Data mining for remuneration of consumers demand response participation

Catarina Ribeiro[1,2,3], Tiago Pinto[1,2], Zita Vale[2], José Baptista[3,4]

[1] GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development
[2] Polytechnic of Porto (ISEP/IPP), Porto, Portugal
{acrib, tcp, zav}@isep.ipp.pt
[3] UTAD – Universidade de Trás-os-Montes e Alto-Douro
[4] CPES - INESCTEC
baptista@utad.pt

Abstract. With the implementation of micro grids and smart grids, new business models able to cope with the new opportunities are being developed. Virtual Power Players are a player that allows aggregating a diversity of entities, to facilitate their participation in the electricity markets and to provide a set of new services promoting generation and consumption efficiency, while improving players` benefits. The elastic behavior of the demand consumption jointly used with other available resources such as distributed generation (DG) can play a crucial role for the success of smart grids. This paper proposes methodologies to develop strategic remuneration of aggregated consumers with demand response participation, this model uses a clustering algorithm, applied on values that were obtained from a scheduling methodology of a real Portuguese distribution network with 937 buses, 20310 consumers and 548 distributed generators. The normalization methods and clustering methodologies were applied to several variables of different consumers, which creates sub-groups of data according to their correlations. The clustering process is evaluated so that the number of data subgroups that brings the most added value for the decision-making process is found, according to players characteristics.

Keywords: Clustering, Distributed Generation, Smart Grid, Demand Response.

NOMENCLATURE

CDR_{calc}	Consumers remuneration
$CDR_{calc}Max$	Maximum remuneration value obtained in a cluster
$CDR_{calc}Med$	Medium remuneration value obtained in a cluster
$CDR_{calc}Min$	Minimum remuneration value obtained in a cluster
$CDR_{calc}Rem$	Remuneration value when affected by a factor
L_c	Consumption value
ML_c	Largest consumption value
$N_{c,h}$	Common normalized load
$P_{RTP\ initial}$	Initial consumption of each load
$P_{RTPDR\ max}$	Maximum value of reduced energy per consumer
$Product_{RTP\ MAX}$	Maximum value of reduced energy per consumer
SML_h	Largest consumption value with customized normalization

$SN_{c,h}$	Customized normalized load
μ_i	Object value
CDI	Cluster Dispersion Index
CDR	Cost of reduction
co	Set of all considered consumers
d	Euclidian distance between two points
Elast	Consumers elasticity
Factor	Influence variable
Gen	System production
h	Hour
Income	Recipe for consumer
J	K-means objective function value
MIA	Mean Index Adequacy
R	Load profile of all consumers
w,x,y,z	Influence variables with specific weights
x	Centroid of the cluster

1 Introduction

The energy system paradigm has changed completely with the massive introduction of renewable energy sources and the introduction of free competition [1]. The uncertainty brought by renewable energy sources' dependency on natural factors, requires the system to use consumers' flexibility to balance the generation variability [2]. However, adequate remuneration schemes for consumers' flexibility are lacking, hence delaying the widespread implementation of demand response programs to incentivize consumers participation. Consumers central role in future power systems, sustained by an active participation in energy markets is therefore, dependent on the models to attract both consumers and aggregators to market transactions.

Important developments concerning market players modelling and simulation including decision-support capabilities can be widely found in the literature [3]; however, these are mostly directed to players participation in the market, and energy resources management, while neglecting the development of flexibility remuneration models that are fair to both consumers and aggregators, while considering the needs from system. Aggregation of small-scale distributed resources, as well as their operation, in a competitive environment leads to the creation of Virtual Power Players (VPP). VPP can aggregate diversity of players and of energy resources, including demand response (DR), making them profitable [4]. The aggregation of players allows the creation of groups (clusters), aiming the capture of common characteristics that better define the resources in a specific context [5],[6].

