Ecological Informatics 9 (2012) 64–68

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Assessing the efficiency of clustering algorithms and goodness-of-fit measures using phytoplankton field data

Alexis Carteron [a,b,*,1], Martin Jeanmougin [a,b,1], Fabien Leprieur [c], Sofie Spatharis [b]

[a] Facultés des Sciences, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier, France
[b] Department of Marine Sciences, University of the Aegean, University Hill, 81100 Mytilene, Greece
[c] Laboratoire Ecologie des Systèmes Marins Côtiers UMR 5119, Université Montpellier 2, cc 093, Place E. Bataillon, 34095 Montpellier Cedex 5, France

ARTICLE INFO

Article history:
Received 6 November 2011
Received in revised form 27 March 2012
Accepted 29 March 2012
Available online 5 April 2012

Keywords:
2-norm
Cophenetic correlation coefficient
Beta diversity
Dendrogram
UPGMA
Ward's algorithm

ABSTRACT

Investigation of patterns in beta diversity has received increased attention over the last years particularly in light of new ecological theories such as the metapopulation paradigm and metacommunity theory. Traditionally, beta diversity patterns can be described by cluster analysis (i.e. dendrograms) that enables the classification of samples. Clustering algorithms define the structure of dendrograms, consequently assessing their performance is crucial. A common, although not always appropriate approach for assessing algorithm suitability is the cophenetic correlation coefficient *c*. Alternatively the 2-norm has been recently proposed as an increasingly informative method for evaluating the distortion engendered by clustering algorithms. In the present work, the 2-norm is applied for the first time on field data and is compared with the cophenetic correlation coefficient using a set of 105 pairwise combinations of 7 clustering methods (e.g. UPGMA) and 15 (dis)similarity/distance indices (e.g. Jaccard index). In contrast to the 2-norm, cophenetic correlation coefficient does not provide a clear indication on the efficiency of the clustering algorithms for all combinations. The two approaches were not always in agreement in the choice of the most faithful algorithm. Additionally, the 2-norm revealed that UPGMA is the most efficient clustering algorithm and Ward's the least. The present results suggest that goodness-of-fit measures such as the 2-norm should be applied prior to clustering analyses for reliable beta diversity measures.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Enhancing our knowledge of the processes that shape variability in community structure (beta diversity) remains one of the fundamental challenges in contemporary community ecology (Condit et al., 2002; Gaston et al., 2007; Tuomisto et al., 2003). To meet this challenge, applications of multivariate statistics in community ecology have expanded significantly during the last two decades (Dray and Legendre, 2008; Gauch, 1982; Legendre and Legendre, 1998). Among multivariate statistical methods, ordination and clustering are now routinely applied by ecologists to explore the spatial or temporal turnover of field communities (e.g. Devictor et al., 2010; Konan et al., 2006; Kreft and Jetz, 2010; Legendre et al., 2005; Leprieur et al., 2009; Winder and Hunter, 2008).

Among clustering methods, those based on a hierarchy of clusters have been used in many research fields (e.g. Bhau et al., 2009; Kniggendorf et al., 2011; Scholz and Sadowski, 2009) and specifically in aquatic ecology (e.g. Jabiol et al., 2009; Leprieur et al., 2009; Torrente-Vilara et al., 2011). Hierarchical clustering analysis is subdivided into agglomerative and divisive methods, the former being most common in ecological studies (Clarke and Warwick, 2001). All agglomerative procedures begin with an initial matrix (**I**) which is then transformed in inter-objects/samples matrix (**D**) using a relevant distance measure whose selection depends on the scientific question (see Fig. 1). In the beginning of the agglomerative process each object/sample is considered as a separate class or cluster. For a set of N initial objects, the first clustering will result in N-1 clusters, the next N-2 and so on until only one cluster contains all the objects, with objects which are most similar fusing together at each step. Based on various clustering algorithms, different dendrograms are formed, associated with their ultrametric distance matrix (**U**) and expressing different degrees of faithfulness with **D** (Legendre and Legendre, 1998).

Selecting an appropriate clustering algorithm is crucial since this can lead to different grouping and levels of linkage between clusters in the resulting dendrograms thus affecting the interpretation of results (Legendre and Legendre, 1998; Parker and Arnold, 2000; van Tongeren, 1987). The cophenetic correlation coefficient *c* based on Pearson's correlation (Legendre and Legendre, 1998; Sokal and Rohlf, 1962) is currently the most commonly used approach (e.g. Blackburn et al., 2005; Kreft and Jetz, 2010; May, 1999) for assessing the most faithful dendrogram (i.e.

