Segmentation d'un parc virtuel de bâtiments par clustering pour la rénovation énergétique

Yunseok Lee, Pierre Boisson, Mathieu Rivallain, Olivier Baverel

▶ **To cite this version:**

Yunseok Lee, Pierre Boisson, Mathieu Rivallain, Olivier Baverel. Segmentation d'un parc virtuel de bâtiments par clustering pour la rénovation énergétique. Conférence IBPSA France, May 2018, Bordeaux, France. ⟨hal-01795043⟩

HAL Id: hal-01795043

https://hal.science/hal-01795043v1

Submitted on 18 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL Authorization

Segmentation d'un parc virtuel de bâtiments par clustering pour la rénovation énergétique

Yunseok Lee*[1], Pierre Boisson[1], Mathieu Rivallain[1], Olivier Baverel[2]

[1] CSTB
84 Avenue Jean Jaurès, 77420 Champs-sur-Marne,
[2] Laboratoire Navier (UMR 8205), École des Ponts ParisTech, ENS Architecture Grenoble
Cité Descartes, 6-8 Avenue Blaise Pascal, 77455 Champs-sur-Marne,
***yunseok.lee@cstb.fr**

RESUME. Dans cette étude, le clustering, une technique d'apprentissage automatique non supervisée, est utilisée pour segmenter un parc de bâtiments en groupes homogènes en termes d'attributs descriptifs et de performance énergétique. Un parc virtuel de bâtiments résidentiels a été généré pour tester une méthode de clustering. Il est constitué de géométries diverses (forme et taille) qui permettent d'identifier différentes morphologies types. Les attributs des bâtiments utilisés pour le clustering sont séparés en deux parties : l'espace de décision des attributs descriptifs, et l'espace objectif des performances énergétiques.

La méthode de clustering qui est développée cherche à satisfaire le critère d'homogénéité dans les deux espaces. L'interaction entre deux espaces a été réalisée par la réduction de la dimension utilisant LDA et l'auto-supervision utilisant l'indice de Rand ajusté (ARI). L'analyse des résultats permet de valider l'approche de la méthode. Les clusters de bâtiments obtenus sont plus ou moins éloignés des typologies traditionnelles.

MOTS-CLÉS : parcs de bâtiments, clustering, rénovation énergétique

ABSTRACT. In this study, clustering, an unsupervised machine learning technique, is used to segment a building stock into homogeneous groups in terms of descriptive attributes and energy performance. A virtual stock of residential buildings has been generated to test a developed clustering method. It consists of various geometries (shape and size) that identify different types of. The attributes of the buildings used for clustering are separated into two parts: the decision space of descriptive attributes, and the objective space of energy performance.

The developed clustering method seeks to satisfy the homogeneity criterion in the two spaces. The interaction between two spaces has been achieved by dimensionality reduction using linear discriminant analysis (LDA) and self-supervision using the adjusted Rand index (ARI). The analysis of the results makes it possible to validate the approach of the method. The obtained clusters of buildings are more or less distant from the traditional typologies (for example in terms of morphology).

KEYWORDS : building stock, clustering, energy retrofit

1. INTRODUCTION

To achieve the energy transition, the French government has set a target of 50% reduction in final energy consumption by 2050 compared to 2012 in the *Loi de transition énergétique pour la croissance verte (LTECV)* in 2015. In the building industry, responsible for nearly 40% of global energy consumption, the thermal regulation for new buildings alone cannot bring about sufficiently rapid and large changes in the entire stock. Retrofit of existing buildings is therefore strategic.

While accurate measurement and efficiency estimation of energy retrofit actions are possible for an individual building, it is necessary to segment the buildings into groups of homogeneous buildings to work on a building stock scale. For this purpose, several descriptive attributes, such as construction periods, building types, and climate zones, have been used to create typologies. The TABULA/EPISCOPE project provided standard typologies for the European building stocks (TABULA Project Team 2012). A study showed that these types of typology might be useful to guess the global final energy demand of a building stock pointing out notable differences by energy usages and sources (Mata, Sasic Kalagasidis, and Johnsson 2014). Although useful for certain purposes, these typologies are not always relevant for characterizing the energy retrofit actions on buildings.

