Essays on Residential Electricity Consumption Profiles:

Weather Effects and Household Behaviour Patterns

Jieyi Kang

Supervisor: Dr. David Reiner

Department of Land Economy
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Sidney Sussex College November 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the regulation length of 80,000 words for the Degree Committee of Land Economy.

Jieyi Kang
November 2020

Essays on Residential Electricity Consumption Profiles: Weather Effects and Household Behaviour Patterns

Jieyi Kang

Abstract

The high temporal resolution data created by smart metering, which has now been deployed in many countries, provides an unprecedented opportunity to examine household consumption behaviour in narrow time windows, whereas past studies could only look at monthly or even yearly consumption. However, most studies that have used smart meter data focused either on load management (load forecasting, theft detection, etc.) or linked electricity usage to demographic and/or building characteristics. Few studies have been conducted on the impacts of weather on intraday consumption behaviour. Better appreciation of the influence of weather could improve pricing designs as well as provide better understanding of household behaviour, which could, for example, potentially increase energy efficiency. With knowledge of weather effects on residential consumption, it could also be valuable for utilities to improve grid stability and reduce operation cost.

To fill the gap, this dissertation analyses the impact of different weather variables as well as consumption patterns through different tools based on smart metering data. This thesis uses a three article format. Chapter 1 provides a general overview of the literature on smart meters and empirical studies using smart metering data. Chapter 2 presents an econometric analysis of the effect of weather factors in Ireland (such temperature, rainfall and sun duration) at different periods of a day, and contrasts the impacts on consumption for workdays versus weekends versus holidays. Chapter 3 employs machine learning methods – clustering algorithms – to categorise households by their electricity demand response to different weather variables. The results demonstrated that some weather sensitivity patterns are closely associated with household characteristics. In Chapter 4, smart meter data was gathered from a very different location, Chengdu, the capital of Sichuan Province in China, which has more extreme weather and greater variability. Three scenarios are analysed in Chapter 4: (1) weekly consumption profiles in different seasons; (2) festival (major holiday) consumption profiles; and (3) consumption patterns during extreme weather. Finally, the thesis is concluded by Chapter 5, which summarises the main empirical and methodological contributions of the three papers and lays out future work in this area.

In memory of my beloved grandmother, Mingzhi Zhang and grandfather, Zehan Li

Acknowledgements

After approximately five years of writing this doctoral thesis, I am at the point to express my gratitude to everybody who supported and accompanied me during this exciting, instructive, and sometimes nerve-racking time. Without the support of my supervisor, friends, and family, this thesis would not have been possible.

First of all, I would like to express my gratitude to my supervisor Dr. David Reiner. His great support, guidance, and encouragement have been unique. He gave me a lot of freedom to work at my own pace and encouraged me to think independently. Furthermore, I thank him for valuable inputs on how to write academic articles and many insights into the world of research. In particular, I am grateful for his willingness to dare the experiment of taking on a PhD outside his department for supervision when my then supervisor left the University during my first year.

I am also grateful to my friends both in Cambridge and elsewhere for all the good times and keeping my spirits up in challenging times. In particular, I would like to thank the many friends I made at Sidney Sussex MCR. My time as a Ph.D. student in Cambridge has been made enjoyable by them. The cozy and friendly atmosphere in Sidney made me feel warm and inclusive throughout the PhD journey. And I am very thankful to have company of my lovely flurry friends, Charles, Camilla, and most importantly, my dearest Maggie, the most gentle and loveable cat one can ever have.

A very special thanks go to my very best friend and soul mate, Yuxiu Wei. During dark periods of the PhD, she was always to put a smile on my face. She has always listened and continuously provided immense moral support. We had so many late-night conversations and so much laughter together. And she has always been there through my ups and downs and has made the downs bearable. I could not have asked for a better company and partner.

On a more personal note, I would like to thank my parents and late grandparents for being a role model each in their own way. They have taught me to be independent and self-sufficient. Most importantly, they taught me that there were no limitations to what I could do and achieve as a woman. My family greatly supported me all my life in everything I have done, even though it requires sacrifices from them. Without their love and continuing support, this thesis would not have been possible.

Finally, I have always admire the curiosity, enthusiasm, perseverance that scientists and researchers show. Their dedication to science inspire generations to drive human progress. Salute and deepest respect to all of them!

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Overview of Smart Meter Development

History

Electricity grids around the world are currently undergoing drastic changes. In the light of dramatic improvements in digital technologies and with the aim of combating climate change and improving energy efficiency, building smart grids has become one of the highest priorities in electricity system reforms. The current grids are mainly electromechanical and there are few sensors through the systems (Wang et al., 2019). Historically, grids have only supported one-way communication. In addition, they generally require manual monitoring and restoration. The nature of these features means that power systems have limited control over demand, so it is a huge challenge for existing grids, particularly with the addition of photovoltaic systems and electric vehicles. The reforms have therefore been prompted to improve the flexibility and efficiency of the grids.

The basic arrangements of traditional power grids has been in place for almost a century (Hoenkamp et al., 2011). In the context of grid modernisation, the integration of advanced sensor information technologies with the communication infrastructure can create an electricity network that operates at high levels of efficiency and security (The European Commission Task Force for Smart Grid, 2010). Advanced metering infrastructure (or AMI commonly known as smart metering) is viewed as the first step in the infrastructure upgrade of the smart grids. Unlike traditional mechanical meters, the smart meter can obtain real-time information from the end-users and provide added information to the utilities. Smart meters offer bidirectional communications between the consumers and the operators and the electricity consumption information transmission can reach a much higher frequency from every minute to every hour.

In the 1980s, the manufacture of meters and communication providers were pitching "the unlimited potential of smart metering" (Sovacool et al., 2017). The highlighted features of smart metering included possible cost saving through less labour required in meter reading and time-of-use pricing schemes that can encourage greater energy efficiency. However, utilities were suspicious of the benefits promoted. They claimed that consumers were generally inflexible and insensitive to electricity pricing. Consequently, interest in the advanced metering system shown from utilities were scarce. They were reluctant to put efforts into large-scale smart meter rollouts in the residential electricity sector (Marvin, Chappells and Guy, 1999). Until the 1990s, smart metering services were firstly and widely introduced in the industry and commercial sectors. The deployment of smart meters in the residential sector has only come into full view in the 2000s onwards with a renewed focus on energy security and climate change (Murphy, 2016).

Opportunities and concerns

In addition to the functions of a traditional meter, opportunities associated with advanced metering infrastructure deployment mainly come from its advanced technical features. When a smart meter is in place, it can implement real-time two-way communications between utilities and residential customers. It measures and records power usage at certain pre-set short intervals, normally every 15 minutes, and the collected data are sent to a central data management system. The immense amount of fine-grained data generated by smart meters brings great potential for both policy-makers and utilities to analyse the electric system from different disciplines and perspectives, such as energy economics, electrical engineering, and environmental psychology.

From the perspective of customers, smart meters with in-home displays and/or proper feedback on consumption enable households to better understand the billings and their consumption habits at a greater detail level, which in the end helps households to manage their energy consumption to reduce their electric bills (Wilson, Hargreaves and Hauxwell-Baldwin, 2017). And for those households who prefer to live in a greener way, AMI can also help to change behaviour and reduce the frequency they use some energy-intensive electric appliances.

From the perspective of utilities and policy-makers, the rich information from smart meters can bring many benefits (Zhou and Matisoff, 2016): the advanced meters not only reduce costs associated with manual labour, for example, meter reading, grid monitoring and maintenance, they also enable more value-added services, such as demand response (DR) scheme, dynamic pricing, and distributed renewable generation (Electric Power Research Institute—EPRI, 2007; Leeds, 2009). Greater knowledge of the details of peak or off-peak periods, demand patterns, higher frequency consumption information can be extremely valuable for many stakeholders in the sector. Early studies used highly aggregated data, normally grid-level data, for developing operational strategies. Better understanding of customers' consumption patterns, for utilities and operators, can promote and enhance the efficiency and sustainability of the demand side; while for policy-makers, such knowledge enables the development of more efficient policies to guide the public to live in a more environmentally sustainable manner, which can help to achieve decarbonisation targets or detect and relieve energy poverty.

However, the benefits also come along with deep concerns of data privacy and security (Asghar et al., 2017). Some studies have shown that high-resolution electricity consumption data may reveal private information or life pattern of a household, for example, economic status, number of the people living in a household, the usage of certain appliances, and household occupancy (Wood and Newborough, 2003; McDaniel and McLaughlin, 2009; Kalogridis et al., 2011). Therefore, the work related to the privacy protection of smart meters as another hot topic has also attracted attention from both industry and the public (Wilson, Hargreaves and Hauxwell-Baldwin, 2017). Studies have focused mainly on two aspects: 1) legislation – for example, how to regulate utilities in terms of storing and employing the data; and 2) technical solutions such as innovations in developing new smart meters with privacy-preserving technologies.

Smart meter deployment status

The rollouts of smart meter in the residential electricity market have been widely discussed in the past decade among all the stakeholders, policy makers, utilities, customers and researchers.

The deployment of smart meters around the world are at different stages. Consider the cases of the European Union, United States and China. In the EU, driven by the European Union Energy End-Use Efficiency and Energy Services Directive 2009/72/EC (European Parliament, 2009) concerning the internal market in electricity, commitments to roll-outs of smart meters in EU member states started to slowly emerge. In January 2018, the average household electricity meter penetration rate was 34.5% in the EU-28 (European Commission, 2020). The residential smart meter penetration rates varied widely within the EU. Some countries have already finished a wide-scale roll-out with over 90% installation rate of SME and household smart meters by 2018, such as in Sweden, Finland, Italy, Estonia, Malta, Spain, and Denmark. On the other hand, the roll-out in some countries such as Croatia, Greece, Hungary and Lithuania was even lower than 2.5% by 2018, although according to the report (European Commission, 2019) most countries will reach a wide-scale penetration (to at least 80%) during 2020-2025.

In the U.S., the Energy Independence and Security Act in 2007 (110th Congress, 2007; Simoes et al., 2012) urged federal, state, and utilities to increase the penetration of smart meters in households (Hmielowski et al., 2019). The act contributed to a wave of Installations of smart meters in the U.S. Until the end of 2017 (EIA, 2017) the ownership had more than doubled since 2010—47% of all U.S. electricity customers now have smart meters. Similar to the situation in the E.U., differences in the smart meter rates among the states was huge (EIA, 2017). Washington, DC, has the highest smart meter deployment rate at 97%, followed by Nevada at 96%, but only 16 states had a residential AMI penetration rate higher than 60% in 2016. Differences in penetration rates are often driven by state legislation and regulation.

China has invested heavily in the smart infrastructure project of modernising its grid system. Back in 2011, State Grid Corporation of China (SGCC) began the deployment of smart electricity meters in various parts of the country. In terms of the number of smart meters installed, China is already the world's largest market (Ngar-yin Mah, Wu and Ronald Hills, 2017). The penetration rate reached 80% in China, although the roll-out rates varied by province depending on the fiscal situation. For example, by 2016, Qinghai in northwest China, one of the poorest provinces, only 45.17% of households had installed smart meters, while Jiangsu, one of the richest provinces in China, had already reached a 98% roll-out rate (SGCC news, 2016). The Chinese government is committed to reaching an overall deployment rate of 90% across the country by 2020 (National Energy Administration, 2015b).

1.2 General Literature review

We present an overview of three main types of studies on smart meters, including smart meter effectiveness, residential demand estimation through econometric tools, and load management. Since the thesis mainly concerns the quantitative analysis of residential consumption data, we will concentrate on the latter two topics. A more specific and detailed review of studies on high-resolution smart meter data in the next section (2.4). In addition, because of out of the scope of this study, this research review does not include cost-benefit analysis studies of smart meters.

1.2.1 Effectiveness of smart meters

The belief and expectation that smart meters can lead to changing customer behaviour may be explained by behavioural change models. In these models, there is a factor named "evaluation of outcomes", which is a part of the evaluation process before people take actions (Martiskainen, 2007). In this case, smart meters provide more transparent consumption information, and with less asymmetrical information the possibility of accurately evaluating the outcomes can be improved. Furthermore, considering another factor, the "facilitating conditions" proposed in Triandis' Theory of Interpersonal Behaviour (Triandis, 1994), smart meter installation establishes part of a smart grid infrastructure and enables two-way communication between utility providers and customers, which offers consumers a self-teaching opportunity to change their behaviour.

The assumption behind the effect of smart meters on customers is that the provision of detailed information on energy consumption will visualise energy use and raise awareness to encourage end users to make rational decisions to reduce their consumption. Darby (2006) believed that there is an association between the level of energy awareness, the probability of efficiency measures installed, and whether or not they regularly check their meter.

The discussion of smart meter effectiveness mainly emerged in the 2000s before large-scale rollouts. Darby (2006) pointed out that there are mainly two types of feedback affecting the effectiveness of smart meters: direct and indirect feedback. The definitions she offers are: direct feedback is from learning from looking at the raw data or paying for bills; whereas indirect feedback comes from reading and reflecting upon the processed data/information provided by the utility. Darby (2010) also summarised some significant trials for smart metering and feedback effects and concluded that the energy saving is between 5% and 15%. While many studies have tested the effect of different types of feedback with ICTs involving a sample size from about 50 to 15,000, a key shortcoming is that, most fail to demonstrate the sustainability of the impacts from the feedback due to a short period of observation. Van Dam et al. (2010) argue that studies using feedback devices mostly have lasted for less than four months, while the very few longer studies (van Houwelingen and van Raaij, 1989; Mountain, 2006) reviewed by Van Dam et al. (2010) present unclear results. Fischer (2008) also conducted a meta-analysis for ten countries over the period 1987–2006 where householders were given consumption feedback. She reports an average 1%–20% reduction from overall feedback and argues that certain forms of feedback were more effective, such as very frequent feedback or providing breakdown by appliance. However, these results should be interpreted with caution, as some groups involved have a small sample size.

1.2.2 Economic studies using residential consumption data

In the energy economics literature, research focused on residential consumption can be divided into two sections based on the type of data used: macro models and micro models.

Macro models use traditional aggregate macroeconomic data, such as GDP, income level, population size, and energy prices to correlate residential electricity demand with macro-data. Weather-related variables are the other important source of data that are frequently used in these studies. The most common weather-related variables used are temperature and related proxies (e.g. heating degree days (HDD) and cooling degree days (CDD)) (Torriti, 2014). However, weather conditions usually act as control variables but are not the focus of these studies. The objectives are mainly on the effects of socio-economic indicators on power demand at aggregate large scales, for example, regional, or even national level. For macro models, household-level smart meter data are rarely used. Instead, monthly, daily, or even higher-resolution grid-level data are usually chosen as input.

Cialani and Mortazavi (2018) used panel data covering 29 EU countries (EU-28 plus Norway) in a partial adjustment model to explain electricity consumption as a function of several economic variables, such as GDP, price, and population, along with temperature-proxy variables. They concluded that income elasticities in EU-29 are slightly higher than price elasticities and that short-term demand in EU-29 is inelastic to its price. In terms of weather effects, the demand seems more sensitive to cold than to hot weather in Europe. Salari and Javid (2016) estimated the residential energy demand in the United States at the state level from 2005 to 2013. Similarly, the variables used in the model included income, population, electricity price, and temperature proxies. Additionally, educational level and general household conditions of states were taken into consideration. They also compared two models in the static analysis – random effects (RE) and fixed effects (FE), which was found to be more robust and better than the RE model. The results showed that demographic characteristics such as household size, educational level, and per capita income have a statistically significant impact on the residential electricity. Other studies with similar objectives rely on a similar group of variables (Holtedahl and Joutz, 2004; Dilaver and Hunt, 2011; Torriti, 2014) to model residential electricity demand. Once again, it should be highlighted that although weather variables are included in these models, the emphasis of the research is given to the economic characteristics or to socio-demographic variables.

Compared with macro studies, micro studies are focused on household socio-economic variables as well as building characteristics. The nature of microscopic studies provides deeper insights into household-level consumption patterns and offers better understanding of consumer responses and the impact of behavioural factors. Before the deployment of smart meters, the electricity usage data used in research were based on monthly bills. Detailed investigation of the connection between household consumptions and weather conditions was almost infeasible and so researchers could only look at the seasonality of residential demand instead.

Kavousian et al. (2013) examine structural and behavioural determinants of residential electricity consumption based on a regression model. In their research, a dataset of 10-minute interval smart meters for 1628 households with an associated detailed survey was used. They identified four major groups of explanatory variables: weather, building characteristics, appliance stock, and occupancy, of which weather and floor size are the most important determinants. Iwafune and Yagita (2016) also employed high-resolution smart meter data for their econometric models to assess impact factors for residential electricity demand. They divided the fine-grained data into four time periods of a day to better analyse the relationship between household attribute data and consumption. The defined periods are midnight load between 23:00 and 7:00, morning load between 7:00 and 10:00, daytime load between 10:00 and 18:00, and evening load between 18:00 and 23:00. The important variables were: household size, floor area, outdoor temperature, and humidity. In terms of electric appliance ownership, the significant determinants were central ventilation systems, water servers, and owning more than one refrigerator. While there is no consensus on the factor importance for electricity demand modelling, Torriti (2014) summarised the five most frequently seen variables in a systematic review of residential demand estimation, which are: type of building, occupants' income, appliance ownership, price of electricity/bills, and number of occupants.

1.2.3 Load management studies

The main research objectives in load management originate from the urgent need to improve operational efficiency and meet increasingly irregular electricity demand. Thereby, the studies can be divided into four categories: 1) demand response scheme (DR) design, 2) load analysis, 3) load forecasting, 4) customer characterisation. We will only briefly discuss the first section here as a general overview of all the smart meter data related research, since the focus of the thesis is not on electricity tariffs. We describe some of the DR schemes and present the other three subjects separately in the next section.

A common definition of DR is given by the U.S. Federal Energy Regulatory Commission (FERC, 2010) as "the changes in electricity usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity". Generally, there are two types of current DR schemes: (i) incentive-based approaches, for example, direct load controls, interruptible tariffs and emergency programs (Albadi and El-Saadany, 2008), and (ii) price-based DR schemes (Shimomura et al., 2014). Since incentives are seen more among commercial and industrial customers, we only focus on the price-based DR schemes.

To encourage customers to respond to pricing signals, flexible pricing structures and available advanced metering technology are the keys to implementing DR in the energy system. Current price structures with fixed price electricity, however, cannot stimulate consumers to reduce consumption or alter their consumption patterns (Borenstein, 2009). As an important part of DR schemes, it is understandable why the pricing tariff design relevant topics have attracted a lot of attention. And the use of smart meter data can be assistance of deciding a suitable tariff. The concept of dynamic pricing schemes was proposed due to serious imbalance between supply and demand. During the early 1970s, the residential electricity consumption in the U.S. soared significantly, especially in space-conditioning loads (Caves, 1984). This pattern of residential usage growth resulted in serious load management issues during peak time. It called for substantial interest and implementation activity in a wide range of rate-incentive designs to replace a flat rate (EPRI, 1985). So far, four main time-varying alternatives have been proposed: real-time pricing (RTP), time-of-use pricing (TOU), critical peak pricing (CPP), and increasing block tariffs (IBT). The most common incentives in the residential sector are time-of use (TOU) and increasing block tariff (IBT) (Sanghvi, 1989), where as later innovative rates, such as critical peak pricing (CPP) and real-time pricing (RTP) are frequently examined after 2000 with the assistance of advanced smart meters. In the household field, the majority of research studying the impact of pricing structures on consumption were undertaken during the period between 2000 and 2010.

Borenstein, Rosenfeld and Jaske (2002) describe RTP as "the most natural or the most extreme approach to price-responsive demand". RTP charges customers differently during different times of the day and on different days. Customers can purchase part of their power through a long-term contract, though not necessarily at the real-time price. However, there are only a few large-scale trials for RTPs and the results are mixed: several researchers criticised that the use of RTP is a less cost-efficient way than TOU and has no significant effect on improving demand response (Goulden et al., 2014; Campillo et al., 2016).

While RTP has not been widely accepted in residential consumption, TOU has been implemented around the world (Albadi and El-Saadany, 2008). Under TOU, a day is divided into large time blocks of several hours, normally super off-peak time, peak-time, off-peak time, and mid-peak time. And in each of the blocks, the price is pre-determined, constant and normally will not be adjusted within 1-2 years. In general, the rate for each time block varies in different seasons, with a higher price in winter compared to summer (Newsham and Bowker, 2010).

Critical peak pricing (CPP) combines the features of RTP with some of the TOU structure. CPP schemes usually adopt a TOU rate structure. Unlike the lack of price signal for peak usage within a price block in TOU, CPP allows suppliers to add one more price to capture "critical" peak hours, which can be imposed on short notice (Herter, 2007).

As price-responsive tools, TOU and RTP are designed to encourage demand shift via differentiated tariffs. IBT, on the other hand, is a price-based instrument for reducing total consumption over a longer time scale, e.g. monthly (Mohsenian-Rad and Leon-Garcia, 2010). The price of electricity depends on total consumption during a billing period (typically a month). The IBT structure divides electricity use into blocks, where the price of electricity rises in line with increased consumption. The higher price in the following blocks are set to induce energy savings among wealthier households with nonessential use while allowing for basic needs of poorer households to be met at a much lower price per unit (He and Reiner, 2016). A similar approach is adopted in water IBT tariffs (Borenstein, 2012). It is clear that IBT is not exclusive from other time-varying schemes. In fact, many countries use a complex price structure with IBT and TOU, or fixed rate and IBT (Joskow and Wolfram, 2012).

1.2.4 Load profile related studies

A load profile represents the usage pattern or shape of a customer over a certain time interval. It could be an hourly or daily profile depending on the context of the analysis. Most policy-makers and utilities base their policies, tariffs, and operation strategies on average load profiles. Before the smart meter era, the decisions and studies of load management largely relied on high-resolution grid-level aggregate data. The technical features of real-time two-way communication brought by smart meters greatly improve the understanding of load profiles and related areas. As mentioned in the last section, load analysis, load forecasting, and customer characterisation are the three types of studies based on load profile analysis. To accord with the scope of the thesis, we will focus on a review of customer characterisation with a summary of employed methods as well as the results while still provide a general review of the other two subjects (load forecasting and load analysis), the technical details of the methodologies are not the main concerns of this review.

Load analysis

In studies of load analysis, the aims are to construct a series of algorithms to monitor the grid and discover any irregularities in the load. Two classes of anomalies most concern utilities and operators: bad data and energy theft. Bad data as discussed here refer to any missing data caused by mechanical faults of smart meters, failures of transmission or communication. Energy theft, as probably one of the most serious concerns, accounts for up to 50% of electricity consumption in developing countries. According to an annual report by the Northeast Group (2014), the annual loss of the world to electricity theft is around USD 89.3 billion.

To detect any anomalies, good knowledge of the fluctuation and uncertainty of the load profiles is the fundamental step to handle the issues. More specifically, the modelling can be divided into three methods (Wang et al., 2019): time series based methods, low-rank matrix technique based methods, and time window based methods.

The implementation of optimally weighted average (OWA) method is based on the time-series feature of smart meter data (Peppanen et al., 2016). The concept is similar to the autoregressive moving average (ARIMA) model for time series, which assumes that the data can be seen as a linear combination of the nearest neighbour data. To look at smart meter data from another angle, since electricity demand is both spatially and temporally correlated, low-rank matrix fitting based methods are proposed, such as Alternating Direction Method of Multipliers (ADMM)-based distributed technique (Mateos and Giannakis, 2013). Time window-based methods, on the other hand, computes the discrepancy between two time windows to continuously detect any dissimilarities within a certain time window rather than the whole load profile. The methods run on segmented load profiles, rather than the entire profile, which could be less computationally complex (Al-Wakeel, Wu and Jenkins, 2016).

Wang et al (2019) summarised energy theft detection methods into two categories: supervised learning and unsupervised learning, which both belong to machine learning methods. To train a system to identify theft by supervised classification methods, feature extraction and classification are the two main stages. For feature extraction, K-means and other clustering algorithms are often used (Jokar, Arianpoo and Leung, 2016; Tong et al., 2016). Clustering algorithms, such as K-means are normally used to identify the similarities among clusters and in this case, could be helpful for those un-labelled datasets to find similar features that can define clusters, where no prior knowledge of similarities have been known. Support Vector Machin (SVM) and decision trees are frequently used algorithms to train classifiers (Depuru et al., 2013; Jindal et al., 2016). Compared to supervised algorithms, unsupervised learning does not require prior knowledge of labels (e.g., a labelled training dataset where each entry is either an energy theft or not). Therefore, unsupervised methods are less expensive and more feasible. K-means, Birch, affinity propagation (AP), discrete Fourier transform (DFT), and Gaussian mixture based models (GMM) are frequently used approaches (Botev et al., 2016; Passos Júnior et al., 2016).

Load forecasting

Load forecasting as the core of the grid operation has always been one of the hot topics in the field of load management. Load forecasting at the regional or state level, or applied to commercial and industrial sectors has been extensively discussed in the literature, since the aggregate high-resolution data was always much easier to obtain, compared to the individual-level data required for the residential sector. Unlike aggregate load profiles representing larger scale loads, the load curves of individual households are much more dynamic and volatile. Forecasting household loads, therefore, has posed greater challenges to the utilities. Clearly, the studies providing better understanding the factors that affect residential demand can improve stability and efficiencies of grid operations.

Thanks to the deployment of residential smart meters, the analysis of behavioural consumption patterns and predicting occupancy activities becomes possible. The forecasting methods may vary depending on how far into the future the modelling is expected to forecast. Generally, the three main objectives of forecasting are: short-term (minutes to daily), medium-term (weekly to monthly), and long-term (yearly to decades) (Khan et al., 2016). Although there are numerous methods proposed for load forecasting, they, however, can be loosely divided into two groups: statistical modelling and machine learning methods.

For statistical modelling, there are generally four specific methods employed: multiple regressions, auto regressive (AR), auto regressive moving average (ARMA) and auto regressive integrated moving average (ARIMA). Regression models are usually used with weather data to predict load demand for certain regions (Barakat et al., 1990; Javeed Nizami and Al-Garni, 1995; Hyde and Hodnett, 1997). In AR models, electricity demand load is treated as time-series data that is correlated with previous usage. AR models can also be adapted to specific modelling demands (El-Keib, Ma and Ma, 1995; Huang, 1997). ARIMA models consider not only the previous linear connections of the current values, the white noise of the values are also taken into account (Huang and Shih, 2003). Similarly to AR and ARMA, ARIMA also treats consumption data as time-series, but it can be used to deal with non-stationary processes (Seymour, Brockwell and Davis, 1997).

Machine learning tools for load forecasting have emerged over the last decades. As one of the most popular machine learning tools, an algorithm called Artificial Neural Network (ANN) is frequently seen in the literature for electricity demand forecasting

(Edwards, New and Parker, 2012; Javed et al., 2012). Apart from ANN, other artificial intelligence tools frequently used include Decision Trees, Fuzzy Logic, support vector regression (SVR), and SVR's variant Least Square Support Vector Regression (LS-SVR), (Chicco, Napoli and Piglione, 2006; Singh, Gao and Lizotte, 2012; Jain et al., 2014). Compared to statistical modelling, weather-related variables are rarely seen in the literature. The models mainly focus on unravelling the consumption patterns.

Consumer characterisation

Likewise, current studies on residential consumer characterisation largely rely on machine learning tools. However, most research aims to understand occupancy behaviour and consumption patterns, rather than power usage itself. The objectives of these studies are to bridge the load profiles to either/both socio-economic status and building characteristics. In turn, greater knowledge of household characteristics associated with demand patterns can help identify DR scheme options and also improve grid efficiency (Chicco et al., 2004; Haben, Singleton and Grindrod, 2016; Ma et al., 2017).

Two methods commonly used in residential load curve analysis are unsupervised algorithms and supervised methods. For unsupervised tools, clustering is the most frequently used approach that can identify and group households with similar load profiles into the same clusters. On the other hand, classification as one major class of supervised methods is usually employed to predict households' load profiles through socio-demographic data as well as building characteristics, verse-visa. According to a systematic review (Tureczek and Nielsen, 2017), unsupervised learnings are more prevalent in related research.

For both clustering and classification methods, there are two main stages: data cleaning (to decide suitable input for the models) and algorithm selection. For data cleaning, there are two types of input in general: raw data and transformed electricity consumption indexes. Some researchers (McLoughlin, Duffy and Conlon, 2013; Yildiz et al., 2018) argue that transformed indices, such as maximum daily usage, a ratio of peak to off-peak consumption, etc, can better cluster consumers with less noise/un-useful information, while others insist that raw data with an hourly averaged day profile can best represent households' consumption patterns in greater detail (Gouveia and Seixas, 2016; Rajabi et al., 2019). In fact, the choice of the form of input depends on the objective of the research, that is, whether the core of the study is to explore the shape of daily profiles or reveal the association between demand patterns and household characteristics.

