

## On Agglomerative Hierarchical Percentile Clustering

Fabrizio Durante<sup>a</sup> and \*Aurora Gatto<sup>b</sup> and Susanne Saminger-Platz<sup>c</sup>

<sup>a</sup> Department of Economic Sciences, Università del Salento,  
Centro Ecotekne, 73100 Lecce (Italy), [fabrizio.durante@unisalento.it](mailto:fabrizio.durante@unisalento.it)

<sup>b</sup> Department of Economic Sciences, Università del Salento,  
Centro Ecotekne, 73100 Lecce (Italy), [aurora.gatto@unisalento.it](mailto:aurora.gatto@unisalento.it)

<sup>c</sup> Institute for Mathematical Methods in Medicine and Data Based Modeling, Johannes Kepler University Linz,  
Altenbergerstrasse 69, Linz (Austria), [susanne.saminger-platz@jku.at](mailto:susanne.saminger-platz@jku.at)

### Abstract

Cluster analysis aims at grouping objects represented by some feature vectors and as such revealing insight into subset structures among the considered objects. However, in many cases, the observations are subject to experimental errors and/or uncertainty. In such a case, a popular way is to summarize first the information about each object and, then, aggregate the objects via some cluster algorithm. The percentile clustering by Janowitz and Schweizer, instead, considers the whole distribution of observed features and, only afterwards, aggregates them. Here, we revisit this approach in an agglomerative clustering perspective. Moreover, we perform a simulation study showing some finite sample performance of the algorithm. Some case studies illustrate the advantages of the whole methodology.

**Keywords:** Cluster analysis, distance distribution function, quantiles.

### 1 Introduction

Cluster analysis is an important tool for unsupervised learning, which consists of finding groups in data, possibly without the help of a response variable. Such groups are generally characterized by the fact that objects within a group tend to be more similar than objects in different groups. In particular, agglomerative hierarchical clustering is a popular class of methods that: (a) provides an intuitive yet meaningful graphical representation of data in the form of a tree (i.e. the dendrogram); (b) requires as an input a dissimilarity (not necessarily distance) matrix among the objects. For an overview, see, for instance, [3, 5].

As a matter of fact, each object to be classified is usually described by  $p$  different features (i.e. attributes) that are here supposed to be quantitative such that each object can be represented by some vector  $\mathbf{x}_i \in \mathbf{R}^p$ . Moreover, a suitable dissimilarity (or distance) among the objects  $\mathbf{x}_i$  and  $\mathbf{x}_j$  ( $i \neq j$ ) is considered (like the Euclidean distance). However, it could be the case that the features are subject to experimental errors and/or are uncertain due to the intrinsic random nature of the phenomenon under consideration. As often witnessed in practice, repetitive measurements, recordings or evaluations of some single object might lead to several vector representations in the data set. In case of measurements a common approach, also in light of data reduction, is to summarize the information gathered for each object, e.g. over some time frame, by means of a suitable aggregated centrality index (like mean, median) and, then apply the clustering algorithm in order to model the grouping of (different) objects rather than the grouping of data vectors as instances of some specific objects.

Clearly, in condensing the data in this fashion, some information may be lost. Information which might be useful or helpful in modeling or understanding the features under considerations as fuzzy sets on some universe or representing uncertainty or trends in data. Therefore, and as already outlined in [9], it could also be desirable “to design cluster methods which, instead of summarizing first and then classifying, classify first and then summarize, i.e. cluster methods that work directly with the distributed data”.

An alternate approach of using all vector representations in the data set as single instances, and as such ignoring knowledge about its assignments to distinct objects beforehand, may lead to different cluster structures on the one hand or, on the other hand, may lead to cluster structures for which it might be difficult to assign the objects under consideration in a reasonable way.

Consider, for instance, four objects that are measured according to one feature ( $p = 1$ ). Due to experimental errors, for each object we have collected 3 to 5 observations of the feature, namely

$$\begin{aligned} \mathbf{x}_1 &= (0, 0, 0)^T, & \mathbf{x}_2 &= (-1, 0, 1)^T, \\ \mathbf{x}_3 &= (-4 - 2, 2, 4)^T, & \mathbf{x}_4 &= (0, -5, 5, -7, 7)^T. \end{aligned}$$

In order to cluster these objects, we may think of summarizing these observations by means of the average operator and, hence, cluster them. In such a way, all four objects will be presented by their mean, i.e. 0, and as such will be associated in single cluster in the first step of any reasonable hierarchical procedure. However, in condensing the data in this fashion, we have somehow lost the information about the different variability of the measurements. In particular, in an alternative agglomerative hierarchical procedure, it could also be desirable that, first, objects  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are clustered together (since the observations have the smallest variability around their mean value), then  $\mathbf{x}_3$  and  $\mathbf{x}_4$  are joined, and, finally, the two groups are coupled together.

