Identifying residential consumption patterns using data-mining techniques: A large-scale study of smart meter data in Chengdu, China

Jieyi
Kang

David M
Reiner

Abstract

The fine-grained electricity consumption data created by advanced metering technologies offers an opportunity to understand residential demand from new angles. Although there exists a large body of research on demand response in short- and long-term forecasting, a comprehensive analysis to identify household consumption behaviour in different scenarios has not been conducted. The study's novelty lies in its use of unsupervised machine learning tools to explore residential customers' demand patterns and response without the assistance of traditional survey tools. We investigate behavioural response in three different contexts: 1) seasonal (using weekly consumption profiles); 2) holidays/festivals; and 3) extreme weather situations. The analysis is based on the smart metering data of 2,000 households in Chengdu, China over three years from 2014 to 2016. Workday/weekend profiles indicate that there are two distinct groups of households that appear to be white-collar or relatively affluent families. Demand patterns at the major festivals in China, especially the Spring Festival, reveal various types of lifestyle and households. In terms of extreme weather response, the most striking finding was that in summer, at night-time, over 72% of households doubled (or more) their electricity usage, while consumption changes in winter do not seem to be significant. Our research offers more detailed insight into Chinese residential consumption and provides a practical framework to understand households' behaviour patterns in different settings.

Reference Details

Identifying residential consumption patterns using data-mining techniques: A large-scale study of smart meter data in Chengdu, China

Jieyi Kang and David M Reiner

Abstract The fine-grained electricity consumption data created by advanced metering technologies offers an opportunity to understand residential demand from new angles. Although there exists a large body of research on demand response in short- and long-term forecasting, a comprehensive analysis to identify household consumption behaviour in different scenarios has not been conducted. The study's novelty lies in its use of unsupervised machine learning tools to explore residential customers' demand patterns and response without the assistance of traditional survey tools. We investigate behavioural response in three different contexts: 1) seasonal (using weekly consumption profiles); 2) holidays/festivals; and 3) extreme weather situations. The analysis is based on the smart metering data of 2,000 households in Chengdu, China over three years from 2014 to 2016. Workday/weekend profiles indicate that there are two distinct groups of households that appear to be white-collar or relatively affluent families. Demand patterns at the major festivals in China, especially the Spring Festival, reveal various types of lifestyle and households. In terms of extreme weather response, the most striking finding was that in summer, at night-time, over 72% of households doubled (or more) their electricity usage, while consumption changes in winter do not seem to be significant. Our research offers more detailed insight into Chinese residential consumption and provides a practical framework to understand households' behaviour patterns in different settings.

www.eprg.group.cam.ac.uk

Identifying residential consumption patterns using data-mining techniques: A large-scale study of smart meter data in Chengdu, China[*]

Jieyi Kang[1,2] and David Reiner[1,3]

[1]Energy Policy Research Group, University of Cambridge
[2]Department of Land Economy, University of Cambridge
[3]Judge Business School, University of Cambridge

Abstract

The fine-grained electricity consumption data created by advanced metering technologies offers an opportunity to understand residential demand from new angles. Although there exists a large body of research on demand response in short- and long-term forecasting, a comprehensive analysis to identify household consumption behaviour in different scenarios has not been conducted. The study's novelty lies in its use of unsupervised machine learning tools to explore residential customers' demand patterns and response without the assistance of traditional survey tools. We investigate behavioural response in three different contexts: 1) seasonal (using weekly consumption profiles); 2) holidays/festivals; and 3) extreme weather situations. The analysis is based on the smart metering data of 2,000 households in Chengdu, China over three years from 2014 to 2016. Workday/weekend profiles indicate that there are two distinct groups of households that appear to be white-collar or relatively affluent families. Demand patterns at the major festivals in China, especially the Spring Festival, reveal various types of lifestyle and households. In terms of extreme

[*]To whom correspondence should be addressed: E-mail: jieyi.kang@hotmail.com

weather response, the most striking finding was that in summer, at night-time, over 72% of households doubled (or more) their electricity usage, while consumption changes in winter do not seem to be significant. Our research offers more detailed insight into Chinese residential consumption and provides a practical framework to understand households' behaviour patterns in different settings.

1 Introduction

Driven by continued rapid growth in urbanisation and the improvement of living standards in China, residential electricity demand has been increasing for the last two decades. Despite slower energy consumption growth from 2014, the growth rate of the residential consumption recovered in 2018 and spiked to a seven-year high of 10.4% (People's Daily, 2019). In addition, China Bureau of Statistics (2020) reported that around 60.6% of the population live in cities in 2019. The residential energy consumption is likely to continue its rapid growth. To achieve the national low-carbon plan, the Chinese government has been devoted to delivering smart energy systems, including smart meters, efficient grid operation, and improved electricity transport network. State Grid Corporation of China (SGCC) has been deploying smart meters along with data acquisition systems since 2009. By the end of 2018, more than 457 million smart meters had been installed by SGCC covering 99.57% of the customers it serves (CPNN, 2018).

In the context of promoting energy efficiency and energy conservation, it is essential for policymakers and the utilities to understand any differences in residential consumption patterns. Before the widespread rollout of smart meters, door-to-door questionnaires were the traditional method that had been widely used to identify household electricity behavioural patterns (Fu et al., 2018). However, the limitations of the method are non-negligible: To avoid statistical biases, large-scale data collection is essential, but the financial and time costs are generally very high. Meanwhile, the reliability of findings highly depends on the quality of the survey

design, sampling method and response rate. Another popular method is the case study, which is a more specific but also time-consuming approach. It often involves detailed investigation and modelling of typical dwellings, however, it is not suitable for comprehensive studies of the behaviour distribution of population since the sample is too small to be representative (An, Yan and Hong, 2018).

With greater availability of digital information and communication technologies (ICTs), especially the deployment of smart meters, obtaining high-resolution energy consumption data becomes realistic. These meters are equipped with real-time or short-interval communication capabilities that enable them to transmit fine-grained consumption information to the utilities or other data aggregators. Establishing the smart grid has fundamentally changed the communication between customers and utilities and creates opportunities for researchers and companies to understand customer demands and offer new services, such as customised tariff structures and energy-efficient demand response programmes.

The sheer volume of data available has naturally led to numerous researchers using data mining techniques in an effort to chcracterise the consumption behaviour of residential customers (Stankovic, L. et al., 2016;Rashid, H. et al., 2019). Clustering is one of the most popular techniques for identifying similar patterns and grouping them into a set of clusters. The most discussed topic in this field is how to improve load forecasting accuracy by analysing load profiles and understanding demand patterns, as well as to design demand response programmes. Both short-term and long-term load forecasting can benefit from studies using high-resolution smart meter data (Mirasgedis et al., 2006; Chaturvedi, Sinha and Malik, 2015; Atalla and Hunt, 2016). Clustering also assists in studies of tariff design, since it can divide different customers based on similar load profiles. Its objective is to separate customers by the shape of daily load profiles over time. For example, those customers with peak loads at the same time can be grouped together and offered a customised tariff (D'hulst et al., 2015; Fu et al., 2018). Some

3

studies combine clustering and questionnaire results. The characteristics of households and of the buildings themselves are common questions asked in the surveys, such as social-economic features of the household (income, education, age, gender, size of the family, etc.), and information of the house (ownership of appliances, size of house, number of bedrooms, etc.) (Firth et al., 2008; Beckel et al., 2016; Satre-Meloy, 2019). The mixed-method approach often offers a deeper and more comprehensive understanding of customer behaviour, although questionnaires are often not available due to privacy and research cost considerations. There is a large family of algorithms applied to electricity demand cluster analysis. K-means, Hierarchical and self-organising map (SOM) clustering are among those most frequently used, although there is no consensus in the literature on the application of clustering approaches (Räsänen et al., 2010; Chicco, 2012; Gouveia and Seixas, 2016). Depending on the research objectives, household electricity consumption has been clustered based on different aggregation levels, including workday/weekend profiles, seasonal profiles and/or even longer periods, such as monthly or yearly profiles (McLoughlin, Duffy and Conlon, 2012; Afzalan and Jazizadeh, 2019). Approaching consumption profiles from different angles can offer a more complete picture of how households behave under different situations and the similarities as well as discrepancies among clusters.

There has been a wide range of studies investigating the residential electricity demand profiles around the world, including in the U.S, Spain, Italy, U.K., and Denmark (Blázquez, Boogen and Filippini, 2013; Rhodes et al., 2014; Alberini et al., 2019; Andersen et al., 2019; Satre-Meloy, Diakonova and Grünewald, 2020). The characteristics of the clustered household profiles among the countries exhibit distinct patterns and are country-wise. For example, the demands of U.K. households were almost insensitive in summer but would respond to winter weather, while the opposite was found in Spain households. These differences are understandable since the profile curves reflect not only the social-economic differences of the countries,

4

but also the cultural and geographical disparities. However, the vast majority of studies using clustering methods have mainly been in advanced economies. The few studies conducted in developing countries, such as Brazil and Turkey (Dilaver and Hunt, 2011; Villareal and Moreira, 2016) have been in a more general filed investigating aggregated residential consumption, which may be due to less penetration rates of smart meters in the less developed countries.

Despite being the leading country in terms of building area, electricity consumption and smart meter deployment, there has been a notable lack of literature that examines household electricity demand profiles in China. The majority of research on China's residential electricity consumption focuses on microeconomic analysis using survey data (Zhou and Teng, 2013; Zheng et al., 2014; Du et al., 2015). The limitation of those studies is because of the reliance on aggregate consumption data. The electricity data involved are mainly monthly billing data or even based on the interviewee's recollection rather than records from the utility. This type of information clearly would not reflect households' daily consumption profiles, due to the low resolution of the data. Although a few studies have investigated residential profiles used data-mining tools, these profiles have been based on either monthly or yearly load curves (Zhou, Yang and Shen, 2017; Guo et al., 2018). A few papers use high-resolution one-minute data to form clustered load curves, but unfortunately, these samples have included electricity data from all sectors and the type of costumers are not distinguishable from the data. A key contribution of this paper is to investigate the residential consumption patterns using higher resolution data taken directly from smart meters of three intra-day periods (super off-peak (23:00-7:00), peak (7:00-11:00 plus 19:00-23:00), and off-peak (11:00-19:00)). We also examine the differences between seasonal workday/weekend profiles, festival profiles, and extreme weather profiles to provide a comprehensive overview of residential demand.

