Using clustering techniques to provide simulation scenarios for the smart grid

Pedro Miguel [a,c,*], José Gonçalves [a,c], Luís Neves [a,b,c], A.Gomes Martins [a,c]

[a] Energy for Sustainability Initiative, University of Coimbra, Coimbra, Portugal
[b] School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal
[c] INESCC—Institute for Systems Engineering and Computers at Coimbra, Coimbra, Portugal

ABSTRACT

The objective of this work is to obtain characteristic daily profiles of consumption, wind generation and electricity spot prices, needed to develop assessments of two different options commonly regarded under the smart grid paradigm: residential demand response, and small scale distributed electric energy storage. The approach consists of applying clustering algorithms to historical data, namely using a hierarchical method and a self-organizing neural network, in order to obtain clusters of diagrams representing characteristic daily diagrams of load, wind generation or electricity price. These diagrams are useful not only to analyze different scenarios of combined existence, but also to understand their individual relative importance. This study enabled also the identification of a probable range of variation around an average profile, by defining boundary profiles with the maximum and minimum values of any cluster prototypes.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Abundant recent developments regarding power systems are focused on the concept of smart-grids (Bundesnetzagentur, 2011; Electric Power Research Institute, 2008; European Union Advisory Council, 2010; Pacific Northwest National Laboratory, 2010; SmartGrids, 2012), aiming to achieve the ultimate goal of balancing supply and demand, not only by acting on a strictly controllable supply side in response to the naturally uncontrollable demand, although predictable to a certain degree, as in the past, but now also using new solutions to make the demand respond to the less predictable variations of supply, that are characteristic of renewable sources, or increasing the ability to store energy, so that energy generated in one period of time can be used in a period of time when demand really occurs.

In this context, there is an obvious need to evaluate the true value of these options so that the required investments can be justified. But to obtain figures of the significance both options can achieve, there is an absolute need of data, namely on the time-dependent behaviour of load demand, renewable generation and electricity price.

The concept of Demand-response (Albadani e and El-Saadany, 2008; Sá, 2011; Mahmoudi, Eghbal, & Saha, 2014), is based on the possibility of inducing consumers to dynamically change their consumption pattern in order to match the pattern that can be most convenient to the grid operator. The partial accomplishment of this objective can result of the use of stimuli as, e.g., a Real Time Price (RTP) scheme (Nunes, 2011) in which the price of electricity seen by the individual consumer changes frequently. New proposed devices can use this stimuli to automatically decide on behalf of the consumer if certain loads can be temporarily stopped, or their start postponed, in order to avoid a price peak. But for the massive deployment of these devices, a careful evaluation is needed in order to justify the heavy investment in all the infrastructure.

Likewise, the search for new storage options led to the concept of distributed storage (Ribeiro, Johnson, Crow, Arsoy e, & Liu, 2001; Divya and Østergaard, 2009; Lassila, Haakana, Tikka, & Partanen, 2012; Sigrist, Lobato e, & Rouco, 2013; Leou, 2012), e.g., using the batteries installed on the electric vehicles that are expected to flood the electric grids in the near future with the so called Vehicle-to-grid (V2G) option (Lassila et al., 2012). But this, or any other distributed storage option, needs also to be assessed in order to justify a significant investment in the supporting infrastructure, and this needs to compare the temporal evolution of demand, prices and

* Corresponding author at: University of Coimbra, Energy for Sustainability Initiative, 3000-033 Coimbra, Portugal.
E-mail addresses: pcmiguel@gmail.com
(P. Miguel), jagoncalves@gmail.com (J. Gonçalves), luis.neves@ipleiria.pt (L. Neves), agmartins@uc.pt (A.Gomes Martins).
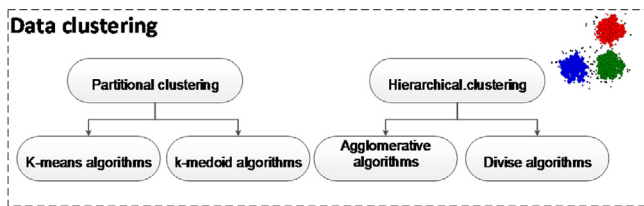
Data clustering

Partitional clustering   Hierarchical.clustering

K-means algorithms   k-medoid algorithms   Agglomerative algorithms   Divisive algorithms

Fig. 1. Data clustering techniques (Pujari et al., 2001).

also renewable generation with time, if the aim is to help improving the match between generation availability and demand.

The development of research work (Miguel, Neves, & Martins, 2014; Gonçalves, Neves, & Martins, 2015) related with the two options above, led to the need of obtaining representative daily diagrams of demand, renewable generation (wind) and electricity gross market prices, in order to generate reliable scenarios of analysis, aiming also to obtain a measure of the representativeness of each diagram regarding the context of a whole year. To achieve that goal, different clustering techniques were applied to the historical data sets of hourly consumption, wind generation and gross market prices of a complete year. The present paper, focuses on this process of obtaining representative daily diagrams or profiles.

The next section makes a short review of clustering techniques, justifying the choice of the particular methods that were used and explaining their application to the problem. Section 3 presents the application of the chosen methods to the data used, representing the consumption of a small town in Portugal, the wind generation data provided by the Independent System Operator (ISO) and the gross market prices obtained from the wholesale electricity market operator website. Section 4 presents the results that were obtained, and finally, Section 5 presents conclusions.

2. Clustering techniques

2.1. Motivation for using clustering methods

The present trend towards the development of smart grids presents an opportunity for data clustering (Benítez, Quijano, Díez, & Delgado, 2014), considering the large data sets for the dynamic analysis of the electricity grid, such as: the annual monitoring of distribution and residential load demand, energy market prices and renewable generation. Although the evaluation of whole years is mandatory, the need to simulate management strategies requires some sort of daily management leading to the need of daily characteristic profiles. e.g., a number of extraordinary days may particularly justify the use of demand response and energy storage, namely when spot prices achieve maximum values and there is potential to shift usage, or when renewable generation is totally decoupled from consumption. But, for assessing their relevance, it is important not only to have models of the diagrams of consumption, generation and market price evolution but also to assess their relevance for a given period of time.

