Customer-Led Network Revolution

Insight Report: Baseline Domestic Profile

Test Cell 1a Customer Subgroup Analysis

DOCUMENT NUMBER
CLNR-L216

AUTHORS
Christian Barteczko-Hibbert, Gavin Whitaker, Durham University
Shane Slater, Joan Groizard, Element Energy

ISSUE DATE
23/02/2015

Contents

1. Executive Summary

Changing electricity demand, the electrification of the transport and heating sectors and the increase in distributed renewable energy sources, all present challenges to distribution networks. The Customer Led Network Revolution project aims to improve our understanding of current and future electricity use patterns of domestic and commercial customers. Data was collected from customers divided into different test cells (TCs) or samples, each with a particular combination of metering type, electricity tariff structure and/or low carbon technology.

TC1a contains half-hourly whole house electricity consumption data from October 2012 to September 2013 for 9201 domestic customers with basic smart metering. This data has been analysed to provide insight into typical domestic electricity use patterns and the relationship between demographic indicators and energy use. Additionally, TC1a is used as the control group or starting point against which the other test cells can be compared. The demographic composition of the participants in this test cell is representative of the UK population.

The dataset displays some expected behaviour:

- The daily energy use profile is broadly divided into night/early morning (low demand), daytime (mid-level demand) and evening (higher demand).
- In the winter months, overall consumption and average daily peak demand are higher than in other seasons. Also in winter there is a greater variability in energy use between customers.
- Most households have their peak daily demand during the 4-8pm period. However, there is a significant minority of households across all demographic groups whose peak demand occurs earlier in the day, around midday.

It is important to note when considering the population statistics below that variability in electricity consumption has been shown to be significant even within any specific group. This means that general trends reported below are not a prediction of the energy consumption of an individual household. However, this level of diversity in electricity consumption is a positive finding for networks as the heterogeneity of the customer base connected to a given network should produce a lower coincident peak.

Household income is the demographic attribute with the clearest impact on energy use, although the correlation is at best weak. On average, high income households (above £30k p.a.) have the highest overall energy consumption and highest average peak power demands, which is in line with broader industry understanding of customer profiles. On average, a high income household was found to consume:

- 40% more energy than a low (<£15k p.a.) income household.
- 15% more energy than a medium income (£15k-£30k) household.

In TC1a, over 34% of the energy was consumed by high income households despite this group forming only 29% of the sample. Additionally, for the high income group, the proportion of electricity use concentrated in the evening peak period (4pm-8pm) is higher than for any other group.

Noting that other correlations are weaker, the following demographic indicators were also found, on average, to be relevant when determining energy consumption:

- The existence of dependents (people aged below 5 or over 65) in the household was found to be correlated with lower overall energy consumption and peak demand.
- Home ownership (rather than renting) was found to be positively correlated with increasing energy consumption and peak demand.
- Living in a rural area was observed to correspond with higher energy use and higher demand peak compared to urban households.

However, it should be noted that the last two attributes above may largely be proxies/secondary to household income, because home ownership and rurality are arguably positively correlated with household income. Additionally, the analysis did not investigate whether the link between income and electricity use was, in turn, related to other characteristics such as household size or behaviour, and of course the trends identified do not necessarily hold for every individual household.

Analysis shows that household thermal efficiency, estimated by using building age, was not strongly correlated with electrical peak demand or overall consumption. However, this analysis does not include gas heating, which is likely to be more closely linked to thermal efficiency of the house.

A categorisation by Experian Mosaic did not generate additional significant insight apart from a confirmation of the importance of household income. Mosaic groups with higher income were shown to be the ones with higher annual electricity consumption. This outcome highlights the continuing challenge of finding reliable indicators of energy use.

While electricity use was found to be linked to demographic indicators, other variables such as ambient temperature and time of the year have a much greater impact on electricity use, in particular when predicting the peak demand on any particular day.

As a comprehensive dataset of residential half-hourly electricity use, TC1a can be used for the following:

- The demand profiles can be compared with and potentially update standard consumption profiles currently used in industry for network planning. More information on this is given in [1] and [2]
- The demographic breakdown of the data could be useful when targeting the deployment of future interventions, for example to identify the customers most responsible for a particular network challenge (such as peak demand), or to understand the distributional impact of energy price tariffs.

2. Introduction

The Customer-Led Network Revolution (CLNR) aims to understand current, emerging and possible future customer energy characteristics, to allow for a more optimal planning of the energy system in the context of increasing electrical demand and deployment of low-carbon forms of energy generation. For this purpose, large numbers of customers were divided into different samples or test cells (TCs), each with a particular combination of metering type, energy tariff and/or low carbon technologies. These tariffs or technologies, referred to as "interventions", are designed to modify customers' energy use characteristics, either directly or through changes in behaviour.

This document details the final analysis of Test Cell 1a and adds to previous work which explored some of the early findings of the CLNR trials.

Test Cell 1a collected electricity usage statistics from over 9000 households across different demographic groups and creates an overall picture of current domestic electrical consumption in the UK. No interventions were applied to TC1a, allowing this to be used as the control group or baseline against which the impacts of interventions (such as low carbon technology or Time of Use tariffs) applied to other domestic test cells can be compared.

The demographic breakdown allowed us to investigate links between different demographic indicators and energy consumption patterns, with a view to validating customer profiles currently used in industry as well as providing a baseline to understand any demographic-specific impacts of the interventions trialled in CLNR.

This report describes the dataset used in TC1a, provides baseline energy consumption characteristics for the different demographic groups and looks at the system peak demand on the days of greatest network stress. The load profiles presented will be of interest to distribution network engineers and designers, as well as DNO operations as a whole, academic bodies and the wider electricity industry. Specifically, the information presented here will be used to direct further work on developing profiles before and after interventions, with a view to updating network design tools.

3. Sample

Test Cell 1a consists of a maximum of 9201 households for which demographic and electricity consumption data were collected. This excluded Economy 7 and other specific high electrical loading customers, and no interventions (other than installation of smart metering) were carried out on these households.

3.1. Metered data

Smart meters were installed prior to commencement of the trial for these customers which recorded electrical energy consumption to a 30 minute resolution (so providing 48 readings per day). The average power demand (in kW) in each half-hourly interval can be calculated from this. Additionally, daily cumulative kWh totals were used to compute customers' annual consumption.

The metering data was collected and analysed for the one year period from October 2012 to September 2013 (inclusive) to be as consistent as possible with the data available for other CLNR studies running in parallel.

Throughout this period, the number of households from which data is available at any one time fluctuates, for example due to missing readings. In any analysis all zero readings were assumed incorrect and thus were removed from the computation, as it was deemed unfeasible that a house would have a zero load for an entire half-hour period unless all appliances were switched off or power outages occurred. These later couple of points would not be classified as normal occurrences. Figure 1 depicts average customer data availability on a weekly basis computed from daily customer numbers. Barring any missing readings at any time instance these customer numbers are by and large a fair representation of the sample size in our calculations. Where relevant, the demographic proportions, which will be significantly less than the global sample size, are noted throughout the report.

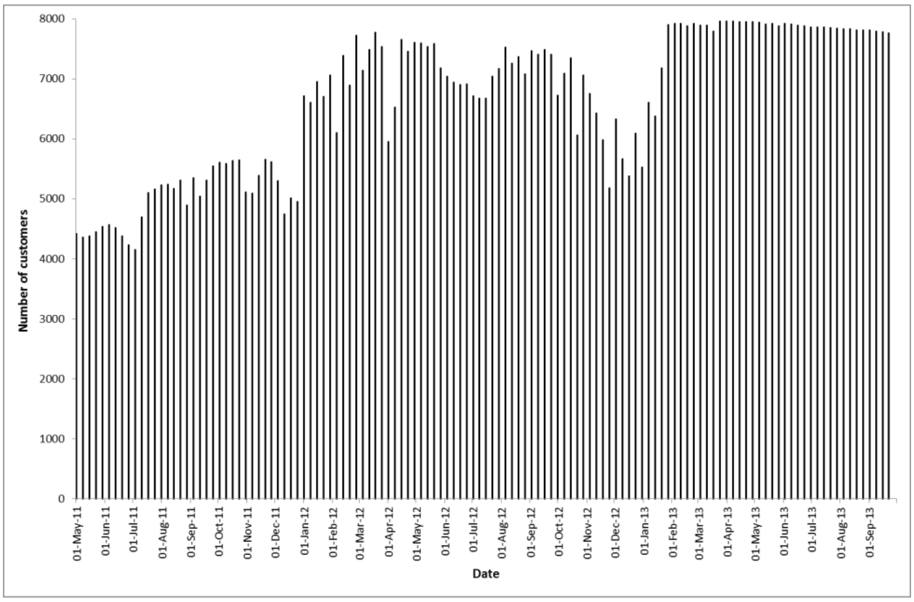


Figure 1: Average customer number participation on a weekly basis from 1st May 2011 - 30 September 2013. Data analysed October 2012 to September 2013.

3.2. Demographic variables

The demographic data collected were used to create two different socio-demographic groupings for TC1a customers: An internally defined set of demographic indicators ("CLNR categorical data", see [3]) and the externally-provided Mosaic categorisation [4].

The CLNR categorisation uses 5 indicators:

1. **Dependents:** does the house have dependents (small children or elderly relatives) or not? (2 categories)
2. **House efficiency:** how efficient is the building? Building age was used as an approximation for this. (3 categories)
3. **Rurality:** in what geographical location is the customer located? (4 categories)
4. **Tenancy:** is the house lived in by renters or the owners? (2 categories)
5. **Income:** what is the household income level? (3 categories)

This leads to a total of 144 possible combinations of socio-demographic categories. For the purposes of benchmarking, the population will be considered according to their labelling in only one category at a time (so a total of 14 categories rather than 144). This provides a more manageable set of population categorisations. **Table 1** provides a detailed picture of the CLNR categorical data for TC1.

Label	Description	Values	Source
Dependents	The presence of under-5s or over-65s in the property.	With <5 or >65	British Gas
		Without <5 or >65	
Energy Efficiency	Relative (proxy of building age)	Low (built before 1919)	British Gas
		Medium (built 1919-1976)	
		High (built after 1976)	
Rurality	Description of the location of the property, roughly corresponding to ONS Rural-Urban classification	Urban	British Gas
		Suburban	
		Rural	
		Rural Off-gas	
Tenancy	Tenancy status	Owner-Occupier	British Gas
		Renter	
Income	Banded	Low (<£15k p.a.)	British Gas
		Med (£15k to £30k p.a.)	
		High (>£30k p.a.)	

Table 1: Categorical stratification data for TC1a

These categories are the Test Cell design sample stratification categories as defined in [3] to provide a good statistical basis for the electrical use for the different demographic groups. This categorical data is used in clustering and analysis of variance exercises (section 5.3).

The second socio-demographic grouping was externally provided by Experian. This 'Mosaic' framework allocated every individual customer to one of 15 possible categories based on various aspects such as census data, credit scores and other surveys.

The Mosaic classification method is not public, and therefore can only be taken as given. This demographic classification method is often used by industry, and is used here as a comparison with the classifications defined by this project. The sample breakdown by Mosaic classifications is shown below in **Table 2**, which also compares this against the UK national average.

Mosaic Category	TC1a Composition	UK National Average [4]
A Alpha Territory	2.08%	3.00%
B Professional Rewards	9.08%	8.00%
C Rural Solitude	2.43%	4.00%
D Small Town Diversity	12.11%	9.00%
E Active Retirement	4.45%	4.00%
F Suburban Mindsets	11.87%	11.00%
G Careers and Kids	4.32%	6.00%
H New Homemakers	2.20%	5.00%
I Ex-Council Community	14.27%	9.00%
J Claimant Cultures	6.34%	5.00%
K Upper Floor Living	1.14%	6.00%
L Elderly Needs	8.13%	6.00%
M Industrial Heritage	10.09%	8.00%
N Terraced Melting Pot	6.63%	8.00%
O Liberal Opinions	2.87%	9.00%

Table 2: MOSAIC composition of TC1a and UK National Average

For both CLNR and Mosaic categorisations, graphs in the analysis have sufficient customer numbers to satisfy sensible confidence intervals (Table 24), with the exception of “rural off gas” customers in the CLNR demographics. For Mosaic groups, the smallest sample size was 70 (for Mosaic group ‘K’).

Additionally,
Table 9 in the appendices gives a breakdown of the age demographic of 3040 British Gas customers used in this analysis. Although this does not include all customers within our dataset, the customers are representative where data was obtained from the meter supplier. The largest customer group is the age range 61-70. There were very few customers in the age range 20-30 which could be attributable to a different set of customers who switch electricity suppliers and have not grown up in an era dominated primarily by one energy supplier.

4. Variables compared

To generate the baseline from the raw meter data, a number of statistics were designed which can be thought of as the 'dependent' variables. These measure various characteristics that are relevant to network operators and generators:

- **Absolute energy consumption**: how much energy has been consumed over a given period of time (measured in kWh).
- **Peak power demand**: the maximum power that was demanded by a group of customers within a specified time frame (measured in kW and reported with the period and time this peak occurred).
- **Variation in energy consumption**: this measures how different a set of customers' consumption is in relation to the different socio demographic groups and within the same subgroup. The results imply how homogenous a group is in terms of its energy consumption or how much potential there is to move consumption.
- **Variation in peak power demand**: this measures how varied the peak demand is within a group of customers, and supports network planning purposes to compute the upper peak demand from a set of customers and therefore how much supply capacity is required.

The themes above are considered across varying time frames, namely annual consumption, monthly demands by weekday and weekend for various demographics and groups. From these the terms defined in Table 3 below are derived, allowing to construct a picture of the UK electrical domestic consumption taking into account quantitative and qualitative data. The parameters in Table 3 are common with other CLNR test cells and therefore allow comparison between test cells.

Additionally, two further elements were investigated:

Demographic drivers
An ANOVA analysis was performed on the full TC1a population to test for significant correlations between CLNR demographic factors and consumption.

A load data driven clustering algorithm was then used to investigate the naturally different daily demand patterns clusters. The clustering approach was applied on the 'raw' demand dataset to test for clusters based on absolute demand, as well as on normalised data to test for clusters of different demand shapes. The clusters were then analysed based on their respective demographic constituents as an alternative means for investigating demographic drivers for load demand.

Analysis of system peak demand
The peak load is a critical parameter in the design of electrical networks. The contribution of the different demographic groups to the system peak on the day of greatest network stress for TC1a was inspected, as well as the impact of other factors to investigate how system peak may be best predicted.

Finally, a correlation study was carried out to highlight any differences between the peak demand patterns based on total consumption. As total consumption data is more readily available, understanding the relationship with peak demand may assist network design.

Term	Description	Mathematical formulation (if applicable)
1. Peak day	The day on which the maximum of the mean demand occurred during a specific time period	$\bar{D}_{max,t} = max\left(\frac{1}{N}\sum_{i=1}^{N} D_{i,t}\right)$
2. Energy consumption	Total energy consumed for a given customer over a specific time period ranging from t_0 to T	$E_i = E_{i,T} - E_{i,t_0}$ Or $E_i = \sum_{t \in S} E_{i,t}$ Where S is the time period considered
3. Peak power demand	Peak power for customer i in the time period given by S	$\hat{P}_i = \max_{j \in S}(P_{i,j})$ Where $P_{i,j}$ are the individual power measurements for customer i
4. Mean peak	The average peak power demand for customer i in the time period S over the number of days M	$\bar{P}_i = \frac{1}{M}\sum_{k=1}^{M}\sum_{j \in S}\max(P_{i,j,k})$
5. Max peak	The maximum peak power demand for customer i in the time period S over the number of days M	$\tilde{P}_i = \max_{j \in S, k \in M}(P_{i,j,k})$
6. Mean mean peak	The mean of the average peak power demand for customer i in the time period S over the number of days M	$\bar{\bar{P}}_i = \frac{1}{N}\bar{P}_i$
7. Mean maximum peak	The mean of the maximum peak power demand for customer i in the time period S over the number of days M	$\bar{\tilde{P}}_i = \frac{1}{N}\tilde{P}_i$
8. Peak time	Peak time H for customer i in the time period S over the number of days M	$H_{i,j,k} = \begin{cases} 1 & if\ P_{i,j,k} = \max_{j \in [S], k \in [M]}(P_{i,j,k}) \\ 0 & otherwise \end{cases}$
9. Modal peak time	The modal value of the time of peak demand accounting for all customers	$\hat{H} = \max_{j \in [S], k \in [M]}(H_{i,j,k})$

Table 3: Mathematical descriptions and definitions

5. Analysis

5.1. Annual Electricity Consumption and Peak Demand

5.1.1. General Overview

Annual consumption gives insights into likely demographic customer energy usage. It is thought that certain customers' lifestyles, because of their daily patterns, family dependencies and wealth will naturally have a different set of needs regarding electricity use. Figure 2 shows annual consumption density plots for TC1a by CLNR demographic classification (dependents, tenancy, income, efficiency and rurality). The data associated with the figure is shown in Table 24.

Low income customers have the lowest annual electrical energy consumption across all CLNR customer groups with a mean of 2955kWh where demand at the 90th percentile (4982.5 kWh) is lower than the mean of the rural off gas customer group (5336.8kWh).

The rural off-gas customer group has the highest mean annual electrical consumption, standing out distinctly from the other groups. However, this group consists of only 39 customers making the variance a potential overestimate which is acknowledged in our analysis. For these customers a wide distribution with a mean annual consumption of 5337kWh is seen and possesses a small secondary hump around 13000kWh; however because of the small sample size this is unlikely to be representative of the whole population.

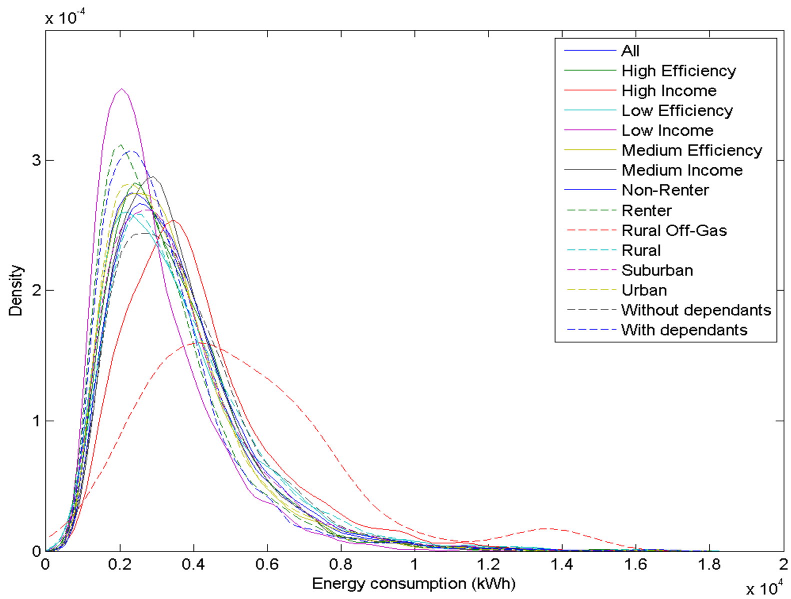


Figure 2: Annual consumption density plots for TC1a customer groups.

High income users yield a mean of 4125kWh per annum and are the highest consuming group other than the rural off-gas customer group.

There is a positive skew in annual consumption with home owners (400kWh) to renters where renters have a lower annual consumption at 3207kWh. Home owners and low efficiency customers consume similar amounts of energy annually.

One aspect of importance to network operators is the likely loads experienced at various times of the day and periods in a year which may affect network operation, design and sizing. With this in mind peak demand on a weekday for a winter and a summer month was explored.

Figure 3 illustrates the distributions of all customer peaks in demand for all weekdays in January and July across all the CLNR subgroups. The two groups that differ the most from each other are the low income and rural off gas groups. The low income customers have a high frequency of lower peak demand compared to any other group whereas the rural off gas customers have a greater spread of peak demand which is positively skewed compared to all other groups. High income customers show the highest peak demands of all other customers when rural off gas customers are not considered. To increase the sample size all weekdays in the month for all customers were considered and so the samples are not point estimates.

Considering the distributions of all samples in the figures, by visual inspection an upper bound of around 25% of all customers appeared to experience an initial peak of 1kW with a further bimodal peak at 1.8kW. Low income customers appear to have the same distributions as dependants and renters, with dependants having the same distribution as renters at the first mode. The demographic with the first mode (i.e. lower peak) could be interpreted to represent families with young children in rented accommodation and/or pensioners. To explore further the distributions would have to be split in turn into low and high peak demand (below 1kW and above 1kW) to try and extract the possible cause and customer types of these bimodal distributions. The same piece of work would be valuable on the Mosaic attributes as evidence suggests these groupings are more consistent in peak demand and electrical consumption separation.

