Ecological Indicators 90 (2018) 305–315

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Unsupervised classification of ecological communities ranked according to their biodiversity patterns via a functional principal component decomposition of Hill's numbers integral functions

Fabrizio Maturo

Department of Management and Business Administration, "G. d'Annunzio" University of Chieti-Pescara, Italy

ARTICLE INFO

Keywords:
Biodiversity
FPCA
Ranking ecological communities
FDA
Clustering biodiversity profiles

ABSTRACT

Biodiversity is recognized as one of the corner-stones of healthy ecosystems, thus its conservation is increasingly becoming one of the most important targets of environmental management. Despite there is a general recognition that it indicates the status of ecosystems, no consensus measure exists because biodiversity is a very complex concept that is intrinsically multidimensional and multivariate. Limiting the attention on the concept of biodiversity as variety, recent studies have highlighted that both the classical indices and biodiversity profiles suffer of some limitations because the former neglect the multivariate nature of biodiversity whereas the latter can not provide precise rankings among ecological communities when profiles intersect. For this reason, each attempt of comparing, ranking, or clustering ecological communities according to their variety is influenced by the limitations of the specific metric that is used to assess biodiversity. To overcome this problem, we propose a new approach that takes advantage of Hill's numbers and functional data analysis for ranking and clustering ecological communities according to their variety. Specifically, we introduce three ecological indicators. The first is the "*Hill's numbers integral function*" for considering the multivariate nature of biodiversity and ranking diversity profiles. Afterwards, a functional principal component decomposition is suggested for interpreting "*Hill's numbers integral functions*" and computing their distance. Finally, an unsupervised classification of ecological communities is proposed by using a functional k-means algorithm based on a semi-metric distance that is computed considering the functional principal component decomposition of the "*Hill's numbers integral functions*". This last indicator provides ranked groups of ecological communities according to their variety by considering the multivariate nature of biodiversity, i.e. both richness and evenness, and all their possible shades. The goal of this research is to provide Ecologists, policymakers, and scholars with additional tools for ranking and detecting areas with high environmental risk and clustering ecological communities with similar biodiversity patterns according to their internal variety.

1. Introduction

The dynamics and functioning of ecosystems, and hence the ability of ecosystems to provide humans with essential goods and services, depends to a great extent on the diversity of life (Sadava et al., 2013). Thus, biodiversity is globally recognised as one of the corner-stones of healthy ecosystems, and its conservation is increasingly becoming one of the most important aims of environmental management (Laurila-Pant et al., 2015; Worm et al., 2006; Kremen, 2005; Feest et al., 2010; Feest, 2013). Despite the great importance that is recognized to biodiversity, many research have stressed that diversity of species, genetics, and communities is being lost at an alarming rate (e.g. UNEP, 2002; UNEP, 2010; Cardinale, 2014. Specifically, it has been estimated that the species extinction rate over the past 400 years has been 100–200 times higher than before; moreover, in 2000, the International Union for Conservation of Nature (IUCN) has estimated that 24% of all mammal species and 12% percent of all bird species were threatened with extinction (Hilton-Taylor and Mittermeier, 2000).

For this reason, biodiversity is receiving an increasing interest of institutions, scholars, and different types of stakeholders (e.g. Balmford and Bond, 2005; van Strien et al., 2009; Vačkář et al., 2012). In 2002 (UNEP, 2002) and 2010 (UNEP, 2010), the Convention of Biological Diversity (CBD) has given the first strong international signals for obtaining a significant decrease of the current rate of biodiversity loss. To meet this target, the use of a set of indicators has been suggested but, despite many indicators have been proposed during the last decades, nowadays, no scientific consensus measure exists due to the multidimensional and multivariate nature of biodiversity (Royal Society,

E-mail address: f.maturo@unich.it.

https://doi.org/10.1016/j.ecolind.2018.03.013
Received 19 January 2018; Received in revised form 22 February 2018; Accepted 7 March 2018
1470-160X/

2003; Di Battista and Gattone, 2003; Gattone and Di Battista, 2009; Di Battista et al., 2017; Maturo and Di Battista, 2018).

The CBD (UNEP, 1992) defines biodiversity as "*the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems*". Therefore, biological diversity indicates the variety, quantity, and distribution of the components of life whether they are species, ecosystems or genes (EASAC, 2005). This wide definition makes biodiversity a multidimensional concept (e.g. see Duelli and Obrist, 2003; Noss, 1990; Santini, 2016; Chao et al., 2014; Hill, 1973), and thus it is obvious that it can not be measured by a single index. Due to this intrinsic characteristic of biodiversity, we stress that, in this context, we refer to biodiversity only as "*variety of living organisms in a delineated study area*" (Patil and Taillie, 1982; Pielou, 1975). However, even if considering the concept of biodiversity only as species variety, a further drawback is given by the multivariate nature of biodiversity. Indeed, the concept of variety combines richness, i.e. the number of different species, with evenness, i.e. the degree to which abundances are equitably divided among species (Ricotta et al., 2003). As a consequence, different indices, which provide diverse weights to these two aspects, may lead to contradictory rankings among ecological communities (e.g. see Di Battista et al., 2016; Di Battista et al., 2017).

