

# Bridging the gap between ecological diversity indices and measures of biodiversity with Shannon's entropy: comment to Izsák and Papp

Carlo Ricotta \*

*Department of Plant Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, 00185 Rome, Italy*

Received 31 May 2001; accepted 11 October 2001

---

## Abstract

Most ecological diversity indices summarize the information about the relative abundances of community species without reflecting taxonomic differences between species. Nevertheless, in the environmental conservation practice, data on species abundances are generally unknown. In such cases, to summarize the conservation value of a given site, so-called 'biological diversity' measures need to be used. Most of these measures are based on taxonomic relations among species and ignore species relative abundances. In a recent paper, Izsák and Papp suggest that the quadratic entropy index ( $Q$ ) is the only diversity index used to date in the ecological practice that incorporates both species relative abundances and a measure of the pairwise taxonomic differences between species in the analyzed data set. I show here that a number of traditional ecological diversity measures can be generalized to take into account a taxonomic weighting factor. Since these new indices violate part of the mathematical properties that an index should meet to be termed an ecological diversity index, I defined this new family of indices 'weak diversity indices'. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Biodiversity measures; Ecological diversity; Quadratic entropy; Shannon's entropy; Species richness

---

## 1. Introduction

Traditional ecological diversity indices, such as the Shannon index and the Simpson index, summarize the information about the relative abundances of species within a species sample or community without regard to species names or differences between species. Nevertheless, for large-scale environmental protection purposes,

data on species abundances are generally unknown. Often the only available data is the number of species. For instance, focusing on conservation problems, species abundances are mostly irrelevant. In addition, the comparison of species abundances is largely meaningless between systematically remote organisms, such as oaks and orchids (Izsák and Papp, 2000).

In these circumstances, to summarize the conservation value of a given site, so-called 'biological diversity' measures are generally preferred. Most of these measures are based on taxonomic

---

\* Tel.: + 39-06-4991-2408; fax: + 39-06-4457-540.

E-mail address: [carlo.ricotta@uniroma1.it](mailto:carlo.ricotta@uniroma1.it) (C. Ricotta).

relations among species and are not linked to traditional diversity indices. To our knowledge, Vane-Wright et al. (1991) were the first to suggest that, for conservation purposes, we should quantify the ‘taxonomic distinctiveness’ of different species. For a given data set, Vane-Wright et al. (1991) proposed a measure of the degree of ‘independent evolutionary history’ that takes into account the topology of the corresponding taxonomic tree, but ignores species relative abundances. Successively, various refinements of this basic idea have been actively pursued (Faith, 1992, 1995). In this view, Izsák and Papp (2000) suggest that the quadratic entropy index (Rao, 1982) is the only ecological diversity index used to date in the ecological practice that incorporates both species relative abundances and a measure of the pairwise taxonomic differences between species in the analyzed data set.

Consider an  $N$ -species community characterized by the relative abundance vector  $\mathbf{P} = (p_1, p_2, \dots, p_N)$  such that  $0 \leq p_i \leq 1$  and  $\sum_{i=1}^N p_i = 1$ . Rao (1982) defined the quadratic diversity index  $Q$  or quadratic entropy as:

$$Q = \sum_{i=1}^N \sum_{j=1}^N d_{ij} p_i p_j, \tag{1}$$

where  $d_{ij}$  is the difference between the  $i$ -th and the  $j$ -th species ( $d_{ij} = d_{ji}$  and  $d_{ii} = 0$ ) and  $Q$  expresses the average difference between two randomly selected individuals. It is easily shown that for  $d_{ij} = \text{constant}$  for all  $i \neq j$ ,  $Q$  reduces to  $d_{ij}(1 - D)$ , where  $D$  is the Simpson index  $\sum_{i=1}^N p_i^2$ . To apply  $Q$  in practice, the differences  $d_{ij}$  need to be specified. Izsák and Papp (2000) propose to generate species differences based on the taxonomic distance between two species, let  $d_{ij}$  stand for the topological distance (i.e. the

number of edges) between the  $i$ -th and the  $j$ -th species in the taxonomic tree. An illustration of an artificial taxonomic tree along with its species distance matrix is given in Fig. 1. For example, in Fig. 1, the taxonomic distance  $d_{ac}$  from species a to species c is 4.

If the species abundance values are unknown (i.e. if the only available data is species richness), we can attribute the same abundance value  $1/N$  to all species present. In this case, Eq. (1) expresses the mean taxonomic distance between two randomly selected species and can be written in the form:

$$J = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N d_{ij}. \tag{2}$$

Izsák and Papp (2000) note that  $J$  does not satisfy set monotonicity, a property generally required for biodiversity measures. This property ensures that the index value will increase by adding a new species to a species set  $A$ . Formally, denoting a biodiversity index by  $I$ ,  $I(A \cup \{x\}) > I(A)$ . Conversely, the extensive counterpart of  $J$ :

$$F = N^2 J = \sum_{i=1}^N \sum_{j=1}^N d_{ij}, \tag{3}$$

where  $F$  is the sum of the taxonomic differences of species in the analyzed data set satisfy set monotonicity.  $F$  is therefore an adequate measure of biodiversity if data on species relative abundances are unknown (Izsák and Papp, 2000). Notice that  $F$  is not simply a (species) richness index insofar as it is not a monotone increasing function of the number of species in the data set. Instead, its values are jointly determined by the number of species and the (taxonomic) differences between the species (Izsák and Papp, 2000).

