Building Archetype Characterization Using K-Means Clustering in Urban Building Energy Models

Orçun Koral İşeri(✉) and İpek Gürsel Dino

Department of Architecture, Middle East Technical University, Ankara, Turkey
{koral.iseri,ipekg}@metu.edu.tr

Abstract. Population growth in cities negatively affects global climate problems regarding environmental impact and energy demand of building stock. Thus, buildings should be examined for energy efficiency by reaching acceptable internal thermal comfort levels to take precautions against climate disasters. Although building energy simulations (BES) are widely used to examine retrofitting processes, the computational cost of urban-scale simulations is high. The use of machine learning techniques can decrease the cost of the process for the applicability of quantitative simulation-based analyses with high accuracy. This study presents the implementation of the *k-means* clustering algorithm in an Urban Building Energy Modeling (UBEM) framework to reduce the total computational cost of the simulation process. Within the scope of the work, two comparative analyses are performed to test the feasibility of the *k-means* clustering algorithm for UBEM. First, the performance of the *k-means* clustering algorithm was tested by using the observations on the training data set with design parameters and performance objectives. The second analysis tests the prediction accuracy under different selection rates (5% and 10%) from the clusters partitioned by the *k-means* clustering algorithm. The predicted and simulation-based calculated results of the selected observations were comparatively analyzed. Analyses show that the *k-means* clustering algorithm can effectively build performance prediction with *archetype characterization* for UBEM.

Keywords: Urban building energy modeling · Archetype characterization · K-means clustering

1 Introduction

More than 50% of the total world population lives in urban areas. However, intensive urbanization has severe consequences on climate change regarding the high energy use and environmental impact [13]. As a result, cities are under transformation to decrease the environmental impact due to climate change. Local governments have already started to reduce greenhouse gas emissions (GHGs) goals by enforcing necessary regulations. For instance, the City of New York committed to decreasing its GHGs by 80% until 2050 [8]. However, the transformation should begin with the quantitative analysis of the

D. Gerber et al. (Eds.): CAAD Futures 2021, CCIS 1465, pp. 222–236, 2022.
https://doi.org/10.1007/978-981-19-1280-1_14

current state and possible intervention actions to improve performance and reduce the environmental footprint in the built environment [6].

City managements started to form datasets for the built environment related to examining the effect of climate change regulations [18]. These datasets are digital representations of the characteristics of the urban building stock, and they can support identifying and analyzing opportunities and corrective actions for sustainable transformation. However, the limitations to data collection make it challenging to analyze retrofit scenarios [14]. For instance, the data sharing process is limited in Turkey [20]. Thus, alternatives are needed for access to urban datasets.

Urban-scale retrofit of buildings is among the climate change adaptation and mitigation strategies. In various studies, high-resolution analyses were applied on urban building stock with different scales and objectives [27], e.g., human-building interaction, micro-climate observation, and building *archetype characterization.* Because building retrofit scenarios should be evaluated from different perspectives for realistic evaluation application of retrofitting process [11, 22], however, a multi-objective approach can be challenging for the consideration of all retrofit alternatives. The evaluation process can be complex, mainly due to the computational cost and a high number of parameters. Therefore, there is a need to examine the urban multi-objective retrofit scenario evaluation process by developing new computational approaches.

The urban-scale retrofit process is evaluated with urban building energy modeling (UBEM) using different approaches such as bottom-up building energy simulations and top-down data-driven algorithms [19, 32]. Data-driven algorithms are preferred in the UBEM process due to their ease of application and evaluation capacity. In particular, one of the critical data-driven approaches in UBEM is *archetype characterization*, which is realized by grouping the building stock according to similar physical and thermal properties [30]. For instance, the grouping criteria can be energy demand values of building units as performance objectives or construction dates as parameters. Although such approaches are preferred in the literature, clustering over a single parameter may be insufficient in evaluating many building stocks in cities. For this reason, the number of evaluation criteria should be increased to understand building stock's properties as explanatory indicators and facilitate the neighborhood analysis with efficient clustering [1, 28].

The data framework in the UBEM compose of different datasets., thus, the analysis process can be laborious for reaching valuable results. Various studies have preferred machine learning (ML) algorithms because of their ability to manage large and heterogeneous data sets [32, 33]. Archetype identification, energy demand prediction, and occupancy pattern detection are purposes for the usage area of ML [16, 17, 19]. Among these methods, clustering algorithms are effective for building *archetype characterization,* which is a suitable approach for pre-processing heterogeneous data [10]. The algorithm can provide acceleration for urban building energy modeling analysis by determining the *archetype characterization* for the building stock in the neighborhood scale.

