Clustering Residential Electricity Consumption Data to Create Archetypes that Capture Household Behaviour in South Africa

Wiebke Toussaint[a, b] iD, Deshendran Moodley[b, c] iD

[a] Engineering Systems & Services Department, Delft University of Technology, Netherlands
[b] Department of Computer Science, University of Cape Town, South Africa
[c] Centre for Artificial Intelligence Research, South Africa

ABSTRACT

Clustering is frequently used in the energy domain to identify dominant electricity consumption patterns of households, which can be used to construct customer archetypes for long term energy planning. Selecting a useful set of clusters however requires extensive experimentation and domain knowledge. While internal clustering validation measures are well established in the electricity domain, they are limited for selecting useful clusters. Based on an application case study in South Africa, we present an approach for formalising implicit expert knowledge as external evaluation measures to create customer archetypes that capture variability in residential electricity consumption behaviour. By combining internal and external validation measures in a structured manner, we were able to evaluate clustering structures based on the utility they present for our application. We validate the selected clusters in a use case where we successfully reconstruct customer archetypes previously developed by experts. Our approach shows promise for transparent and repeatable cluster ranking and selection by data scientists, even if they have limited domain knowledge.

Keywords: clustering, external clustering validation measures, competency questions, daily household electricity patterns, customer segmentation

Categories: • Computing methodologies ∼ Cluster analysis • Applied computing ∼ Engineering

Toussaint, W. and Moodley, D. (2020). Clustering Residential Electricity Consumption Data to Create Archetypes that Capture Household Behaviour in South Africa. *South African Computer Journal 32*(2), 1–34. https://doi.org/10.18489/sacj.v32i2.845

1   INTRODUCTION

Energy planning requires insights into the electricity consumption behaviour of customers to predict long term demand. Unlike commercial and industrial customers who consume electricity predictably, the daily consumption behaviour of residential households is highly variable (Swan & Ugursal, 2009). In South Africa economic volatility, income inequality, geographic and social diversity contribute to increased variability of daily household consumption behaviour (Heunis & Dekenah, 2014). Understanding this variability is important for policy and planning decisions such as tariff design, network planning and operation, and demand response programmes (Yilmaz et al., 2019).

The aggregate consumption behaviour, or representative load profiles, of residential customers has been modelled extensively to yield standard consumption patterns, or *archetypes*, for dominant groups of households that have common attributes (Swan & Ugursal, 2009). These archetypes consolidate expert knowledge and represent the electricity consumption of typical customer classes. They are an essential tool for demand planning, but are difficult and tedious to construct and do not cater for changes in household behaviour. This is a serious limitation that impacts energy demand planning. Daily consumption behaviour can vary drastically for individual households over time (Dent et al., 2014b). In addition, several years may pass before archetypes are updated. The dominant customer classes can become outdated as new groups of households emerge which may not correspond to the current archetypes. An example of this is households in rural areas in South Africa, where thatch roof huts with limited appliances have gradually transitioned to brick buildings with modern appliances, resulting in a significant change in electricity consumption.

Cluster analysis can be used to identify dominant daily energy consumption patterns for different types of households. Current approaches usually aggregate households (Dang-Ha et al., 2017; McLoughlin et al., 2015) or assume that their consumption behaviour is static. This limits them in their ability to create archetypes that consider variability in daily consumption patterns over time. Another challenge is cluster validation. As acknowledged in the data mining community, clustering problems are notoriously difficult to evaluate. Internal validation measures are frequently limited to specific application scenarios, and insufficient on their own (Liu et al., 2010). External evaluation measures can be used instead, but they require ground-truth labels that indicate true clusters (Song & Zhang, 2008). These are often not available because clustering is an unsupervised learning problem and true clusters are typically unknown. Evaluation by visual inspection is thus relied on, but can be biased by the interpretation of the visual representation (Gogolou et al., 2019). In the electricity domain these challenges are evident. While internal metrics are well established, cluster rankings produced by internal metrics yield conflicting results and are usually insufficient for discriminating between different cluster sets (Jin et al., 2017). Cluster selection is thus typically done through visual inspection, which can be time consuming, adhoc, subjective and difficult to reproduce. When archetypes are updated, the visual evaluation process also has to be repeated. External validation measures based on domain knowledge are sometimes used to rank and

guide the selection of a useful cluster set (Xu et al., 2017). However, there are no standard external metrics or guidelines for choosing such measures.

The data mining literature suggests that cluster quality is best evaluated against the specific objectives of the application (Aggarwal, 2015). Even so, data scientists often have limited domain knowledge, which can hinder them from identifying useful clustering structures. The field of ontology engineering provides structured methods for acquiring and representing knowledge from domain experts. One such method, competency questions, is widely used by ontology engineers to acquire application requirements and to compare and evaluate candidate conceptualisations of domain knowledge for a specified context (Grüninger & Fox, 1995). Using an application case study in South Africa, we present an approach that uses informal interviews to derive competency questions, which we operationalise as external evaluation measures to identify a cluster set that represents the most useful daily consumption patterns in our dataset for analysing variability in household behaviour. This cluster set then presents a library of dominant daily consumption patterns that can be used to generate customer archetypes and analyse variability in national residential electricity demand in South Africa.

We build on previous work where we compared and analysed different clustering techniques for generating daily electricity consumption patterns (Toussaint & Moodley, 2019) and developed external clustering evaluation measures from competency questions (Toussaint & Moodley, 2020). We validate our results in a use case where we use the pattern library to create customer archetypes for South African households. The archetypes generated by the approach are compared against equivalent benchmark archetypes developed by experts. We show that combining internal and external cluster validation measures enables the selection of a cluster set that is useful for our application. In particular, we found competency questions to be a promising technique for eliciting and representing application requirements. Our approach has potential to enable transparent and repeatable cluster selection by data scientists with limited domain knowledge.

The paper reviews relevant literature in Section 2, and presents the dataset and clustering experiments in Section 3. In Section 4 we outline our approach to elicit competency questions and formalise application requirements to specify the clustering objective. The clustering results are presented in Section 5. In Section 6 we demonstrate how the pattern library can be applied in a use case to create customer archetypes. Finally, we discuss our findings in Section 7 and conclude in Section 8.

2   LITERATURE REVIEW

This section describes previous work on data processing, algorithms and evaluation measures for clustering electricity consumption patterns, based on the systematic analysis and comparison of clustering approaches in 25 load profile clustering studies published before and in 2018, listed in Table 12 in Appendix A. We discuss internal and external clustering measures used in the domain, present and discuss the limitations of existing work on cluster analysis for constructing customer archetypes, and introduce competency questions as a method to illicit

and specify domain knowledge and application requirements.

2.1   Clustering Residential Load Profiles

Long term energy end-use models in the electricity sector are based on information about customers and the dynamics of change (Feinberg & Genethliou, 2006). Customer behaviour is frequently approximated with load profiles or load curves, which are time-varying electricity consumption patterns. A daily load profile describes the electricity consumption pattern of a household over a 24 hour period. A representative daily load profile (RDLP) characterises the electricity consumption of a customer archetype (Chicco et al., n.d.), e.g. a rural household on a weekday in winter. RDLPs are obtained by aggregating individual household load profiles that share common attributes, such as socio-demographic characteristics and temporal attributes at varying granularity like a year, season, month and daytype.

Table 1: Frequency of use and performance of clustering algorithms for clustering domestic load profiles, based on 25 studies captured in Appendix A Table 12.

Abbreviation	Algorithm	Frequency of use	Best
	k-means	19	4
HC	Hierarchical Clustering	12	2
SOM	Self-Organising Maps	7	2
	kmedoids	4	2
MFTL	Modified Follow-The-Leader	4	2
	fuzzy k-means	4	
	SOM + k-means	3	1
	AKM + HC	3	
	fuzzy c-means	2	
AKM	Adaptive kmeans	1	1
WFAKM	Weighted Fuzzy Averages kmeans	1	1
IRC	Iterative Refinement Clustering	1	1
MKM	Modified kmeans	1	
	SAX k-means	1	
	Spherical k-means	1	
AVQ	Adaptive Vector Quantisation	1	
	DBSCAN	1	
FTL	Follow-The-Leader	1	
GMM	Gaussian Mixture Model	1	
	Random Forests	1	
	Voronoi decomposition	1	
	SOM + HC	1	

Cluster analysis is an unsupervised machine learning approach that partitions data points into groups of similar data points (Aggarwal, 2015). It is widely used in the electricity domain to generate RDLPs and construct customer archetypes. We performed an extensive literature survey to find the most widely used and most effective algorithms for clustering electricity consumption patterns. Table 1 summarises common clustering algorithms, how frequently they have been used to cluster load profiles and how often they were found to be the best algorithm when multiple algorithms were evaluated, based on the 25 studies that we surveyed. 16 studies compared different algorithms, but 4 of them did not identify the algorithm that was best suited for their application. We assume that this was due to challenges with evaluating

clustering structures. For our context and application we concluded that variants of kmeans, self-organising maps (SOM) and hierarchical clustering are the most widely used algorithms. We implemented three of the algorithms: kmeans, SOM, and a combination of the two. We selected the Euclidean distance measure, which was used as similarity measure in the majority of studies and is appropriate in our application where sequences have the same length and a one-to-one mapping between data points.

Clustering residential load profiles has been well explored to understand customer behaviour (Jin et al., 2017), for demand side management (Yilmaz et al., 2019), and to create customer archetypes for tariff development (Chicco et al., 2002) and small scale renewable generation (Xu et al., 2017) in the electricity sector. Many studies from developed countries in the northern hemisphere cluster relatively homogeneous populations, where domain experts expect that the variability in electricity demand is primarily influenced by seasonal and weekday effects, rather than by extreme variance in consumption between households. Splitting the input data along temporal dimensions prior to clustering is thus common in the literature, for example Cao et al. (2013) split data for summer and winter seasons, and Dang-Ha et al. (2017) additionally split data for weekdays and weekends. Xu et al. (2017) cluster highly variable households spread across the United States. They found pre-binning along a consumer demand dimension, first clustering load profiles by overall consumption and then by load shape, an effective method to improve clustering results in this context. We use a similar two-stage pre-binning with integral kmeans to stratify our variable population along a consumer demand dimension before clustering.

2.2   Cluster Evaluation

Despite the practical potential of cluster analysis, evaluating clustering structures remains a challenge for applications in the residential electricity sector. Over 18 different validity metrics were used across the studies we reviewed. Most of them are internal measures, but 3 studies used external measures.

