Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles

Ramon Granell, Colin J. Axon, *Member, IEEE*, and David C. H. Wallom

Abstract—There is growing interest in discerning behaviors of electricity users in both the residential and commercial sectors. With the advent of high-resolution time-series power demand data through advanced metering, mining this data could be costly from the computational viewpoint. One of the popular techniques is clustering, but depending on the algorithm the resolution of the data can have an important influence on the resulting clusters. This paper shows how temporal resolution of power demand profiles affects the quality of the clustering process, the consistency of cluster membership (profiles exhibiting similar behavior), and the efficiency of the clustering process. This work uses both raw data from household consumption data and synthetic profiles. The motivation for this work is to improve the clustering of electricity load profiles to help distinguish user types for tariff design and switching, fault and fraud detection, demand-side management, and energy efficiency measures. The key criterion for mining very large data sets is how little information needs to be used to get a reliable result, while maintaining privacy and security.

Index Terms—Classification algorithms, clustering algorithms, data mining, energy consumption, machine learning, power demand, smart grids.

I. INTRODUCTION

T HE proliferation of smart meters in many countries will enable the creation of time series of power demand in both residential and commercial premises. Most meters will be capable of generating data with high temporal resolution, but customers and network operators have different requirements, expectations, and priorities for using this data. This paper addresses the question of how much information is sufficient. Data mining and pattern recognition techniques have been applied to power demand time series [1]. The application of these techniques can be useful for many purposes such as detecting failures and fraudulent usage or grouping users that present similar consumption patterns [2], [3]. Grouping directly the time series is not a practical task since there may be very long and contain too many unnecessary details. Usually, the information of the time series during a specific period (e.g., season, weekdays of a year) is aggregated in a representative profile that has a fixed temporal duration such as a day [4]–[6].

In this work, the impact of temporal resolution on clustering electricity profiles is investigated. Temporal resolution or granularity refers to the data sampling rate, which may be equal to or less than the acquisition rate by the meter. The maximum acquisition rate is determined by technical parameters of the meter such as its capacity to store the averaged information [7]. In practical terms, high resolution is data measurement at a rate of every few seconds to 30 min. Low-resolution profiles present data over longer time segments, e.g., few hours. The key is to select a resolution that is a trade-off between the level of detail that represents the essential consumption behavior of the user, and the necessity to store and process data.

Little attention has been paid to the effects of data granularity on the performance and efficacy of statistical methods for analyzing power demand profiles. For example, in [5] only the typical temporal resolutions and the existence of techniques to reduce it are mentioned. Clustering experiments have been performed over data with their original granularity, where 15 min is the most typical resolution [5], [6], [8]–[17]. Studies by [18]–[20] were performed using 30-min data, and on 60-min data by [2], [21]. The present authors have used a data set of 1-min resolution to perform clustering using a Dirichlet process mixture model [22]. Other studies investigated dimensional reduction using symbolic aggregate approximation [23], principal component analysis, the curvilinear component analysis and the Sammon map [4].

To understand the advantages and disadvantages of temporal resolution on the clustering of representative load profiles, this paper uses a data set with an 8-s sampling rate and emulates lower resolution data sets. In this way, it is certain that the time-series being used in each experiment has been gathered under the exact same conditions and thus is directly comparable. Three different types of clustering algorithm were employed to perform the experiments: the k-means algorithm, hierarchical algorithms, and the Dirichlet process mixture model algorithm. Several validity measures were employed to evaluate the quality and similarity of the clusters obtained over profiles with distinct resolutions. The efficiency of these algorithms was analyzed with respect to the data resolution.

The paper is organized in the following way. In Section II the process to obtain profiles with different resolution is explained.

R. Granell and D. C. H. Wallom are with the Oxford e-Research Centre, University of Oxford, Oxford, OX1 3QG, U.K. (e-mail: ramon.granell@oerc.ox.ac.uk; david.wallom@oerc.ox.ac.uk).

C. J. Axon is with the School of Engineering and Design, Brunel University, Uxbridge, London, UB8 3PH, U.K. (e-mail: colin.axon@brunel.ac.uk).

In Section III the clustering algorithms and validity indicators employed for the experiments are introduced. The experiments and their results are analyzed in Section IV. Lastly, in Section V, conclusions are drawn.

II. CREATING PROFILES WITH DISTINCT TEMPORAL RESOLUTION

In general, the resolution of the profiles may be different from the meter acquisition frequency. From the acquisition frequency, it is possible to recompute profiles employing different resolutions. Previous studies have used techniques to reduce the dimensionality of the input data set [4], [23]. However, applying these dimensional reduction techniques may produce different profiles than the ones obtained directly from the readings. This may be due to two factors: the nature of the reduction technique [1], or intrinsic data properties such as differing number of readings in each time interval. Although for a single day, this error will be small, the cumulative effect over the large number of days needed to compute the representative profile becomes significant. Therefore the present authors did not employ any technique to reduce the dimension from an established profile, but obtained the values of the profiles for each one of the different resolution are directly obtained from the set of readings.

The daily representative profiles employed in this study were created from the Dehems data set [24]. It comprises electrical power consumption data, with an acquisition rate of 7–8 s, during 2010 from Bulgarian and English dwellings. Each customer record had a different number of readings. After performing a pre-process [22] where records that did not contain enough readings were removed, data for 197 customers was used in this study. The daily profile for each of the customers with temporal resolution T minutes (implying a dimension of $D = (60 * 24)/T$ of the vector of power values $\{v_0, \ldots, v_{D-1}\}$) was directly created from its set of readings in the following way:

- For each weekday j, each value $v_i^{(j)}, 0 \leq i \leq D - 1$ was computed by averaging the readings that were in the temporal interval $i * T$ to $(i + 1) * T$ minutes of the day (there may be not the same number of readings at each interval). This dimensional reduction technique—piecewise aggregate approximation [1]—is analyzed for load profiles in [25].
- Value $v_i, 0 \leq i \leq D - 1$ of the representative profile is obtained by averaging the values $v_i^{(j)}$, for all weekdays (j), that are not equal to zero.
- A linear normalization process, using the respective minimum and maximum of the profile is carried out, obtaining values between zero and one [11].

