Developing urban residential reference buildings using clustering analysis of satellite images

Article

Accepted Version

Li, X., Yao, R. ORCID: https://orcid.org/0000-0003-4269-7224, Liu, M., Costanzo, V., Yu, W., Wenbo, W., Alan, S. and Baizhan, L. (2018) Developing urban residential reference buildings using clustering analysis of satellite images. Energy and Buildings, 169. pp. 417-429. ISSN 0378-7788 doi: 10.1016/j.enbuild.2018.03.064 Available at https://centaur.reading.ac.uk/76547/

To link to this article DOI: http://dx.doi.org/10.1016/j.enbuild.2018.03.064

Publisher: Elsevier

Accepted Manuscript

Developing urban residential reference buildings using clustering analysis of satellite images

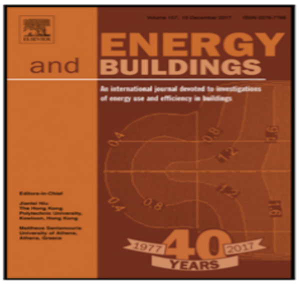
Xinyi Li , Runming Yao , Meng Liu , Vincenzo Costanzo , Wei Yu , Wenbo Wang , Alan Short , Baizhan Li

PII: S0378-7788(18)30417-1
DOI: 10.1016/j.enbuild.2018.03.064
Reference: ENB 8450

To appear in: *Energy & Buildings*

Received date: 2 February 2018
Revised date: 21 March 2018
Accepted date: 25 March 2018

Please cite this article as: Xinyi Li , Runming Yao , Meng Liu , Vincenzo Costanzo , Wei Yu , Wenbo Wang , Alan Short , Baizhan Li , Developing urban residential reference buildings using clustering analysis of satellite images, *Energy & Buildings* (2018), doi: 10.1016/j.enbuild.2018.03.064



Developing urban residential reference buildings using clustering analysis of satellite images

Xinyi Li [1,2] , Runming Yao[1,2*], Meng Liu[1], Vincenzo Costanzo[2], Wei Yu[1] , Wenbo Wang[2], Alan Short[3], Baizhan Li[1*]

[1,] Joint International Laboratory of Green Buildings and Built Environments, Ministry of Education, Chongqing University, Chongqing, 400045, China;

[2,] The School of the Built Environment, University of Reading, Whiteknights, Reading, Whiteknights PO Box 219, RG6 6AW, UK;

[3,] The Department of Architecture, University of Cambridge, CB2 1PX, UK

Corresponding Author: Runming Yao,

Email: r.yao@reading.ac.uk;

Postal Address: The School of the Built Environment, University of Reading, Whiteknights, Reading, RG6 6UD, UK

Abstract

Built-up areas tend to comprise a variety of buildings with diverse and complex shapes, functions and construction characteristics. This variety is the source of significant challenges when calculating building energy use at the building stock level. Moreover, the process of developing stock models usually requires large amounts of data that are frequently scarce, nonexistent or at least not publicly available. Under these circumstances, defining a limited set of reference buildings representing the stock is useful to study the actual energy consumption and the potential effects of different energy conservation measures. This paper presents a new method for developing typical residential reference buildings at district level for bottom-up energy modeling purposes. By means of widely and freely available satellite images, an information database of building shapes is created and a clustering analysis of the geometrical features is performed to define a number of archetypes representative of the heating and cooling energy demand of the district. The method is tested and demonstrated through the case study of the Yuzhong District in Chongqing (China) by comparing the Energy Use Intensity (EUI) of the archetypes derived in this way against detailed dynamic simulations. Results show very small differences in the estimated stock energy consumption (+0.03% in heating energy consumption and +2.97% in cooling energy consumption).

Keywords: Residential building benchmark; building shape; cluster analysis; built-up area; bottom-up approach

1 Introduction

The residential building stock, defined as a group or population of residential buildings according to Kohler and Hassler [1], accounts for 48% of total building energy consumption in China in 2015 [2]. China has imposed an ambitious carbon reduction target of reducing carbon dioxide emissions by 60-65% per unit of GDP based on the 2005 baseline, by 2030 [3] whilst simultaneously promoting improvements in the population's living standards [4]. Local authorities face difficult challenges in decision-making about planning for energy retrofit for residential buildings at a community-scale to meet both objectives.

Bottom-up modeling '*calculates the energy consumption of individual or groups of houses and then extrapolates these results to represent the region or nation*'[5]. The approach has been widely employed for evaluating the impact of energy conservation measures at both national and local scales. It can assess the effect of the implementation of building energy policies on different types of buildings. On the other hand, this approach needs extensive building information data and heavy computation time, in particular for the stock-level energy assessment.

Developing a set of prototypical reference buildings to act as a building benchmark, enables the representation of a reasonable percentage of buildings with detailed specifications [6]. This will help policymakers to examine the effect of national energy policies on building energy consumption more effectively.

A building benchmarking approach is popularly employed for building stock energy simulation. Nationwide large-scale stock surveys are usually the first source of data about building characteristics. Such information includes building types, year of construction, floor dimensions, and number of occupants. These information ideally is coupled with a breakdown of the energy consumption features derived from Energy Performance Certificates for generating residential stock energy models [7-10]. Examples are the Residential Energy Consumption Survey (RECS) carried out in the US by the Energy Information Administration [11] and the English House Survey for England's residential stock reported by the Department for Communities and Local Government [12]. However, for many countries, this kind of national scale residential survey does not exist, as is the case for China and due to historical reasons connected with statistical data collection, no official statistical data for residential building operational energy consumption in China exists even on a macroeconomic level [13].

To close this gap, the aim of this research is to develop an easy-to-follow building archetype approach to construct a residential building stock model for energy consumption calculation and evaluation purposes using the limited information available.

The main novelty of the method is given by the use of a purely geometric characterization of the buildings derived using freely-available satellite images. In this way, not only are the input data required and the time needed for modeling significantly reduced, but the method's applicability is very wide since it only requires satellite images. The method developed in this paper is tested in the Yuzhong district in Chongqing, China.

1.1 Archetype approach

Two main building stock aggregation approaches for bottom-up models, namely, building-by-building and archetypes, are classified [14]. The detailed building-by-building aggregation approach usually involves GIS data [15-17] and is very precise in retaining the peculiarities of every building, but it is very often impractical because of the burden of collecting the detailed data needed [18] and GIS data is not publicly available in many countries, including China. Thus, its applicability is rather limited.

According to Mata*, et al.* [19], archetypes (or archetypical buildings) are *'statistical composites of the features found within a category of buildings in the stock'*. The archetype aggregation approach aims at defining typical buildings able to represent the studied stock, which is a preferential option under the circumstances of scarce/unavailable data.

Even though the representative archetype selection has been criticized by Brøgger and Wittchen [20] for a lack of reasoning, the archetype approach for residential building stock aggregation has still been widely utilized [7-10, 21-32] due to its practical advantages. The indexes used for archetype classification include very different features, such as construction period, building type, building size, and HVAC systems characteristics, resulting in a varying number of archetype definitions. However, the building shape characteristics selection for archetypes still mainly remains in the gray zone. Some archetypes [7-9, 22, 26] do not present any information about the shape characteristics, while others [10, 21, 23-25, 27-31, 33] are based on authors' expertise and assumptions. Apart from this, Monteiro*, et al.* [32] used the average geometric information from buildings which belongs to the archetype for building shape definition. The TABULA project [30, 33] takes advantage of the available statistical data, their archetypes' shape characteristics remain the same as the real buildings, which show similar shape characteristics to the mean geometrical features of the statistical sample, or they define "virtual" buildings by using properties statistically detected in the stock. There is no evidence to support the idea that statistically averaged shape characteristics can fully represent the residential building stock given its huge variations from the energy consumption evaluation point of view. However, building shape characteristics are of paramount importance to define the building energy demand since the amount of externally exposed surfaces determines the magnitude of the heat exchanged between indoor and outdoor environments.

