a georeferenced shapefile with the use, surface, year of construction, building adjoining, surfaces, quantity of floors, material properties, and electricity consumption for each building.

The process works with the database's variables to generate new information that can help to understand the energy consumption conditions of the buildings. New parameters are calculated to determine shape relations (i.e. compactness, slenderness, settlement, and abutment) and thermal flows in the building (Serra & Coch, 2001). This step relates to the Building Energy Model (BEM) methodology, defined schematically as the heat exchanges of the building with the environment. Knowledge of the

[1] Personal communication.

new parameters can help to determine which design aspects are more efficient in energy consumption. Other relations such as electric consumption per household or per square meter are calculated in order to compare the behaviour between buildings.

As the building inventory is composed of many cases with their own particularities, modelling these characteristics with low granularity is not possible. The number of types of buildings analysed should be limited to reduce the time and resources required and to make it easier to compress and communicate the outcome. The use of archetypes reduces granularity and allows buildings to be grouped in a smaller number of representative ones.

There are two steps in the group definition: segmentation and characterisation. The variables included to perform the segmentation are cadastral use, building footprint area, total area, quantity of households per building, compactness, slenderness, settlement, abutment, envelope area, total transmittance, surface transmission coefficient, electricity consumption in 2017, electricity consumption per household in 2017, and surface electricity consumption in 2017. A dissimilarity matrix calculation[2] was performed using R Cluster Package through gower metric, in order to include mixed types. Segmentation is based on Ward's method (Ward, 1963) using the hierarchical clustering function[3] included in the R Stats Package. The result is a dendrogram, generated using dendrogram function[4] included in the R Stats Package, that represents successive divisions in groups according to existing differences and similarities (Figure 3).

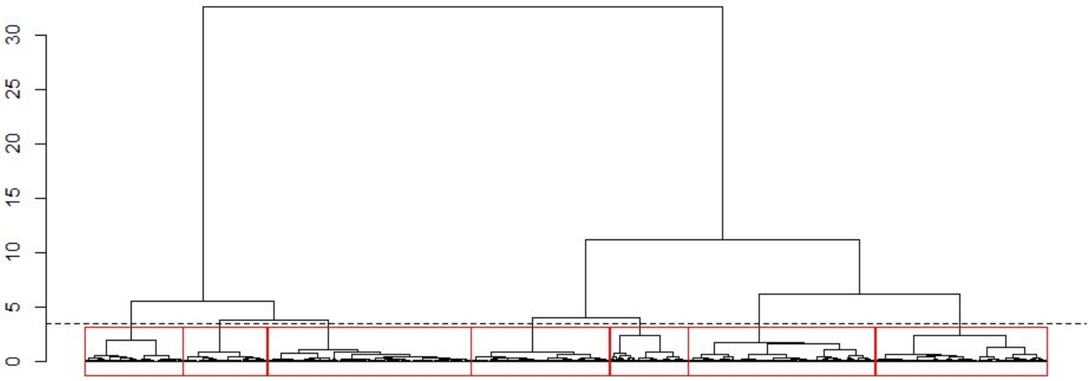

Figure 3. Dendrogram obtained by applying hierarchical clustering to residential buildings.

In characterisation, differences between groups and their particular characteristics are explained. A tree with characteristics similar to the dendrogram is constructed (Figure 4) to determine the dominant parameter in each division. Two steps are applied to determine the dominant parameter. Firstly, through graphic analysis by colouring the values of the variables in the dendrogram, and secondly, by contrasting the numerical values of the resulting groups. Out of the initial 833 cases, one first division segregates single-family (334 cases) and multi-family (399 cases) dwellings. Single-family cases are divided, firstly, into attached and isolated (detached) houses and, subsequently, isolated houses into old and new ones (i.e. less than 30 years old). Multi-family cases have an early division into old and new houses, and subsequently, the old ones are divided again into attached and isolated; here the new ones are segregated into small and large scale. Finally, we obtain 7 well-defined groups seen as archetypes, which represent the total stock of residential buildings.

