This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

Cluster-based characterisation of Australian apartment electricity demand and its implications for low-carbon cities

Mike B Roberts [a,b,1], Navid Haghdadi [a,c], Anna Bruce [a,b] and Iain MacGill [a,c]

[a] Centre for Energy and Environmental Markets,

[b] School of Photovoltaics & Renewable Energy Engineering,

[c] School of Electrical Engineering and Telecommunications,

University of New South Wales, Sydney, NSW 2052, Australia

Abstract

Understanding of residential electricity demand has application in efficient building design, network planning and broader policy and regulation, as well as in planning the deployment of energy efficiency technologies and distributed energy resources with potential emissions reduction benefits as well societal and household cost savings. Very few studies have explored the specific demand characteristics of apartments, which house a growing proportion of the global urban population.

We present a study of apartment electricity loads, using a dataset containing a year of half-hourly electricity data for 6000 Australian households, to examine the relationship between dwelling type, demographic characteristics and load profile. The focus on apartments, combined with the size of the data set, and the representative seasonal load profiles obtained through clustering full annual profiles, is unique in the literature. We find that median per-occupant household electricity use is 21% lower for apartments than for houses and that, on average, apartments have lower load factor and higher daily load variability, and show greater diversity in their daily peak times, resulting in a lower coincidence factor for aggregations of apartment loads. Using cluster analysis and classification, we also show the impact of dwelling type on the shape of household electricity load profiles.

Keywords

Apartments; Residential electricity demand; Load profiles; Cluster analysis; Load aggregation; low-carbon cities

1. Introduction

The Paris Agreement that emerged from COP 21 in December 2015 commits the world to "pursuing efforts" to limit average global temperature to 1.5ºC above pre-industrial levels by reducing anthropogenic greenhouse gas (GHG) emissions [1]. Globally, residential buildings are responsible for 13% of end energy use and 28% of electricity use [2], while the IEA estimates that keeping average temperature rises to 1.5ºC requires "virtually all" residential and commercial buildings to achieve net-zero emissions by 2040 [3].

In the absence of coherent national policies, numerous Australian cities have adopted ambitious carbon reduction targets [4-6], but achieving these reductions in the residential sector is complex, involving consideration of embodied and operational energy in residential buildings, as well as wider energy use to deliver services such as transport. The residential sector as a whole is currently responsible for an estimated 11% of Australia's total final energy use [7]. As in many countries, increasing population is driving compact city planning strategies across Australia, with the number of occupied residential dwellings forecast to rise to almost ten million by 2020, an increase of 61% since 1990 [8]. While 14% of all Australian dwellings are currently

Abbreviations: ABS, Australian Bureau of Statistics; ADMD, After Diversity Maximum Demand; BOM, Bureau of Meteorology; CER, Commission for Energy Regulation; CF, Coincidence Factor; CV, Coefficient of Variation; HVAC, Heating, ventilation, and air conditioning; LF, Load Factor; MLR, Multinomial Logistic Regression; NMI, National Meter Identifier; NSW, New South Wales; PCA, Principal Component Analysis; PCA, Profile Class; PV, Photovoltaic; RFE, Recursive Feature Elimination; SCM, Self-Consumption Metric; SGSC, Smart Grid Smart City.

[1] *Corresponding author:* m.roberts@unsw.edu.au

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

apartments[2] [9] and these house 10% of the population, apartments now comprise one third of all new residential buildings [10]. Although much has been written [11-14] about the implications for energy use and carbon emissions of increasing housing density including this move to greater apartment living, there is still a lack of clear consensus, due in part to a lack of detailed analysis of operational energy loads across the diverse range of multi-occupancy residential buildings, and their comparison with stand-alone housing.[3]

A greater understanding of operational energy use in this sector, particularly of the temporal variability of electricity loads, diversity across different building types and between households, and the differential effects of demand aggregation between stand-alone homes and apartments, can better inform regulation and design strategies for residential buildings and urban developments. One clear opportunity is to assist network operators in forecasting capacity requirements for augmentations, replacements and new developments. More generally, it could also help to identify suitable opportunities for distributed energy resources such as photovoltaic (PV) generation, storage, energy efficiency and flexible residential loads to contribute to shaping individual or aggregated household demand in order to reduce household and network energy expenditure, while reducing urban climate change emissions and supporting the transition to low carbon cities.

Increased availability of individual household load data with high temporal resolution, as residential accumulation meters are replaced by smart or interval meters, offers an opportunity to extend this understanding. In Australia, however, policies for deployment of these meters are inconsistent between states and between electricity utilities [16, 17]. Even where suitable meters are widespread and high quality data is available to Network Service Providers and retailers, concerns about customer consent, confidentiality and data protection [18] can result in data that is hard to access or dislocated from information about household characteristics or dwelling type.

Smart Grid Smart City (SGSC) [19] was a major Australian project aimed at assessing the impact of a range of demand response products, including tariff structures and energy monitoring tools, on residential customers' energy use. Household level 30-minute interval load data was collected and made publicly available for some 13,700 residential customers across New South Wales over a three-year period and a subset of this dataset containing a complete year of data (allowing analysis of seasonal variations) for approximately 6000 households has been utilised for this study. Crucially, this subset includes some 2000 apartment households, enabling a detailed comparison between dwelling types, an assessment of the diversity of load profiles between different apartments and an analysis of the effects of aggregating loads for each type of dwelling. This combination of a large household dataset, associated demographic survey information, and the use of clustering to derive representative load profiles from 12 months of interval meter data has not been applied to multi-occupancy dwellings in the literature to date.

The rest of this paper is structured as follows. Section 2 provides a brief review of previous studies analysing household load profiles, both internationally and in Australia. In light of the global growth in apartment living, it is interesting to note that the literature contains little specific analysis of apartment electricity demand, either in comparison to stand-alone housing or with respect to the diversity of building and household characteristics found in this housing sector. Our emphasis on the distinct characteristics of apartment demand is therefore somewhat novel and may be useful in drawing out detail pertinent to electricity networks that are adapting to evolving urban landscapes as well as transitioning towards distributed and low-carbon generation. Section 3 introduces the dataset used for this study and summarises the processes used to select households and to prepare the data. In section 4, we compare average (total and per capita) energy demand for apartments and houses and examine features of their average daily load profiles. In section 5, we assess the temporal variability of the annual load profiles, using a range of metrics pertinent to assessing network impacts and to designing PV and battery storage systems, and examine the effect of aggregation on these metrics for apartments and for houses. In section 6, we present the results of a cluster analysis, carried out to categorise the annual household

[2] The terms 'Apartment' and 'Unit' are used interchangeably in this paper

[3] It is important to note that apartment building electricity use combines common property loads and individual apartment loads. Common property loads are highly variable across Australia's diverse building stock 15. Roberts, M.B., et al., *Using PV to help meet common property energy demand in residential apartment buildings*, in *Australian Summer Study in Energy Productivity*. 2016: Sydney. and, although negligible for many two- or three-storey 'walk-ups', can be responsible for a large proportion of total building energy use in high-rise tower blocks, where they may include lifts, carpark ventilation and lighting, water heating and pumping for pools and centralised HVAC and water heating for apartments, as well as lighting for stairwells, corridors and other common areas.

load profiles and relate them to a range of household characteristics. Finally, in Section 7, we discuss our results, draw some tentative conclusions regarding the energy and hence emissions performance of apartments, aggregate demand network impacts as well as their potential suitability for PV and other distributed energy options compared to stand-alone housing, and suggest possible areas for further study.

2. Previous work

Approaches to the analysis of residential electricity consumption are generally categorised according to the type and level of data used. Top-down approaches use high level aggregated data at the scale of substations or the wider network, while bottom-up approaches use metered electricity load data at a household level or synthesise household loads from appliance usage statistics and physical models of dwelling performance.

A comprehensive review of models for predicting household energy loads is outside the scope of this paper, and a number of such reviews can be found in the literature [20-26] . Instead, the following summary is focussed on studies that apply cluster analysis to high resolution load profiles and those that are specific either to apartment loads or to the Australian context.

2.1 Energy use and household characteristics

Annual household load profiles are dependent on multiple factors including dwelling characteristics, climate, household demographics, appliance ownership and cultural modes of energy use [27]. A considerable number of studies have examined this relationship but, whether by necessity (reliance on low temporal resolution data, such as electricity bills) or by design, many of these studies [28-35] characterise household energy use by the total (or average daily) energy demand, without consideration of the temporal variation of the load. More recently, widespread deployment of interval meters for residential customers has resulted in increasing availability of high-resolution household load data [20] with sampling periods ranging from one minute to an hour, which can be used to identify the contribution of individual appliances to overall load [36-38] for forecasting, including to predict the contribution of individual appliances to peak electricity demand, measured either at the time of the customer's peak, or relative to aggregate or network peaks [39], and for load management, both customer classification and implementation of demand response [40].

In Australia, energy bills and Australian Bureau of Statistics (ABS) household expenditure data have been used to model household energy use in Sydney [41], and to estimate New South Wales aggregated electricity consumption at the census collection district level [42]. Due to the sporadic deployment of residential smart meters across Australia, as well as privacy concerns (see Section 0), more detailed analysis of residential electricity load profiles in Australia has been limited by the small number of published datasets containing residential customer load data at high temporal resolution.

The Smart Grid Smart Cities dataset used for this paper (see Section 3) has been the subject of a number of other studies [20, 29, 39, 43, 44]. Fan et al. [29, 39] used the data to create a linear regression model relating household characteristics to average daily demand and to peak demand. One of their conclusions was that detached and semi-detached houses have higher average daily electricity demand than apartments [29] and that, during the 12 maximum peak demand periods in the financial year 2013, average detached house loads were approximately twice those of apartments[4] [39]. Elsewhere, the dataset has been used to propose a strategy for demand response to address peak load caused by air conditioning [38]. Looking at temporal variability of the load profiles, Motlagh et al. [44] applied clustering and principal component analysis (PCA) techniques to 30-day profiles for 1800 of the SGSC customers to assess the impact of the dynamic peak pricing and dynamic peak rebate tariffs on the load profiles.