Clustering methods may provide a response for this need by performing an automatic classification of daily diagrams over a long period of time.

Cluster analysis is a convenient method for identifying homogenous groups of objects or clusters (Mahmoudi et al., 2014; Mooi and Sarstedt, 2011), aiming to group objects by similarity, but keeping a significant difference between groups. Clustering techniques can be divided in two main groups, Partitional clustering methods and Hierarchical clustering methods (Pujari, Rajesh, & Reddy, 2001), as presented in Fig. 1.

Partitional algorithms use an interactive optimization paradigm. They start with an initial partition and, in each iteration, objects

are swapped if the quality of the clustering is improved, by applying an interactive control strategy. The definition of the clusters will strongly depend of the initially selected partition. Another relevant issue is the representation of clusters. In the K-mean algorithms a cluster is represented by its center of gravity while the K-medoid algorithms use the closest object from the center as the center representation. However, according to Sousa (2006) the limitations of this method are the dependence on the prototypes initially assigned to clusters and also the lack of a clear assignment of objects to clusters, creating difficulties to the definition of prototypes.

The Hierarchical methods (HM) create sequenced partitions, each one containing the previous one in a hierarchical way. There are two types of hierarchical methods (Maimon and Rokach, 2005), the agglomerative clustering algorithms and the divisive clustering algorithms, the former performing a bottom-up approach and the latter a top-down approach, using a similar clustering rule based on similarity or distance (Maimon and Rokach, 2005). According to Sousa (2006), the agglomerative clustering algorithms are probably the most used hierarchical methods, and therefore one of the choices made for the current work was based on such methods. A more complete description is then provided in Section 2.2.

An alternative procedure introduced by Teuvo Kohonen in the early1980's, the self-organizing maps (SOM), is a type of artificial neural network (ANN) with an unsupervised learning process and unknown output partitions (Kohonen, 1998). The Kohonen SOM could be compared to conventional clustering methods, considering the internal allocation rules and their performance (Sousa, 2006).

Other specialized procedures can be found in the literature, as the weighted fuzzy average (WFA) K-means algorithm and the "Modified Follow the Leader" (Mahmoudi-Kohan, Moghaddam, & Bidaki, 2009).

According to Mahmoudi-Kohan et al. (2009) WFA K-means is well suited to design electricity tariffs, while "Modified Follow the Leader" is suited for the selection of demand response policies.

For the current work the authors chose to implement the HM and SOM methods due to their adequacy to the objective of clustering of load diagrams, as shown by Sousa (2006).

2.2. Hierarchical method (HM)

According to Sousa (2006), the hierarchical clustering method investigates possible groupings of data by creating a structure similar to a hierarchical tree. Such tree is comprised not only by a simple set of clusters, but rather by a multiple level hierarchy, where clusters at one level are grouped in clusters of a higher level. The basic procedure for developing the cluster prototypes is performed according to the following steps (Sousa, 2006):

The input data is evaluated and a decision is made on what kind of clustering results are valuable to obtain later on. For the cases studied on this paper, this decision is based on either the amplitude difference between cluster members (defined as the difference between extreme values of a daily profile) or on the diagram shape. For the price diagrams it was assumed that what is relevant to differentiate prototypes is the amplitude difference, while for electricity profiles (consumption and generation) what is relevant is the diagram shape. As stated earlier, a possible approach of the hierarchical method consists in merging objects according to their linkage distance, using one of multiple possible definitions, namely: single, complete, average, centroid or ward distance, the latter being used in the present case. After calculating distances between data members, it is possible to group objects according to their proximity (the shortest distance). As new higher level clusters are created, comprised by more members, it is necessary to determine the distances between them until the hierarchical tree is composed by only one

single cluster. More information regarding this technique can be found in Everitt, Landau, Leese, and Stalh (2011).

2.3. Artificial neural network (ANN) method

The ANN method is based on the Kohonen self-organizing map (Kohonen, 1998).This method is able to convert nonlinear statistical relationships between elements of data into a simple geometric relationship on a low dimensional display, based on similarity and topology. Similarly to the hierarchical method, this clustering method is also an example of unsupervised learning because the classes for the output vectors are not initially known. The neural network contain two layers: an input layer of p-dimensional observations and an output layer representing k nodes for the k clusters. The clustering process occurs when an input vector is assigned to an output node with a p-dimensional vector with w weight. According to Everitt et al. (2011), a random weight is assigned to the output node which changes with the learning process, with neurons being temporarily assigned to clusters. The stabilization of the iterative process occurs when weights correspond to cluster centers, in such a way that clusters that are similar to one another are situated close together on the map.

3. Clustering load, price, and renewable generation profiles

The development of research work regarding the simulation of the impact of a large

deployment of a residential energy management system in the electricity grid (Miguel et al., 2014) and the choice of the best location of distributed storage devices, considering associated impacts, both of economic and technical nature and possible management schemes of storage units (Gonçalves et al., 2015), led to the need of generating representative daily diagrams of demand, renewable generation (wind) and electricity wholesale market prices, in order to generate reliable scenarios of analysis, aiming also the determination of a measure of representativeness of each diagram regarding the consumption of a year. To achieve that goal, different clustering techniques were applied to the historical data sets of hourly consumption, wind generation and gross market prices of a complete year.

The proposed methodology, represented in Fig. 2, deals with two types of input data, namely, one year on a quarter hour time step for load and wind power generation diagrams and one year of hourly electricity wholesale market prices.

In order to define the appropriate use of gathered data, an experimental setup was performed using two approaches for the data clustering techniques, one using the input data without any treatment and a second one where the input data were subjected to a normalization procedure such that the sum of all the elements of each vector equals 1, in order to cross-check their validity (Sousa, 2006).