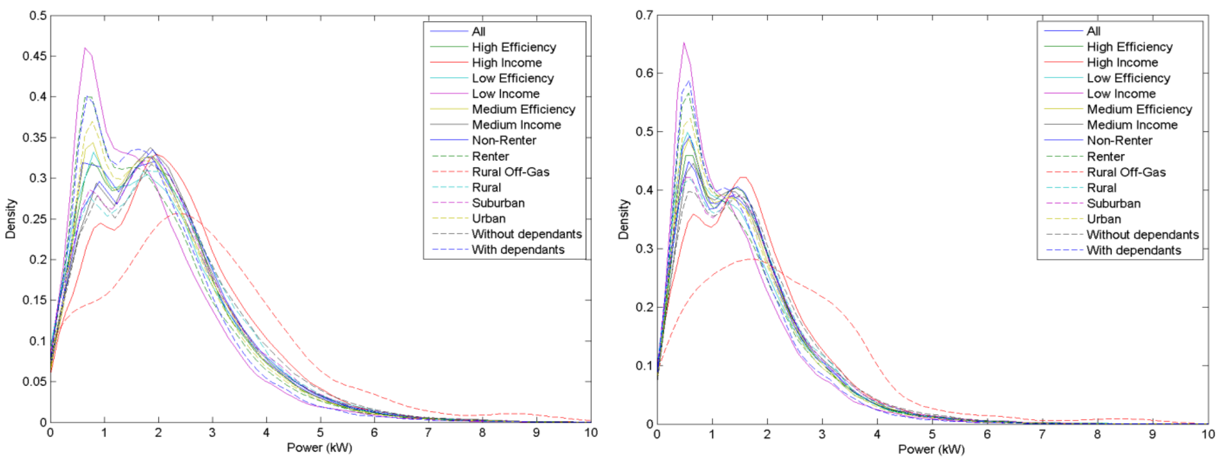


Figure 3: Distributions of all customer peaks in demand for all weekdays in January (left) and July (right) where the sample is composed of daily single peaks for all customers.

A more stratified income demographic may not provide any more relevant information as there are other factors that affect electrical energy consumption.

Taking the low and high income groups the seasonality of the mean peak demands was explored (see Figure 4). Mean peak demand for high income customers shows a clear shift between summer (modal peak of ~1.5kW) and winter (~2kW). For low income customers the summer modal mean peak is less than 1kW only increasing to ~1.2kW during the winter months.

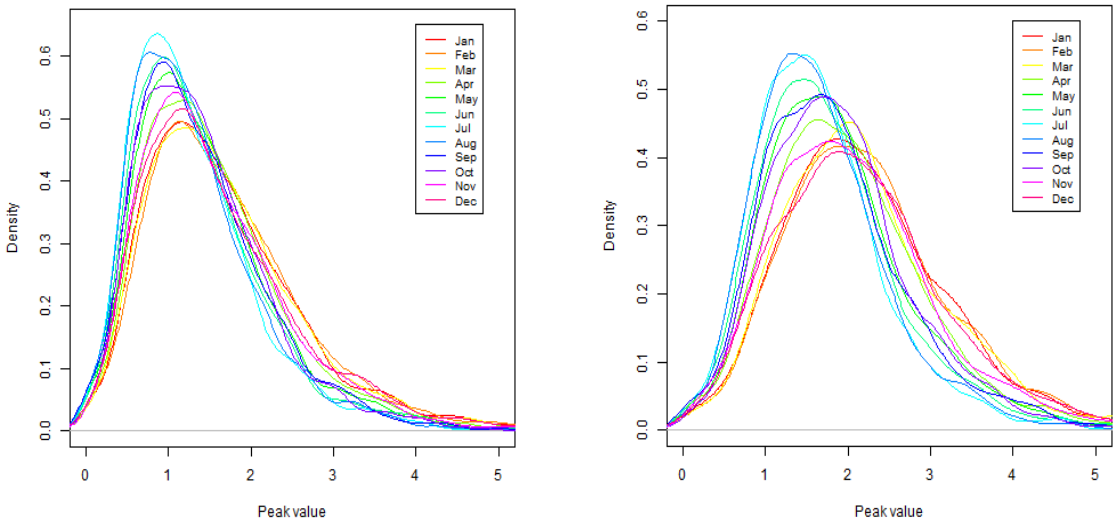


Figure 4: Weekday mean peak density plots for low (left) and high income (right) customers over all 12 months.

Similarly, low efficiency customers exhibit a similar modal demand to those on low income. In the winter months, however, low efficiency customers show some form of 'shoulder' in demand which is not seen in the high efficiency or either of the income groups. This could potentially suggest some form of auxiliary electric heating. The modal demand is positively skewed for high efficiency customers with a modal peak of ~1.2 and ~1.8kW in the summer and winter period respectively.

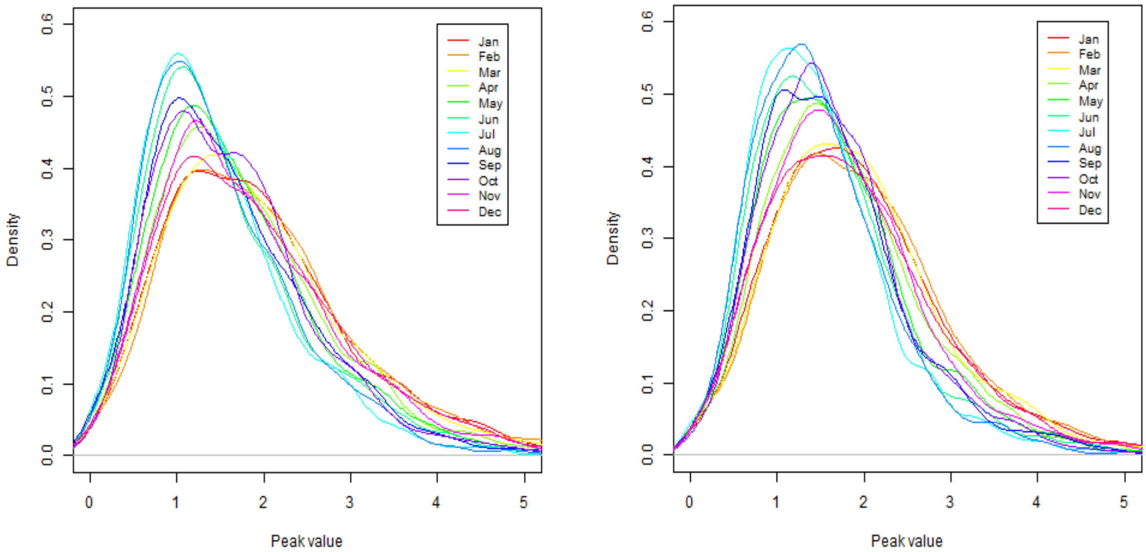


Figure 5: Weekday mean peak density plots for low (left) and high efficiency (right) customers over 12 months.

Looking across the year, Test Cell 1 provides some insight into the seasonal nature of electricity demand, with the winter months showing not only highest energy consumption but also higher peaks in demand (Tables 12-19).

This is in line with insight from the **Social Science Report April 2014, Section 4.2.2:**

"Diversity of levels of electricity consumption increases in the winter months. This can be seen as the interquartile range (the gap between the 25% percentile and the 75% percentile) increases when demand rises in winter, which has the effect of widening the gap between mean and median consumption. The qualitative data reveals some evidence of changes in how practices are performed relating to the seasons, most notably, laundering and heating (thermal comfort) practices:

> *Through the summer months it can be 7 o'clock in the morning put it [washing machine] on, quite early so I can get them out. Usually maybe two loads.* **(ML20)**
>
> *Obviously, in the summer I never, ever use my tumble dryer. I always put them out on the line. 'Cause I prefer it, they smell's nicer when it comes in from the air. […] Winter- obviously, I do put the tumble dryer on.* **(MJRTL07)**
>
> *We tend to use dryer during the winter, but once it gets Spring we wouldn't use the dryer. Even in the winter I only use the dryer for towels and sheets 'cause I've got airers for the clothes. But then in the summer we don't really use the dryer at all really … we've got a couple of outside lines. Also, 'cause it's a washer dryer you find its tied up doing the washing.* **(EPJ012)**
>
> *Yes, we'd have different settings [spring, autumn, winter], we'd change it very often. Well, we don't change the start time and the finish time but during the day we'd change it when it comes on so depending on sunshine basically.* **(ML23)**"

5.1.2. Group Variability

To determine the spread of demand peaks and annual consumption for the different customer groups, distributions of the two metrics on each sub-group were carried out. The mean annual consumption and the mean peak demand for the different groups were computed following the equations 1 and 2 below. Figure 6 and Figure 7 depict the results.

$$E_i = \sum_{j\in[S_y]} E_{i,j} \tag{1}$$

$$\hat{P}_{i,k} = \max_{j\in[S_{R1}]}\left(P_{i,j,k}\right) \tag{2}$$

From the above, i, E and $\hat{P}$ represent annual energy consumption and peak demand for a group j and analysis period k. The analysis period is defined by month and weekday or weekend, where the analysis period is the whole year k is simply the whole period. These two approaches explore whether the variability within a group is greater than the variability between different groups considering both the CLNR and Mosaic demographics.

An interval estimate of the mean around each group is desirable because the estimate of the mean varies from sample to sample and thus gives an indication of how much uncertainty there is in the estimate of the true mean. A confidence interval (CI) of 95% was computed. Supporting tables for the figures can be found in the appendices (Table 25 and Table 26).

In both, the annual consumption and peak demand of the groups show the highest variability around the mean for the rural off-gas group, with a lower and upper confidence limits for the annual consumption of 4392.3 – 6131.9kWh. All the other groups show a narrower CI reflecting higher precision in the mean estimate of the group. Rural, high income, renters and customers without dependants show that their mean interval estimate annual consumption lies outside of the whole data sample. Because of no overlap between the CI there appears to be a difference in the population means, specifically high income compared to all others when rural off gas grid is excluded. Besides rural off gas customers, high income customers have the highest energy annual energy consumption of all other groups.

Simply looking by inspection reveals information on the statistical significance nature of the groups. Low income customers have the lowest annual energy consumption of all groups. Since no overlap occurs with any other demographic, we can say that there is a statistically significant difference in the population means of this group compared to all other groups. In a similar way the rural off-gas, high income, with dependants and renters groups give further sets that also show statistical significance to all other groups.

Exploring the peak demand of the customers in a similar manner it is noted that the CI within each group is slightly narrower than for annual consumption, since it is expected that the peak of each customer within the group, and across all the year, will show less variance. The average mean half hour peak for any customers will show diversity in when that peak occurred, as explained through the after diversity of maximum demand approach [7] where within each household there are times, mainly the 4-8pm period, in which electricity is used simultaneously. A possible explanation could be offered by speculating that low income customers have less aplliances than all other customers or use power in a more sparing manner.

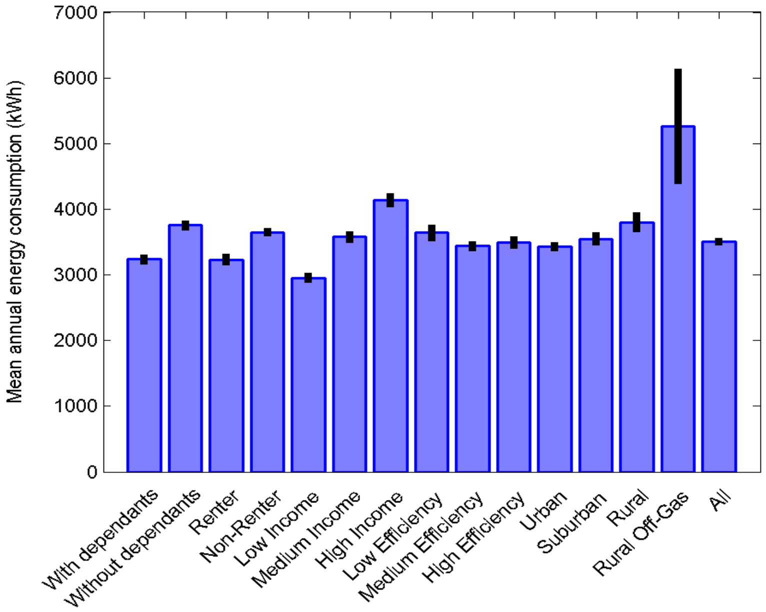


Figure 6: Mean annual energy consumption by customer group with associated confidence interval.

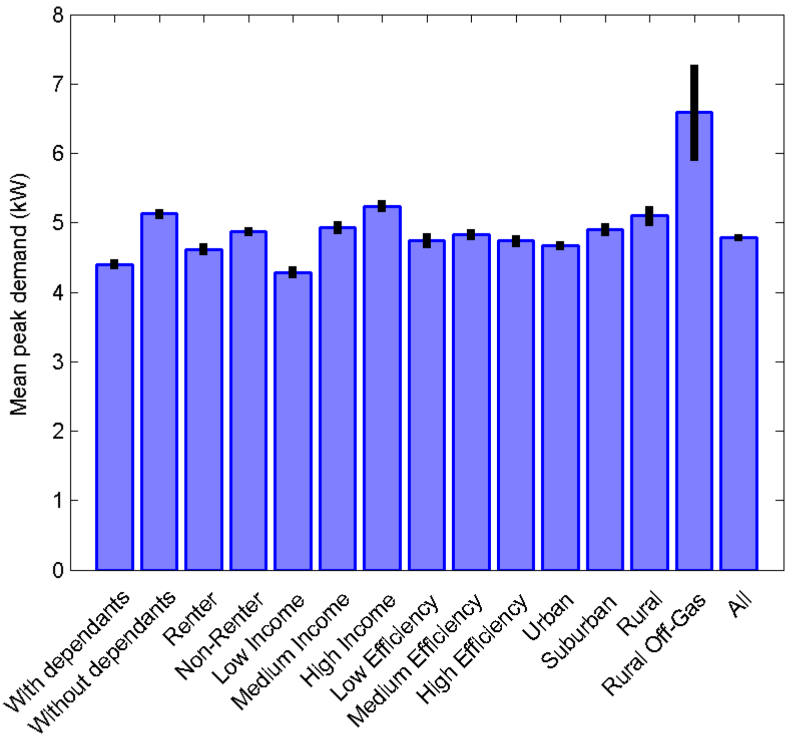


Figure 7: Mean peak demand for all customers within each demographic group. The demand data for the customer was considered across all the year.

Similarly, the same metrics were computed for the Mosaic categories. The results are depicted in Figure 8 and Figure 9. The variability around the mean in both cases is larger than the subgroups seen in the CLNR attributes showing a reduced customer number in each mosaic group.

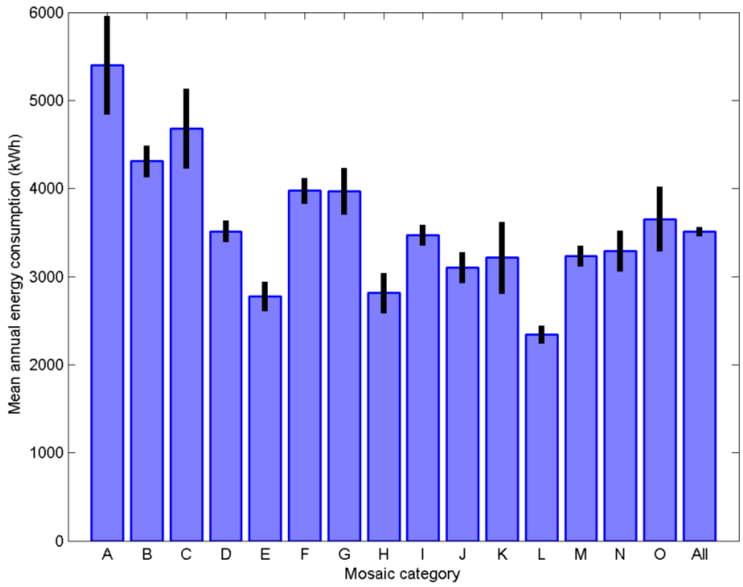


Figure 8: Mean annual consumption for all customers within each Mosaic group. The demand data for the customer was considered across all the year.

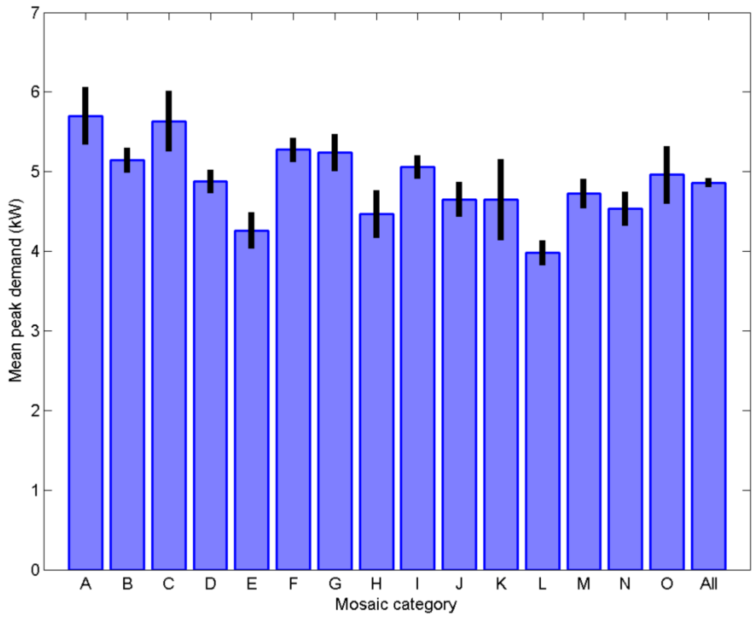


Figure 9: Mean peak demand for all customers within each Mosaic group. The demand data for the customer was considered across all the year.

For annual consumption (Figure 8), the group L "Elderly Needs" is a standalone group with groups A "Alpha Territory" and C "Rural Solitude" showing that at the 95th percentile these two groups have statistically similar annual consumptions. Groups B, C, F, G and O are statistically similar at the 95th percentile level, encompassing groups from "Professional Rewards" to "Liberal opinions" where groups F and G ("Suburban Mind-sets" and "Careers and Kids") use almost exactly the same mean annual energy as each other. These 5 groups account for 30% of the sample makeup.

Exploring the peak demand (Figure 9), Mosaic group A still possesses the highest mean of all groups with group C having a very similar mean peak demand. Groups F, G and O can be classified as statistically similar to these two groups also. L "Elderly Needs" carries a common mean peak to group E "Active Retirement" and also to H "New home makers" but as Section 5.2.2 eludes, the modal peak time isn't consistent (Elderly Needs usually peak at 11:30 whereas Active Retirement usually peaks at 6pm). However the electrical activities between these two groups could be said to be very similar. Besides these 2 subgroups, all others possess a common mean significant at the 95th percentile to subgroup O "Liberal opinions".

In terms of statistical significance group L may be different to all groups except to group E, however because of the 5% misclassification error when doing multiple testing it is difficult to quantify whether the visual level of separation here is big enough to recognise a genuine difference.

To explore seasonal variation, the mean peak demand for January and July with upper and lower CI is shown in Figure 10 and Figure 11 respectively.

The difference in the magnitudes of the peak demands indicate that January was not when the maximum electrical demand was witnessed for any of the mosaic groups with the mean peak between 1 and 1.5kW below that of the whole year mean peak demand. The intention here however was to compare the arbitrary seasonal periods between January and July.

The difference in the magnitude of the mean peak demands in July reveals a reduction by 2kW on average from the whole year. Because the mean peak demand of group A reduced to ~3.3kW, a higher reduction than all other groups more groups are statistically similar to this high consumption and peak power group in July. Nine groups shared the same mean to the 95th percentile (groups B, C, F, G, I, K, M, N and O) with groups E and L (Elderly Needs and Active Retirement) having a common mean.

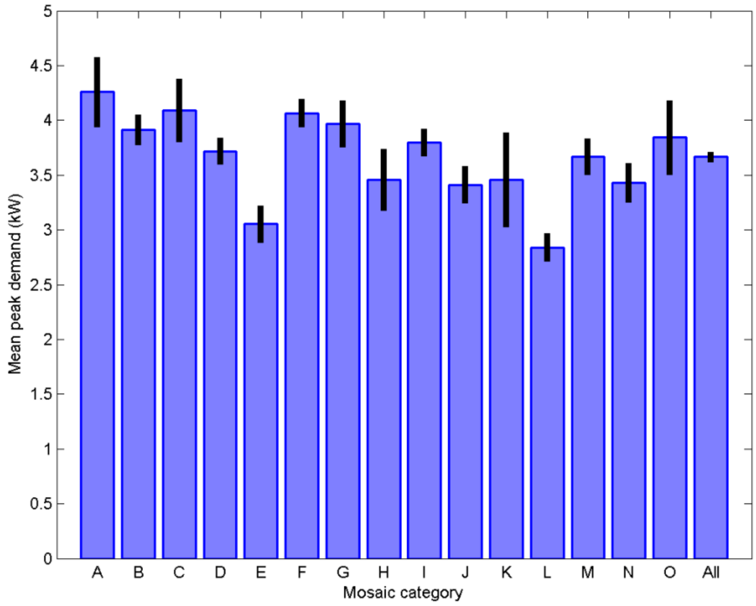


Figure 10: Mean peak demand for all customers within each Mosaic group. The demand data for the customer was considered across January.