2.2 Cluster analysis

Cluster analysis is a commonly-used data mining technique that can be used to group energy load profiles in order to identify characteristics common to groups of households. Overviews of different clustering techniques, as commonly applied to aggregated residential loads or to individual commercial loads, are available in the literature [20, 45]. However, the application of clustering to individual residential loads is less common due to reliance on a small group of published datasets. In a US study [46], daily profiles of 85 customers were clustered and then each customer was characterised by the proportion of their daily profiles to be found in each cluster. Irish researchers [47] have used a similar approach, applying clustering techniques to each daily profile of 3941

[4] Detached house loads were between 98% and 107% more than apartment loads, depending on the use and type of air-conditioning in the household

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

households, drawn from a dataset collected by the Commission for Energy Regulation (CER), over a six-month period, assigning each customer to one of ten profile classes (PC) based on the modal PC of its daily profiles, and then describing the relationship between profile class and household characteristics.

Rhodes et al. [48], using one-minute load data from 103 Texan households, clustered the seasonal profiles (rather than daily profiles) of customers (a technique elsewhere applied to the CER dataset [49]) and applied a regression analysis to relate the clusters to household characteristics. However, as the paper does not reference dwelling type, it is not clear whether any of the households were apartments.

In their CSIRO report for Queensland electricity network operator Energex, Berry et al. [50] examined 5 datasets. Load profiles for 1871 customers across three of the datasets were grouped into clusters according to daily load shape and seasonal trends, identifying five core clusters and three smaller outliers. Separately, 921 households from a single dataset were segmented according to demographic characteristics (dwelling type, household composition and income), their load profiles clustered within each segment, and a single profile chosen as being representative of each cluster. With up to three clusters within each of eight segments, group sizes were small, although some of these intra-segment clusters showed strong alignment with the clusters derived from the larger dataset. However, as the sample contained only 34 households not living in detached houses, the report is not able to draw strong conclusions about the relationship between profile and dwelling type, except to say that the largest cluster of non-detached dwellings has the lowest typical demand of all the segments. Also in Australia, Yildiz et al. [51] developed a clustering approach to classify households with rooftop PV according to their daily load and PV generation profiles, in order to assist in forecasting self-consumption, import and export.

2.3 Apartment energy use

Most of the studies mentioned above are either based on houses or do not specifically consider dwelling type. A small number of studies have directly addressed the relationship between household loads and dwelling type, but the conclusions drawn are limited and sometimes contradictory.

An international review of factors affecting domestic energy consumption [52] cites the conclusion of a number of studies in the UK, the Netherlands, Ireland, Portugal and the USA, that apartments use less energy than detached dwellings. Amongst studies that include dwelling type as a potential factor in energy use, many have found that total electricity consumption increases with the degree of detachment [26, 53-57], although one [58] did not include apartments in the analysis. Others have been less conclusive, finding no relationship between dwelling type (other than mid-terrace houses) and high electricity consumption in UK dwellings [52], or lower consumption of gas or other fuels in apartments, but no significant difference in electricity consumption [59, 60]. Others [61] argue that, because dwelling size is correlated to size of household, the effect of dwelling type on energy use is covered by these factors (houses use more energy because they are larger than apartments), and so does not require explicit consideration. However, a study of dwellings in urban and rural Finland [62] found that, although average energy use in apartments is significantly lower than in houses, apartment *per capita* energy use is only slightly lower in rural areas, and in urban areas is higher in apartments than *per capita* energy use in houses.

Although the Irish CER dataset used in a number of studies [27, 47, 63, 64] included only 67 apartments amongst over 4000 residential dwellings, one of the studies [27], examining the dependence of four load parameters on household characteristics for 4200 households, found apartments to have a significant negative impact on total load and maximum demand. However, no significant correlation was found between load factor or time of peak demand and apartment dwellings, (although terraced and semi-detached houses showed a significant negative impact on load factor, compared to detached dwellings).

Two commonly cited Australian studies give conflicting views of the relationship between dwelling type and energy use. A 2010 IPART study [32] suggests that apartments use less total energy (gas and electricity) than detached dwellings. Although this study does not consider apartment building common property loads, these are likely to be significant only for the 8% of the apartments surveyed that were in buildings of three storeys or more [65]. A 2005 study by Energy Australia [35] is very often quoted as finding that high rise apartments and detached houses have higher energy use than low or medium rise apartments and that *per-capita* energy emissions are highest in high-rise apartment buildings and higher in mid- and low-rise apartment buildings than in detached houses, although it is unclear whether these emissions calculations are based solely on total energy use or consider the energy sources utilised.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

In contrast, a detailed analysis of datasets from several studies, including the IPART report, [65] estimated that energy use in low-rise attached dwellings is 15-20% lower than in detached dwellings, but allowed the possibility that dwelling type may, to some degree, act as a proxy for dwelling size. The same analysis also suggests that the apparent high energy use of high-rise apartments attributed to the Energy Australia report may be due to the presence of "luxury common area features". Previous work by the authors [15] highlighted the diversity of common property demand which can present a significant opportunity for renewable generation and demand management.

Given the lack of concrete conclusions on apartment loads, it is unsurprising that there is high variability in the consideration given to dwelling type by Australian Network Service Providers when estimating 'After Diversity Maximum Demand' (ADMD), the average customer contribution to aggregated peak demand. For example, in Western Australia, Western Power use ADMD values of 3.1 – 5.4kVA for apartments in buildings with more than ten dwellings, compared to 4.7 – 8.7 kVA for detached dwellings [66], while in NSW, Endeavour Energy use values of 6.5 or 7.5kVA for apartments (in areas with and without gas supplies, respectively), the same as the value for small (and in most areas for medium) detached houses [67].

3. Data source and preparation

3.1 Dataset

This study utilises a dataset [68] of 30-minute interval load data from households across eight local government areas in New South Wales (NSW), collected for the Smart Grid Smart City (SGSC) Project. Details of the project and dataset are provided by Motlagh et al. [44], and in the SGSC Executive Report [19] and SGSC Technical Compendia [69-71], with additional information about the customer selection process contained in the 'Futura Report' [72]. Of the 78,000 households involved in the study, 13,735 provided some interval load data between 2011 and 2014, with the highest concentration of data collected during the calendar year 2013. This period was therefore chosen for the analysis, with customers excluded from the dataset if more than 10% of their data was missing or zero[5] over the period.

For the SGSC study, the type of dwelling for each customer was inferred from the number of 'National Meter Identifiers' (NMI)[6] at the core address – if there were six or more NMIs, the dwelling was designated as 'Unit'; if less than six, as 'Not Unit' (i.e. detached or semi-detached houses). As these inferences were correct in 96% of cases where a sales agent visited the dwelling [71], the 'assumed dwelling type' has been used for much of this analysis.

3.2 Data selection

The SGSC project involved trialling a number of tariff structures (seasonal time of use, dynamic peak pricing and dynamic peak rebate) and customer feedback products (online portal, home energy monitor and appliance control) to investigate their impact on customer load and peak demand. Some of the load data was therefore influenced by participation in the trials. In order to reduce this influence, three groups of households have been selected from the SGSC trials: the 'control' group of 1809 customers who were supplied with a smart meter but not offered any of the trial products; 3164 customers supplied with a smart meter who declined all products offered to them; and 1150 customers who were supplied with smart meters and access to an online portal to monitor energy use, but no other products. (For this latter group, the SGSC study found that the impact of the trial on energy use and peak demand were only 0.1% and -0.7% respectively [69]. Of these customers, a further 89 (including 23 from the control group) who were identified as having net metering for solar generation were removed from the dataset. The geographic area of the dataset includes two climate zones, with 79% of households in the Warm Temperate regions near the NSW coast and 21% in the Mild Temperate regions further inland.

Although precautions were taken to ensure the initial SGSC customer group was representative of the wider population [71], the various methods used to sell customers each of the demand response products are likely to have resulted in selective biases within each of the product groups. The control group (of 1786 customers) is therefore the only group that can be considered to be more broadly representative of the customer population,

[5] Zero values in the dataset represent loads less than 2W and are most likely to be empty dwellings or missing data as most households have at least one appliance on standby which would exceed this value 73. Equipment Energy Efficiency (E3) Committee, *Standby Power – Current Status*. 2006.. Additionally, the dataset contains blocks of zero data which appear to be periods without data recording (for example, immediately after meter connection).

[6] The NMI is a unique identifier assigned to each electricity connection point within the Australian National Electricity Market

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

and so has been used for the first part of this study (sections 4, 5 and 6.3). However, as this group has no associated demographic or household survey information, all three groups were used in the exploration of household characteristics (section 6.4). This larger dataset comprises 6034 households, of which 2081 are apartments or units. Of these, customers from 220 units and 1150 houses participated in a survey, providing information regarding household demographics, dwelling characteristics and energy and appliance usage.

