Energy Policy 50 (2012) 419–427

Contents lists available at SciVerse ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

A housing stock model of non-heating end-use energy in England verified by aggregate energy use data

Stephen Lorimer*

University College London, UCL Energy Institute, Central House, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom

HIGHLIGHTS

- ▸ Housing stock energy model was developed for end-uses outside of heating for UK context.
- ▸ This entailed changes to the building energy model that serves as the bottom of the stock model.
- ▸ The model is adaptable to reflect rapid changes in consumption between major housing surveys.
- ▸ Verification was done against aggregated consumption data and for the first time uses a measured size of the housing stock.
- ▸ The verification process revealed spatial variations in consumption patterns for future research.

ARTICLE INFO

Article history:
Received 28 February 2012
Accepted 22 July 2012
Available online 2 September 2012

Keywords:
Energy
Domestic
Model

ABSTRACT

This paper proposes a housing stock model of non-heating end-use energy for England that can be verified using aggregate energy use data available for small areas. These end-uses, commonly referred to as appliances and lighting, are a rapidly increasing part of residential energy demand. This paper proposes a model that can be verified using aggregated data of electricity meters in small areas and census data on housing. Secondly, any differences that open up between major collections of housing could potentially be resolved by using data from frequently updated expenditure surveys. For the year 2008, the model overestimated domestic non-heating energy use at the national scale by 1.5%. This model was then used on the residential sector with various area classifications, which found that rural and suburban areas were generally underestimated by up to 3.3% and urban areas overestimated by up to 5.2% with the notable exception of "professional city life" classifications. The model proposed in this paper has the potential to be a verifiable and adaptable model for non-heating end-use energy in households in England for the future.

1. Introduction

This paper focuses on the modelling of non-heating end-use energy in the housing stock of England. "End-use energy" is defined in this paper as the energy delivered to a user within a building (Sørensen, 2011). End-uses are the ways that energy can be used by people within the context of a domestic home – for example, a user can turn on (or set a timer for) heating, or turn on a light. A non-heating end-use for domestic homes in the United Kingdom is defined by this author as energy used for appliances, lighting, electronics, and cooking and not for space heating and cooling nor water heating.

Data from recent housing energy fact files for Great Britain reveals that from 1970 to 2006, the energy consumption of lighting and appliances grew by 148 per cent compared to 23 percent overall (Shorrock and Utley, 2008, Department of Energy and Climate Change, 2011d). In the same time frame, electricity use in the UK housing stock has risen by 161 per cent (Department of Energy and Climate Change, 2011a). This is supported by other data released by the Office of National Statistics in the UK (Office for National Statistics, 2010).

Examining future forecasts, this trend is set to continue and even accelerate. The main driver of domestic energy consumption outside of heat demand is information and communications technology (ICT) and consumer electronics (CE) as illustrated in Fig. 1.

In the period 1990–2030, the average unit consumption of electricity for these uses in Organisation for Economic Co-operation and Development (OECD) countries is projected by the International Energy Agency to rise 75 per cent. There is also the multiplying effect of having more of them – the average person will have three times as many of these 'gadgets' at the end of the 1990–2030 timeframe as the beginning. For illustration,

* Present address: UCL Centre for Transport Studies, University College London, Department of Civil, Environmental, and Geomatic Engineering, Chadwick Building Gower Street, London WC1E 6BT, United Kingdom.
E-mail address: s.lorimer@ucl.ac.uk

0301-4215
http://dx.doi.org/10.1016/j.enpol.2012.07.037

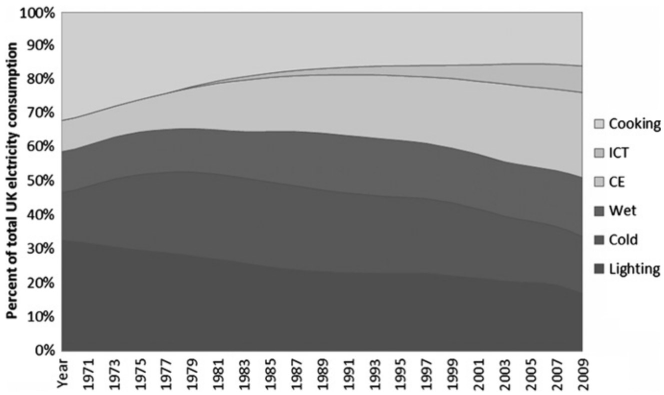

Fig. 1. Total UK electricity consumption by household domestic appliances 1970 to 2010 (Department of Energy and Climate Change, 2011c).

a person using 100 kW h (kWh) per year in 1990, for consuming ICT and CEs, is likely to be using 525 kWh per year in 2030 (International Energy Agency, 2009). Clearly, although the climate change impact caused by lights, appliances, and electronics is less than for heating and future planning is underway for changing the mix of electricity generation (Department of Energy and Climate Change, 2009), this growth is significant and needs further investigation as the potential for energy demand reduction in the medium term is greater than for gas-fuelled heating (Market Transformation Programme, 2008).

This study is a secondary data analysis of electricity meter readings and fuel expenditure surveys in England. It is a distinctive setting for study as the majority of homes have natural gas-fed, single household heating systems. For the researcher, this creates a significant benefit which removes electric space heating use from large numbers of homes as a confounder when correlating electricity consumption with non-heating end-uses. Energy fuels in the home in Great Britain have continuously diverged to natural gas as the fuel for space and water heating end-uses and electricity for all other end-uses (with cooking split between the two) since 1970. By 2008, centrally heated dwellings comprised 90 percent of the total housing stock, with 90 percent of central heating systems fuelled by natural gas (Shorrock and Utley, 2008). This equated to 81 percent of all households using natural gas for space and water heating.

