Electricity consumption and household characteristics: Implications for census-taking in a smart metered future

Ben Anderson (b.anderson@soton.ac.uk) **

Sharon Lin (X.Lin@soton.ac.uk) *

Andy Newing (A.Newing@leeds.ac.uk) ^

AbuBakr Bahaj (A.S.Bahaj@soton.ac.uk) *

Patrick James (P.A.James@soton.ac.uk) *

** Corresponding Author

* Sustainable Energy Research Group, Energy & Climate Change Division, University of Southampton, Southampton, UK

^ School of Geography, University of Leeds, Leeds, United Kingdom

Abstract

This paper assesses the feasibility of determining key household characteristics based on temporal load profiles of household electricity demand. It is known that household characteristics, behaviours and routines drive a number of features of household electricity loads in ways which are currently not fully understood. The roll out of domestic smart meters in the UK and elsewhere could enable better understanding through the collection of high temporal resolution electricity monitoring data at the household level. Such data affords tremendous potential to invert the established relationship between household characteristics and temporal load profiles. Rather than use household characteristics as a predictor of loads, observed electricity load profiles, or indicators based on them, could instead be used to impute household characteristics. These micro level imputed characteristics could then be aggregated at the small area level to produce 'census-like' small area indicators. This work briefly reviews the nature of current and future census taking in the UK before outlining the household characteristics that are to be found in the UK census and which are also known to influence electricity load profiles. It then presents descriptive analysis of two smart meter-like datasets of half-hourly domestic electricity consumption before reporting on the results from a multilevel modelling-based analysis of the same data. The work concludes that a number of household characteristics of the kind to be found in UK census-derived small area statistics may be predicted from particular load profile indicators. A discussion of the steps required to test and validate this approach and the wider implications for census taking is also provided.

Keywords: census, smart meter, transactional data, big data, households

Paper accepted for publication as:

Anderson, B., Lin, S., Newing, A., Bahaj, A., and James, P., (2016) Electricity consumption and household characteristics: Implications for census-taking in a smart metered future, *Computers, Environment and Urban Systems*, DOI: 10.1016/j.compenvurbsys.2016.06.003

Author Highlights

- Temporal electricity consumption patterns (profiles) are known predictors of some household characteristics;
- Such patterns could be used to estimate census characteristics at household level for aggregation to 'normal' census output areas;
- Results suggest standard profile indicators can predict the number of people, the presence of children and also employment status of the household response person.

1 Energy monitoring for a 'Smart Census'

Area based population statistics in the United Kingdom (UK) have historically been derived from the decadal census of housing and population. In addition to basic demographic statistics, the socio-economic information collected is used to produce robust small area estimates of a range of characteristics for every neighbourhood. Representing 'a definitive snapshot of the nation' (Calder and Teague 2013) this data provides a backbone for commercial, academic and social research as well as policy analysis, a decadal 're-grouping' and 're-basing' of all small area population projections statistics (Norman 2013) and, crucially, national and local resource allocation (Eurostat 2011; Norman 2013). Nonetheless, the UK census has also faced criticism as a costly and frequently outdated source of population statistics, with a time lag of at least two years between data collection and reporting (Keith Dugmore et al. 2011b)

Currently considered approaches for the future provision of population statistics include decennial census-taking, more frequent social surveys or administrative (Government held) data linkage and aggregation (ONS 2013). In contrast, this work explores the possibility of deriving small area estimates of traditional socio-economic indicators from 'digital trace' or transactional data collected by utility (or other) services as part of normal service provision. As a number of recent authors have noted large-scale geo-coded transactional datasets, such as those collected in the retail, telecommunications, finance and utilities sectors could offer opportunities to supplement census based small area statistics by supporting the delivery of area-based population statistics, and generating novel indicators at a neighbourhood level (Keith Dugmore et al. 2011b; Struijs, Braaksma, and Daas 2014; Deville et al. 2014). For the United Kingdom Statistics Authority, via its executive office the Office for National Statistics (ONS) in England and Wales, the use of commercial data to support census taking may therefore help address census users' requests for more frequent and timely reporting of census-type statistics in the intercensal periods.

Recent related work suggests that commercial 'big data' could both support near real time census taking and also provide unique insights into household or individual behaviours (Carroll, Lyons, and Denny 2014; Claxton, Reades, and Anderson 2012; Keith Dugmore et al. 2011b; Pucci, Manfredini, and Tagliolato 2015; Deville et al. 2014; Douglass et al. 2015). In this work we consider household level data held by a range of utility companies before focusing in particular on smart meter derived electricity consumption data. Compared to a number of other forms of potentially useful 'big data', a grid-connected electricity supply is almost universally available in the UK, almost universally connected to domestic dwellings and metering of consumption is mandatory. Furthermore the planned universal rollout of electricity smart meters collecting at least half-hourly consumption data (DECC 2013) means that consideration of the value of suitably anonymised and aggregated smart meter data in the production of official statistics is now timely.

The use of this kind of data for market segmentation and other electricity related services has been noted in the literature (McKenna, Richardson, and Thomson 2012) and was noted by Dugmore et al. (Keith Dugmore et al. 2011b) in the context of future census data collection. However, as far as we are aware only one published study has investigated its potential in the development of official and/or small area statistics (Carroll, Lyons, and Denny 2014). A growing literature suggests that household level electricity load data, collected via smart metering, could provide considerable opportunities to infer household characteristics (Beckel, Sadamori, and Santini 2012; Newing et al. 2015; Struijs, Braaksma, and Daas 2014) The link between household characteristics and household energy consumption is long established and the literature recognises that household characteristics will give rise to different load profiles and subsequent demand on the electricity supply network (e.g. see (F McLoughlin, Duffy, and Conlon 2013) for a summary). Consequently, the energy sector uses household or area based indicators of household composition and characteristics to predict electricity 'demand' in order to manage networks and target interventions designed to reduce or time-shift peak loads (e.g. see (Elexon 2013; Hamidi, Li, and Robinson 2009; Wright and Firth 2007))

The purpose of this work is to explore the value of inverting this approach to assess the feasibility of using observed high temporal resolution electricity consumption data to infer household characteristics as a first step in the aggregation of household characteristics to form 'normal' area level population statistics. It should be emphasised therefore that the overall objective is not to characterise or 'profile' individual households, rather we seek to aggregate inferred household characteristics to develop area based 'neighbourhood' indicators similar to or in combination with Census estimates or other appropriate datasets.

This work briefly reviews the future provision of area based statistics in the UK, recognising the opportunities to enhance or supplement the census taking process with digital trace data. It then considers the extent to which digital trace data from the commercial sector could represent a novel tool to generate census type small-area statistics, before focusing on the use high resolution electricity consumption monitoring data collected via smart metering. Based on preliminary analyses of a 'smart meter-like' dataset the research highlights the potential value of the approach and then discusses significant challenges and concludes by setting out a research programme which could systematically test the value of the approach.

2 Future provision of area based population statistics in the UK

As a consistent and robust source of small area population statistics, the United Kingdom census is used to allocate billions of pounds of government and commercial investment at the local level. It represents a fundamental tool for market research, policy making, commercial decision making, resource allocation and for academic research (ONS 2013; Watson 2009). Estimates of population counts by age and sex are a key census output, yet the detailed attribute information related to households and their usual residents - determining characteristics such as ethnic composition, education, socio-economic status, religion and employment – offer greatest value to the academic and commercial sector.