Data standardisation is one of the key questions that researchers need to consider. Similarly, the decision of whether and how to standardise data needs to be made based on the scope of research. To focus on the shape of load curves of each household, rather than the magnitude, the standard process is to normalise each profile to its maximum value (Panapakidis, Alexiadis and Papagiannis, 2012; Rhodes et al., 2014) where the values lie over (0,1] or [0,1], depending on whether values with zero consumption are dropped in data cleaning process. The alternative method is to subtract minimum values before dividing by the maximum value of the daily profile, which ensures the value domain is [0,1]. The latter technique therefore excludes the minimum values and cannot represent the ratio of shape differences between maximum and minimum values.

In addition, load shapes based on raw data can be investigated from different temporal perspectives. For instance, profiles have been defined by day of the week, weekend versus workday, holidays, or seasons (Kwac, Flora and Rajagopal, 2014; Hsiao, 2015). Examining residential demand patterns on these different time scales can provide important insights about the seasonality and periodicity of consumption behaviour (Kavousian, Rajagopal and Fischer, 2013; Gouveia and Seixas, 2016). Furthermore, more detailed studies of load shapes and the advent of smart meter data allows for the analysis of daily profiles at different intervals, such as midnight, morning, afternoon, and peak, etc (Quilumba et al., 2015; Haben, Singleton and Grindrod, 2016). The authors attempted to link consumption behaviour in specific periods with household characteristics, for example, whether air conditioning ownership is associated with peak demand. In summary, the best method to choose for profile segmentation completely depends on the aims of research and no optimal choices exist in the selection.

Another problem with daily profile clustering is that when the resolution is too high, such as every 15-minutes or half-hourly, it could result in larger uncertainty and less accuracy, since the useful information could be hidden in the high dimensional datasets. That is when dimensionality reduction techniques like Fourier series, Wavelet decomposition, and Principle Composition Analysis can help (Abreu, Câmara Pereira and Ferrão, 2012; McLoughlin, Duffy and Conlon, 2013; Ozawa, Furusato and Yoshida, 2016). The tools remove correlated structures to combine and reduce features into new sets of attributes. The methods can efficiently lower computational burden and are generally preferred when the size of datasets makes them difficult to handle. However, the cost of a lower computational requirement is interpretability of the new attributes. It could be difficult to infer the relationship between the revised attributes and the original features.

In terms of algorithm selection, K-means and its variants (such as Fuzzy K-means, K medians, etc.) are the most prevalent clustering algorithms in profile-clustering-related research (Tureczek and Nielsen, 2017). The simplicity and generally stable performances attribute to the popularity of K-means. One limitation of K-means is that prior knowledge of cluster numbers is needed. The decision largely depends on practical needs and previous experience. Hierarchical clustering, on the other hand, does not require a pre-set number for clustering. As another popular choice, it offers the attractive feature of being able to graphically display the class aggregation/disaggregation process (depending on whether the algorithm is an agglomerative or a divisive method). Unlike K-means, hierarchical clustering requires link functions (such as Euclidian, Wald and average) to decide how to link measured distances (Tureczek and Nielsen, 2017).

The more technical details of clusterings will be discussed separately in the following chapters, so here we mainly provide an overview of the clustering process.

1.3 Connections of the thesis

In reviewing the residential demand research described above, we noticed two main gaps in the field of energy economics and load management work respectively.

Firstly, two aspects have been rarely explored in previous economic research. Although many studies have used economic modelling to identify the relationship between weather conditions and residential consumption, the majority were based on macro-economic variables, such as GDP, income, and population. In addition, the scope was frequently to examine residential demand at the grid-operational or national level, rather at the household level. Furthermore, the data for the response variable used, residential consumption, was not high-frequency and normally daily or even monthly data. This set of studies would not be able to interpret the links between weather and demand at specific periods of a day. Better understanding of impacts of weather on electricity consumption would be beneficial for both utilities and policy makers. It could increase the accuracy of electricity transmission models. In addition, residential consumer responses to weather could help the design of policy instruments targeting energy efficiency improvement. Using economic models to examine the relationship between weather and household consumption would provide a general picture of how people respond to weather changes.

Secondly, weather factors have been left out in those studies based on household-level demand. The scope of previous efforts were to discover how power demand was associated with either socio-economic background (e.g., family income, education level, social class, etc.) or building characteristics (floor area, heating type, bedroom number, etc.). Similarly, this second type of study rarely concerns temporal consumption. It can be seen that research based on high-resolution panel data could provide deeper insight that how households' electricity consumption patterns respond to weather changes, specifically at different time of a day, for example, overnight, morning, afternoon, and peak demand. Without the knowledge on household-level and pattern differences during different intra-day periods, it could be difficult to design appropriate and efficient policies to combat energy waste. If research findings were available on the matter, it could show more insights on dynamic pricing schemes and energy efficiency policies.

The study in Chapter 2 therefore aims to fill the gap in the economic studies. In the analysis, we used a panel dataset from the Commission for Energy Regulation (CER) of Ireland, which includes over 3,000 Irish households over a period of 2009 to 2010. This dataset was compiled for the Electricity Customer Behaviour Trial, which consisted of a very-detailed survey and high-resolution consumption with smart meter recordings taken every 15 minutes. Using a series of Fixed-Effects models, we examined the relationship between intra-day residential demand and five weather variables – sun duration, outdoor temperature, precipitation, wind speed, and relative humidity. In terms of household demand, we divided daily consumption into 9 usage periods, defined as follows: early morning(6:00-8:00), day_1(8:00-10:00), day_2(10:00-12:00), day_3(12:00-15:00), afternoon/day_4(15:00-17:00), peak(17:00-19:00), early evening/evening_1 (19:00-21:00), evening_2 (21:00-23:00), and night (23:00-3:00). The period (3:00-6:00) was not included in the research, since there was very limited activities and consumption fluctuation occurred in the time scale.

Regarding the load profiling/customer characterisation, a similar issue to the econometric studies is that the inclusion of weather-related information in the load profile clustering is rare. When included, it is mainly used in load forecasting research and primarily for controlling exogenous impacts, rather than being the highlight of the work. For customer characterisation, the work mainly focuses on whole-day daily profile clustering, although seasonality was considered in some research by providing separate daily load curves for each season. Studies investigating intra-day profiles dividing up daily consumption into segments (e.g. morning, afternoon, late evenings, etc.) have not been fully explored.

In fact, residential electricity consumption patterns are closely associated with the timing of active occupancy. The traditional method of detecting household occupancy is either through surveys or very high-resolution data, such as minutely. Both methods can be particularly time- consuming and costly. Privacy is another barrier to obtaining what can be highly sensitive data. However, there is potential that households' demand sensitivity to weather at temporal periods could be a good indicator for occupancy detection and can therefore be used to infer the occupancy status at specific peridos about certain combinations of household characteristics. Using this method can be less intrusive and costly for policy makers and utilities to understand residential sector from another and/or deeper perspective.

Chapters 3 is based on machine learning methods to fill the gap that limited research have studied the weather effect at temporal periods. Similarly, the research in Chapter 3 also used the CER datasets, although from the clustering perspective. In Chapter 3, we clustered household responses to three weather attributes, temperature, precipitation, and sun duration separately. The household demand profiles of a day contain the usage in 9 periods and the definition of the periods was the same as in Chapter 2. In addition, we further divided profiles by season, and weekend versus workday. By clustering the demand change levels to different weather variables, we expected to unravel the customers' daily activity patterns in various weather conditions. The underlying assumptions were as follow:

1. Demand response to temperature may indicate the seasonality of activities occurred during selected periods
2. The sensitivity to rainfall implies whether regular outdoor activities occur in that period or the household is used to going out during that period
3. The sun sensitivity can be seen as an indicator of whether the household is flexible/does not have fixed schedules and has spare time so that they would be able to respond to a sunny day

The correlation between weather sensitivity clusters and household features was also examined by statistical tools.

Similarly, Chapter 4 used clustering methods, but focused on another geographic location, Chengdu, the capital of Sichuan province in southwestern China, with a sample size of smart meter readings from 2,000 households over a period of three years from 2014 to 2016. Geographic zones, as mentioned in the literature, can massively affect the demand response to weather conditions. By contrast, Ireland, in the previous chapter, has a milder and less variant weather whereas the climate in Chengdu is more extreme than in Ireland with much hotter summers and colder winters. Due to the limitation of data resolution, only three usage period in a day was included in the dataset, and so the focus and method employed in Chapter 4 was different from Chapter 3. There were three main objectives: 1) we investigated clustered day-of-week profiles in three periods separately in different seasons, in order to understand weekly residential consumption patterns. 2) how the demand profiles of the Chinese household changed in two major festivals, the Spring Festival and the National Festival, and what the changes meant. 3) how the different clusters of households responded to extreme hot weather. The study provided a better understanding of the Chinese household consumption habits in different scenarios.

Finally, Chapter 5 summarised some major conclusions and future work in Chapter 2 to 4. In Chapter 2, through fixed-effects models, we demonstrated that in general, rain and sunshine duration have bigger potential to affect people's behaviour and daily routines, while temperature has robust and relatively flat impacts. Followed by Chapter 3, We proposed a novice method of using the weather sensitivities as proxies to identify the household daily life patterns. Similarly, temperature clusters, compared to the rain and sun clusters, could reveal least information about the household life styles. Chapter 4 focused on another geographical location, Chengdu. We identified various consumption patterns and possible demographic groups associated with the patterns.

Reference

110th Congress (2007) Enengy Independence and Security Act of 2007, Public Law.

Abreu, J. M., Câmara Pereira, F. and Ferrão, P. (2012) 'Using pattern recognition to identify habitual behavior in residential electricity consumption', Energy and Buildings. doi: 10.1016/j.enbuild.2012.02.044.

Albadi, M. H. and El-Saadany, E. F. (2008) 'A summary of demand response in electricity markets', Electric Power Systems Research. doi: 10.1016/j.epsr.2008.04.002.

Asghar, M. R. et al. (2017) 'Smart meter data privacy: A survey', IEEE Communications Surveys and Tutorials, 19(4), pp. 2820–2835. doi: 10.1109/COMST.2017.2720195.

Barakat, E. H. et al. (1990) 'Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics', IEEE Transactions on Power Systems. doi: 10.1109/59.65910.

Borenstein, S. (2012) 'The redistributional impact of nonlinear electricity pricing', American Economic Journal: Economic Policy. doi: 10.1257/pol.4.3.56.

Botev, V. et al. (2016) 'Detecting non-technical energy losses through structural periodic patterns in AMI data', in Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016. doi: 10.1109/BigData.2016.7840967.

Campillo, J. et al. (2016) 'Is real-time electricity pricing suitable for residential users without demand-side management?', Energy. doi: 10.1016/j.energy.2016.04.105.

Chicco, G. et al. (2004) 'Load pattern-based classification of electricity customers', IEEE Transactions on Power Systems. doi: 10.1109/TPWRS.2004.826810.

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

Cialani, C. and Mortazavi, R. (2018) 'Household and industrial electricity demand in Europe', Energy Policy. Elsevier, 122, pp. 592–600. doi: 10.1016/J.ENPOL.2018.07.060.

Van Dam, S. S., Bakker, C. A. and Van Hal, J. D. M. (2010) 'Home energy monitors: Impact over the medium-term', Building Research and Information. doi: 10.1080/09613218.2010.494832.

Darby, S. (2006) 'the Effectiveness of Feedback on Energy Consumption a Review for Defra of the Literature on Metering , Billing and', Environmental Change Institute University of Oxford. doi: 10.4236/ojee.2013.21002.

Darby, S. (2010) 'Smart metering: What potential for householder engagement?', Building Research and Information, 38(5), pp. 442–457. doi: 10.1080/09613218.2010.492660.

Depuru, S. S. S. R. et al. (2013) 'High performance computing for detection of electricity theft', International Journal of Electrical Power and Energy Systems. doi: 10.1016/j.ijepes.2012.10.031.

Dilaver, Z. and Hunt, L. C. (2011) 'Modelling and forecasting Turkish residential electricity demand', Energy Policy. doi: 10.1016/j.enpol.2011.02.059.

Edwards, R. E., New, J. and Parker, L. E. (2012) 'Predicting future hourly residential electrical consumption: A machine learning case study', Energy and Buildings. doi: 10.1016/j.enbuild.2012.03.010.

El-Keib, A. A., Ma, X. and Ma, H. (1995) 'Advancement of statistical based modeling techniques for short-term load forecasting', Electric Power Systems Research. doi: 10.1016/0378-7796(95)00987-6.

European Commission (2020) 'European smart metering benchmark - Benchmarking Smart Metering Deployment in the EU-28'. Accessed from May 2020. Available at: https://op.europa.eu/en/publication-detail/-/publication/

European Parliament (2009) Directive 2009/72/EC of the European Parliament and of the Council, Official Journal of the European Union.

Fischer, C. (2008) 'Feedback on household electricity consumption: A tool for saving energy?', Energy Efficiency. doi: 10.1007/s12053-008-9009-7.

Goulden, M. et al. (2014) 'Smart grids, smart users? the role of the user in demand side management', Energy Research and Social Science. doi: 10.1016/j.erss.2014.04.008.

Gouveia, J. P. and Seixas, J. (2016a) 'Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys', Energy and Buildings. Elsevier, 116, pp. 666–676. doi: 10.1016/J.ENBUILD.2016.01.043.

Haben, S., Singleton, C. and Grindrod, P. (2016) 'Analysis and clustering of residential customers energy behavioral demand using smart meter data', IEEE Transactions on Smart Grid. IEEE, 7(1), pp. 136–144. doi: 10.1109/TSG.2015.2409786.

He, X. and Reiner, D. (2016) 'Electricity demand and basic needs: Empirical evidence from China's households', Energy Policy. Elsevier, 90, pp. 212–221. doi: 10.1016/J.ENPOL.2015.12.033.

Hernández, L. et al. (2012) 'A Study of the Relationship between Weather Variables and Electric Power Demand inside a Smart Grid/Smart World Framework', Sensors. Molecular Diversity Preservation International, 12(9), pp. 11571–11591. doi: 10.3390/s120911571.

Hmielowski, J. D. et al. (2019) 'The social dimensions of smart meters in the United States: Demographics, privacy, and technology readiness', Energy Research and Social Science. Elsevier, 55, pp. 189–197. doi: 10.1016/J.ERSS.2019.05.003.

Hoenkamp, R. et al. (2011) 'Smart Meter Rollout in the N The Neglected Consumer T', Renewable Energy Law and Policy Review, 2(4), pp. 269–282.

Holtedahl, P. and Joutz, F. L. (2004) 'Residential electricity demand in Taiwan', Energy Economics. doi: 10.1016/j.eneco.2003.11.001. van Houwelingen, J. H. and van

Raaij, W. F. (1989) 'The Effect of Goal-Setting and Daily Electronic Feedback on In-Home Energy Use', Journal of Consumer Research. doi: 10.1086/209197.

Hsiao, Y. H. (2015) 'Household electricity demand forecast based on context information and user daily schedule analysis from meter data', IEEE Transactions on Industrial Informatics. doi: 10.1109/TII.2014.2363584.

Huang, S. J. and Shih, K. R. (2003) 'Short-term load foJ recasting via ARMA model identification including non-Gaussian process considerations', IEEE Transactions on Power Systems. doi: 10.1109/TPWRS.2003.811010.

Huang, S. R. (1997) 'Short-term load forecasting using threshold autoregressive models', IEE Proceedings: Generation, Transmission and Distribution. doi: 10.1049/ip-gtd:19971144.

Hyde, O. and Hodnett, P. F. (1997) 'An adaptable automated procedure for short-term electricity load forecasting', IEEE Transactions on Power Systems. doi: 10.1109/59.574927.

Iwafune, Y. and Yagita, Y. (2016) 'High-resolution determinant analysis of Japanese residential electricity consumption using home energy management system data', Energy and Buildings. Elsevier, 116, pp. 274–284. doi: 10.1016/J.ENBUILD.2016.01.017.

Jain, R. K. et al. (2014) 'Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy', Applied Energy. doi: 10.1016/j.apenergy.2014.02.057.

Javed, F. et al. (2012) 'Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting', Applied Energy. doi: 10.1016/j.apenergy.2012.02.027.

Javeed Nizami, S. and Al-Garni, A. Z. (1995) 'Forecasting electric energy consumption using neural networks', Energy Policy. doi: 10.1016/0301-4215(95)00116-6.

Jindal, A. et al. (2016) 'Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid', IEEE Transactions on Industrial Informatics. doi: 10.1109/TII.2016.2543145.

Jokar, P., Arianpoo, N. and Leung, V. C. M. (2016) 'Electricity theft detection in AMI using customers' consumption patterns', IEEE Transactions on Smart Grid. doi: 10.1109/TSG.2015.2425222.

Joskow, P. L. and Wolfram, C. D. (2012) 'Dynamic pricing of electricity', in American Economic Review. doi: 10.1257/aer.102.3.381.

Kalogridis, G. et al. (2011) 'ElecPrivacy: Evaluating the privacy protection of electricity management algorithms', IEEE Transactions on Smart Grid. doi: 10.1109/TSG.2011.2160975.

Kavousian, A., Rajagopal, R. and Fischer, M. (2013) 'Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior', Energy. Pergamon, 55, pp. 184–194. doi: 10.1016/J.ENERGY.2013.03.086.

Khan, A. R. et al. (2016) 'Load forecasting, dynamic pricing and DSM in smart grid: A review', Renewable and Sustainable Energy Reviews. Pergamon, 54, pp. 1311–1322. doi: 10.1016/J.RSER.2015.10.117.

Kwac, J., Flora, J. and Rajagopal, R. (2014) 'Household energy consumption segmentation using hourly data', IEEE Transactions on Smart Grid. doi: 10.1109/TSG.2013.2278477.

Ma, Z. et al. (2017) 'The role of data analysis in the development of intelligent energy networks', IEEE Network. doi: 10.1109/MNET.2017.1600319.

Marvin, S., Chappells, H. and Guy, S. (1999) 'Pathways of smart metering development: shaping environmental innovation', Computers, Environment and Urban Systems. Pergamon, 23(2), pp. 109–126. doi: 10.1016/S0198-9715(99)00011-3.

Mateos, G. and Giannakis, G. B. (2013) 'Load curve data cleansing and imputation via sparsity and low rank', IEEE Transactions on Smart Grid. doi: 10.1109/TSG.2013.2259853.

McDaniel, P. and McLaughlin, S. (2009) 'Security and privacy challenges in the smart grid', IEEE Security and Privacy. doi: 10.1109/MSP.2009.76.

McLoughlin, F., Duffy, A. and Conlon, M. (2013) 'Evaluation of time series techniques to characterise domestic electricity demand', Energy. doi: 10.1016/j.energy.2012.11.048.

Mohsenian-Rad, A. H. and Leon-Garcia, A. (2010) 'Optimal residential load control with price prediction in real-time electricity pricing environments', IEEE Transactions on Smart Grid. doi: 10.1109/TSG.2010.2055903.

Murphy, M. H. (2016) 'Technological solutions to privacy questions: what is the role of law?', Information & Communications Technology Law. Routledge, 25(1), pp. 4–31. doi: 10.1080/13600834.2015.1134148.

Newsham, G. R. and Bowker, B. G. (2010) 'The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review', Energy Policy. Elsevier, 38(7), pp. 3289–3296. doi: 10.1016/J.ENPOL.2010.01.027.

Ngar-yin Mah, D., Wu, Y.-Y. and Ronald Hills, P. (2017) 'Explaining the role of incumbent utilities in sustainable energy transitions: A case study of the smart grid development in China', Energy Policy. Elsevier, 109, pp. 794–806. doi: 10.1016/J.ENPOL.2017.06.059.

Ozawa, A., Furusato, R. and Yoshida, Y. (2016) 'Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles', Energy and Buildings. Elsevier, 119, pp. 200–210. doi: 10.1016/J.ENBUILD.2016.03.047.

Panapakidis, I. P., Alexiadis, M. C. and Papagiannis, G. K. (2012) 'Electricity customer characterization based on different representative load curves', in 9th International Conference on the European Energy Market, EEM 12. doi: 10.1109/EEM.2012.6254761.

Passos Júnior, L. A. et al. (2016) 'Unsupervised non-technical losses identification through optimum-path forest', Electric Power Systems Research. doi: 10.1016/j.epsr.2016.05.036.

Peppanen, J. et al. (2016) ‘Handling bad or missing smart meter data through advanced data imputation’, in 2016 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2016. doi: 10.1109/ISGT.2016.7781213.

Quilumba, F. L. et al. (2015) ‘Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities’, IEEE Transactions on Smart Grid, 6(2), pp. 911–918. doi: 10.1109/TSG.2014.2364233.

Rajabi, A. et al. (2019) ‘A pattern recognition methodology for analyzing residential customers load data and targeting demand response applications’, Energy and Buildings. Elsevier, 203, p. 109455. doi: 10.1016/J.ENBUILD.2019.109455.

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

Salari, M. and Javid, R. J. (2016) ‘Residential energy demand in the United States: Analysis using static and dynamic approaches’, Energy Policy. Elsevier, 98, pp. 637–649. doi: 10.1016/J.ENPOL.2016.09.041.

Seymour, L., Brockwell, P. J. and Davis, R. A. (1997) ‘Introduction to Time Series and Forecasting.’, Journal of the American Statistical Association. doi: 10.2307/2965440.

Shimomura, Y. et al. (2014) ‘A method for designing customer-oriented demand response aggregation service’, CIRP Annals - Manufacturing Technology. doi: 10.1016/j.cirp.2014.03.117.

Simoes, M. G. et al. (2012) ‘A comparison of smart grid technologies and progresses in Europe and the U.S.’, IEEE Transactions on Industry Applications. doi: 10.1109/TIA.2012.2199730.

Singh, R. P., Gao, P. X. and Lizotte, D. J. (2012) ‘On hourly home peak load prediction’, in 2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm 2012. doi: 10.1109/SmartGridComm.2012.6485977.

Sovacool, B. K. et al. (2017) ‘Vulnerability and resistance in the United Kingdom’s smart meter transition’, Energy Policy. Elsevier, 109, pp. 767–781. doi: 10.1016/J.ENPOL.2017.07.037.

Tong, X. et al. (2016) ‘Cross-domain feature selection and coding for household energy behavior’, Energy. doi: 10.1016/j.energy.2016.03.135.

Torriti, J. (2014) ‘A review of time use models of residential electricity demand’, Renewable and Sustainable Energy Reviews. Pergamon, 37, pp. 265–272. doi: 10.1016/J.RSER.2014.05.034.

Tureczek, A. M. and Nielsen, P. S. (2017) ‘Structured Literature Review of Electricity Consumption Classification Using Smart Meter Data’, Energies. Multidisciplinary Digital Publishing Institute, 10(5), p. 584. doi: 10.3390/en10050584.

Wang, Y. et al. (2019) ‘Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges’, IEEE Transactions on Smart Grid, 10(3), pp. 3125–3148. doi: 10.1109/TSG.2018.2818167.

Wilson, C., Hargreaves, T. and Hauxwell-Baldwin, R. (2017) 'Benefits and risks of smart home technologies', Energy Policy. Elsevier, 103, pp. 72–83. doi: 10.1016/J.ENPOL.2016.12.047.

Wood, G. and Newborough, M. (2003) 'Dynamic energy-consumption indicators for domestic appliances: Environment, behaviour and design', Energy and Buildings. doi: 10.1016/S0378-7788(02)00241-4.

Yildiz, B. et al. (2018) 'Short-term forecasting of individual household electricity loads with investigating impact of data resolution and forecast horizon', Renewable Energy and Environmental Sustainability. doi: 10.1051/rees/2018003.

Zhou, S. and Matisoff, D. C. (2016) 'Advanced Metering Infrastructure Deployment in the United States: The Impact of Polycentric Governance and Contextual Changes', Review of Policy Research. John Wiley & Sons, Ltd, 33(6), pp. 646–665. doi: 10.1111/ropr.12203.

Chapter 2

What is the effect of weather on household electricity consumption? Empirical evidence from Ireland

2.1 Introduction

In recent years, there has been an increase in residential smart meter installations in many jurisdictions as they move to modernise their electricity networks (Eid, Hakvoort and De Jong, 2016). The old mechanical metering systems usually record monthly energy consumptions of households, which limit the possibility of understanding residential electricity consumptions in depth. Besides, dynamic pricing of electricity is impossible using current metering infrastructures, due to the technical constraints of having no real-time usage data. In light of these concerns, the deployment of Advanced Metering Systems can potentially be part of the solution to achieve greater energy efficiency. There is one significant advantage of smart metering that is widely accepted — The new technologies record high-resolution data of household electricity usage and increase the visibility of energy consumption. As a consequence, the availability of high volumes of data enables more fine-grained studies of residential behaviour and consumption patterns (Razavi et al., 2019).

Thus, one area that particularly benefits from the installation of smart meters is the study of the effects of pricing structures on electricity consumption. Previous studies in this area have focused either on longer-time frames, such as monthly household usage, or relatively shorter periods (daily consumption) but at the regional level (Pardo, Meneu and Valor, 2002; Davies, 2010; Atalla and Hunt, 2016; Trotter et al., 2016). High-frequency individual usage data makes it possible to examine the price effects during a specific short period during a day rather than using daily or monthly time steps. Although the results of the efficiency of different price schemes can be contradictory, increasingly studies have been done in the field to examine the effects from different perspectives.

However, the influence of weather in residential electricity consumption is one area that has not been extensively studied, although it has been widely accepted as an important factor affecting energy demand. The exploration of the relationship between energy consumption and weather is often seen in two sets of studies: weather as control variables in models focusing on price or on socio-economic effects (Wangpattarapong et al., 2008; Newsham and Bowker, 2010; Di Cosmo and O 'hora, 2017). Alternatively weather has been used as the main independent variables but only when investigating the relationship between daily or even monthly regional demand and weather variables (Moral-Carcedo and Vicéns-Otero, 2005; Costa and Kahn, 2010; Blázquez, Boogen and Filippini, 2013). Weather variables such as temperature, precipitation, relative humidity, wind speed, cloud cover, and sun duration are the most common variables used in both types of research. In spite of interest in the relationship between electricity consumption and weather, few studies have studied the possible association during different periods of the day due to limitations on the frequency of energy use data (Davies, 2010). Would specific findings hold in every period? For example, will residential customers reduce their consumption in every period of a sunny day? Are the weather responses, in fact, period-dependent? A better understanding of the weather impact on electricity can assist researchers, policymakers and energy companies. A study of how residential customers respond to weather in different periods can provide insights into daily patterns of household behaviour, e.g during which periods a family is more likely to be active or often go out.

We examine here the weather response at different times of day using fixed-effects models on high-frequency usage data from Ireland's Smart Metering Electricity Behavioural Trial (CER, 2012a) combined with weighted weather data from five weather stations in Ireland. Due to the half-hourly data available from smart meters, we are able to investigate the household response to weather during different periods. We aim to provide evidence that the weather sensitivities are indeed period-dependent and that weather factors may be good proxies for household behaviour patterns in different periods of a day. In addition, this chapter explores the impact of weather on the differences in electricity demand between weekends and workdays, thereby demonstrating that the relationships between weather and energy demand are not universal.

The chapter continues by reviewing the related literature of weather effects in Section 2 and the details of the dataset used are specified in Section 3. The two main models used and the explanation of variable selection are described in Section 4. Results of the models are presented in two parts in Section 5, and Section 6 provides a discussion of potential implications and offers some conclusion.

2.2 Literature review

2.2.1 Weather effects on demand in general

The discussion of weather variables often appears in two sets of studies in this field: one is model establishments for electricity consumption forecasting and usually at an aggregated regional/national level. For example, Mirasgedis et al. (2006) summarise the studies paying particular attention to short-term forecasting and the role of weather variability. They claim that based on the experience of utilities, the main weather factors affecting electricity consumption are temperature, humidity, wind, and precipitation in decreasing order of importance, while wind speed and solar radiation is not significant for the Greek mainland. Therefore, they only include the two weather variables (temperatures and relative humidity) in the models predicting the mid-term electricity consumption in Greece. Instead of using outdoor temperatures directly, heating degree days (HDD) and cooling degree days (CDD) are used to reflect the non-linear relationship between temperature and demand, which is particularly common in electricity demand studies (Bessec and Fouquau, 2008; Alberini and Filippini, 2011). However, in these studies the effects of weather are based on total consumption including all sectors, not just on residential consumption specifically. Thus, the importance of these factors still needs to be examined with a particular focus on residential electricity demand. The second type of research where weather variables are often included is in studies of the determinants of regional electricity consumption. Such studies rarely focus solely on residential electricity demand but rather on total regional consumption. Trotter et al. (2016) examined the relationship between climate and daily electricity demand in Brazil where the only weather factors used are CDD, HDD, and the lag effects of CDD and HDD. One compelling argument they make is that the effect of temperature on the weekend is slightly different than on working days.

