2015-03-01

A Clustering Approach to Domestic Electricity Load Profile Characterisation Using Smart Metering Data

Fintan McLoughlin
Technological University Dublin, fintan.mcloughlin@tudublin.ie

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

Michael Conlon
Technological University Dublin, michael.conlon@tudublin.ie

Recommended Citation

McLoughlin F., Duffy, A., Conlon, M. "A clustering approach to domestic electricity load profile characterisation using smart metering data", *Applied Energy*, Vol. 141 March 2015. doi:10.1016/j.apenergy.2014.12.039

Funder: Fiosraigh

AUTHOR QUERY FORM

ELSEVIER	**Journal: APEN** **Article Number: 5957**	**Please e-mail or fax your responses and any corrections to:** **E-mail: corrections.esch@elsevier.sps.co.in** **Fax: +31 2048 52799**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact r.balan@elsevier.com immediately prior to returning your corrections.
Q2	Please confirm that given name(s) and surname(s) have been identified correctly.
	Please check this box if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Applied Energy xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Energy

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

A clustering approach to domestic electricity load profile characterisation using smart metering data

Fintan McLoughlin [a,*], Aidan Duffy [a], Michael Conlon [b]

[a] School of Civil Engineering and Dublin Energy Lab, Dublin Institute of Technology, Bolton St, Dublin 1, Ireland
[b] School of Electrical & Electronic Engineering and Dublin Energy Lab, Dublin Institute of Technology, Kevin St, Dublin 4, Ireland

H I G H L I G H T S

- We characterise diurnal, intra-daily, seasonal and between customer electricity use.
- A series of profile classes reflective of home electricity use are constructed.
- We examine the influence of household characteristics on patterns of electricity use.

A R T I C L E I N F O

Article history:
Received 1 August 2014
Received in revised form 4 December 2014
Accepted 21 December 2014
Available online xxxx

Keywords:
Domestic electricity load profile
Segmentation
Clustering

A B S T R A C T

The availability of increasing amounts of data to electricity utilities through the implementation of domestic smart metering campaigns has meant that traditional ways of analysing meter reading information such as descriptive statistics has become increasingly difficult. Key characteristic information to the data is often lost, particularly when averaging or aggregation processes are applied. Therefore, other methods of analysing data need to be used so that this information is not lost. One such method which lends itself to analysing large amounts of information is data mining. This allows for the data to be segmented before such aggregation processes are applied. Moreover, segmentation allows for dimension reduction thus enabling easier manipulation of the data.

Clustering methods have been used in the electricity industry for some time. However, their use at a domestic level has been somewhat limited to date. This paper investigates three of the most widely used unsupervised clustering methods: k-means, k-medoid and Self Organising Maps (SOM). The best performing technique is then evaluated in order to segment individual households into clusters based on their pattern of electricity use across the day. The process is repeated for each day over a six month period in order to characterise the diurnal, intra-daily and seasonal variations of domestic electricity demand. Based on these results a series of Profile Classes (PC's) are presented that represent common patterns of electricity use within the home. Finally, each PC is linked to household characteristics by applying a multi-nominal logistic regression to the data. As a result, households and the manner with which they use electricity in the home can be characterised based on individual customer attributes.

1. Introduction

Throughout the European Union, there has been a move towards smarter electricity networks, where increased visibility over electricity generation and consumption has been achieved with the installation of Advanced Metering Infrastructure (AMI). Smart metering is part of this and is seen as a necessary component to achieve EU 20-20-20 energy policy goals by the year 2020: to cut greenhouse gas emissions by 20%, to improve energy efficiency by 20% and for 20% of EU energy demand to come from renewable energy resources [1].

In recent years, smart meter installations have increased worldwide in a bid to modernise aging electricity networks [2]. Furthermore, improvements in the regulatory environment, particularly within the residential sector in Europe has resulted in a number of smart metering pilot programmes [3]. As a consequence, a wealth of new data exists for utilities, giving detailed electricity consumption at increased granularity for a large number of customers within the residential sector [4]. The availability of this source of data can potentially be used by utilities to create customised electricity load Profile Classes (PC) and can assist in areas such

* Corresponding author. Tel.: +353 (0)1 4027912; fax: +353 (0)1 4024035.
E-mail address: fintan.mcloughlin@dit.ie (F. McLoughlin).

http://dx.doi.org/10.1016/j.apenergy.2014.12.039
0306-2619/© 2014 Published by Elsevier Ltd.

as: improved load planning and forecasting; Time of Use (ToU) tariff design; electricity settlement; and Demand Side Management (DSM) strategies [5].

This paper presents a new methodology for electricity load profile characterisation. In doing so, a series of domestic electricity PC's are constructed that are reflective of the varied manner with which electricity is used within the home. Currently, PC's are derived based on aggregating many dissimilar patterns of electricity use together [6]. The application of this type of approach, where individual households which may use electricity in very different ways get lumped together, results in the formation of highly aggregated load profiles. However, in reality this is not a true reflection of how electricity is actually consumed and which can change considerably between different customers [7]. The paper proposes an alternative method which uses clustering to identify similar patterns of electricity use before any aggregation processes are applied. In this way, information pertaining to the electricity load profile shape is not lost. In addition, the paper also presents a method of linking PC's to individual customers so that a household and the manner with which they use electricity within the home can be characterised based on their individual customer attributes.

The paper is structured as follows. Section 2 illustrates existing methods used for electricity load profile characterisation and their limitations in dealing with smart metering data. Section 3 presents the structure of the data on which the analysis was carried out. Section 4 provides the methodological approach for the paper which is divided into three distinct sections: clustering; electricity load profile characterisation; and customer profile classification. Section 5 presents and discusses results with Section 6 containing concluding remarks.

2. Domestic electricity load profile characterisation

Based on the literature, existing methods used to characterise domestic electricity use can generally be divided into four categories: statistical; engineering; time series and clustering. Statistical methods have been widely used in de-regulated electricity markets to form standard load PC's [6]. Standard load PC's are used for the purposes of settlement and provide an estimate as to the quantity and Time of Use (ToU) of electricity being used. A series of PC's are produced for different segments of the market (e.g. residential, commercial, industrial) and are derived based on the average for all customers contained within a single customer class [8]. The UK electricity market has two domestic PC's; Unrestricted and Economy 7. In Ireland, four PC's exist for the domestic sector; 24 h and Night Saver which are split by urban and rural divide [9]. Although PC's are suitable for the purposes of settlement, in reality they are not reflective of how electricity is actually consumed within the home on a daily basis and merely represent the average for all customers contained within the same class. Other statistical techniques consist of using descriptive statistics and probability [10–16] as well as regression [17–22] to describe electricity use within the home. Similar to that stated above, these methods produce highly diversified load profile shapes, a result of combining many dissimilar patterns of electricity use together [10].

