Ordination of measures of biodiversity in reduced space: sense or nonsense?

Olga Lyashevska[1]*, Keith Farnsworth[1]

[1] School of Biological Sciences, Queen's University Belfast, UK, BT9 7BL

Abstract

Biodiversity measures the degree of difference within biological systems. Ideally, a single measure would unify biodiversity assessments across science. Given the multidimensional nature of biodiversity, in practice, we search for a minimal necessary and sufficient set ($\mathbf{M}^*$) of metrics. Starting with a decomposition of biodiversity indices, we constructed the manifold permutation matrix from which $\mathbf{M}^*$ must be found. Simulated ecological communities were generated from lists of benthic marine species found around Ireland and selected by taxonomic sampling. We tested the sensitivity and relationships among various indices of biodiversity using multivariate statistics with a population of simulated communities. The main goal of this analysis is to show how an orthogonal set of marine biodiversity metrics can be formed from which to select the most sensitive and easily measured diversity properties of the community. Having quantified an individual indicator performance, we will test whether currently used biodiversity indicators can be sensibly reduced to a smaller number and what we might miss out by doing it. Modelling results of this kind will help make practical the measuring of biodiversity in marine ecosystems, in line with ICES mission.

Keywords: Biodiversity, ordination, indicators, community composition

Introduction

Biodiversity, generally referred to as 'the variety of life on earth', in fact 'may itself have a diversity of meanings' (Begon et al., 2006). As a term, biodiversity appeared in the middle of the 80s in the policy context. Soon enough it became a widely used scientific jargon and its popularity might seem as 'a successful breakthrough in some new field of science' (Ghilarov, 1996). However it is rather a convenient myth as the usage of the term 'biodiversity' is value laden (Gaston and Spicer, 2004) and 'biodiversity is a buzzword' (Noss, 1990). 'The need to quantify biodiversity drives its fundamental meaning' (DeLong, 1996). At present biodiversity literature contains confusing and incredibly extensive 'lexicon zoo' (Marcot, 2007). Confusion as to what exactly is meant by biodiversity (Hamilton, 2005) aggra-vates with the number of publications. Biodiversity is rather understood as a diversity of its meanings which can never be captured by a single number (Purvis and Hector, 2000). Clearly, diversity of meanings encompass a diversity of measures, each of them representing various facets of biodiversity. Applying single measure, we describe only one facet and therefore some loss of information will inevitably occur. But before we can even try to answer the question what measure of biodiversity has better performance we need to put stronger emphasis on the meaning of biodiversity. The nub of the problem of measuring biodiversity is that it can mean anything: from genetic and phenotypic variability, to variability in species numbers, ecosystem properties and patterns and functional heterogeneity. To reflect on this it has been suggested by Farnsworth et al. (2010) to recognize biodiversity as a hierarchy at several levels and include interaction

*Contact author. E-mail: olyashevska01@qub.ac.uk

Table 1: A nine-level hierarchy of biocomplexity. (Source: Farnsworth et al., 2010)

Organization Level	Interactions
ecological communities	competition, predator-prey etc.
populations - species	reproduction, migration, mortality
multi-cellular organisms	environmental interactions, behaviour
tissues, organs and organ systems	cell-interactions and organ function
cells	specialist behaviour and reproduction
sub-cellular structures	the 'machinery' of the cell
molecular networks	biochemical engines and 'factories'
molecular surfaces	lock and key - e.g. enzymes
DNA sequences: codons to genes	coding and expression control

between them (see Table 1). Various organization levels represent various components and conceptually, there is a disagreement among ecologists which components of biodiversity should be quantified. Even though the use of indices based on species is a firmly established tradition in ecology, scientists are in constant search of new indices of biodiversity, indices which would perform better and unify all aspects, but also better suit personal needs. To define biodiversity objectively, we need to find the degree of difference between each of these levels - multiple biological systems. Addressing the question posed in the title, we want to find out whether it is feasible to reduce the number of marine biodiversity metrics so that the most sensitive and easily measured diversity properties of the community could be selected. In practice, given the multidimensional nature of biodiversity, it means that we need to search for a minimal necessary and sufficient set of metrics which can be done testing indices derived from the literature in a controlled environment (i.e. simulated ecological communities).