2.2.1   Internal Clustering Validation Measures

Internal metrics rely only on information in the data to measure the quality of a clustering and typically evaluate clusters for their compactness and distinctness (Liu et al., 2010). The five internal metrics that have been used most frequently in the studies we surveyed are listed in Table 2.

Jin et al. (2017) have found that the ranking of experiments can be inconsistent across internal metrics. They also present a trade-off in their ability to capture both the distinctness and compactness of clusters. Dang-Ha et al. (2017) compare a number of metrics including the CDI, MIA and DBI, and have found that these indicators do not penalise the formation of large, noisy clusters sufficiently, while also tending to create unique clusters for outliers. These limitations have been studied extensively in the data mining literature (Liu et al., 2010), where it is well known that internal metrics tend to be biased towards an algorithm, and that

Table 2: Frequency count of the five most used clustering metrics for clustering domestic load profiles, based on the 25 studies captured in Appendix A Table 12.

Abbreviation	Clustering index	Frequency count
DBI	Davies-Bouldin Index	13
CDI	Cluster Dispersion Index	11
MIA	Mean Index Adequacy	11
SMI	Similarity Matrix Indicator	5
SilhI	Silhouette Index	4

outliers affect clustering outcomes (Aggarwal, 2015). Furthermore, a single metric on its own is insufficient to adequately measure the performance of clustering algorithms (Bezdek & Pal, 1998). We selected the Davies-Bouldin Index (Davies & Bouldin, 1979), the Mean Index Adequacy described in Chicco et al. (n.d.) and the Silhouette Index (Jiawei Han et al., 2012) as internal metrics based on their frequency of use in the domain and ease of implementation for the evaluation of a large dataset.

2.2.2   External Clustering Validation Measures

External evaluation measures can be used for additional evaluation when ground truth of the true clusters is known (Aggarwal, 2015). Entropy is a widely used external clustering evaluation measure which intuitively measures how class labels are distributed across clusters. Song and Zhang (2008) define two types of entropy: cluster-based cross-entropy measures the consistency of class labels with respect to clusters; and class-based cross-entropy measures clustering consistency with respect to classes. In the electricity domain Kwac et al. (2014) use entropy as a metric for capturing the daily variability in electricity consumption of households. To evaluate the result of segmenting a large number of daily load profiles into interpretable consumption patterns, Xu et al. (2017) use peak overlap, percentage error in overall consumption and entropy as metrics. We use cluster-based cross-entropy, peak overlap and percentage error to specify the clustering objective for our application. Unlike Kwac et al. (2014) and Xu et al. (2017) who use entropy to measure demand variability within a household, we use entropy to evaluate how consistently clusters are used at specific times and for specific consumption groups.

2.3   From Clusters to Customer Archetypes

To associate consumption patterns with characteristic household attributes, current approaches use a two stage process that first clusters load profiles, and then classifies the resultant representative load profiles according to the socio-demographic characteristics of the households that use them (McLoughlin et al., 2015; Rhodes et al., 2014; Viegas et al., 2016). Viegas et al. (2016) and Rhodes et al. (2014) apply context filtering and cluster the average seasonal load profiles of households to derive seasonal patterns. They then perform binary regression to classify the seasonal load curves based on survey data. The study by Rhodes et al. (2014) only considered a small population of 103 college-educated households, with the majority of house-

holds earning well above the national mean income. The study by Viegas et al. (2016) was much larger and considered 1972 households from the Irish CER dataset. The same dataset was also used by McLoughlin et al. (2015), who used a different approach to cluster the daily hourly load profiles of 3941 households and derive consumption patterns by averaging the load profiles of cluster members. They then captured the consumption pattern used by every household on every day in a Customer Class Index (CCI), and assigned the most frequently occurring pattern of the CCI to each household. Finally, they used multinomial logistic regression to classify the CCI by household attributes.

These approaches for creating customer archetypes capture a very coarse grained temporal dynamic, which does not represent the variability in electricity consumption of households across days, weeks and months. Cao et al. (2013), Jin et al. (2017) and Yilmaz et al. (2019) present approaches that cluster households' individual daily load profiles, and are thus able to observe variability of household electricity consumption across time and across households. Kwac et al. (2014) cluster individual daily load profiles and build a pattern library to characterise individual household consumption as variable or stable, but the study does exclude low consuming households. While these studies aim to create patterns that can be used to create customer archetypes or identify households with particular attributes, they do not extend the work to characterise the patterns and evaluate the extent to which they are able to achieve this. This is a typical challenge, as socio-demographic survey data is often unavailable and costly to obtain. We draw on the work of Jin et al. (2017) to capture consumption variability and McLoughlin et al. (2015) to characterise the selected clustering structure in our case study application.

2.4   Specifying Application Requirements for External Validation

It is commonly known that the performance of clustering algorithms depends on both the data characteristics and the clustering objective (Aghabozorgi et al., 2015). Aligning cluster evaluation measures with the goal of the specific application thus makes intuitive sense, and is important to generate clusters that are useful (Aggarwal, 2015). Ultimately, the clustering process must yield a cluster set that is useful for creating customer archetypes which represent distinguishing socio-demographic attributes of households in the country and the temporal variation in energy consumption. The choice and granularity of these properties are informed by the current practises of experts in the energy sector and the diversity in the population. While some external metrics such as peak overlap, percentage error in overall consumption and entropy have been proposed (see section 2.2.2), visual inspection by domain experts is commonly suggested and applied as an additional validation step (Dang-Ha et al., 2017). Visual inspection has inherent challenges as suggested by Gogolou et al. (2019), who have found that the assessment of time series similarity is subjective and depends on its visual representation. Moreover, the application requirements and expert knowledge that are used to evaluate and guide cluster set selection may be implicit and qualitative and difficult to specify explicitly. It would be useful to evaluate clustering structures against the qualitative re-

quirements of domain experts that understand the nuances of the population being clustered, even if they are not familiar with the technicalities of clustering. To do this, methods for formalising qualitative expert knowledge are required.

The ontology engineering community uses competency questions to acquire context-specific requirements and to compare candidate ontologies (Grüninger & Fox, 1995). Competency questions can be used to represent a set of problems that characterise microtheories in a rigorous manner, enabling more precise evaluation of different conceptualisations of a domain (Fox & Grüninger, 1994). Brainstorming, expert interviews and consulting established sources of domain knowledge can be used to identify competency questions (De Nicola et al., 2009). The techniques for developing competency questions and the questions themselves can be formal or informal. Informal competency questions can be expressed in natural language and connect a proposed ontology to its application scenarios, thus providing an informal justification for the ontology (Uschold & Gruninger, 1996). In this study we propose to use unstructured expert interviews to derive competency questions and elicit application requirements. The requirements are then used to guide the evaluation of clustering structures based on their ability to represent the variable daily load profiles of households for the purpose of creating customer archetypes.

3   LOAD PROFILE CLUSTERING

In this section we compare and analyse different clustering techniques for generating daily electricity consumption patterns in South Africa. We provide an overview of the application case study and the dataset, the experimental setup which includes clustering algorithms, parameters and preprocessing, and internal clustering validation measures.

3.1   Application Case Study and Dataset

The Domestic Electrical Load Metering Hourly (DELMH) dataset (Toussaint, 2019a) is the largest and most comprehensive database of household electricity consumption in South Africa. South Africa's electricity utility uses this dataset for long term residential electricity planning. It contains 3 295 194 daily load profiles for South African households over a period of 20 years from 1994 to 2014 (Toussaint, 2020).

The daily load profile h for a particular household j on a given day d is a 24 element array containing the mean electricity consumption of the household for each hour of the day. For example, the first element, l_0, is the household's mean consumption for the first hour of the day, i.e. 00:00:00 - 00:59:59.

$$h_d^{(j)} = \begin{bmatrix} l_0 & l_1 & \dots & l_{23} \end{bmatrix} \tag{1}$$

$H^{(j)}$ is a two-dimensional array ($n\_days \times 24$) that contains all daily load profiles, $h_d^{(j)}$, for all

days, n_days_j, where data is available for household j.

$$H^{(j)} = \begin{bmatrix} h_0^{(j)} \\ h_1^{(j)} \\ \vdots \\ h_{n\_days_j}^{(j)} \end{bmatrix} \tag{2}$$

Y is a two dimensional array (3 295 194 $\times$ 24) which concatenates $H^{(j)}$ for all 14 945 households observed over the 20 year period.

$$Y = \begin{bmatrix} H^{(1)} \\ H^{(2)} \\ \vdots \\ H^{(14\ 945)} \end{bmatrix} \tag{3}$$

Each row, y_i, in Y represents one of the 3 295 194 daily load profiles, $h_d^{(j)}$, across all households j in the data set. We can then use clustering to find an optimal clustering structure k, given the input dataset Y. A single cluster k_x is representative of individual daily load profiles that capture similar daily energy consumption behaviour. The centroid of k_x is used to construct the representative daily load profile (RDLP) of k_x, which represents the mean daily consumption pattern of all y_i assigned to the cluster. Collectively, the RDLPs of a cluster set represent the dominant daily consumption patterns across all households in the data set and can be used to generate customer archetypes for long term energy modelling applications.

3.1.1   Zero Consumption Values and Outliers

A significant percentage of households in our dataset are low income and rural consumers. In South Africa as in many developing countries, energy access is a priority and must be considered alongside energy security. Through conversations with experts we gathered that the treatment of outlier and extremely low-valued profiles should be reconsidered when clustering electricity consumers in this context. Very low consumption profiles that are close to zero typically belong to consumers living in rural or informal settings. The reasons for low or no consumption are not necessarily due to technical errors, as is assumed in other studies, but because households cannot afford to buy electricity or they choose a different fuel type. The inclusion of these households is important if energy access is a concern.

3.2   Experiment Setup

An experiment run n takes input array Y to produce cluster set $k^{(n)}$ and assigns each normalised daily load profile y_i' to a cluster $k_x^{(n)}$. The RDLP of each cluster, $r_x^{(n)}$, is the mean of all denormalised daily load profiles (i.e. y_i) assigned to $k_x^{(n)}$. A pattern library $\{r_1^{(n)}...r_{x_n}^{(n)}\}$ is the

set of RDLPs for all clusters in $k^{(n)}$. The objective of the load profile clustering experiments is to select the clustering structure $k^{(n)}$ for Y that produces the most useful pattern library for creating customer archetypes. Given the high variance in our dataset, preprocessing was an important component of the clustering process. Different normalisation and pre-binning algorithms were set up for comparison alongside clustering algorithms.