To overcome the lack of explicitly-stated evidence or reasoning that supports the building shape archetype selection for stock energy consumption aggregation, clustering was considered for its inherent advantages. Clustering is an unsupervised machine-learning approach that automatically divides data into sub-groups (clusters) [34], and it has been widely used in the building energy research field for different purposes, such as identifying typical occupancy profiles [35], behavior patterns [36], load profiles [37], key building energy efficiency explanatory factors [38], and energy performance benchmarking [39]. It has also been applied in building archetypes development. For example, Ghiassi, *et al.* [40] applied the hierarchical agglomerative clustering method (with Euclidean distance as the distance function and Ward's method as the similarity measure) to cluster residential and office buildings by means of physical and contextual properties. These properties included effective average envelope U-value, effective window-to-wall ratio, thermal compactness, heated volume, and effective floor height. Ballarini, *et al.* [41] proposed the development of archetypes by using hierarchical clustering techniques (using the Mahalanobis distance in conjunction with the Centroid Linkage algorithm). The variables considered in their cluster analysis included the energy needs for space heating, obtained from Energy Performance Certificates; primary energy for space heating; net floor area; opaque envelope average U-value, and window average U-value. Ghiassi and Mahdavi [42], [43] instead used multivariate cluster analysis to generate a reductive bottom-up urban energy computing model. Multivariate cluster analysis methods, including hierarchical agglomerative clustering, K-means clustering, and model-based clustering, are used to identify archetype buildings. Descriptive indicators considered include geometry, solar gains, thermal quality, and operational parameters such as net volume, effective floor height, thermal compactness, effective glazing ratio, effective average envelope U-value, fraction of the year used, etc. Monthly heating demand calculated by simple steady-state computation [42] and simplified annual heating demand calculation [43] are used respectively in comparison evaluation of the clustering methods. These aforementioned studies confirmed the suitability of using the cluster analysis method for building archetype generation. However, they are heavily reliant on existing GIS data or an energy performance certification database to provide building shape characteristics information. Moreover, they only considered heating energy needs without accounting for the cooling needs, maybe because of the geographic location of the study areas where heating is of major concern.

For residential buildings located in the Hot Summer and Cold Winter (HSCW) climate zone in China, both heating and cooling are required for maintaining indoor comfort conditions. Moreover, the GIS data or other types of building information database are not available. Therefore, this study will develop an easy-to-apply approach to collect building shape characteristics from freely-available satellite images. Dynamic building simulation had been utilized to replace simplified heating demand calculation to account for both heating and cooling energy consumption, as cooling energy consumption requires an iteration process to calculate [44]. Representative buildings of the studied residential stock are defined through cluster analysis. The validation of the method involves comparing the energy consumption calculated using the representative buildings with that of the building-by-building calculation method.

2 Methodology

The building benchmark for the residential stock energy calculation will include building shape, the glazing ratio, building envelope properties, occupancy pattern and heating/cooling equipment. The method proposed to develop the residential building shape archetype for the community in this paper consists of the following five steps:

Step 1: General study of the area: obtaining building functions, building window-to-wall ratio (WWR), and ages of construction from publicly available sources and a ground survey;

Step 2: Geometric characteristics database: developing a 3D building information model inferred from satellite images;

Step 3: Characteristics of building shapes: creating geometric information containing building shape characteristics;

Step 4: Clustering analysis: building shape variables correlation analysis and clustering for representative buildings for clusters;

Step 5: Energy analysis for clustering performance evaluation: Evaluation of clustering performance by comparing the results of energy consumption from the clustered archetype reference building aggregation methods and the building-by-building simulation method.

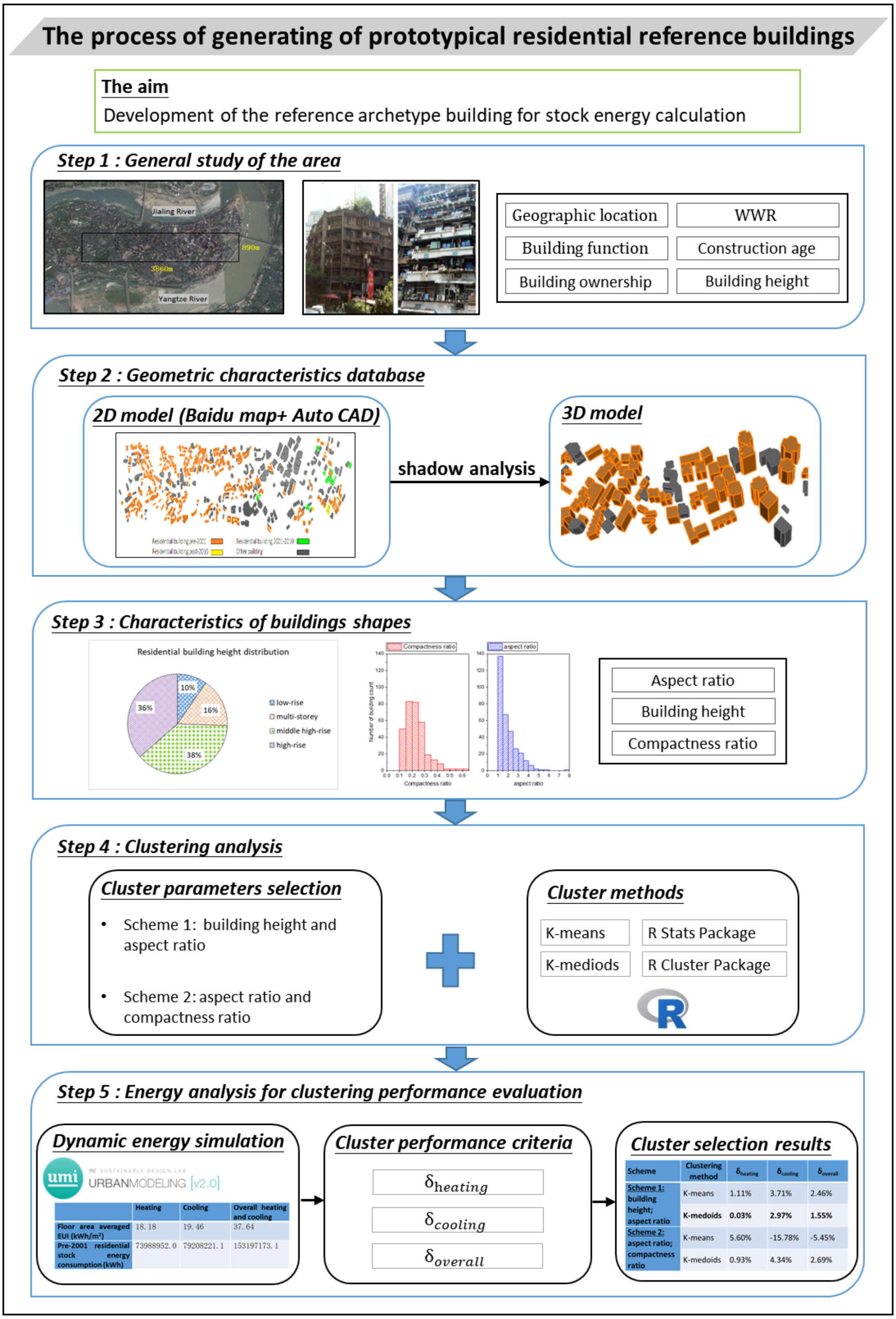


Figure 1: Research framework: the process for generating typical residential reference buildings.

The detailed description of the method is presented in the following section using a case study of the Yuzhong District in Chongqing, China, as an example.

3 General study of the area

Yuzhong district is a well-established built area with a population of around 649.5 thousand people in 2015 [45] that hosts a complex variety of residential and public buildings. More than 60% of residential buildings with a combined floor area of 9,679,167m^2 [45, 46] were built before the release of the first energy conservation standard [47] in 2001. There is a huge energy savings potential in retrofitting old residential buildings in this district because of the poor quality of building envelopes and of the mechanical systems in use. So the local government plans to inject financial support of 90 million RMB to refurbish 4.3 million m^2 floor area of the existing residential buildings [48].

Within the framework of a UK-China collaborative research project entitled 'Low Carbon Climate Responsive Heating and Cooling of Cities'(LoHCool) funded by the Natural Science Foundation China (NFSC) and the Engineering and Physical Sciences Research Council (EPSRC), UK, a stripe within the district has been randomly chosen for analysis (see Figure 2). The total study area is 3.4km^2, accounting for about 17% of the total land area of the district (20.08km^2 [49]).

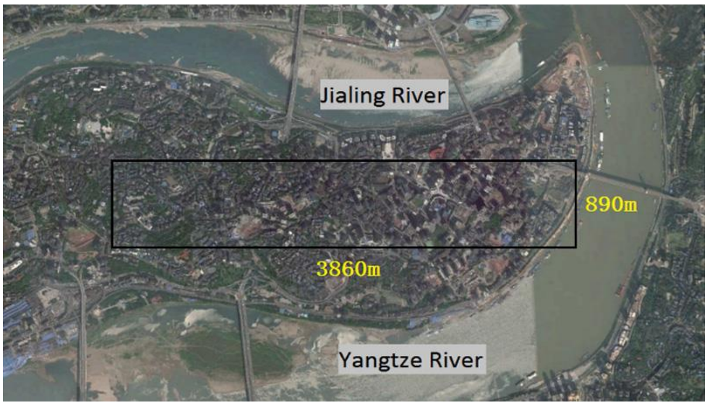


Figure 2: GoogleMap view of the Yuzhong district with the case study stripe highlighted in a black box.

