Generating Prototypical Residential Building Geometry Models Using a New Hybrid Approach

Deng Wu, Yuanli Ma, Jing Xie, Tim Heath

Abstract

Building prototyping has regularly been used in building performance analyses with statistically feasible models. The novelty of this research involves a new hybrid approach combining stratified sampling and k-means clustering to establish building geometry prototypes. The research focuses on residential buildings in Ningbo, China. Seventeen small residential districts (SRDs) containing 367 residential buildings were systemically selected for survey and data collection. The stratified sampling used building construction year as the main parameter to generate stratification. Floor numbers, shape coefficients, floor areas, and window-to-wall ratios were used as the four observations for k-means clustering. Based on this new approach, nine building geometry prototypes were identified and modelled. These statistically representative prototypes provide building geometrical information and characteristic-based evaluations for subsequent building performance analysis.

Keywords: *building prototyping, geometry models, new hybrid approach, Ningbo China*

1. Introduction

Generally, building typology deals with a classification of parameters commonly found in buildings, such as building sizes, building styles, and construction materials (Loga, 2011). Although the typology is predetermined, it has a dialectical relationship with technology, history, polity, social proceedings, building individuality, and commonality (Theodoridou et al., 2011). Lang et al. (2018) suggest that building typology is a prerequisite for defining a thorough building classification scheme. By definition, a prototype is a copy or imitation following the object sample (Arís, 1993). Each geographical region can contain one or more typical prototypical buildings representing the limited building types within that region (Ye et al., 2019).

Typology is the study of types or common elements within a particular field of study or discipline. For example, in architecture, it can be used to organise and classify either form or function of buildings. Indeed, in past decades, building typology has been widely used to create a harmonised structure of different building stock scenarios (Dascalaki et al., 2011; Kragh and Wittchen, 2014). The development of building typology can be briefly summarised into three stages. The first stage mainly involves the classification of individual buildings by the design concept associated with architectural forms, internal layouts, and functions. For example, Durand (1799) established architectural types by classifying and integrating building layout, elevation, and basic components into geometric figures. He proposed 72 categories of

building types in the book *Compilation and Comparison of Various Ancient and Modern Materials (Recueil Et Parallèle Des Édifices de Tout Genre Anciens Et Modernes)*, which was considered a pioneering work in the field. In the second stage, the classification scope can be expanded to cities and towns instead of individual buildings. Aldo Rossi (1982) initiated new building typology research in his work *The Architecture of the City*, expanding the research objects from individual buildings to the forms of urban elements, the organisation of urban spaces, and dwellers' living habits. In the third stage, a further expansion enabled research to focus on building design and performance evaluation, such as thermal comfort, acoustics, and optics, rather than only the architectural forms. For example, in the book *The Architecture of the Well-tempered Environment*, Reyner Banham (1984) analysed building typologies from the perspective of users, including daylighting, heating, ventilation, and air-conditioning systems.

Unlike typology, a prototype is an original model of something from which other forms can be developed or copied. Building prototyping is a significant development from building typology, representing a standard or a type of building. Typology can be considered more as a process, while the prototype is more of an outcome. Retrospectively, the first prototypical building models were proposed by Synergic Resource Corporation in the 1980s from the study of ten office buildings that were subgrouped as new or existing buildings (Shahrestani et al., 2014). Prototypical building models can be divided into three categories: econometric and technological models, physical measurement models, and statistical models. The econometric and

technological model was mainly developed during the energy crisis of the late 1970s and was used to help formulate national energy plans (de Sa, 1999). This model is not geometric, as it attempts to explain the changing relationship between energy use and the adjustment of energy supply and energy prices. The model is usually designed to examine the relationship between large-scale, long-term energy consumption or carbon dioxide emissions and urban economic and social parameters, such as GDP development (Natarajin, 2007). The physical model and the statistical model can be used to analyse the correlation between building characteristics and associated building performance, such as thermal comfort, energy consumption, and life cycle costs (Hong et al., 2020). In urban building energy consumption analysis, modelling techniques based on physical characteristics are widely used. Most physical measurement models calculate each device's terminal energy consumption based on the type, quantity, and power usage characteristics of the building equipment to estimate the building's energy consumption (Enkvist et al, 2007). For example, Granade et al. (2009) use physical characteristics such as construction equipment components, indoor temperature settings, building shape, and innovative design technologies to investigate the energy consumption impact of renovation and transformation technologies. The statistical model relies on various input parameters to obtain impact parameters by mathematical methods (Zhang, 2004). These data are mainly derived from building energy bill data and field survey results (Swan et al., 2009; Kecebas and Yabanova, 2012) to provide flexible and reasonable predictions using macro data simulation (Booth et al., 2012; Wilson and Swisher, 1993). However,

each type of building prototype model has limitations, which are summarised in Table

1.

Model types	Econometric and Technological Modelling	Statistical Modelling	Physical Measured Modelling
Data sources	Long-term statistics, and macro data for model calibration.	Large and integrity existing data samples.	Field measurements.
Sample scales	Large-scale sampling at district or city level.	Can be used for both large-scale and small-scale study.	Usually at a small-scale sampling, such as buildings from one plot.
Modelling results	The relationship between macroeconomic index and energy demand.	The impact of economic and user behavior on energy consumption.	The real effect of different technology packages.
Limitations	Only supports macro and large-scale forecasting, and relies on long-term statistics. Not suitable for simulating the performance of a specific technology and user behavior.	Reliance on data integrity and large data samples. Limited predictive ability for new building performance.	High computational cost, large amount of manpower and financial expenditure. Long-term data acquisition is needed for a large-scale survey.

Table 1. Advantages and limitations of different building prototyping models

In China, there is a lack of building prototyping research for its residential building stock. Earlier studies have mainly focused on either urban spatial layouts (e.g., Li and Yang, 2019; Zhang et al., 2019) or building energy assessments and indoor environmental quality (e.g., Ma, 2017; Hong et al 2020). For example, Zhou (2008) investigated the energy consumption of different building types in Chongqing. Dong (2013) employed building construction years, building types, and space layout as parameters to classify and summarise residential buildings in Guangzhou. Additionally, Hong et al. (2020) used a hierarchical method to investigate low-rise office buildings in Shanghai based on building dimensions and floor areas.