By *percentile clustering* a series of different methods as developed in [9] is understood as to cope with uncertain and multiple measurements of the same features. It is grounded on the theory of probabilistic metric spaces [11]. The underlying assumption is summarized in the statement that “distributions are the numbers of the future”, as said by Berthold Schweitzer in 1984 (see also [2]). Thus, the dissimilarity among objects should be first represented in terms of a *distance distribution function* and, then, classical dissimilarity indices could be obtained by means of a suitable percentile (i.e. quantile) associated with the distribution function.

Although the idea of percentile clustering has been grounded on an inspiring mathematical framework (see, e.g., [10, 8]), it has only received little attention in the literature, perhaps also due to the computational complexity too high for the time when it had been introduced.

Here, we would like to revisit such an approach by focusing on a particular method that we call *Agglomerative Hierarchical Percentile Clustering* (AHPC, in short). As the common hierarchical algorithms, the method is based on a linkage function, but the dissimilarity matrix is computed according to the percentile approach described in [9]. Note that by considering percentiles it is reflected that the dissimilarity coefficient of a clustering method frequently has only ordinal significance by ranking the objects w.r.t. their dissimilarity (compare also, e.g., [7], but also [12, 1]). In particular, we perform a simulation study to show the per-

formance of AHPC under different linkage functions and different percentile levels. Various suggestions for practical applications of AHPC are hence derived.

## 2 The APHC algorithm

Suppose that we would like to cluster  $d$  objects that are measured according to  $p$  different features. Moreover, each object  $i$  is associated with a set of  $n_i$  observations related to the features, that is with a given  $(n_i \times p)$  data matrix  $\mathbf{X}_i$ .

For a fixed linkage method `link` chosen among single, average and complete linkage, and a fixed  $c \in (0, 1)$ , the AHPC can be summarized as follows:

1. For each object  $i$  and  $j$ ,  $i \neq j$ , represented by the data matrices  $\mathbf{X}_i$  and  $\mathbf{X}_j$  calculate the Euclidean distance  $d_{k\ell}^{ij}$  between the  $k$ -th row of  $\mathbf{X}_i$  and the  $\ell$ -th row of  $\mathbf{X}_j$  for every  $k = 1, \dots, n_i, \ell = 1, \dots, n_j$ .
2. Return the dissimilarity between object  $i$  and  $j$ ,  $i \neq j$ , to be equal to the  $c$ -quantile of the vector

$$(d_{k\ell}^{ij} : k = 1, \dots, n_i, \ell = 1, \dots, n_j), \quad (1)$$

denoted by  $p_{ij}$ .

3. Create the  $(d \times d)$ -dissimilarity matrix  $P$  among all the objects by setting  $P = (p_{ij})$  with  $p_{ii} = 0$  and  $p_{ij} = p_{ji}$ .
4. Apply a hierarchical clustering algorithm with linkage method `link` starting with the input given by the matrix  $P$ .

Summarizing, the AHPC is a classical agglomerative hierarchical algorithm whose input matrix is obtained from the distance distribution function given in (1) among the objects under consideration. However, because of its more general formulation, it may enjoy some novel features. For instance, in AHPC, the number of observations related to each object can be different, i.e.  $n_i \neq n_j$ , when  $i \neq j$ . This fact could be of potential interest, e.g. when one object can be deterministically obtained from a single feature vector, while other objects are subject to various experimental errors or when repetitive measurements for one object might be more costly. Moreover, it allows to incorporate prior (expert) knowledge about the assignment of single feature vectors to objects as such seeding resp. directing the clustering process of the objects as indicated, e.g. also in [12].

Notice that the dissimilarity matrix  $P$  is by construction symmetric and assigns value 0 on its main diagonal. This initialization step is fundamental since agglomerative clustering requires that, in the first step, each object is assigned to its own group.