Engineering approaches to domestic load profile characterisation are varied but generally characterise electricity use as a function of parameters such as occupancy or appliance ownership [23–28]. These methods are considered to be a bottom up approach where multiple profiles are constructed for different households and therefore do not suffer from the same problem highlighted above for statistical approaches. However, engineering methods are difficult to generalise and require detailed knowledge of household occupant and appliance Time Use (TU) [29]. In contrast time series approaches have been limited in their application to domestic households, but this is most likely due to a historical lack of available data for the sector [7]. The methods have been used extensively to describe electricity use at a Transmission System Operator (TSO) level [30–34]. However, these approaches suffer from a similar problem to that highlighted above for statistical techniques when many dissimilar profiles are aggregated together resulting in diversified electricity load profile shapes [35].

Finally data mining techniques such as cluster analysis have been used to group customers which exhibit similar electrical behaviour through ToU smart meter data, but have mostly been applied at an aggregated level [36–38]. Furthermore, customers have also been clustered based on aggregated parameter values such as annual electricity use or features relating to the electricity load profile shape (e.g. load factor) [39,40]. Similarly, load profiles have been constructed for commercial, industrial and mostly aggregated residential customers based on clustering methods: Self Organising Maps (SOM), k-means and Follow the Leader [41–43]. In particular, one large study of approximately 3000 residential customers was monitored over a period of a single year and used methods: SOM; k-means; and hierarchical to cluster and construct load profiles [44]. However, the analysis was restricted to only a small portion of the time series (5%) due to computational demands. Clustering methods do not suffer from many of the problems highlighted above particularly when it is applied prior to carrying out any statistical analysis. Furthermore with improvements in computer hardware tasks such as clustering, which can be computationally intensive have become easier to implement.

This paper fills a gap in the literature by clustering based on ToU for a large sample of residential customers over a period of six months. This allows for load PC's to be derived based on individual patterns of electricity use within the home over this period and does not suffer from some of the same aggregation problems highlighted above. Furthermore, as the entire dataset is clustered, diurnal, intra-daily and seasonal patterns to electricity use can be characterised, as well as between customer variations. Moreover, as dwelling, occupant and appliance characteristics are correlated with each PC's it also provides a method of assigning patterns of electricity use to individual customers. Finally, as the sample size is relatively large the PC's can be considered to be representative of the wider population in Ireland. A similar method could also be used in other electricity markets outside of Ireland.

3. Data structure

The smart metering trial carried out by Commission for Energy Regulation (CER) provided the necessary information to segment the domestic electricity market in Ireland based on ToU [45]. The trial was conducted between 2009 and 2010 and consisted of installing smart meters in over 4000 residential dwellings in Ireland. Electricity demand at half hourly intervals as well as detailed information on dwelling, occupant and appliance characteristics for a representative sample of dwellings in Ireland was recorded [46,47]. The data provided was in anonymised format in order to protect personnel and confidential information relating to the customer.

The data used in the analysis was taken over the period 1st July to 31st December 2009. The sample size was trimmed to 3941 customers in total on account of missing information due to technology communication problems. Matlab and its respective statistical (ver. 7.3) and neural network toolboxes (ver. 6.0.4) were used to carry out manipulation and analysis of the data [48]. SPSS was used to analyse dwelling, occupant and appliance characterises with a unique service ID providing the link between the two software programs [49].

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

4. Methodology

The smart metering data described in Section 3 was used to segment customers based on patterns of electricity use within the home using clustering. A series of PC's were produced and linked to dwelling and household characteristics, such as Head of Household (HoH) age and Household (HH) composition, through multi-nominal logistic regression. The methodology used is shown in Fig. 1 and can be divided into three distinct parts: clustering; electricity load PC characterisation; customer PC classification.

4.1. Stage 1 – Clustering

Firstly, each clustering technique was evaluated as to the suitability for segmenting the data. Three of the most widely used clustering algorithms for the electricity industry were investigated: k-means; k-medoid and Self Organising Maps (SOM) [50–52,42]. Secondly, a suitable number of clusters was identified to segment the data. In both cases, a Davies–Bouldin (DB) validity index was used to identify the most suitable clustering method and appropriate number of clusters [53]. This is a commonly used measure to evaluate how well a dataset has been segmented [54]. The index was evaluated over three separate random days and the average taken. This was done so as to ensure that the index was not calculated against an atypical day. Finally, once a suitable clustering method and number of clusters was identified, each day was clustered separately on a 24 h basis over a six month period. This ensures that the diurnal, intra-daily and seasonality components to electricity use within the home can be captured by the characterisation process.

4.2. Stage 2 – Electricity load PC characterisation

Electricity demand for an individual cluster on a particular day was averaged (as it represents a similar pattern of electricity use) to create a daily electricity load profile for a cluster. Clusters that were small in size and that differed slightly in terms of both magnitude and timing of electricity use were combined together (thus reducing the number of similar shaped profiles) to produce a series of PC's. This results in a vector size of 48 × 184 data points for each class representing average half hourly electricity use for each day over a six month period respectively. Fig. 2 shows an illustration of a single customer and the manner in which PC's are used to characterise daily electricity use within the home.

4.3. Stage 3 – Customer PC classification

The PC that each customer used on a particular day was recorded in a Customer Class Index (CCI). The data structure of the CCI index can be seen on the right hand side of Fig. 1. As customers tend to use electricity differently on a daily basis, as was shown in Fig. 2, often customers use multiple PC's over a period. Therefore, the statistical *Mode* of the CCI index was used to determine which PC each customer used for the majority of the time across the six month period. This was done so that a multi-nominal logistic regression could be used to determine the likelihood of a customer with individual characteristics (e.g. dwelling type, number of bedrooms, etc.) using a particular PC.

Eq. (1) describes the likelihood or odds ratio *Exp* [*B*] of using a particular PC where: β_0 is a constant; $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients that explain the association of each explanatory variable $X_1, X_2, \ldots, X_n$ (customer characteristics) on the response variable (PC). $P(x)$ describes the probability of using a particular PC when compared against a reference class $p'(x)$ [55]. The explanatory variables were chosen based on a linear multivariate regression model (shown in a previous paper) which described the key characteristics that influenced electricity use within the home [56].

$$
\begin{aligned}
Exp[B](odds\ ratio) &= \log\left[\frac{p(x)}{p'x}\right] \\
&= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad (1)
\end{aligned}
$$

Table 1 shows the sample size for each explanatory variable with base categories highlighted in bold italics. For electric water heating and cooking the base category was households that use non-electric means to heat water and cook. Similarly for each appliance type the