Furthermore, a model based on aggregated monthly or annual data might not be able to reveal differences between the two cases. In addition to temperature, rainfall is another common variable examined. Hor et al. (2005) investigated monthly electricity demand from 1983 to 1995 in the UK and found a very weak negative relationship between rainfall and monthly demand. However, they argued that the correlation between demand and rainfall should be stronger but that the weak unexpected negative coefficient is mainly because they used only national-level data, while rainfall is very location-specific. Davies (1958)'s work considered aggregated country-level electricity demand in England and Wales arguing that five meteorological elements affect demand: temperature, wind speed, cloudiness, visibility, and precipitation. Temperature allied with wind speed determines the need for heat, while the remaining variables determine the level of daylight illumination, affecting lighting demand. The study divides daily demand into eight three-hour periods of demand to verify whether the effect of weather is the same across different periods of a day. The results show that temperature has a peak influence on demand around 9:00 and a lower coefficient during the 17:00 period. However, the direct effects of rainfall are only evident at 17:00. The findings indicate that the effect of a weather variable is not constant through a day, and it could be interesting to examine the differences in the residential sector specifically. Many researchers (Pardo, Meneu and Valor, 2002; Räsänen et al., 2010; Albert and Rajagopal, 2013) agree with Davies (1958) that weather indices such as humidity, wind speed, cloudiness, and barometric pressure are suitable explanatory factors for weather sensitivity, although those variables may have less significant influence on electricity demand than temperature, rainfall and sun duration.

As discussed above, studies involving weather effects have paid more attention to total electricity consumption in a region. There has been a lack of panel data to support deeper studies of the residential electricity sector – current panel studies concerning weather and residential electricity are primarily based on aggregated regional panel data. Atalla and Hunt (2016) looked at the residential electricity demand in six Gulf Cooperation Council countries using a panel dataset of annual demand in slightly different periods from country-to-country. CDD and HDD are the only weather indicators used but do not necessarily have significant impacts on demand. It depends on geographic locations and whether there is variation in the weather variable. Blázquez et al. (2013) used aggregate monthly panel data at the province level for 47 Spanish provinces from 2000 to 2008. The authors acknowledge that in panel data analysis, fixed-effects models (FE) or random-effects models could be helpful to control unobserved heterogeneity, however, neither of these was appropriate for their study since they include a lagged dependent variable in their model. Again, CDD and HDD are also the only weather conditions considered, which is common in panel studies of regional residential electricity consumption. Due to the lack of data at household level, very little research has been done based on non-aggregate residential consumption. Henley and Peirson (1998) studied residential energy demand and the interaction of price and temperature based on a Time-of-Use (TOU) trial with 150 households between April 1989 and March 1990. Through a fixed-effects model, they found that the effect of temperature is negative and non-linear, and the magnitudes vary for different periods. Alberini and Towe (2015) attempted to estimate residential electricity usage savings from energy efficiency programs. They assembled a panel dataset of monthly electricity usage and bills for a sample of about 17,000 households in Maryland from 2008 to 2012. They used Difference-in-Difference" and fixed-effects models to capture annual and seasonal household effects, and season-by-year effects. Weather effects are not the focus of the study, but CDD and HDD were included for monthly consumption control.

2.2.2 Weather effects in studies using smart metering data

In light of the trend of smart meter installation around the world, availability of household-level consumption data has begun to change. One of the main innovations brought by smart meters is that electric utilities can obtain huge volumes of high-resolution household usage data. A daily load profile of a household that depicts daily consumption trends from midnight to 11:59 p.m can now be easily drawn. High sampling frequencies provide operators with the opportunity to better understand consumption patterns of their residential customers. The availability of household consumption data enables researchers to identify the determinants of residential demand and the difference of effects on the demand of different periods of a day in more depth. One main strand of the literature using smart meter data investigates the effects of socio-economic and house-specific variables on load profiles. Anderson et al. (2016) summarised the existing evidence of household characteristics linked to load profiles and categorisd those variables into three subgroups: 1) household features, such as number of persons, number of children, and age distribution (Yohanis et al., 2008; Beckel et al., 2015); 2) dwelling status: e.g. dwelling type, household tenure, number of rooms (Firth et al., 2008; McLoughlin, Duffy and Conlon, 2012); and 3) householder characteristics: employment status, social status, age and gender. Other Householder variables, such as education level, ethnic group, marital status and household income are also found to have significant impact on demand and load profiles (McLoughlin, Duffy and Conlon, 2012; Carroll, Lyons and Denny, 2014). Nevertheless, research in5o electricity demand and household features have rarely paid attention to weather variables. There is little evidence of weather effects on residential demand from household-level data. Kavousian, Rajagopal and Fischer (2013) examine structural and behavioural determinants of residential consumption using a dataset of 10-minute interval smart meter readings from 1628 households in California. They prove that weather and location are among the most immmportant determinants of residential electricity use. However, the only weather variables, they include in their models are outdoor temperature and climate zone.

Another set of studies use smart metering data and consider weather variables to identify the effectiveness of time-of-use tariffs. Torriti (2012) took advantage of data from a TOU and smart metering trial in Northern Italy involving quarter-hourly readings from 1446 households from 1 July 2009 to 30 June 2011. The findings show that peak load shifting took place for morning peaks and created a split into two peaks for evening periods, while total consumption increased by 13.69%. The only weather variable, temperature, is used to control for the effect of weather variation, but the effect is not discussed in details. Other studies have used data from the large-scale trial smart metering experiment or Consumer Behavioural Trial (CBT) carried out by the Irish Commission for Energy Regulation (CER). Di Cosmo et al. (2014) utilise the CBT panel data of over 4000 households to explore whether the designed TOU is efficient in reducing peak demand. Two weather variables – sunshine duration and heating degree days – are included. Their results show that HDD are positively associated with consumption, while the opposite relationship is found for sunshine duration for the three periods considered (day, peak, and night). They only used the weather data from Dublin Airport weather station, as detailed information on household location is not available. However, considering that the selected households were drawn from across the country, a population-weighted weather dataset from different weather stations would be more accurate for a study of weather effects. In addition, the time periods may be too long since weather effects could change dramatically over the course of a period lasting as long as 10 hours.

From the review above It can be easily seen that little research has focused on weather effects and weather influences are usually introduced as control variables for other research objectives. Generally, temperature is the main weather factor considered and other variables, such as precipitation, wind speed, and sun duration, have not been explored extensively. Furthermore, weather impacts are commonly discussed at an aggregate level, e.g., daily or monthly level. However, as proposed in some studies (Davies, 1958; Henley and Peirson, 1998), the impact of weather indices might differ depending on the time of a day. The lack of research could be due to the limited availability of high-frequency household-level data. Even with greater access to detailed household usage data, the focus of studies using smart meter data has been on time-of-use tariffs, rather than weather effects. Therefore, a comprehensive study of the weather effects on residential electricity demand and household behaviour patterns during different periods of the day would be helpful to filling the gap.

2.3 Data

2.3.1 Residential electricity consumption data

The smart meter dataset used in this chapter was collected as part of Ireland's Electricity Smart Metering Consumer Behavioural Trial, which includes 4000 residential customers (CER, 2012a).

Half-hourly readings of usage were recorded by meters installed in the trial from 15 July 2009 to 31 December 2010. During the benchmark period (July 2009 to Dec 2009) all households were charged a fixed tariff. From 1 January 2010, those who were selected into treatment groups were charged time-of-use (TOU) tariffs. There were 4 TOU tariff periods: peak (17:00–18:59 Monday-Friday, excluding public holidays), day (08:00–16:59; 19:00–22:59 Monday-Friday, plus 17:00–18:59 public holidays, Saturday and Sunday) and night (23:00–07:59) periods. The tariff structure is shown in Table 2.1. In order to control the effects caused by the price incentives, our research only includes data recorded from 1 January 2010 onward, where tariffs are constant across all households during that period. In addition, homes with average daily consumption of more than 54 kWh are also removed because these outliers may not be residential consumers, but are more likely to be small enterprises or home-based enterprises [1]

2.3.2 Weather data

To generate daily weather observations at specific times of day, hourly weather data provided by the Irish Meteorology Office (Met Éireann) are matched with household

[1] Furthermore, the impacts of daylight saving time (31st October 2010 and 25th March 2010) are taken into account. The data from the second 2 am (end of DST) is deleted from the dataset to avoid double counting.

cents per kWh	**Night** 23.00-8.00	**Day** 8.00-17.00 19.00-23.00 weekdays 17.00-19.00 weekends and bank holidays	**Peak** 17.00-19.00 (Monday to Friday), excluding bank holidays
Tariff A	12.00	14.00	20.00
Tariff B	11.00	13.50	26.00
Tariff C	10.00	13.00	32.00
Tariff D	9.00	12.50	38.00

Table 2.1 Residential Time-of-Use tariffs 1st January to 31st December 2010

electricity consumption data. Recorded hourly observations are dry-bulb (air) temperature (°C), relative humidity (%), wind speed (kph), and fraction of sunshine per hour (%). Since location information is not provided by CER due to privacy concerns, a population-weighted weather dataset should be considered to reflect consumption response to weather from families living across Ireland (Valor et al., 2001; Auffhammer and Aroonruengsawat, 2012). As a result, four weather stations, Dublin Airport, Valentia Observatory, Belmullet and Cork Airport, were chosen. The first three synoptic stations are the choices of Met Éireann for regular Irish weather statements (Met Éireann, 2018) and Cork was selected to ensure enough sufficient regional representation because a significant number of participating households live in Cork (See Figure 2.1). Since the distribution of the final acceptances onto the trial was similar to the total population at county-level (Figure 2.1), the population ratios around the four stations are aggregated and calculated as weights to create a new dataset to match with the consumption data.

The weights of the population ratios used are 0.535 for Dublin, 0.175 for Cork, 0.16 for Belmullet and 0.11 for Valentia. The method of how to draw the boundaries of each station is not ideal due to the absence of household location information. For example, County Clare (CE in Figure 2.1) can be associated with Cork station or Belmullet station. However, as the weather in Ireland is relatively similar, the boundaries/weights hardly change the final results as we tried different weights for the analysis. The datasets from different observatory stations have similar correlations between weather variables and household consumptions (see Appendix A.1) and the Dublin and weighted weather data have a higher and better correlation with the residential demand. The descriptive statistics for the weather variables are described in Table 2.2.

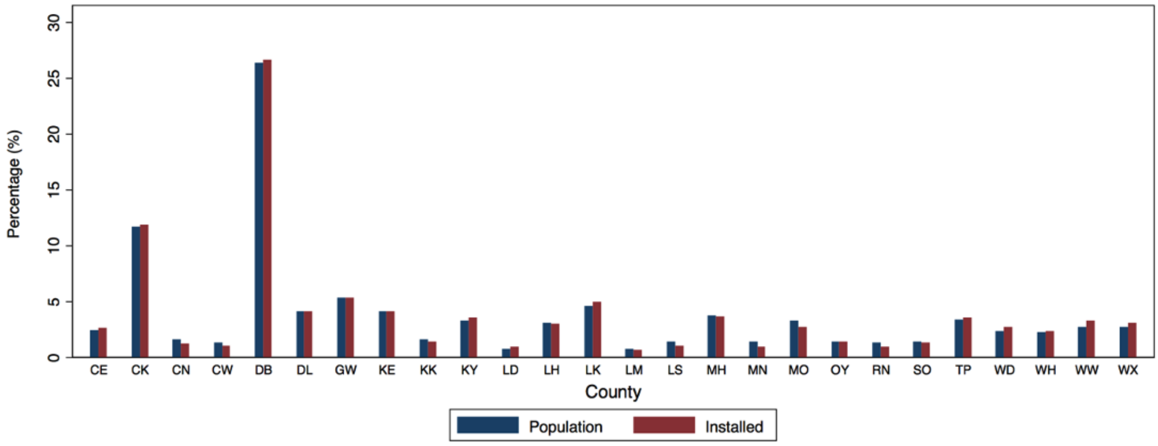

Fig. 2.1 Comparison of county level distribution between acceptances and total population

2.3.3 Time Use Study data

In order to help divide hourly data into several discrete periods, the Irish National Time-Use Survey 2005 (Economic and Social Research Institute, 2005) is used. It collected detailed national time-use statistics on over 1000 adult participants' daily activities, which includes two complete diaries of their activities over a 24-hour period — one for a weekday and another for a weekend day. It provides a comprehensive view of daily life in Ireland and possible behaviour during every 15-minute slot of a day. As a result, the findings of the survey can be particularly helpful in two ways: 1) to divide hourly data more accurately and avoid splitting one major daily activity into two periods, which may distort the actual response by either exaggerating or underestimating the effects. For instance, separating 12:00-14:00 into two different periods may cancel out part of the impacts of lunchtime; 2) to better understand how people respond to weather changes. For example, if people are less sensitive to rainfall during 12.00-14.00, it could be a lunchtime effect. Therefore, this survey data provides a supplementary tool to explain and confirm the results obtained from the proposed models.

2.4 Methodology

As seen in the literature review in Section 2, studies of the effects of weather variables have mainly focused on relationships between daily electricity consumption and daily weather change. It is not clear that how households respond to weather change at

		Morning	Day_1	Day_2	Day_3	Day_4	Peak	Evening_1	Evening_2	Night
Temperature (ºC)	Mean	7.23	8.64	10.18	11.03	10.75	9.83	8.75	7.91	7.20
	Std. Dev.	5.42	5.56	5.37	5.33	5.71	5.88	5.63	5.32	5.25
	Min	-7.49	-7.49	-5.16	-2.84	-3.60	-4.98	-5.20	-5.85	-6.33
	Max	17.52	18.09	19.94	20.57	20.78	19.82	18.69	17.52	17.09
Rel. Humidity (%)	Mean	89.26	84.43	77.98	74.21	74.97	78.61	83.04	86.48	88.75
	Std. Dev.	5.08	7.54	9.74	10.48	11.10	10.39	8.42	6.21	4.75
	Min	64.67	62.83	51.77	46.53	47.57	51.18	57.25	64.74	71.44
	Max	97.52	97.14	96.15	96.32	96.53	96.91	97.27	97.22	97.28
Wind speed (knots)	Mean	8.24	9.04	10.06	10.67	10.46	9.60	8.61	8.05	8.02
	Std. Dev.	3.56	3.55	3.72	3.77	3.84	3.87	3.82	3.89	3.78
	Min	2.03	2.54	2.91	3.46	2.66	2.86	1.75	1.47	2.06
	Max	27.16	27.87	27.07	29.03	28.93	29.79	28.83	27.93	24.34
Sun duration (% per hour)	Mean	0.14	0.36	0.48	0.46	0.33	0.20	0.07	0.00	0.00
	Median	0.04	0.31	0.50	0.45	0.25	0.03	0.00	0.00	0.00
	Std. Dev.	0.23	0.30	0.31	0.30	0.31	0.29	0.15	0.00	0.00
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	0.97	0.98	0.98	0.98	0.98	0.98	0.91	0.03	0.00
Rainfall (mm)	Mean	0.08	0.07	0.08	0.08	0.10	0.13	0.12	0.10	0.10
	Median	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.01
	Std. Dev.	0.19	0.18	0.18	0.18	0.23	0.31	0.29	0.23	0.22
	Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Max	1.40	1.34	1.32	1.71	1.96	2.27	1.93	1.84	1.82

Table 2.2 Descriptive Statistics for the weather variables

different times of day. To investigate the weather sensitivities in different periods of a day, it is reasonable to assume that households will not change their behaviour immediately when the weather changes. In order to capture the lagged effects, the hourly data is aggregated and divided into periods based on patterns of daily activities, rather than using raw hourly data directly. Although autoregressive models can be used on hourly data to control lagged effects, it might complicate the situation and the lag lengths suitable for weather effects are not clear and there is no agreed lag time in the literature.

Two rules are employed in separating the time periods: 1) To control for possible price effects caused by the TOU tariffs, the time periods chosen should not cross over two different tariffs (i.e., the tariff structure shown in Table 2.1); and 2) A period does not split major activities. On the basis of these rules, the tariffs provides natural breaks at the early morning, peak and night periods. However, the day price period (see Table 1) is much longer than the other periods, which may obscure the real response, and so needs to be sub-divided. In the end, 9 periods are set as follow with the help of the time use study: early morning (6:00-8:00), day_1 (8:00-10:00), day_2 (10:00-12:00), day_3 (12:00-15:00), afternoon/day_4 (15:00-17:00), peak (17:00-19:00), early evening/evening_1 (19:00-21:00), evening_2 (21:00-23:00) and night (23:00-3:00). The period from 3:00 to 6:00 is not included in the night period, since no major activities occur during that period and that could bias the analysis. The descriptions of the four main activities with the highest proportions of people doing on workdays and weekends are shown in Table 2.3. The numbers in each cell represent the minimum and maximum percentages of people doing the activities during each period of a day.

As panel data allows for the exploitation of both time and cross-section dimensions, it has the potential to eliminate unobserved heterogeneity in the data (Asteriou and Hall, 2011). As a result, given the nature of the panel dataset, two fixed-effect models are employed. Although random-effects (RE) models are also used in the related literature, fixed effects (FE) models better suit the purposes of this study.

With FE models, the focus is given to weather variables, while the effects of variables whose values are consistent across time (Wooldridge, 2013), such as demographics, housing conditions, and electric appliance ownership, are captured in a single fixed-effects estimator since the focus of the study is not on household characteristics. In addition, the results of the Hausman test imply that FE models are more suitable, since the null hypotheses of RE models is rejected (p-values of 0.0000).

	1st		2nd		3rd		4th	
	Workdays	Weekends	Workdays	Weekends	Workdays	Weekends	Workdays	Weekends
Morning 6.00-8.00	***Sleep*** 46.3~91.9	***Sleep*** 72.9~92.4	***Personal Care*** 1.7~12.4	***Personal care*** 1.2~6.2	***Eating*** 0.3~9.7	***Paid Employ*** 0.6~4	***Travel*** 0.6~9	***Travel*** 0.4~3.1
Day_1 08.00-10.00	***Paid Employ*** 11.6~35.8	***Sleep*** 26.4~61.5	***Sleep*** 8.8~32	***Eating*** 5.5~14.2	***Travel*** 9~16	***Perscare*** 8.1~11.5	***Personal care*** 4.6~15.9	***Paid Employ*** 5.6~10.9
Day_2 10.00-12.00	***Paid Employ*** 33~39	***Sleep*** 6.7~18	***Cleaning*** 7.4~9.1	***Eating*** 4.5~12.4	***Eating*** 3.1~9	***Paid Employ*** 9.1~10.8	***Travel*** 6.3~8.3	***Cleaning*** 6.2~10.3
Day_3 12.00-15.00	***Paid Employ*** 17.9~38.7	***Eating*** 5.6~23.7	***Eating*** 3~26.2	***Paid Employ*** 6.8~12.4	***Breaks*** 1.8~14.5	***Cooking*** 7.2~11.9	***Shopping*** 3.4~5.7	***Travel*** 5.3~8.9
Day_4 15.00-17.00	***Paid Employ*** 34.9~38.6	***Eating*** 3.8~16.3	***Eating*** 2.2~9.6	***Paid Employ*** 9.2~11	***Travel*** 5.3~7.1	***Travel*** 7~9.1	***Cleaning*** 4.2~6.6	***Shopping*** 7.3~8.5
Peak 17.00-19.00	***Paid Employ*** 18.2~35.9	***Paid Employ*** 9.1~10.5	***Travel*** 8~13.9	***Chatting*** 6.6~9.6	***Cooking*** 4.3~11.9	***Travel*** 6.1~8.8	***Eating*** 2.6~9.7	***TV/Vdieos*** 7.6~8.8
Evening_1 19.00-21.00	***Eating*** 6.5~20.3	***TV/Videos*** 10~18.3	***TV/Videos*** 7.8~19.2	***Eating*** 6.6~17.8	***Paid Employ*** 8.3~12.7	***Travel*** 5.2~9.8	***Travel*** 4.9~11.9	***Chatting*** 7.2~9
Evening_2 21.00-23.00	***TV/Videos*** 18.4~33.2	***TV/Videos*** 16~26.8	***Resting*** 8.6~10.8	***Eating out*** 5.2~12.4	***Chatting*** 6.8~8.5	***Resting*** 8.7~11.3	***Eating*** 2.4~6.9	***Chatting*** 8~9.7
Night 23.00-03.00	***Sleep*** 37.2~90.1	***Sleep*** 26.1~81.3	***TV/Videos*** 0.4~18.3	***TV/Videos*** 10~17.7	***Resting*** 0.8~6.6	***Eating out*** 6.5~16.3	***Chating*** 0.3~5.3	***Resting*** 0.9~6.5

Table 2.3 Percentage of four main activities for workdays and weekends

The first model (Model 1) explores the effects of selected weather variables on electricity consumption and is as follows:

$$\log(q_{i,t}^h) = \alpha_i^h + \zeta^h \cdot H + \sum_{p=1}^{5} \delta_p^h \cdot W_{p,t} + \sum_{m=2}^{12} \lambda_m^h \cdot M_{m,t} + \sum_{j=2}^{7} \theta_j^h \cdot D_{j,t} + \varepsilon_{i,t}^h \tag{2.1}$$

where $h = 1 - 9$ for each of the 9 periods. $\log(q_{i,t}^h)$ denotes the logarithm of household i's daily electricity consumption in kilowatt-hours during one period of day t. As discussed above, there are 9 periods in a day. The model therefore is run for each period separately; $W_{p,t}$ are the five weather variables; $M_{m,t}$ are dummies of month indicators, and January is selected as the baseline (when m=1) ; $D_{j,t}$ indicate day of week and the reference category is Monday (where j=1); the coefficient δ represents the expected weather effect on consumption, while the coefficients λ and θ quantify the consumption differences between the expected effect (the month i and the day j) and the baseline (January and Monday); H is the public holiday dummy; α_i^h are household fixed effects and $\varepsilon_{i,t}^h$ is a stochastic disturbance term. There may also be unobserved household-specific differences in consumer demand, for example, presence of electric dryers or other appliances. The fixed-effects estimator used can handle it well as this household-level heterogeneity is constant over time.

Although weather has been identified in many studies as an essential factor, no agreement has been reached on which weather variables and in what form they should be added into the modeling. However, heating/cooling degree days, hours of sunshine, rainfall, wind speed and relative humidity are five leading variables that have been used in the relevant research. Model 1 employs all these variables, apart from heating degree days (HDD) and cooling degree days (CDD), which are replaced by air temperature in the equation. The reason for this substitution is that HDD and CDD are used to reflect the non-linear relationship between daily electricity demand and daily temperature. However, although a non-linear response is found in other studies, there is no clear non-linear relationship, but rather a linear correlation in Irish houses (Figure 2.2). One reason may be that the temperature range in Ireland is relatively flat and air conditioning uncommon in Ireland. In addition, Model 1 examines the weather during different periods of the day, rather than daily changes, so the using CDD and HDD would not suit the case.

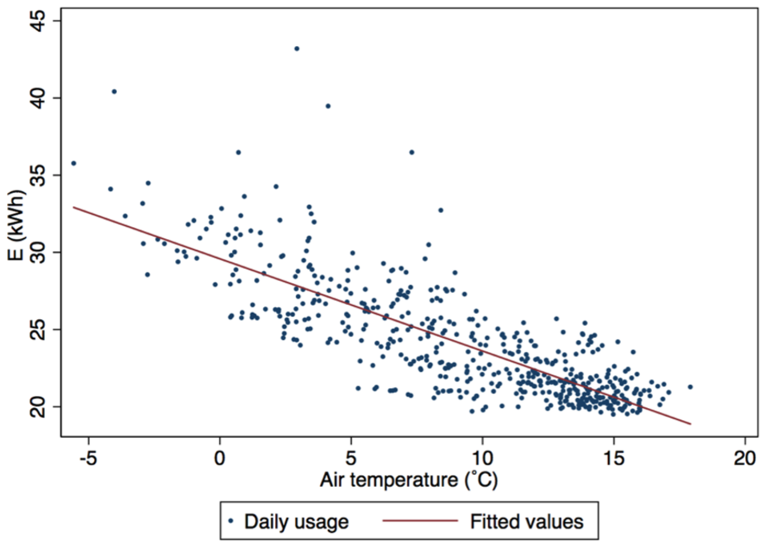

Fig. 2.2 Average daily electricity consumption per household

The second model (Model 2) is based on Model 1 but streamlined to focus on estimating how differently households respond to weather changes on weekends and weekdays. The different consumption patterns can be seen in Figure 2.3.

To estimate the differences, the following model is tested:

$$\log\left(q_{i,t}^{h}\right) = \alpha_i^h + \vartheta^h \cdot D_x + \zeta^h \cdot H + \sum_{p=1}^{3} \delta_p^h \cdot W_{p,t} + \sum_{p=1}^{3} \beta_p^h \cdot W_{p,t} \cdot D_x + \sum_{m=2}^{12} \lambda_m^h \cdot M_{m,t} + \varepsilon_{i,t}^h \quad (2.2)$$

The model is similar to Model 1, apart from the following changes:

1. Day of week dummies are replaced by a workday dummy D_x to estimate the difference between workdays and weekends. It should be highlighted that the definition of working days varies depending on the period of the day. Before 19:00 (peak period), the definition remains the same as the typical sense that Monday to Friday are working days. However, the definitions of working days from 19:00-03:00 are slightly different. For the evening_1 and evening_2, periods workdays are defined as Monday to Thursday, which means 3 days for each weekend because it is sensible to treat Friday evening as the start of a weekend. Additionally, before 23:00 on a Sunday can also be regarded as part of a weekend. However, it may be logical to assume that the behaviour/life pattern for the night period (23:00-3:00) on a Sunday is more similar to a workday. During late evenings/evening_2 on weekends eating out is still the second most common

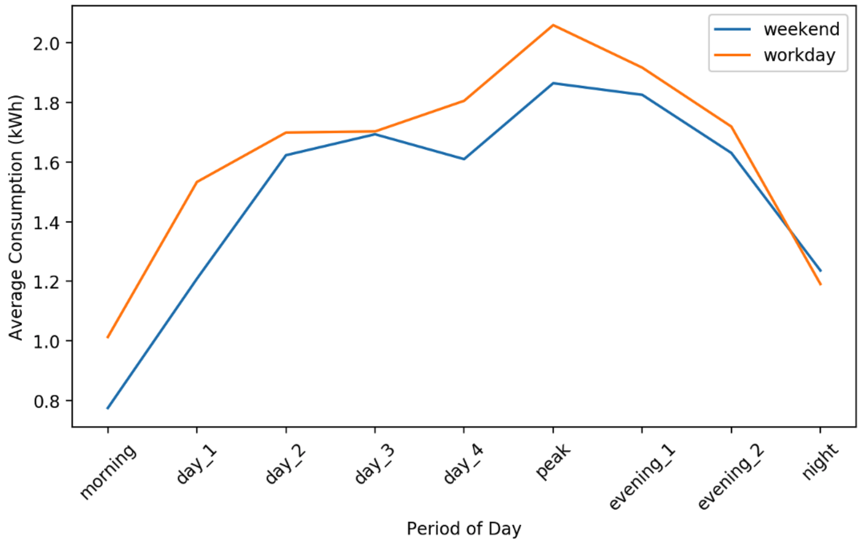

Fig. 2.3 Average daily electricity consumption per household

activity (Table 2.3) and so Sunday evenings should not be treated as workdays. Therefore, the definition of workdays for the night period is Sunday to Thursday. The analysis for holidays/public holidays applies the same rule.

2. Only three weather variables are included in this model. Wind speed and relative humidity are excluded as they have less impact on demand. In addition, the objectives of this model are to examine the differences in response in the main weather factors between weekdays and weekends. Adding variables that have limited effect can overfit the model and may lead to biased results. As a result, Model 2 only keeps three weather variables. It is because: a) the results shown in Model 1 prove that humidity and wind speed have the least and almost negligible effects on demand. A model including the two variables would weaken the model 2) we tested the model with all five weather variables and their interactions, which shows that with or without relative humidity and wind speed included, the results for the other three variables remain almost the same. Therefore, the more concise model with better explanatory power is employed.

In Model 2, the main coefficients of concern are β_p^h and δ_p^h. β_i^h represents the difference in demand between workdays and weekends/holidays caused by weather variable $W_{p,t}$. δ_p^h indicates the possible effects of weather variable $W_{p,t}$ on weekends/holidays. Note that together holidays plus weekends act as the reference category and that holidays are not separated from weekends because less than 10 days in a year are treated as holidays. Results from the interaction between the holiday dummy and weather factors may be biased due to the limited sample size.