3.3 Data preparation

Prior to analysis of the load data, missing data was filled to ensure an equal number (17520) of general load readings for each customer. Following the method used in the CSIRO-Energex report [50], the timestamp with missing data ($t_{missing}$) for a given customer *c* was compared to all other timestamps in the period to find the 'most similar' timestamp ($t_{similar}$), and the load data copied from the 'most similar timestamp' to the missing timestamp $E_{c,t_{missing}} = E_{c,t_{similar}}$. Similarity between two timestamps is defined by Equation(1) and found by calculating the sum across the whole dataset of the square of the difference between the load data for the two timestamps. For customers with missing data at the 'most similar' timestamp, the second most similar timestamp was used, and so-on. As all customers have some general electricity load, this method was used to fill all missing *general load* data.

$$t_{similar} = \left\{ t_i | 1 \leq i \leq 17520 : min\left(\sum_{\forall c} \left(E_{c,t_i} - E_{c,t_{missing}} \right)^2 \right) \right\} \qquad \textit{Equation(1)}$$

Of the 6034 households, 3531 have interval data for a controlled load (commonly hot water heating) in addition to metered general load. Although this controlled load was excluded from the exploration of load variability and daily profiles, it was included in calculations of total energy use (Section 4.1). Because the controlled load is inherently intermittent and time-specific, gaps in this part of the dataset were only filled for timestamps with a simultaneous gap in the general load data, and the 'most similar' timestamp for controlled load was constrained to the same time of day as the timestamp for the missing data. Finally, the annual load profiles were normalised, using the 99th percentile of the data to ensure robustness against outliers. The processed dataset, along with household characteristics and the results of the cluster analysis will be made publicly available.

4. Average Daily Energy

4.1 Average total daily energy use

Using the total (general + controlled) load for each customer, daily average loads were calculated for the whole dataset, divided into 'units' and 'not units'. Figure 1(a) shows that average median household loads are twice as large (17.7 kWh/day compared to 8.9 kWh/day) for detached and semi-detached houses as for units. This is in part due to the relatively low occupancy rates of apartments compared to houses, averaging 1.9 compared to 2.7 [9].

As shown by the frequency distribution of average daily energy *per occupant*, plotted for the smaller surveyed dataset (Figure 1(b)), the median daily load *per occupant* for houses is 7.16 kWh while that for units is 21% lower at 5.69 kWh. In order to assess the significance of the difference in these values, a statistical *t*-test was applied to the two distributions. Figure 1 suggests it is at least possible that the two distributions have different variances, so Welch's (unequal variance) *t*-test [74, 75] was applied instead of the more common Student's *t*-test. This gave a p-value of 0.000005% which allows rejection of the null hypothesis that the two distributions have equal averages and confirms that the average daily household load, normalised for occupants, is significantly higher for houses than for units, while the greater difference in summer (29%) compared to winter (21%) is indicative of greater cooling loads in houses.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

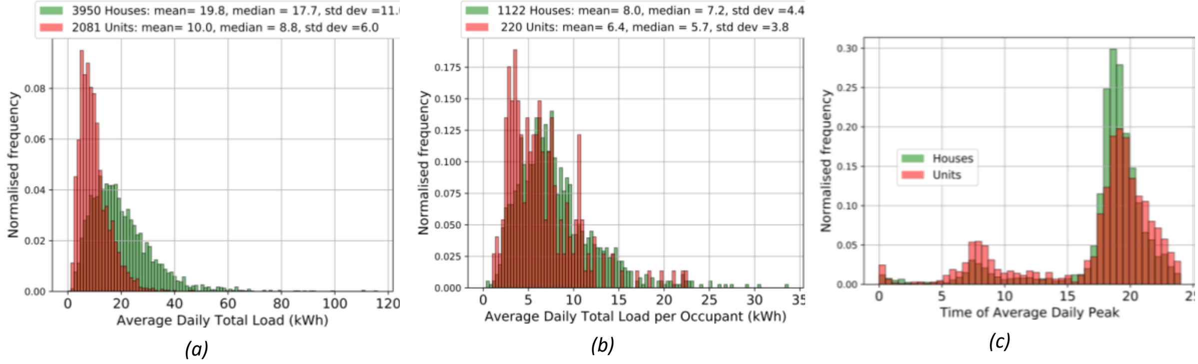


Figure 1 - Frequency distribution of (a) average daily total load, (b) average daily load normalised for occupants and (c) average time of daily peak

4.2 Average daily load profiles

Figure 2 shows the normalised average daily load profile for units and for houses. Compared to houses, units show a smaller evening peak, particularly in summer. The frequency distributions for the peak load time (Figure 1(c)) show a higher proportion of units have their daily peak in the morning; consequently, the mean peak time period is earlier for units (17:00) than for houses (17:30), with a broader distribution (standard deviation is 5.3 for units, 4.1 for houses).

In all seasons, average unit daytime loads are flatter than houses (see also Section 5.2), suggesting potentially greater suitability for PV self-consumption.

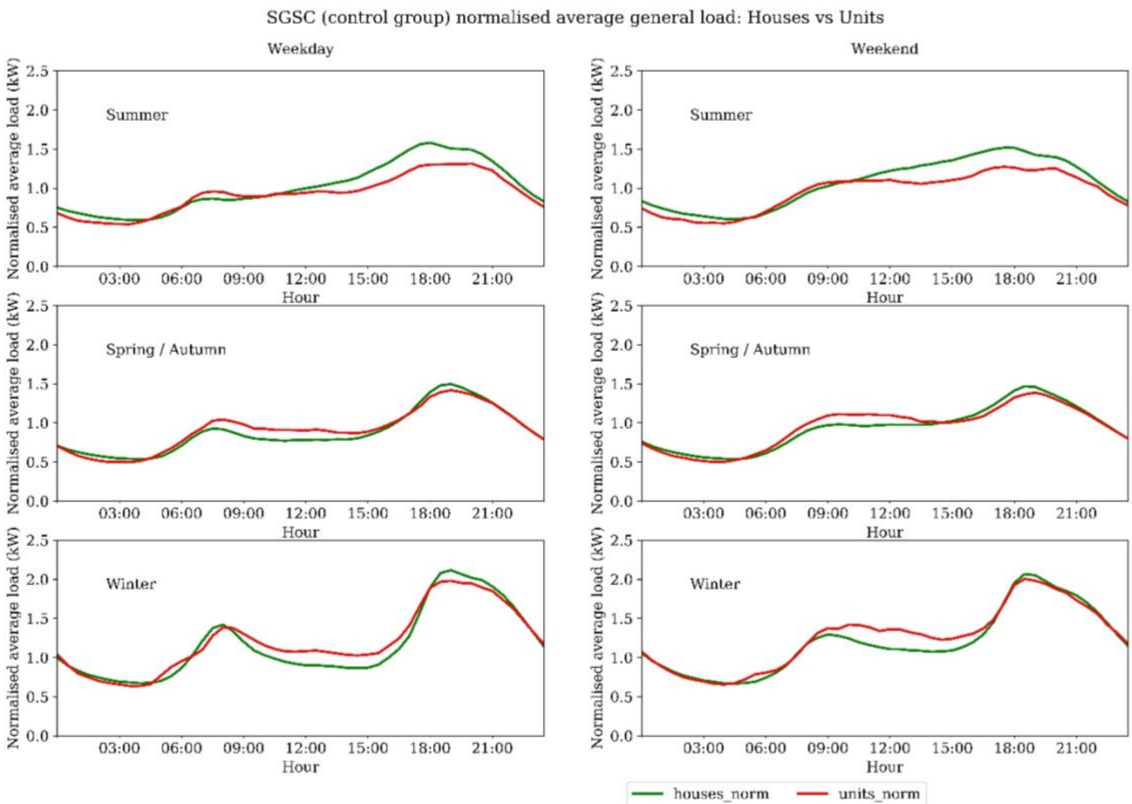


Figure 2 – Normalised average daily general load profiles for houses and units in control group

5. Variability and load aggregation

5.1 Variability metrics

It is useful to understand the diurnal and annual variability of load profiles for the calculation of network capacity requirements, load management, suitability assessment for PV and design of appropriately sized storage. A range of metrics were used to measure variability as a means of comparing general loads of apartments to those of houses (controlled loads were omitted as their time specificity would distort the results), and to explore the effects of load aggregation.

The average *coefficient of variation* (CV) is the ratio of the standard deviation of the daily load to its mean, as shown in Equation(2), where $E_{d,j}$ is the *jth* 30-minute energy reading on day *d.*

$$\overline{CV} = \frac{1}{365}\sum_{d=1}^{365}\frac{\sqrt{\left(\sum_{j=1}^{48}\left(E_{d,j}-\bar{E}_d\right)^2/48\right)}}{\frac{1}{48}\sum_{j=1}^{48}E_{d,j}} \qquad \text{Equation(2)}$$

The load factor (LF), a useful metric for appropriately sizing storage capacity, is the mean energy divided by the peak energy and can be calculated on a daily or annual basis as shown in Equation(3) and Equation(4).

$$\overline{LF}_{daily} = \frac{1}{365}\sum_{d=1}^{365}\frac{\sum_{j=1}^{48}E_{d,j}/48}{\max_{1\le j\le 48}\{E_{d,j}\}} \qquad \text{Equation(3)}$$

$$LF_{annual} = \frac{\sum_{d=1}^{365}\sum_{j=1}^{48}E_{d,j}/17520}{max\{E_{d,j}, 1\le d\le 365, 1\le j\le 48\}} \qquad \text{Equation(4)}$$

Because of the disparity between electricity tariffs for imported and exported energy, the potential for PV self-consumption is a factor to consider in assessing the suitability of different loads to utilisation of on-site PV generation. A range of parameters can be found in the literature [76-78] for measuring PV self-consumption but, in essence, this metric refers to the amount or proportion of on-site PV generation that is consumed by a load, and therefore depends on the size, orientation and output characteristics of the PV system, as well as the weather conditions during the period under consideration, and the size and shape of the load profile.