Census data are not made available at the individual household level but are published as non-disclosive aggregated counts within a hierarchy of 'output zones' or areas. These are built from unit postcodes, designed for the release of aggregate population statistics and represent small areas ranging from Output Areas (OAs – typically containing around 125 households) through to local authority districts (LADs) or Unitary Authorities (UAs). The former represents an important analytical unit for resource allocation and policy making at the local level, especially within the commercial sector (K. Dugmore 2013; ONS 2014a). It is this combination of universal geographic coverage at the small area level coupled with detailed attribute data that represents a major strength of the census (House of Commons Treasury Committee 2008).

However, inevitably increasing costs, difficulties of ensuring full response, concerns over the decadal reporting cycle and the two year time-lag between census-taking and the delivery of initial outputs has given rise to a search for alternatives (Keith Dugmore et al. 2011b). This work has been conducted by the ONS 'Beyond 2011' programme (ONS 2014a) and, together with subsequent reviews of international census taking practice (see for example Dugmore et al (2011a); and Martin (2006)), has highlighted a variety of approaches to collecting area based statistics including the use of governmental administrative sources (e.g. Netherlands and Denmark) or a rolling census (France). However the work also showed that a number of options under consideration by 'Beyond 2011', particularly those driven by administrative data, were unable to provide the level of socio-economic attribute data that many census users rely upon for commercial analysis, policy making and resource allocation (Calder and Teague 2013; ONS 2014a). Additionally, concerns have been raised over the likely success and practicalities of a census based on an administrative or register based system given the lack of a population register within the UK (Skinner, Hollis, and Murphy 2013).

Based on the recommendations of the Beyond 2011 program (ONS 2014a) on, extensive user consultation (ONS 2014b) and an independent review (Skinner, Hollis, and Murphy 2013), the UK Statistics Authority recommended to parliament that a 'traditional' decadal census should be carried out in 2021 (Dilnot 2014). They also noted that this should be primarily carried out online and that the considerable potential of utilising administrative data and larger scale household surveys as a supplement to census based statistics should be developed further (Dilnot 2014).

Whilst recognising that data held by commercial organisations may offer more cost effective or timely reporting (Keith Dugmore et al. 2011b), this avenue has received far less attention and discussion has tended to refer only to 'customer information' recorded in customer service databases and/or retail transaction data. As far as we are aware, commercial data does not currently feature within the national statistical census taking or population statistics of any nation and has not yet been explicitly considered by the ONS' Beyond 2011 programme. As Struijs, Braaksma, and Daas (2014) note such data could be used to provide substantial additional data over and above basic address listings.

3 Smart Meters for a Smart Census

The nascent roll-out of domestic electricity 'smart meters' in a number of major markets including the US, China, Brazil, India and Japan (Deloitte 2011) and the UK (DECC 2012) provides an opportunity for the exploration of precisely the scenario described above.

In the UK, smart meters incorporate communication infrastructure allowing them to transmit near real-time energy usage data to in home display units (IHDs), to energy demand service operators selected by the customer and to a centralised data retrieval service to extract half-hourly data from all smart meters for use by energy suppliers (billing and fraud prevention); network operators (network management) or other authorised third parties (e.g. switching agencies). Unlike other household level transactional data sources, the universal coverage and singular data access gateway suggests that electricity monitoring data could represent a valuable data source for research and policy making.

A growing literature suggests that household level electricity load data, collected via smart metering, could provide considerable opportunities to infer household characteristics (Beckel, Sadamori, and Santini 2012; Newing et al. 2015; Struijs, Braaksma, and Daas 2014). The link between household characteristics and household energy consumption is long established and the literature recognises that household characteristics will give rise to different load profiles and subsequent demand on the electricity supply network (Wright and Firth 2007; Hamidi, Li, and Robinson 2009; Elexon 2013; F McLoughlin, Duffy, and Conlon 2013). In this work we assess the feasibility of inverting this approach and use observed high temporal resolution electricity loads in order to infer household characteristics. It should be noted, however, that the overall intention is not to characterise or 'profile' individual households, but to develop area based 'neighbourhood' indicators, which are themselves important for policy making and the delivery of population statistics and social indicators.

We have previously used the term 'Smart Censuses' (Newing et al. 2015) in reference to the potential generation of area based population statistics and area based indicators inferred via smart metering. First, however, it is necessary to demonstrate that smart metered electricity load data can be used as a tool to accurately infer household characteristics. In this work we make use of a smart-meter like dataset, introduced in the following section. We use this dataset to generate a series of summary statistics and indices ('profile indicators') to describe the shape and characteristics of household load profiles. We then use multi-level regression modelling techniques to test the extent to which such load profile indicators have the potential to predict key household characteristics.

4 A Smart meter-like dataset

Preliminary work conducted with a small smart-meter like dataset from a University of Southampton energy demand reduction project showed some evidence of differences in load profiles for different types of households (Newing et al. 2015). However the small sample size and relative homogeneity of the household sample led us to follow (F McLoughlin, Duffy, and Conlon 2012) and exploit a much larger smart-meter like dataset from the Irish Commission for Energy Regulation's (CER) Smart Metering Electricity Customer Behaviour Trials (CBTs)[1]. The purpose of the trials, which took place from 2009 to 2010 was to assess the impact of various tariff regimes and feedback methods on consumers' electricity consumption. In order to do this over 3,000 Irish households were recruited and equipped with electricity consumption monitors before being allocated to control and intervention groups. All households were surveyed during the baseline (2009) and again during the post-trial (2010) stages to gather information on household composition, appliance ownership and usage and socio-economic status. Overall, as Table 1 shows, some 4232 households completed the initial survey but only 3487 had full functioning half-hourly consumption monitors in October 2009 at the start of the baseline period. In total 3143 completed both surveys and had full consumption records for the entire trial period.

Table 1: Irish CER Smart Meter Trial Household samples

Sample	N
Households who completed the 2009 pre-trial survey	4232
Households with valid consumption data in October 2009	3488
Households who completed the 2010 survey	3422
Households who completed both surveys and had valid consumption records	3144

In order to reduce the processing and analysis time, avoid major holidays and exclude potentially confounding seasonal variation, four weeks of data from mid-week days (Tues, Wednesday and Thursday) in late September and early October 2009 was selected for analysis. The selection of

[1] Accessed via the Irish Social Science Data Archive - http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

mid-week days was guided by previous work suggesting that there is greater differentiation in load profiles for different kinds of households during the week compared to weekends (Newing et al. 2015) .

Table 2 shows summary statistics for the mid-week household consumption for this period and indicates the extent to which consumption is positively skewed. Table 3 extends this analysis to show that variation in half hourly consumption increases as the number of residents increases, so does the mean total consumption over the period and also the mean consumption per half hour.

