2016

Statistical models to infer gas end-use efficiency in individual dwellings using smart metered data

Ronan Oliver
Technological University Dublin, ronan.oliver@tudublin.ie

Aidan Duffy
Technological University Dublin, aidan.duffy@tudublin.ie

Ian Kilgallon
Gas Networks Ireland

Recommended Citation

Ronan Oliver, Aidan Duffy, Ian Kilgallon, Statistical models to infer gas end-use efficiency in individual dwellings using smart metered data, Sustainable Cities and Society, Volume 23, 2016, Pages 1-10, ISSN 2210-6707, DOI: 10.1016/j.scs.2016.01.009.

Funder: Technological University Dublin and Gas Networks Ireland

Statistical models to infer gas end-use efficiency in individual dwellings using smart metered data

Ronan Oliver[a], Aidan Duffy[a,*], Ian Kilgallon[b]

[a] Dublin Energy Lab, Dublin Institute of Technology, Ireland

[b] Gas Networks Ireland, Cork, Ireland

[*] Corresponding author: tel.: +353 (0)1 402 3940, e-mail: aidan.duffy@dit.ie

Abstract

Residential buildings can significantly contribute to the European Union's 2020 efficiency energy targets. For this reason, energy distributors and suppliers are required to provide assistance to householders to reduce energy end-use. This paper develops statistical modelling methods that can be used by suppliers to infer the gas fuel efficiency of buildings in their residential portfolio, in order to deliver improved energy management services to consumers. The study begins by estimating individual statistical building energy models for a sample of consumers and presents the resulting distribution of independent parameters. These parameter distributions are then characterised by regression models using descriptive household data that is generally known by the consumer and can be easily gathered by the energy supply company. These models are then used to compare the inferred energy end-use efficiency of the household (cooking, hot-water and space heating) compared to similar dwellings. Buildings with higher-than-expected gas consumption can be targeted for energy efficiency programmes.

Keywords: Energy suppliers, residential gas consumption, energy efficiency, smart meters, degree days.

1 Introduction

In the European Union (EU), residential buildings are responsible for 26% of annual energy consumption and 37% of this energy is consumed as gas (European Commission, 2014). Domestic gas consumers can therefore make a significant contribution to the EU's 2020 targets of: 1) a 20% reduction in greenhouse gas emissions from 1990 levels; 2) a 20% increase in energy from renewable resources; and 3) a 20% improvement in energy efficiency (European Commission, 2009); and thus help to meet the objective of decarbonising energy end-use in Europe.

To help realise such improvements and a reduction in fossil fuel imports, the EU has mandated that smart meters are made available to residential gas consumers in each member state, except those states where an adverse cost benefit has been established (Official Journal of the European Union, 2009). This has resulted in the on-going installation of these meters in many countries across the EU. These include the United Kingdom (UK) where 22 million are planned for installation by 2019 and France, where 11 million could be in place before 2020 (Hierzinger et al., 2013). In such countries, consumers will have access to high resolution time-of-use consumption data. Sampling intervals for smart meters are typically hourly (or less) compared to monthly (or more) for traditional manually-read meters. Access to such high-frequency data will enable consumers to manage their gas consumption more effectively and identify readily achievable energy savings.

The EU has also recommended that energy distributors and/or suppliers provide assistance to consumers to help reduce their energy consumption. In this regard, each EU member state can implement an 'Energy Efficiency Obligation Scheme' to ensure that suppliers achieve energy savings each year from 2014 to 2020 that are at least equivalent to 1.5% of their consumers' average annual energy consumption between 2010 and 2012 (Official Journal of the European Union, 2012). However, because gas is used in the home for space heating, hot water production and cooking purposes, and since this consumption is dependent on factors such as dwelling size and occupancy, it is difficult for suppliers to identify appropriate energy efficiency measures for individual consumers based on their gas consumption data alone.

Abbreviations

A	area (m^2)
C	gas consumption (kWh)
F	fuel consumption (kWh)
HLC	overall heat loss coefficient (kW/ °C)
HDD	heating degree day (°C·day)
MLS	multinomial logistic regression
N	number of air changes per hour (1/h)
NLS	non-linear least squares
Q	heat (kW)
T	temperature (°C)
U	U-value ($W/m^2 \cdot °C$)
V	volume of the heated space (m^3)

Subscripts

B	base
D	day
G	gain
IN	indoor
MP	metered period
O	outdoor
SP	set-point

Greek symbols

ε	model error
η	heating system efficiency (%)

This study therefore develops and demonstrates a methodology that can be used to compare a household's gas consumption end-uses to those of other households with similar characteristics. Smart-metered gas consumption and household data (e.g. number of bedrooms and dwelling type) are used to develop statistical models which estimate energy end-use (e.g. space heating, cooking and hot water) for individual dwellings and compare these to a benchmark for dwellings with similar characteristics. This information allows energy suppliers to screen their customers and target appropriate energy efficiency measures at the most appropriate households. The methodology is demonstrated using daily gas consumption and household data collected for a sample of over 500 residential dwellings in Ireland.

The paper is organised as follows. It begins with a section on the current methods used to benchmark building energy efficiency using metered energy data. Because heating degree days (HDDs) are widely used in these methods and since they are used in the approach later described in this paper, a brief review of HDD theory is then given. The Methodology section describes data sources and two statistical inferential models. The first of these, based on non-linear least squares (NLS) estimates dwelling gas consumption based on parameters which we relate to gas end-uses. The second, based on multinomial logistic regression (MLR), estimates the relationships between these end uses and household characteristics. This latter model is then used to compare the relative energy end-use performances of households of similar characteristics. Following this Methodology section, a Results and Discussion section presents the statistical models, and by way of example, assesses the relative energy end-use efficiency of a sample of consumers with similar household characteristics. Conclusions and Recommendations are then presented.

2 Benchmarking

Benchmarking is the process of comparing an individual performance against a relevant standard, or benchmark. A wide variety of benchmarking methods have been developed for assessing household energy efficiencies using metered energy consumption data and, in almost all cases, these are based on HDDs. The HDD variable is a parameter based on outdoor temperature data that is traditionally used to estimate building heating system fuel consumption; the approach is described in detail in the next section.