The current housing stock model being used by the UK government during 2012, the Cambridge Housing Model (CHM), is a bottom-up model of the energy use of the residential sector using a variation of the Standard Assessment Procedure (SAP)[1] of the BRE Domestic Energy Model (BREDEM) as the building energy model that forms the "bottom" of the model (Cambridge Architectural Research Ltd. & DECC, 2011). The model guide released by UK Department of Energy and Climate Change (DECC) describes it as a simulation of first the size, and then the energy consumption, of the entire housing stock. The building stock in CHM is the 2009 English Housing Survey, but the data only relates to the physical characteristics of dwellings and does not include energy use data. The model in this paper will be using households that have both their physical characteristics and their energy consumption surveyed. The CHM verified against national energy consumption measured from electricity and gas meters, but using a modelled size of the housing stock. This paper also proposes to create a model verified against energy use measured by meters, but using a measured size of the housing stock.

[1] SAP (Standard Assessment Procedure) is a simplified version of BREDEM that is officially adopted by the Department for Communities and Local Government to analyse the energy consumption of a dwelling.

The underlying equation for lights and appliances use within BREDEM (Anderson, 2002, BRE, 2010) is:

$$E_i = 267.53 \times (TFA \times N)^{0.4714} \quad (1)$$

where E_i is the predicted energy use due to lights and appliances in kilo-watt hours per year, *TF* is the usable floor area of the building in square meters, and *N* is the number of occupants of the household (which is based on the usable floor space if it is not known).

Correction factors based on seasonality and on the physical attributes of low-energy lightbulbs and the daylighting of the building can alter this prediction if known. Usually, this only applies in new buildings where the algorithm would be used as part of a building permission application.

There are aggregated data sources available at the level of the English Lower Layer Super Output Area (LLSOA). An LLSOA is a local area with around 3,000 households created as a geographic unit by the Office of National Statistics for use in conjunction with the 2001 Census. Physical household size is available measured, not as usable floor space and its consequent number of assumed occupants, but as the numbers of rooms and occupants, from the 2001 United Kingdom Census. The total electricity use per LLSOA is available from the DECC. Some of these LLSOAs have a monoculture of single-household natural gas-fed heating systems such that electricity use can represent non-heating end-use energy. Housing stock models of non-heating energy should ideally take advantage of this aggregated data to validate the model against actual usage. This paper proposes such a model and validates the results over selected ranges of LLSOAs in England.

2. Data sources

Measurement of non-heating end-use energy in England at the national level is measured in this paper using housing surveys and small-area aggregate statistics. There is one recent housing survey that asked participants to record metered electricity data along with recording the heating fuel of the household – the 1996 English House Condition Survey (Department of the Environment Transport and the Regions, 2000). The amount of electricity use of small areas is collected by energy settlement and balance companies and released by the UK Department of Energy and Climate Change (Department of Energy and Climate Change, 2010b).

2.1. Individual household electricity data

The source of individual household data of energy use at the national level comes through housing surveys. The English Housing Survey is the main current survey of both physical (dwelling) and household (occupant) characteristics (Department for Communities and Local Government., 2010), but directly measured meter data is still unavailable. Other major databases such as the Homes Energy Efficiency Database (HEED) contain a wealth of physical dwelling characteristics in homes that have installed energy efficiency installations, but without directly measured energy consumption (Energy Saving Trust, 2008). The National Energy Efficiency Database (NEED), currently under development by DECC matches the dwellings in HEED with meter data (Department of Energy and Climate Change, 2011e), but as it contains home addresses, access is severely limited until work is done to anonymise all of the records (Steadman, 2012).

Unfortunately for the energy research community, the last large-scale national housing survey with open data from meter readings was the 1996 English House Condition Survey (EHCS) (McIntyre, 2011). A fuel sub-sample was taken from this survey where participants kept a 27-month diary of their gas and electricity meters between 1997 and 1999. Combined with household and occupancy characteristics, this dataset is the richest and most complete picture of energy use in English households to date (Department of the Environment Transport and the Regions, 2000). However, the data is becoming out of date, and there are other surveys, notably the Living Costs and Food Survey (LCFS), available that ask respondents about their last household energy bills with additional household data that can enable an updated estimate of energy use in the present day when harmonised with the 1996 EHCS (Office for National Statistics, 2010 and Department for Environment Food and Rural Affairs, 2010).

Unlike the Living Costs and Fuel Survey which is a simple random sample, the 1996 English House Condition Survey (EHCS) (Department of the Environment Transport and the Regions, 2000) developed a stratified sampling method based around the strata of building age (pre- or post-1945), tenure (private or public), dwelling type (house or flat) and English government office region (eight regions). More common combinations of these characteristics were undersampled, and conversely less common combinations were oversampled. The fuel sample, of which this survey concerns itself, consisted of 3,676 homes. Annual fuel use for electricity and natural gas were calculated using up to 9 consecutive quarters of metered fuel data. The 2,531 homes with the most reliable data were given gross weighting factors in proportion to the individual homes' representativeness in the population.

The Department for Communities and Local Government later provided additional location data for homes that were visited for an interview in the 1996 EHCS in the form of the Office for National Statistics Area Classification (Office for National Statistics, 2008) of the LLSOA in which the dwelling was located (McIntyre, 2011). The spatial independence (Longley and Tobon, 2004) of the EHCS cases was tested, and the null hypothesis that area classification and government office region of the homes in the 1996 EHCS are correlated is rejected ($\chi^2 = 1471$, $p < 0.01$).

Heating systems powered by electricity increases the amount of electricity used in a household approximately six-fold (Department of the Environment Transport and the Regions, 2000). Therefore the final dataset derived from the EHCS in this secondary data analysis was required to consist only of homes that do not use electricity for space or water heating. Homes that reported in interviews conducted as part of the EHCS using electricity as a primary or secondary heating source are removed from the dataset. EHCS interviewees that either did not have central heating or did not know the fuel source for their home, or did not report any use of natural gas are also removed from this analysis, as it was possible that electricity could be used as a heating fuel. Finally, homes that did not respond to the energy survey with electricity or gas meter data are removed. This reduces the number of cases down to 2,399 of which 1,776 had a gross weighting factor attached to it.

The 2008 Living Costs and Food Survey (LCFS) (Office for National Statistics and Department for Environment Food and Rural Affairs, 2010) interviewed 5,843 households and asked them about the last expenditure they made on electricity. Again, the analysis in this paper required that the final dataset consisted of only homes surveyed in the LCFS that did not use electricity for space or water heating. Those who reported in the LCFS using electricity as a primary or secondary heating fuel are excluded. The dataset also needs the electricity expenditure to be over a defined time period, meaning that those paying quarterly or monthly were included, but those prepaying for electricity are removed. This reduces the number of households considered in this analysis to 4,929.