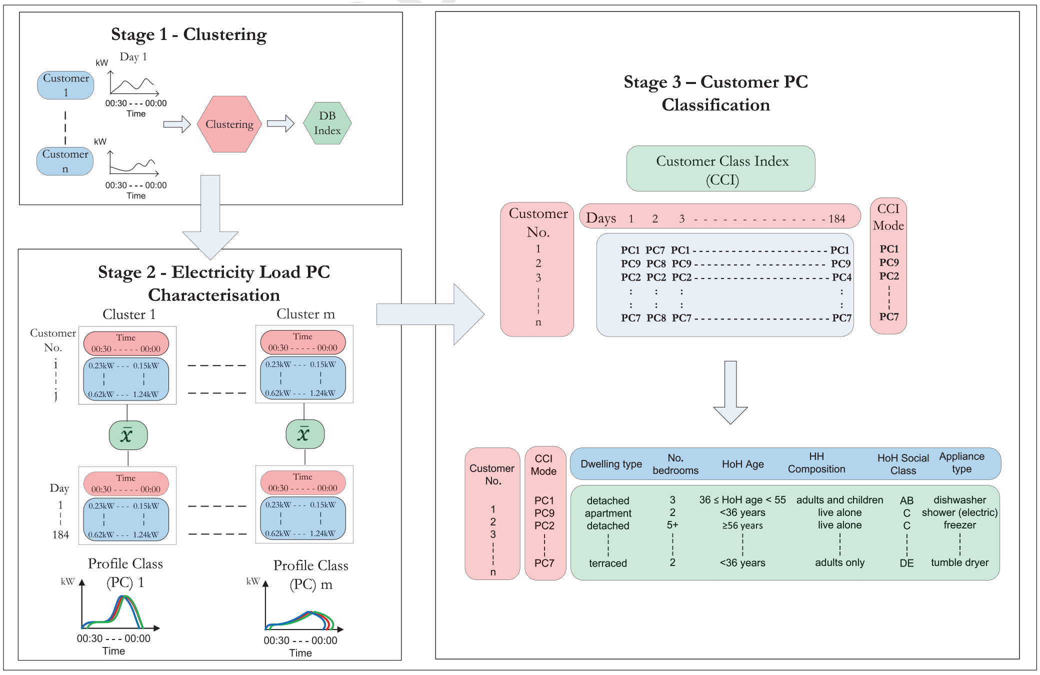

Fig. 1. Methodological approach to electricity load profile characterisation through clustering: Stages 1, 2 and 3 are described in Sections 4.1–4.3 respectively.

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

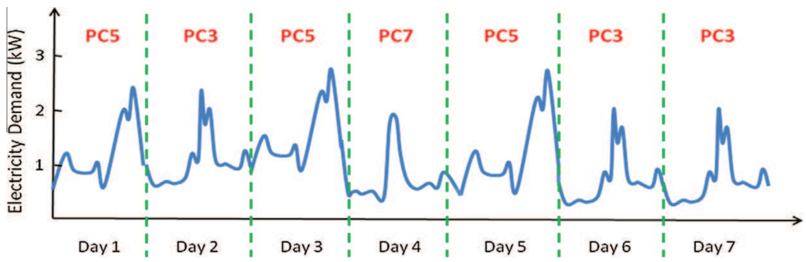

Fig. 2. Illustration of a single customer's characterised electricity use within the home using Profile Classes (PC).

Table 1
Dwelling, occupant and appliance characteristic sample sizes.

Explanatory variable	Explanatory variable explanation	Sample size (N)
Dwelling type – detached	Dwelling is detached (includes bungalows)	2068
Dwelling type – semi-detached	Dwelling is semi-detached	1230
Dwelling type – terraced	Dwelling is terraced	569
Dwelling type – apartment	Dwelling is apartment	67
No. of bedrooms – 1	Dwelling has one bedroom	42
No. of bedrooms – 2	Dwelling has two bedrooms	333
No. of bedrooms – 3	Dwelling has three bedrooms	1748
No. of bedrooms – 4	Dwelling has four bedrooms	1367
No. of bedrooms – 5+	Dwelling has five plus bedrooms	451
HoH age < 36 years	Head of household age less than 36 years	390
HoH age betw. 36 & 55 years	Head of household age between 36 and 55	1776
HoH age ⩾ 56 years	Head of household age above 56	1753
HH comp. – live alone	Household composition – live alone	756
HH comp. – with adults only	Household composition – live with adults only	2064
HH comp. – with adults and children	Household composition – live with adults and children	1121
HoH social class – AB	High and intermediate managerial, administrative or professional	593
HoH social class – C	Supervisory and clerical and junior managerial, skilled manual workers	1697
HoH social class – DE	Semi-skilled and unskilled manual workers, state pensioners, unemployed	1505
HoH social class – F	Farmers	107
Water heating – electric	Water is heated by electricity	2237
Cooking type – electric	Cooking is mostly done by electricity	2749
Washing machine	Appliance type washing machine is present	3873
Tumble dryer	Appliance type tumble dryer is present	2693
Dishwasher	Appliance type dishwasher is present	2638
Shower (instant)	Appliance type shower (instant) is present	2726
Shower (pumped)	Appliance type shower (pumped) is present	1150
Electrical cooker	Appliance type electrical cooker is present	3039
Heater (plug in convective)	Appliance type heater is present	1199
Freezer (stand alone)	Appliance type freezer is present	1961
Water pump	Appliance type water pump is present	772
Immersion	Appliance type immersion is present	3022
TV < 21 in.	Appliance type TV < 21 in. is present	2583
TV > 21 in.	Appliance type TV > 21 in. is present	3309
Computer (desktop)	Appliance type computer (desktop) is present	1864
Computer (laptop)	Appliance type computer (laptop) is present	2107
Game consoles	Appliance type game console is present	1310

base category was compared against households that do not own that particular appliance.

5. Results and discussion

The following section presents results and discussion for each stage of the methodology described in Section 4.

5.1. Clustering

The DB validity index was calculated for each clustering technique (k-means, k-medoid, SOM) and for varying number of clusters (2–16) over three separate random days with the average shown in Fig. 3. SOM showed a consistently lower DB index overall across varying number of clusters, and therefore was selected to segment the data further. The optimal number of segments used to divide the data was chosen at between 8 and 10 clusters as after this point any further decrease in DB index was minimal. It is important to note that the DB index was lowest overall for two clusters, however, as this would lead to highly aggregated PC's like that described in Section 2, more than two segments was sought.

The dataset was divided into nine clusters based on 3 × 3 hexagonal lattice structure shown on the left hand side of Fig. 4. Cluster centres are shown to be visually separated by Euclidean distance indicated by different colours. The brighter colours show clusters that are close together whereas the darker colours represent cluster centres that are further apart. It can be seen that clusters c6 and c9 are most similar to each other compared to any other cluster pair.

The cluster size is shown as a percentage of total sample size in Fig. 4. Clusters c6 and c9 combined represent nearly two thirds of the entire sample and therefore these were further divided using sub-clustering. This approach was used most recently by Lo et al.

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

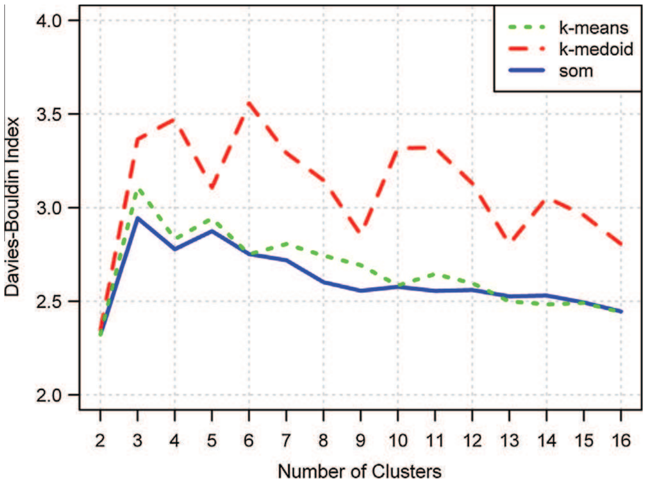

Fig. 3. Average DB index for clustering methods k-means, k-medoid, and SOM.

and Zainal et al. to break up larger clusters [57,58]. C6 and c9 were divided into four additional clusters each as shown on the right hand side of Fig. 4.

