International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 1063

GENERATING CONSUMER LOAD PROFILES TO ASSESS DEMAND SIDE MANAGEMENT POTENTIAL OF INDUSTRIES

Javier Valdes[1*], Luis Ramirez Camargo [1] , Wolfgang Dorner[1]

1 Institute for Applied Informatics, Technologie Campus Freyung, Technische Hochschule Deggendorf, Freyung, Germany
(javier.valdes@th-deg.de)

ABSTRACT

The generation of electricity demand profiles is a fundamental task to determine the potential flexibility that could be introduced to an electrical system when applying demand side management. In this study we compare the results of two approaches, classical time series analysis and unsupervised clustering, to generate synthetic and anonymized electricity demand profiles. The objective is to retain the statically characteristics and descriptive value of real profiles without compromising company confidentiality. In this case the methods are applied to data obtained from a Chilean paper industry. The results reflect the complexity of the task and show the advantages and disadvantages of both methods.

Keywords: Demand side management, Time series clustering, Time series regression, electricity load profiles

1. INTRODUCTION

One of the major challenges in the transformation of our energy systems, from fossil to renewables based, is to match the electricity generation from variable renewable energy sources (VRES) and the demand. While the electricity generation of e.g. photovoltaic and wind power installations variates in its availability (following local and temporal weather patterns), electricity demand follows societal and industrial production requirements. Technologies that should align this mismatch in systems with high shares of VRES include distributed generation, storage systems and demand side management (DSM). For the optimal planning of these technologies one of the main requirements is high quality data that allow an accurate calculation of optimal sites, system sizes and operation thresholds. In the particular case of DSM planning one key requirement is the availability of electricity demand profiles in high temporal resolution that reflect high and low demand peaks, periods of flexibility and common event chains.

The acquisition of electricity demand data, which is useful for DSM planning, is not a trivial task that depends on many factors that range between the reliability of data loggers and secrecy needs of particular users. In the present study we address the generation of synthetic and anonymized electricity demand profiles for industrial users. The objective is to generate data sets that retain statistical characteristics and the descriptive value of measured data but are sufficiently simplified to be easily integrated in optimization models and to not compromise company confidentiality. To achieve this, twenty-five alternative methods from classical time series analysis and time series clustering are applied to a time series of electricity demand data from a Chilean paper industry. The alternative methods variate e.g. in the type of clustering (hierarchical and partitional), the type of centroid, the distance matrixes and the linkage methods.

The rest of this paper is structured as follows: Section 2 presents the methodology including an overview of the theoretical background, section 3 describes the available input data, sections 4 present the results by emphasizing in the advantages and disadvantages of the methods, and conclusions are drawn in section 5.

2. METHODOLOGY

Existing methods used to characterize electricity use can generally be divided into four categories: statistical; engineering; time series and clustering [1]. Within these methodologies there is a large number of methods for the characterization of data. Each of these methods offers a number of advantages and its use is linked to the type of data and the objective of the analysis. While the analysis is normally done with time series, the possibilities of using certain methods are reduced due to the fact that their feature values change as a function of time. This means that the values of each point of a time-series are one or more observations that are made chronologically.

Among the most used methods clustering methods stand out for their capacity to deal with a large amount of data and without the need to establish hypothesis a priori on the time series. That is why, time series clustering is an active research area with applications in a wide range of disciplines and usually covers one or more of the following aims: data reduction, hypothesis generation, hypothesis testing or prediction based on groups. Our aim is to reduce the complexity of the data creating representative profiles of electricity consumption. Formally, given a dataset of i time-series data D = {S1, S2, S3,. ., Sj} , the process of unsupervised partitioning of D into K = {K1, K2, K3, .., Kj}, in such a way that time-series are grouped together. Therefore, cluster analysis allows a population to be classified into a certain number of groups, based on similarities and discrepancies in the existing profiles between the different elements of the population. In our analysis we will use two different methodologies, on the one hand a descriptive analysis together with regression techniques and on the other, clustering analysis.