According to the National Education Association [50]and Robert and Daryle [51], the representative sample size of a given population can be calculated using the following formula:

$$S = \frac{X^2 * N * p * (1-p)}{ME^2 * (N-1) + X^2 * p * (1-p)} \qquad (1)$$

Where S is the suggested sample size, X^2 is the table value of chi-square for 1 degree of freedom at the desired confidence level, N is the population size, ME is the designed margin of error (%), p is the population proportion (%) (assumed as 50% to provide the maximum sample size [52]).

Given a confidence level of 99% and an ME of 1%, S is calculated as 16,560m^2 for N equals 9,679,167m^2. Meanwhile, the pre-2001 residential buildings within the studied area reached 4,069,813m^2, which is 245 times the suggested sample size. Therefore, the sampling of the selected strip is deemed representative and appropriate.

A team of ten people carried out a field survey from July 2015 to September 2015;

building information including building geographic location (longitude and latitude), building function, building construction age, number of floors, window-to-wall ratio (WWR), and building ownership were collected for the purpose of generating a building information database.

There are 575 buildings within the stripe and 334 of them are residential buildings, accounting for 60% of the total, while the remaining are public buildings including offices, shopping malls, hotels, hospitals, and schools. Given the predominance of residential buildings, the method is employed to develop residential archetypes, although it is applicable to every building type.

4 Geometric characteristics database

The paucity or lack of digital information for existing buildings in most countries prevents performing a numerical simulation that could be beneficial to understand fully the current stock characteristics and weaknesses, especially in terms of energy efficiency and renovation policies.

Although a significant effort has been made in recent years in the field of digital building reconstruction techniques, there still exist challenges to overcome in terms of i) conversion from captured building data to semantic rich information, ii) possibility of updating the information and iii) definition of an acceptable uncertainty level according to the specific modeling task [53].

It therefore emerges how a case-by-case approach should be followed when tackling the issue of building a semantic 3D model for clusters of buildings. To this aim, several different techniques can be successfully employed, broadly divided into on-site data acquisition (aerial photographs, city building images, 3D laser scanning, and mobile applications) and data from existing building documentation (architectural sketches, 2D scanned paper plans, and CAD plans)[54].

On-site data techniques are usually preferred because of their broad availability and ease of use if compared to data extraction and manipulation from existing building documentation. GIS [55-57] and 3D laser scanning [58-61] are the most employed.

In this work, building geographic location data collected from the ground survey helped locate the buildings on the Baidu map. The web map version used in this study is dated as May 2016. Building footprints have thus been drawn by superimposing Baidu map screenshots in the AutoCAD software, and then scaling them according to a real field measurement. The location of the residential buildings in the stripe is shown in Figure 3, where the different colors stand for different construction ages: orange (pre-2001 buildings), green (2001-2010 buildings) and yellow (post-2010 buildings) respectively. The grey-shaded ones are non-residential buildings that are outside the scope of this study. The construction period classification adopted is consistent with the residential building energy standards in force in the Hot Summer Cold Winter (HSCW) zone to which Chongqing belongs [47, 62]. By looking at the map in Figure 3, it is easy to notice how the pre-2001 buildings form the vast majority (321 premises), accounting for around 95% of the whole residential building stock in the selected sample area. The pictures taken during the ground survey for two pre-2001 residential buildings are shown in Figure 4 to give an example of the appearance of the old residential buildings.

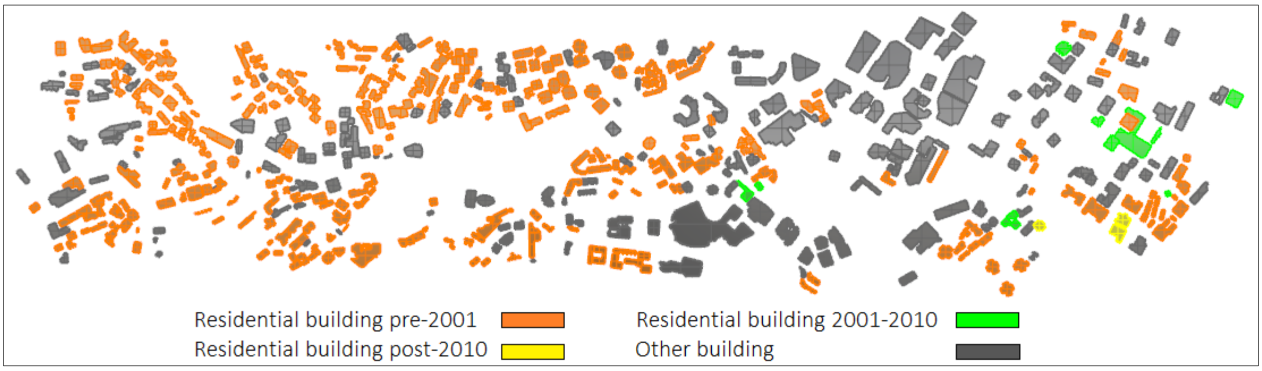


Figure 3: Age-band distribution of residential buildings within the study area

Figure 4: The example views of two pre-2001 residential buildings

A MATLAB script based on the Corner Shadow Length Ratio (CSLR) method described in [63, 64] has been used for estimating the height of the buildings. This approach makes use of shadow analysis from Google Earth images, and derives the building height by linking the corner shadow lengths with satellite position information from the acquisition tool of Google Earth and with the help of traditional solar astronomy relationships for the study site [65]:

$$\sin h_s = \sin\phi \sin\delta + \cos\phi \cos\delta \cos\Omega \tag{2}$$

$$\cos h_s = \frac{\sin h_s \sin\phi - \sin\delta}{\cos h_s \cos\phi} \tag{3}$$

$$\tan h_s = \frac{\cos\delta \sin\Omega}{\cos h_s} \tag{4}$$

where h_s = solar elevation, Φ= latitude, δ = solar declination, and Ω= solar hour angle. (See Figure 5).

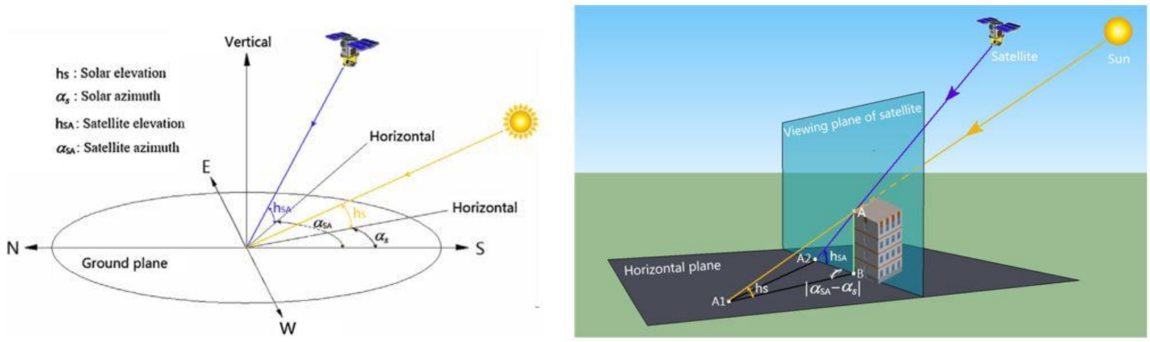


Figure 5: Relations between satellite and sun angles (on the left) and application to a generic building (on the right) [64]

Once one real building height is known, it is possible to calculate the ratio of the building height (red line in Figure 6) to the shadow length of the roof corner (yellow line in Figure 6), known as the *Rcs* ratio. Because this ratio is proved a constant for a given satellite image [63], the other building heights can be obtained by just multiplying the corner height measured in Google Earth (red line in Figure 6) by the *Rcs* ratio.

Figure 6: Test buildings in London (on the left) and Chongqing (on the right)

This approach has been tested first by comparing the real building height of a building located in a regular urban layout in London (Figure 6, left), namely the Holiday Inn Kensington Forum building (LAT. 51° 29’N, LON. 0° 11’ E), with the one estimated by using the CSLR method.

For the sake of evaluating the variations in the computed height due to different satellite picture parameters, three different Google Earth shots have been analyzed and processed under the conditions reported in Table 1.