The temporal resolutions of the daily profiles employed in this work were (in minutes): 0.5, 1, 2, 4, 8, 15, 30, 60, 120, and 240. An example of the same profile of our data set represented with six different resolutions is shown in Fig. 1. The upper profile has the lowest resolution (a value every 2 h, 12 values in total) and the lower profile the highest resolution (one minute i.e., 1440 values). Some details of the profile shape are lost with decreasing resolution, however, these details may or may not be important for comparing and clustering profiles.
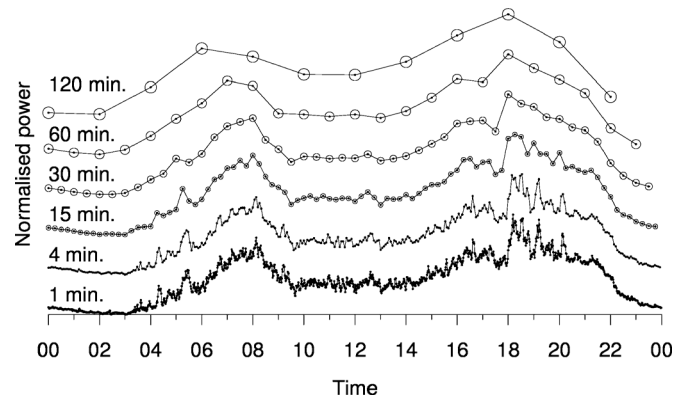

Fig. 1. Example profile with six different resolutions.

III. CLUSTERING ALGORITHMS AND EVALUATORS

The results of clustering depend on both the algorithm and the resolution of the data. The main aim of this paper is not to compare the performance of the algorithms, but to compare the clustering results when the data resolution is varied. Thus, three popular types of algorithm were selected:

- A partitioning algorithm, k-means is one of the most common methods [3]–[5], [8], [11], [12], [17], [18], [21], [22]. From an initial partitioning, a converging process in which data elements are moved from one group to another is carried out until stable partitions are achieved. The convergence of the algorithm depends on the initial partitioning. Therefore such algorithms must be run several times with different initializations.
- Agglomerative hierarchical algorithms [4], [5], [11]–[13], [18], [21]. These bottom-up algorithms create a new cluster for each one of the data elements then successively merge the closest subgroups until the specified number of clusters is achieved. There are different variations depending on the criterion used to compute the distance to merge cluster. If two clusters C_i and C_j are merged to form a new cluster C_q, then the distance of this new cluster C_q from any other existing cluster $C_l, l \neq i \land l \neq j, d(C_q, C_l)$, can be computed in several ways:

— Single link algorithm:

$$d(C_q, C_l) = \min\{d(C_l, C_i), d(C_l, C_j)\} \quad (1)$$

— Complete link algorithm:

$$d(C_q, C_l) = \max\{d(C_l, C_i), d(C_l, C_j)\} \quad (2)$$

— Unweighted pair group method average algorithm (UPGMA):

$$d(C_q, C_l) = \frac{|C_i| * d(C_l, C_i) + |C_j| * d(C_l, C_j)}{|C_i| + |C_j|} \quad (3)$$

where $|C_a|$ is the number of elements of cluster C_a.

— Weighted pair group method average algorithm (WPGMA):

$$d(C_q, C_l) = (d(C_l, C_i) + d(C_l, C_j)) * 0.5 \quad (4)$$

— Unweighted pair group method centroid algorithm (UPGMC):

$$d(C_q, C_l) = \frac{|C_i| * d(c_l, c_i) + |C_j| * d(c_l, c_j)}{|C_i| + |C_j|}$$
$$- |C_i| * |C_j| * \frac{d(c_i, c_j)}{(|C_i| + |C_j|)^2} \quad (5)$$

where $d(c_a, c_b)$ is the Euclidean distance between the centroids of the clusters C_a and C_b

— Weighted pair group method centroid algorithm (WPGMC):

$$d(C_q, C_l) = \frac{d(c_l, c_i) + d(c_l, c_j)}{2} - \frac{d(c_i, c_j)}{4} \quad (6)$$

— Ward or minimum variance algorithm (WARD):

$$d(C_q, C_l) = ((|C_l| + |C_i|) * d'(C_l, C_i)$$
$$+ (|C_l| + |C_j|) * d'(C_l, C_j)$$
$$- |C_l| * d'(C_i, C_j)) * (|C_l| + |C_i| + |C_j|)^{-1} \quad (7)$$

where $d'(c_a, c_b) = |C_a| * |C_b| * d(c_a, c_b) * (|C_a| + |C_b|)^{-1}$

- Bayesian non-parametric statistics: the Dirichlet Process Mixture Model (DPMM) [22], [26]. The DPMM algorithm creates a separation that best adapts to the nature of the data with a hierarchical model of Dirichlet and Multinomial distributions. Profiles are represented as draws from a multinomial distribution whose parameters are obtained from a Dirichlet distribution of dimension D (number of readings per 24-h demand profile). Clusters are computed with the Chinese restaurant process [26]. Contrary to what happens with the two previous types of algorithms, the number of clusters is not an input parameter for the DPMM. A Gibbs sampling process is used to estimate the concentration parameter β of the Dirichlet distribution.