2.5 Results

2.5.1 General relationships between weather factors and demands

Analysing the weather sensitivities on periods of day basis will allow us to answer different questions. First, do consumers change their behaviour alongside changes in weather and the seasons? And if so, which weather variables affect the consumption behaviour most significantly? Is there any particular period in which the effect of one specific weather factor dominates? The weather effects on electricity consumption in Ireland do not reflect behaviour change related to heating demand since natural gas is the main heating source in Ireland and electric heating appliance ownership is low, around 10% (CER, 2012a). Instead, the changes in demand reflect how weather factors will affect households' daily behaviour for electricity-intensive activities such as lighting, cooking, and other household appliances (washing machines, dishwashers, dryers, televisions, etc) and so will reflect both variation in household chores and activities as well as whether people are at home or whether have gone outside or away from home.

As the models employ log-linear form, the coefficients describe the percentage change in demand for a one unit increase in that variable. The table presents the estimated coefficients from the models for the nine periods (Table 2.4).

Temperature

Temperature always has a negative effect on household electricity consumption. This is in line with many previous research findings (Blázquez, Boogen and Filippini, 2013; Cosmo et al., 2014) that daily electricity demand decreases when the daily temperature rises. This result holds across all periods, not just for average daily consumption. The reduction in demand with rising temperature could be caused by various drivers including enaging in more outdoor activities and lower heating demand. Considering that the Irish heating system largely depends on natural gas (CER, 2012b), with a higher possibility that the reduction from temperature is from spending more time outside. By contrast, the reason for the negative effect on mornings (6:00-8:00) may be different, since most households should be still asleep or in the bed. The negative sign indicates that people tend to get up slightly later or spend less time on personal care and cooking on warmer days. The response to temperature can be seen as the sensitivity of the activities in this period to temperature change (warm/cold weather. Hence, a higher sensitivity represents the activities/behaviour in that period are more likely to be outdoor activities. From Table 2.4, it can be seen that night (23:00-03:00), and especially early morning (6:00-8:00) are far less sensitive than other periods with less than a 2% reduction. The highest coefficient is in the early afternoon (12:00-15:00), which indicates that the activities in that period can be most sensitive to warmer weather.

Rain

In terms of rainfall, our prior expectations were that higher rainfall could be associated with an increase in electricity demand for all periods. It seems reasonable that the heavier it rains the less likely that people would go outside. As expected, all periods show a negative relationship between rainfall and consumption, except for mornings (6:00-10:00) and late nights. The reversed sign in the early mornings (6:00-8:00) may indicate that the households wake up later when it is raining outside. Moreover, the electricity usage in mornings (8:00-10:00) and late nights are rarely affected by rain. By the time many people have left home to work, while those who stay at home may not be ready to go out immediately for shopping or exercises after breakfasts. Relatively few households are awake after 23:00, most households tend to go to bed earlier on workdays and even on weekends many households won't stay up beyond midnight. This assumption can be verified by the Time-Use Survey in Ireland (Economic and Social Research Institute, 2005), which showed that more than 50% percent of people are sleeping at 23:00-23:59. The figure soars to 85% for 0:00-1:00.

Variables	(1) **Morning** 6.00 -- 8.00	(2) **Day_1** 8.00 -- 10.00	(3) **Day_2** 10.00 -- 12.00	(4) **Day_3** 12.00 --15.00	(5) **Day_4** 15.00 -- 17.00	(6) **Peak** 17.00 -- 19.00	(7) **Evening_1** 19.00 -- 21.00	(8) **Evening_2** 21.00 -- 23.00	(9) **Night** 23.00 -- 03.00
Weather									
Temperature	-0.001792***	-0.005751***	-0.01291***	-0.01402***	-0.01326***	-0.01307***	-0.01242***	-0.01032***	-0.008721***
Rainfall	-0.0002214	0.04434***	0.04876***	0.01696***	0.04374***	0.01897***	0.01339***	0.01001***	0.005665*
Sun duration	0.01921***	0.01038***	0.002493	0.01257***	-0.03527***	-0.04913***	-0.09351***	\	\
Relative humidity	-0.001364***	-0.0007425***	0.001294***	0.002896***	0.001511***	0.001598***	0.001433***	0.0007125***	0.002579***
Wind speed	-0.0006325**	-0.00001235	0.002379***	0.003416***	0.003632***	0.002928***	0.003362***	0.002778***	0.005031***
Time									
Sunday	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
Monday	0.2519***	0.06085***	-0.3173***	-0.2831***	-0.06142***	0.1199***	0.07649***	0.07878***	0.01315***
Tuesday	0.2891***	0.08829***	-0.3536***	-0.3214***	-0.08644***	0.1056***	0.06792***	0.08234***	0.01210***
Wednesday	0.2855***	0.09368***	-0.3522***	-0.3289***	-0.1049***	0.07815***	0.04795***	0.06488***	0.01540***
Thursday	0.2928***	0.09713***	-0.3717***	-0.3506***	-0.1207***	0.06379***	0.03456***	0.06429***	0.03191***
Friday	0.2727***	0.09912***	-0.3376***	-0.3145***	-0.1003***	0.03893***	0.009116***	0.02324***	0.1083***
Saturday	0.05649***	0.08121***	-0.07385***	-0.09508***	-0.003968	0.05539***	0.004411*	-0.02828***	0.1030***
Non-holiday	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
Holiday	-0.08155***	-0.04648***	0.1033***	0.09791***	0.01649***	-0.06722***	-0.04036***	-0.02831***	0.04457***
Month									
January	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
February	0.01436***	-0.003099	-0.07754***	-0.09040***	-0.1293***	-0.1381***	-0.05003***	-0.05716***	-0.07583***
March	-0.008113**	-0.03512***	-0.08563***	-0.09903***	-0.1810***	-0.2989***	-0.09901***	-0.08046***	-0.1120***
April	-0.08523***	-0.08503***	-0.07490***	-0.1029***	-0.2316***	-0.4113***	-0.3476***	-0.1098***	-0.08673***
May	-0.08081***	-0.09148***	-0.09447***	-0.1194***	-0.2253***	-0.4209***	-0.4191***	-0.2003***	-0.07620***
June	-0.07489***	-0.08472***	-0.04169***	-0.06096***	-0.1848***	-0.4117***	-0.4339***	-0.2984***	-0.03452***
July	-0.1195***	-0.1150***	-0.01284**	-0.04466***	-0.2042***	-0.4413***	-0.4692***	-0.2919***	-0.05109***
August	-0.1254***	-0.1164***	-0.003654	-0.03843***	-0.2074***	-0.4392***	-0.4411***	-0.2108***	-0.05412***
September	0.006025	-0.05519***	-0.06729***	-0.07816***	-0.1684***	-0.3890***	-0.2401***	-0.1364***	-0.09618***
October	-0.02453***	-0.04794***	-0.07490***	-0.09983***	-0.1810***	-0.3057***	-0.1242***	-0.1359***	-0.1128***
November	0.01848***	-0.01011***	-0.09065***	-0.1065***	-0.06440***	-0.05878***	-0.08547***	-0.09943***	-0.1045***
December	0.05976***	0.1034***	0.1129***	0.05702***	0.09065***	0.01957***	0.007466**	0.02586***	0.03491***
Constant	-0.2103***	0.2800***	0.5056***	0.9000***	0.5277***	0.8759***	0.8964***	0.7818***	0.4845***
Obeservations	1495843	1496318	1495849	1496196	1496368	1496948	1497084	1497236	1497783
R_2	0.5637	0.4846	0.4322	0.4735	0.4702	0.5116	0.523	0.5553	0.6319

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4 Estimated coefficients from Model 1 of weather effects

Sun

The results of sun duration suggest two clear patterns over the course of a day. The turning point seems to be 15:00, which is consistent with the findings of Harold, Lyons and Cullinan (2015) and Di Cosmo and O 'hora (2017). The more sunshine observed In that period, the less electricity consumed by households. Before 1500 it has an opposite effect with the model producing a positive coefficient. The positive effect may be due to the different nature of the activities during the two periods. Since, as discussed, electricity consumption does not reflect heating demand, the results would indicate that more indoor activities (e.g chores, DIY, gardening) tend to occur over 6:00-15:00 whereas there was greater chance of outdoor activities (e.g., shopping, sports) occurring in the late afternoon and early evening. Furthermore, the larger coefficients in the early evening (17:00-21:00) reveal that for the half year that has sunshine in the early evening (mainly late Spring and summer), willingness to go out is particularly sensitive to sunshine during that period.

Humidity and wind speed

Relative humidity and wind speed show similar patterns in affecting residential electricity demand. They increase demand for electricity for all periods after 10:00. In terms of wind speed, it has limited impact on electricity demand in the early mornings (6:00-10:00) with less than a 0.005% reduction in demand during 6:00-8:00 and with an insignificant coefficient at 8:00-10:00. On the other hand, relative humidity has a negative relationship with consumption during the same period. Humidity have a compounding effect with temperature, where air temperature with higher humidity may give a colder apparent temperature. However, all the impacts from humidity and wind speed are of negligible magnitude with under 0.5% change in demand. Therefore, these two variables will be removed in the following model where the focus is to identify the differences between weekdays and weekends for each of the main weather factors.

2.5.2 Behaviour difference between weekends and workdays

Based on the overview of the effects of the weather factors, this section attempts to identify and answer the following questions: are there differences in demand between weekends and workdays in weather sensitivities? What differences in daily routine between weekends and workdays can cause any discrepancies? Estimated results are shown in Table 2.5.

Temperature

As expected, the results suggest that temperature have a negative effect on the demand among both weekends and workdays of all periods. However, the exception is weekend early mornings. The reason for this unexpected impact of temperature is not particularly clear. However, it may be that while most families are asleep during that period, early birds are willing to get up earlier on warmer weekend days. Of all periods, early weekend mornings (6:00-10:00) have the least impact, which suggests that the early morning is the most insensitive period. The behaviour during that period is robust and less likely to be changed by temperature.

Furthermore, weekends are in general more sensitive to temperature change than weekdays. This difference can be explained by more activities occurring indoors. However, the difference in the early evening (19:00-21:00) seems negligible. It could be explained by that limited activities would occur during the post-dinner time on both weekends and workdays, since many would enjoy an indoor relaxing time after dinner. The largest difference appears at night (23:00-3:00), which is in line with expectations. People would be more likely to go out later and stay out later on weekends/holidays, especially on warmer days, whereas people tend to go to sleep earlier on workdays even on warmer days.

Rainfall

The effects of rain represent to what percentage the electricity demand would change by the rainfall. From the results it is possible to infer how flexible plans or activities are in a given period. A period with higher sensitivity to rain that there may be more outdoor activities or households prefer to go out during that period.

The midday (10:00-15:00) and early evening (19:00-21:00) periods on workdays are the only time slots with greater sensitivities than their counterparts on weekends. It is noteworthy that the sensitivities during the midday period (10:00-15:00) on weekdays are exceptionally higher than any other period in either weekends or workdays. It indicates that stay-at-home family members tend to go out during that period on weekdays. While on weekends, the households may not be able to go out early due to more chores and family care. This period actually shows the largest gap between weekends and workdays, which indicates the large underlying difference in daily routines between weekends and workdays are in the midday period. For example, people might regularly go out on workdays while stay at home on weekends during this period. Likewise, the significant high coefficients of over 0.1 are also found on weekend mornings (8:00-10:00) and nights (23:00-3:00). The unusually high sensitivities on weekend mornings may be due to more chores done or sports activities.

Interestingly, in spite of a smaller difference compared to the 10:00-15:00 period, workdays in the early evening (19:00-21:00) are surprisingly more sensitive than on weekends, whereas a plausible hypothesis would be that evenings should be more sensitive on weekends. It could be a result of the timing of outdoor activities on weekdays since people would only be able to go out during that period on weekdays while they could choose other time periods on weeknds. In addition, households may have dinner at a slightly later period on weekends. For many, this period may be post-dinner on workdays (19:00-21:00) but may actually be dinner time for weekends. Therefore, whether there is rain or not may have a greater effect on workdays, due to the lower probability of going out in the evenings on workdays.

The only negative effects are for early mornings (6:00-10:00) on workdays. It is possible that the heavier it rains, the earlier people may feel to leave houses to avoid the traffic jam, although the effect significantly drops from -3.3% to -0.8% at 8:00-10:00. It is a solid proof that the negative effect mainly comes from the behaviour of workers in the house since the effect falls to nearly zero when it reaches the start of work hours.

Variables	(1) **Morning** 6.00 -- 8.00	(2) **Day_1** 8.00 -- 10.00	(3) **Day_2** 10.00 -- 12.00	(4) **Day_3** 12.00 --15.00	(5) **Day_4** 15.00 -- 17.00	(6) **Peak** 17.00 -- 19.00	(7) **Evening_1** 19.00 -- 21.00	(8) **Evening_2** 21.00 -- 23.00	(9) **Night** 23.00 -- 03.00
Weather									
Temperature	0.002524***	-0.003289***	-0.01518***	-0.01767***	-0.01343***	-0.01192***	-0.01074***	-0.009082***	-0.01038***
Temperature * Workday	-0.004587***	-0.003138***	0.002516***	0.005478***	0.001465***	0.0008027**	0.0003072	0.001683***	0.005085***
Rainfall	0.04134***	0.1401***	0.02043**	0.03680***	0.08389***	0.04746***	0.03440***	0.04573***	0.1062***
Rainfall * Workday	-0.07466***	-0.1499***	0.1075***	0.09673***	-0.02118***	-0.01767***	0.01490***	-0.02919***	-0.09189***
Sun duration	0.04544***	0.04637***	0.03262***	-0.004638	-0.07735***	-0.08104***	-0.09148***	\	\
Sun duration * Workday	-0.02227***	-0.04787***	-0.05150***	-0.01570**	0.02682***	0.009996	-0.04168***	\	\
Time									
Weekend	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
Workday	0.3123***	0.1146***	-0.3311***	-0.3451***	-0.1178***	0.04710***	0.05132***	0.06453***	-0.1198***
Holiday	0.04419***	-0.01954***	-0.03182***	-0.02297***	-0.01975***	-0.02584***	-0.01376***	0.000299	-0.0006169
Month									
January	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*	*Ref*
February	0.004319	-0.008879**	-0.07065***	-0.09739***	-0.1371***	-0.1472***	-0.05878***	-0.06146***	-0.08921***
March	-0.004226	-0.02927***	-0.1076***	-0.1486***	-0.2024***	-0.3194***	-0.1171***	-0.08663***	-0.1314***
April	-0.1016***	-0.08898***	-0.07714***	-0.1317***	-0.2542***	-0.4451***	-0.3794***	-0.1315***	-0.1149***
May	-0.1026***	-0.09503***	-0.09476***	-0.1480***	-0.2462***	-0.4544***	-0.4558***	-0.2277***	-0.1127***
June	-0.1000***	-0.09125***	-0.03643***	-0.08592***	-0.2083***	-0.4538***	-0.4787***	-0.3347***	-0.07309***
July	-0.1447***	-0.1189***	0.002915	-0.05617***	-0.2208***	-0.4777***	-0.5048***	-0.3240***	-0.07985***
August	-0.1503***	-0.1238***	0.0003798	-0.05829***	-0.2266***	-0.4726***	-0.4770***	-0.2396***	-0.09119***
September	-0.01313***	-0.05696***	-0.05770***	-0.09163***	-0.1853***	-0.4187***	-0.2734***	-0.1618***	-0.1238***
October	-0.04151***	-0.05426***	-0.06378***	-0.1088***	-0.1861***	-0.3289***	-0.1444***	-0.1533***	-0.1291***
November	0.007114**	-0.01446***	-0.08299***	-0.1136***	-0.06349***	-0.06272***	-0.09159***	-0.1012***	-0.1115***
December	0.04578***	0.09784***	0.1216***	0.06538***	0.08648***	0.01402***	0.001041	0.02436***	0.02734***
Constant	-0.3460***	0.2122***	0.6075***	1.1735***	0.7033***	1.0694***	1.0534***	0.8668***	0.8850***
Observations	1495843	1496318	1495849	1496196	1496368	1496948	1497084	1497236	1493682
R_2	0.5647	0.4844	0.4323	0.4732	0.4699	0.5106	0.5227	0.5549	0.6315

*** p<0.01, ** p<0.05, * p<0.1

Table 2.5 Estimated coefficients from Model 2 Workday and Weekend differences

Sun

The prior expectation was that longer sun duration should be associated with decreased electricity demand, as people are more likely to go out on a sunny day. However, contrary to this expectation, the opposite findings are found in the weekend mornings (6:00-12:00) and the early mornings on workdays (6:00-8:00). The increased consumption in sunny early mornings for both weekends and workdays could be partly explained by a relatively early wake-up times. It may be that sunshine gives individuals the feeling of being energetic and increases the possibility of going out later. This effect is especially clear on weekends since on workdays people are more likely to maintain their routines in the early evenings and may not change their behaviour easily in response to greater sunlight.

In the mornings, from 8:00 until 12:00, interesting and unusual differences between weekends and workdays appear. While sunshine hours now have a negative effect on workdays, the positive effects continue on weekends during this period. The increased demand reverse the common idea that families or individuals are more willing to spend time outside, especially on a sunny weekend. However, this may be capturing an effect of preferences of specific activities/routines on weekend mornings. The positive results could be due to the fact that households have propensities to carry out housework on weekend mornings, before heading out in the afternoons. Additionally, some types of chores are more likely to give a rise to electricity consumption on a sunny day. For example, roughly 30% of households do not own a dryer (Leahy, Lyons and Walsh, 2012), so they would choose to do laundry on a sunny day and even households with dryers might choose to reduce their bills and dry their clothes outside. Thus, the positive effects may reflect the behavioural habits on weekend mornings. It should be highlighted that the positive impacts are decreasing from the 8:00-10:00 morning period and becomes insignificant by mid-afternoon (12:00-15:00), which is the only insignificant period. The reason may be that on weekends, family meals are common at lunchtime and sun duration does not affect these behaviour patterns. After that point, sun sensitivities on weekends gradually increase from -7.6% to -9% during the 15:00—21:00 period. This may indicates more sun-related outdoor activities later on weekend days, compared to "housework mornings".

On the other hand, negative effects are seen during almost every weekday period. However, it is still important to note that compared to a relatively constant sensitivity of -1.8% in the period 8:00-15:00 on workdays, a clear increasing pattern is shown for the period after 15:00. The sensitivities are much higher than during the first half of day, with values over 5%. This provides strong evidence that similar to behavioural patterns on weekends, afternoons are generally more flexible on workdays. Higher negative coefficients imply that the households have more free or flexible time and are more likely to go out. Nevertheless, it should be pointed out that a weaker sensitivity to sunshine duration does not necessarily mean that people are less likely to go out during that period. Unlike for rain, whether or not there is sunshine would generally not affect people's movements or activities. For instance, if one is used to shopping for groceries for the family in the morning, he/she would not cancel or delay the shopping just because of cloudy weather. Therefore, a relatively smaller sensitivity should be interpreted as a higher possibility that one's time is occupied by regularly scheduled plans, which could be either indoors or outdoors.

Similar to the results shown in the rain effects above, early evening (19:00-21:00) is the only period when workdays are more sensitive than holidays. It should be kept in mind that only half of a year (mainly late spring and summer) has sunshine during the period. The findings in this period therefore largely limit and reflect the behaviour in summer. As suggested in the rain section, only in this evening period are people still able and more willing to go out on workdays, compared to late evenings and nights. Another interesting finding is that sun duration in this period (19:00-21:00) of workdays has the largest effect among all other sunshine effects on encouraging people to go out. Note that as no sunshine exists after 21:00, no sun effect can be tested for those periods.

2.6 Discussion and conclusion

This study set out to examine the behaviour of residential customers exposed to different weather conditions in different periods of a day using unbalanced panel data from the Irish Smart Metering Electricity Consumer Behavioural Trial (CER, 2012a). To conduct the analysis, half-hourly electricity consumption data from 3827 household meters over one year were aggregated into daily usage for every period of a day. Together with the weather variables, fixed-effects models with robust standard errors clustered at the household level were used to control for unobserved household-specific factors, which gives a better understanding of households' response to weather factors at different times of the day.

Overall, this chapter has demonstrated from the first model that in general although temperature has robust and relatively flat effects on electricity demand across all periods, rain and sunshine duration show greater potential to affect individual behaviour and daily routines. The demand response to temperature could be interpreted as warm/cold sensitivities of the activities in that period. As expected, the periods from 10:00-21:00 present higher sensitivities than early mornings and nights, since more activities occur in those periods. Although night time periods (21:00-3:00) have smaller sensitivities than daytime, they are still much more sensitive than early mornings. Not many activities occur over 6:00-10:00 when most people are getting up and going off to work. The rainfall sensitivity may act as an indicator of whether outdoor activities occur more often in that period. It should be noted that the results mainly reflect the behaviour of the households who are in the house during day-time, and the proportion of these households account for over 68% of the sample. One of the lowest rainfall sensitivities appears at 12:00-15:00 which is cooking and lunchtime that the possibility of going outdoor would be relatively small. This finding is consistent with the Irish Time Use Survey (Table 3) that for those who are not working (61 82%), eating is one of the main activities. Apart from the similar pattern of lower sensitivities at the start and end of a day, the relatively high coefficients at two periods 10:00-12:00 and 15:00-17:00 reveal that individuals could be more used to or prefer going out during these periods. On the other hand, the impact of sunshine on households' behaviour differs from rainfall, although both affect the chances of going out. Negative sunshine sensitivities represent the time availability and willingness to go out of households, in other words, how flexible the period so that one can response to good weather. The results strongly support the interpretation that the sensitivity gradually increases from late afternoon (15:00) and peaks in early evening (19:00-21:00), compared to the small and positive sensitivities shown in the mornings.

The responses to weather factors for weekends and workdays are tested in the second model. The differences are caused by different household patterns between weekends and workdays. In terms of temperature sensitivities, weekends are more sensitive than workdays because households have more available time to spend, while the sensitivity difference is minimal. The biggest difference is seen on the night period (23:00-3:00), where people would care less about the cold weather and be more likely to go out on weekends. Moreover, the rainfall results suggest two clear patterns: before 15:00 workdays are more sensitive than weekends, although the effects of workday mornings are insignificant; after 15:00, weekends show higher sensitivities than workdays, apart from 19:00-21:00. The findings imply that more rain-sensitive activities occur before mid-afternoon during weekdays, while these activities (e.g. outdoor activities) occur at 15:00 afterward in general. The difference in life patterns between workdays and weekends are also revealed by sun duration. In the mornings (6:00-12:00), while sunlight has positive effects on weekends: the longer the duration of sunshine, the greater the consumption during those periods on the weekend. It could be associated with more sun-sensitive chores on the weekend mornings. The pattern changes after 15:00 – households seem more flexible at this time on weekends. And both weekends and workdays reveal increasing sensitivities during that period, especially workdays, which soars after 15:00 from -1.6% to a maximum of 13%. One especially interesting finding is that early evening (19:00-21:00) is the one period when weekends are less sensitive to all weather factors than workdays. This may be unexpected but could be explained by the fact: Compared to weekends, early evenings on weekdays might be the most flexible time where outdoor activities are possible, especially for those employed households, so the period is more sensitive to weather.

The study could be instructive for understanding household energy consumption behaviour. First, the weather sensitivity analysis provides an overview of households' behaviour/life pattern without the assistance of a survey. Especially, sunshine and rain sensitivities may be considered as proxies of whether a period is with more flexibility and whether people tend to leave home (or use less) at certain periods respectively. Furthermore, analysing the differences in patterns between weekdays and weekends can help identify which periods on weekends or workdays are more sensitive and flexible. With more knowledge of people's life pattern among different periods the tariff structure design could be more efficient in shifting energy demand. Secondly, with deeper analysis on individual level, for example, combing the attitude and behaviour data in the survey with the weather sensitivity patterns, it could create an initial profile of a family's daily activities. For instance, if a family displays relatively higher weather sensitivities, this may reflect greater flexibility in their living patterns. A target tariff aimed at those families may help shift peak electricity demand. These potential implications lead to possible future research in improving residential customers' consumption profiles. It would be interesting to categorise the households by their weather sensitivities and to examine if the weather sensitivities are associated with certain demographic factors, which may provide a cheaper and faster means of understanding a household's social-economic profile. Data mining tools are helpful in this case to cluster and classify the residential customers, which offers a new angle to summarise and depict customers' activity patterns by using only weather and consumption data. By using only weather indicators this approach can be faster and simpler than traditional methods —such as surveys or questionnaires — in identifying which period are more flexible at the household level.

Reference

Albert, A. and Rajagopal, R. (2013) ‘Building dynamic thermal profiles of energy consumption for individuals and neighborhoods’, in Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013. doi: 10.1109/BigData.2013.6691644.

Anderson, B. et al. (2017) ‘Electricity consumption and household characteristics: Implications for census-taking in a smart metered future’, Computers, Environment and Urban Systems. Elsevier Ltd, 63, pp. 58–67. doi: 10.1016/j.compenvurbsys.2016.06.003.

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

Auffhammer, M. and Aroonruengsawat, A. (2012) ‘Simulating the Impacts of Climate Change, Prices and Population on California’s Residential Electricity Consumption’, Climate Change, 109(S1).
Available at: https://ei.haas.berkeley.edu/research/papers/ (Accessed: 5 January 2018).

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

Bessec, M. and Fouquau, J. (2008) ‘The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach’, Energy Economics, 30(5), pp. 2705–2721. doi: 10.1016/j.eneco.2008.02.003.

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.

Carroll, J., Lyons, S. and Denny, E. (2014) ‘Reducing household electricity demand through smart metering: The role of improved information about energy saving’, Energy Economics. doi: 10.1016/j.eneco.2014.07.007.

Commission for Energy Regulation (CER). (2012a). CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0012-00. www.ucd.ie/issda/CER-electricity

Commission for Energy Regulation (CER). (2012b). CER Smart Metering Project - Gas Customer Behaviour Trial, 2009-2010. [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0013-00. www.ucd.ie/issda/CER-gas

Cosmo, V. Di, Lyons, S. and Nolan, A. (2014) 'Estimating the impact of time-of-use pricing on irish electricity demand', Energy Journal, 35(2), pp. 117–136. doi: 10.5547/01956574.35.2.6.

Cosmo, V. and O 'hora, D. (2017) 'Nudging electricity consumption using TOU pricing and feedback: evidence from Irish households', Journal of Economic Psychology, 61, pp. 1–14. doi: 10.1016/j.joep.2017.03.005.

Costa, D. L. and Kahn, M. E. (2010) Why has California's Residential Electricity Consumption been so flat since the 1980s?: A Microeconometric Approach, NBER Working Paper No. 15978. 15978. Cambridge, MA.

Davies, M. (2010) 'The relationship between weather and electricity demand', Proceedings of the IEE Part C: Monographs, 106(9), p. 27. doi: 10.1049/pi-c.1959.0007.

Eid, C., Hakvoort, R. A. and de Jong, M. (2017) 'Global Trends in the Political Economy of Smart Grids', The Political Economy of Clean Energy Transitions, 1(May 2019), pp. 250–270.

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.

Harold, J., Lyons, S. and Cullinan, J. (2015) 'The determinants of residential gas demand in Ireland', Energy Economics. 51, pp. 475–483. doi: 10.1016/J.ENECO.2015.08.015.

Henley, A. and Peirson, J. (1998) 'Residential energy demand and the interaction of price and temperature: British experimental evidence', Energy Economics. 20(2–6), pp. 157–171. doi: 10.1016/s0140-9883(97)00025-x.

Kavousian, A., Rajagopal, R. and Fischer, M. (2013) 'Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior', Energy, 55, pp. 184–194. doi: 10.1016/j.energy.2013.03.086.

Leahy, E. and Lyons, S. and Walsh, S. (2012). 'Electrical Appliance Ownership and Usage in Ireland,' Papers WP421, Dublin: Economic and Social Research Institute (ESRI).

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. Elsevier, 48, pp. 240–248. doi: 10.1016/J.ENBUILD.2012.01.037.

Met Éireann. (2018). Weather Statements. Accessed from: https://www.met.ie/climate/past-weather-statements

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.

Moral-Carcedo, J. and Vicéns-Otero, J. (2005) ‘Modelling the non-linear response of Spanish electricity demand to temperature variations’, Energy Economics. 27(3), pp. 477–494. doi: 10.1016/J.ENECO.2005.01.003.

Newsham, G. R. and Bowker, B. G. (2010) ‘The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review’, Energy Policy, 38, pp. 3289–3296. doi: 10.1016/j.enpol.2010.01.027.