Objectives

The objectives of this paper are threefold:

- to construct a set of measures of fundamental biodiversity ($\mathbf{M}$) by collecting and decomposing biodiversity measures found in the literature;
- to test the relationships among various measures of biodiversity and their sensitivity to known changes in community structure and composition using simulated ecological communities;
- to reduce a set of measures of biodiversity to a minimal set of necessary and sufficient measures ($\mathbf{M}^*$) using multivariate statistics;

Materials and methods

Measures of biodiversity. Biodiversity is manifested at different levels of biological organization, and therefore it has different resolutions. In the search for precise definition of biodiversity metrics we disintegrate each measure of biodiversity in some descriptors ($\mathbf{D}$) taking place at certain organizational levels ($\mathbf{L}$)

$$\mathbf{D}_{i,1} = \begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_i \end{pmatrix} \qquad \mathbf{L}_{1,j} = \begin{pmatrix} L_1 & L_2 & \dots & L_j \end{pmatrix}$$

We recognize a finite set of empirical descriptors (e.g. numbers, feature, pattern, distance or function). A measure of biodiversity can be defined as a scalar combination of one descriptor at one level specifying a component of biodiversity $\mathbf{M}_{i,j} \equiv (\mathbf{D}_i|\mathbf{L}_j)$. If we want to describe biodiversity as a whole and take into account all facets one measure apparently is not sufficient, we need a set or a matrix to accommodate all possible measures.

$$\mathbf{D}_{i,1} \cdot \mathbf{L}_{1,j} = \mathbf{M}_{i,j} = \begin{pmatrix} D_1 L_1 & D_1 L_2 & \dots & D_1 L_j \\ D_2 L_1 & D_2 L_2 & \dots & D_2 L_j \\ \vdots & \vdots & \ddots & \vdots \\ D_i L_1 & D_i L_2 & \dots & D_i L_j \end{pmatrix}$$

Note that $\mathbf{M}_{i,j}$ may be null because not all combinations $\mathbf{D}|\mathbf{L}$ are present in the literature, but using modelling approach we can investigate properties of the complete set and to find minimal necessary and sufficient set $\mathbf{M}^*$.

Simulation. To test various measures of biodiversity we created theoretical datasets displaying the properties of real marine communities (composition and structure). Each dataset was generated by a rigorous taxonomic sampling process applied on taxonomic data made available by ITIS[1] (Bisby et al., 2009). These datasets will be referred to as ecological communities. To ensure the plausibility of taxa content within community, theoretical population has been truncated at a class level and restricted to a list of benthic marine taxonomic groups found around Ireland (BioMar project, Picton et al., 1992). All low-resolution taxa falling into these classes were included. To represent taxonomic sampling process schematically we use plate notation (see Figure 1).
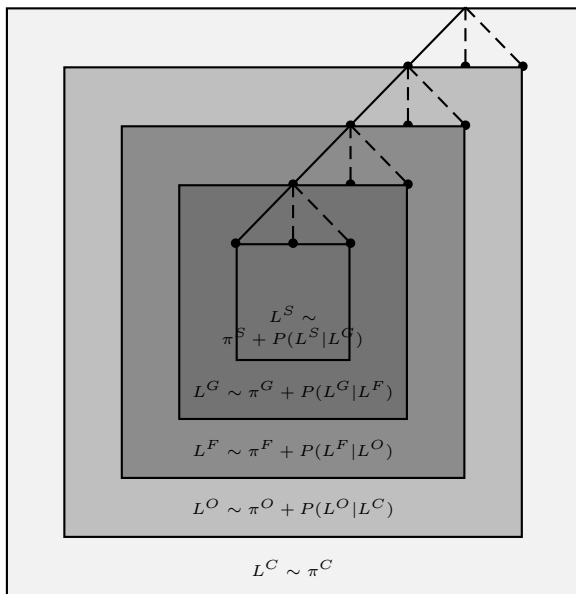

Figure 1: Taxonomic sampling process as a plate notation with a tree representation $T = \{L^C, L^O, L^F, L^G, L^S\}$

[1]Integrated Taxonomic Information System

The taxonomic tree T is presented at several taxonomic levels L^C, L^O, L^F, L^G, L^S shown by different shades of gray (the darkest colour corresponds to the highest number of taxonomic units). We take a stepwise top-down approach and select a number of classes as the first step, then within the classes that were selected we select a number of orders, etc. We repeat this algorithm in a similar fashion all the way down to the species level. On the graph the solid lines indicate further selection, and dashed lines mean that taxonomic units were dropped. Complete model run results in a subtree T_1, T_2, $\ldots T_n$ ($n = 1000$) being generated and described as $T = L^C, L^O, L^F, L^G, L^S$. Each T is a separate ecological community, taxonomic elements of which are defined by some prior probabilities π^C, such that $L^C \sim \pi^C$. For instance the probability of an entity to be selected at a species level is defined by some prior probability π^S (species weights) and probability that appropriate taxonomic group has been selected earlier - conditional probability P or $L^S \sim \pi^S + P(L^S|L^G)$.