* Corresponding author at: Laboratoire Insulaire du Vivant et de l'Environnement, University of New Caledonia, 145 Avenue James Cook, BP R4, 98851 Nouméa Cedex, New Caledonia. Tel.: +33 687 19 02 55; fax: +33 687 25 48 29.
E-mail address: alexis.carteron@gmail.com (A. Carteron).
[1] Co-first authors.

1574-9541/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2012.03.008

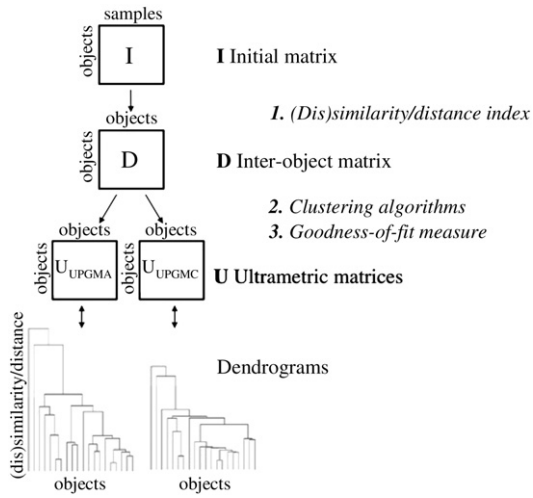

Fig. 1. Scheme of the methodological steps followed in clustering analysis, from the initial matrix to the formation of dendrograms.

the dendrogram that preserves most faithfully **D**). However, it is known that *c* may not always be a reliable measure of distortion generated by algorithms (Farris, 1969; Holgersson, 1978; Mérigot et al., 2010). For instance, Mérigot et al. (2010) raised three criticisms on the reliability of information given by *c* measure: (i) it is only a measure of intensity of the monotonic linear relationship between **D** and **U**; (ii) it is sensitive to extreme values; and (iii) a *c* close to 1 indicates a perfect correspondence of **D** and **U** whereas the match between the two matrices can in fact be weak. In order to find the appropriate clustering algorithm, Mérigot et al. (2010) introduced a goodness-of-fit measure based on the greatest singular value of the matrix of **D**–**U**, called the 2-norm measure.

Although selecting an optimal agglomerative algorithm is a crucial step, the majority of studies are based on subjective criteria and do not employ a goodness-of-fit measure to check which algorithm most faithfully preserved **D** (Frontalini and Coccioni, 2008; Moya-Anegon et al., 2006; Primpas et al., 2008; Scholz and Sadowski, 2009; Verfaillie et al., 2009). For instance, Ward's linkage method has been reported to form "exceptionally well defined clusters" (Parker and Arnold, 2000) and is in some cases preferred to other algorithms for being already applied in previous studies and/or enhancing the discrimination between clusters (Frontalini and Coccioni, 2008; Parker and Arnold, 2000; Verfaillie et al., 2009). Mérigot et al. (2010) used simulated assemblages to compare the suitability of 2-norm and *c* in selecting the best agglomerative algorithm and found that Ward's algorithm produced by far the least faithful dendrogram. In addition, the results obtained with *c* and 2-norm were not totally in agreement, revealing possible misinformation due to *c*. Several other goodness-of-fit measures are available for algorithm selection (e.g. Gower distance, correspondence matrices tests, and the Monte Carlo process) (Clatworthy et al., 2007; Gower, 1983; Mantel, 1967). However, the current approach focuses on the comparison between the novel 2-norm measure as it has never been applied using community field data and *c* for being so far the most commonly applied measure for algorithm selection.

This study aims to compare the efficiency of the 2-norm and *c* measures in selecting the most faithful dendrogram among those produced by seven commonly used agglomerative algorithms. This comparison was undertaken for 15 (dis)similarity and distance indices to assess whether the best selected agglomerative algorithm by the 2-norm measure and/or *c* differed among indices. To this aim, the (dis)similarity and distance indices were applied using data on phytoplankton assemblages sampled in the Gulf of Kalloni, an enclosed coastal area in the Aegean Sea, Greece.

2. Methodology

2.1. Dataset

In the present study we used available phytoplankton species-abundance data originally obtained to assess the effect of terrestrial runoff on an enclosed coastal ecosystem, namely the Gulf of Kalloni, Aegean Sea, Greece (Spatharis et al., 2007a, 2007b). These data have been collected on a monthly basis over a period of one annual cycle (August 2004 to July 2005) from four stations in the interior of the Gulf (K3, K4, K5, and K7), each station containing information from two depths (1 and 5 m). Detailed information on the dataset, sampling methodology, and sample analysis are provided in Spatharis et al. (2007a, 2007b). The initial matrix (**I**) used for the analysis contained 109 phytoplankton species by 96 samples and was analyzed using 15 (dis)similarity and distance indices (Table 1).