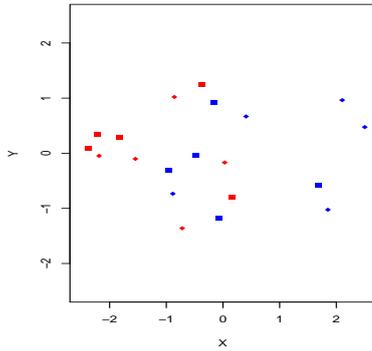


Figure 1: Illustration of four 2-dim objects (each with 5 observations) belonging to two different groups. Each color corresponds to a group.

### 3 The simulation study

In order to provide an overview of the performance of the algorithm we simulate three different scenarios.

#### Study 1: Bivariate independent Gaussian case

First, we consider a set of  $D$  objects ( $D \in \{20, 40\}$ ) divided into two equally sized groups. Each object is associated with two features whose values are generated by a 2-dim Gaussian distribution with the following mean  $\mu$  and covariance matrix  $\Sigma$ :

**G1:**  $\mu = (\mu_1, 0)$ ,  $\Sigma_{11} = \Sigma_{22} = 1$ , and  $\Sigma_{12} = \Sigma_{21} = 0$ ;

**G2:**  $\mu = (-\mu_1, 0)$ ,  $\Sigma_{11} = \Sigma_{22} = 1$ , and  $\Sigma_{12} = \Sigma_{21} = 0$ ;

for  $\mu_1 \in \{0.25, 0.5, 1\}$ . Hereinafter, this case is called the *bivariate independent Gaussian case*. The number of observations associated with each object is equal to  $N \in \{5, 10, 25\}$ . An illustration is given in Figure 1.

Given the previous setup, we first associate the distance distribution function between all the pairs of objects. Then we compute the dissimilarity matrix according to a related  $c$ -percentile, with  $c \in \{0.25, 0.50, 0.75\}$ .

Once the dissimilarity matrix is obtained, we perform the hierarchical cluster algorithm with single, average and complete linkage, and we calculate the agreement of the clustering result with the group composition of the data generating process by means of the *adjusted Rand index*, shortly ARI (see, for instance, [6]). We recall that  $ARI = 1$  corresponds to the perfect agreement of two cluster composition, herein the perfect agreement of the original grouping and the obtained cluster structure.

In Figures 2–4, we illustrate the values of ARI indices obtained from  $B = 250$  simulations. As can be seen,

- as the value of the mean  $\mu_1$  increases (i.e. the cluster centers tends to be more separate), the ARI assumes higher values;
- as the number of observations associated with each object increases, the performance increases;
- comparing the linkage functions, the single linkage seems to perform worse than the other two methods;
- in all cases, selecting the percentile at 0.75 level seems to work better than using the percentile at 0.25 or 0.50 level.

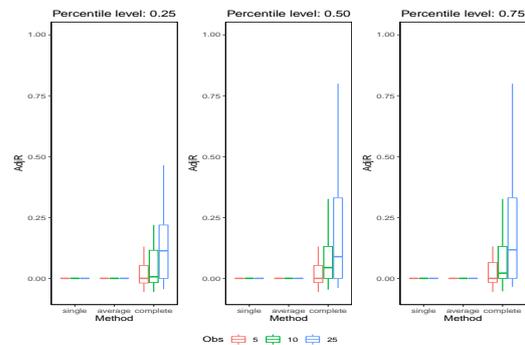


Figure 2: ARI index for 20 objects belonging to two groups, where the values are generated from the bivariate independent Gaussian case, with  $\mu_1 = 0.25$ .

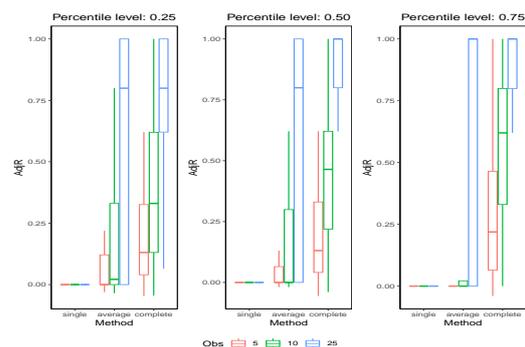


Figure 3: ARI index for 20 objects belonging to two groups, where the values are generated from the bivariate independent Gaussian case, with  $\mu_1 = 0.50$ .