2. METHODOLOGY

2.1. THE GENERATION OF A VIRTUAL BUILDING STOCK

2.1.1. Background

While "real building stock" data from actual building diagnoses have reliable values, the comparison between before and after energy retrofit of buildings in a large building stock is hardly possible. Since building energy data with both pre-retrofit and post-retrofit status were rare, only scores of buildings could be available in studies (Deb and Lee 2018). Alternatively, a virtual building stock saves the effort and the resource required for collecting the data (Nikolaou et al. 2009). In this study, a virtual building was generated with three attribute groups, i.e. morphology of buildings, energy features such as envelope and energy system, and energy performance.

2.1.2. Morphology

Various different typologies based on the morphology have been proposed depending on research interests, and historical and regional contexts. LSE and EIFER adopted five types related to the urban context (LSE Cities and EIFER 2014). Regarding the energy features, the previously mentioned TABULA/EPISCOPE project harmonized different typologies of member countries (Stein et al. 2014). In France, several morphological typologies were compared focusing on the urban forms and the solar gains (Gauthier 2014).

	Type	Width (m)	Length (m)	L/W	Stories	Height (m)	H/L	Number
MI	Individual house	5-10	5-20	1.0-3.0	1-2	2.8-9.4	0.14-1.88	1000
LC	Small collective housing	7.5-25	15-80	1.0-10.7	3-21	9-39	0.5-2.0	1000
LCT	High-rise tower	7.5-30	15-50	1.0-6.7	5-33	30-100	2.0-6.7	1000
LCB	Low-rise block	7.5-25	16.7-150	1.0-20	3-13	9-39	0.26-0.5	1000

Table 1 : The dimensions bounds set-up for the virtual building stock

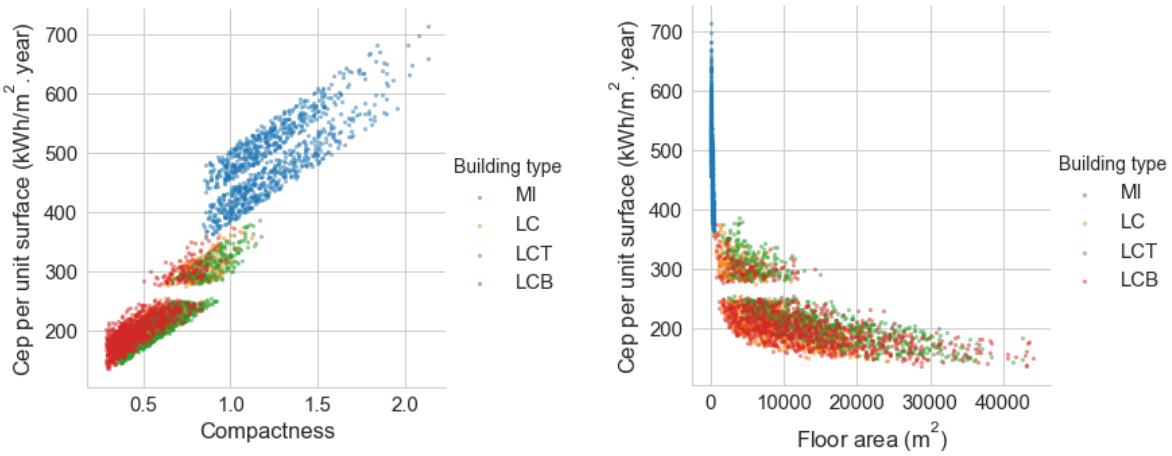
Based on these morphological analyses, four residential building types were selected for the virtual building stock shown in Table 1. The types were separated by the limits of dimensions (width, length, and the number of stories), and the aspect ratios, such as height-to-length (H/L) and width-to-length (L/W). Within the limits, 4000 buildings were generated in a uniformly random manner, using a cityGML modeling engine, Random3Dcity (Biljecki, Ledoux, and Stoter 2016). The generated building stock did not reproduce the real statistics where individual houses are more common than collective housings.