The chapter is organised as follows: Section 2 provides a review of the application of clustering to residential electricity as well as past studies of China's residential sector. The datasets

and the methodology including data preparation and clustering techniques used are presented in Section 3. Section 4 shows the comparison of two clustering approaches (K-means and Hierarchical) and a possible explanation of the results based on the cluster analysis. Section 5 provides discussion and conclusions.

2   Literature review

This section firstly surveys the application of data mining to the study of electricity demand. Then, a more specific review of the studies of electricity consumption of the residential sector in China follows. In that part, the characteristics of household consumption in literature are presented as well as the important features that affect Chinese electricity consumption behaviour. Finally, limited research based on the clustering technique is discussed.

The large-scale deployment of smart meters around the world has opened up possibilities for researchers to analyse residential electricity consumption on an unprecedented and fine-grained scale. Traditionally, high-resolution data are only stored at the grid level, which has meant that although utilities and researchers could run forecasting models at the very short term (minute or hour) level, uncertainties in those models are large. The fluctuations are often caused by differences in customer behaviour. However, more detailed consumption information for individual households was not available. Now equipped with advanced metering technologies, the scope of residential electricity demand research have broadened (Yildiz, B. et al., 2017). The technique that has attracted the most attention is residential consumption pattern clustering (Räsänen et al., 2010; Ramos et al., 2015; Gouveia and Seixas, 2016) since it allows for more detailed analysis of customer demand. There are two key technical steps needed for consumption pattern clustering: 1) Algorithm selection and 2) Cluster number decision by cluster validity index. Researchers agree that there is no single standard optimal algorithm or cluster validity index for all scenarios. K-means and its algorithm family (K-medoids, K-medians,

etc.) are the most popular method (McLoughlin, Duffy and Conlon, 2012; Al-Wakeel, Wu and Jenkins, 2016; Razavi et al., 2019) although hierarchical algorithm is another common choice (Chicco, Napoli and Piglione, 2006; Razavi et al., 2019). Possible choices of cluster validity indexes include Davies–Bouldin (DB) validity index, which measures the average similarity of each cluster with its most similar cluster (McLoughlin, Duffy and Conlon, 2012; Ozawa, Furusato and Yoshida, 2016; Viegas et al., 2016), and Silhouette scores, which defines how similar a object is to the objects in the same cluster compared to other clusters (Yilmaz et. al, 2019).

The applications using smart meter data and data mining techniques include both short-term and long-term forecasting model improvement, tariff structure design, consumption pattern modelling through electric appliance use detection, and classification of new customers. For instance, for clustering-based load forecasting, Fu et al. (2018) applied Fuzzy C-means algorithm to cluster the daily household-level data of 533 households from Quanzhou city, over the period April 2014 to February 2015 under increasing-block pricing, which achieved a high accuracy of load forecasting through better customer consumption profile classifications than the traditional K-means method; Chaturvedi et al. (2018) adopted hybrid clustering methods—Artificial Neural Network (ANN) and wavelet transform and fuzzy system on 1-hour level data from India to optimise the short-term load forecast performance. There have been a number of other load forecast studies based on clustering. Kavousian et al. (2015), which used a hierarchical algorithm on smart meter data to identify appliance energy efficiency based on a 30-min interval dataset of 4231 households in Ireland. Mahmoudi-Kohan, et al. (2010) optimised selling price to each cluster of customers to maximise the annual profits for utilities based on a profit function using a weighted fuzzy K-means algorithm. Flath et al. (2012) employed a K-means algorithm on customers in Germany and a segment-specific rate design of different prices for each group of the customers; A study by Viegas et al. (2016) combined smart meter data and survey data to classify new customers using a K-means clustering algorithm on the representative curves.

They advise that age, employment status and appliances are crucial factors for the classification.

Knowledge of domestic energy use behaviour is one of the key pieces of the puzzle to better understand Chinese residential electricity consumption. In many countries, regular surveys of household residential energy by government or other institutions provide one part of the knowledge base. For example, the US Energy Information Administration has tracked energy data of US households since 1978. In China, by contrast, similar surveys of its kind, are rare and the most comprehensive survey of residential energy consumption to date was carried out by Zheng et al. (2014) conducted in 2012, involving 1450 total observations from 26 provinces. They found that space heating and cooking were the most energy-intensive activities for the Chinese families, accounting for 54% and 23% of total energy consumption respectively. They also compared the international differences of energy consumption by end-use activities. One extraordinary difference is that the share of cooking in China is far bigger versus nearly 0% to 6% in other developed countries, such as the US and EU-27. Chinese households mainly use gas for cooking purposes and only families in Southern China use electricity for space heating. Since the survey was for all energy types used, not exclusively for the electricity, the findings may only be partial referential to the residential electricity usage in China. Zhou and Teng (2013) also used survey data to estimate the urban residential electricity in Sichuan Province. In the study, they found that both price and income were inelastic to electricity demand. The results also show that on a per capita basis, smaller households seem to consume more electricity. Another finding was that the households that included those aged 50 years or more consumed more electricity, because older people generally tend to stay at home longer. In terms of examined appliances (refrigerator, computer, TV, washing machine, and air conditioner), although refrigerators currently are the largest consumers of electricity due to the highest ownership rate, demand from air conditioners and computers will increase substantially as their penetration rate and utilisation grows.

Hekkenberg et al (2009) pointed out that geography is one of the main factors influencing electricity demand, since weather conditions are mainly determined by region. Murata et al. (2008) advances a similar argument in comparing household electricity consumption from 13 cities among 5 different climate conditions, cold (Class I), moderately cold (Class II), hot in summer and cold in winter (Class III), hot in summer and warm in winter (Class IV), and warm all year (Class V). They confirmed that the variation in demand for space cooling and hot-water supply lead to the differences in electricity consumption across the classes. For example, cities in Class IV usually have more than one room air-conditioner and are used more frequently, while the consumption for space cooling in Class V households is very low. In addition, Class IV cities have much higher unit consumption of electricity for water heaters than any other classes. Another stream of research study could also be helpful for assessing the electricity demand behaviour from a unique angle – The findings of residential building occupancy rate could be good indicators for the possibility of presence of the family members at home and their activities at different time of the day. Hu et al. (2019) investigated the occupancy schedules of different room types in residential buildings in 3 cities in China –Beijing, Chengdu, and Yinchuan – representing different climate zones. The authors conducted a survey for half-hourly occupancy rate for living rooms, bedrooms, study(rooms) and kitchens. The results are a useful guide for the time use of end-users of electricity demand at home. It can be seen that the daytime occupancy rate of around 50% in Beijing is higher than in Chengdu or Yinchuan, which are in the range of 20-40%. Each room, nevertheless, has a similar shape of the occupancy schedules regardless of location. The living room is mainly used from 6:00 to 23:00, and most intensively during the evening period from 18:00 to 22:00. Bedroom occupancy is usually during the night time, from 22:00 to 6:00. Meal times can be surmised based on the kitchen occupancy schedule: 6:00 to 8:00 (breakfast), 11:00 to 13:00 (lunchtime), and 17:00 to 19:30 (dinner). Another regional difference is that kitchen occupancy or meal time schedule is later

in Chengdu, which indicate a potential different life style existed.

Studies of household consumption patterns in China using data mining are rare, especially true in terms of intra-daily residential load profiles. A case study from Shanghai by Pan et al. (2017) collected 15-minute residential consumption data from 138 households in Shanghai between May and December 2013. Shanghai is in the hot summer and cold winter zone (Class III) and no large central heating systems (normally run jointly by the State companies and local governments) is operated in Shanghai. They summarised that 4 cluster algorithm families are often used in the field of residential consumption, including K-means, fuzzy K-means clustering, Hierarchical clustering, and SOM. And K-means is one of the most popular cluster algorithms and the one that they adopted. They divided customers into 10 clusters where different profiles indicate differences in lifestyle. For instance, three of the 10 groups with double peaks and low morning and longer evening consumption levels are categorised as mostly white-collar workers. Apart from the routine analysis of the hourly consumption, their study also reports the results based on seasonal load and differences between weekdays and weekends. They categorised the ten consumption patterns into four sub-groups: (i) dominated by heating period, (ii) dominated by the cooling period, (iii) dominated by transitional seasons; and (iv) no distinguished features. The limitation of the research is that the sample of 138 families is small and so the clustering results may be biased and not stable, especially given that the cluster number is large.

Additional clustering research on the Chinese residential sector is based on lower-resolution data, e.g daily usage. Guo et al. (2018) collected daily household electricity demand data from January 1, 2014 to December 31, 2014 for 3,000 households in Nanjing and 1,399 households in Yancheng city, which are both in Jiangsu Province, China. They employed the K-means algorithm and attempted to depict the clustered household profiles at two levels: 1) The daily electricity consumption patterns during three Chinese major festivals, the Spring Festival, the National Day holiday and the Labour Day; 2) Daily residential profiles of a month for each

season. The households were divided into 9 groups for each festival. The nine groups can be summarised in to three types of customers on the Spring Festival were found and represented different life-styles of the households. While the patterns in the National Holiday and the Spring Festival are diverse, the load curves during the Labour Day holiday are relatively flat and less diverse. The seasonal curves reveal that the fluctuation in winter is much higher than in spring. A similar kind of the research but only focused on general residential consumption profiles was conducted by Zhou et al. (2017) using Fuzzy K-means algorithm based on daily consumption data from 1,312 households in Jiangsu Province during the month of December 2014. They found 6 and 9 are the appropriate cluster numbers for two different scenarios. However, one issue of these studies should be pointed out is that the monthly profile including every day of a month could be less insightful to distinguish the activity differences of customers, since household activities normally do not follow a cyclic pattern based on the day of a month, while a week profile could reflect more about the consumption patterns. Besides, the number of the clusters may be too large which may lead to a biased and less meaningful result, given that the sizes of the samples are not large enough. Another problem is that the authors used raw consumption data to cluster, while the cluster results more reflect the magnitude differences of consumption, rather than the fluctuation of consumption behaviour. The number of the pattern groups in the similar literature conducted in other countries is around 3 to 6 with larger sample sizes (Viegas et al., 2016; Ramos et al., 2007).

3 Data and methodology

3.1 Overview

As seen in the review above, studies on residential consumption patterns in China are scarce. Research into higher-resolution data normally involves a limited number of households participating, whereas studies based on over 1,000 households unfortunately have lower time resolu-

tion (normally monthly). The objectives of this paper are to fill this gap by using a large dataset (2000 households) and greater time resolution. We will structure three-level residential consumption patterns using three intra-day period usage data (peak, offpeak and super-offpeak): 1) The daily consumption patterns of a week in summer and winter for each intra-day period; 2) holiday load profiles; and 3) load profiles for extremes of hot and cold weather.