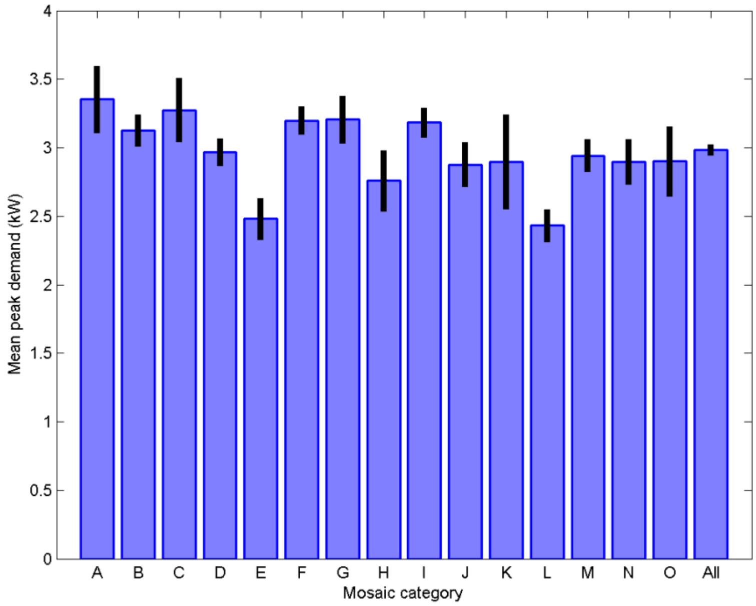


Figure 11: Mean peak demand for all customers within each Mosaic group. The demand data for the customer was considered across July.

5.1.3. Social science cross-reference

This section contains some of the findings of the qualitative and social science work carried out regarding energy use patterns by demographic.

Tenure

"Electricity demand has been found to differ by tenure, with owner-occupiers exhibiting, on average, higher demand than renters".

- **Domestic LO1 & LO2 Qualitative Paper, Research Findings, pg. 3**

Income

"Social Science team reported that of all the socio-demographic attributes analysed to date, income has the strongest association with electricity demand with higher income households (combined household income of more than £30,000) consuming on average 2.9 kWh per day in June and July and 4.7kWh per day in December more than lower income households (combined household income of less than £14,999)"

- Social Science Report April 2014 Section 3.1.2.

"High income is associated with considerably higher demand. With the exception of rural off gas households in the rurality analysis, of the socio demographic variables recorded for Test Cell 1 it is income that is associated with the most divergent load profiles.
Furthermore we found out that the high income groups had more peak intensive loads than lower income groups; this is true for almost all comparisons in all months (the exception is that high income groups have lower ECM than medium income groups in May). The difference between income groups becomes more pronounced as months become colder and darker, other than in December."

- Social Science Report April 2014 3.1.2.4 Income

Rural off-gas

"Households in rural off-gas areas have a substantially increased demand for electricity throughout the year compared to gas-connected households."

- **Domestic LO1 & LO2 Qualitative Paper, Research Findings, pg. 3**

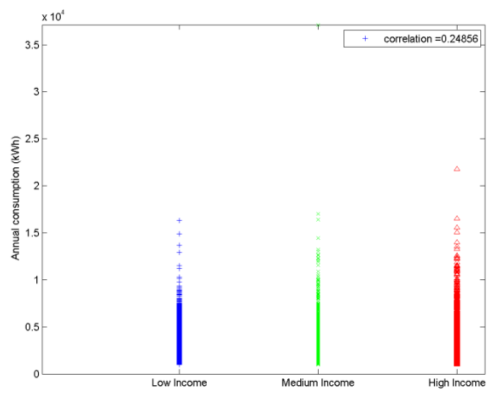


Figure 12: Plot of annual consumption vs income demographic variables (p value = 2.237 *10-84)

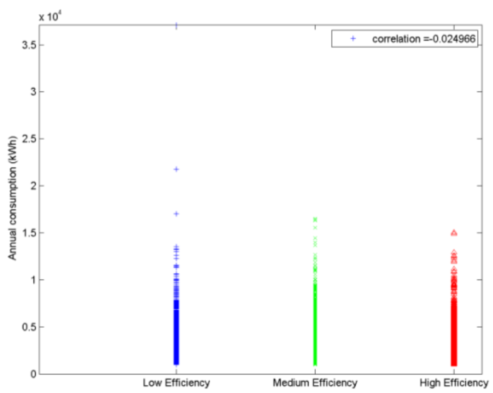


Figure 13: Plot of annual consumption vs efficiency demographic variables (p value = 0.0543)

5.2. Daily variation

The analysis in the previous section considers year-round trends and differences between the demographic groups. However, the time of day of electricity use is of importance when considering the potential for demand side response whilst exploring the options for alternative low carbon technologies for the different customer groups.

For instance, if consumption is higher during the midday period then solar power could be well suited for these customers. Also, electricity networks tend to be more heavily loaded in the evening period, so shifting energy consumption away from these times can delay the need for upgrades to the system.

5.2.1. Social science cross-reference

"Electricity use also has a daily pattern. Dividing the day into three time periods, we can see that although the four hour evening peak period (4pm – 8pm) appears to account for the smallest amount of energy consumption, this should be interpreted as being a period of higher energy intensity given that, on average, Test Cell 1 participants used 25% of their energy in just 17% (1/6) of the day. This translates to a demand for 1.69 times as much electricity in the peak period (4pm – 8pm) as at other times of the day"

\- **Social Science Report April 2014, 3.1.1 Energy consumption and the intensity of energy use**

"The proportion of electricity consumption concentrated in the evening period was also highest and most variable amongst high-income households and lowest and least varied amongst low-income households. Because of their overall contribution to demand in the peak period and the variability in their demand high-income households appear to be a key target group for future DSR"

\- **Social Science Report April 2014, Research Findings LO1**

"Renters also consume a lower proportion of their total electricity use during the evening peak hours, whereas owners tend to consume more during this period. Owners also exhibit greater variation in the proportion of total electricity consumption that happens during the evening period."

- Domestic LO1 & LO2 Qualitative Paper, Research Findings, pg. 3

"We found no evidence that the proportion of electricity consumed in the evening peak period was more or less varied amongst households in terms of their tenure, thermal efficiency or urban or rural location"

- Social Science Report April 2014 3.1.2.5 Thermal efficiency of the home

"As well as having the greatest average daily demand, rural off gas households who tend to use electricity for heating and hot water also consume a higher proportion of their total electricity in the evening period. The potential that new technologies will increase electricity demand in the early evening will need to be considered carefully if plans to shift away from gas to electricity as a source of energy for domestic heating move ahead."

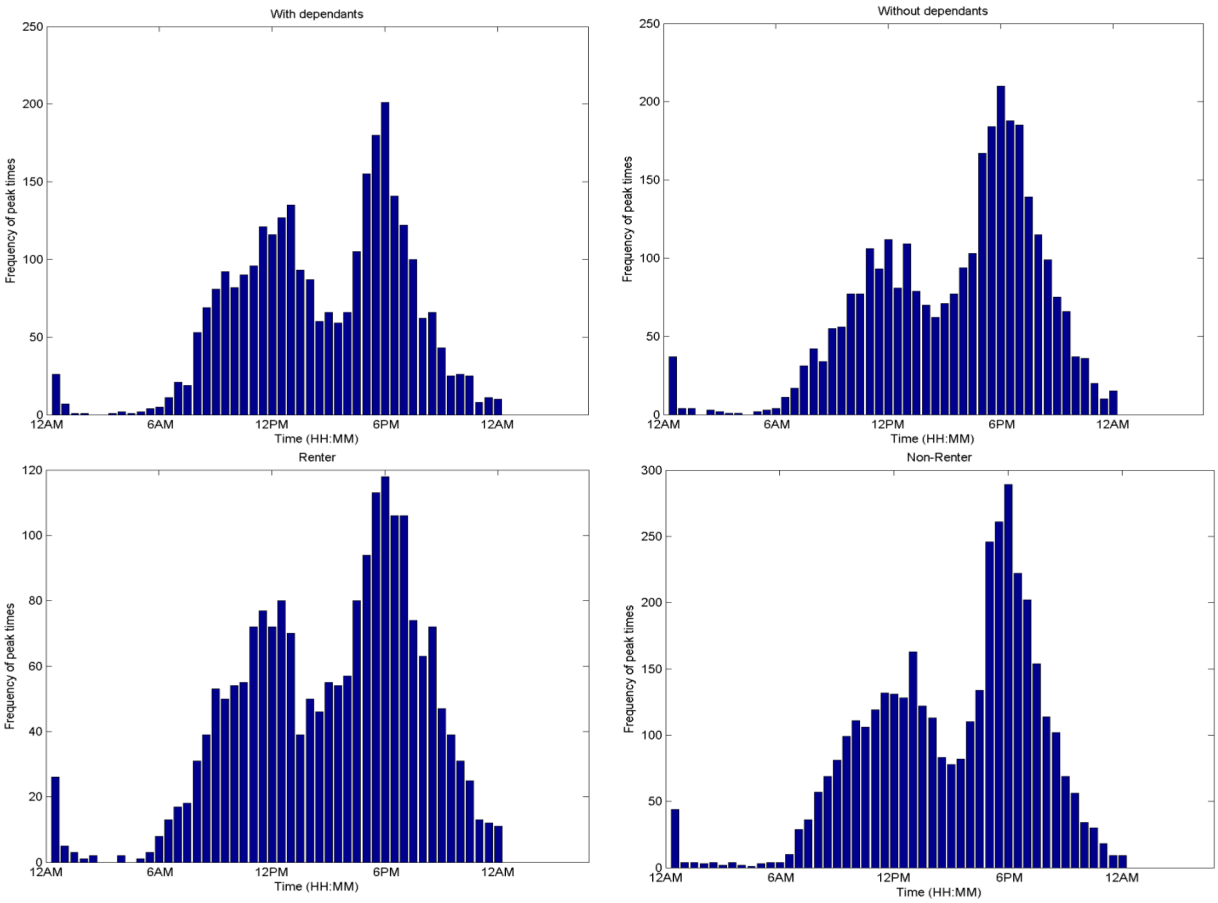
\- **Domestic LO1 & LO2 Qualitative Paper, Research Findings, pg. 3**

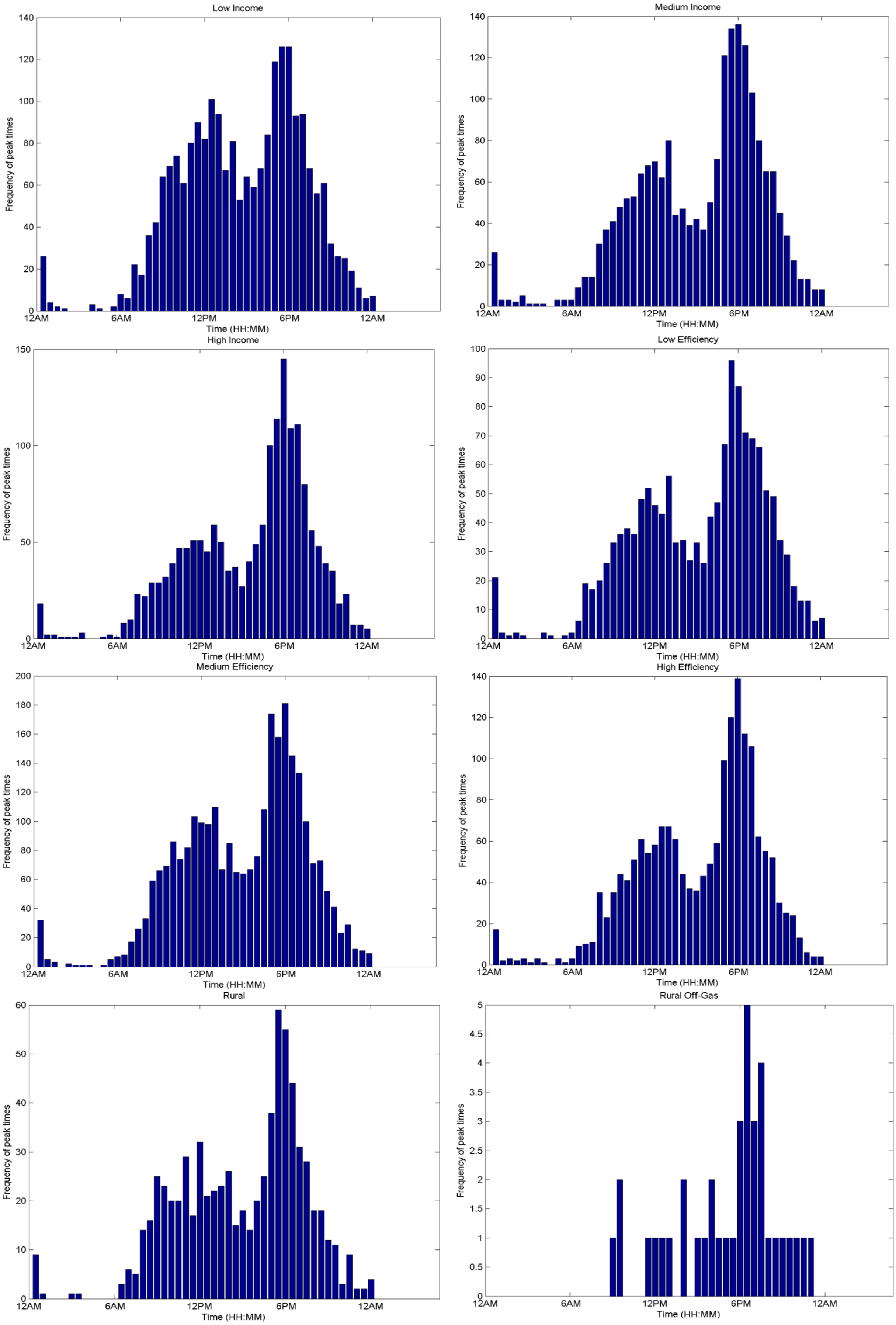
5.2.2. Peak demand time of use

This section considers the time of the day at which different users have their peak energy demand. The times of use of peak demand for the different subgroups are represented in histograms for a typical weekday in January is shown in Figure 14 for the CLNR demographic groups and Figure 15 for the Mosaic groups. These are computed from terms 8 and 9 from Table 3.

> Note that this does **not** consider different individual or demographic group contributions to the overall system peak (investigated in Section 5.4). The histograms should not be confused with daily demand profiles: the bars indicate the number of households whose daily peak is in that half-hour slot, but contain no information as to what the individual or collective demand during those times was.

For the CLNR groups, all sub-groups barring rural off gas yield what appears to be a set of bimodal distribution of times of peak usage at half hour intervals. In January peak times were mostly concentrated around a narrow window where the earliest peak occurred at 17:00 and the latest peak occurred at 18:30 (rural off gas only). In the summer month of July the peak time span didn't alter suggesting that peak demand is immoveable. Only the rural sub-group offered a change in when the modal peak occurred, from 18:30 in the winter to 17:30 in July.




Low Income
Medium Income
High Income
Low Efficiency
Medium Efficiency
High Efficiency
Rural
Rural Off-Gas
Frequency of peak times
Time (HH:MM)
12AM
6AM
12PM
6PM
12AM

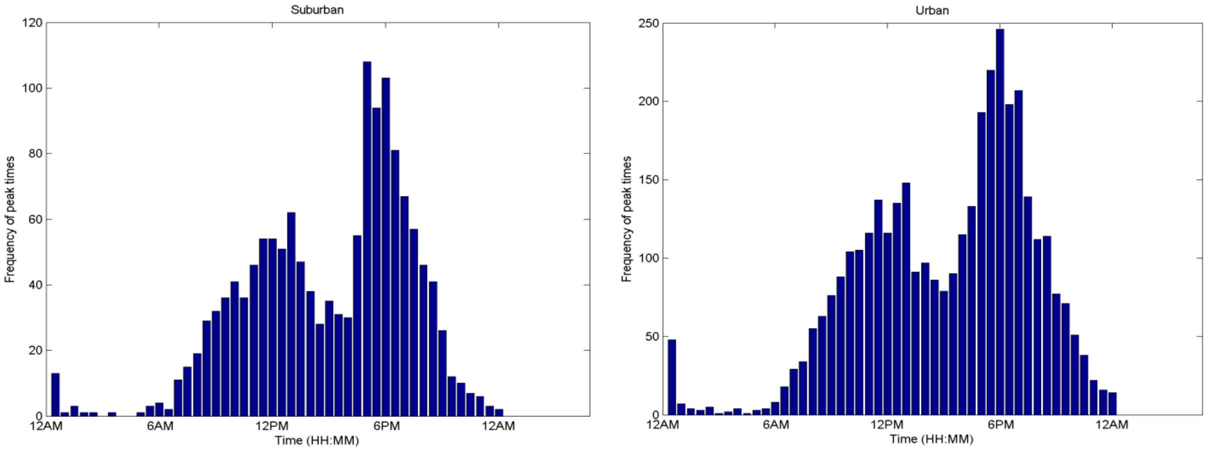
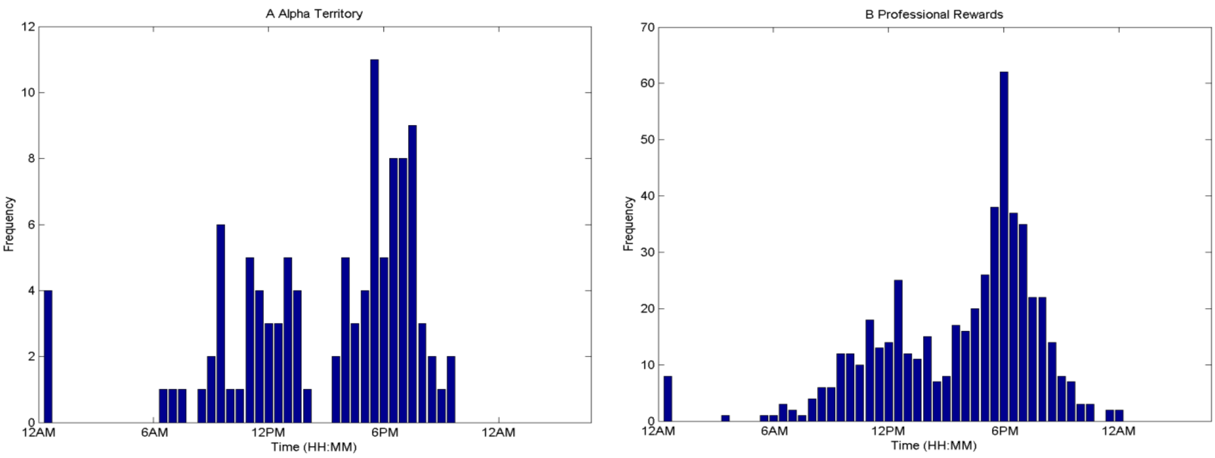


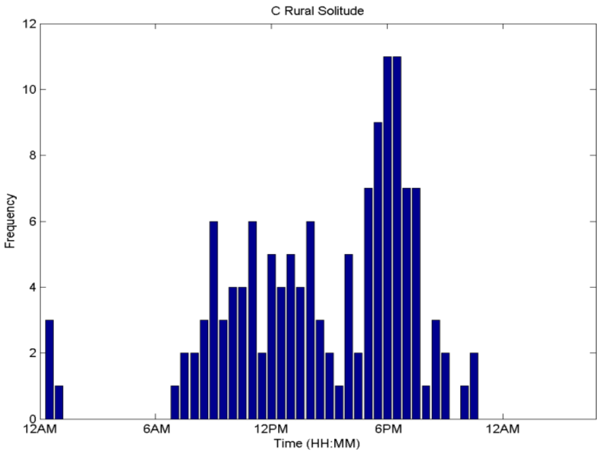
Figure 14: Mass function of peak times for the CLNR demographic groups. The distributions are counts of each customers peak demand for January weekday.

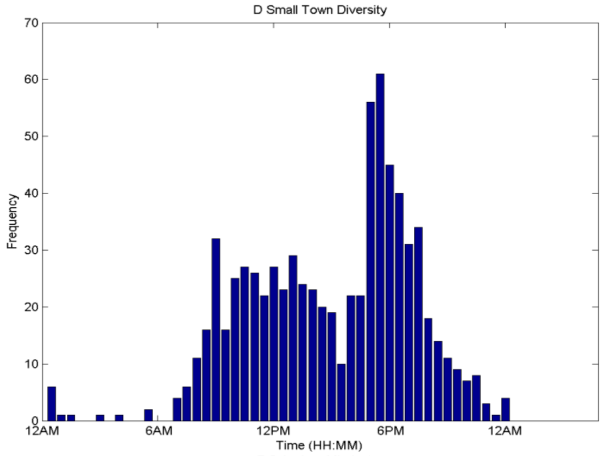
The Mosaic groups offered a larger time window for the modal peak than the CLNR sub-groups. Modal peak times across all groups occurred at 17:00 (Mosaic H – New Home Makers and Mosaic group I – Ex-Council Community) and 19:00 (Mosaic group N – Terraced Melting Pot) with the exception of Mosaic group K – Upper Floor Living whose modal peak occurred at 20:30 and Elderly Needs (Mosaic group L) with a modal peak at 11:30. The most frequent maximum peak time across all groups was 6pm from 6 demographic groups.

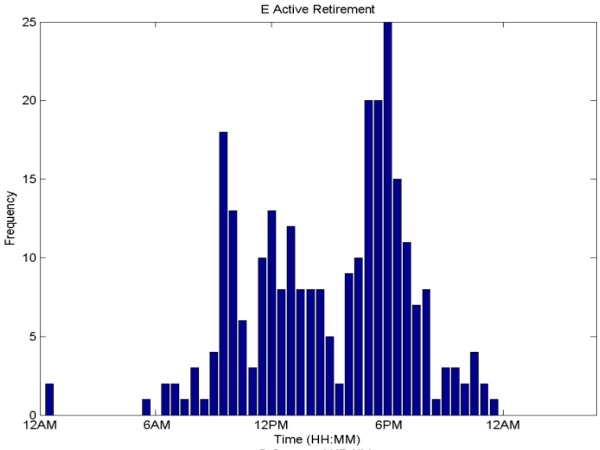
Claimant Cultures and Elderly Needs (Mosaic groups J and L) showed a very characteristic bimodal distribution which is easily explainable through the daily habits of these groups. New Home Makers show the lowest occurrence of peak demand during the day which seems to infer firm indication of the household being in work as would be expected. Alpha Territory (group A) shows a low peak demand count during the day also.