A. Validity Indicators

It is necessary to separately evaluate the different algorithms performance in terms of the quality of the clustering, and the robustness of the cluster composition (cluster membership). Assessing the quality of the clusters is an internal process, while assessing the consistency of the cluster membership requires an external reference with which to compare.

1) Internal Evaluators: The present authors selected six internal evaluators [27]: the clustering dispersion indicator (CDI), Davies-Bouldin index (DBI), modified Dunn Index (MDI), mean index adequacy (MIA), scatter index (SI), and variance ratio criterion (VRC). As these evaluators use the intra-point distance, each set of clustering results was evaluated using the average of the evaluator computed over the profiles with all possible resolutions.

2) External Evaluators: This set of evaluators has not been used in previous work on clustering electricity data, but they were employed in other clustering studies [27], [28]. Being a set of n data elements $S = \{o_1, \ldots, o_n\}$ and two clustering results

over this set, $C = \{C_1, \ldots, C_k\}$ and $C^* = \{C_1^*, \ldots, C_{k^*}^*\}$, where

$$\bigcup_{1 \le i \le k} C_i = \bigcup_{1 \le i \le k^*} C_i^* = S \quad (8)$$
$$\forall C_a, C_b \in C, a \ne b \to C_a \cap C_b = \emptyset \quad (9)$$
$$\forall C_a^*, C_b^* \in C^*, a \ne b \to C_a^* \cap C_b^* = \emptyset. \quad (10)$$

$C(o_i)$ and $C^*(o_i)$ indicate the cluster that contains data element o_i in C and C^*, respectively. The external evaluators that will determine the similitude of C employing as a reference C^* define the following:

- True positives (TP) as the number of pairs of elements that are in the same cluster for both C and C^*. Let $\delta(\text{cond})$ a function that is one if the condition *cond* is fulfilled, zero otherwise

$$\text{TP} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(C(o_i) = C(o_j) \wedge C^*(o_i) = C^*(o_j)). \quad (11)$$

- True negatives (TN) as the number of pairs of elements that are not in the same cluster for both C and C^*:

$$\text{TN} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(C(o_i) \ne C(o_j) \wedge C^*(o_i) \ne C^*(o_j)). \quad (12)$$

- False positives (FP) as the number of pairs of elements that are not in the same cluster in C^*, but are in C:

$$\text{FP} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(C(o_i) = C(o_j) \wedge C^*(o_i) \ne C^*(o_j)). \quad (13)$$

- False negatives (FN) as the number of pairs of elements that are in the same cluster in C^*, but are not in C:

$$\text{FN} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(C(o_i) \ne C(o_j) \wedge C^*(o_i) = C^*(o_j)). \quad (14)$$

The external evaluators used in this study were:
- Rand Index represents the pairs that are correctly clustered over the total of pairs:

$$\text{Rand Index} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100. \quad (15)$$

- Pair-counting precision accounts for the pairs that are together in C:

$$\text{Pair-counting precision } (P) = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100. \quad (16)$$

- Pair-counting recall accounts the pairs that are together in the reference C^*:

$$\text{Pair-counting recall } (R) = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100. \quad (17)$$

- Pair-counting F-score combines the precision and recall:

$$\text{Pair-counting F-score} = \frac{2 \cdot P \cdot R}{P + R}. \qquad (18)$$

These external evaluators operate as a percentage, 100% being the best outcome. In the present experiments, the results obtained with all other analyzed resolutions for the same number of clusters ($k = k^*$) are individually selected as the reference C^*. The mean is computed to obtain the final evaluator score.

IV. EXPERIMENTS AND RESULTS

Experiments were performed combining the proposed algorithms with the ten different resolution versions of the data set. The algorithms were implemented in C++. Experiments were performed on an Intel Core2 Quad CPU Q9650 at 3.00 GHz with 4 Gb of memory.

For the k-means experiments, results are averaged across one hundred repetitions with the same configuration (k and resolution) employing random initialisation for the centroids of the clusters. Euclidean distance was the distance used to compute the distance between profiles and centroids. The hierarchical algorithms used an extension of the Euclidean distance applied between data point and set of data points, and then between two sets of points [2], [11]. The number of clusters (an input parameter) used for both algorithms ranged from two to twenty.

For the DPMM algorithm, the number of clusters is not fixed by the user. The unique user parameters is the concentration parameter, β, that has an influence in the number of clusters and their composition. Therefore the consequences of changing the resolution of the input data in the resulting number of clusters for same values of β was investigated. The different values scanned for the concentration parameter ranged from 10^{-11} to 10^{-2}. The effect of the resolution on the number of clusters obtained when using the same values for the concentration parameter is shown in Fig. 2. Larger numbers of clusters require a very much greater value of the concentration parameter (N.B. log scale, Fig. 2). Furthermore, as the resolution decreases, generally the value of the concentration parameter also increases. This is due to the nature of the statistical algorithm that employs a combination of multinomial and Dirichlet distributions, and the Chinese Restaurant Process (see Section III). Any value of the concentration parameter leads to a different number of clusters for any chosen resolution (data dimension D). This demonstrates a clear dependence on resolution which is not discernible from the algorithm structure. This means that the DPMM is not robust with data resolution.

A. Clustering Quality

As this work is focused on the evolution of the results when the resolution is modified, values of each of the internal evaluators are averaged for the clustering results obtained with all the possible number of clusters for a particular resolution. In the case of the DPMM, results are averaged only over results with the number of clusters obtained for each resolution.