For the purposes of this study, we propose two alternative PV self-consumption metrics (SCMs) that are independent of both the size of the load and the PV system size, measuring the degree to which the *shape* of an annual load profile matches the *shape* of the generation profile of an optimally orientated, co-located PV system over the same period. They are given by Equation(5) and Equation(6) where $\boldsymbol{p} = (P_1, P_2 \ldots P_T)$ is the normalised PV generation and $\boldsymbol{e} = (E_1, E_2 \ldots E_T)$ is the normalised load for timestamps $\boldsymbol{t} = 1, 2, \ldots T$. The PV profiles were generated using NREL's System Advisor Model (SAM) [79] with weather files for 2013 generated from the Australian Bureau of Meteorology (BOM) gridded satellite insolation data [80] and temperature and wind speed from the BOM automatic weather stations (AWS).[7]

$$SCM_1 = \frac{\boldsymbol{p}\cdot\boldsymbol{e}}{\boldsymbol{p}\cdot\boldsymbol{p}} \times 100\% \qquad \text{Equation(5)}$$

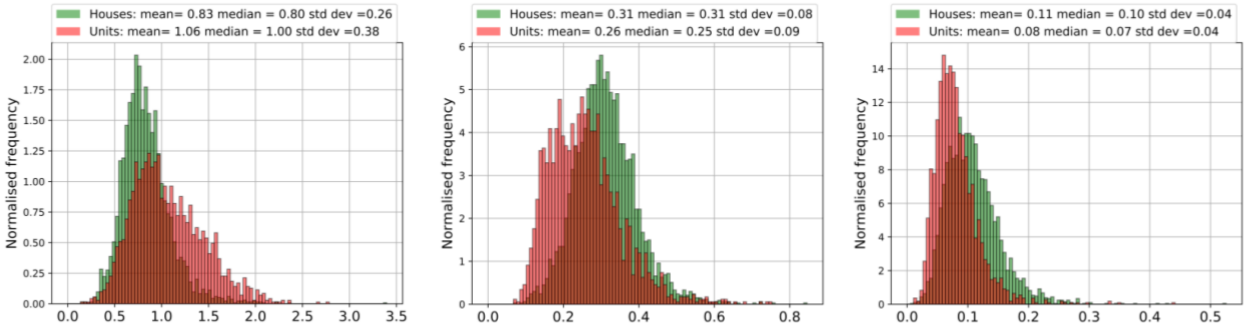
$$SCM_2 = 100\% - \frac{1}{2}\sum_{\boldsymbol{t=1}}^{\boldsymbol{17520}}|P_t - E_t| \qquad \text{Equation(6)}$$

5.2 Variability results

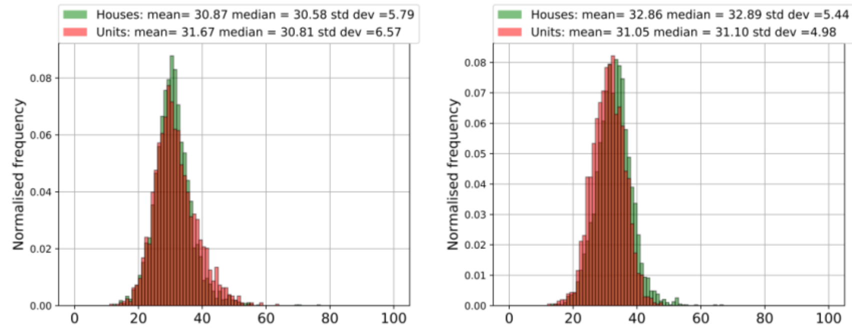
Figure 3 shows the distribution of the variability metrics discussed above for houses and for units, with average values shown in Table 1. Welch's *t*-test, applied to the distributions of each metric, all returned p-values below 0.0003%, showing a high degree of certainty that the variabilities of the house and unit loads are significantly different on all these metrics.

[7] As the exact locations of the SGSC customers are not known, weather files were generated using data for the geographical centre of the postcode area for each customer.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222



(a) Coefficient of variation (b) Daily load factor (c) Annual load factor



(d) PV Self-consumption SCM_1 *(e) PV Self-consumption SCM_2*

Figure 3 - Frequency distribution of variability and self-consumption metrics

On average, individual unit load profiles show higher coefficient of variation and lower load factor (whether calculated on a daily or an annual basis) than load profiles for houses. This greater temporal variability in average unit loads may be related to the lower proportion of households with children, and therefore lower daytime energy use. Alternatively, lower variability in houses may be due to the smoothing effect of higher occupancy. Additionally, Figure 3 shows that apartment households have a broader distribution of temporal variability between customers.

For all dwelling types, mean load factor calculated on an average daily basis is lower in winter than summer (by 9% for units and 12% for houses), while mean coefficient of variation is higher in winter (15% for units, 20% for houses). This suggests that, on the hot summer days when it is used, air conditioning is used for longer duration compared to winter electric heating use (so that days with high peak cooling loads also have high mean loads, whilst high heating peak loads have less impact on the daily mean load) and may be related to households with gas heating and either sparingly-used supplementary electric heating or electric ovens (common even where cooktops are gas). Conversely, load factor calculated across the whole season is lowest in spring / autumn when there are less extreme heating or cooling loads, and is *higher* in winter than summer (by 7% and 9% for units and houses respectively), as a single summer peak day can disproportionately lower the seasonal value. In all seasons, average weekend loads show slightly earlier peak time, higher load factor and lower coefficient of variability than weekday loads for all dwelling types, due to higher levels of daytime occupancy.

Table 1 Average variability metrics for household loads

	Unit Load Profiles			**House Load Profiles**			**Welch's *t*-test**	
	Mean	**Median**	**Std Dev**	**Mean**	**Median**	**Std Dev**	**t**	**P**
Coefficient of Variation	1.05	0.98	0.38	0.82	0.80	0.26	24.9	9.8×10^{-125}
Average Daily Load Factor	0.26	0.25	0.09	0.31	0.31	0.08	-22.7	7.8×10^{-107}
Annual Load Factor	0.08	0.08	0.04	0.11	0.10	0.04	-26.1	1.2×10^{-140}
Average Daily Peak Time	17:00	18:30	5.3	17:30	18:30	4.1		
Self-Consumption SCM_1 (%)	31.7	30.8	6.6	30.9	30.6	5.8	4.7	3.4×10^{-6}
Self-Consumption SCM_2 (%)	31.1	31.1	5.0	32.9	32.9	5.4	-13.0	4.9×10^{-38}

Although the t-test indicates a significant difference in the average self-consumption for houses and units, the values are close with houses or units having the greater value depending on the metric chosen.

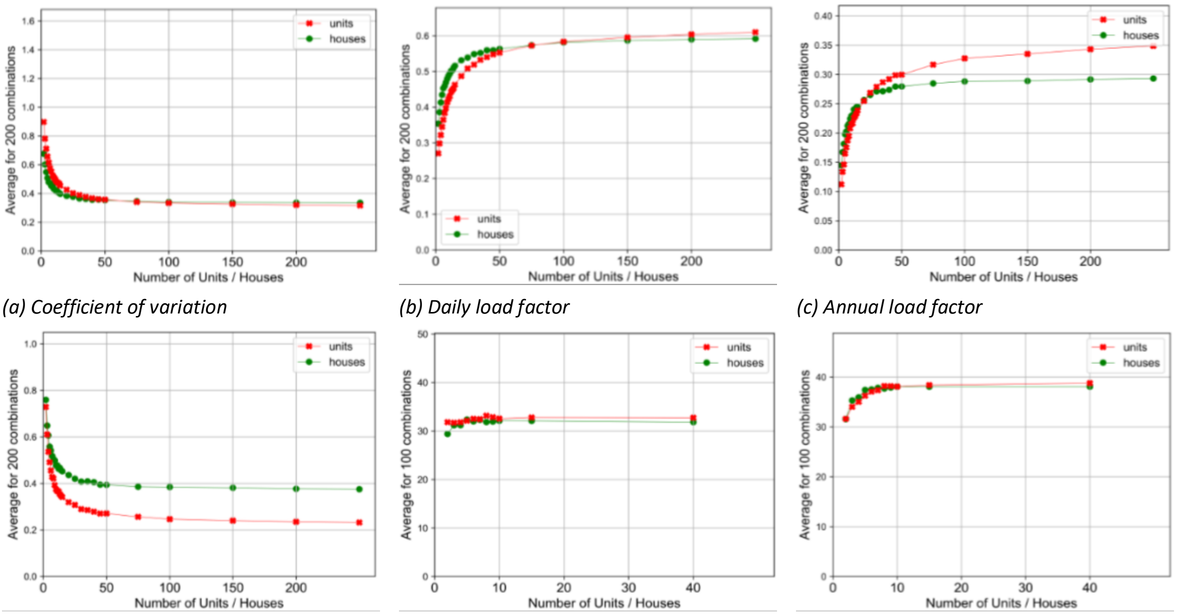
This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

5.3 Load aggregation

For each of the datasets ('units' and 'not units') annual load profiles were selected at random from the whole dataset and combined to create aggregations of between two and 250 households, and variability metrics were calculated for the aggregated load. This process was repeated 100 or 200 times for each size of aggregation and the average of each metric was calculated. Figure 4(a) to (c) show that, for all these metrics, load variability is reduced by aggregation of up to 50 household loads. The higher coefficient of variation and lower *daily* load factor of unit loads compared to house loads is apparent in aggregations of 50 or below, but above this, these metrics do not distinguish between house loads and unit loads. However, *annual* load factor for large aggregations of unit loads is 35% compared to aggregated house loads of 30%, due to the higher likelihood of outlier peak cooling loads in large houses.

On the first metric, self-consumption is not affected by aggregation of household loads, while on the second, there is a significant increase but only for small groups. Using either metric, self-consumption calculated solely on the shape of the aggregate profile is not affected by aggregation above ten households (Figure 4(e) and (f)).

The coincidence factor of a collection of loads can be defined as the ratio of the system or aggregate peak to the sum of the individual, non-co-incident peaks, as shown in Equation (7). This metric has relevance in the calculation of potential demand charge components in retail tariffs, and more generally in assessing network impacts.