This paper introduces a methodology for dynamic definition of consumers remuneration for demand response participation. An optimize remuneration method is proposed, considering the potential economic benefit for both the aggregators and the consumers. The proposed model also considers the consumers characteristics (consumption, elasticity, participation in demand response programs) and the needs from the system, such as the volume of generation from renewable sources. The proposed model is applied to a set of consumers, using a data mining process. The case

study considers several variables of different consumers of a real smart grid. In this way dynamic remuneration schemes are defined, which contribute to reduce consumption peaks, increase the use of renewable energy and reflect the wholesale market price on consumers.

2 Proposed methodology

The proposed model is directed to the demand response management from an aggregator that may participate in energy markets. Initially, the aggregator identifies the aggregated consumers and creates groups of consumers according to their similarity, so that the remuneration process may be facilitated depending on players' characteristics, such as consumption, elasticity and energy cost. Clustering methodologies are used in this work to determine the optimal consumer groups to be considered for the application of these methodologies. Although a wide variety of clustering algorithms can be found in the literature, there is no single algorithm that can, by itself, discover all sorts of cluster shapes and structure [7]. K-means clustering algorithm [8] has been used, as it proves to be a robust model for distinct applications: K-means minimizes the distance from each point to the centre of the respective cluster, as defined in (1).

$$J = \min \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||^2 \tag{1}$$

where μ_i is the mean of points in C_i, *i.e.* the cluster *centroid*. To determine the quality of partition of players into different clusters, the clusters validity indices MIA and CDI [9] have been used, as formalized in (2) and (3) respectively.

$$MIA = \sqrt{\frac{1}{K} \sum_{k=1}^{K} d^2(x^{(k)}, \mu^{(k)})} \tag{2}$$

$$CDI = \frac{\sqrt{\frac{1}{K} \sum_{k=1}^{K} \left[\frac{1}{2.n^{(k)}} \sum_{n=1}^{n^{(k)}} d^2(x^{(m)}, \mu^{(k)}) \right]}}{\sqrt{\frac{1}{2K} \sum_{k=1}^{K} d^2(x^{(k)}, R)}} \tag{3}$$

The K-means algorithm has been used to normalized and non-normalized values of several different consumers' characteristics, in order to allow taking conclusions on the quality of partition. The considered normalization methods are defined in (4)(5) for Regular Normalization and (6)(7) for Customized Normalization. The Regular Normalization process is defined as:

$$N_{c,h} = \frac{L_{c,h}}{ML_c}, \forall c \in co \tag{4}$$

$$ML_c = \max(L_c), \forall c \in co \tag{5}$$

The Customized Normalization method normalizes data using each consumers' load value at each period divided by the largest recorded value of all loads in all periods, it is formalized in (6) and (7).

$$SN_{c,h} = \frac{L_{c,h}}{SML_h}, \forall c \in co \tag{6}$$

$$SML_h = \max(L_{co,h}), \forall c \in co \tag{7}$$

The remuneration method is based on different consumers' characteristics. The calculated value of the remuneration associated with demand response is influenced by a factor that comprises the different variables that influence the remuneration calculation. The formulation is summarized as follows.

Equations presented in (8) and (9) are used to represent the remuneration calculation for each consumer.

$$CDR_{calc} = CDR \times Factor \tag{8}$$

$$Factor = w \times P_{RTPR_{MAX}} + x \times P_{RTP_{initial}} - y \times Gen - z \times Elast \tag{9}$$

The parameters w, x, y, z influence variables with different weights and different combinations, represented in Table 1, in order to represent the relative importance of each variable to the remuneration calculation; where $w + x + y + z = 1$.

Table 1: Factors different values, defined by the author

Combination	w	x	y	z
C1	0.25	0.25	0.25	0.25
C2	0.7	0.1	0.1	0.1
C3	0.1	0.7	0.1	0.1
C4	0.1	0.1	0.7	0.1
C5	0.1	0.1	0.1	0.7

Once the remuneration calculation is made, the evaluation of the income that results for the consumer is made. Following are the different approaches to evaluate the income, associated to the application of each of the remuneration methods.