The values for three of the internal evaluators for the different algorithms are shown in Fig. 3. The k-means algorithm
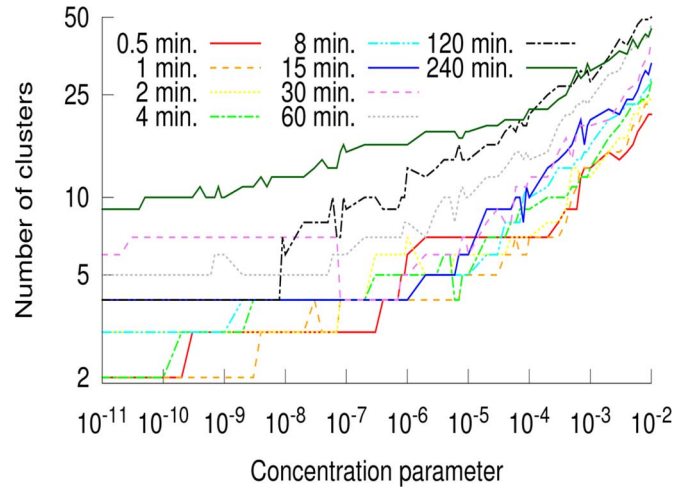

Fig. 2. Number of clusters obtained with the DPMM depending on resolution and concentration parameter.

performed best over the 4- to 60-min resolutions, but worst for the extreme resolutions. This tendency can be clearly seen for the VRC and MDI evaluators. The hierarchical algorithms do not show such a clear behavior.

The WARD distance presents the best scores for 8-min resolution. Best results for the UPGMA distance are in 4, 8, 15, and 30 min, depending on the evaluator. The other distances presented heterogeneous behavior from a flat pattern such as the UPGMC and WPGMC using the VRC evaluator, to a saw-tooth shape by the WPGMA using the MDI evaluator. A general tendency common for all these hierarchical algorithms is that they seem to start declining from 60- to 120-min resolution data results, obtaining usually the worst scores for 240 min. Scores with 0.5- and 1-min resolution are usually slightly worse than the ones obtained with 15- or 30-min resolution, but not for all the evaluators and distances. The DPMM also displayed heterogeneous behavior; sometimes a linear behavior such as for MIA, but was irregular for the MDI evaluator. The best results for the VRC evaluator are with data with 1- or 2-min resolution.

The other three internal evaluators (CDI, DBI, and SI) show similar behavior for all the algorithms, but not in such a marked way. Generally, the hierarchical algorithms presented a flat shape until 60-min data, then showing declining performance. The k-means showed again a U-shape behavior for the MIA and MDI evaluators. The DPMM did not present a clear pattern.