3.2 Datasets

The electricity data was provided by the Electric Power Company of Sichuan Province, SGCC. The daily electricity consumption data collected contains three points representing three intraday periods (super off-peak (23:00-7:00), peak (7:00-11:00 plus 19:00-23:00), off-peak (11:00-19:00)). The analysis period is from January 2014 to January 2017 and includes 2,000 randomly selected households from Chengdu. Chengdu City is a sub-provincial city and the capital of Sichuan, a southwestern province of China, which is known for being a major agricultural heartland. Chengdu is the fifth-most populous agglomeration in China. As of 2018, the resident population of Chengdu was over 14.76 million, and the city's total number of households was over 5.63 million (Chengdu, 2019). The average size of the resident household was 2.76 people per household in 2011 (Chengdu Bureau of Statistics, 2011). In 2018, Chengdu was ranked the best-performing city in China in terms of economic growth, with great potential of electricity demand growth (Bloomberg, 2018).

Chengdu is located in the southern monsoon climate zone and within humid subtropical climate. From both Murata et al. (2008) and Hu et al., (2019), Chengdu was categorised as being in Class III (hot summer and cold winter). The weather is generally warm with high relative humidity all year and it has four distinct seasons. Due to the high humidity, summer can be extremely uncomfortable and hot. However, no centralised heating supply is operated in Chengdu. The space heating in winter, if needed, is normally done by electrical appliances.

Statistics from the Chengdu branch of the National Bureau of China (2019) shows that air-conditioner ownership per 100 urban households in 2017 is 148.8, which demonstrates that the air-conditioner related consumption in summer could be very high, given around 1.5 air conditioners per household.

The weather dataset for the extreme weather analysis include six variables: minimum daily temperature, maximum daily temperature, average daily temperature, precipitation, average wind speed, and average cloud cover. And the data source is the land SYNOP (surface synoptic observations) alphanumeric messages that are managed by the World Meteorological Organization (WMO) (Meteomanz, 2019).

3.3 Data Preparation

The data we requested from the SGCC was for residential customers as defined in their system. However, a small number of the sample, which report an unusual and extreme high consumption, may be small businesses run out of the family home. For example, online store owners. Through the boxplots, the outliers that those households used more than 350kWH for an average monthly consumption were removed from the dataset. The households who have more than 5% of the values missing were also removed.

A common approach for data preparation before undertaking the clustering is to create average daily load profiles for each household, depending on the objectives of the research. Alternatively, profiles can be created to distinguish seasons or workdays from weekends. In our case, average load profiles were calculated separately at three levels:

1. Seasonal weekend/workday profiles: The Monday to Sunday profile for every household during different seasons were created. For each season, it consisted of three load profiles representing intra-day period separately. In total, each household had 3*4 load curves. To better reveal the differences between workdays and weekends, we standardised the

13

consumptions using the following formula:

$$D_{i,j} = \frac{U_i - \overline{U}_w}{\overline{U}_w} \times 100\% \quad \text{and} \quad \overline{U}_w = \sum_{i=1}^{5} U_i \tag{1}$$

where U_i is the consumption on the i_{th} day of a week for the family j, represents the average usage of weekdays. The clustering was based on the profile D_j for each household $(D_{1,j}, D_{2,j}, \ldots, D_{7,j})$ which reflects the change percentage of consumption between each day of a week with the average usage.

2. Holiday profiles. We chose the Spring Festival and National Day holidays since these are the two longest and most important holidays in China. The potential changes in behaviour or electricity consumption during those holidays would be very different from other holidays. To identify the consumption change, we extend the clustering period to three weeks – 7 days before the holidays, 7 days during the holidays, and 7 days after the holidays.

- The Spring Festival period: the holiday dates are not fixed, since the date of the Spring Festival changes every year

- The National Day Holiday period: 23th September to 14th October

Similarly to the seasonal profile standardisation, we wanted to compare the usage changes between a normal day and a festival day. In this scenario, we used the average daily consumptions in January and in September as the baselines for the Spring Festival and the National Day holidays respectively. The calculation was as follow:

$$F_{i,j} = \frac{F_i - \overline{F}_h}{\overline{F}_h} \times 100\% \quad \text{and} \quad \overline{F}_h = \sum_{i=1}^{21} F_i \tag{2}$$

where F_i is the consumption on the i_{th} day of the 21-day observation period for the family j, $\overline{F}_h$ represents the average daily usage before the holidays. The clustering input for each

14

household is a vector $H_j(H_{1,j}, H_{2,j}, \ldots, H_{21,j})$ which reflects the change percentage of consumption between holidays and the average consumption.

3. Extreme weather event profiles: Since there is no universally agreed definition of extreme heatwaves or cold-waves in Chengdu, we used the following rule for this paper:

   - For heatwaves: the consumption data of a day that the maximum or average temperature of a day is over the top 95th percentile in July and August.

   - For cold-waves: the consumption data of a day that the maximum or average temperature of a day is lower than the 5th percentile in January and February.

   - For the baselines: the day with average temperature falls in between the 45th and 55th percentile.

   The standardisation process is similar to the 1st and 2st scenarios and the final input will be a daily profile of consumption changes including three periods (off-peak, super off-peak, and peak) between extreme weather days and average summer/winter days.

Although the only papers on Chinese residential consumption pattern clustering by Zhou et al. (2014) and Guo et al. (2018) used raw data, normalisation is necessary as a standard step (Panapakidis et al., 2012; Rhodes et al., 2014a), especially given the samples were not large enough to eliminate biases and outliers. In addition, the primary focus of this research is to examine the households with similar behavioural change in electricity use (i.e. variation in profile shape). The direct use of raw data results in clusters, such as done by Guo et al. (2018), only reflect load magnitudes and cannot reflect the consumption variations within a day. The raw data method would also be much more sensitive to the outliers.

3.4 Algorithm selection

There is no consensus on the most suitable clustering approaches to residential metering data. The selection of algorithms should be based on the objectives of the study and the data structure. Yildiz et al. (2017) and Zhang et al. (2012) provide a detailed discussion of clustering methods. Among the reviewed methods of K-means, fuzzy c-means, and SOM, K-means is considered the most consistent clustering method based on their analysis. The K-means clustering method is one of the most widely used algorithms in the residential sector, due to its fast computation time and applicability to large datasets. This algorithm starts with the desired number of clusters K and randomly inserts the K data pattern into the initial centroids for each cluster (Hernández et al., 2012). The algorithm then iterates until the local minimum Euclidean distance between pattern xi and its closest cluster centroid is reached. The obvious disadvantage is that the results are affected by the initial set-up for each cluster.

Hierarchical clustering has also been explored in many clustering studies of electricity data (Chicco, Napoli and Piglione, 2006; Gounveia and Seixas, 2016). The method starts with each object as a separate cluster and in each successive iteration it merges the clusters with minimum distances in the distance matrix, until no cluster can be merged, or a termination condition is triggered. The advantage compared to K-means is that hierarchical algorithms do not need to pre-set the number of the clusters. Although the number of clusters is not necessary for hierarchical clustering preparation, a distance metric and a linkage criterion need to be decided before running the algorithms. Different distance metrics calculate the distances between each pair of data points through various formulations. The distance metrics we tested included: Euclidean, sqeuclidean (squared Euclidean), Chebyshev, cityblock, seuclidean (squared Euclidean), and Minkowski. Linkage criteria defines the dissimilarity between sets of the clusters and we examined the following criteria: single, complete, average, weighted, centroid, median, and ward To obtain the optimal performance of the clustering results, both K-means and Hierarchical ap-

proaches were examined to decide which technique would be most suitable. The effectiveness of the data partition as well the decision of the number of clusters were measured by clustering validity indexes. It is important to note that none of them prevails over the others uniformly compared by Chicco (2012). For cross-validation we used two indexes in this study, including DB and Silhouette score. The experience from the similar research show that the appropriate number of clusters of residential customers are usually between 3 and 10. Although more clusters can distinguish small groups with unusual consumption patterns, over-clustering should be avoided since results with few observations may be biased and less meaningful.

The data preparation and clustering were processed using Python.

4 Results and Discussion

In this section, we aim to explore the results from the clustering analysis describing the household electricity consumption profiles from the following perspectives: 1) Work/Weekend consumption patterns in different seasons; 2) Major festival demand patterns, including the National Day holidays and the Spring Festival; 3) Patterns of consumption changes associated with extreme weather.

In comparing the K-means and Hierarchical algorithms, we found that K-means was more suitable in this case based on the clustering validity indexes. Although the hierarchical algorithm using the the linkage method of Ward and with distance defined by Sqeuclidean and Chebyshev matrix has similar performance to K-means. However, this approach was slower and less robust than K-means. Therefore, we adopted K-means in our study. In order to focus on the clustering results, we did not show the comparison results between K-means and Hierarchical here but include them in the Appendix 1 to 7.

4.1 Seasonal workday/weekend profiles

The profiles shown in Figure 1 describe residential consumption trends over the course of a week. One distinct seasonal difference in the patterns is that almost all groups have an increasing demand on weekends in summer, apart from Cluster 2. However, around half of the consumers use less-than-average-workday electricity on weekends (C0 and C4) in winter. In addition, it can be seen that both in summer and winter, the majority of people falls into the group with the least fluctuation. In summer, for the majority there is a slight increase of total consumption on weekends, while the opposite is found in winter. The differences between the seasons could be explained by the use of space cooling appliances. In summer, the longer the households are in the home, the more electricity they may use on the appliances. The slightly less consumption in winter weekends could be led by the outdoor activities, while the summer in Chengdu is muggy and uncomfortable and people would tend to stay inside at home when possible. It should be highlighted that despite the stuffy hot weather, there was still a small part of households (Cluster 2) that would go outside during the summer. To understand the reasons and the specific differences in demand, the patterns divided to the intraday periods can be helpful.

