Using Energy Smart Meter Data and Socio-Demographic Indicators to Understand Consumption in Relation to Deprivation Levels

Ellen Talbot

18/04/2018

## Abbreviations

CDRC – Consumer Data Research Centre  
DEP – Domestic Energy Provider  
PCS – Postcode Sector (e.g. OL9 0)  
SM – Smart Meter

# Abstract

To do.

# Introduction

Domestic energy consumption continues to grow (Hargreaves, Nye, and Burgess 2010), and during the period 1990 and 2004, energy use in the domestic sector increased by 18%, meaning that this accounted for just over 30% of the UKs overall energy consumption is accounted for (Faiers, Cook, and Neame 2007). In 2007, Perez-Lomard et al (2008) situated the UK in a global context; The then current figure of 28% was higher than that of both Spain and the USA (22% and 15% respectively). Compared to the EU average of 26%, the UK was again higher. More recent figures show that figure is still increasing; in 2013 it accounted for 35% of energy consumption.

The 2008 Climate Change Act in the UK requires a 34% reduction in 1990 greenhouse gas emissions by 2020 and an 80% reduction by 2050. Dwellings are an important target area for reductions and the UK government set the goal of a 29% decrease in this area by 2020 (HM Government 2009). As such, the installation of smart meter technology that enable the collection of digital records of gas and electricity consumption are viewed as integral to encouraging greater efficiency through the measurement and monitoring of consumer behaviour alongside more efficiently producing and storing resources to meet the demand (UK CCC 2010; Guerreiro et al. 2015; “Britain’s Smart Grid” 2018).

This research has two related objectives; the first is to explore the representativeness and spatio-temporal signatures of aggregate residential energy consumption as recorded by a national extract of smart meter data provided by one of the 5 largest energy providers in the UK; and secondly, to evaluate the socio-demographic determinants of recorded energy consumption; and their relationship to fuel poverty. This analysis capitalises on data provided to the Consumer Data Research Centre by the Domestic Energy Provider (DEP) for the financial year 2015 - 16 across the UK. To the authors knowledge this is the first dataset of it’s kind to be analysed at the half hourly cadence and the first energy study in the UK to expore both gas and electricity consumption in tandem at this magnitude.

## Energy, Monitoring and Consumer Behaviour

Energy meters have been an essential but modest element of the infrastructure in UK households since the early 20th century - firstly gas, and then with the invention of alternating current, electricity - both being produced on a large, saleable scale. By the turn of the century many improvements had been made to the original model inside peoples homes; reductions in weight and dimension; compensation of changes of power factor, voltage and temperature, elimination of friction, and improving longterm stability. These induction meters are still widely produced today in their modern form, and are popular due to their low production price and excellent reliability (**???**).

Yet with ever increasing and changing power demands, the UK is upgrading it’s infrastructure to a ‘Smart Grid’, which “has the potential to solve a lot of our energy problems” (“Britain’s Smart Grid” 2018; Stern 2011). It will improve reliability, better match supply and demand and enable more efficient planning for both short and long term future; by being more secure, planning ahead for unexpected power outages and informing the number of power stations the UK is likely to need in the future. Smart meters enable the smart grid by providing real-time data on the consumption of energy (“Britain’s Smart Grid” 2018; Guerreiro et al. 2015). Smart meters are quickly replacing conventional meters in many parts of the world and the move to smart metering in the UK is predicted to be complete for both gas and electricity if all goes as planned by 2020(Haben, Singleton, and Grindrod 2016). This deadline was proposed to align with the targets set out in the 2008 Climate Change Act (HM Government 2009). It is no small feat and the largest change faced by the energy industry since the changeover to North Sea Gas (Darby 2010) where the project to refit 40 million appliances took 10 years and cost £500m (Gas 2018).

Of the literature reveiwed there was a clear disparity in the meaning of the word ‘energy’. Of the 23 titles which included the words ‘smart meter’, ‘energy’ or ‘electricity’, over half focus explicitly on electricity and feature it in the title, [Jones and Lomas (2015); McLoughlin, Duffy, and Conlon (2015); Ardakanian et al. (2014); Viegas et al. (2016); Rhodes et al. (2014); Kavousian, Rajagopal, and Fischer (2013); López et al. (2011); Tso and Yau (2007); Huebner et al. (2016); Figueiredo2005] while seven more use ‘energy’ as a reference to electricity (Hargreaves, Nye, and Burgess 2013; Kwac et al. 2013; Alahakoon and Yu 2013; Flath et al. 2012a; Haben, Singleton, and Grindrod 2016; Parra, Quilumba, and Arcos 2016; Yohanis et al. 2008). Huebner et al. (2015), Druckman and Jackson (2008), Faiers, Cook, and Neame (2007), Wyatt (2013) and Brounen, Kok, and Quigley (2012) are the only authors to review both gas and electricity - but the sample sizes were significantly smaller than in our study. This is oft the case in these energy studies, and of all the studies reviewed, the largest sample of housing was Brounen, Kok, and Quigley (2012) with 300,000 Dutch homes. As far as the authors can establish, there are not yet any known studies with a focus purely on gas consumption and smart meters.

Within this sphere of ‘energy’, tangents included the categorisation of new customers based on existing load profiles (McLoughlin, Duffy, and Conlon 2015; Figueiredo et al. 2005; Haben, Singleton, and Grindrod 2016; Viegas et al. 2016; Albert and Rajagopal 2013; Rhodes et al. 2014; Lavin and Klabjan 2016), the reduction of energy consumption through smart meter installation (Abrahamse and Steg 2011; Brandon and Lewis 1999; Brounen, Kok, and Quigley 2012) and issues of privacy, which are arising now the data is of a high enough granularity to be able to predict significant individual occupancy patterns within households, increasing the potential for reidentification through external and internal invasions of privacy (McKenna, Richardson, and Thomson 2012; Buchmann et al. 2013). Buchmann et al. (2013) ’s analysis shows that re-identification rates are particularly high, even using only simplistic methods, whereas McKenna, Richardson, and Thomson (2012) states that though 15 minute to 1 hour intervals in the data give a strong indication of occupancy, there is little potential for revealing individual appliance usage.