Pardo, A., Meneu, V. and Valor, E. (2002) ‘Temperature and seasonality influences on Spanish electricity load’, Energy Economics. 24(1), pp. 55–70. doi: 10.1016/S0140-9883(01)00082-2.

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. Elsevier, 87(11), pp. 3538–3545. doi: 10.1016/J.APENERGY.2010.05.015.

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.

Torriti, J. (2012) ‘Price-based demand side management: Assessing the impacts of time-of-use tariffs on residential electricity demand and peak shifting in Northern Italy’, Energy. 44(1), pp. 576–583. doi: 10.1016/J.ENERGY.2012.05.043.

Trotter, I. M. et al. (2016) ‘Climate change and electricity demand in Brazil: A stochastic approach’, Energy. Elsevier Ltd, 102, pp. 596–604. doi: 10.1016/j.energy.2016.02.120.

Valor, E., Meneu, V. and Caselles, V. (2001) ‘Daily air temperature and electricity load in Spain’, Journal of Applied Meteorology. American Meteorological Society, 40(8), pp. 1413–1421. doi: 10.1175/1520-0450(2001)040<1413:DATAEL>2.0.CO;2.

Wangpattarapong, K. et al. (2008) 'The impacts of climatic and economic factors on residential electricity consumption of Bangkok Metropolis', Energy and Buildings, 40(8), pp. 1419–1425. doi: 10.1016/j.enbuild.2008.01.006.

Yohanis, Y. G. et al. (2008) 'Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use', Energy and Buildings. Elsevier BV, 40(6), pp. 1053–1059. doi: 10.1016/j.enbuild.2007.09.001.

Chapter 3

Machine Learning and residential electricity consumption: Which households are more responsive to weather?

3.1 Introduction

Past quantitative studies of residential energy consumption have mainly focused on energy tariff pricing, explanations for differences in energy consumption, and models to predict consumption. However, many of these studies are based on aggregate levels of consumption, particularly those using econometrics (Bianco, Manca and Nardini, 2009; Karanfil, 2009; Sanquist et al., 2012). The objectives of these studies are various, including the research on relationships between daily consumption and household social-economic background (Hackett and Lutzenhiser, 1991; Druckman and Jackson, 2008; Jones, Fuertes and Lomas, 2015); studies focused on effects of weather variables on the total regional electricity consumption (Valor, Meneu and Caselles, 2001; Pardo, Meneu and Valor, 2002; Hor, Watson and Majithia, 2005). Due to the limitations in the resolution of their data however, these researchers were unable to conduct more detailed studies. As installation of smart metering in households has increased in recent years, analysis of high-resolution electricity consumption data becomes possible. The nature of high-frequency data brings opportunities to understand energy consumption with a granularity that would have been unimaginable even a few years ago. As a result, there are now new areas of research available arising from this high-resolution data in the energy sector.

One main area of focus is load management, especially the prediction of electricity consumption, which is of interest to both utilities and policymakers. Previous prediction models have generally been based on aggregated grid consumption data. Now with the technological advancements in metering, the high-frequency load data has the potential to help related parties to understand consumer behaviour better to achieve higher efficiencies. Prediction models can massively benefit from such data and significantly improve their accuracy (Beccali et al., 2008; Ghofrani et al., 2011). Another research question which has been constantly discussed is dynamic pricing of electricity (Faruqui and Malko, 1983; Sanghvi, 1989; Herter, McAuliffe and Rosenfeld, 2007; Faruqui and Sergici, 2010; Alberini and Filippini, 2011). The deployment of smart meters enables utilities to set dynamic pricing structures than flat prices. There have been a great number of trials for various types of pricing schemes (Newsham and Bowker, 2010; Haider, See and Elmenreich, 2016), e.g. time-of-use tariffs, critical peak prices and etc. By introducing fluctuating prices, it could be helpful to reduce energy consumption and save the environment to some extent. Another motivation for dynamic pricing is that utilities intend to encourage customers to shift away from the peak times to reduce the power load during critical periods (Herter, 2007; Faruqui and Sergici, 2010). To analyse the efficiency of the tariff design, the effects of the pricing structures can be investigated thoroughly using high-resolution data.

Furthermore, understanding the correlations between customers' social-economic profiles and electricity consumption is also one of the classic applications of smart metering data (McLoughlin, Duffy and Conlon, 2012; Beckel et al., 2015). Accurate segmentation of electricity customers can assist in higher energy efficiencies and lower operation losses. Previously, using only aggregated daily or monthly household consumption data, it was difficult to look into the details of how and why households' consumption behaviours differ during specific time periods (Cramer et al., 1984; Silk and Joutz, 1997; Kaza, 2010). Previous studies therefore could only focus on longer periods of household consumption to explore differences in social-economic backgrounds or of property characteristics of houses. With smart metering data, utilities and researchers can finally create and understand the customer demand curves and the habits and behavioural patterns underlying the profiles. Due to the huge volumes of data involved and the complexity of the data processing, machine learning techniques, such as clustering and classification, have been increasingly adopted rather than more traditional econometric tools.

The majority of research featuring data mining techniques is focused on how to cluster customers based on their load curves (Räsänen et al., 2010; McLoughlin, Duffy and Conlon, 2012; Razavi et al., 2019). The objectives of those studies are to identify the connections between the demand curves and the characteristics of households. However, there have been few studies using these tools on smart metering data for behavioural studies, and in particular there has been limited work on the effects of weather on household behavioural patterns. Load curves can partially reflect households' consumption patterns, for instance, identifying peak times or the amount consumed during a specific period. Nevertheless, an aggregated curve cannot reveal household preferences in any detail. Greater understanding of household behavioural patterns could benefit both policymakers and utilities.

To fill the gap, our main objectives are to understand how household daily life patterns are reflected in the demand response to weather sensitivities. This study brings together the smart metering and the survey data from the Irish Electricity Smart Metering Customer Behaviour Trials in 2012 with the weather data during the trial. Weather sensitivities of electricity consumption of each household during different periods of the day are the core of this study. They are used as proxies to discover household behaviour patterns, for example, when members of a household normally need to go out and at what time of day people are more likely to have spare time (or at least when they are most flexible in terms of their behaviour). The work proposes a novel method using machine learning techniques to identify the patterns using a two-step process: 1) Define the demand change indexes under different weather conditions; and 2) Employ clustering techniques on the indexes defined in Step 1 to generate representative sensitivity curves. With this method, the study offers a new perspective on the differences in household responses to weather changes drawing on time of use preferences derived from smart metering data. Combining these results with load curves can provide a better understanding of daily residential electricity consumption patterns.

This remainder of the paper is organised as follows: Section 2 presents a literature review of past studies of smart metering data, especially works on electricity consumer segmentation and the correlations between weather and electricity consumption. Both the scope and methods are discussed. Section 3 includes the data and the methodology used here, consisting of data prepossessing, definition of demand change indexes, and the algorithms and a brief description of the clustering process. In Section 4, the results of the weather sensitivity curves are presented and explained in detail. A summary of the work and the conclusions are drawn in Section 5.

3.2 Literature review

Among all the grid operation improvements, the deployment of smart meters particularly benefit short-term load management. While short-term (hourly to daily) load forecast plays a critical role in load management, it has been rather difficult to model the demands by low-frequent data. Under this circumstance, forecasting for residential electricity demand particularly benefits from smart meter data. With household level load data, the models for both long-term and short-term demand forecasting have been well-established. Quilumba et al. (2015) propose improving the accuracy of short-term load forecasts by considering customer behaviour. Using clustering techniques on smart meter data, they create models for load forecasts from 30 minutes up to one day-ahead predictions. Taieb et al. (2016) prove that for disaggregated demand, an additive quantile regression model outperforms the traditional model with an normality assumption, based on the smart metering data from a trial in Ireland over a period of 1.5 years. Ghofrani et al. (2011) combines traditional Gauss-Markov process modelling with automatic meter readings to achieve a higher prediction accuracy, although it increases the computational cost. Some other studies using smart meter data focus on identifying the efficiencies of domestic appliances. Firth et al. (2008) attempt to reveal the trends in the use of appliances from a high-resolution dataset. They found that a 10.2% rise of "standby" appliances (such as consumer electronics) consumption accounts for the largest share of the overall demand increase. Weiss et al. (2012) prove that disaggregation of individual appliances is possible by using a set of algorithms on smart metering data.

Apart from the modelling-related research, another stream of studies using smart meter focus on the effects of household characteristics on residential consumption. Gouveia and Seixas (2016) combine the meter readings with a door-to-door survey of 110 questions administered to 265 households in Portugal to unravel residential consumption profiles. They carried out clustering analysis using daily consumption and formed three profiles. The main variables used for profile analysis included: dwelling location and type, age, gender and educational attainment of household members. A U-shape pattern with higher consumptions at the beginning and end of a year and lower demand in the middle was found to be the most common type accounting for 77% of the households. Beckel et al. (2014) attempted to reveal household characteristics purely from consumption data with a supervised method. In their work the Irish CER trial data was used and the household's socio-economic status, appliance stock, properties of the dwelling, and the consumption behaviour of the occupants were considered as class labels in the research. The electricity consumption profiles were firstly formed through different indexes of consumption behaviour, for instance, the ratio of peak to off-peak. Then by a supervised-learning process, with input of electricity consumption data, the model would be able to identify household characteristics only depending on the consumption patterns. The experimental results show that among all other household characteristics, the occupancy state of the house, the number of persons in the house and the appliance stock can be identified directly from the consumption load profiles very well with an accuracy of more than 70%.

3.2.1 Data mining methods in smart metering data

The rich information brought by high resolution real-time smart meter data can improve the efficiency of grid operations. However, such massive data flows pose a major challenge for the utilities to store and extract knowledge from the data (Viegas et al., 2015). Traditional tools such as database software are inadequate when dealing with huge amounts of data. In response, computational techniques, particularly, machine learning, have become increasingly appealing.

The applications can vary from operations, such as load forecasting, simulating Demand Side Management (DSM), and detecting bad data, to marketing – tariff design and potential customer identification. The core of the implementation is the segmentation of electricity consumers and load clustering. Wijaya et. al. (2015) use a cluster-based method to achieve short-term (1 hour and 24 hours ahead) electricity demand forecasting. Some argue that Support Vector Regression (SVR) is one of the most effective models to forecast electricity consumption (Chen, Chang and Lin, 2004; Sapankevych and Sankar, 2009; Cao and Wu, 2016; Chen et al., 2017). Wijaya et al. (2015) compared different algorithms including SVR, linear regression, and cluster-based aggregate forecasting (CBAF) and they suggest there is no single best algorithm in forecasting and use their own algorithm for clustering. From a review of load forecasting studies, it can be seen that Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the three most widely used metrics to evaluate the accuracy of forecasting algorithms. However, it should be noted that those effectiveness metrics are not the only basis for algorithm measurement. Data structure and other relevant practical issues should also be considered.

McLoughlin et al. (2015), also drawing on Irish CER trial data, use clustering approaches to explore household load profile information. Unlike forecasting studies, they focused on the effectiveness of customer clustering by their household characteristics. The diurnal, intra-daily and seasonal consumption patterns were all examined. In order to evaluate the different clustering techniques, three of the most widely used algorithms are investigated: k-means; k-medoid and Self Organising Maps (SOM). They use a Davies–Bouldin (DB) validity index to compare the effectiveness of the algorithms as well as to determine an appropriate number of clusters. Across the three techniques, the number of clusters was varied between 2 and 16. The results show that SOM and K-means have similar higher clustering power and that 8 to 10 clusters are the optimal numbers in this case. It is important to note that the optimal cluster number can vary depending on the objectives of the research, the features of datasets and even the selection of validity indices. Hierarchical clusterings are another popular set of algorithms for residential customer segmentation (Chicco, Napoli and Piglione, 2006; Chicco, 2012). Al-Wakeel et al. (2017) use k-means cluster analysis for load estimation study on the CER trial data. They suggest that compared to other algorithms, the significant advantages of k-means are simplicity and efficiency, particularly when considering the computational cost. Four different varieties of K-means distance functions are compared: Average Euclidean distance, Average Manhattan distance, Average Canberra distance, and Average Pearson correlation distance. The results of MAPE and RMSE indexes reveal that Canberra produces more accurate forecasts and the smallest error distributions. Gouveia and Seixas (2016) employ hierarchical clustering using Ward's Method based on the Squared Euclidean distance. They tested a range of numbers for cluster from 3 to 12. Since they conclude that increasing the number of clusters captures more information, they opted for the 10 clusters variant. One difference in conducting their cluster analysis is that they used mean daily consumption data to create a year profile, rather than using hourly data for a daily profile in other studies. In addition, the raw data was not normalised and the shapes of the profiles mainly reflect the magnitude of the consumption.

In summary, the key process of applying clustering analysis is to determine suitable algorithms and the number of clusters. No single algorithm outperforms the others in all situations with regard to residential electricity customer clustering. Any decision or selection must be based on the aims of the research and nature of the data structures. In particular, K-means and its relevant algorithms are mainstream choices that have been abundantly discussed for the case of residential load profile clustering. However, clustering on the basis of weather sensitivities has rarely been explored in the past. Traditional consumption load profiles mainly focus on load forecasting (and forecast accuracy). On the other hand, clustering on weather sensitivities, might offer new perspectives and approaches to customer behaviour pattern studies that the weather response may be a good indicator for behavioural patterns.

3.2.2 Weather factors in residential electricity demand

Due to the lack of high-resolution consumption data at household level, past weather studies have mainly been conducted at the regional level using aggregated data. There are two types of research typically incorporating the weather factors: (i) demand forecast models and (ii) econometric models to identify the effect of different factors on electricity demand. Taieb et al. (2016) conducted a quantile regression to improve forecasting accuracy based on the Irish CER trial. In their model, the only weather variable included is outdoor temperature to control its effect in forecasting demand. Beccali et. al (2008) assessed the weather sensitivity on short-term household electricity consumption using cooling and heating degree-days (CDDs and HDDs) as temperature proxies. These two proxies are commonly employed when a non-linear relationship between temperature and demand is assumed (Fan and Hyndman, 2011; Blázquez Gomez, Filippini and Heimsch, 2013). Other weather factors that have been considered are relative humidity, humidex index, global solar radiation, wind speed, and atmospheric pressure. Some researchers (Albert and Rajagopal, 2012; Fikru and Gautier, 2015) claim that the main contributors are humidity index, CDDs, and HDDs while other variables are negligible in affecting residential electricity consumption. Henley and Peirson (1998) modelled the relationship between residential demand and price and temperature in the UK using a fixed-effects model and found a negative correlation. The opposite result was also shown in Wangpattarapong et al. (2008) in examining the impacts of climatic and economic factors on energy consumption in Bangkok. They use cooling-degree days as their temperature variable and find that a significant positive relationship exists. Although the results in these two studies seem contradictory, the different effects of temperature may be the product of geographical location or climate zone. Whereas peak consumption in Bangkok is from air conditioning on the hottest summer days, peak UK electricity demand comes in winter. Hor, Watson and Majithia (2005) find a very weak negative effect of rainfall for monthly demand from 1983 to 1995 in the UK. Apart from temperature, which is usually seen as the main driver of a weather effect, humidity, wind speed, degree of cloudiness, and barometric pressure are also often discussed in related studies (Pardo, Meneu and Valor, 2002; Albert and Rajagopal, 2013).

However, to the extent they are considered, weather conditions have been treated as exogenous variables to control the effects of interest. Perhaps surprisingly, investigating the impacts of weather changes on households' daily life patterns through the electricity demand response to weather variables have been rarely seen. In the next section on Methodology, a novel approach for using weather sensitivities of consumption as proxies for household behaviour pattern will be explored in detail.

3.3 Data and methodology

In this study, three datasets are used for the clustering analysis of customers' life patterns: meter readings and survey results from the Customer Behaviour Trials (CBT) conducted by the Commission for Energy Regulation (CER) in Ireland, and hourly weather data collected by the Irish Meteorology Office. We begin with an overview of the data sources followed by a discussion of the data pre-processing needed. Finally, the algorithms and the performance measures used are presented.

3.3.1 Data preparation

The metering data from the CBT contains 15-minute consumption data from over 4,000 respondents during the period from July 2009 to December 2010. Since around 1,000 commercial customers participated, we only selected the sample of 3000 which are defined as residential customers.

Before the cluster analysis for load profiles, higher-resolution data (e.g. quarter-hourly) is often aggregated into hourly consumption profiles of 24 hours as part of data pre-processing. However, in this study, due to the different objectives, the data cleaning process is different than in most previous studies. The weather sensitivities of electricity consumption in households are not real-time responses but involve lags. Therefore, the quarter-hourly data is aggregated into larger chunks of time to accommodate the lag effects. Another consideration is the effect of various time-of-use (TOU) tariffs during the trial. The periods of the tariffs are: off-peak (8:00-17:00 and 19:00-23:00 weekdays and 17:00-19:00 weekends and bank holidays), peak (17:00-19:00 Monday to Friday, excluding bank holidays), and super off-peak (23:00-8:00). To control for the effects of the tariffs, all data aggregation should be within the period division with the same TOUs. Apart from the two-hour peak period, the off-peak and super-off-peak periods are much longer, which might hide some weather effects in specific sub-periods.

For example, the demand response at lunch time might be clearer or stronger if the lunchtime from 12.00-14.00 were to be analysed separately. Otherwise, the demand change would seem minimal during the longer 9 hour off-peak period from 8.00 to 17.00. As a result, subdivisions of these two periods are created by considering the time use of the households during different periods of a day: Morning (6:00-8:00), Day_1 (8:00-10:00), Day_2 (10:00-12:00), Day_3 (12:00-15:00), Day_4(15:00-17:00), Peak (17:00-19:00), Evening_1 (19:00-21:00), Evening_2 (21:00-23:00) and Night (23:00-3:00). We exclude the period 3:00-6:00 from the analysis for two reasons: 1) From the Irish time use survey, it can be seen that over 99% of the households are sleeping after 2:00. 2) The demand response is minimal in that period. The metering data are then aggregated accordingly and transformed to indexes for the clustering.

Since the location of each household is not provided for confidentiality reasons, it is impossible to match exact local weather data with the households. The half-hourly weather data at Dublin airport from the Irish Meteorology Office is used because the participants were concentrated around Dublin according to the CER report (CER, 2012). Moreover, Ireland is a relatively small country and the weather variations across the country are limited (Ben Taieb et al., 2016). We assume that the weather at Dublin airport is sufficient to be representative for the weather elsewhere in the country at any given time. A weighted-average approach was also explored using several Irish weather stations but the differences from the Dublin-only approach were relatively minor (the weather data from the two datasets was compared by the t-test and the results is shown in Appendix B.1). Three weather variables in the downloaded dataset, temperature, precipitation, and sun duration, are selected for the weather sensitivity estimations. As discussed in the literature review, the impacts of other weather variables, such as wind speed and humidity, might be negligible for which are not included in the analysis. The weather data is then aggregated and prepared for the clustering model.

3.3.2 Clustering input

To identify how people respond differently to weather variations we use a novel index to measure electricity demand changes under different weather conditions, including temperature, rainfall, and sun duration. Considering the behavioural response to weather may vary seasonally and on different days of week, we explore four combinations for each weather variable: summer workdays (SW), summer rest-days/weekends (SR), winter workdays (WW), and winter rest-days/weekends (WR). To ensure seasonality is as representative as possible, we define summer as from May to August and the winter from December to February.

In order to ensure a wide-enough fluctuation in weather variables while maintaining a relatively robust sample size, we selected the top 20% and bottom 20% of days in each period for all weather variables. It should be noted that in choosing days of two ends of temperatures, we excluded the days with medium or heavy rain before the day selection to control the precipitation effect. In addition, Christmas and New Year holidays are not included in the selection pools for weather variations in the winter rest-day scenarios, since it is expected that the behavioural sensitivity to weather would be completely different than usual weekends. The statistical summaries of the three weather variables are shown as below in Table 2.2.

The new index for clustering is defined as follows and is calculated for each scenario separately:

$$I_{w,p,i} = \frac{\overline{E}_{w,p,i,low} - \overline{E}_{w,p,i,high}}{\overline{E}_{w,p,i,high}} \times 100\% \tag{3.1}$$

$I_{w,p,i}$ denotes the revised demand change index for the i_{th} household for the weather variable w in period p. It shows by what percentage that energy demand changes towards the weather changes in the p_{th} period. $\overline{E}_{w,p,i,low}$ indicates the average electricity demand for household i in the p_{th} period on days with the bottom 20% of values for weather variable w. For example, for temperature in the morning period, we first selected the bottom 20% of the days in terms of temperature value and then calculated the average morning demand for those selected days based on household consumption. Similarly, $\overline{E}_{w,p,i,high}$ represents the demand of the p_{th} period for the top 20% of days for weather variable w. Therefore, the vector $C_{w,i}(I_{w,p_1,i}, I_{w,p_1,i}, \ldots, I_{w,n,i})$ consists of the weather sensitivities of household i of periods of day in certain scenario and the vectors are then directly used as inputs for the household clustering.

3.3.3 Algorithms and performance measures

The aim of cluster analysis is to identify weather sensitivity patterns. The sensitivity of the three weather variables can be regarded as different proxies for households' daily patterns:

1. The demand response to temperature may indicate the seasonality of activities during a certain period
2. The sensitivity to rainfall may imply that regular outdoor activities occur in that period or that the household is used to going out during that period
3. The sun sensitivity can be seen as an indicator of whether the household has spare time and to what extent their behaviour or activities are sensitive to sunshine over a given period

The weather sensitivity profile of a day for each household is obtained and consists of nine coefficients for each period of a day. The cluster analysis generates representative pattern curves for each weather variable.

The literature review shows that K-means is among the most widely used techniques for analysing load profiles. K-means have significant advantages in terms of being simpler and demanding less computational capacity. In addition, cluster analysis using K-means on index-based clustering results have been widely discussed in the past studies and have proved efficient. Hierarchical algorithms can be helpful to determine cluster numbers and also as cross-validation.

There are two important issues which must be addressed before clustering: how to decide on the number of clusters for the algorithms, and the effectiveness of the data partitioning. Performance can be measured using different clustering validity indicators but the indices used in previous studies vary. Moreover, Chicco (2012) finds that no single measure consistently prevails over the others. Therefore, most previous research into electricity consumption clustering adopts at least two different indices in order to address concerns over robustness and obtain a reliable and valid result (Yang and Sun, 2013; Räsänen et al., 2010; Ramos et al., 2015). We use two indicators, the Silhouette score and Davies–Bouldin index (DBI), which are widely used in electricity demand clustering studies, to assist in the selection. Silhouette score is defined by the mean intra-cluster distance and the mean nearest-cluster distance for each observation. A higher Silhouette score means clusters are farther apart and less dispersed, while values near 0 indicate overlapping clusters. The DBI score measures the average similarity of each cluster with its most similar cluster. The similarity is calculated using the ratio of within-cluster distances to between-cluster distances. A lower value indicates a better clustering.

Considering the load profile clusterings in the literature (Gouveia and Seixas, 2015), a series of numbers of clusters from 5 to 15 are examined. It should be noted that no absolute optimal number exists in the cluster analysis. The choice of the final number of clusters is based on the indicators and the practical experience.

3.3.4 Statistical inferences

After the representative weather sensitivity curves are created by the clustering algorithms, we want to investigate the relationships between household background variables and clustering of weather sensitivity patterns from two perspectives: 1) whether the household features could affect the clustering in each scenario, in other words, whether the clusters are independent of each household feature; and 2) whether certain dominated profiles are correlated with a particular cluster of weather sensitivity or daily behaviour patterns. To answer the first question, Chi-square tests of independence are employed to identify whether in a certain group the distribution/structure of one social variable is different from each other. On the second question, Chi-square goodness of fit test is adopted to examine whether the clusters statistically differ from that of the population as a whole. Effect sizes are calculated for both to compare which variable potentially has more effect on the sensitivity pattern segmentations. The demographic variables we are interested in are gender, age, employment status, social class, whether they live with other people, how many adults/children are in the household, education, and income (see Table 3.1).

3.4 Results and Discussion

In this section, we start by comparing algorithms and selecting a suitable number of clusters for each weather sensitivity analysis. The results for sunshine duration, rainfall and temperature are discussed separately and followed by the results of the statistical tests for the relationships between demographics and weather. With the assistance of DBI and Silhouette analysis, we chose seven as the cluster number for the workday scenarios for sun as well as all the scenarios for temperature, while six was the optimal number for the sun and rain weekend scenarios. In order to stay focused on the results here, the DBI and Silhouette results are included in the Appendix B.2 to B.5.

3.4.1 Clustering results

In each sub-section, the weather sensitivity patterns for both workdays and weekends are briefly described. In the legends, the first number describes the cluster number, the second is how many households are categorised into that group and the third reflects the ratio of the households in that group to the whole sample.

Sun

The usage change in the Figure 3.1 represents the percentage change in electricity demand changes from the bottom 20% to the top 20% of sunny days. If the number is positive, it indicates that households use less electricity on a sunny days. As discussed under Methodology, the sun duration sensitivities represent the availability of discretionary time in certain periods. The sensitivity curves therefore can be seen as indicators of the extent to which households are able to allocate their time freely through a day. The sensitivity patterns are presented in Figure 3.1. Given the shortened days, in Winter (December to February), there is no direct sunlight for any periods after the Peak. Therefore, the sensitivity curves for sun duration only include periods from Morning through Day_4 in the winter scenarios.

In general, for all four scenarios, afternoons (Day_3 and Day_4) are more responsive to sunlight than mornings, indicating that people tend to have more discretionary time during the afternoon. In terms of the seasonal difference between workdays and weekends, mornings in summer are more sensitive to sun duration changes, while households are more responsive during afternoons in winter than in summer. The only exception is the morning period where responses are drastically stronger in winter. It could be explained by people tending to get up earlier on sunny days, since most of winter mornings would be dark and sunny mornings might wake people up earlier. In summer, especially on weekends, households respond even more dramatically in the evenings. In terms of winter, the trends are generally similar to the shapes in summer, although the responsive curves in the mornings of summer are more diverse. The possible explanation could be that people have less variety of outdoor/indoor activities in winter in general. For example, even if the weather conditions are good, households would be relatively unlikely to plan a picnic for weekend mornings in winter.

From the distribution/number of households in the clusters, we found that the segmentation is more dense and less balanced when dividing up into clusters on weekends. Group 0, the largest and least sensitive group, accounts for around 40%

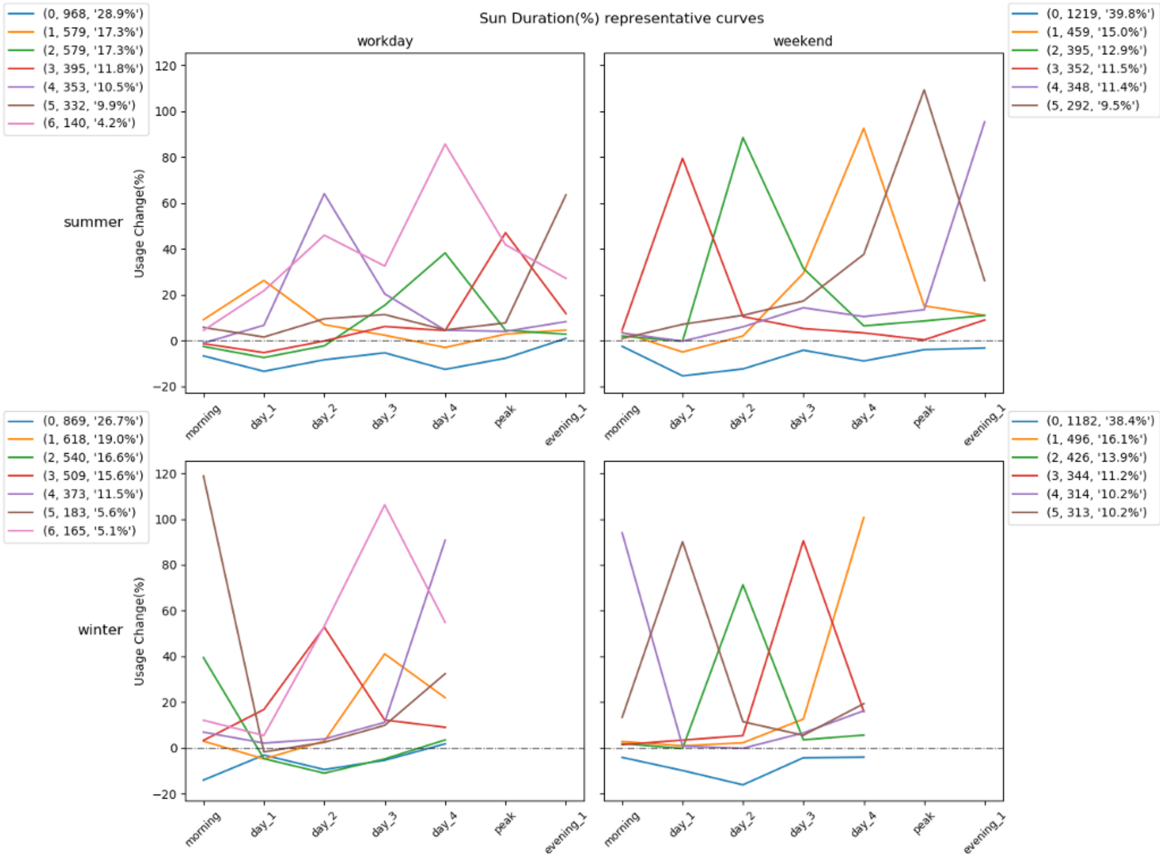

Fig. 3.1 Sun duration sensitivity

of households in the clusterings for both summer and winter. However, Group 0 on weekends is still less flat and more sensitive than its counterpart group on workdays, especially for the mid-day periods (Day_2 to Day_4). It could be caused by less flexibility on workdays for employed households. Although in both summer and winter the largest group gives a negative score, which indicates the households use more electricity during a sunny day, the reason or behaviour behind it could be different. For winter, the increase in energy demand could be driven by the sun-related indoor activities, such as laundry or car washing or gardening. Because sunny days are rarer in winter people would take advantage of the weather to plan for weather-related activities. For summer with plenty of sunlight and better weather, the increase is likely to be from hosting parties especially on weekends or enjoying sunlight at home, rather than rushing to arrange the chores because of a good weather. And it should also be noted that even as the largest group, it still only accounts for 30% of the whole sample and the majority is positively sensitive to sunlight and would prefer outdoor activities in a sunny day.

