Research Article

The Weighted Gini-Simpson Index: Revitalizing an Old Index of Biodiversity

Radu Cornel Guiasu¹ and Silviu Guiasu²

¹ *Environmental and Health Studies Program, Department of Multidisciplinary Studies, Glendon College, York University, 2275 Bayview Avenue, Toronto, ON, Canada M4N 3M6*

² *Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3*

Correspondence should be addressed to Silviu Guiasu, guiasus@pascal.math.yorku.ca

Received 19 September 2011; Revised 22 November 2011; Accepted 6 December 2011

Academic Editor: Jean-Guy Godin

Copyright © 2012 R. C. Guiasu and S. Guiasu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The distribution of biodiversity at multiple sites of a region has been traditionally investigated through the additive partitioning of the regional biodiversity into the average within-site biodiversity and the biodiversity among sites. 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, such as the Gini-Simpson index, for instance. Recently, it was noticed that the widely used Gini-Simpson index does not behave well when the number of species is very large. The objective of this paper is to show that the new weighted Gini-Simpson index preserves the qualities of the classic Gini-Simpson index and behaves very well when the number of species is large. The weights allow us to take into account the abundance of species, the phylogenetic distance between species, and the conservation values of species. This measure may also be generalized to pairs of species and, unlike Rao's index, this measure proves to be a concave function of the joint distribution of the relative abundance of species, being suitable for use in the additive partitioning of biodiversity. The weighted Gini-Simpson index may be easily transformed for use in the multiplicative partitioning of biodiversity as well.

1. Introduction

Measuring biodiversity is a major, and much debated, topic in ecology and conservation biology. The simplest measure of biodiversity is the number of species from a given community, habitat, or site. Obviously, this ignores how many individuals each species has. The best known measures of biodiversity, that also take into account the relative abundance of species, are the Gini-Simpson index GS and the Shannon entropy H . Both measures have been imported into biology from other fields. Thus, Gini [1] introduced his formula in statistics, in 1912. Much later, after 37 years, Simpson [2] pleaded convincingly in favour of using Gini's formula as a measure of biodiversity. Shannon was an engineer who introduced his discrete entropy in information theory [3], in 1948, as a measure of uncertainty, inspired by Boltzmann's continuous entropy from classical statistical mechanics [4], defined half a century earlier. Shannon's

formula was adopted by biologists about 17 years later [5–8], as a measure of specific diversity. This import of mathematical formulas has continued. Rényi, a probabilist, introduced his own entropy [9], in order to unify several generalizations of the Shannon entropy. He was a pure mathematician without any interest in applications, but later, Hill [10] claimed that by taking the exponential of Rényi's entropy we obtain a class of suitable measures of biodiversity, called Hill's numbers, which were praised by Jost [11–13] as being the “true” measures of biodiversity. In 1982, Rao [14], a statistician, introduced the so-called quadratic entropy R , which in fact has nothing to do with the proper entropy and depends not only on the relative abundance of species but also on the phylogenetic distance between species. This function has also been quickly adopted by biologists as a measure of dissimilarity between the pairs of species. In the last 20 years, a lot of other measures of diversity have been proposed. According to Ricotta [15], there is currently

a “jungle” of biological measures of diversity. However, as mentioned by S. Hoffmann and A. Hoffmann [16], there is no unique “true” measure of diversity.

Starting with MacArthur [5], MacArthur and Wilson [7], and Whittaker [8], the distribution of biodiversity at multiple sites of a region has been traditionally investigated through the partitioning of the regional or total biodiversity, called γ -diversity, into the average within-site biodiversity, called α -diversity, and the between-site biodiversity or diversity turnover, called β -diversity. All these diversities, namely, α -diversity, β -diversity, and γ -diversity, should be nonnegative numbers. Unlike α -diversity and γ -diversity, there is no consensus about how to interpret and calculate β -diversity. According to Whittaker [8], 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 [5], Lande [17], and, more recently, Veech et al. [18], β -diversity is the difference between γ -diversity and α -diversity. This is the additive partitioning of diversity.