2.1.3. Energy features

Focusing on the variation of the morphology of buildings, the thermal properties of exterior walls and openings and the energy system (natural ventilation, electric convectors for heating, and constant

lighting power) were assumed identical for all buildings. DHW system was not included because of the lack of information at the moment. Simplified formulae were utilized to determine features, such as heating demand (Delphine and Pierre 2016) and ventilation capacity (CSTB 2008).

Aiming to be used for the management of energy retrofit, the virtual building stock requires the energy performance, such as energy demands and consumptions. The energy performance was simulated considering the features dealt with in the previous clauses, using COMETH, a calculation engine for the dynamic energy simulation developed by CSTB (Da Silva et al. 2016).



(a) Compactness (exterior wall / floor area) *(b) Total floor area*

Figure 1 : Primary energy consumption (Cep) of the generated buildings by building types

2.2. CLUSTERING IN TWO SPACES

2.2.1. Background

Clustering of buildings is quite recent research topic, and some studies compared different clustering algorithms (Geyer, Schlüter, and Cisar 2017). Concerning clustering in two spaces, clustering in a space of all features was formerly examined as a method (Lee et al. 2016). In this study, a new method for clustering in two spaces was designed as shown in Figure 2. This method comprises (1) constitution of objective and decision spaces, (2) clustering in the objective space, (3) dimensionality reduction of the decision space, (4) clustering in the decision subspaces, (5) pairwise comparison of objective and decision clusters, and (6) selection of final clusters.

2.2.2. Objective space and decision space

The descriptive features, such as construction years, morphologies and functions, are comparatively easy to obtain by analyzing documents and conducting surveys. However, they are only restrictively capable of telling about the energy performance of the buildings. The energy performance itself, such as energy consumption, energy demand, or CO_2 emission, is difficult to acquire as precise measurement or computational estimation are required.

These two groups of building features were named as decision features and objective features, respectively. Decision space comprises descriptive features, and they are usually available with the facility. On the other hand, an objective space consists of the rarely available energy performance. As the generated virtual buildings had very limited energy usages (heating and lighting) and source (electricity), three features associated with building shapes and two features of the energy performance were selected as the decision features and the objective features, respectively. The features were standardized to be used for the clustering application.

Decision features	Objective features
•Floor area •SSE (South equivalent glazing area by 3CL-DPE method ("RT Existant: Outils et Guides Pour Le DPE" n.d.)) to floor area ratio •Compactness (exterior wall to floor area ratio)	•Heating energy demand per unit area •Primary energy consumption per unit area

Table 2 : Feature selection for objective and decision spaces

2.2.3. Objective clustering

For the clustering, we adopted k-means algorithm which segments the objects into k clusters according to the distance between objects and centroids of clusters. The clusters are determined by iteration of (1) cluster assignment of objects to the nearest centroids, and (2) centroid update with the new assignment until the convergence. To determine the number of clusters k in advance, the Silhouette coefficient was calculated for various k's.

In the developed clustering method, the 3-dimensional decision space was reduced into multiple 1-dimensional decision subspaces which can reproduce the best each objective cluster. Consequently, every objective cluster would have a corresponding decision subspace. Then decision clustering was performed in each decision subspace with all buildings in the stock. In the decision clustering, k-means algorithm and the silhouette coefficient were used as well.

2.2.4. Evaluation of clusters and selection of final clusters

Rand index is the ratio between the number of agreed pairs of objects and the total number of pairs of objects. The agreed pairs mean that the objects are either in the same group or the different groups in both partitions. At present, the Adjusted Rand Index (ARI), a corrected Rand index with chance normalization, is practically used. ARI is 1 for exactly same partitions, 0 for perfectly random partitions.

