Electricity market price analysis using time series clustering

Ana Martins
ISEL/ADM
Lisbon, Portugal
anamartins@deea.isel.ipl.pt

João Lagarto
ISEL/ADEEEA; INESC-ID
Lisbon, Portugal
jlagarto@deea.isel.ipl.pt

Margarida G. M. S. Cardoso
ISCTE-IUL; BRU-IUL
Lisbon, Portugal
margarida.cardoso@iscte-iul.pt

Abstract— **The creation of the internal market of electricity has long been a goal of the European Union, for which it has established common rules through the directive 2009/72/EC. In this context, the analysis of electricity markets operation of the different countries that will form the internal market is of the utmost importance.**
In this work, we use clustering techniques to analyze 26 time series of day-ahead electricity prices from European markets between 2015 and 2018 in order to identify different price patterns. The cluster technique proposed uses a combination of three dissimilarity measures for time series: Euclidean, Pearson correlation based and periodogram based.
Results show that there is a clear distinction between Northern markets, especially Nord Pool, and Southern markets, MIBEL and Italy. Moreover, results also show that despite some market prices presenting similar behaviors, a full integrated European electricity market is yet to be accomplished.

Index Terms-- **Clustering, Electricity markets, Market integration, Time series.**

I. Introduction

The 2009/72/EC directive stresses the importance of a European internal market in electricity, which aims to enable European citizens and businesses to choose their supplier, to create new business opportunities and enhance cross-border trade with the purpose of ensuring efficiency gains and competitive prices and contribute to security of supply and sustainability. Therefore, one of the main purposes of the 2009/72/EC is the development of an internal electricity market through an interconnected network.

Many works have devoted their attention to the degree that electricity markets of the European countries have been integrated. As an example of such works, [1] analyzes the degree of market integration among 25 European electricity markets between January 2010 and June 2015. Firstly, the authors use cointegration analysis to study long-run price relationships between adjacent and non-adjacent markets. The results show that market integration increased from 2010 to 2012 and then decreased until 2015.

In [2], the authors analyze if the APXUK, Nord Pool and Phelix electricity markets became more efficient and integrated after the introduction of the directive 2009/72/EC. To achieve these goals, the authors use a combination of econometric models such as Variance Ratio test, cointegration techniques and the GARCH model. The results suggest that the analyzed markets have a segmented behavior, thus suggesting the need of more action from the governments to achieve full integrated internal market.

Reference [3] also assesses if a set of European electricity market prices were converging and if special events on the supply side, such as new interconnections, market coupling and increasing share of renewable production, could have affected the speed of price convergence. The analysis is performed for nine electricity spot markets and for four one-month-ahead markets. The authors follow a time-varying fractional cointegration analysis and the results show that one-month-ahead prices have become more resilient to shocks during the analyzed period and that the studied electricity markets are increasingly integrated.

In a different approach, [4] analyzes the prices of five different electricity markets (OMEL, NEM, Nord Pool, Ontario and Austria), comprising a set of fourteen hourly price time series from 2005 to 2009. The authors use a time series clustering method to group time series according to the level of dependency among them. The clustering approach combines a permutation-based coding of time series with several distance measurements for discrete distributions and different hierarchical clustering methods. The analysis highlights the fact that the classification results agree with the degree of integration of the analyzed markets.

In general, clustering methods aim to organize a set of items into homogeneous (with similar items within a given group) and well separated groups (with dissimilar items in different groups).

The clustering methods for time series data have been used in many scientific areas [5]. In electricity markets, [6] applies clustering algorithms to historical data of consumption, wind generation and electricity prices to obtain daily profiles to be

This work was supported by national funds through FCT – Fundação para a Ciência e Tecnologia, under projects UID/CEC/50021/2019 and UID/GES/00315/2019.