2.1 Cluster Analisis

Cluster analysis is a task which deals with the creation of homogenous groups of objects, where each group is called a cluster. Ideally, all time series of the same cluster are similar to each other, but are as dissimilar as possible from time series in a different cluster. As there is no single definition of a cluster, and the characteristics of the time series to be clustered vary, there are different algorithms to perform clustering. Each one defines specific ways of defining what a cluster is. Additionally, each application might have different goals, so a certain clustering algorithm should be chosen depending on the type of clusters the Researcher is interested in. Moreover, since no single clustering algorithm can be said to outperform the others, different clustering methods must be tested and compared. In these study we will consider two broad clustering methods: partitional and hierarchical clustering.

A partitional clustering is simply a division of the set of data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset. A hierarchical clustering is a set of nested clusters that are organized as a tree [2]. Moreover, we focus on four general-purpose popular distances, three different distance metrics and seven linkage methods fort hierarchical clustering (Table 1).

Clustering algorithm	Distance measure
Partitioning Around Medoids	Euclidean distance
Partitioning Around Medoids	Dynamic Time Wrapping
Partitioning Around Medoids	Shape-based distance
Partitioning k-shape	Dynamic Time Wrapping
Hierarchical with single linkage	Euclidean distance
Hierarchical with average linkage	Euclidean distance
Hierarchical with complete linkage	Euclidean distance
Hierarchical with ward.d2 linkage	Euclidean distance
Hierarchical with ward linkage	Euclidean distance
Hierarchical with centroid linkage	Euclidean distance
Hierarchical with median linkage	Euclidean distance
Hierarchical with single linkage	Dynamic Time Wrapping
Hierarchical with average linkage	Dynamic Time Wrapping
Hierarchical with complete linkage	Dynamic Time Wrapping
Hierarchical with ward.d2 linkage	Dynamic Time Wrapping
Hierarchical with ward linkage	Dynamic Time Wrapping
Hierarchical with centroid linkage	Dynamic Time Wrapping
Hierarchical with median linkage	Dynamic Time Wrapping
Hierarchical with single linkage	Shape-based distance
Hierarchical with average linkage	Shape-based distance
Hierarchical with complete linkage	Shape-based distance
Hierarchical with ward.d2 linkage	Shape-based distance
Hierarchical with ward linkage	Shape-based distance
Hierarchical with centroid linkage	Shape-based distance
Hierarchical with median linkage	Shape-based distance

Table 1. Clustering methods used in this study

2.2 Time-series Analisis

Time series analysis has been one of the main occupations of economists over the last decades. Due to the limitations to carry out controlled experiments, economic science and with it econometrics has been based on the study of processes over time. The analysis of time series covers a wide range of areas of study, ranging from market behavior, national accounts to the analysis of specific industries. Due to continuous effort, time series analysis has been extensively refined and there are a multitude of approaches depending on the type of phenomena to be studied as well as the research question and hypothesis. In the area of energy economics, time series approaches have been used extensively for modelling electricity demand.

Reference [1] provides a short review of statistical techniques consist of using descriptive statistics and probability (e.g. [3]) as well as regression (e.g. [4]) to describe electricity use. These methods produce highly diversified load profile shapes. In this paper we will apply a mixture of descriptive analysis and regression techniques to identify the main aspects of the data. Descriptive analysis aims at establishing the main characteristics of the time series and establishing hypotheses and then developing a regression analysis.

3. DATA

The available data corresponds to one year of measurements in hourly temporal resolution from a Chilean paper industry. Data from this type of industry is particularly valuable for DSM planning since the paper industry is recognized as one of the industries with the highest DSM potential [5]. The data were provided directly by the company and feature quality requirements of an energy management certification program. No particular information is provided here and data is only presented graphically (Fig. 1.) and in a normalized way (to 1MW) in order to comply with the confidentiality agreement with the company.