Approach 1 - based on the value of the demand response remuneration of all the players that make up a particular cluster, which will be the cost of reduction when the consumer is paid to reduce consumption. The methods are formalized in (10), (11), (12) and (13).

$$\text{DMax } Income_{(\mathcal{C})} = CDR_{calc}MAX_{(u.m./kWh)} \times Preduct\_RTPMAX_{(kWh)} \tag{10}$$

$$\text{DMin } Income_{(\mathcal{C})} = CDR_{calc}MIN_{(u.m./kWh)} \times Preduct\_RTPMAX_{(kWh)} \tag{11}$$

$$\text{DMed } Income_{(\mathcal{C})} = CDR_{calc}MED_{(u.m./kWh)} \times Preduct\_RTPMAX_{(kWh)} \tag{12}$$

$$\text{DRem } Income_{(\mathcal{C})} =$$
$$CDR_{calc}Rem_{(u.m./kWh)} \left(Preduct\_RTPMAX_{(kWh)} \times \frac{CDR_{calc}\,Rem_{(u.m./kWh)}}{CDR_{(u.m./kWh)}} \right) \tag{13}$$

In methods DMax, DMin and DMed, CDR_{calc} represents the remuneration value (max, min or med) obtained in a certain cluster. $Product\_RTPMAX$ represents the reduction of maximum individual consumption of each cluster, to participate in *real-time pricing* program. The DRem method considers that the value of the remuneration $CDR_{calc}Rem$, is calculated considering that the value of the remuneration associated to the demand response is affected by a factor.

Approach 2 - is based on remuneration processes (13), but now adjusted according to elasticity on a proportional base. Again, this calculation is performed for all players that form each cluster. The methods are formalized in (14), (15), (16) and (17).

$$\text{EMax } Income_{(€)} = \tag{14}$$
$$CDR_{calc}MAX_{(u.m./kWh)} \times \left(ProductRTPMAX_{(kWh)} \times \left(\frac{CDR_{calc}MAX_{(u.m/kWh)}}{CDR_{8u.m/kWh)}} \right) \right)$$

$$\text{EMin } Income_{(€)} = \tag{15}$$
$$CDR_{calc}Min_{(u.m./kWh)} \times \left(ProductRTPMAX_{(kWh)} \times \left(\frac{CDR_{calc}Min_{(u.m/kWh)}}{CDR_{8u.m/kWh)}} \right) \right)$$

$$\text{EMed } Income_{(€)} = \tag{16}$$
$$CDR_{calc}Med_{(u.m./kWh)} \times \left(ProductRTPMAX_{(kWh)} \times \left(\frac{CDR_{calc}Med_{(u.m/kWh)}}{CDR_{8u.m/kWh)}} \right) \right)$$

$$\text{ERem } Income_{(€)} = \tag{17}$$
$$CDR_{calc}Rem_{(u.m./kWh)} \times \left(ProductRTPMAX_{(kWh)} \times \left(\frac{CDR_{calc}Rem_{(u.m/kWh)}}{CDR_{8u.m/kWh)}} \right) \right)$$

3 Case study

This case study shows the suitability of the proposed clustering methodology and normalization approaches to solve the problem of players remuneration, considering players with heterogeneous technologies and behaviors. In order to test the adequacy of the method the clustering algorithm has been applied to values obtained from a scheduling methodology of a real network with 30 buses, supplied by one high voltage substation (60/30 kV) with a total number of 937 buses and 464 MV/LV transformers, with 20310 consumers and 548 distributed generators. Fig. 1 shows the summarized scheme of the distribution network [10].