Many building energy efficiency benchmarking tools have been developed that apply HDDs. For example, the US Environmental Protection Agency (US-EPA) has developed an Energy Star Score system for a range of commercial buildings that applies a regression based benchmarking tool (Energy Star, 2014a). The first step in this scoring system calculates an energy efficiency ratio for a building by dividing its annual energy use intensity (both electricity and gas) by that predicted by a regression model for the building type (Energy Star, 2014a). For example, the regression model applied for multifamily housing (or apartment) buildings has been fitted using a reference dataset of such buildings and is based on the number of dwellings per $1000ft^2$, the number of bedrooms per dwelling, the total HDDs and cooling degree days for the year, and the number of levels in each building (Energy Star, 2014b). The probability or percentile of the building's energy efficiency ratio is then found using a lookup table developed using energy efficiency ratios for the reference dataset (Energy Star, 2014a). The Energy Star Score for the building is simply 100 minus this percentile value. For example, a building with an Energy Star Score of 75 is bettered by only 25% of the reference dataset.

Home Energy Yardstick is an online tool that has been developed as part of the US-EPA's Energy Star program (Energy Star, 2015a). This tool benchmarks residential building energy efficiency using a 1 to 10 scoring system, where a score of 10 represents a home with the best energy efficiency level (Energy Star, 2015a, 2015b). This score is based on a statistical method and requires users to provide utility bill consumption data for electricity and gas, and their building's location, floor area and number of occupants (Energy Star, 2015a, 2015b). Energy suppliers in the US are encouraged to host this tool on their own web sites (Energy Star, 2015c).

In Europe, a Display Energy Certificate system is applied to large public buildings. These certificates are also based on metered energy consumption and building floor area and are used to present a building's annual energy use intensity ($kWh/m^2/year$) on an A1 to G scale, where an E1 rating corresponds to a typical building in the relevant building class (SEAI, 2015). These energy intensities are based on building floor area. Such normalised energy consumption parameters are a very common way of benchmarking building energy efficiency (Wang, Yan, & Xiao, 2012).

Each of the above benchmarking tools is based on energy intensity parameters normalised by building floor area, which presupposes that floor area data are readily available. However, it has been observed that many householders are unable to provide their building's floor area when surveyed – 75% in the case of a previous Irish housing quality survey (Watson & Williams, 2003) and 59% in the case of the smart metering survey used here. Accurate area data would therefore be difficult to collect for an energy supply company. Moreover, many variables other than floor area contribute to household energy use; these include occupancy patterns, no. of occupants and dwelling type (detached, semi-detached, etc.). These, too, should be considered in a comprehensive gas consumption benchmarking method. Therefore, instead of using an area-related energy intensity parameter, this study develops an alternative regression-based benchmarking method based on multiple household variables which are known to householders and can be easily obtained through phone interview.

3 Heating Degree Days

Heating degree days form the basis of almost all energy efficiency benchmarking models. They are based on the concept that the instantaneous heat demand for a building may be estimated as the product of the building's overall heat loss coefficient (*HLC*) and the temperature differential between the heated space and the surrounding environment. HDDs estimate the integral of this temperature differential over time, so that the fuel consumption of the building's heating system, is approximated by the sum (CIBSE, 2006):

$$F = HLC\left(\sum_{i=1}^{n} HDD_i\right)\left(\frac{24}{\eta}\right) \quad (1)$$

where: *F* is fuel consumption (kWh); *n* is the number of days in the relevant time period; *HDD* is the heating degree day parameter (°C·day); 24 is a conversion factor to kWh units; η is a conversion factor to fuel consumption units that is given by the efficiency of the building's heating system (%); and *HLC* (kW/°C) is given by (CIBSE, 2006):

$$HLC = \left(\sum UA + 0.33NV\right)/1000 \quad (2)$$

where: ΣUA is the building's fabric loss coefficient (W/°C), given by the sum of the products of *U*-values (W/m^2·°C) and areas A (m^2) for each of the external building fabric elements; 0.33*NV* is the building's air-infiltration coefficient (W/°C), given by N the number of air changes per hour for the building (1/h) and V the volume of the heated space (m^3); and 0.33 and 1000 are conversion factors required to ensure that the units of the HLC are kW/°C.

While there are alternative methods to calculate a HDD that depend on the resolution or format of the available temperature data (CIBSE, 2006), this paper will apply the following internationally accepted function (EN ISO 15927-6, 2007):

$$HDD = max(0; T_B - \overline{T}_O) \quad (3)$$

where: $\overline{T}_O$ is the average outdoor temperature for the day (°C); and T_B is the building's base temperature parameter (°C).

This base temperature parameter is used to estimate the average internal temperature in the building during the heating season, less the equivalent temperature effect of incidental heat gains, as follows (CIBSE, 2006):

$$T_B = \overline{T}_{IN} - T_G \quad (4)$$

where: $\overline{T}_{IN}$ is the building's average indoor temperature (°C), and T_G is the equivalent temperature effect of incidental heat gains (°C), given by (CIBSE, 2006):

$$T_G = Q_G / HLC \quad (5)$$

where: Q_G is the useful heat gain to the heated space (kW).

Based on these concepts, the HDD variable can be used to model monthly (or bi-monthly) gas meter readings by employing the following regression model:

$$C_{MP} = b_o Days_{MP} + b_1 \sum HDD_{MP} + \varepsilon_{MP} \quad (6)$$

where: C_{MP} is the building's gas consumption (kWh) for each metering period (MP); b_0 is an estimate of the building's daily base or weather-independent gas consumption (kWh/day); $Days_{MP}$ is the number of days in each metering period; b_1 is an estimate of the building's gas consumption response to HDDs (kWh/°C·day); $\sum HDD_{MP}$ is the sum of HDDs in each metering period; and ε_{MP} is the model error for each metering period.

Such HDD regression models are generally fitted using HDD data published by the local meteorological service that is calculated using the traditional base temperature adopted for that region – for example, 15.5°C in the UK (CIBSE, 2006) and Ireland. However, if instead outdoor temperature data are applied the true (or individualised) base temperature for the building can be estimated, and a more representative building energy model will result. Many calls have been made in this regard for the adoption of building-specific base temperatures (CIBSE, 2006).

Traditionally, the true base temperature for a building has been estimated using alternative 'trial and error' techniques for monthly or daily metered fuel consumption data (CIBSE, 2006). For monthly data, a quadratic HDD regression model is applied that estimates a building's base temperature by the value which yields a zero squared-HDD coefficient (Day et al., 2003). For daily metered data, however, a building's base temperature is estimated either by: 1) visually identifying the point of inflection in a scatter plot of fuel consumption vs. temperature; or 2) identifying the upper temperature limit in the data that yields the maximum coefficient of determination (R^2 value) for a linear model of fuel consumption based on the lower temperatures (CIBSE, 2006).