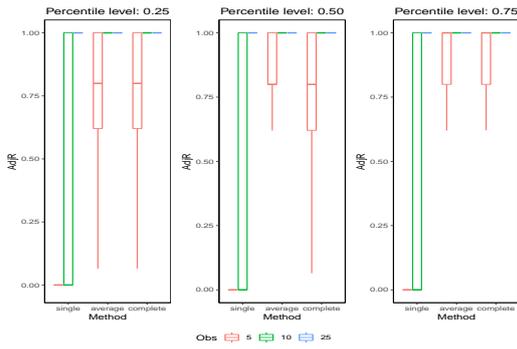


Figure 4: ARI index for 20 objects belonging to two groups, where the values are generated from the bivariate independent Gaussian case, with  $\mu_1 = 1$ .

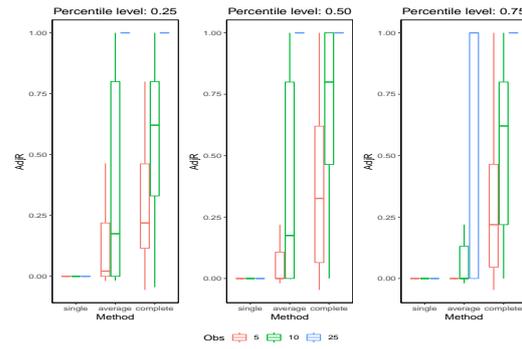


Figure 5: ARI index for 20 objects belonging to two groups, where the values are generated from the bivariate dependent Gaussian case, with  $\mu_1 = 0.5$  and  $\Sigma_{12} = 0.50$ .

### Study 2: Bivariate dependent Gaussian case

Next, we consider the case when the features are not independent, but related with some given correlation. Again we consider a set of  $D$  objects ( $D \in \{20, 40\}$ ) divided into two equally sized groups. Each object is associated with two features whose values are generated by a bivariate Gaussian distribution with the following mean  $\mu$  and covariance matrix  $\Sigma$ :

**G1:**  $\mu = (\mu_1, 0)$ ,  $\Sigma_{11} = \Sigma_{22} = 1$ , and  $\Sigma_{12} = \Sigma_{21} \in \{0.5, 0.75\}$ ;

**G2:**  $\mu = (-\mu_1, 0)$ ,  $\Sigma_{11} = \Sigma_{22} = 1$ , and  $\Sigma_{12} = \Sigma_{21} \in \{-0.5, -0.75\}$ ;

with  $\mu_1 \in \{0.25, 0.5, 1\}$ . Hereinafter, this case is called the *bivariate dependent Gaussian case*. The number of observations associated with each object is equal to  $N \in \{5, 10, 25\}$ .

In Figures 5 and 6 we show the values of ARI obtained from  $B = 250$  simulations when  $\mu_1 = 0.5$ ; the other cases are available upon request. Here, we find a better performance as the correlation value increases, especially when we consider  $c$ -percentiles for  $c \in \{0.25, 0.75\}$ . However, comparing this case with the independent case, we see no relevant differences.

### Study 3: Five groups

Finally, we consider the case when the dimension of the object is larger than in the previous cases. Specifically, we take into account a set of  $D = 100$  objects divided into five equally sized groups, each group related to the Gaussian independent case. From Figures 7 and 8 we see that the performance improves as the number of observations increases, being the 0.75 percentile level the better choice. Again the complete

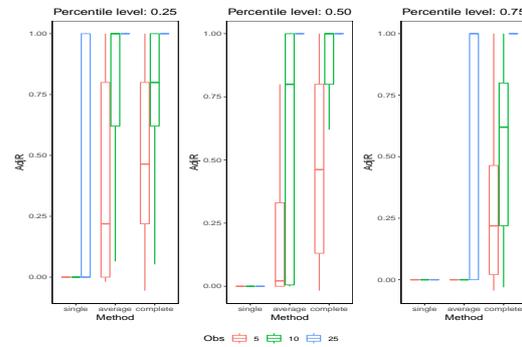


Figure 6: ARI index for 20 objects belonging to two groups, where the values are generated from the bivariate dependent Gaussian case, with  $\mu_1 = 0.5$  and  $\Sigma_{12} = 0.75$ .

method outperforms the other two linkage methods and the ARI value increases significantly when we consider  $\mu_1 = 1$  instead of  $\mu_1 = 0.5$ . Moreover, in general, it seems that the increase of the dimension of the problem does not worsen the performance of AHPC compared to the previous cases.