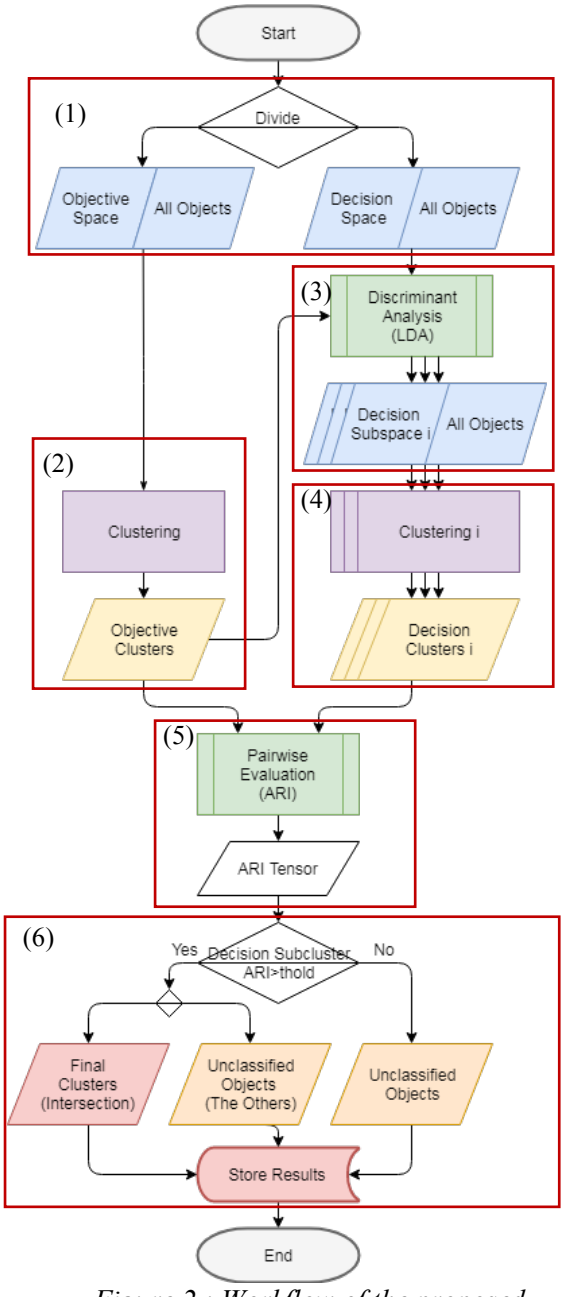


Figure 2 : Workflow of the proposed clustering method

For the evaluation of clustering in this study, ARI between binary classes of objective clusters and those of decision clusters. For example, the cluster member objects were assigned as 1 and the non-member objects as 0. If certain pairs of objective clusters and decision subspace clusters have ARI values close to 1, it means that the two clusters are in agreement.

From the result of ARI evaluation, the best-fit pairs of objective clusters and decision subspace clusters would be found on objective clusters basis. Therefore, intersections of each objective cluster and the corresponding decision cluster with the highest ARI was chosen as the final cluster.

3. RESULTS AND DISCUSSION

3.1. OBJECTIVE CLUSTERING

3.1.1. The decision of the number of clusters

Bearing in mind that the *diagnostic de performance énergétique (DPE)* and the class *gaz à effet de serre (GES)* and evaluating silhouette coefficients of clustering analyses with the number of clusters from 5 to 20, nine clusters were selected as the number of objective clusters.

The result of the objective clustering is shown in Figure 3. The objective clusters were named from C01 to C09 according to the energy consumption per unit area. While the four most energy consuming clusters, i.e. C05 to C09 comprise principally individual houses, C01 to C04 consists of the mixture of three collective housing types.