The clustering algorithm's performance is essential. It can also be modified according to the selection of feature types of training data because the selection of the features is related to the clustering algorithm' performance. There are examples in the literature that

propose statistical sensitivity analysis to evaluate the building features' impact on the building performance criteria [26, 38]. The analysis ranks critical parameters. Sensitivity analysis is commonly applied in building energy modeling to quantify the impact of design parameters on the performance objectives [25]. On the other hand, the calculation cost can be reduced by fixing the features that do not affect the model outputs with the analysis.

This study proposes an approach to predict building performance objectives, which can accelerate the neighborhood-scale building energy simulation process with high accuracy. *K-means* clustering algorithm is used for the partitioning of the residential building stock based on their (a) physical/thermal properties and (b) performance objectives. Building simulations were conducted for the whole neighborhood model to calculate the latter dataset. The main reason for choosing the clustering technique is to predict energy use and indoor thermal comfort on the neighborhood scale rapidly by selecting from the partitioned clusters using different selection rates (5% and 10%). This method can provide advantages for analyzing the current condition of the building stock and the quantitative performance evaluation of retrofit alternatives. Since the *k-means* clustering algorithm is sensitive to data distribution (particularly to outliers), two comparative analyses were performed to understand the performance of the clustering algorithm for *archetype characterization*. In the first analysis, two clustering models were separately trained using physical/thermal properties and performance objectives. The first analysis indicates that the selected design parameters can be used to characterize archetypes using clustering, which can be used for the performance objective prediction. In the second analysis, the performance objectives were used as input features of the training data for the *k-means* clustering algorithm. Random selection was applied from the clusters formed previously, then two different selection rates (5% and 10%) were applied from these representative clusters. These selections were simulated and were compared with full model simulation results for the prediction accuracy of the clustering algorithm with the selections from the partitioned clusters. Consequently, the results of the analyses indicated that the random selections from these clusters successfully represent the performance of the studied neighborhood. Thus, clustering algorithm preference before the neighborhood simulation could contribute to the acceleration of the neighborhood-scale building energy simulations.

2 Materials

The *archetype characterization* with the *k-means* algorithm was tested in multiple neighborhoods to measure the success of the process. *Bahçelievler*, *Yukarı Bahçelievler*, and *Emek* neighborhoods in Ankara's *Çankaya* district were included as the study area. Ankara generally has a cold and arid climate, so the ASHRAE climate zone is included in the 4B classification ($CDD10^oC \leq 250$, $HDD18^oC$ *(Heating Degree Days)* ≤ 3000) [3]. Heating energy demand (Q_H) has a high proportion of total energy demand in the building stock of a region. Therefore, cooling energy demand was not calculated for the simulation process.

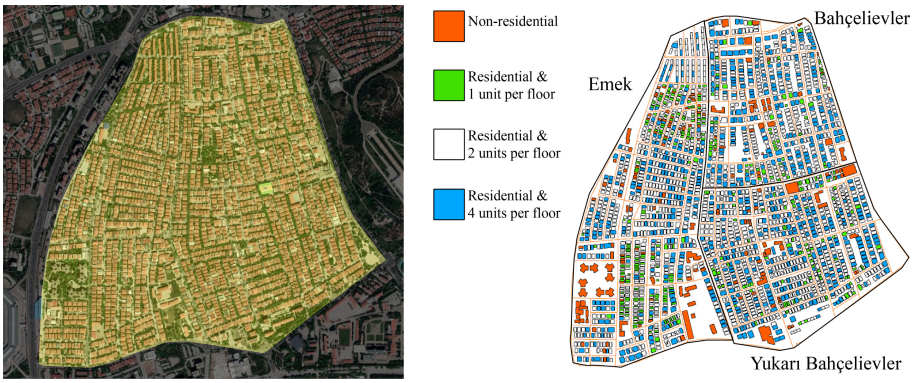

Fig. 1. Selected Neighborhoods in Ankara (left) and 2614 Buildings Based on Building Functions and number of floors.

In Fig. 1, the boundary of the three neighborhoods is shown on the left, and a color-coded representation for the building function with the number of floors is shown on the right. 93% of the study area buildings are residential units, and the remaining buildings are commercial buildings in which are generally located on the ground floors of the buildings. Since occupancy information is not among the data provided by official institutions, it has been obtained from national and city statistical reports by adapting it to the region [35]. During the field visits, the total number of floors and window-wall ratio values of the buildings were collected and entered into the physical properties datasets. Consequently, the building's energy simulations generated the residents' daily energy usage patterns with building energy and comfort standards [4, 5, 34].

3 Methodology

This section presents the proposed methodology of UBEM for the building stock in the selected neighborhoods. The process includes geometrical operations, building simulation (i.e., energy and comfort performance results generation), clustering with machine learning with hyper-parameter tuning, and two-step comparative analysis (i.e., clustering for *archetype characterization*) (Fig. 2).

