The weighted quadratic index of biodiversity for pairs of species: a generalization of Rao's index

Radu Cornel Guiasu[1*], Silviu Guiasu[2]

[1]Environmental and Health Studies Program, Department of Multidisciplinary Studies, Glendon College, York University, Toronto, Canada; [*]Corresponding Author: rguiasu@glendon.yorku.ca
[2]Department of Mathematics and Statistics, York University, Toronto, Canada; guiasus@pascal.math.yorku.ca

ABSTRACT

The distribution of biodiversity at multiple sites of a region has been traditionally investigated through the additive partitioning of the regional biodiversity, called γ-diversity, into the average within-site biodiversity or α-diversity, and the biodiversity among sites, or β-diversity. The standard additive partitioning of diversity requires the use of a measure of diversity which is a concave function of the relative abundance of species, like the Shannon entropy or the Gini-Simpson index, for instance. When a phylogenetic distance between species is also taken into account, Rao's quadratic index has been used as a measure of dissimilarity. Rao's index, however, is not a concave function of the distribution of relative abundance of either individual species or pairs of species and, consequently, only some nonstandard additive partitionings of diversity have been given using this index. The objective of this paper is to show that the weighted quadratic index of biodiversity, a generalization of the weighted Gini-Simpson index to the pairs of species, is a concave function of the joint distribution of the relative abundance of pairs of species and, therefore, may be used in the standard additive partitioning of diversity instead of Rao's index. The replication property of this new measure is also discussed.

Keywords: Additive Partitioning of Biodiversity; Biodiversity Measures; Rao's Index of Dissimilarity; Replication Property; Weighted Alpha-, Beta-, and Gamma-Diversities; Weighted Gini-Simpson Index

1. INTRODUCTION

The amount of turnover among species assemblages is an important component for the conservation of biodiversity. The diversity turnover is called β-diversity while the regional diversity and the mean of the local diversities are called γ-diversity and α-diversity, respectively. Unlike the α-diversity and the γ-diversity, there is no consensus about how to interpret and calculate the β-diversity. According to Whittaker [1,2] who introduced the terminology, β-diversity is the ratio between γ-diversity and α-diversity. This is the multiplicative partitioning of diversity. According to MacArthur [3], MacArthur and Wilson [4], and Lande [5], β-diversity is the difference between γ-diversity and α-diversity. This is the additive partitioning of diversity. Initially, the diversity measures used in the partitioning of diversity (such as the classic Shannon's entropy [6] and the Gini-Simpson index [7,8]) depended only on the relative abundance of species. Later, Rao [9] introduced a dissimilarity measure that takes into account both the relative abundance of species and an arbitrary distance between species (for example, the phylogenetic distance). When Rao's index is used as a measure in the additive partitioning of biodiversity, β-diversity reflects the dissimilarity between the diversities of the sites of the respective region.

Within the last decade or so, many published studies attempted to use Rao's index in the additive partitioning of biodiversity. Some papers have looked for special types of distance matrices for which Rao's index is a concave function of the distribution of the relative abundance of individual species. Other papers have looked for nonstandard, particular additive partitioning of diversity in α-, β-, and γ-diversities. The main difficulty comes from the fact that, for an arbitrary distance between species, Rao's index is a quadratic, but not concave, function of the distribution of the relative abundance of individual species and a linear, but not quadratic, function of the joint distribution of the relative abundance of distinct pairs of species and, as a consequence, it is not suited for the general standard additive partitioning into α-, β-, and γ-diversities. The present paper proposes the use of a

weighted quadratic indicator instead of Rao's index. The weighted quadratic indicator, a generalization of the weighted Gini-Simpson index to the pairs of species, proves to be a concave function of the joint distribution of the relative abundance of pairs of species and is suitable for use in the additive partitioning of biodiversity induced by the pairs of species when a distance between species is taken into account. The formula for calculating the β-diversity, as a measure of dissimilarity among the diversities of the sites, is given. A numerical example is presented in order to illustrate how the mathematical formalism may be applied. In Section 4, a simple algebraic transformation is presented, which allows the use of the weighted quadratic index in the multiplicative partitioning of biodiversity, and the corresponding replication property is discussed.