2.2. Data analysis

The methodology followed for the creation and evaluation of dendrograms is summarized step by step in Fig. 1. Firstly, the initial matrix (**I**) is created based on species-abundance or transformed presence-absence data. Then, 15 common (dis)similarity and distance indices are computed on **I** to create the (dis)similarity/distance matrix (**D**) (e.g. Euclidean, Jaccard; see Table 1). The application of these indices between samples generated a triangular matrix **D** of pairwise similarities among samples. Subsequently, 7 clustering algorithms used in the methodological framework of Mouchet et al. (2008) were applied to generate ultrametric matrix (**U**) (e.g. UPGMA, Ward; see Table 1). Each algorithm has a specific aggregation criterion that distorts **D** to some degree (Legendre and Legendre, 1998). For each (dis)similarity and distance index, we explored the faithfulness of **U** over the triangular matrix **D** using the 2-norm and *c* approaches. It is worth noting that the 2-norm does not allow comparing the goodness-of-fit of dendrograms obtained by different distance indices as it is directly dependent on the scale of the employed distance index (see Mérigot et al., 2010). The R software (version 2.12.2) was used for all calculations together with the Clue package (Hornik, 2009) for the 2-norm computation.

Table 1
Indices of similarity, dissimilarity, and distance. For indices based on presence-absence data see Gower and Legendre (1986) and for indices based on species-abundance data see Legendre and Legendre (1998). For clustering algorithms (acronyms in brackets) see Legendre and Legendre (1998).

(Dis)similarity/distance index	
Presence-absence data	Species-abundance data
Dice or Sorensen	Bray-Curtis
Gower & Legendre coefficient	Canberra
Hamann coefficient	Chi-2
Jaccard	Euclidean distance
Ochiai	Log of Bray Curtis
Phi of Pearson	
Rogers & Tanimoto	
Simple matching coefficient	
Sokal & Sneath 1	
Sokal & Sneath 2	
Clustering algorithm	
Ward's linkage (Ward)	
Simple linkage (Simple)	
Complete linkage (Complete)	
Unweighted pair group method using arithmetic averages (UPGMA)	
Weighted pair group method using arithmetic averages (WPGMA)	
Unweighted pair group centroid method (UPGMC)	
Weighted pair group centroid method (WPGMC)	

2.3. Goodness-of-fit measures

Two goodness-of-fit measures were used. Firstly *c*, based on Pearson's *r* correlation, is the linear correlation coefficient between cophenetic distances of **U** and the inter-objects distances of **D** (Sokal and Rohlf, 1962). When *c* is close to 1, **U** is considered to respect the intrinsic information of **D** without significantly distorting it (Legendre and Legendre, 1998). The range value is between -1 and 1, which expresses increasing faithfulness of the clustering algorithm. The **U** having the highest *c* is therefore assigned the highest rank (see next section). Secondly, the 2-norm measure is based on the greatest singular value of the 2-norm matrix of **D**–**U**. A perfect match between **D** and **U** will result in a 2-norm value equal to 0. Consequently the lowest value is attributed to the most faithful clustering method and the rank 1 is assigned. For comparative reasons, the maximum 2-norm value observed in the present analysis was 2730.41 (see Table 2) but the maximum is theoretically not bounded. For detailed information on the 2-norm see Mérigot et al. (2010).

2.4. Performance of clustering algorithms

In order to compare the performance of clustering algorithms a general ranking was established based on the results of each algorithm for each index. When an algorithm was the best performing (i.e. the lowest 2-norm and/or the highest *c* values) it was assigned the rank 1 for the particular goodness-of-fit measure. If an algorithm was the second best, then the rank 2 was assigned and so on. Therefore each algorithm corresponded to a rank number for all indices. These "scores" were cumulated to determine the overall performance of each of the 7 algorithms. The worst score would be 7 (lowest rank) times 15 (the number of indices) i.e. 105.

