

Contents lists available at [ScienceDirect](#)

## Journal of Asia-Pacific Biodiversity

journal homepage: <http://www.elsevier.com/locate/japb>Journal of  
Asia-Pacific  
Biodiversity

## Short Communication

## Scaling the chord and Hellinger distances in the range [0,1]: An option to consider

Anxo Conde\*, Jorge Domínguez

Department of Ecology and Animal Biology, University of Vigo, Vigo E-36310, Spain

## ARTICLE INFO

## Article history:

Received 24 October 2017  
 Received in revised form  
 20 December 2017  
 Accepted 9 January 2018  
 Available online xxx

## Keywords:

Beta diversity  
 Euclidean distance  
 Methodological unification  
 Multivariate methods  
 Resemblance measures

## ABSTRACT

The chord and Hellinger distances are commonly used as measures of resemblance in ecological studies. Both distances are bound within the range  $[0, \sqrt{2}]$ . We propose to scale them within the range  $[0, 1]$ . The scaling is mainly justified to report beta diversity values in the range  $[0, 1]$  properly. Moreover, results for both unscaled distances in multivariate techniques such as cluster analysis or ordinations are not directly comparable with similar graphical displays obtained with indices bound in the range  $[0, 1]$ . Although comparability and/or interpretability of values are compromised, the used of the unscaled Hellinger and chord distances do not void their validity in ecological studies. Nonetheless, we have found one exception when comparing clustering models using the Gower distance criterion.

© 2018 National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA), Publishing Services by Elsevier. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Measures of resemblance, which are widely used in ecology, facilitate the analysis of multivariate data. Resemblance measures are applied to clustering, ordination, and statistical methods used for hypothesis testing or just for comparison between pairs of samples (Anderson et al 2008; Legendre and Legendre 1998). Ecological resemblance is calculated by using mathematical expressions that account for similarity or dissimilarity between sites depending on their species composition considered as binary (presence/absence) or abundance data (Pielou 1984).

Some analytical techniques in ecology are based on the Euclidean distance that is mainly used in methods involving ordination such as Principal Components Analysis and Redundancy Analysis (Legendre and Gallagher 2001). The Euclidean distance would be a natural measure of resemblance if it would be fully informative in terms of dissimilarity between sites. This problem may arise when the distance between sites with no species in common is smaller than the distance between sites sharing species. Orlóci (1978) exemplified this paradox. Conversely, many resemblance measures overcome the major drawback of the Euclidean

distance by displaying their upper bound when sites do not share any species, in contrast to the Euclidean distance which has no upper limit (Legendre and Legendre 1998).

The chord and Hellinger distances are commonly used as measures of resemblance in ecological studies. Both distances are bound within the range  $[0, \sqrt{2}]$ . These distances gained importance after Legendre and Gallagher (2001) proposed transformation of species data matrices for use in ordination methods based on the Euclidean distance. Thereafter, the Hellinger and chord distances have become more popular because they can be used in Principal Components and Redundancy Analysis and in other multivariate methods which optimize least-squares loss functions (such as K-means partitioning). Legendre and Gallagher (2001) have shown that these distances increase monotonically in value across a simulated ecological gradient and are maximally related to spatial distances along geographic gradients. The Hellinger and chord distances have also good properties to be used in beta diversity studies (Legendre and De Cáceres 2013).

The Hellinger and chord distances imply the use of vectors (sites, defined by variables that are species) with a norm of value 1, resulting in a maximum distance of  $\sqrt{2}$ . However, the values of these two distances within the range  $[0, \sqrt{2}]$  do not enable direct comparison against values of other dissimilarity measures in the range  $[0, 1]$ . Additionally, the range of the ordination or clustering axes and some statistical multivariate parameters are also dependent on the maximum value that a dissimilarity measure may take. Preferably, resemblance measures should increase smoothly from

\* Corresponding author. Tel.: +34 986 814099; fax: +34 986 812556.  
 E-mail address: [aconde@uvigo.es](mailto:aconde@uvigo.es) (A. Conde).

Peer review under responsibility of National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA).

<https://doi.org/10.1016/j.japb.2018.01.001>

pISSN2287-884X eISSN2287-9544/© 2018 National Science Museum of Korea (NSMK) and Korea National Arboretum (KNA), Publishing Services by Elsevier. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

some fixed minimum to a fixed maximum, as the two community samples become more similar/dissimilar (Krebs 1999). Nonetheless, this maximum value may differ between resemblance measures. Yet, any variable defined within a given range may be rescaled into the range [0,1] by the ranging method (Sneath and Sokal 1973, see next section).