However, monthly gas meter readings are generally only applied to large commercial (or high consumption) consumers. For example, in the United Kingdom, monthly meter readings are only recorded for consumers with an annual gas requirement greater than 293,000 kWh (or 10,000 therms, an order of magnitude greater than typical domestic consumption), and annually for consumers below this threshold (Joint Office of Gas Transporters, 2015). In France, bi-annual meter readings are recorded for consumers (including households) with an annual requirement less than 300,000 kWh (Commission de Régulation de L'Énergie, 2012). While, such data limitations have made it difficult for energy suppliers to apply the above HDD modelling methods to domestic consumers, the increasing availability of domestic smart-metering data means that they can now be applied to the residential sector.

4 Methodology

This section first describes the data used in this work. A non-linear least squares (NLS) statistical inferential model which uses HDDs as well as other independent parameters to estimate daily dwelling gas consumption is then described. We then explain how the NLS estimator parameters can be used to infer gas consumption related to cooking/hot water, envelope heat losses and heating controls. The section concludes by describing a multinomial logistic regression modelling method which is used to relate the inferred end-use consumptions to household characteristics (both physical and occupancy-related), and how this method can be used to compare the relative end-use efficiencies of individual households of similar characteristics. The benefit of this approach is that it estimates the relative end-use fuel consumption for each customer compared to other similar households, rather than comparing buildings based on floor area only, which takes no account of dissimilar household characteristics.

4.1 Data

This study is based on smart-metered gas consumption data and household survey data, recorded between December 2009 and May 2011, for a sample of over 500 Irish dwellings which participated in gas smart-metering trials (Commission for Energy Regulation, 2011). Participants were selected to be representative of the Irish gas consumer population, and were located in either in the largest city, Dublin (64%), or in urban centres no more than approximately 250km from Dublin. Due to anonymity requirements, the locations of households were not known. For this reason, and given the small geographic spread of participants, the models estimated in this study were fitted using outdoor temperature data for the most representative single location, Dublin Airport.

The household survey data were collected using a telephone questionnaire survey and are listed later in Table 2 under ‘Survey Data Collected’. This survey also collected data on building floor area, wall insulation and building occupancy. However, it was found that a significant proportion of consumers did not provide information for several variables. For example, 59% did not know their building’s floor area, 27% did not know whether or not wall insulation was present in their building, and 26% did not state whether or not their building was occupied by adults during the day. Therefore these explanatory factors were not used in the development of the logistic regression models, as their inclusion would significantly limit the usable sample size. Data relating to the presence of attic insulation were not used for similar reasons.

4.2 Non-Linear Least Squares Regression Modelling

Because daily gas consumption data will soon be widely available for domestic consumers from smart meters this study has developed a more direct method to estimate the b_0, b_1 and T_B parameters of the HDD regression model than the methods reviewed in the literature. Such daily data allows the HDD regression model in Eq. (6) to reduce to the following form:

$$C_D = b_o + b_1 HDD + \varepsilon_D \tag{7}$$

where: C_D is daily gas consumption (kWh), ε_D is the model error for each day (D), b_0 and b_1 are as before in Eq. (6) but can now be referred to as the intercept and slope parameters of the HDD regression model respectively.

By substituting Eq. (3) for the HDD parameter this model can be re-expressed as follows:

$$C_D = b_o + b_1 \max(0; T_B - \bar{T}_O) + \varepsilon_D \tag{8}$$

This expression permits the use of the non-linear squares model fitting method described later in this section. The resulting model parameters can be interpreted as follows.

4.2.1 Intercept parameter (b_0)

The b_0 parameter is the building's daily base gas consumption, and for residential consumers this is typically used for hot water and cooking purposes. Therefore, the b_0 parameter may be used to identify buildings in need of a hot water heating system upgrade or a reduction in hot water set-point temperature (Raffio et al., 2007).

4.2.2 Slope parameter (b_1)

The b_1 parameter is related to the building's heat loss coefficient and heating system efficiency as follows (CIBSE, 2006):

$$b_1 \approx HLC\left(\frac{24}{\eta}\right) \quad (9)$$

and may be used to identify buildings in need of building fabric or heating system upgrades (Raffio et al., 2007).

4.2.3 Base temperature parameter (T_B)

The T_B parameter is related to the average indoor temperature and useful heat gain in the building, as shown in Eq. (4) and Eq. (5). This average temperature is in turn related to the building's heating system set-point temperature, as follows (CIBSE, 2006):

$$\bar{T}_{IN} \approx \frac{T_{SP}(On)+\sum_{h}^{(24-On)} T_{IN,h}}{24} \quad (10)$$

where: T_{SP} is the heating system's set-point temperature (°C) – which is assumed to be representative of the building's indoor temperature during heating periods; *On* is the number of heating system operating hours each day; and $T_{IN,h}$ is the indoor temperature at hour *h* in the day when the heating system is off.

The T_B parameter may be used to assess a consumer's thermal comfort requirement, as buildings with high base temperatures must respond to more HDDs during each heating season than those with lower base temperatures. This may either be the result of increased set-point temperatures and heating system operating hours or poor heat gain retention in the building. Such buildings are targets for behavioural programmes or improved heating system control systems, for example programmable thermostats (Raffio et al., 2007).

4.2.4 Model Fitting

The parameters of the non-linear regression model in Eq. (8) are estimated for each consumer in the sample using the Levenberg-Marquardt non-linear least-squares algorithm, available in the statistical computing software, R (R Core Team, 2013). This local NLS modelling method was used in preference to a global NLS algorithm that is also available in R, as it is more robust to stochastic changes in the modelled series.

Each NLS model is estimated using daily gas consumption data for the final year in the smart-meter trial (31st May 2010 - 30th May 2011), as only a single heating season is required to estimate the base temperature parameter. To help convergence to a local NLS solution, starting values and limits have been stipulated for each parameter as shown in Table 1. Alternative starting values were trialled to assess the sensitivity of the models, but this resulted in a slight decrease in the number of successfully converged models and no observable change to the intercept, slope and base temperature parameter distributions presented in Figs. 2 - 4.