Table 1: Satellite pictures parameters for the London test building

	Date of shooting	**Solar azimuth angle (°)**	**Satellite azimuth angle (°)**	**Computed height (m)**
Picture 1	8/4/2017	149.52	108.91	85.5
Picture 2	19/7/2013	138.42	110.6	84.6
Picture 3	27/6/2010	104.83	338.16	86.1

average height	*85.4*

The average computed height from the three pictures is 85.4m, which is less than 2% higher than the real value of 84m and can be considered accurate enough for clustering purposes. In order to see if the method is also applicable to a dense and non-regular urban layout such as that of Chongqing, a building located within the study stripe (right side of Figure 6, LAT. 29° 33'N, LON. 103° 33'E) has been selected and tested. During the shooting date of 27/11/2016, the azimuth angle was 165.45°, the satellite azimuth angle was 176.73° and the resulting computed height is 73.2m, just 1.2m bigger than the real figure of 72m obtained from the ground survey.

The above-mentioned tests demonstrate that the method can be successfully applied in building up a database of building heights for stock modeling purposes when actual data is not available, although the reader is invited to refer to refs. [63, 64] for a thorough understanding of the underlying hypotheses and limitations.

5 Characteristics of building shapes

The shape of the Chinese residential buildings is more regular than that of public buildings, with most of them being rectangular and parallelepiped [63]. So building shape variables including building height, aspect ratio and compactness ratio are selected as critical factors defining the surfaces and volume of a building and, consequently, its energy consumption [66-70]. The building compactness ratio (*CR*) is calculated using Equation (5):

$$CR = S / V \quad (5)$$

Where, S is the building surface area;

V is the enclosed volume.

The building's aspect ratio (*AR*) can be calculated using Equation (6):

$$AR = L / M \quad (6)$$

where, L is the longer side width of the building floor plan and M is the width of the shorter side.

The real shape of the residential building had been rounded to a rectangular shape to measure these widths.

According to the Code for Design of Civil Buildings [71], residential buildings can be classified into 4 types according to the number of floors: low-rise (from one to three floors), multi-storey (from four to six floors), middle high-rise (from seven to nine floors) and high-rise (ten and more floors). Middle high-rise (38%) and high-rise (36%) typologies account for more than 70% of the total pre-2001 residential buildings within the studied area, with low-rise and multi-storey accounted for only 10% and 16% respectively (see Figure 7).

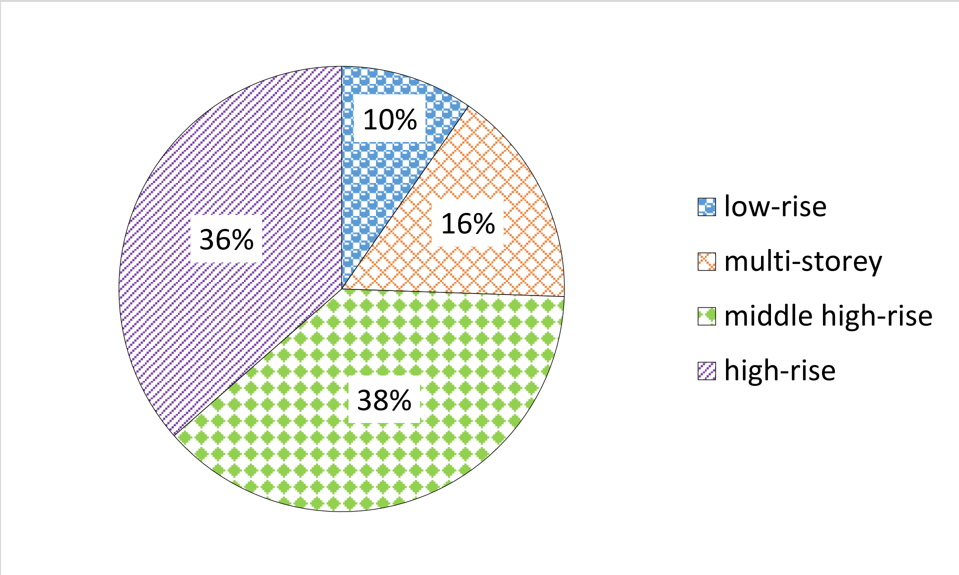


Figure 7: Distribution of pre-2001 residential building height characteristics

The aspect and compactness ratios distribution of these buildings are presented in Figure 8, where it emerges that most of them have a compactness ratio ranging from 0.1 to 0.4, with more than 50% in the range of 0.15 to 0.25. For the aspect ratio, the majority of the buildings are found in the range of 1 to 2, although the highest value recorded is 7.63.

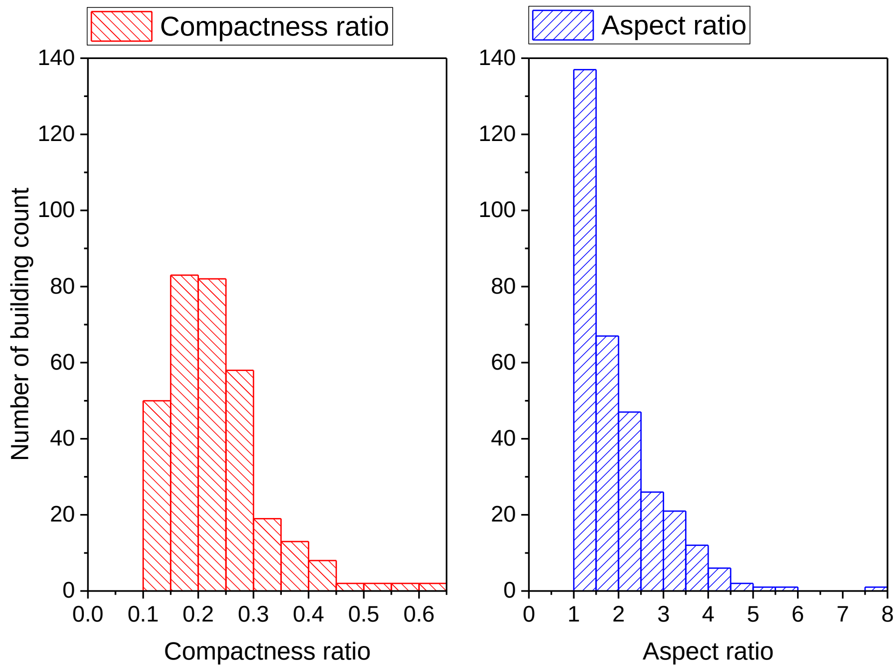


Figure 8: Compactness and aspect ratios distribution of pre-2001 residential buildings

The database has been generated including building's name (building number codes are used), geographic location, building function, number of floors, building height, construction age, total floor area, aspect ratio, and compactness ratio. Clustering analysis will be based on the collected characteristics of the buildings' shapes.

6 Clustering analysis

6.1 Correlation analysis for cluster variable selection

As highly correlated variables cause problems for clustering analysis results, variable elimination should be done to discount redundant variables before clustering [72]. Therefore, the aforementioned three building shape variables:

building height, aspect ratio, and compactness ratio, are tested through correlation analysis. Correlation analysis is a method to measure the strength of association between two variables and the direction of the relationship. A numeric index ranging from -1 to +1, called a correlation coefficient, is used to quantify the correlation strength. A value close to -1 or +1 indicates a strong negative/positive correlation between the variables, while a value close to 0 indicates a very weak correlation. The Shapiro-Wilk normality test has been conducted as a test for the normal distribution of the variables and the results are shown in Table 2. From the table, we can see that three selected variables all reject the normal distribution assumption, as the p-value is smaller than the commonly used significance level of 0.05. Therefore, Spearman correlation, typically used for evaluating the monotonic relationship between two continuous or ordinal variables, is employed instead of traditional Pearson correlation that only works with normally distributed variables.

Table 2: Results of the Shapiro–Wilk normality test

Building shape variable	**p-value**
Building height	0
Building aspect ratio	0
Building compactness ratio	1.58E-14

The Spearman correlation coefficient p can be calculated using Equation 7 [73]:

$$p = \frac{\sum (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 (S_i - \bar{S})^2}} \tag{7}$$

Here, R_i is the rank of the i^{th} x;

S_i is the rank of the i^{th} y;

$\bar{R}$ and $\bar{S}$ are the average of R_i and S_i. While x and y are the two variables subject to correlation analysis.

The results of the Spearman correlation calculation for the three variables of building height, building aspect ratio and building compactness ratio are shown in Table 3. A p-value lower than 0.05 indicates that the correlation coefficient is significantly not zero, i.e. the correlation between the two parameters is statistically significant (these cases are marked in bold in the table). Correlation coefficient values over 0.9 or lower than -0.9 indicate very high correlation, while correlation coefficients between 0.7 to 0.9 or -0.9 to -0.7 indicate high correlation [74]. The correlation analysis shows that the building height and the building compactness ratio are highly correlated (Spearman coefficient of -0.74251), whereas the building aspect ratio is not significant correlated with the other variables. So, to eliminate correlated variables from the cluster analysis, two non-correlated variables selection schemes were used, including:

- Scheme 1: building height and building aspect ratio; and
- Scheme 2: building aspect ratio and building compactness ratio.