Rainfall

From the negative response to rainfall one can infer whether a period is normally occupied by outdoor activities. The positive response reflects households using more electricity during the bottom 20% of rainy days (normally non-raining days). Thus, one can identify whether the period is typically used for rain-sensitive activities. Figure 3.2 shows that on summer workdays there is no single preferred outdoor period for all groups. The preferences are more evenly spread throughout the daytime, although there are slightly more groups affected during the before and after lunchtime periods (Day_2 and Day_4). People in winter workdays are clearly more responsive to rain. And the majority of groups, except Groups 4 and 5, prefer to go out during the midday periods of 10:00-15:00. One possible reason to prefer the mid-day periods may be that for stay-at-home family members those periods are more flexible and less likely to be occupied by fixed house-bound activities, such as picking up children and preparing meals for families. The sensitivity decreases as it is getting late and it could be argued that as evening approaches, more indoor activities/house chores take place and households are less likely to choose these later periods to go out.

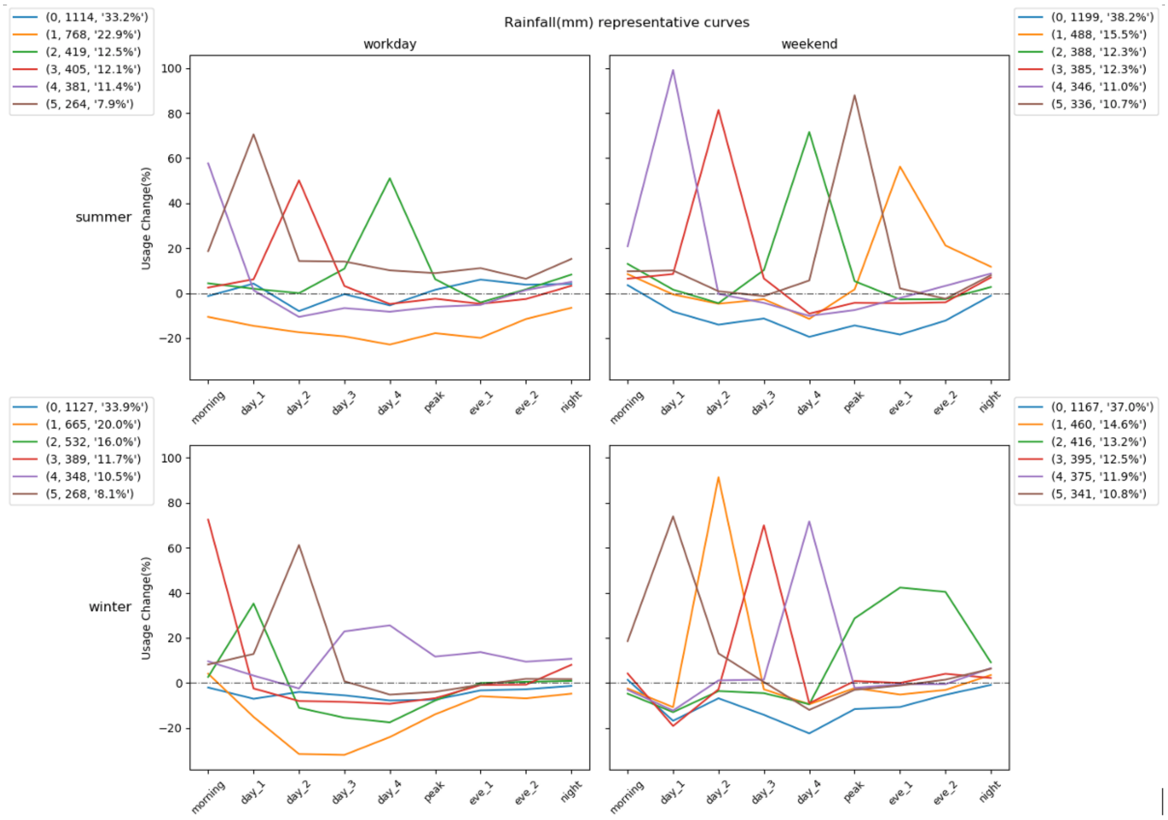

Fig. 3.2 Rain sensitivity

In terms of seasonal differences, it can be seen that all groups in the mornings of winter workdays have non-negative responses, while the changes in summer varies. One possible reason could be that people are more likely to be affected by rain on winter workdays and tend to leave home earlier to avoid possible traffic jams. However, on winter weekends almost all groups, apart from Group 5, display slightly non-positive changes. This could be caused by those who typically have outdoor morning plans, such as jogging, where rainfall interrupt their schedules, although the responses are still minimal compared to Day_1. The magnitude of the sensitivity demonstrates that Day_1 is the period most affected, which indicates a preference of going out during mornings on winter weekends. It is clear from the Figure that workdays evenings in summer are slightly more likely to be affected. This may be because that people usually tend not to go out on workday evenings, especially in winter, and rain would be less disruptive in the winter evenings. The greater fluctuation in responses on summer weekends could reflect the fact that more outdoor activities are planned at those times and people would be more willing to go out than winter. Relatively bad weather in winter is expected and households would not easily alter their behaviour due to rainy weather.

To explore the differences between workdays and weekends, we examined the household distributions into clusters in all the scenarios. One finding is that, regardless of season, people are much more densely segregated into one group on weekends. With over 37% of households clustered into one group, which is also the most sensitive group, it could reflect the fact that most families would generally spend some time outdoors during weekends, since weekends are more flexible than workdays. However, it can be seen that there is no obvious peak/preferred period for the majority group since the curve is smoother than its counterpart in winter. On the other hand, Day_2 and Day_4 are preferred by many households in the largest group on winter weekends. The more significant positive responses at different periods of time could mean that many households may prefer specific time periods for outdoor activities on weekends, especially rain-sensitive work, for instance, washing cars or gardening.

Temperature

In general, response on workdays are less sharp than on weekends due to more limited flexibility (as can be seen in Figure 3.3). The sensitivities in winter are more vibrant than those in summer. The differences among clusters in winter is much bigger – for example, the difference between the top 20% and the bottom 20% of day in summer is much smaller than in winter. The temperature difference is around 3-4 ºC in every period in summer versus 7-8 ºC in winter. As an island in the North Atlantic with mild summers and moderate winters, the maximum summer temperature in Ireland is only above 23 ºC, while the minimum in winter is -8ºC. It might be imagined that such a narrow range of summer temperature fluctuations would result in fairly limited behaviour changes moderating any swings in electricity demand. As expected, temperature response curves in winter are much more stronger than in summer. The majority group (G0) in summer show a relatively flat response. Meanwhile, the household distribution in the clusters confims the hypothesis that individuals are less likely to be affected by temperature changes in summer, notably the largest and the least sensitive groups on both workday and weekends accounts for over 32% of all households. The situation in winter is more evenly spread and even the largest group is more responsive than in summer.

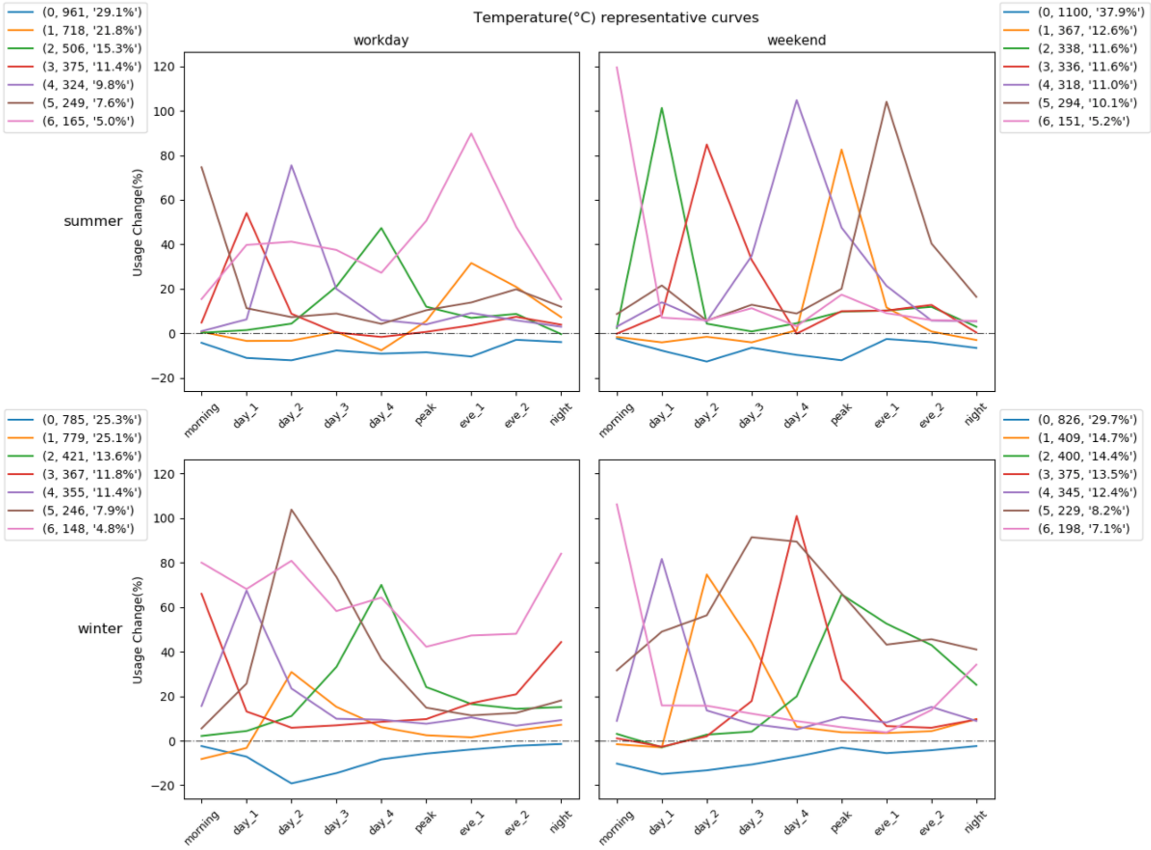

Fig. 3.3 Temperature sensitivity

Another seasonal difference is that whereas the peaks/the most sensitive periods in winter fall during midday periods in winter, the counterparts in summer do not see a clear trend and can occur at any periods throughout the day. Unlike summer, almost all groups are sensitive to temperature in winter, especially the periods from Day_3 onward. For winter workdays, Day_2 and Day_4 appear to be the most responsive periods where people tend to go out if it is not extremely cold; On the other hand, we see all groups respond to Evening_1 on summer workdays. In addition, the sensitivities of temperature show a non-increasing trend after Evening_1 (19:00-21:00) in summer, but the responses in winter reflect an opposite tendency of non-decreasing. This contrary result is even more non-considerable on weekends. The possible explanation could be that due to the limited change of temperature in summer, the chance that temperature affects peoples' outdoor plans at night could be minimal, compared to the effect in winter. People would be more willing to go out during a warm day in winter.

3.4.2 Statistical results

In this part, we focused on the questions at two levels: 1) In general, whether the clustering is associated with some of the selected socio-economic variables and whether the variables are more connected to a season or workdays/weekends. 2) At the cluster level, is there clear household profiles behind some groups?

3.4.3 Features reflecting weather sensitivity clustering

To answer the first question, we used chi-square tests of independence as well as effect sizes to identify the variables that affect the clustering.

The Chi-squared test of independence is used to determine whether a relationship exists between two nominal/categorical variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable. Here, we compare the frequency distribution of each social-economic variable for each cluster category separately. The observed frequencies are the total counts for each level of one variable at each level of the cluster category. The expected frequency counts are computed separately for each level of one categorical variable at each cluster[1]. The

[1] $E_{r,c} = \frac{n_r * n_c}{n}$, where $E_{r,c}$ is the expected frequency count for level r of a social-economic variable A and level c of Cluster C, n_r is the total number of sample observations at level r of Variable A , n_c is the total number of sample observations at level c of Cluster C, and n is the total sample size.

chi-square test is then performed based on the expected and observed frequencies[2]. For example, to investigate the relationship between education levels and clusters, the observed frequencies are the counts of each education level for each cluster. The expected frequencies are the calculated frequencies of each education level at each cluster based on the distribution of the total sample.

The list of variables we tested can be seen in Table 3.1. Table 3.2 shows the results that with p-value lower than 0.05 and the questions are ranked in descending order of effect size, which is indicated in parentheses.

Code	Variables
Q300	**Age**: 18-25, 25-35, 36-45, 46-55, 56-65, 65+, Retired
Q310	**Employment Status:** Employee, Self-employed, Unemployed, Retired, Carer: Looking after relative family
Q401	**Social Class:** AB, C1, C2, DE and below, F(Record all farmers)
Q402	**Income level:** Less than €15k,15k to 30k, 30k to 50k, 50k to 75k, 75k+
Q410	**Living status:** Live alone, All people over 15yrs, Both adults and Children
Q420	**How many people over 15?**: 1, 2, 3, 4+
Q430	**How many people over 15 in house during day time?** (If Q410 is not "Live alone") 1, 2, 3+
Q4312	**How many under 15 in house during day time?** (If Q410 is "Both adults and Child") 1, 2, 3+
Q5418	**Education level:** Primary and below, Secondary to Intermediate Cert, Secondary to Leaving Cert, Third level

Table 3.1 Code list

From a quick glance at the number of variables in each column in Table 3.2, it can be seen that in general more socio-economic features are associated with rain sensitivities. In other words, compared to other weather variables, the behaviour patterns affected by precipitation are more related to multiple household characteristics. In terms of seasonal differences, the clustering in the workday scenarios are associated with more household demographic variables during winter although a number of the significant variables overlap. By contrast, for rest-days there is no consistent pattern or many significant variables that repeat for different seasons.

[2]The test statistic is a chi-square random variable (χ^2) defined as: $\chi^2 = \sum \frac{(O_{r,c}-E_{r,c})^2}{E_{r,c}}$, where $O_{r,c}$ is the observed frequency count at level r of Variable A and level c of Cluster C, and $E_{r,c}$ is the expected frequency count at level r of Variable A and level c of Cluster C.

	Sun	Rain	Temp
SW **(Summer Workdays)**	**Q410**(0.085), **Q430**(0.066), **Q420**(0.056), Q310(0.055)	**Q410**(0.076), **Q420**(0.068), **Q430**(0.066), **Q300**(0.059), **Q310**(0.058)	**Q410**(0.076), Q310(0.067), **Q300**(0.064), **Q420**(0.057), Q5418(0.057)
SR **(Summer Rest-days)**	**Q410**(0.06), **Q420**(0.058)	Q310(0.059), Q401(0.057), **Q410**(0.056), **Q300**(0.051),	**Q410**(0.078), **Q300**(0.07), Q310(0.069), Q5418(0.061)
WW **(Winter Workdays)**	Q310(0.094), **Q430**(0.081), Q401(0.084), **Q300**(0.073), **Q410**(0.07), **Q420**(0.058)	**Q410**(0.072), Q402(0.07), **Q430**(0.063), **Q300**(0.061), **Q310**(0.058), Q401(0.055), **Q420**(0.055),	**Q410**(0.11), **Q420**(0.078), **Q300**(0.074), **Q310**(0.074), **Q430**(0.071)
WR **(Winter Rest-days)**	Q402(0.076), Q5418(0.063), **Q430**(0.061), Q401(0.056), **Q310**(0.053),	**Q300**(0.066), **Q410**(0.06), Q401(0.055), **Q420**(0.053)	**Q420**(0.064), **Q300**(0.061)

Table 3.2 Significant variables for each scenario

Regarding the specific variables, living status (Q410) is the most relevant variable to affect clustering and is statistically significant in 10 out of 12 scenarios (across all weather variables in summer and winter workdays). For sun sensitivity clustering, how many people in the household are over 15 years old (Q420), whether they are at home during daytime (Q430), and employment status (Q310) are the next three most commonly significant variables. For rain sensitivities, age (Q300) affects all scenarios, although Q310 and Q420 also often play roles. Likewise, age (Q300) and employment status(Q310) are significant for certain temperature profiles. However, the effect of living status dominates other variables with the highest effect sizes regardless of seasons or workdays/weekends.

The effects of income-related variables are not as significant as the occupancy-related variables in the clustering of weather sensitivity. Yet with lower differentiating power, it stills has a role in pattern segmentations. In general, social-class variables are least likely to affect temperature sensitivity: Education level (Q5418) is the only income-related variable that is linked with the temperature sensitivity clustering and even then only exists in summer scenarios. Rain patterns, on the other hand, are frequently associated with a number of variables of this kind, for example social class (Q401) and income level (Q402), especially Q401, which affects three out of four rainfall scenarios. In terms of sun clustering, it appears that only winter scenarios are related to income-related variables. However, it should be highlighted that the scenario of sun sensitivity in winter weekends is mostly associated with socio-economic variables, notably income, social class, and education level.

3.4.4 Clusters with clear background profiles

To investigate if any groups are dominated by specific demographic profiles of households, we look into the distributions of each variable to identify the differences in household characteristics between each group and the population using Chi-square tests. The variables we chose to test are the same as used in the last section and listed in Table 3.1. We selected the top two distinguished groups (the representative curves shown in in Figure 3.4) that have at least three profile variables statistically significant (p values less than 0.05) for the weather sensitivity clusterings. The meaning of the labels in the legends is as follows: SW for summer workdays, WW for winter workdays, and WR for winter weekends. The latter part starting with "g" indicates the group number in that scenario.

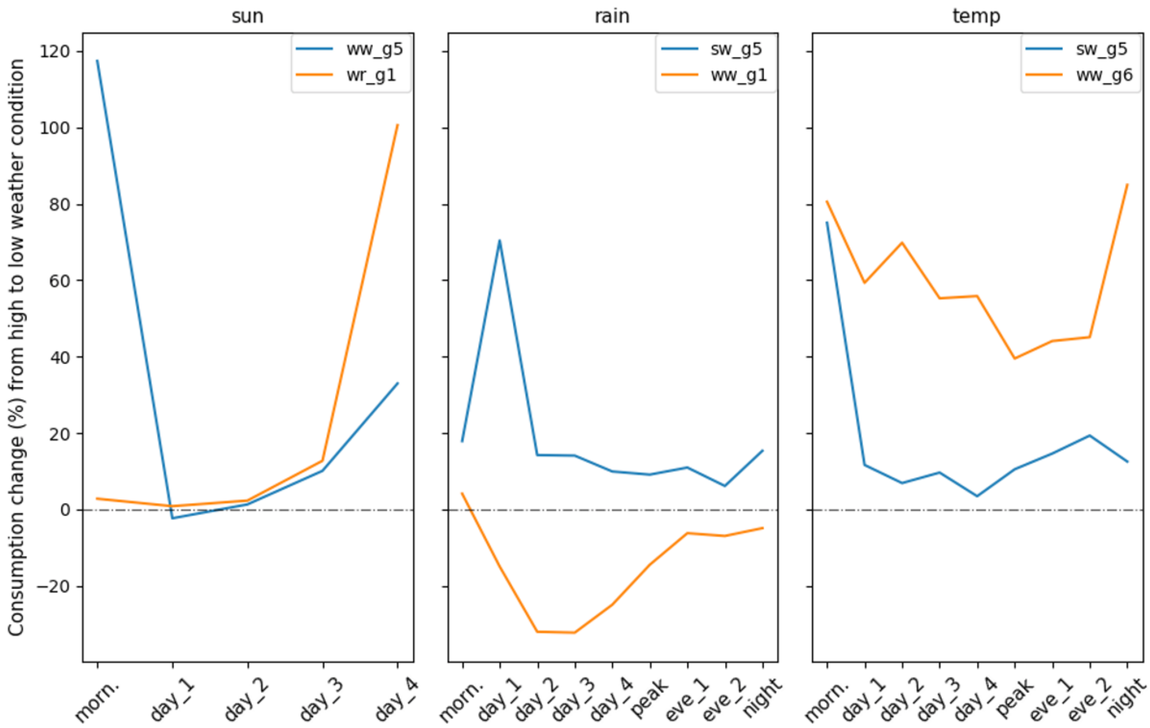

Fig. 3.4 Representativ curves for selected groups

Table 3.3 shows which questions are statistically significant for each group. We further examined how the group differs from the overall sample by comparing the distribution of the variables that are listed in Table 3.3. In terms of sun sensitivity, Group 5 on the winter workdays is notably sensitive to sunlight during early mornings. In looking at the demographic break-down, this group has the largest number of full-time workers. In addition, it includes the highest share among all groups of those belonging to higher managerial and professional class (AB) and supervisory and junior managerial (C1), as well as lowest percentage of those in the DE social class. The group also includes both moe 18-35 and 36-45 year olds and is also the group with largest percentage of households with no person staying in the house during the daytime. Meanwhile, Group 1 on the winter weekends seems more likely to have their entertainment/spare time during the later periods of the day. Indeed, G1 has the highest ratio of younger employed/self-employed individuals. In addition, it is also a relatively affluent group, since it includes the largest percent of people in AB and C1 social class, as well as the lowest in DE.

	Sun duration	Rainfall	Temperature
SW		G5: **Q430**(0.041); **Q420**(0.052); **Q410**(0.061); **Q310**(0.053); **Q300**(0.11)	G5: **Q5418**(0.042); **Q420**(0.061); **Q410**(0.069); **Q401**(0.032); **Q310**(0.043); **Q300**(0.063)
WW	G5: **Q430**(0.14); **Q401**(0.12); **Q310**(0.17); **Q300**(0.088)	G1: **Q5418**(0.01); **Q401**(0.011); **Q310**(0.013); Q402(0.012)	G6: **Q420**(0.18); **Q410**(0.21); **Q310**(0.16) **Q300**(0.23)
WR	G1: **Q401**(0.026); **Q310**(0.017); **Q300**(0.02)		

Table 3.3 Statistically significant variables

For the rainfall clustering, Group 5 is more likely to include employed young families or singles (within the 18-35 or 36-45 age categories) compared to other groups on summer workdays. The demographic analysis shows that the households include mainly those who live alone or live with children, but with the fewest number of families where all people are over 15 years old. Moreover, it has the highest percent of respondents where no one or at most one adult person remain in the house during the daytime. It also appears that this group would prefer to arrange their external activities during the mornings. On winter workdays, Group 1 have a wider and more sensitive response to rainfall, which could indicate more regular outdoor time for those households. This group is also more likely to have a higher educational degree and higher social class (AB and C1) with the fewest in the lowest social class (DE). In addition, it has the highest percentage of households that are employed or self-employed. Group 1 is most likely to include those with incomes above €50,000 and least likely to include those with incomes of less than €15,000.

There are also some interesting differences between groups for the temperature clustering. Group 5 in the summer workday scenarios appears to be most sensitive during the early mornings. This group is most likely to have the highest education level, a family structure with significantly lower possibility of living alone or consisting of less than 2 adults and a greater likelihood of children and adults living together. Looking to age and employment status, it is seen that the group has an extremely high ratio of middle-age people (36-45) and fewer older members, as well as a much larger proportion of members of households bing employed full time. By contrast, Group 6 is sensitive throughout the day on winter workdays. Households in this group have a greater possibility of including those living alone and in low income groups. The ratio of being older than 65 years old and retired is dramatically higher in Group 6 and it also shows the highest number belonging to the DE social class. The structure explained why the group could be sensitive to temperature changes all day, since they are mainly staying at home.

3.5 Conclusion

The introduction of smart meters has brought opportunities for both utilities and policymakers to understand residential electricity consumption in greater depth. Due to the extremely large volume of high-resolution data, machine learning techniques have been used to investigate the information buried in metering data. Most studies have focused on load management, especially for demand forecasting and customer load profile segmentation and most implementation of clustering algorithms has been applied directly to metering data. There have been, however, few studies using the techniques to study daily life patterns within households. We introduce a novel method to detect household behaviour/daily patterns using clustering algorithms applied to weather variables. Our analysis proposes using the weather sensitivities as proxies for the household daily life patterns, for instance, when a household tends to go out and at which periods of a day they have more spare time. The clusterings are not applied to meter readings but to the weather sensitivity coefficients. To reflect the differences in behaviour patterns between workdays and weekends in different seasons, the clusterings were conducted separately for seasons as well as for weekends versus workdays and for three weather variables – sun duration, temperature, and rainfall.

We are able to characterize clear differences in the daily patterns between workdays and weekends in summer and winter and how households respond to changing weather patterns. Based on the sun sensitivities, households are found to be less flexible and have less spare time during the middle of a workday while enjoying greater freedom during the afternoons. Households are more responsive to sun on summer weekends, which indicates greater discretionary time and outdoor activities during summer. The rain sensitivity profile curves tell us that stay-at-home family members tend to go out during late mornings and early afternoons regardless of season, probably due to having fewer fixed housework commitments such as cooking dinner and picking up children. Meanwhile, people are more likely to arrange outdoor activities in the evenings on summer weekends compared to workdays. The profiles which yield the fewest noteworthy differences are the temperature sensitivity curves. The statistical tests suggest that demographic features are most connected to rain sensitivities. In terms of seasonal differences, the clustering in workday scenarios reflect more about the household features in winter.

Loooking across all factors, the effect of social class in the clustering of weather sensitivity was not as significant as the occupancy-related variables. Living status, employment status, and the number of adults of the household are the main classifiers for all the clusterings. Among all weather variables, rain patterns are relatively more associated with variables of this kind, especially social class and income level.

This analysis could also serve as a starting point for classifying customers by their daily life patterns. Understanding during which periods individuals may prefer to be outside of the home and when they are more likely to have spare time or be more flexible in their behavior patterns could be important when designing customised electricity price schemes. This work and the methods presented herein could be the basis of a new prediction model to classify existing or new customers' behaviour patterns and responses to weather conditions.

Reference

Al-Wakeel, A., Wu, J. and Jenkins, N. (2017) 'k-means based load estimation of domestic smart meter measurements', Applied Energy. Elsevier, 194, pp. 333–342. doi: 10.1016/J.APENERGY.2016.06.046.

Alberini, A. and Filippini, M. (2011) 'Response of residential electricity demand to price: The effect of measurement error', Energy Economics, 33 (5), 889–895. doi: 10.1016/j.eneco.2011.03.009.

Albert, A. and Rajagopal, R. (2013) 'Building dynamic thermal profiles of energy consumption for individuals and neighborhoods', in Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013. pp. 723-728. doi: 10.1109/BigData.2013.6691644.

Beccali, M. et al. (2008) 'Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area', Renewable and Sustainable Energy Reviews. Pergamon, 12(8), pp. 2040–2065. doi: 10.1016/J.RSER.2007.04.010.

Beckel, C. et al. (2014) 'Revealing household characteristics from smart meter data', Energy. Pergamon, 78, pp. 397–410. doi: 10.1016/J.ENERGY.2014.10.025.

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

Bianco, V., Manca, O. and Nardini, S. (2009) 'Electricity consumption forecasting in Italy using linear regression models', Energy. Pergamon, 34(9), pp. 1413–1421. doi: 10.1016/J.ENERGY.2009.06.034.

Blázquez Gomez, L. M., Filippini, M. and Heimsch, F. (2013) 'Regional impact of changes in disposable income on Spanish electricity demand: A spatial econometric analysis', Energy Economics. North-Holland, 40, pp. S58–S66. doi: 10.1016/J.ENECO.2013.09.008.