3. Results

3.1. Performance of clustering algorithms: c versus 2-norm

After carrying out the 105 combinations between (dis) similarity/distance indices and clustering algorithms, both the 2-norm and *c* goodness-of-fit measures selected UPGMA as the first and WPGMA as the second most faithful clustering algorithms, followed by complete linkage (Table 2). UPGMC was found to be the second least faithful algorithm according to both measures. Disparities in the general ranking were also observed and these are represented graphically in Fig. 2. Ward's method was selected by *c* as the fourth most faithful algorithm whereas 2-norm considered it as the least faithful (score of 103 out of 105). Simple linkage was ranked fifth for *c* and fourth for 2-norm. The same trends were revealed when looking at the ranking results for each distance measure separately (Table 2). Only two distance measures revealed a disagreement between the selected algorithms by the 2-norm and *c*, namely the Euclidean distance and the Sokal & Sneath 1 index. When comparing between different distance measures it is worth noting that values of *c* were often close and expressed low variation between different algorithms in contrast to the 2-norm values (Table 2).

3.2. Potential incongruence between 2-norm and c: the case of Euclidean distance

An example of potential incongruence between the two goodness-of-fit measures (2-norm and *c*) is presented for the Euclidean distance measure and the UPGMA and UPGMC algorithms in the matrix plot (or Shepard-like diagram) of Fig. 3. In order to qualitatively assess the distortion between the inter-object matrix **D** and the ultrametric matrix **U**, matrix plots are very effective in visualizing all the information (Legendre and Legendre, 1998). In the plot of UPGMA points tends to follow the first bissectrix (Fig. 3a, c = 0.90), however this is not the case for UPGMC (Fig. 3b) although *c* is slightly higher

Table 2
Ranks of values obtained by the 2-norm measure and the cophenetic correlation coefficient *c* based on 7 clustering algorithms and 15 (dis)similarity and distance indices (calculated values in parentheses). The ranking was performed for each distance measure separately. The same rank was assigned to similar values obtained by the 2-norm or *c* measure.

	Ward		Simple		Complete		UPGMA		WPGMA		UPGMC		WPGMC	
Distance	*c*	2-norm	*c*	2-norm	*c*	2-norm	*c*	2-norm	*c*	2-norm	*c*	2-norm	*c*	2-norm
Euclidean	7 (0.41)	7 (2730.41)	3 (0.86)	3 (371.11)	6 (0.73)	5 (510.64)	2 (0.90)	1 (100.08)	4 (0.82)	2 (293.45)	1 (0.91)	6 (544.91)	4 (0.82)	4 (462.84)
Bray-Curtis	6 (0.69)	7 (470.63)	6 (0.69)	4 (28.23)	3 (0.71)	3 (21.19)	1 (0.80)	1 (6.24)	2 (0.79)	2 (7.21)	4 (0.70)	6 (35.71)	4 (0.70)	5 (31.94)
Log of Bray Curtis	4 (0.61)	7 (299.85)	4 (0.61)	4 (19.71)	2 (0.78)	3 (17.46)	1 (0.79)	1 (3.56)	3 (0.74)	2 (4.85)	7 (0.46)	6 (27.70)	6 (0.52)	5 (25.27)
Canberra	5 (0.62)	7 (230.07)	3 (0.78)	4 (15.83)	2 (0.82)	3 (9.85)	1 (0.84)	1 (1.92)	3 (0.78)	2 (3.18)	6 (0.47)	5 (37.45)	7 (0.18)	6 (37.90)
Jaccard	3 (0.80)	7 (15.11)	5 (0.77)	4 (2.24)	4 (0.78)	3 (2.14)	1 (0.87)	2 (0.53)	1 (0.87)	1 (0.52)	7 (−0.17)	6 (8.04)	6 (0.09)	5 (7.45)
Sorensen	3 (0.74)	7 (16.71)	5 (0.72)	3 (2.58)	3 (0.74)	4 (2.74)	1 (0.84)	2 (0.66)	1 (0.84)	1 (0.65)	6 (0.18)	6 (6.95)	7 (0.08)	5 (6.57)
Simple matching	3 (0.77)	7 (11.51)	5 (0.71)	4 (1.86)	3 (0.77)	3 (1.42)	1 (0.84)	1 (0.39)	2 (0.82)	2 (0.47)	6 (0.34)	6 (4.57)	7 (0.23)	5 (4.30)
Sokal & Sneath 1	5 (0.80)	7 (11.45)	3 (0.81)	4 (1.64)	3 (0.81)	3 (1.46)	1 (0.89)	2 (0.38)	1 (0.89)	1 (0.37)	6 (−0.08)	6 (8.82)	7 (−0.16)	5 (7.99)
Rogers & Tanimoto	4 (0.78)	7 (12.77)	5 (0.72)	4 (2.00)	3 (0.79)	3 (1.55)	1 (0.86)	1 (0.41)	2 (0.84)	2 (0.45)	7 (0.09)	6 (5.94)	6 (0.15)	5 (5.32)
Hamann	4 (0.77)	7 (15.95)	5 (0.71)	4 (2.56)	3 (0.78)	3 (2.08)	1 (0.84)	1 (0.53)	2 (0.82)	2 (0.59)	6 (0.32)	6 (6.45)	7 (0.19)	5 (6.01)
Ochiai	3 (0.74)	7 (16.81)	5 (0.72)	3 (2.60)	4 (0.73)	4 (2.82)	1 (0.84)	1 (0.65)	1 (0.84)	1 (0.65)	6 (0.21)	6 (6.92)	7 (−0.02)	5 (6.74)
Sokal & Sneath 2	3 (0.79)	7 (16.71)	4 (0.75)	3 (2.43)	5 (0.72)	4 (2.64)	1 (0.86)	2 (0.59)	2 (0.85)	1 (0.58)	7 (0.05)	6 (7.46)	6 (0.06)	5 (6.82)
Phi of Pearson	4 (0.79)	7 (19.18)	5 (0.75)	4 (2.83)	3 (0.83)	3 (2.76)	1 (0.85)	2 (0.68)	1 (0.85)	1 (0.67)	6 (0.21)	6 (7.82)	7 (−0.03)	5 (7.57)
Gower & Legendre	5 (0.59)	5 (2.37)	3 (0.68)	3 (0.49)	4 (0.62)	4 (0.60)	1 (0.76)	1 (0.16)	2 (0.74)	2 (0.22)	7 (−0.60)	7 (9.70)	6 (−0.21)	6 (7.23)
Chi-2	7 (0.77)	7 (2317.16)	6 (0.95)	4 (241.47)	4 (0.97)	6 (308.33)	1 (0.98)	1 (65.25)	4 (0.97)	3 (233.10)	1 (0.98)	5 (286.81)	1 (0.98)	2 (196.97)
Total rank	66	103	67	55	52	54	16	20	31	25	83	89	88	73