The data clustering techniques were applied in two different modes: one using the input data without any treatment and a second one where the input data was subjected to a normalizing procedure. This experiment showed that in the case of the electricity profiles, the shape of the diagram is the relevant attribute while in the case of the price diagrams, the amplitude similarity between diagrams should be used. The diagram shape comparison is performed by using the Euclidian distance between individual normalized vectors or/and intermediate vectors that represent the data clusters. In both cases, it was necessary to initially adjust the data to the daylight savings clock changes of the daylight savings scheme, namely on the last Sunday of March and October.

At the clustering stage, the hierarchical method helps in choosing the appropriate number of clusters, graphically, using the tree
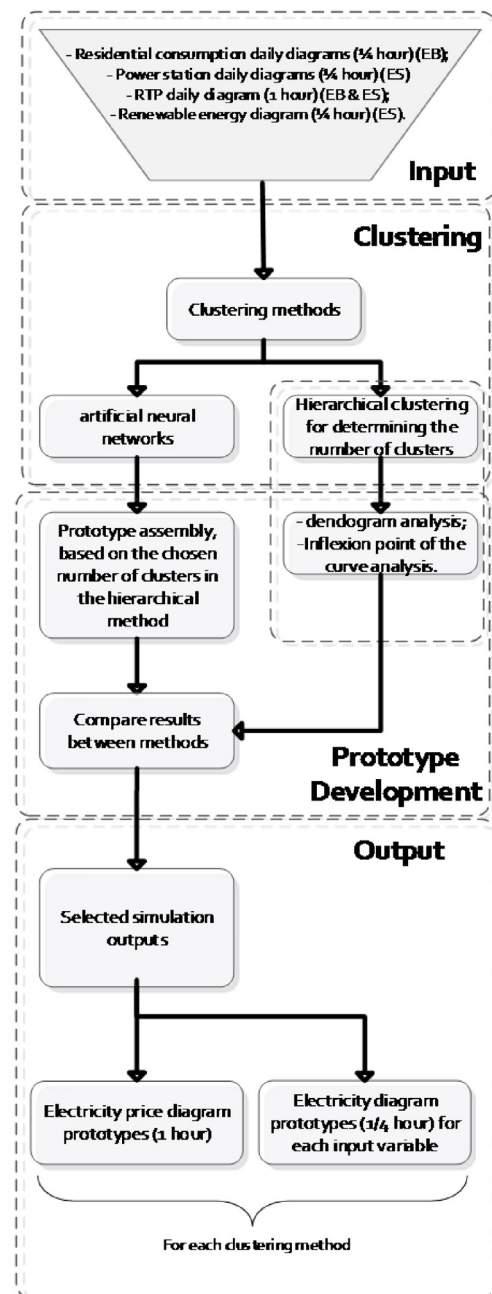

Fig. 2. Methodology for profile cluster prototypes of price, load demand electricity consumption and renewable generation.

dendogram and using the inflection point of the curve that relates the distances between clusters with the number of clusters (further detailed in Section 4) (Salvador and Chan, 2004). A comparison was made of the output of both methods regarding the assignment of day types (working day, Saturday, Sunday and holiday) to clusters.

With this information, it is also possible to determine the range of variation (maximum and minimum) for the profiles, as well as to define a range of variation in reference to the average profile.

To test the proposed methodology, the authors used data from different sources pertaining to the Portuguese market. The information regarding market prices was extracted from "The Iberian Energy Derivatives Market Exchange or MIBEL" website. The high voltage (HV) load demand data, derived from an annual measurement at the Pinheiros substation located in the city of Leiria in Portugal, and the low voltage (LV) household diagrams were kindly
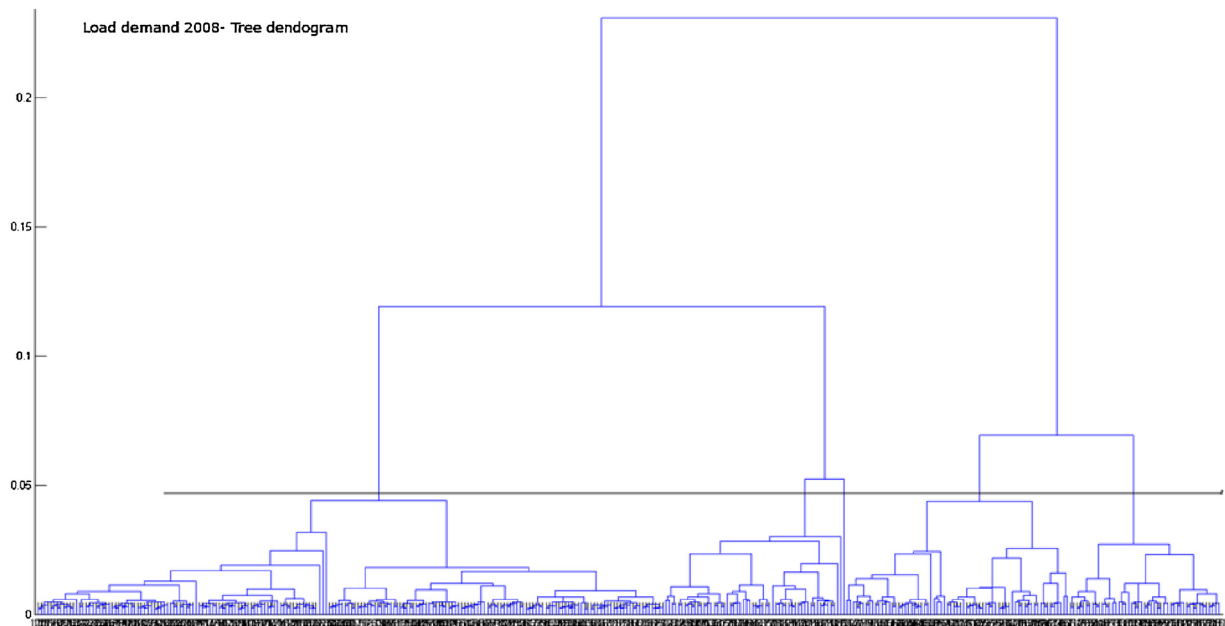
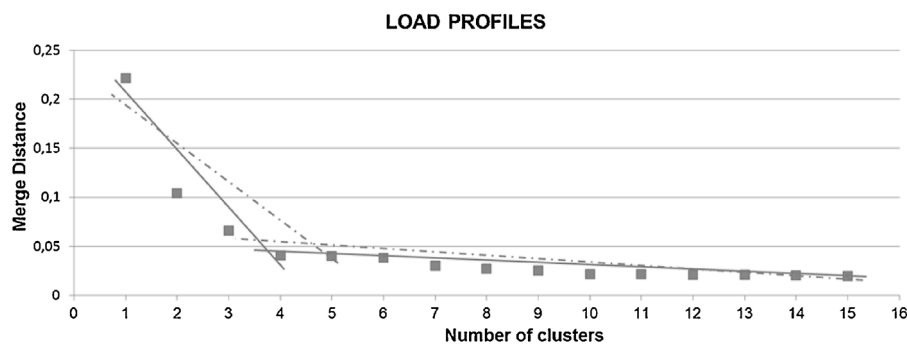**Fig. 3.** Tree dendogram representing the clustering of LD daily diagrams for 2008.