2. RAO'S INDEX OF DISSIMILARITY

If there are n species, let $\theta = (p_1, \cdots, p_n)$ be a vector such that:

$$p_i \geq 0, (i = 1, \cdots, n); \sum_i p_i = 1,$$

where p_i is the relative abundance of species i. The classic measures of biodiversity are the Shannon entropy ([6]):

$$H(\theta) = -\sum_i p_i \ln p_i,$$

(where $0 \ln 0 = 0$, extending by continuity the function $-x \ln x$ to be equal to zero at the origin), and the Gini-Simpson index ([7,8]):

$$GS(\theta) = 1 - \sum_i p_i^2.$$

Recently, Jost [10,11], and Jost *et al.* [12] gave some examples showing that both the Shannon entropy and the Gini-Simpson index do not behave well when the number of species n is very large. Guiasu and Guiasu [13] showed, however, that the Rich-Gini-Simpson index:

$$RGS(\theta) = n \sum_i p_i (1 - p_i),$$

which depends explicitly on the number of species (species richness), preserves all the properties of the classic Gini-Simpson index GS but, unlike GS, behaves very well when n is large. Another measure of diversity is the weighted Gini-Simpson index (Guiasu and Guiasu [14]):

$$GS_w(\theta) = \sum_i w_i p_i (1 - p_i),$$

where $w = (w_1, \cdots, w_n)$ is the vector whose components are some nonnegative weights assigned to the species, such as the conservation values of the respective species, for instance. Obviously, GS_w becomes RGS if $w_i = n$ for each $i = 1, \cdots, n$.

Let $\boldsymbol{D} = \left[d_{ij} \right]$ be a square matrix whose entries are the distances between species, such as the phylogenetic distances, for instance. Then, we have:

$$d_{ij} \geq 0, d_{ii} = 0, (i, j = 1, \cdots, n)$$

Rao's index [9], also called quadratic entropy or dissimilarity measure, is:

$$R_D(\theta) = \sum_{ij} d_{ij} p_i p_j.$$

Let us assume that in a certain region there are n species and m sites. In what follows, the subscripts i and j refer to species ($i, j = 1, \cdots, n$) and the subscripts k and r refer to sites, ($k, r = 1, \cdots, m$).

Let $\theta_k = (p_{1,k}, \cdots, p_{n,k})$ be the vector whose components are the relative abundances of the individual species within site k, such that:

$$p_{i,k} \geq 0, (i = 1, \cdots, n); \sum_i p_{i,k} = 1,$$

for each $k = 1, \cdots, m$. A measure of diversity μ may be used in the standard additive partioning of diversity induced by individual species if it is a concave function, which means that it satisfies the inequality:

$$\mu \left(\sum_k \lambda_k \theta_k \right) \geq \sum_k \lambda_k \mu(\theta_k) \qquad (1)$$

for arbitrary parameters such that:

$$\lambda_k \geq 0, (k = 1, \cdots, m), \sum_k \lambda_k = 1. \qquad (2)$$

In such a case, as pointed out by Lande [5], the right-hand side of (1) is the α-diversity, denoted by α, measuring the average local diversity, and the left-hand side of (1) is the γ-diversity, denoted by γ, measuring the regional or global diversity. In the additive partitioning of diversity, the β-diversity is the diference between the γ-diversity and the α-diversity, $\beta = \gamma - \alpha$, representing the variation, changes and dissimilarity between the diversities of the sites within the given region.

The measures of biodiversity H, GS, RGS, and GSW are concave functions and satisfy the inequality (1), for arbitrary parameters (2), and, therefore, are suitable to be used in the additive partioning of diversity induced by individual species.

Viewed as a function of the relative abundance of individual species, R_D is a quadratic function of the relative abundance of individual species. Unfortunately, if the distance matrix $\boldsymbol{D}$ is arbitrary, Rao's index R_D, taken as a diversity measure μ, does not satisfy the inequality (1) for arbitrary parameters (2) and, therefore, cannot be used in a standard additive partitioning of the diversity. There is a vast recent literature (Pavoine *et al.* [15], Ricotta [16], Ricotta and Szeidel [17], Hardy and Senterre [18], Villéger and Mouillot [19], Hardy and Jost [20], Ricotta and Szeidel [21], Sherwin [22], De Bello *et al.* [23], and Tuomisto [24,25] about how to use Rao's index R_D in the additive partioning of diversity. Some research

focused on finding special kinds of distance matrices **D** for which the corresponding Rao's index R_D is a concave function, such as the matrix $\sqrt{D}$ assumed to be Euclidean, for instance. Some other research focused on how to use Rao's index for getting a nonstandard additive partitioning of diversity, which means determining whether some special parameters (2) could be used in order to define analog α-, β-, and γ-diversities corresponding to these particular parameters.

For instance, Hardy and Senterre [18] proposed the following nonstandard statistical framework for partitioning the phylogenetic γ-diversity into α- and β-components using Rao's index.