Cao, G. and Wu, L. (2016) 'Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting', Energy. Pergamon, 115, pp. 734–745. doi: 10.1016/J.ENERGY.2016.09.065.

Chen, B.-J., Chang, M.-W. and Lin, C.-J. (2004) ‘Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001’, IEEE Transactions on Power Systems, 19(4), pp. 1821–1830. doi: 10.1109/TPWRS.2004.835679.

Chen, Y. et al. (2017) ‘Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings’, Applied Energy. 195, pp. 659–670. doi: 10.1016/J.APENERGY.2017.03.034.

Commission for Energy Regulation (CER). (2012). CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010 [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0012-00. www.ucd.ie/issda/CER-electricity

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.

Cramer, J. C. et al. (1984) ‘Structural-behavioral determinants of residential energy use: Summer electricity use in Davis’, Energy, 9(3), pp. 207–216. doi: 10.1016/0360-5442(84)90108-7.

Druckman, A. and Jackson, T. (2008) ‘Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model’, Energy Policy. 36(8), pp. 3177–3192. doi: 10.1016/J.ENPOL.2008.03.021.

Fan, S. and Hyndman, R. J. (2011) ‘The price elasticity of electricity demand in South Australia’, Energy Policy. 39(6), pp. 3709–3719. doi: 10.1016/J.ENPOL.2011.03.080.

Faruqui, A. and Sergici, S. (2010) ‘Household response to dynamic pricing of electricity: A survey of 15 experiments’, Journal of Regulatory Economics. doi: 10.1007/s11149-010-9127-y.

Faruqui, A. and Malko, J R. (1983). The residential demand for electricity by time-of-use: a survey of twelve experiments with peak load pricing. Energy. 8(10). pp.781-795.

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.

Ghofrani, M. et al. (2011) ‘Smart meter based short-term load forecasting for residential customers’, in 2011 North American Power Symposium. IEEE, pp. 1–5. doi: 10.1109/NAPS.2011.6025124.

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. Elsevier, 116, pp. 666–676. doi: 10.1016/j.enbuild.2016.01.043.

Hackett, B. and Lutzenhiser, L. (1991) ‘Social structures and economic conduct: Interpreting variations in household energy consumption’, Sociological Forum. Kluwer Academic Publishers-Plenum Publishers, 6(3), pp. 449–470. doi: 10.1007/BF01114472.

Haider, H. T., See, O. H. and Elmenreich, W. (2016) ‘A review of residential demand response of smart grid’, Renewable and Sustainable Energy Reviews. Pergamon, 59, pp. 166–178. doi: 10.1016/J.RSER.2016.01.016.

Henley, A. and Peirson, J. (1998) ‘Residential energy demand and the interaction of price and temperature: British experimental evidence’, Energy Economics. North-Holland, 20(2), pp. 157–171. doi: 10.1016/S0140-9883(97)00025-X.

Herter, K. (2007) ‘Residential implementation of critical-peak pricing of electricity’, Energy Policy, 35 (4), pp. 2121–2130. doi: 10.1016/j.enpol.2006.06.019.

Herter, K., McAuliffe, P. and Rosenfeld, A. (2007) ‘An exploratory analysis of California residential customer response to critical peak pricing of electricity’, Energy, 32(1), pp. 25-34. doi: 10.1016/j.energy.2006.01.014.

Hor, C. L., Watson, S. J. and Majithia, S. (2005) ‘Analyzing the impact of weather variables on monthly electricity demand’, IEEE Transactions on Power Systems, 20(4), pp. 2078-2085. doi: 10.1109/TPWRS.2005.857397.

Irish Social Science Archive. (2005). ‘Irish National Time Use Survey 2005’. Accessed from: https://www.ucd.ie/issda/data/irishnationaltimeusesurvey.

Jones, R. V., Fuertes, A. and Lomas, K. J. (2015) ‘The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings’, Renewable and Sustainable Energy Reviews. Pergamon, pp. 901–917. doi: 10.1016/j.rser.2014.11.084.

Karanfil, F. (2009) ‘How many times again will we examine the energy-income nexus using a limited range of traditional econometric tools?’, Energy Policy, 37(4), pp. 1191–1194. doi: 10.1016/j.enpol.2008.11.029.

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.

Kaza, N. (2010) 'Understanding the spectrum of residential energy consumption: A quantile regression approach', Energy Policy, 38(11), pp. 6574–6585.
doi: 10.1016/j.enpol.2010.06.028.

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.

McLoughlin, F., Duffy, A. and Conlon, M. (2015) 'A clustering approach to domestic electricity load profile characterisation using smart metering data', Applied Energy. 141, pp. 190–199. doi: 10.1016/J.APENERGY.2014.12.039.

Newsham, G. R. and Bowker, B. G. (2010) 'The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review', Energy Policy. 38(7), pp. 3289–3296. doi: 10.1016/J.ENPOL.2010.01.027.

Pardo, A., Meneu, V. and Valor, E. (2002) 'Temperature and seasonality influences on Spanish electricity load', Energy Economics. 24(1), pp. 55–70. doi: 10.1016/S0140-9883(01)00082-2.

Quilumba, F. L. et al. (2015) 'Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities', IEEE Transactions on Smart Grid, 6(2), pp. 911–918. doi: 10.1109/TSG.2014.2364233.

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), pp. 3538-3545. doi: 10.1016/j.apenergy.2010.05.015.

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.

Sanghvi, A. P. 1989. Flexible strategies for load/demand management using dynamic pricing. IEEE Transactions on Power Systems. 4(1), 83-93.

Sanquist, T. F. et al. (2012) 'Lifestyle factors in U.S. residential electricity consumption', Energy Policy. 42, pp. 354–364. doi: 10.1016/J.ENPOL.2011.11.092.

Sapankevych, N. and Sankar, R. (2009) 'Time series prediction using support vector machines: A survey', IEEE Computational Intelligence Magazine, 4(2), pp. 24-38. doi: 10.1109/MCI.2009.932254.

Silk, J. I. and Joutz, F. L. (1997) 'Short and long-run elasticities in US residential electricity demand: a co-integration approach', Energy Economics, 19(4), pp. 493–513. doi: 10.1016/S0140-9883(97)01027-X.

Ben Taieb, S. et al. (2016) 'Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression', IEEE Transactions on Smart Grid, 7(5), pp. 2448–2455. doi: 10.1109/TSG.2016.2527820.

Valor, E., Meneu, V. and Caselles, V. (2001) 'Daily Air Temperature and Electricity Load in Spain', Journal of Applied Meteorology, 40(8), pp. 1413-1421. doi: 10.1175/1520-0450(2001)040<1413:DATAEL>2.0.CO;2.

Viegas, J. L. et al. (2015) 'Electricity demand profile prediction based on household characteristics', in International Conference on the European Energy Market, EEM. August. doi: 10.1109/EEM.2015.7216746.

Wangpattarapong, K. et al. (2008) 'The impacts of climatic and economic factors on residential electricity consumption of Bangkok Metropolis', Energy and Buildings, 40(8), pp. 1419–1425. doi: 10.1016/j.enbuild.2008.01.006.

Weiss, M. et al. (2012) 'Leveraging smart meter data to recognize home appliances', in 2012 IEEE International Conference on Pervasive Computing and Communications, PerCom 2012, pp. 190–197. doi: 10.1109/PerCom.2012.6199866.

Wijaya, T. K. et al. (2015) 'Cluster-based aggregate forecasting for residential electricity demand using smart meter data', in 2015 IEEE International Conference on Big Data (Big Data). IEEE, pp. 879–887. doi: 10.1109/BigData.2015.7363836.

Wooldridge, J. M. (2013) Introductory Econometrics, Cengage Learning. doi: 10.1016/j.jconhyd.2010.08.009.

Chapter 4

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

4.1 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. 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 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, 2016a). 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, 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.

4.2 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. 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. (2019) 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).

4.3 Data and methodology

4.3.1 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 resolution (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.

4.3.2 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 intra-day 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).

4.3.3 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 busineses 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 consumptions using the following formula:

$$D_{i,j} = \frac{U_i - \overline{U}_w}{\overline{U}_w} \times 100\% \quad \text{and} \quad \overline{U}_w = \sum_{i=1}^{5} U_i \tag{4.1}$$

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:

$$F_{i,j} = \frac{F_i - \overline{F}_h}{\overline{F}_h} \times 100\% \quad \text{and} \quad \overline{F}_h = \sum_{i=1}^{21} F_i \tag{4.2}$$

    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 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.

4.3.4 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 approaches 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.

4.4 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 C.1 to C.7.

4.4.1 Seasonal workday/weekend profiles

The profiles shown in Figure 4.1 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.

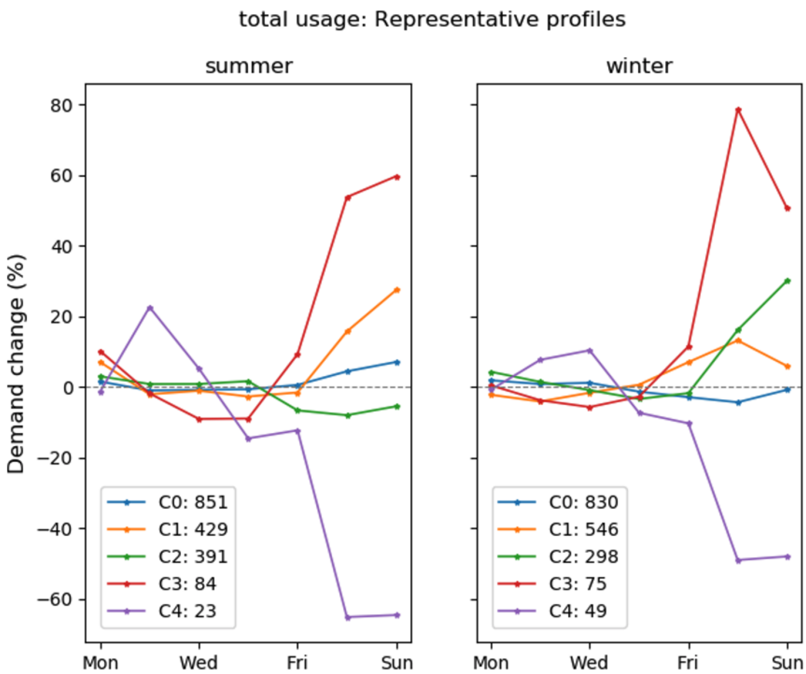

Fig. 4.1 Consumption patterns of total usage

Off-peak (11:00-19:00) usage patterns are shown in Figure 4.2. 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. 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.

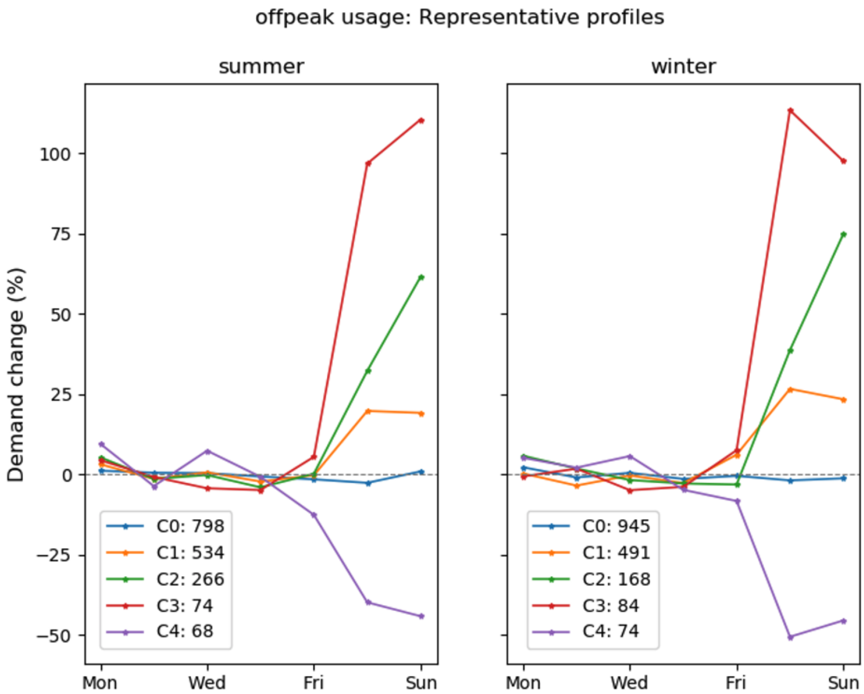

Fig. 4.2 Consumption patterns for off-peak (11:00-19:00)

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 4.3). 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 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).

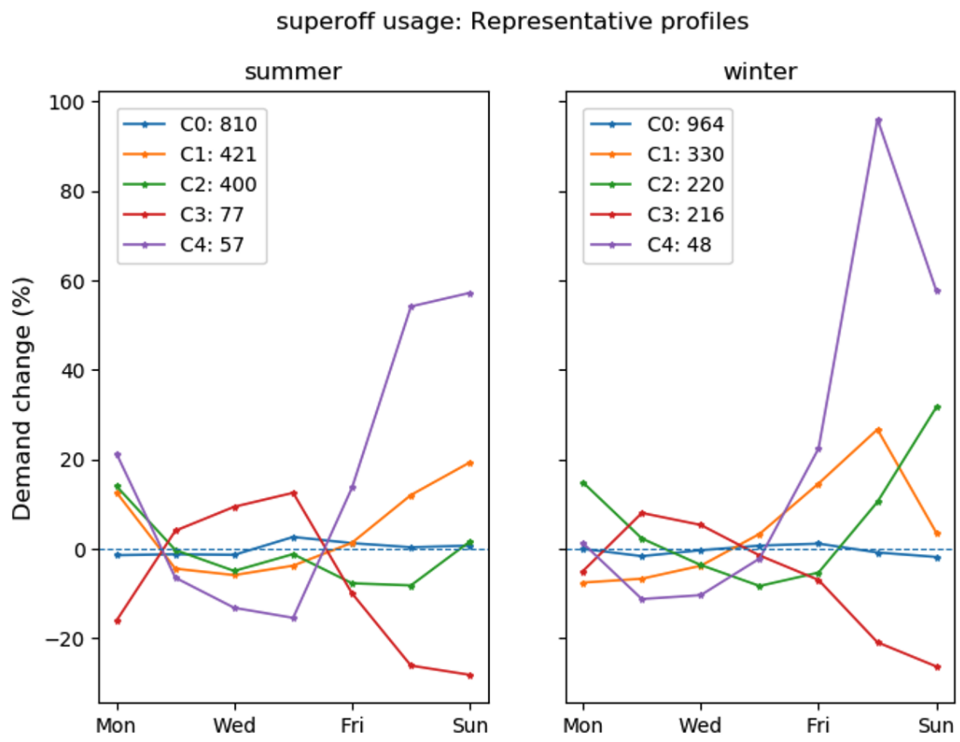

Fig. 4.3 Consumption patterns for Super off-peak(23:00-7:00)

The peak time profiles resemble the total consumption patterns (Figure 4.4). 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 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.

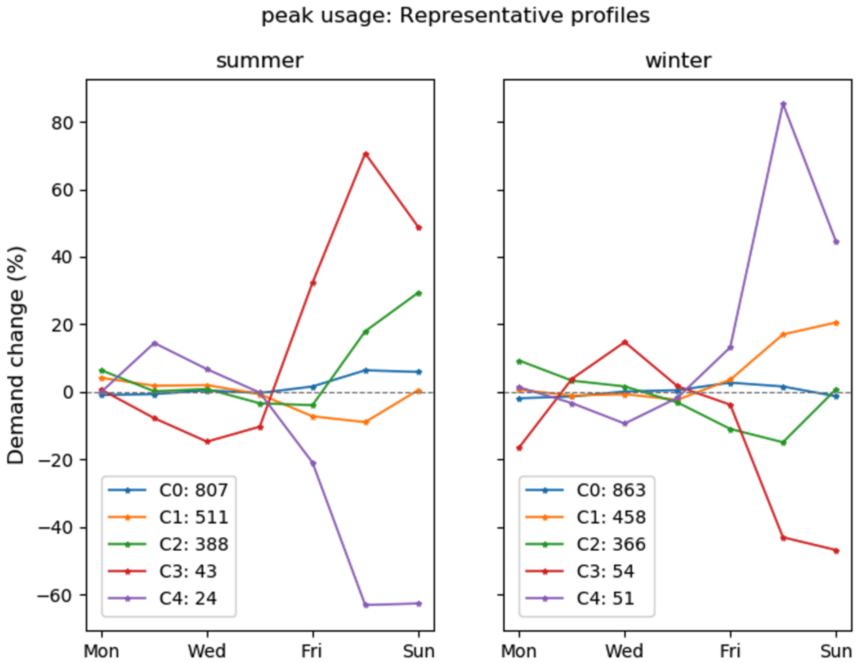

Fig. 4.4 Consumption patterns for Peak (7:00-11:00 and 19:00-23:00)

4.4.2 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 costumery to sit in front of the television and watch the Spring Festival Gala and TV programs with families. From Figure 4.5 and Figure 4.6 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

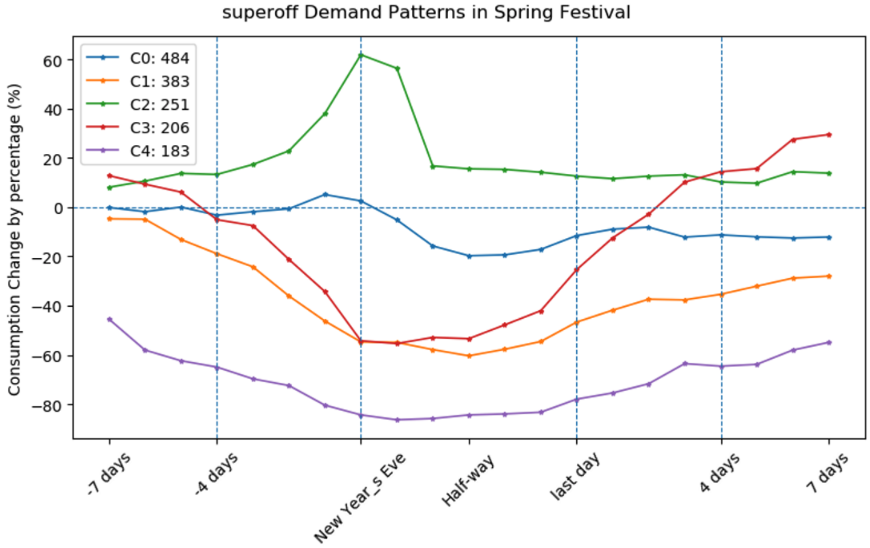

Fig. 4.5 Super off-peak demand for Spring Festival

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 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 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).

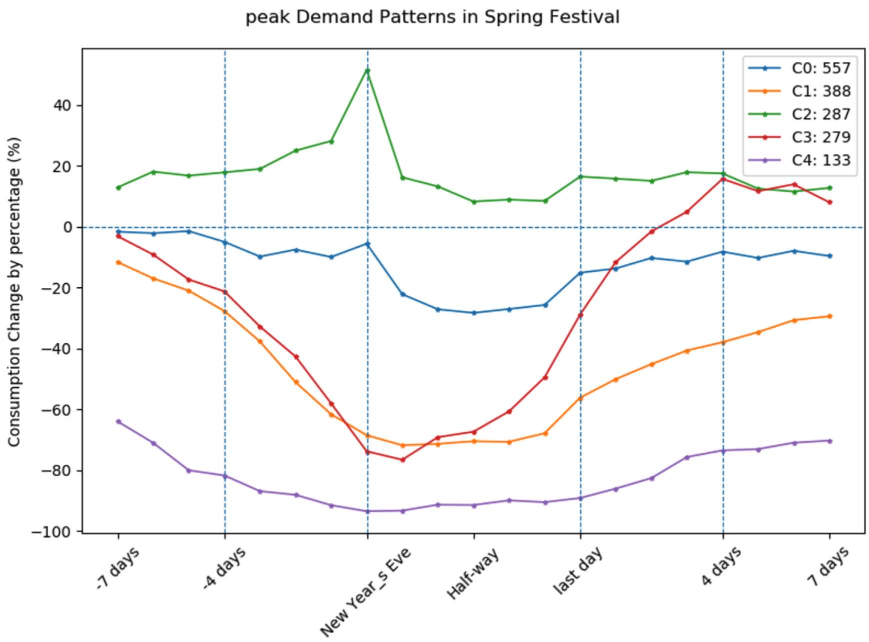

Fig. 4.6 Peak demand for Spring Festival

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 to the Spring Festival. One of the reasons behind the lower

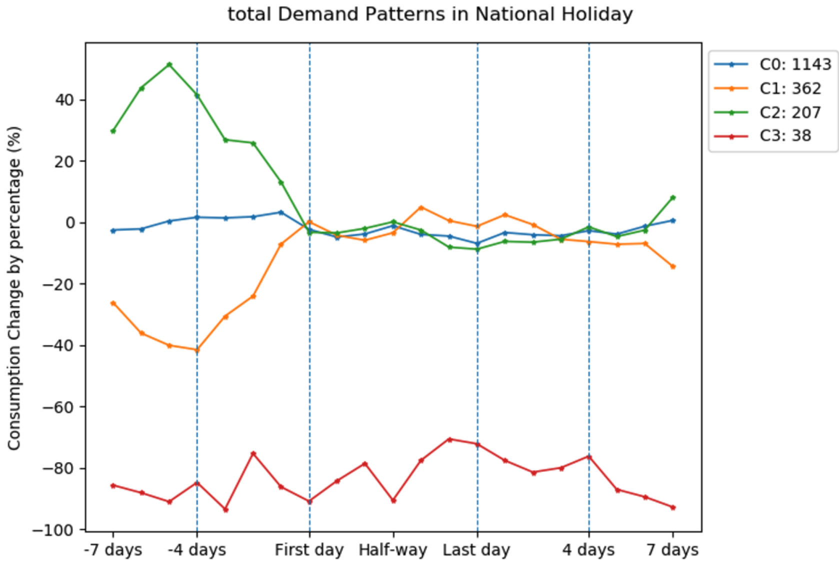

Fig. 4.7 Total consumption profiles in National Holidays

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 4.7, 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.

In order to identify the daytime activities during the holidays, we examined consumption patterns in the off-peak period (Figure 4.8). 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.

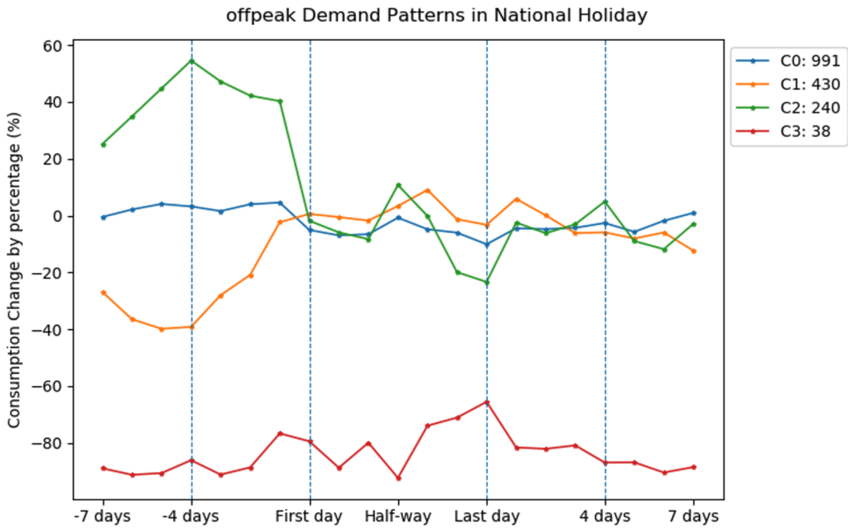

Fig. 4.8 Offpeak consumption profiles in National Holidays

4.4.3 Extreme weather profiles

Figure 4.9 and Figure 4.10 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).

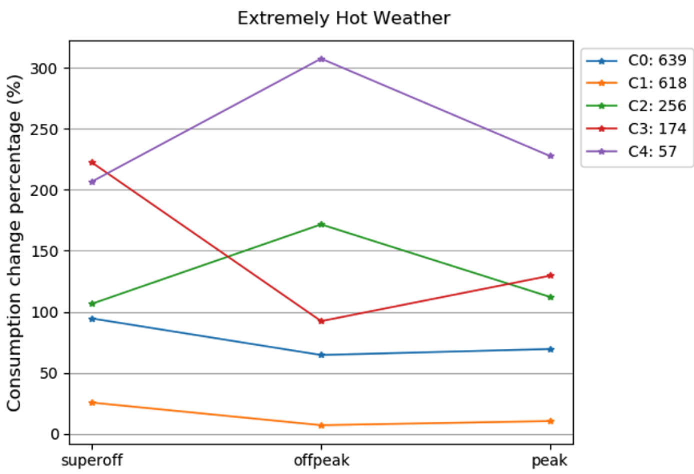

Fig. 4.9 Extreme hot days

In general, all customers in summer increase their consumptions rapidly, although to a different degree (Figure 4.9). During extremely hot weather, the most affected period is the night-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.

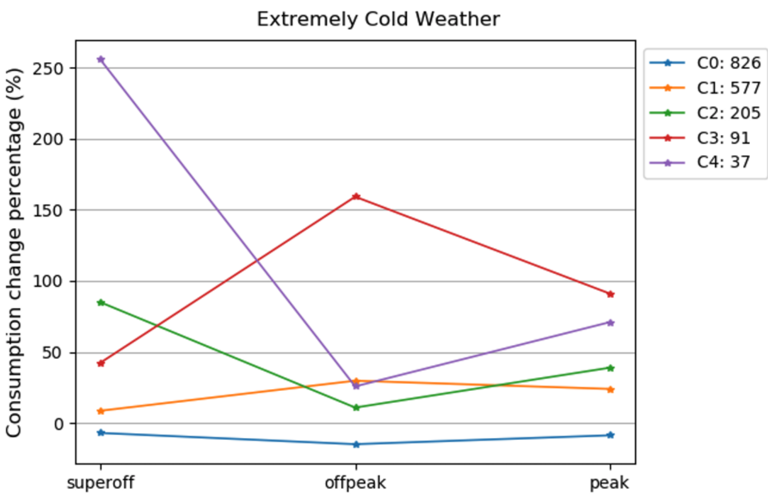

Fig. 4.10 Extreme cold days

The consumption patterns in winter differ significantly from the summer profiles (Figure 4.10). 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 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.

4.5 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 assistance 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 night-time, 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 algorithms 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.

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. (2016a) '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.

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 household'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. 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. doi: 10.1016/j.apenergy.2010.05.015.

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.

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.

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. Elsevier, 61, pp. 394–402. doi: 10.1016/J.ENPOL.2013.06.092.

Zhang T, Zhang G, Lu J, Feng X, Yang W. A new index and classification approach for load pattern analysis of large electricity customers. IEEE Trans Power Syst 2012;27(1):153e60.

Chapter 5

Conclusion

Electricity demand and weather have always been closely tied together but the widespread deployment of smart meters finally begins to allow for a detailed exploration of the relationship between the two. In particular, the thesis focuses on residential electricity consumption and weather. As one of the main drivers of electricity demand, residential consumption has long been at the centre of demand-side studies (Albadi and El-Saadany, 2008; Haider et.al, 2016). Many past studies have examined the influence of weather conditions on residential electricity demand (Torriti, 2014; Alberini, A. et al., 2019). However, using high-resolution data to better understand weather-related impacts on changing household consumption behavioural patterns has received limited attention. In Chapter 1, we provided a comprehensive overview of the literature including the history and current status of smart meter development, which has enabled and led to an explosion of studies focused on the residential sector. The review revealed several gaps in the current literature and so the following chapters have tried to address some of those gaps.

Over the past decades, econometric techniques have been preferred to study residential demand at a macro level. In such analyses, climatic variables were included purely to improve the model (Cialani and Mortazavi, 2018; Dilaver and Hunt, 2011). On the other hand, in any micro studies, weather attributes were usually completely neglected (Iwafune and Yagita, 2016; Kavousian et al., 2013). Instead, the focus was almost exclusively on household-level characteristics, such as socio-economic or building variables. Notably, panel data analysis for weather impacts on residential demand has been missing in the jigsaw.

Unlike most studies using longitudinal datasets to study the effects of socio-economic factors on energy consumption, in Chapter 2 a high-resolution panel dataset (obtained from CER, the Irish electricity regulator) was used to examine the weather effects on household consumers at different periods in time. In order to control for the endogenous and time-invariant variables at household level, we used Fixed-Effects models to differentiate socio-economic factors (education, income, etc.), building characteristics (floor size, bedroom number, etc.), and electric appliance ownership (TVs, kettles, etc.). We demonstrated that in general, rain and sunshine duration have a greater potential to affect people's behaviour and daily routines, while temperature has robust and relatively small impacts.