(a) Coefficient of variation *(b) Daily load factor* *(c) Annual load factor*

(d) Coincidence factor *(e) PV Self-consumption, SC1* *(f) PV Self-consumption, SC2*

Figure 4 - Variability metrics for aggregated loads

As shown in Figure 4(d), the coincidence factor for unit load profiles tends towards 0.23 for aggregations over 200 customers, while for houses the figure is 0.37. This suggests relatively large benefits in aggregating unit loads (for example in an embedded network) in terms of reduced demand charges and network impacts.

$$CF = \frac{\max\limits_{1\leq i\leq 17520}\left\{\sum_{c=1}^{n} E_{c,i}\right\}}{\sum_{c=1}^{n}\left(\max\limits_{1\leq i\leq 17520}\left\{E_{c,i}\right\}\right)} \qquad \textit{Equation (7)}$$

6. Clustering and classification

6.1 Clustering method

In order to explore the range of load profiles for different customers in different dwellings, a cluster analysis was firstly carried out for ***a***ll the dwellings ('units' and 'not units') in the ***c***ontrol group only (1786 households), as this is the group most representative of the wider population. This dataset (labelled ***ca***) was divided into clusters of similarly-shaped profiles, irrespective of dwelling type or other household characteristics.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

Because there are 17520 features (30-minute timestamps) for each household, the dataset is too sparse to cluster without first reducing the number of features. In order to retain information regarding diurnal and seasonal variability, the profiles were first grouped by month and by the type of day (weekday or weekend), reducing the features to 1125 (number of months x number of day types x number of readings per day = 12 x 2 x 48). Principle component analysis (PCA) was then applied to further reduce the number of features, whilst minimising the loss of useful data. A final figure of 180 features was chosen as optimal for greatly increased computation with only 10% loss of data (see Figure 5).

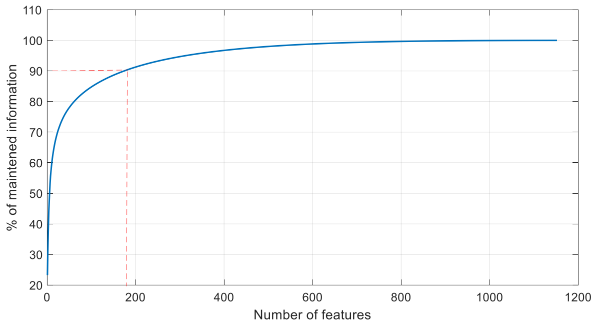


Figure 5- Optimisation of number of features through PCA

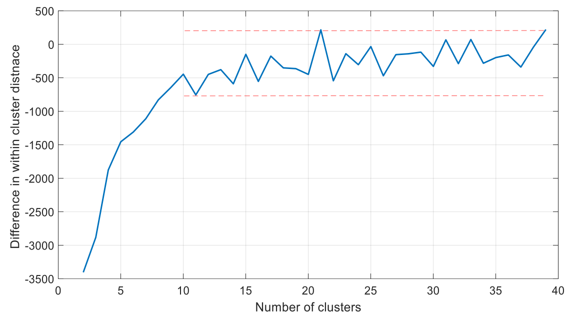


Figure 6 - SUM validity index versus number of clusters

The dataset was then clustered using k-means clustering to partition the whole dataset into up to 40 clusters. The SUM validity index was used to assess the similarity between and within clusters. From Figure 6, it can be seen that increasing the number of clusters above ten gives minimal variation in the index, suggesting ten clusters are sufficient to capture the variability of the dataset.

As the demographic and household data available for the ***c***ontrol group is extremely limited (consisting only of assumed values for dwelling type, household income, electricity use and gas use, as well as the climate zone and geographical region of the dwelling), further clustering was carried out, again using 10 profile classes, for the 2000 ***u***nits (apartments) in all ***3*** data groups (the ***c***ontrol group, those that were offered but declined demand management products, and those that accepted only home energy monitors - see Section 3.2), dataset ***3u***.

6.2 Classification method

A classification process was used to explore the relationship between the profile classes and a mixture of assumed and known household characteristics[8] (for the ***c***ontrol group), with additional household survey data[9] (for the larger dataset). Prior to classification, each of the datasets (***ca*** and ***3u***) was randomly divided into a learning set containing 80% of the customers and a testing set containing the remaining 20%.

In order to examine the dependence of membership of each profile class on customer demographics and household characteristics, a multinomial logistic regression (MLR) was applied to the data. Given *J* profile classes ($Y = 1 \dots J$), the probability of a household with *n* explanatory variables (given by $\boldsymbol{x} = X_1, X_2 \dots X_n$) belonging to profile class *j* is $\pi_j(\boldsymbol{x}) = P(Y = j | X_1, X_2 \dots X_n)$ and the "logit" or log odds ratio" relative to a reference class J is given by Equation(8) [81].

$$log\left[\frac{\pi_j(x)}{\pi_J(x)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n \text{ where } \sum_{j=1}^{J} \pi_j = 1 \qquad \textit{Equation(8)}$$

The coefficients β_i were calculated from fitting an MLR to the training data, using an iterative *Maximum Likelihood* method implemented in Python through sci-kit learn [82] with *one-versus-rest* and *multi-class* training algorithms, and *sag* and *saga* solvers [83]. The β_i coefficients were then applied to the test data to predict the profile classes of each customer. Accuracy scores (the proportion of customers allocated to the correct profile

[8] The data fields for the control group are TRIAL_REGION_NAME, POSTCODE, LOCAL_GOV_AREA_NAME, SUBURB_NAME, LOCATION_TYPE_CD, SERVICE_TYPE, GENERAL_SUPPLY_CNT, CONTROLLED_LOAD_CNT, NET_SOLAR_CNT, GROSS_SOLAR_CNT, OTHER_LOAD_CNT, ASSRTD_HHOLD_INCOME_GROUP_CD, ASSRTD_CLIMATE_ZONE_CD, ASSRTD_CLIMATE_ZONE_DESC, ASSRTD_DWELLING_TYPE_CD, ASSRTD_GAS_USAGE_GROUP_CD, ASSRTD_ELECTRICITY_USE_GRP_CD.

[9] The additional data fields from the household survey are DWELLING_TYPE_CD, HHOLD_INCOME_GROUP_CD, DRYER_USAGE_CD, REDUCING_CONSUMPTION_CD, AIRCON_TYPE_CD, NUM_OCCUPANTS, NUM_CHILDREN_0_10, NUM_CHILDREN_11_17, NUM_OCCUPANTS_70PLUS, HAS_CHILDREN, NUM_REFRIGERATORS, NUM_ROOMS_HEATED, HAS_GENERATION, HAS_INTERNET_ACCESS, HAS_GAS, HAS_GAS_HEATING, HAS_GAS_HOT_WATER, HAS_GAS_COOKING, HAS_SOLAR, HAS_POOLPUMP, HAS_AIRCON, IS_RENTING, HAS_GAS_OTHER_APPLIANCE, IS_HOME_DURING_DAYTIME.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

class) were calculated for the training data and for the test data to determine the optimum number of features to use in the regression.

Before fitting the MLR, the available features were first converted to binary variables by creating 'dummy' variables relative to a base variable which was removed from the list. For the non-surveyed ***c**ontrol group (***ca***), the features were further reduced by removing those with a very low variance (HAS_GROSS_SOLAR) or those likely to have a high degree of co-linearity with retained features (TRIAL_REGION_NAME, HAS_GROSS_SOLAR and ASSRTD_ELECTRICITY_USE), leaving the final list of variables shown in Table 2 below.

For the remaining dataset (***3u***), which includes surveyed customers, a process of feature selection was applied in order to reduce the likelihood of overfitting due to the large number of features in the household survey data. After removing variables with over 90% homogeneity ('HAS_SOLAR', 'LOCATION_TYPE', 'HAS'GENERATION', 'HAS_GROSS_SOLAR', 'HAS_INTERNET_ACCESS') and the 'assumed' and 'assigned' variables (which are likely to have a high level of co-linearity with the equivalent surveyed variables), recursive feature elimination (RFE) was applied to progressively eliminate the least important remaining variables. At each stage, an iterative MLR process was applied to the features and training and test scores were recorded in order to determine the optimum number of features.

6.3 Control group, all dwellings (*ca*)

Figure 7 shows the average profile for each of the 10 profile classes (PCs) for all dwelling types in the ***c**ontrol group (1786 customers). PCs *ca9*, *ca1*, and *ca3* (shown in blue) have a higher proportion of houses than units while PC *ca10* and *ca2* and (to a lesser extent) PC *ca4* (all shown in red), have a higher proportion of units. Table 2 shows some characteristics of the PCs. Those with more units include the cluster with the lowest overall energy use and are seen to have more pronounced peaks (morning and evening) in winter than summer, suggesting greater use of heating or electric cooking – but for shorter periods - than summer air conditioning in apartments. The PCs with more houses than units include PC *ca9* with flat daytime load all year (suggesting high air conditioning and heating use, with residents home during the daytime), PC *ca3* with high summer daytime load but lower in winter (possibly due to high air conditioning use in households with gas or non-existent heating), as well as PC *ca1* with high overnight load in both winter and summer.