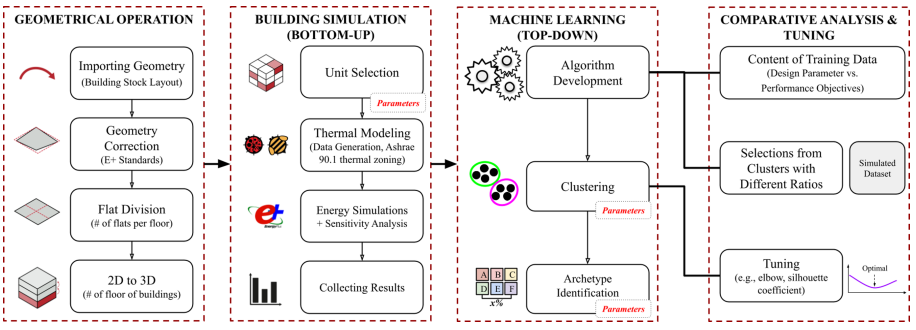

Fig. 2. Flowchart of the proposed method

3.1 Thermal Modeling

The workflow starts by importing the building physical (e.g., layout, # of floors) data to the algorithm. The building footprint curves were simplified into four-edged convex polylines to decrease the building energy simulations' computing cost based on the energy modeling standards defined in [21]. The 2-dimensional building footprints transform to the 3-dimensional thermal zones with the knowledge of the number of floors (Fig. 3).

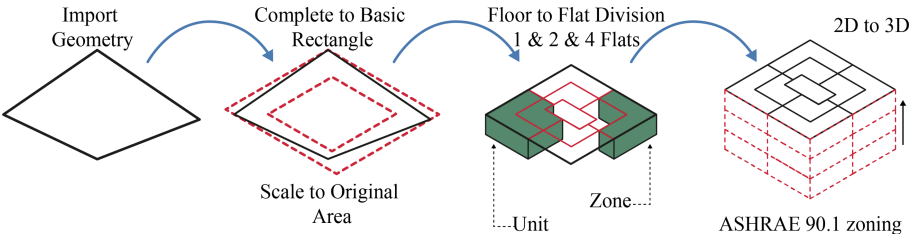

Fig. 3. Geometry correction for thermal modeling

The spatial layout of residential units affects buildings' thermal balance; therefore, it is essential to model in detail. However, the layout data of all buildings in the neighborhood cannot be available in the study regions. For that reason, authors have developed three different types of layouts for units that consist of one, two, and four thermal zones (Fig. 1). The simulation units were divided into different zone types, e.g., bedroom (*B*), living room (*L*), service (*S*). According to the building function, living rooms and bedrooms are the default for all the units. Service areas in which include hallways and bathrooms, do not have external windows (Fig. 4). Each unit has different thermal loads and occupancy schedules compatible with its usage. The distribution of layouts for the building stock was developed by random distribution in parallel with the data obtained from national statistics to the study area.

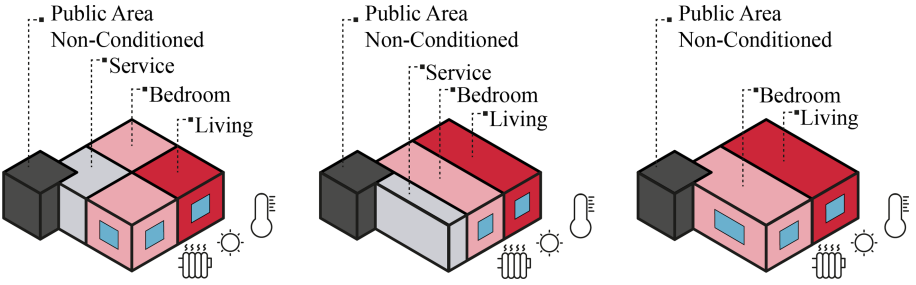

Fig. 4. Zone division of simulated units

More than 25000 residential units were simulated, the simulation results were used for the training data in the clustering algorithm. Three performance objectives are calculated annually by building performance simulations (Fig. 4). These are heating energy

demand (Q_H) and lighting energy demand (Q_L), and degrees of overheating (*OHD*). The heating and lighting demand are two parts of the total energy use calculation that vary according to different design parameter values (Table 1). Since there is no mechanical cooling system, natural ventilation through windows is the only way to cool the zones in residential units. Overheating ratings (*OHD*) are used to measure thermal disturbance in the summer season. *OHD* is calculated using a fixed upper-temperature limit for each zone type. The threshold of *OHD* is 28 °C for the living room and 26 °C for the bedroom [9].