Most of the earlier studies adopted a small-scale survey of building samples, which was not statistically feasible (SRC, 1985; Akbari et al., 1989; Hernandez, Burke, and Lewis 2008; Hong et al 2020). This would make the building prototypes less representative of the entire building stock within the study area and create difficulties in developing a building benchmark. Significantly, investigating thousands of buildings in a Chinese city involves an enormous input of time and resources, which is impossible in many cases. This limitation has promoted the need for an efficient approach to generate specific prototypical building information through a wider range of building samples. To fill this gap, this research proposes a hybrid classification approach to establish prototypical building geometry models. The proposed approach involves four steps: site selection; building sampling; classification parameter selection; and k-means clustering, which is the core of the research (see Fig. 1). In this research, 367 residential buildings from 17 selected small residential districts (SRDs)

in Yinzhou District, Ningbo, China, were investigated as examples to expound the approach.
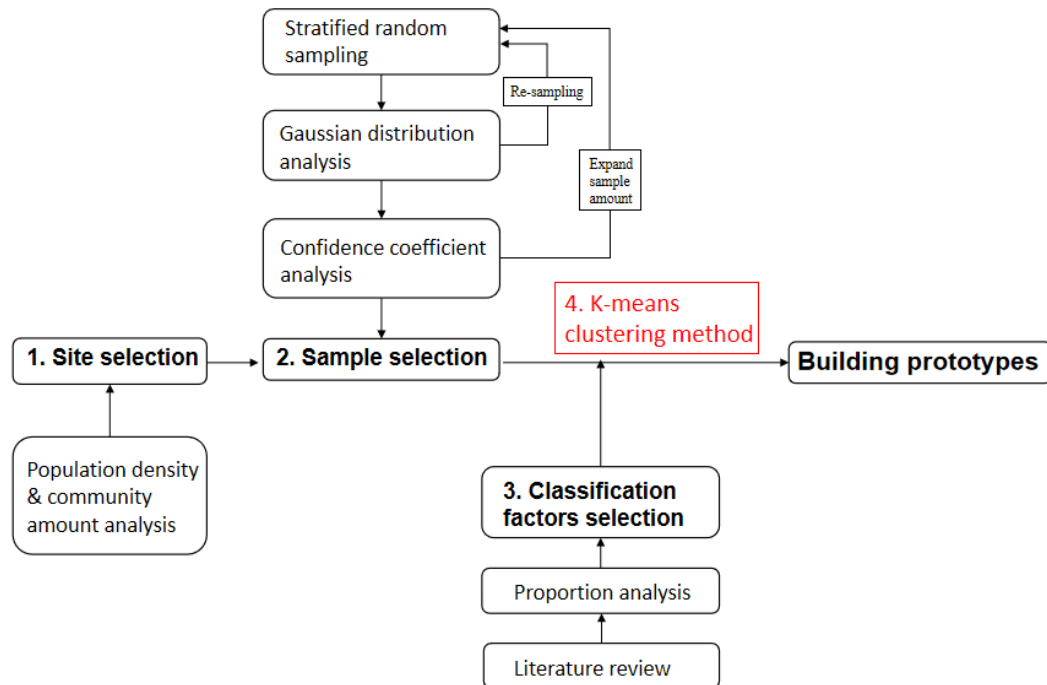

Fig. 1. Research process flow chart

2. Data Collection and Analysis

2.1 Site Selection

Ningbo is an important port city along the southeast coast of China and the economic centre south of the Yangtze River Delta (see Fig. 2). The city covers an area of $9,816 km^2$ with a population of 8.542 million (Ningbo Statistics Bureau, 2020). Climatically, Ningbo is located in the hot summer and cold winter zone in China. It is a transition zone between the cold and hot climates, characterised by high humidity, heating demands in winter, cooling demands in summer, and inadequate solar flux.
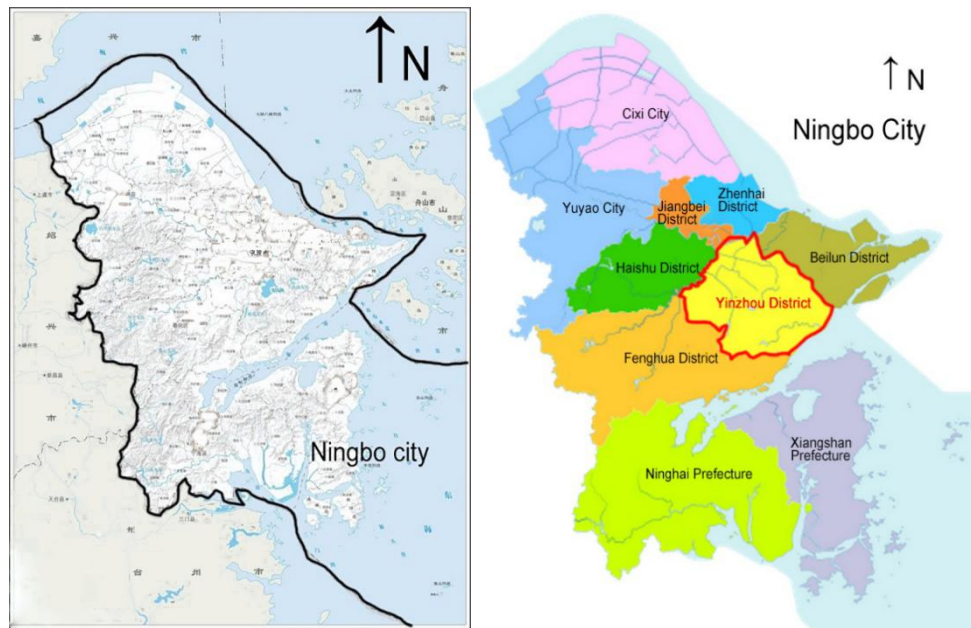
Fig. 2. Ningbo city (left) and Yinzhou District (right)

The municipal area of Ningbo is comprised of six districts, two prefectures, and two county-level cities. In this research, the small residential district (SRD, also known as "xiaoqu" in Chinese) was used as the basic spatial unit for the investigation. Designed by professional planners and architects, an SRD is a planned neighbourhood where residential buildings are integrated with communal facilities such as kindergartens, convenience shops, and communication infrastructure, all operated by a professional property management company. Currently, a significant proportion of urban residents in China live in an SRD (Bray 2006). Generally, the geographic boundaries and internal layouts of an SRD can be identified from Google Maps. In addition, data related to the number of SRDs in Ningbo, including the locations, the number of floors, and the number of residential buildings within an SRD were obtained from Anjuke's online database (Anjuke is a distinguished real estate information service

group with over 69 million independent users monthly). Table 2 shows the demographic and SRD information of each urban district in Ningbo. In total there are 4,775 SRDs in the city. Yinzhou is one of the six urban districts in Ningbo, which has the largest number of SRDs, approximately four times larger than Jiangbei District or Zhenhai District. In addition, Yinzhou has a long history of development, with the earliest SRD dating back to 1990. With many new SRDs developed in the last ten years, Yinzhou provides sufficient and diverse residential building types for sampling and analysis. Therefore, it is reasonable to focus the research on the Yinzhou District for further investigation.