3.2.1   Normalisation

We compared four normalisation techniques from the literature (Table 3) against a baseline with no normalisation.

Table 3: Data normalisation algorithms and descriptions

Normalisation	Equation	Comments
Unit norm	$y_i' = \frac{y_i}{|y_i|}$	Scales input vectors individually to unit norm
De-minning	$y_i' = \frac{y_i - y_i{}^{min}}{|y_i - y_i{}^{min}|}$	Subtracts daily minimum demand from each hourly value, then divides each value by deminned daily total; proposed by Jin et al. (2017)
Zero-one	$y_i' = \frac{y_i}{y_i{}^{max}}$	Scales all values to a range [0, 1]; retains profile shape but is very sensitive to outliers; also known as min-max scaler
SA norm	$y_i' = \frac{y_i}{\frac{1}{24} \times \sum_{t=0}^{23} y_i[t]}$	Normalises all input vectors to mean of 1; retains profile shape but very sensitive to outliers; introduced for comparison, as it is frequently used by South African domain experts

3.2.2   Pre-binning

We implemented two different approaches to pre-bin all daily load profiles in Y. To pre-bin by average monthly consumption (AMC), we selected 8 expert-approved bin ranges based on South African electricity tariff ranges (see Appendix A for ranges). All the daily load profiles $H^{(j)}$ of household j were assigned to one of the 8 bins based on the value of the household's average monthly consumption, $AMC^{(j)}$. Individual household identifiers were removed from Y after pre-binning. AMC for household j over one year is:

$$AMC^{(j)} = \frac{1}{12} \sum_{month=1}^{12} \sum_{d=1}^{month_{end}} \sum_{t=0}^{23} 230 \times h_d^{(j)}[t] \text{ kWh} \qquad (4)$$

Pre-binning by integral k-means followed these steps:

1. For each $h_d^{(j)}$:

   (a) Normalise with unit norm and construct a 24 element vector of the cumulative sum

(b) Append the daily maximum consumption value $h^{(j)}_{d\ peak}$

2. Concatenate all vectors constructed in step 1 into an array Y' (dim 3 295 194 x 25)

3. Cluster Y' into $k = 8$ bins (same as bins created for AMC) with kmeans

4. Assign all y_i in Y to bins to replicate the clustering structure of Y'
   i.e. a daily load profile $h^{(j)}_d$ of household j on day d should share cluster membership with the same daily load profiles in Y and Y'

3.2.3   Clustering Algorithms

Variations of kmeans, self-organising maps (SOM) and a combination of the two algorithms were implemented to cluster Y. The kmeans algorithm was initialised with a range of m clusters. The SOM algorithm was initialised as a square map with dimensions $s_i \times s_i$ for s_i in range s. Combining SOM and kmeans first creates a $s \times s$ map, which acts as a form of dimensionality reduction on Y. For each s, kmeans then clusters the map into m clusters. The mapping only makes sense if s^2 is greater than m. m and s are the algorithm parameters.

3.2.4   Summary of Clustering Experiments

Table 4 summarises the algorithms, parameters and pre-processing steps for each experiment. Each experiment was executed with all normalisation approaches. Experiments with pre-binning were clustered independently in each bin. $Zeros = True$ indicates that zero consumption values were retained in the input dataset.

Table 4: Summary of experiments.

Exp.	Algorithm	Parameters	Pre-bin	Zeros
1	kmeans	$m\{5, 8, 11, ...136\}$	none	True
	SOM	$s\{5, 7, 9, ...29\}$	none	True
	SOM + kmeans	$s\{30, 40, ...90\}, m\{5, 8, 11, ...136\}$	none	True
2	kmeans	$m\{5, 8, 11, ...136\}$	none	False
	SOM	$s\{5, 7, 9, ...29\}$	none	False
	SOM + kmeans	$s\{30, 40, ...90\}, m\{5, 8, 11, ...136\}$	none	False
3	kmeans	$m\{2, 3, ...10\}$	AMC	True
	SOM	$s\{2, 3, 4, 5\}$	AMC	True
	SOM + kmeans	$s\{4, 7, 11, ...20\}, m\{2, 3, ...10\}$	AMC	True
4	kmeans	$m\{2, 3, ...19\}$	AMC	True
	SOM + kmeans	$s\{4, 7, 11, ...20\}, m\{2, 3, ...19\}$	AMC	True
5	kmeans	$m\{2, 3, ...19\}$	AMC	False
6	kmeans	$m\{2, 3, ...19\}$	integral kmeans	True
7	kmeans	$m\{2, 3, ...19\}$	integral kmeans	False

3.3   Internal Evaluation Measures and the CI Score

The Mean Index Adequacy (MIA), Davies-Bouldin Index (DBI) and the Silhouette Index were combined into a Combined Index (CI) score so that clustering performance can be evaluated

across these internal metrics (see Appendix B for details on the metrics). CI is used as a relative index to enable simultaneous interpretation of multiple metrics. Distances between cluster centroids and cluster members were computed using Euclidean distance. The CI is the weighted average Ix score for all bins and calculated as follows:

$$CI = log\Big(\sum_{bin=1}^{bins}\big(Ix_{bin} \times \frac{N_{bin}}{N}\big)\Big) \tag{5}$$

where N_{bin} is the number of daily load profiles in a bin (as specified in Section 3.2.2), and N is the total number of daily load profiles in Y.

$$Ix_{bin} = \begin{cases} \text{undefined} & \text{if } DBI_{bin}, MIA_{bin} \text{ or } Silhouette\ Index_{bin} \leq 0 \\ \\ \dfrac{DBI_{bin} \times MIA_{bin}}{Silhouette\ Index_{bin}} & \text{otherwise} \end{cases} \tag{6}$$

Ix_{bin} is an interim score that computes the product of the DBI, MIA and inverse Silhouette Index in each bin. CI is the log of the weighted sum of Ix_{bin} across all bins. DBI and MIA measure cluster compactness. Both metrics increase as cluster compactness deteriorates, thus increasing Ix_{bin} and CI if this is the case. The Silhouette Index has a range between {-1, 1} and is a measure of cluster distinctness and compactness. The Silhouette Index is close to 1 when clusters are both distinct and compact. The closer the Silhouette Index is to 0, the greater Ix_{bin} and CI become. A lower CI score is desirable and an indication of a better clustering structure. The logarithmic relationship between Ix_{bin} and CI means that CI is negative when Ix_{bin} is between 0 and 1, 0 when $Ix_{bin} = 1$ and greater than 0 otherwise. For experiments with pre-binning, the experiment with the lowest Ix_{bin} score in each bin was selected, as it represents the best clustering structure for that bin. For experiments without pre-binning, $bins = 1$ and $N_{bin} = N$. We weighted the Ix_{bin} of each bin to account for the cluster membership size in that bin.

4   FORMALISING APPLICATION REQUIREMENTS

In this section we describe how we elicited and formalised implicit domain knowledge from experts who understand the application objective but have limited technical knowledge about clustering techniques. We formulated competency questions from domain knowledge and operationalised them as external evaluation measures, which we implemented as a cluster scoring matrix to provide a qualitative ranking of cluster sets in terms of the application requirements and clustering objective.

4.1   Competency Questions and Clustering Objective

We analysed existing standards and conducted unstructured interviews with domain experts to formulate informal competency questions expressed in natural language. The Geo-based Load

Forecasting Standard (2012) is used as design standard by South Africa's electricity utility and contains manually constructed load profiles and guiding principles for load forecasting in the country. The competency questions were developed after analysis of this standard and continuous engagement with a panel of five industry experts. There were initial interviews with all experts to elicit the usage requirements. Preliminary competency questions were presented at a workshop with key stakeholders in the community. The final version of the competency questions incorporated the feedback from the stakeholders. The following five competency questions were identified and expressed in natural language:

1. Does the load shape deduced from clusters represent expected energy demand?
2. Do clusters distinguish between low, medium and high demand consumers?
3. Can clusters represent specific loading conditions for different day types and months?
4. Can a zero-consumption profile be represented in the cluster set[1]?
5. Is the number of households assigned to clusters reasonable, given the sample population?

Based on these questions, we define a good cluster set as having expressive clusters and being usable within the context of the intended application. An expressive cluster must convey specific information related to particular socio-economic and temporal energy consumption behaviour. A usable cluster set must represent energy consumption behaviour that makes sense in relation to the application context, and carry the necessary information to make it pertinent to domain users.

4.1.1 Cluster Expressivity

Current domain knowledge suggests that daily electricity consumption behaviour is strongly influenced by daily routines, seasonal climatic variability and the quantity of electricity consumption (low, medium, high) of a household. Beyond producing patterns that exhibit specific features typically associated with load profiles (question 1), it is desirable that individual clusters convey specific information about the demand profiles of different types of consumers (question 2), on different days of the week and months (question 3). Expressivity thus requires firstly that the RDLP that a cluster produces is *representative* of the energy consumption behaviour of the individual daily load profiles that are members of that cluster. Secondly, clusters must be *specific* to known temporal and consumption contexts, e.g. low demand households on Sundays in June. The choice and granularity of the temporal and consumption features for enumerating the different contexts must be aligned with and support the accepted practise in the expert community for categorising and analysing the daily load demand of residential households.

---

[1]Important for energy access in low income contexts (see Section 3.1.1)

4.1.2   Cluster Usability

The attribute of cluster usability was derived from competency questions 4 and 5. Question 4 is evaluated as being either true or false. Question 5 is calculated as the percentage of clusters whose membership exceeds a threshold value. Moreover, while we anticipate a relatively large number of clusters to represent the large variety of consumers, the following two factors should also be considered:

1. Fewer clusters typically ease interpretation and are thus preferable to larger numbers of clusters

2. The maximum number of clusters is limited to 220, based on population diversity and existing expert models which account for 11 socio-demographic groups, 2 seasons, 2 daytypes and 5 climatic zones

4.2   External Cluster Evaluation Measures

We now translate the clustering objectives into quantifiable external evaluation measures. For *representative* clusters the *mean demand errors* of the *total and peak consumption* values measure the average deviation between the RDLP and the cluster members' load profiles. The *mean peak coincidence ratio* measures the deviation of the peak usage time between the RDLP and the daily load profiles in the cluster. Together these measures express the extent to which a RDLP is representative of the shape and demand of the cluster's member profiles. To measure the degree to which a cluster maps to a $specific$ context, cluster entropy can be used to establish the information embedded in a cluster and thus its specificity. We calculate *day type* and *monthly entropy* to establish *temporal specificity*, and *total* and *peak daily consumption entropy* to establish *demand specificity*.