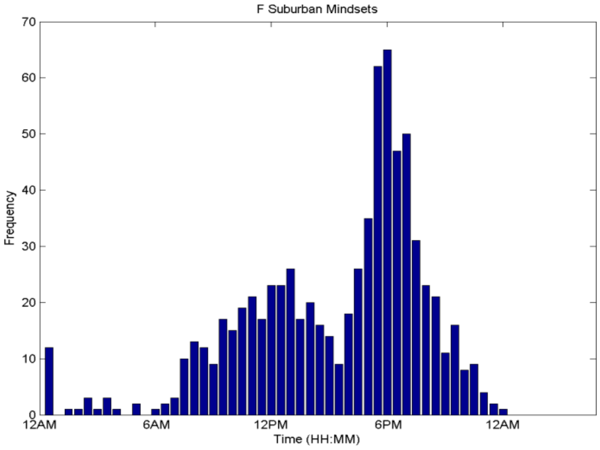
Twelve of the 15 groups in the Mosaic demographic proportions revealed a peak time shift between winter and summer with the largest shift occurring for Upper Floor Living (Mosaic group K) where the modal peak was 20:30 in January and 12:30 in July, Claimant Cultures closely followed with a peak July occurring also at 12:30. Eight groups (groups A, C, E, G, H, I, N and O) showed an increase in the model peak time during the summer as opposed to an earlier peak (J, K, L and M).

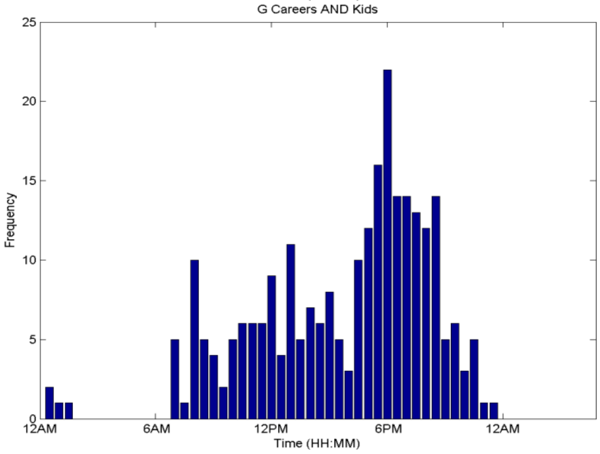


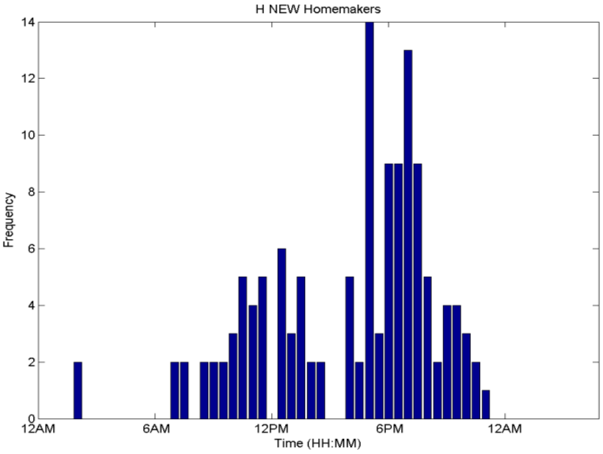

C Rural Solitude
Frequency
0 2 4 6 8 10 12
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)

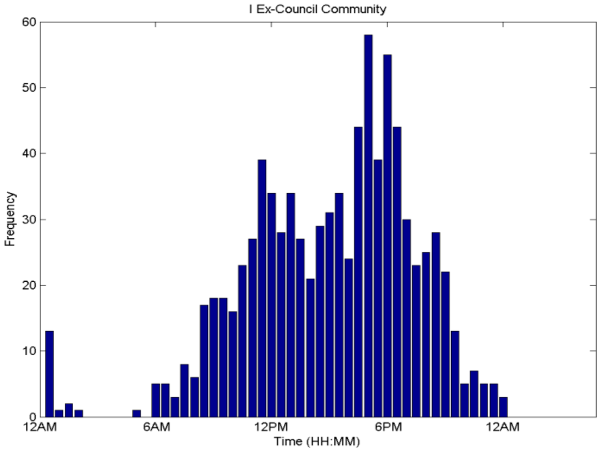

D Small Town Diversity
Frequency
0 10 20 30 40 50 60 70
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)

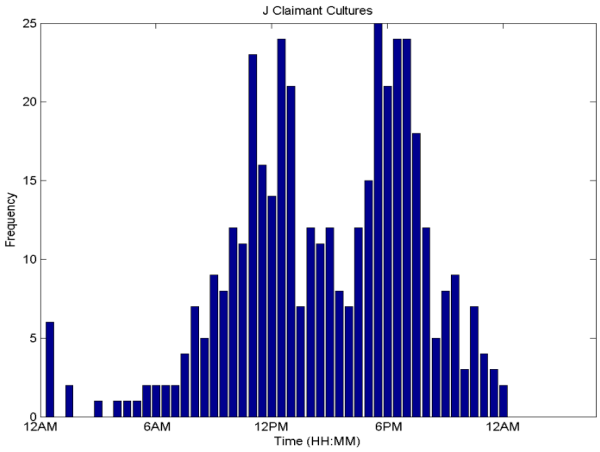

E Active Retirement
Frequency
0 5 10 15 20 25
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)


F Suburban Mindsets
Frequency
0 10 20 30 40 50 60 70
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)


G Careers AND Kids
Frequency
0 5 10 15 20 25
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)


H NEW Homemakers
Frequency
0 2 4 6 8 10 12 14
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)


I Ex-Council Community
Frequency
0 10 20 30 40 50 60
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)


J Claimant Cultures
Frequency
0 5 10 15 20 25
12AM 6AM 12PM 6PM 12AM
Time (HH:MM)

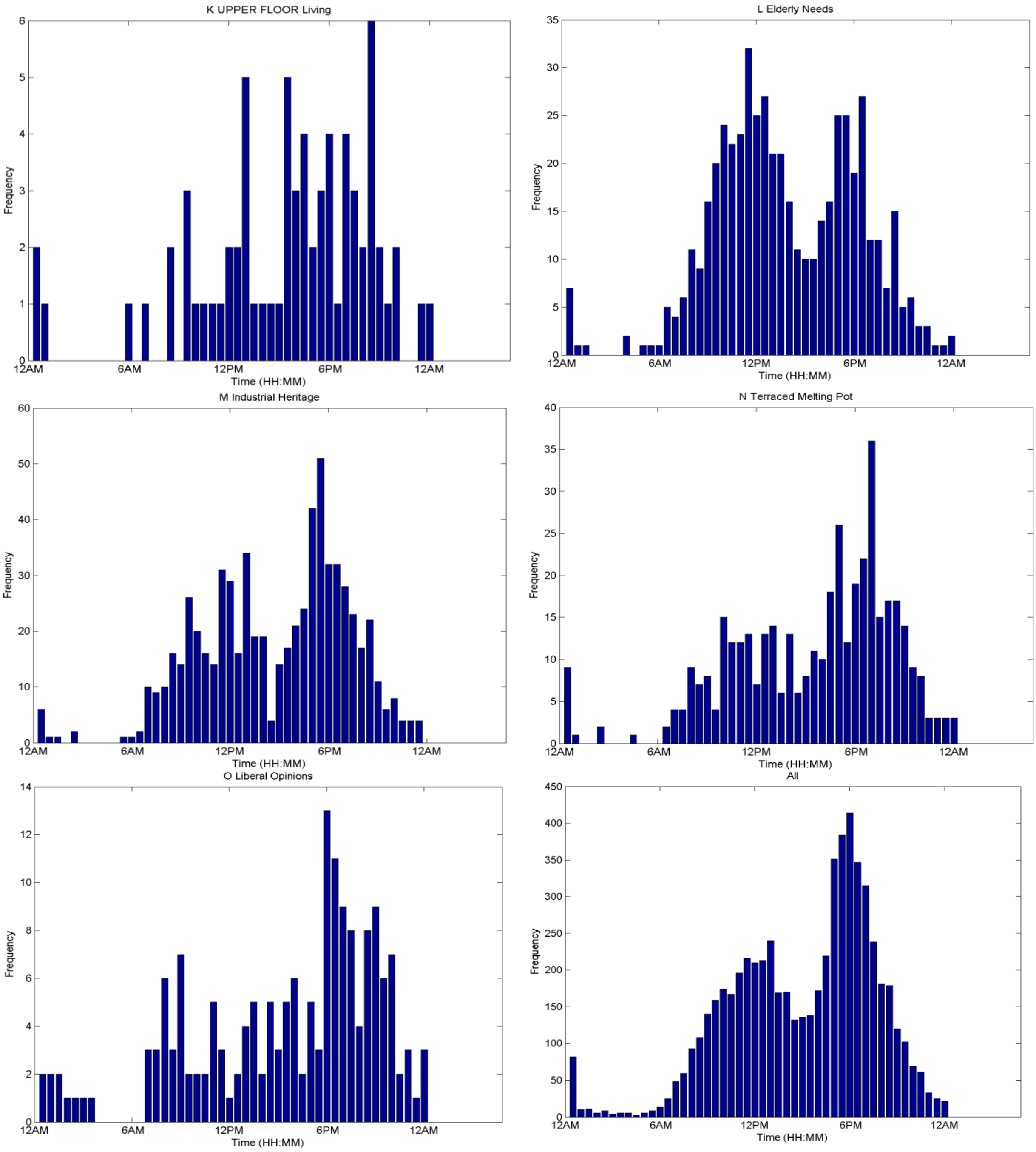


Figure 15: Mass function of peak times for the different Mosaic demographic groups. The distributions are counts of each customers peak demand for January weekday.

5.3. Demographic Effects on Electrical Demand and Consumption

The analysis presented thus far has not managed to infer which of the CLNR subgroups annual consumption and demand is most dependent on. In this section we focus our attention on which subgroups are the most useful independent variables when inferring either of these two metrics. Analysis of Variance (ANOVA) and clustering are the two techniques used.

Through ANOVA we gain insights into which sub-groups are the most influential in predicting peak demand and energy consumption which help to explain the usefulness of the independent (subgroup) variables. However when considering the CLNR subgroups only it is important to see which, from a network design perspective, the demographic make-up of natural demand clusters. From such insight, it may be possible to determine likely locations of retrofit required for the future or new electrical infrastructure sizes, based on the likely local demographics. Thus networks could be designed to more natural clusters of the various subgroups.

This section explores which of the CLNR sub-groups possess a common mean, either individually or simultaneously, looking at the interactions between groups. A two way ANOVA is carried out to elicit insights into which group carries the greatest influence in predicting both annual energy consumption and peak demand (annual energy consumption shows the highest R squared value to peak demand). A linear regression model was then fitted to test these categorical predictor variables against the dependent variables (annual consumption and peak).

5.3.1. Demographic Group Significance: ANOVA analysis

Because of the interdependencies of the CLNR subgroups it is possible to explore the demographic groups for a common mean using these. The Mosaic groups have no interdependencies and are therefore omitted from this analysis.

A *2*-way ANOVA was performed to explore the single and interaction effects of all the variables, cross compared together, in order to find out which of the single and shared groups have a common mean. In doing so we use two dependent variables, mean annual energy consumption and the yearly peak half hour demand computed using the formulae in Table 3. The result with the lowest *p*-value will be ranked the highest in terms of its significance in determining the dependent variable.

Because of the number of independent variables under consideration the ANOVA was performed instead of a *t*-test owing to the chance of committing a type I error. Exploring single comparisons between the 5 independent variables would give a 40% chance of committing this type of error. If analysed on a seasonal basis the *p* statistic would be divided by the number of seasonal periods to limit the risk of a type I committal error, i.e. it is important to “spread your *p*’s”.

The normal-model based ANOVA analysis assumes the independence, normality and homogeneity of the variances of the residuals. A parametric test such as the analysis of variance assumes the underlying source population(s) to be normally distributed, where the variances in each population are similar, the data was log normalised. The quantile plot, Figure 16, shows the transformation of the data to be reasonably normal compared to the absolute raw data.

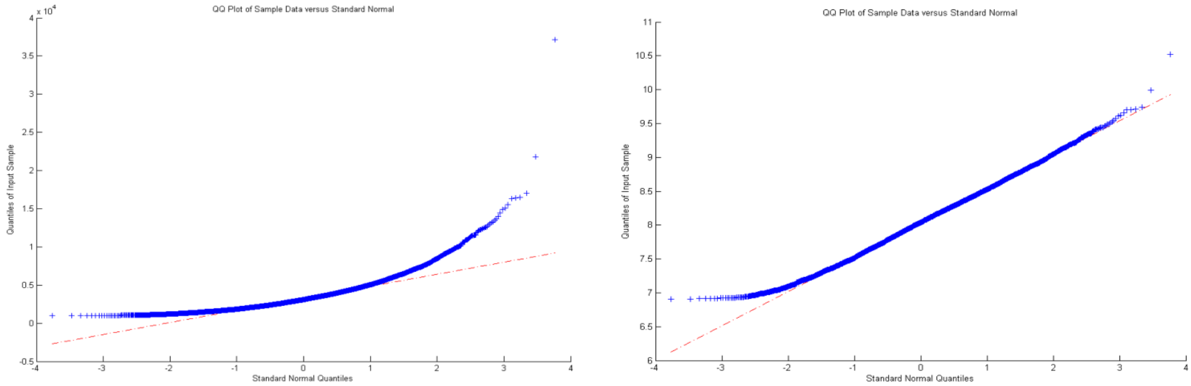


Figure 16: Quantile – Quantile plot of absolute demand (left) and log normalised demand (right) for all customers in the sample brom October 2012 to September 2013 inclusive.

Initially 15 groups were presented to the model with the 5 categorical variables and the 10 interaction variables ($2N$categorical variables) with the dependent variable being annual consumption. Once an ANOVA test was performed the model was further pruned to reduce the number of variables, either single or interactive, depending on the significance of the categorical variables. The least significant variable (high *p*-value) was eliminated and the model was run again with *N*-1 reduced terms until all factors proved significant. The ANOVA model is shown as:

$$y_{ijklmk} = \mu + \alpha_i + \beta_j + \gamma_j + \delta_j + \varphi_j + \omega_{ij} +, \dots, + \omega_{mn} + \varepsilon \quad (3)$$

Where:

$\mathrm{y_{ijklm}}$ is a matrix of customers annual consumption observations, μ is the overall mean response and α_i is a matrix whose columns are the deviations of each households annual consumption from μ that are attributable to the factor household dependency with i levels. The different subgroups within a categorical factor constitute different levels. Likewise $\beta_\mathrm{j}, \gamma_\mathrm{k}, \delta_\mathrm{l}, \varphi_\mathrm{m}$ represent the categorical factors, renters, income, efficiency and rurality respectively at $\mathrm{j, k, l, m}$ levels. ω_ij is a matrix of interactions and ε is a matrix of random disturbances between all levels.

Table 4 details the model outputs, namely the degrees of freedom, F factor and *p* value where the categorical factors were significant (*p* value < 0.05). The F ratio compares the amount of symmetric variation in the data to the amount of unsymmetrical variation and is given by the ratio of the model mean squares by the residual mean squares where a value < 1 leads to an insignificant result. Because our ANOVA specifies interactions we are interested in more than the grand mean of the variables, namely the marginal means, the combined cell means of one variable given a specific level of another variable (demographic) is now important. For this reason the degrees of freedom is an important parameter to report and indicates the number of parameters available to vary.

It was shown that all but one of the categorical variables (efficiency) are factors in predicting the annual consumption with the interaction between households with and without under 5 and over 65 dependants being significant alongside income. Income is by far the most significant factor when predicting income with a p value 1.22041e-91 with rurality also significant at 4.71989e-06 marginally higher than the interaction terms.

Table 4: ANOVA results for the model in equation (3) when the dependent variable was annual consumption

Demographic variable	*df*	F factor	P value
Dependencies	1	132.26	2.75407e-30
Renter	1	68.87	1.29181e-16
Income	2	216.92	1.22041e-91
Rurality	3	9.17	4.71989e-06
Dependencies*Income	2	13.31	1.7134e-06

In the first instance a 2-way ANOVA was performed on all CLNR demographic variables with full interaction effects considered. The initial run revealed the interaction term between dependencies and renter yielded the lowest *F*-ratio of 0.04 closely followed by the single variable effect – efficiency, with an *F*-ratio of 0.33. The model was re-evaluated after eliminating the variable (single or interactive) with the lowest p value in turn until all remaining variables showed an appropriate level of significance (p value < 0.05).

A Multiple linear regression model was constructed to assess the predictive power of the identified significant variables in the ANOVA to both annual consumption and peak demand. The coefficients were determined in each group based on the order in which they were presented to the model; those presented first were defaulted to zero with respect to all other parameters within that group. The coefficients are therefore vectors of the group in which is being measured. The regression fit is depicted in equation (4).

$$y = \mu + \beta_0 Dependent + \beta_1 Renter + \beta_2 Income + \beta_3 Rurality + \beta_4 Dependency \cdot Income + \epsilon \qquad (4)$$

The model yielded an R^2 value of 0.118, which shows a very low goodness of fit.

The same analysis was conducted for the same categorical variables but considering the peak demand of each customer. The factor Efficiency wasn't discarded in the analysis this time but it was shown that this variable was the least significant.

Table 5: ANOVA results for the model in equation (3) when the dependent variable was peak demand

Demographic variable	*df*	F factor	P value
Dependencies	1	175.02	0
Renter	1	22.14	0
Income	2	150	0
Efficiency	2	6.37	0.0017
Rurality	3	19.59	0
Dependencies*Income	2	8.19	0.0003

Similarly a regression model for the peak demand was also fitted with the regression equation highlighted in (5) below yielding an **R^2 value of 0.0771.**

$$y = \mu + \beta_0 Dependent + \beta_1 Renter + \beta_2 Income + \beta_3 Efficiency + \beta_4 Rurality + \beta_5 Dependency \cdot Income + \epsilon \quad (5)$$

The low goodness of fit in both cases indicates that these categorical variables cannot be used on their own in predicting peak demand or annual energy consumption.

The regression study in section 5.4.2 indicates system peak demand can be predicted to reasonable accuracy through daylight hours and ambient temperature. The same study could be studied further in predicting each demographic group's peak demand. A richer investigation would be required in predicting annual consumption.

5.3.2. Demographic Clustering

It has been shown that all the variables are significant factors in determining the peak demand and all but one of the variables (efficiency) are significant factors in determining annual consumption. Both ANOVA models regarded the interaction between dependants and income as highly significant. The ANOVA model did yield a very poor fit when a linear regression model was fitted and did not reveal the composition of customer types (through demographic subgroups) on varying levels of demand. In this work, demand levels were extracted through clustering. More precisely natural demand profiles were grouped through a "$k - means$" clustering exercise.

Because the peak demand of customers, when considering the CLNR demographics, showed very small variations in their mean, and annual energy consumption is not the major parameter of consideration in network design, the monthly mean consumption of a peak demand month was the parameter of choice. January was taken as the month of highest peak demand.

Normalised Clustering

The analysis is based on three clusters observed from the dendrogram in Figure 17. The dendrogram presents a hierarchical branch and leaf structure of the data where the leaves indicate the data sub clusters with the branches to other sub clusters. The largest distance measure is indicative of the number of clusters. The distance measure used is based on Ward's linkage, which aims to have the smallest within cluster distances (i.e. the 'tightest' clusters). Figure 17 shows the hierarchical cluster tree for the normalised consumption. Normalising the monthly mean energy consumption by that customer's maximum consumption in the month of January produced two clusters whereas the absolute data showed 3 clusters present, Figure 18. On examination of customer attributes there was no difference between analysing the normalised demand data with two or three clusters, for this reason the preceding analysis will use three clusters. Figure 19 depicts the normalised cluster centres.

Because the clusters represent mean monthly usage patterns compared to that customer's maximum monthly demand the clusters in themselves show the distribution of a set of customer's usage patterns.