An attempt for solving this problem was proposed by Hill (1973), who suggested the use of a unifying diversity formulation by introducing a parametric family of diversity indices called "*diversity profiles*". Specifically, they are functions dependent on a parameter that reflects the sensitivities to rare and abundant species, and hence they provide a continuum of possible biodiversity measures yielding a faithful graphical representation of community diversity (Leinster and Cobbold, 2012; Chao et al., 2014). If these curves do not intersect, the graph gives an immediate biodiversity-ranking among ecological communities but many studies have underlined that, when profiles intersect, a unique ranking is not achievable (e.g. seeMaturo and Di Battista, 2018).

To overcome this drawback, recent research have proposed the use of functional data analysis applied to diversity profiles (Di Battista and Gattone, 2003; Maturo et al., 2017); particularly, Di Battista et al. (2017) recommended the use of the area under the diversity profiles but this approach loses the immediacy of the graphical representation and, in addition, it starts from functional data but then provides only a scalar measure. For this reason, this study aims adding to the existing literature regarding biodiversity assessment a new functional instrument, namely "*the Hill's numbers integral function*"; this tool can be obtained from Hill's diversity profiles and allows us reaching a unique ordering among ecological communities (or other possible rankings according to specific values of the integral function).

In a period of rapid global change, monitoring biodiversity variations is key to detect early warning signals of decline, infer the causes of such decline, and develop effective strategies to mitigate it (Buckland et al., 2005; Balmford and Bond, 2005; Santini, 2016). For this reason, numerous statistical techniques have been recently proposed for assessing changes in biodiversity (e.g. Gregory et al., 2005; van Strien et al., 2012; van Strien et al., 2009; Maturo and Di Battista, 2018). Particularly, in last decades, many scholars have highlighted the role of cluster analysis for improving the possible information given by biodiversity indicators and discovering other interesting insights about ecological communities (e.g. Arima et al., 2013; Coro et al., 2015; Levatić et al., 2015; Watts and Worner, 2009; Dietrich et al., 2013; Gregorius, 2006; Rashidi et al., 2015).

This research shares the basic idea of adopting cluster analysis as an useful indicator of ecosystems' condition, and proposes a new approach for grouping ranked ecological communities according to their variety pattern. Particularly, a functional k-means approach based on the distance between functional principal components is proposed. This methods allows us obtaining a fixed number of ranked groups in which ecological communities are similar with respect to their internal variety.

In the literature, many metrics for computing distances between functional objects have been proposed. In effect, several scholars have suggested that the use of proximity measures which are based on derivatives, functional principal components, partial least squares, B-spline representation, and Fourier representation (e.g. Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Febrero-Bande and de la Fuente, 2012; Aguilera and Aguilera-Morillo, 2013; Escabias et al., 2014) could be suitable in specific circumstances. In the following, we refer to a semi-metric distance based on functional principal components due to their informative power in the interpretation of "*Hill's numbers integral functions*".

The final aim of this method is to provide policy makers and stakeholders with additional robust statistical tools to identify possible signals of decline in the quality of the environment by considering the multivariate nature of biodiversity, i.e. all of the infinite possible combinations of richness and evenness (the whole domain of diversity profiles).

This paper is structured as follows. The introduction presents the main issues regarding biodiversity assessment and the purposes of our proposal. The second Section shows the main classical biodiversity indices, their drawbacks, the "*Hill's number integral function*" and its use for ranking ecological communities. The third Section explains the functional principal components decomposition and functional k-means algorithm, which are applied to this new function. The fourth Section proposes an application to a real dataset regarding the fish biodiversity of the rivers of the Lazio Region in Italy. The paper ends with the conclusions and perspectives of research.

2. Classical biodiversity indices and the integral function of Hill's numbers

2.1. Classical biodiversity indices and Hill's numbers

A common method for measuring biodiversity is to analyse trends of α-diversity indices. Traditional biodiversity metrics rely on species counts (richness index) and composite indices such as the Shannon-Wiener (Shannon, 1948) or Simpson indices (Simpson, 1949). They are widely used in ecology due to their ease of understanding and application (Lamb et al., 2009).

The richness index is given by the total number of species living in an ecological community; it is the simplest metric for representing biodiversity, and thus the most common (Whittaker, 1972; Magurran, 2004). The greatest drawbacks of species richness are that it does not take into account relative abundances of species, and it is highly sensitive to sample size; indeed, biodiversity samples are usually incomplete, and the observed number of species is a biased underestimate of true species richness (Chao et al., 2014). Furthermore, it is strongly affected by the presence of rarities (Hill, 1973); therefore, this metric is inadequate for detecting early warning signals of biodiversity change (Santini, 2016).

The Simpson index contemplates both the number of species and their relative abundances (see Eq. (1)), and can be expressed as follows:

$$\triangle_{K_1} = 1 - \sum_{i=1}^{k} f_i^2 \tag{1}$$

where f_i is the relative frequency of the i-th species and k is the total number of species in an ecological community. It can range from zero to $(k-1)/k$. Its maximum is a function of the number of species, and occurs when they are equally distributed (evenness) whereas its minimum (maximum homogeneity) happens when only one species is present. Both evenness and richness contribute to a higher value, and it is a good indicator of the dominance of one or few species on the others; however, it is not a good predictor of richness because it is particularly sensitive to changes in the relative abundances of the most dominant