## Demographics

Past literature has identified key factors that influence energy consumption in a domstic setting, both dwelling and socio-demographic (Ramos and Vale 2008; Jones and Lomas 2015; Kavousian, Rajagopal, and Fischer 2013; Huebner et al. 2015; Huebner et al. 2016; Druckman and Jackson 2008; beckel2014revealing; Wyatt 2013).

Time and again the existing literature looks at demographic predictors of energy usage. Frederiks, Stenner, and Hobman (2015), [pp576] provide a thorough review of this literature, arguing that despite many inconsistencies within the framework, there is now general agreement that “several broad yet interrelated categories of variables may explain individual differences in household energy use”. These explanatory variables encompass three main fields; socio-demographic (e.g., income, education, household size, dwelling type and tenure), psychological factors (e.g., knowledge, values, attitudes) and external factors (e.g., economics, political and legal). Whilst it is important to recognise the potential impact of macro-level predictors such as external factors, it is beyond the scope of this paper to include them. This is also true of the psychological factors, which require extensive qualitative research to corroborate any findings.

Our literature review undertaken to establish those factors previously identified as having an effect on energy demand in a domestic setting was used to influence the framework of the current study. Overwhelmingly, the literature pinpoints household income as the most positively related socio-demographic variable in regards to domestic energy consumption [Gatersleben, Steg, and Vlek (2002); ONeill2002; Poortinga2004; Abrahamse2011; Rhodes2014; Jones2015; Druckman2008]. However, a study by Kavousian, Rajagopal, and Fischer (2013) found that there was neither a positive or negative relationship between income and energy consumption, instead finding external factors such as the weather conditions and location as more influential.

Income is intrinsically linked to other factors such as employment status, education and household size, all of which may facilitate or constrain energy related behaviours. As an example; those in full time employment (and therefore have a higher income) are more likely to spend less time at home than the retired and unemployed, but they are also more likely to own and use more electrical appliances than lower income households [@Gatersleben, Steg, and Vlek (2002); Holloway2006]. Employment status can be used as a proxy for household income but there is only limited evidence to show how it indirectly affects energy consumption (Frederiks, Stenner, and Hobman 2015).

The literature also suggest tenure as a highly influential variable on energy consumption, if only indirectly. Sardianou (2007) argues that because tenants in privately rented accommodation are constrained, they are the least likely to undertake energy conservation home improvements as they will see very little return on their investment. Likewise, private landlords are often unwilling to invest because they are not responsible for the bills. Home-owners are more likely to invest in energy conservation as they are less transient, have greater financial security and receive greater return on energy efficiency investments in the long term (Frederiks, Stenner, and Hobman 2015). However, Jones and Lomas (2015), found that households had the same likelihood of being high consumers irrespective of mode of tenure. Other earlier studies have also concluded that tenure type has no significant effect on electricity use (Kavousian, Rajagopal, and Fischer 2013 ; Bedir, Hasselaar, and Itard 2013; Leahy and Lyons Sean 2010; Tso and Yau 2007).

“The stages of a family’s life cycle – typically defined as a combination of criteria such as family members’ age, marital status, and size/type – appears to be one of the strongest predictors of household energy consumption” (Frederiks, Stenner, and Hobman 2015, p585). It is suggested that energy usage is likely to peak during the child-rearing years due to the difference in household needs such as; housework and entertainment as well as occupied rooms (Frey and LaBay 1983). However, this can be hard to measure, as some of these values are not explicitly measured by the census and proxy variables may be needed. Some studies take particular interest in the number of appliances in a household, which may be indicative of a families life cycle - but sample sizes are small and generally only one survey is done for the explicit puropse of the study (McLoughlin, Duffy, and Conlon 2015; viegas2016classification; Rhodes2014; weiss2012leveraging; Leahy and Lyons Sean 2010).

Even though family life cycle is considered one of the strongest indicators of household energy consumption, age and gender when considered as standalone variables have an insignificant effect, with only weak statistical links to variability in energy consumption (O’Neill and Chen 2002). There is evidence in the literature of both those in the middle life stages and those in younger and older households showing peak energy consumption [Poortinga, Steg, and Vlek (2004); Abrahamse2011]. Other factors considered in research include psychological and behavioural factors, but go beyond the remit of this paper, using extensive qualitative research to corroborate their findings (Abrahamse and Steg 2011).

Characteristics associated with the dwelling itself have also been linked to occupant’s variations in energy consumption. Dwelling age, type and size are influential, as is the existing equipment within the dwelling. The number of features such as floors, room and windows as well as the installation of insulation and efficient heating all contribute [Raaij and Verhallen (1983); Jones2015]. Schipper et al. (1989) suggest that up to half of total household energy use depends on these characteristics.

As discussed, many of the papers make links between socio-demographic factors and energy usage, but only Druckman and Jackson (2008) refer directly to statistical measures of deprivation deprivation, such as IMD or Townsend scores. They take the IMD and OAC categorisation of small areas, yet they do not use actual consumption measures but a proxy measure for energy from the expenditure and food survey.

There are also several studies which use the magnitude of smart meter data to test new exploratory data techniques. The focus of the discovery in all cases uses data mining techniques to build demand profiles, before clustering the profiles using the maximum and minimum usage features to establish different user groups. None of these studies however, link the demand profiles with socio-demographic data and instead focus on comparing the performance of their chosen data minig techniques [Parra, Quilumba, and Arcos (2016); alahakoon2013advanced; Figueiredo2005; lavin2016clustering; kwac2013utility]. Ardakanian et al. (2014) argues that generating consumption profiles is one of the fundamental data mining operations achievable through smart meter data - using household features captured through the profiles to understand different categories of consumers.