2.2. Aggregated data of electricity use in the residential sector

The Department for Energy and Climate Change has produced aggregate data for domestic energy use since 2005 (Department of Energy and Climate Change, 2010b). These are produced for all 34,378 LLSOAs in England. The 2001 United Kingdom Census contains data on the possession of central heating in households. If there were a small but significant number of people that reported not having central heating, then that census area is considered likely to have a significant uplift on the number of homes that use electricity for heating end-uses. 31% of homes in 2001 without central heating used electricity as the heating fuel. Only 10% of homes with central heating in 2001 used electricity as the fuel; by 2006, this proportion had decreased to 5% (Shorrock and Utley, 2008).

Three options for a "significant" number of households not using central heating were assessed at 1, 5, and 10 percent of each LLSOA in England. To ensure that small numbers of LLSOAs remaining did not have undue influence, each option was required to have at least four LLSOAs remaining per ONS LLSOA area classification group. The 1 percent option was rejected, but the 5 and 10 percent options passed this criteria. Therefore the 5 percent option was adopted as it kept the higher number of LLSOAs: 10,350 out of 34,378.

3. Method

The model uses linear regression to estimate the annual household non-heating end-use energy of the building energy model that makes up the "bottom" of a bottom-up housing stock model. There are some steps to the teasing out of complex relationships between a dependent variable and multiple independent variables. Multiple regression, simple regression by category, and simple regression using an interaction term were considered. The simple regression method using an interaction term was selected due to its current adoption in this area and its flexibility.

The traditional method of dealing with more than one independent variable is to give each of them their own independent term x. This multiple regression equation using individual household data with two independent variables would then be

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \quad (2)$$

where y_i is the outcome variable of non-heating end-use energy in a household, β_0 is the intercept, β_1 and β_2 are two slope coefficients relating to independent variables x_{1i} and x_{2i} and $e_i$is the residual, or the difference between the actual y-value and the one that is predicted by the x-value.

There are other methods available to provide an alternative conclusion. The most obvious is to split the sample and fit a regression for each category of household size. However, there are several problems that can occur. The sample size can become unacceptably small. There can be more than one categorical predictor, and the effects of other independent variables can vary for each possible group. There can be several independent variables predicting the dependent variable, but not all of these variables will vary across the grouping, leading to redundant and separate regressions. Finally, and crucially, hypothesis tests cannot be carried out to compare regression coefficients across the different regressions for each grouping.

One solution, and this has been implemented in the BREDEM model currently in use in the United Kingdom, is the fitting of a pooled sample or interaction term as the product of the two independent variables. The pooled variable $x_1\_x_2 = x_1 \times x_2$ in this case replaces the previous two-variable multiple regression

$$y_i = \beta_0 + \beta_3 x_1\_x_{2i} + e_i \quad (3)$$

$x_1\_x_2$ is known as the interaction effect between x_1 and x_2 and allows the effect of x_1 to differ for each interval of x_2. An interaction effect is found if the effect of x_2 multiplies the effect of x_1.

An interaction effect is "the differing effect of one independent variable on the dependent variable, depending on the particular level of another independent variable" (Cozby, 1997). The interaction effect between two independent variables used in SAP2009 and in previous versions of the BRE Domestic Energy Model is $\{TFA \times N\}$. This interaction term was used by modellers of domestic energy use in the 1980s to indicate that an increase in occupants multiplies the effect of increased floor space on energy consumption in residential buildings (Henderson, 2009). In the context of non-heating end-use energy, this is a plausible hypothesis: buildings do not use electricity without demand from people, and people need devices that "inhabit" buildings with them to operate.

This interaction effect also has practical benefits for researchers and creators of algorithms in past BREDEM and the current SAP2009 models. The distribution of annual energy consumption is skewed and likely exhibits non-linear behaviour. These can be estimated using simple linear regression methods with an interaction term when there is a "long tail" of larger households and larger energy-consumption households in the dataset that results in the data failing the parametric test. Linear modelling of positively skewed datasets transformed with a logarithmic or square root of the dependent variable of non-heating end-use energy can produce a meaningful algorithm, if there is only one independent term, such as an interaction term or a single independent variable.

4. Calculation

There are two important aspects to consider for the statistical calculation of the non-heating end-use energy of the housing stock in England – the composition of the interaction term in the building energy model and making sure that all the data refers to a single year when the bottom-up model is built up to the entire residential sector. The interaction term in SAP2009 is composed of occupants and floorspace. For validation, it is desirable to compose the interaction term as the product of occupants and habitable rooms as defined by the 2001 United Kingdom Census (Office for National Statistics, 2004) and collected by Lower Layer Super Output Area.

4.1. Correlation test of different measures of household size

The assumption that the higher resolution variable, total usable floorspace, leads to a better correlation with non-heating end-use energy was tested using the Spearman correlation. The Spearman correlation coefficient can be defined as the Pearson correlation coefficient between two ranked variables. This is a correlation test that is independent of the distribution of the two variables considered (Myers et al., 2010, Conover, 1999).

Both correlation procedures measured both the strength and the direction of a linear relationship. If one variable X is an exact linear function of another variable Y, a positive relationship exists if the correlation is 1 and a negative relationship exists if the correlation is -1. If there is no linear predictability between the two variables, the correlation is 0.

The Spearman correlation coefficient ρ is the covariance, or the measurement of how much two variables move together, of the two variables x and y that are the ordinal rankings of the raw scores X and Y over the distribution of all individual cases i divided by the product of their standard deviations.

As shown in Table 1, both proposed interaction terms are statistically significant. However, there is no discernible advantage predicted in using floor area over the number of rooms as the measurement of household size in the interaction term. As the number of habitable rooms and occupants are available in aggregate at the level of the Lower Layer Super Output Area in the 2001 United Kingdom Census, this paper will use the number of rooms and occupants as the measurement of physical household size instead of usable floor space and the resultant calculated number of occupants.

Table 1
Correlation of two possible interaction terms with non-heating end-use energy.