5.2. Electricity load PC characterisation

In total, ten PC's were produced using the methodology described in Section 4.2 which represent different patterns of electricity use both in terms of magnitude and timing. Fig. 5 shows the sample size for each PC as a total percentage of all classes.

Fig. 6 shows diurnal patterns of electricity use for each PC's over the six month period (note *y*-axes differ between subplots). In the majority of classes, a characteristic 'primary peak' and a smaller 'secondary peak' of electricity use is apparent. If a 'primary peak' occurs in the morning then the 'secondary peak' tends to be smaller in magnitude in the evening. Similarly, the converse is also true. It must be noted that PC8 shows characteristics quite different to any other class in terms of magnitude of electricity use across a 24 h period and most likely corresponds with a vacant dwelling.

Fig. 7 illustrates the intra-daily effects of electricity use for PC1 and is shown by Weekday, Saturday and Sunday. A similar effect is also observed across all classes but is unable to be shown due to space constraints. A clear distinction can be made between Weekends and Weekdays, where the majority of PC's show electricity use earlier in the morning for the latter. This earlier use of electricity during the Weekdays is most likely due to employment and schooling commitments for some or all of the occupants. Similarly, an earlier morning peak is apparent on Saturdays compared to Sundays, with the latter showing more electricity use across the afternoon period. An outlier is also evident for this particular class which corresponds to Christmas day.

The seasonal component to the classes is illustrated in Fig. 8. PC4 is presented, but like before a similar effect is observed across all classes. The brighter colours represent mid/late summer through to the darker colours indicating mid/late winter. The change in profile shape between seasons (particularly mornings and evenings) is likely to be influenced by sunrise and sunset times with the switching of lights on within the home. However, this could also be related to a change in occupancy between Summer and Winter. Similarly a change in profile shape during early morning/afternoon is apparent over the Summer which may also be related to changes in occupancy (e.g. children being at home during school holidays). However, this could also be related to an increase in external temperatures during the summer thus resulting in a greater cycling of cold appliances. A similar increase is also observed during the night (01:30–05:30) for the Summer suggesting that it is temperature rather than occupancy influencing its use during these times.

5.3. Customer PC classification

As discussed in Section 4.3 the statistical Mode was used to determine which PC customers used for the majority of time over the six month period. A multi-nominal logistic regression was then applied to determine the likelihood of households with certain characteristics using electricity in a similar manner to each PC. Table 2 presents results for the regression and shows the strength of the association for each explanatory variable with each individual PC's by way of an *Exp*(*B*) value. Table 2 also shows standard errors and levels of statistical significance for each explanatory variable. Standard errors indicate variation within the explanatory variable and where large errors exist, it corresponds with small sample sizes within the sub-category. This was mitigated by combining clusters that showed similar patterns of electricity use as described in Section 4.2. However, in some instances particularly for apartments and one bedroom dwellings the total overall sample size is small (67 and 42 respectively) which contributes to large standard errors for some classes. Furthermore, this also has a bear-

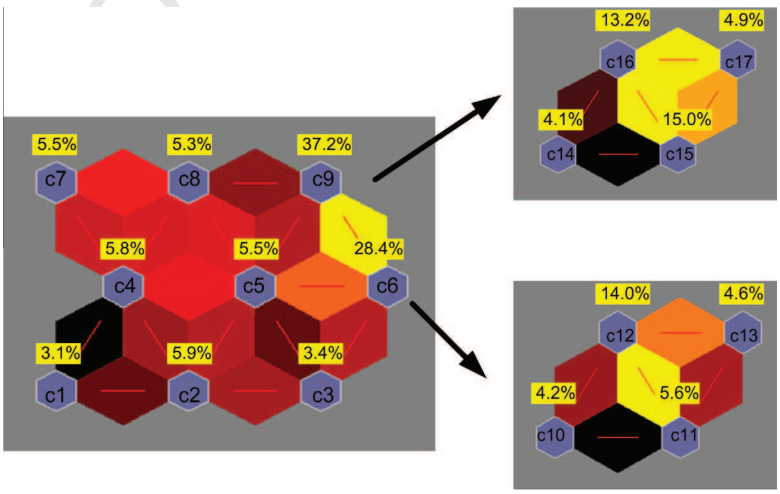

Fig. 4. Hexagonal lattice structure for SOM clusters.

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

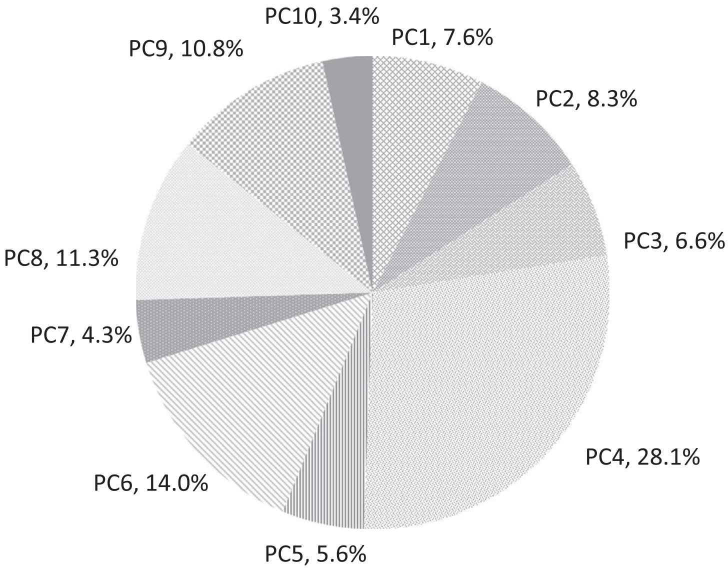

Fig. 5. Profile Class (PC) by sample sizes.