Let A_{ik} be the abundance (number of individuals) of species i within site k. The relative abundance of species i within site k is:

$$p_{i,k} = A_{ik} / \left(\sum_i A_{ik} \right),$$

and the relative abundance of species i over all sites is defined as:

$$p_i = \sum_k A_{ik} / \left(\sum_{i,k} A_{ik} \right).$$

Let d_{ij} be the distance between the species i and j. Using Rao's index, Hardy and Senterre [18] defined the γ-diversity to be:

$$E_T = \sum_{i,j} d_{ij} p_i p_j,$$

measuring the total or regional diversity. The diversity within-site k is:

$$E_k = \sum_{i,j} d_{ij} p_{i,k} p_{j,k}.$$

The average within-site diversity:

$$E_S = (1/m) \sum_k E_k$$

is called α-diversity. According to the additive partitioning of diversity, β-diversity is defined by them to be:

$$\beta = \gamma - \alpha = E_T - E_S.$$

Villéger and Mouillot [19] generalized the model proposed by Hardy and Senterre, following an approach outlined by Ricotta [16]. They kept E_T and E_k unchanged but replaced E_S by the new:

$$E_S = \sum_k \lambda_k E_k,$$

where

$$\lambda_k = \sum_i A_{ik} / \left(\sum_{i,k} A_{ik} \right), (k = 1, \cdots, m),$$

which is also a particular case of parameters $\lambda_1, \cdots, \lambda_m$ satisfying (2).

According to Hardy and Jost [20], the approaches proposed by Hardy and Senterre [18] and by Villégere and Mouillot [19] are both adequate in their specific contexts but they provide, however, only a nonstandard additive partitioning of biodiversity.

3. WEIGHTED QUADRATIC INDEX

Rao's index is important because it measures diversity taking into account both the dissimilarity between species, as induced by a distance between them, such as a phylogenetic distance for instance, and the relative abundance of species. Let $D = \left[d_{ij} \right]$ be an arbitrary matrix whose entries are the distances between the pairs of n species. Let $\theta_k = \left(p_{1,k}, \cdots, p_{n,k} \right)$ be the vector whose components are the relative abundances of the individual species within site k, such that:

$$p_{i,k} \geq 0, (i = 1, \cdots, n); \sum_i p_{i,k} = 1,$$

for each $k = 1, \cdots, m$.

As mentioned in the previous section, Rao's index R_D is a quadratic function of the relative abundances of the individual species. It, however, may be also viewed as being a linear function of the joint distribution of the relative abundance of the pairs of species. Indeed, let us take the joint distribution of the pairs of species within site k, induced by the distribution of the relative abundance of the individual species θ_k within site, namely, $\Theta_k = \left[\pi_{ij,k} \right]$, where:

$$\pi_{ij,k} = p_{i,k} p_{j,k}, (i, j = 1, \cdots, n). \quad (3)$$

In such a case, Rao's index for site k is:

$$R_D \left(\Theta_k \right) = \sum_{i,j} d_{ij} p_{i,k} p_{j,k} = \sum_{i,j} d_{ij} \pi_{ij,k} \quad (4)$$

and it is indeed a linear function of the joint distribution $\Theta_k = \left[\pi_{ij,k} \right]$ of the relative abundance of pairs of species.

In dealing with species diversity, a good measure of the difference or dissimilarity among the sites in a certain region has to be nonnegative and equal to zero if and only if there is no such difference. Dealing with pairs of species, a measure of diversity μ may be used in the standard additive partitioning of diversity if it is a concave function, which means that it satisfies the inequality:

$$\mu \left(\sum_k \lambda_k \Theta_k \right) \geq \sum_k \lambda_k \mu \left(\Theta_k \right) \quad (5)$$

for arbitrary parameters $\lambda_1, \cdots, \lambda_m$ satisfying (2). In dealing with the diversity induced by the pairs of species, the right-hand side of (5) is the α-diversity, denoted by α, the left-hand side of (5) is the γ-diversity, denoted by γ, and the β-diversity is $\beta = \gamma - \alpha$. As R_D is a linear function of Θ_k, if we take $\mu = R_D$, (5) becomes an equality. But in such a case the corresponding β-diversity induced by the pairs of species is equal to zero, for any distance matrix **D** and any relative abundance of species. Therefore, Rao's index is not suitable for use in the addi-

tive partitioning of diversity induced by pairs of species when a dissimilarity distance between species is taken into account.