Dependent variable	Interaction term	ρ	p-value
Non-heating end-use energy[a]	Usable floor area x occupants	0.48	< 0.001
Non-heating end-use energy	Rooms x occupants	0.49	< 0.001

[a] Where non-heating end-use energy is the electricity use of homes in housing surveys that do not report electricity use as their central heating fuel nor report not having a central heating system.

4.2. Update of 1996 survey homes to 2008 electricity levels

The sample of homes in the 1996 English House Condition Survey has to be converted to the year 2008. This will enable the direct comparison of a bottom-up housing stock model of non-heating end-use energy with the small area statistics on electricity consumption in all homes which have only been collected by the Department of Energy and Climate Change since 2005. The 2008 Living Costs and Fuel Survey (LCFS) asks participants the amount of their last electricity bill and the value of any rebate from the previous billing cycle to produce a total electricity liability for either the last month or quarter (monthly direct debits have become more popular since the liberalisation of the energy market). The month or quarter of electricity use was estimated to be the month or end on the month before the date of the LCFS questionnaire.

These electricity bills are then converted into kilowatt-hours of energy using data on the average regional electricity rates for credit and direct debit ordinary electricity customers. Unfortunately, prepayment customers again are not part of the dataset as their latest electricity payment is not connected to a time period in the 2008 LCFS. Dual-rate electricity customers "Economy7" were also removed from the LCFS dataset as this tariff is designed for electricity customers with electric heating (Department of Energy and Climate Change, 2010c).

Electricity unit costs in pounds per kilowatt-hour for the year 2008 by city and payment type were accessed via the Department for Energy and Climate Change (Department of Energy and Climate Change, 2010a). The data for each city were assigned to government office region, with a mean taken of the unit costs if there were multiple cities in a region. All the cases in the 2008 LCFS then have their energy bill per month converted into kilowatt-hours per month using the unit cost for ordinary electricity for the region.

The quarterly electricity consumption data in the 1996 EHCS is also converted into monthly electricity use across the three months before the meter reading date for that quarter. Most of the participants in the fuel survey took nine consecutive quarters of meter readings over the course of just over two years. If there is repeat observation for a month in the following two years of the fuel survey, then the mean of the repeated observations is taken to represent the energy use for that month. This assumes that seasonality in electricity use for non-heating end-uses does not vary based on year-to-year differences in weather for the same month (Fig. 1).

An estimate for the electricity usage in the year 2008 can now be derived for the participants in 1996 EHCS. First, this assumes that the homes in 2008 are the same composition as they were in 1996, and that the factors that explain the differences between homes are constant over time. The monthly electricity data for the 1996 homes are standardised around a mean of zero and a standard deviation of 1 for each month of the year. The 2008 estimate is then computed as:

$$E_{2008} = \sum_{m=1}^{12} E_{m,2008} \quad (4)$$

where E_{2008} is the annual electricity use for non-heating end-uses for a domestic household and $E_{m,2008}$ is the electricity use in month m in the year 2008. Each monthly energy use estimate $E_{m,2008}$ is computed as:

$$E_{m,2008} = \overline{E}_{m,2008} + (z_{m,1996}\sigma_{m,2008}) \quad (4)$$

where $\overline{E}_{m,2008}$ is the mean electricity use for non-heating end-uses in all domestic households surveyed in the 2008 LCFS for month m, $z_{m,1996}$ is the z-score, or the number of standard deviations from the mean of the distribution of electricity use for non-heating end-uses in households surveyed in the 1996 EHCS for month m, and $\sigma_{m,2008}$ is the standard deviation of electricity use for non-heating end-uses in all domestic households surveyed in the 2008 LCFS for month m. The process of updating the 1996 EHCS data so that it represents the energy use in 2008 is summarised in Fig. 2.

The result of this process is new annual estimates of non-heating end-use energy for homes in the 1996 EHCS for the year 2008. This procedure can be repeated for subsequent expenditure surveys until the next round of housing surveys, census data, and area classifications are made available around 2014 (McIntyre, 2011).

5. Results from the linear regression model

A single-level linear regression model was run using the PROC REG procedure in SAS (S.A.S. Institute, 2011) using the modified dataset with values of non-heating end-use energy for the year 2008 in homes surveyed in the 1996 EHCS. There was a square root transformation of the dependent variable of non-heating end-use energy. Outliers and high leverage points more than two standard deviations from the mean were removed as they would have undue influence on the model resulting in 2,002 cases in the final dataset. As the dataset has been heavily modified, the decision was taken to not use the weightings from the EHCS in this regression model. This decision was confirmed by finding a similar frequency distribution of household sizes in the unweighted modified dataset to the national population.

The final regression equation for predicting non-heating end-use energy from the size of the household was:

$$\sqrt{e} = (1.06\ mn + 42.7) \quad (6)$$

where e is non-heating end-use energy measured as annual electricity consumption measured in kilowatt-hours, m is the number of rooms, and n is the number of occupants.

this formula is squared on both sides to back-transform from the square root to:

$$e = (1.06\ mn + 42.7)^2 = 1.12\ m^2n^2 + 90.5\ mn + 1820$$

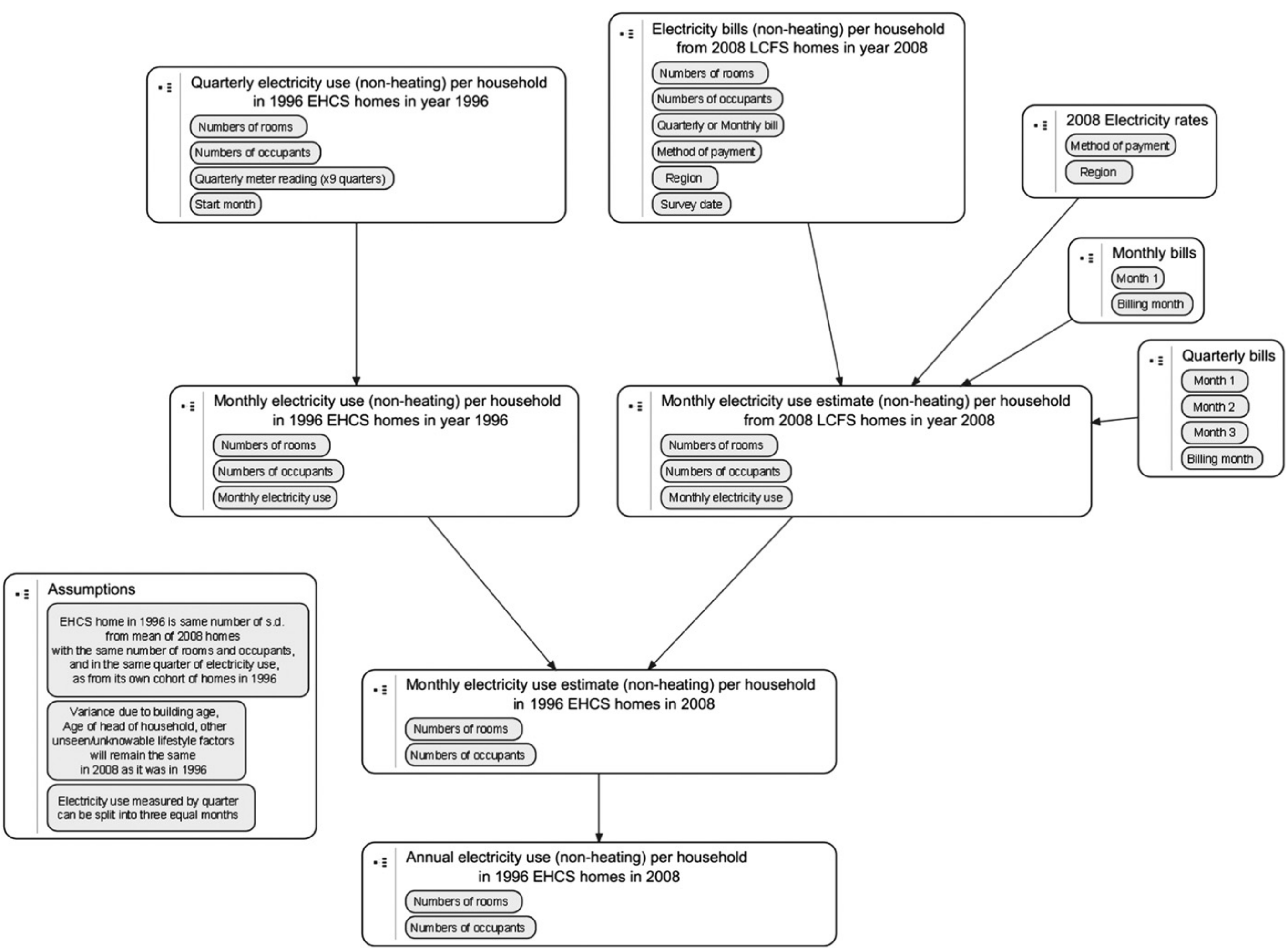

Fig. 2. Process of updating 1996 EHCS electricity data to 2008 levels.

5.1. Validation

This model was validated against census data from the 2001 UK Census for each LLSOA that had less than 5 percent of households without central heating. The number of residents and rooms are available for each LLSOA and it is possible to calculate the number of households for each group of m rooms with n occupants to build up the energy use of the entire residential sector of the LLSOA using the regression model. The model is validated both at the national level and by each LLSOA area classification supergroup.

Each Lower Layer Super Output Area has two datasets that relate to the size of the household. The first of these is a dataset that details the number of households in a census area by the number of habitable rooms that they contain (Office for National Statistics, 2005b). The second of these datasets is the number of people living in a household (Office for National Statistics, 2005a). A cross-tabulation of the two is available in the 2001 Census Aggregate Statistics Database from the Census Dissemination Unit with the following categories for household size – 1 person, 2 persons, 3 to 4 persons, and 5 or more persons, and 1 room, 2 rooms, 3 or 4 rooms, 5 or 6 rooms, and 7 or more rooms (Census Dissemination Unit, 2011). These categories were created to protect the anonymity of respondents to the census.

The number of rooms and number of people could be treated as ordinal instead of categorical variables if the spacing between some of them is calculated. In other words, in order to proceed, 3 to 4 rooms needed to be converted to a single point between 3 and 4. The averages in each of the categories is calculated in Table 2 using all cases in the unmodified, weighted 1996 English House Condition Survey. For the largest category, medians were considered instead of means because of the skew caused by larger homes. In this case, the value of the median is the lower bound because of the preponderance of these homes in the population. The final representative average was placed between the mean and the median to ensure that the size of the household was not overestimated in an LLSOA.

An estimate for electricity in 2001 can be made for each Lower Layer Super Output Area using the template in Table 3 once the two measures of household size are considered as ordinal variables. The grey area is different for every LLSOA. Table 4

The cross-tabulation was updated for 2008 using the estimate for households made by the DECC and the BRE for the total number of households in 2008 as part of fuel poverty research using data from the 2007 and 2008 English Housing Surveys (Department of Energy and Climate, Change, 2010 and BRE, 2010). The new number of households was assumed to have the same distribution of household sizes represented by the number of rooms and the number of occupants as in 2001. This enables the estimation of energy use for the year 2008 for each household to build a bottom-up model of electricity energy use of each of the 10,350 LLSOAs to compare to aggregated electricity use from that year.

Using these cross-tabulations for each census area, the total ordinary electricity use for these areas is estimated to be 1.5% away from the modelled non-heating energy consumption of all census areas, although plotting the two against each other is rather inconclusive (see Fig. 3 and 4). An examination of the differences by area classification supergroup name shows that there were some differences between the statistically modelled and actual electricity use for non-heating end-uses.

Table 2
Representative average calculated for occupants per category of household size measured by numbers of rooms and number of occupants.

Category	Weighted Mean (1996 EHCS)	Weighted Median (1996 EHCS)	Representative Average (1996 EHCS)
3 to 4 rooms	3.73	N/A	3.73
5 to 6 rooms	5.52	N/A	5.52
7 or more rooms	7.61	7.00	7.30
3 to 4 people	3.50	N/A	3.50
5 or more people	5.52	5.00	5.52

6. Conclusions

This model of non-heating end-use energy is able to be validated against actual energy use consumption. A bottom-up housing stock model was built using the product of the number of

Table 3
Template for calculating the non-heating energy-use of each LLSOA using the census data for household sizes as ordinal variables.