4.2.1   Mean Demand Error

The total daily demand h_{total} and peak daily demand h_{peak} of a daily load profile are its sum and maximum value respectively. Thus, the total demand $r_{x\ total}^{(n)}$ and peak demand $r_{x\ peak}^{(n)}$ of a RDLP $r_x^{(n)}$ are its sum and maximum value. Likewise, $h_{l\ total}^{x,n}$ and $h_{l\ peak}^{x,n}$ are the total and peak daily demand of the member profiles $h_l^{x,n}$ in $k_x^{(n)}$, with l and index counting through all cluster member profiles. Four error metrics are used to calculate the mean deviation between a RDLP's peak and total demand, and that of the member profiles. Mean absolute percentage error (MAPE) and median absolute percentage error (MdAPE) are well known error metrics. The median log accuracy ratio (MdLQ) (Morley, 2016) overcomes some of the drawbacks of the absolute percentage errors. The median symmetric accuracy (MdSymA) can be interpreted as a percentage error similar to MAPE, making it more intuitive than MdLQ. The equations for calculating the total demand errors are shown below. Both mean and median values are calculated across all $N_x^{(n)}$ cluster members. Equivalent equations are used to calculate peak demand error.

Absolute Percentage Error

$$mape = 100 \times \frac{1}{N_x^{(n)}} \sum_{l=1}^{l=N_x^{(n)}} \frac{|h_{l\ total}^{x,n} - r_{x\ total}^{(n)}|}{h_{l\ total}^{x,n}} \tag{7}$$

$$mdape = 100 \times median_l \left(\frac{|h_{l\ total}^{x,n} - r_{x\ total}^{(n)}|}{h_{l\ total}^{x,n}} \right) \tag{8}$$

Median Log Accuracy ratio

$$Q = \frac{r_{x\ total}^{(n)}}{h_{l\ total}^{x,n}} \tag{9}$$

$$mdlq = median_l \big(log(Q) \big) \tag{10}$$

Median Symmetric Accuracy

$$mdsyma = 100 \times (\mathbf{exp}\,(median_l(|log(Q)|)) - 1) \tag{11}$$

4.2.2   Mean Peak Coincidence Ratio

We defined peaks as all those values that are greater than half the maximum daily load profile value h_{peak}. Peak coincidence is the count that the time of $h_{l\ peak}^{x,n}$ coincides with the time of $r_{x\ peak}^{(n)}$, the peak demand of the RDLP. Mean peak coincidence averages peak coincidence for all member profiles of $k_x^{(n)}$. The mean peak coincidence ratio is the fraction of mean peak coincidence over the count of peaks in $r_x^{(n)}$. It has a value between 0 and 1. The magnitude of the peak is not considered in the mean peak coincidence ratio.

4.2.3   Entropy as a Measure of Cluster Specificity

Entropy S is used to quantify the specificity of clusters and is calculated as follows:

$$S_x(F) = - \sum_{i=1}^{n} p(f_i) \mathbf{log}_2(p(f_i)) \tag{12}$$

F is a feature vector with possible values $f_1, ..., f_n$. $p(f_i)$ is the probability that daily load profiles with value f_i are assigned to cluster k_x. For day type entropy $S_x(daytype)$ expresses the specificity of a cluster with regards to day of the week. Thus $F = daytype$ has possible values $f_i = \{Mon, Tues, Wed, Thurs, Fri, Sat, Sun\}$. $p(Sun)$ is the likelihood that daily load profiles that are used on a Sunday are assigned to cluster k_x. $F = month$ has possible values $f_i = \{January, ..., December\}$ and is used to calculate monthly entropy $S_x(month)$. To calculate

peak and total daily demand entropy, we created percentile demand bins. Thus the possible values of feature $F = peak\_demand$ are $f_i = \{1, ..., 100\}$. $p(60)$ is the likelihood that daily load profiles with peak demand corresponding to that of the 60th peak demand percentile are assigned to cluster k_x. The lower the entropy, the more information is embedded in the cluster, the more specific the cluster, the better the cluster.

4.3   Cluster Scoring Matrix

The external evaluation measures operationalise the clustering objectives and competency questions as quantifiable scores that can be used to rank experiments. We ranked experiments by each measure and weighted the ranks by the relative importance that experts assigned to that measure. We then calculated a cumulative score for each experiment by summing its weighted ranks. The lower the total score, the better the cluster set meets our application requirements. Table 5 summarises the objectives, competency questions, qualitative measures and corresponding weights, which we implemented as a cluster scoring matrix. The total score of an external measure for cluster set k is the mean of the individual measures of all clusters k_x with more than 10490 members[2]. Clusters with a small member size were excluded when calculating mean measures, as they tend to overestimate the performance of poor clusters. Moreover, cluster scores were weighted by cluster size to account for the overall effect that a particular cluster has on the set. For the mean demand error, experiments were ranked against all four error metrics and the mean rank used in the cluster scoring matrix was calculated across all of them.

Table 5: Clustering objectives, competency questions and external clustering evaluation measures

Objective	CQ	External measure		Weight
usable	4	zero-profile representation		1
	5	membership threshold ratio		2
expressive	1	mean demand error	total	6
representative	1		peak	6
	1	mean peak coincidence		3
expressive	3	temporal entropy	day type	4
specific	3		monthly	4
	2	demand entropy	total daily	5
	2		peak daily	5

5   EVALUATION OF CLUSTERING RESULTS

Experiment runs were conducted using the parameter values in Table 4. Each run was first evaluated with the CI score. The best runs of the best experiments were then further evaluated with the external evaluation measures in the cluster scoring matrix. We implemented

---

[2]Threshold selected as a value approximately equal to 5% of households using a particular cluster for 14 days.

our experiments in python 3.6.5 using k-means algorithms from scikit-learn (0.19.1) and self-organising maps from the SOMOCLU (1.7.5) libraries[3].

5.1   Clustering Validation with Internal Measures

The CI scores for all experiments range from 2.282 to 9.627. Lower scores are better. Almost two thirds (65.5%) of experiments have a score below 4. These experiments have been normalised with unit norm, de-minning or zero-one. The remaining experiments have scores above 5 and have not been normalised, or normalised with SA norm. The top 10 ranked experiment runs based on the CI score are shown in Table 6. The percentage point difference between the scores of the first and tenth experiment is only 3.2%, making it difficult to conclusively select an experiment that is useful for our application.

Table 6: Top 10 runs ranked by CI score

Rank	CI	DBI	MIA	Sil.	Exp.	Alg.	m	Norm.
1	2.282	2.125	0.438	0.095	1	kmeans	47	unit
2	2.289	1.616	1.220	0.262	4	kmeans	17	0-1
3	2.296	1.616	1.220	0.260	3	kmeans	17	0-1
4	2.301	2.152	0.485	0.119	5	kmeans	82	unit
5	2.316	2.115	0.447	0.093	1	kmeans	35	unit
6	2.320	2.199	0.486	0.121	4	kmeans	71	unit
7	2.349	2.152	0.481	0.143	6	kmeans	49	unit
8	2.351	2.189	0.434	0.090	1	kmeans	50	unit
9	2.354	2.111	0.476	0.128	7	kmeans	59	unit
10	2.355	2.173	0.453	0.093	1	kmeans	32	unit

Closer analysis of the results confirms that normalisation significantly impacts clustering results. Almost all of the top experiments have been normalised with unit norm, with the exception of two experiments normalised with zero-one. The effects of pre-binning are less clear. Both pre-binning approaches and runs without pre-binning are represented in the top results. Kmeans is the uncontested best clustering algorithm. Four runs belong to Experiment 1 (kmeans, unit norm), but were initialised with different numbers of clusters ($m = \{32, 35, 47, 50\}$). For both the kmeans and SOM algorithms the batch fit time increases linearly with dimensionality. For SOM+kmeans the SOM is used for dimensionality reduction and the dimensions explored are thus considerably greater. This significantly increases experiment run times, as shown in Table 7.

5.2   Clustering Validation with External Measures

The results after external evaluation with the cluster scoring matrix are presented in Table 8 for the top runs of the top experiments. The lower the score, the better. The rank by CI score is shown in the last column for comparison. Despite being ranked 9th by CI score,

---

[3]The codebase is available online at https://github.com/wiebket/delarchetypes

Table 7: Summary of algorithm CI scores and run times

Algorithm	Mean CI score	Mean run time (s)
k-means	2.59	44.79
SOM	4.11	39.42
SOM + k-means	3.17	1498.77

Experiment 7 (kmeans, unit norm) is now ranked 1st. Table 9 shows a detailed view of the cluster scoring matrix, with rankings for individual external measures. The second best run, Experiment 4 (kmeans, unit norm), ranks highly for entropy and demand error measures, but has a poorer peak coincidence ratio. Experiment 5 (kmeans, unit norm) ranks third for most measures. While the top two runs lie only 8 points apart, they comfortably outperform the third best run, which has double the score.