There are several examples of authors highlighting the importance of using indices to quantify diversity in the range [0,1] (Legendre and De Cáceres 2013). The informative value of an index in the range [0,1] is a useful criterion to choose an index for functional diversity studies in accordance to Mason et al (2003) because it facilitates an intuitive interpretation about the ecological performance of communities. Szava-Kovats and Pärtel (2014) also proposed unifying beta diversity measures by applying a logistic transformation that relies on the standardization of beta diversity indices in the range [0,1] to facilitate comparison of studies on ecological gradients. Likewise, Chiu et al (2014) used an upper bound to normalize phylogenetic beta diversity in the range [0,1] so that it can be compared across regions with different number of assemblages. Similarly, the aim of this article is to remind ecologists that the Hellinger and chord distances are easily scaled within the range [0,1]. The scaling enables direct comparison of outputs based on the Hellinger or chord distances with those of other resemblance measures defined within the range [0,1] as well as facilitate their interpretation. Thus, we compare results between the scaled and unscaled version of these distances extended to graphical outputs to show and persuade researchers to use the Hellinger and chord distances in the range [0,1] if either of these resemblance measures is decided to be used in their studies.

## Materials and methods

### Theory and calculation: Scaling the Hellinger and chord distances in the range [0,1]

Ranging any quantitative variable within the range [0,1] is reached by using the extremes of its current range (Sneath and Sokal 1973):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where  $x'$  is the ranged value of  $x$ . When the minimum value of the variable equals zero, the previous equation is reduced to the following simpler expression:

$$x' = \frac{x}{\max(x)} \quad (2)$$

so that the ranging is achieved by dividing the variable by its maximum value. Because the Hellinger and chord distances are defined in the range  $[0, \sqrt{2}]$ , scaling in the range [0,1] is attained by dividing the current values by the maximum upper limit of their range,  $\sqrt{2}$ . We consider this transformation a scaling because the former ranges of the distances are just established on a new scale.

For clustering purposes, the square of the Hellinger and chord distances may be useful, but in this case, the distances lose their metric properties and are less appropriate for ordination purposes (Legendre and Legendre 1998; Pielou 1984). If the square is applied on the Hellinger and chord distances in the range  $[0, \sqrt{2}]$ , the scaling in the range [0,1] is now:

$$x'^2 = \frac{x^2}{2} \quad (3)$$

where  $x'^2$  is the scaled value of  $x^2$ . In this regard, the scaled semi-metric expression of the Hellinger or chord distances used with presence–absence data is equal to the one-complement expression of the Ochiai similarity index applied to a binary data set (Legendre and De Cáceres 2013). Conversely, one minus the scaled semi-metric expression of the Hellinger or chord distances based on presence–absence data is equal to the Ochiai similarity.

The species composition data may be also transformed to yield the chord and Hellinger distances in the range [0,1] by computing the Euclidean distance on the modified data set. The Euclidean distance ( $d_{ij}$ ) between two sites ( $i$  and  $j$ ) involving  $p$  species is:

$$d_{ij} = \sqrt{\sum_{k=1}^p (y_{ik} - y_{jk})^2} \quad (4)$$

The equations provided by Legendre and Gallagher (2001) to transform species composition data should therefore be modified as follows:

$$y'_{ij} = \frac{y_{ij}}{\sqrt{2 \sum_{j=1}^p y_{ij}^2}} \quad (5)$$

for the chord distance and,

$$y'_{ij} = \sqrt{\frac{y_{ij}}{2y_i}} \quad (6)$$

for the Hellinger distance, where  $y_{ij}$  represents species abundance data, with  $i=\{1...n\}$  sites and  $j=\{1...p\}$  species;  $y_i$  represents the sums of species abundances per site;  $y'_{ij}$  represents the transformed value of the original abundance per species and site  $y_{ij}$ . Both expressions are identical to those originally proposed by Legendre and Gallagher (2001), though the denominators within the square root are now multiplied by 2 to guarantee that the scaling is in the range [0,1], after applying Equation (4) on the modified biotic matrix.

### Analytical approach

The “transpecies” data set freely available in the R package BiodiversityR (Kindt 2015) is used to show the practical consequences of applying the Hellinger and chord distances in the range [0,1]. This artificial data set which comprises abundance data for nine species and 19 sites was created by Legendre and Gallagher (2001). It is not an extensive species by site matrix in which species replace each other along an artificial gradient, resembling real case studies. All of the following calculations can be reproduced using the vegan package (Oksanen et al 2015), based on R language (R Core Team 2013).

The Hellinger and chord distances were obtained by applying the Euclidean distance (function “dist”) to the transformed biotic matrix (function “decostand” with “hell” or “norm” as arguments) (Oksanen et al 2015). Both distances are less susceptible to the arch effect in analysis of community gradients, and they also assign low weighting to rare species (Legendre and Gallagher 2001). The Bray–Curtis dissimilarity was also computed with function “vegdist” (Oksanen et al 2015).

Clusters were formed on the distance matrices using the Ward method (vegan’s function “hclust” with argument “ward.D2”; Murtagh and Legendre 2014; Oksanen et al 2015) as agglomerative method. The Gower distance and the correlation between the original dissimilarity matrix and the correlative matrix were computed to compare clustering results (Borcard et al 2011). The

Download English Version:

<https://daneshyari.com/en/article/8848700>

Download Persian Version:

<https://daneshyari.com/article/8848700>

[Daneshyari.com](https://daneshyari.com)