Let us assume that in a certain region there are n species, m sites, and θ_k is the distribution of the relative abundance of species at site k . Let $\lambda_k > 0$ be an arbitrary parameter assigned to site k , such that $\lambda_1 + \dots + \lambda_m = 1$. These parameters may be used to make adjustments for differences (in size, altitude, etc.) between the sites. If no adjustment is made, we take these parameters to be equal, that is, $\lambda_k = 1/m$, for every k . If μ is a nonnegative measure of diversity, which assigns a nonnegative number to each distribution of the relative abundance of the n species, then the corresponding γ -diversity is $\gamma = \mu(\sum_k \lambda_k \theta_k)$ and the α -diversity is $\alpha = \sum_k \lambda_k \mu(\theta_k)$. The β -diversity is taken to be $\beta = \gamma - \alpha$, in the additive partitioning of diversity, and $\beta = \gamma/\alpha$, in the multiplicative partitioning of diversity. In general, a measure of biodiversity ought to be nonnegative, in which case the corresponding α -diversity and γ -diversity calculated by using such a measure are also nonnegative, as they should be. From a systemic point of view, the β -diversity shows to what extent the total, or regional diversity differs from the average diversity of the communities/habitats/sites taken together, as a system, reflecting the dissimilarity, or differentiation between communities/habitats/sites of the region with respect to the individual species. If the measure of biodiversity is a concave function of the distribution of the relative abundance of species θ_k , then the corresponding β -diversity is $\beta \geq 0$, in the additive partitioning of diversity, and $\beta \geq 1$, in the multiplicative partitioning of diversity, for arbitrary parameters $\lambda_k > 0$, $\lambda_1 + \dots + \lambda_m = 1$. If a measure of diversity μ is not a concave function of the distribution of the relative abundance of species θ_k , then the corresponding β -diversity could be negative, in the additive partitioning of biodiversity, or less than 1, in the multiplicative partitioning of biodiversity, for some parameters satisfying $\lambda_k > 0$, $\lambda_1 + \dots + \lambda_m = 1$, which is absurd. As discussed by Jost [11, 12], if a measure of diversity μ is not a concave function of the distribution of the relative abundance of species θ_k , we can still attempt the partitioning of biodiversity into α -, β -, and γ -diversity if a new kind of α -diversity may be introduced. This new type of α -diversity would be

based on a different way of averaging the diversities of the individual communities/habitats/sites instead of the simple, golden mean value $\alpha = \sum_k \lambda_k \mu(\theta_k)$ from statistics, which works so well for the concave measures of diversity. However, finding such an unorthodox, nonstandard α -diversity when the measure of biodiversity μ is not concave is not easy. It is also difficult to find a mathematical interpretation for such a new kind of α -diversity. In spite of the passage of time, the most popular measures of biodiversity are still GS, H , and R . Both GS and H are concave functions of the distribution of the relative abundance of species and therefore can be used for doing the additive partitioning of biodiversity. The first two Hill's numbers are mathematical transformations of GS and H , namely, $Hl_1 = 1/(1 - GS)$ and $Hl_2 = \exp(H)$, and are used in the multiplicative partitioning of biodiversity. Recently, however, it was noted [13, 19] that both Shannon's entropy and the Gini-Simpson index do not behave well when the number of species is very large. On the other hand, when a distance between species, such as the phylogenetic distance for instance, is also taken into account along with the relative abundance of species, Rao's index [14] is a widely used measure of dissimilarity. But, unfortunately, Rao's index is not a concave function of the distribution of the relative abundance of species, for an arbitrary distance matrix between species. Consequently, it proves to be suitable for use in the standard additive partitioning of diversity only in some special cases, but not in general. The objective of this paper is to show that the weighted Gini-Simpson quadratic index, a generalization of the classic Gini-Simpson index of biodiversity, offers a solution to both of the drawbacks just mentioned. Unlike Shannon's entropy and the classic Gini-Simpson index, this new weighted measure of biodiversity behaves very well even if the number of species is very large. The weights allow us to measure biodiversity when a distance between species and/or conservation values of the species are taken into account, along with the abundance of species. When the phylogenetic distance between species is taken as the weight, the corresponding weighted Gini-Simpson index, unlike Rao's index, is a concave function of the distribution of the relative abundance of the pairs of species, being suitable for use in the additive partitioning of biodiversity. A simple algebraic transformation makes the weighted Gini-Simpson index suitable for use in the multiplicative partitioning of biodiversity as well.