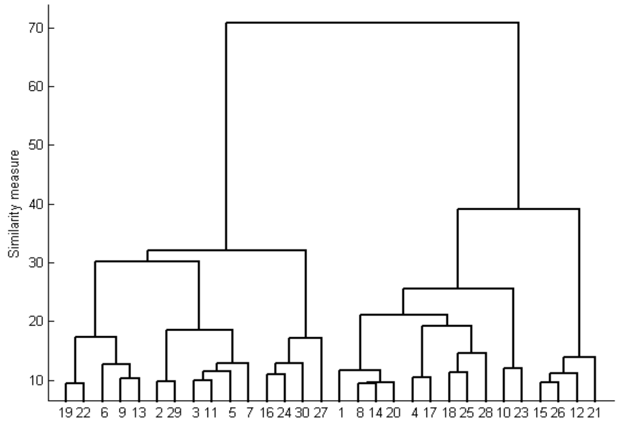


Figure 17: Hierarchical cluster tree for normalised consumption

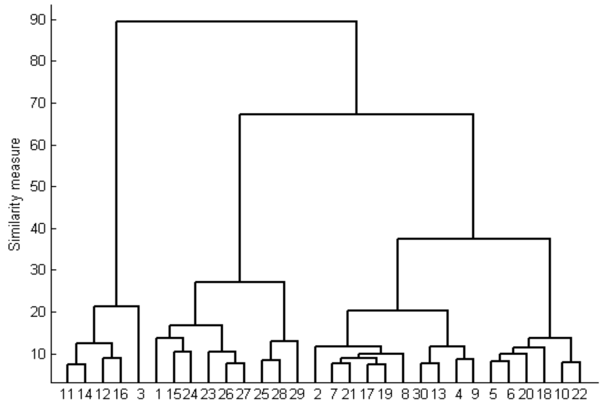


Figure 18: Hierarchical cluster tree for absolute consumption

Clustering on the normalised data extracted the shapes of energy usage more so than their magnitude. Customers in Clusters 2 and 3 represented high and low energy consumers.

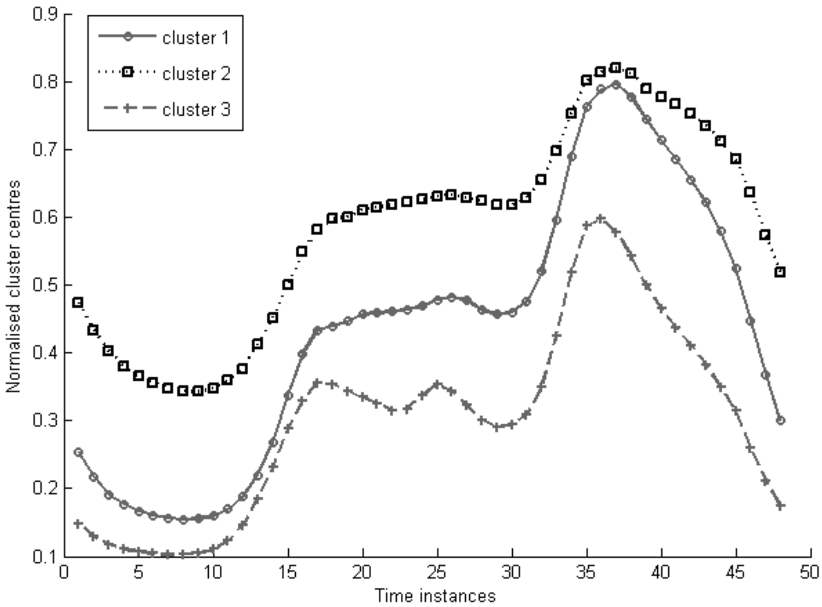


Figure 19: Normalised cluster centres

It was hypothesised that demographics do have an effect on energy consumption. We stipulated a hypothesis to say high income earners generally show higher energy consumption than those on low incomes. Using the demographic indicators in the dataset the proportion of customers that reside in the relevant groups for each cluster is assessed. From this a probability is inferred for the types of demographic groups present in each cluster given our sample size.

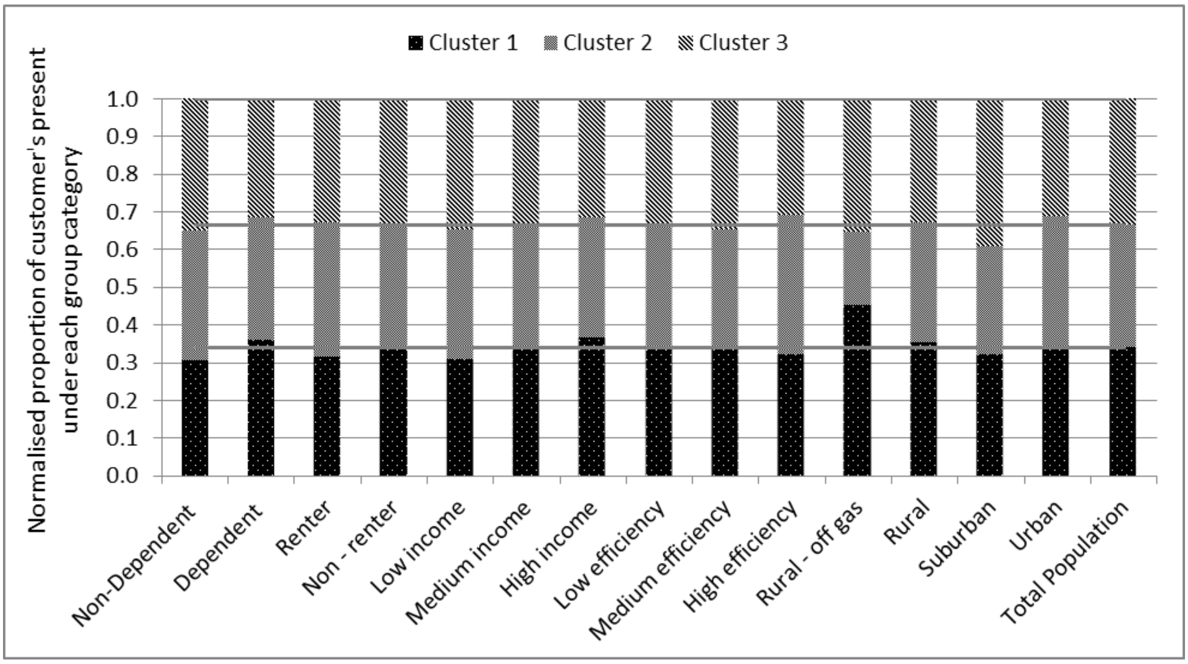
$$P(Group \mid cluster = i) \qquad \forall\ groups$$

Because of the disparity in customer numbers between clusters a relative proportion measure was constructed to draw out the contributions of all demographic categories. The relative proportion is computed by obtaining the percentage contribution for each category in each cluster and then normalising against each clusters percentage contribution.

$$R_i = \frac{P^c{}_i}{\sum_{i=1}^{n} P^c{}_i} \qquad \forall c$$

where P^c indicates the proportion of customers for each demographic category c. R is the relative proportion in relation to other all other clusters of the same category.

Figure 20



represents the relative proportions of customers in each demographic variable to the total number of customers across all the clusters. For the normalised demand we see that no demographic group dominates, or put another way there is no dependency between demographic and cluster.

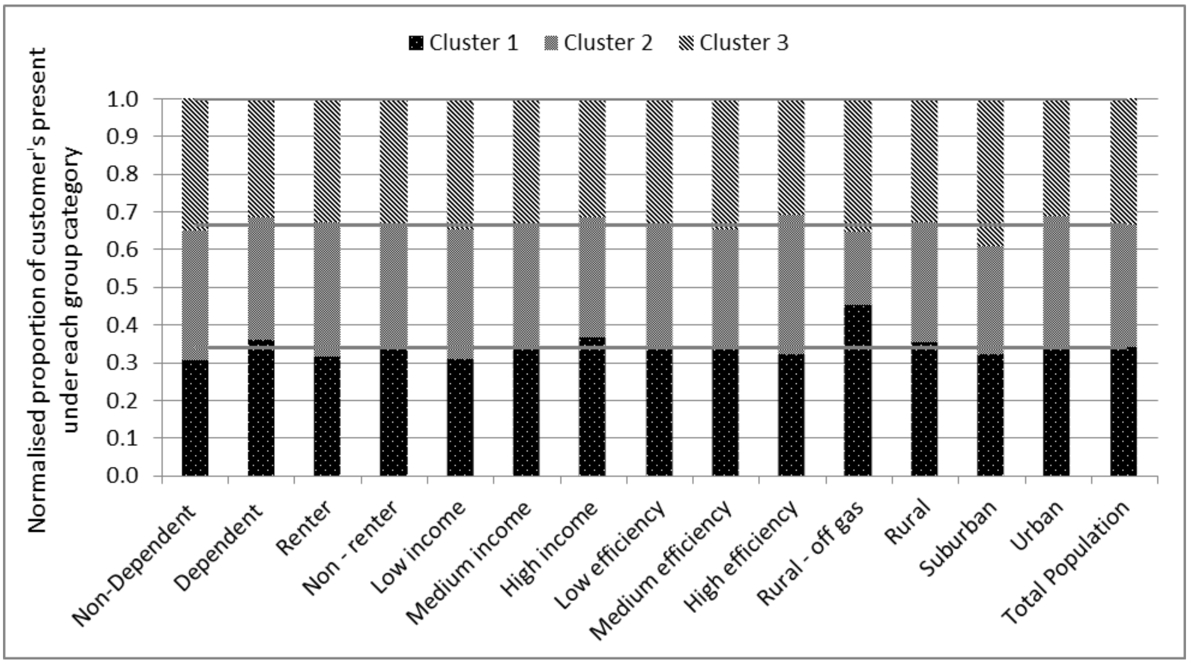


Figure 20: Relative proportions of customer groups present for each cluster when normalised under each category

Absolute Clustering

We repeat the aforementioned process on the absolute data. Clustering on the normalised dataset did not reveal any specific information as to whether a particular attribute in a cluster dominated. In contrast to the normalised clusters a large shift in the number of customers present within each cluster was found when clustering on absolute data. Here an assignment of customers was based on the magnitude of their energy consumption. The proportion of customers that remained part of the same cluster is presented in Table 6.

Cluster 1	Cluster 2	Cluster 3
41.4% (1460)	13.2% (200)	68.7% (1494)

Table 6: Proportion of customers as a percentage (and number) that remained in their respective cluster from normalised to absolute

Figure 21 depicts the cluster centres related to the absolute loads. Table 7 shows the proportions of customers in each cluster under each demographic category. Within cluster 1, rural off gas customers dominated, with equal proportion attributed to the highest and lowest demand clusters, this is of no significance owing to the small sample set. Cluster 2 represented the proportion of customers with the highest consumption whilst cluster 3 was the lowest consumption group.

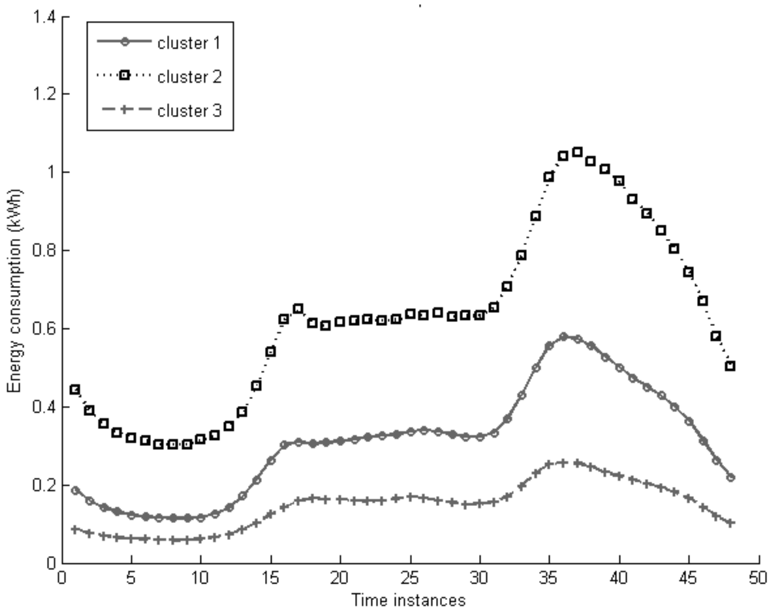


Figure 21: Three cluster centres for the absolute load profiles

Demo-graphic category	Household composition		Tenure		Household income			Household efficiency (proxy for house age)			Rurality measure			
Demo-graphic variable	Non-Dependent	Dependent	Renter	Non - renter	Low income	Medium income	High income	Low efficiency	Medium efficiency	High efficiency	Rural - off gas	Rural	Suburban	Urban
cluster 1	978	1608	669	1917	620	940	1026	792	1217	577	25	327	653	1581
cluster 2	167	313	120	360	101	139	240	126	210	144	10	76	114	280
cluster 3	2248	1900	1658	2490	1919	1352	877	1238	1932	978	10	485	976	2677

Table 7: Customer numbers present within each demographic category for each cluster on absolute data

The total customers in each cluster can be deduced by the total in each demographic category. Out of the total customers present in the data sample of 7214, cluster 2 had the least percentage of customer's, 6.65% of the whole sample, where 50% were high income earners. The majority of customers (57.5%, cluster 3) exhibited low mean energy consumption with 46% low income earners. Clusters 1 had 35.85% of the population with almost an equal proportion of medium and high income earners in the cluster at 36 and 49% respectively. Low income earners made up 23% of the group respectively. Figure 22 depicts the relative proportions of each demographic variable for all clusters. A large displacement is indicative of a higher proportion of customers being present in that category relative to the number of customers in that cluster.

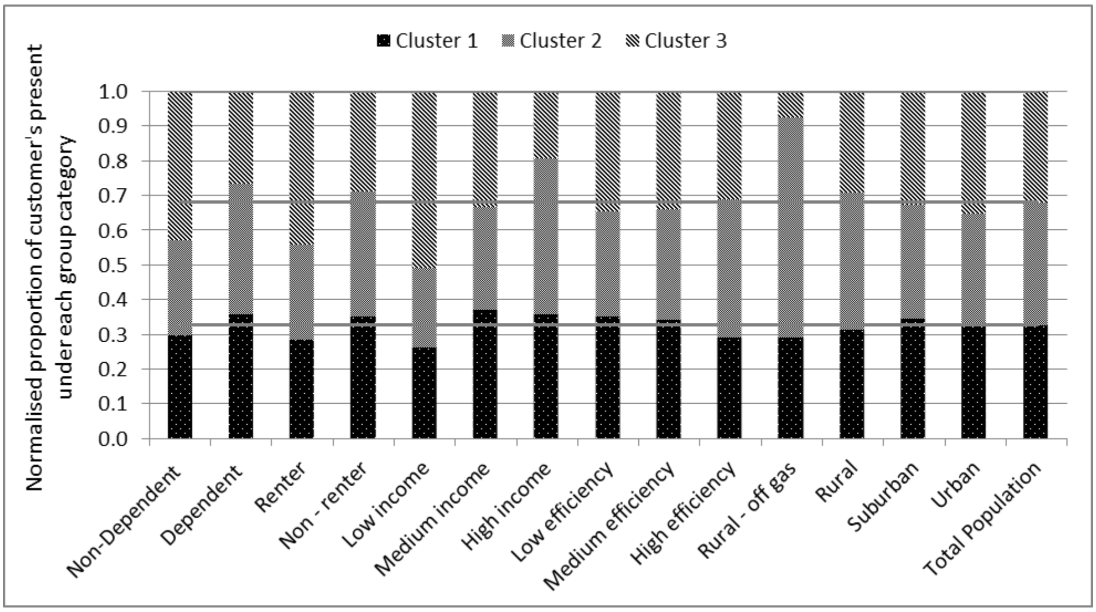


Figure 22: Relative proportions of customer groups present for each of the absolute clusters

Out of all demographic categories income appeared to be the most significant attribute that relates energy consumption. In the dataset a bias exists towards home ownership with twice as many home owners. As indicated by Table 7, home owners are dominant in cluster 1 and 2 ratio with a 3:1 ratio to that of renters. Cluster 3 gave a ratio of renters to non-renters by 3:2 inferring in the CLNR dataset that low energy uses are more likely to rent than own their own house.

Cluster 3 had the largest relative proportion of customers than clusters 1 and 2; similarly high income uses in cluster 3 remained the smallest. With only the efficiency measure to act as a proxy to the household building fabric and age we cannot explore this relationship further. On a cluster by cluster basis the medium efficiency variable was the dominant of the three categorical variables.

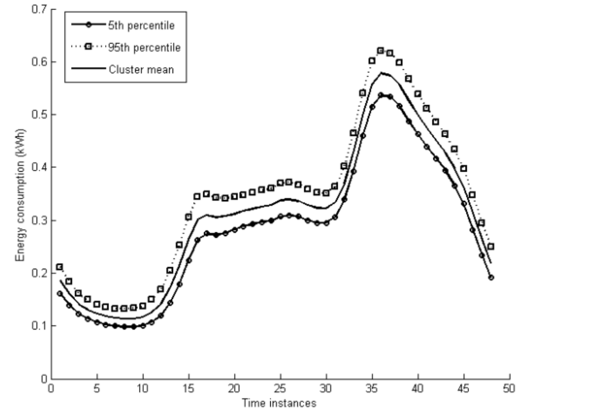
Load Profile Characterisation

Monte Carlo Simulations

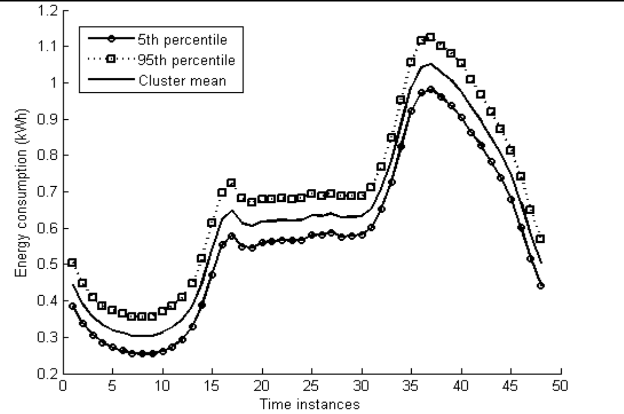
Characteristics of the extremes in demand between clusters were explored. To evaluate the variability within each cluster the cluster members were analysed by repeated sampling. Different possible load characteristics to determine the maximum and minimum variability of each cluster could infer information useful to network designers.

Monte-Carlo simulation was used to generate a set of M random variations of mean customer load profiles. Performing N repeated sampling over the 48 half hour period the maximum and minimum profiles around each cluster was computed. M was taken to be 50 whilst N = 100,000. Figure 23 depicts the 95th and 5th percentiles of the sampled load profiles.

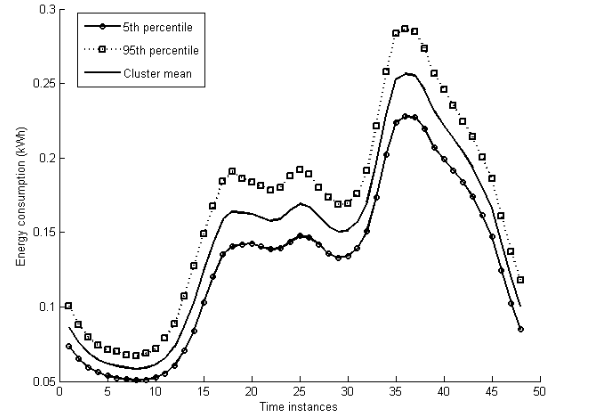
The tighter the variance the more homogeneous the group of customers are. Cluster 1 showed the most homogeneous group of all the clusters. In absolute terms cluster 2 showed the largest variance in early morning. The highest variance is shown in cluster 3 through midday and early evening. The gradients either side of morning or evening peaks show the least variance for all clusters.



Cluster 1 random sampled percentile extremes



Cluster 2 random sampled percentile extremes



Cluster 3 random sampled percentile extremes

Figure 23: 5th & 95th percentiles with mean loads of each cluster through random sampling

Aggregated Load Profile Construction

Having affirmed the relevant proportions of the demographic variables that contribute to the centroids the range of customer's demand within a cluster based on their attributes are explored. Specifically we examine how amalgamating two cluster loads from specific demographic variables varies the consumption profiles. The significant proportion of two clusters was combined in a weighted analysis. We use significance to denote either a large or small customer proportion in a given cluster. The weighted proportions varied from 10% to 90% in 20% increments. Figure 24 depicts variations for a mix of customer types by assigning different cluster proportions to the two clusters. Through observation clusters 2 and 3 were chosen based on their difference measures.

The series of black dashed lines present the higher proportions of cluster 2 and the lower weighted proportions of cluster 3. The composite load profiles of both the higher and lower income customers in cluster 2 reveal the same characteristics as each other. Cluster 2 low income (90%) customers do have a slightly higher night - very early morning and afternoon consumption than all others but lower evening consumption and lower morning (5am to 9am) consumption. This could be attributable to behavioural patterns in the home (TV watching). The whole cluster set and high income customers show effectively the same profile but when the mix is dominated by high income earners, higher evening consumption is witnessed.