Table 2: Descriptive statistics for mid-week electricity consumption in kWh for the household sample (mid-week days) – October 2009

Variable	N	Mean	SD	Median	min	max	skew	kurtosis
All half-hours	2,009,088	0.48	0.64	0.24	0	10.44	3.15	14.1
02:00 – 05:00 (Baseload half-hours)	334,848	0.19	0.21	0.14	0	6	6.19	70.9
16:00 – 20:00 (Evening Peak half-hours)	418,560	0.71	0.82	0.4	0	10.44	2.34	7.51
Daily sum per household	55,808	17.35	14.87	15.74	0	158.13	0.94	1.53

Table 3: Descriptive statistics for half hourly mid-week electricity consumption in kWh for the household sample (mid-week days) by number of residents – October 2009

	N households	N half-hours	Mean total consumption per household	Mean (half-hours)	SD	Median	skew	kurtosis
1 person	699	402,624	161.50	0.28	0.44	0.14	4.41	29.02
2 people	134	77,184	247.81	0.43	0.60	0.21	3.45	17.67
3 people	1,702	980,928	285.52	0.50	0.64	0.26	3.01	12.72
4 people	511	294,336	325.22	0.56	0.69	0.32	2.94	12.34
5+ people	441	254,016	384.35	0.67	0.79	0.37	2.68	10.15

Some 2% of households used electricity storage heaters as their main means of heating while 28% used gas, 42% used oil and 26% used solid fuels. It was therefore considered unlikely that the use of electricity for heating would substantially affect the analysis. On the other hand, at least 19% of the households used electric immersion heaters as their source of hot water. We acknowledge that the use of such appliances could introduce artefacts into the kinds of consumption patterns we discuss below. However in the future census collection context we propose, the presence of hot water immersion heaters would be unknown and so must be accepted as a potential source of error in the estimation of household attributes. More detailed analysis of individual dwelling level consumption patterns could attempt to identify and control for electricity consumption of this kind but due to its computational complexity (Zoha et al. 2012) this was considered outside the scope of the current paper.

Other sources of variation in electricity consumption include external temperature, the ownership of more (or less) energy efficient appliances and on-site electricity generation through solar panels or domestic wind turbines. By using a single month in the Autumn of 2009 in Ireland we attempt to control for temperature fluctuations between households and therefore assume that all dwellings were exposed to the same climatic conditions. We do not take into account ownership of different kinds of appliances as this could not be known at the household level in the context we propose and so would contribute to 'error' in our estimations. Finally only 1% of households reported using "Renewable (e.g. solar)" for heating and 2% for hot water and we take this as an indicator that the depression of measured power import from the grid due to within-dwelling generation is unlikely to cause problems in the analysis.

4.1 Electricity Load Profiles

The half-hourly resolution electricity consumption data collected during the four week period in October 2011 corresponds to the default temporal reporting interval specified by the UK smart meter roll-out programme (Energy UK 2013) and is generally considered to be adequate for load profile analysis (Beckel, Sadamori, and Santini 2013).

As an example, Figure 1 shows temporal load profiles for different kinds of households using this data and exhibits a familiar shape, with pronounced morning (for those in work) and evening peaks potentially driven by active household occupancy and use of household appliances at these times of the day. The literature suggests that a households' consumption profile tends to be fairly consistent on a day-to-day basis (Ning and Kirschen 2010), no doubt driven by similar routines, behaviours and occupancy patterns.

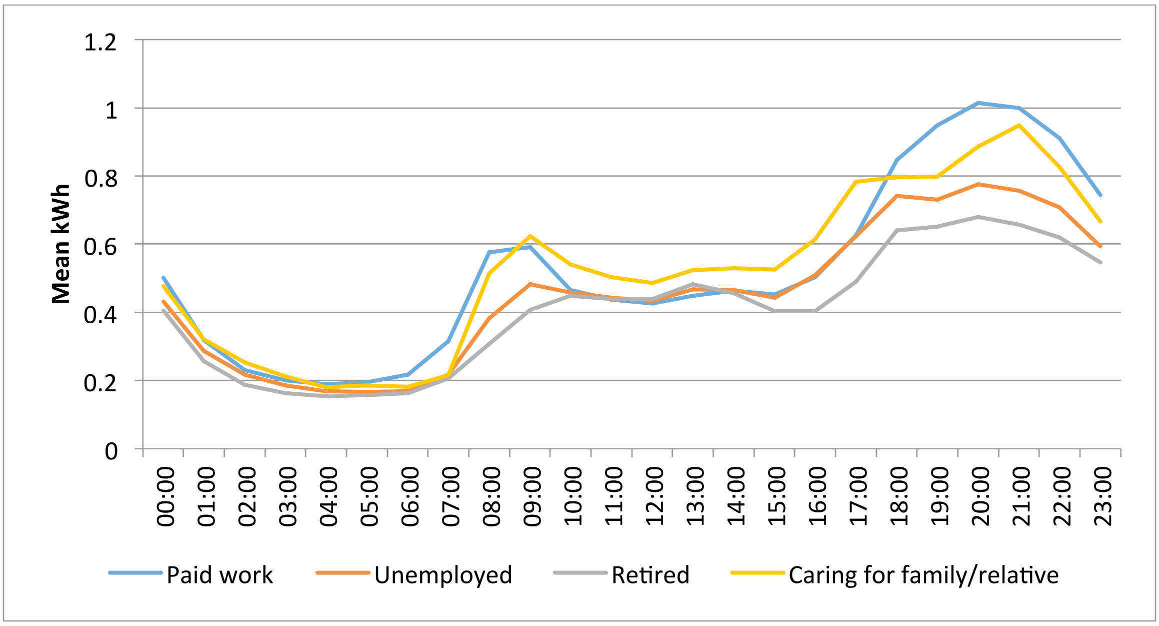

Figure 1: Mean half-hourly electricity consumption per half hour (Tuesday – Thursday) by self-reported employment status of household response person. Source: Authors' calculation using Irish CER Smart Meter Trial data October 2009 (n = 3,488).

A number of studies have confirmed the impact of household composition, dwelling characteristics and householder behaviours on such load profiles (Beckel, Sadamori, and Santini 2012; Druckman and Jackson 2008; Firth et al. 2008; Owen 2012; Wright 2008). Temporal periods where the greatest inter-household variability in load profiles may be evident, such as the evening peak period, could offer greatest value in identifying household characteristics, on the assumption that it is differences in those characteristics which are likely to drive differences in household behaviour and routines, and thus loads at these times of the day. Recent findings (Newing et al. 2015) suggest that the household load profiles for our study households exhibit a number of key features which may assist in differentiating between households based on their characteristics.

4.2 Linked household survey data

The CER electricity consumption dataset is linked to survey data that incorporates a number of household characteristics of potential interest in the production of small area population statistics. These include householder and dwelling characteristics directly comparable with existing area based population statistics (such as dwelling type, number of residents and employment status) alongside indicators not currently part of small area data collection but of considerable relevance to policy makers (such as income), as Table 4 makes clear.

Table 4: Selected household characteristics collected or potentially collected by the Census together with evidence of their relationship to load profiles.