The solution proposed here is to replace Rao's index (4) by the weighted Gini-Simpson quadratic index for pairs of species:

$$GS_D(\Theta_k) = \sum_{i,j} d_{ij} p_{i,k} p_{j,k} (1 - p_{i,k} p_{j,k}) \quad (6)$$

This is a quadratic function of $\Theta_k = [\pi_{ij,k}]$ defined by (3), which satisfies the inequality (5) for arbitrary parameters $\lambda_1, \cdots, \lambda_m$ for which (2) holds. Therefore, GS_D is suitable for use in the additive partitioning of diversity induced by the pairs of species when both a distance between species and the relative abundance of species are taken into account.

In order to prove that GS_D satisfies the inequality (5), which mathematically means that it is a concave function of Θ_k, we move to a more general context which, however, makes the proof simpler and more elegant. Thus, let $\Theta_k = [\pi_{ij,k}]$ be an arbitrary joint probability distribution of the pairs of species, where $\pi_{ij,k}$ is the probability of the pair of species $\{i, j\}$ within site k, such that:

$$\pi_{ij,k} \geq 0, (i,j = 1, \cdots, n); \sum_{i,j} \pi_{ij,k} = 1. \quad (7)$$

In a more general context, let $W = [w_{ij}]$ be a square matrix whose components are arbitrary nonnegative weights assigned to the pairs of species. Denote by:

$$GS_W(\Theta_k) = \sum_{i,j} w_{ij} \pi_{ij,k} (1 - \pi_{ij,k}) \quad (8)$$

the weighted quadratic index of site k.

Remarks: 1) Let us note that if $w_{ij} = 1$, for all i and j, then (8) becomes the classic Gini-Simpson index assigned to the pairs of species.

2) If $w_{ij} = n^2$, for all i and j, where n^2 is the number of pairs of n species, then (8) becomes the Rich-Gini-Simpson index assigned to the pairs of species.

3) If the weights are the distances between species, $w_{ij} = d_{ij}$, and the species are independent, $\pi_{ij,k} = p_{i,k} p_{j,k}$, then $GS_W(\Theta_k)$ from (8) becomes $GS_D(\Theta_k)$ from (6).

4) Another case of interest is when the species are independent, namely $\pi_{ij,k} = p_{i,k} p_{j,k}$, and the weights are $w_{ij} = \frac{1}{2}(v_i + v_j) d_{ij}$, where d_{ij} is the distance between the species $\{i, j\}$ and v_i is the value, such as the conservation value for instance, of species i, in which case (8) becomes:

$$GS_{v,D}(\Theta_k) = \sum_{i,j} \frac{1}{2}(v_i + v_j) d_{ij} p_{i,k} p_{j,k} (1 - p_{i,k} p_{j,k}).$$

If $v_i = 1$ for all i, then $GS_{v,D}(\Theta_k)$ becomes $GS_D(\Theta_k)$ given by (6).

The weighted quadratic index $GS_W(\Theta_k)$ given by (8) is a concave function of the joint probability distribution $\Theta_k = [\pi_{ij,k}]$ given by (7). Indeed, let:

$$\lambda_k \geq 0, (k = 1, \cdots, m); \sum_k \lambda_k = 1$$

be arbitrary parameters. Taking into account that:

$$-\lambda_k^2 \pi_{ij,k}^2 + \lambda_k \pi_{ij,k}^2 = \lambda_k (1 - \lambda_k) \pi_{ij,k}^2$$
$$= \lambda_k (\lambda_1 + \cdots + \lambda_{k-1} + \lambda_{k+1} + \cdots + \lambda_m) \pi_{ij,k}^2$$
$$= (\lambda_1 \lambda_k + \cdots + \lambda_{k-1} \lambda_k + \lambda_k \lambda_{k+1} + \cdots + \lambda_k \lambda_m) \pi_{ij,k}^2,$$

for every $1 \leq k \leq m$, we get:

$$\beta = \gamma - \alpha = GS_W \left(\sum_k \lambda_k \Theta_k \right) - \sum_k \lambda_k GS_W(\Theta_k)$$
$$= \sum_{i,j} w_{ij} \left(\sum_k \lambda_k \pi_{ij,k} \right) \left(1 - \sum_k \lambda_k \pi_{ij,k} \right)$$
$$- \sum_k \lambda_k \sum_{i,j} w_{ij} \pi_{ij,k} (1 - \pi_{ij,k})$$
$$= \sum_{i,j} w_{ij} \left(\sum_k \lambda_k (1 - \lambda_k) \pi_{ij,k}^2 - \sum_{k \neq r} \lambda_k \lambda_r \pi_{ij,k} \pi_{ij,r} \right) \quad (9)$$
$$= \sum_{i,j} w_{ij} \left[\sum_{k<r} \lambda_k \lambda_r (\pi_{ij,k}^2 - 2\pi_{ij,k} \pi_{ij,r} + \pi_{ij,r}^2) \right]$$
$$= \sum_{i,j} w_{ij} \sum_{k<r} \lambda_k \lambda_r (\pi_{ij,k} - \pi_{ij,r})^2.$$