Off-peak (11:00-19:00) usage patterns are shown in Figure 2. Of the three periods, behaviour patterns during the off-peak (daytime) period is least divided and more similar among all households. For both summer and winter, there is a group with a significant drop in demand on weekends (C4 in two seasons). This decline may indicate the nature of the property or the household: Those people are more likely to be local white-collar workers that are relatively richer with more than one property in the city, rather than immigrant workers. Because it appears that they only or mainly live in those properties during workdays and probably go back to their real home or another property for weekends. On the other hand, C3 in winter and summer show the exact opposite pattern with the lowest demand on workdays and highest on weekends.
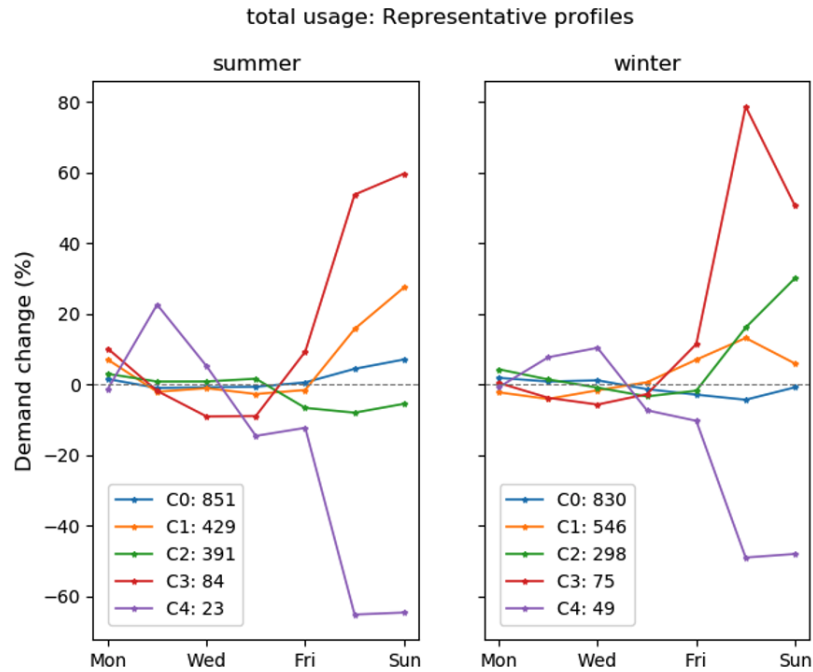
Fig. 1 Consumption patterns of total usage

They are possibly the "holiday" properties for the people that work (and live) at another location during the workweek and only go home during weekends. Those clusters may also include richer households since they can afford the cost of living in two properties.

To continue the analysis of intra-day patterns, evidence from super-off peak (night-time) (23:00-7:00) shows even more clearly that there are residents (C3 or C4) who may only be living in the property during either weekend or workday (Figure 3). For Cluster 3 and 4, the greater differences in demand at bedtime between workdays and weekends, compared to other clusters, indicate the possible non-occupancy during some days of a week, since they are not sleeping at the property and the least electricity usage is needed. One interesting seasonal difference is on Monday night. The weekend effect of delayed bed-time appears to extend to Monday in summer, where half of the households (C1, C2, C4) still have a significant higher demand. However, the similar trend is not found in winter. This could be explained by the summer effect

19
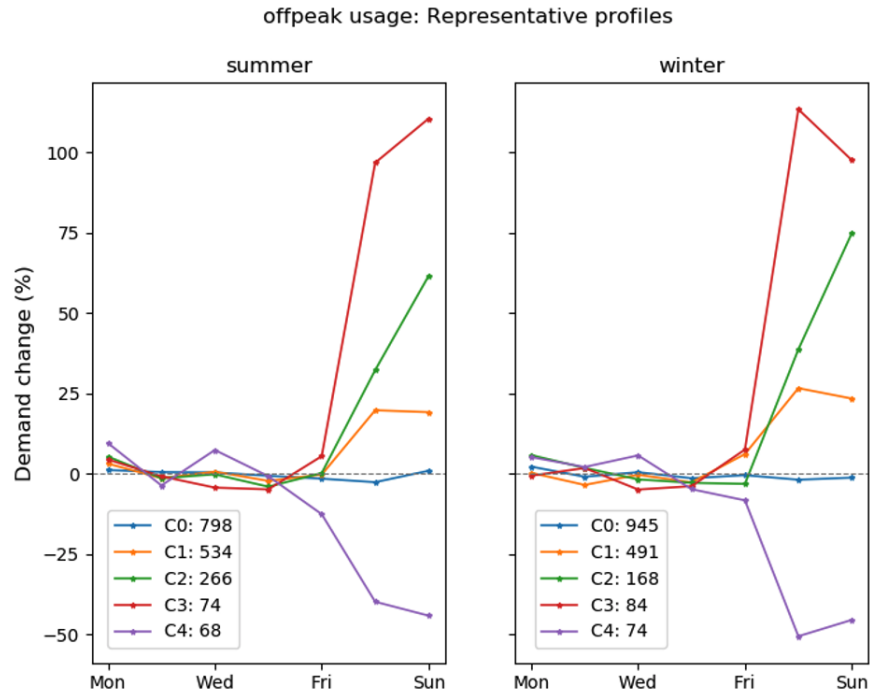
Fig. 2 Consumption patterns for off-peak (11:00-19:00)

that people would tend to sleep later at night, especially on weekends. Furthermore, the increase in demand on weekends in night-time exist in other groups as well. It may be due to the delayed bed-time and more activities after 23:00 during the weekends including watching TV, playing computer games, etc. The household with the larger magnitude difference could be the young adult group, for example, C1 in summer, compared with the older people that would tend not to stay up late even at weekends (C0).

The peak time profiles resemble the total consumption patterns (Figure 4). In both seasons, the majority of households have relatively flat consumption on weekdays, while experiencing higher demand on weekends. However, C1 (summer) and C2 (winter) show an opposite trend. In winter, households who prefer to go outside during weekend peak times (C2) tend to stay outdoors longer than their counterparts in summer (C1), since the magnitude of the fall in winter is much bigger than it is in summer. One reason people may tend to stay inside in summer rather
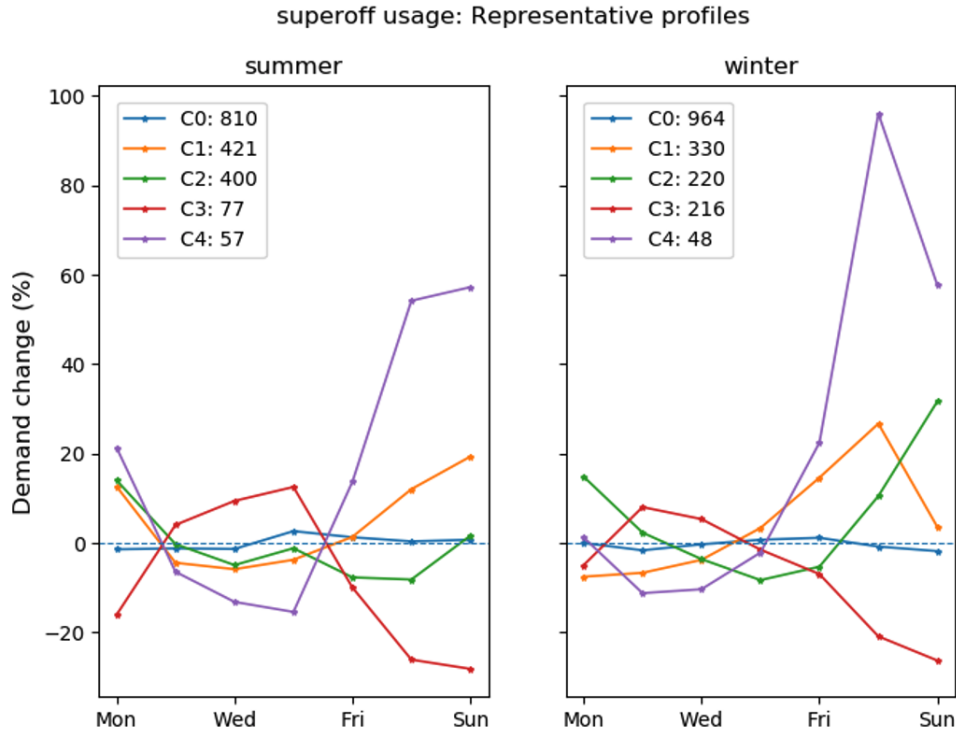
Fig. 3 Consumption patterns for Super off-peak(23:00-7:00)

than enjoy outdoor activities during their spare time could be the uncomfortable humid summer weather in Chengdu. Nevertheless, it should be noted that C1 in summer still includes more households than C1 in winter, which demonstrates that despite the hot weather, a larger fraction of households would prefer to go outside at weekend peak times, even if they stay out for shorter than those in winter.

4.2 Festival consumption patterns

Spring Festival

The Spring Festival is the most important family gathering holiday for the Chinese and it is also one of only two continuous 7-day public holidays in China. Electricity consumption fluctuates dramatically between the start and end of the Festival, which reflect the different holiday patterns during that period. On the first evening of the Spring Festival holidays, it has become

peak usage: Representative profiles
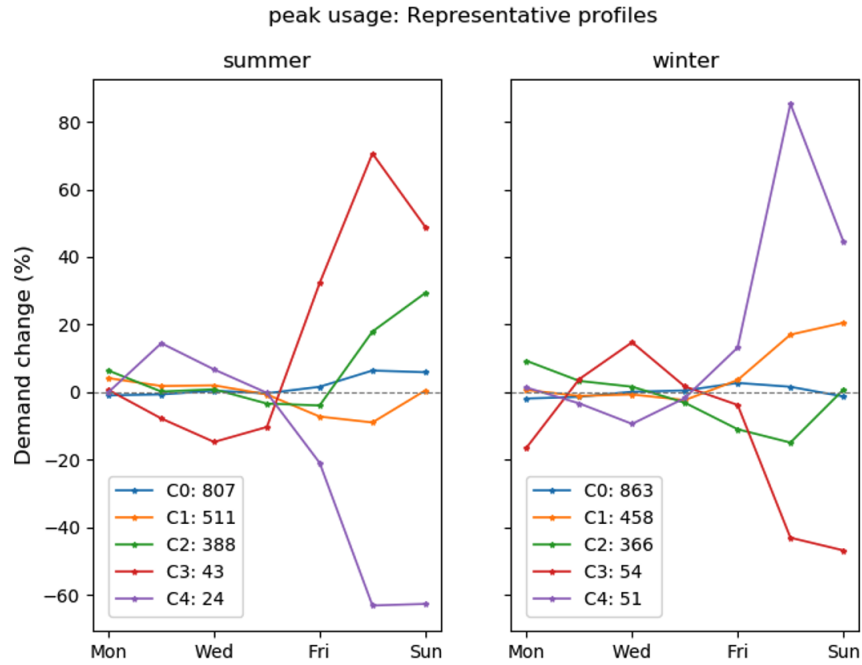
Fig. 4 Consumption patterns for Peak (7:00-11:00 and 19:00-23:00)

costumery to sit in front of the television and watch the Spring Festival Gala and TV programs with families. From Figure 5 and Figure 6 of the consumption patterns in peak and super off-peak times, we can identify four distinct types of households:

Type I: Cluster 0 (in both plots) is likely to be local families that their children live with or closer to them. It is highly likely that the peak of Cluster 0 on the first day comes from watching the gala with their family members. However, due to no extra people added in the house, the increase is not significant, compared to the other clusters.