These time series show in general two periods of large differences, winter and summer. This reflects already the response of the industry to the tariff based mechanism that penalizes with additional costs high power requirements during winter [6]. Apart from this general difference it is also possible to recognize from the figure 1 a cyclical behavior during the last hours of the day with emphasis during the winter months and multiple days with very low demand. The latter can be attributed to production interruptions related to holidays or programmed maintenance.

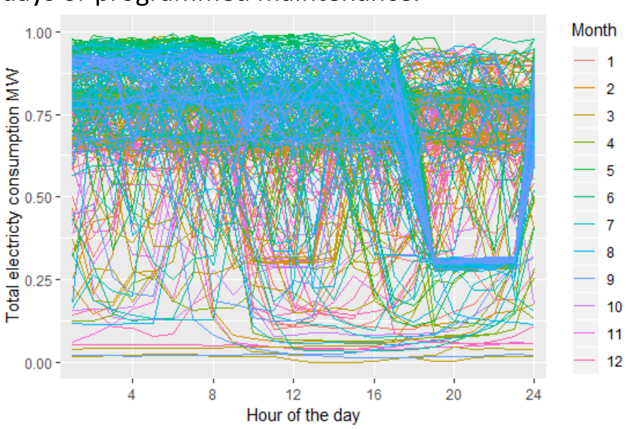

Fig. 1. Hourly energy demand by day of the year in a paper industry

4. RESULTS

Time series analysis is carried out using Ordinary Least Squares method on a linear model with dummy with an intercept term and dummy variables to indicate the main temporal aspects. Dummy or hot encoding variables take the 0 or 1 value to indicate the absence or presence of some categorical effect that may be expected to shift the outcome. In this case we use these variables to calculate various models of linear regression that include: months, days of the month, days of the week and hours. These models are calculated for the set of time series as well as for each of the months (see Fig. 2).

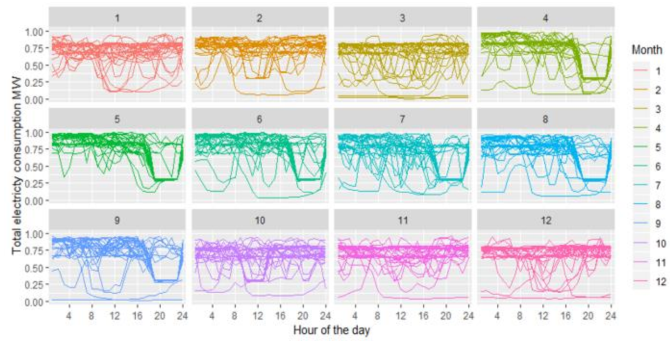

Fig. 2. Hourly energy demand by day of the year and month in a paper industry

The results show that the plant's total electricity consumption is significantly negatively affected in the winter months, particularly between 6 p.m. and 11 p.m., and that this effect is less pronounced on Fridays and Saturdays. Taking May as the representative month, the regression model has an Adjusted R-squared: 0.71. As for the coefficients, we found levels of significance above 99% between 18 and 23 with negative coefficients and 24 with positive coefficients. As for the days, Thursday and Friday also present negative coefficients and levels of significance above 99%. Detailed results are available upon request. Based on the descriptive data analysis and these results it is possible to identify four well differentiated groups of profiles (Fig. 3). A first group in which the consumption is significantly reduced during most of the day and that can correspond to the maintenance of the plant. A second group consists of the winter months in which the energy consumption presents a homogeneous body in which no pattern is identifiable. A third group identified by the winter months in which consumption is significantly lower during the afternoon and evening hours. And a last group, reduced from days on Friday and Saturday in which consumption is stable throughout the day.

We finish the analysis based on regressions here even when it will be possible to continue the analysis analyzing the results of the winter group in more detail. We will use this results as benchmark for the cluster methods.