We can see that $\beta \geq 0$ and β is equal to zero if and only if the sites have identical joint relative distribution, namely, $\pi_{ij,k} = \pi_{ij}$, for all $k = 1, \cdots, m$.

If the weight is the distance between species and the species are independent as far as their relative abundance is concerned, namely:

$$w_{ij} = d_{ij}, \pi_{ij,k} = p_{i,k} p_{j,k},$$

for all $i, j = 1, \cdots, n$ and $k = 1, \cdots, m$, where $p_{i,k}$ is the relative abundance of species i within site k, then the corresponding α-, γ- and β-diversities, with respect to the parameters $\lambda_1, \cdots, \lambda_m$, become:

$$\alpha = \sum_k \lambda_k GS_D(\Theta_k)$$
$$= \sum_k \lambda_k \sum_{i,j} d_{ij} p_{i,k} p_{j,k} (1 - p_{i,k} p_{j,k}), \quad (10)$$

$$\gamma = GS_D \left(\sum_k \lambda_k \Theta_k \right)$$
$$= \sum_{i,j} d_{ij} \left(\sum_{i,j} \lambda_k p_{i,k} p_{j,k} \right) \left(1 - \sum_k \lambda_k p_{i,k} p_{j,k} \right), \quad (11)$$

where, according to (9), we have:

$$\beta = \gamma - \alpha = \sum_{i,j} d_{ij} \sum_{k<r} \lambda_k \lambda_r (p_{i,k} p_{j,k} - p_{i,r} p_{j,r})^2 \quad (12)$$

We have $\beta \geq 0$ and $\beta = 0$ if and only if the relative abundance of the pairs of species $p_{i,k} p_{j,k}$ is the same for each site k.

Numerical example: Let A_{ik} be the abundance (number of individuals) of species i within site k. Using an example given by Villéger and Mouillot [19], let us assume that there are three species ($n = 3$), three sites ($m =$

3), and the absolute frequencies of the species are:

$$A_{11} = 1, \quad A_{21} = 1, \quad A_{31} = 2,$$
$$A_{12} = 28, \quad A_{22} = 1, \quad A_{32} = 1,$$
$$A_{13} = 1, \quad A_{23} = 1, \quad A_{33} = 2.$$

Therefore, the respective relative frequencies are:

$$P_{1,1} = 0.25, \quad P_{2,1} = 0.25, \quad P_{3,1} = 0.50,$$
$$P_{1,2} = 0.934, \quad P_{2,2} = 0.033, \quad P_{3,2} = 0.033,$$
$$P_{1,3} = 0.25, \quad P_{2,3} = 0.25, \quad P_{3,3} = 0.50.$$

The distance between species is assumed to be defined by: $d_{12} = 1, d_{13} = 2, d_{23} = 2$. For the parameters:

$$\lambda_1 = \lambda_2 = \lambda_3 = 1/3,$$

the Formulas (10)-(12) give: $\alpha = 0.361$, $\gamma = 0.678$, $\beta = 0.317$.

Remark: If the species are independent, namely $\pi_{ij} = p_i p_j$, where p_i is the relative abundance of species i, and the weights are $w_{ij} = n^2 \frac{1}{2} (v_i + v_j) d_{ij}$, where d_{ij} is the distance between the species $\{i, j\}$ and v_i is the value of species i, such as its conservation value for instance, in which case the weighted quadratic index:

$$GS_W (\Theta) = \sum_{i,j} w_{ij} \pi_{ij} (1 - \pi_{ij})$$

becomes the weighted Rich-Gini-Simpson quadratic index:

$$RGS_{v,D} (\Theta_k) = n^2 \sum_{i,j} \frac{v_i + v_j}{2} d_{ij} p_{i,k} p_{j,k} (1 - p_{i,k} p_{j,k}) \quad (13)$$

a measure of diversity which depends not only on the distance between species, the relative abundance of the distinct pairs of species, and the value of the species, but also on the explicit number of distinct pairs of species (the richness induced by the pairs of species). If all species have the same value, $v_i = 1, (i = 1, \cdots, n)$, then (13) is the weighted Rich-Gini-Simpson version of the weighted Gini-Simpson quadratic index (6).