Type II: Similarly, Cluster 2 (in both plots) also spike at that time with a much sharper rise, compared to Cluster 0. The extremely high demand indicates that Cluster 2 may also be older adult households but have guests and relatives come to visit, which add to the household demand. The spike is caused by more people at home as well as more electrical equipment usage. For example, for snack preparation/cooking and space-lighting demand.

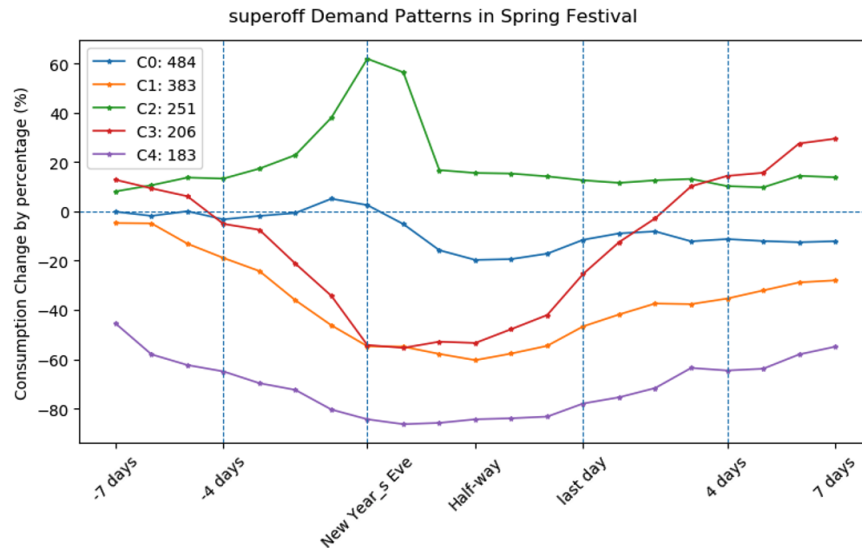Type III: Cluster 3 (in both plots) represents another classic holiday pattern in China and

22

Fig. 5 Super off-peak demand for Spring Festival

could be the younger white-collar workers, while Cluster 1 could be the older or senior workers. The drop in electricity consumption at night on the days before the Spring Festival may largely be explained by workers leaving their residences for their hometown or to travel. The Spring Festival travel season can be extremely hectic and many migrants will choose to leave days before the public holidays start to avoid the terrible traffic. The differences between Cluster 1 and Cluster 3 are mainly at the timing they leave the residence (where the drop starts) and when they return (when the consumption resumes to normal). It can be easily seen that Cluster 3 has a latter leaving date — around three days before the Festival, while Cluster 1 leaves the town earlier at about 5 days before. The difference could be largely explained by the fact that it would be almost impossible for the younger/junior workers in China to leave their work many days before the Spring Festival, while the senior employees would be more likely to be approved to leave earlier before the holiday. Another piece of evidence is that Cluster 3 starts to return to their residences days before the Festival ends and the demand rapidly resumes to the normal level after the Spring Festival holidays. Meanwhile Cluster 1 returns to the normal day

23

consumptions much slower.

Type IV: Cluster 4 could be retirees who tends to travel or labourers. By contrast with other groups, the cluster has a much earlier departure date and a longer period to get back to normal after the Festival. Compared to Type III, they seem to be able to leave their residences much earlier and are not rush to back for work. Thus, two plausible hypothese are that this cluster reflects either 1) retiree households leaving to visit their adult migrant children; or 2) labourers who normally leave their work one or even two weeks before the Spring Festival and would not return until the Lantern Festival (14 days after the Spring Festival).
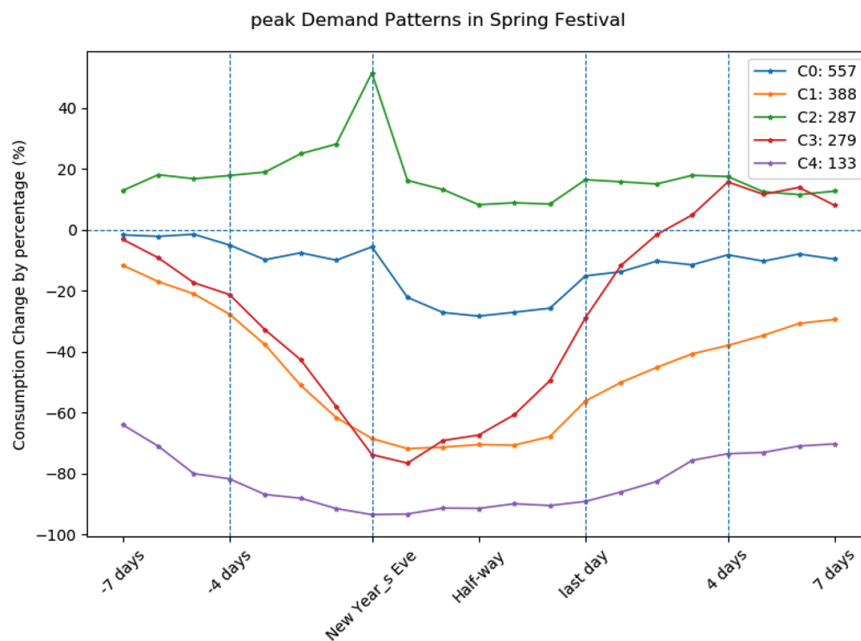
Fig. 6 Peak demand for Spring Festival

National Day holidays

National Day holidays are the other public holidays that last for seven consecutive days. However, the behaviour patterns are completely different from the Spring Festival. In general, the behaviour patterns among households are relatively similar and notably less diverse, compared
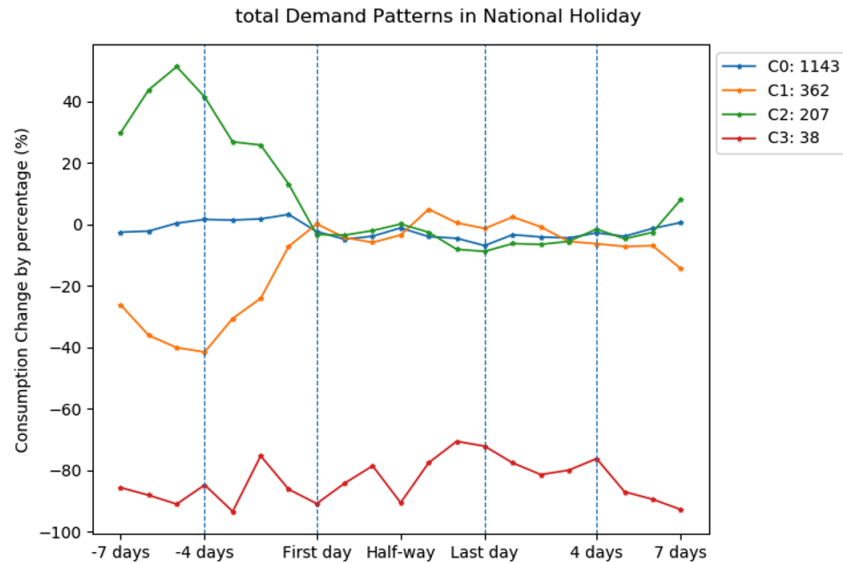
24

Fig. 7 Total consumption profiles in National Holidays

to the Spring Festival. One of the reasons behind the lower dissimilarity is that fewer immigrants will choose to return to their hometown during National Day Holidays. Most Chinese treat the National Holidays as an opportunity for relaxation whereas the Spring Festival is the most important time for family gatherings. It would be expected therefore that residential demand during these holidays would not see dramatic fluctuations.

As we can see from the total consumption patterns in Figure 7, the shapes of the clusters are similar during the 7-day national holiday and consumption remains almost unchanged. Unlike the Spring Festival holiday, the majority of households fall into one group (C0), which accounts for over 65% of the sample, which demonstrates that the consumption patterns are much less diverse and more concentrated over the National Day Holiday. The fact that over 95% of households (i.e., all clusters apart from Cluster 3) do not show a dramatic drop in total consumption confirms the relatively minimal travel during the Holiday. This may reflect shorter duration trips, apart from the much larger decrease in Cluster 3, which may reflect relatively longer-distance journeys.
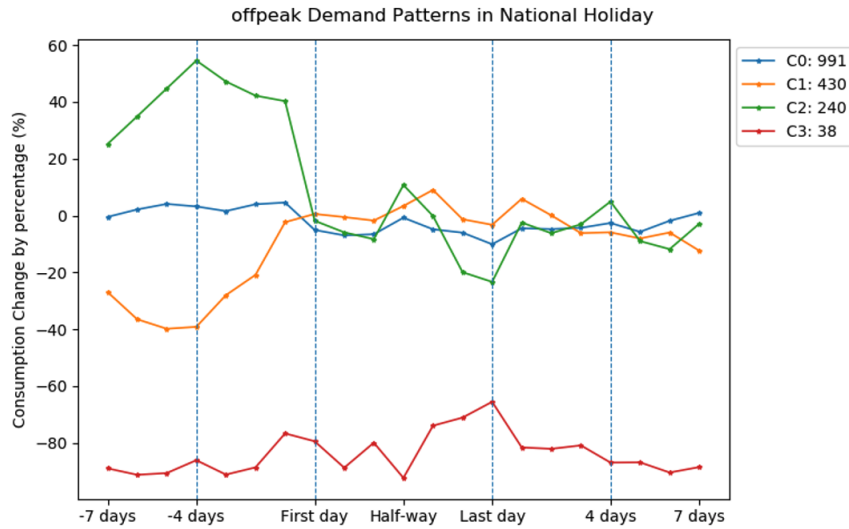
25

Fig. 8 Offpeak consumption profiles in National Holidays

In order to identify the daytime activities during the holidays, we examined consumption patterns in the off-peak period (Figure 8). The trends in demand across the groups are similar to the total consumption curves: Three clusters are similar while Cluster 3 may indicate long-distance and/or longer-duration travel. One interesting finding is that a temporary rise in demand occurs in the middle of the holidays for all groups.