B. Cluster Membership Consistency

The evolution of the cluster membership with respect to other resolutions is the important part of this analysis. The values for the external validity measures are shown in Fig. 4 (averaged for all numbers of clusters). There are three algorithms that show high scores for all the evaluators and resolutions: the single link, UPGMC and WPGMC. This is due to two reasons. First, these hierarchical algorithms tend to create disproportional clusters with respect to the number of elements. There is always a very large cluster with the other clusters formed from very few data elements, which ensures these scores since most of the points are always together in the same cluster. Secondly, the algorithms criteria to merge two clusters is robust to changes in the data
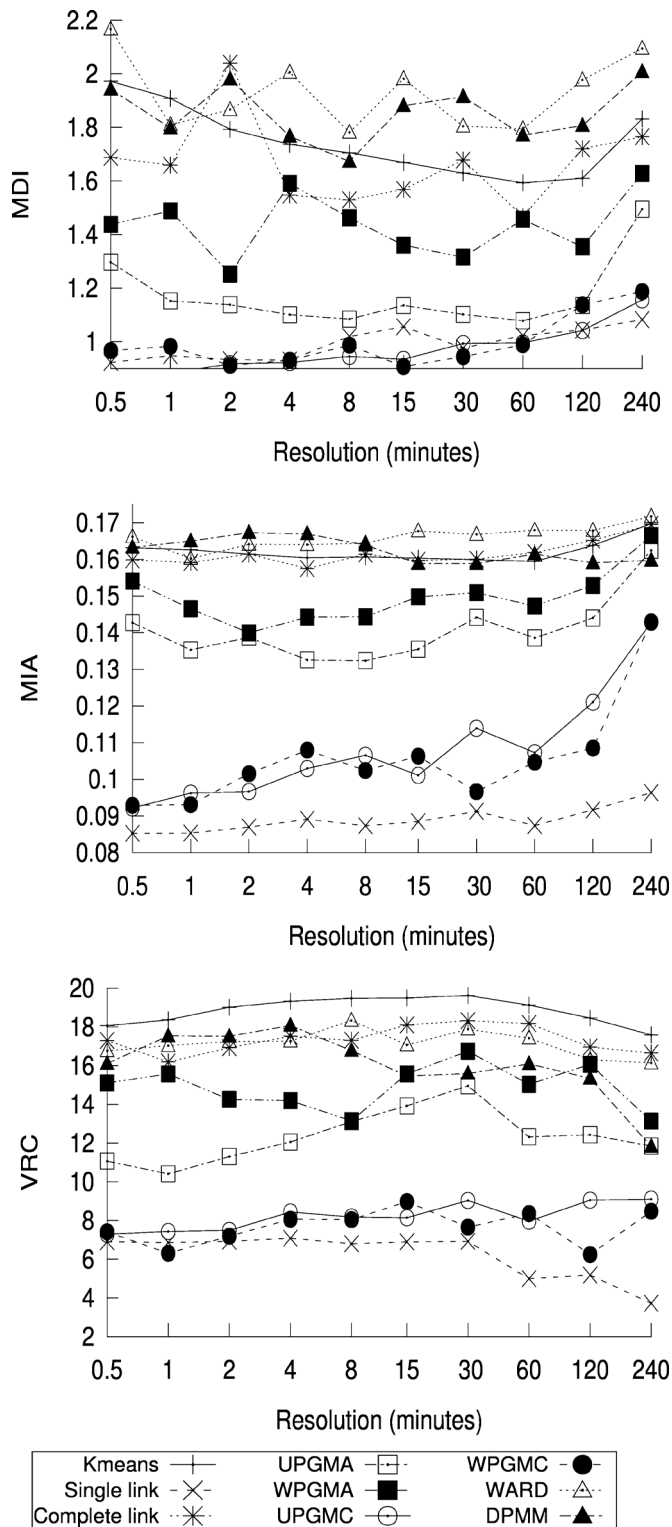
Fig. 3. Values for the internal evaluators for all the algorithms. For all except the VRC, lower values suggest better clustering result.
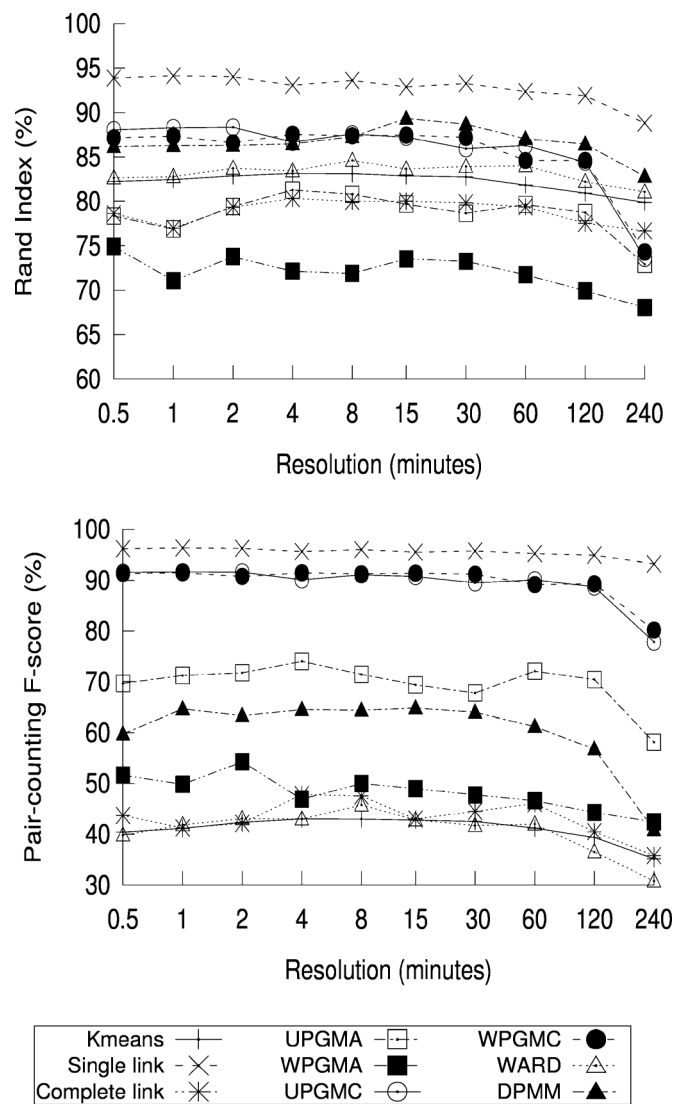

Fig. 4. Values for the external evaluators for all the algorithms.

points, and the smaller ones with three and one elements. In this case, the clusters of eight profiles is the same for both results being dissimilar only in the smaller ones and in one or two elements of the largest cluster. Therefore, the number of FN and FP is very small; similarly for other resolutions. For the other algorithms, DPMM and UPGMA present middling values and the k-means, complete link, UPGMA, WPGMA and WARD the lowest values. For all these algorithms the better scores are obtained for higher resolutions, with the worst results for resolutions greater than 1 h. The behavior of the evaluators is in line with previous work [4], [5], [11].

Values of the F-score (by number of clusters) for four of the algorithms are shown in Fig. 5. The profiles with resolutions closer to the reference resolution obtain more similar clusters (i.e., better F-score) than profiles with lower or higher resolutions. For the hierarchical algorithm UPGMA, there are two clear states of operation, switching at eight to eleven clusters depending on the resolution. The reason is that these agglomerative algorithms tend to form one large cluster that contains most of the data points. For the references, this large cluster is