Table 1. Design parameters and performance objectives of the training data

Thermal and physical properties			
Property	Value	Unit	Type
U-value, Wall*	{0.60, 1.88}	W/m^2-K	Pre-defined
U-value, Roof*	{1.88, 3.12}	W/m^2-K	Pre-defined
U-value, Floor*	{0.93, 1.92}	W/m^2-K	Pre-defined
U-value, Window*	{5.1, 2.1}	W/m^2-K	Pre-defined
Heating set point/set back	25.0, 20.0	°C	[6]
Ventilation type	only natural, one-sided	–	Pre-defined
Ventilation limits	21.0, 24.0	°C	Pre-defined
Infiltration	0.0002, 0.0003	m^3/s-m^2	Extra
Window opening ratio	[0.25–0.5]	–	Pre-defined
Occupancy schedule	29 types,1 to 5 people	–	[26]
Window-to-wall-ratio	[0.15–0.30]	–	Extra
Equipment load	{2, 3, 5}	W/m^2	Pre-defined
Lighting density	{5, 7, 10}	W/m^2	Extra
Performance objectives			
Property		Unit	Type
Heating demand (Q_H)		kWh/m^2	Pre-defined
Lighting demand (Q_L)		kWh/m^2	Pre-defined
Overheating degrees (OHD)		°C	Pre-defined

*Before/after 1980

A residential archetype unit can be characterized as related to building physical, thermal, or occupancy properties. A part of the residential unit features is obtained from official institutions for this study [24], e.g., building footprint or per floor unit number. The rest of the building features were generated to the extent specified by national statistics and building energy modeling standards [4, 5], [42], for instance, occupancy properties. However, in several cases, the number of building parameters in the training

data may not be sufficient for the clustering algorithm; therefore, the authors added extra parameters to the simulation algorithm.

The extra design parameters are included in the data pool by performing sensitivity analysis according to their impact on energy use performance objectives. Within the scope of this study, the calculation of impact was achieved with sensitivity analysis. Morris's analysis is used as a sensitivity analysis. Morris Sensitivity analysis is a screening local sensitivity analysis with the elementary-effect method based on a finite distribution of input parameters. The analysis works to rank the input factors' relative importance, namely the first-order main effect (*Si*), by influencing the output parameters [22].

Table 2. Results of Morris sensitivity analysis

Parameter	Range	mu*	Type
U-value, wall	{0.6, 1.2, 1.8, 2.4}	33.386	Pre-defined
Window-to-wall ratio	{0.1, 0.2, 0.3, 0.4}	31.844	Extra
Infiltration rate	{0.0002, 0.0003, 0.0004, 0.0005}	28.452	Extra
Lighting density	{5, 7.5, 10, 12.5}	22.252	Extra

Morris sensitivity analysis was applied to test a pre-defined and three extra design parameters regarding the influence (*mu**) on a performance objective, which is selected as the heating demand (Table 2). The *U-value Construction* is a pre-defined design parameter, and it was included in the analysis to compare the impact of the three extra design parameters as a proxy [34]. The authors manually defined ranges for these parameters in the building energy simulations (Table 1). Based on the first-order (*Si*) main effect index results, window-to-wall ratio, infiltration rate, and lighting density for zone parameters were highly influential on Q_H. Consequently, these design parameters were added to the training dataset.

3.2 Occupancy Modeling

Occupancy modeling for residential buildings is one of the critical features for building performance. The subject previously studied residential buildings to monitor occupant actions and cluster activity schedules from performance objectives [7, 12]. Many uncertainties exist with a high degree of influence for energy demand and indoor thermal comfort as occupants interact with building systems (e.g., heating setpoints, natural ventilation) [41]. Nevertheless, most modeling approaches use default occupancy schedules, and they ignore the different occupant profiles and their specific ways of space use and system interaction. The writers of this study have proposed to use a new method for realistic occupancy modeling. The process is a combination of datasets from different resources and different statistical techniques (Fig. 5).

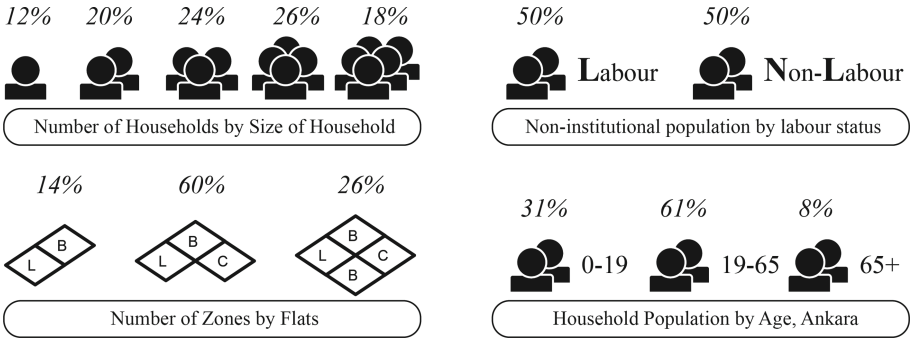

Fig. 5. Occupancy by national statistics

The proposed approach consists of national statistics for occupancy and location-based address registrations for unit information [24]. For the selected neighborhoods, occupancy scenarios are modeled based on three different statistical information and different unit layout modeling, where is mentioned in Sect. 3.1 (Fig. 5). The statistical data were used to map the daily occupant activities for residential units, i.e., household size, labor status, age [36, 37]. The simulation units were divided into different zone types, i.e., bedroom (*B*), living room (*L*), service (*S*). Thirty-one occupancy schedules were matched with the simulation zones based on occupant preferences. Each schedule was assigned randomly to the units, which helped to produce randomly distributed objective performance results in the training dataset.