The K-means algorithm has been used to perform the clustering process using non-normalized values of load and also normalized values, using both the regular normalization and customized normalization, for k=3 and 4 for a total of five different groups. Other tests carried out in previous studies, [9], [10], have revealed that the use of the clustering methodology associated to the Customization methodology allows for relevant results to be obtained. In this case study, the lower indices values of MIA/CDI, were achieved with customized normalization for k=3 in group G5_Elast, as it is possible to see in Fig.2.
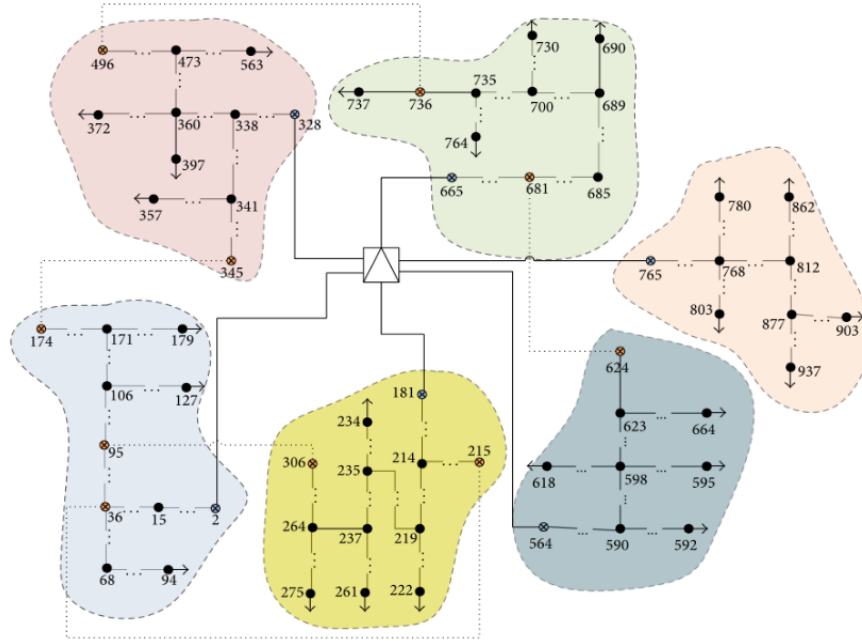
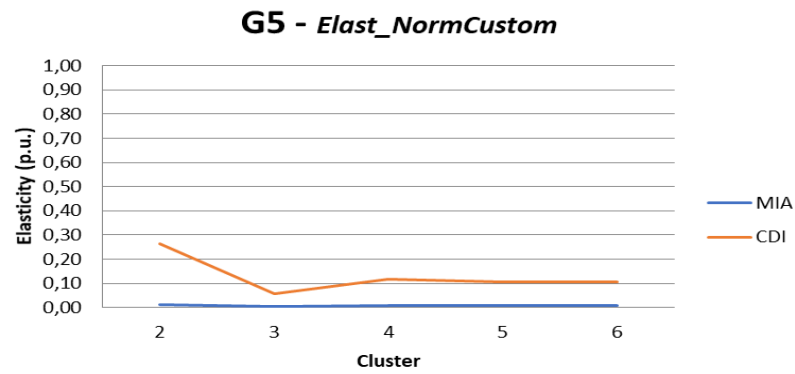**Fig. 1**. Distribution network used for the SG simulation.

Fig.2. MIA/ CDI indices for G5_elast with customized normalization.