	Census 2011 household level variables*	**Existing evidence for links to load profiles**
Household	Number of persons	(Beckel, Sadamori, and Santini 2013)
	Presence of person with limiting long term illness	
	Number of children	(Yohanis et al. 2008)
	Age distributions of all persons	
Dwelling	Household dwelling type	(Firth et al. 2008; F McLoughlin, Duffy, and Conlon 2012)
	Household tenure	(Druckman and Jackson 2008)
	Number of (bed)rooms	dwelling floor area as a proxy
	Number of cars/vans	
	Presence of and fuel used for heating	(F McLoughlin, Duffy, and Conlon 2013)
Householder	Ethnic group/country of birth of HRP/main language	
	Age of HRP	(F McLoughlin, Duffy, and Conlon 2013)
	NS-SEC of household reference person (HRP)	(Druckman and Jackson 2008; M. Hughes and Moreno 2013; F McLoughlin, Duffy, and Conlon 2013)
	Economic activity of HRP/hours worked	(Yohanis et al. 2008; F McLoughlin, Duffy, and Conlon 2013)
	HRP Education level	
	Marital Status	
	Other	
	Dwelling floor area	(Beckel, Sadamori, and Santini 2013; Craig et al. 2014; F McLoughlin, Duffy, and Conlon 2013)
	Household Income	(Beckel, Sadamori, and Santini 2013; Craig et al. 2014; F McLoughlin, Duffy, and Conlon 2013)
	Daily consumption profile shape	(Haben et al. 2014)

For policy application, dwelling size (and in particular the number of bedrooms) is an important indicator of household level overcrowding, used primarily by local authorities to tackle housing issues. Following the 2011 census, this information has been reported as an 'occupancy rating' (relating the number of bedrooms to the number of usually resident occupants) and notions of household overcrowding and under-occupancy have also become policy relevant in the wake of the welfare reforms in the UK, whereby available benefits are cut if claimants have a spare bedroom within a council or housing association provided home (Ramsden 2014). The existing literature provides evidence that dwelling floor area is linked to electricity consumption (Table 4) and we assess the extent to which, based on our sample, load profiles can be used to infer household floor area. This could in turn be used to estimate the number of bedrooms in order to generate a more policy-relevant indicator.

Household income does not form part of small area population statistics, yet represents an indicator of considerable value to policy makers and the commercial sector, with its link to electricity loads well established (Table 4). In spite of frequent calls for its inclusion, plans to collect this information within the 2011 UK census were dropped amidst concerns of under-response driven by the perceived intrusion posed by an income question. Income is an important indicator of economic well-being and the lack of information on income available via population statistics is frequently cited as a weakness (See e.g. Keith Dugmore et al. 2011b). The CER survey recorded household response person's self-reported net annual income via income bands.

Employment status is reported via population statistics in relation to economic activity, and forms an important tool at the local, regional and national level for policy making and intervention, enabling household classification and acting as a predictor for behaviours and routines. The literature clearly identifies that employment status impacts upon the timing of electricity loads (Table 4). Household response person (HRP) employment status formed part of the CER survey with over 59% of study HRPs in employment (incorporating full time, part time, freelance and self-employment), almost 30% retired, with the remaining 11% representing HRPs not in active employment through unemployment, study or full time care duties. The latter categories have been combined with retired households for subsequent analysis, giving two groups; '*Employed'* and '*Not in active employment'*. Householder employment status should, however, be treated with some caution. Self-reported employment status must be treated as an indicator only as response categories provided by the survey did not account for the full range of nuanced employment patterns that may exist, such as homeworking and flexible working arrangements which would impact considerably on behaviours, routines and domestic electricity loads.

It would, however, be an oversimplification to suggest that characteristics such as these could be predicted solely on the basis of household electricity consumption. The literature clearly identifies that load profiles are also a function of the number of household residents, and the household composition, the latter referring to the age structure and presence of children which may drive routines associated with education, for example (Druckman and Jackson 2008; Firth et al. 2008; Owen 2012; Wright 2008; Zimmerman et al. 2012). The survey dataset collected information on household composition, noting the presence of children and presence of seniors, plus a count of the number of household residents. Information of this nature is commonly collected via the Census, household social surveys and a range of administrative datasets. ONS recommendations to parliament following the 'Beyond2011' programme noted the important role of administrative data as a future source of information on household composition with potential to provide population counts and basic household composition at the small-area or address level (ONS, 2013, 2014a). Thus within this analysis we do not attempt to predict these characteristics; rather they represent predictors that we assume would be available at the small area geography. In our analysis, we therefore incorporate basic household composition alongside energy monitoring data to infer additional household characteristics of interest (income, floor area and employment status). The following section outlines a series of indicators that can be used to summarise electricity loads for use in subsequent analysis.

5 Profile indicators and household characteristics

Smart-meter like datasets such as the CER study present a number of challenges related to data storage, manipulation and analysis (Graham and Shelton 2013). Manipulating and processing smart-meter derived datasets often requires specialist high performance computing equipment and tools (See e.g. Thumim, Wilcox, and Roberts 2013) or aggregation and summary of time series data prior to analysis (Carroll, Lyons, and Denny 2014; Fintan McLoughlin 2013). Even for just 3,488 households over a four-week period, the thirty minute resolution measurement generated over 4.6 million records. To simplify the analysis we used a series of parameters or 'profile indicators' to summarise some of the temporal and magnitudinal features of household load profiles, facilitating comparison between households whilst also reducing the volume of data to be processed.

The literature provides a number of examples of indicators derived from load profiles as listed in Table 5. These indicators consider characteristics including load magnitude (base load, peak load), summary statistics (e.g. mean load), temporal properties such as the timing and duration of key features (e.g. time of use [max]) and ratios of, for example, peak to off-peak loads. Thus profile indicators provided a series of summary measure for each household whilst also helping preserve household privacy and removing redundant data. The process also considerably smoothed data on a household-by-household basis, reducing the impact of very rare or atypical high load events (Williams 2013). Nevertheless, the literature suggests that profile indicators maintain the ability to differentiate between households based on key features of their loads, such as magnitude or timing of their peak load (Fintan McLoughlin 2013). The use of profile indicators could thus offer considerable advantages if this form of analysis were up-scaled to incorporate far larger samples of households and time series of the order of months rather than weeks, with a commensurate increase in the volume of data to be stored, manipulated and handled.

Table 5: Parameters or 'profile indicators' to describe magnitude and temporal characteristics of load profiles

Parameter	Description	Source(s)	Possible predictor of ..
Base Load	Mean load 2am-5am	(Yohanis et al. 2008)	Number of residents, size of dwelling
97.5th Percentile Load	97.5th Percentile of ranked load – used rather than peak load which often represents an extreme peak value, driven by very short-term use of high power equipment	(Price 2010)	Income, employment status
Load Factor	Ratio of mean daily load to maximum daily load	(Carroll, Lyons, and Denny 2014; F McLoughlin, Duffy, and Conlon 2012)	Employment status, presence of children
Lunchtime load	Mean load between midday and 2pm	(Chicco et al. 2001)	Presence of seniors
Mean Load	Mean load across all timestamps	(Beckel, Sadamori, and Santini 2012; Yohanis et al. 2008)	Number of residents
Morning Maximum	Maximum load between 6am and 10.30am	(Carroll, Lyons, and Denny 2014)	Presence of children
Evening Consumption Factor (ECF)	Mean load during the evening peak (4pm-8pm) relative to the mean load at all other times of the day	(Powells et al. 2014)	Employment status
Total power consumed	Total power consumption (kWh) during the study period	(Fintan McLoughlin 2013)	Number of residents, income

As noted above we calculated the profile indicators listed in Table 1 over the midweek day (Tuesday – Thursday) periods based on the assumption that habits and routines associated with employment or study, which could reveal important household characteristics, will be more evident on weekdays. We have excluded Mondays and Fridays as these represent transition points with the weekend and households may thus exhibit atypical weekday behaviours.

Since all indicators summarise characteristics of the same load profiles, there may be a tendency for indicators to be strongly associated with each other, especially where they represent similar measures of magnitude. The 'Morning Maximum', 'Total Power Consumed' and '97.5th percentile load' are likely to be strongly correlated and therefore care is used when applying these indicators in subsequent analysis, ensuring that highly correlated indicators are not incorporated together within regressions or classifications. However, no indicators have been discounted as both the literature and prior exploratory analyses suggests that these indicators may reveal different household characteristics. In the following section we assess the potential of these indicators as predictors of key household characteristics of interest.