## 4 Two real case studies

In order to illustrate the methodology, we analyze two different real case studies. The first one – already discussed [9] and reconsidered here – describes monthly combat deaths in the Vietnam War for six years from 1966 to 1971. The other one is related to the annual shares of the generation of renewable energy sources on the total of all possible energy sources for the 27 countries of the European Union over a period of five years.

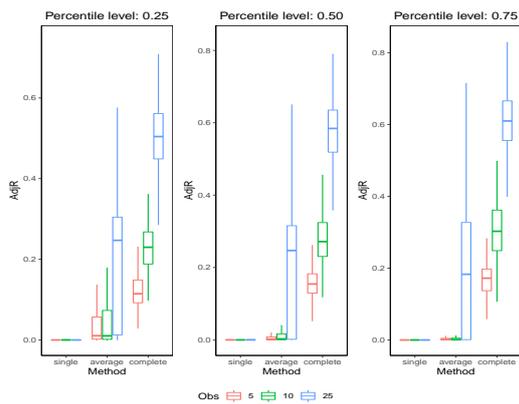


Figure 7: ARI index for 100 objects belonging to five groups, where the values are generated from the bivariate independent Gaussian case, with  $\mu_1 = 0.5$

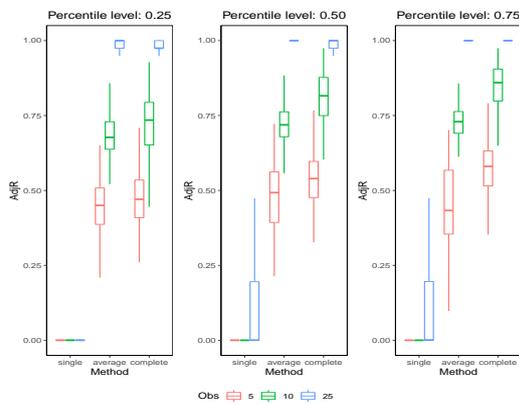


Figure 8: ARI index for 100 objects belonging to five groups, where the values are generated from the bivariate independent Gaussian case, with  $\mu_1 = 1$

#### 4.1 The first case study

First, we analyze the dataset from [4] related to monthly combat deaths in the Vietnam War over a period of six years. The data has been also considered in [9] for the purpose of detecting different phases of U.S. involvement in the war. We have been interested whether by our algorithms we can re-identify the therein achieved results. To this end, following [9], we consider successive six-month time intervals (which are indicated with the letters from A to L).

First, according to [9], we focus only on U.S. versus South Vietnam deaths. Taking into account the simulation study, we perform the AHPC with complete linkage and percentile level equal to 0.75. The corresponding dendrogram is visualized in figure 9. As can be seen, the results are quite similar to the results presented in [9]. In particular, in this case three groups

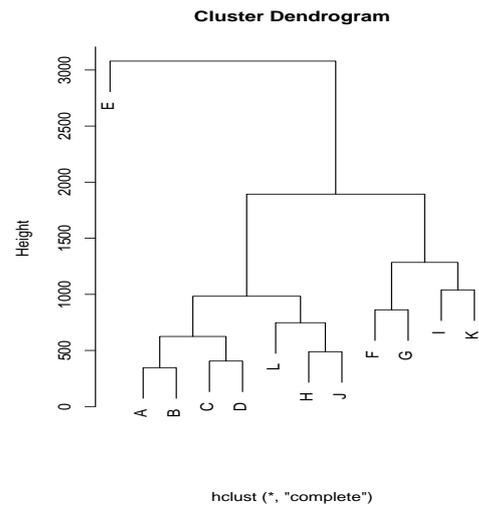


Figure 9: Dendrogram obtained from AHPC applied to Hartigan dataset considering U.S. versus South Vietnam deaths.

are selected, A,B,C,D (first two years of the war) belong to one group, while E (first semester of the third year) plays the role of a singleton (outlier) group.

It should also be noted that Hartigan's data included U.S., South Vietnam, Third World and Enemy deaths. If we elaborate a similar analysis by considering also these two additional feature, the resulting dendrogram is visualized in Figure 11. As can be seen, there are common behaviour in the two dendrograms, especially when three groups are considered as an optimal cutting tree. In particular, A, B, C, and D are also grouped together, while E plays the role of an outlier group.

#### 4.2 The second case study

We analyze the production of renewable energy sources for the countries of the European Union in the period from 2015 to 2019<sup>1</sup>.