3.3 Building Energy Simulation

The selected district's digital models were built according to Turkish TS-825, ASHRAE 55, ASHRAE 90.1 standards [4, 5, 34]. Ladybug/Honeybee Visual Coding tools were the simulation tools for the building energy simulations with the EnergyPlus engine [29, 39]. All residential unit simulations were separately simulated. For each simulation, random construction, internal load, and occupancy schedules were assigned (Fig. 6).

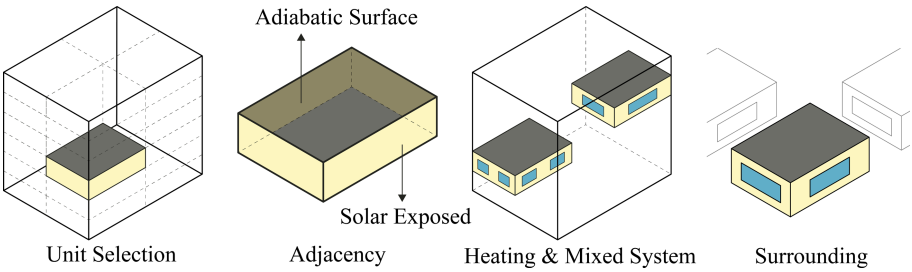

Fig. 6. Unit selection and surface adjacency

Surface types differ for vertical and horizontal positions based on the residential unit position. The internal walls are adjacent, and they are set to adiabatic surfaces. The surrounding geometries were introduced as environmental context surfaces. They are essential due to solar radiation reflections.

3.4 Clustering

Clustering is an unsupervised machine learning algorithm that works for unlabeled data structures—the algorithm search for similarity between the values of parameters. The similarity is a valuable measure for the qualitative data features. However, the distance calculation works better to recognize the numeric data's relationship. The algorithm defines the distances of the instances from each other according to their similarities or dissimilarities [40].

The process starts with the selection of features and feature extraction from training data. Then, the algorithm proceeds with the design of the clustering algorithm suited explicitly to the problem. It evaluates the results to improve the algorithm's performance. Lastly, it completes with the realization and comparison of the results based on statistical formulas [31].

3.5 Partial Clustering

The partitional clustering algorithm defines the center points in the data for non-overlapping clusters [23], e.g., *k-means*, *k-medoids*. The *k-means* approach begins with the random selection of k-different center points for each cluster [2]. The *k-means* algorithm updates the center points by iterative computation. At the same time, the expectation-maximization step repeats until the centroid positions reach a pre-defined convergence value. While the expectation step arranges each point for its nearest center point of the cluster, the maximization step computes all the points for each group and sets the new centroid.

In clustering, there is a trade-off between prediction accuracy and cluster stability [10, 15]. Therefore, the tuning process is essential for the algorithm's performance in terms of accuracy. The process of parameter tuning consists of sequentially altering one of the algorithm's parameters' input values. The elbow method and the silhouette coefficient are implemented during tuning. Lastly, *k-means* clustering algorithms are sensitive to the data type. Thus, two different input data types are tested for this study, i.e., physical and thermal design parameters vs. performance objectives.

4 Results

In this section, the *k-means* clustering algorithm was tested in two different ways for UBEM. The first test was to compare the *k-means* clustering with two different training data types, i.e., design parameters and performance objectives. The second test was realized with a different amount of training data (5% and 10%) from the clusters partitioned by the *k-means* clustering algorithm. The training data sets were taken from the same generated data of the selected built environment for each step.

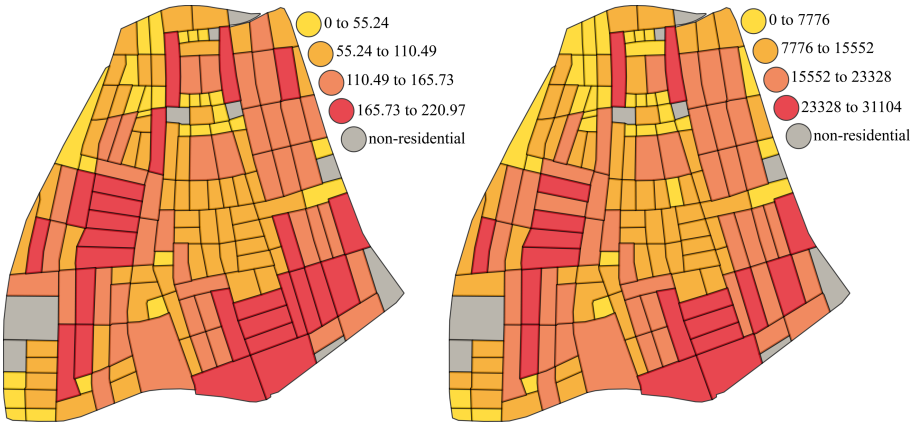

Fig. 7. Spatial distribution of the performance objectives; (left) heating demand (kWh/m^2), (right) overheating degrees