6 Estimating household attributes from load profiles

The literature, industry practice and our own exploratory analysis suggests that key features of household load profiles, as summarised via the profile indicators presented in Table 5, may be able to estimate a number of household attributes. We argue that exploring the value of such indicators for this purpose requires three main steps. The first is to use a multi-level mixed effects framework to identify whether or not known household characteristics can predict load profile indicators where those indicators are measured multiple times for each household. The second is to use the results of this step to select the most likely profile indicators and reverse the modelling direction to use them to estimate the household attributes. The final step is to assess the classification accuracy of these estimates at the household level.

6.1 Predicting profile indicators using household characteristics

The first step in this process used a multilevel regression modelling approach to allow for the appropriate modelling of repeated profile indicator measurements over the 3 day * 4 week = 12 days observed for each household. The models were constructed using a mixed effects framework as follows:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \cdots \beta_5 x_{5it} + u_{0i} + \varepsilon_{it}$$
$$u_{0i} \sim N(0, \sigma_{0u}^2)$$
$$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$$

where the dependent variable y_{it} is a profile indicator measurement for household i at time t, x_{1it} ,..., x_{5it} are the explanatory variables (see Table 6) for household i at time t. β_1 to β_5 are FE coefficients, representing the FE part of the model. u_{0i} is the RE on the intercept, which represents how an individual household differs from the average household, it is normally distributed with mean 0, variance σ_{0u}^2 . ε_{it} is the within-group residual with normal distribution of mean 0 and variance σ_ε^2.

Table 6: Coding of explanatory variables for multilevel models

Explanatory variable	Coding scheme

Income	6 income bands: < 7,500 euro < 22,500 < 40,000 < 62,500 < 92,500 >92,500
Self-reported employment status of HRP	0 = in paid work, 1 = not in paid work (unemployed, retired or caring role)
Presence of children	0 = no children, 1 = 1+ child
Number of residents	0 = 1 or 2 residents of any age, 1 = 3+ residents

Overall the results, which are summarised in Table 7, suggest that the number of residents and the presence of children are both statistically significant predictors of several mid-week consumption profile indicators, especially those reflecting the overall magnitude of consumption. Further, more detailed analysis using nested models (not shown) suggest that the inclusion of the Household Response Person's employment status modifies the effect of the number of residents and the presence of children for several profile indicators.

Income and (un)employment status are clearly correlated and so as we would expect the inclusion of one to largely mask the effects of the other for most profile indicators although this is not the case for the average baseload which is significantly predicted by income even when the number of residents is controlled.

(Un)employment status appears to predict the evening consumption factor as we might expect from Figure 1 and Table 5. However the relatively high residual value for this model (64%) suggests that it is less robust than the load factor model (47%).

Table 7: Effectiveness of household characteristics in predicting electricity consumption 'profile indicators' (values in bold are significant at the 95% level)

	Daily peak time		Daily peak 06:00 to 10.30		Daily average baseload (02:00 - 05:00)		Daily average	
	beta	Z	beta	Z	beta	Z	beta	Z
Constant	39.9	7.74	-0.80	-3.10	-0.27	-2.66	-0.44	-2.02
Number of residents	-0.19	-0.22	**0.21**	**4.94**	**0.06**	**3.73**	**0.24**	**6.63**
Income band	-0.72	-1.45	**0.10**	**3.88**	**0.04**	**3.75**	**0.07**	**3.10**
Number of children	0.27	0.42	**0.17**	**5.32**	0.00	0.38	**0.11**	**4.21**
Employment status of HRP	-1.01	-0.92	0.13	2.43	**0.05**	**2.44**	**0.11**	**2.27**
Marginal R2	0%		15%		7%		20%	
Conditional R2	20%		63%		64%		81%	
Residual R2	80%		37%		36%		19%	
	Daily sum		Daily 97.5th percentile		Evening Consumption Factor		Load factor	

	beta	Z	beta	Z	beta	Z	beta	Z
Constant	-21.69	-2.05	-0.03	-0.05	1.72	4.24	0.08	1.10
Number of residents	**11.64**	**6.69**	**0.83**	**6.77**	0.06	0.90	0.00	-0.09
Income band	**3.17**	**3.12**	**0.10**	**1.42**	-0.02	-0.49	**0.01**	**1.69**
Number of children	**5.49**	**4.22**	**0.44**	**4.81**	0.08	1.59	0.00	0.21
Employment status of HRP	**5.01**	**2.22**	0.06	0.36	**-0.18**	**-2.14**	**0.04**	**2.14**
Marginal R2	20%		16%		1%		1%	
Conditional R2	81%		66%		36%		53%	
Residual R2	19%		34%		64%		47%	

6.2 Predicting household characteristics using profile indicators

The second step was then to reverse the modelling process and test that ability of the load profile indicators to correctly predict household attributes. As noted above it was assumed that the number of residents and the number of children was already known through potentially available administrative data sources and the work reported here focuses only on the household response person's employment status as an exemplar.

A logistic regression approach was therefore used to estimate the probability that a Household Response Person (HRP) was not in paid work on the basis of number of residents, the number of children and the profile indicators selected as being most likely to be of value in Table 7 based on their ability to predict the HRP work status in the absence of other factors (ECF and LF). By applying a success threshold of 50% an estimate of the percentage of correctly classified HRPs could then be calculated as a simple within-sample validation test.

The results of this initial model (model 1) are shown in Table 8 and they suggest that whilst the evening consumption factor and load factor both had statistically significant predictive effects, the model was only able to correctly predict around 65% of HRP unemployment status.

Table 8: Logistic regression modelling results for HRP unemployment status (model 1)

	beta	t	p value
Number of residents < 3	0.42	2.30	0.02
Number of residents >= 3	-0.70	-7.90	0.00
Children present	-1.87	-16.82	0.00
Evening Consumption Factor (ECF)	-0.15	-2.59	0.01
Load Factor (LF)	1.93	3.38	0.00
Correct prediction:	65.42%		

In order to improve the performance and to test the relative value of the profile indicators and the 'known' demographic variables (number of residents and number of children), we estimated a series of increasingly more complex models. Thus base model 2.1 (see Table 9) included just the evening consumption factor (ECF) and the Load Factor (LF) as was the case for model 1 but did not include the number of residents or presence of children. Despite this the results suggest that nearly 60% of HRPs were correctly classified.

In attempt to improve classification performance we drew on McLoughlin et al (McLoughlin, Duffy, & Conlon, 2013) to develop clusters of households with similar consumption profiles. Cluster membership was calculated via a weighted least squares and k-means clustering process using only the half-hour consumption profiles. This produced six clusters of households of which two captured the majority (33% and 28% respectively) with the remainder distributed roughly evenly across the remaining four (model 2.2). As Table 9 shows the inclusion of these clusters increased the performance of the model by just under 5 percentage points with only membership of cluster 3 proving not to be a statistically significant predictor.