rate. For example, results for the UPGMC algorithm for a resolution of half a minute with $k = 4$ is composed of a large cluster with 187 load profiles, another one formed with 8 profiles and the last two with just an element each one. Clustering with the same algorithm and number of clusters for a resolution of 15 min produce a cluster of 185 load profiles, another with 8 data
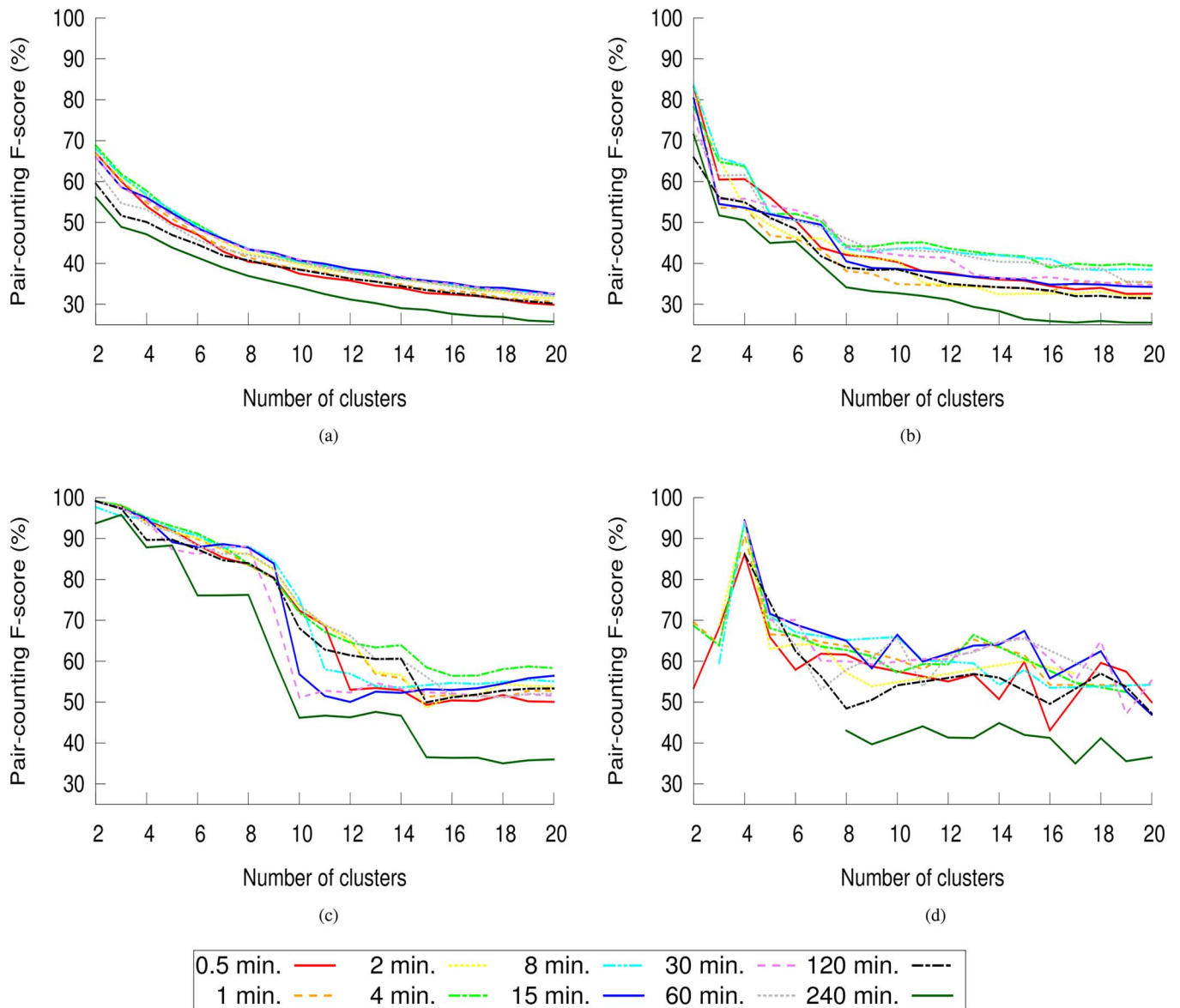
Fig. 5. Pair-counting F-score depending on the number of clusters for four clustering algorithms. (a) K-means. (b) Complete link. (c) UPGMA. (d) DPMM.

also formed. Therefore, for fewer than eight clusters, both the reference and the compared resolution clusterings present this large cluster and a higher score. Similarly for the WPGMA algorithm, but the transition is at four clusters.

What it means for these algorithms is that decreasingly the number of clusters, results may be different until a particular point that all the results present the large cluster. A similar effect but not with one unique large cluster but with two clusters, happens for the Complete link algorithm at seven number of clusters. The k-means algorithm presents continuous decreasing scores with the number of clusters. A large number of clusters implies more combinations and fewer elements together (true positives) than for the reference case. The DPMM shows irregular behavior (with gaps in the results) because not all numbers of clusters are obtained for both the reference and compared resolutions.

Table I shows the average pair-counting F-scores over all of the algorithms used and all possible combinations of resolu-

TABLE I
SYMMETRIC MATRIX WITH AVERAGED F-SCORE FOR ALL THE ALGORITHMS
AND ALL POSSIBLE COMBINATIONS OF RESOLUTIONS

Res	0.5	1	2	4	8	15	30	60	120	240
0.5	-									
1	76	-								
2	74	75	-							
4	68	69	70	-						
8	66	67	69	73	-					
15	64	64	65	69	72	-				
30	62	62	64	67	69	71	-			
60	62	62	63	65	67	67	69	-		
120	60	60	61	61	63	63	64	69	-	
240	53	53	53	54	55	55	56	58	60	-

tions. Although there is some variability, this symmetric matrix shows that similar resolutions obtain more similar results. It also shows that the results obtained with highest resolutions (0.5, 1, and 2 min) are more similar than with the lower resolutions.
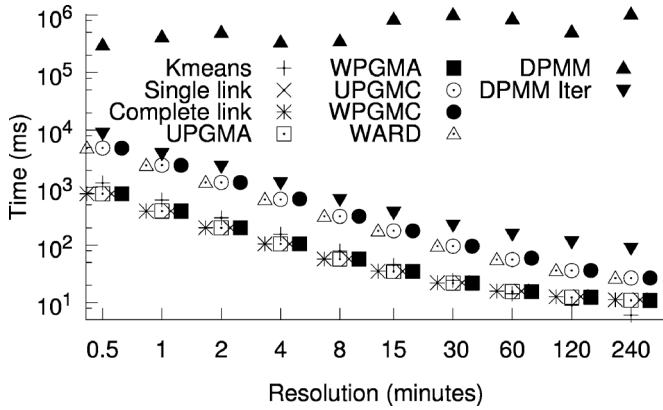
Fig. 6. Averaged time for the algorithms to cluster the *real* data set with different resolutions. Some points are shifted on the x-axis for improved visualization.
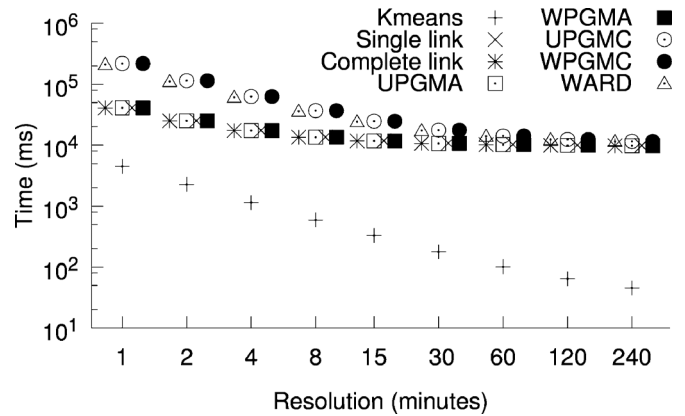