Figure 7 shows the spatial distribution of the two performance objectives in the clustering algorithm's training data for the selected region, i.e., Q_H and *OHD*. The colors represent the region-based the generated data that is the average results of all buildings inside the area. The spatial distribution of the two performance objectives has resulted differently.

4.1 Comparative Analysis for the Qualities of Training Data for Clustering Algorithm

In this section, the *k-means* clustering algorithm is trained with two different data types for the training dataset. Firstly, design parameters were introduced in the *k-means* algorithm, e.g., the residential units' physical and thermal properties and occupancy data (Table 1). The number of clusters resulted in seven clusters using elbow and silhouette coefficient tuning techniques. Secondly, performance objectives were introduced, and after the tuning process, the number of clusters was four.

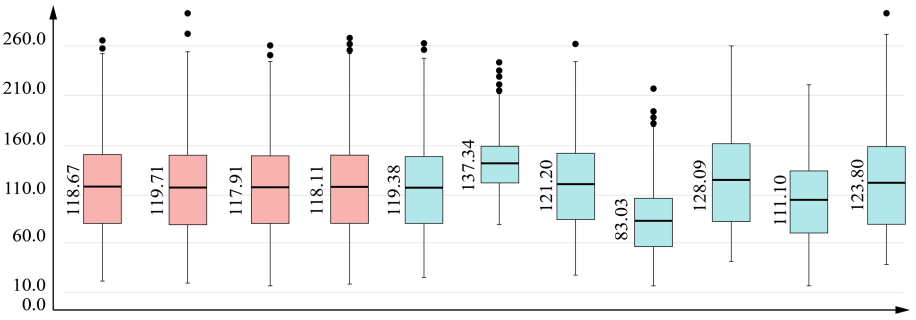

Fig. 8. Energy demand (kWh/m^2) averages for cluster outputs of two clustering processes; Performance objective-based clusters (left, red), Design parameter-based clusters (right, blue) (Color figure online)

Figure 8 shows the box plot values of energy demand averages for two clustering processes in terms of mean and distribution. The number of features of the training dataset was more than the objectives, and the algorithm performed more segmentation for the training dataset. While the energy demand averages of the clustering algorithm groups trained with the training data containing the performance objectives were similar, the clustering results trained with the design parameter differed.

4.2 Partitional Clustering for Facilitation of UBEM

In this part, the model was used in a comparative study to validate how the *k-means* perform with a lower number of selections of residential units. The metrics of comparative analysis were the average and standard deviation for objectives Q_H, Q_L, and *OHD*. The clustering algorithm was coded with the sckit-learn library of *Python 3.6*. The number of clusters was seven clusters after the hyper-parameter tuning. Finally, *the complete model* (i.e., brute-force simulations) in Table 3 is the simulation results of all residential units in the selected neighborhoods.

Table 3. The comparative analysis of different selection rates (5% and 10%) from clusters and complete model simulation

Metrics	Five percent sample	Percentage change	Ten percent sample	Percentage change	Complete model
$\bar{x}_{QH}$	124.97 ∓ 43.70 SD	%5.87	120.1 ∓ 45.00 SD	%1.75	118.04 ∓ 44.36 SD
$\bar{x}_{QL}$	14.97 ∓ 5.54 SD	%0.00	14.8 ∓ 5.60 SD	%1.14	14.97 ∓ 5.69 SD
$\bar{x}_{OHD}$	14986 ∓ 6882 SD	%2.75	15699 ∓ 7096 SD	%1.88	15410 ∓ 7150 SD

Five-percent and *ten-percent* selections were the ratios of the random selection data instances from seven clusters. These selected instances were simulated to generate performance objectives. Then, they have compared with the complete model results in terms of the percentage change. The *k-means* algorithm divided the dataset of generated objectives into different clusters. As seen in Table 3, three performance objectives were compared with average, standard deviation, mean percent ratio. *Ten-percent* sample performed more accurately compared to *five-percent* sample for all performance objectives. However, the values were close to less than a 5% confidence interval ratio even for *five-percent* sample of clustering. In conclusion, the performance of the clustering algorithm showed that it could be used for UBEM to decrease the computation time.

5 Discussion

In this study, comparative analyses were carried out to observe the performance of the *k-means* algorithm for *archetype characterization* of the UBEM performance dataset.