Our interest depends on the fact that, in the recent years, the EU has invested, on average, almost 20 billion Euro per year in clean energy research and innovation, which is considered a priority by the Energy Union<sup>2</sup>. In fact, Europe's new growth strategy (*European Green Deal*<sup>3</sup>) aims to transform the EU economy into a sustainable one by 2050, with the aim of reducing greenhouse gas emissions by at least 55% by 2030. Achieving the EU's climate targets will require signif-

<sup>1</sup>For further details, see <https://ec.europa.eu/eurostat/web/energy/data>

<sup>2</sup>COM(2015) 80 final; JRC SETIS <https://setis.ec.europa.eu/publications/setis>

<sup>3</sup>COM(2019) 640 final

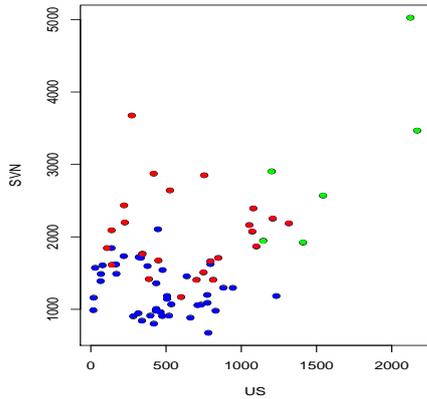


Figure 10: Cluster composition (three groups) related to AHPC applied to Hartigan dataset considering U.S. versus South Vietnam deaths. Each color corresponds to a different group

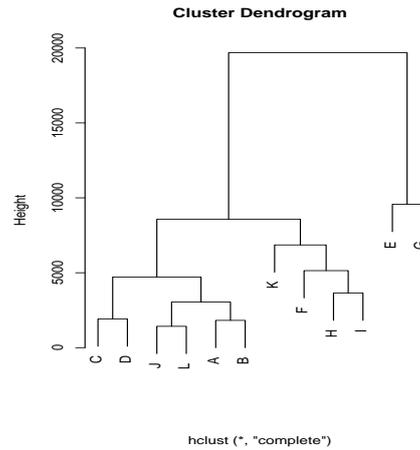


Figure 11: Dendrogram obtained from AHPC applied to Hartigan dataset considering U.S., South Vietnam, Third World and Enemy deaths.

icant decarbonisation and an integrated energy system largely based on renewable energies.

EU electricity production from renewable sources is already expected to rise from the level of 32% in 2020 to around 65% by 2030 <sup>4</sup> and to 80% by 2050 <sup>5</sup>. Evidence of the fact that the energy system is gradually shifting towards clean energy technologies is that the share of renewable energy in final energy consumption has risen from 10% in 2005 to the 20% target for 2020 and the share of energy renewables in the electricity sector rose to just over 32% <sup>6</sup>. Furthermore, despite the COVID 19 pandemic, the European energy system has shown a greener energy mix, with a 34% decrease in the production of electricity from coal in the EU and an increase in electricity generation from renewable sources equal to 43% in the second quarter of 2020 <sup>7</sup>.

The importance of renewable energy in the European Union supports the choice of considering this crucial issue as a case study. Specifically, we analyze the shares of renewable energy sources on the total of all possible energy sources for the countries of the European Union from 2015 to 2019 (see *Eurostat* <sup>8</sup>). In particular, we consider the total amount of electricity

generation from all energy sources and the shares of hydro, wind, solar, solid biofuels power generations and also the share of other sources that includes electricity generation from gaseous and liquid biofuels, renewable municipal waste, geothermal, and tide, wave and ocean. Notice that it is also important to consider minor renewable sources as they affect the total production of renewable energy. Each country is hence associated with a data matrix formed by the share values per each year in the considered period.

Thus, we apply the AHPC to the data under consideration algorithm with complete linkage and percentile level equal to 0.75. The resulting dendrogram is shown in the Figure 12. Taking into account a possible grouping into three clusters, the composition of each group is visualized in Table 1.

Cluster 1	Cluster 2	Cluster 3
Austria	Belgium	Denmark
Croatia	Bulgaria	Germany
Latvia	Czech Rep	Ireland
Romania	Cyprus	Portugal
Slovenia	Estonia	Spain
Sweden	Finland	
	France	
	Greece	
	Hungary	
	Italy	
	Lithuania	
	Luxembourg	
	Malta	
	Netherlands	
	Poland	
	Slovakia	

Table 1: Cluster composition related to Figure 12 when three groups are considered.