The application of the clustering methods has as output a large number of clusters due to the multiple combinations possible between the methods summarized in table 1. In order to select only the best results of the multiple iterations we calculate the following five cluster validity indices (cvi) using the function cvi() from R package dtwclust v3.1.1 [7,8]: the Silhouette index, Dunn index, COP (the Context-independent Optimality and Partiality properties) index, Davies-Bouldin index, Calinski-Harabasz index and the Score Function. Moreover, as some clustering methods may be affected by the initial choice of centroid locations, we run 10 times each clustering algorithm to the data restricting the maximum number of clusters to 15, finally the clusters with maximizing or minimizing the cvi are selected. The clustering results selected using the different cvi function are summarized in tables 2 and 3.

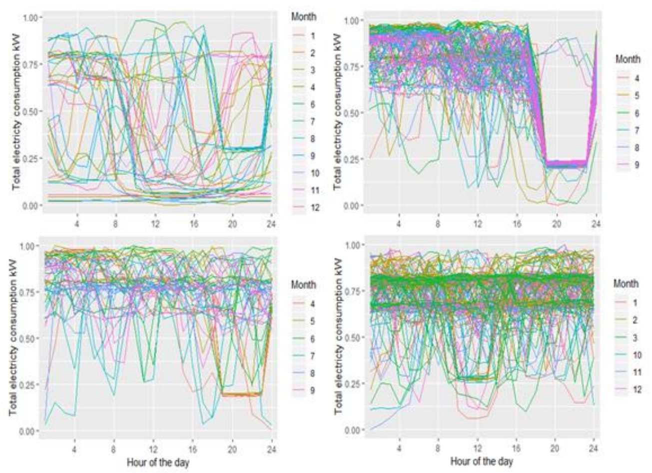

Fig. 3. Results from the regression analysis

The results of the partitioning clustering show that there is a tendece to generate three groups, these results are consistent with the results of the hierarchical cluster and the previous regression analysis. However, it is important to note that in some of these cases the clusters have only a series of time, which makes the method actually only able to detect outliers and not representative groups of consumption patterns. In spite of these, based on these results and the rule of majority voting seems logical to think that the structure of the data responds to three or four groups, something that corresponds with the previous results.

Distance	Centroid	Sil	Dunn	COP	DB	CH	SF
Euclidean	k-medoids	11	4	15	3	15	12
Dtw	k-medoids	3	15	15	3	3	3
Dtw	dba	3	9	15	4	3	3
Sbd	shape	3	8	15	4	3	3

Table 2. Cvi results for the partitional clustering

Distance	Centroid	Sil	Dunn	COP	DB	CH	SF
Euclidean	k-medoids	3 ave	3 sing	15 w.d2	3 cent	3 ward	5 sing
Dtw	k-medoids	3 cent	5 cent	15 com	3 cent	3 w.d2	5 sing
Sbd	k-medoids	3 ave	3 sing	15 w.d2	4 med	3 sing	7 sing

Table 3. Cvi results for the hierarchical clustering

To assess these results and their agreement with the previous ones, figure 4 and 5 present the results of two different clusters with 3 and 15 clusters selected by the DB and Dunn index respectively. The results are satisfactory considering the knowledge about the time series gathered during the time-series analysis and the objective of the research: the creation of consumption profiles for later use in demand management models.

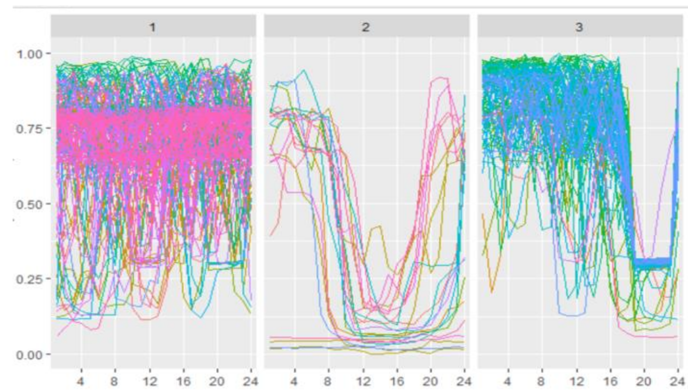

Fig. 4. Results for hierarchical clustering, ward linkage, k-medoids and euclidean distance