	As a categorical variable	1 room	2 rooms	3 to 4 rooms	5 to 6 rooms	7 or more rooms
As a categorical variable	**As an ordinal variable**	1 room	2 rooms	3.73 rooms	5.52 rooms	7.30 rooms
1 person	1 occupant					
2 people	2 occupants					
3 to 4 people	3.50 occupants					
5 or more people	5.52 occupants					

Table 4
Validation of the model against aggregated electricity use by area classification.

ONS 2001 Area Classification Supergroup Name	Number of LLSOAs <5% without central heating	Model estimate of electricity use in 2008 (kWh)	Actual electricity use recorded in 2008 (kWh)	Difference between estimate and actual use (kWh)	Difference between estimate and actual use (percent)
Countryside	673	1,552,265,666	1,604,585,375	−52,319,710	−3.3
Disadvantaged Urban Communities	1,933	3,912,787,843	3,736,427,124	38,540,686	1.0
Miscellaneous built up areas	3,068	6,690,162,789	6,356,415,994	329,603,350	5.2
Multicultural City Life	1,479	2,942,719,802	2,836,110,254	105,202,641	3.7
Professional City Life	827	1,837,506,949	1,909,557,762	−75,538,266	−4.0
Urban Fringe	1,149	2,639,688,748	2,631,764,847	−56,820,327	−2.2
White Collar Urban	1,221	2,653,822,364	2,595,908,579	34,494,415	1.3
TOTAL	**10,350**	**22,228,954,160**	**21,670,769,936**	**323,162,789**	**1.5**

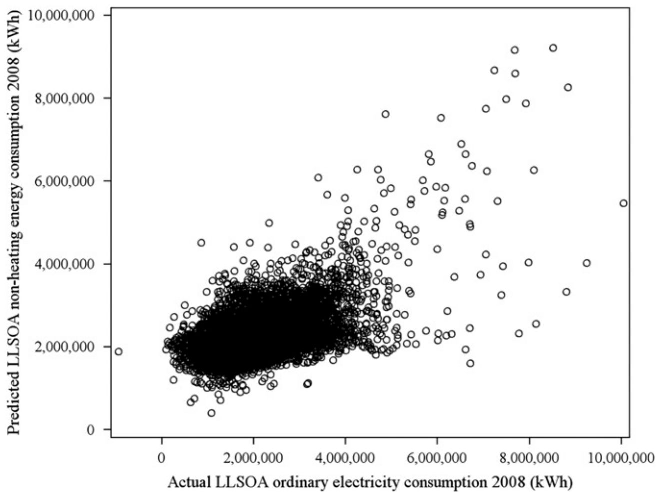

Fig. 3. Scatter plot of predicted non-heating consumption against actual ordinary electricity consumption by LLSOA.

occupants and the number of rooms to predict non-heating end-use energy, which facilitated the housing stock simulation of total energy use to be validated against actual energy consumption. The measurement of dwelling size as the number of rooms enabled the model to be validated against a specific housing stock with both known total electricity consumption data and minimal electricity use due to heating end-uses. The total modelled and actual energy use was found to be very close, though different types of areas performed appreciably differently against actual energy use. This simple, flexible housing stock model for non-heating end-use energy has the potential to both simulate consumption of the housing stock and adapt quickly to new data as it is made available.

There may be an opportunity in the future to use this simple interaction term with new data to refine this housing stock model because its component terms are widely collected, and they decrease privacy concerns that limit access to data. The current housing stock model for non-heating end-use energy, the Cambridge Housing Model, measures household size in usable floorspace and number of occupants. The number of habitable rooms and occupants, however, are measures that have been kept in the United Kingdom Census, regional planning, and building control for many decades (Martin, 2001, Abercrombie and Forshaw, 1943). The use of the number of rooms and the number of occupants in a household can also encourage more data to become available from households whilst protecting their privacy. Homes could be surveyed using a simple questionnaire for the resident to count habitable rooms instead of sending in a surveyor to measure usable floorspace. Using these measures of household size can make it easier in the current implementation of smart meters in the UK to opt-in residents to share electricity meter data to help refine domestic energy stock models of non-heating end-uses without compromising privacy – the current proposal from the UK government is to opt-out all residents from sharing most data outside of trials (Department of Energy and Climate Change, 2012).

This model will also be adaptable in future years as a yearly update using the methodology set out in this paper, as the 2011 Census and future energy follow-up surveys of the EHS available in 2014 will be derived from the same core set of questions as the LCFS as part of the Integrated Housing Survey introduced in 2008 (Department for Communities and Local Government, 2011). There have been long gaps between surveys that include meter data, and this method future-proofs the model from an irregular programme of housing surveys that include electricity and gas meter data.[2] This may impact on the scope and frequency of surveys commissioned by governments. It will also impact on the assessment of how the domestic sector is performing against five-year carbon budgets proposed by the UK government (Department of Energy and Climate Change, 2011b).

There was a difference in accuracy in this stock model between different types of area classifications. In general, the model underestimates the energy demands of rural and suburban areas and conversely overestimates urban areas. There are numerous possible causes, including travel to work and leisure, online and offline sociability, and device ownership that can be explored further (Energy Saving Trust, 2011, U.S. Energy Information Administration, 2011). If the influence of area type can be measured and is significant, this may aid the targeting of energy efficiency measures to optimise the effectiveness of the UK Green Deal programme.

The model proposed in this paper has the potential to be a verifiable and adaptable housing stock model for non-heating end-use energy. It was found to be a good predictor of non-heating consumption at the national level and for groups of similar areas, but not for many individual LLSOAs. It can be updated easily with new data, it can alleviate concerns over sharing data from smart meters once they become ubiquitous, and could even help effectively target energy efficiency measures for electricity as part of the UK Green Deal programme. Future modelling of these end-uses should be able to use much more diverse sources of data to simulate the housing stock at the national and sub-national scales.