Districts	Land area (km^2)	Number of permanent residents (*thousand people*)	Population density (*person/km^2*)	Number of SRDs
Haishu District	595	90.9	152.8	880
Jiangbei District	208	37.5	180.3	335
Beilun District	599	66.8	111.5	480
Zhenhai District	246	44.3	180.1	325
Yinzhou District	814	129.4	159.0	1,225
Fenghua District	1,268	51.1	40.3	240
Yuyao City	1,501	109.5	72.9	430
Cixi City	1,361	150.2	110.4	495
Xiangshan Prefecture	1,382	52.6	38.1	220
Ninghai Prefecture	1,843	68.2	37.0	145
Total	9,816	800.5	81.6	4,775

Table 2. Regional population and building information for Ningbo (source: Ningbo Municipal People's Government, 2019)

2.2   Sampling

The number of SRDs in Yinzhou District (1,225) is too large to investigate; therefore, a stratified random sampling method was used to select the SRDs and their buildings. The stratification method is a microcosm of the population where each element has the same selection probability. The age of the buildings has a significant impact on their performance. Buildings constructed in the same time frame usually have similar thermal characteristics (such as the heat transfer coefficient of the building envelope, solar heat gain coefficient of windows, and airtightness) since they had to meet the building regulations issued by the government concurrently (Theodoridou et al., 2011; Hong et al., 2020). Generally, building regulations are updated every five to ten years in China. Hence, this research adopted the major revision for building regulations, *Design Standards for Energy Conservation of Residential Buildings in Hot Summer and Cold Winter Climatic Zones (JGJ134),* as the preliminary basis for its stratified random sampling of all SRDs in Yinzhou District. Since its enactment, JGJ134 has stipulated boundary design standards for residential buildings in hot summer and cold winter (HSCW) zones; it was updated in 1993, 2005, and 2015 (MOHURD, 2020). Therefore, the SRDs were divided into four time frames: before 1993; from 1994 to 2005; from 2006 to 2015; and after 2016.

The 1,225 SRDs in Yinzhou District were then assembled for further statistical analysis. The construction year of the SRDs and their proportions were used as coordinates to derive the SRD distribution in Ningbo. The results (see Fig. 3) show

that the building distribution conforms with the normal distribution. Therefore, each SRD can be considered a statistically independent normal random variable.
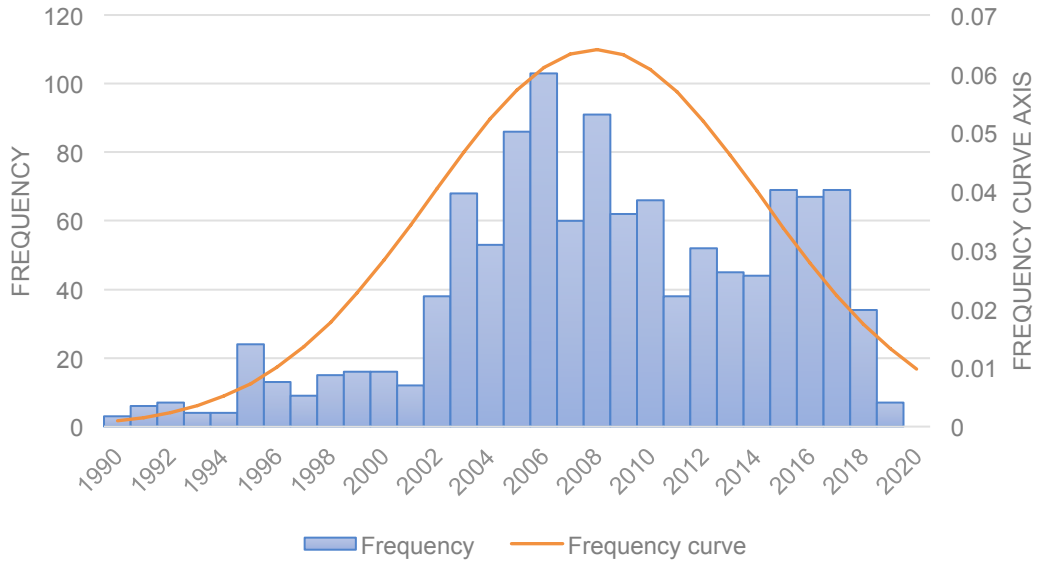

Fig. 3. SRD distribution from 1990 to 2020

In statistics, the confidence level refers to the authenticity degree of an individual parameter to a given proposition (Fisher, 1956). This research was carried out at the municipal level, and a 95% confidence level and a 10% allowable error (see examples from Freire et al., 2019; Suchmacher and Geller, 2012) were selected for the prototyping sample selection. As calculated using Equation 1, at least 17 SRD samples were needed to make them statistically representative. Then, the number of SRDs for each stratification in the Yinzhou District was determined using Equation 2.

$$e = 10\% \times (2020 - 1990) = 3; \quad n = \frac{z^2 \sigma^2}{e^2 + \frac{z^2 \sigma^2}{N}} = 16.27 \quad \text{(Equation 1)}$$

where,

z is the corresponding z-value; σ is the standard deviation; e is the sample

allowable error; N is the total sample size (all SRDs in Ningbo city); and n is the selected sample size (the number of SRDs in Yinzhou District).

$n_s = fn$ (Equation 2)

where,

n_s is the number of SRDs of each stratification that needs to be selected; and f is the ratio of the number of SRDs of each stratification from Yinzhou District to all SRDs in Ningbo.

Given that the distribution of buildings after 2015 is slightly more than the normal distribution, two more buildings constructed after 2015 were added as a supplement in the analysis. The locations of the 17 selected SRDs are illustrated in Fig. 4.