4.3 Extreme weather profiles

Figure 9 and Figure 10 describe the patterns of change in residential electricity demand under the extreme weather in both summer and winter. The consumption changes in winter and summer can be easily distinguished from each other. And the differences in the patterns are largely driven by the popularity of air conditioners and limited ownership of space-heating appliances, due to local climate conditions (typical temperatures in Chengdu in January are 9 °C for typical high and 2 °C for the typical low).

In general, all customers in summer increase their consumptions rapidly, although to a different degree (Figure 9). During extremely hot weather, the most affected period is the night-
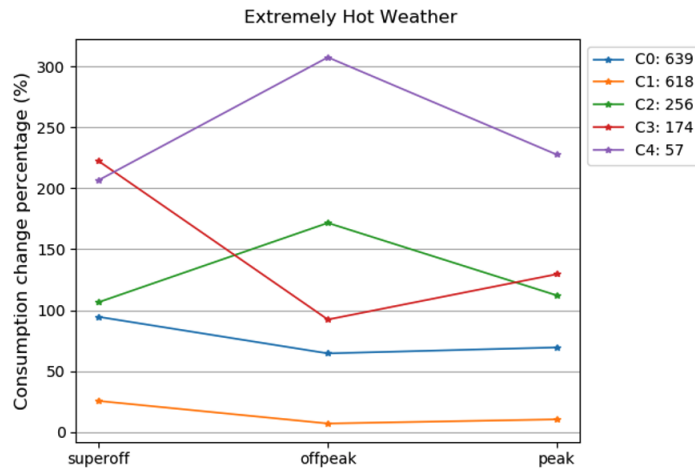
Fig. 9 Extreme hot days

time when around 65% of households doubled their usage, apart from Cluster 0. The cluster (C0) with the smallest rise in the super-off peak may be those who are most concerned with energy saving or relatively poorer families, which do not own air conditioners. While other clusters have the highest increase in super off-peak time, Cluster 3 have slightly different patterns that experience a higher increment in the off-peak/day-time. The group is likely to include those who are retired or self-employed and financially free, leading to the growth of the electricity during the daytime. Cluster 4, on the other hand, could be those who are more affluent than other groups. The highest surge in consumptions, where sees an almost tripled demand, could be led by either larger house sizes or/and more air conditioners.

The consumption patterns in winter differ significantly from the summer profiles (Figure 10). The categorisation is heavily concentrated on two clusters (C0 and C1) accounting for over 80% of households. In other words, the majority of families share similar patterns of consumption change during the top 5% coldest days. C0, the largest group, even have a slightly lower than usual winter consumption during the bed-time, which could be caused by an earlier bed-time. It should be noted that the ownership of space-heating appliances is not common in Chengdu. Although portable heaters have become increasingly popular in recent years possibly
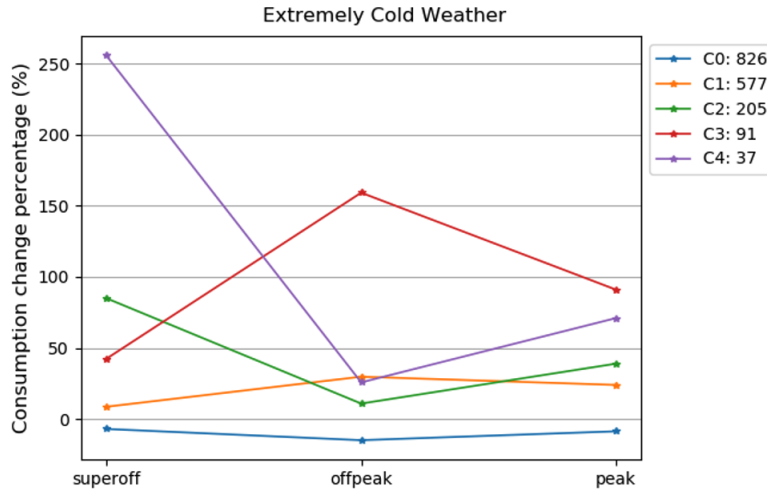
Fig. 10 Extreme cold days

led by the rising household income, the possibility of leaving the heaters on for the whole night is low due to safety concerns. In addition, it is also because although the sensible temperature could be very cold due to the high relative humidity, the 5th percentile of winter temperatures in Chengdu is still much higher than for major cities in Northern China – around average minimum temperature of 3ºC versus -15ºC at night. Therefore, no such large central heating system run in the northern cities has been operated in Chengdu. The people in Chengdu generally have used to the winter coldness and heating appliances are not seen as necessities among most Chengdu residents.

5 Conclusion

This paper presented a data-mining based approach to explore and structure a group of electricity demand profiles for 2,000 households in Chengdu, China. The clustering analysis was applied to average household electricity profiles in three different contexts (weekday/weekend; holidays; and extreme weather). Our innovative approach allows us to unravel or infer the life style and household characteristics from residential electricity demand profiles, without the as-

sistance of traditional survey tools. Our study addressed the problem of the lack of studies on intra-day clustering on the Chinese residential sector, with the majority of the literature focusing on monthly profiles.

First, the results of the weekend/workday profiles show that there are two groups of households that appear to be following a pattern of moving between properties within the week. We surmise those clusters are white-collar or relatively affluent families. In terms of the seasonal differences between the weekend/workday, the summer weekend consumption for most of the households is up to some degree, while the counterpart in winter remains unchanged or even slightly drops. Furthermore, the demand patterns in the major festivals in China unveil various types of lifestyle and behaviour, especially for the Spring Festival. For one group of older adults living with or close to their offspring and close family will see a limited increase in electricity use during the Spring Festival's Eve. Compared to the Spring Festival, the patterns found during the National Day Holidays are less diverse and more similar to each other. In terms of the demand changes resulting from extreme weather, we learned that most strikingly, at nighttime, over 72% of households doubled their electricity usage. We expect that the huge increase is driven by air conditioners due to the high penetration rate of space-cooling appliances. The consumption changes in cold days, however, does not seem to be significant, which might be explained by the limited popularity of space-heating appliances in Chengdu and less harsh winter weather than colder regions such as Northern China.

This chapter extends the current knowledge of Chinese residential behaviour patterns. Further research on customer classifications and load forecasting could benefit from the study. For example, a better understanding of customer consumption patterns during festivals and under extreme weather conditions would assist in load management on special occasions or under special circumstances. Implementation could be extremely helpful because the critical demand peak in Chengdu is in summer driven by air-conditioner spikes. In addition, the clustering algo-

29

rithms based on weather sensitivities can reveal new information about household consumption patterns which cannot be revealed by surveys alone. Although the results from Chengdu cannot be directly generalised to other areas due to different climatic and cultural backgrounds, the methodology proposed in the study can be applied to any region and to build the geographic-specific knowledge of the consumption behaviour in local areas for further studies. Meanwhile, the clustered results can be used as the base for future customer classifications. The different clustering methods offer a unique approach to classifying (new) customers and it could help build better/specific tariffs. For instance, classification based on the temperature sensitivity in summer could create a new tariff scheme that aims for shift in the critical peak, and result in better load management.

There are, of course, some limitations of the current study and further investigation is needed. First, hourly (or at least higher resolution) consumption data could undoubtedly offer more detailed information on consumption behaviour. The non-time continuous data would conceal important information. For example, peak-time consumption includes both usage in the 7:00-9:00 and 19:00-23:00 time slots, although consumption in early morning should be far lower than in the evening period. Second, it would be very useful to have customer-related attributes and that could assist with new customer classification with greater accuracy. Not having socio-demographic or dwelling characteristic data available hinders the possibility of exploring the consumption patterns and the underlying reasons in more depth. The absence of reliable electricity consumption data that can be associated with household information in China has been a logstanding problem. We encourage more studies aimed at identifying the residential consumption patterns without the need of such data.

Appendix

To compare K-means with hierarchical clustering, linkages and distance matrixes need to be selected for hierarchical algorithms before performing clustering. The next sections present how the selection process was done. The hierarchical algorithms with the most suitable linkages and distance matrixes were then picked to compare with K-means.

Hierarchical alogirthm: Linkage selection

This appendix contains the linkage selection process. As a first step we compared the clustering results produced by different linkages. The standardised total demands of families on the selected periods, Saturday and Monday is shown on Figure 1. It can be seen that only Ward and Average linkages are not sensitive to outliers and can divide households better.
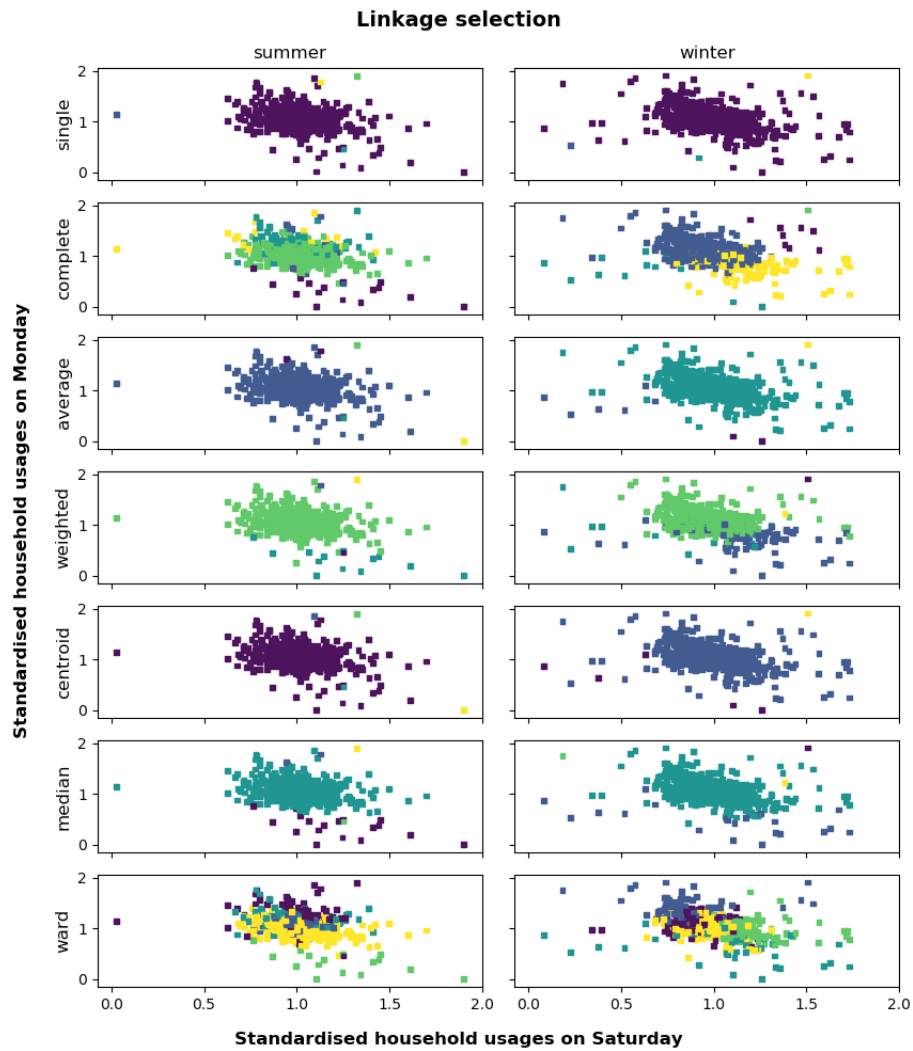
Fig. 1 household demand between two periods – Linkage selection