Table 1 Parameter starting values and limits

Parameter	Starting Value	Lower Limit	Upper Limit
Intercept (b_0)	0	0	None
Slope (b_1)	0	0	None
Base temperature (T_B)	15.5	5	25

4.3 Multinomial Logistic Regression Modelling

MLR modelling is a well-known method used to model categorical variables (Field, 2013). It is used in this study to categorise the intercept, slope and base temperature parameter estimates of individual household NLS models as 'low', 'medium' or 'high' relative to other similar households. This allows consumers with higher-than-expected NLS modelling parameters to be identified so that relevant energy saving advice can be tailored to their needs. The approach is demonstrated in the Results and Discussion.

Three MLR models have been developed for this purpose. These are initially used to characterise each of the intercept, slope and base temperature parameter distributions resulting from the NLS models. They are fitted to low, medium and high categories of these distributions using the household survey data collected for the consumer sample. The relationship between the intercept, slope and base temperature parameters of the NLS regression model and this survey data are described in Table 2.

Each of the resulting MLR models comprise low and high sub-models based on a medium reference category. The most frequently occurring categorical explanatory variable (listed under 'Categories' in Table 2) has been specified as a reference category. Small sample categories of some explanatory variables have been merged into alternative categories or removed from the logistic regression models were appropriate. Each of these models is fitted using the 'multinom' algorithm in R.

Table 2 Survey data collected and their relationship to the NLS regression model

Survey Data Collected		**Relationship to the NLS Regression Model**	
Variable	Categories	Parameter	Description and likely effect on the associated parameter
No. of adults	1, 2, 3, 4, 5 or ≥ 6.	b_0	Building occupancy positively affects hot water and cooking gas requirements.
No. of children	0, 1, 2, 3, 4, 5 or ≥ 6.		Children (less than 15 years old) are likely to consume less hot water than adults.
Hot water system	Timed gas fuelled, untimed gas fuelled or alternatively fuelled system.[(a)]		Alternatively fuelled hot water systems should result in reduced base consumption, while timed gas fuelled systems should consume less gas than untimed systems.
Cooking system	Gas fuelled or alternatively fuelled system.[(a)]		Alternatively fuelled (e.g. electrical) cooking systems should result in reduced base consumption.
Bedrooms	1, 2, 3, 4 or ≥ 5.	b_1	This is a simple metric known to consumers that can be used as a proxy measure of building floor area, which in turn is related to the building's exposed fabric area that is used to determine a building's heat loss coefficient.
Dwelling type	Apartment, terrace, semi-detached, detached or bungalow.		In general terms, these alternative building types have increasing proportions of exposed building fabric area, which in turn is related to the building's heat loss coefficient.
Construction year	Pre-1935, 1935-1979, 1980-1999, 2000-2004 or 2005-2010.[(b)]		These construction years generally relate to increasing levels of insulation as required by Irish building standards. And this relates to the fabric U-values used to determine a building's heat loss coefficient.
Boiler service frequency	Annually, every 2-3 years or never.		This relates to heating system efficiency, which in turn is used to determine a building's heat loss coefficient.
Temperature set-point	< 18°C, 18-20°C, 21°C, 22-24°C, >24°C, not known by the consumer, or no thermostat control system.	T_B	This set-point together with the heating system operating hours is related to a building's average indoor temperature, and this is in turn is related to the building's base temperature.
Timer control	Separate zones, single zone, not known by the consumer, or either the timer system is not used or no time control system is present.[(a)]		This relates to heating system operating hours and whether or not a consumer can control the set-point temperature in different zones of their building in order to facilitate decreased average indoor temperatures. All of which is related to the building's base temperature.
Operating hours	$0 <$ hours/day ≤ 8, $8 <$ hours/day ≤ 10, $10 <$ hours/day ≤ 12 or $12 <$ hours/day ≤ 24.[(c)]		See temperature set-point description above.

Notes:

(a) The levels (or categories) of this explanatory factor incorporate alternative categories or answers allowed in the survey questionnaire. For example, there were three alternative answers in the survey which described a gas fuelled hot water system: 1) central heating system, 2) combination boiler (no hot water cylinder) or 3) gas fuelled system.

(b) Construction year is reported in the survey either by the actual construction year or by the categories given in the table, thus any actual construction years reported in the survey have been also been categorised.

(c) Heating system operating hours have been determined using each consumer's hourly resolution smart-metered gas consumption data. For simplicity, this metric has been evaluated for each consumer by the average daily number of gas consumption hours during the second week of January. During this week, it is assumed that buildings are likely to be occupied and heating systems are likely to consume gas during each timed operating hour. Any suspected pilot light consumption in the sample has been accounted for by applying a nominal 0.5 kWh gas consumption threshold to the hourly gas consumption data.

5 Results and Discussion

The results of this study begin with a presentation of the R^2 distribution resulting from the individual NLS models for the consumer sample. Models which poorly fit the data are removed. A R^2 value threshold of 0.6 was chosen resulting in the removal of 66 dwellings.

5.1 NLS Regression Results

The distribution of R^2 values resulting for each of the household NLS models are shown in Fig. 1. From this distribution it has been found that 15% and 72% of the models have strong and moderately-strong R^2 values above 0.8 and between 0.6 and 0.8 respectively. However, 13% of the R^2 values are weak to moderate between 0 and 0.6, and as result these models or consumers have been eliminated from the subsequent NLS model analysis. These consumers gas consumption was frequently zero during the heating season, indicating they were either unoccupied, or intermittently occupied. Consequently, they would not represent a good opportunity for energy savings. In addition, two consumers from the total sample (524) are not included in the R^2 distribution in Fig. 1 or in the subsequent NLS model analysis, as the algorithm failed to converge using these consumers' gas consumption series. Again, both of these consumers had numerous zero consumption days during wintertime.

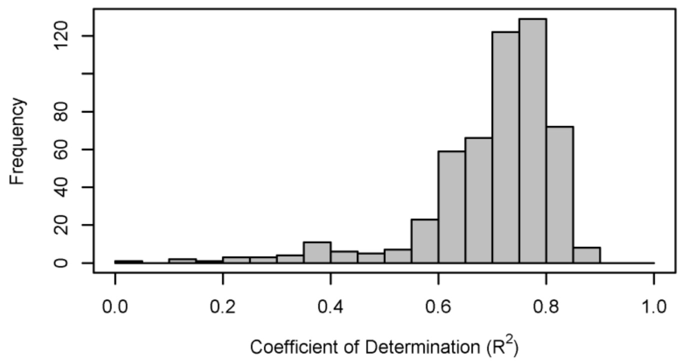

Fig. 1 Distribution of coefficient of determination (R^2) values for the NLS models (522 sample size).