A multinomial logistic regression was carried out using the assumed characteristics of the ***c**ontrol group household as explanatory variables and the MLR coefficients obtained are shown in Table 2. The highest coefficients are for ASSRTD_DWELLING_TYPE_CD_Unit showing a strong positive correlation with PCs *ca10, ca2* and *ca4*, as expected, some negative correlation with *ca9* and only weak or negligible negative correlation with *ca1 and ca3*. Bearing in mind that the clustered profiles exclude controlled loads (nearly all overnight water heating), the strong negative correlation between HAS_CONTROLLED_LOAD and the flat profile of PC *ca10* may be because these households have non-off-peak electric water heating, while the positive correlation with *ca4* (with high evening peaks) may be because these households with controlled (likely water heating) load are less likely to have gas and therefore more likely to use electricity for cooking. The regression gave a relatively low training score of 0.22 and a test score of 0.20. It is unsurprising that these assumed variables alone are inadequate to reliably predict the shape of a customers' annual load profile, but the score is twice that expected from chance for ten PCs.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

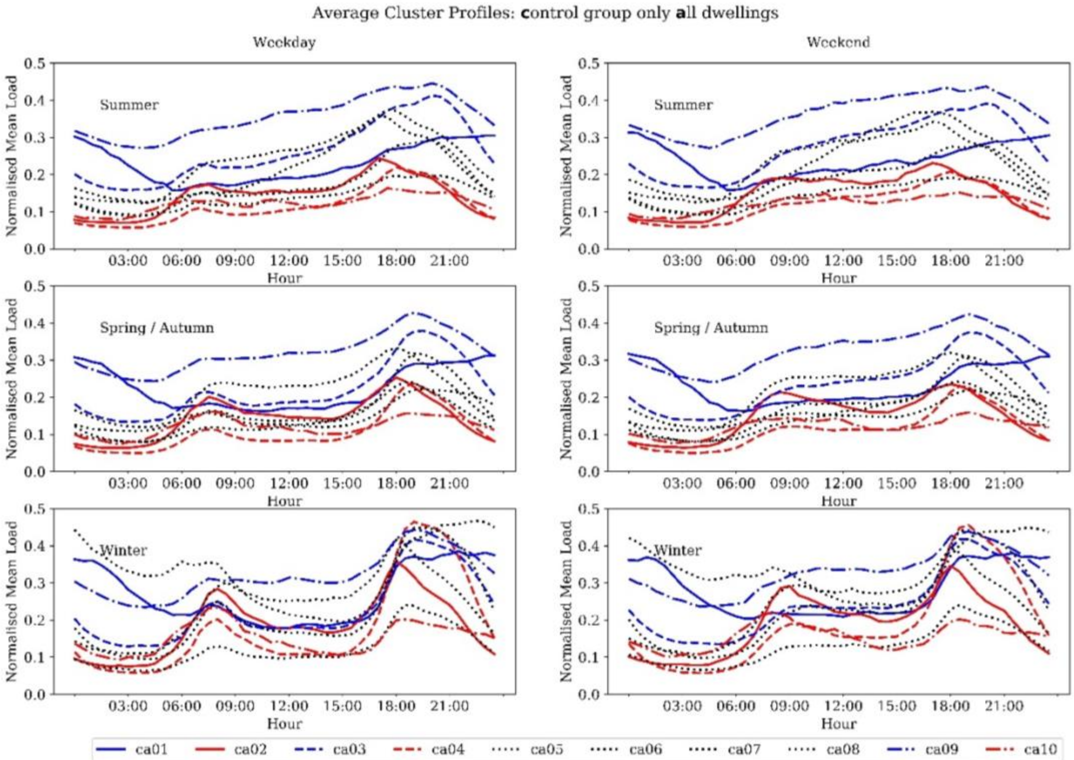


*Figure 7 – Average cluster profiles (ca1 to ca10) for **a**ll dwellings in **c**ontrol group*

*Table 2 Profile classes and classification coefficients for control group, all dwellings (**ca**)*

				MLR Coefficients						
Cluster (ca)	Cluster Size	Units as % of cluster	Average daily E_{tot} (kWh) (Summer / Winter)	ASSRTD_DWELLING_TYPE_CD_Unit	HAS_CONTROLLED_LOAD	ASSRTD_HHOLD_INCOME_GROUP_CD_LOW	ASSRTD_GAS_USAGE_GROUP_CD_LOW	ASSRTD_CLIMATE_ZONE_DESC_Mild temperate	ASSRTD_GAS_USAGE_GROUP_CD_HI	ASSRTD_HHOLD_INCOME_GROUP_CD_HI
ca1	106	20.8	18.4 / 22.9	0.00	-0.08	-0.33	0.24	0.00	0.01	0.00
ca2	199	50.3	12.8 / 17.5	0.89	0.00	0.67	0.00	0.22	-0.11	-0.08
ca3	211	21.3	20.0 / 19.1	-0.10	0.00	-0.05	-0.11	-0.11	0.00	0.14
ca4	258	34.9	11.1 / 17.6	0.84	0.58	-0.04	-0.47	-0.39	0.00	-0.01
ca5	130	27.7	13.5 / 31.9	0.00	0.00	0.00	0.26	0.18	-0.06	-0.24
ca6	289	22.8	15.0 / 20.5	0.06	0.00	-0.26	-0.07	-0.26	0.10	0.16
ca7	171	22.2	19.3 / 13.4	-0.07	0.14	0.00	0.04	-0.23	-0.64	0.00
ca8	190	22.6	19.4 / 22.3	-0.11	0.00	0.08	-0.05	0.00	0.35	0.00
ca9	132	20.5	18.6 / 19.1	-0.34	0.00	0.09	0.00	0.10	0.00	-0.30
ca10	100	59.0	8.0 / 10.9	0.95	-0.82	0.28	0.46	0.29	0.10	0.10
		Max absolute		0.95	0.82	0.67	0.47	0.39	0.64	0.30
		Mean absolute		0.39	0.25	0.23	0.22	0.21	0.20	0.12

6.4 Household characteristics of apartments (*3u*)

In order to better examine the range of load profiles exhibited by apartment households, the apartments from the larger dataset were clustered independently into ten clusters. Figure 8 shows the average daily profiles for these classes, while Table 3 shows the distribution of unit profiles between the clusters and selected household and demographic information. It should be noted that, although clustering was applied to load profiles for 2081 units, survey data was only available for 225 of these customers, including an atypically low proportion of renters (seven, compared to a national average of 60% of apartment residents [9]).

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

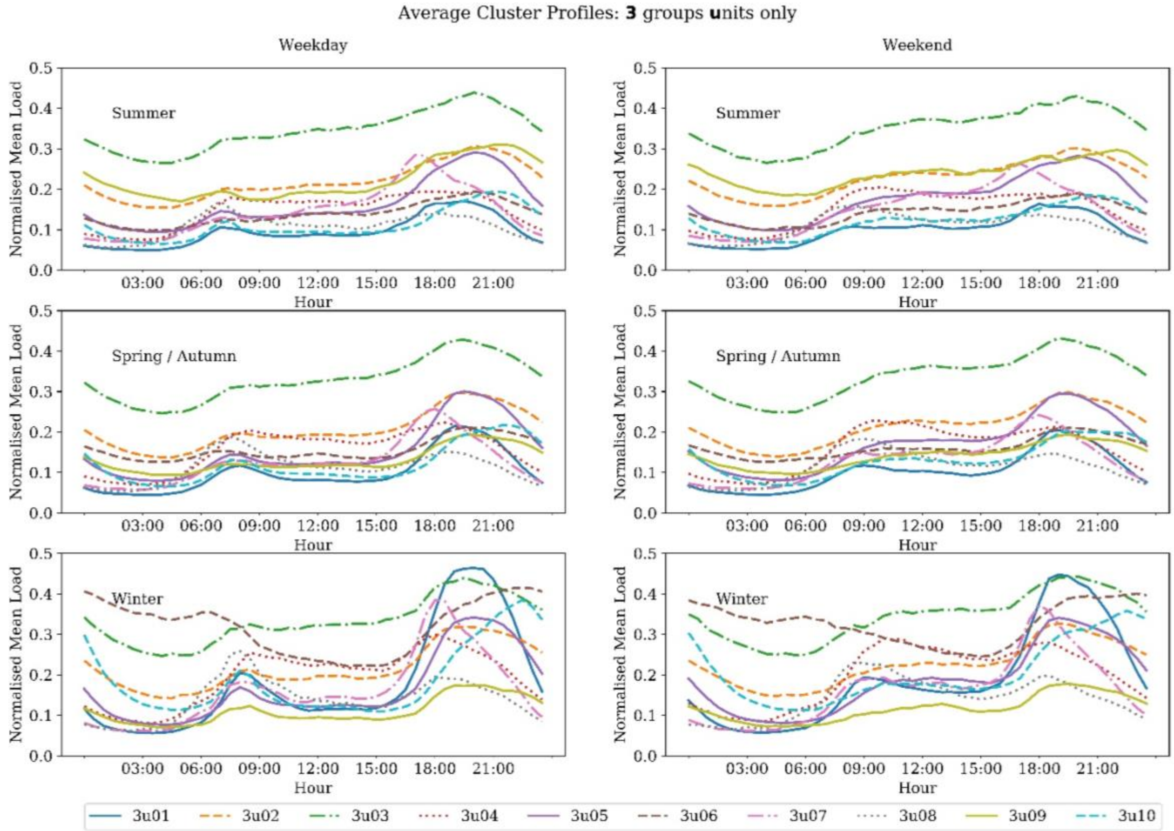


Figure 8 - Average profiles for 10 clusters of units from 3 customer groups

*Table 3 Household characteristics of profile classes for **3u***

Cluster (3u)	% of Dataset	Average daily total energy (kWh) (summer / winter)	Mean time period of peak	Average Occupancy	% with residents over 70	% with children	Average Income 1=low, 2=med, 3=high	% with aircon	% with Gas	% with gas hot water	% with gas cooking	% with gas heating	% renting
3u1	13%	6.9 / 13.1	18:30	1.7	46%	9%	2.3	49%	49%	23%	40%	14%	0%
3u2	12%	9.3 / 9.7	18:30	2.0	4%	16%	2.4	20%	72%	36%	68%	20%	12%
3u3	4%	12.0 / 13.0	18:00	1.7	27%	0%	1.5	9%	82%	55%	73%	9%	9%
3u4	12%	10.9 / 15.3	14:00	1.3	33%	10%	1.7	36%	36%	23%	27%	5%	0%
3u5	16%	9.1 / 9.7	19:30	1.8	3%	9%	2.7	28%	87%	31%	69%	38%	0%
3u6	6%	9.3 / 21.6	16:00	2.0	0%	0%	2.2	33%	67%	44%	67%	22%	11%
3u7	11%	10.0 / 11.8	17:30	1.6	57%	3%	1.5	63%	40%	33%	20%	13%	0%
3u8	11%	8.0 / 10.8	10:30	1.5	34%	9%	1.6	26%	11%	6%	6%	3%	2%
3u9	5%	14.3 / 7.3	18:30	2.6	11%	0%	2.4	67%	78%	33%	67%	11%	11%
3u10	9%	7.4 / 12.7	19:00	1.8	22%	20%	1.8	50%	60%	20%	30%	0%	0%

Figure 9 shows the average MLR accuracy scores achieved as recursive feature elimination was applied to the survey data for these units. The low test score compared to training scores, even for a low number of features, suggests that the regression model is overfitting to the small, sparse dataset. However, some characteristics of the profile classes can be deduced from their shapes and from the data presented Table 3.