Fig. 7. Averaged time for the algorithms to cluster the *synthetic* data with different resolutions. Some points are shifted on the x-axis for improved visualization.

The high-resolution clustering results have a F-score of about 70% (or better), while results for over 60-min resolution show less than 70%. Results obtained for the 240-min resolution are always less than 60%, indicating that they present very different cluster memberships. In general, the F-score values are quite high since the single link, UPGMC and WPGMC algorithms always achieve more than 90%. For other algorithms e.g., k-means there is just one value over 50%, but the same trends that appear in Table I were observed.

C. Efficiency Analysis

The efficiency of each algorithm depends on its nature and implementation. Theoretically, the order of time complexity of the k-means is $\mathcal{O}(n \cdot k \cdot D \cdot I)$, where n is the number of data points, k is the number of clusters, D is the resolution of the data, and I is the number of iterations to converge the algorithm. For the single link, complete link, WPGMC, WPGMA hierarchical algorithms the complexity is $\mathcal{O}(n^2 \cdot D)$. The rest of hierarchical algorithms (UPGMC, WPGMC, and WARD) present an order of complexity of $\mathcal{O}(n^3 \cdot D)$. The complexity of the DPMM algorithm is $\mathcal{O}(I^* \cdot n \cdot k^* \cdot D^2 \cdot c^*)$, where I^* is the number of iterations to converge the algorithm, k^* is the maximum number of clusters during all the iterations, and c^* is the maximum of the counters of the multinomial distribution (see Section III).

The execution time experiments for each of the algorithms are averaged for the same value of time resolution (Fig. 6). For all of the algorithms except the DPMM, execution time decreases with decreasing resolution, i.e., decreasing of D. The decreases in execution time and resolution are approximately linear over two orders of magnitude. The DPMM algorithm is the slowest method for all resolutions. This is because the most important factor for this method is the number of iterations (I^*) that it takes to converge the algorithm, which rises as the resolution decreases. The stop criterion of the algorithm is that the elements of the cluster stay the same for 25 consecutive iterations, or that 500 iterations have occurred. Thus for low resolutions (more clusters) this criterion takes more time to be achieved than for high-resolution data.

The averaged time for iteration of the DPMM algorithm (points labelled with "DPMM iter" in Fig. 6) behaves similar to the other algorithms. The group of algorithms that are faster is formed by the single link, complete link, WPGMC, WPGMA, and k-means algorithms. The other hierarchical algorithms are slower because the representative centroids of the clusters need to be computed.

1) Performance Tests: To analyse the performance over larger data sets, synthetic data were generated. Each profile was obtained by simulating the electricity consumption of a dwelling during one year based on probabilistic occupancy, lighting and appliances models [29]. Each one of the dwellings has a particular occupancy (from one to six people) and a number of appliances with different power demands. The probabilities of the models are based on real data [29]. Data are generated with one minute resolution and lower resolutions were obtained by averaging the values of each interval.

The data sets generated ranged between 30 and 5000 profiles, and the clustering algorithms were applied to obtain clusters with k from two to twenty. The results were averaged for all the instances that employ the same resolution. For the k-means algorithm, results were also averaged over ten experiments with different initialization.

Fig. 7 displays the running times for all the algorithms, except the DPMM algorithm owing to its high complexity. The range of times for all the algorithms is consistent with those using real data (Fig. 6). The difference in complexity of the algorithms can be seen in Fig. 7, and for all hierarchical algorithms the improvement in performance levels out for data set with 30-min resolution and greater, where convergence is observed. However, for k-means the improvement was continuous across the whole range of resolutions used. In addition, the k-means was 1–2 orders of magnitude faster than the hierarchical algorithms.

V. CONCLUSION

This work analyses the impact of the temporal resolution when clustering electricity load profiles. Several algorithms have been systematically tested by changing the resolution of the input data (of real household consumption). The results are evaluated with internal and external validity measures, and the efficiency was computed using a large-scale synthetic data set.

The quality of the clustering was measured using a variety of internal evaluators. Some hierarchical algorithms, notably single link, WPGMC, and UPGMC derived unbalanced clusters. This shows that these algorithms might only be suitable for clustering profiles when the question posed about the data set is about which behavior types do not belong to a defined majority group. The reliability of the cluster membership was measured using the Rand Index and Pair-counting F-score external evaluators. A band of temporal resolutions between 4–60 min showed better performance than either of the extreme resolutions. The k-means algorithm was shown to be robust to data resolution effects in this this temporal band. The efficiency was tested using synthetic data with up to 5000 profiles which showed that for high resolution data there was a difference in execution time of as much as two orders of magnitude. The k-means algorithm was faster. The DPMM displayed interesting properties that were strongly dependent on the resolution of the data. It was the slowest, but usefully it did not need to specify the number of clusters *a priori*. We conclude that it is used best as a way to establish the number of clusters, but switch to using k-means for the remainder of the analysis.