---

[2] http://stat.ethz.ch/R-manual/R-devel/library/cluster/html/daisy.html

[3] http://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html

[4] http://stat.ethz.ch/R-manual/R-devel/library/stats/html/dendrogram.html

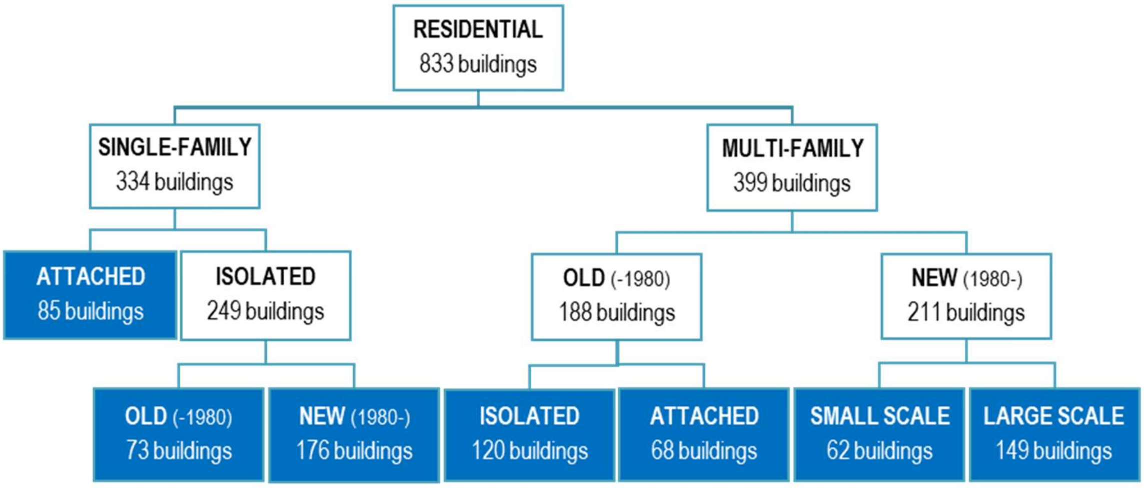

Figure 4. Descriptive division of groups using hierarchical clustering.

Household energy consumption results from factors related to thermal comfort (heating, ventilation and cooling systems) that use different energy types, and other related to electricity consumption, lighting, electric appliances, etc. As it is not possible to obtain real data on how the consumption of each house is segregated and there is no information on the type of energy consumed, we define two key parameters to characterise the behaviour of each group. These parameters are the global transmittance coefficient, which shows the thermal response of the building to outside temperature variation, and the electric consumption per square meter, which is usually related to other than thermal uses. Both parameters are expressed in minimum, maximum, and average values to maintain the diversity of each group.

RESULTS AND DISCUSSION

To evaluate the reliability of the methodology presented in this paper, we compared it with a deterministic methodology applied in the same area two years previously (Borges, 2016). This methodology was based on the method applied in the European project TABULA (Typology Approach for Building Stock Energy Assessment), which created a national residential building classification for 13 European countries according to age, size, and further building parameters (IEE, 2012). This methodology, as in the statistical method presented in this paper, was based on the archetype bottom-up approach. When the deterministic methodology was carried out, limited building information was available in Escaldes-Engordany. This limitation was overcome by exhaustive fieldwork to collect data that affected the building's energy behaviour, which was used in the building stock fragmentation.

Data collected during the fieldwork included the type of building, use, number of floors, and the number of households. This information was then entered and organised in a georeferenced cartographic base provided by the local municipality. It also included the building's year of construction, which unfortunately was only available for some buildings. The residential building stock was then divided according to the building type (i.e., single-family/multi-family and isolated/attached), and the construction period. Four construction periods (i.e., before 1955, 1956–1980, 1981–1995, after 1996) were considered based on a previous study carried out by local experts from the Andorran College of Architects (COAA and SaAS, 2012).

As can be seen in Figure 5, sixteen typologies were obtained, in which 50% of the entire building stock corresponds to only four of them.