Table 8: Top runs ranked by the total score obtained through validation with external measures

Rank	Score	Exp.	Norm.	Pre-binning	Zeros	CI rank
1	57.0	7	unit	int. kmeans	False	9
2	65.0	4	unit	AMC	True	6
3	117.5	5	unit	AMC	False	4
4	143.5	6	unit	int. kmeans	True	7
5	150.0	1	unit	none	True	1
6	205.0	4	0-1	AMC	True	2
7	208.0	3	0-1	AMC	True	3

Table 9: Cluster Scoring Matrix that ranks experiments by each external evaluation measure

Experiment Normalisation Qualitative measures	Weight	1 unit	3 0-1	4 unit	4 0-1	5 unit	6 unit	7 unit
threshold ratio	2	1	5	3	5	7	4	1
peak coincidence ratio	3	1	7	4	6	2	5	3
peak demand error	6	5.50	5.50	2.00	5.05	4.00	3.00	1.50
total demand error	6	5.00	6.25	2.00	6.00	3.25	3.75	1.00
peak demand entropy	5	5	7	2	6	3	4	1
total demand entropy	5	5	6	1	6	3	4	2
day type entropy	4	4	6	1	6	3	5	2
monthly entropy	4	4	6	1	6	3	5	2
SCORE		150.0	214.5	65.0	205.0	117.5	143.5	57.0

Figure 1 visualises the likelihood ($p(f_i)$) that the clusters of Experiment 7 (kmeans, unit norm) are used on a particular day of the week. This is indicative of the entropy (see Eq. 12) and gives an intuition of the expressivity and usability of the cluster set. The higher the peak of a line, the more likely that profiles assigned to that cluster are used on that day of the week. The lower the peak, the less likely that this is the case. Cluster 15 (C15) is a good example

of a cluster that has a very high likelihood of being used on a Sunday, and a lower likelihood of being used on a Saturday or weekday. This cluster is thus specific to the Sunday day type, which is desirable.
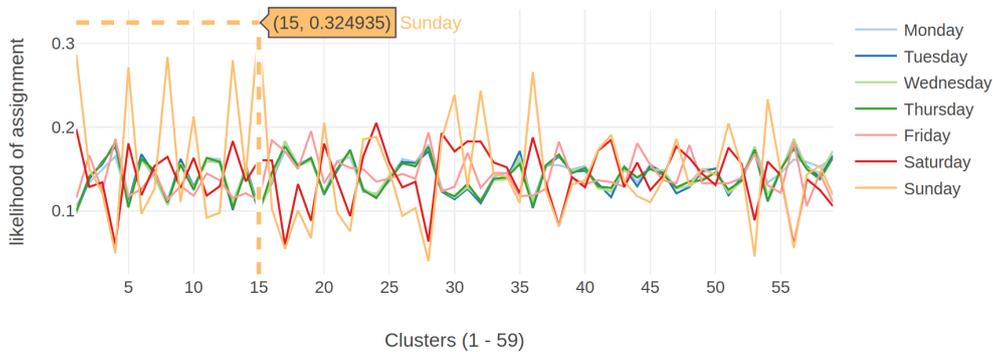

Figure 1: Likelihood of day type assignment for Experiment 7 (kmeans, unit norm): the higher (lower) the peak, the more (less) likely that a cluster is used on that day of the week

5.3   Contrasting Results of Internal and External Measures

The internal metrics provide a useful tool for identifying the most distinct and compact cluster sets, but the CI score is limited for analysis and comparison within the application context. We visually illustrate the strength of the external measures for application-specific evaluation by contrasting the patterns of Experiment 4 (kmeans, zero-one) with those of Experiment 7 (kmeans, unit norm). Experiment 4 (kmeans, zero-one) ranked 2nd based on the CI score, but 6th based on the cluster scoring matrix. Experiment 7 (kmeans, unit norm) on the other hand ranked 9th by CI score, yet ranked 1st after evaluation with external measures. Comparing the patterns in Figures 2 and 3 clearly shows that the latter have greater potential for generating customer archetypes that represent variability in daily electricity consumption behaviour.
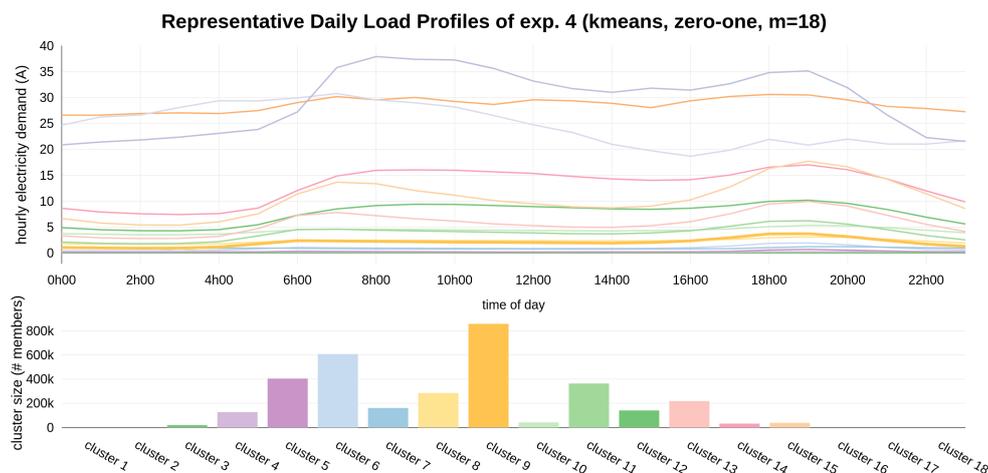

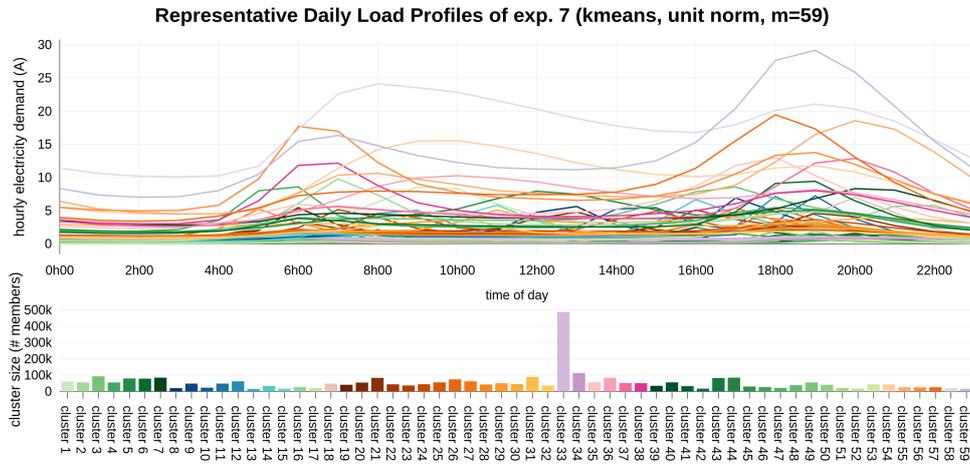Figure 2: RDLPs (top) and cluster membership (bottom) of Experiment 4 (kmeans, zero-one)

Figure 3: RDLPs (top) and cluster membership (bottom) of Experiment 7 (kmeans, unit norm)

Experiment 4 (kmeans, zero-one) has only 18 clusters. The five smallest clusters combined have fewer than 1500 member profiles and appear invisible in the bar chart at the bottom of Figure 2. The ragged shapes of the patterns of cluster 16 (C16), C17 and C18 are an indication that very few profiles were aggregated in these RDLPs. Over half of all load profiles belong to only three clusters: C5, C6 and C9. As a whole, the individual RDLPs lack distinguishing features, making them neither expressive nor useable, and thus poor candidates for creating customer archetypes. Experiment 7 (kmeans, unit norm) on the other hand has 59 clusters. With the exception of C33 which accounts for roughly 15% of all daily load profiles, cluster membership for the remaining clusters varies in a range from 15 000 to 100 000 members. C33 is one of only two clusters in a bin with large membership, due to the high number of low consumption households represented in our dataset. Collectively, the individual patterns are representative and specific, which promises that they will be useful for constructing customer archetypes.

6   APPLICATION OF CLUSTERS TO CONSTRUCT CUSTOMER ARCHETYPES

To validate our selected clustering structure and pattern library for constructing customer archetypes, we benchmark it against customer archetypes currently used by experts in industry. First we describe the benchmark, then we show how our system can be used to create equivalent archetypes. Finally, we illustrate how our pattern library can be used to develop new customer archetypes that capture changing household behaviour.

6.1   Benchmark archetype created by experts

The benchmark customer archetypes represent the aggregated, average electricity consumption of a given type of household, distinguished by its building structure and socio-demographic

characteristics. These archetypes are used by South Africa's electricity utility to model the residential load observed at the medium-voltage substation level. We use the archetype of a lower middle class, long term electrified household in KwaZulu-Natal (KZN), South Africa for demonstration purposes. Figure 4 depicts a customer archetype developed by experts for such a household. KZN lies in the East of South Africa, and subsequently has an earlier sunrise and sunset than most other parts of the country. Work day morning peaks are expected between 5am and 7am, and evening peaks between 5pm and 7pm. The climate is subtropical, with humid summers and warm winters.
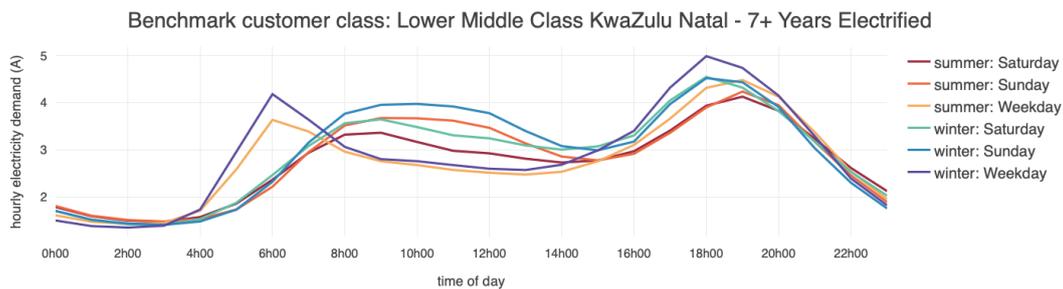

Figure 4: Expert archetype for medium-term electrified lower middle class households in KwaZulu-Natal

Table 10 shows the specific characteristics of different household types identified by experts. The lower middle class household described above will be a household that has piped water access (tap in house), a floor area between 80m^2 and 150m^2 with walls constructed from asbestos, blocks or bricks and a monthly income between R7 800 and R11 600.