The comparison between these two linkages is followed (See Figure 2 and Figure 3). The linkages with higher Silhouette scores and/or lower DBI scores were selected for different scenarios separately.
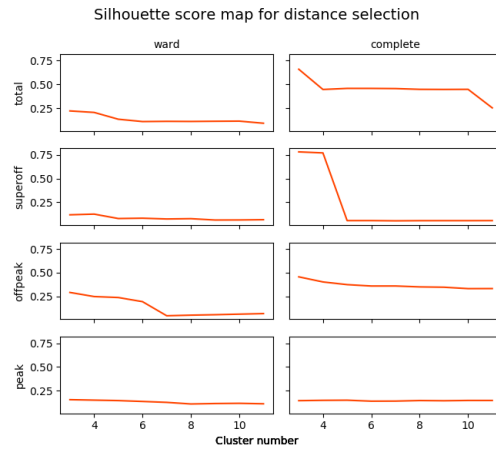
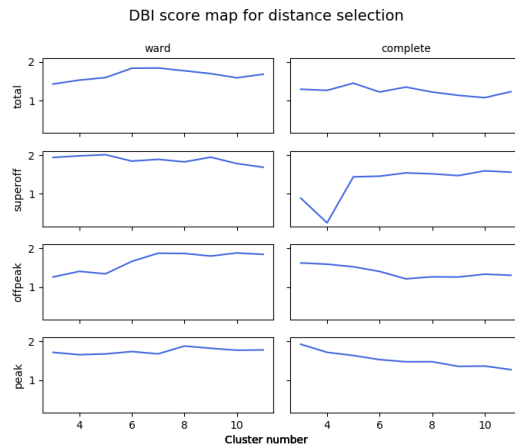Fig. 2 Silhouette score for selected linkages

Fig. 3 DBI score for selected linkages

Hierarchical alogirthm: Distance matrix selection

In this section, we used two methods to decide distance matrixes for hierarchical clustering. Firstly, the Cophenet score was adopted (Figure 4) and a distance matrix with a higher Cophenet score is regarded as a better-performed choice. To cross-validate the results, we also employed DBI scores to help the decision process (Figure 5). Similarly, a lower DBI score would suggest a higher clustering quality.
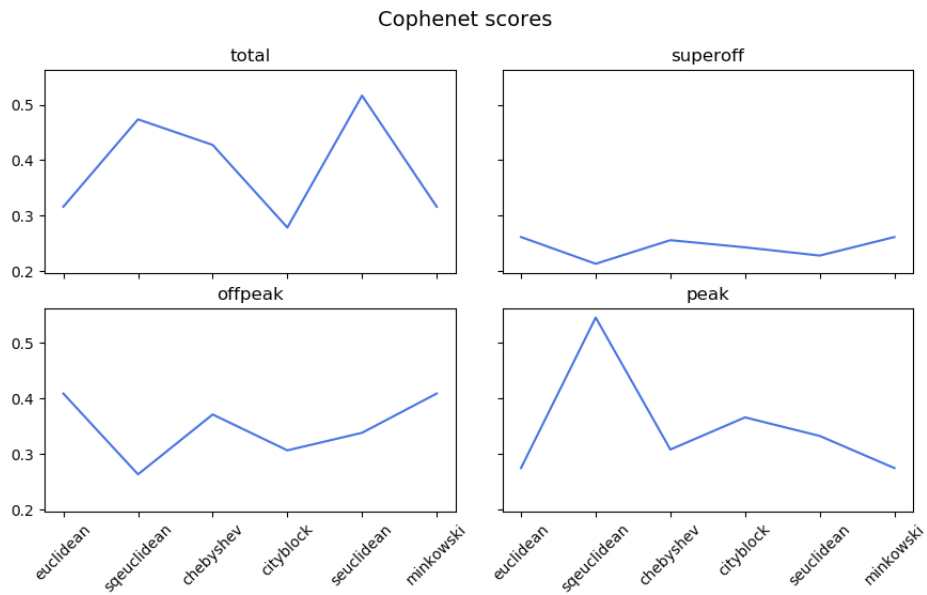
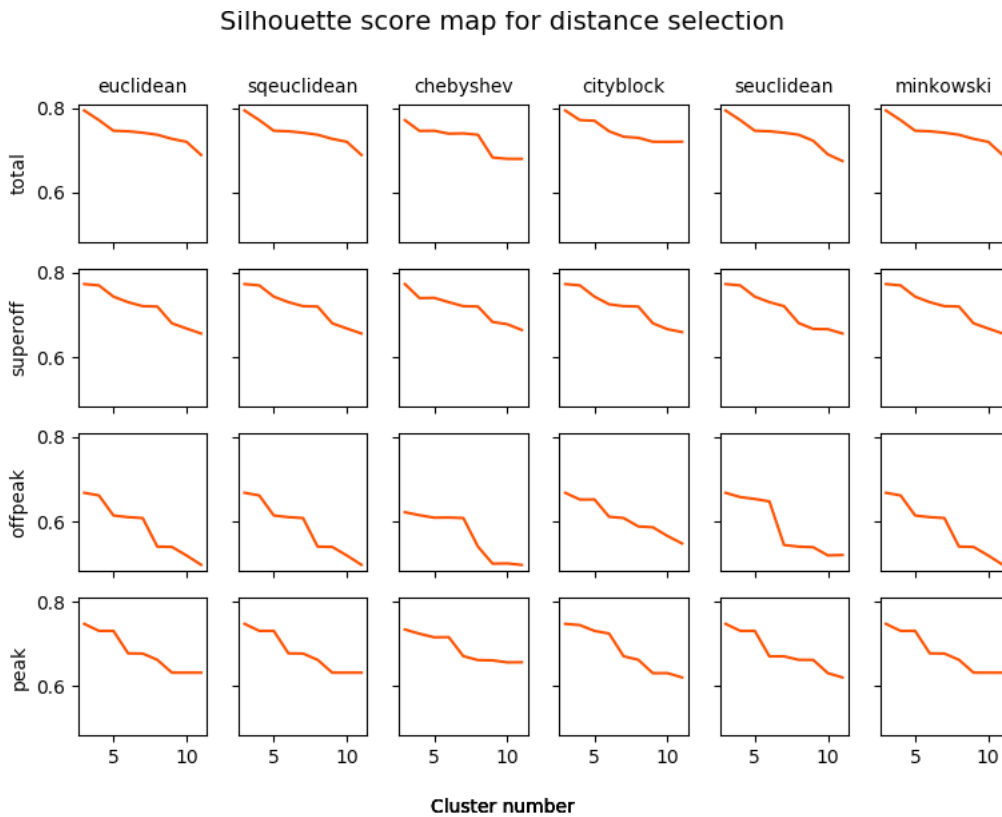Fig. 4 Cophenet score for distance matrixes

Fig. 5 Silhouette score for distance matrixes

Hierarchical and K-means comparison

Based on the selection process shown above, we compared the selected hierarchical algorithms with K-means through the clustering validity indexes, Silhouette scores, and DBI scores. Figure 6 and 7 demonstrated that irrelevant to cluster numbers, K-means outperformed the hierarchical clustering (A higher Silhouette core or/and a lower DBI score means clusters are well apart from each other and clearly distinguished). Therefore, K-means was chosen for the analysis. Here we only present the clustering results for total demand profiles, because the conclusion and findings remain the same to the intra-day period profiles.
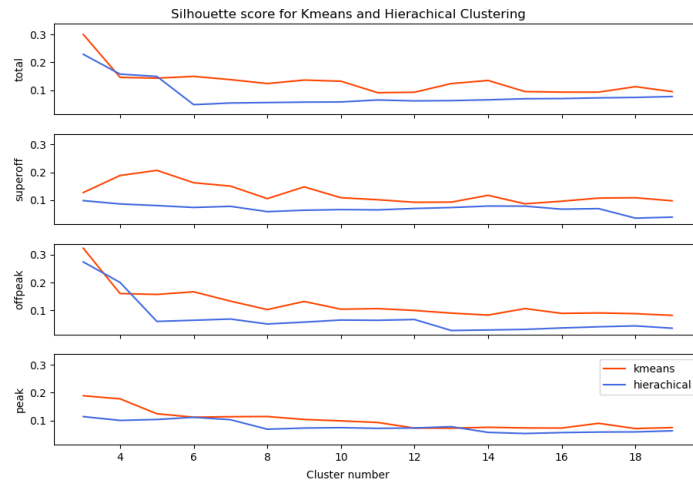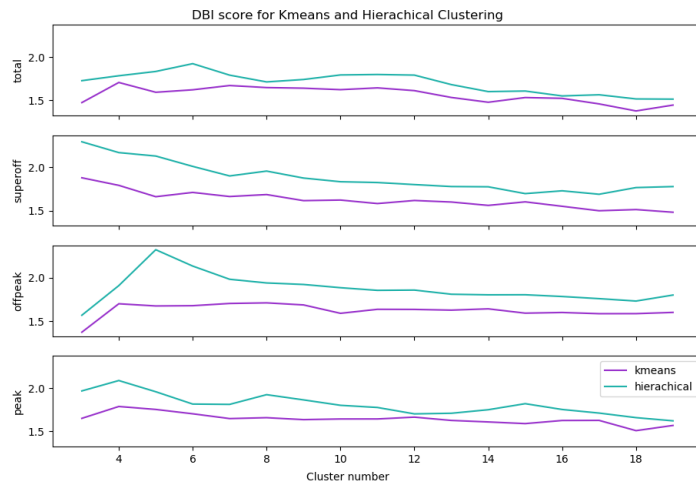

Fig. 6 Silhouette score

Fig. 7 DBI score

Reference

Afzalan, M. and Jazizadeh, F. (2019) 'Residential loads flexibility potential for demand response using energy consumption patterns and user segments', Applied Energy. 254, p. 113693. doi: 10.1016/J.APENERGY.2019.113693.