Remuneration analysis

The methodologies for the definition of the remuneration value of consumers have been tested and the results are now analysed.
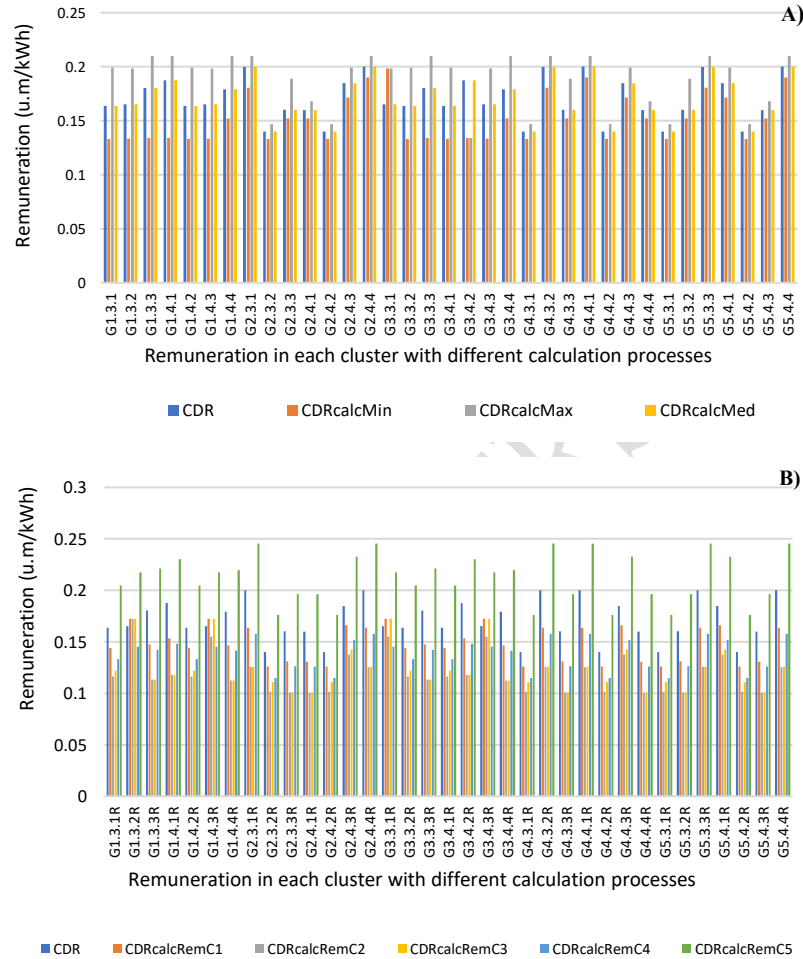

Fig. 3. Remuneration values obtained with methods Min, Max, Med in grafic A) and method Rem in grafic B).

Figure 3 A) shows that the results of new CDRcalc have the expected behavior when compared to the base CDR values. When the remuneration is calculated through the minimum value, it is smaller than the base CDR, when it is calculated by the maximum value it is greater, and when we use the average of the values it will be equal to the base CDR.

From a practical point of view, these results are useful for the aggregator (or the entity that defines the remuneration), since allows to manage the overall remunera-

tion to be defined for the group, depending on its strategy and objectives. To maximize their own profits, they can use the minimum amount of remuneration (with consumers having a higher starting value of remuneration); but to attract more customers and increase competitiveness they can use the maximum value (where consumers who have a lower value of remuneration will benefit). In the base it is possible to simply use the average compensation value, which will be the most balanced.

In graph B) a more careful analysis of the compensation values obtained is necessary since the influence factors w, x, y and z in the different combinations (C1, C2, C3, C4 and C5) represented in Table 1, are considered to affect the system in several ways. When assigning different weights to the components w, x, y and z, we intend to analyse how there is an adjustment in the remuneration, related to the changes in the system. Analysing B), we would have to do a very extensive analysis in each group, so in order to make the analysis more assertive we will focus on the results obtained for the G5_Elast group, k = 3. In this case, allocated to cluster 1 were 147 loads, 9312 loads in cluster 2 and 10250 in cluster 3. In order to do this analysis, it is necessary to bear in mind the values that in this case study we are considering for generation and for consumption. For the generation we have a value considered high in relation to what is normal, total daily production = 1.2303 p.u, in the case of consumption we are considering a total daily value that corresponds to the total average of each cluster, in the different scenarios.
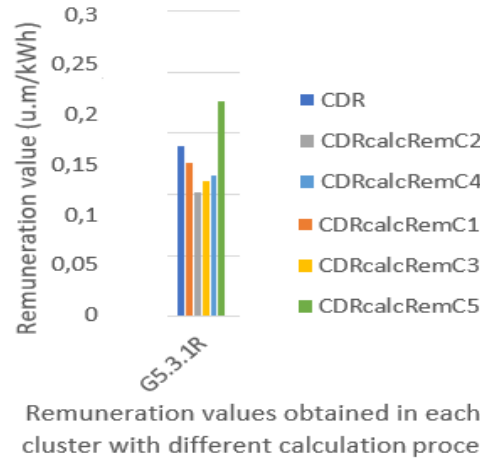

Fig. 4. Remuneration values obtained with method Rem.