Table 3: Spearman correlation analysis for building shape characteristics

		Building height	*Building aspect ratio*	*Building compactness ratio*
Building height	Spearman Corr.	1	**-0.18753**	**-0.74251**
	p-value	—	7.33E-04	0
Building aspect ratio	Spearman Corr.	**-0.18753**	1	**0.21289**
	p-value	7.33E-04	—	1.21E-04
Building compactness ratio	Spearman Corr.	**-0.74251**	**0.21289**	1
	p-value	0	1.21E-04	—

6.2 Cluster methods selection

Clustering methods can be divided into four different types, namely partitional clustering, hierarchical clustering, density-based clustering, and grid-based clustering [75]. For the sake of generating building archetypes, Ghiassi and Mahdavi [42], [43] investigated multivariate cluster analysis using K-means clustering, hierarchical agglomerative clustering, and model-based clustering to generate archetypes for a neighborhood in Vienna (Austria), finding out that K-means performs the best in clustering for energy demand prediction. In more detail, K-means is a partitional clustering technique that assigns objects to clusters to minimize the distance from objects to the cluster center after a user-defined number of clusters (K) is selected [76]. The process of K-means clustering begins with K points selected as initial centroids; clusters are formed by assigning every point to the closest centroid, and then the cluster centroids are recomputed with the new points assigned. The aforementioned assignment and re-computation steps are iterated until every point stays unchanged [76]. When Ghiassi and Mahdavi [42, 43] use K-mean clustering, they define the representative buildings as the ones that have the closest distance to their cluster center, in order to avoid the virtual representative building being used for energy calculation. This approximation will be followed in this study when K-means clustering is used.

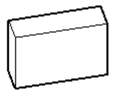
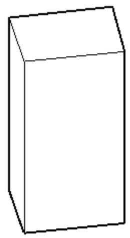
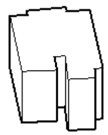
Apart from K-means, K-medoids is widely used and this technique uses the most representative point for a group of data, the medoid, as a measure of the center of the cluster. This ensures real data points are selected as prototypes for the clusters [77], while the centroid used in the K-means method rarely corresponds to an actual data point [76].

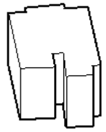
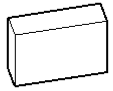
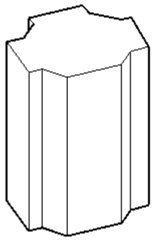
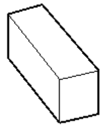
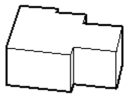
This study attempts to explore the cluster performance of both K-means and K-medoids techniques. Because both techniques utilize distance for cluster assignment, Z-score standardization is performed to rescale the features and to reach a mean of zero and a deviation of one. This helps avoid clustering variables with larger variation ranges dominating the clustering process. The scaled data is thus clustered using the R programming language and software environment [78] together with its nbClust package [79] to choose the optimal number of clusters.

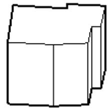
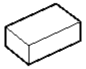
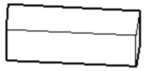
The test range for number of clusters, K, is set from two to ten in order to avoid too many archetype definitions with too many building shape clusters. This will benefit the development of a building energy stock model due to the simplicity of the archetype.

The optimal number of clusters determined by the nbClust package for the two above-mentioned non-correlated variable selection schemes is three, which is finally selected as the K number for both the K-means and K-medoids clustering techniques. Squared Euclidean distance and Euclidean distance are used in the K-means and K-medoids techniques respectively for calculating dissimilarity distances. The clustering results obtained by applying different variable selection schemes and clustering methods are shown in Table 4. It is noted that when using K-means under scheme 1, two buildings (Building B253 and B263) share the same closest distance from the virtual center generated by K-means in cluster 1. The selected representative building shape, as well as the total floor area of buildings belonging to each of the clusters, are shown in Table 4.

Table 4: Clustering results under different variable selection schemes and clustering methods

Schemes	Clustering method	Cluster	Aspect ratio	Building height（m）	Compactness ratio	Selected representative building number code	Selected representative building shape	Total floor area in the cluster (m^2)
Scheme 1: building height； aspect ratio	K-means	Cluster1	3.19	24		B257 & B263*		717,045.00
		Cluster2	1.36	81		B401		1,382,253.13
		Cluster3	1.56	24		B488		1,970,515.33

	K-medoids	Cluster1	1.56	24		B488		1,947,149.04
		Cluster2	3.19	24		B263		740,411.29
		Cluster3	1.16	78		B179		1,382,253.13
<u>Scheme 2:</u> aspect ratio； compactness ratio	K-means	Cluster1	3.33		0.22	B571		726,060.08
		Cluster2	1.53		0.19	B27		3,268,974.18
		Cluster3	2.05		0.38	B543		74,779.20

K-medoids	Cluster1	1.43	0.18	B193		3,158,510.21
	Cluster2	1.75	0.32	B104		152,391.83
	Cluster3	3.12	0.22	B103		758,911.42

2 * These two buildings share the same closest distance from the virtual center of cluster 1.

7 Energy analysis for clustering performance evaluation

From Table 4 we can see that three reference buildings were developed for each scheme under the different clustering methods. To answer the question which scheme and its associated clustering method has the best performance, comparison studies were conducted using building-by-building energy simulation using the recently-developed Urban Modeling Interface (UMI)[80] as a benchmark.

7.1 Energy consumption simulations – Building-by-building approach

UMI is a Rhino-based modeling software package which is able to simulate energy consumption of the building stock at individual building as well as neighborhood and city levels, it uses EnergyPlus [81] as the simulation core engine. To speed-up the simulation process, UMI uses an algorithm that automatically creates thermal zones named 'shoeboxes' based on the definition of perimeter and core zones as reported in ASHRAE 90.1 Appendix G [82], as well as on a detailed solar insolation analysis of the facades. The details of this procedure can be found in [83-85], where validation tests show mean percentage errors in the range 2-5% when shoebox models are compared against their traditional EnergyPlus whole-building models. Apart from a 3D building model, UMI needs all the other input parameters required by EnergyPlus, such as the building envelope thermal physical characteristics, to simulate building heating and cooling energy use intensities.

The building information such as envelope U-values, HVAC equipment performance coefficient and heating/cooling set points, and internal gains is set with reference to the residential buildings design standards JGJ 134-2001 [47] (presented in Table 5). The summer period in which cooling was required was assumed to be from June 15th to August 31st, while the winter period requiring heating was assumed to be from December 1st to February 28th [62]. Occupants' air-conditioning operation schedules were based on survey results from Hu, et al. [86]. Heating is available for one hour in the morning (from 7 to 8a.m.) and four hours before sleep (from 6 to 10p.m.), while cooling is always available except during working hours (from 8a.m. to 5p.m.). All buildings are simulated at their real geographic location, considered real orientation, WWR, and shading from other buildings.

Table 5: The energy consumption simulation parameters used in the study

Envelope U-values (W/m^2K)	Walls	1.97
	Roof	1.62
	Slab	3.74
	Windows (U value/SHGC)	5.74/0.85
HVAC	Heating/Cooling set point (°C)	18/26
	Heating COP/ Cooling EER (-)	1/2.2
Internal gains (W/m2)	Equipment and occupancies	4.3
	Lighting	6

Air change rate (/h)	2
Window to wall ratio (WWR)	Every building retained the real WWR collected from the field survey.

Using this intensive simulation method, 321 pre-2001 residential buildings with a total floor area of 4,069,813.46m^2 were simulated. The color-coded cooling and heating EUIs for every pre-2001 residential buildings was shown in Figure 9.