This paper extends the current state of the art by not only using a dataset of a magnitude never seen before, but also encompasses statistical representations of deprivation to link these to energy usage.

## Smartmeter Energy Readings

Data mining is defined as the process that integrates a number of areas; extract, process and obtain useful information from a given dataset and is comprised of five basic actions: sorting, pre-processing, transformation, mining and interpretation (Parra, Quilumba, and Arcos 2016).

Because the Smart Meter data was provided in a raw format from the DEP, data preparation and data cleaning steps were undertaken to ensure usability and accuracy in the later analysis. Data pre-processing is an essential aspect of the data analysis in order to avoid influencing the data with atypical values (Ramos and Vale 2008). Those rows which were an exact duplicate of another were removed and assumed to be a computational error and rows which were missing categorical variables were also removed as these cannot be accurately imputed without biasing the overall result (Field, Miles, and Field 2012).

Outliers were determined as those variables with a value over 3SD away from the mean, as these can be considered extreme outliers (Field, Miles, and Field 2012). Those instances which were identified as outliers, but below 3SD were kept on the basis that they could be accounted for by different energy consuming devices - for instance, a household with an electric car will use significantly more energy during overnight hours but should not be considered an outlier.

With all outliers removed, the frequencies of missing data across the whole dataset and the aggregated to postcode sector (PCS) level were investigated. Data could be missing for a number of reasons - as the infrastructure behind smart meters starts to span more and more technical components, the opportunity for these to fail also increases and leads to gaps in the data (Flath et al. 2012b). The spread of missing data was mapped to understand if there was a spatial aspect to the missing data - rurality and social factors such as people not feeling engaged with smart metering could both have an effect. It has been confirmed by the DEP that there was no specific rollout programme in place in the beginning and users only had to be at home during the day to have a smart meter fitted, therefore it is important to consider the bias implications of those included within the dataset.

Data imputation was then carried out on all missing values. It was decided that imputation by the mean was likely to present the most accurate results (Lavin and Klabjan 2016; Field, Miles, and Field 2012). The data was aggregated to the PCS level and the mean for each half hourly reading per PCS was imputed, resulting in a complete dataset.

As part of the data transformation process, the data dimensions were manipulated to preserve the anonymity of the users and prevent re-identification by aggregating the results from individual user to PCS level. This also acted as a data minimisation method, to make the analysis less expensive, in both time and computational sense (Jiawei et al. 2012). Table 1 visualises this data mining process and shows the number of records removed overall.

Table 1: The Data Cleaning Process

|  |
| --- |
| *Step 1: Prior to data cleaning*  Total number of individual records: 292,855,095 |
| *Step 2: Remove missing values*  Number of individual records removed: 1,275  Total number of records remaining: 292,853,820 |
| *Step 3: Remove missing data information*  Number of individual records removed: 66  Total number of baskets remaining: 292,853,754  *Step 4: Aggregation to Postcode Sector*  Total number of records: 6,141,494  **Final number of records to be included in the analysis: 6,141,494** |

**Do up to here Friday afternoon**

The remainder of the paper should then present empirical results as:

1. Energy Profiles – this would look at the profiles in aggregate and be mainly descirpive – include the OAC stuff here…. This would lead into the interesting patterns with deprived areas
2. Fuel Poverty – this would present the towsend work; this needs introduction and definition plus citation of where it has been used in other fields. You should include a much brief discussion of the census data aggregation as part of this

**Census Data Reweighting**

To get deeper insight into the characteristics of the Smart Meter users census data was appended to the energy usage data. Before this could be done it was necessary to undertake a reweighting of the census variables to PCS level to ensure they were apportioned correctly as they are collected at different geographical scales.

This was done using the postcode headcount dataset **(reference - how to cite a dataset?)** aggregated to postcode sector level and an OA to PCS lookup table to establish the proportion of the OA population contained within each PCS. When compared, this method is preferable to the GIS method of creating a shapefile based on PCS areas which overlap and intersect OAs as this is prone to inaccuracies such as slither polygons and is computationally much more intensive. A PCS is made up of on average 300 houses, of which there were at the last official count in 2014, 12,381 unique postcode sectors (Statistics 2018).

Further to this a Townsend score was calculated for each postcode sector and proportions of OAC group categories were also assigned to each PCS. A Townsend score is a measure of relative deprivation and can be calculated for any areas where information is available for the four index variables; percentage of homes with access to a car or van, percentage of homes with more than 1 person per room, percentage of homes not owner occupied and percentage of people economically active yet unemployed (Townsend 1987; Morris and Carstairs 1991). The overall Townsend score is a total of all four, converted to a z-score and is useful for understanding deprivation as it adheres most closely to the concepts of material disadvantage (Morris and Carstairs 1991). Because these variables are provided at OA level, it was necessary to reweight them to PCS level before ascribing a score based on the weighting of the OA population in each PCS. This is done by multiplying the census variable by the ‘OA proportion’ variable and recalculating a new total count for each PCS by it’s sometimes multiple OAs.

# Data analysis

## Descriptive analysis

Because of the innovative nature of the dataset, it is first imperative to grasp what characteristics are underlying (Lavin and Klabjan 2016). The overall trend was that the number of Smart Meters included increased incrementally over the 12 month period which the data spans, with slightly more electricity smart meters overall than gas. Rates of electricity Smart Meters also increased more sharply than gas over the 12 months.

Overall usage for the entire dataset at a diurnal granularity is presented in Figure 1, showing that as a whole, energy usage peaks during the evening at around 19:30 - when typically most members of the family are at home - therefore most appliances, heating and lighting are likely to be in use. It drops off very quickly after 9pm, most likely because as people go to bed for the night they make the most effort to turn off appliances and not use energy unnecessarily. This overnight rate is sometimes known as the ‘standby rate’ or ‘vampire power’ (Wyatt 2013).