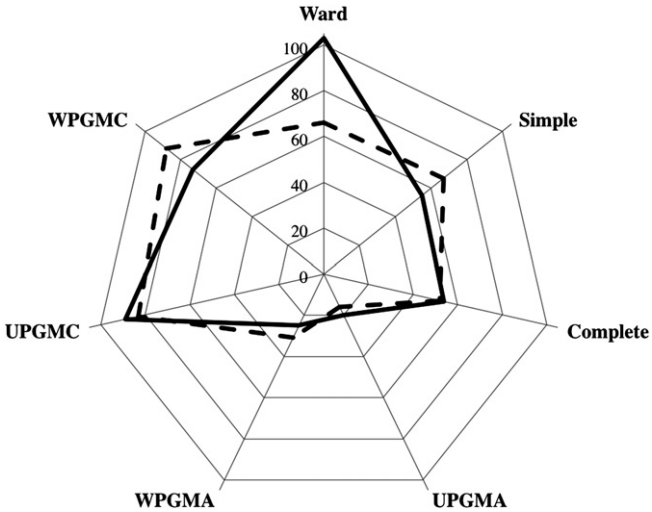

Fig. 2. Spider chart of the performance of the 7 clustering algorithms for the 15 inter-object matrices generated by the 15 (dis) similarity/distance indices from the unique initial matrix. Algorithms performing best according to the 2-norm and cophenetic correlation coefficient *c* are closer to the center of the chart being often near to the first rank. The vertical axis represents the total score of algorithms based on the 2-norm (solid lines) and *c* (dotted lines) measures.

(c = 0.91). The ranking is clearly different for the 2-norm that selected UPGMA as the most faithful algorithm (2-norm = 100.08). The 2-norm considered UPGMA to have the lowest distortion between elements of the inter-objects matrix d_{ij} and elements of the ultrametric matrix u_{ij}, UPGMC being the second least faithful (2-norm = 544.91). In addition, Ward's method was ranked as the least faithful with a value of 2-norm and *c* equal to 2730.41 and 0.41 respectively exemplifying its recurring weak faithfulness.

4. Discussion

Cluster analyses are widely used in many scientific disciplines and even more so in community ecology (Leprieur et al., 2009; Petchey and Gaston, 2009). In the case of field phytoplankton data such as those used in the present study, important ecological questions can be explored related to the spatial or temporal organization of assemblages (Spatharis et al., 2007b; Winder and Hunder, 2008). Despite the value of cluster analyses in revealing beta diversity patterns, the use of goodness-of-fit measures for obtaining more reliable results is still scarce. The present investigation indicates that the cophenetic correlation coefficient *c* is mostly congruent with the 2-norm, revealing nonetheless important disparities. Consequently, a question rises on whether the wide use of (linear) correlation coefficients as goodness-of-fit measures might have mislead studies on the choice of the most faithful algorithms and resulting hierarchical classifications (e.g. Hale and Dougherty, 1988; Kreft and Jetz, 2010; Malik and Husain, 2006).