978-1-7281-1257-2/19/$31.00 ©2019 IEEE

used in residential demand response and small-scale energy storage simulations. The clustering algorithms use the Ward hierarchical method and a self-organizing neural network. Also [7] uses a combination of Artificial Neural Networks (ANN) and clustering algorithms to forecast day-ahead electricity market prices without data pre-processing. The use of clustering algorithms aims to improve the ANN's capability by generating smaller training sets with more similar patterns. The clustering algorithm uses K-means with Euclidean distance to group the training sets.

In this work, we use clustering techniques to analyze 26 electricity market price time series of five day-ahead electricity markets (MIBEL, Nord Pool, Italy, Germany and France). The aim of this paper is twofold. Firstly, we cluster time series regarding day-ahead electricity market prices of several European countries in distinct periods of time; secondly, the temporal stability of the clustering solutions obtained is analyzed.

In the literature one can find different proposals for measuring the distance or dissimilarity among time series. Each one of these measures is based only on some characteristics of the data. In this work, we propose to use a combination of three dissimilarity measures in order to consider more diversity of data features in the clustering task.

In order to consider simultaneously the relational data of multiple dissimilarity matrices, [8] proposed a fuzzy clustering algorithm, in [9] a hard-clustering approach is proposed. In this work, we use the Ward hierarchical algorithm. In order to complement the use of dendrograms to determine an adequate number of clusters, we also resort to classical MDS and to a cohesion-separation index.

This paper is organized as follows. Section II presents a brief description of the electricity markets analyzed. Section III deals with the methodological approach and section IV presents the main results of the application. Finally, section V draws the main findings of this work.

II. Brief Description of the Analyzed Electricity Markets

A. The Iberian Electricity Market (MIBEL)

The MIBEL entered into force on the 1st of July 2007 by the integration of the Portuguese electricity market with the Spanish market which had already existed since 1998. During 2018, 276 TWh of energy were traded in the day-ahead market, which corresponded to 83% of the consumption of Portugal and Spain [10]. Moreover, the average prices in the Portuguese and Spanish areas were 57.45 €/MWh and 57.29 €/MWh, respectively. On average, the interconnection capacity between the Portuguese and the Spanish power systems was about 3100 MW from Spain to Portugal and about 2200 MW in the opposite direction [11]. This led to a price difference between the two areas of less than 1 €/MWh in 95.8% of the hours. The same did not happen with France where the price difference was less than 1 €/MWh in only 27.5% of the hours. The interconnection capacity between the Spanish and French power systems was about 2600 MW from France to Spain and about 2200 MW from Spain to France.

B. The Italian electricity market

The Italian market is constituted by 19 price zones (six geographical, one pole of limited production and twelve foreign virtual zones). In 2017, The average spot price was 53.95 €/MWh and the total volume traded was 211 TWh, that is, 65% of the total demand [12]. Mainly, the Italian market has interconnections with France, Switzerland, Austria and Slovenia. The interconnection capacity between Italy and France was about 2500 MW for export and about 1000 MW for imports, while with Austria was much lower, 250 MW for imports and 100 MW for exports.

C. Nord Pool

The Nord Pool is formed by the electricity markets of the Nordic countries (Norway, Sweden, Denmark and Finland) and of the Baltic countries (Latvia, Lithuania and Estonia) and comprises 16 bidding areas. In 2018, the average system price was 43.99 €/MWh and the total volume traded was 396 TWh, which corresponded to 94% of total consumption. The interconnection capacity between Nord Pool (Sweden and Denmark) and the German power system was roughly 1900 MW towards Germany and 2100 MW from Germany [13].

D. France and Germany electricity markets

France and Germany are part of the EPEX power market [14], which comprises also the markets of Austria, United Kingdom, the Netherlands, Belgium, Luxemburg and Switzerland. In 2017, the total volume traded in the EPEX market was 517 TWh, of which 105.7 TWh was traded in the French spot market, with an average price of 44.97 €/MWh. The traded volume in the French market corresponded to 23% of total consumption. Besides the above-mentioned interconnection capacity values with Spain and Italy, the interconnection capacity between France and Germany (including Belgium) is 7000 MW from France to Germany and 9200 MW in the opposite direction.