Table 10: Attributes of South African residential electricity customer archetypes defined by experts

Archetype	Water	Wall material	Floor area	Income
rural	river/dam	daub/mud/clay	0-50	R0-R1.8k
informal	street taps, tap in yard	corr.iron/zinc	0-50	R1.8-R3.2k
township	tap in house	asbestos, blocks, brick	50-80	R3.2k-R7.8k
lower middle	tap in house	asbestos, blocks, brick	80-150	R7.8k-R11.6k
upper middle	tap in house	brick	150-250	R19k-R24.5k

6.2   Archetype reconstructed with pattern library

We used the clusters from Experiment 7 (k-means, unit norm) to reconstruct the above archetype with a simple multi-class regression model that maps socio-demographic attributes[4] of cluster members to their clusters. To train the model, we created a feature input vector of the socio-demographic and temporal attributes (i.e. day type and season) for each daily load profile belonging to a cluster. The socio-demographic household data was discretised into the ranges recommended by domain experts as shown in Table 10. Each input vector was labelled

---

[4]The socio-demographic attributes were obtained from the DELSKV (Toussaint, 2019b) dataset

with the cluster to which the daily load profile was assigned. The model outputs odds ratios that indicate the likelihood that a particular feature value (i.e. socio-demographic or temporal attribute) is correlated with a cluster. We associated attributes with clusters if the odds ratio was equal to or greater than 1.05. The model was trained with WEKA's[5] multinomial logistic regression algorithm, but any appropriate classification method can be used. The full implementation details and additional archetypes have been presented in previous work (Toussaint, 2019c).
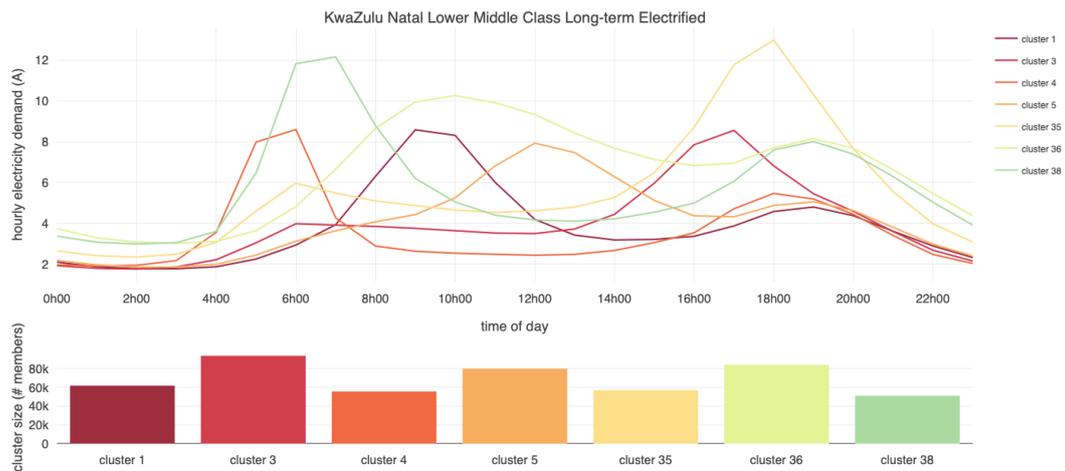

Figure 5: RDLPs (top) and cluster membership (bottom) for medium-term electrified lower middle class households in KwaZulu-Natal

Table 11: Temporal attributes of clusters in Fig 5

Winter Cluster	Daytype	Summer Cluster	Daytype
3	weekday	1	Saturday, Sunday
35	weekday	4	weekday, Friday
36	Saturday, Sunday	5	Saturday, Sunday
38	weekday, Friday		

Seven clusters showed a strong correlation with the socio-demographic attributes of this archetype. Table 11 shows the day type and seasonal attributes of the 7 clusters. Each day type in each season is represented by at least one cluster. Full temporal coverage like this is desirable. Work day and weekend clusters, and winter and summer clusters are mutually exclusive. There are 3 winter weekday clusters (C3, C35, C38), one summer weekday cluster (C4), 1 winter weekend cluster (C36) and two summer weekend clusters (C1 and C5). Figure 5 shows the RDLPs for the clusters. All weekday patterns resemble a typical 'out of home' shape, with either a high morning or evening peak and lower consumption throughout the day. This is expected for a lower middle class household, where adults are typically blue collar workers that have a fixed work routine. The patterns of C1, C5 and C36 show a strong correlation

---

[5]https://www.cs.waikato.ac.nz/ml/weka/

with weekends. C1 and C36 are indicative of a slow starting day when there is no job to rush to. C5 with its peak at 12pm is typical for families that have a strong tradition of a shared family lunch on weekends. The summer weekday pattern of C4 has an earlier morning peak than those of the winter weekday clusters. The weekday patterns of C3, C4, and C35, show an earlier evening peak. With the exception of C3, the winter patterns have a higher energy demand throughout the day than the summer patterns.

As a whole, the patterns of this archetype were found to resemble expected customer behaviour. However, some discrepancies exist in relation to the expert archetype. In contrast to the expert RDLPs in Figure 4, the shapes of our patterns have only one distinct peak, either in the morning or evening. While the peak times correspond between the archetypes, the peak demand values of the expert archetype are approximately half the value of those of our patterns. A plausible explanation for this is that the expert archetype represents the aggregate consumption of a group of households and has only one RDLP for each temporal energy usage context. If we were to aggregate all our patterns for common temporal contexts, for example the three winter weekday RDLPs, the single resultant profile shape and its peak demand would more closely resemble those of the expert archetype.

6.3   Towards an understanding of dynamic household behaviour

The method presented in this paper lays the foundation for analysing changing household behaviour and developing new customer archetypes that can inform long term electricity planning and policy interventions. Consider four hypothetical clusters, C1, C2, C3 and C4. After cluster analysis, in a random week, let us assume that household j is assigned to C1 on day d_1 and d_5, to C2 on d_2, d_3 and d_4, and to C3 on d_6 and d_7, based on its historical consumption data. A year later, the electricity consumption pattern of household j changes on weekdays, and C4 is now assigned to replace C1 and C2 for d_1 to d_5. This shift in electricity consumption behaviour is meaningful. Depending on the pattern represented by C1, C2 and C4, it could, for example, be indicative of a high consumption household installing PV panels and reducing grid demand, or of a low consumption household having purchased additional appliances and becoming a medium consumption household.

The duration and frequency of the shift in consumption is also important and can be used as further evidence to reason about the cause of the change. Do pattern changes happen sporadically, for short periods of time in a season? This may point to extreme weather conditions, like a heat wave. Do they happen consistently at the end of the month? This may be indicative of a household cutting costs due to financial circumstances. Do they coincide with the national lockdown due to Covid-19 that shut down businesses and forced people to stay at home? This could show changes in consumption behaviour that indicate job-loss, or the work-from-home phenomenon. In this manner, external climatic, economic and social circumstances can be linked to the electricity consumption of an individual household using the clusters and pattern library. Exploring this in more detail is an interesting avenue for future work.

7   DISCUSSION

This study combines internal and external validation metrics in the cluster evaluation process to identify a cluster set that represents the daily electricity consumption patterns that best capture variability in behaviour of households in South Africa. As observed in related studies in the electricity domain, we found internal clustering evaluation measures insufficient to capture the nuances and implicit knowledge of domain experts that are needed to identify a useful cluster set. Even though some previous studies have incorporated external validation measures, like entropy, domain knowledge and visual inspection by experts was still required for effective cluster evaluation. Drawing from these studies and the use of competency questions for knowledge elicitation in the ontology engineering community, we operationalised competency questions as external evaluation measures to identify a cluster set that satisfies the expressivity and usability criteria set by domain experts.

We conducted unstructured interviews with experts to identify essential characteristics of daily load profiles and customer archetypes for comparing and analysing different cluster sets. The informal nature of unstructured interviews was a good approach for eliciting expert knowledge, as this facilitated an inherently exploratory process through which the pertinent characteristics of RDLPs emerged. We distilled these characteristics into five competency questions for identifying the required expressivity and usability requirements of the application. The competency questions were highly effective for engaging with domain experts, but they lack intrinsic support for specifying the clustering objective. We therefore introduced a collection of external evaluation measures and a cluster scoring matrix to translate the competency questions into a ranking system for evaluating and comparing cluster sets. Unlike previous studies that conducted secondary evaluation steps informally through visual inspection or with evaluation measures that the authors found interesting, the use of competency questions justified our choice of objective external evaluation measures and grounded them in the application requirements. As a whole, the competency questions made assumptions explicit, the external evaluation measures made the clustering objective explicit, and the cluster selection matrix made it easy to apply and repeat the method. By applying the cluster selection matrix during the evaluation step, a data scientist with limited domain knowledge could produce more useful clusters, with limited involvement of domain experts.

To validate that the clusters selected through our method align with residential load profiles that would be accepted and used by domain experts, we evaluated them in a use case. The use case shows that our clusters can be used successfully to produce customer archetypes that are equivalent to those created by experts. A notable distinction between the RDLPs produced by experts and the RDLPs we derived from our clusters, is that we frequently obtained several patterns for a single day type and season. Each pattern has a distinct shape, peak time and peak demand value. This is a strong indication that our load profiles are more fine grained than the expert ones and better capture the variability in daily consumption of individual households. The profiles of the expert archetypes on the other hand tend to have a morning and evening peak, and lower maximum demand. This is indicative of the aggregate nature

of these patterns, which average electricity consumption over a large number of households with different consumption patterns. We observe that our profiles can reconstruct the essential patterns represented in the expert archetypes, which gives us confidence in their usefulness. Additionally, they offer insights about the variability of individual household demand, which opens the door to understanding changing daily consumption behaviour in the residential electricity sector. To our knowledge this paper presents the first end-to-end approach for creating customer archetypes from electricity consumption and survey data in a highly heterogeneous population. The daily consumption pattern library that we have produced provides a mechanism for domain experts to further study the dynamics of household consumption behaviour in relation to household characteristics, like the transition from low to medium electricity consumption, the shift to off-grid renewable generation and potential connections between volatility in electricity use and vulnerability due to social and economic circumstances.

The competency questions, weights and threshold values are subjective, but as our aim was to formalise the domain knowledge that experts use to select clusters and not to create an objective evaluation process, we do not consider this as a constraint. A limitation of the approach however is that eliciting competency questions through unstructured interviews and existing knowledge artifacts requires the synthesis of a large amount of information, which can be tedious and challenging. The initial time investment for creating competency questions and associating them with external evaluation measures is high, and subject to the experts and knowledge sources consulted. However, as utility companies require ongoing insights into customer behaviour on a quarterly or annual basis, this time investment is well spent, as it reduces inconsistencies and the repeated effort required for manually evaluating and selecting clusters. While the audio recordings and notes taken during the informal interviews were sufficient for gathering expert knowledge, it would be worth exploring alternative approaches for eliciting domain knowledge and distilling it into competency questions. Our approach is promising for similar applications in different geographic locations and adjacent domains such as residential water consumption, though the competency questions and external evaluation measures will need to be adapted to suit their objectives.

8   CONCLUSION

In this paper we use an application case study to illustrate an approach for eliciting and representing expert domain knowledge and application requirements to formalise clustering objectives and guide the evaluation and selection of clustering structures. We conducted unstructured interviews with experts to identify essential characteristics of daily load profiles and customer archetypes, which we distilled into competency questions. The competency questions were operationalised as external evaluation measures and a cluster scoring matrix. By combining internal and external validation measures we were able to evaluate clustering structures against application requirements to select the clustering structure that best represents the variable daily electricity consumption behaviour of South African households. The selected cluster set was used to create a pattern library and generate customer archetypes that we

evaluated against an expert benchmark. Our approach has potential to enable transparent and repeatable cluster ranking and selection by data scientists with limited domain knowledge.