The distribution of the intercept, slope and base temperature parameters for the NLS models for the retained consumer sample are shown in Figs. 2 - 4. Each of these parameter distribution have been categorised by low and high quartiles and a medium interquartile range. These categories are shown using boxplots in the figures and are used as a basis in which to develop the following MLR models. This limitation to quartiles allows simple classifications of each distribution and reduces the size of the resulting MLR models in Tables 3 - 5.

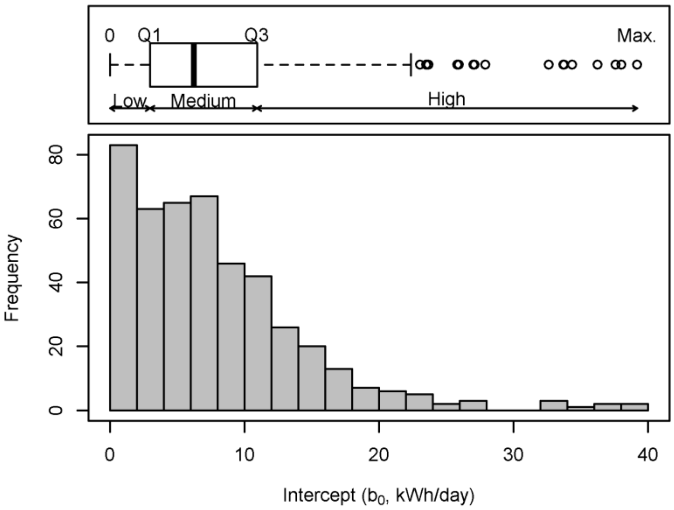

Fig. 2 Distribution of intercept (b_0) parameters for NLS models with an $R^2 \geq 0.6$ (456 sample size).

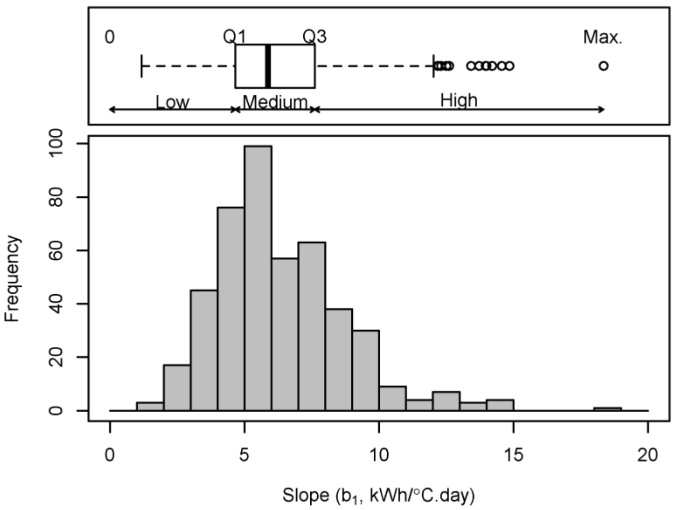

Fig. 3 Distribution of slope (b_1) parameters for NLS models with an $R^2 \geq 0.6$ (456 sample size).

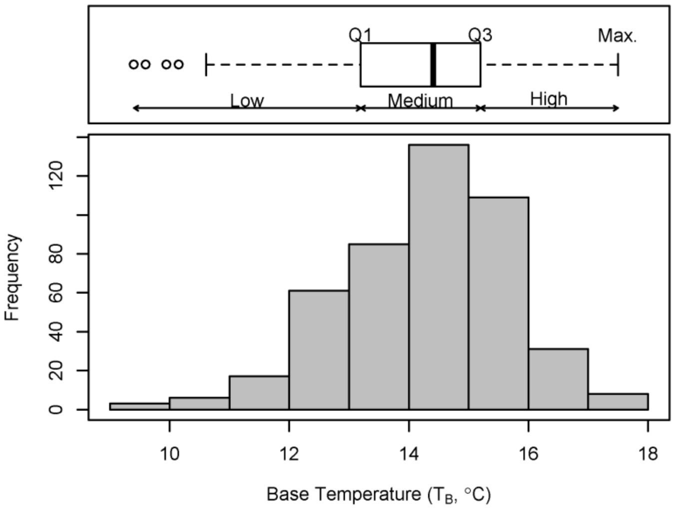

Fig. 4 Distribution of base temperature (T_B) parameters for NLS models with an $R^2 \geq 0.6$ (456 sample size).

It has been found that the mean value of the base temperature parameter distribution is 14.23°C. This is over a degree lower than the 15.5°C traditionally assumed for HDD modelling in the UK. This is unsurprising as this 15.5°C value was recommended in 1934 (CIBSE, 2006), since when improvements have been made to heating control systems and building insulation standards.

5.2 Multinomial Logistic Regression Models

The MLR models for the intercept, slope and base temperature parameter distributions are shown in Tables 3 - 5. Likelihood ratio (or X^2) tests for these models show that each model rejects the test's null hypothesis (see Note (a) in Table 3), and that most explanatory factors are significant in this regard. Although, some explanatory factors did not significantly contribute to their respective models, including: the number of children, boiler service frequency and temperature set-point. By comparing the pseudo-R^2 value (see Note (d) in Table 3) for each model, it is seen that the slope and base temperature models have the best and weakest overall fits, respectively.