Usage (kWh)

Half hour readings 24hr

Figure 1: Total energy consumption by half hourly reading

There is also a clear AM peak between 07:30 and 08:00, confirming that energy usage increases as people begin their daily tasks, but dips off again as people go out to work. The lowest energy usage occurs between the hours of 01:00 and 04:30, but does not go as low as zero because of the ‘standby’ rate of energy usage, where only those appliances which are not switched off such as the fridge freezer or central heating are consuming energy (Wyatt 2013). The fact that energy usage does not dip this low at any other time of the day indicates that in some cases energy is consumed throughout the day because people are at home through illness, unemployement, caring duties or shift work.

Figures 2 and 3 display energy usage at higher temporal granularities and shows both weekly and seasonal variations; lower energy usage during the typical working week and high usage at weekends when people tend to be at home, and very high usage in the winter months when heating requirements are much greater than in the summer. This is also likely to reflect shorter hours of darkness resulting in less lighting being used during the summer months, and other energy saving measures such as the ability to dry clothes outside in warmer weather.



Day of the week

Usage (kWh)

Figure 2: Weekly energy usage pattern



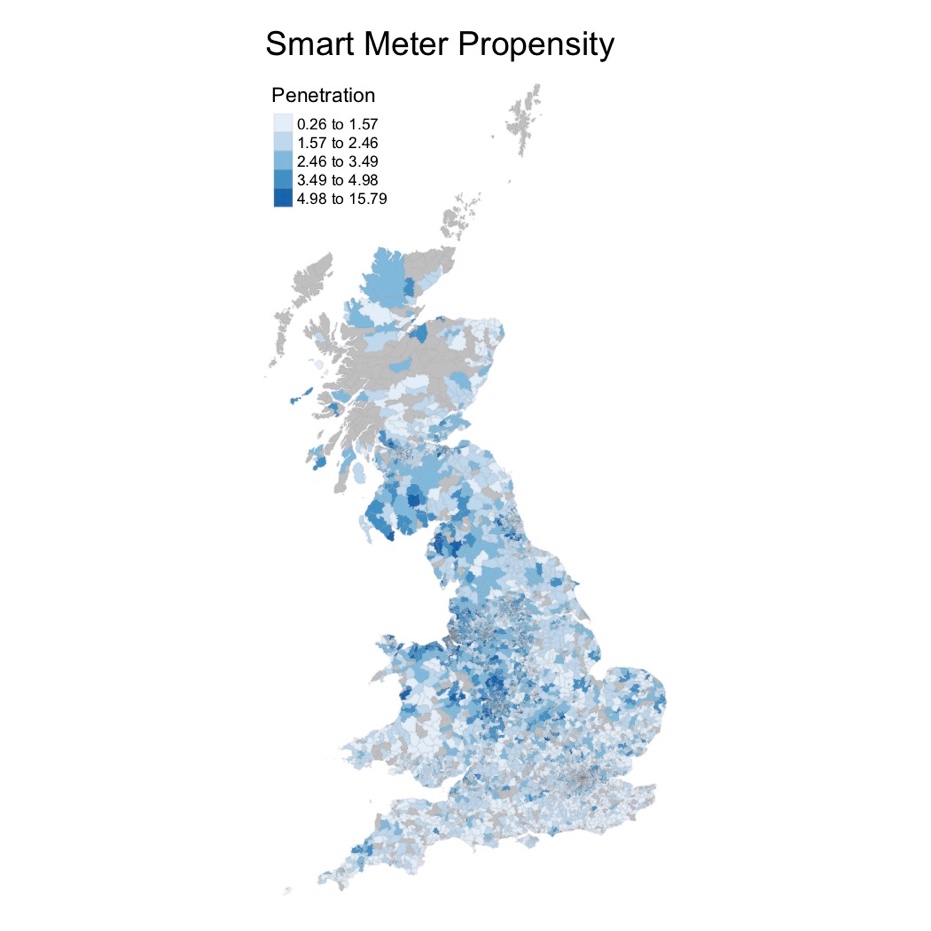
Month

Usage (kWh)