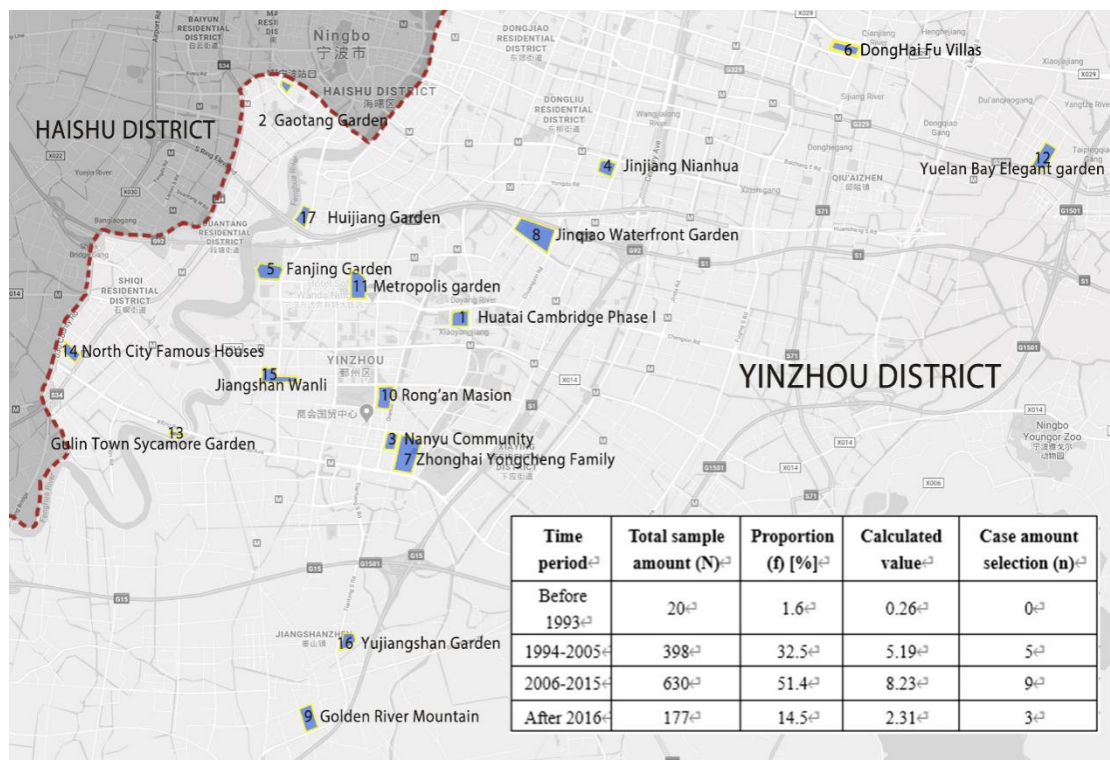| Time period	Total sample amount (N)	Proportion (f) [%]	Calculated value	Case amount selection (n)
Before 1993	20	1.6	0.26	0
1994-2005	398	32.5	5.19	5
2006-2015	630	51.4	8.23	9
After 2016	177	14.5	2.31	3

Fig. 4. The locations of the 17 SRDs selected in Yinzhou District

The buildings' geometrical information within the selected SRDs, such as orientation, building size, and building shape, was collected from the blueprints provided by the Yinzhou Building and Urban-Rural Development Bureau. In addition, to ensure a practical and accurate classification, three strategies were considered for simplifying the building prototypes:

- Insignificant minor details (e.g., a decorative feature on the surface, attached features, and balconies) were ignored.

- The irregular buildings with complicated forms were disassembled into smaller separate simple geometries.

- All the residential buildings were represented by a combination of simple geometries using their dimensions of height (h), length (l) and depth (d).

The data collection was conducted between 25 September 2019, and 31 May 2020. After an initial screening, 367 residential buildings in these 17 SRDs were selected. The statistical analysis of these buildings is provided in Table 3.

Building types	Number of buildings	Mean value					
		l (*m*)	d (*m*)	h (*m*)	Cf	Number of building floors	Number of dwelling units
Low-rise buildings	179	32.99	8.67	9.37	0.39	2.94	3.62
Multi-storey buildings	62	42.22	13.53	15.96	0.37	5.02	4.19
Mid-rise buildings	39	43.10	12.02	21.86	0.29	7.05	4.00
High-rise buildings	87	36.54	16.75	63.77	0.19	20.57	5.75

Note: "l" is building length, "d" is building width, "h" is building height, and "Cf" is building shape coefficient. According to *GB50352-2015*, a low-rise building has one to three storeys, a multi-storey building has four to six storeys, a mid-rise building has seven to nine storeys, and a high-rise building has over ten storeys,.

Table 3. Statistical information of the 367 residential buildings investigated

2.3 Selection of the Classification Parameters

In the first stage of building prototyping development, a single parameter was used to classify the building stock. For example, in Germany, the Federal Statistic Authority