Table 3 Multinomial logistic regression model for the intercept parameter

Intercept Model		X^2 test of -2LL (df) [c]					pseudo-R^2 [d]				
Overall Model [a]		74.5	(16)	***			0.22				
Explanatory Factors [b]											
No. of Adults		17.34	(4)	**							
No. of Children		2.73	(6)								
Hot Water		27.95	(4)	***							
Cooking		30.9	(2)	***							
Sub-models [e]	Low					Med. [f]	High				
	n [g]	β [h]	SE [i]		Exp(β)	n	n	β	SE		Exp(β)
Intercept		-1.72	0.37	***	0.18			-0.46	0.28		0.63
No. of Adults:											
2 [j]	61					107	53				
1	6	0.22	0.56		1.24	11	1	-1.82	1.06	˙	0.16
≥ 3	13	-0.94	0.37	*	0.39	59	36	0.28	0.28		1.33
No. of Children:											
0 [j]	43					104	51				
1	14	0.22	0.40		1.25	36	17	-0.07	0.35		0.93
2	18	0.52	0.40		1.69	23	14	0.36	0.40		1.44
≥ 3	5	-0.16	0.58		0.85	14	8	0.15	0.49		1.16
Hot Water:											
Untimed gas system [j]	26					75	45				
Timed gas system	24	-0.21	0.35		0.81	71	38	-0.09	0.28		0.91
Alternative system	30	1.13	0.38	**	3.11	31	7	-1.09	0.46	*	0.34
Cooking:											
Gas cooker [j]	25					105	60				
Alternative system	55	1.43	0.31	***	4.18	72	30	-0.39	0.28		0.67

Notes:
(a) Chi-squared (X^2) test to ascertain the significance of the decrease in unexplained variance from an intercept only model to the overall model (Field, 2013), based on the null hypothesis that each regression coefficient in the model is zero (Andrews & Krogmann, 2009). The -2 log likelihood (-2LL) statistic used in this test is given by -2(LL(intercept model) - LL(overall model)) (Andrews & Krogmann, 2009; Field, 2013). This X^2 test is based on model's corresponding degrees of freedom (df) (Andrews & Krogmann, 2009).
(b) X^2 test to ascertain the significance of explanatory factors to the overall model (Field, 2013). This -2LL statistic is given by -2(LL(overall model) - LL(overall model without the factor under test) (Field, 2013). This X^2 test is based on explanatory factor's corresponding degrees of freedom (df).
(c) See notes (a) and (b).
(d) Nagelkerke's pseudo-R^2 statistic is a measure of the improvement in fit of the overall model compared to a model with no independent variables. This statistic has a range of 0 to 1 and is analogous to the coefficient of determination (R^2) statistic used in ordinary least squares regression (Andrews & Krogmann, 2009).
(e) Sub-model categories: 0≤Low<Q1 and Q3<High≤Max; where Q1, Q3 and Max are the first and third quartiles and the maximum value of the modelled distribution.
(f) Reference category level. (g) Sample size (n). (h) Coefficient (β). (i) Standard Error (SE).(j) Reference factor level.
˙, *, ** and *** significance at 0.1, 0.05, 0.01 and 0.001 levels respectively.

Table 4 Multinomial logistic regression model for the slope parameter

Slope Model		X^2 test of -2LL (df) [c]					pseudo-R^2 [d]				
Overall Model [a]		157.4	(24)	***			0.37				
Explanatory Factors [b]											
Bedrooms		62.57	(6)	***							
Dwelling Type		32	(6)	***							
Construction Year		46.34	(8)	***							
Boiler Service Freq.		5.2	(4)								
Sub-models [e]	Low					Med.[f]	High				
	n [g]	β [h]	SE [i]		Exp(β)	n	n	β	SE		Exp(β)
Intercept		-0.82	0.28	**	0.44			-0.96	0.30	**	0.38
Bedrooms:											
3 [j]	67					122	29				
≤2	19	1.52	0.47	**	4.59	9	1	-1.26	1.13		0.28
4	15	-0.91	0.35	**	0.40	62	55	1.21	0.32	***	3.34
≥ 5	1	-1.12	1.13		0.33	6	13	2.19	0.61	***	8.97
Dwelling Type:											
Semi-detached [j]	54					115	51				
Apartment/Terrace	37	0.06	0.30		1.07	54	5	-1.63	0.54	**	0.20
Detached	8	-0.18	0.49		0.83	24	36	1.09	0.38	**	2.98
Bungalow	3	-0.45	0.79		0.64	6	6	0.84	0.67		2.32
Construction Year:											
1935-1979 [j]	31					77	53				
<1935	11	0.32	0.51		1.38	14	13	0.47	0.52		1.60
1980-1999	34	0.56	0.32	·	1.75	64	21	-1.28	0.37	***	0.28
2000-2004	18	0.79	0.39	*	2.20	32	10	-1.71	0.48	***	0.18
2005-2010	8	0.92	0.56	·	2.51	12	1	-2.93	1.12	**	0.05
Boiler Service Freq.											
Annually [j]	63					106	54				
2-3 years	32	-0.40	0.28		0.67	76	38	0.35	0.31		1.42
Never	7	-0.53	0.51		0.59	17	6	-0.29	0.60		0.75
Notes: see Table 3											

Table 5 Multinomial logistic regression model for the base temperature parameter

Base Temperature Model		X^2 test of -2LL (df) [c]				pseudo-R^{2}[d]					
Overall Model [a]		40.26	(22)	*		0.11					
Explanatory Factors [b]											
Temperature Set-point		10.28	(12)								
Timer Control		10.07	(4)	*							
Operating Hours		17.91	(6)	**							
Sub-models [e]	Low					Medium[f]	High				
	n[g]	β[h]	SE[i]		Exp(β)	n	n	β	SE		Exp(β)
Intercept		-0.50	0.29	·	0.61			-1.21	0.34	***	0.30
Temperature Set-point:											
18 - 20°C [j]	33					51	26				
< 18°C	8	0.08	0.54		1.08	11	1	-1.50	1.09		0.22
21°C	6	-0.61	0.53		0.55	18	10	-0.02	0.48		0.98
22 - 24°C	7	-0.14	0.53		0.87	12	10	0.53	0.51		1.69
> 24°C	2	-0.50	0.87		0.60	5	5	0.79	0.70		2.20
No Thermostat	32	-0.29	0.31		0.75	70	31	-0.12	0.33		0.88
Unknown	14	-0.14	0.41		0.87	25	15	0.27	0.42		1.31
Timer Control:											
Single Zone [j]	58					130	71				
Separate Zones	20	0.65	0.36	·	1.92	23	5	-0.95	0.54	·	0.39
No Timer/Not Used	24	0.29	0.31		1.34	39	22	-0.01	0.32		0.99
Operating Hours:											
0 < hours/day ≤ 8 [j]	48					81	23				
8 < hours/day ≤ 10	28	0.00	0.31		1.00	49	32	0.80	0.33	*	2.23
10 < hours/day ≤ 12	13	-0.36	0.39		0.70	28	14	0.62	0.41		1.85
12 < hours/day ≤ 24	13	-0.47	0.38		0.62	34	29	1.11	0.35	**	3.04
Notes: see Table 3											