Fig. 4. Relation between the selected number of clusters and the merge distance for the case of load demand.

provided by EDP, the Portuguese electric utility. Finally, the renewable wind generation was extracted from the information center website of the Portuguese Transmission Operation (TSO)—Redes Energéticas Nacionais (REN).

4. Application of the methodology and analysis of results

The first step is to choose the appropriate number of clusters to consider. This can be attained by analyzing the tree dendogram in the case of HM, as depicted in Fig. 3, or by testing the results for different numbers of output nodes in the ANN and analyzing the observed relation between the number of clusters and the distances between the centers of the obtained clusters, as visible in Fig. 4.

Using both figures the selection of the number of clusters can be better justified as the single use of any of them results in a degree of uncertainty: the threshold in the case of the dendogram, as depicted in Fig. 3, and the location of the inflexion point of the curve, as observed in Fig. 4. This procedure was applied to select a set of admissible scenarios, and a total of 5 clusters were selected for the three data sets.

As both methods are based on non-supervised learning, a first assessment was made to the meaning of the obtained clusters by attempting to match with day types, namely weekdays, Saturdays, Sundays and holidays.

4.1. Clustering of load profiles

Regarding energy consumption, both methods presented a significant correspondence of clusters with weekdays and weekend days, although with a few differences regarding the number of allocated days to each cluster. For different distributions, another measurement method may be used (e.g. complete linkage). The holidays were not represented in any particular cluster, as they present patterns similar to weekends.

The results for both methods, regarding the two datasets, the 2008HV load demand and the 2012–2013 household (HH) load demand, are presented in Table 1 and Table 2.

The prototype of each cluster was obtained with an average of the diagrams allocated to it. The prototypes concerning the 2012–2013 household load demand are presented in Fig. 5 and Fig. 6, representing the output of each method.

The obtained prototypes are similar for both methods, especially in the residential profiles. This may be explainable, because the residential consumption has a more stable and defined pattern than HV load demand representing a substation that includes residential, services and industrial consumers.

The output of both methods provided similar prototypes which confirms the individual reliability of such methods. As an example, the following prototypes present similarities: HM #1 (32 hits) with ANN #3 (87 hits), HM#2 (110 hits) with ANN#4 (80 hits),

Table 1
Data allocation for 2008HV load demand and 2012–2013HH load demand with the hierarchical method.

Cluster	Total hits	5 Cluster's analysis for HM − 2008HV LD				Total hits	5 Cluster's analysis for HM − 2012–2013HH LD			
		Week days	Saturdays	Sundays	Holidays		Week days	Saturdays	Sundays	Holidays
1	24	1	2	17	4	32	31	0	0	1
2	53	0	18	33	2	110	110	0	0	0
3	37	0	32	0	5	59	4	25	26	4
4	55	55	0	0	0	76	18	27	26	5
5	197	195	0	2	0	88	88	0	0	0
Total	366	251	52	52	11	365	251	52	52	10

Table 2
Data allocation for 2008HV load demand and 2012–2013HH load demand with the artificial neural network method.

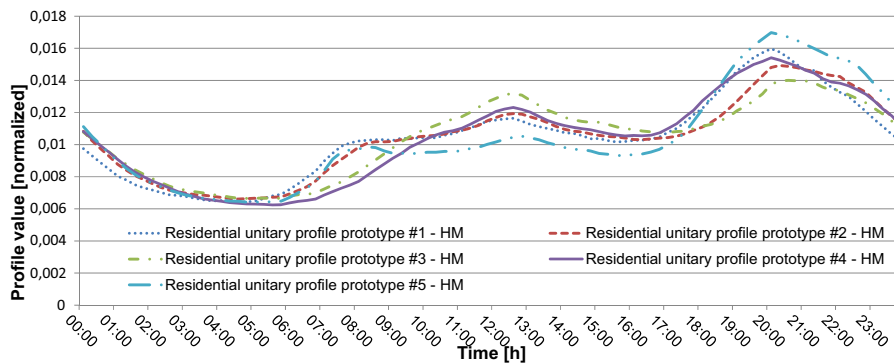| Cluster	Total hits	5 Cluster's analysis for ANN − 2008HV LD				Total hits	5 Cluster's analysis for ANN 2012–2013HH LD			
		Week days	Saturdays	Sundays	Holidays		Week days	Saturdays	Sundays	Holidays
1	52	1	43	3	5	88	88	0	0	0
2	61	0	8	47	6	57	11	22	20	4
3	92	92	0	0	0	87	86	0	0	1
4	108	105	1	2	0	80	65	10	4	1
5	53	53	0	0	0	53	1	20	28	4
Total	366	251	52	52	11	365	251	52	52	10

Fig. 5. HM 2012–2013 cluster prototypes for the residential load profiles.