In 2017, in the German electricity market, which is integrated with the Luxemburgish and, until September 2018, with the Austrian market, 233.2 TWh were traded in the day-ahead market. This traded volume corresponded to 50% of the total consumption, with an average price of 34.19 €/MWh. Also in 2017, the German and French markets had the same price in 37% of the hours, which is an increase from the 36% and 27%, in 2016 and 2015, respectively.

III. Methodological Aproach

A. The Distance Between Time Series

Selecting a dissimilarity or a distance measure is a critical issue in clustering time series. In the literature one can find different proposals for measuring the distance among time series. [5] presents a review on these measures. The measure of similarity between the time series can be defined by considering the raw time series data, some feature vector of lower dimension previously extracted from the data or, by comparing the parameters of underlying models assumed for each one of the time series. Naturally, the choice of a dissimilarity measure depends heavily on the characteristics of the specific time series considered.

In this work, three measures of distance are considered, each of them reflecting different characteristics of the data. Let $X_t = (x_1, x_2, \ldots, x_T)'$ and $Y_t = (y_1, y_2, \ldots, y_T)'$ be two real univariate time series.

The first distance measure considered is the Euclidean distance:

$$d_{Eucl} = \left(\sum_{t=1}^{T} (x_t - y_t)^2 \right)^{1/2}. \quad (1)$$

This is a one-to-one measure that considers the closeness of the observations indexed in time. It is very sensitive to signal transformations [15].

The Pearson correlation-based measures also compare data profiles in time. In this work we consider the Rooted Normalized One-Minus-Correlation distance measure proposed by [16]:

$$d_{RNOMC} = \sqrt{\frac{1 - \rho_{X,Y}}{2}}, \quad (2)$$

with $\rho_{X,Y}$ representing the Pearson correlation between the time series X_t and Y_t This distance is invariant to scale and $0 \leq d_{RNOMC} \leq 1$.

The third distance measure is based on the periodogram of time series [17]. The periodogram expresses the contribution of the various frequencies, or cyclical components, to the variability of the series. Let $P_x(w_j) = (1/n)\left|\sum_{t=1}^{T} x_t e^{-itw_j}\right|^2$ and $P_y(w_j) = (1/n)\left|\sum_{t=1}^{T} y_t e^{-itw_j}\right|^2$ be the periodograms' for X_t and Y_t, respectively, at frequencies $w_j = 2\pi j/T, j = 1,2,\ldots,[T/2]$ (where $[T/2]$ is the largest integer less or equal to $T/2$). The Euclidean distance between the periodograms is defined by:

$$d_{Period} = \left(\sum_{j=1}^{[T/2]} \left(P_x(w_j) - P_y(w_j)\right)^2 \right)^{1/2}. \quad (3)$$

These three distances capture different time series features. The Euclidean distance compares one-to-one values of the two data series; the correlation takes into account the association of the data, that is if the increase and decrease of the data is followed by the other data series, and finally the periodogram compares the cyclical behavior of the time series data.

B. Combining the Distance Measures

In the clustering procedure, we propose to combine the three distances presented in III.A by:

$$d_{combined} = \alpha_1 d_{NEucl} + \alpha_2 d_{RNOMC} + \alpha_3 d_{NPeriod}, \quad (4)$$

with $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and being d_{NEucl} and $d_{NPeriod}$ normalized distances obtained by:

$$(d - min(d))/(max(d) - min(d)), \quad (5)$$

where d is the initial distance (d_{Eucl} or d_{Period}) and $min(d)$ and $max(d)$ are the minimum and maximum of d, respectively, guaranteeing that all the three distances in (4) range from 0 to 1.

C. The Clustering Procedure

In the hierarchical clustering methods, the goal is to obtain a nested sequence of partitions of all the objects. In these methods, the number of groups is not defined in advance. In fact, the output, a hierarchy structure represented by a dendrogram, provides a visual aid that enables to reasonably establish an adequate number of clusters.

In this work we resort to the agglomerative Ward method for clustering - [18].