UPGMA was selected by both goodness-of-fit criteria (2-norm and *c*) as the most faithful algorithm, providing the best fit between **D** and **U**. This agglomerative algorithm may be able to handle simultaneously highly similar and dissimilar data (Kniggendorf et al., 2011). It is worth noting that the closely-related WPGMA algorithm may be more appropriate in some cases, e.g. when using the Jaccard or Sorensen dissimilarity indices (Table 1). On the contrary, Ward's method seems to be the least faithful, never achieving the best fit. These results disagree with previous studies (Cao et al., 1997; Clatworthy et al., 2007; Primpas et al., 2008). For instance, in the study by Primpas et al. (2008) Ward's linkage algorithm was selected as the most appropriate for eutrophication studies based on its high resolution and discriminative power, however not using a goodness-of-fit criterion. Based on the results of the present study, we anticipate that the use of a goodness-of-fit measure such as the 2-norm could have promoted the selection of an algorithm other than Ward's.

Selection of appropriate algorithms for a given dataset should be carried out with great caution on a case-by-case basis. Furthermore, algorithm selection should not entirely rely on criteria such as discriminative ability or cluster configuration (e.g. Parker and Arnold, 2000; Primpas et al., 2008) but should be based on a goodness-of-fit measure. Disparities between clustering algorithms in different studies may have been generated because datasets were assumed to have a predetermined structure and obvious patterns, assessing the performance of clustering methods with an a priori knowledge of the grouping (Cao et al., 1997; Primpas et al., 2008). To clarify these issues there is a need for comparative studies considering other types of data (e.g. phylogenetic data). The intrinsic nature of data (e.g. vectors or scalars) is also of importance (Kniggendorf et al., 2011). Finally the performance of an algorithm could depend on the clustering level of similarity or the dataset size. For instance, Cao et al. (1997) suggested that UPGMA performs less well at higher levels of similarity.

Although our results showed an overall agreement between the 2-norm and *c* goodness-of-fit measures in selecting the most faithful and appropriate clustering algorithm, several issues should be considered. First, our results revealed incongruence between the 2-norm and *c*

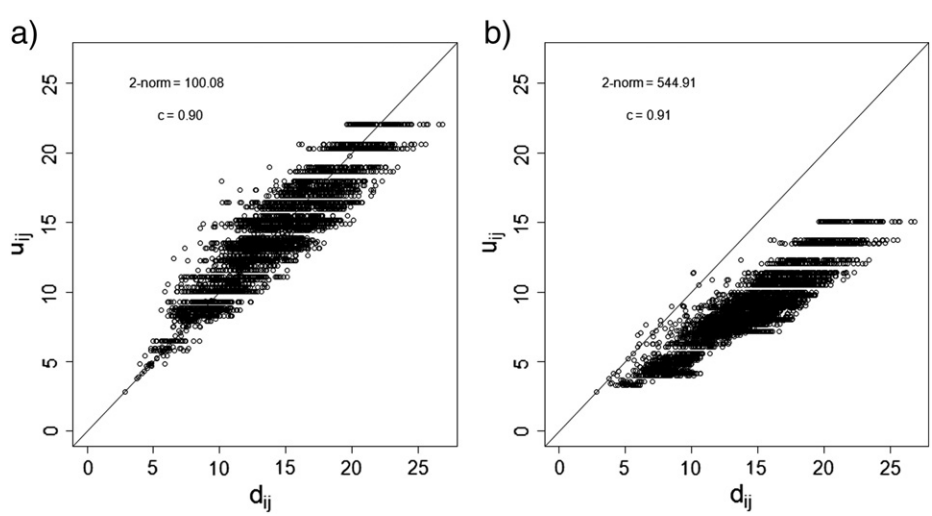

Fig. 3. Matrix plot of inter-object (**D**) and ultrametric (**U**) matrices for clustering methods UPGMA (a) and UPGMC (b) computed with the Euclidean distance matrix for the 96 samples. The solid line corresponds to the bissectrix line ($x = y$) i.e. when the goodness-of-fit of **D** to **U** is perfect (**D** = **U**).

when using the Euclidean distance and the Sokal & Sneath 1 index. This implies that *c* may lead to an inadequate choice of clustering algorithms when using these two distance measures since the selected clustering algorithm is not always the most faithful. Second, in contrast to the 2-norm, values of *c* tend to display the same rank or to be very similar between different clustering algorithms. This trend of *c* is in agreement with a previous study (Kreft and Jetz, 2010) and was observed for most of the distance measures examined in the present study. Furthermore *c* is known to be very sensitive to extreme values (Mérigot et al., 2010). Consequently, any effect of sampling effort could lead to the selection of another clustering algorithm. The above issues may render the selection of a given algorithm difficult and subjective. In contrast, the 2-norm provided a clear indication for most of the employed distance measures showing a higher discriminating potential. Therefore, considering the incorrect and unclear measurements of *c* we recommend the use of 2-norm as a more appropriate goodness-of-fit measure.