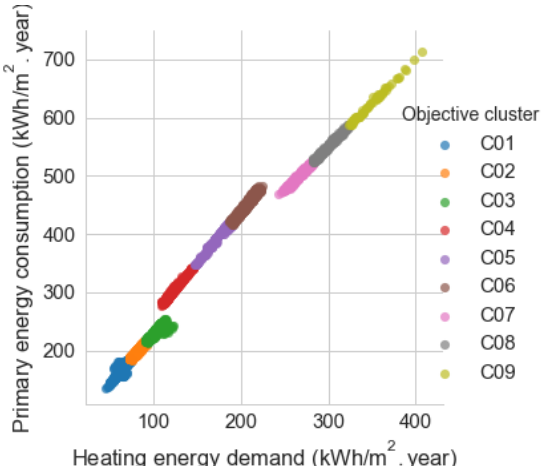


Figure 3 : Objective clusters

3.1.2. Feature importance of objective clustering

To understand which features are more significant for the distinction of objective clusters, F values of ANOVA (the ratio between the variance of group means, and the mean of the within-group variances) was compared. With greater F values, the selected decision features, i.e. floor area, window-to-floor area ratio, and compactness appeared evidently important.

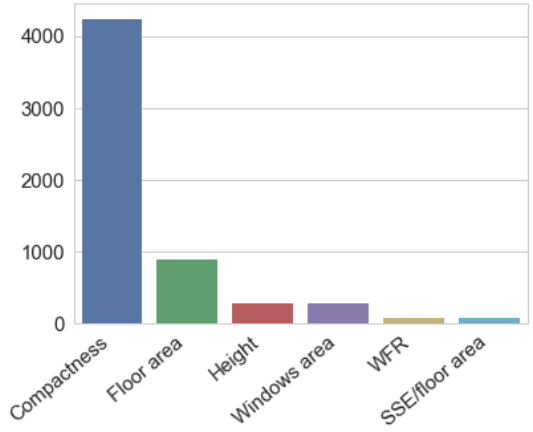


Figure 4 : F value of decision features

3.2. DECISION CLUSTERING

Nine decision subspaces were determined by binary-class LDA for each objective cluster. Due to the binary-class LDA, the decision subspaces were limited to one-dimensional space. Decision clustering analysis was performed on each decision subspace. Figure 5 shows some examples of decision clustering and the corresponding objective cluster. The reduced subspaces by LDA are on the *x*-axes and the floor area is added as the *y*-axes to improve the visibility. As the consequence of the decision clustering, 111 clusters were obtained. Unlike the objective clusters, the decision clusters could be superposed if originated from different decision clustering. For example, a building can belong to the cluster 1 of the decision clustering 1, and to the cluster 2 of the decision clustering 9 at the same time.

3.3. FINAL CLUSTERS

3.3.1. Evaluation of ARI

ARI of all possible pairs of nine objective clusters and 111 decision clusters were calculated. Table 3 shows the pairs of objective clusters and the corresponding decision clusters of maximum ARI

Objective cluster	Maximum ARI	Decision cluster of maximum ARI
C01	0.336	Subspace 1, Cluster 2
C02	0.188	Subspace 1, Cluster 3
C03	0.228	Subspace 8, Cluster 8
C04	0.385	Subspace 8, Cluster 1
C05	0.315	Subspace 6, Cluster 6
C06	0.283	Subspace 9, Cluster 5
C07	0.286	Subspace 8, Cluster 4
C08	0.398	Subspace 2, Cluster 7
C09	0.550	Subspace 8, Cluster 7

Table 3 : Cluster pairs selected by ARI

values. The highest ARI was 0.550, and certain objective clusters did not have corresponding decision clusters of high ARI values. Because of the lack of a threshold, the maximum ARI for each objective cluster was selected to determine the corresponding decision cluster. Figure 5 shows that the best-fit decision clusters for the objective clusters 4 and 9 were found in the LDA subspace 8.