In Methodology, the weighted Gini-Simpson quadratic index is defined both for individual species and for pairs of species. This new measure of biodiversity is used for calculating the average within-site biodiversity (α -diversity), the intersite biodiversity (β -diversity), and the regional or total biodiversity (γ -diversity). It is also shown that the weighted Gini-Simpson quadratic index may be easily modified, by a simple algebraic transformation, to get a measure of biodiversity suitable for use in the multiplicative partitioning of biodiversity as well. In Section 3, a numerical example is presented, which illustrates how the mathematical formalism should be applied from a practical standpoint.

2. Methodology

2.1. The Weighted Measure of Diversity with Respect to Individual Species. Let us assume that there are n species in a certain community/habitat/site and let p_i be the relative abundance of species i (the number of individuals of species i divided by the total number of individuals in that community/habitat/site). We have diversity if species i is present at that location but other species are found there as well. The probability that the species i is present and there are other species present as well is $p_i(1 - p_i)$. If we take all possible values of p_i from the unit interval $[0, 1]$ into account, the wave function $p_i(1 - p_i)$ corresponding to the species i is a nonnegative, symmetric, bell-shaped, concave function, reaching its maximum value $1/4$ at $p_i = 1/2$. If we sum up these wave functions, for all n species, we obtain the classic Gini-Simpson index $GS(\theta)$ corresponding to the given distribution of the relative abundance of the species $\theta = (p_1, \dots, p_n)$. Since 1949, this has been considered to be a very good measure of biodiversity. In order to generalize it, we may assign an amplitude $w_i \geq 0$ to the wave function of the species i , and the resulting new wave function $w_i p_i(1 - p_i)$ continues to be a nonnegative, symmetric, bell-shaped, concave function of p_i , but this time its maximum value is $w_i/4$. Summing up these wave functions for all the species, we get the weighted Gini-Simpson index:

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

which depends both on the distribution of the relative abundance of species θ and on the nonnegative weights $w = (w_1, \dots, w_n)$. The concavity of $GS_w(\theta)$ was proven in [20, 21]. The weight w_i could be anything which contributes to the increase in the diversity induced by the species i . However, the weights may not depend on the relative abundance of species. If $w_i = n$, for each i , then (1) becomes the so-called Rich-Gini-Simpson index $GS_n(\theta)$, introduced in [22], which is essentially dependent on the species richness of the respective community/habitat/site. If there are some conservation values assigned to the species $v = (v_1, \dots, v_n)$, which are positive numbers on a certain scale of values, and the weights are $w_i = nv_i$, the corresponding weighted Gini-Simpson index is denoted by $GS_{n,v}(\theta)$. Obviously, if $w_i = 1$, for each species i , then (1) is the classic Gini-Simpson index $GS(\theta)$. An upper bound for $GS_w(\theta)$, which depends only on the maximum weight and the number of species, is

$$0 \leq GS_w(\theta) \leq \left(\max_i w_i \right) \sum_i p_i(1 - p_i) \leq \left(\max_i w_i \right) \left(1 - \frac{1}{n} \right). \quad (2)$$

Denoting by B_1 the bound from the right-hand side of the inequality (2), the relative weighted Gini-Simpson index for individual species is $0 \leq GS_w(\theta)/B_1 \leq 1$. In Appendix A, another bound of $GS_w(\theta)$ is given, denoted by B_2 , which depends on all the weights assigned to the species. If $w_i = 1$, for each species i , we get $\max_\theta GS(\theta) = 1 - 1/n$, and this maximum biodiversity is obtained when all species have the same relative abundance $p_i = 1/n$. The fact that the

maximum value of $GS(\theta)$ is almost insensitive to the increase of the number of species, tending very slowly to 1 when n increases, allowed Jost [12, 13] and Jost et al. [19] to give some examples showing that the Gini-Simpson index does not behave well when the number of species n is very large. R. C. Guiasu and S. Guiasu [22] showed, however, that the Rich-Gini-Simpson index $GS_n(\theta)$ has no such problem. Indeed, if we take $w_i = n$, for each species i , we get from (2), $\max_\theta GS_n(\theta) = n - 1$, whose value sensibly increases when the number of species increases, which makes inapplicable the criticism of the classic Gini-Simpson index.