4. REPLICATION PROPERTY

Dealing with the multiplicative partitioning of diversity, Whittaker ([1,2]) suggested the use of the exponential of the Shannon entropy and the algebraic inverse of the Gini-Simpson index as measures of biodiversity. Following this line of thought and taking the exponential transformation of a general entropy studied by Rényi [26], Hill [27] introduced a unified index of diversity suitable to use in the multiplicative partitioning of diversity. Hill [27] and Jost [28] noticed that these multiplicative measures have the so-called doubling property (or replication property) according to which a measure of diversity should double when two identically distributed but distinct communities (with no shared species) are added together in equal proportions. In a personal correspondence, Professor C. Ricotta asked whether there is a

reasonable index transformation of the weighted Gini-Simpson index that possesses this doubling property. The answer is yes.

4.1. Dealing with Individual Species

If in a community we have n species, such that the distribution of the relative abundance of these species is $\theta = (p_1, \cdots, p_n)$ and the weights assigned to these species are $w = (w_1, \cdots, w_n)$, then the corresponding weighted Gini-Simpson index:

$$GS_w (\theta) = \sum_i w_i p_i (1 - p_i),$$

which can be used in the additive partitioning of diversity induced by individual species, may be transformed into the measure of diversity:

$$1 / \left[\sum_i w_i p_i - GS_w (\theta) \right] = \left(\sum_i w_i p_i^2 \right)^{-1}, \quad (14)$$

which can be used in the multiplicative partitioning of diversity induced by the individual species. The measure (14) has the doubling property. Indeed, if we have two communities, A and B, having n species each, all being different but having the same distribution of the relative abundance $\theta = (p_1, \cdots, p_n)$ and the same qualitative weights $w = (w_1, \cdots, w_n)$, assigned to the individual species, then the union C of these two communities will have $2n$ species with the distribution of the corresponding relative abundance of these species $(q_1, \cdots, q_{2n})$, where $q_i = q_{i+n} = p_i / 2$, and the corresponding weights $(u_1, \cdots, u_{2n})$, where $u_i = u_{i+n} = w_i$. The diversity of C, as measured by (14), applied to the $2n$ species of C is:

$$\left(\sum_{\ell} u_{\ell} q_{\ell}^2 \right)^{-1} = 2 \left(\sum_i w_i p_i^2 \right)^{-1}, \text{ where } (\ell = 1, \cdots, 2n) ; \textit{i.e.}$$

twice the diversity of A or B.

Chao *et al.* [29] claim that the doubling property is "an important requirement for species-neutral diversity". This replication property refers, however, to a very singular situation because, practically, it is almost impossible to see two communities having the same number of entirely different species with exactly the same relative abundance. But there is another problem here. Thus, Jost ([10,28]) states that: "The diversity of a community (say community C) should double if every species is divided into two equal groups, say males (community A) and females (community B) and each group is considered to be a distinct species…A measure of diversity doubles when two identically distributed but distinct communities, with no shared species, (like communities A and B, respectively) are added together in equal proportions". We notice, however, that the simple union of the identically diverse communities A and B does not reflect the diversity of community C, because C contains indeed the diversities of A and B, taken separately, but has now the additional gender diversity, which is missing in A and in

B. The diversity of the union (C) is more than the sum of the diversities of A and B, taken separately.

4.2. Dealing with Pairs of Species

If there are n species in a community with the joint distribution of relative abundance $\Theta = \left[\pi_{ij} \right]$ of the pairs of species and the weights $W = \left[w_{ij} \right]$ assigned to the pairs of species, then the corresponding weighted Gini-Simpson index :

$$GS_W(\Theta) = \sum_{i,j} w_{ij} \pi_{ij} (1 - \pi_{ij}),$$

which can be used in the additive partitioning of diversity induced by pairs of species, may be transformed into the measure of diversity:

$$1 \Big/ \Big[\sum_{i,j} w_{ij} \pi_{ij} - GS_w(\Theta) \Big] = \Big(\sum_{i,j} w_{ij} \pi_{ij}^2 \Big)^{-1}, \quad (15)$$