Several biodiversity indices such as community composition and structure indices (e.g. Shannon, Jaccard), taxonomic diversity indices (e.g. taxonomic distinctness and distinctiveness according to Clarke and Warwick, 1998) and functional diversity indices (Petchey and Gaston, 2006) across several taxonomic levels have been calculated for each community. While indices reflecting community composition and structure and taxonomic diversity were calculated in a rather straightforward manner, for functional diversity indices an additional information on functional traits was required. To fill this gap, functions on a subset of organisms have been obtained from the biological traits information catalogue - BIOTIC, which is a database of biological traits on selected benthic species (BIOTIC, 2010).

Dimension reduction. Matrix $\mathbf{M}$ includes every kind of biodiversity metrics and it can be best presented as a multidimensional space with axes (measures) pointing in different directions and clouds of point estimates - ecological communities. To reduce the high-dimensional bio-

diversity metric space into low-dimensional one we need to conduct principle component analysis (PCA) which will help to detect projections of measures of biodiversity onto biodiversity space of maximal variability. Principle components by definition will be orthogonal to each other, the proportion of variance explained by each principle component will show the explanatory power of each of the measures. This type of analysis will reduce the multitude of measures of biodiversity ($\mathbf{M}$) into a smaller set of measures (their linear combinations) ($\mathbf{M}^*$) elements of which are strictly necessary and sufficient. The degree of difference between the origin and the point estimate in $\mathbf{M}^*$ will be estimated using Manhattan distance. This measure of difference will give us an Algorithmic Information Criterion (AIC) of each ecological community as it has been explained by Farnsworth and Reid (2010).

Results

The organizational discipline of a relational database (RDB) is ideal support for meta-analysis of biodiversity literature and a good guide for categorizing biodiversity studies, measures and results in a rigorous and consistent way. To analyze patterns in measures of biodiversity the meta-data on relationships between them must be recorded. We organized empirical knowledge on biodiversity measures into RDB which implies the minimal set of orthogonal measures of biodiversity. As a result of a careful review of biodiversity literature we selected 53 quantitative studies that give 180 measures of biodiversity. Studies spread across several study systems were preferred. Various assumptions have been tested using sql queries and matrix of elements $\mathbf{M}$ has been extracted using cartesian join. By mapping data from publications on biodiversity measures into a database we also convey the notion of interrelatedness between different concepts of biodiversity (i.e. $\mathbf{D}$ and $\mathbf{L}$). Correlation between various indices and their sensitivity to known changes in ecological community structure and composition are shown on the Figure 2.
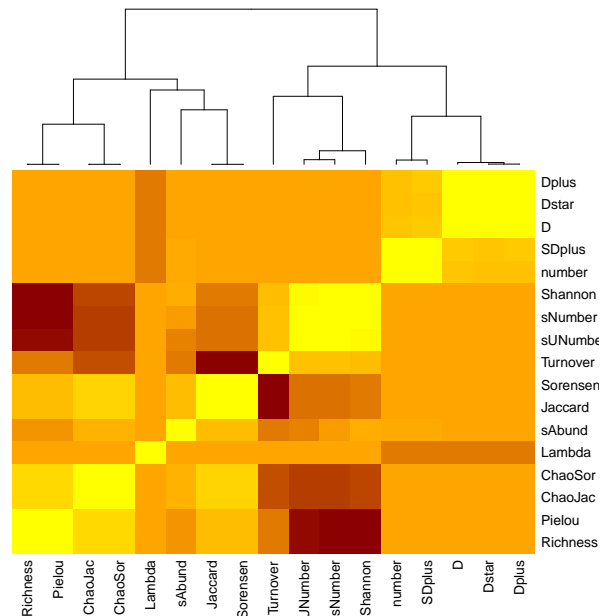

Figure 2: Correlation matrix as a heatmap. The colour intensity indicates the degree of correlation between each pair of indices calculated for a set of communities. The darkest colour corresponds to the negative correlation and the lightest colour is positive correlation, while all colour in between demonstrate low correlation.

Spearman correlation between measures of biodiversity revealed a high correlation between Richness and Pielou ($\rho = 1.00$), Richness and Shannon ($\rho = -1.00$). Interestingly, taxonomic diversity indices have very low correlation with community structure and composition indices (ρ fluctuate around 0 ± 0.04). The dendrogram placed on the top of the figure is used to illustrate the clustering of the variables. Depending on a cut-off point several clusters can be clearly distinguished. The first three axis of the PCA carried out on 17 variables accounted for 68.8% of total inertia. All variables were scaled to have unit variance before the analysis took place. The calculation was done by a singular value decomposition of the centered data matrix. There is a clear division between variables contributing to principle components: while PC1 is mostly explained by community structure and composition indices, PC2 by taxonomic diversity indices (see Figure
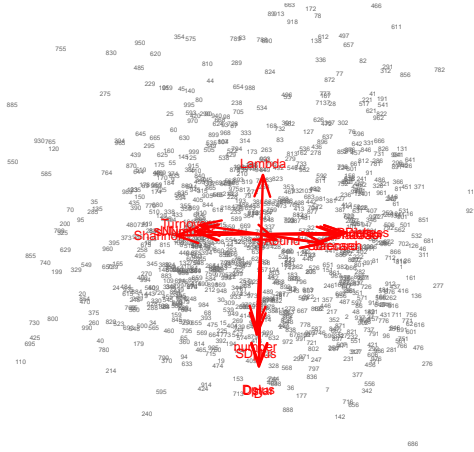
4

3).