In the result of the graph corresponding to G5.3.1R, Fig.4, an equal weight is considered for all the parameters that make up the CDRcalc remuneration (0.25, 0.25, 0.25, 0.25), the generation value is considered high and the consumption has a low value, 0,13995pu. In this situation, when we compare the value of CDRcalc with the average value of CDR base in cluster 1, we find that the value of the remuneration decreases slightly with respect to the base value. When compared to CDRcalcRemC2, where a fairly high weight was attributed to the initial reduction P_RTPDR_MAX, the remuneration drops even further. In this case it drops substantially to1/4 of the initial value of the remuneration. This phenomenon occurs since it

is expected that when the system has already a high reduction value, the remuneration to be given to consumers is lower since there is no need for an incentive.

Analysing CDRcalcRemC3 we find that by assigning a high weight to the factor associated with consumption (x), the value of remuneration rises slightly in relation to CDRcalcRemC2 although it remains lower when compared to the initial remuneration, since the initial value considered for consumption was low, although the factor makes it rise slightly, the value of the remuneration remains low. In the case of the CDRcalcRemC4 scenario, a greater weight is attributed to the generation, which already has a high value and will increase the value of the remuneration. In the case of CDRcalcRemC5 it is verified that the value of the remuneration goes up considerably, here a higher weight is attributed to the component of the elasticity.

When we analyse the CDRcalcRem5 result, we find that by increasing the elasticity component (z) in the different scenarios, the value of the remuneration increases considerably compared to the base CDR. In the other groups represented in the graph this trend is maintained for all the results obtained in this simulation.

After calculating the remuneration and respective analysis of results, the income results for the consumer is evaluated. The following are the different approaches to revenue evaluation, associated to the application of each of the remuneration methods.

Income results of methods Max, Min e Med (D and E)

The graphs in Fig.5 show the results obtained for the different scenarios considering the sum and the total average of the cluster, in graphs C) and D), respectively.