2.2. The Additive Partitioning of Biodiversity with Respect to the Individual Species. 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 ($i, j = 1, \dots, n$) and the subscripts k and r refer to sites, ($k, r = 1, \dots, m$). Let $\theta_k = (p_{1,k}, \dots, p_{n,k})$ be the vector whose components are the relative abundances of the individual species at site k , such that $p_{i,k} \geq 0$, ($i = 1, \dots, n$), $\sum_i p_{i,k} = 1$, for each $k = 1, \dots, m$. Let $w = (w_1, \dots, w_n)$ be nonnegative weights assigned to the species. In dealing with species diversity, a good measure of the differentiation, 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. We assign a parameter λ_k to each site k , such that

$$\lambda_k \geq 0, \quad (k = 1, \dots, m), \quad \sum_k \lambda_k = 1. \quad (3)$$

These parameters may be used to make adjustments for differences (in size, altitude, etc.) between the sites, as shown in [23]. If no adjustment is made and we focus only on the species abundance, we take these parameters to be equal, that is, $\lambda_k = 1/m$, for every ($k = 1, \dots, m$). As $GS_w(\theta)$ is a concave function of the distribution of the relative abundance θ , it may be used in the additive partitioning of biodiversity. The corresponding γ -diversity, reflecting the total or regional biodiversity, the α -diversity, interpreted as the within-site diversity or the average diversity of the sites, and the β -diversity, as a measure of between-site diversity, are given by

$$\gamma = GS_w \left(\sum_k \lambda_k \theta_k \right), \quad \alpha = \sum_k \lambda_k GS_w(\theta_k), \quad \beta = \gamma - \alpha. \quad (4)$$

The β -diversity may be interpreted as a measure of dissimilarity or differentiation between the sites of the respective region with respect to the individual species. As shown in Appendix C, taking into account (4), the β -diversity has the expression

$$\beta = \sum_i w_i \sum_{k < r} \lambda_k \lambda_r (p_{i,k} - p_{i,r})^2 < 2 \sum_i w_i \sum_{k < r} \lambda_k \lambda_r. \quad (5)$$

The upper bound in (5) is quite loose. As α is nonnegative and cannot exceed γ , a better upper bound for β is the value B_2 from (A.1), corresponding to n and the given weights w . Thus, the relative between-site diversity, or the dissimilarity between sites is $0 \leq \beta/B_2 \leq 1$, with respect to the individual

species. From (5), we can see that if the species have the same abundance in each site, which means that $p_{i,k} = p_i$ for each site k and each species i , the β -diversity is equal to zero, reflecting the fact that in such a case there is no dissimilarity between the sites.

2.3. The Multiplicative Partitioning of Biodiversity with Respect to Individual Species. Dealing with the multiplicative partitioning of diversity, Whittaker [8] suggested the use of the exponential of the Shannon entropy as a measure of biodiversity. The weighted Gini-Simpson index $GS_w(\theta)$, given by (1), which can be used in the additive partitioning of diversity induced by individual species, may also be transformed into the measure of biodiversity (R. C. Guiasu and S. Guiasu [21]):

$$\frac{1}{[\sum_i w_i p_i - GS_w(\theta)]} = \left(\sum_i w_i p_i^2 \right)^{-1}, \quad (6)$$

which can be used in the multiplicative partitioning of diversity induced by the individual species. This measure of biodiversity may be viewed as being the weighted version of the classic Hill number of first degree from [10]. The corresponding multiplicative γ -diversity and α -diversity are

$$\gamma = \left[\sum_i w_i \left(\sum_k \lambda_k p_{i,k} \right)^2 \right]^{-1}, \quad \alpha = \left[\sum_k \lambda_k \sum_i w_i p_{i,k}^2 \right]^{-1}. \quad (7)$$

Due to the convexity of the function $\sum_i w_i p_i^2$, as a function of the distribution of the relative abundance of species, the γ -diversity cannot be smaller than the α -diversity, as it should be, and, consequently, the multiplicative β -diversity satisfies the inequality: $\beta = \gamma/\alpha \geq 1$. In the additive partitioning of diversity, the γ -diversity, α -diversity, and β -diversity are entities of the same kind and may be expressed in the same units. In the multiplicative partitioning of diversity, the β -diversity is simply a ratio between the total, regional biodiversity γ and the average within-site biodiversity α , a numerical indicator showing to what extent the regional biodiversity, as a whole, exceeds the average biodiversities of the sites of the respective region. Obviously, if the sites have the same species and the same abundance of these species, which means that $p_{i,k} = p_i$ for each species i , then $\beta = 1$.