classifies residential building stock into eight classes by their construction years (German Federal Statistic Authority, 2006). Similar categorisation schemes were also implemented in Switzerland (Hassler, 2009), England (English Housing Survey, 2010), Italy (Ballarini and Corrado, 2009), and Denmark (Danish Energy Agency, 2007; Tommerup and Svendsen, 2006; Wittchen, 2009). In recent years, scholars have gradually adopted multiple parameters for prototyping buildings. Theodoridou et al. (2011), for example, classified the dominating multifamily building prototypes in Greece by construction year, material, and building type. Dascalaki et al. (2011) performed a similar classification to evaluate various energy conservation measures. Furthermore, Santos et al. (2013) selected building types, internal systems of access, building layouts, and risk mitigation as parameters to present a synthesis of construction forms in the old city centre of Seixal, Portugal. The classification parameters in this research were based on an extensive review of previous studies, combining the keywords "prototype" and "residential building" for searching relevant research projects. These studies included peer-reviewed articles from journal databases and emerging research databases (e.g., Passive House Institute). Additionally, other sources consulted included academic books, dissertations, government documents, and conference proceedings.
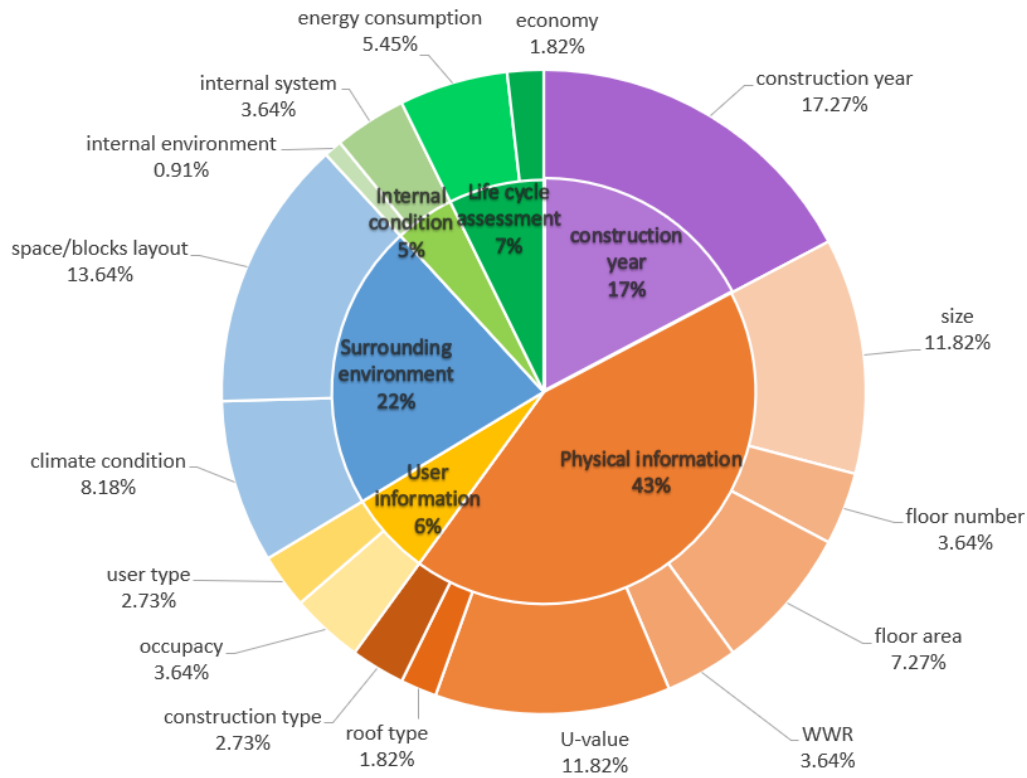
Fig. 5. Proportion of residential building classification parameters based on an extensive literature review

In total, fifty-eight publications were identified, and an analysis of various building classification parameters adopted in these studies and associated frequencies of use can be seen in Fig. 5. These parameters can be grouped into six main areas (see the inner circle in Fig. 5): building construction year; building physical information; user information; surrounding environments; internal conditions; and life cycle assessment. Among them, building physical information was most commonly used for building classification. More specifically, building physical information includes the general building features (e.g., window-to-wall ratio and shape coefficient), geometric data, and thermal properties of the envelope. The Typology Approach for Building Stock Energy Assessment (TABULA) is an Intelligent Energy Europe Programme (IEE) project to create a harmonised structure for European residential buildings. The

TABULA project used *construction year*, *building size*, *type and age of the supply system*, and *regional location* as classification parameters. The construction year was already selected for SRD stratified sampling in this research. Given that our project's research aim was to develop building geometry models, the classification parameters proposed were the number of floors, shape coefficient, average floor area, and window-to-wall ratio.

● Floor numbers:

A building can be divided into different types according to its number of floors, which is affected by each floor's height limitation and the whole building. Residential buildings can be categorised by the number of floors based on the *Uniform Standards for Residential Building Design (GB 50352—2005)*. According to this document, there are four types: low-rise buildings (1-3 stories); multi-storey buildings (4-6 stories); mid-rise buildings (7-9 stories); and high-rise buildings (over 10 stories).

● Shape coefficient (C_f):

The building surface area and volume determine the building shape coefficient. Buildings can also be classified into point-style ($\frac{l}{d} < 2$) and stripe-style ($\frac{l}{d} > 2$) buildings by the ratio of length (l) to depth (d) with different C_f limitations.

● Average floor area:

The average floor area of a building is selected as the primary form of floor area, an essential building characteristic widely used for building typology (Wan and Yik, 2004; Peri et al., 2013). Each floor's building layout is similar in a residential building;

thus, it is mainly determined by its length and depth.

● Window-to-wall ratio (WWR):

WWR is the ratio of glazed area to the building elevation area. This parameter has a significant effect on building energy consumption performance. This parameter classification is based on the *Design Standards for Energy Conservation of Residential Buildings in Hot Summer and Cold Winter Climatic Zones (JGJ134)*, which gives the boundary values for different building facades depending on the orientation.

2.4 K-means Clustering

As a basic machine-learning algorithm, the k-means method can effectively realise the clustering of the population with multiple defined variables (Shi et al, 2020). This research used k-means arithmetic methods to perform clustering based on the selected building samples' classification parameters. The elbow method was then used to discriminate the optimal cluster number expressed by distortion. This method aims to minimise the squared error between the sample and the centroid. The distortion degree is the sum of the square of the distance error between each sample point and the centroid in the same cluster. A smaller distortion means a higher degree of aggregation and vice versa. In addition, when the selected cluster amount (k) value is insufficient, the distortion will sharply decrease when k increases by one. After a certain critical point, the distortion degree would be vastly improved and decrease

slowly. This critical point can be determined by the elbow method to select the cluster amount. At that time, the information of the centroids generated can be directly used for prototype modelling. The distortion of the k value can be calculated using Equation 3:

$$D_k = \sum_{k=1}^{k} \sum dist(x, c_k)^2 \quad \text{(Equation 3)}$$

where,

D_k is the distortion of its corresponding k value;

$\sum dist(x, c_k)^2$ is the sum of the distances' square of each point to the cluster centroid;

x is a clustering parameter that can be considered as one variant.

In this research, four parameters were proposed for classifying and developing building geometry models. These parameters were treated as multidimensional vectors, which were also called observations in the k-means clustering.