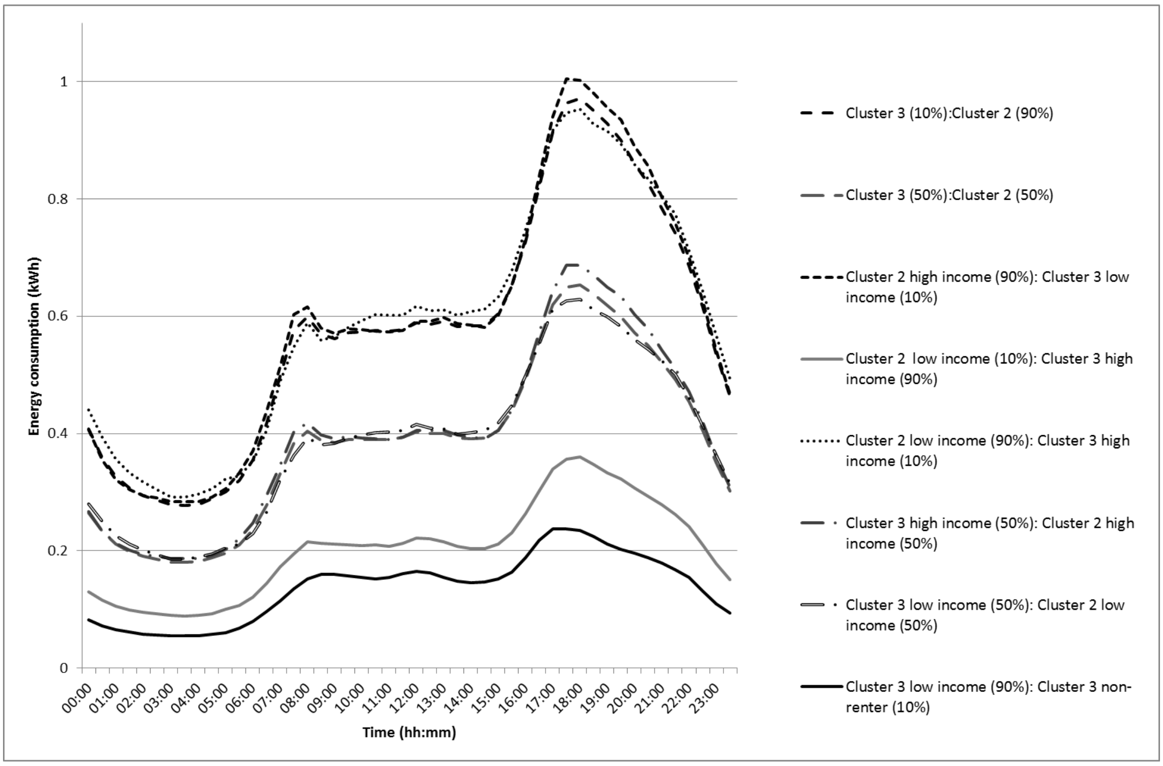


Figure 24: Composite load profiles from different groups in clusters 1 and 4

Income was shown to be the most contributing factor. High income customers have a higher early morning and evening peak, this trend is noticed throughout all weighted means of the two clusters between both extreme income bands. Reducing the weighted proportion of customers from cluster 2 to that of cluster 3 reduces consumption magnitude.

High energy uses tend to show greater 'peakiness' behaviour where low energy customers show a greater dispersion of energy consumption across the peak and into the later evening.

The lowest weighted consumption profile observed was that of a single profile in cluster 3 where 90% was borne from low income and 10% non-renter customers. Increasing the proportion of non-renters to low income users for this cluster yielded a marginal increase in peak consumption.

5.3.3. Conclusion: main demographic factors

It is shown that energy consumption seems weakly related to income where stratifications of low and high income produce the largest effect. The clusters produced did not highlight any other major demographic category. Home occupancy was another variable that had a mild effect on consumption. No correlation could be coupled to home efficiency which indicates that gas heating prevails in our sample set.

Greater granularity would be required to explore such analysis further and in particular see where income levels began to show a difference in energy consumption. The same argument can be made for other demographic variables. It could be possible that if this was the case then refining the clusters further may give a greater insight.

5.4. System peak

Network planners traditionally considered peak demand of base domestic customers, excluding those with low carbon load bearing technologies, around the central winter period to evaluate the sufficiency of distribution network capacity bring around the concept of headroom. Headroom refers to the difference between the load experienced on a network, and the rating (i.e. cable laid direct in ground) of that network. If the rating exceeds the load, then there is a positive amount of headroom and reinforcement is not required. However, once load exceeds the rating then the headroom figure becomes negative and reinforcement to release additional headroom must be undertaken (see [8]).

The analysis within the work explores peak usage from a customer and temporal perspective since traditionally electrical networks are analysed on a peak day arising in midwinter. Analysing on a customer basis we obtain the variability between the different customer groups. Peak demand generally arises through the simultaneous loads from central winter and is the peak of the average demands when all customers are considered together, thus accounting for the diversity of demands.

The peak demand is defined in Table 3, term 1, where the computation was performed during a 2 year period from May 2011 to April 2013 inclusive. In 2013 the peak daily mean demand occurred on the 18th January 2013 at 17:30 (see Figure **25**). Where peak demand months are required, January is chosen as the period of choice with July (six month split) the arbitrary month to assess the peak demand in the summer. The peak demand in the summer was analysed to see if there were significant differences between the groups. The authors acknowledge that for network design the mean minimum demand across customer groups is of most interest considering localised generations on the network through photovoltaics.

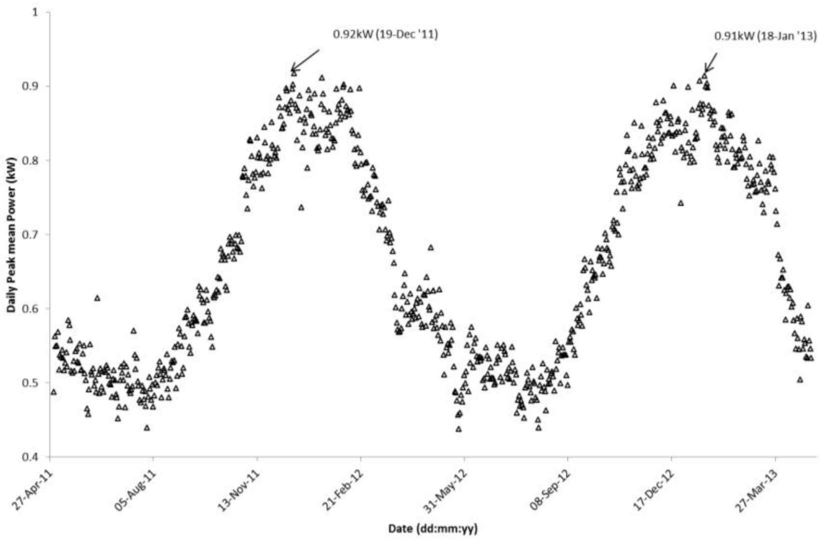


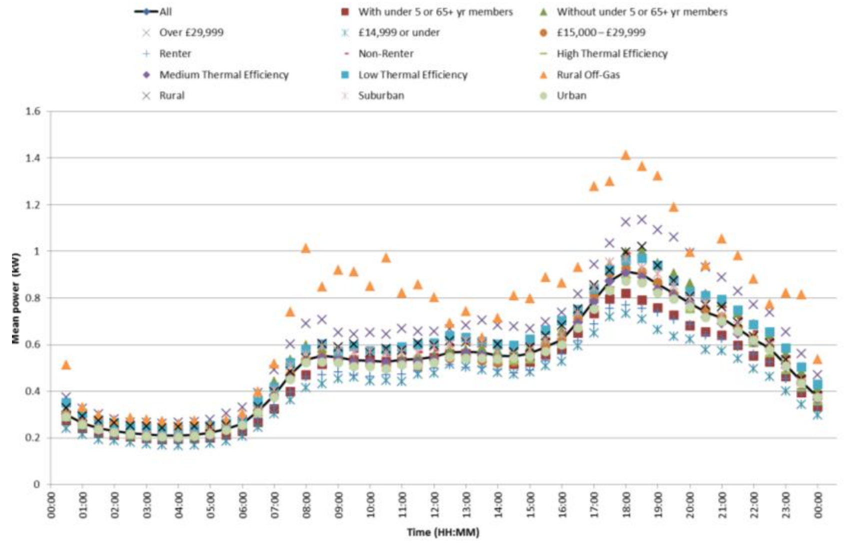
Figure 25: Daily mean peak power demands considering all customers in TC1a across two years (1st May 2011 – 30th April 2013).

The mean demand of the customers for the month in which the highest network stress generally occurred was computed and used in the clustering analysis. Although an adequate approach for choosing the sample peak demand is to explore the difference between groups, it is equally adequate to define the network peak considering all customers in the population since this is the network wide peak.

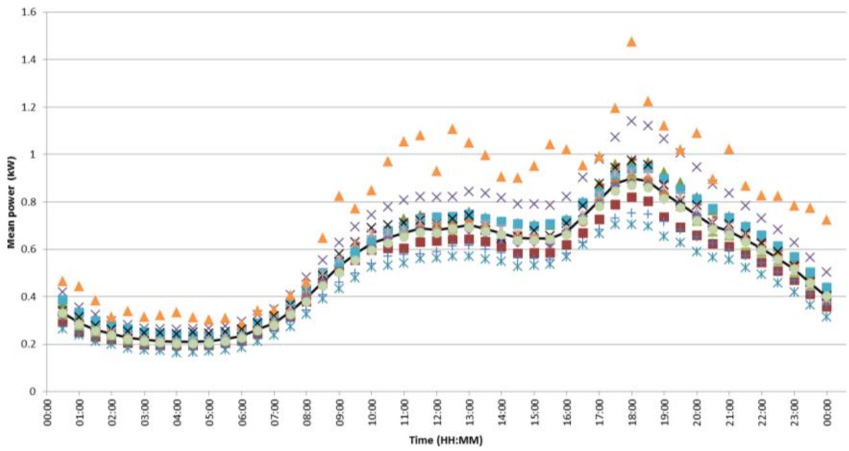
5.4.1. Demographic breakdown

This analysis investigates the applicability of the load curves akin to that produced in the ACE 49 Report [6] to represent the way that the characteristics and behaviour of present customers affects the shape of their energy use. As in the ACE 49 method, each customer's load at each half hour during the day is normalised by their annual consumption since this is seen as an appropriate indicator of peak demand.

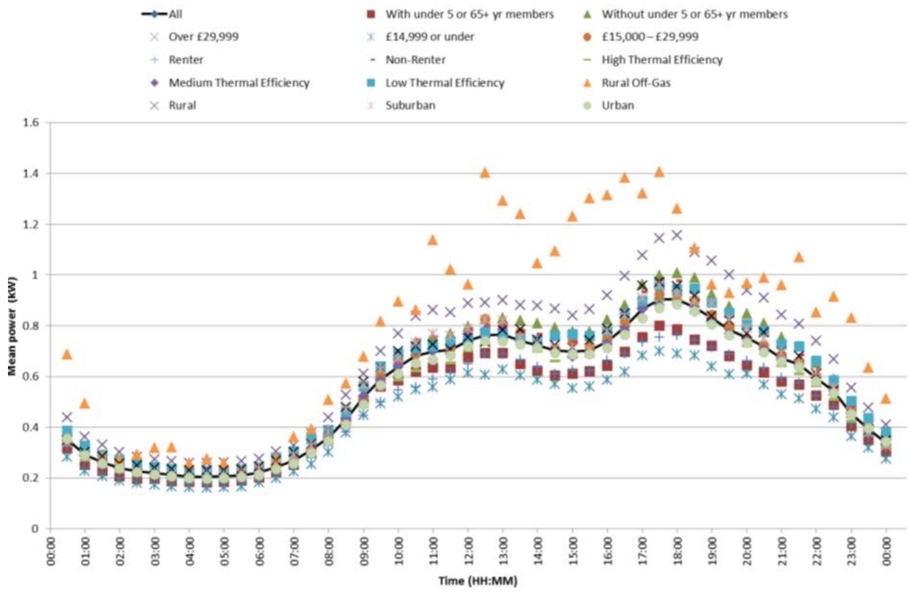
There are two points of interest here, the difference in demand between groups on the day of the group peak demand and the difference in magnitudes between the days. In the 3 days the mean peak demand of the population is given by the solid black line and resides in the envelope of low income and high income customers with the exception of rural off-gas customers. Rural off-gas customers have higher demands owing to electrical loads being required to supply space heating and hot water assuming no oil heating is present. Low thermal efficiency customers also have higher demands, an indication of auxiliary heating.



Peak loads for the 18th January 2013 across all CLNR subgroups



Peak loads for the 19th January 2013 (Saturday) across all CLNR subgroups

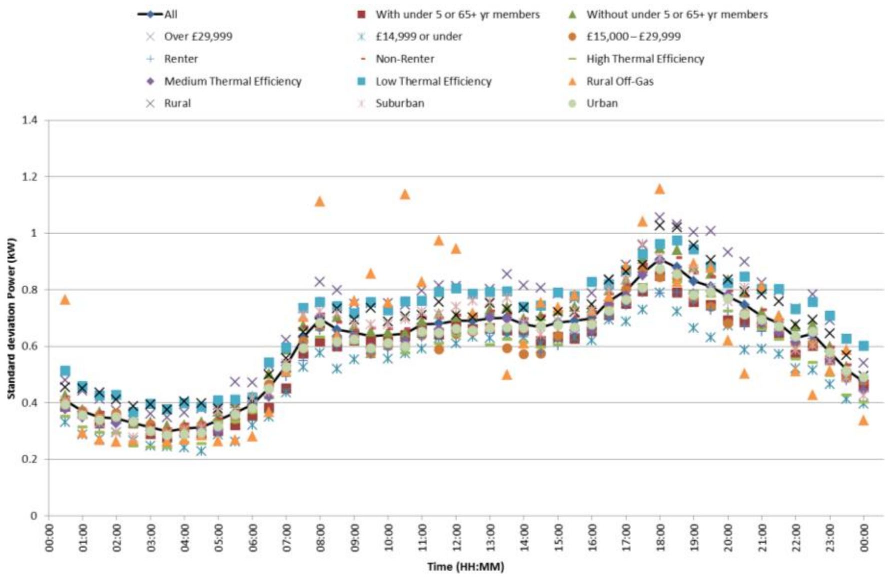


Peak loads for the 20th January 2013 (Sunday) across all CLNR subgroups

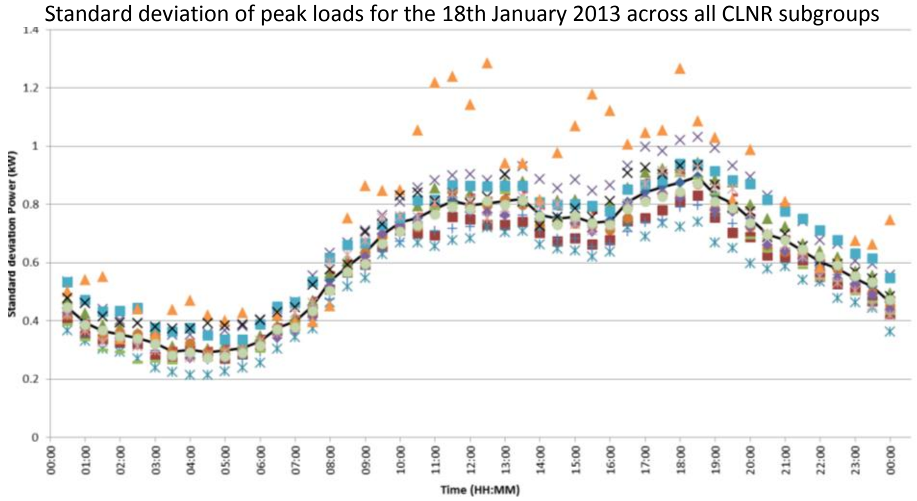
Figure 26: Peak demands a weekday, and Saturday and Sunday in January 2013 as indicated below the figures.

The peak demand of the population occurs at 6pm on the 18th and 19th January and 5:30pm on the 20th January with mean demands of 0.91, 0.90 and 0.90kW respectively. The corresponding standard deviations are similar and are 0.91, 0.87 and 0.90kW. The daytime demand on the Saturday and Sunday experience a higher demand in the daytime which is attributable to higher activity in the household on these days. Sunday carries a higher demand overall. The relevant means and standard deviations for the population over the 3 days across the daytime periods from 10am to 4pm were 0.62, 0.70, 0.77 and 0.70, 0.81 and 0.88kW.

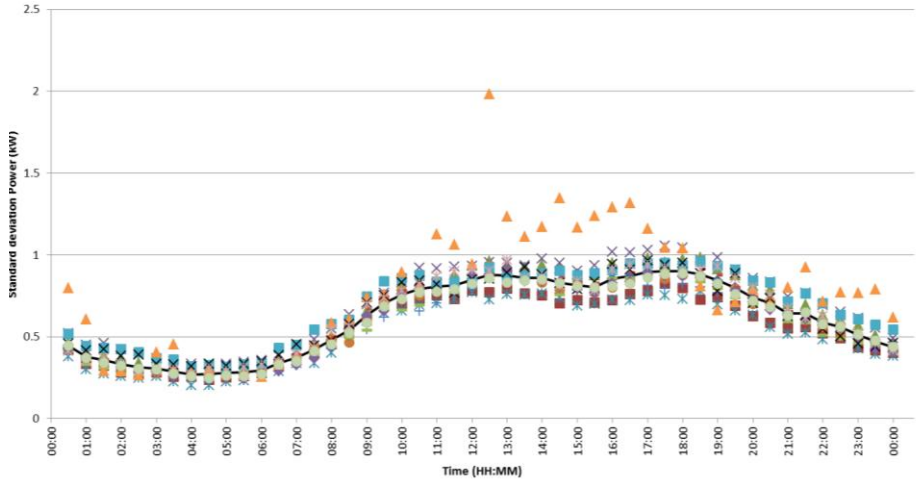
One observation to note regarding peak demand is that although the highest combined demand for the overall TC1a population occurred in January, when looking at the daily peak demands (averaged across any given month, and across each demographic group), February was the month with the highest mean weekday peak demand for all demographic groups except rural off-gas (see Table 15). However, this analysis does not take into account the fact that different customers' peak demands do not necessarily occur simultaneously. This highlights the importance of looking at the aggregate or combined population demand as well as trying to identify patterns for individual households, as it is the cumulative effect that has the greatest impact on the energy system.



Standard deviation of peak loads for the 18th January 2013 across all CLNR subgroups



Standard deviation of peak loads on the 19th January 2013 (Saturday) across all CLNR subgroups



Standard deviation of peak loads for the 20th January 2013 (Sunday) across all CLNR subgroups

Figure 27: Peak demands a weekday, and Saturday and Sunday in January 2013 as indicated below the figures.

5.4.2. Predicting the system peak

The average ambient temperature over the winter period was -0.18°C, this compares to 3.32°C in the winter of 2013 obtained from the Met Office [9].The peak day analysis did not account for any increases in demand due to colder periods in which people are more likely to use more energy. In terms of electrical demand an exploration of meteorological dependencies on peak load demand is presented since having only two years' worth of data the severity of a winters peak demand is unlikely to be tested.

Considering the severity of winter a network operator may wish to know how meteorological factors affect peak demand. An exploration of meteorological dependencies on peak load demand is presented since having only two years' worth of data the severity of a winters peak demand is unlikely to be tested.

The peak demand on each day in the period of study is defined by the maximum of the mean demand. Greenwich mean time (GMT) was the period of study as this is when peak demand is most likely to exist and fell in the range 30/10/2011 to 25/03/2012. From previous studies the peak demand arose on the 15/01/2012 at 17:30 where peak demand $\mathrm{D_t}$ (random variable) is the demand d of customer i on a given day type at peak time. The categorical variables Friday, Saturday and Sunday were listed separately to all other days which were referenced as weekday.

To correlate demand at peak time to a series of predictor ambient temperature and sunrise and sunset variables the following methodology was approached.

The ambient temperature variable was chosen be a four hour moving average of ambient temperature at the current time $\mathrm{T_t}$ and is given by:

$$\mathrm{T_0} = \mathrm{T_t} + \mathrm{T_{t-1}} + \mathrm{T_{t-2}} + \mathrm{T_{t-3}} \tag{6}$$

The constructed moving average temperature variable $\mathrm{T_0}$ was incorporated in a 24 hour lagged variable given by:

$$TE_t = \frac{TE_{t-24}+T_0}{2} \tag{7}$$

where TE at time t is a mean between TE 24 hours before, and current T_0

Sunrise and sunset times will be given as minutes after midnight, $\mathrm{S_{rise}}$and $\mathrm{S_{set}}$ respectively.

The general multiple linear regression equation was formed:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{8}$$

Where β are the regressor variable coefficients and ϵ is the error term

Because demand was regressed over days of the week, days of the week comprised the following categorical variables:

$\mathrm{d_{week}} = \mathrm{Monday} - \mathrm{Thursday\ weekday}$

$\mathrm{d_{Fri}} = \mathrm{Friday}$

$\mathrm{d_{Sat}} = \mathrm{Saturday}$

$\mathrm{d_{Sun}} = \mathrm{Sunday}$

Expanding (8) and incorporating the dummy variables to the regressor derived in (7) we obtain the following full equation where the weekday categorical variable acts as the reference. If there are p categories, we have p − 1 dummy regressor variables. The following equation would denote a change in intercept from the weekday regression line for each categorical variable.

$$\mathrm{Y} = \beta_0 + \beta_1 \mathrm{TE_t} + \beta_2 \mathrm{S_{rise}} + \beta_3 \mathrm{S_{set}} + \beta_{\mathrm{Fri}} \mathrm{d_{Fri}} + \beta_{\mathrm{Sat}} \mathrm{d_{Sat}} + \beta_{\mathrm{Sun}} \mathrm{d_{Sun}} + \epsilon \qquad (9)$$

Where $\beta_0, \ldots \beta_3$ are the intercept and coefficients for the respective regressor variables, S_{rise} and S_{rise} are the variables sunrise and sunset.

The model yielded a goodness of fit with an R-squared value of 0.88. Figure **28** shows the observed and predicted demands with the following coefficients of the regressor variables depicted in **Table 8**.