It can be seen that each statistically significant coefficient (β) estimate in the MLR models is consistent with the residential gas consumption dynamics described in Table 2. This is confirmed by the following characterisations of the intercept, slope and base temperature parameter distributions:

- Dwellings with low b_0 intercepts (which are inferred to use little gas for cooking and hot water) are unlikely to be occupied by three or more adults, given this factor's low odds-ratio (Exp(β)) value, and are highly likely to use alternative hot water and cooking systems, given these factors high odds-ratios. Those with high intercepts are unlikely to be occupied by a single adult and to use an alternative hot water system.
- Dwellings with low b_1 slopes (which are inferred to have low exposed envelope areas and/or low U-values) are likely to have no more than two bedrooms, and to have been built since 1980. Those with high slopes are likely to have four or more bedrooms, are likely to be detached dwellings rather than apartment or terrace type dwellings and are unlikely to have been built since 1980.
- Dwellings with low T_B base temperatures are likely to use zoned time control systems. High base temperature dwellings are unlikely to use zoned time control systems, and are likely to have their heating systems operated for over eight hours each day, although this characterisation is not statistically significant for the ten to eleven hours category.

5.3 Energy Efficiency Assessments

In this section the MLR models presented in Tables 3 - 5 are used to compare the relative energy end-use levels of consumers with the same household characteristics in order to identify buildings with unexpectedly high intercept, slope and base temperature parameter estimates.

In Table 6, intercept parameters are presented for three sample consumers – Consumer No. 1, 2 and 3. It is seen that these consumers have low, medium and high intercept parameter estimates, respectively, even though they share the same household characteristics. Based on these characteristics, 9%, 58% and 33% probabilities have been predicted for the low, medium and high intercept categories, respectively, using MLR probability formulae (Field, 2013) and the relevant β coefficients in Table 3. Therefore, Consumer No. 3 has an unexpectedly high intercept parameter estimate; thus indicating unusually high hot water and cooking consumption. This may be due to an inefficient hot water heating system, poor hot water cylinder insulation, or high hot water consumption by the occupants, relative to the other consumers in the Table. Energy saving opportunities should be explored for this consumer in this regard. For example, this consumer could: 1) decrease the number of operating hours set by their hot water system’s timer, 2) upgrade their hot water cylinder’s insulation, and/or 3) decrease its temperature set-point, if such a control system is present. In addition, it is estimated that Consumer No. 3 spends approximately €425/year on cooking and hot water (14.51kWh/day (intercept) x 365days/year x €0.08/kWh) at current Irish gas market rates. This estimate may be used to assess the viability of installing a solar hot water heating system or boiler upgrade based on current cost estimates.

Table 6 Energy efficiency assessments

Consumer	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9
Parameter	Intercept (b_0)			Slope (b_1)			Base temperature (T_B)		
Estimate	2.73	8.21	14.51	3.18	5.84	7.63	11.66	14.0	15.54
Standard Error	1.72	1.87	2.65	0.12	0.16	0.26	0.31	0.62	0.62
Category	Low	Med.	High	Low	Med.	High	Low	Med.	High
Characteristics	2 adults			3 bedrooms			18-20°C temp. set-point		
	0 children			Semi-detached			Single zone timer		
	Timed gas fuelled hot water			1980-1999 construct. year			0 - 8 operating hours		
	Gas cooker			Annual boiler service					
Category	Probability								
Low	9%			41%			32%		
Medium	58%			53%			52%		
High	33%			6%			16%		

In Table 6, the estimated slope parameters are presented for another three consumers – Consumer No. 4, 5 and 6. It is seen that Consumer No. 6 has an unexpectedly high slope parameter estimate. This indicates that this dwelling may have an inefficient space heating system or a building fabric with poor thermal insulation levels, relative to the other consumers in the Table. Therefore, this consumer may benefit from a boiler or building fabric upgrade. It is estimated that this consumer spent approximately €660 on space heating for the previous year (7.63kWh/°C·day (slope) x 1078.72°C·day/year x €0.08/kWh, where the total HDDs for the year is estimated using the dwelling’s base temperature). This estimate may be used to assess the viability of boiler or building fabric upgrades based on current cost estimates.

In Table 6, the estimated base temperature parameters are presented for another three consumers – Consumer No. 7, 8 and 9. It is seen that Consumer No. 9 has an unexpectedly high base temperature parameter estimate, relative to the other consumers in the Table. Such consumers could be targeted with behavioural change programmes or zoned heating control systems. If for example, behavioural change or zoning results in a nominal 1°C reduction in base temperature, a saving of approximately €140 was possible in the modelled year for this consumer (5.53 kWh/°C·day (slope) x (2365.45-2050.41)°C·day/year x €0.08/kWh, where the reduction in HDDs is estimated using the total HDDs for a 1°C reduction in the dwelling's base temperature parameter).

Table 7 summarises the advice which could be given to individual customers with high inferred gas end-uses or NLS modelling parameters, such as Consumer No. 3, 6 and 9. Possible energy saving interventions are also given in Table 7 and these can be tailored to individual customers based on household data gathered through phone interviews or internet survey.

Table 7 Potential causes of high inferred gas end-uses and possible energy saving interventions

High Parameter	Potential Causes	Possible Interventions
Intercept (b_0)	Inefficient hot water heating system	Install timer and thermostatic control system to regulate hot water temperature. Reduce hot water cylinder set-point temperature and the operating hours of the hot water heating system. Install water softener to reduce limescale deposits in the hot water heating system. Service the gas boiler to increase its efficiency.
	Poor hot water cylinder insulation	Fit lagging jacket and/or insulation to reduce hot water cylinder and pipework heat losses.
	Excessive hot water consumption	Install water saving devices such as spray head taps and mixer showers in place of power (or pumped) showers.
Slope (b_1)	Inefficient space heating system	Install zoned timer and thermostatic control systems to regulate room temperatures by occupancy profile. Reduce set-point temperatures on room thermostats to limit heating system operating hours. Move poorly located thermostats such as those obstructed by furniture or those close to heat emitters in order to improve their effectiveness. Service the gas boiler to increase its efficiency.
	Poor building fabric insulation and excessive air-infiltration	Upgrade insulation and windows to decrease envelope *U*-values. Install and/or maintain draught stripping devices on windows and doors in order to limit air-infiltration.
Base temperature (T_B)	Excessive indoor temperature and heating system operating hours	See above advice on inefficient space heating systems. Ensure that window blinds and curtains allow solar gains during the day and limit heat losses at night.