Figure 13 shows, instead, the results of a ‘classical’ clustering procedure performed by previous averaging the values observed in the considered time period. Comparing the results we see some differences in the composition of the groups. In particular, in the latter

<sup>4</sup>COM(2020) 562 final

<sup>5</sup>COM(2018) 773 final

<sup>6</sup>For further details, see

<https://ec.europa.eu/eurostat>

<sup>7</sup>Quarterly Report on the European Electricity Markets, Volume 13, N. 2.

<https://ec.europa.eu/energy/data-analysis>

<sup>8</sup>For further details, see

<https://ec.europa.eu/eurostat/web/energy/data>

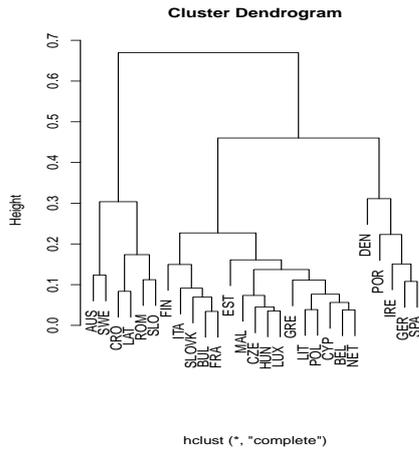


Figure 12: Dendrogram obtained from AHPC applied to Shares Renewable dataset considering the shares of renewable power generation of the different Countries of the European Union on the total of all possible energy sources (linkage method: complete, percentile level 0.75). Period: 2015–2019.

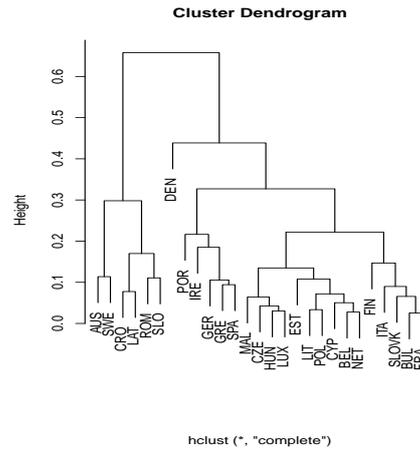


Figure 13: Dendrogram obtained from hierarchical clustering applied to Shares Renewable dataset considering the shares of renewable power generation of the different Countries of the European Union on the total of all possible energy sources (linkage method: complete). Average values in the period 2015-2019 are considered.

case Germany, Ireland, Portugal and Spain shift from the third group to the second one.

In order to provide an interpretation of the obtained groups, for the sake of presentation, we can consider Figure 14, which shows the boxplots of the shares of renewable power generation on the total of all possible energy sources for all countries in a group (i.e. the distribution of the values of each group feature). It seems that, on average, the countries belonging to the first group and the third group are more focused on a single energy source, while the countries of the second group tends to focus on a multi-source system.

As a follow-up of our study, we restrict to the share of renewable power generation of the different countries of the European Union with respect to the total value of renewable energy production (i.e. ignoring the production by non-renewable energy). Such an analysis tends to group together countries that have focused on similar types of renewable energies. As can be seen from Figure 15, four clusters seem to be suitable to consider. In particular, we notice that Malta forms a singleton group, since its production of renewable energy is almost totally focused on solar energy.

### 5 Additional comments

Clustering methods as unsupervised machine learning technique are often applied for detecting (sub)structures within data samples for obtaining deeper insights into a problem or process to be

modelled. Nowadays hybrid approaches in machine learning try to incorporate and combine expert knowledge available beforehand and available measurement or simulation data. The APHC algorithm as discussed in this paper may be seen as a step in this direction where the aim is to cluster objects represented by several observations/measurements w.r.t. a certain amount of different features within one data set. As already indicated in [12] some directing resp. seeding of the clustering process by considering initial classification results, which would correspond to the assignment to objects in our approach, would be worth considering instead of single data items.

By starting from objects, rather than instances, we may incorporate expert knowledge into the clustering process at an initial level of the process. We may even allow that objects may be represented by different amount of data instances in the data set as this might appear more often in practical situations where measurements for different objects might not be available to the same extent or where measurements are simple recorded according to different frequency regimes.