HM#3 (59 hits) with ANN#5 (53 hits), HM#4 (76 hits) with ANN#2 (57 hits) and HM#5 (88 hits) with ANN#1 (88 hits). The difference in the number of hits for specific prototypes is probably due to outliers (frontier values) that could be assigned differently in the clustering process of each method (Salvador and Chan, 2004). Nevertheless, the two clustering methods followed different processes and resulted in different groupings, which are visible in terms of matching profiles to week days and weekend days, thus justifying these differences in number of hits.

In order to determine the range of variation of the profiles, the average profile for all data was compared to all five obtained pro-

totypes. This process determined an average variation range of the profiles, in absolute and relative values, as presented in Figs. 7 and 8, respectively.

Taking into consideration the differences between extreme values, and based on one year of data, Fig. 7 suggests that the average profile may be used as a reference for DR studies in the residential sector.

The information provided by Fig. 8 can be used to assess the range of variation based on the average profile, e.g., when estimating the value of Demand Response.
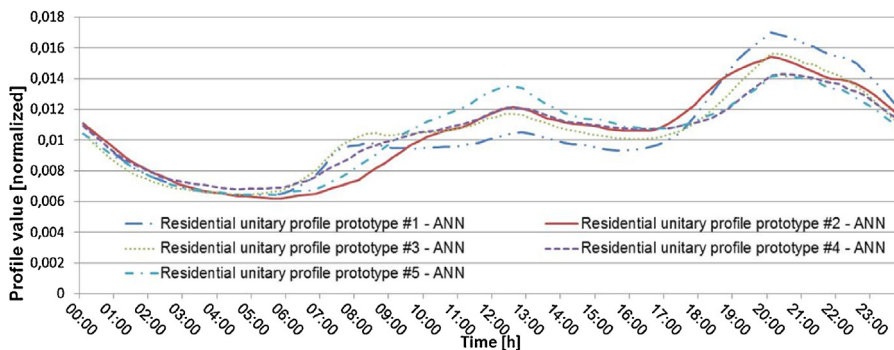

Fig. 6. ANN method 2012–2013 cluster prototypes for the residential load profiles.
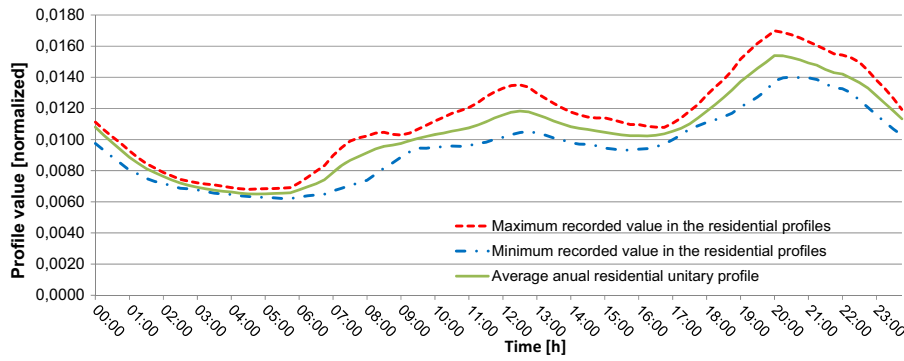
Fig. 7. Clustering evaluation of household load demand for the 2012–2013 cluster profiles.
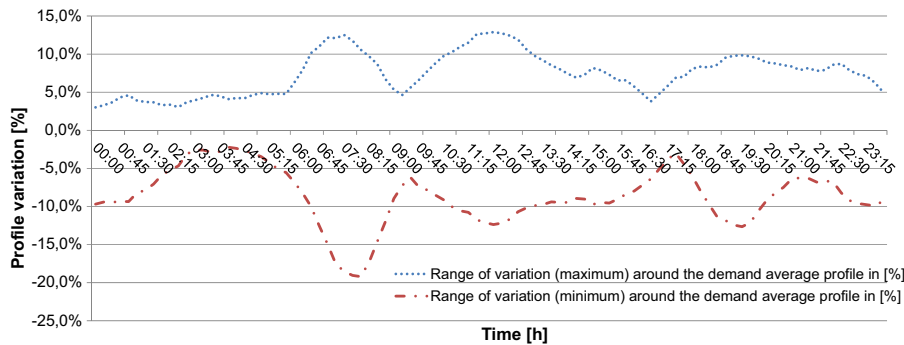

Fig. 8. Range of variation for household load demand for the 2012–2013 cluster profiles.
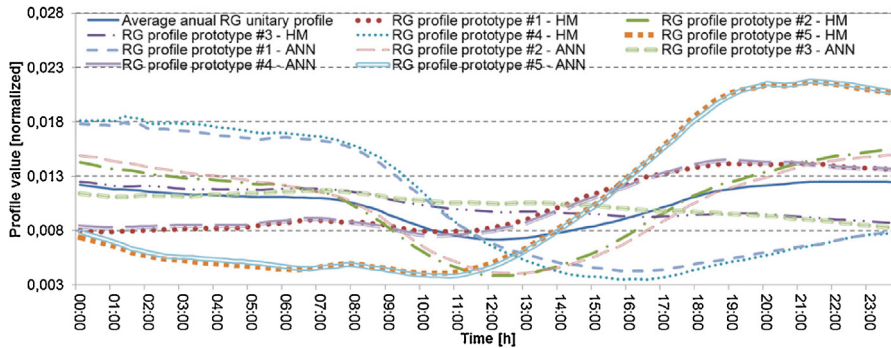

Fig. 9. Renewable HM and ANN prototypes for the year of 2008.

4.2. Clustering of renewable generation profiles

As one of the major motivations for storage is to increase the usefulness of renewable energy, by decoupling the periods of usage from the periods of generation, the clustering of historical data from renewable energy generation allows to obtain characteristic daily patterns and to understand their relevance. As expected, the obtained prototypes were quite different, due to the unpredictability of this energy source, as shown in Fig. 9.

Both methods provided similar prototypes in shape. The HM#1 (62 hits) prototype matches the ANN#4 prototype (77 hits), the HM#4 prototype (46 hits) matches the ANN#1 prototype (63 hits). Corresponding HM and ANN prototypes #2 (83 hits, 81 hits, respectively), #3 (129 hits, 97 hits, respectively) and #5 (46 hits, 48 hits, respectively) match, as perceivable in Fig. 9.