In order to improve the model still further (model 2.3) we then included an indicator of 'habitual behaviour' by calculating an autocorrelation coefficient for the 24 hour lag of each half hourly consumption for each household on mid-week days after the hours of sleep (00:00 – 06:00) were removed to avoid artificially increasing the lag correlation. This coefficient is therefore an indicator of the degree to which mid-week consumption between 06:00 and 00:00 is replicated at the same time on subsequent days for each household and, based on exploratory analysis (not shown) we expected lower autocorrelation (less 'regular habits') for those not in paid work. In general as Figure 2 shows the coefficients followed an expected 24 hour profile with the highest being the immediately following half hours before a gentle decline and then rise to a higher correlation at the 24 hour lag (i.e. the same time the next day) which in this case is represented by the 36th lag due to the removal of sleep hours. The model included the coefficients at lags 36 (the same time tomorrow) and 72 (the day after tomorrow) however as Table 9 shows, the inclusion of this habituality indicator produced a marginal improvement in performance with only the lag 36 coefficient proving to be statistically significant.

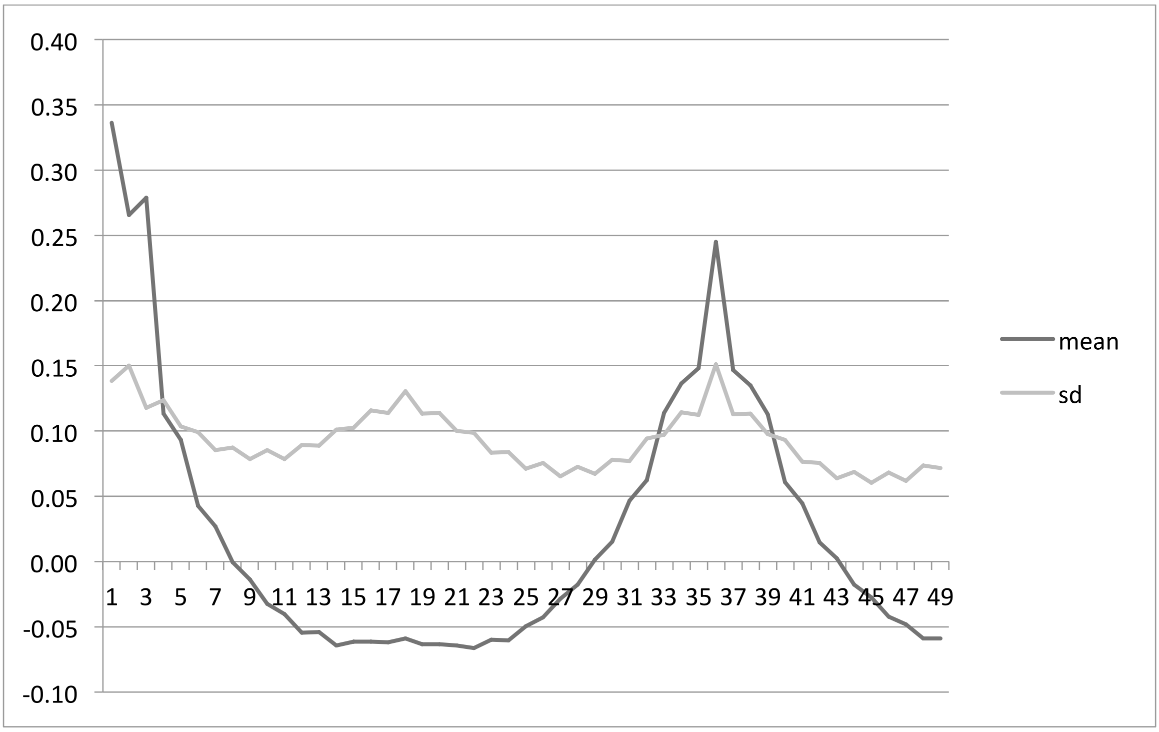

Figure 2: Mean and standard deviation of AR coefficient by lag for mid-week days in October 2009

Finally (model 2.4) we re-introduced the presence of children and household size variables to understand the additional value of this potentially administratively sourced data. As Table 9 shows, the increasingly complex models were significantly different from the simpler versions (LR test results) while the adjusted pseudo r-squared scores (McFadden) also increased as the additional variables were added reflecting increased improvement over the intercept model in each case. Re-adding the presence of children and dummy for larger households increased the pseudo r-squared score substantially and the classification success by 6.2 percentage points and, as might be expected, reduced the predictive power of the ECF as well as most cluster membership.

Table 9: Logistic regression modelling results for HRP employment status

	Model 2.1: Base model		Model 2.2: with cluster membership		Model 2.3: with 24 hour autocorrelation coefficient		Model 2.4 with presence of children and 3+ persons	
Explanatory variable	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p
Intercept	-0.312	0.074	-1.199	0.000	-1.238	0.000	-0.164	0.514
Evening Consumption Factor	-0.260	0.000	-0.235	0.000	-0.176	0.001	-0.062	0.260
Load Factor	1.947	0.000	3.038	0.000	3.477	0.000	3.324	0.000
Cluster 2 membership (compared to Cluster 1)			-0.354	0.047	-0.251	0.164	-0.500	0.011
Cluster 3 membership			-0.346	0.173	-0.318	0.212	-0.525	0.058
Cluster 4 membership			0.702	0.000	0.728	0.000	0.040	0.823
Cluster 5 membership			0.846	0.000	0.829	0.000	0.477	0.011
Cluster 6 membership			1.039	0.000	1.045	0.000	0.019	0.916
Lag 36 coefficient					-0.278	0.574	-0.641	0.218
Lag 72 coefficient					-1.351	0.010	-1.010	0.069
Presence of children							-1.471	0.000
3+ persons							-0.860	0.000
N	3160		3160		3160		3160	
McFadden Pseudo R^2	0.024		0.067		0.075		0.150	
LR tests:								
Model 2.1 v 2.2	$p < 0.0001$							
Model 2.2 v 2.3	$p = 0.0001$							
Model 2.3 v 2.4	$p < 0.0001$							
Classification rate	59.3%		63.8%		63.9%		70.1%	

Overall, these results suggest that model 2.4 is most comprehensive among the list of models tested, with significant predictors approximating household energy usages, energy usage behaviours and administrative variables. Our simplest model (model 2.1) indicates that although the absence of administrative data reduced the ability of electricity consumption profile indicators to predict HRP employment status, the success rate was still close to 60%. This, together with the relatively unchanging regression coefficients for the profile indicators in each model suggests that most of the differentiation captured by the profile clusters and all of that captured by the 'habitual behaviour' indicator may already be embodied in the profile indicators used in model 2.1.

7 Conclusions and next steps

This paper has started the process of assessing the feasibility of using household electricity load profiles as a tool to infer key household characteristics. Using a smart meter-like dataset, we generated a series of load profile indicators that summarise key features of household load profiles, enabling differentiation between households. These indicators, coupled with household composition, offered a degree of predictive potential for the characteristic tested and, when compositional data was excluded but other consumption indicators included, this potential was still substantial especially when membership of twenty four hour demand profile clusters was included. This suggests that electricity consumption data of this kind could be used to independently estimate the employment status of HRPs as a means of validating (or contributing to) Census estimates and it could also be used to produce slightly less robust estimates in the absence of administrative data. This may be especially pertinent to mid-census period estimates or to situations where only smart meter data is available which may well be the case for stakeholders who do not have access to Government held administrative data sources.

However it must be recognised that the size of the sample used in this work has precluded the testing of the differential performance of the process in different sub-populations or in different regions. In both cases we would expect reduced heterogeneity and thus increased performance of the estimation process. Whilst it would in principle be possible to develop the work to test larger sub-groups of this sample, we concluded that the consequential reduction in statistical power would mean that such work requires a much larger representative population sample. Further the lack of any geo-coding in the data also precludes analysis of regional differences within the sample itself.