2.4. The Weighted Measure of Diversity with Respect to the Pairs of Species. Let $D = [d_{ij}]$ be an $n \times n$ matrix whose entries are the distances between the pairs of n species, such that $d_{ij} \geq 0, d_{ii} = 0, (i, j = 1, \dots, n)$. This could be the matrix of the phylogenetic distance between species, for instance. When Rao introduced his quadratic index, improperly called quadratic entropy [14]:

$$R_D = \sum_{i,j} d_{ij} p_i p_j, \quad (8)$$

he had to focus on the pairs of species instead of the individual species, using the distance between the pairs

of distinct species, along with their relative abundance, in order to measure the dissimilarity between species. Rao's indicator is very simple and may be easily interpreted as the average dissimilarity between two individuals belonging to two different species when the phylogenetic distance is taken into account. There have been numerous attempts, such as [24], for instance, at using Rao's index R_D in the additive partitioning of diversity. Unfortunately, R_D is not a concave function of the distribution of the relative abundance of species $\theta = (p_1, \dots, p_n)$ for an arbitrary distance matrix D . Thus, R_D can be applied to the additive partitioning of biodiversity only for some special kinds of such matrices D (as mentioned in [24], for instance), but not in general. On the other hand, there is no generally accepted proposal of a new kind of nonstandard α -divergence which could be defined for such a measure which is not a concave function of the distribution of the relative abundance of species. This paper shows that the generalization of the weighted Gini-Simpson index to the pairs of species provides a concave measure of diversity which could indeed be used both for the additive partitioning and the multiplicative partitioning of diversity when the phylogenetic distance between species is taken into account. Therefore, this provides a suitable replacement of Rao's index in the partitioning of biodiversity.

Let $\Theta = [\pi_{ij}]$ be an $n \times n$ matrix where π_{ij} is a joint probability of the pair of species (i, j) , in this order. As π_{ji} is the probability of the pair (j, i) , in this order, the probability of the subset of species $\{i, j\}$ is $\pi_{ij} + \pi_{ji}$. We have $\pi_{ij} \geq 0, (i, j = 1, \dots, n); \sum_{i,j} \pi_{ij} = 1$. Let $W = [w_{ij}]$ be an $n \times n$ matrix whose entries are arbitrary nonnegative weights assigned to the pairs of species. However, these weights may not depend on the joint distribution $[\pi_{ij}]$. The weighted Gini-Simpson quadratic index of the pairs of species is

$$GS_W(\Theta) = \sum_{i,j} w_{ij} \pi_{ij} (1 - \pi_{ij}), \quad (9)$$

If $w_{ij} = 1$, for all pairs of species, then (9) becomes the generalization of the classic Gini-Simpson index to the pairs of species and is denoted by $GS(\Theta)$. As a function of the joint distribution Θ , the weighted Gini-Simpson index $GS_W(\Theta)$ is nonnegative and concave, as shown in [20, 21]. An upper bound for $GS_W(\Theta)$ which depends only on the maximum weight and the number of species is

$$0 \leq GS_W(\Theta) \leq \left(\max_{i,j} w_{ij} \right) \sum_{i,j} \pi_{ij} (1 - \pi_{ij}) \leq \left(\max_{i,j} w_{ij} \right) \left(1 - \frac{1}{n^2} \right). \quad (10)$$

Denoting by B_3 the bound from the right-hand side of the inequality (10), the relative weighted Gini-Simpson index for pairs of species is $0 \leq GS_W(\Theta)/B_3 \leq 1$. In Appendix B, another bound for $GS_W(\Theta)$, denoted by B_5 , is given, which depends on all the weights assigned to the species. If $w_{ij} = 1$, for each pair of species (i, j) , we get $\max_{\Theta} GS(\Theta) = 1 - 1/n^2$, and this maximum biodiversity is obtained when all species have the same relative abundance $\pi_{ij} = 1/n^2$. The special cases of interest are the following ones.

- (a) If the species are independent, which means $\pi_{ij} = p_i p_j$ and the weights are $w_{ij} = d_{ij}$, then the weighted Gini-Simpson index (9) is denoted by

$$GS_D(\Theta) = \sum_{i,j} d_{ij} p_i p_j (1 - p_i p_j), \quad (11)$$

which generalizes Rao's index R_D given by (8).