D. Visualizing the distances using classical multidimensional scaling (MDS)

We also propose to use classical MDS-Multidimensional Scaling, [19], as a complementary tool for the evaluation of the clustering results and the visualization of segments constituted.

In this work we resort to the R implemented method described in [20].

Classical MDS goal is to represent, in a Euclidean space, a matrix of distances between points. The points coordinates are found in a way that the original distances are preserved. To be of practical use, only two-dimensional representation is considered. Thus, classical MDS represents a reduction of data dimensionality and the proportion of variation of original distances explained by the representation using only two dimensions is used as a goodness-of-fit measure.

E. Measuring cohesion-separation of the clustering solution

The Dunn index [21] measures the between-within distances ratio. The best partitions (with K clusters) should exhibit the largest index values corresponding to compact and well separated clusters. A variant of this index is used in this work which incorporates more information than the original index [22] and may be therefore less sensitive to noise in the data:

$$Dunn = \min_{p \neq q \text{ and } p,q \in \{1 \ldots K\}} \left\{ \frac{\delta_{pq}}{\max_{l \in \{1 \ldots K\}} \Delta_l} \right\} \quad (6)$$

In the original version δ_{pq} is defined as a minimum distance between a pair of objects across clusters p and q whereas Δ_l is defined as the maximum distance between a pair of objects in cluster l. In the variant adopted δ_{pq} is an average of distances between elements in clusters p and q and Δ_l is also defined as average of distances between all pair of objects within cluster l.

IV. CLUSTERING ELETRICITY MARKET PRICES

A. Data

In this work, we have considered 26 hourly day-ahead market prices observed in five European electricity markets, MIBEL, Nord Pool, Italy, Germany and France over the years 2015 to 2018. The time series of prices considered were for the MIBEL: Portugal (MI_PT), Spain (MI_ES); for Nord Pool: Denmark 1 (Np_DK1), Denmark 2 (Np_DK2), Sweden 1 (Np_SE1), Sweden 2 (Np_SE2), Sweden 3 (Np_SE3), Sweden 4 (Np_SE4), Oslo (Np_Oslo), Kristiansand (Np_Kr.sand), Bergen (Np_Bergen), Trondheim (Np_Tr.heim), Molde (Np_Molde), Tromso (Np_Tromso), Finland (Np_FI), Estonia (Np_EE), Latvia (Np_LV), Lithuania (Np_LT); for Italy: Northern (IT_NORD), Central Northern (IT_CNOR), Central South (IT_CSUD), South (IT_SUD), Sardinia (IT_SARD), Sicily (IT_SICI); and, also, Germany (DE) and France (FR). Data regarding these markets were obtained through the respective market operators' website.

B. The Results

Clustering is conducted using the combined distance presented in (4) with $\alpha_1 = \alpha_2 = \alpha_3 = 1/3$, i.e., considering the same weight to all the three distance measures: Euclidean, Pearson correlation based and periodogram based distance.

In the implementation, we resort to software R and also use specific packages like TSclust [15] and fpc [23].

Considering the 2015 data, the dendrogram representing the clustering result is presented in Fig. 1. Resorting to the modified Dunn index (see *III.E*), the solution with three clusters is considered.

The MDS representation of the distances is presented in Fig. 2. This two-dimensional map explains 78.6% of the total variation of the data.