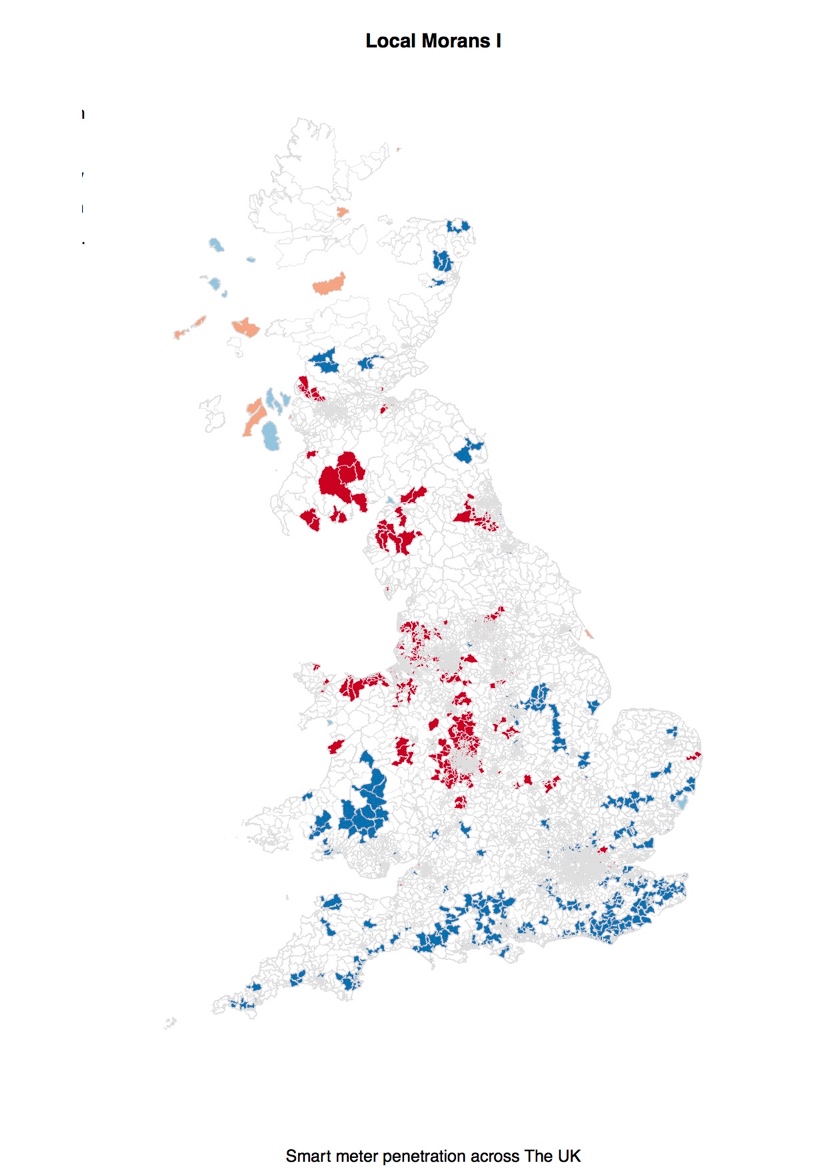
Figure 3: Seasonal energy usage pattern

The spatiality of the distribution of Smart Meters was also investigated and Figure 4 shows the spread as a percentage of total households per PCS. The North West and Midlands are the most prevalent area, while the Scottish Highlands are particularly sparse. This does suggest that there are some urban/rural variations in either the ease of installation or availability of the smart meter infrastructure into homes. Nowhere at the time of this data collection in the UK has higher than 16% of homes with a smart meter (whose energy is provided by this particular DEP).

To get a better understanding of the spatial correlation within the dataset, both a Global Morans I and a ‘local indices of spatial correlation analysis’ (LISA) were undertaken, which in turn determined that the penetration data was spatially correlated, as shown in Figure 5. There is an clear north south divide, and whilst any reasoning for this would be purely speculative, it is true to say that in the North of the UK, high penetration areas are much more likely to be surrounded by high penetration areas, whereas in the South of the UK the exact opposite is true, with the pattern being especially prominent along the southern coastlines. The Global Morans I statistic has a positive value of 4.98, significant to 0.05, suggesting that these patterns have not occurred randomly. Again, Scotland is particularly sparse.



*Figure* 4*: Seasonal energy usage pattern*



*Figure 5: LISA map of Smart Meter Penetration across England, Scotland and Wales*

**Results and discussion**

**OAC groups and energy usage.**

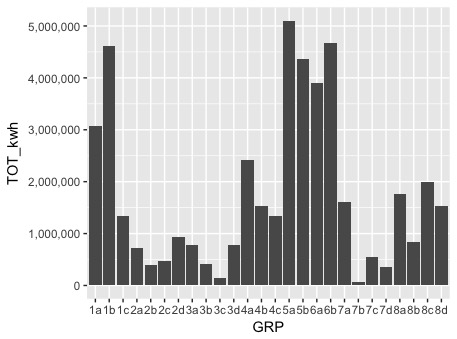
Total energy usage was calculated and appended to the OAC categorisation to understand if households area classification has an effect on their energy usage patterns. Figure 6 shows this relationship.

Figure 6: OAC groups and their overall energy usage

Table 2 provides a brief description of the groups. The most interesting patterns occur in the highest and lowest usage values. Group 5a; The urban professionals and families use the most energy overall. The pen portrait for this group as provided by the Office for National Statistics states that;

“The population of this group shows a noticeably higher proportion of children aged 0 to 14 than the parent supergroup and a lower proportion aged 90 and over. There is also a higher proportion of people with mixed ethnicity. Households in this group are more likely to live in terraced properties and to live in privately rented accommodation. Unemployment is slightly higher than for the parent supergroup.”

|  |  |
| --- | --- |
| *Table 2: OAC subgroup descriptive names* | |
| OAC subgroup | Description |
| 1a | Farming Communities |
| 1b | Rural Tenants |
| 1c | Ageing Rural Dwellers |
| 2a | Students Around Campus |
| 2b | Inner-City Students |
| 2c | Comfortable Cosmopolitans |
| 2d | Aspiring and Affluent |
| 3a | Ethnic Family Life |
| 3b | Endeavouring Ethnic Mix |
| 3c | Ethnic Dynamics |
| 3d | Aspirational Techies |
| 4a | Rented Family Living |
| 4b | Challenged Asian Terraces |
| 4c | Asian Traits |
| 5a | Urban Professionals and Families |
| 5b | Ageing Urban Living |
| 6a | Suburban Achievers |
| 6b | Semi-Detached Suburbia |
| 7a | Challenged Diversity |
| 7b | Constrained Flat Dwellers |
| 7c | White Communities |
| 7d | Ageing City Dwellers |
| 8a | Industrious Communities |
| 8b | Challenged Terraced Workers |
| 8c | Hard-Pressed Ageing Workers |
| 8d | Migration and Churn |

This suggests that family size and having children living at home is linked to energy usage. It also suggests that energy usage is higher due to being at home consistently during the day through unemployment and child care duties. It may also back up the literature that states that privately rented accommodation are the most likely to use the greatest amounts of energy due to them being comparatively poorly insulated and energy efficient, as landlords know they will see little return on this kind of investment when tenants are typically responsible for the utility bills. The tenants are also unwilling to make improvements of this nature as they are also unlikely to see any long-term benefit from investment on a house they do not own.

Another group worthy of noting is Group 1b; Rural tenants, whose pen portrait reads:

“The age structure is very similar to the supergroup^, though people are less likely to live in communal establishments. Compared with the parent supergroup, there is a higher proportion of households living in semi-detached, terraced properties and flats, with a higher proportion socially renting. People are less likely to work in the agriculture industry than for the parent supergroup.” ^Middle aged to older and retired.