The results of this investigation suggest that:

- cophenetic correlation coefficient did not provide a clear indication on clustering algorithm efficiency;
- 2-norm seems to be efficient as an alternative goodness-of-fit measure;
- both measures revealed UPGMA as the most faithful algorithm;
- 2-norm showed Ward's as the least faithful algorithm; and
- a goodness-of-fit measure should be applied for reliable clustering analyses.

Acknowledgements

We would like to thank Dezerald Olivier, Juhel Jean-Baptiste, and Velazquez Graciela for their help. We are also grateful to all BIODIV (Biodiversity Conservation) M.Sc. staff for their support. This work has been initiated by A.C. and M.J., who are two graduate students of BIODIV, a bilateral M.Sc. program between the University of Montpellier 2 (France) and the Aegean University (Greece). F.L. and S.S. are part of BIODIV M.Sc. teaching staff.

References

Bhau, B.S., Medhi, K., Das, A.P., Saikia, S.P., Neog, K., Choudhury, S.N., 2009. Analysis of Genetic Diversity of Persea bombycina "Som" Using RAPD-Based Molecular Markers. Biochemical Genetics 47, 486–497.

Blackburn, T.M., Petchey, O.L., Cassey, P., Gaston, K.J., 2005. Functional diversity of mammalian predators and extinction in island birds. Ecology 86, 2916–2923.

Cao, Y., Williams, W.P., Bark, A.W., 1997. A comparison of clustering methods for river benthic community analysis. Hydrobiologia 347, 25–40.

Clarke, K.R., Warwick, R.M., 2001. Change in Marine Communities: An Approach to Statistical Analysis and Interpretation (PRIMER-E).

Clatworthy, J., Hankins, M., Buick, D., Weinman, J., Horne, R., 2007. Cluster analysis in illness perception research: A Monte Carlo study to identify the most appropriate method. Psychology & Health 22, 123–142.

Condit, R., Pitman, N., Leigh, E.G., Chave, J., Terborgh, J., Foster, R.B., Nunez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E., Hubbell, S.P., 2002. Beta-diversity in tropical forest trees. Science 295, 666–669.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., Mouquet, N., 2010. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. Ecology Letters 13, 1030–1040.

Dray, S., Legendre, P., 2008. Testing the species traits-environment relationships: the fourth-corner problem revisited. Ecology 89, 3400–3412.

Farris, J.S., 1969. On the cophenetic correlation coefficient. Systematic Biology 18, 279–285.

Frontalini, F., Coccioni, R., 2008. Benthic foraminifera for heavy metal pollution monitoring: A case study from the central Adriatic Sea coast of Italy. Estuarine, Coastal and Shelf Science 76, 404–417.

Gaston, K.J., Evans, K.L., Lennon, J.L., 2007. The scaling of spatial turnover: pruning the thicket. In: Storch, D., Marquet, P., Brown, J.M. (Eds.), Scaling biodiversity. Cambridge University Press, Cambridge, pp. 181–222.

Gauch, H.G., 1982. Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge.

Gower, J.C., 1983. Comparing classifications. In: Felsenstein, J. (Ed.), Numerical taxonomy. NATO ASI series, Vol. G-1. Springer, Berlin, pp. 137–155.

Gower, J.C., Legendre, P., 1986. Metric and Euclidean properties of dissimilarity coefficients. Journal of Classification 3, 5–48.

Hale, R.L., Dougherty, D., 1988. Differences between ward's and UPGMA methods of cluster analysis: implications for school psychology. Journal of School Psychology 26, 121–131.

Holgersson, M., 1978. The limited value of cophenetic correlation as a clustering criterion. Pattern Recognition 10, 287–295.

Hornik, K., 2009. CLUE: cluster ensembles. R package version 0.3-27. http://CRAN.R-project.org/package=clue.

Jabiol, J., Corbara, B., Dejean, A., Cereghino, R., 2009. Structure of aquatic insect communities in tank-bromeliads in a East-Amazonian rainforest in French Guiana. Forest Ecology and Management 257, 351–360.