which can be used in the multiplicative partitioning of diversity induced by the pairs of species. The measure (15) has the quadrupling property. Indeed, assume that we have two communities, A and B, with n species each, all being different (only males $\{M_1, \cdots, M_n\}$ in A and only females $\{F_1, \cdots, F_n\}$ in B, for instance) with the same joint distribution Θ of the relative abundance of the pairs of species and the same weights W assigned to the pairs of species. Thus w_{ij} is the same weight assigned to the pairs $\{M_i, M_j\}, \{M_i, F_j\}, \{F_i, M_j\}, \{F_i, F_j\}$, which happens, for instance, when the distance d_{ij} between species i and j, taken as the weight w_{ij}, assigned to the pair of species $\{i, j\}$, does not depend on sex. Let C be the union of communities A and B. It has $2n$ species $\{M_1, \cdots, M_n, F_1, \cdots, F_n\}$ with the joint $2n \times 2n$ distribution of the relative abundance of the pairs of species from C:

$$\boldsymbol{\Theta} = \left[\boldsymbol{\pi}_{\ell s} \right] = \begin{bmatrix} \Theta/4 & \Theta/4 \\ \Theta/4 & \Theta/4 \end{bmatrix}, (\ell, s = 1, \cdots, 2n),$$

and the $2n \times 2n$ matrix of the weights assigned to the the pairs of species from C:

$$\boldsymbol{W} = \left[\boldsymbol{w}_{\ell s} \right] = \begin{bmatrix} W & W \\ W & W \end{bmatrix}, (\ell, s = 1, \cdots, 2n).$$

Then, the diversity of C, as measured by (15), is:

$$\Big(\sum_{\ell s} \boldsymbol{w}_{\ell s} \boldsymbol{\pi}_{\ell s}^2 \Big)^{-1} = 4 \Big(\sum_{i,j} w_{ij} \pi_{ij}^2 \Big)^{-1},$$

which shows that the community C has a diversity four times larger than the diversity of A or B. But the same comment may be made as above: community C, the union of identical communities A and B, has not only four times the diversity of A or B (which is the quadrupling replication property for the pairs of species), but also has

the supplemental gender diversity which is missing in A or B.

5. CONCLUSIONS

Rao [9] introduced a dissimilarity measure (4) that takes into account both the relative abundance of species and an arbitrary distance between species. When Rao's index is used in the additive partitioning of diversity, the β-diversity reflects the dissimilarity between the diversities of the sites of the respective region.

Rao's dissimilarity measure takes into account both the relative abundance of species and an arbitrary distance between species (for example, the phylogenetic distance). There is a large number of papers published in the last decade dealing with different attempts at using Rao's index in the additive partitioning of biodiversity. Some papers have looked for special types of distance matrices for which Rao's index is a concave function of the distribution of the relative abundance of individual species. Other papers have looked for nonstandard, particular additive partitioning of diversity into analog α-, β-, and γ-diversities. The main difficulty comes from the fact that, for an arbitrary distance between species, Rao's index is a quadratic, but not concave, function of the distribution of the relative abundance of the individual species and a linear, but not quadratic, function of the joint distribution of the relative abundance of the distinct pairs of species and, as a consequence, it is not suitable for use in the general standard additive partitioning of diversity into α-, β-, and γ-diversities. The present paper proposes the use of a weighted quadratic indicator (6) instead of Rao's index (4). The weighted quadratic indicator, a generalization of the weighted Gini-Simpson index to the pairs of species, proves to be a concave function of the joint distribution of the relative abundance of the pairs of species and is suitable for use in the additive partitioning of biodiversity induced by the pairs of species when a distance between species is taken into account. There is a simple Formula (12) for calculating the β-diversity, as a measure of dissimilarity among the diversities of the sites. A simple algebraic transformation is given which allows the use of the weighted Gini-Simpson index in the multiplicative partitioning of biodiversity induced by the individual species (13) or by the pairs of species (14), and the corresponding replication properties are discussed.

REFERENCES

[1] Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213-251. doi:10.2307/1218190

[2] Whittaker, R.H. (1977) Evolution of species diversity in land communities. In: Hecht, M.K. and Steere, B.W.N.C.,

Eds., *Evolutionary Biology*, Plenum Press, New York, **10**, 1-67.

[3] MacArthur, R.H. (1965) Patterns of species diversity. *Biological Review*, **40**, 510-533. doi:10.1111/j.1469-185X.1965.tb00815.x

[4] MacArthur, R.H. and Wilson, E.O. (1967) The theory of island biogeography. Princeton University Press, Princeton.

[5] Lande, R. (1996) Statistics and partitioning of species diversity and similarity among multiple communities. *Oikos*, **76**, 5-13. doi:10.2307/3545743

[6] Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379-423, 623-656.