As outlined in, e.g. [1], it may be of practical and theoretical interest that the final cluster size distribution is approximately uniform, i.e. leading to clusters of similar size, however this has not been the main interest of our present analysis. Moreover, we assume that in case of imbalanced data sets it might be reasonable not to impose any additional conditions on the final cluster size beforehand.

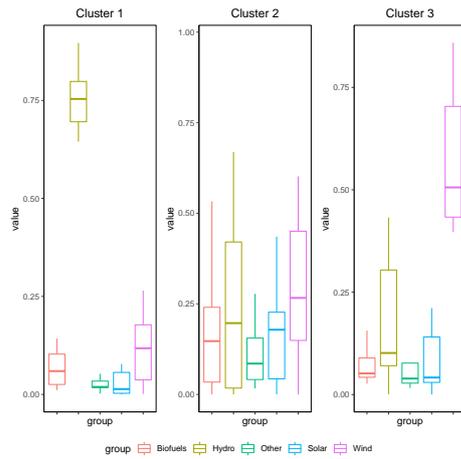


Figure 14: The shares of renewable power generation on the total of all possible energy sources for the different groups.

By calculating a dissimilarity matrix on the basis of the  $c$ -percentile of the a distance distribution function assigned to two objects, we employ an ordered weighted aggregation function at the second step of the APHC algorithm, similarly as discussed in [12] and [1]. The investigation of more general OWA-based linkages, as also discussed in [12] and [1], has not been applied (yet) in step 4 of the APHC algorithm, but will be of interest for future studies.

**Acknowledgement**

We thank the anonymous referees for their comments which have helped to further improve our contribution. The first and second author acknowledge the support by MIUR-PRIN 2017, Project “Stochastic Models for Complex Systems” (No. 2017JFFHSH).

**References**

[1] A. Cena, M. Gagolewski, Genie+owa: Robustifying hierarchical clustering with owa-based linkages, *Information Sciences* 520 (2020) 324–336.

[2] E. Diday, M. Vrac, Mixture decomposition of distributions by copulas in the symbolic data analysis framework, *Discrete Appl. Math.* 147 (1) (2005) 27–41.

[3] B. S. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster analysis*, 5th Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 2011.

[4] J. A. Hartigan, *Clustering algorithms*, John Wiley & Sons, New York-London-Sydney, 1975, wiley Series in Probability and Mathematical Statistics.

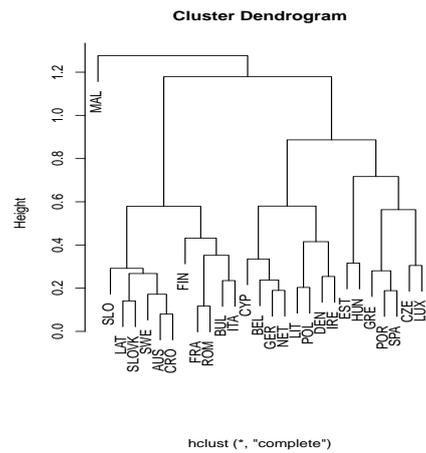


Figure 15: Dendrogram obtained from AHPC applied to Shares Renewable dataset considering the shares of renewable power generation of the different Countries of the European Union within the renewable production (linkage method: complete, percentile level 0.75).

[5] C. Hennig, M. Meila, F. Murtagh, R. Rocci (Eds.), *Handbook of cluster analysis*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, FL, 2016.

[6] L. Hubert, P. Arabie, Comparing partitions, *J. Classif.* 2 (1985) 193–218.

[7] M. F. Janowitz, An order theoretic model for cluster analysis, *SIAM J. Appl. Math.* 34 (1) (1978) 55–72.

[8] M. F. Janowitz, R. C. Powers, A duality between fuzzy and percentile clustering, *Fuzzy Sets and Systems* 50 (1) (1992) 51–58.

[9] M. F. Janowitz, B. Schweizer, Ordinal and percentile clustering, *Math. Social Sci.* 18 (2) (1989) 135–186.

[10] R. Powers, Flatness and continuity in the percentile clustering model, *Math. Social Sci.* 21 (1) (1991) 53–66.

[11] B. Schweizer, A. Sklar, *Probabilistic metric spaces*, North-Holland Series in Probability and Applied Mathematics, North-Holland Publishing Co., New York, 1983.

[12] R. Yager, Intelligent control of the hierarchical agglomerative clustering process, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 30 (6) (2000) 835–845.