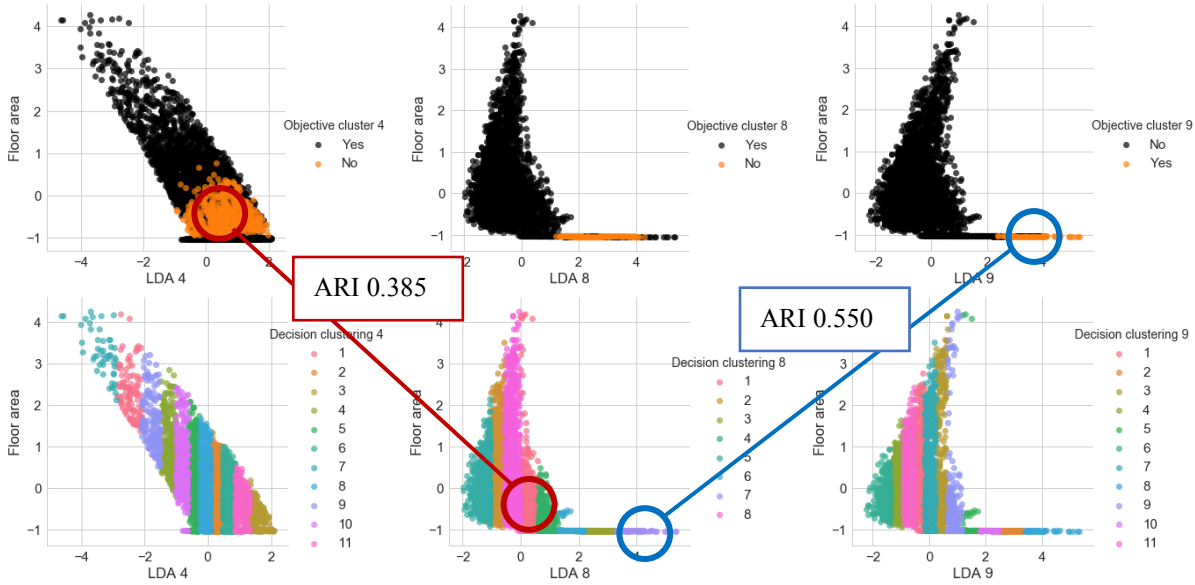
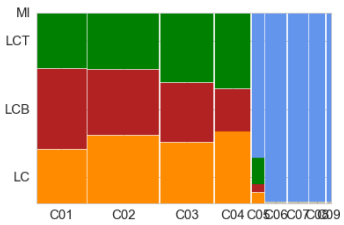


Figure 5 : Objective clusters and decision clusters on some decision subspace

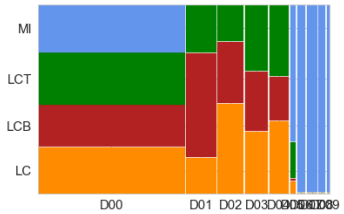
3.3.2. Intersection as final clusters

The intersection of the objective cluster and the corresponding decision cluster was determined as the final cluster. The buildings excluded from the intersection were considered as unclassified buildings. The final clusters were named from D01 to D09 which correspond to the objective clusters C01 to C09 and D00 for the unclassified. The final clusters have 1755 buildings (43.9%) and the others (2245 buildings, 56.1%) were assigned to the unclassified buildings.

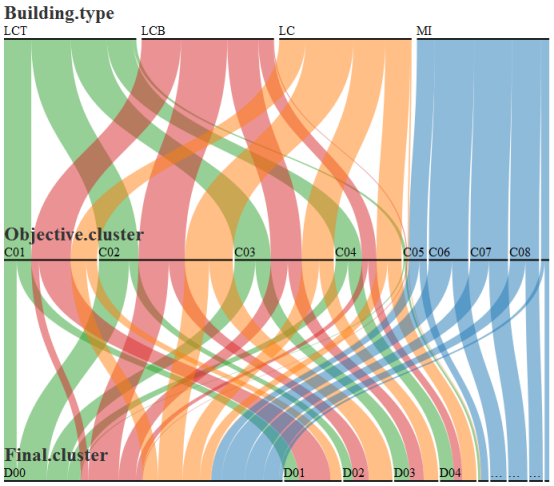
3.3.3. Comparison of objective clusters and final clusters