The use of load profile indicators may also be over-simplifying the nuanced detail within the load profiles. Given that the range of household types available within the dataset is relatively narrow, detailed temporal electricity consumption behaviour not captured by the profile indictors may be useful in order to discriminate between households. Whilst evidence from the literature suggests that profile indicators are frequently used to extract meaningful information from load profiles, it may be beneficial to work with more of the time-series data in order to ensure that the profile indicators do not mask habits which may prove useful in differentiating between households. Such an approach may provide opportunities to build regression models or generate clusters for different days of the week, recognising that load profiles may be very different on weekdays and at weekends, and that the difference between weekday and weekend profiles may, for example, allow inferences to be made about household characteristics whilst differences between seasons and school vs non-school holiday periods may also be instructive.

Overall the analysis confirms existing literature suggesting that profile indicators are potentially useful summaries of key features of household electricity load profiles. The findings suggest that analytic approaches such as regression and classification could offer potential in inferring key small area household characteristics from load profile indicators and basic household composition. This approach could add considerable additional 'value' to domestic smart metering, enabling remote and quasi real-time estimation of small area population statistics.

However this approach has the potential for impact beyond the delivery of enhanced population statistics. Unlike existing small area statistics or periodic household sample surveys, quasi-real time observation of energy use and behaviours could be used to both *target* and assess the *impact* of neighbourhood or household level (energy) policy or market interventions. Thus these datasets offer the potential to target area based (energy) policy and to remotely evaluate impacts in near-real time without the need for household sample surveys.

The next steps in this work must be to explore additional statistical methods to more robustly estimate attributes from consumption profiles and to test the performance of these models using not only out-of-sample validation of household level estimates but also by aggregating and validating against real Census data.

In the case of statistical methods there may be considerable scope to develop multi-level hierarchical models of the kind proposed for use in the estimation of historical climate characteristics from sparse proxy observations (M. K. Hughes and Ammann 2009; Tingley et al. 2012). Such development should include the development of appropriate uncertainty measures for both point (household) and aggregated area level estimates. Future work should also consider the potential value of including area level and temporal co-variates to enable the function linking the consumption profiles to the household attributes to vary spatially and over time.

In terms of household level validation, it is possible that the Irish CER data set used in this paper may be sufficiently large to support out-of-sample validation but doing so is likely to considerably increase uncertainty. It is likely that such validation will only become possible when substantially larger samples of suitably linked consumption and household attribute data become available. Such data would also support the kind of sub-population and sub-regional analysis discussed above.

Finally, area level validation of aggregated estimates would require access to anonymised large-scale smart meter data extracts from either all households or a representative sample of them in known small area geographies which could be used as the basis for model-based estimation of household characteristics. These estimates could then be aggregated to current Census geographies and validated against recently observed Census-derived population statistics. Unfortunately as far as we are aware, such datasets do not currently exist in the UK.

8 Acknowledgements

This work is part of the activities of the Energy and Climate Change Division at the University of Southampton (www.energy.soton.ac.uk) and was supported by the UK Economic and Social Research Council (ESRC) under grant ES/L00318X/1.

The data used in this paper was collected by and is copyright the Irish Commission for Energy Regulation (CER) and was accessed via the Irish Social Science Data Archive (ISSDA - www.ucd.ie/issda/data). Neither CER nor ISSDA bear any responsibility for the further analysis or interpretation of the data.

9 References

Beckel, Christian, Leyna Sadamori, and Silvia Santini. 2012. 'Towards Automatic Classification of Private Households Using Electricity Consumption Data'. In *Proceedings of the Fourth ACM*

Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 169–76. ACM. http://dl.acm.org/citation.cfm?id=2422562.

———. 2013. 'Automatic Socio-Economic Classification of Households Using Electricity Consumption Data'. In *Proceedings of the Fourth International Conference on Future Energy Systems*, 75–86. ACM. http://dl.acm.org/citation.cfm?id=2487175.

Calder, A., and A. Teague. 2013. 'The Census and Future Provision of Population Statistics in England and Wales: Presentation Delivered at RGS "The Future of Small Area Population Statistics" 21st October 2013'. Newport: Office for National Statistics.

Carroll, James, Seán Lyons, and Eleanor Denny. 2014. 'Reducing Household Electricity Demand through Smart Metering: The Role of Improved Information about Energy Saving'. *Energy Economics* 45: 234–43.

Chicco, G., R. Napoli, P. Postolache, M. Scutariu, and C. Toader. 2001. 'Electric Energy Customer Characterisation for Developing Dedicated Market Strategies'. In .

Claxton, R, J Reades, and B Anderson. 2012. 'On the Value of Digital Traces for Commercial Strategy and Public Policy: Telecommunications Data as a Case Study'. In *The Global Information Technology Report 2012*, edited by S Dutta and B Bilbao-Osorio. Geneva: World Economic Forum.

Craig, Tony, J. Gary Polhill, Ian Dent, Carlos Galan-Diaz, and Simon Heslop. 2014. 'The North East Scotland Energy Monitoring Project: Exploring Relationships between Household Occupants and Energy Usage'. *Energy and Buildings* 75 (June): 493–503. doi:10.1016/j.enbuild.2014.02.038.

DECC. 2012. 'Smart Metering Implementation Programme - Programme Update April 2012'. London: Department of Energy and Climate Change.

———. 2013. 'Smart Metering Equipment Technical Specifications Version 2'. London: Department of Energy and Climate Change.

Deloitte. 2011. 'Empowering Ideas 2011: A Look at Ten of the Emerging Issues in the Power and Utilities Sector'. Cleveland, Ohio: Deloite Center for Energy Solutions.

Deville, Pierre, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. 2014. 'Dynamic Population Mapping Using Mobile Phone Data'. *Proceedings of the National Academy of Sciences* 111 (45): 15888–93.

Dilnot, A. 2014. 'The Census and Future Provision of Population Statistics in England and Wales. Letter to Rt. Hon. Francis Maude MP from the Chair of the UK Statistics Authority, Sir Andrew Dilnot CBE, 27th March 2014'. London: UK Statistics Authority.

Douglass, Rex W., David A. Meyer, Megha Ram, David Rideout, and Dongjin Song. 2015. 'High Resolution Population Estimates from Telecommunications Data'. *EPJ Data Science* 4 (1): 1–13.

Druckman, A, and T Jackson. 2008. 'Household Energy Consumption in the UK: A Highly Geographically and Socio-Economically Disaggregated Model'. *Energy Policy* 36 (8): 3177–92.

Dugmore, K. 2013. 'The Census and Future Provision of Population Statistics in England and Wales - Public Consultation September 2013. Response by the Demographics User Group (DUG)'. London: Demographics User Group.

Dugmore, Keith, Peter Furness, Barry Leventhal, and Corrine Moy. 2011a. 'Beyond the 2011 Census in the United Kingdom: With an International Perspective'. *International Journal of Market Research* 53: 619. doi:10.2501/ijmr-53-5-619-650.

———. 2011b. 'Information Collected by Commercial Companies: What Might Be of Value to ONS?' *International Journal of Market Research* 53 (5): 619–50.

Elexon. 2013. 'Load Profiles and Their Use in Electricity Settement'. London. http://www.elexon.co.uk/wp-content/uploads/2013/11/load_profiles_v2.0_cgi.pdf.