The implication of the work presented is that to be useful to electricity retailers in discerning differences between consumers time series power use data needs to be at a frequency of least 30 min and ideally 8 or 15 min. Distribution network operators may have recourse to higher resolution data for various applications, however, 8-min data would provide a useful basis. Data collected at frequencies slower than 30 min, and certainly 60 min, is not sufficiently reliable for most purposes.

REFERENCES

[1] T. Fu, "A review on time series data mining," *Eng. Appl. Artif. Intell.*, vol. 24, no. 1, pp. 164–181, 2011.

[2] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 381–387, Feb. 2003.

[3] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data—An implementation in practice," *Business Inf. Syst. Eng.*, vol. 4, no. 1, pp. 31–39, 2012.

[4] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 933–940, May 2006.

[5] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, 2012.

[6] S. Verdu, M. Garcia, C. Senabre, A. Marin, and F. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006.

[7] G. Chicco, "Challenges for smart distribution systems: Data representation and optimization objectives," in *Proc. 12th OPTIM*, 2010, pp. 1236–1244.

[8] V. Figueiredo, F. Rodrigues, Z. Vale, and J. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, May 2005.

[9] C. Chen, M. Kang, J. Hwang, and C. Huang, "Synthesis of power system load profiles by class load study," *Int. J. Elect. Power Energy Syst.*, vol. 22, no. 5, pp. 325–330, 2000.

[10] I. H. Yu, J. K. Lee, J. Ko, and S. Kim, "A method for classification of electricity demands using load profile data," in *Proc. 4th ICIS*, 2005, pp. 164–168.

[11] G. Tsekouras, N. Hatziargyriou, and E. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120–1128, Aug. 2007.

[12] G. Chicco and I. Ilie, "Support vector clustering of electrical load pattern data," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1619–1628, Aug. 2009.

[13] S. Ramos, Z. Vale, J. Santana, and J. Duarte, "Data mining contributions to characterize MV consumers and to improve the suppliers-consumers settlements," in *Proc. IEEE PES General Meeting*, 2007, pp. 1–8.

[14] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 548–555, May 2005.

[15] R. Lamedica, L. Santolamazza, G. Fracassi, G. Martinelli, and A. Prudenzi, "A novel methodology based on clustering techniques for automatic processing of MV feeder daily load patterns," in *Proc. IEEE PES Summer Meeting*, 2000, vol. 1, pp. 96–101.

[16] J. Nazarko, A. Jurczuk, and W. Zalewski, "Arima models in load modelling with clustering approach," in *Proc. IEEE Power Tech.*, 2005, pp. 1–6.

[17] D. Marques, K. de Almeida, A. de Deus, A. da Silva Paulo, and W. da Silva Lima, "A comparative analysis of neural and fuzzy cluster techniques applied to the characterization of electric load in substations," in *Proc. IEEE PES T&D Conf. Expo.*, 2004, pp. 908–913.

[18] J. Williams, "Clustering household electricity use profiles," in *Proc. ACM MLSDA*, 2013, pp. 19–26.

[19] B. D. Pitt and D. Kitschen, "Application of data mining techniques to load profiling," in *Proc. 21st IEEE PICA*, 1999, pp. 131–136.

[20] C. Beckel, L. Sadamori, and S. Santini, "Towards automatic classification of private households using electricity consumption data," in *Proc. 4th ACM BuildSys*, 2012, pp. 169–176.

[21] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Appl. Energy*, vol. 87, no. 11, pp. 3538–3545, 2010.

[22] R. Granell, C. J. Axon, and D. C. H. Wallom, "A Dirichlet process mixture model for clustering of household electricity load profiles," in *9th MLDM, Poster Proc.*, 2013, pp. 1–6, IBaI Publishing.

[23] A. Notaristefano, G. Chicco, and F. Piglione, "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," *IET Gener., Transm., Distrib.*, vol. 7, no. 2, pp. 108–117, 2013.

[24] R. Sowden and M. Tommis, "D7.7 project cycle analysis report for cycle 3," 2011, project Dehems.

[25] A. Wright and S. Firth, "The nature of domestic electricity-loads and effects of time averaging on statistics and on-site generation calculations," *Appl. Energy*, vol. 84, no. 4, pp. 389–403, 2007.

[26] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing clusters among related groups: Hierarchical Dirichlet processes," in *NIPS*. Cambridge, MA, USA: MIT Press, 2005, pp. 1385–1392.

[27] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, no. 2–3, pp. 107–145, Dec. 2001.

[28] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[29] I. Richardson, M. Thomson, D. Infield, and C. Clifford, "Domestic electricity use: A high-resolution energy demand model," *Energy Build.*, vol. 42, pp. 1878–1887, 2010.

Ramon Granell is a Research Assistant at the Oxford e-Research Centre, University of Oxford, U.K. He is working on machine learning and data mining applications for power, energy, and dynamic tariffs.

Colin J. Axon (M'14) is a Lecturer in the Institute of Energy Futures at Brunel University, London, U.K. His research interests are broadly in the use energy in buildings and by transport. This includes the of data in LV networks and household power demand modeling.

David C. H. Wallom is the Associate Director Innovation of the Oxford e-Research Centre, U.K., where he leads Energy and ICT, Cloud Computing and Volunteer Computing. He is a member of the Cybersecurity ACE and the EPSRC DTC in Cybersecurity at the University of Oxford.