Their energy usage is high overall, but also significantly higher than those within their parent supergroup. Again this can be attributed to tenure and the fact that some of these residents are retired and therefore are in their homes most of the day, but these homes may not be the most energy efficient. It may also look comparatively high to those in the other categories of the supergroup as they are less likely to live in communal housing such as retirement homes.

The lowest consumption groups are also worth consideration. Group 7b: Constrained flat dwellers have the lowest energy usage of any group. Looking at the pen portrait:

“This group is characterised by people living in flats, with a higher proportion living in socially rented accommodation than for the supergroup. Ethnic groups generally have a similar representation as for the supergroup, persons of mixed ethnicity are underrepresented. There is a lower proportion of households with two or more cars.”

This could be for several resaons. As they are most likely to be living in flats they have overall less space to heat and lights, there are likely to be fewer people living in them to consume energy but also they are considered constrained, and fewer households have two or more cars – an indication that they are poor – and are therefore likely to be frugal and less wasteful with their energy for economic reasons.

Group 3C: Ethnic Dynamics use only slightly more energy than the lowest group.

“In this group non-White ethnic groups are not represented as highly as in the parent supergroup and there is a higher proportion of people born in the UK or Ireland. Households are more likely to live in a flat and to socially rent. There is a higher proportion of unemployed in the group but those in employment are more likely to work in the manufacturing industry, and to use private transport to travel to work.”

This low usage can again be attributed to living in smaller accommodation, but it is possible that they use slightly more as some are unemployed and therefore more likely to use energy consistently throughout the day.

## Accommodation variables

Table 3 summaries the accommodation variable used in conjunction with the Townsend variables for the final deprivation model.

# INSERT VARIABLES TABLE

## Models

### Townsend scores, housing types and energy usage.

To gain a deeper understanding of the nations characteristics, the full dataset was aggregated to give an average usage over various temporal granularities; half hourly readings over a day, average usage per day of the week and seasonal usage displayed as average monthly overall usage. To achieve this, the energy usage profiles were combined with the reweighted Townsend scores and accommodation types, and a linear regression was calculated. Only England and Wales were investigated because of the very sparse nature of Scotland’s data.

The total daily usage was calculated in each case. The largest group of accommodation type in the dataset was semi-detached, and so this was used as the reference category, allowing for accommodation type to be controlled for. Because the groups for “flats” and “shared accommodation” were so small they were combined into one category before modelling the daily usage total against them and the Townsend score.

### Deprivation model

The deprivation model explained R2 = 15.54% (adjusted R2 = 15.49%) of the variability in domestic energy consumption, p < 2.2e-16.

A detached and typically larger house than the reference category (semi-detached) would use on average an extra 1953.2 kW per day in comparison, whereas a smaller terrace house would use less to the sum of 135.5 kW per day. Detached houses typically have a greater number of rooms and lower levels of overcrowding, therefor more rooms need heating and lighting (which account for X amount of average consumption). This model also suggests that as Townsend scores increase, energy usage decreases, which as a measure of deprivation is to be expected – the higher the level of deprivation the less likely people are to be able to afford to be wasteful and consume unnecessary energy. They are also likely to live in smaller houses.

## Conclusions and future work

In conclusion, this research has revealed that the innovative dataset has elements which replicate the existing data in terms of the most relevant socio-demographic indicators when it comes to understanding energy usage, especially in terms of recognising different levels of deprivation. We have proved that housing type is significant in terms of understanding the amounts of energy use, as are deprivation levels, indicated by the Townsend score introduced in the regression model and the investigation of energy usage by OAC groups.

## Bibliography

Abrahamse, Wokje, and Linda Steg. 2011. “Factors Related to Household Energy Use and Intention to Reduce It: The Role of Psychological and Socio-Demographic Variables.” Society for Human Ecology. doi:[10.2307/24707684](https://doi.org/10.2307/24707684).

Alahakoon, Damminda, and Xinghuo Yu. 2013. “Advanced analytics for harnessing the power of smart meter big data.” In *Intelligent Energy Systems (Iwies), 2013 Ieee International Workshop on*, 40–45. IEEE.

Albert, Adrian, and Ram Rajagopal. 2013. “Smart meter driven segmentation: What your consumption says about you.” *IEEE Transactions on Power Systems* 28 (4). IEEE: 4019–30.

Ardakanian, Omid, Negar Koochakzadeh, Rayman Preet Singh, Lukasz Golab, and Srinivasan Keshav. 2014. “Computing Electricity Consumption Profiles from Household Smart Meter Data.” In *EDBT/Icdt Workshops*, 14:140–47.

Bedir, Merve, Evert Hasselaar, and Laure Itard. 2013. “Determinants of electricity consumption in Dutch dwellings.” *Energy and Buildings* 58: 194–207. doi:[10.1016/j.enbuild.2012.10.016](https://doi.org/10.1016/j.enbuild.2012.10.016).

Brandon, Gwendolyn, and Alan Lewis. 1999. “Reducing household energy consumption: A qualitative and quantitative field study.” *Journal of Environmental Psychology* 19 (1): 75–85. doi:[10.1006/jevp.1998.0105](https://doi.org/10.1006/jevp.1998.0105).

“Britain’s Smart Grid.” 2018. <https://www.smartenergygb.org/en/smart-future/britains-smart-grid>.

Brounen, Dirk, Nils Kok, and John M. Quigley. 2012. “Residential energy use and conservation: Economics and demographics.” doi:[10.1016/j.euroecorev.2012.02.007](https://doi.org/10.1016/j.euroecorev.2012.02.007).

Buchmann, Erik, Klemens Böhm, Thorben Burghardt, and Stephan Kessler. 2013. “Re-identification of smart meter data.” *Personal and Ubiquitous Computing* 17 (4). Springer: 653–62.