Following the steps for determining the centroid, multidimensional k-means clustering is generated for prototyping:

1. Select the k initial centroid randomly, and each centroid is determined as a cluster.

2. Assign each observation to its nearest centroid and form a new cluster.

3. Recalculate a new centroid of each cluster. This new centroid is the average vector of all observations in a cluster.

4. Repeat step two and step three until the centroid does not change or reach the maximum number of iterations.

3. Clustering and Prototyping Results

By calculating the distortion against the number of clusters using Python software, the distortion value converges after k reaches nine (see Fig. 6). This means that at least nine clustering groups need to form for this sampling population. The centroid value of each observation can represent the characteristics of nine clusters.
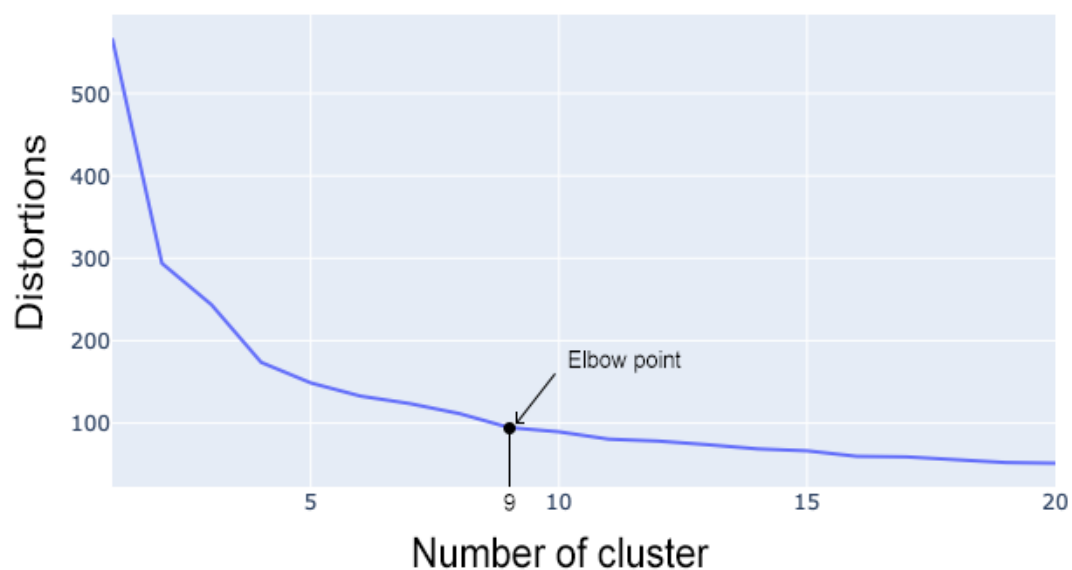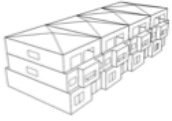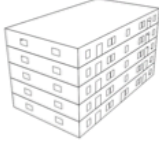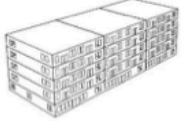

Fig. 6. Results of the elbow method analysis using Python software
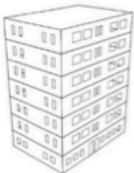
A small number of outliers appeared during the clustering process. These outliers were insufficient to form a new cluster and had a weak impact on the cluster results. Table 4 indicates the nine building models generated by using the IES-VE software and the associated building characteristics. These models comprised three low-rise prototypes: two multi-storey building prototypes, one mid-rise prototype, and three

high-rise building prototypes.

Clusters	Building types	Average floor area (m^2)	Length (m)	Width (m)	Height (m)	Number of floors	Model scheme diagrams
Model 1	Low-rise building; detached; point-style	132.88	14.62	9.09	8.99	2.90	
Model 2	Low-rise building; bilocular; stripe-style	219.47	26.65	8.23	9.23	3.00	
Model 3	Low-rise building; terraced; stripe-style	572.36	50.79	11.27	9.75	3.00	
Model 4	Multi-storey building; point-style	375.75	29.31	12.82	14.91	4.89	
Model 5	Multi-storey building; strip-style	745.69	52.85	14.11	16.83	5.20	
Model 6	Mid-rise building; stripe-style	499.34	43.54	11.47	21.62	7.09	

Model 7	High-rise building; point-style	414.12	26.67	15.53	32.45	10.64
Model 8	High-rise building, strip-style	631.08	39.10	16.14	52.92	17.35
Model 9	High-rise building, strip-style	658.91	35.93	18.34	96.23	31.55

Note: The data listed above is clustering results of each centroid; the data for the number of storeys were rounded as an integer when generating building models.

Table 4. Residential building information of prototyping models

The most significant factors affecting prototyping are the average floor area for low-rise buildings and multi-storey buildings, the building height, and the number of floors for high-rise buildings. Additionally, the shape coefficient of the low-rise prototypes is generally larger than that of the other prototypes. Table 5 shows a comparison between the shape coefficients and WWRs of the building models and the specifications required in *JGJ134* (2010). The nine prototypes all met those requirements, which means that the clustering results were reasonable, and the prototypical building models were statistically representative of the collection of buildings being investigated. They are therefore useful as reference buildings to further evaluate the building performance and effectiveness of building policies.

Models	Model information					Specifications in JGJ134			
	C_f	WWR				C_f	WWR		
		E	S	W	N		E, W	S	N
Model 1	0.47	0.08	0.18	0.31	0.28	0.55	0.35	0.45	0.40
Model 2	0.41	0.09	0.28	0.08	0.23	0.55	0.35	0.45	0.40
Model 3	0.32	0.10	0.27	0.07	0.22	0.55	0.35	0.45	0.40
Model 4	0.38	0.09	0.39	0.08	0.20	0.40	0.35	0.45	0.40
Model 5	0.36	0.05	0.25	0.05	0.16	0.40	0.35	0.45	0.40
Model 6	0.29	0.04	0.39	0.03	0.19	0.40	0.35	0.45	0.40
Model 7	0.24	0.09	0.31	0.07	0.23	0.40	0.35	0.45	0.40
Model 8	0.19	0.09	0.35	0.09	0.25	0.35	0.35	0.45	0.40
Model 9	0.17	0.17	0.52	0.19	0.40	0.35	0.35	0.45	0.40

Table 5. Model shape coefficients and WWR ratios and related regulatory specifications

4. Discussion

Residential buildings with different construction periods, shapes, or equipped with various appliances and building systems may perform differently with the same design technologies (Filogamo et al., 2014). Significantly, the modern built environment has become more complex in terms of building typologies and environmental systems. This complexity causes difficulties in selecting appropriate technologies to optimise building performance for different building typologies. The geometric characteristics of residential buildings in different regions can also be influenced by local contextual conditions (e.g., climate, culture, building regulations).