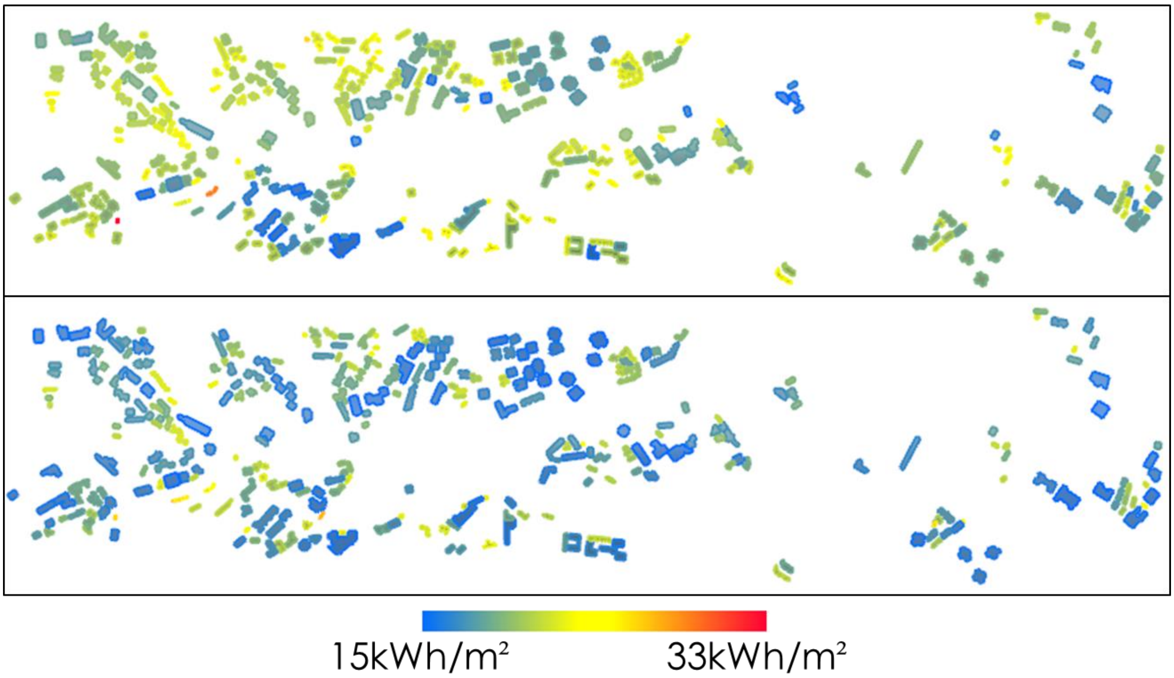


Figure 9: the cooling (top) and heating (bottom) EUIs of every pre-2001 residential buildings

The stock energy consumption characteristics are shown in Table 6, with floor area averaged EUI and pre-2001 residential stock total energy consumption for heating, cooling as well as overall heating and cooling.

Table 6: The energy consumption for the studied stock using the building-by-building approach

	Heating	Cooling	Overall heating and cooling
Floor area averaged EUI (kWh/m^2)	18.18	19.46	37.64
Pre-2001 residential stock energy consumption (kWh)	73,988,952.0	79,208,221.1	153,197,173.1

It is worth noting that, even though UMI is able to simulate all the studied buildings, the process of creating an urban 3D model including all studied buildings, attaching the required information for thermal simulation purposes to every building, and running a stock energy simulation is labor intensive and time consuming, which limited its usability.

7.2 Energy consumption simulations – archetype reference building

From the building-by-building stock simulation in section 7.1, individual building's EUI can be identified with their building number code. The EUIs of the archetype representative buildings are shown in Table 7 for different schemes and clustering methods.

Table 7: Archetype representative buildings EUIs (kWh/m^2)

Scheme	*Clustering method*	*Cluster*	*Heating EUI*	*Cooling EUI*	*Overall heating and cooling EUI*
Scheme 1: building height； aspect ratio	K-means	Cluster1*	20.42	24.15	44.57
		Cluster2	17.53	18.82	36.35
		Cluster3	18.24	19.70	37.94
	K-medoids	Cluster1	18.24	19.70	37.94
		Cluster2	20.57	24.30	44.87
		Cluster3	16.84	18.24	35.07
Scheme 2: aspect ratio； compactness ratio	K-means	Cluster1	18.95	20.27	39.22
		Cluster2	19.14	15.47	34.60
		Cluster3	24.30	19.08	43.38
	K-medoids	Cluster1	18.19	20.05	38.24
		Cluster2	19.79	22.11	41.90
		Cluster3	18.72	21.00	39.72

* These two buildings share the same distance from the center of cluster1, so the representative heating and cooling EUI of cluster1 is set as the average value of buildings 253 and 263.

The EUIs of archetype reference buildings are used to calculate the stock heating and cooling as well as the overall heating and cooling energy consumption by using the following equations 8-10.

$$\mathrm{RSE}_{heating} = \sum_{i=1}^{3} EUI_{heating,i} \times A_i \quad (8)$$

$$RSE_{cooling} = \sum_{i=1}^{3} EUI_{cooling,i} \times A_i \quad (9)$$

$$RSE_{overall} = \sum_{i=1}^{3} EUI_{overall,i} \times A_i \quad (10)$$

Where, $\mathrm{RSE}_{\mathrm{heating}}$, $\mathrm{RSE}_{\mathrm{cooling}}$, and $\mathrm{RSE}_{\mathrm{overall}}$ are the calculated heating, cooling, and overall heating and cooling energy consumption of the studied stock using archetype representative buildings;

$\mathrm{EUI}_{\mathrm{heating,i}}$, $\mathrm{EUI}_{\mathrm{cooling,i}}$, and $\mathrm{EUI}_{\mathrm{overall,i}}$ are the heating, cooling, and overall heating and cooling EUIs of the representative building of cluster i;

A_{i} is the total floor area of buildings belonging to cluster i, in this study, i ranges from 1 to 3.

The energy consumption of the pre-2001 residential stock by using archetype representative buildings is shown in Table 8.

Table 8: Pre-2001 residential stock energy consumption under different schemes and clustering methods (Unit: kWh).

Scheme	*Clustering method*	*$RSE_{heating}$*	*$RSE_{cooling}$*	*$RSE_{overall}$*
Scheme 1: building height; aspect ratio	K-means	74,808,358.8	82,149,893.8	156,958,252.6
	K-medoids	74,014,424.1	81,560,641.8	155,575,065.9
	K-means	78,131,956.6	66,711,884.6	144,843,841.2

Scheme 2: aspect ratio; compactness ratio	K-medoids	74,678,790.4	82,645,530.7	157,324,321.0

7.3 Clustering performance analysis

The representative building capability of accurately representing the energy consumption of the pre-2001 residential building stock is investigated using the following three criteria, presented in equations 11-13:

$$\delta_{heating} = \frac{RSE_{heating} - BSE_{heating}}{BSE_{heating}} \quad (11)$$

$$\delta_{cooling} = \frac{RSE_{cooling} - BSE_{cooling}}{BSE_{cooling}} \quad (12)$$

$$\delta_{heating} = \frac{RSE_{overall} - BSE_{overall}}{BSE_{overall}} \quad (13)$$

Where, δ_{heating} , δ_{cooling} , and δ_{overall} are the relative error of using archetype representative buildings to calculate heating, cooling, and overall heating and cooling energy consumption of the studied stock.

BSE_{heating}, BSE_{cooling}, and BSE_{overall} are the heating, cooling, and overall heating and cooling energy consumption of the studied stock using the building-by-building stock simulation approach, which is shown in Table 6. The calculation results of the abovementioned three criteria are shown in Table 9 for different schemes and clustering methods.

Table 9: Analysis of the performance of different schemes and clustering methods

Scheme	*Clustering method*	δ_{heating}	δ_{cooling}	δ_{overall}
Scheme 1: building height; aspect ratio	K-means	1.11%	3.71%	2.46%
	K-medoids	**0.03%**	**2.97%**	**1.55%**
Scheme 2: aspect ratio; compactness ratio	K-means	5.60%	-15.78%	-5.45%
	K-medoids	0.93%	4.34%	2.69%

When choosing aspect ratio and compactness ratio as the cluster variables together with the K-means clustering method (Scheme 2-K-means), the relative error of overall heating and cooling energy consumption is -5.45%. This is because it overestimates heating by 5.60% while underestimating cooling by -15.78%. All other cluster analysis combination scenarios tend to overestimate cooling, heating, as well as overall heating and cooling energy consumption, but to a lower degree.

Compared with compactness ratio, building height is a better variable to be used in clustering analysis for building shape archetype reference building generation for representing energy consumption and for aggregation purposes. Scheme 1 always has a lower stock scale relative error compared to scheme 2 in heating, cooling as well as overall heating and cooling energy consumption.

The K-medoids technique is considered a better clustering analysis method option as it achieved lower stock level relative error in heating, cooling, as well as overall heating and cooling energy consumption. Moreover, the K-medoids clustering process is more straightforward and convenient given that cluster centers are real buildings instead of virtual '*average*' ones as in the case of the K-means method. This means that the iterative process of calculating the distance between every building and its virtual cluster center to find the closest building as the adjusted cluster center, which is a key feature in K-means analysis, is no longer needed and thus the process is significantly improved.