Darby, Sarah. 2010. “Smart metering: what potential for householder engagement?” *Building Research & Information* 38 (5). Taylor & Francis: 442–57.

Druckman, A., and T. Jackson. 2008. “Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model.” *Energy Policy* 36 (8): 3167–82. doi:[10.1016/j.enpol.2008.03.021](https://doi.org/10.1016/j.enpol.2008.03.021).

Faiers, Adam, Matt Cook, and Charles Neame. 2007. “Towards a contemporary approach for understanding consumer behaviour in the context of domestic energy use.” *Energy Policy* 35 (8): 4381–90. doi:[10.1016/j.enpol.2007.01.003](https://doi.org/10.1016/j.enpol.2007.01.003).

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. Vol. 58. doi:[10.1111/insr.12011\_21](https://doi.org/10.1111/insr.12011_21).

Figueiredo, V, F Rodrigues, Z Vale, and J B Gouveia. 2005. “An electric energy consumer characterization framework based on data mining techniques.” 2. Vol. 20. doi:[10.1109/TPWRS.2005.846234](https://doi.org/10.1109/TPWRS.2005.846234).

Flath, Christoph, David Nicolay, Tobias Conte, Clemens van Dinther, and Lilia Filipova-Neumann. 2012a. “Cluster analysis of smart metering data.” *Business & Information Systems Engineering* 4 (1). Springer: 31–39.

———. 2012b. “Clusteranalyse von Smart-Meter-Daten.” *Wirtschaftsinformatik* 54 (1). Gabler Verlag: 33–42.

Frederiks, Elisha R., Karen Stenner, and Elizabeth V. Hobman. 2015. “The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review.” doi:[10.3390/en8010573](https://doi.org/10.3390/en8010573).

Frey, Cynthia J., and Duncan G. LaBay. 1983. “A Comparative Study of Energy Consumption and Conservation Across Family Life Cycle.” <http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=6198>.

Gas, British. 2018. “The History of British Gas.” <https://www.britishgas.co.uk/business/history>.

Gatersleben, Birgitta, Linda Steg, and Charles Vlek. 2002. “Measurement and Determinants of Environmentally Significant Consumer Behavior.” *Environment and Behavior* 34 (3). Sage PublicationsSage CA: Thousand Oaks, CA: 335–62. doi:[10.1177/0013916502034003004](https://doi.org/10.1177/0013916502034003004).

Guerreiro, Susana, Susana Batel, Maria Lu’isa Lima, and Sérgio Moreira. 2015. “Making energy visible: sociopsychological aspects associated with the use of smart meters.” *Energy Efficiency* 8 (6). Springer Science & Business Media: 1149.

Haben, Stephen, Colin Singleton, and Peter Grindrod. 2016. “Analysis and clustering of residential customers energy behavioral demand using smart meter data.” *IEEE Transactions on Smart Grid* 7 (1). IEEE: 136–44.

Hargreaves, Tom, Michael Nye, and Jacquelin Burgess. 2010. “Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors.” *Energy Policy* 38 (10). Elsevier: 6111–9.

———. 2013. “Keeping energy visible? Exploring how householders interact with feedback from smart energy monitors in the longer term.” *Energy Policy* 52. Elsevier: 126–34.

HM Government. 2009. *The UK Low Carbon Transition Plan: National strategy for climate and energy*. July 2009. doi:[10.1007/BF02393883](https://doi.org/10.1007/BF02393883).

Huebner, Gesche M., Ian Hamilton, Zaid Chalabi, David Shipworth, and Tadj Oreszczyn. 2015. “Explaining domestic energy consumption - The comparative contribution of building factors, socio-demographics, behaviours and attitudes.” *Applied Energy* 159: 589–600. doi:[10.1016/j.apenergy.2015.09.028](https://doi.org/10.1016/j.apenergy.2015.09.028).

Huebner, Gesche, David Shipworth, Ian Hamilton, Zaid Chalabi, and Tadj Oreszczyn. 2016. “Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes.” doi:[10.1016/j.apenergy.2016.04.075](https://doi.org/10.1016/j.apenergy.2016.04.075).

Jiawei, H, Micheline Kamber, Jiawei Han, Micheline Kamber, and Jian Pei. 2012. *Data Mining: Concepts and Techniques*. doi:[10.1016/B978-0-12-381479-1.00001-0](https://doi.org/10.1016/B978-0-12-381479-1.00001-0).

Jones, Rory V., and Kevin J. Lomas. 2015. “Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics.” *Energy and Buildings* 101: 24–34. doi:[10.1016/j.enbuild.2015.04.052](https://doi.org/10.1016/j.enbuild.2015.04.052).

Kavousian, Amir, Ram Rajagopal, and Martin Fischer. 2013. “Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants’ behavior.” *Energy* 55: 184–94. doi:[10.1016/j.energy.2013.03.086](https://doi.org/10.1016/j.energy.2013.03.086).

Kwac, Jungsuk, Chin-Woo Tan, Nicole Sintov, June Flora, and Ram Rajagopal. 2013. “Utility customer segmentation based on smart meter data: Empirical study.” In *Smart Grid Communications (Smartgridcomm), 2013 Ieee International Conference on*, 720–25. IEEE.

Lavin, Alexander, and Diego Klabjan. 2016. “Clustering time-series energy data from smart meters.” *arXiv Preprint arXiv:1603.07602*.

Leahy, Eimear, and S. Lyons Sean. 2010. “Energy use and appliance ownership in Ireland.” *Energy Policy* 38 (8): 4265–79. doi:[10.1016/j.enpol.2010.03.056](https://doi.org/10.1016/j.enpol.2010.03.056).

López, José J, José A Aguado, F Mart’in, F Munoz, A Rodr’iguez, and José E Ruiz. 2011. “Hopfield–K-Means clustering algorithm: A proposal for the segmentation of electricity customers.” *Electric Power Systems Research* 81 (2). Elsevier: 716–24.