Energy UK. 2013. 'About Smart Meters'.

Eurostat. 2011. 'EU Legislation on the 2011 Population and Housing Censuses: Explanatory Notes'. Luxembourg: European Commission (Eurostat).

Firth, Steven, K. Lomas, A. Wright, and R. Wall. 2008. 'Identifying Trends in the Use of Domestic Appliances from Household Electricity Consumption Measurements'. *Energy and Buildings* 40 (5): 926–36.

Graham, Mark, and Taylor Shelton. 2013. 'Geography and the Future of Big Data, Big Data and the Future of Geography'. *Dialogues in Human Geography* 3 (November): 255–61. doi:10.1177/2043820613513121.

Haben, Stephen, Jonathan Ward, Danica Vukadinovic Greetham, Colin Singleton, and Peter Grindrod. 2014. 'A New Error Measure for Forecasts of Household-Level, High Resolution Electrical Energy Consumption'. *International Journal of Forecasting* 30 (2): 246–56. doi:10.1016/j.ijforecast.2013.08.002.

Hamidi, Vandad, Furong Li, and Francis Robinson. 2009. 'Demand Response in the UK's Domestic Sector'. *Electric Power Systems Research* 79: 1722–26. doi:10.1016/j.epsr.2009.07.013.

House of Commons Treasury Committee. 2008. 'Counting the Population: Eleventh Report of Session 2007-08. Volume 1, Report, Together with Formal Minutes,oral and Written Evidence.' London: The Stationery Office Limited.

Hughes, M. K., and C. M. Ammann. 2009. 'The Future of the Past—an Earth System Framework for High Resolution Paleoclimatology: Editorial Essay'. *Climatic Change* 94 (3-4): 247–59.

Hughes, M., and G. Moreno. 2013. 'Further Analysis of Data from the Household Electricity Usage Study: Consumer Archetypes'. Cambridge: Element Energy Ltd.

Martin, D. 2006. 'Last of the Censuses? The Future of Small Area Population Data'. *Transactions of the Institute of British Geographers* 31: 6–18.

McKenna, Eoghan, Ian Richardson, and Murray Thomson. 2012. 'Smart Meter Data: Balancing Consumer Privacy Concerns with Legitimate Applications'. *Energy Policy* 41 (February): 807–14. doi:10.1016/j.enpol.2011.11.049.

McLoughlin, F, A Duffy, and M Conlon. 2012. 'Characterising Domestic Electricity Consumption Patterns by Dwelling and Occupant Socio-Economic Variables: An Irish Case Study'. *Energy and Buildings*. http://www.sciencedirect.com/science/article/pii/S0378778812000680.

———. 2013. 'Evaluation of Time Series Techniques to Characterise Domestic Electricity Demand'. *Energy* 50 (1): 120–30.

McLoughlin, Fintan. 2013. 'Characterising Domestic Electricity Demand for Customer Load Profile Segmentation [THESIS]'. Dublin: Dublin Institute of Technology.

Newing, Andy, Ben Anderson, AbuBakr Bahaj, and Patrick James. 2015. 'The Role of Digital Trace Data in Supporting the Collection of Population Statistics - the Case for Smart Metered Electricity Consumption Data'. *Population, Space and Place*, July, EarlyView. doi:10.1002/psp.1972.

Ning, Z., and D. Kirschen. 2010. 'Preliminary Analysis of High Resolution Domestic Load Data'. Manchester: School of Electrical & Electronic Engineering, University of Manchester.

Norman, P. 2013. 'The Case for Small Area Data. Presentatopn Delivered at the Beyond 2011 Research Conference, Universiry of Southampton, 30th April-1st May 2013'. Leeds: University of Leeds.

ONS. 2013. 'Beyond 2011: Options Report'. London.

———. 2014a. 'Beyond 2011: Final Options Report'. London.

———. 2014b. 'The Census and Future Provision of Population Statistics in England and Wales: Report on the Public Consultation'. Newport: Office for National Statistics.

Owen, P. 2012. 'Powering the Nation: Household Electricity Habits Revealed'. London: Energy Saving Trust.

Powells, Gareth, Harriet Bulkeley, Sandra Bell, and Ellis Judson. 2014. 'Peak Electricity Demand and the Flexibility of Everyday Life'. *Geoforum* 55: 43–52. doi:10.1016/j.geoforum.2014.04.014.

Price, Phillip. 2010. 'Methods for Analyzing Electric Load Shape an Its Variability'. Sacramento: Lawrence Berkeley National Laboratory & California Energy Commission.

Pucci, Paola, Fabio Manfredini, and Paolo Tagliolato. 2015. *Mapping Urban Practices Through Mobile Phone Data*. SpringerBriefs in Applied Sciences and Technology. Cham: Springer International Publishing. http://link.springer.com/10.1007/978-3-319-14833-5.

Ramsden, S. 2014. 'Briefing: Size Criteria ('Bedroom Tax')'. London: National Housing Federation.

Skinner, C., J. Hollis, and M. Murphy. 2013. 'Beyond 2011: Independent Review of Methodology'. London: Independent review for the UK Statistics Authority.

Struijs, P., B. Braaksma, and P. J. Daas. 2014. 'Official Statistics and Big Data'. *Big Data & Society* 1 (1): 2053951714538417–2053951714538417. doi:10.1177/2053951714538417.

Thumim, J., T. Wilcox, and S. Roberts. 2013. 'Managing and Mining Smart Meter Data - at Scale. Presentation Delivered at the CSE Project Showcase, 9th July 2013'. Bristol: Centre for Sustainable Energy.

Tingley, Martin P., Peter F. Craigmile, Murali Haran, Bo Li, Elizabeth Mannshardt, and Bala Rajaratnam. 2012. 'Piecing Together the Past: Statistical Insights into Paleoclimatic Reconstructions'. *Quaternary Science Reviews* 35 (March): 1–22. doi:10.1016/j.quascirev.2012.01.012.

Watson, G. 2009. 'Making the Case for the 2011 Census. Presentation Delivered by at the DUG Annual Conference, 8th October 2009'. Newport: Office for National Statistics.

Williams, John. 2013. 'Clustering Household Electricity Use Profiles'. In *Proceedings of Workshop on Machine Learning for Sensory Data Analysis - MLSDA '13*, 19–26. ACM Press. doi:10.1145/2542652.2542656.

Wright, Andrew. 2008. 'What Is the Relationship between Built Form and Energy Use in Dwellings?' *Energy Policy* 36 (12): 4544–47. doi:10.1016/j.enpol.2008.09.014.

Wright, Andrew, and Steven Firth. 2007. 'The Nature of Domestic Electricity-Loads and Effects of Time Averaging on Statistics and on-Site Generation Calculations'. *Applied Energy* 84 (4): 389–403.

Yohanis, YG, JD Mondol, A Wright, and B Norton. 2008. 'Real-Life Energy Use in the UK: How Occupancy and Dwelling Characteristics Affect Domestic Electricity Use'. *Energy and Buildings* 40 (6): 1053–59.

Zimmerman, Jean-Paul, Matt Evans, Jonathan Griggs, Nicole King, Les Harding, Penelope Roberts, and Chris Evans. 2012. 'Household Electricity Survey: A Study of Domestic Electrical Product Usage'. Milton Keynes.

Zoha, Ahmed, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar. 2012. 'Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey'. *Sensors* 12 (12): 16838–66.