[7] Gini, C. (1912) Variabilità e mutabilità. In: Pizzetti, E. and Salvemini, T., Eds., *Rome*: *Libreria Eredi Virgilio Veschi*, Memorie di Metodologica Statistica.

[8] Simpson, E.H. (1949) Measurement of diversity. *Nature*, **163**, 688. doi:10.1038/163688a0

[9] Rao, C.R. (1982) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, **21**, 24-43. doi:10.1016/0040-5809(82)90004-1

[10] Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427-2439. doi:10.1890/06-1736.1

[11] Jost, L. (2009) Mismeasuring biological diversity: Response to Hoffmann and Hoffmann. *Ecological Economics*, **68**, 925-928. doi:10.1016/j.ecolecon.2008.10.015

[12] Jost, L., DeVries, P., Walla, T., Greeney, H., Chao, A. and Ricotta, C. (2010) Partitioning diversity for conservation analyses. *Diversity and Distributions*, **16**, 65-76. doi:10.1111/j.1472-4642.2009.00626.x

[13] Guiasu, R.C. and Guiasu, S. (2010) The Rich-Gini-simpson quadratic index of biodiversity. *Natural Science*, **2**, 1130-1137. doi:10.4236/ns.2010.210140

[14] Guiasu, R.C. and Guiasu, S. (2010) New measures for comparing the species diversity found in two or more habitats. *International Journal of Uncertainty*, *Fuzziness and Knowledge-Based Systems*, **18**, 691-720. doi:10.1142/S0218488510006763

[15] Pavoine, S., Ollier, S. and Pontier, D. (2005) Measuring diversity from dissimilarities with Rao's quadratic entropy: Are any dissimilarities suitable? *Theoretical Population Biology*, **67**, 231-239. doi:10.1016/j.tpb.2005.01.004

[16] Ricotta, C. (2005) Additive partitioning of Rao's quadratic diversity: A hierarchical approach. *Ecological Modelling*, **183**, 365-371. doi:10.1016/j.ecolmodel.2004.08.020

[17] Ricotta, C. and Szeidel, L. (2006) Towards a unifying approach to diversity measures: Bridging the gap between the Shannon entropy and Rao's quadratic index. *Theoretical Population Biology*, **70**, 237-243. doi:10.1016/j.tpb.2006.06.003

[18] Hardy, O.J. and Senterre, B. (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology*, **95**, 493-506. doi:10.1111/j.1365-2745.2007.01222.x

[19] Villéger, S. and Mouillot, D. (2008) Additive partitioning of diversity including species differences: A comment on Hardy and Senterre (2007). *Journal of Ecology*, **96**, 845-848.

[20] Hardy, O.J. and Jost, L. (2008) Interpreting measures of community phylogenetic structuring. *Journal of Ecology*, **96**, 849-852. doi:10.1111/j.1365-2745.2008.01423.x

[21] Ricotta, C. and Szeidel, L. (2009) Diversity partitioning of Rao's quadratic entropy. *Theoretical Population Biology*, **76**, 299-302. doi:10.1016/j.tpb.2009.10.001

[22] Sherwin, W.B. (2010) Entropy and information approaches to genetic diversity and its expression: Genomic geography. *Entropy*, **12**, 1765-1798. doi:10.3390/e12071765

[23] De Bello, F., Lavergne, S., Meynard, C.N., Lepš, J. and Thuiller, W. (2010) The partitioning of diversity: Showing Theseus a way out of the labyrinth. *Journal of Vegetation Science*, **21**, 992-1000. doi:10.1111/j.1654-1103.2010.01195.x

[24] Tuomisto, H. (2010) A diversity of beta diversities: Straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, **33**, 2-22. doi:10.1111/j.1600-0587.2009.05880.x

[25] Tuomisto, H. (2010) A diversity of beta diversities: Straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, **33**, 23-45. doi:10.1111/j.1600-0587.2009.06148.x

[26] Rényi, A. (1961) On measures of entropy and information. In: Neyman, J., Ed., 4*th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, **1**, 547-561.

[27] Hill, M. (1973) Diversity and evenness. A unifying notation and its consequences. *Ecology*, **88**, 2427-2439.

[28] Jost, L. (2006) Entropy and diversity. *Oikos*, **113**, 363-375. doi:10.1111/j.2006.0030-1299.14714.x

[29] Chao, A., Chiu, C.H. and Jost, L. (2010) Phylogenetic diversity measures based on Hill numbers. *Philosophical Transactions of the Royal Society Biological Sciences*, **365**, 3599-3609. doi:10.1098/rstb.2010.0272