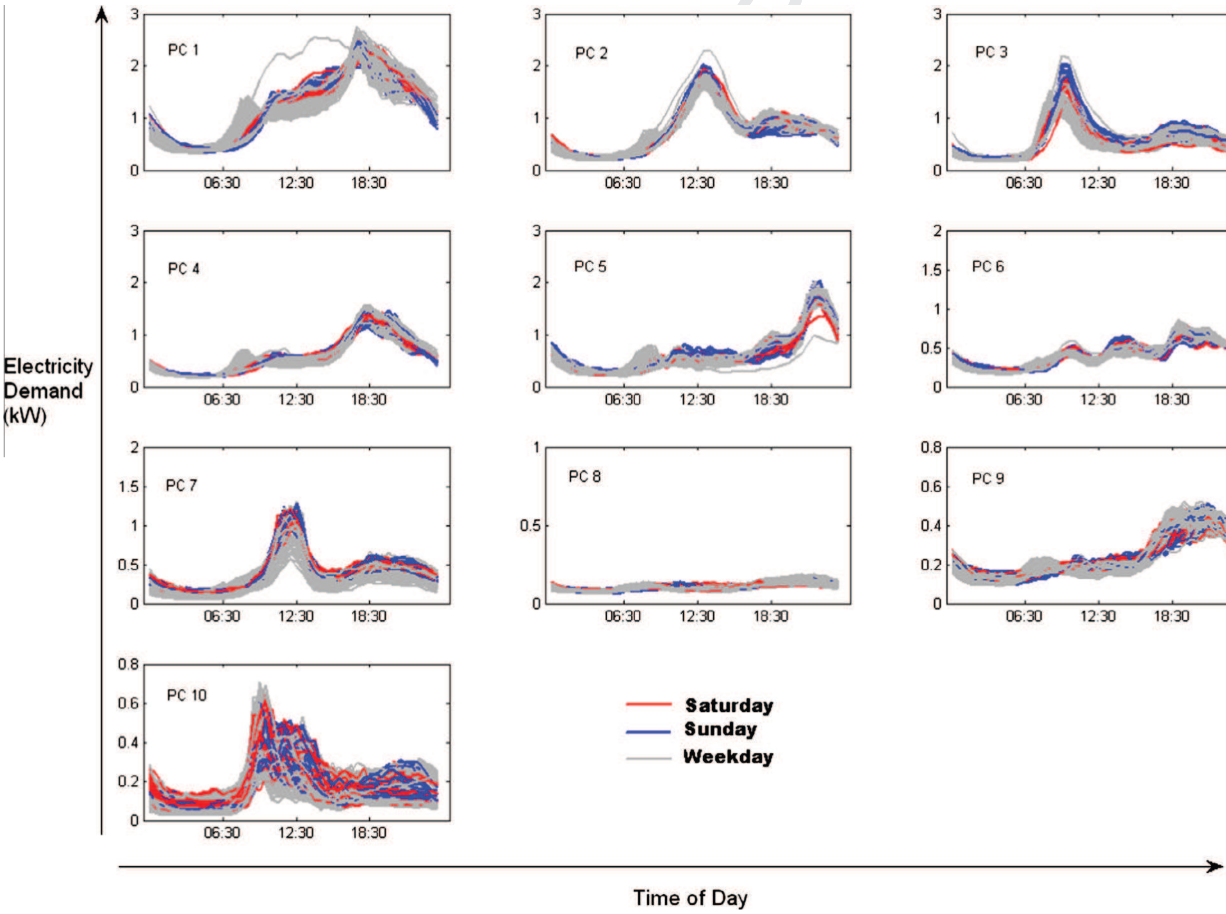

Fig. 6. PC's 1–10 over the six month period.

ing on the statistical significance within some sub-categories in the regression model. Therefore when comparing classes, the degree with which each characteristic either positively or negatively influences use of a particular PC is additionally reported in instances where it is informative.

In the following text, each PC is discussed in terms of the influence that individual customer characteristics have on its use within the home. PC4 was used as the reference class as it corresponded with the largest number of households (28% as was shown in Fig. 5).

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

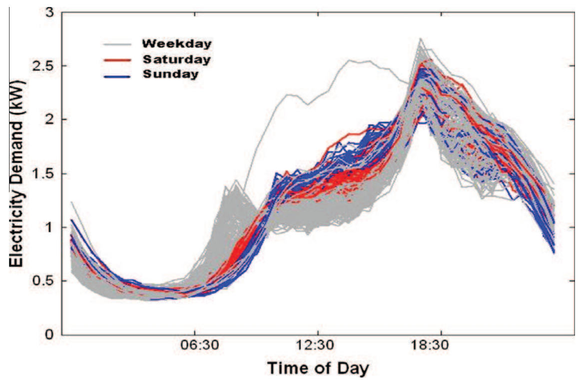

Fig. 7. PC1 by day type over the six month period.

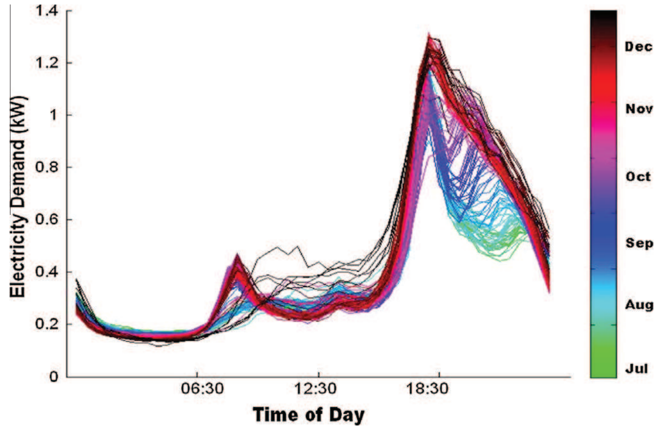

Fig. 8. PC4 for weekdays over the six month period.

5.3.1. Profile Class 1 (PC1)

This class reflects a heavy user of electricity across a 24 h period and therefore it is not surprising that occupiers of dwellings with 5+ bedrooms were more likely to use this class, with all other variables showing strong negative association within this category. Older (HoH $\geqslant$ 56 years) and middle aged (36 $\leqslant$ HoH < 55 years) were also more likely to use this PC compared to the base category, although the former was only statistically significant at the 10% level and latter not at all. A HoH social class of 'F' showed the greatest positive association for this PC but again was only shown to be significant at the 10% level. Finally, not surprisingly households that owned high energy intensive appliances such as tumble dryers and dishwashers were also more likely to use this class.

5.3.2. Profile Class 2 (PC2)

PC2 describes a high use of electricity centred around midday, with a considerably smaller evening peak compared to the previous class. The class showed poor statistically significant results within the regression model, however, it is still possible to discuss the effect. In particular, water heating showed a high association for this class which may explain the increase in electricity use around midday. Similarly, dwelling occupants which had a HoH age ($\geqslant$56 years) showed the greatest positive association. Finally, appliance types: tumble dryer, instant electric showers and water pumps all showed strong positive association.

5.3.3. Profile Class 3 (PC3)

This class showed a large morning peak with considerably less electricity used during the evening time. Similar to PC2, older HoH age ($\geqslant$56 years) showed strong positive association but this was not statistically significant. Strong positive association was also apparent for HH composition for occupants that lived alone. A strong positive association with households that use electricity for cooking was also evident but this was only statistically significant at the 10% level. Households that did not own a tumble dryer, TV > 21 inch and a desktop computer were also more likely to use this class.

5.3.4. Profile Class 4 (PC4)

As PC4 was used as the reference class all other profiles were compared against this. The class showed a similar pattern of electricity use to PC1 but with a smaller magnitude component.