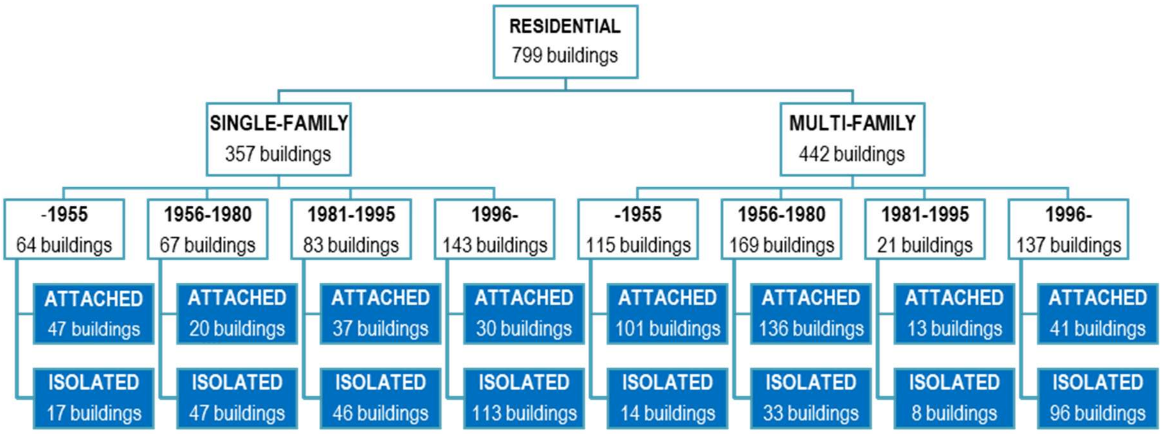

Figure 5. Descriptive division of groups using deterministic method.

In order to compare both methods systematically, a detailed analysis of the process was conducted. Buildings identification (ID), inputs, processing, and obtained results for both methods were compared and presented in Table 2.

	Deterministic method	Statistical method
ID	• ID for the project • Coordinates	• Cadastral ID • Electric contract ID coordinates
Inputs	• Cadastral use • Year of construction • Building adjoining	• Cadastral use • Year of construction • Building adjoining • Surfaces • Levels • Shape relations • Material properties • Transmittance • Electric consumption • ...
Processing	• Pre-processing • Matrix	• Pre-processing • Processing • Hierarchical clustering
Results	16 archetypes • Estimated total energy demand per building	7 archetypes • Values of each variable are known for every building included in the groups • Key values as the global transmittance coefficient and the consumption per square meter

Table 2. Comparison between the deterministic and the statistical methods.

ID

According to identification variables, the statistical method improved on the way the information is connected with other information sources or systems. Even when both are georeferenced, working with the same identification code used by both the municipality and the local electricity company allows them to incorporate outputs generated by the statistical method and use it in the energy planning process.

Inputs

According to input data, the deterministic method requires less variables to conduct the analysis, making it ideal in cases where there is little information available. On the other hand, the statistical method completes some aspects that are not considered before, but introduces major dependency on data availability. For this reason, its application is limited in cases where there is poor data availability for some critical variables such as metered energy consumption.

The main innovation in terms of inputs for the statistical method is that it works with real electricity consumption data, allowing the consumer factor to be incorporated in the equation. This means that buildings can be distinguished not only by their characteristics but also by their inhabitant behaviour. At this point, it should be highlighted that metered data on energy consumption used in this case study refers only to electricity. This introduces a high level of uncertainty in the model, as it is estimated that a significant part of the building stock uses fuel for heating.

Processing

As shown in Table 1, both methods start with a pre-processing that involves assigning a row to each building with all the existing data collected in a line. Cadastral use, year of construction, and building adjoining are included in both methods, but the statistical method adds additional information such as surfaces, levels, shape relations, material properties, transmittance, and electric consumption. Additionally, the statistical method incorporates a second step to calculate relations or new information using the existing one.

The final archetype construction is where the processes vary the most. The deterministic method works with three criteria to fragment the building stock, all of them designated by local expertise. The result is a matrix that groups buildings with common characteristics, without taking into account particularities such as size or material properties. The statistical method uses hierarchical clustering to determine the number of groups and the segmentation. It results in a tree diagram with divisions at different heights based on the distance between the groups. The diagram is drawn by applying multi-criteria analysis to identify the main components of each group and their characteristics. The cutting height is determined by the researcher's judgement according to the accuracy required. More groups mean more accuracy, however it makes harder to identify any differences between them. The workflow of this methodology incorporates machine-learning principles in the standardization and systematisation process, allowing the applicability in new cases and the implementation of unsupervised processes.

Taking into consideration new trends in data collecting and processing for decision-making, the statistical methodology is more appropriate in supporting these processes.

Results

The results from both methodologies are representative archetypes of a large number of buildings to avoid build models for each individual case. The representativeness of archetypes of the real building stock is key in constructing reliable energy models.

On analysis of the residential building fragmentation diagrams for both methods (Figure 4 and Figure 5), the deterministic method seems to be more rigid, as the divisions are defined from the very start of the process. The statistical method works with similitudes and differences between the cases that determine the divisions, emphasising the ones that establish the character of each group. This introduces a certain level of freedom in the process, such as in the number of groups considered and the key variables used in each case.

The energy-related values of each archetype follow different logics in both methodologies. The deterministic method works with an estimated total energy consumption, while the statistical works with global transmittance coefficient and electric consumption per square meter, due to the lack of data to determine the energy required for heating. This situation makes a direct comparison between the results obtained in both methods impossible, as different aspects are calculated. On the other hand, the accuracy of the final values cannot be verified with real data because the information available is not specific enough to complete the task nor is it disaggregated.