The *k-means* clustering models were separately trained with performance objectives (e.g., energy demand and overheating degrees) and design parameters (e.g., physical and thermal properties of building stock). Even though the number of instances was the same in the two training datasets, the training performed with the parameter-based dataset consisted of more features with higher accuracy. The *k-means* has achieved more clusters with the design parameters included training dataset, and each cluster was reliable to differentiate itself from other clusters according to their dissimilarities (Fig. 8). According to these results, it was seen that the design parameters of residential building stock can play significant role for the building performance prediction in the *archetype characterization* process without simulating all residential units in the selected urban regions. Because building energy simulation is an expensive method in terms of computational cost, and the real building performance data may not always be available for building stock analysis.

Secondly, the performance of *k-means* was tested with different selection ratios from the clusters partitioned by the *k-means* clustering, i.e., *complete model* vs. two different sampling ratios (*five-percent* and *ten-percent*). *Ten-percent* selection resulted more accurately to predict to cluster the energy demand and overheating degree values. However, *five-percent* selection ratio also can be used as an alternative instead of simulating all residential units in the selected region. Thus, it has been seen that the partitions formed by the k-means clustering algorithm are successful in representing the performance data of the entire study area. Nevertheless, each clustering process should be tuned to reach an optimal number of clusters with high accuracy. Elbow and silhouette coefficients were applied and tested multiple times as hyperparameter tuning. Because *centroid positioning* of clusters may result differently between trials due to random initiation. Otherwise, this situation may cause false interpretations during the use of the clustering algorithm.

6 Conclusion

UBEM has capacity for analyzing the urban building stock's performance objectives by collecting, managing, and producing large amounts of real or synthetic data. In addition, advanced machine learning algorithms can be suitable for clustering or estimating these performance objectives. This study proposes a methodology to apply the *k-means* clustering algorithm for the UBEM process. Instead of simulating the entire building stock in the neighborhoods, the clustering method was applied for the clusters of similar features of the building stock. However, the qualities of training data are essential in clustering algorithms. Therefore, two different comparative analyzes were realized for the qualities of training data and the prediction performance of clustering algorithm. For the first analysis, two clustering models were trained with the training data consists of design parameters and performance objectives, separately. The clustering algorithm split the design parameters included training dataset into more groups than the performance objective-based training dataset with high accuracy. The comparative analyses results indicated that physical and thermal parameters of residential building stock could be used as training data content in the clustering process for the UBEM *archetype characterization.* In the second analysis, the methodology consists of the comparative analysis

for the energy simulation with the different selection ratios from the cluster of buildings partitioned by *k-means* clustering algorithm—energy demand averages were compared between the different number of samples from clusters and complete model. The results showed that the clustering algorithm might be suitable for urban building energy modeling to reduce simulations' computational costs. For further studies, the proposed methodology will be tested in an automated process for different climatic zones without simulating entire settlements.

Acknowledgements. This research is supported by the Scientific and Technological Research Council of Turkey (TUBITAK), Grant No. 120M997.

References

1. Aksoezen, M., et al.: Building age as an indicator for energy consumption. Energy Build. **87**, 74–86 (2015). https://doi.org/10.1016/j.enbuild.2014.10.074
2. Arvai, K.: K-Means Clustering in Python: A Practical Guide – Real Python
3. ASHRAE: ASHRAE climatic design conditions 2009/2013/2017
4. ASHRAE: ASHRAE Standard 55-2004 – Thermal Comfort (2004). https://doi.org/10.1007/s11926-011-0203-9
5. ASHRAE: ASHRAE Standard 90.1-2013 – Energy Standard For Buildings Except Low-rise Residential Buildings (2013)
6. World Bank: Cities and climate change: an urgent agenda. Urban development series knowledge papers. World Bank, Washington DC (2010)
7. Bedir, M.: Occupant behaviour and energy consumption in dwellings: an analysis of behavioral models and actual energy consumption in the Dutch housing stock (2017)
8. Chen, Y., et al.: Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Appl. Energy. **205**, 323–335 (2017). https://doi.org/10.1016/j.apenergy.2017.07.128
9. CIBSE: Guide a - Environmental design. The Chartered Institution of Building Services Engineers (2006)
10. Deb, C., Lee, S.E.: Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data. Energy Build. **159**, 228–245 (2018). https://doi.org/10.1016/j.enbuild.2017.11.007
11. El Gindi, S., Abdin, A.R., Hassan, A.: Building integrated Photovoltaic Retrofitting in office buildings. Energy Procedia **115**, 239–252 (2017). https://doi.org/10.1016/j.egypro.2017.05.022
12. Guerra-Santin, O.: Relationship between building technologies, energy performance and occupancy in domestic buildings. In: Keyson, D.V., Guerra-Santin, O., Lockton, D. (eds.) Living Labs, pp. 333–344. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-33527-8_26
13. Hong, T., et al.: CityBES: a web-based platform to support city-scale building energy efficiency (2016)
14. Hong, T., et al.: Ten questions concerning occupant behavior in buildings: the big picture. Build. Environ. **114**, 518–530 (2017). https://doi.org/10.1016/j.buildenv.2016.12.006
15. Hsu, D.: Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Appl. Energy. **160**, 153–163 (2015). https://doi.org/10.1016/j.apenergy.2015.08.126