5.3.5. Profile Class 5 (PC5)

PC5 shows an evening peak much later than any other class at 10:30 pm. In contrast to previous classes, younger HoH age < 36 years as well as households with a social class of 'AB' were more likely on account of negative association between all other categories for this variable, although neither were shown to be statistically significant. There was strong positive association for HH composition for people living alone although this was only shown to be significant at the 10% level. Households that did not use electricity for heating water were also more likely to use this class. Finally households that owned TV > 21 in. showed strong positive association but again was only significant at the 10% level.

5.3.6. Profile Class 6 (PC6)

This class showed three distinct electricity peaks occurring during morning, lunch and evening periods respectively, with a smaller magnitude component compared to previous classes. People living in apartments and dwellings of two and three bedrooms showed a high likelihood for using this class; however, none were shown to be statistically significant. Older households with a HoH age $\geqslant$ 56 years showed strong positive association. HH composition of live alone showed strong positive association indicating that single occupants were most likely to use this class. Households that do not use electricity to cook and/or heat water were more likely as indicated by the negative association for these categories. Finally, households that did not own a dishwasher or an instant electric shower were also more likely to use this class.

5.3.7. Profile Class 7 (PC7)

This class showed a large peak around midday but similar to PC2 showed poor statistically significant results. Comparable to PC2, this class also showed strong positive association for using electricity to heat water. Mid-sized dwellings of three and four bedrooms were more likely compared to the base category as well as households that lived with adults only. There was also strong negative association for households that did not own a dishwasher, computer or game console for this particular class.

5.3.8. Profile Class 8 (PC8)

As alluded to earlier, PC8 showed a pattern of electricity use that was quite different to the other classes in terms of the magnitude of electricity used across a 24 h period and most likely reflects an empty dwelling. Similar to PC6, people who lived in apartment dwellings showed a high likelihood of using this class, although this showed not to be statistically significant. Two bed dwelling occupants were strongly associated with this class. In contrast to PC6, younger HoH age < 36 years were more likely, with the other two age categories showing negative association. Similar to PC6, households that lived alone showed very strong positive association. Social classes 'AB' and 'F' were more likely amongst this class as well as households that did not own a tumble dryer, dishwasher or a stand-alone freezer.

Table 2
Multi-nominal logistic regression results for dwelling, occupant and appliance characteristics.

Dwelling, occupant and appliance characteristics	Profile classes																	
	PC 1		PC 2		PC 3		PC 5		PC 6		PC 7		PC 8		PC 9		PC 10	
	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error
Dwelling type - detached																		
Dwelling type – semi-detached	.753	.180	0.513**	.244	1.286	.237	.573	.347	.950	.130	.609	.423	.932	.152	1.088	.151	1.421	.417
Dwelling type – terraced	1.437	.256	.637	.342	1.793*	.298	.409	.641	.841	.174	1.192	.469	.783	.193	1.156	.185	1.801	.445
Dwelling type – apartment	6.920E−08	4871.374	2.530E−08	6701.038	1.477	1.112	2.663E−08	9835.488	1.512	.480	8.915E−08	7.184E+03	1.179	.439	1.106	.459	8.110E-09	0.000
No. bedrooms – 1	1.527E−08	6737.317	1.694	1.149	2.704E−09	0.000	2.572E−08	0.000	.403	.850	4.826E−08	9.858E+03	.709	.617	1.672	.615	4.268E+07	3.534E+03
No. bedrooms – 2	0.227**	.552	.759	.599	0.351*	.591	.538	1.122	1.453	.293	.502	1.280	2.122**	.330	2.448**	.355	4.810E+07	3.534E+03
No. bedrooms – 3	0.236***	.221	.932	.311	.595	.344	.763	.506	1.230	.206	1.781	.780	.968	.277	1.628	.299	3.472E+07	3.534E+03
No. bedrooms – 4	0.370***	.170	.854	.289	.621	.334	1.279	.434	.879	.203	1.669	.780	.787	.280	1.067	.303	7.514E+06	3.534E+03
No. bedrooms – 5+																		
HoH age < 36 years																		
HoH age between 36 and 55 years	1.538	.272	.588	.355	1.127	.402	.840	.437	1.002	.226	1.430E+07	3.142E+03	0.581**	.221	0.607**	.223	.580	.863
HoH age ⩾ 56 years	1.714*	.313	1.737	.382	1.338	.432	.565	.543	1.864**	.244	2.339E+07	3.142E+03	0.592**	.244	.706	.244	1.864	.884
HH composition – live alone	1.451	.382	.900	.462	5.432***	.379	3.915*	.579	2.433***	.224	.746	.885	25.838***	.253	17.687***	.251	3.202	.727
HH composition – live with adults only	.860	.171	1.051	.284	1.761*	.304	1.200	.335	1.053**	.160	1.487	.713	2.293***	.223	2.337***	.216	.620	.700
HH composition – live with adults & children																		
HoH social class – AB																		
HoH social class – C	.960	.182	.719	.299	.647	.288	.591	.346	.824	.184	.779	.823	0.591**	.205	0.691*	.205	1.485	1.081
HoH social class – DE	.980	.221	1.015	.308	0.576*	.320	.715	.419	1.215	.192	1.795	.783	0.571**	.221	0.669*	.221	2.947	1.062
HoH social class – F	2.025*	.412	2.509*	.491	1.341	.676	.724	1.074	1.593	.379	2.793	1.085	2.800**	.398	1.234	.455	13.634*	1.339
Water heating – electric	.761	.248	1.320	.438	.900	.444	0.347**	.401	0.390***	.195	3.362	1.036	.756	.274	0.493**	.241	1.623	1.051
Cooking type – electric	1.157	.173	1.501	.252	1.575*	.267	0.568*	.313	0.713**	.135	.988	.417	.788	.155	0.564***	.151	1.314	.448
Washing machine	.792	1.078	9.862E+05	0.000	1.309	1.084	8.361E+05	0.000	1.511	.590	1.101E+06	1.250E+03	.529	.468	1.078	.509	1.647	1.148
Tumble dryer	1.690**	.229	1.452	.260	0.524**	.219	.690	.343	0.802*	.126	.644	.363	0.477***	.138	0.427***	.136	.649	.349
Dishwasher	1.884**	.251	.712	.234	.868	.234	.836	.388	0.620***	.126	0.341**	.373	0.445***	.142	0.545***	.140	0.399**	.382
Shower (instant)	.823	.164	1.198	.236	.933	.231	1.017	.324	0.783**	.122	1.029	.384	.832	.142	0.749**	.138	.990	.383
Shower (pumped)	1.029	.151	.866	.220	.964	.234	1.271	.299	1.131	.126	1.012	.392	1.065	.154	1.047	.150	.689	.492
Electrical cooker	.943	.295	.793	.499	.488	.476	.733	.445	0.493**	.205	2.490	1.433	0.364***	.288	0.465**	.254	.299	1.079
Heater (plug in convective)	1.155	.154	.900	.208	1.394	.212	1.291	.303	1.009	.123	1.737	.343	.889	.143	.913	.142	1.138	.357
Freezer (stand alone)	1.093	.149	.987	.196	.856	.206	1.004	.288	1.094	.113	1.314	.346	0.596***	.135	.855	.130	0.390**	.397
Water pump	1.209	.165	1.395	.225	.754	.297	1.046	.331	1.096	.140	.790	.475	0.628**	.188	.925	.172	.950	.522
Immersion	.981	.217	.861	.303	1.335	.351	1.549	.406	0.693**	.147	1.010	.492	1.054	.180	.926	.168	0.422*	.470
TV < 21inch	1.026	.158	1.116	.211	.918	.227	1.037	.302	.899	.123	.943	.369	0.568***	.143	0.699**	.139	.645	.381
TV > 21 in.	1.043	.267	1.185	.313	0.584**	.265	3.537*	.743	.777	.157	.860	.451	0.529***	.171	.833	.177	.656	.427
Computer (desktop)	1.233	.158	.888	.199	0.624**	.216	1.367	.307	1.011	.116	0.469**	.376	0.543***	.143	0.677**	.137	.515	.417
Computer (laptop)	1.300	.161	.758	.202	.896	.221	1.338	.322	1.020	.120	0.244**	.445	0.788*	.143	.895	.141	.603	.421
Game console	1.994***	.171	.807	.258	.733	.274	1.086	.329	.878	.145	0.255*	.793	0.651**	.196	0.496***	.192	.272	.808