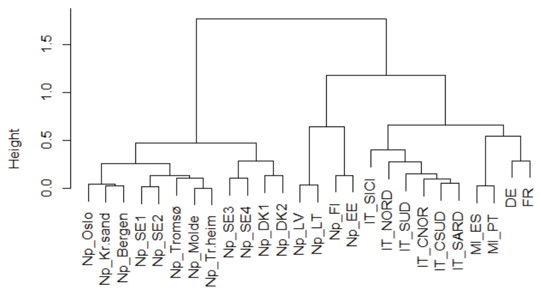

Figure 1. 2015 dendrogram

The Nord Pool areas are divided into two groups: one includes the Baltic countries, Lithuania, Latvia, and Estonia, and Finland, the remaining areas corresponding to Norway, Sweden and Denmark, are grouped together in another group. The third group includes the Italian areas, the MIBEL countries, France and Germany. Moreover, although the clusters considered divide Nord Pool in two groups, the fact is that there is a division between Northern markets (Nord Pool) and Southern markets (MIBEL and Italy) with Germany and France in between. This division is somehow consistent with the fact that Nord Pool markets, especially Denmark, Norway and Sweden have higher shares of renewable production leading to lower market prices, whereas, markets such as the MIBEL and Italy, but also, Germany, Finland and the Baltic countries have higher shares of production from fossil fuels leading to higher market prices.

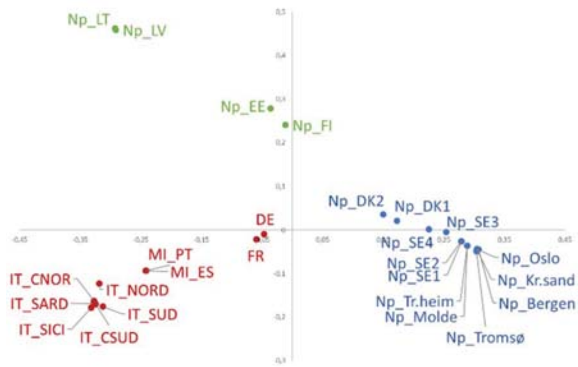

Figure 2. MDS representation of 2015 distance matrix

For the 2016, and also according to the modified Dunn index, a solution with only two clusters is considered. The MDS representation (Fig. 3) explains 68.7% of the total variation of the distances data.

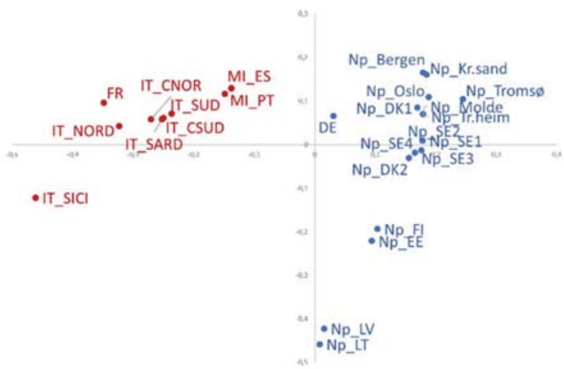

Figure 3. MDS representation of 2016 distance matrix

Again, in 2016, the clustering procedure separates the Northern from the southern markets, now with the difference of including Germany in the group of Nord Pool and France in the group of MIBEL and Italy. This separation can also be seen in Fig. 4 and Fig. 5 where the MDS representation of 2017 and 2018 is presented, respectively. The dendrogram for 2018 can be seen in Fig. 6.