Al-Wakeel, A., Wu, J. and Jenkins, N. (2016) 'State estimation of medium voltage distribution networks using smart meter measurements', Applied Energy. 184, pp. 207–218. doi: 10.1016/J.APENERGY.2016.10.010.

Alberini, A. et al. (2019) 'Hot weather and residential hourly electricity demand in Italy', Energy. 177, pp. 44–56. doi: 10.1016/J.ENERGY.2019.04.051.

An, J., Yan, D. and Hong, T. (2018) 'Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings', Energy and Buildings. 174, pp. 214–227. doi: 10.1016/J.ENBUILD.2018.06.035.

Andersen, F. M. et al. (2019) 'Long-term projections of the hourly electricity consumption in Danish municipalities', Energy. 186, p. 115890. doi: 10.1016/J.ENERGY.2019.115890.

Atalla, T. N. and Hunt, L. C. (2016) 'Modelling residential electricity demand in the GCC countries', Energy Economics, 59, pp. 149–158. doi: 10.1016/j.eneco.2016.07.027.

Beckel, C. et al. (2016) 'Automated customer segmentation based on smart meter data with temperature and daylight sensitivity', 2015 IEEE International Conference on Smart Grid Communications, SmartGridComm 2015, pp. 653–658. doi: 10.1109/SmartGridComm.2015.7436375.

Blázquez, L., Boogen, N. and Filippini, M. (2013) 'Residential electricity demand in Spain: New empirical evidence using aggregate data', Energy Economics. 36, pp. 648–657. doi: 10.1016/J.ENECO.2012.11.010.

Bloomberg. 2019. China Best-Performing Cities Ranked: Chengdu Overtakes Shenzhen – Bloomberg News. 23 October 2019.
Access from: https://www.bloomberg.com/news/articles/2019-10-23/chengdu-overtakes-shenzhen-as-china-s-best-performing-city

Chaturvedi, D. K., Sinha, A. P. and Malik, O. P. (2015) 'Short term load forecast using fuzzy

logic and wavelet transform integrated generalized neural network', International Journal of Electrical Power & Energy Systems, 67, pp. 230–237. doi: 10.1016/j.ijepes.2014.11.027.

Chengdu Bureau of Statistics. 2019. Annual Statistics. Access from: http://csp.stats.gov.cn

Chicco, G. (2012) 'Overview and performance assessment of the clustering methods for electrical load pattern grouping', Energy. 42(1), pp. 68–80.
doi: 10.1016/J.ENERGY.2011.12.031.

Chicco, G., Napoli, R. and Piglione, F. (2006) 'Comparisons among clustering techniques for electricity customer classification', IEEE Transactions on Power Systems, 21(2), pp. 933–940. doi: 10.1109/TPWRS.2006.873122.

China Bureau of Statistics (2020). China Urbanisation Statistics.
Accessed from: http://www.stats.gov.cn/was5/web/

CPNN(China Power Network News). The State Grid Nearly Completed Smart Metering Installation (2019). (Accessed: 8 December 2019)
Available at: http://www.cpnn.com.cn/zdyw/201805/t20180524_1071233.html .

D'hulst, R. et al. (2015) 'Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium', Applied Energy. 155, pp. 79–90. doi: 10.1016/J.APENERGY.2015.05.101.

Du, G. et al. (2015) 'Residential electricity consumption after the reform of tiered pricing for household electricity in China', Applied Energy. 157, pp. 276–283.
doi: 10.1016/J.APENERGY.2015.08.003.

Firth, S. et al. (2008) 'Identifying trends in the use of domestic appliances from household electricity consumption measurements', Energy and Buildings. 40(5), pp. 926–936. doi: 10.1016/J.ENBUILD.2007.07.005.

Flath, C. et al. (2012) 'Cluster analysis of smart metering data: An implementation in practice', Business and Information Systems Engineering, 4(1), pp. 31–39. doi: 10.1007/s12599-011-0201-5.

Fu, X. et al. (2018) 'Clustering-based short-term load forecasting for residential electricity under the increasing-block pricing tariffs in China', Energy. 165, pp. 76–89. doi: 10.1016/J.ENERGY.2018.09.156.

Gouveia, J. P. and Seixas, J. (2016) 'Unraveling electricity consumption profiles in households

through clusters: Combining smart meters and door-to-door surveys', Energy and Buildings. 116, pp. 666–676. doi: 10.1016/J.ENBUILD.2016.01.043.

Guo, Z. et al. (2018) 'Data mining based framework for exploring household electricity consumption patterns: A case study in China context', Journal of Cleaner Production. 195, pp. 773–785. doi: 10.1016/J.JCLEPRO.2018.05.254.

Hausmann, J. A., Kinnucan, M. and McFaddden, D. (1979) 'A two-level electricity demand model. Evaluation of the connecticut time-of-day pricing test', Journal of Econometrics, 10(3), pp. 263–289. doi: 10.1016/0304-4076(79)90085-X.

Hekkenberg, M. et al. (2009) 'Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands', Energy Policy. 37(4), pp. 1542–1551. doi: 10.1016/J.ENPOL.2008.12.030.

Hu, S. et al. (2019) 'Investigation and analysis of Chinese residential building occupancy with large-scale questionnaire surveys', Energy and Buildings. 193, pp. 289–304. doi: 10.1016/J.ENBUILD.2019.04.007.

Kavousian, A., Rajagopal, R. and Fischer, M. (2015) 'Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings', Energy and Buildings. 99, pp. 220–230. doi: 10.1016/J.ENBUILD.2015.03.052.

Mahmoudi-Kohan, N., Parsa Moghaddam, M. and Sheikh-El-Eslami, M. K. (2010) 'An annual framework for clustering-based pricing for an electricity retailer', Electric Power Systems Research. 80(9), pp. 1042–1048. doi: 10.1016/j.epsr.2010.01.010.

McLoughlin, F., Duffy, A. and Conlon, M. (2012) 'Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study', Energy and Buildings, 48, pp. 240–248. doi: 10.1016/j.enbuild.2012.01.037.

Meteomanz. (2019). http://www.meteomanz.com/index?l=1

Mirasgedis, S. et al. (2006) 'Models for mid-term electricity demand forecasting incorporating weather influences', Energy, 31(2–3), pp. 208–227. doi: 10.1016/j.energy.2005.02.016.

Murata, A. et al. (2008) 'Electricity demand in the Chinese urban household-sector', Applied Energy. 85(12), pp. 1113–1125. doi: 10.1016/J.APENERGY.2007.05.008.

Ozawa, A., Furusato, R. and Yoshida, Y. (2016) 'Determining the relationship between a house-

hold's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles', Energy and Buildings. 119, pp. 200–210.
doi: 10.1016/J.ENBUILD.2016.03.047.

Pan, S. et al. (2017) 'Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in Shanghai residences', Building Simulation. 10(6), pp. 889–898.
doi: 10.1007/s12273-017-0377-9.

Ramos, S. et al. (2015) 'A data-mining-based methodology to support MV electricity customers' characterization', Energy and Buildings. 91, 16–25.
doi: 10.1016/j.enbuild.2015.01.035.

Räsänen, T. et al. (2010) 'Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data', Applied Energy, 87(11), 3538–3545. doi: 10.1016/j.apenergy.2010.05.015.

Rashid, H., Singh, P., Stankovic, V., and Stankovic, L. (2019). Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour? Applied Energy, 238, 796–805. doi: 10.1016/j.apenergy.2019.01.061

Razavi, R. et al. (2019) 'Occupancy detection of residential buildings using smart meter data: A large-scale study', Energy and Buildings. 183, pp. 195–208.
doi: 10.1016/J.ENBUILD.2018.11.025.

Rhodes, J. D. et al. (2014) 'Clustering analysis of residential electricity demand profiles', Applied Energy. 135, pp. 461–471. doi: 10.1016/J.APENERGY.2014.08.111.

Satre-Meloy, A. (2019) 'Investigating structural and occupant drivers of annual residential electricity consumption using regularization in regression models', Energy. 174, pp. 148–168.
doi: 10.1016/J.ENERGY.2019.01.157.

Satre-Meloy, A., Diakonova, M. and Grünewald, P. (2020) 'Cluster analysis and prediction of residential peak demand profiles using occupant activity data', Applied Energy. 260, p. 114246. doi: 10.1016/J.APENERGY.2019.114246.

Stankovic, L., Stankovic, V., Liao, J., and Wilson, C. (2016). Measuring the energy intensity of domestic activities from smart meter data. Applied Energy, 183, 1565–1580.
doi: 10.1016/j.apenergy.2016.09.087

People's Daily. The electricity of whole society achieved a new high in 2018 (2019). Available at: http://energy.people.com.cn/n1/2019/0119/c71661-30578230.html (Accessed: 8

December 2019).

Viegas, J. L. et al. (2016) 'Classification of new electricity customers based on surveys and smart metering data', Energy. 107, pp. 804–817.
doi: 10.1016/J.ENERGY.2016.04.065.

Yildiz, B., Bilbao, J. I., Dore, J., and Sproul, A. B. (2017) Recent advances in the analysis of residential electricity consumption and applications of smart meter data. Applied Energy. 208, 402–427. doi:10.1016/j.apenergy.2017.10.014.

Yilmaz, S., Chambers, J. and Patel, M. K. (2019) 'Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management', Energy. 180, pp. 665–677. doi: 10.1016/J.ENERGY.2019.05.124.

Zheng, X. et al. (2014) 'Characteristics of residential energy consumption in China: Findings from a household survey', Energy Policy. 75, pp. 126–135.
doi: 10.1016/J.ENPOL.2014.07.016.

Zhou, K., Yang, C. and Shen, J. (2017) 'Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China', Utilities Policy. 44, pp. 73–84. doi: 10.1016/J.JUP.2017.01.004.

Zhou, S. and Teng, F. (2013) 'Estimation of urban residential electricity demand in China using household survey data', Energy Policy. 61, pp. 394–402.
doi: 10.1016/J.ENPOL.2013.06.092.

Zhang, Tiefeng, Guangquan Zhang, Jie Lu, Xiaopu Feng, and Wanchun Yang. (2012). "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers." IEEE Transactions on Power Systems 27 (1): 153–60.
doi:10.1109/TPWRS.2011.2167524.