The difference between maximum and minimum values in Fig. 9 shows that an average profile is of limited use. Cluster #1 (17.21%) and cluster #5 (13.11%), representing 30.33% of the wind pattern profiles in one year time, exhibit an attractive possibility for the

study of the integration of energy storage systems. Moreover, and as an example, cluster #1 presents an appealing profile for storing energy from 00h00 m to 08h00m, while in cluster #5 this possibility would only occur from 18h30 m to 0h00m, if this shifting had any usefulness.

The study of the deployment of electric storage systems needs real data regarding the relation between daily demand and renewable generation patterns, as visible in Fig. 10.

In fact, cluster #1 of the renewable generation and cluster #1 of load demand, representing 17.21% and 14.21% of the number of occurrences in one year time, respectively, show situations of particular interest of energy storage to maximize the benefits of renewable energy due to the visible decoupling between generation and consumption.

4.3. Clustering of wholesale market price profiles

Tables 3 and 4 show the correspondence of electricity price clusters with the type of day for both methods.
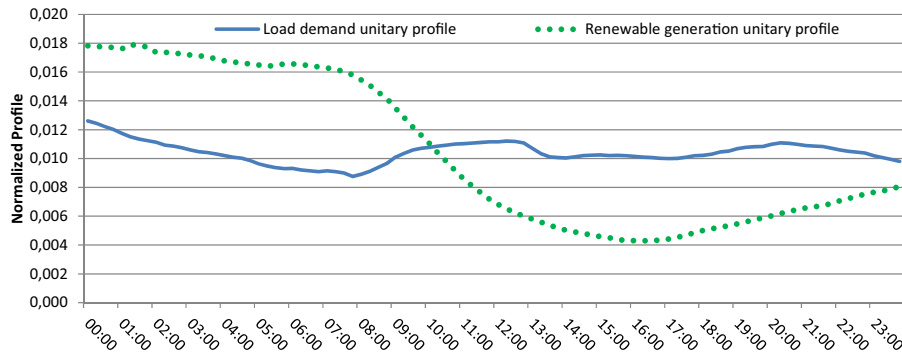
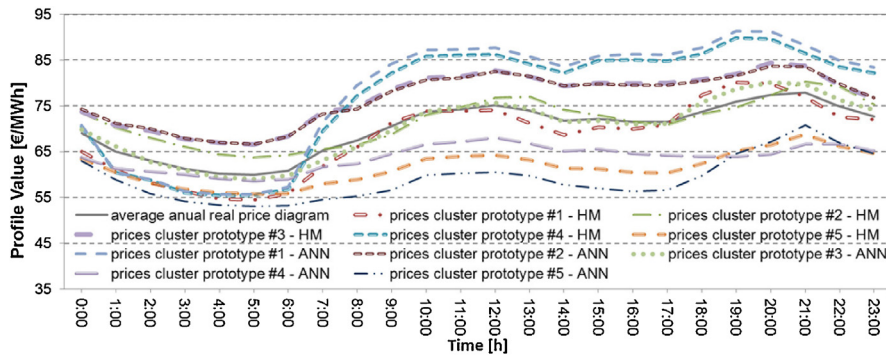**Fig. 10.** Cluster#1 for HV load demand and renewable energy generation.

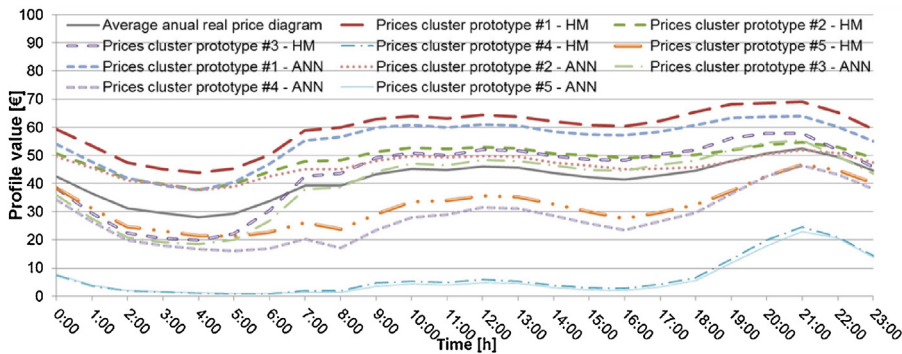**Fig. 11.** Clustering price prototypes for the HM and the ANN method, for the 2008 year.

Fig. 12. Clustering price prototypes for the HM and the ANN method, for the 2012 year.
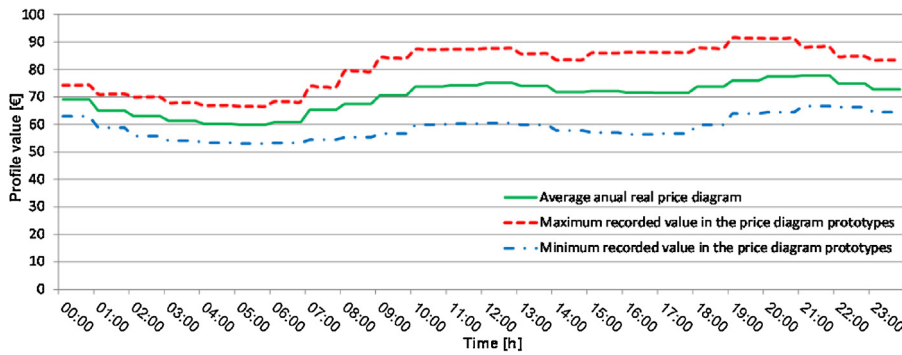

Fig. 13. Prototypes price range variation, for the year of 2008.