Analysing the graphs A) and B), is visible that in general, regardless of the scenario considered, the results tend to assume a very similar behavior within the same method. Taking the analysis in detail, we focus on the results of the G5_Elast group, the group already selected in the previous study and for example, in the graph A) of the sum, since the behavior of the results is very similar in both graphs. Thus in graph A) we can verify that: with the DMin and EMin, in the scenarios G5.3.1, G5.3.2, G5.4.2 and G5.4.3 the results of the revenue are larger when the elasticity is not considered, whereas in the other scenarios the opposite is true. From a practical point of view, it should be remembered that the elasticity corresponds to the value that the consumer can reduce or increase the reduction. If there is no elasticity it will always reduce the same at whatever price. But with elasticity, if the price is good (higher than the base CDR), it will reduce more, and then make more money, but if the price is low, it will reduce less because it no longer interests you so much to be having the disadvantage of reducing consumption for a less good price, and consequently having less income. In the mentioned scenarios it is verified that the value of CDRcalc is smaller than the CDRbase. Regarding the results of DMax and EMax, it is verified that in G5.3.1 and G5.4.2 the value of the revenue is higher when the elasticity is not considered, in the remaining scenario the revenue is always greater in method E, when compared to the method D. Being significantly larger in scenarios G5.3.3, G5.4.1 and G5.4.4. In the case of DMed and EMed results, similar to what happened in the Max method only in scenario G5.3.1 and G5.4.2, the remuneration without elasticity is higher than the remuneration that considers the elasticity. In the other scenarios value is always higher when considering Method E, and in scenarios G5.3.2 and G5.4.3 the values are very similar, with greater discrepancy in scenarios G5.3.3, G5.4.1 and G5.4.4.
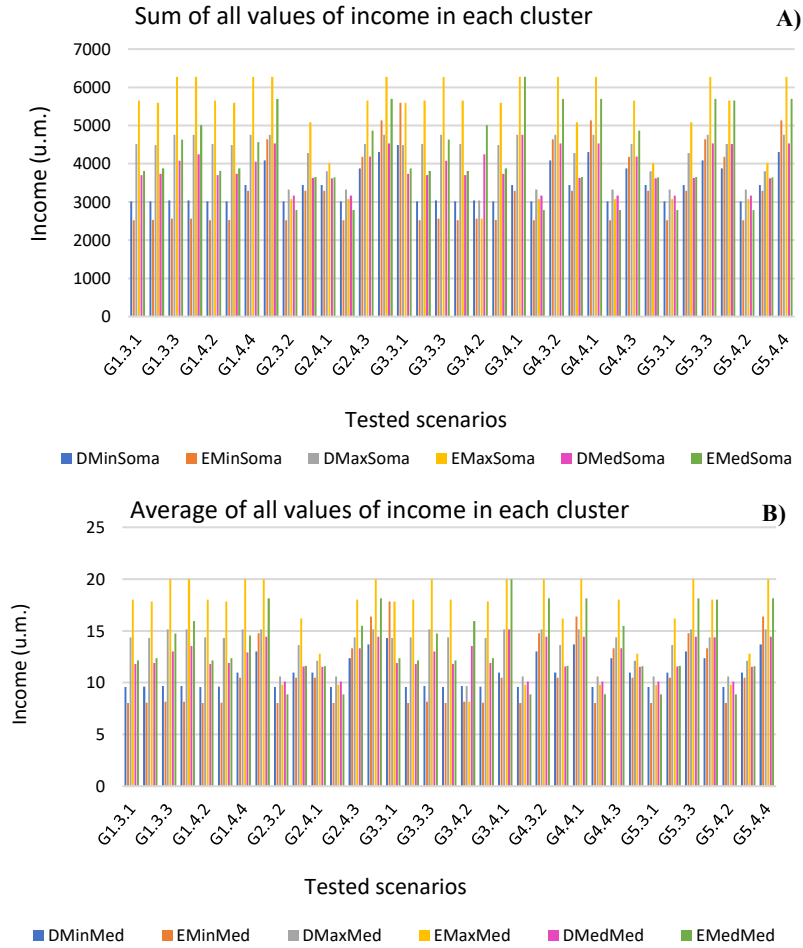
Fig. 5 – Results for Sum A) and Average B) of income in each cluster, considering the scenarios defined for methods D and E, Min, Max and Med.

Income results of method Rem (D and E)

Since we have a large number of data to the graphics are too large to present in this paper. Analysing the results, we find that in general, regardless of the scenario considered, the results tend to assume a very similar behaviour within the same method. Now, let's focus on the results of the G5_Elast group, the group already selected in the previous study, we can thus verify that: when analysing the results obtained with DSomaC1 and ESomaC1, in the scenarios G5.3.1, G5.3.2, G5.4.2 and G5.4.3 the results of the revenue are higher when the elasticity is not considered, while in the other scenarios the contrary. In this scenario C1 is assigned the same weight to all the parameters that make up the revenue calculation. This phenomenon may explain

the fact that the results are similar to the previous methods; when we analyse scenario C2, where a much larger weight is applied in the initial participation or initial value of the reduction, we find that the results obtained for the compensation through method D are higher than with the use of method F in all groups tested; the previous phenomenon occurs again in scenario C3, when a much higher weight is attributed to the value of consumption; for scenario C4, the component y is assigned a very high weight, which affects the generation parameter. In this case, earnings without elasticity will be higher, although in G5.3.3 G5.4.1 and G54.4 the difference is minimal, obtaining very similar values through the two methods; in scenario C5 a high weight is attributed to the component that affects the elasticity parameter. In this scenario, in all the groups considered, the value of the remuneration is always greater in the method E, in relation to the method D.