However, they tend to have similar geometric characteristics if constructed in the same time frame and the same city. As a case study for building prototyping, the Yinzhou District comprises the largest number of SRDs in Ningbo and covers a wide range of residential buildings built in different time frames. The nine building models generated from the research are based on a statistical analysis of the SRDs and buildings that were investigated.

Importantly, this research proposes a new hybrid approach for generating prototypical building models. It is a bottom-up approach that involves building geometrical parameters for clustering and prototyping. Compared with other prototyping approaches, the hybrid approach is more convenient, effective, and can be widely used. It does not need long-term statistics and can generate statistically reliable building models. Regarding accuracy, the plotted curve is smooth, as shown in Fig. 6, which is subjective when determining the optimal cluster number corresponding to the elbow point. The change rate of the distortion of k from nine to ten is 1%, which means that the nine models developed are statistically acceptable.

Based on a new hybrid approach, this research has developed nine prototypical building geometry models. The nine models can simultaneously express their corresponding building populations' characteristics in terms of geometric characteristics, which are related to building performance. Additionally, the approach

can be expanded to other cities to generate prototypical local residential building models. In short, this research:

- Provides an effective way to develop statistically sound building models for evaluating building performance using simulation software. For example, retrofitting strategies can be customised and optimised for a particular building prototype.

- Generates a database for the building geometric characteristics of building stock.

- Provides a reference for undertaking life cycle assessments for new city block subdivisions and city regeneration, which includes changes in the proportion of regional buildings and the adoption and optimisation of regional functional systems.

5. Conclusion

This research proposes a hybrid approach for developing building geometry models that combine statistical, mathematical and physical measurement methods. In this research, nine residential building models were identified within the case study city of Ningbo, China. With these building models, further performance evaluation can be performed with simulation software by adding additional information such as solar information, building system information, and occupancy profiles.

This study is part of a research project to provide an integrated framework for developing specific retrofitting strategies for residential buildings. The establishment

of building models will facilitate further building performance evaluation and optimisation. Future research will enable the existing building prototyping models to be further improved or subdivided through more data support. A sensitivity analysis of the correlation between various building technologies and associated energy performance can be performed using these building models, which can help provide planning assistance for future sustainable city development. The challenges of selecting optimised design technologies for different building types again highlights the importance of creating building prototypes.

Author contribution statement

Wu Deng conceived and designed the research analysis and revised the paper. Yuanli Ma contributed to the research analysis design and partook in data collation and processing and wrote the paper. Jing Xie, Professor Tim Heath and Yuanda Hong contributed to the design of the research analysis and paper revision. Yeyu Xiang contributed to the data support.

References

Arís, C. M. (1993). Las variaciones de la identidad. Ensayo sobre el tipo en arquitectura. Universitat Politècnica de Catalunya (UPC).

Akbari, H., J. Eto, S. Konopacki, and A. Afzal. (1994). Integrated Estimation of Commercial Sector End-Use Load Shapes and Energy Use Intensities. Berkeley, CA: Lawrence Berkeley Laboratory for Pacific Gas and Electricity Company.

Asimakopoulos, D., Santamouris, M., Farrou, I., Laskari, M., Saliari, M., Zanis, G., Giannakidis, G., Tigas, K., Kapsomenakis, J. & Douvis, C. (2012). Modeling the Energy Demand Projection of the Building Sector in Greece in the 21st Century. *Energy and Buildings,* 49, 488-498.

Ballarini, I. & Corrado, V. (2009). Application of Energy Rating Methods to the Existing Building Stock: analysis of some residential buildings in Turin. *Energy and Buildings,* 41, 790-800.

Banham, R. (1984). *The Architecture of the Well-tempered Environment.* The University of Chicago Press; 2nd. ISBN-10: 0226036987; ISBN-13: 978-0226036984.

Booth, A., Choudhary, R., & Initiative, E. E. C. (2012, February). Calibrating micro-level models with macro-level data using bayesian regression analysis. In *Proceedings of the 12th IBPSA Conference, November* (pp. 14-16).

Bray, D. (2006). Building 'Community': new strategies of governance in urban China. *Economy and Society*, *35*(4), 530-549.

Code for Design of Civil Buildings (GB50352-2005), Chief editorial department: Ministry of construction of the People's Republic of China, Approved department: Ministry of Construction of the People's Republic of China, date of implementation: July 1st, 2005.

Danish Energy Agency. (2007) The energy in Denmark, National Building Research Institute, Aalborg University, Department of Energy and Environment, *Energy Consumption and Savings* (in Danish).

Dascalaki, E. G., Droutsa, K. G., Balaras, C. A. & Kontoyiannidis, S. (2011). 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, 3400-3409.

de Sa, P. (1999). Mapping the Energy Future: energy modeling and climate change policy. *Energy Policy*, *27*(7), 433-434.

Design Standard for Energy Efficiency of Residential Buildings in Hot Summer and Cold Winter Zone (JGJ134-2010). Approved department: Ministry of Construction of the People's Republic of China, date of implementation: August 1st 2010.

Dong, Y. (2013). Research on the Development of Residential Buildings in the Centre City of Contemporary Guangzhou. The South China University of Technology.

Durand. J. (1799). *Recueil Et Parallèle Des Édifices de Tout Genre Anciens Et Modernes (in French*). Hachette Livre – BNF. ISBN-10: 2013600259; ISBN-13: 978-2013600255.

English Housing Survey (2010) Department for Communities and Local Government, Headline Report 2008-09, London.

Enkvist, P., Nauclér, T., & Rosander, J. (2007). A Cost Curve for Greenhouse Gas Reduction. *McKinsey Quarterly*, *1*, 34.

Filogamo, L., Peri, G., Rizzo, G. & Giaccone, A. (2014). On the Classification of Large Residential Buildings Stocks by Sample Typologies for Energy Planning Purposes. *Applied Energy,* 135, 825-835.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, Hafner Publishing Company.

Freire, A. P. C. F., Elkins, M. R., Ramos, E. M. C. & Moseley, A. M. (2019). Use of 95% confidence intervals in the reporting of between-group differences in randomized controlled trials: analysis of a representative sample of 200 physical therapy trials. *Brazilian Journal of Physical Therapy*, 23, 302-310.

German Federal Statistic Authority, Bauen und Wohnen – Wohnsituation, (2006). Available from: http://www.destatis.de/jetspeed/portal/cms/Sites/ destatis/Internet/DE/Content/Statistiken/BauenWohnen/Wohnsituation/ Tabellen/Content75/WohneinheitennachBaujahr.psml (in German).