Variables	Coefficients
Intercept	0.26968
TE	-0.00377
Sunset	-0.00003
Sunrise	0.00044
Friday	-0.00719
Saturday	-0.00580
Sunday	0.00797

Table 8: Regression model coefficients

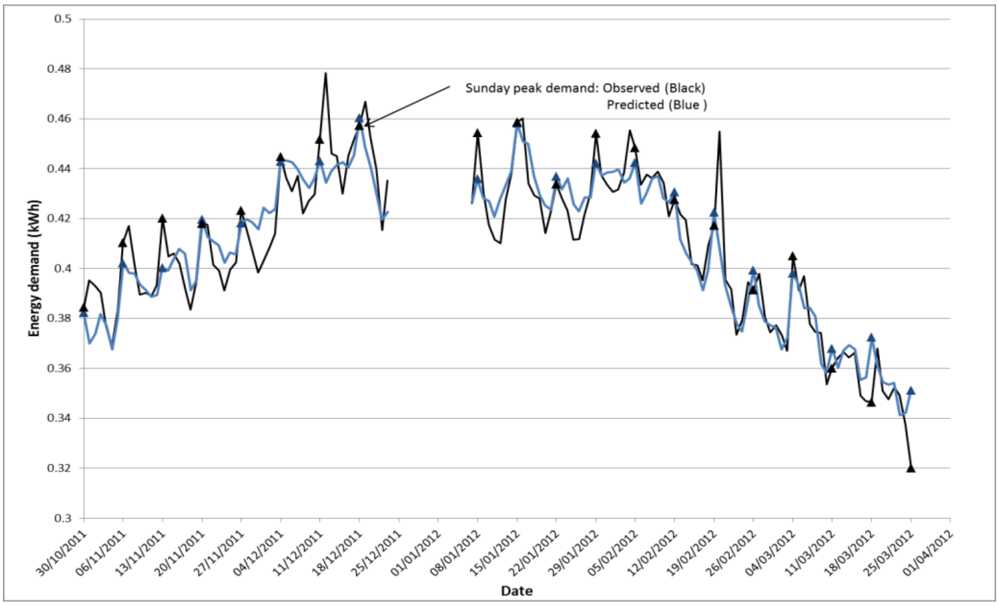


Figure 28: Peak demand of Test Cell 1a and its corresponding predicted demand obtained from regression.

5.5. Correlation of Electrical Demand and Consumption

Electrical network load analysis sometimes uses customer annual electricity consumption as a normalising and / or scaling factor when computing generic load curves, or when calculating predictive loading profiles [6]. It is thus useful to know what degree peak electrical demand is correlated with total electrical energy consumption, over some period of time. To do this several scatter plots were obtained for each of the CLNR and mosaic subgroups where a linear Pearson correlation $\rho_{x,y}$was obtained between annual consumption and peak demand(X, Y):

$$\rho_{x,y} = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

Any customers with demand less than 1000kWh were discarded from the analysis because very low consumptions were regarded as outliers that could not be authenticated. In our sample set consumptions of less than 200kWh were witnessed, so for increased certainty an appropriate filter was constructed to cut off low-demand data points. Annual consumption greater than 15000kWh and peak demand greater than 15kW were discarded for this analysis. **Error! Reference source not found.**Figure 29 shows annual consumption vs peak power customers in Test Cell 1a with a correlation of 0.578.

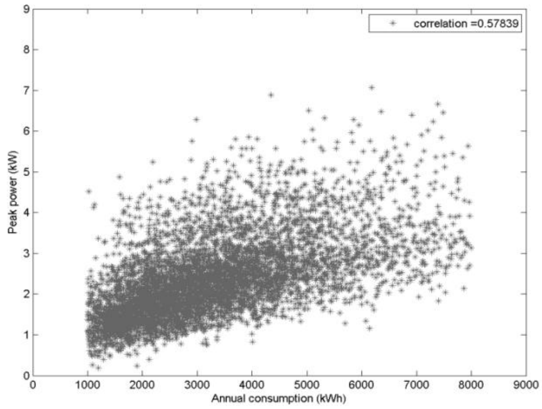


Figure 29: Scatter plot of annual consumption vs peak power for all customers in TC1a.

Figure 30**Error! Reference source not found.** shows scatter plots for four CLNR customer groupings with two of the lowest and highest correlations of peak to annual consumption. Low efficiency customers show the highest correlation of all subgroups where compared to the rural customers (correlation coefficient of 0.562) shows the variance of the outliers to be less. Rural off gas customers showed the peak demand to be least correlated at 0.427 with annual consumption of all plots. Low efficiency and suburban customers showed similar levels of correlation (Suburban correlation = 0.62337, Low efficiency = 0.63357). Of these subgroups 20% of customers were mutually inclusive.

The Mosaic categories (Figure 31) have the potential to predict peak demand slightly better based on annual consumption with higher correlations of 0.683 for Alpha Territory and 0.653 for Rural Solitude. In contrast, the group Rural Solitude had a greater positive correlation than the rural CLNR customer base and ultimately showed demand to be better predicted for more affluent people using the Mosaic criteria. Only 47 Rural Solitude customers were classified as rural in the CLNR group with 15 customers part of the rural off gas customer group the majority of customers were actually suburban when compared to the CLNR rurality group. Elderly needs customers show the least correlation of all Mosaic groups with a correlation coefficient of 0.455.

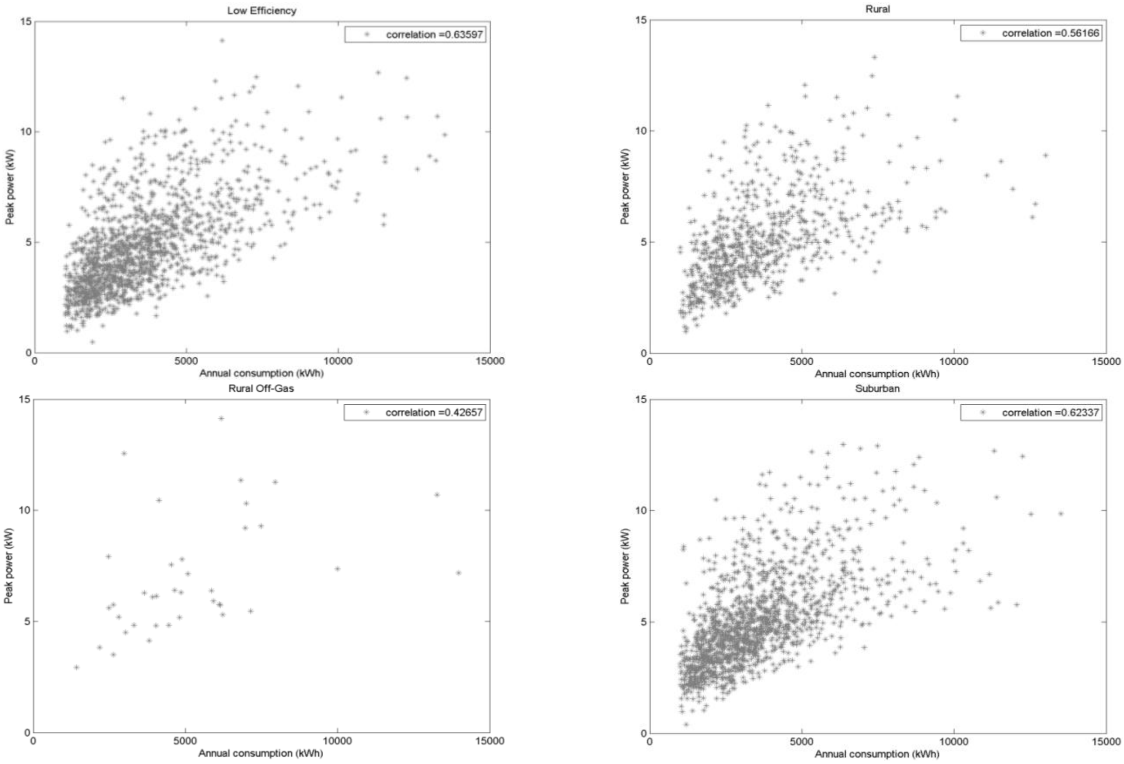
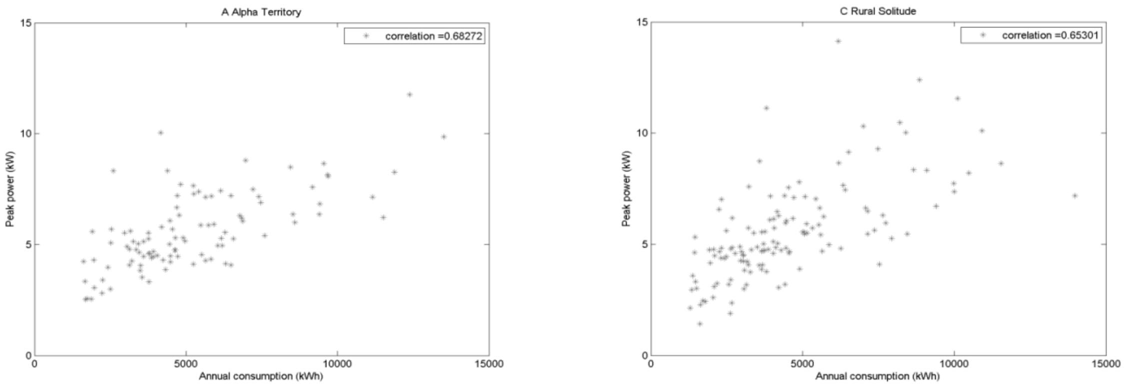


Figure 30: Scatter plot with corresponding correlation of annual consumption vs peak power for TC1a customers when categorised by CLNR sub-groups.



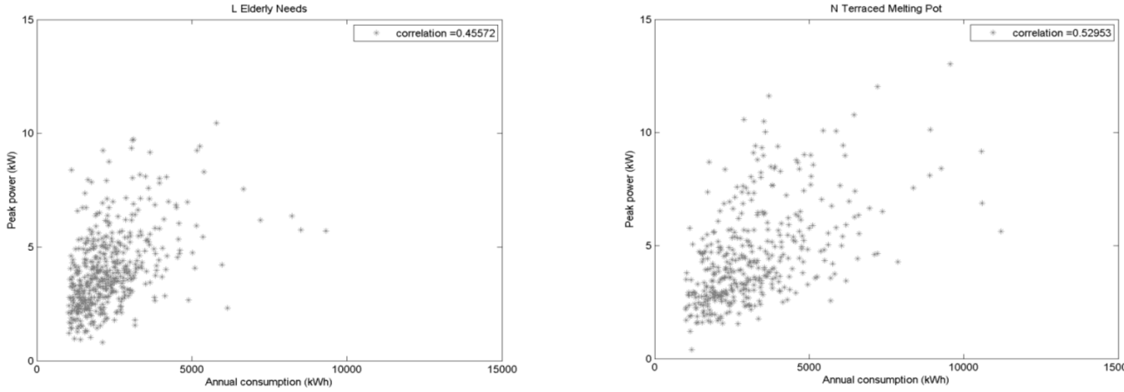


Figure 31: Scatter plot with corresponding correlation of annual consumption vs peak power for TC1a customers when categorised by Mosaic demographics.

6. Discussion

The consistent outcome from the statistical analysis is that the observed standard deviation is typically about half the mean. Given that for a normal distribution, 95% of the data lie within two standard deviations, this highlights the very large variance consistently observed throughout the various groups analysed (both as defined internally and the externally Mosaic defined groups). Within these groups, the only discriminatory factors appear to be income based and rural off gas.

The link between customer consumption and demographics is primarily driven by income. A similar pattern is also observed in the peak demand for these customer groups. Other demographic factors are important in predicting electrical demand and consumption, in the ANOVA test however efficiency wasn't a significant factor in predicting annual consumption but is deemed important when predicting peak demand. The apparent difference in factors could be explained through the use of electrical heaters in times of small ambient temperatures which coincide with peak electrical demand and/or increased electrical activity in this period which subsequently leads to internal gains in the household. For instance when the house is cold, extra activity through the use of electrical appliances could mitigate the need for electric heaters but the effects on peak demand are the same. However this activity is not picked up when explored in annual electrical consumption.

Looking into the consumption patterns based on demographic breakdown provides a mostly consistent picture. Most customers consume about 3500kWh during the year, with a range of roughly 200kWh either side of that value. Notable exceptions here are the high income group (ca 4100kWh) and rural off gas (ca 5300kWh). Further, the low income group (ca 3000kWh) and renter group (ca 3200kWh) are outliers to this picture. There is even less variance in the mean peak demand across groups, with the exception of rural off gas group. On the other hand, the maximum peak demand follows the annual consumption pattern more closely.

The final analysis investigated in more detail the correlation between annual consumption and peak demand. This was done using a scatter plot of these two measurements. What can be seen here is that there is an envelope that bounds the majority of the data. This envelope is consistent across all demographics: it forms a line passing through (1000kWh, 2.5kW) and (5000kWh, 5kW). Most data points lie below this line. This potentially enables estimating of worst case peak demand scenario based on meter consumption data (thereby not requiring a smart meter to measure demand at half hourly intervals).

7. Conclusion

Overall, the demand analysis of Test Cell 1a customers reveals a relatively consistent average demand profile. However, this masks a highly variable underlying demand data set. This high variability is seen both in total consumption and in peak demand. The standard deviation across all groupings (both internally defined and the Mosaic social groupings defined by Experian) is roughly half the mean value for each group. Given that for a normal distribution roughly 95% of all data is expected to lie within two standard deviations of the mean, this takes the lower bound to near zero: this does not provide an informative lower bound. With a similar range above the mean, it becomes very difficult to characterise individual customers based on demand profiles. This variance is seen not only within customer groups, but also within individual customers.

The key determinant to both annual consumption and peak demand are driven by income, as recorded in the demographic data. However, there does not appear to be a 'smooth' transition through income levels: most demographic groups behave very similarly (in terms of annual consumption and peak demand). The exceptions to this are the high income and rural off gas (higher than average), and with low income and renters groups (lower than average).

A number of analysis methods were used on the demand data with the aim to link demographics to load profile:

1. Experian's Mosaic groupings do separate customers into different consumption and peak behaviours than the CLNR demographic groupings.
    a. In exploring the variability between different groups the Mosaic groups give a larger variance around the mean compared to the CLNR demographics. This can be explained due to the lower sample sizes and is not indicative of better defined demographics.
    b. Relationships of peak demand to annual electricity consumption are more consistent and show higher correlations in more socio-demographic groups to that of the CLNR datasets.
2. Higher income related demographics show generally higher electrical demand than other customer types with Mosaic group ‘Alpha territory’ and the CLNR high income the high consumer groups in their respective demographic types.
3. Income is the primary driver for electricity use, although the correlation is weak. This is evident throughout all aspects of the analysis and especially evident in the clustering analysis.
4. Renters consume less electricity than owners which is likely to be income related.

Beyond the income based link, other demographic characteristics were found to have an impact on energy consumption, although on their own they were not able to predict annual consumption or demand peak. From the demand data, the variance on the various metrics could not be reduced by the groups that were either defined a priori (internally define groups and Mosaic), or by groups defined by clustering algorithms.

8. References

[1] CLNR-L052 - "CLNR Durham University Social Science Research - April 2014 Report". 2014

[2] Domestic LO1 and LO2 Qualitative Paper

[3] M. Blell, "Customer Led Network Revolution: Revised Selection Criteria for Domestic Test Cells," DEI-CLNR-R008, Sep. 2011

[4] http://www.experian.co.uk/assets/business-strategies/brochures/Mosaic_UK_2009_brochure.pdf

[5] F. McLoughlin, A. Duffy, and M. Conlon, "A parametric analysis of domestic electricity consumption patterns in Ireland," in *Environment and Electrical Engineering (EEEIC), 2011 10th International Conference on*, 2011, pp. 1–4.

[6] "Report on Statistical Method for Calculating Demands and Voltage Regulations on LV Radial Distribution Systems," Energy Networks Association, A.C.E. Report No. 49, 1981.

[7] CLNR-L217 After Diversity Maximum Demand Report.

[8] CLNR-L185, (2014) "Review of the Distribution Network Planning and Design Standards for the Future Low Carbon Electricity System Reliability Evaluation of Power Systems". 2.

[9] http://www.metoffice.gov.uk/climate/uk/datasets/

9. Glossary of terms

CLNR	Customer-Led Network Revolution
DEI	Durham Energy Institute
DSR	Demand Side Response
HP	Heat Pump
NEA	National Energy Action
PV	Photovoltaic
TC	Test Cell
ToU	Time of Use (Tariff)
WWG	Wet White Goods

10. Appendices

Age Range	Customer numbers	%
20-30	94	3.09%
31-40	268	8.82%
41-50	460	15.13%
51-60	549	18.06%
61-70	669	22.01%
71-80	526	17.30%
81-90	227	7.47%
91-100	36	1.18%
Unknown	211	6.94%
Total	3040	100.00%

Table 9: Contribution of customer ages in a portion of British Gas' business portfolio

Table 10: Annual consumption for each Experian demographic group

Demographic	Mean (kW)	Standard deviation
A, Alpha territory	5399.123	2878.068
B, Professional rewards	4308.41	2054.347
C, Rural solitude	4677.403	2659.781
D, Small town diversity	3514.447	1678.426
E, Active retirement	2777.237	1314.677
F, Suburban mind-sets	3975.683	1938.237
G, Careers and kids	3966.406	2137.707
H, New homemakers	2812.009	1260.839
I, Ex-council community	3468.685	1702.071
J, Claimant cultures	3104.43	1721.833
K, Upper floor living	3213.897	1738.529
L, Elderly needs	2343.82	1099.721
M, Industrial heritage	3232.075	1455.913
N, Terraced melting pot	3289.976	2366.982
O, Liberal opinions	3653.654	2364.19

Table 11: Annual consumption

Annual consumption (means)	Mean	Standard deviation
Low income	2954.967	1578.56
Medium income	3581.738	1945.291
High income	4134.928	2115.547
Low efficiency	3640.461	2258.28
Medium efficiency	3444.168	1828.649
High efficiency	3496.438	1789.479
Non renter	3653.169	1895.898
Renter	3232.034	1967.13
Rural	3732.484	2038.59
Rural off gas	5336.846	2739.336
Suburban	3542.134	1877
Urban	3429.101	1907.053

Table 12: Max peak demand for each Experian demographic group, month dependent for weekday MAX(P_D,WD,JAN)

Demographic	Month	Mean (kW)	Standard deviation
A, Alpha territory	Dec	4.62875	1.673
B, Professional rewards	Dec	3.701094	1.798
C, Rural solitude	Dec	4.065246	1.634
D, Small town diversity	Dec	4.386864	1.951
E, Active retirement	Dec	3.807649	1.728
F, Suburban mind-sets	Mar	3.171523	1.734
G, Careers and kids	Dec	4.097372	1.718
H, New homemakers	Dec	4.04344	1.844
I, Ex-council community	Jan	3.45879	1.583
J, Claimant cultures	Mar	3.901371	1.832
K, Upper floor living	Mar	3.437136	1.699
L, Elderly needs	Mar	3.561353	2.06
M, Industrial heritage	Mar	2.914519	1.426
N, Terraced melting pot	Jan	3.667517	2.074
O, Liberal opinions	Mar	3.467157	1.833

Table 13: Max peak demand for each CLNR demographic group, month dependent for weekday MAX(P_D,WD,JAN)

Demographic	Month	Mean	Standard deviation
With dependants	Jan	3.246	1.597
Without dependants	Dec	3.833	1.905
Low income	Mar	3.1464	1.680
Medium income	Dec	3.674161	1.926
High income	Dec	3.95223	1.705
Low efficiency	Dec	3.515668	1.947
Medium efficiency	Dec	3.576367	1.771
High efficiency	Mar	3.531796	1.686
Non renter	Mar	3.67376	1.695
Renter	Mar	3.429533	1.784
Rural	Mar	3.788498	1.731
Rural off gas	Feb	4.52641	2.324
Suburban	Jan	3.634948	1.752
Urban	Feb	3.44956	1.686

Table 14: Mean peak demand for each Experian demographic group, month dependent for weekday.

Demographic	Month	Mean (kW)	Standard deviation
A, Alpha territory	Feb	2.656066	1.154
B, Professional rewards	Feb	2.234768	1.001
C, Rural solitude	Feb	2.359253	1.194
D, Small town diversity	Feb	2.08728	1.026
E, Active retirement	Feb	1.629468	0.797
F, Suburban mind-sets	Feb	2.31143	1.065
G, Careers and kids	Feb	2.269127	1.076
H, New homemakers	Feb	1.910464	0.999
I, Ex-council community	Feb	2.125088	1.099
J, Claimant cultures	Feb	1.83574	0.993
K, Upper floor living	Feb	1.919257	1.236
L, Elderly needs	Feb	1.421348	0.725
M, Industrial heritage	Feb	2.008716	1.0
N, Terraced melting pot	Feb	1.892213	1.07
O, Liberal opinions	Feb	2.145184	1.34

Table 15: Mean peak demand for each CLNR demographic group, month dependent for weekday.