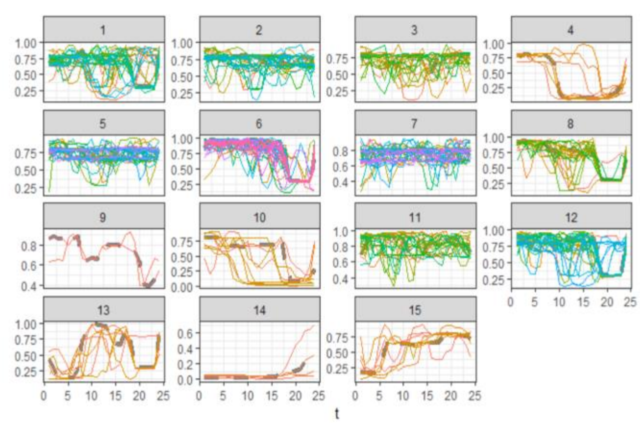

Fig. 5. Results for partitional clustering with dtw distance and k-medoids

5. CONCLUSIONS AND FUTURE WORK

The results show that although clustering techniques for time serires present great advances and versatility, their application to real data characterized by high volatility presents important barriers. The results show the advantages of some clustering techniches over others. Given these results, a regression analysis supported by a descriptive analysis and hypothesis on consumption presents a great ability to identify key elements in consumption patterns and generate a benchmark for the clustering methods. Therefore, the combination of both methods should be further explored in order to refine the analysis, for example, creating further subsets of data to find other significant consumption patterns as the ones identified in figure 5. Nevetheles, the results are sufficient to, by means of a statistical analysis of the different clusters, generate consumption ranges and to establish the flexibility capacity of the industry. On the other hand, one avenue for future research will be the application of other unsupervised learning techniques based on neural networks that can cope with the complexity of the data collected.

ACKNOWLEDGEMENT

The study was conducted within the project INCREASE "Increasing renewable energy penetration in industrial production and grid integration through optimized CHP energy dispatch scheduling and demand side management" (grant number BMBF150075) funded by the German Federal Ministry of Education and Research (BMBF) and the Chilean National Commission for Scientific Research and Technology (CONICYT).

REFERENCE

[1] F. McLoughlin, A. Duffy, M. Conlon, Evaluation of time series techniques to characterise domestic electricity demand, Energy. 50 (2013) 120–130. doi:10.1016/j.energy.2012.11.048.

[2] P.-N. Tan, M. Steinbach, A. Karpatne, K. Vipin, Introduction to Data Mining, 2nd Edition, Pearson, 2019. https://www.pearson.com/us/higher-education/program/Tan-Introduction-to-Data-Mining-2nd-Edition/PGM214749.html (accessed May 20, 2019).

[3] M. Aydinalp, V. Ismet Ugursal, A.S. Fung, Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks, Applied Energy. 71 (2002) 87–110. doi:10.1016/S0306-2619(01)00049-6.

[4] E. Gabreyohannes, A nonlinear approach to modelling the residential electricity consumption in Ethiopia, Energy Economics. 32 (2010) 515–523. doi:10.1016/j.eneco.2009.08.008.

[5] M. Paulus, F. Borggrefe, The potential of demand-side management in energy-intensive industries for electricity markets in Germany, Applied Energy. 88 (2011) 432–441. doi:10.1016/j.apenergy.2010.03.017.

[6] International Energy Agency, Energy Policies Beyond IEA Countries: Chile 2018, 2018. http://www.iea.org/publications/freepublications/publication/EnergyPoliciesBeyondIEACountriesChile2018Review.pdf (accessed May 18, 2018).

[7] S. Alexis, dtwclust package, 2017. https://www.rdocumentation.org/packages/dtwclust/versions/3.1.1 (accessed May 20, 2019).

[8] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, I. Perona, An extensive comparative study of cluster validity indices, Pattern Recognition. 46 (2013) 243–256. doi:10.1016/j.patcog.2012.07.021.