(a) Objective clusters



(b) Final clusters



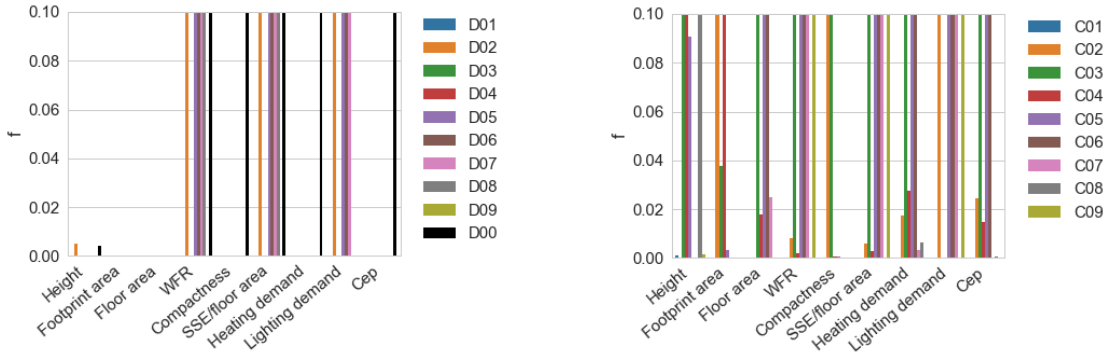
(c) The transition from building types to clusters

Figure 6 : Comparison between clustering results and building types of morphology

The comparison of the final clusters with the objective clusters and the four building types of morphology was shown in Figure 6. The proportion of building types in the resulting clusters kept constant in general. The LCBs were slightly less found in the unclassified cluster (D00) than the others. While the proportion of LCBs increased in D01 than in C01 (low-energy-consuming), it decreased in

D05 than in C05 (medium-energy-consuming). The collective buildings tended to be more unclassified than the individual houses.

To find which features were significant or not to determine the clusters, significance probability (p-value) of features were evaluated. The significance probability is defined as a probability for a given statistical model that the sample difference is same or greater than the actual observation. A common threshold of p-value is 0.05 (5% of probability) but smaller values are recommendable.



(a) Final clusters *(b) Unclassified objects of each final cluster*

Figure 7 : Significance probability of features (the shorter is the more significant)

Figure 7 shows that the height, the areas, the compactness, the heating energy demand and the primary energy consumption are significant (f less than 0.05) features for the most clusters. On the other hand, the other features (WFR, SSE per floor area, and lighting energy demand) appeared more or less significant for certain clusters, particularly the individual houses (D02, D05, D06, D07, and D08). Concerning the unclassified objects, C01 divided into D01 and D00 by all the features. On the other hand, D03 was selected from C03 by the lighting demand. In the same way, the significant features for the classified objects and the unclassified ones for each objective cluster could be analyzed.

4. CONCLUSION

The feasibility of the introduced method for clustering in two spaces was verified. The results showed that the dimensionality of decision space could be reduced by the application of LDA, which is usually used as a classifier for a supervised learning. Some decision clusters in decision subspaces obtained by LDA showed a close connection with the objective clusters. The considerable part of objective clusters, particularly collective housings, were reproduced in decision subspaces. On the other hand, individual houses were found in unclassified buildings, which was the largest part of final clusters. More detailed analysis of the unclassified buildings might allow us interesting knowledge. The division of two spaces made ARI, which is usually used for the classification, a supervised learning technique, usable the performance estimation of clustering, which is a sort of unsupervised learning due to the division of two feature spaces, i.e. the objective space and the decision space. Lastly, the characterization of clusters, which are usually possible in a qualitative manner, was tried in a quantitative way through ANOVA and significance probability test, and the different tendency of clusters could be observed depending on the clusters.