Base variables: Dwelling type detached; No. bedrooms – 5+; HoH age < 36 years, HH composition – live with adults and children; HoH social class – AB; Water heating – non-electric; and Cooking type – non-electric, no washing machine, no tumble dryer, no dishwas.

*** $P < 0.0$.
** $P < 0.05$.
* $P < 0.1$.

5.3.9. Profile Class 9 (PC9)

Similar to PC5 this class also shows a late evening peak but differs in terms of a much smaller magnitude component to electricity use across a 24 h period. Dwellings with a smaller number of bedrooms were more likely, particularly those with two bedrooms. A HoH age < 36 years was more likely, as indicated by negative association for the other two categories. People who lived alone were also particularly likely to use this class as indicated by strong positive association. It was also likely for people not to use electricity for heating and cooking. Households that did not own appliance types tumble dryers, dishwashers and instant electric showers were also more likely to use this PC as indicated by strong negative association.

5.3.10. Profile Class 10 (PC10)

This class shows a morning peak time use of electricity that continues until lunch time. Households, with HoH age $\geqslant$ 56 years were more likely to use this class as well as those that lived alone although neither were shown to be statistically significant. Electric water heating and cooking was also likely but was not statistically significant. Appliance types that were least likely to be owned by users of this class were: dishwasher and stand alone freezer.

The PC's described above are characterised based on dwelling, occupant and appliance characteristics and have a number of practical applications as introduced in Section 1. For example, electricity demand for new residential developments may be estimated based on knowledge of dwelling characteristics and demographics for a particular area. Similarly, by understanding how electricity is actually used within the home, new tariff structures can be tailored to suit customer lifestyles and new standard load profiles introduced for residential settlement based on ToU within the market. Finally customers that are most likely to use electricity at peak times can be targeted by utilities for demand reduction schemes.

The application of the approach described in this paper is applicable to any smart metering dataset. However, depending upon the usage profile within the electricity market the number of clusters may vary. Furthermore, the Irish smart metering trials collected detailed information on dwelling, occupant and appliance characteristics for each of the participants. It is unlikely that an electricity utility will hold this level of detailed information for each of their customers. However, information such as location (which was excluded from the Irish smart metering trial on anonymity grounds) and building type etc could be used to carry out a similar analysis. Finally, a balance was sought in this research paper between over fitting and producing a series of load profiles that were reflective of the varied manner with which electricity is used within the home.

6. Conclusions

This paper presented a clustering methodology for creating a series of representative electricity load PC's for the domestic sector in Ireland. Clustering methods: k-means, k-medoid and SOM were evaluated against a DB validity index for segmenting the data into disparate patterns of electricity use within the home. SOM proved to be the most suitable and therefore was used to segment the data prior to carrying out any aggregation. In this way characteristic information pertaining to the load profile shape is maintained.

Ten PC's for each day across a six month period were presented thus preserving the diurnal; intra-daily; and seasonality components to electricity use within the home. A multi-nominal logistic regression was then used to link PC's to dwelling, occupant and appliance characteristics. In most cases, individual customer characteristics showed either a positive or negative association with each class indicating which pattern of electricity use was more or less likely to be used within a household. As a result, it is possible to classify customers and the manner with which they use electricity based on their individual characteristics, and without prior knowledge of household electricity consumption.

Acknowledgements

The authors would like to acknowledge the support of Electric Ireland for access to the dataset and to the Higher Education Authority of Ireland for funding (TSR – Strand III and Fiosraigh).

References

[1] European Smart Metering Industry Group (ESMIG). A guide to smart metering. Brussels; 2011.
[2] Navigant Research. Smart meters – global market analysis and forecasts. <http://www.navigantresearch.com/research/smart-meters>; 2014 [accessed 07.07.14].
[3] Hierzinger R, Albu M, van Elburg H, Scott AJ, Łazicki A, Penttinen L, Puente F, Sæle H. European smart metering landscape report 2012. Vienna; 2012.
[4] Sustainable Energy Authority of Ireland (SEAI). Press Release – Full Data from National Smart Meter Trial Published. <http://www.seai.ie/News_Events/Press_Releases/2012/National_Smart_Meter_Trial_Data_Release.pdf>; 2012 [accessed 08.08.13].
[5] Greentech Media Inc and e Meter. Understanding the potential of smart grid data analytics; 2012.
[6] Eurelectric. Metering, load profiles and settlement in deregulated markets system tariff issues working group; 2000.
[7] McLoughlin F, Duffy A, Conlon M. Evaluation of time series techniques to characterise domestic electricity demand. Energy 2013;50:120–30.
[8] Elexon. Load profiles and their use in electricity settlement. <http://data.ukedc.rl.ac.uk/browse/edc/Electricity/LoadProfile/doc/Load_Profiles.pdf> [accessed 15.10.14].
[9] Retail Market Design Service. Standard load profiles. <http://www.rmdservice.com/guidance/standard_load_profiles.htm> [accessed 14.08.13].
[10] Yohanis YG, Mondol JD, Wright A, Norton B. Real-life energy use in the UK: how occupancy and dwelling characteristics affect domestic electricity use. Energy Build 2008;40(6):1053–9.
[11] Firth S, Lomas K, Wright A, Wall R. Identifying trends in the use of domestic appliances from household electricity consumption measurements. Energy Build 2008;40(5):926–36.
[12] Hart M, de Dear R. Weather sensitivity in household appliance energy end-use. Energy Build 2004;36(2):161–74.
[13] Parker DS. Research highlights from a large scale residential monitoring study in a hot climate. Energy Build 2003;35(9):863–76.
[14] Schick IC, Usoro PB, Ruane MF, Hausman JA. Residential end-use load shape estimation from whole house metered data. IEEE Trans Power Syst 1987;3(3):986–92.
[15] Heunis SW, Herman R. A probabilistic model for residential consumer loads. IEEE Trans Power Syst 2002;17(3):621–5.
[16] Capasso A, Invernizzi A, Lamedica R, Prudenzi A. Probabilistic processing of survey collected data in a residential load area for hourly demand profile estimation. In: IEEE/NTUA Athens power tech conf.; 1993. p. 866–70.
[17] O'Doherty J, Lyons S, Tol RSJ. Energy-using appliances and energy-saving features: determinants of ownership in Ireland. Appl Energy 2008;85(7):650–62.
[18] Leahy E, Lyons S. Energy use and appliance ownership in Ireland. Energy Policy 2010;38(8):4265–79.
[19] Parti M, Parti C. The total and appliance-specific conditional demand for electricity in the household sector. Bell J Econ 1980;11(1):309.
[20] Aigner C, Sorooshian C, Kerwin P. Conditional demand analysis for estimating residential end-use profiles. Energy J 1984;5:81–97.
[21] Bartels R, Fiebig R, Garben DG, Lumsdaine M. An end-use electricity load simulation model (DELMOD). Util Policy 1992;2(1):71–82.
[22] Tiedemann KH. Using conditional demand analysis to estimate residential energy use and energy savings. Eur Council Energy Effic Econ Summer Study 2007:1279–83.
[23] Capasso A, Grattieri W, Lamedica R, Prudenzi A. A bottom-up approach to residential load modeling. IEEE Trans Power Syst 1994;9(2):957–64.
[24] Yao R, Steemers K. A method of formulating energy load profile for domestic buildings in the UK. Energy Build 2005;37(6):663–71.
[25] Widén J, Wäckelgård E. A high-resolution stochastic model of domestic activity patterns and electricity demand. Appl Energy 2010;87(6):1880–92.
[26] Richardson I, Thomson M, Infield D, Clifford C. Domestic electricity use: a high-resolution energy demand model. Energy Build 2010;42(10):1878–87.
[27] Widén J, Lundh M, Vassileva I, Dahlquist E, Ellegård K, Wäckelgård E. Constructing load profiles for household electricity and hot water from time-use data—modelling approach and validation. Energy Build 2009;41(7):753–68.
[28] Walker JL, Pokoski CF. Residential load shape modeling based on customer behavior. IEEE Trans Power Appar Syst 1985;7:1703–11.