K-medoids under Scheme 1 is the best performing clustering option. Indeed, the relative errors of heating, cooling as well as overall heating and cooling energy consumption are only 0.03%, 2.97% and 1.55% respectively; which proves the feasibility of using a clustering method to generate building shape archetypes to aggregate the energy consumption of the whole stock. The archetype representative buildings can be used for modeling the energy consumption of the residential buildings built before 2001 in the Yuzhong district.

8 Conclusions

This research has developed a new comprehensive methodology to generate residential building benchmarks (representative buildings of the stock) based on building shape archetypes to serve building stock energy calculation at district level. The main contributions of this research to the body of knowledge are described as follows:

- The method for developing a 3-Dimensional building information database of the studied sample area including building footprint and building height is introduced using freely available 'satellite images'- accompanied with a shadow analysis technique. Computer coding is programmed using the MATLAB platform. This method resolves the problems of scarce publicly available urban digital building information.
- The generation process for the 'Building-shape-based' archetype representative building is proposed. Building characteristics including compactness ratio, aspect ratio, and building height are identified as the shape variables in the clustering analysis. Two clustering analysis input variable selection schemes of the shape variables are determined through correlation analysis.
- Two partitional clustering techniques, namely K-means and K-medoids clustering, are employed for the two preliminary proposed schemes for determining the representative archetype buildings. The energy consumption of the studied stock calculated by using representative archetype buildings is analyzed and compared with that from the 'building-by-building' simulation approach. Performance analysis indicates that the K-medoids clustering technique is more accurate than K-means, by showing a lower relative error of stock energy consumption – compared with dynamic stock simulations of real building shapes – of just 0.03% for space heating and of 2.97% for space cooling respectively.

The building shape archetypes generated are the most representative buildings chosen by clustering analysis; it can be used for investigating the energy consumption of the residential buildings in the area and evaluating the effectiveness of different refurbishment schemes for building energy conservation purposes. This study bears a limitation of manually superimposing map to get building footprints, as this process can be improved by utilizing image processing techniques. Future work lies on the area of how to precisely recognize the building floor plan based on the map screenshot.

Acknowledgements

This research is financially supported by the Natural Science Foundation of China (grant numbers NSFC 51561135002] and the UK Engineering and Physical Sciences Research Council (grant number EPSRC EP/N009797/1) for the collaborative China-UK project LoHCool Low carbon climate responsive Heating and Cooling of cities. Miss Xinyi Li would like to thank the financial support from the China Scholarship Council (grant number 201506050035) for her 12-month academic visit at the University of Reading, UK. Data will be available on request.

References

[1] N. Kohler, U. Hassler, The building stock as a research object, Building Research & Information, 30 (4) (2002) 226-236.

[2] THUBERC, Annual Report on China Building Energy Efficiency, China Architecture & Building Press, Beijing, 2017.

[3] Department of Climate Change, Enhanced actions of climate change: China's intended nationally determined contributions, in, National Development & Reform

Commission of China, 2015.

[4] The State Council of the People's Republic of China. Report on the 19th National Congress of the Communist Party of China. 2017; http://www.gov.cn/zhuanti/2017-10/27/content_5234876.htm (Accessed 2017.11.23)

[5] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews, 13 (8) (2009) 1819-1835.

[6] M. Shahrestani, R. Yao, G.K. Cook, A review of existing building benchmarks and the development of a set of reference office buildings for England and Wales, Intelligent Buildings International, 6 (1) (2013) 41-64.

[7] E. Wilson, C. Christensen, S. Horowitz, H. Horsey, A high-granularity approach to modeling energy consumption and savings potential in the U.S. residential building stock, in: ASHRAE and IBPSA-USA SimBuild 2016, Building Performance Modeling Conference, Salt Lake City, UT, 2016.

[8] A.A. Famuyibo, A. Duffy, P. Strachan, Developing archetypes for domestic dwellings—An Irish case study, Energy and Buildings, 50 (2012) 150-157.

[9] V. Cheng, K. Steemers, Modelling domestic energy consumption at district scale: A tool to support national and local energy policies, Environmental Modelling & Software, 26 (10) (2011) 1186-1198.

[10] S.K. Firth, K.J. Lomas, A.J. Wright, Targeting household energy-efficiency measures using sensitivity analysis, Building Research & Information, 38 (1) (2010) 25-41.

[11] EIA. Residential Energy Consumption Survey (RECS). 2017; https://www.eia.gov/consumption/residential/ (Accessed 2017.10.16)

[12] DCLG. English Housing Survey. 2017; https://www.gov.uk/government/collections/english-housing-survey#history (Accessed 2017.12.24)

[13] CABEE, China Building Energy Consumption Report, in, 2016.

[14] A. Mastrucci, A. Marvuglia, U. Leopold, E. Benetto, Life Cycle Assessment of building stocks from urban to transnational scales: A review, Renewable and Sustainable Energy Reviews, 74 (2017) 316-332.

[15] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, Energy and Buildings, 134 (2017) 11-24.

[16] C. Cerezo Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy, 117, Part 1 (2016) 237-250.

[17] M. Osterbring, E. Mata, L. Thuvander, M. Mangold, F. Johnsson, H. Wallbaum, A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model, Energy and Buildings, 120 (2016) 78-84.

[18] C.F. Reinhart, C.C. Davila, Urban building energy modeling - A review of a nascent field, Building and Environment, 97 (2016) 196-202.

[19] É. Mata, A. Sasic Kalagasidis, F. Johnsson, Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK, Building and Environment, 81 (2014) 270-282.

[20] M. Brøgger, K.B. Wittchen, Estimating the energy-saving potential in national building stocks – A methodology review, Renewable and Sustainable Energy Reviews, 82 (2018) 1489-1496.

[21] Z. Wang, Z. Zhao, B. Lin, Y. Zhu, Q. Ouyang, Residential heating energy consumption modeling through a bottom-up approach for China's Hot Summer–Cold Winter climatic region, Energy and Buildings, 109 (2015) 65-74.

[22] Y. Shimoda, Y. Yamaguchi, T. Okamura, A. Taniguchi, Y. Yamaguchi, Prediction of greenhouse gas reduction potential in Japanese residential sector by residential energy end-use model, Applied Energy, 87 (6) (2010) 1944-1952.

[23] Y. Shimoda, T. Fujii, T. Morikawa, M. Mizuno, Residential end-use energy simulation at city scale, Building and Environment, 39 (8) (2004) 959-967.

[24] Y. Shimoda, T. Asahi, A. Taniguchi, M. Mizuno, Evaluation of city-scale impact of residential energy conservation measures using the detailed end-use simulation model, Energy, 32 (9) (2007) 1617-1633.

[25] J. Kragh, K.B. Wittchen, Development of two Danish building typologies for residential buildings, Energy and Buildings, 68, Part A (2014) 79-86.

[26] N. Heeren, M. Jakob, G. Martius, N. Gross, H. Wallbaum, A component based bottom-up building stock model for comprehensive environmental impact assessment and target control, Renewable and Sustainable Energy Reviews, 20 (2013) 45-56.

[27] L. Filogamo, G. Peri, G. Rizzo, A. Giaccone, On the classification of large residential buildings stocks by sample typologies for energy planning purposes, Applied Energy, 135 (2014) 825-835.

[28] E.G. Dascalaki, K.G. Droutsa, C.A. Balaras, S. Kontoyiannidis, Building typologies as a tool for assessing the energy performance of residential buildings – A case study for the Hellenic building stock, Energy and Buildings, 43 (12) (2011) 3400-3409.

[29] P. Caputo, G. Costa, S. Ferrari, A supporting method for defining energy strategies in the building sector at urban scale, Energy Policy, 55 (2013) 261-270.

[30] I. Ballarini, S.P. Corgnati, V. Corrado, Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project, Energy Policy, 68 (2014) 273-284.

[31] J. An, D. Yan, T. Hong, K. Sun, A novel stochastic modeling method to simulate cooling loads in residential districts, Applied Energy, 206 (2017) 134-149.

[32] C.S. Monteiro, A. Pina, C. Cerezo, C. Reinhart, P. Ferrão, The Use of Multi-detail Building Archetypes in Urban Energy Modelling, Energy Procedia, 111 (2017) 817-

825.

[33] TABULA. TABULA Building Typologies - Country Pages. 2016; http://episcope.eu/building-typology/country/ (Accessed 2017.11.15)

[34] B. Lantz, Machine Learning with R, Southeast University Press, Nanjing, 2017.

[35] G. Buttitta, W. Turner, D. Finn, Clustering of Household Occupancy Profiles for Archetype Building Models, Energy Procedia, 111 (2017) 161-170.