6 Conclusions and Recommendations

This paper presents a NLS regression model to estimate intercept, slope and base temperatures of individual dwellings using daily gas consumption and ambient external temperature data. These consumption data will become available as smart metering infrastructure is deployed across Europe. The resulting model parameter estimates can be used to infer gas consumption end-use due to cooking/water heating, envelope energy efficiency and heating control performance

This study also demonstrated a multinomial logistic regression modelling method based on the resulting intercept, slope and base temperature parameter distributions and various household characteristics. This was used to compare the inferred gas end-uses of individual dwellings to other dwellings with similar characteristics. These models have been presented as an alternative to energy intensity metrics based on building floor area. By way of example, the multinomial logistic regression models were used to compare the inferred gas end-use efficiency of similar buildings based on their intercept, slope and base temperature parameter estimates. It was shown that households with high (and low) relative consumptions can be identified using this approach

Acknowledgement

This study has been supported by Dublin Institute of Technology (DIT) and Gas Networks Ireland under DIT's Fiosraigh Enterprise Scheme. The authors also wish to acknowledge the Irish Commission for Energy Regulation and the Irish Social Science Data Archive for providing the residential gas consumption data, and Met Eireann for providing the temperature data.

References

Andrews, C. J. and Krogmann, U. (2009). Explaining the adoption of energy-efficient technologies in U.S. commercial buildings, Energy and Buildings, vol. 41, pp. 287-294. DOI: http://dx.doi.org/10.1016/j.enbuild.2008.09.009.

Chartered Institution of Building Services Engineers (CIBSE, 2006). Degree-days: theory and application, CIBSE, TM41.

Commission de Régulation de L'Énergie (2012). Deliberation of the French Energy Regulatory Commission of 28 February 2012 forming a decision on the equalised tariff for use of GrDF public natural gas distribution networks. Available: http://www.cre.fr/en/documents/deliberations/decision/grdf-public-natural-gas-distribution-networks/deliberation-of-the-french-energy-regulatory-commission-of-28-february-2012-forming-a-decision-on-the-equalised-tariff-for-use-of-grdf-public-natural-gas-distribution-networks. Access Date: 21/04/2015.

Commission for Energy Regulation (2011). Smart Metering Information Paper - Gas Customer Behaviour Trial Findings Report. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/. Access Date: 21/04/2015.

Day, A. R., Knight, I., Dunn, G., and Gaddas, R. (2003). Improved methods for evaluating base temperature for use in building energy performance lines, Building Services Engineering Research and Technology, vol. 24, pp. 221-228. DOI: http://dx.doi.org/10.1191/0143624403bt073oa.

Energy Star (2014a). Score - Technical Reference. Available: https://portfoliomanager.energystar.gov/pdf/reference/ENERGY%20STAR%20Score.pdf. Access Date: 21/04/2015.

Energy Star (2014b). Score for Multifamily Housing in the US. Available: http://www.energystar.gov/sites/default/files/tools/Multifamily.pdf. Access Date: 21/04/2015.

Energy Star (2015a). Home Energy Yardstick. Available: https://www.energystar.gov/index.cfm?fuseaction=HOME_ENERGY_YARDSTICK.showGetStarted. Access Date: 21/04/2015.

Energy Star (2015b). How the Yardstick works? Available: https://www.energystar.gov/index.cfm?fuseaction=HOME_ENERGY_YARDSTICK.showHowItWorks . Access Date: 21/04/2015.

Energy Star (2015c). Home Improvement: Resources for Utilities, State Energy Offices, and Efficiency Program Sponsors. Available: https://www.energystar.gov/ia/partners/univ/download/HIPResources_for_Utilties.pdf?0544-2a1e. Access Date: 21/04/2015.

EN ISO 15927-6 (2007). Hygrothermal Performance of Buildings – Calculation and Presentation of Climatic Data. Part 6: Accumulated Temperature Differences (Degree-Days).

European Commission (2009). The 2020 climate and energy package. Available: http://ec.europa.eu/clima/policies/package/index_en.htm. Access Date: 21/04/2015.

European Commission (2014). Energy balance sheets 2011-2012. Available: http://ec.europa.eu/eurostat/documents/3217494/5785109/KS-EN-14-001-EN.PDF. Access Date: 21/04/2015.

Field, A. (2013). Discovering statistics using IBM SPSS statistics, Sage.

Hierzinger, R., Albu, M, van Elburg, H., Scott, A. J., Łazicki, A., Penttinen, L., et al (2013). European Smart Metering Landscape Report 2012 – update May 2013. Available: www.smartregions.net. Access Date: 21/04/2015.

Joint Office of Gas Transporters (2015). Uniform Network Code – Transportation Principal Document Section M - Supply Point Metering, Version 4.64 09 April 2015. Available: http://www.gasgovernance.co.uk/sites/default/files/TPD%20Section%20M%20-%20Supply%20Point%20Metering_24.pdf. Access Date: 21/04/2015.

Official Journal of the European Union (2009). Directive 2009/73/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC.

Official Journal of the European Union (2012). Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC.

Raffio, G., Isambert, O., Mertz, G., Schreier, C., and Kissock, K. (2007). Targeting residential energy assistance, ASME 2007 Energy Sustainability Conference, pp. 489-495, American Society of Mechanical Engineers. DOI: http://dx.doi.org/10.1115/ES2007-36080.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: http://www.R-project.org/. Access Date: 21/04/2015.

Sustainable Energy Authority Ireland (SEAI, 2015). Display Energy Certificates for Large Public Buildings. Available: http://www.seai.ie/Your_Building/BER/Large_Public_Buildings/DEC_FAQ/. Access Date: 21/04/2015.

Wang, S., Yan, C. and Xiao, F. (2012). Quantitative energy performance assessment methods for existing buildings, Energy and Buildings, vol. 55, pp. 873-888. DOI: http://dx.doi.org/10.1016/j.enbuild.2012.08.037.

Watson, D. and Williams, J. (2003). Irish National Survey of Housing Quality 2001-2002, Economic and Social Research Institute (ESRI) Research Series.