Please cite this article in press as: McLoughlin F et al. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy (2014), http://dx.doi.org/10.1016/j.apenergy.2014.12.039

[29] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew Sustain Energy Rev 2009;13(8):1819–35.
[30] Aydinalp M, Ugursal VI, Fung AS. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Appl Energy 2002;71:87–110.
[31] Aydinalp M. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Appl Energy 2004;79(2):159–78.
[32] Riddell AG, Manson K. Parametrisation of domestic load profiles. Appl Energy 1996;54(3):199–210.
[33] Pappas S, Ekonomou L, Karamousantas D, Chatzarakis G, Katsikas S, Liatsis P. Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. Energy 2008;33(9):1353–60.
[34] MohammadZadeh S, Masoumi AA. Modeling residential electricity demand using neural network and econometrics approaches. In: The 40th international conference on computers & industrial engineering; 2010. p. 1–6.
[35] McLoughlin F. Characterising domestic electricity demand for customer load profile segmentation. PhD thesis, Dublin Institute of Technology; 2013.
[36] Verdú SV, García MO, Senabre C, Marín AG, Franco FJG. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. IEEE Trans Power Syst 2006;21(4):1672–82.
[37] Cagni A, Carpaneto E, Chicco G, Napoli R, Elettrica I, Torino P. Characterisation of the aggregated load patterns for extra-urban residential customer groups. IEEE Melecon 2004;2004:1–4.
[38] Espinoza M, Joye C, Belmans R, De Moor B. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Trans Power Syst 2005;20(3):1622–30.
[39] Räsänen T, Ruuskanen J, Kolehmainen M. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. Appl Energy 2008;85(9):830–40.
[40] Verdú SV, García MO, Franco FJG, Encinas N, Marín AG, Molina A, Lázaro EG. Characterization and identification of electrical customers through the use of self-organizing maps and daily load parameters. In: IEEE PES power syst conf expo, vol. 2; 2004. p. 899–06.
[41] Chicco G, Napoli R, Piglione F, Postolache P, Scutariu M, Toader C. Load pattern-based classification of electricity customers. IEEE Trans Power Syst 2004;19(2):1232–9.
[42] Dent I, Aickelin U, Rodden T. Application of a clustering framework to UK domestic electricity data. In: UKCI 2011, 11th annu work comput intell. Manchester; 2011. p. 161–6.
[43] Carpaneto E, Chicco G, Napoli R, Scutariu M. Electricity customer classification using frequency–domain load pattern data. Int J Electr Power Energy Syst 2006;28(1):13–20.
[44] Räsänen T, Voukantsis D, Niska H, Karatzas K, Kolehmainen M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87(11):3538–45.
[45] Irish Social Science Data Archive. Data from the Commission for Energy Regulation (CER) – smart metering project. http://www.ucd.ie/issda/data/commissionforenergyregulationcer/; 2012 [accessed 14.08.13].
[46] Commission for Energy Regulation (CER). Electricity smart metering customer behaviour trials (CBT) – findings report Part 1. CER; 2011.
[47] Commission for Energy Regulation (CER). Electricity smart metering customer behaviour trials (CBT) – findings report Part 2; 2011.
[48] MathWorks. MATLAB. p. Version. 7.10.0.499; 2010.
[49] IBM. SPSS statistics. p. Version 22; 2013.
[50] Bidoki SM, Mahmoudi-Kohan N, Sadreddini MH, Jahromi MZ, Moghaddam MP. Evaluating different clustering techniques for electricity customer classification. Electr Power Distrib Netw 2011:1–5.
[51] Chicco G, Napoli R, Piglione F. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 2006;21(2):933–40.
[52] Chicco G, Napoli R, Piglione F. Application of clustering algorithms and self organising maps to classify electricity customers. In: IEEE bologna power tech conference; 2003.
[53] Davies D, Bouldin D. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;PAMI-1(2):224–7.
[54] Tsekouras GJ, Hatziargyriou ND, Dialynas EN. Two-stage pattern recognition of load curves for classification of electricity customers. IEEE Trans Power Syst 2007;22(3):1120–8.
[55] SAGE. Logistic regression. Extension chapters on advanced techniques. <http://www.uk.sagepub.com/burns/website material/Chapter 24-Logistic regression.pdf> [chapter 24].
[56] McLoughlin F, Duffy A, Conlon M. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: an Irish case study. Energy Build 2012;2010.
[57] Lo KL, Zakaria Z, Sohod MH. Determination of consumers ' load profiles based on two-stage fuzzy C-means. In: Proc. 5th WSEAS int conf power syst electromagn compat, vol. 2005; 2005. p. 212–7.
[58] Zainal A, Samaon DF, Maarof MA, Shamsuddin SM. Fuzzy c-means sub-clustering with re-sampling in network intrusion detection. In: 2009 Fifth int conf inf assur secur; 2009. p. 683–6.