McKenna, Eoghan, Ian Richardson, and Murray Thomson. 2012. “Smart meter data: Balancing consumer privacy concerns with legitimate applications.” *Energy Policy* 41. Elsevier: 807–14.

McLoughlin, Fintan, Aidan Duffy, and Michael Conlon. 2015. “A clustering approach to domestic electricity load profile characterisation using smart metering data.” *Applied Energy* 141. Elsevier: 190–99.

Morris, Russell, and Vera Carstairs. 1991. “Which deprivation? A comparison of selected deprivation indexes.” *Journal of Public Health Medicine* 13 (4): 318–26. [https://watermark.silverchair.com/13-4-318.pdf?token=AQECAHi208BE49Ooan9kkhW{\\_}Ercy7Dm3ZL{\\_}9Cf3qfKAc485ysgAAAcEwggG9BgkqhkiG9w0BBwagggGuMIIBqgIBADCCAaMGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMq6paVTi7sAhCrUvKAgEQgIIBdBA84wEmDW-oSLjoUIhdUGgG0HTVZqUe0yVbUhyl9C2{\\_}fC](https://watermark.silverchair.com/13-4-318.pdf?token=AQECAHi208BE49Ooan9kkhW%7b\_%7dErcy7Dm3ZL%7b\_%7d9Cf3qfKAc485ysgAAAcEwggG9BgkqhkiG9w0BBwagggGuMIIBqgIBADCCAaMGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMq6paVTi7sAhCrUvKAgEQgIIBdBA84wEmDW-oSLjoUIhdUGgG0HTVZqUe0yVbUhyl9C2%7b\_%7dfC).

O’Neill, Brian C, and Belinda S Chen. 2002. “Demographic Determinants of Household Energy Use in the United States.” *Population and Development Review* 28: 53–88. doi:[10.2307/3115268](https://doi.org/10.2307/3115268).

Parra, Juan E, Franklin L Quilumba, and Hugo N Arcos. 2016. “Customers’ demand clustering analysisA case study using smart meter data.” In *Transmission & Distribution Conference and Exposition-Latin America (Pes T&D-La), 2016 Ieee Pes*, 1–7. IEEE.

Pérez-Lombard, Luis, José Ortiz, and Christine Pout. 2008. “A review on buildings energy consumption information.” *Energy and Buildings* 40 (3): 394–98. doi:[10.1016/j.enbuild.2007.03.007](https://doi.org/10.1016/j.enbuild.2007.03.007).

Poortinga, Wouter, Linda Steg, and Charles Vlek. 2004. “Values, Environmental Concern, and Environmental Behavior.” *Environment and Behavior* 36 (1): 70–93. doi:[10.1177/0013916503251466](https://doi.org/10.1177/0013916503251466).

Raaij, W. Fred van, and Theo M M Verhallen. 1983. “Patterns of residential energy behavior.” *Journal of Economic Psychology* 4 (1-2): 85–106. doi:[10.1016/0167-4870(83)90047-8](https://doi.org/10.1016/0167-4870(83)90047-8).

Ramos, Sérgio, and Zita Vale. 2008. “Data mining techniques to support the classification of MV electricity customers.” In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 Ieee*, 1–7. IEEE.

Rhodes, Joshua D., Wesley J. Cole, Charles R. Upshaw, Thomas F. Edgar, and Michael E. Webber. 2014. “Clustering analysis of residential electricity demand profiles.” *Applied Energy* 135: 461–71. doi:[10.1016/j.apenergy.2014.08.111](https://doi.org/10.1016/j.apenergy.2014.08.111).

Sardianou, Eleni. 2007. “Estimating energy conservation patterns of Greek households.” *Energy Policy* 35 (7): 3778–91. doi:[10.1016/j.enpol.2007.01.020](https://doi.org/10.1016/j.enpol.2007.01.020).

Schipper, Lee, Sarita Bartlett, Dianne Hawk, and Edward Vine. 1989. “Linking Life-Styles and Energy Use: A Matter of Time.” *Annual Review of Energy and the Environment* 14 (1): 273–320. doi:[10.1146/annurev.energy.14.1.273](https://doi.org/10.1146/annurev.energy.14.1.273).

Statistics, Office for National. 2018. “Postal Geography.” Accessed April 28. <https://www.ons.gov.uk/methodology/geography/ukgeographies/postalgeography>.

Stern, Stephanie M. Chicago-Kent. 2011. “Smart-Grid: Technology and the Psychology of Environmental Behavior Change.” *Law Review* 86 (1): 139. doi:[10.1525/sp.2007.54.1.23.](https://doi.org/10.1525/sp.2007.54.1.23.)

Townsend, Peter. 1987. “Deprivation.” *Journal of Social Policy* 16 (02): 125. doi:[10.1017/S0047279400020341](https://doi.org/10.1017/S0047279400020341).

Tso, Geoffrey K.F., and Kelvin K.W. Yau. 2007. “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks.” *Energy* 32 (9): 1761–8. doi:[10.1016/j.energy.2006.11.010](https://doi.org/10.1016/j.energy.2006.11.010).

UK CCC. 2010. “Fourth Carbon Budget. London (UK): UK Committee on Climate Change; 2010.”

Viegas, Joaquim L, Susana M Vieira, Rui Mel’icio, V M F Mendes, and João M C Sousa. 2016. “Classification of new electricity customers based on surveys and smart metering data.” *Energy* 107. Elsevier: 804–17.

Wyatt, Peter. 2013. “A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England.” *Energy Policy* 60: 540–49. doi:[10.1016/j.enpol.2013.05.037](https://doi.org/10.1016/j.enpol.2013.05.037).

Yohanis, Yigzaw G., Jayanta D. Mondol, Alan Wright, and Brian Norton. 2008. “Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use.” *Energy and Buildings* 40 (6): 1053–9. doi:[10.1016/j.enbuild.2007.09.001](https://doi.org/10.1016/j.enbuild.2007.09.001).