Granade, H. C., Creyts, J., Derkach, A., Farese, P., Nyquist, S., & Ostrowski, K. (2009). Unlocking Energy Efficiency in the US Economy. *McKinsey & Company*.

Hassler, U. (2009). Long-term Building Stock Survival and Intergenerational Management: the role of institutional regimes. *Building Research & Information,* 37, 552-568.

Hernandez, P., K. Burke, and O. Lewis. (2008). Development of Energy Performance Benchmarks and Building Energy Ratings for Non-Domestic Buildings: An Example for Irish Primary Schools. Energy and Buildings 40(3): 249‑254. DOI: 10.1016/j.enbuild.2007.02.020

Hong, Y., Ezeh, C. I., Deng, W., Hong, S. H., Peng, Z., & Tang, Y. (2020). Correlation Between Building Characteristics and Associated Energy Consumption: Prototyping low-rise

office buildings in Shanghai. *Energy and Buildings*, 109959.

Keçebaş, A., & Yabanova, I. (2012). Thermal Monitoring and Optimization of Geothermal District Heating Systems using Artificial Neural Network: A case study. *Energy and Buildings*, *50*, 339-346.

Kragh, J. & Wittchen, K. B. (2014). Development of Two Danish Building Typologies for Residential Buildings. *Energy and Buildings,* 68, 79-86.

Lang, D. H., Kumar, A., Sulaymanov, S. & Meslem, A. (2018). Building Typology Classification and Earthquake Vulnerability Scale of Central and South Asian Building Stock. *Journal of Building Engineering,* 15, 261-277.

Li, X. & Yang, X. (2019). Planning Of Ciyun Ancient Street - A Practical Exploration Based On Typo-Morphological Approach. *Urban Development Studies*, 26 (04),12-16, 21.

Loga, T. & Diefenbach, N. (Eds.), *Use of building typologies for energy performance assessment of national building stocks, existent experiences in European countries and common approach*, First TABULA Synthesis Report, June 2010. Available from: www.building-typology.eu/downloads/public/docs/report/TABULA SR1.pdf.

Loga, T., et al. (2011). Use of Building Typologies for Modeling the Energy Balance of the Residential Building Stock. *Project Partner, 2*.

Ma, Y. (2017). *Study on air-conditioning energy consumption and design optimization of a residential building in Hangzhou based on Architectural Typology*. Zhejiang University.

MOHURD. (2020). Design Standard for Energy Conservation of Residential Buildings in Hot Summer and Warm Winter Areas. Beijing, China Building Industry Press (in Chinese). [online] Available at: http://www.mohurd.gov.cn/wjfb/201508/t20150831_224551.htm l [Accessed on: 29 February 2020].

Natarajan, S., & Levermore, G. J. (2007). Domestic Futures - Which Way to a Low-carbon Housing Stock?. *Energy Policy*, *35*(11), 5728-5736.

Ningbo Statistics Bureau. (2020) Available at: http://vod.ningbo.gov.cn:88/nbtjj/tjnj/2019nbnj/indexch.htm [Accessed on: 10 March 2020]

Peri, G., Foresta, F., Inzerillo, L. & Rizzo, G. (2013). Environmentally Assessing Buildings Characterized by Complex Shape and Innovative Materials. Advanced Materials Research, 2013. *Trans Tech Publ,* 409-414.

Rossi, A., Eisenman, P., Ghirardo, D. Y. & Ockman, J. (1982). The Architecture of the City, *MIT Press Cambridge, MA*.

Santos, C., Ferreira, T. M., Vicente, R. & Da Silva, J. R. M. (2013). Building Typologies Identification to Support Risk Mitigation at the Urban scale - Case study of the old city centre of Seixal, Portugal. *Journal of Cultural Heritage*, 14, 449-463.

Shahrestani, M., R. Yao, and G.K. Cook. (2014) 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): 41-64.

Shi, C., Wei, B., Wei, S. & Wang, W. (2020) A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm. EURASIP Journal of Wireless Communications and Networking.DOI:10.21203/rs.3.rs-58011/v1.

SRC. (1985). New Office Buildings End-Use Energy Consumption Survey – for Northeast Utilities Service Company. Berlin, Connecticut: Synergic Resources Corp.

Suchmacher, M. & Geller, M. (2012). Practical Biostatistics: A Friendly Step-by-step Approach for Evidence-based Medicine, *Academic Press.*

Swan, L., Ugursal, V. I., & Beausoleil-Morrison, I. (2009). Implementation of a Canadian Residential Energy End-use Model for Assessing New Technology Impacts. *Proceedings of Building Simulation, Glasgow*, 1429-1436.

Theodoridou, I., Papadopoulos, A. M. & Hegger, M. (2011). A Typological Classification of the Greek Residential Building Stock. *Energy and Buildings,* 43, 2779-2787.

Theodoridou, I., Papadopoulos, A. M. & Hegger, M. (2011). Statistical Analysis of the Greek Residential Building Stock. *Energy and Buildings,* 43, 2422-2428.

Tommerup, H. & Svendsen, S. (2006). Energy Savings in Danish Residential Building Stock. *Energy and Buildings,* 38, 618-626.

Wan, K. & Yik, F. (2004). Building Design and Energy End-use Characteristics of High-rise Residential Buildings in Hong Kong. *Applied Energy*, 78, 19-36.

Wilson, D., & Swisher, J. (1993). Exploring the Gap: Top-down versus bottom-up analyses of the cost of mitigating global warming. *Energy Policy*, *21*(3), 249-263.

Wittchen, K. (2009). Potential Energy Savings in Existing Residential Buildings. SBi 2009: 05. Danish Building Research Institute, Aalborg University, Department of Energy and Environment.

Ye, Y., Hinkelman, K., Zhang, J., Zuo, W. & Wang, G. (2019). A Methodology to Create Prototypical Building Energy Models for Existing Buildings: A case study on US religious worship buildings. *Energy and Buildings,* 194, 351-365.

Zhang, Q. (2004). Residential Energy Consumption in China and its Comparison with Japan, Canada, and USA. *Energy and Buildings*, *36*(12), 1217-1225.

Zhou, Y (2008). Study on the Influence on Energy-saving by the Residential Plan in Chongqing. Chongqing University.