## 2. Linking ecological diversity and biodiversity with Shannon’s entropy

In a similar manner, it is also possible to bridge the gap between traditional ecological diversity indices based on species relative abundances and conservation-oriented biodiversity indices starting from Shannon’s entropy  $H = -\sum_{i=1}^N p_i \log p_i$ . Shannon’s entropy  $H$  of a given community is a information-theoretic measure of uncertainty in

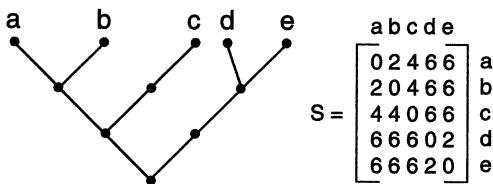


Fig. 1. Illustration of an artificial taxonomic tree along with its species distance matrix to clarify how to measure taxonomic distances.

predicting the relative abundance of species. Since uncertainty is maximal when entropy is highest, information theory forms one of the basic foundations of ecological diversity theory (Orlóci, 1991). Within this framework, Casquilho et al. (1998) proposed a generalized version of Shannon's entropy

$$H_G = - \sum_{i=1}^N w_i p_i \log p_i, \quad (4)$$

where  $w_i$  is a weighting factor that embodies the 'quality' of the  $i$ -th species in the data set. Based on the species distance matrix of Fig. 1, a simple way to compute a taxonomic weight  $w_i$  associated to the  $i$ -th species in the data set is to add all  $d_{ij}$  elements along row  $i$  or column  $i$  of the species matrix. The result of this summation represents the total taxonomic distance from the  $i$ -th species to all other species in the data set. For example, in Fig. 1, the taxonomic weight  $w_a$  associated to species  $a$  is 18.

If the species abundance values are unknown, two expression analogous to  $J$  and  $F$  can be derived (see Eqs. (2) and (3)):

$$J_G = \frac{1}{N} \sum_{i=1}^N w_i \log N \quad (5)$$

and

$$F_G = NJ_G = \log N \sum_{i=1}^N w_i, \quad (6)$$

where  $\sum_{i=1}^N w_i = \sum_{i=1}^N \sum_{j=1}^N d_{ij}$  and  $\log N$  is the entropy introduced by Hartley (1928), who wanted the entropy to depend only upon the number of species, not upon their relative abundances. Therefore, since  $F_G$  takes into account species richness and the topology of the taxonomic tree, but ignores species relative abundances, it can be defined a biological diversity index with a meaningful statistical pedigree *sensu* Izsák and Papp (2000).

It is easy to show that  $Q$  and  $H_G$  violate the usual diversity axiom that for a given number of species  $N$  the maximal diversity arise for an equiprobable species distribution (i.e. a distribution where  $p_i = p_j$  for all  $i \neq j$ ). Another diversity axiom that is violated by  $Q$  and  $H_G$  is the permutation invariance (Pielou, 1975). This postulates that the diversity values corresponding to the relative abundances  $p_1,$

$p_2, \dots, p_N$  and to a  $p'_1, p'_2, \dots, p'_N$  permutation of those are identical. Therefore, I propose to define these indices that derive from traditional ecological diversity indices but that have their greatest value for non-completely even communities 'weak diversity indices'. Notice that the weak diversity formulation of Shannon's entropy reported in Eq. (4) can be easily extended to parametric diversity indices such as the  $\Delta_\beta$  diversity family (Patil and Taillie, 1982)

$$\Delta_\beta = \frac{1}{\beta} \sum_{i=1}^N w_i p_i (1 - p_i^\beta) \quad \text{or to Aczél and Daróczy}$$

$$(1975) \text{ generalized entropy } H^\alpha = \frac{1}{2^{1-\alpha} - 1}$$

$\sum_{i=1}^N w_i (p_i^\alpha - p_i)$  to obtain a vector representation of community diversity as suggested by Patil and Taillie (1982). Will weak diversity indices be able to effectively summarize the important information contained in the analyzed species assemblages? This is a critical question and their answer will provide valuable insights into the practice of nature conservation.

## References

- Aczél, J., Daróczy, Z., 1975. On Measures of Information and their Characterizations. Academic Press, New York.
- Casquilho, J., Neves, M., Rego, F., 1998. Extensões da função de Shannon e equilíbrios de proporções—uma aplicação ao mosaico de paisagem. Anais do Instituto Superior de Agronomia 48, 77–99.
- Faith, D.P., 1992. Conservation evaluation and phylogenetic diversity. Biol. Conserv. 61, 1–10.
- Faith, D.P., 1995. Phylogenetic pattern and the quantification of organismal biodiversity. In: Hawksworth, D.L. (Ed.), Biodiversity Measurement and Estimation. Chapman & Hall, London, pp. 45–58.
- Hartley, R.V., 1928. Transmission of information. Bell Syst. Techn. J. 7, 535–563.
- Izsák, J., Papp, L., 2000. A link between ecological diversity indices and measures of biodiversity. Ecol. Model. 130, 151–156.
- Orlóci, L., 1991. Entropy and Information. SPB Academic Publishing, The Hague, NL.
- Patil, G.P., Taillie, C., 1982. Diversity as a concept and its measurement. J. Am. Stat. Assoc. 77, 48–567.
- Pielou, E.C., 1975. Ecological Diversity. Wiley Interscience, New York.
- Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theor. Popul. Biol. 21, 24–43.
- Vane-Wright, R.I., Humphries, C.J., Williams, P.M., 1991. What to protect: systematics and the agony of choice. Biol. Conserv. 55, 235–254.