The resulting clusters for the electricity prices did not show a particular correspondence to the type of day, no cluster being formed more with a type than other, meaning that the electricity prices were decoupled from the type of day. Although the current market situation creates conditions for this to happen, e.g. a path to a liberalized market with an increase in renewable energy gen-
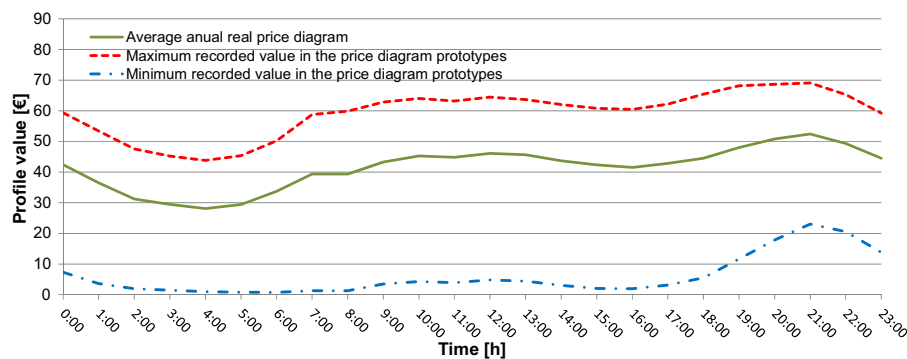
Fig. 14. Prototypes price range variation, from October 2012 to September 2013.

Table 3
Data allocation for 2008 electricity prices performed with the hierarchical method.

5 Cluster's analysis for hierachical method					
Cluster N.°	Total hits	Week days	Saturdays	Sundays	Holidays
1	42	35	3	2	2
2	71	27	20	22	2
3	93	81	9	3	0
4	44	44	0	0	0
5	116	64	20	25	7
Total	366	251	52	52	11

Table 4
Data allocation for 2008 electricity prices performed with the artificial neural network method.

5 Cluster's analysis for artificial neural networks					
Cluster N.°	Total hits	Week days	Saturdays	Sundays	Holidays
1	36	36	0	0	0
2	101	84	9	8	0
3	108	63	24	17	4
4	69	49	9	8	3
5	52	19	10	19	4
Total	366	251	52	52	11

eration, it suggests also that the natural consumption variation between workdays and weekends has not a noticeable influence on prices.

The resulting clustering price prototypes can be visualized in Figs. 11 and 12. Both methods present similar types of price prototypes and the approximate same number of hits. This resemblance ensures that any of the clustering methods may be used for generating reliable simulation prototypes.

From the analysis of Fig. 11 it is possible to recognize prototypes that originate from the two different methods but that are similar, namely: HM#1 (42 hits) with ANN#3 (108 hits), HM#2 (71 hits) with ANN#4 (69 hits), HM#3 (93 hits) with ANN #2 (101 hits), HM#4 (44 hits) with ANN#1 (36 hits) and HM#5 (116 hits) with ANN#5 (52 hits).

Larger hourly price differences are more likely to stimulate active load shifting and the period between October 2012 and September 2013 provided a particular cluster of price daily profiles, the #3 from the HM matched by the #2 from ANN, that seems interesting for the test of DR. Prototype #3 presents approximately the price of 20€/MWh at 4 h00 m and almost 60€/MWh at 20 h00 m, with an impressive number of hits, namely: 74/365 in the HM and 65/365 in the ANN method. Profiles with smaller hourly differences are expected to be less interesting for DR application.

As for the evaluation of the fluctuation of the maximum and the minimum hourly values as regards to the average price diagram, a significantly increased amplitude variation can be verified between the 2008 and the 2012–2013 prototypes, as shown in Figs. 13 and 14 and . In Fig. 13, it is possible to identify that electricity reaches higher hourly prices in both the maximum and the minimum obtained prototypes for 2008, than in the 2012–2013 diagrams. In addition, there is an increased difference in price amplitude of the 2012–2013 prototypes than in the 2008 diagrams, due to market liberalization (since January 2008) and the increase of the integration of renewable power generation (PORDATA, 2015).

Fig. 14 shows very low minimum prices of almost 0€/MWh, between 3 h00 m and 9 h00 m, probably due to the high contribution of must run renewable generation during that period and the simultaneous low consumption. This grid problem may become a market problem or an opportunity (Bundesnetzagentur, 2011) being one of the issues that both Demand Response and Distributed Storage aim to solve.

5. Conclusions

This work was caused by the need of reliable data to assess the deployment of new technologies in the smart grid, based on demand response and distributed energy storage systems.

The use of the two clustering methods enabled the comparison of results. This methodology shows that both methods return similar data, although with slightly different groupings, which is useful for prototype development.

The prototypes obtained from the average annual real prices clustering reveled a more competitive market between October 1st, 2012, and September 30th, 2013, than in the entire year of 2008. The average gross energy prices have been decreasing since 2008, pos-

Table 5
Summary of results for the cluster distributions in the ANN method, according to the number of hits and percentage for all studied profiles.

Type of profile	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	# hits	% hits	# hits	% hits	# hits	% hits	# hits	% hits	# hits	% hits
Household load demand	88	24.11%	57	15.62%	87	23.84%	80	21.92%	53	14.52%
HV load demand	52	14.21%	61	16.67%	92	25.14%	108	29.51%	53	14.48%
Renewable generation	63	17.21%	81	22.13%	97	26.50%	77	21.04%	48	13.11%
Price (2008)	36	9.84%	101	27.60%	108	29.51%	69	18.85%	52	14.21%
Price (2012–2013)	103	28.22%	115	31.51%	65	17.81%	47	12.88%	35	9.59%

sibly due to the market liberalization introduced with the Iberian electricity market settlement (MIBEL) and the corresponding operational agreements of January 2008, as well as the increase of integrating renewable generation combined with a demand reduction.

A higher similarity between the obtained prototypes and the average load profile for the household sector was also verified, contrarily to what happens when comparing to the substation's load diagram. This may be explainable, since that particular substation, serves not only residential clients, but also industrial customers.