Kniggendorf, A.K., Gaul, T.W., Meinhardt-Wollweber, M., 2011. Hierarchical Cluster Analysis (HCA) of Microorganisms: An Assessment of Algorithms for Resonance Raman Spectra. Applied Spectroscopy 65, 165–173.

Konan, F.K., Leprieur, F., Ouattara, A., Brosse, S., Grenouillet, G., Gourène, G., Winterton, P., Lek, S., 2006. Spatio-temporal patterns of fish assemblages in coastal West African rivers: a self-organizing map approach. Aquatic Living Resources 19, 361–370.

Kreft, H., Jetz, W., 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography 37, 2029–2053.

Legendre, P., Legendre, L., 1998. Numerical ecology. Elsevier Science BV, Amsterdam, The Netherlands.

Legendre, P., Borcard, D., Peres-Neto, P., 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs 75, 435–450.

Leprieur, F., Olden, J.D., Lek, S., Brosse, S., 2009. Contrasting patterns and mechanisms of spatial turnover for native and exotic freshwater fish in Europe. Journal of Biogeography 36, 1899–1912.

Malik, R.N., Husain, S.Z., 2006. Classification and ordination of vegetation communities of the Lohibehr reserve forest and its surrounding areas, Rawalpindi, Pakistan. Pakistan Journal of Botany 38, 543–558.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27, 209–220.

May, A.C.W., 1999. Towards more meaningful hierarchical classification of amino acid scoring matrices. Protein Engineering 12, 707–712.

Mérigot, B., Durbec, J.P., Gaertner, J.C., 2010. On goodness-of-fit measure for dendrogram-based analyses. Ecology 91, 1850–1859.

Mouchet, M., Guilhaumon, F., Villeger, S., Mason, N.W.H., Tomasini, J.A., Mouillot, D., 2008. Towards a consensus for calculating dendrogram-based functional diversity indices. Oikos 117, 794–800.

Moya-Anegon, F., Herrero-Solana, V., Jimenez-Contreras, E., 2006. A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library science research and information. Journal of Information Science 32, 63–77.

Parker, W.C., Arnold, A.J., 2000. Quantitative methods of data analysis in foraminiferal ecology. In: Sen Gupta, B.K. (Ed.), Modern Foraminifera. Kluwer Academic Publisher, pp. 71–89.

Petchey, O.L., Gaston, K.J., 2009. Dendrograms and measures of functional diversity: a second instalment. Oikos 118, 1118–1120.

Primpas, I., Karydis, M., Tsirtsis, G., 2008. Assessment of clustering algorithms in discriminating eutrophic levels in coastal waters. Global Nest Journal 10, 359–365.

Scholz, M., Sadowski, A.J., 2009. Conceptual classification model for Sustainable Flood Retention Basins. Journal of Environmental Management 90, 624–633.

Sokal, R.R., Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. Taxon 9, 33–40.

Spatharis, S., Danielidis, D.B., Tsirtsis, G., 2007a. Recurrent Pseudo-nitzschia calliantha (Bacillariophyceae) and Alexandrium insuetum (Dinophyceae) winter blooms induced by agricultural runoff. Harmful Algae 6, 811–822.

Spatharis, S., Tsirtsis, G., Danielidis, D.B., Chi, T.D., Mouillot, D., 2007b. Effects of pulsed nutrient inputs on phytoplankton assemblage structure and blooms in an enclosed coastal area. Estuarine, Coastal and Shelf Science 73, 807–815.

Torrente-Vilara, G., Zuanon, J., Leprieur, F., Oberdorff, T., Tedesco, P.A., 2011. Effects of natural rapids and waterfalls on fish assemblage structure in the Madeira River (Amazon Basin). Ecology of Freshwater Fish 20, 588–597 (Online ISSN: 1600-0633).

Tuomisto, H., Ruokolainen, K., Yli-Halla, M., 2003. Dispersal, environment, and floristic variation of western Amazonian forests. Science 299, 241–244.

van Tongeren, O.F.R., 1987. Clustering analysis. In: Jongman, R.H., ter Braak, C.J.F., van Tongeren, O.F.R. (Eds.), Data Analysis in Community and Landscape Ecology. Pudoc Wageningen, pp. 180–183.

Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W., Van Lancker, V., 2009. A protocol for classifying ecologically relevant marine zones, a statistical approach. Estuarine, Coastal and Shelf Science 83, 175–185.

Winder, M., Hunter, D.A., 2008. Temporal organization of phytoplankton communities linked to chemical and physical forcing. Oecologia 156, 179–192.