[36] L. Diao, Y. Sun, Z. Chen, J. Chen, Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation, Energy and Buildings, 147 (2017) 47-66.

[37] Z. Ma, R. Yan, N. Nord, A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings, Energy, 134 (2017) 90-102.

[38] E. Wang, Decomposing core energy factor structure of U.S. commercial buildings through clustering around latent variables with Random Forest on large-scale mixed data, Energy Conversion and Management, 153 (2017) 346-361.

[39] X.F. Gao, A. Malkawi, A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm, Energy and Buildings, 84 (2014) 607-616.

[40] N. Ghiassi, K. Hammerberg, M. Taheri, U. Pont, O. Sunanta, A. Mahdavi, An enhanced sampling-based approach to urban energy modelling, in: Proc. of 14th IBPSA Conf. IBPSA, Hyderabad, 2015.

[41] I. Ballarini, S.P. Corgnati, V. Corrado, N. Talà, Improving energy modeling of large building stock through the development of archetype buildings, (2011).

[42] N. Ghiassi, A. Mahdavi, Reductive bottom-up urban energy computing supported by multivariate cluster analysis, Energy and Buildings, 144 (2017) 372-386.

[43] N. Ghiassi, F. Tahmasebi, A. Mahdavi, Harnessing buildings' operational diversity in a computational framework for high-resolution urban energy modeling, Building Simulation, 10 (6) (2017) 1005-1021.

[44] S.C. Hui, K. Cheung, Application of building energy simulation to air conditioning design, in: Proceedings of the Mainland-Hong Kong HVAC seminar, 1998, pp. 12-20.

[45] Chongqing Minicipal Bureau of Statistics & NBS Survey Office in Chongqing, Chongqing Statistical Yearbook, China Statistic Press, Beijing, 2016.

[46] Chongqing Statistics Bureau, TABULATION ON THE 2010 POPULATION CENSUS OF CHONGQING MUNICIPALITY China Statistics Press, 2012.

[47] MOHURD, Design standard for energy efficiency of residential buildings in hot summer and cold winter zone JGJ 134-2001 (in Chinese), in, 2001.

[48] Chongqing Yuzhong District Government. Initiatives to promote energy-saving construction work in Yuzhong District. 2013;

http://www.cq.gov.cn/publicinfo/web/views/Show!detail.action?sid=3611466 (Accessed 2017.11.26)

[49] Chongqing Yuzhong District Government. Physical Geography 2017; http://en.cqyz.gov.cn/content.jsp?id=2c9f813d5a6da45a015a6dda47dd0070&classid=8bcfe211962a4109be84bb516869189d (Accessed 2017.12.16)

[50] National Education Association, Small-Sample Techniques., The NEA Research Bulletin, 38 (1960) 32.

[51] V.K. Robert, W.M. Daryle, Determining Sample Size for Research Activities, Educational and Psychological Measurement, 30 (3) (1970) 607-610.

[52] The Research Advisors. Sample Size Table. 2006; http://research-advisors.com/tools/SampleSize.htm (Accessed 2017.12.20)

[53] R. Volk, J. Stengel, F. Schultmann, Building Information Modeling (BIM) for existing buildings — Literature review and future needs, Automation in Construction, 38 (2014) 109-127.

[54] L. Gimenez, J.-L. Hippolyte, S. Robert, F. Suard, K. Zreik, Review: reconstruction of 3D building information models from 2D scanned plans, Journal of Building Engineering, 2 (2015) 24-35.

[55] L. Truong-Hong, D.F. Laefer, Quantitative evaluation strategies for urban 3D model generation from remote sensing data, Computers & Graphics, 49 (Supplement C) (2015) 82-91.

[56] G. Meinel, R. Hecht, H. Herold, Analyzing building stock using topographic maps and GIS, Building Research & Information, 37 (5-6) (2009) 468-482.

[57] S. Du, F. Zhang, X. Zhang, Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach, ISPRS Journal of Photogrammetry and Remote Sensing, 105 (Supplement C) (2015) 107-119.

[58] F. Lafarge, C. Mallet, Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation, International journal of computer vision, 99 (1) (2012) 69-85.

[59] Y. Arayici, An approach for real world data modelling with the 3D terrestrial laser scanner for built environment, Automation in Construction, 16 (6) (2007) 816-829.

[60] N. Haala, M. Peter, A. Cefalu, J. Kremer, Mobile LiDAR mapping for urban data capture, in: Proceedings of the 14th International Conference on Virtual Systems and Multimedia, Limassol, Cyprus, 2008, pp. 95100.

[61] C. Sahin, A. Alkis, B. Ergun, S. Kulur, F. Batuk, A. Kilic, Producing 3D city model with the combined photogrammetric and laser scanner data in the example of Taksim Cumhuriyet square, Optics and Lasers in Engineering, 50 (12) (2012) 1844-1853.

[62] MOHURD, Design standard for energy efficiency of residential buildings in hot summer and cold winter zone JGJ 134-2010 (in Chinese), in, 2010.

[63] F. Qi, Y. Wang, A new calculation method for shape coefficient of residential building using Google Earth, Energy and Buildings, 76 (2014) 72-80.

[64] F. Qi, J.Z. Zhai, G. Dang, Building height estimation using Google Earth, Energy and Buildings, 118 (2016) 123-132.

[65] A. Bhatnagar, W. Livingston, Fundamentals of solar astronomy, World Scientific, Singapore 2005.

[66] J. Parasonis, A. Keizikas, D. Kalibatiene, The relationship between the shape of a building and its energy performance, Architectural Engineering and Design Management, 8 (4) (2012) 246-256.

[67] T. Catalina, J. Virgone, V. Iordache, Study on the impact of the building form on the energy consumption, in: Proceedings of building simulation, 2011.

[68] P. McKeen, A. Fung, The Effect of Building Aspect Ratio on Energy Efficiency: A Case Study for Multi-Unit Residential Buildings in Canada, Buildings, 4 (3) (2014) 336-354.

[69] L. Wei, W. Tian, J. Zuo, Z.-Y. Yang, Y. Liu, S. Yang, Effects of Building Form on Energy Use for Buildings in Cold Climate Regions, Procedia Engineering, 146 (2016) 182-189.

[70] M.N. Inanici, F.N. Demirbilek, Thermal performance optimization of building aspect ratio and south window size in five cities having different climatic characteristics of Turkey, Building and Environment, 35 (1) (2000) 41-52.

[71] MOHURD, Code for design of civil buildings GB 50352—2005, in, Ministry of Housing and Urban-Rural Development, People's Republic of China., 2005.

[72] R. Sambandam, Cluster analysis gets complicated, Marketing Research, 15 (1) (2003) 16-21.

[73] W. Lu, SPSS for Windows: Statistical Analysis Guide, Publishing House of Electronics Industry, Beijing, 2006.

[74] M.M. Mukaka, A guide to appropriate use of Correlation coefficient in medical research, Malawi Medical Journal : The Journal of Medical Association of Malawi, 24 (3) (2012) 69-71.

[75] G. Sheikholeslami, S. Chatterjee, A. Zhang, Wavecluster: A multi-resolution clustering approach for very large spatial databases, in: VLDB, 1998, pp. 428-439.

[76] P.-N. Tan, Introduction to data mining, Pearson Education India, 2006.

[77] A. Flynt, N. Dean, A Survey of Popular R Packages for Cluster Analysis, Journal of Educational and Behavioral Statistics, 41 (2) (2016) 205-225.

[78] R.I. Kabacoff, R in Action, Second Edition ed., manning, New York, 2015.

[79] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set, 2014, 61 (6) (2014) 36.

[80] Sustainable Design Lab, Umi, in, 2017.

[81] DOE. EnergyPlus Energy Simulation Software. 2017;

http://apps1.eere.energy.gov/buildings/energyplus/ (Accessed 2017.12.14)

[82] ASHRAE, Standard 90.1-2013, Energy standard for buildings except low-rise residential buildings. , in, ASHRAE, Atlanta, 2013.

[83] T. Dogan, C. Reinhart, Automated conversion of architectural massing models into thermal ‘shoebox’models, Proceedings of BS2013, (2013).

[84] T. Dogan, C. Reinhart, P. Michalatos, Autozoner: an algorithm for automatic thermal zoning of buildings with unknown interior space definitions, Journal of Building Performance Simulation, 9 (2) (2016) 176-189.

[85] T. Dogan, C. Reinhart, Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation, Energy and Buildings, 140 (2017) 140-153.

[86] T. Hu, H. Yoshino, Z. Jiang, Analysis on urban residential energy consumption of Hot Summer & Cold Winter Zone in China, Sustainable Cities and Society, 6 (2013) 85-91.