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

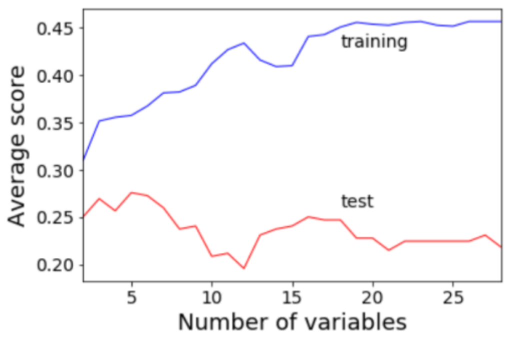


*Figure 9 – Accuracy scores vs number of explanatory variables for dataset **3u***

Almost half of the households in ***3u1*** have residents over 70, and this is reflected by flat profiles except for high evening heating peaks in winter. ***3u2*** is also flat, with low evening peaks and similar loads summer and winter, likely due to the high proportion of households with gas, particularly used for cooking. The smallest class, ***3u3***, has a moderately high 24-hour load with a low early-evening peak. ***3u4*** has high, flat daytime load, particularly in winter, with very low peaks, having a low % of households with gas heating. Households in ***3u5*** and ***3u6*** contain relatively affluent couples with a low proportion of children and aged people. In ***3u5***, this is reflected by the later evening peak, while ***3u6*** households have very high overnight heating loads in winter, resulting in a very early average peak time. Over half the households in ***3u7*** contain elderly residents and their profiles show an early evening peak that is more pronounced and narrower in winter. ***3u8*** is the only class with a morning peak, which is most pronounced in winter and may be explained by the very low incidence of gas heating. ***3u9*** has the largest average households, though without children, and is the only class to have higher summer loads than winter, while the high proportion of households having gas cooking makes the profile relatively flat. Finally, ***3u10*** has relatively flat profiles in summer but high overnight heating loads in winter.

7. Discussion and conclusions

Excluding common property loads, the average electricity demand per occupant is lower for apartment residents than for houses, although, because of the absence of data on the size of the dwellings in the study, it is not clear whether this is simply due to the typically smaller floor area of apartments or to other features such as shared walls (hence improved thermal performance) or lower levels of appliance ownership or use. It would be useful in future studies to collect more physical data pertaining to the dwellings, including floor area and number and orientation of external walls.

Although the average daily profile of all the apartments is flatter than that for houses, the diurnal variability of individual apartment loads is, on average, higher than that for houses. Their wider distribution of peak time and of daily variability also suggests greater diversity between apartment load profiles than between those of houses. For all dwelling types, aggregating loads in groups of up to 50 households increases load factor and reduces variability, but this effect is more pronounced for apartments due to the greater diversity of their load profiles. While dwelling type makes no discernible difference to daily load variability for larger aggregations of households, load factor calculated on an annual basis is higher for aggregated apartment loads than for houses, and the coincidence factor of aggregated apartment loads is lower.

Based solely on the shape of the load profile, aggregation in groups of up to ten households may increase self-consumption of PV, but, perhaps surprisingly, no further benefit is shown for larger aggregations.

Although the study does not reveal distinct apartment or house load profiles, the application of cluster analysis to annual load profiles does show that there are groups of profiles that are more likely to belong to apartment households and those more likely to be from houses. Indeed, dwelling type is shown to be one of the more important indicators of load profile, along with elderly occupants or having a gas cooker or a pool pump.

Although the SGSC dataset has a high incidence of apartments, both proportionally and in absolute quantity, compared with other published studies, the significance of the classification analysis is affected by the low number of apartment residents who participated in the household survey, although it is possible that more complex classification methods (such as ensemble decision tree or support vector machine), or the use of k-fold cross-validation would yield more conclusive results. A larger dataset, with household survey data for all apartments, would enable a more detailed analysis of the relationship between load profiles and household characteristics and could usefully facilitate an examination of the effects on load shape and variability of

This is a pre-print of the published article Mike B. Roberts, Navid Haghdadi, Anna Bruce, Iain MacGill, Characterisation of Australian apartment electricity demand and its implications for low-carbon cities, Energy, 2019, DOI: 10.1016/j.energy.2019.04.222

aggregating loads within and between clusters to explore the benefits of household diversity on aggregated loads. The potential applications of this research include design of distributed PV and battery storage for apartment buildings and distribution network planning, whether installation of new infrastructure for greenfield residential developments or augmentation of existing assets in urban areas where housing density is increasing. More generally, our findings highlight some of the opportunities as well as challenges that a focus on higher density urban form through apartment living pose for future low-carbon cities.

Acknowledgements

This research is supported by a project grant (AP841) from Energy Consumers Australia. The lead author gratefully acknowledges the support of a Research Training Program Stipend from the Australian Government and of a scholarship funded by the CRC for Low Carbon Living. The authors would like to thank Dr Tess Stafford for her insights on the SGSC dataset.

Declarations of Interest: None.

References

1. United Nations Framework Convention on Climate Change (UN FCCC), *FCCC/CP/2015/L.9/Rev.1 - Adoption of the Paris Agreement*. 2015.
2. US Energy Information Administration. *International Energy Outlook*. 2016 18/11/2018]; Available from: https://www.eia.gov/outlooks/aeo/data/browser/#/?id=15-IEO2016&region=4-0&cases=Reference&start=2010&end=2040&f=A&linechart=Reference-d021916a.2-15-IEO2016.4-0~~~~~~~~~Reference-d021916a.34-15-IEO2016.4-0~Reference-d021916a.35-15-IEO2016.4-0~Reference-d021916a.36-15-IEO2016.4-0~Reference-d021916a.37-15-IEO2016.4-0~Reference-d021916a.38-15-IEO2016.4-0~Reference-d021916a.39-15-IEO2016.4-0~Reference-d021916a.40-15-IEO2016.4-0~Reference-d021916a.41-15-IEO2016.4-0&map=&sourcekey=0.
3. IEA, *World Energy Outlook*. 2016, International Energy Agency.
4. City of Sydney. *Sustainable Sydney 2030: Carbon Reduction*. 2017 27/2/2018]; Available from: http://www.cityofsydney.nsw.gov.au/vision/towards-2030/sustainability/carbon-reduction.
5. City of Melbourne. *Net Zero Emissions Strategy*. 2018 27/2/2018]; Available from: http://www.melbourne.vic.gov.au/about-council/vision-goals/eco-city/Pages/zero-net-emissions-strategy.aspx.
6. Australia, G.o.S. *Carbon Neutral Adelaide*. 2015 27/2/2018]; Available from: http://www.climatechange.sa.gov.au/carbon-neutral-adelaide.
7. IEA, *Energy Efficiency Indicators - Highlights* 2016, International Energy Agency.
8. DEWHA, *Energy Use in the Australian Residential Sector 1986-2020*. 2008.
9. ABS, *Census of Population and Housing*. 2016, Australian Bureau of Statistics.
10. ABS, *Monthly building approvals statistics*. 2015, Australian Bureau of Statistics.
11. Guneralp, B., et al., *Global scenarios of urban density and its impacts on building energy use through 2050.* Proc Natl Acad Sci U S A, 2017.
12. Pears, A., *Does Higher Density Really Reduce Household Energy Requirements? It Depends ….* Urban Policy and Research, 2005. **23**(3): p. 367-369.
13. Perkins, A., et al., *Transport, Housing and Urban Form: The Life Cycle Energy Consumption and Emissions of City Centre Apartments Compared with Suburban Dwellings.* Urban Policy and Research, 2009. **27**(4): p. 377-396.
14. Rickwood, P., G. Glazebrook, and G. Searle, *Urban Structure and Energy—A Review.* Urban Policy and Research, 2008. **26**(1): p. 57-81.
15. Roberts, M.B., et al., *Using PV to help meet common property energy demand in residential apartment buildings*, in *Australian Summer Study in Energy Productivity*. 2016: Sydney.
16. Victoria State Government. *Smart meters: End of Rollout*. 2015 28/3/2015 31/8/2017]; Available from: http://www.smartmeters.vic.gov.au/about-smart-meters/end-of-rollout.
17. Dickers, J. *Australia's first market-led smart meter rollout*. Utility Magazine, 2016.
18. Australian Government Productivity Commission, *Data availability and use inquiry report.* 2017: Canberra.
19. AusGrid, et al., *Smart Grid, Smart City: Shaping Australia's Energy Future - Executive Report*. 2014.
20. Yildiz, B., et al., *Recent advances in the analysis of residential electricity consumption and applications of smart meter data.* Applied Energy, 2017.