The results offer the possibility of further studies as well. Firstly, different dimensionality reduction technique can be considered. More dimensionality reduction techniques are worth to be compared with LDA. Secondly, other metrics of evaluation of clusters on two spaces can be reviewed. While ARI tends to stress on the overall agreement of two clusters, other metrics are able to find the agreement when a

cluster is a total subset of the other cluster. It would be possible to improve the final clusters by introducing more appropriate metrics. At last, the energy retrofit scenarios and their economics still remain as the interesting and important subject of further studies.

5. BIBLIOGRAPHY

Biljecki, F., H. Ledoux, and J. Stoter. 2016. "Generation of Multi-Lod 3D City Models in Citygml With the Procedural Modelling Engine Random3Dcity." *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-4/W1 (September): 51–59. https://doi.org/10.5194/isprs-annals-IV-4-W1-51-2016.

CSTB. 2008. "Méthode de Calcul TH-C-E Ex : Annexe à l'arrêté Portant Approbation de La Méthode de Calcul TH-C-E Ex."

Deb, Chirag, and Siew Eang Lee. 2018. "Determining Key Variables Influencing Energy Consumption in Office Buildings through Cluster Analysis of Pre- and Post-Retrofit Building Data." *Energy and Buildings* 159. Elsevier B.V.: 228–45. https://doi.org/10.1016/j.enbuild.2017.11.007.

Delphine, Destruel, and Boisson Pierre. 2016. "Rapport : Règles d ' Estimation Des Quantitatifs Pour Le Module de Calcul de CASIE^2."

Gauthier, Noémie. 2014. "Analyses Morphologiques de Formes Urbaines et Etude de l'impact Des Formes Urbaines Sur Les Gains Energetiques Solaires."

Geyer, Philipp, Arno Schlüter, and Sasha Cisar. 2017. "Application of Clustering for the Development of Retrofit Strategies for Large Building Stocks." *Advanced Engineering Informatics* 31. Elsevier Ltd: 32–47. https://doi.org/10.1016/j.aei.2016.02.001.

Lee, Yunseok, Pierre Boisson, Mathieu Rivallain, and Olivier Baverel. 2016. "Application de Techniques de Clustering Pour La Segmentation de Parcs de bâTiments à Rénover." In *Conférence IBPSA France*. Marne-la-Vallée.

LSE Cities, and EIFER. 2014. "Cities and Energy: Urban Morphology and Heat Energy Demand." London.

Mata, É, A. Sasic Kalagasidis, and F. Johnsson. 2014. "Building-Stock Aggregation through Archetype Buildings: France, Germany, Spain and the UK." *Building and Environment* 81. Elsevier Ltd: 270–82. https://doi.org/10.1016/j.buildenv.2014.06.013.

Nikolaou, T., I. Skias, D. Kolokotsa, and G. Stavrakakis. 2009. "Virtual Building Dataset for Energy and Indoor Thermal Comfort Benchmarking of Office Buildings in Greece." *Energy and Buildings* 41 (12): 1409–16. https://doi.org/10.1016/j.enbuild.2009.08.011.

R. A. Fisher. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7 (2): 179–88. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

"RT Existant: Outils et Guides Pour Le DPE." n.d. Accessed April 20, 2018. http://www.rt-batiment.fr/batiments-existants/dpe/outils-et-guides-pour-le-dpe.html.

Silva, David Da, Jean-marie Alessandrini, Jean-baptiste Videau, and Jean-robert Millet. 2016. "Evaluation et Perspectives Du Modèle Thermique de COMETh , Le Cœur de Calcul de La Réglementation Thermique Des bâTiments Neufs." In *Conférence IBPSA France*. Marne-la-Vallée.

Stein, Britta, Tobias Loga, Nikolaus Diefenbach, Bogdan Atanasiu, Aleksandra Arcipowska, Eleni Kontonasiou, Gašper Stegnar, et al. 2014. "Inclusion of New Buildings in Residential Building Typologies Steps Towards NZEBs Exemplified for Different European Countries National Observatory of Athens." Darmstadt, Germany.

TABULA Project Team. 2012. "Typology Approach for Building Stock Energy Assessment - Main Results of the TABULA Project," no. June 2009: 43.