CONCLUSION

This paper describes a statistical methodology to determine representative archetypes in the process of implementing an UBEM. In addition, it presents a comparison between this methodology and a deterministic method implemented previously in the same location.

Nowadays, there are a wide variety of methodologies to estimate the building's energy performance with very different purposes, ranging from intuitive categorization of the entire building stock to very precise and detailed simulation models. To determine which is the most appropriate for each situation, the environmental conditions (i.e., data availability, staff training, and financial and computer resources) must be evaluated in detail.

Some methods such as the deterministic approach reviewed in this paper are more appropriate for cases where there is little data available or there are restrictions in data access. The results are more intuitive and less precise, however such methods do not require great staff training or a large investment of resources to implement them.

The application of more complex methodologies such as the statistical method presented here has an important dependency on data access and availability. The precision of the model depends on the quantity and quality of the information available, but the results are more robust and precise than in deterministic methodologies. Given that the accuracy of the archetypes is a key factor in UBEM's generation, the methodology presented here makes a step in this direction, addressing variables that are key in the building's energy behaviour. In this regard, the lack of validation with real data is a limitation of the presented method. The next steps in the process involve quantifying the accuracy of the model in different scales (i.e., city and building scale).

ACKNOWLEDGMENT

This work has been developed as part of a project of the Observatory for the Sustainability of Andorra (OBSA) and the Sustainability Measurement and Modelling Lab of the Universitat Politècnica de Catalunya (SUMMLAB-UPC) focused on building an UBEM in Andorra. The OBSA gratefully acknowledges the Government of the Principality of Andorra for the grant on Andorran issues, APTA0014-AND/2017.

REFERENCES

Borges Martins, P. 2016. *Canvi climàtic i consums energètics del parc edificat a Andorra: el cas d'Escaldes-Engordany.* Master Thesis. Research Institute for Sustainability Science and Technology (ISST.UPC), Universitat Politècnica de Catalunya, Spain.

Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A., & Hajiah, A. 2017. "Comparison of Four Building Archetype Characterization Methods in Urban Building Energy Modelling (UBEM): A Residential Case Study in Kuwait City". *Energy and Buildings*, 154, 321–334.

Cerezo, C., Sokol, J., Reinhart, C., & Al-Mumin, A. (2015). "Three Methods for Characterizing Building Archetypes in Urban Energy Simulation. A Case Study in Kuwait City". *14th Conference of International Building Performance Simulation Association*, 2873–2880.

COAA and SaAS, 2012. *Quantificació del potencial d'eficiència energètica en el sector de l'edificació a Andorra* (Private Report).

Comú d'Escaldes-Engordany, 2017. *Local cadastral information*. Andorran Department of Urban Development.

Garrido Soriano, N. 2010. *Eficiència i certificació energètica d'edificis residencials: avaluació del potencial d'estalvi energètic i d'emissions dels edificis residencials a Catalunya segons diferents escenaris normatius. Definició d'una metodologia de certificació energètica d'edificis residencials, CEPEC, i comparació d'aquesta amb la metodologia oficial CALENER VyPe*. PhD Thesis. Available at: http://www.tdx.cat/handle/10803/21795

Guillaumet, M.P. 2018. *Modelización urbana para la planificación energética: el caso de Escaldes-Engordany*. Master Thesis. University Research Institute for Sustainable Science and Technology (ISST.UPC), Universitat Politècnica de Catalunya, Spain.

IEE, 2012. *Intelligent Energy Europe Project TABULA (Typology Approach for Building Stock Energy Assessment)*. Available at: http://www.episcope.eu [Retrieved March 30, 2018]

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(7), 1683–1697.

Reinhart, C. F., & Cerezo Davila, C. 2016. "Urban Building Energy Modelling - A Review of a Nascent Field". *Building and Environment*, 97, 196–202.

Serra, R., & Coch, H. 2001. *Arquitectura y Energía Natural*.

UN-Habitat, 2011a. *United Nations Human Settlements Programme Cities and Climate Change: Global Report on Human Settlements 2011*.

UN-Habitat, 2011b. Cities and Climate Change: Global Report on Human Settlements 2011 – Policy Directions – ABRIDGED.Ward, J. H., 1963. “Hierarchical Grouping to Optimize an Objective Function”. *Journal of the American Statistical Association*, 58(301), 236–244.