The findings from Chapter 2 have the potential to contribute towards not only greater appreciation of residential consumption behaviour, but can inform the operational strategies of utilities. Firstly, our study provides a comprehensive picture of how weather conditions interact with household occupancy and lifestyle patterns in different scenarios. Secondly, gaining greater understanding of consumer behaviour patterns without needing to employ relatively intrusive and costly approaches like surveys could be very attractive to, for example, utilities who could use such an approach as a preliminary tool before conducting smaller, more targeted surveys for certain value-added services. One shortcoming is the limited longitudinal data – the dataset only contains one-year of consumption data and the weather variation in Ireland is generally small. With data from more than one year, the models would be more robust. In addition, additional studies could be conducted on how people would respond to specific weather scenarios, such as extreme cold/hot weather. With more and more extreme weather events occurring, the findings for different weather conditions would be insightful for both researchers and policy makers to understand the effect of climate change on residential electricity consumption.

The traditional approach of using econometric tools would be better suited to providing preliminary results on the impacts of weather in general, however, it cannot be used to unravel differences in consumption patterns. Over the last decade as more and more jurisdictions have deployed smart meters at scale, the analysis of residential demand has greatly improved, allowing for the possibility of focusing on differences in consumption patterns for households, rather than only at the grid level. However, customer characterisation has not been explored extensively for load management or load forecasting, due to the dependence on the availability of survey information. Instead, most work on customer profiling has been based on the relationship between electricity demand and household socio-economic background. Furthermore, even less research has been carried out on clustering residential customers by their weather sensitivities. In Chapter 3 and 4 we contributed to the academic literature by conducting clustering techniques to explore the issue from different perspectives.

In Chapter 3, we used the same dataset as in Chapter 2, the CER trial data. Unlike the econometric models employed in Chapter 2, we introduced clustering algorithms to detect household behaviour/daily patterns using weather variables. We adopted K-means for the task of weather sensitivity clustering due to its proven capability of handling high-resolution household consumption data. We proposed a novel method of using the weather sensitivities as proxies to identify the daily patterns in the household. Apart from profiling customers using different weather variables, we analysed the profiles from two additional perspectives: seasonality and workday/weekend differences. In addition, the correlations between weather sensitivity clusters and socio-economic attributes were also examined using statistical tests. One main finding was that living status (i.e., whether living alone or with only adults or with young children), employment status, and the number of adults in the household are the main variables that can help explain the differences in the consumption patterns.

The major contribution of Chapter 3 is the novel approach that enables utilities or other researchers to understand the consumption patterns from a brand new angle by weather responses and it could be used to reveal the different patterns of occupancy based on the weather sensitivity clusters. Having a better grasp of the main types of residential customer occupancy patterns could serve as the basis for adopting various demand-side management strategies, from better peak control to electricity price design.. Traditionally, occupancy patterns were detected either by small-scale trials with higher resolution/electrical appliance level data or surveys. Examining weather sensitivity groups offers a quicker way of identifying occupancy and behavioural patterns, and so can deliver insights in a much less time-consuming and much more cost-efficient manner, for example, if a utility wanted to gain a general picture of its customers. In terms of limitations, two main aspects should be pointed out. Most importantly, the relatively short time series affects the robustness of the clustering. Secondly, the relationship between the socio-demographic profiles and weather sensitivities were not significant. There could be two reasons: 1) the unbalanced dataset is not representative and has too many respondents from certain socio-economic profiles, such as retirees, but also far too few high-income and younger consumers; and 2) Irish weather is generally mild and so there are only moderate weather fluctuations both on a daily and seasonal basis, and so would not be expected to have a large impact on people's behaviour compared to countries or regions with greater variation in temperature and other weather variables.

In Chapter 4, we examined a different location with more extreme weather conditions, Chengdu, the capital of Sichuan in southwestern China. The data we used included smart meter recordings from 2,000 households over a much longer period of 4 years, compared to the CER dataset. The larger sample size offered more robust clustering results by limiting the impacts of outliers. Unlike Chapter 3, the objectives did not focus on the sensitivity clustering of different weather variables. Instead, there were three main aims to build up a richer analysis of residential consumption patterns in a humid subtropical area of China. Firstly, we compared weekly profiles at different temporal periods of a day in summer and winter. Secondly, the consumption patterns in two major holidays in China (the Spring Festival and the National Day) were examined. While the demand clusters were extremely divided during the Spring Festival, changes during the National Day holidays were not significant. Lastly, we identified how households responded to extremely hot weather and concluded that essentially all households would heavily increase their consumption, especially at bedtime.

By investigating the usage habits from these three perspectives, we were able to structure and analyse urban households' behaviour patterns on different days and under extreme weather conditions without the assistance of socio-economic data. It is a common dilemma when analysing Chinese residential consumption that very few if any surveys have been done of electricity consumption behaviour. Chapter 4 solves the problem by using clustering methods and provides better approaches to unravel household behaviour patterns. Research relating to residential demand in China has been highly concentrated on macro-level analysis, that is, how residential electricity consumption is influenced by macroeconomic data, such as GDP, population, income level, etc. Few studies have used aggregate usage data at finer than monthly resolution. In micro-level studies, due to data availability considerations, most studies used recalled monthly bill data gathered in surveys, which gives rise to concerns over the accuracy of the residential consumption data. The methodology employed in Chapter 4 proposed a new direction to cluster the consumption patterns by comparing the fluctuations within a given period. In addition, the innovation of using new indexes to cluster (reflecting demand change percentages) produces more robust results than from either employing direct usage data or standardising the analysis using maximum/minimum usage. One shortcoming of this study comes from needing to rely on lower-resolution data, compared to the CER dataset. The daily consumption dataset only included three data points per day, which explains why it would also be unrealistic to examine the weather sensitivities in a similar way to Chapter 2 and 3. As more refined data becomes available either in China or elsewhere there will be opportunities to reinsert weather sensitivities into analysis of residential consumption.

Based on the research in this thesis, some interesting future work can be done if data becomes available. 1) whether people facing different electricity pricing schemes would respond to weather changes differently at various periods of day? The current trials normally lasted less than 2 years, which limits the possibility of doing such research in the scope of this research. The findings of the suggested topic could shed light on designing efficient and equal pricing strategies that would not punish households in energy poverty; 2) this analysis could also serve as a starting point for classifying customers by their daily life patterns. Understanding during which periods individuals may prefer to be outside of the home and when they are more likely to have spare time or be more flexible in their behaviour patterns could be important when keeping load peak under control. This work and the methods presented herein could be the basis of a new prediction model to classify existing or new customers' behaviour patterns and responses to weather conditions; 3) the methodology here can be applied to other regions and enhance understanding of local electricity consumption behaviour, since the findings here could not be generalised and the responses to weather largely depends on cultural, economic, and geographic factors. The model can then serve as a module for an integrated model of consumer classification and load management.

Reference

Albadi, M. H. and El-Saadany, E. F. (2008) 'A summary of demand response in electricity markets', Electric Power Systems Research. 11, pp. 1989–1996. doi: 10.1016/j.epsr.2008.04.002.

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.

Cialani, C. and Mortazavi, R. (2018) 'Household and industrial electricity demand in Europe', Energy Policy. 122, pp. 592–600. doi: 10.1016/J.ENPOL.2018.07.060.

Dilaver, Z. and Hunt, L. C. (2011) 'Modelling and forecasting Turkish residential electricity demand', Energy Policy, 39(6), pp. 3117–3127. doi: 10.1016/j.enpol.2011.02.059.

Haider, H. T., See, O. H. and Elmenreich, W. (2016) 'A review of residential demand response of smart grid', Renewable and Sustainable Energy Reviews. 59, pp. 166–178. doi: 10.1016/j.rser.2016.01.016.

Torriti, J. (2014) 'A review of time use models of residential electricity demand', Renewable and Sustainable Energy Reviews. 37, pp. 265–272. doi: 10.1016/j.rser.2014.05.034.

Appendix A

Chapter 2 appendices

A.1 Weather dataset comparison

We investigated which datasets could better reflect the consumption responses to weather changes through correlation analysis. First, we calculated the correlation coefficients of each household to the weather variables separately. Secondly, the average correlation coefficients for the weather factors were obtained for each weather dataset and shown in Figure A.1. The weighted method seems the most balanced dataset that is with higher correlations across different weather variables.

Observatory stations	Rain	Temperature	R. Humidity	Wind Speed	Sun Duration
Belmullet	0.051	-0.274	0.088	-0.031	-0.088
Cork Airport	0.016	-0.274	0.091	0.049	-0.090
Dublin Airport	0.047	-0.267	0.194	0.041	-0.132
Valentia	0.024	-0.265	-0.026	0.000	-0.076
Weighted	0.051	-0.271	0.165	0.020	-0.129

Fig. A.1 Average correlations between weather variables and household demands

Appendix B

Chapter 3 appendices

B.1 Weather data comparison

We tried two weather datasets, 1) from the Dublin Airport station; 2) from the weighted weather dataset from four weather observatory stations (see Chapter 2 Section 2.3.2). We examined the data similarity with the t-test. The result in Figure B.1 demonstrated that the two datasets were highly similar. We further looked at the representative curves generated from these datasets and found that the trends are similar and comparable. Therefore, we decided to use the Dublin dataset for the following reasons:

- The differences between using the two datasets were not statistically significant.
- The Dublin dataset could retain the information of extreme weather conditions, while this information could be cancelled out when calculating the weighted dataset.

Weather variables	p-value
Temperature	0.998
Wind speed	0.995
Rain	0.998
Sun Duration	0.998
Relative humidity	0.998

Fig. B.1 Weather data t test results

B.2 Cluster number selection

To cross-validate the suitable cluster numbers for different scenarios, we used combined techniques, including silhouettes scores, DBI scores, and silhouettes analysis. In general, the higher the silhouette score/the lower the DBI score it is, the better the clustering performance it means. It should be noted that no absolute optimal cluster number exists and it largely depends on the objectives of the research as well as the selection of validity indices. Based on these rules, we chose seven as the cluster number for the workday scenarios for the sun duration as well as all the scenarios for temperature, while six was the optimal number for the sun and rain weekend scenarios.

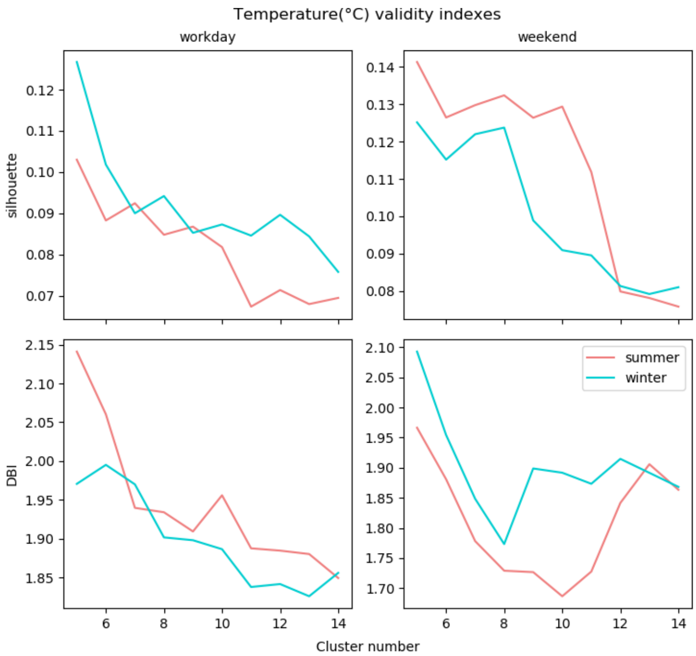

Fig. B.2 Clustering validity indexes for temperature

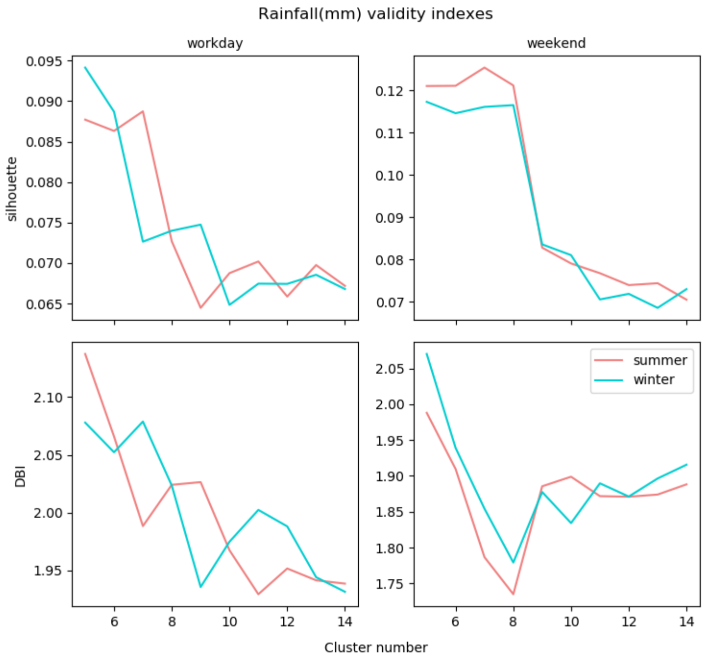

Fig. B.3 Clustering validity indexes for rainfall

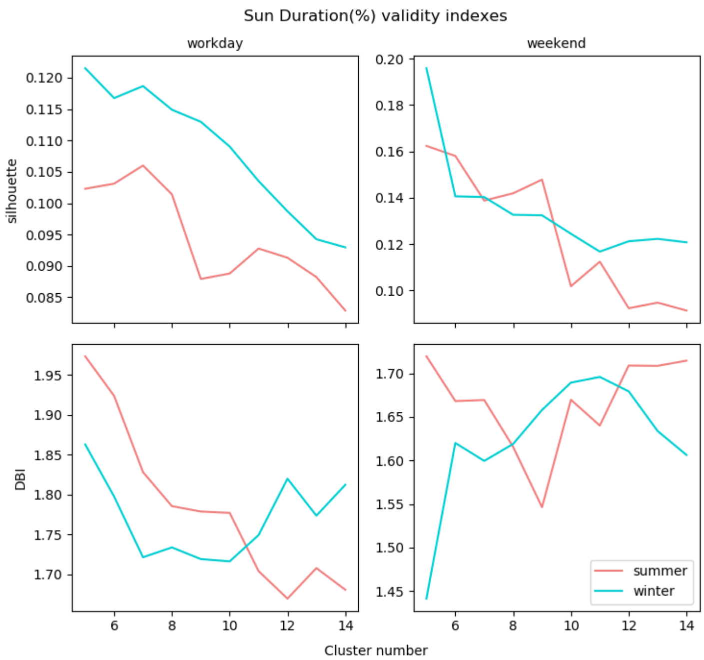

Fig. B.4 Clustering validity indexes for sun duration

We only showed one of the silhouette analysis as an example here (Figure B.5), since the actual analysis was significantly longer and remained irrelevant to the objective of this paper. In principle, a silhouette plot with evenly distributed areas across clusters and a high silhouette coefficient would be ideal. In this case, the right plot (cluster=7) would be better than the left (cluster = 6).

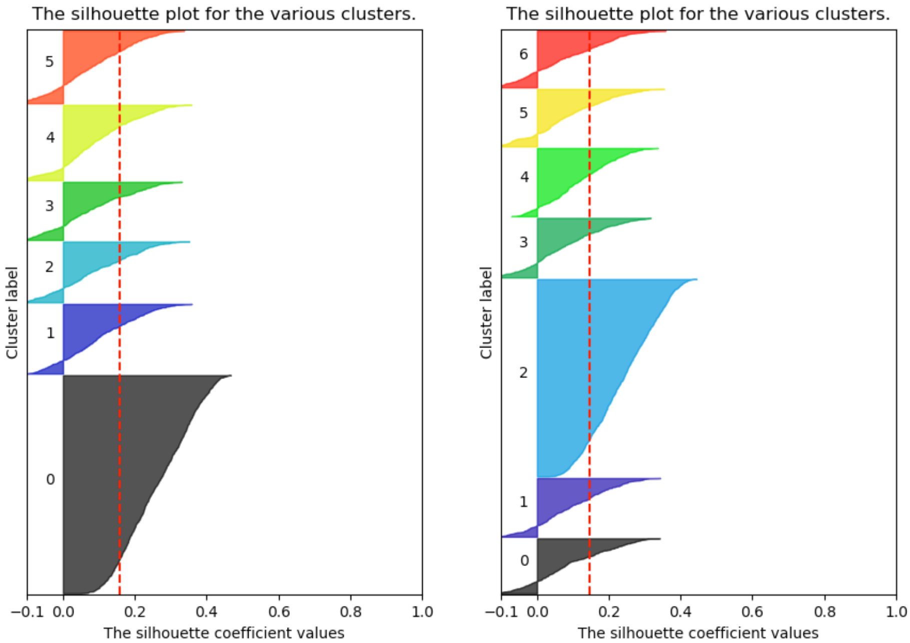

Fig. B.5 Silhouette analysis for daily profiles, when cluster = 6 (left) and 7 (right)

Appendix C

Chapter 4 appendices

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.

C.1 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 C.1. It can be seen that only Ward and Average linkages are not sensitive to outliers and can divide households better.

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

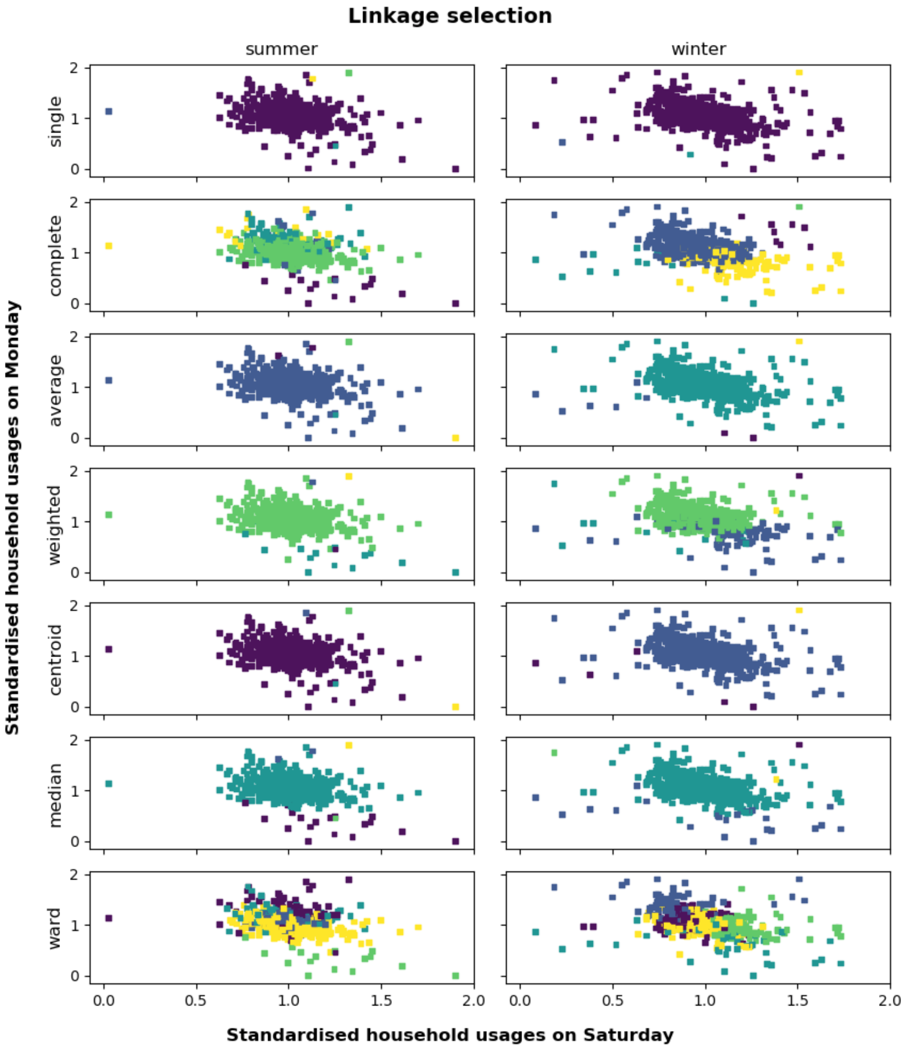

Fig. C.1 household demand between two periods – Linkage selection

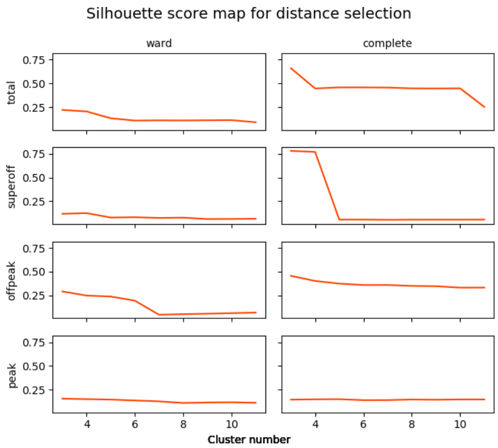

Fig. C.2 Silhouette score for selected linkages

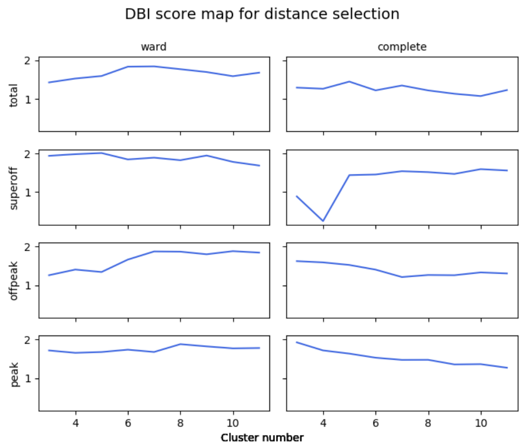

Fig. C.3 DBI score for selected linkages

C.2 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 C.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 C.5). Similarly, a lower DBI score would suggest a higher clustering quality.

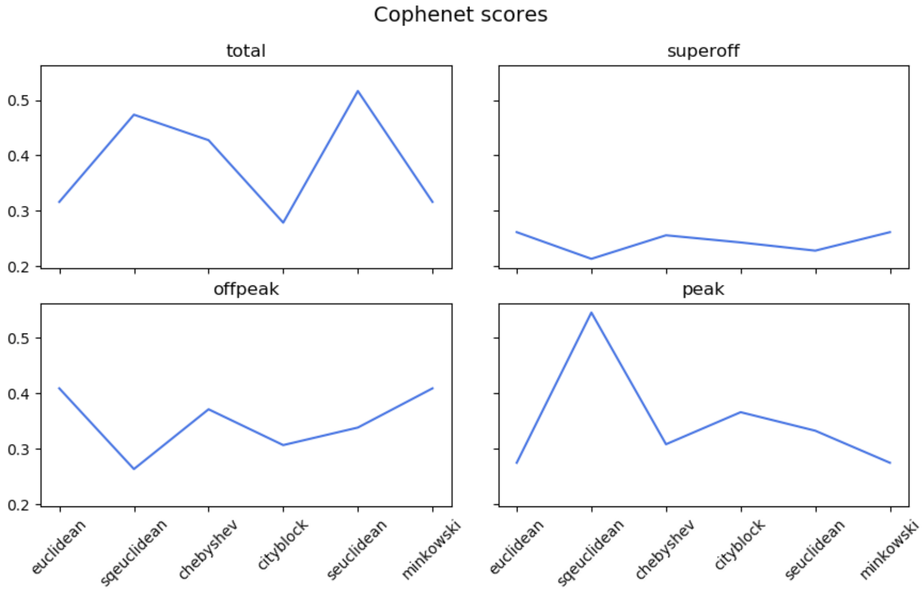

Fig. C.4 Cophenet score for distance matrixes

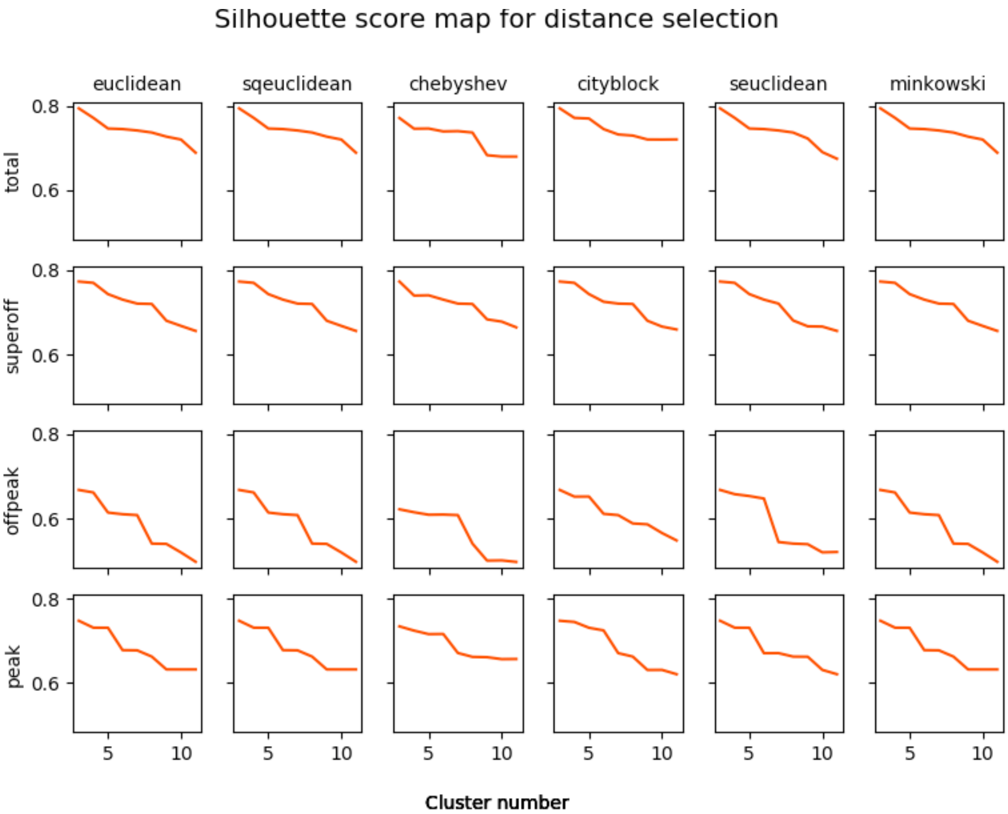

Fig. C.5 Silhouette score for distance matrixes

C.3 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 C.6 and C.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.

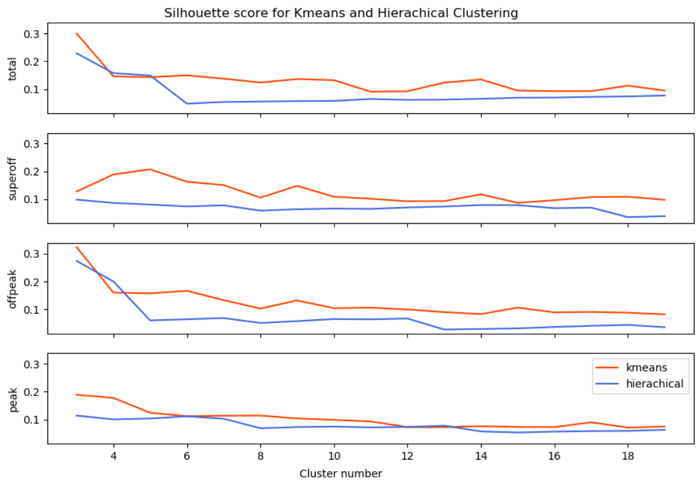

Fig. C.6 Silhouette score

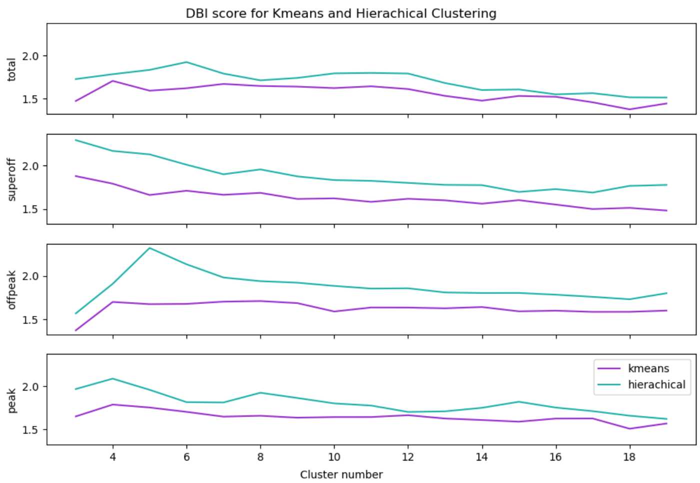

Fig. C.7 DBI score