Acknowledgements

I wish to thank the reviewers for their detailed comments to help this paper reach this audience. I also wish to thank the UK Engineering and Physical Sciences Research Council for funding this work, Robert Lowe and Philip Steadman of the UCL Energy Institute for their help reviewing this research, and Terry McIntyre of the UK Department of Communities and Local Government for matching location data to the 1996 EHCS Energy sub-sample.

Appendix A. This appendix details some of the statistical outputs in the development of the model using regression analysis using SAS.

The first set of diagnostics when running linear regression is an analysis-of-variance table:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	130896	130896	1503.90	< .0001
Error	2000	174076	87.03796		
Corrected Total	2001	304972			

The analysis of variance table contains the following (U.C.L.A., 2011):

- Source – the variance of the dependent variable (energy) is broken down into the categories of Model, Error, and Corrected

[2] If heating within the housing stock moves increasingly towards electricity as predicted by the *40% House* report by the Oxford Environmental Change Institute (Boardman, 2005) to allow for more renewable energy sources, this modelling approach will become increasingly difficult to apply.

Total. The analysis of variance partitions the total variance into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Error).
- DF - These are the degrees of freedom associated with the sources of the variance. The total variance has N-1 degrees of freedom (in this model $N=2002$). The model degrees of freedom corresponds to the number of coefficients estimated minus 1. Including the intercept, there are 2 coefficients, so the model has 2-1=1 degrees of freedom. The Error degrees of freedom is the DF total minus the DF model, 2001–1=2000.
- Sum of Squares - These are the Sum of Squares associated with the three sources of variance, Total, Model and Error.
- Mean Square - These are the Mean Squares, the Sum of Squares divided by their respective DF.
- F Value - This is the F-statistic. It is the Mean Square Model (130896) divided by the Mean Square Error (87.04), yielding $F=1503.90$.
- Pr > F - This is the p-value associated with the above F-statistic. It is used in testing the null hypothesis that all of the model coefficients are 0 and have no predictive power. As the p-value is less than 0.001, then the F-statistic, and therefore the interaction term has predictive power on the electricity use of a household.

The overall model fit is assessed by the following statistics:

Root MSE	9.32941	**R-Square**	0.4292
Dependent Mean	56.11073	**Adj R-Sq**	0.4289
Coeff Var	16.62679		

- Root MSE - Root MSE is the standard deviation of the error term, and is the square root of the Mean Square Error.
- Dependent Mean - This is the mean of the dependent variable (the square root transformation of the annual non-heating end-use energy consumption of a household, represented by the electricity use of a household where it does not use electricity for heating end-uses).
- Coeff Var - This is the coefficient of variation, which is a unitless measure of variation in the data. It is the root MSE divided by the mean of the dependent variable, multiplied by 100: (100*(9.33/56.11)=16.63.
- R-Square - R-Squared is the proportion of variance in the dependent variable (non-heating end-use energy) which can be explained by the independent variables (size of household measured by the number of rooms and number of occupants). This is an overall measure of the strength of association of all the independent variables with the dependent variable. As R-Squared is only 0.43, this is a first alert that there may be further outliers and high leverage points based on the residuals as opposed to the raw score.
- Adj R-Sq - This is an adjustment of the R-squared that penalizes the addition of extraneous additional predictors to the model. As the independent variables are combined into an interaction term, this statistic is not necessary.

Thirdly, the parameter estimates in PROC REG form the algorithm that predicts the dependent variable from the independent variables:

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	42.67870	0.40428	105.57	<.0001
rooms_hsize96	Interaction term of the number of rooms and number of occupants (1996)	1	1.06322	0.02742	38.78	<.0001

The diagnostic statistics for this run of the linear regression model is as follows:

- Variable - This column shows the independent variables (in this case, the interaction term that represents the two independent variables of the number of rooms and numbers of occupants). The first refers to the model intercept or the height of the regression line when it crosses the Y axis. In other words, this is the predicted value of non-heating end-use energy when all other variables are 0.
- Label - This column gives the label for the variable.
- DF - This column give the degrees of freedom associated with each independent variable. All continuous variables, such as the interaction term, have one degree of freedom.
- Parameter Estimates - These are the values for the regression equation for predicting the dependent variable from the independent variable. For the interaction term, this is its regression coefficient with a value of 1.06.
- Standard Error - of the coefficient.
- t Value - These are the t-statistics used in testing whether the coefficient generated as a parameter estimate is significantly different from zero.
- Pr > |t|- The *p*-value used in testing the null hypothesis that the coefficient (parameter) is 0. Using an alpha, or confidence of 95%, of 0.05,the coefficient for the interaction term is significantly different from zero because its p-value is <0.001, which is smaller than 0.05, and the intercept is significantly different from zero.

Reference

Abercrombie, P.S., Forshaw, J.H., 1943. County of London Plan prepared for the London County Council. MacMillan and Co, London.

Anderson, B.R. 2002. *BREDEM-12: model description,* 2001 *update,* Garston, Watford, UK, BRE.

BRE 2010. The Government's Standard Assessment Procedure for Energy Rating of Dwellings: incorporating RdSAP 2009. 2009 ed. Garston: BRE Group.

Cambridge Architectural Research Ltd. & DECC. 2011. *A Guide to the Cambridge Housing Model* [Online]. Available: 〈https://www.yousendit.com/sharedFolder?phi_action=app/orchestrateSharedFolder&id=yoztP438Ra-fjVwTVEhtK1nx6UGwpjDleJ63mmD0GU8〉 [Accessed 10 May 2012].

Census Dissemination Unit. 2011. *Casweb* [Online]. Available: 〈http://casweb.mimas.ac.uk/〉 [Accessed 11 March 2011].

Conover, W.J., 1999. Practical nonparametric statistics. John Wiley, New York; Chichester.

Cozby, P.C. 1997. *Methods in Behavioral Research,* Mountain View, Calif., USA, Mayfield.

Department for Communities and Local Government. 2010. *English House Condition Survey (EHCS)* [Online]. London: Department for Communities and Local Government. Available: 〈http://www.communities.gov.uk/housing/housingresearch/housingsurveys/englishhousecondition/〉 [Accessed 9 March 2010].