References

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer.

Aghabozorgi, S., Shirkhorshidi, A. S. & Wah, Y. (2015). Time-series clustering—A decade review. *Information Systems*, *53*, 16–38. https://doi.org/10.1016/j.is.2015.04.007

Batrinu, F., Chicco, G., Napoli, R., Piglione, F., Postolache, P., Scutariu, M. & Toader, C. (2005). Efficient iterative refinement clustering for electricity customer classification. *2005 IEEE Russia Power Tech*, 1–7. https://doi.org/10.1109/PTC.2005.4524366

Bezdek, J. C. & Pal, N. R. (1998). Some New Indexes of Cluster Validity. *28*(3), 301–315.

Bidoki, S. M., Mahmoudi-Kohan, N., Sadreddini, M. H., Jahromi, M. Z. & Moghaddam, M. P. (2010). Evaluating different clustering techniques for electricity customer classification. *IEEE PES T&D 2010*, 1–5. https://doi.org/10.1109/TDC.2010.5484234

Cao, H. A., Beckel, C. & Staake, T. (2013). Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns. *IECON 2013—39th Annual Conference of the IEEE Industrial Electronics Society*, 4733–4738. https://doi.org/10.1109/IECON.2013.6699900

Chelmis, C. (2015). Big data analytics for demand response: Clustering over space and time. *2015 IEEE International Conference on Big Data (Big Data)*, 2223–2232. https://doi.org/10.1109/BigData.2015.7364011

Chicco, G., Napoli, R., Postolache, P., Scutariu, M. & Toader, C. (2002). Customer characterization options for improving the tariff offer. *IEEE Power Engineering Review*, *22*(11), 60. https://doi.org/10.1109/MPER.2002.4311841

Chicco, G., Napoli, R. & Piglione, F. (2003). Application of clustering algorithms and Self Organising Maps to classify electricity customers. *2003 IEEE Bologna Power Tech Conference Proceedings*, *1*, 373–379. https://doi.org/10.1109/PTC.2003.1304160

Chicco, G., Napoli, R. & Piglione, F. (2006). Comparison among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, *21*(2), 1–7. https://doi.org/10.1109/TPWRS.2006.873122

Chicco, G., Napoli, R. & Piglione, F. (n.d.). A review of concepts and techniques for emergent customer categorisation. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.7270&rep=rep1&type=pdf

Dang-Ha, T.-H., Olsson, R. & Wang, H. (2017). Clustering methods for electricity consumers: An empirical study in Hvaler-Norway. *NIK-2017*. http://arxiv.org/abs/1703.02502

Davies, D. L. & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

De Nicola, A., Missikoff, M. & Navigli, R. (2009). A software engineering approach to ontology building. *Inf. Syst.*, *34*(2), 258–275. https://doi.org/10.1016/j.is.2008.07.002

Dent, I., Craig, T., Aickelin, U. & Rodden, T. (2014a). An approach for assessing clustering of households by electricity usage. https://doi.org/10.2139/ssrn.2828465

Dent, I., Craig, T., Aickelin, U. & Rodden, T. (2014b). Variability of behaviour in electricity load profile clustering; Who does things at the same time each day? *Advances in Data Mining. Applications and Theoretical Aspects*, 70–84. https://doi.org/10.1007/978-3-319-08976-8_6

Du Toit, J., Davimes, R., Mohamed, A., Patel, K. & Nye, J. M. (2016). Customer segmentation using unsupervised learning on daily energy load profiles. *Journal of Advances in Information Technology*, 7(2), 69–75. https://doi.org/10.12720/jait.7.2.69-75

Feinberg, E. A. & Genethliou, D. (2006). Load forecasting. *Applied mathematics for restructured electric power systems* (pp. 269–285). https://doi.org/10.1007/0-387-23471-3_12

Figueiredo, V., Rodrigues, F., Vale, Z. & Gouveia, J. B. (2005). An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems*, 20(2), 596–602. https://doi.org/10.1109/TPWRS.2005.846234

Fox, M. S. & Grüninger, M. (1994). Ontologies for enterprise integration. *CoopIS*, 82–89.

Gogolou, A., Tsandilas, T., Palpanas, T. & Bezerianos, A. (2019). Comparing similarity perception in time series visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 523–533. https://doi.org/10.1109/TVCG.2018.2865077

Grüninger, M. & Fox, M. S. (1995). The role of competency questions in enterprise engineering. In A. Rolstadås (Ed.), *Benchmarking — theory and practice* (pp. 22–31). Springer US. 10.1007/978-0-387-34847-6_3

Heunis, S. & Dekenah, M. (2014). Manual for Eskom Distribution Pre-Electrification Tool (DPET). https://doi.org/10.25375/uct.7246673.v1

Jiawei Han, Kamber, M. & Pei, J. (2012). *Data Mining Concepts & Techniques* (3rd). Morgan Kaufmann Publishers. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Jin, L., Lee, D., Sim, A., Borgeson, S., Wu, K., Spurlock, C. A. & Todd, A. (2017). Comparison of clustering techniques for residential energy behavior using smart meter data. *AAAI Workshops—Artificial Intelligence for Smart Grids and Buildings*, 260–266.

Jin, L., Spurlock, A., Borgeson, S., Fredman, D., Hans, L., Patel, S. & Todd, A. (2016). Load shape clustering using residential smart meter data: A technical memorandum. https://emp.lbl.gov/publications/load-shape-clustering-using

Kwac, J., Flora, J. & Rajagopal, R. (2014). Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1), 420–430. https://doi.org/10.1109/TSG.2013.2278477

Kwac, J., Tan, C.-W., Sintov, N., Flora, J. & Rajagopal, R. (2013). Utility customer segmentation based on smart meter data: Empirical study. *2013 IEEE Int. Conf. Smart Grid Commun.*, (October), 720–725. https://doi.org/10.1109/SmartGridComm.2013.6688044

Liu, Y., Li, Z., Xiong, H., Gao, X. & Wu, J. (2010). Understanding of internal clustering validation measures. *2010 IEEE International Conference on Data Mining*, 911–916. https://doi.org/10.1109/ICDM.2010.35

McLoughlin, F., Duffy, A. & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. https://doi.org/10.1016/j.apenergy.2014.12.039

Morley, S. K. (2016). Alternatives to accuracy and bias metrics based on percentage errors for radiation belt modeling applications. https://doi.org/10.2172/1260362

Panapakidis, I. P. & Christoforidis, G. C. (2018). Optimal selection of clustering algorithm via Multi-Criteria Decision Analysis (MCDA) for load profiling applications. *Applied Sciences*, *8*(2), 237. https://doi.org/10.3390/app8020237

Ramos, S., Duarte, J. M. M., Soares, J., Vale, Z. & Duarte, F. J. (2012). Typical load profiles in the smart grid context: A clustering methods comparison. *2012 IEEE Power and Energy Society General Meeting*, 1–8. https://doi.org/10.1109/PESGM.2012.6345565

Räsänen, T., Voukantsis, D., Niska, H., Karatzas, K. & Kolehmainen, M. (2010). Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Applied Energy*, *87*(11), 3538–3545. https://doi.org/10.1016/j.apenergy.2010.05.015

Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F. & Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, *135*, 461–471. https://doi.org/10.1016/j.apenergy.2014.08.111

Song, M. J. & Zhang, L. (2008). Comparison of cluster representations from partial second- to full fourth-order cross moments for data stream clustering. *Eighth IEEE International Conference on Data Mining*, 560–569. https://doi.org/10.1109/ICDM.2008.143

Swan, L. G. & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, *13*(8), 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033 excellent overview of residential sector energy modelling!

Teeraratkul, T., O'Neill, D. & Lall, S. (2018). Shape-based approach to household electric load curve clustering and prediction. *IEEE Transactions on Smart Grid*, *9*(5). https://doi.org/10.1109/TSG.2017.2683461

Toussaint, W. (2019a). Domestic Electrical Load Metering, Hourly Data 1994-2014. version 1. https://doi.org/10.25828/56nh-fw77

Toussaint, W. (2019b). Domestic Electrical Load Survey - Key Variables 1994-2014. version 1. https://doi.org/10.25828/mf8s-hh79

Toussaint, W. (2019c). Evaluation of clustering techniques for generating household energy consumption patterns in a developing country. http://hdl.handle.net/11427/30905

Toussaint, W. (2020). Domestic Electrical Load Data Descriptor. https://doi.org/10.25375/uct.11774691.v1

Toussaint, W. & Moodley, D. (2019). Comparison of clustering techniques for residential load profiles in South Africa. *Proceedings of the South African Forum for AI Research*. CEUR-WS.org/Vol-1//Vol-2540/FAIR2019_paper_55.pdf

Toussaint, W. & Moodley, D. (2020). Identifying optimal clustering structures for residential energy consumption patterns using competency questions. *Conference of the South*

African Institute of Computer Scientists and Information Technologists 2020, 66–73. https://doi.org/10.1145/3410886.3410887

Tsekouras, G. J., Hatziargyriou, N. D. & Dialynas, E. N. (2007). Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transactions on Power Systems, 22*(3), 1120–1128. https://doi.org/10.1109/TPWRS.2007.901287

Uschold, M. & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review, 11*, 93–136.

Viegas, J. L., Vieira, S. M., Melício, R., Mendes, V. M. & Sousa, J. M. (2016). Classification of new electricity customers based on surveys and smart metering data. *Energy, 107*, 804–817. https://doi.org/10.1016/j.energy.2016.04.065

Viegas, J. L., Vieira, S. M., Sousa, J. M., Melício, R. & Mendes, V. M. (2015). Electricity demand profile prediction based on household characteristics. *2015 12th International Conference on the European Energy Market (EEM), 2015-Augus*, 0–4. https://doi.org/10.1109/EEM.2015.7216746

Xu, S., Barbour, E. & González, M. C. (2017). Household segmentation by load shape and daily consumption. *Proceedings of ACM SigKDD 2017 Conference*, 1–9. https://doi.org/10.475/123_4

Yilmaz, S., Chambers, J. & Patel, M. (2019). Comparison of clustering approaches for domestic electricity load profile characterisation—Implications for demand side management. *Energy, 180*, 665–677. https://doi.org/10.1016/j.energy.2019.05.124

A   LOAD RESEARCH LITERATURE

Over twenty different algorithms are used for clustering daily load profiles. Many studies have found a particular algorithm to perform better than others. Algorithm abbreviations, their frequency counts and the number of studies that indicate the algorithm as one of the best are listed in Table 1. Best performing algorithms have been denoted with a * in Table 12.