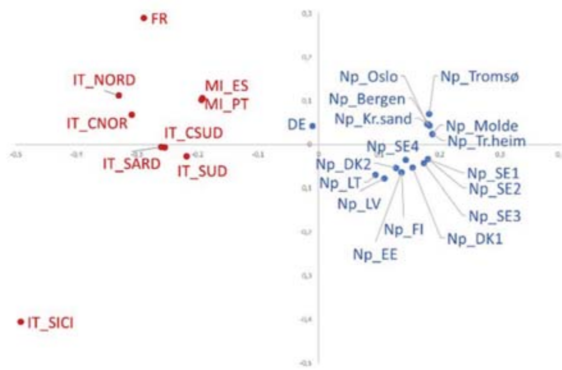

Figure 4. MDS representation of 2017 distance matrix

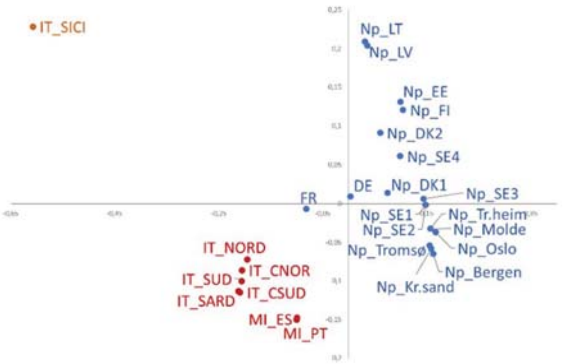

Figure 5. MDS representation of 2018 distance matrix

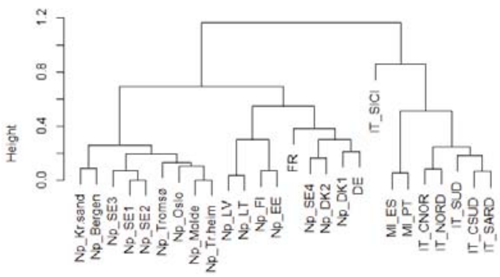

Figure 6. 2018 dendrogram

For 2017, the MDS representation accounts for 60,5% of the total variation of the distances data while for 2018 explains 61.3%.

The major difference in 2017 and 2018 is the fact that the solution considers three distinct clusters of which one is composed by the Sicily zone (IT_SICI) alone. The split of Sicily in a third group is explained by the fact that in 2015 and 2016, Sicily's prices were under Resolution n. 521/2014/R/eel of the Italian Regulatory Authority for Electricity, Gas and Water [24], under which the supply and payment of generating units were subject to regulation. The end of this resolution in 2017 when the Sorgente-Rizziconi interconnection cable entered in operation, along with a decrease in wind energy availability lead to an increase of the price spread between Sicily and mainland Italy when compared to 2015 and 2016 [12].

Another aspect that can be observed in the MDS representations is the grouping of Nord Pool markets and the evolution of the Baltic countries and Finland compared to the rest of the Nordic countries. In 2015, the Baltic countries and Finland were separated in a different cluster and although from 2016 on, they had been grouped with the rest of the countries there are some dynamics inside this group. From 2015 until 2017, the Baltic countries and Finland were approaching the Nordic countries but in 2018 they are further apart again. Moreover, among the Baltic countries, Latvia and Lithuania are the countries furthest away from the other Nord Pool countries. This evolution can be understood if we bear in mind that the major drivers for price differences between markets are the interconnection capacity and the cost difference between marginal units in each market. In fact, the entry in operation, in 2016, of the NordBalt interconnection between Sweden and Lithuania allowed electricity transactions between the two countries and an approximation of prices. Moreover, the interconnection capacity available for the day-ahead market between Swedish area SE1 and Finland has increased from about 1400 MW in 2015 and 2016 to about 1500 MW in 2017 and 2018. Along with this increase in interconnection capacities, the production from renewables, especially wind and hydro in Latvia and Lithuania has increased from 2015 to 2017. This made that the more expensive generating units had to work few hours in these two countries decreasing the need for imports from the other markets. However, in 2018, wind and hydro production had declined again in Latvia and Lithuania, moving away prices of these areas from the prices from the other Nord Pool markets.

V. CONCLUSIONS

Among the benefits of the development of an internal market of electricity at European level are the possibility of citizens and businesses to choose their supplier, the creation of new business opportunities and the improvement of cross-border trade, ensuring efficiency gains and competitive prices and, the contribution to security of supply and sustainability.

The aim of this work was to analyze 26 time series of electricity prices of European markets between 2015 and 2018. The analysis was performed using a combination of three dissimilarity measures: Euclidean, Pearson correlation-based and one based on the periodogram of time series.

The clustering solutions obtained for the analyzed period show a clear division between Northern European markets, in particular Nord Pool, but also Germany, and Southern European markets, MIBEL and Italy, with the French market having an intermediate behavior sometimes closer to Northern markets and other times closer to southern markets. Results also showed that on the matter of market integration more efforts have to be done in order to accomplish a full functioning internal electricity market for the European Union.

REFERENCES

[1] K. Gugler, A. Haxhimusa, and M. Liebensteiner, "Integration of European Electricity Markets: Evidence from Spot Prices," *The Energy Journal*, Vol. 39, N. 2, pp 41-66, 1988.