Department for Communities and Local Government. 2011. *English Housing Survey (EHS)* [Online]. London: Department for Communities and Local Government. Available: 〈http://www.communities.gov.uk/housing/housingresearch/housingsurveys/englishhousingsurvey/〉 [Accessed 15 March 2011].

Department of Energy and Climate Change 2009. *The UK low carbon transition plan: national strategy for climate and energy,* London, Stationery Office.

Department of Energy and Climate Change 2010a. Average annual domestic electricity bills for selected towns and cities in the UK and average unit costs. *In:* QEP223.XLS (ed.). London: Department of Energy and Climate Change.

Department of Energy and Climate Change 2010b. Guidance note for the DECC MLSOA/IGZ and LLSOA electricity and gas consumption data London: Department of Energy and Climate Change.

Department of Energy and Climate Change. 2010c. *Variation in Tariff Types and Energy Bills* [Online]. Available: 〈http://www.decc.gov.uk/assets/decc/statistics/publications/trends/articles_issue/1_20100324125048_e_@@_variationtarifftypes.pdf〉 [Accessed 10 May 2012].

Department of Energy and Climate Change. 2011a. *2010 Provisional UK Figures* [Online]. Available: 〈http://www.decc.gov.uk/en/content/cms/statistics/climate_stats/gg_emissions/uk_emissions/2010_prov/2010_prov.aspx〉 [Accessed 16 November 2011].

Department of Energy and Climate Change. 2011b. *The Carbon Plan* [Online]. London: Department of Energy and Climate Change. Available: 〈http://www.decc.gov.uk/Media/viewfile.ashx?FilePath=What%20we%20do/A%20low%20car bon%20UK/1358-the-carbon-plan.pdf&filetype=4&minwidth=true〉 [Accessed 10 May 2011].

Department of Energy and Climate Change. 2011c. *Energy consumption in the UK: Domestic data tables 2011 update* [Online]. Available: 〈http://www.decc.gov.uk/assets/decc/statistics/publications/ecuk/269-ecuk-domestic-2010.xls〉 [Accessed 21 November 2011].

Department of Energy and Climate Change 2011d. Great Britain's housing energy fact file. London: Department of Energy and Climate Change.

Department of Energy and Climate Change. 2011e. *National Energy Efficiency Data-Framework: Report on the development of the data framework and initial analysis* [Online]. Available: 〈http://www.decc.gov.uk/assets/decc/11/stats/energy/energy-efficiency/2078-need-data-framework-report.pdf〉 [Accessed 11 May 2012].

Department of Energy and Climate Change. 2012. *Smart Metering Implementation Programme: Data access and privacy consultation draft* [Online]. Available: 〈http://www.decc.gov.uk/assets/decc/11/consultation/smart-metering-imp-prog/4933-data-access-privacy-con-doc-smart-meter.pdf〉 [Accessed 12 April 2012].

Department of Energy and Climate Change and BRE. 2010. *Fuel Poverty Sub-Regional Statistics* [Online]. Available: 〈http://www.decc.gov.uk/en/content/cms/statistics/fuelpov_stats/regional/regional.aspx〉 [Accessed 25 July 2011].

Department of the Environment Transport and the Regions 2000. English House Condition Survey 1996 User Guide. London: Department of the Environment Transport and the Regions.

Energy Saving Trust, 2008. Homes Energy Efficiency Database (HEED): An overview. Energy Saving Trust, London.

Energy Saving Trust, 2011. The elephant in the living room: How our appliances and gadgets are trampling the green dream. Energy Saving Trust, London.

Henderson, J., 2009. Review of auxiliary energy use and the internal heat gains assumptions in SAP. Garston. BRE Group, Watford, UK.

International Energy Agency 2009. *Gadgets and Gigawatts - Policies for Energy Efficient Electronics,* Paris.

Longley, P.A., Tobon, C., 2004. Spatial dependence and heterogeneity in patterns of hardship: An intra-urban analysis. Annals of the Association of American Geographers 94, 503–519.

Market Transformation Programme 2008. Policy Analysis and Projections 2006/09.

Martin, D., 2001. Geography for the 2001 Census in England and Wales, Office of National Statistics.

McIntyre, T. 5 April 2011. *RE: Energy Follow Up Survey*.

Myers, J.L., Well, A., Lorch, R.F., 2010. Research design and statistical analysis. Routledge Academic, London.

Office for National Statistics 2004. *Census 2001: key statistics for urban areas in England and Wales: laid before Parliament pursuant to section 4(1) Census Act 1920,* London, TSO.

Office for National Statistics, 2005. Number of People Living in Households (UV51). Office for National Statistics, London.

Office for National Statistics, 2005. Number of Rooms (UV57). Department for Communities and Local Government, London.

Office for National Statistics, 2008. Guidance notes for the 2001 Area Classification of Super Output Areas and Data Zones. Office for National Statistics, London.

Office for National Statistics, 2010. Social Trends, 40 ed. Office for National Statistics, London.

Office for National Statistics and Department for Environment Food and Rural Affairs 2010. Living Costs and Food Survey, 2008. Colchester, Essex: UK Data Archive.

S.A.S. Institute, 2011. Base SAS(R) 9.2 Procedures Guide: Statistical Procedures. SAS Institute, Cary, North Carolina, USA.

Shorrock, L.D., Utley, J.I., 2008. Domestic energy fact file. Garston, BREbookshop 2008.

Sørensen, B., 2011. Renewable energy: physics, engineering, environmental impacts, economics & planning. Academic, London.

Steadman, J. P. 2 May 2012. *RE: Meeting over access to meter data at address level.*

U.C.L.A. 2011. *SAS Annotated Output: Regression Analysis* [Online]. Los Angeles: UCLA Academic Technology Services. Available: 〈http://www.ats.ucla.edu/stat/sas/output/reg.htm〉 [Accessed 27 June 2011].

U.S. Energy Information Administration 2011. 2009 Residential Energy Consumption Survey. *In:* 〈http://www.eia.doe.gov/consumption/residential/reports/images/mainheatingsrc.xls〉 (ed.). Washington, D.C.: U.S. Energy Information Administration.