Authors	Input patterns	Customers	Dataset	Notes
Batrinu et al. (2005)	234	234 non-residential	unspecified	IRC developed by authors, builds on Chicco (2003)
Bidoki et al. (2010)	127	127 non-residential	unspecified	performance objective dependent
Cao et al. (2013)		4225 households	Irish CER dataset	k-means pre-binning
Chelmis (2015)		115 buildings	USC campus microgrid data	EVI developed by authors
Chicco et al. (2002)	471	471 non-residential	Romanian national electricity distribution company	
Chicco et al. (2003)		234 non-residential	unspecified	suggests potential for pre-binning
Chicco et al. (2006)		234 non-residential	unspecified	
Dang-Ha et al. (2017)	3 090	3090 households	Hvaler dataset	
Dent et al. (2014a)	180	180 households	NESEMP	clusters & segments by peak time flexibility
Dent et al. (2014b)	204	204 households	NESEMP	assesses variability of energy demand
Du Toit et al. (2016)		11kV & 22kV feeders	Eskom	shows that PCA dim reduction & NUBS centroids improve results & run time
Figueiredo et al. (2005)		165 small consumers	Portuguese Distribution Company	
Jin et al. (2016)	32 611 421	residential	unspecified	
Jin et al. (2017)	104 673	325 households	unspecified	clustering for preprocessing & data reduction to segment customers
Kwac et al. (2013)		85 households	PG&E	investigates consistency of consumption
Kwac et al. (2014)	44 949 750	123 150 households	PG&E	AKM developed by authors
McLoughlin et al. (2015)		3941 households	Irish CER dataset	
Panapakidis and Christoforidis (2018)	365	1 small industrial user	unspecified	develops a cluster algorithm selection framework
Ramos et al. (2012)		208 non-residential	Portuguese Distribution Company	
Räsänen et al. (2010)	3989	3989 small consumers	unspecified	
Rhodes et al. (2014)	103	103 households	Pecan Street Project	
Teeraratkul et al. (2018)	23 254	1057 households	Opower Corporation (Oracle)	
Tsekouras et al. (2007)		94 non-residential	Greek Public Power Cooperation	applies two-stage clustering
Viegas et al. (2015)		1972 households	Irish CER dataset	
Xu et al. (2017)	19 070	residential	Pecan Street Project	best results when applying two-stage clustering

Table 12: Overview of literature on clustering electricity consumers

	Data representation				Data variability	
Authors	**RDLP aggregation**	**Dimensions**	**Interval**	**Norm**	**Time range**	**Spatial cover**
Batrinu et al. (2005)	weekday (spring)	96	15min	[0,1]		
Bidoki et al. (2010)	annual	96	15min	[0,1]	1 year	
Cao et al. (2013)	weekday	48, 18	30min, features	[0,1]	4 weeks	Ireland
Chelmis (2015)	none	96	15min		spring semester Monday	University of Southern California, US
Chicco et al. (2002)	weekday (winter)	4	features	[0,1]	3 weeks	
Chicco et al. (2003)	weekday (spring)	96	15min	[0,1]		
Chicco et al. (2006)	weekday (spring)	2 - 6	features			
Dang-Ha et al. (2017)	summer, winter, weekday, weekend	96	60min	[0,1]	1 year	Hvaler, Norway
Dent et al. (2014a)	weekday evening peak	2 - 8	features	[0,1]		North East Scottland
Dent et al. (2014b)	spring weekday evening peak	48, 42	5min, motifs	[0,1]	3 months	North East Scottland
Du Toit et al. (2016)	none	48, 8	30min, features	standardise	2 summer months x 8 years	
Figueiredo et al. (2005)	summer, winter, weekday, weekend	96	15min	[0,1]		
Jin et al. (2016)	none	24	60min	de-minning	1 year	California, US
Jin et al. (2017)	none	24	60min	de-minning	1 year	California, US
Kwac et al. (2013)	none	96, 24	15min, 60min	unit norm	3 summer/ autumn months	City in San Francisco Bay Area, US
Kwac et al. (2014)	none	24	60min	unit norm	13 months	California, US
McLoughlin et al. (2015)	none	24	60min		6 months	
Panapakidis and Christoforidis (2018)	none	24	60min	[0,1] custom	1 year	
Ramos et al. (2012)	weekday	96	15min	[0,1]	6 months	
Räsänen et al. (2010)		489	features		1 year	Northern Savo, Finland
Rhodes et al. (2014)	summer, autumn, winter spring	24	60min	[0,1]	1 year	Austin, US
Teeraratkul et al. (2018)	none	24	60min	unit norm	22 days	
Tsekouras et al. (2007)		96	15min	[0,1] custom	10 months	
Viegas et al. (2015)	summer, autumn, winter spring	48	30min		18 months	
Xu et al. (2017)	none	96	15min	unit norm	1 month	18 cities, US

Authors	Distance measure	Algorithms	Clustering	
			Cluster range * is best performing	Evaluation
Batrinu et al. (2005)	Euclidean	*IRC, HC, k-means, fuzzy k-means, MFTL	2-15	ScatI, VRC, MIA, CDI
Bidoki et al. (2010)	Euclidean	k-means, *WFAKM, MFTL, SOM, HC		MIA, CDI
Cao et al. (2013)	Manhatten, Euclidean, *correlation, cos	HC, *k-means, SOM + k-means	5- *14	peak overlap, Hamming distance, Wiener filter
Chelmis (2015)	Euclidean, Hausdorff	k-means, HC, k-medoids, Voronoi decomposition	inconclusive	DI, EVI, VRC
Chicco et al. (2002)	w-Euclidean	FTL	7, 9	MIA, CDI
Chicco et al. (2003)	Euclidean, w-Euclidean	k-means, *MFTL, SOM, HC(ward), HC(avg), fuzzy k-means	10-20; inconclusive	MIA, SMI, CDI, DBI
Chicco et al. (2006)	Euclidean, w-Euclidean	k-means, *MFTL, SOM, HC(comp), HC(ward), *HC(avg), fuzzy k-means	10-30; inconclusive	CDI, DBI, MDI, ScatI
Dang-Ha et al. (2017)	Euclidean, cos, Minkowski	k-means, SKM, *SOM, HC (ward), HC(avg), HC(single)	2-50; inconclusive	CDI, DBI, MDI, MIA
Dent et al. (2014a)		k-means	2-10	CDI, DBI, MIA, SMI, BH
Dent et al. (2014b)	Euclidean	k-means, fuzzy c-means, SOM, HC, Random Forests	8	MIA, CDI
Du Toit et al. (2016)	Euclidean^2, DTW, PCC, cos	k-means	*5	DBI, SilhI
Figueiredo et al. (2005)		SOM + k-means	6-9	MIA
Jin et al. (2016)		AKM + HC		DBI
Jin et al. (2017)	Chebyshev, Euclidean	*k-means, *kmedoids, *AKM, *HC(ward), HC(avg), HC(comp), GMM, DBSCAN	10-500	CDI, DBI, MIA, SilhI, SMI, VRSE
Kwac et al. (2013)		AKM + HC		threshold
Kwac et al. (2014)		AKM + HC		threshold, entropy
McLoughlin et al. (2015)		k-means, kmedoid, *SOM	2-16	DBI
Panapakidis and Christoforidis (2018)		k-means, MKM(1), MKM(2), various HC, fuzzy c-means, SOM, others	2-30	CDI, DBI, MIA, SMI, ScatI, VRC, others
Ramos et al. (2012)	Euclidean	*k-means, HC (ward), HC (avg), HC(comp), HC(norm cut)	2-30	DBI, DI, SilhI, J, others
Räsänen et al. (2010)	Euclidean	*SOM + k-means, SOM + HC(comp)	2-30	DBI, IA
Rhodes et al. (2014)	Euclidean	k-means	*2	
Teeraratkul et al. (2018)	Euclidean, *DTW	k-means, *kmedoids, E&M		WCS, WB, WCBCR
Tsekouras et al. (2007)	Euclidean	k-means, AVQ, fuzzy k-means, HC	5-25	CDI, MIA, SMI, DBI, WCBCR, J
Viegas et al. (2015)		k-means	*2-7	DBI, DI, SilhI
Xu et al. (2017)	Euclidean	*k-means, AKM + HC, SAX k-means	3 (stage 1), 4 (stage 2)	peak overlap, consumption error, entropy, WCSS

B   SUPPLEMENTARY TABLES FOR CLUSTERING EXPERIMENTS

B.1   Bin ranges AMC pre-binning

bin	AMC	
1	0 - 1 kWh	no consumption
2	2 - 50 kWh	lifeline tariff - free basic electricity
3	51 - 150 kWh	
4	151 - 400 kWh	
5	401 - 600 kWh	
6	601 - 1200 kWh	
7	1201 - 2500 kWh	
8	2501 - 4000 kWh	

Table 13: AMC bins based on South African electricity tariffs

B.2   Clustering metrics

The Silhouette Index for an individual pattern p in the dataset is:

$$silhouette(p) = \frac{distinctness(p) - compactness(p)}{max\{distinctness(p), compactness(p)\}} \tag{13}$$

Compactness is the average distance between p and all other patterns in the same cluster. Distinctness is the average distance between p and all remaining patterns that are not in the same cluster.

The Davies Bouldin Index (DBI) for two clusters is calculated as the ratio of the sum of cluster dispersions, and the distance between the two cluster centroids.

$$DBI(i, j) = \frac{dispersion(i) + dispersion(j)}{distance(i, j)} \tag{14}$$

Cluster dispersion can be calculated using different measures. A simple method for computing it is as the average distance between the centroid of a cluster and each pattern in the cluster. The DBI for the dataset is obtained by averaging the similarity measure of each cluster and its most similar cluster, $DBI(i, j)_{max}$, for all clusters. A small DBI value indicates that cluster dispersions are small and distances between clusters are large, which is desirable. When plotting the DBI against the number of clusters, the optimal number of clusters can be visually identified. It is possible for the DBI to have several local minima (Davies & Bouldin, 1979).