21. Swan, L.G. and V.I. Ugursal, *Modeling of end-use energy consumption in the residential sector: A review of modeling techniques.* Renewable and Sustainable Energy Reviews, 2009. **13**(8): p. 1819-1835.
22. Grandjean, A., J. Adnot, and G. Binet, *A review and an analysis of the residential electric load curve models.* Renewable and Sustainable Energy Reviews, 2012. **16**(9): p. 6539-6565.
23. Kavgic, M., et al., *A review of bottom-up building stock models for energy consumption in the residential sector.* Building and Environment, 2010. **45**(7): p. 1683-1697.
24. Zhao, H.-x. and F. Magoulès, *A review on the prediction of building energy consumption.* Renewable and Sustainable Energy Reviews, 2012. **16**(6): p. 3586-3592.
25. Foucquier, A., et al., *State of the art in building modelling and energy performances prediction: A review.* Renewable and Sustainable Energy Reviews, 2013. **23**: p. 272-288.
26. Wiesmann, D., et al., *Residential electricity consumption in Portugal: Findings from top-down and bottom-up models.* Energy Policy, 2011. **39**(5): p. 2772-2779.
27. McLoughlin, F., A. Duffy, and M. Conlon, *Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study.* Energy and Buildings, 2012. **48**: p. 240-248.
28. Sanquist, T.F., et al., *Lifestyle factors in U.S. residential electricity consumption.* Energy Policy, 2012. **42**: p. 354-364.
29. Fan, H., I.F. MacGill, and A.B. Sproul, *Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia.* Energy and Buildings, 2015. **105**: p. 9-25.
30. Wallis, H., M. Nachreiner, and E. Matthies, *Adolescents and electricity consumption; Investigating sociodemographic, economic, and behavioural influences on electricity consumption in households.* Energy Policy, 2016. **94**: p. 224-234.
31. IPART, *Determinants of residential energy and water consumption in Sydney and surrounds*. 2011.
32. IPART, *Residential energy and water use in Sydney, the Blue Mountains and Illawarra: Results from the 2010 household survey*. 2010, Independent Pricing and Regulatory Tribunal.
33. IPART, *Residential energy and water use in Sydney, the Blue Mountains and Illawarra*. 2006, Independent Pricing and Regulatory Tribunal.
34. IPART, *Residential energy use in Sydney, the Blue Mountains and Illawarra*. 2004.
35. Myors, P., R. O'Leary, and R. Helstroom, *Multi unit residential buildings energy and peak demand study*. 2005, Energy Australia / NSW Dept of Infrastructure, Planning & Natural Resources.
36. Cetin, K.S., P.C. Tabares-Velasco, and A. Novoselac, *Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use.* Energy and Buildings, 2014. **84**: p. 716-726.
37. Firth, S., et al., *Identifying trends in the use of domestic appliances from household electricity consumption measurements.* Energy and Buildings, 2008. **40**(5): p. 926-936.
38. Smith, R., et al., *Demand response: a strategy to address residential air-conditioning peak load in Australia.* Journal of Modern Power Systems and Clean Energy, 2013. **1**(3): p. 223-230.
39. Fan, H., I.F. MacGill, and A.B. Sproul, *Statistical analysis of drivers of residential peak electricity demand.* Energy and Buildings, 2017. **141**: p. 205-217.
40. Wang, Y., et al., *Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges.* IEEE Transactions on Smart Grid, 2018.
41. Lenzen, M., C. Dey, and B. Foran, *Energy requirements of Sydney households.* Ecological Economics, 2004. **49**(3): p. 375-399.
42. Boulaire, F., et al., *Statistical modelling of district-level residential electricity use in NSW, Australia.* Sustainability Science, 2013. **9**(1): p. 77-88.
43. Fan, H., I.F. MacGill, and A.B. Sproul, *Impact of feedback interventions on residential electricity demand in Australia's first large scale smart grid project.* 2015: p. 1-6.
44. Motlagh, O., G. Foliente, and G. Grozev, *Knowledge-Mining the Australian Smart Grid Smart City Data: A Statistical-Neural Approach to Demand-Response Analysis*, in *Planning Support Systems and Smart Cities*, S. Geertman, et al., Editors. 2015, Springer International Publishing: Cham. p. 189-207.
45. Chicco, G., *Overview and performance assessment of the clustering methods for electrical load pattern grouping.* Energy, 2012. **42**(1): p. 68-80.
46. Kwac, J., et al. *Utility Customer Segmentation Based on Smart Meter Data: Empirical Study*. in *IEEE SmartGridComm 2013 Symposium - Support for Storage, Renewable Resources and Micro-grids*. 2013.

47. McLoughlin, F., A. Duffy, and M. Conlon, *A clustering approach to domestic electricity load profile characterisation using smart metering data.* Applied Energy, 2015. **141**: p. 190-199.
48. Rhodes, J.D., et al., *Clustering analysis of residential electricity demand profiles.* Applied Energy, 2014. **135**: p. 461-471.
49. Viegas, J.L., et al., *Classification of new electricity customers based on surveys and smart metering data.* Energy, 2016. **107**: p. 804-817.
50. Berry, A., et al., *Energex customer load profile market segmentation and clustering - Final Report*. 2015, CSIRO.
51. Yildiz, B., et al., *A method for classifying households to help forecasting their photovoltaic electricity self-consumption patterns*, in *Asia Pacific Solar Research Conference*. 2017: Melbourne.
52. Jones, R.V., A. Fuertes, and K.J. Lomas, *The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings.* Renewable and Sustainable Energy Reviews, 2015. **43**: p. 901-917.
53. Hamilton, I.G., et al., *Energy efficiency in the British housing stock: Energy demand and the Homes Energy Efficiency Database.* Energy Policy, 2013. **60**: p. 462-480.
54. Guerra Santin, O., L. Itard, and H. Visscher, *The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock.* Energy and Buildings, 2009. **41**(11): p. 1223-1232.
55. Wyatt, P., *A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England.* Energy Policy, 2013. **60**: p. 540-549.
56. Gram-Hanssen, K., C. Kofod, and K.N. Petersen, *Different everyday lives: different patterns of electricity use*, in *American Council for an Energy Efficient Economy 2004 Summer Study*. 2004.
57. Kavousian, A., R. Rajagopal, and M. Fischer, *Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior.* Energy, 2013. **55**: p. 184-194.
58. Yohanis, Y.G., et al., *Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use.* Energy and Buildings, 2008. **40**(6): p. 1053-1059.
59. Brounen, D., N. Kok, and J.M. Quigley, *Residential energy use and conservation: Economics and demographics.* European Economic Review, 2012. **56**(5): p. 931-945.
60. Leahy, E. and S. Lyons, *Energy use and appliance ownership in Ireland.* Energy Policy, 2010. **38**(8): p. 4265-4279.
61. Hayn, M., V. Bertsch, and W. Fichtner, *Electricity load profiles in Europe: The importance of household segmentation.* Energy Research & Social Science, 2014. **3**: p. 30-45.
62. Heinonen, J. and S. Junnila, *Residential energy consumption patterns and the overall housing energy requirements of urban and rural households in Finland.* Energy and Buildings, 2014. **76**: p. 295-303.
63. McLoughlin, F., A. Duffy, and M. Conlon, *Evaluation of time series techniques to characterise domestic electricity demand.* Energy, 2013. **50**: p. 120-130.
64. Beckel, C., et al., *Revealing household characteristics from smart meter data.* Energy, 2014. **78**: p. 397-410.
65. Rickwood, P., *Residential Operational Energy Use.* Urban Policy and Research, 2009. **27**(2): p. 137-155.
66. Western Power. *Maximum demand calculator*. 5/09/2017]; Available from: https://westernpower.com.au/technical-information/calculators-tools/maximum-demand-calculator/.
67. Endeavour Energy, *Mains design instruction : Method of calculating voltage drop in low voltage mains*. 2014.
68. Australian Government Department of the Environment and Energy, *Smart Grid Smart City Customer Trial Data*. 2014.
69. Ausgrid, *Modelling Inputs Compendium: SGSC Technical Compendium*. 2014, Smart Grid Smart City.
70. Ausgrid, *Smart Metering Infrastructure: SGSC Technical Compendium*. 2014, Smart Grid Smart City.
71. Ausgrid, *Customer Applications: SGSC Technical Compendium*. 2014, Smart Grid Smart City.
72. Futura and Taylor Fry, *Smart Grid Smart City (SGSC) Detailed Sampling Design: Final Report for Ausgrid*. 2011.
73. Equipment Energy Efficiency (E3) Committee, *Standby Power – Current Status*. 2006.

74. Ruxton, G.D., *The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test.* Behavioral Ecology, 2006. **17**(4): p. 688-690.

75. Zimmerman, D.W. and B.D. Zumbo, *Rank Transformations and the Power of the Student t-test and Welch t-test for Non-Normal Populations With Unequal Variances.* Canadian Journal of Experimental Psycology, 1993. **47**(3): p. 523-539.

76. Salom, J., et al., *Understanding net zero energy buildings: Evaluation of load matching and grid interaction indicato*, in *12th Conference of International Building Performance Simulation Association*. 2011: Sydney.

77. Luthander, R., et al., *Photovoltaic self-consumption in buildings: A review.* Applied Energy, 2015. **142**: p. 80-94.

78. Cao, S., *Matching analysis for onsite building energy systems involving energy conversion, storage and hybrid grid connections*, in *Department of Energy Technology, Building Energy Efficiency and HVAC Technology*. 2014, Aalto University: Finland.

79. Gilman, P. and A. Dobos, *System Advisor Model, SAM 2011.12.2: General Description*. 2012, NREL.

80. Bureau of Meteorology, *Australian Hourly Solar Irradiance Gridded Data* 1990-2016, BOM.

81. Agresti, A., *Categorical Data Analysis*. 2013, Somerset, UNITED STATES: John Wiley & Sons, Incorporated.

82. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.

83. sci-kit learn developers. *Sci-Kit Learn User Guide: 1.1.11. Logistic regression*. 2007 - 2017 23/12/2018]; Available from: https://scikit-learn.org/stable/modules/linear_model.html.