Demographic	Month	Mean	Standard deviation
With dependants	Feb	1.802	0.965
Without dependants	Feb	2.162	1.110
Low income	Feb	1.705949	0.977
Medium income	Feb	2.076326	1.076
High income	Feb	2.256643	1.048
Low efficiency	Feb	1.996344	1.148
Medium efficiency	Feb	1.985418	1.028
High efficiency	Feb	1.980228	1.023
Non renter	Feb	2.050981	1.04
Renter	Feb	1.866999	1.077
Rural	Feb	2.057069	1.047
Rural off gas	Mar	2.64626	1.262
Suburban	Feb	2.05664	1.088
Urban	Feb	1.938336	1.038

Table 16: Max peak demand for each Experian demographic group, month dependent for weekend.

Demographic	Month	Mean (kW)	Standard deviation
A, Alpha territory	Jan	4.037154	1.633
B, Professional rewards	Mar	3.30682	1.646
C, Rural solitude	Mar	3.492249	1.482
D, Small town diversity	Mar	3.691182	1.748
E, Active retirement	Mar	3.346896	1.578
F, Suburban mind-sets	Mar	2.770255	1.387
G, Careers and kids	Mar	3.710251	1.617
H, New homemakers	Mar	3.693802	1.646
I, Ex-council community	Jan	3.097626	1.50
J, Claimant cultures	Mar	3.428909	1.667
K, Upper floor living	Mar	3.056199	1.66
L, Elderly needs	Mar	3.193912	1.831
M, Industrial heritage	Mar	2.509146	1.372
N, Terraced melting pot	Mar	3.208602	1.652
O, Liberal opinions	Mar	3.130693	1.714794

Table 17: Max peak demand for each CLNR demographic group, month dependent for weekend.

Demographic	Month	Mean	Standard deviation
With dependants	Mar	3.031	1.597
Without dependants	Mar	3.648	1.806
Low income	Mar	2.947485	1.685
Medium income	Mar	3.487457	1.776
High income	Mar	3.719626	1.639
Low efficiency	Mar	3.29945	1.788
Medium efficiency	Mar	3.37084	1.718
High efficiency	Mar	3.351266	1.714
Non renter	Mar	3.428279	1.702
Renter	Mar	3.198055	1.784
Rural	Mar	3.485055	1.778
Rural off gas	Mar	4.434256	1.896
Suburban	Mar	3.492472	1.812
Urban	Mar	3.252499	1.683

Table 18: Mean peak demand for each Experian demographic group, month dependent for weekend.

Demographic	Month	Mean (kW)	Standard deviation
A, Alpha territory	Jan	2.737878	1.260
B, Professional rewards	Feb	2.069247	1.065
C, Rural solitude	Jan	2.292716	1.037
D, Small town diversity	Feb	2.460717	1.208
E, Active retirement	Feb	2.123886	1.039
F, Suburban mind-sets	Mar	1.645173	0.894
G, Careers and kids	Feb	2.371661	1.081
H, New homemakers	Jan	2.289625	1.149
I, Ex-council community	Jan	1.976779	1.051
J, Claimant cultures	Feb	2.130741	1.096
K, Upper floor living	Feb	1.856266	1.014
L, Elderly needs	Mar	1.955352	1.12
M, Industrial heritage	Feb	1.455789	0.764
N, Terraced melting pot	Feb	2.007195	0.993
O, Liberal opinions	Feb	1.961939	1.03

Table 19: Mean peak demand for each CLNR demographic group, month dependent for weekend.

Demographic	Month	Mean	Standard deviation
With dependants	Feb	1.887	1.028
Without dependants	Feb	2.320	1.202
Low income	Feb	1.801829	1.05
Medium income	Feb	2.208453	1.175
High income	Feb	2.401097	1.12
Low efficiency	Feb	2.113894	1.225
Medium efficiency	Feb	2.112606	1.112
High efficiency	Feb	2.097938	1.114
Non renter	Feb	2.17307	1.122
Renter	Feb	1.988658	1.166
Rural	Feb	2.196743	1.146
Rural off gas	Feb	2.83177	1.363
Suburban	Feb	2.183113	1.192
Urban	Feb	2.054385	1.112

Table 20: Annual consumption confidence limits (95%) for each of the CLNR customer groups.

customer	Lower limit (kWh)	Upper limit (kWh)
With dependencies	3170.16	3310.34
Without dependencies	3686.54	3821.98
Renter	3147.31	3317.61
Non-Renter	3590.44	3709.70
Low Income	2888.64	3019.91
Medium Income	3493.08	3664.14
High Income	4033.60	4233.42
Low Efficiency	3523.47	3757.63
Medium Efficiency	3373.05	3510.24
High Efficiency	3410.92	3577.40
Urban	3366.08	3488.74
Suburban	3449.59	3643.65
Rural	3652.48	3948.12
Rural Off-Gas	4392.34	6131.87
All	3456.14	3554.41

Table 21: Half hour peak demand confidence limits (95%) for each of the CLNR customer group. The CI for the whole group was 4.74 and 4.83 kW.

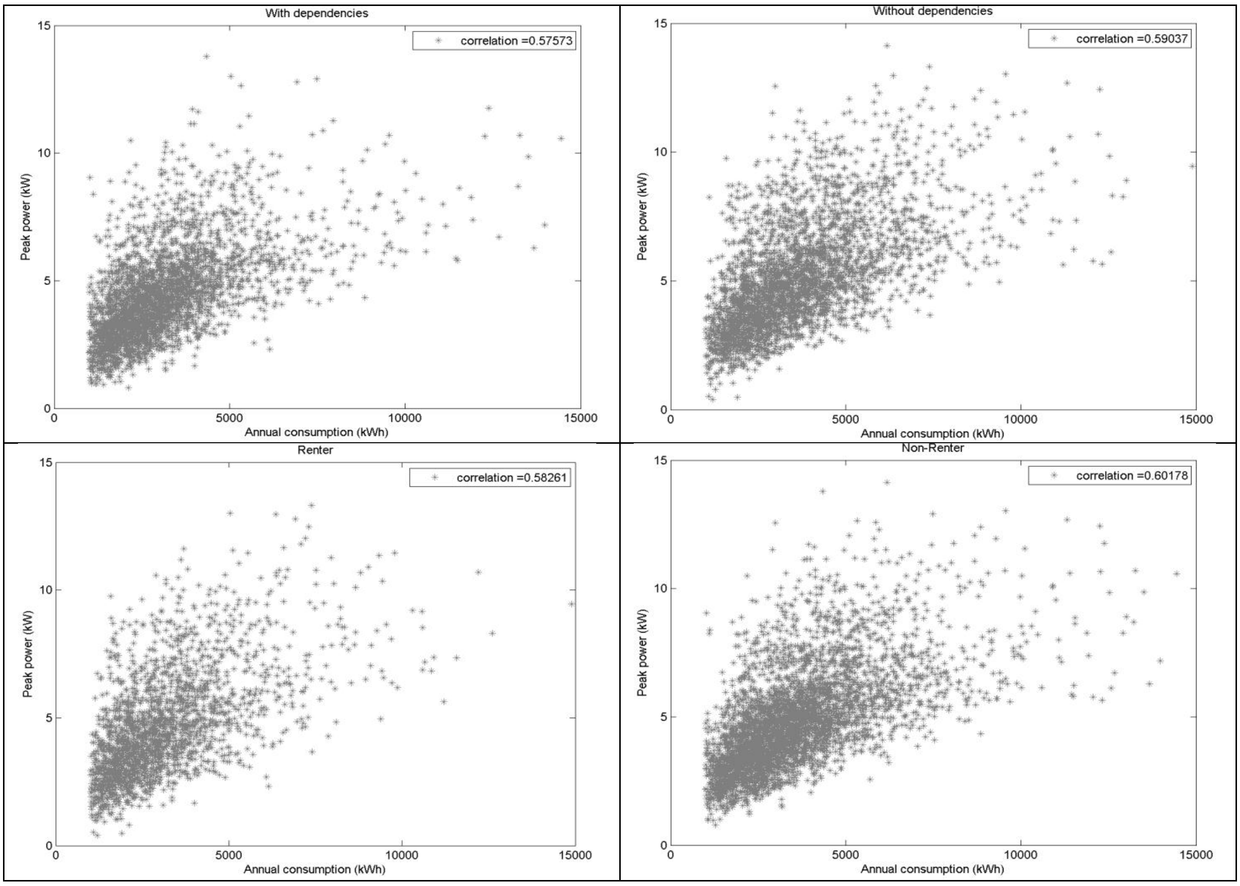
customer	Lower limit (kW)	Upper limit (kW)
With dependencies	4.34	4.46
Without dependencies	5.06	5.20
Renter	4.54	4.70
Non-Renter	4.82	4.93
Low Income	4.21	4.36
Medium Income	4.85	5.01
High Income	5.16	5.32
Low Efficiency	4.65	4.84
Medium Efficiency	4.76	4.90
High Efficiency	4.66	4.82
Urban	4.61	4.73
Suburban	4.81	4.99
Rural	4.96	5.24
Rural Off-Gas	5.90	7.28

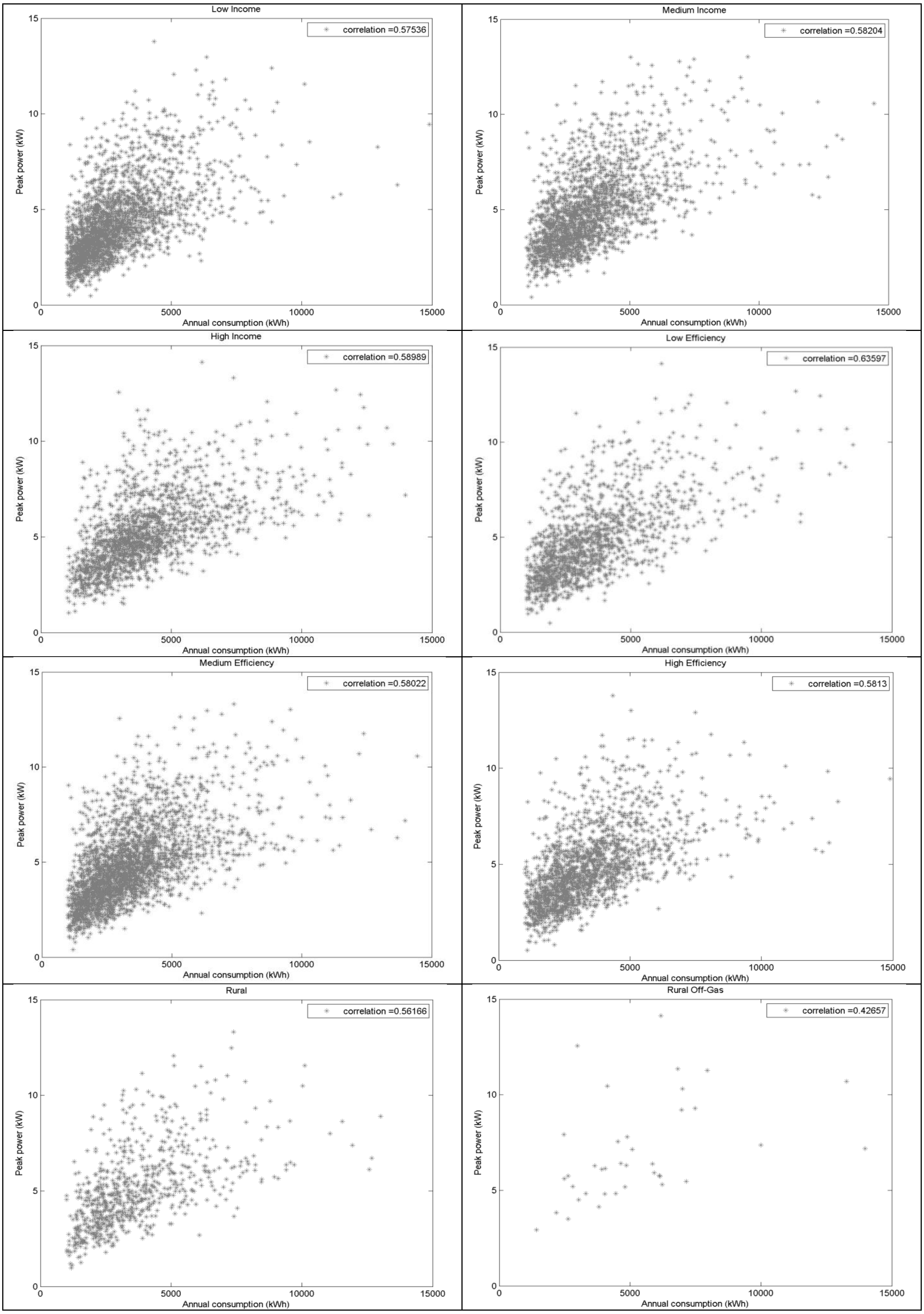
Table 22: Intercept 8.1319

$\beta_1 = \begin{bmatrix} 0 \\ -0.082599 \end{bmatrix} \equiv \begin{bmatrix} No\ dependencies \\ \text{with dependencies} \end{bmatrix}$
$\beta_2 = \begin{bmatrix} 0 \\ -0.10664 \end{bmatrix} \equiv \begin{bmatrix} Non-Renter \\ \text{Renter} \end{bmatrix}$
$\beta_3 = \begin{bmatrix} -0.10941 \\ 0 \\ 0.13924 \end{bmatrix} \equiv \begin{bmatrix} \text{Low income} \\ Medium\ income \\ \text{High income} \end{bmatrix}$
$\beta_4 = \begin{bmatrix} 0.060523 \\ 0.32246 \\ 0.015905 \\ 0 \end{bmatrix} \equiv \begin{bmatrix} \text{Rural} \\ \text{Rural Off} - \text{Gas} \\ Suburban \\ Urban \end{bmatrix}$
$\beta_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.1362 \\ 0 \\ -0.017994 \end{bmatrix} \equiv \begin{bmatrix} \text{No dependencies} \cdot \text{Low income} \\ \text{No dependencies} \cdot \text{Medium income} \\ \text{No dependencies} \cdot \text{High income} \\ \text{with dependencies} \cdot \text{Low income} \\ \text{with dependencies} \cdot \text{Medium income} \\ \text{with dependencies} \cdot \text{High income} \end{bmatrix}$

Table 23: model intercept 1.6096 and error term

$\beta_1 = \begin{bmatrix} 0 \\ -0.10284 \end{bmatrix} \equiv \begin{bmatrix} No\ dependencies \\ \text{with dependencies} \end{bmatrix}$
$\beta_2 = \begin{bmatrix} 0 \\ -0.050239 \end{bmatrix} \equiv \begin{bmatrix} Non-Renter \\ \text{Renter} \end{bmatrix}$
$\beta_3 = \begin{bmatrix} -0.095427 \\ 0 \\ 0.071585 \end{bmatrix} \equiv \begin{bmatrix} \text{Low income} \\ Medium\ income \\ \text{High income} \end{bmatrix}$
$\beta_4 = \begin{bmatrix} -0.043667 \\ 0 \\ -0.023259 \end{bmatrix} \equiv \begin{bmatrix} \text{Low efficiency} \\ Medium\ efficiency \\ \text{High efficiency} \end{bmatrix}$
$\beta_5 = \begin{bmatrix} 0.0191 \\ 0.25815 \\ 0 \\ -0.053135 \end{bmatrix} \equiv \begin{bmatrix} \text{Rural} \\ \text{Rural Off} - \text{Gas} \\ Suburban \\ \text{Urban} \end{bmatrix}$
$\beta_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.082911 \\ 0 \\ 0.0026112 \end{bmatrix} \equiv \begin{bmatrix} \text{No dependencies} \cdot \text{Low income} \\ \text{No dependencies} \cdot \text{Medium income} \\ \text{No dependencies} \cdot \text{High income} \\ \text{with dependencies} \cdot \text{Low income} \\ \text{with dependencies} \cdot \text{Medium income} \\ \text{with dependencies} \cdot \text{High income} \end{bmatrix}$




Low Income
correlation =0.57536
Medium Income
correlation =0.58204
High Income
correlation =0.58989
Low Efficiency
correlation =0.63597
Medium Efficiency
correlation =0.58022
High Efficiency
correlation =0.5813
Rural
correlation =0.56166
Rural Off-Gas
correlation =0.42657
Peak power (kW)
Annual consumption (kWh)
0
5
10
15
0
5000
10000
15000

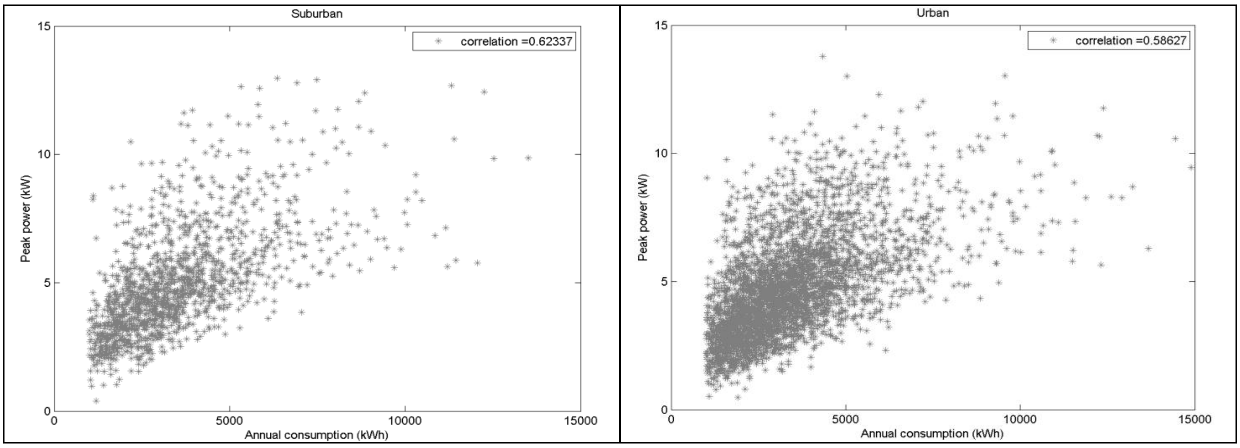


Figure 32: Scatter plot with corresponding correlation of annual consumption vs peak power for TC1a customers when categorised by CLNR sub-groups.

Table 24: Summary statistics for the CLNR stratification variables with corresponding sample size.

DEI stratefication variable	mean	std	90th percentile	Sample size
All	3498.4	1866.6	5845.9	5941
Renter	3206.6	1774.1	5387.8	2059
Non-Renter	3653.2	1895.9	6053.9	3882
With dependants	3228.3	1808.1	5270.5	2876
Without dependants	3751.8	1885.3	6280.7	3065
Low Income	2955.0	1578.6	4982.5	2230
Medium Income	3564.9	1794.2	5722.4	1987
High Income	4124.7	2073.0	6835.6	1724
Low Efficiency	3604.4	2022.6	6244.8	1430
Medium Efficiency	3444.2	1828.6	5654.1	2731
High Efficiency	3496.4	1789.5	5708.8	1780
Rural Off-Gas	5336.8	2739.3	8083.8	39
Rural	3732.5	2038.6	6351.4	744
Suburban	3542.1	1877.0	5981.8	1447
Urban	3415.1	1800.5	5637.2	3711

Table 25: 95th percentile confidence intervals for the peak demand and annual consumption of the CLNR subgroups.

	Peak demand (kW)		Yearly consumption (kWh)	
DEI stratification variable	**Lower CI**	**Upper CI**	**peaks_CI**	**peaks_CI**
With dependants	4.3	4.5	3170.2	3310.3
Without dependants	5.1	5.2	3686.5	3822.0
Renter	4.5	4.7	3147.3	3317.6
Non-Renter	4.8	4.9	3590.4	3709.7
Low Income	4.2	4.4	2888.6	3019.9
Medium Income	4.8	5.0	3493.1	3664.1
High Income	5.2	5.3	4033.6	4233.4
Low Efficiency	4.6	4.8	3523.5	3757.6
Medium Efficiency	4.8	4.9	3373.0	3510.2
High Efficiency	4.7	4.8	3410.9	3577.4
Urban	4.6	4.7	3366.1	3488.7
Suburban	4.8	5.0	3449.6	3643.7
Rural	5.0	5.2	3652.5	3948.1
Rural Off-Gas	5.9	7.3	4392.3	6131.9
All	4.7	4.8	3456.1	3554.4

Table 26: 95th percentile confidence intervals for the peak demand and annual consumption for the mosaic categories

	Peak demand (kW)		Annual consumptio (kWh)	
Mosaic demographic groups	**Lower CI**	**Upper CI**	**Lower CI**	**Upper CI**
A Alpha Territory	5.3	6.1	4846.0	5952.3
B Professional Rewards	5.0	5.3	4131.3	4485.5
C Rural Solitude	5.3	6.0	4228.7	5126.1
D Small Town Diversity	4.7	5.0	3393.3	3635.6
E Active Retirement	4.0	4.5	2613.3	2941.2
F Suburban Mindsets	5.1	5.4	3833.6	4117.8
G Careers AND Kids	5.0	5.5	3704.5	4228.3
H NEW Homemakers	4.2	4.8	2590.1	3033.9
I Ex-Council Community	4.9	5.2	3354.4	3583.0
J Claimant Cultures	4.4	4.9	2933.8	3275.1
K UPPER FLOOR Living	4.1	5.2	2806.6	3621.2
L Elderly Needs	3.8	4.1	2245.6	2442.0
M Industrial Heritage	4.5	4.9	3119.0	3345.1
N Terraced Melting Pot	4.3	4.7	3059.4	3520.5
O Liberal Opinions	4.6	5.3	3287.3	4020.0