4 Conclusions

The decision support model proposed in this paper enables attracting consumers to assume an active role in the system, through fairer remuneration strategies. The best clustering results occur for k=3, in any of the groups relative to the five variables. With MIA and CDI validation was verified that the lowest errors were verified for 3 clusters. When applied to the data, the Customized Normalization was verified that MIA/CDI error is lower when compared to the results obtained with Regular Normalization process.

The application of the remuneration methodologies allowed to show that it is possible to define remuneration values with different natures for the aggregator. It is possible to group consumers according to their similarity, to identify the basic values of remuneration taking into account the a priori values of each consumer and to calculate average minimum and maximum values of remuneration, appropriate for each group of consumers. It is also possible to verify that by the proposed methodology it is possible to define dynamic remuneration values, which depend not only on the preferences of the aggregator and the consumer in terms of prices, but also on the variation of the generation in each moment, and the incentive to participate in programs of taking into account the basic elasticity of each consumer. It allowed evaluate the impact from the dynamic remuneration schemes by assessing the potential incomes of each consumer when providing consumption flexibility. From the presented results it is possible to verify that the value of remuneration decreases when the elasticity increases and decreases when the generation also increases, being an automatic way to stimulate the participation of the consumers and at the same time to guarantee a fair remuneration for both the participants and the aggregators.

Acknowledgements

References

1. Ciarreta, et al, "Has renewable energy induced competitive behavior in the Spanish electricity market?," Energy Policy, 104, 171–182, 2017.
2. L. Tjørring, et al, "Increasing the flexibility of electricity consumption in private households: Does gender matter?," Energy Policy, 118, 9–18, 2018.
3. P. Ringler, D. Keles, and W. Fichtner, "Agent-based modelling and simulation of smart electricity grids and markets - A literature review," Renew. Sustain. Energy Rev., 57, 205–215, 2016.
4. T. Pinto, Z. A. Vale, H. Morais, I. Praça, and C. Ramos, "Multi-agent based electricity market simulator with VPP: Conceptual and implementation issues," in Proc. IEEE Power & Energy Society General Meeting, 2009, Jul. 26–30, 2009, pp. 1–9.
5. L. Kaufman, & P. Roussesseeuw, "Finding groups in data: an introduction to cluster analysis". New Yoek, Wiley, 1990.
6. V. Figueiredo et al., "an electric energy consumer characterization framework based on data mining techniques". IEE Trans. on Power Syst., vol.20, no.2, pp.596-602, May 2005.
7. Anil K. Jain et. al., (1999) "Data Clustering: A Review." ACM Computing Surveys, 31 (3). pp. 264-323.
8. Anil K. Jain, "Data Clustering: 50 years beyond K-Means". Pattern Recognition Letters, Elsevier, Vol. 31, Issue 8, pp.651-666, June 2010
9. Chicco et al., "Support Vector Clustering of Electrical Load Pattern Data". IEEE Transactions on Power Systems, vol.24, no.3, pp.1619-1628, August 2009.
10. P. Faria, J. Soares, Z. Vale et al., "Modified particle swarm optimization applied to integrated demand response and DG resources scheduling", IEEE Trans. Smart Grid, vol. 4, no. 1, pp. 606-616, 2013.
11. Catarina Ribeiro, Tiago Pinto, Zita Vale, "Customized Normalization Method to enhance the Clustering process of Consumption Profiles", Ambient Intelligence - Software and Applications – 7th International Symposium on Ambient Intelligence (ISAMI'16), pp 67-76, 2016, ISSN 2194-5357.
12. Catarina Ribeiro, Tiago Pinto, Zita Vale, José Baptista. "Data Mining for Prosumers Aggregation considering the Self-Generation". Distributed Computing and Artificial Intelligence, 14th International Conference. DCAI 2017. Advances in Intelligent Systems and Computing, vol 620. Springer, Cham.