[2] L. Morales, and J. Hanly, "European power markets–A journey towards efficiency," *Energy Policy,* Vol. 116, pp. 78-85, Feb. 2018.

[3] L. M. de Menezes, and M. A. Houllier, "Reassessing the integration of European electricity markets: A fractional cointegration analysis," *Energy Economics*, Vol. 53, pp. 132–150, Jan. 2016.

[4] M. C. Ruiz-Abellón, A. Gabaldón, and A.Guillamón, "Dependency-Aware Clustering of Time Series and Its Application on Energy Markets," *Energies*, Vol. 9(10), 809, Oct. 2016.

[5] S. Aghabozorgi, A. Seyed Shirkhorshidi, and T. Ying Wah, "Time-series clustering - A decade review," *Inf. Syst.*, vol. 53, pp. 16–38, 2015.

[6] P. Miguel, J. Gonçalves, L. Neves, and A. G. Martins, "Using clustering techniques to provide simulation scenarios for the smart grid," *Sustainable Cities and Society*, Vol. 26, pp. 447–455, Apr. 2016.

[7] I. P. Panapakidis, and A. S. Dagoumas, "Day-ahead electricity price forecasting via the application of artificial neural network-based models," *Applied Energy*, Vol. 172, pp. 132-151, Mar. 2016.

[8] H. Frigui, C. Hwang, and F. Chung-Hoon Rhee, "Clustering and aggregation of relational data with applications to image database categorization," *Pattern Recognit.*, vol. 40, no. 11, pp. 3053–3068, 2007.

[9] F. de A. T. de Carvalho, Y. Lechevallier, and F. M. de Melo, "Partitioning hard clustering algorithms based on multiple dissimilarity matrices," *Pattern Recognit.*, vol. 45, no. 1, pp. 447–464, 2011.

[10] OMIE, Main results of the electricity market 2018. [Online]. Available Iberian Electricity Market Operator web site: http://www.omie.es/files/omie_diptico_pantalla_en.pdf.
[11] OMIE. [Online]. Data available Iberian Electricity Market Operator web site: http://www.omie.es.
[12] GME, Annual Report 2017. [Online]. Available Italian Electricity Market Operator web site: http://www.mercatoelettrico.org/En/MenuBiblioteca/documenti/20180801RelazioneAnnuale2017_En.pdf.
[13] Nord Pool. [Online]. Data available Nord Pool market operator: https://www.nordpoolgroup.com/historical-market-data/.
[14] EPEX. [Online]. Available web site: https://www.epexspot.com.
[15] P. Montero and J. Vilar, "TSclust: An R Package for Time Series Clustering," *JSS J. Stat. Softw*., vol. 62, no. 1, pp. 1–43, 2014.
[16] P. Rodrigues, "Hierarchical clustering of time-series data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 1–13, 2008.
[17] J. Caiado, N. Crato, and D. Peña. "A periodogram-based metric for time series classification." *Computational Statistics & Data Analysis*, vol. 50, no. 10, pp. 2668-2684, 2006.
[18] J. Joe H. Ward, "Hierarchical Grouping to Optimize an Objective Function," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.
[19] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp.401-419, Dec. 1952.
[20] T. Cox and M. Cox, "Multidimensional Scaling," in *Handbook of Data Visualization* 2nd Ed., Springer 2000, pp 315-347.
[21] J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, pp. 32–57, 1973.
[22] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 3, pp. 301–315, 1998.
[23] C. Hennig. Package 'fpc': Flexible Procedures for Clustering. [Online] Available: http://cran.r-project. org/web/packages/fpc/fpc.pdf.
[24] Italian Regulatory Authority for Electricity, Gas and Water, Annual Report 2015. [Online]. Available Council of European energy Regulators web site: https://www.ceer.eu/documents/104400/3737359/C15_NR_Italy-EN.pdf/7bf89bf6-05a3-7404-0139-a277790bbd69?version=1.0