16. El Kontar, R., Rakha, T.: Profiling occupancy patterns to calibrate urban building energy models (UBEMs) using measured data clustering. Technol. Archit. Des. **2**(2), 206–217 (2018). https://doi.org/10.1080/24751448.2018.1497369
17. Kontokosta, C.E., et al.: A dynamic spatial-temporal model of urban carbon emissions for data-driven climate action by cities (2018)
18. Kontokosta, C.E.: Energy disclosure, market behavior, and the building data ecosystem. Ann. N. Y. Acad. Sci. **1295**(1), 34–43 (2013). https://doi.org/10.1111/nyas.12163
19. Kordas, O., et al.: Data-driven building archetypes for urban building energy modelling. Energy **181**, 360–377 (2019). https://doi.org/10.1016/j.energy.2019.04.197
20. KVKK, K.V.K.K.: Kişisel verilerin Korunması ve işlenmesi Politikası, Ankara (2018)
21. LBNL, L.B.N.L.: Input Output Reference. EnergyPlus (2009)
22. Ma, Z., Cooper, P., Daly, D., Ledo, L.: Existing building retrofits: methodology and state-of-the-art. Energy Build. **55**, 889–902 (2012). https://doi.org/10.1016/j.enbuild.2012.08.018
23. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, pp. 281–297. University of California Press, Berkeley (1967)
24. NVI, T.C.İ.B.N. ve V.İ.G.M.: Yerleşim Yeri Sorgulama / Adres Sorgulama / Adres Doğrulama - Vatandaş Sorgu İşlemleri
25. Østergård, T., et al.: A stochastic and holistic method to support decision-making in early building design. Proc. Build. Simul. Tian 2013, 1885–1892 (2015)
26. Østergård, T., et al.: Building simulations supporting decision making in early design - a review. Renew. Sustain. Energy Rev. **61**, 187–201 (2016). https://doi.org/10.1016/j.rser.2016.03.045
27. Reinhart, C.F., Davila, C.C.: Urban building energy modeling - a review of a nascent field. Build. Environ. **97**, 196–202 (2016). https://doi.org/10.1016/j.buildenv.2015.12.001
28. Pérez, M.G.R., Laprise, M., Rey, E.: Fostering sustainable urban renewal at the neighborhood scale with a spatial decision support system. Sustain. Cities Soc. **38**, 440–451 (2018). https://doi.org/10.1016/j.scs.2017.12.038
29. Roudsari, M.S., Pak, M.: Ladybug: a parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design. In: Proceedings of BS2013: 13th Conference of International Building Performance Simulation Association, pp. 3128–3135 (2013)
30. Sokol, J., et al.: Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. Energy Build. **134**, 11–24 (2017). https://doi.org/10.1016/j.enbuild.2016.10.050
31. Sola, A., et al.: Simulation tools to build urban-scale energy models: a review. Energies **11**, 12 (2018). https://doi.org/10.3390/en11123269
32. Swan, L.G., Ugursal, V.I.: Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew. Sustain. Energy Rev. **13**(8), 1819–1835 (2009). https://doi.org/10.1016/j.rser.2008.09.033
33. Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., Finn, D.P.: Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. Build. Environ. **140**, 90–106 (2018). https://doi.org/10.1016/j.buildenv.2018.05.035
34. TSE: Ts 825: Binalarda Isı Yalıtım Kuralları (2008)
35. TÜİK: TÜRKİYE İSTATİSTİK KURUMU Turkish Statistical Institute (2010)
36. TUIK, T.S.I.: Employment status and participation rate (2020)
37. TUIK, T.S.I.: Indicators related with disability and old age, 2012, 2014, 2016, 2019 (2019)
38. Westermann, P., Evins, R.: Surrogate modelling for sustainable building design – a review. Energy Build. **198**, 170–186 (2019). https://doi.org/10.1016/j.enbuild.2019.05.057

39. Crawley, D.B., Pedersen, C.O., Lawrie, L.K., Winkelmann, F.C.: EnergyPlus: energy simulation program. ASHRAE J. **42**, 49–56 (2000)
40. Xu, D., Tian, Y.: A comprehensive survey of clustering algorithms. Ann. Data Sci. **2**(2), 165–193 (2015). https://doi.org/10.1007/s40745-015-0040-1
41. Yan, D., et al.: Occupant behavior modeling for building performance simulation: current state and future challenges. Energy Build. **107**, 264–278 (2015). https://doi.org/10.1016/j.enbuild.2015.08.032