A clear identification of the clusters with the weekday type (workday, Saturday, Sunday and holiday) was obtained for the demand profiles. As for the renewable generation, due to its high unpredictability, neither a weekday type or by season allocation was possible to attain. Moreover, the electricity prices did not present a weekday type correspondence, despite the current market situation intending to create conditions for this to happen, which suggests that the natural consumption variation is not influencing prices.

The existence of outliers which constrain the accuracy of the data allocation differs in the number of hits within each prototype. This subject presents different methods of analysis (Salvador and Chan, 2004; Sousa, Jorge, & Neves, 2013; Pujari et al., 2001), which tackle totally or partially such constraints. However, the annual representativeness of each cluster that was obtained and the stable nature of the data led the authors to focus in the overall diversity provided by the prototypes of both methods.

In Table 5 a summary is shown of the distribution of all profiles under study and that are independent of each other, including the representativeness of each cluster, in one year time span.

The proposed methodology can be used for developing plausible scenarios, using real and up-to-date data, thus enabling the development of prototypes for simulating the smart grid environment. The classification process enabled the perception of how market prices, consumption and renewable generation profiles may be grouped, in one year time.

It is thus possible to study innovative technologies under the smart grid paradigm, such as the use of energy management systems or electrical energy storage systems.

Acknowledgments

References

Albadani e, M. H., & El-Saadany, E. F. (2008). A summary of demand response in electricity markets. *Electric Power Systems Research, 78*, 1986–1996.

Benítez, I., Quijano, A., Díez, J.-L., & Delgado, I. (2014). Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Electrical Power and Energy Systems, 55*, 437–448.

Bundesnetzagentur. (2011). *Smart grid und smart market, bundesnetzagentur.* Germany: Bonn [December].

Divya, K. C., & Østergaard, J. (2009). Battery energy storage technology for power systems—an overview. *Electric Power Systems Research, 79*, 511–520.

Electric Power Research Institute. (2008). *The green Grid—Energy savings and carbon emissions reductions enabled by a smart grid.* California, USA: EPRI. June.

European Union Advisory Council. (2010). *Smartgrids—Strategic deployment document for europe's electricity networks of the future.* EUAC, EU. April.

Everitt, B. S., Landau, S., Leese, M., & Stalh, D. (2011). *Cluster analysis* (5th ed.). Wiley—A John Wiley and Sons, Ltd., Publication.

Gonçalves, J., Neves, L., & Martins, A. (2015). Multiobjective assessment of distributed energy storage location in electricity networks. *International Journal of Sustainable Energy*, 1–15. http://dx.doi.org/10.1080/14786451.2015.1066787

Kohonen, T. (1998). The self-organizing map. *Neurocomputing, 21*, 1–6.

Lassila, J., Haakana, J., Tikka, V., & Partanen, J. (2012). Methodology to analyze the economic effects of electric cars as energy storages. *IEEE Transactions on Smart Grid, 3*(March), 506–516.

Leou, R.-C. (2012). An economic analysis model for the energy storage system applied to a distribution substation. *Electrical Power and Energy Systems, 34*, 132–137.

Mahmoudi, N., Eghbal, M., & Saha, T. K. (2014). Employing demand response in energy procurement plans of electricity retailers. *Electrical Power and Energy Systems, 63*, 455–460.

Mahmoudi-Kohan, N., Moghaddam, M., & Bidaki, S. M. (2009). Evaluating performance of WFA K-means and modified follow the leader methods for clustering load curves. *In IEEE/PES Power Systems Conference and Exposition, 2009.*

Maimon, O., & Rokach, L. (2005). *Data mining and knowledge discovery handbook.* New York: Springer-Verlag.

Miguel, P., Neves, L., & Martins, A. (2014). Methodology to simulate the impact of a large deployment of a residential energy management system in the electricity grid. *Electric Power Systems Research, 116*, 399–407.

Mooi, E., & Sarstedt, M. (2011). *A Concise Guide to Market Research.* Berlin, Heidelberg: Springer.

Nunes, P. S. (2011). *Automatic meter reading systems as a precondition for enhancing energy efficiency and increasing competion in the energy sector.* pp. 43–50. Energia e Futuro. October–December.

PORDATA. (2015). *Base de dados Portugal contemporâneo, DGEG/MAOTE/Pordata.* [Online] Accessed 05.09.15. Available:. http://www.pordata.pt/Portugal/Ambiente+de+Consulta/Tabela/5690459

Pacific Northwest National Laboratory. (2010). *The smart grid: an estimation of the energy and CO2 benefits.* Washington, USA: PNNL. January.

Pujari, A. K., Rajesh, K., & Reddy, D. S. (2001). 'Clustering techniques in data mining—a survey. *IETE Journal of Research, 47*, 19–28.

Ribeiro, P. F., Johnson, B. K., Crow, M. L., Arsoy e, A., & Liu, Y. (2001). Energy storage systems for advanced power applications. *Proceedings of the IEEE, 89*, 1744–1756.

Sá, J. L. P. (2011). *Um périplo pelos roteiros alternativos nas Smartgrids.* Energia e Futuro.

Salvador, S., & Chan, P. (2004). Determining the number of Clusters/Segments in hierarchical Clustering/Segmentation algorithms, em tools with artificial intelligence. In *ICTAI 2004. 16th IEEE international conference.*

Sigrist, L., Lobato e, E., & Rouco, L. (2013). Energy storage systems providing primary reserve and peak shaving in small isolated power systems: an economic assessment. *Electrical Power and Energy Systems, 53*, 675–683.

SmartGrids, E. T. P. (2012). *SmartGrids SRA 2035 strategic research agenda update of the SmartGrids SRA 2007 for the needs by the year 2035.* European SmartGrids Platform. March. pp. 74.

Sousa, J. C., Jorge, H. M., & Neves, L. P. (2013). Short-therm load forecasting based on support vector regression and load profiling. *International Journal of Energy Research, 38*(3), 350–362. http://dx.doi.org/10.1002/er.3048

Sousa, J. M. C. (2006). *Metodologia para determinação de perfis de consumidores de Baixa Tensão.* Coimbra: Faculdade de Ciências e Tecnologia da Universidade de Coimbra.