Figure 3: Principal component analysis (PC1 and PC2) carried out on artificial communities generated from bentic marine list of species found around Ireland using taxonomic sampling algorithm.

Measures of biodiversity that give the highest proportion of inertia explained form a new low-dimensional space of measures of biodiversity, this new biodiversity space is effectively a minimal necessary and sufficient set of measures of biodiversity $\mathbf{M}^*$. Having specified M^* and as it holds all kinds of metrics but excludes correlation among them, an objective valuation of biodiversity is now possible using AIC.

Conclusions

In this paper we addressed biodiversity in a more integrated and formal way starting with an attempt to give a precise definition of biodiversity. We constructed the manifold permutation matrix $\mathbf{M}$ accommodating all possible measures of biodiversity - combinations $\mathbf{D}|\mathbf{L}$. Matrix $\mathbf{M}$ has been populated with measures of biodiversity found in the literature and reduced to a minimal necessary and sufficient set $\mathbf{M}^*$ using multivariate analysis. We used taxonomic sampling algorithm applied to a list of benthic marine species found around Ireland to generate large number of artificial ecological communities with a plausible taxa content. Using these communities we tested the relationship and sensitivity of biodiversity indicators to known changes in community structure

and composition. It has been shown that some of the indicators are closely related and similar enough to form clusters. We know that biodiversity measures represent various facets of biodiversity and using single measure may result in a loss of information. To minimize this loss, we reduced multidimensionality of biodiversity by eliminating measures of biodiversity that contribute least in multidimensional biodiversity space. Finally, we constructed a limited set of biodiversity metrics which proved to be a good approximation of total biodiversity. As a follow-up research it would be interesting to address the comparability between the simulated ecological communities and empirical communities by means of model validation.

Acknowledgement

References

Begon, M., Townsend, C., and Harper, J. (2006). *Ecology: From individuals to ecosystems*. Oxford: Blackwell, 4 edition.

BIOTIC (2010). The biological traits information catalogue. Digital resource at http://www.marlin.ac.uk.

Bisby, F., Roskov, Y., Orrell, T., Nicolson, D., Paglinawan, L., Bailly, N., Kirk, P., Bourgoin, T., Baillargeon, G., and eds (2009). Species 2000 and itis catalogue of life: 2009 annual checklist. Digital resource at www.catalogueoflife.org/annual-checklist/2009/ Species 2000: Reading, UK.

Clarke, K. R. and Warwick, R. M. (1998). A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, 35(4):523–531.

DeLong, D. C. (1996). Defining biodiversity. *Wildlife society bulletin*, 24(4):738–749.

Farnsworth, K., Longo, A., and Lyashevska, O. (2010). Substitutability and the value of biodiversity as information. Unpublished (under review: Environmental and Resource Economics).

Farnsworth, K. D. and Reid, D. (2010). ICES CM 2010 Q:09. In *Theme Session Q: Marine Biodiversity - the science and management needed to meet the 2010 commitments.*

Gaston, K. and Spicer, J. (2004). *Biodiversity: An Introduction.* Blackwell Publishing, 2nd edition.

Ghilarov, A. (1996). What does 'biodiversity' mean – scientific problem or convenient myth? *Trends in Ecology and Evolution*, 11(7):304–306.

Hamilton, A. J. (2005). Species diversity or biodiversity? *Journal of Environmental Management*, 75(1):89–92.

Marcot, B. G. (2007). Biodiversity and the lexicon zoo. *Forest Ecology and Management*, 246(1):4–13.

Noss, R. F. (1990). Indicators for monitoring biodiversity - a hierarchical approach. *Conservation Biology*, 4(4):355–364.

Petchey, O. L. and Gaston, K. J. (2006). Functional diversity: back to basics and looking forward. *Ecology Letters*, 9(6):741–758.

Picton, B., Emblow, C., Morrow, C., Sides, E., Tierney, P., McGrath, D., McGeough, G., McCrea, M., Dinneen, P., Falvey, J., Dempsey, S., Dowse, J., and Costello, M. J. (1992). Marine sites, habitats and species data collected during the biomar survey of ireland. Digital resource.

Purvis, A. and Hector, A. (2000). Getting the measure of biodiversity. *Nature*, 405(6783):212–219.