- (b) If there are the positive numbers $v = (v_1, \dots, v_n)$, representing conservation values of the individual species, the species are independent, which means $\pi_{ij} = p_i p_j$, and the weights are

$$w_{ij} = \frac{n(n-1)}{2} \frac{1}{2} (v_i + v_j) d_{ij}, \quad (12)$$

where $n(n-1)/2$ is the number of distinct pairs of species (i, j) , such that $i < j$, or the number of the pairs of species $\{i, j\}$, and $(1/2)(v_i + v_j)$ is the average value of the pair of species (i, j) , then the corresponding weighted Gini-Simpson index (9) is denoted by:

$$GS_{n,v,D}(\Theta) = \frac{n(n-1)}{2} \sum_{i < j} (v_i + v_j) d_{ij} p_i p_j (1 - p_i p_j), \quad (13)$$

which takes into account all the information available, namely, the species richness n , the relative abundance θ of species, the matrix D of the distance between species, and the conservation values v of the species. As the measure given in (13) is a nonnegative concave function of the distribution of the relative abundance of the pairs of species Θ , for an arbitrary distance matrix D , and also depends explicitly on the species richness, the distance between species, and the conservation values of the species, all these are sufficient reasons to suggest that it could more than adequately replace the use of Rao's index (8).

2.5. The Additive Partitioning of Biodiversity with Respect to the Pairs of Species. Let us assume that in a certain region there are n species and m sites. Again, in what follows, the subscripts i and j refer to species ($i, j = 1, \dots, n$) and the subscripts k and r refer to sites ($k, r = 1, \dots, m$). Let $\Theta_k = [\pi_{ij,k}]$ be an arbitrary joint probability distribution of the pairs of species within site k , where $\pi_{ij,k}$ is the probability of the pair of species (i, j) , in this order, within site k , such that $\pi_{ij,k} \geq 0$, $\sum_{i,j} \pi_{ij,k} = 1$. Let $W = [w_{ij}]$ be the matrix whose entries are nonnegative weights assigned to the pairs of species. We assign a parameter λ_k to each site k , satisfying (3). As $GS_W(\Theta)$, given by (9), is a concave function of the joint distribution Θ assigned to the pairs of species, it may be used in the additive partitioning of biodiversity. The corresponding γ -diversity, reflecting the total or regional biodiversity, the α -diversity, interpreted as the within-site

diversity or the average diversity of the sites, and the β -diversity, as a measure of between-site diversity, with respect to the pairs of species, are given by

$$\gamma = GS_W \left(\sum_k \lambda_k \Theta_k \right), \quad \alpha = \sum_k \lambda_k GS_W(\Theta_k), \quad \beta = \gamma - \alpha. \quad (14)$$

The β -diversity may be interpreted as a measure of dissimilarity or differentiation between the sites of the respective region with respect to the pairs of species. As shown in [20, 21], taking into account (14), the β -diversity has the following expression:

$$\beta = \sum_{i,j} w_{ij} \sum_{k < r} \lambda_k \lambda_r (\pi_{ij,k} - \pi_{ij,r})^2 < \sum_{i,j} w_{ij} \sum_{k < r} \lambda_k \lambda_r. \quad (15)$$

The upper bound in (15) is quite loose. As α is nonnegative and cannot exceed γ , a better upper bound for β is the value B_5 from (B.4) of Appendix B, corresponding to n and the given weights W . Thus, the relative between-site diversity, or the relative dissimilarity between sites, is $0 \leq \beta/B_5 \leq 1$, with respect to the pairs of species.

Let $\theta_k = (p_{1,k}, \dots, p_{n,k})$ be the vector whose components are the relative abundances of the individual species at site k . If the species are independent, $\pi_{ij,k} = p_{i,k} p_{j,k}$. Let also $D = [d_{ij}]$ be the matrix of the distances between species. Then, for the weights (12), the corresponding β -diversity (15) is

$$\beta = \frac{n(n-1)}{2} \sum_{i < j} (v_i + v_j) d_{ij} \sum_{k < r} \lambda_k \lambda_r (p_{i,k} p_{j,k} - p_{i,r} p_{j,r})^2, \quad (16)$$

measuring the between-site diversity with respect to the distinct pairs of species when the species richness, the distance between species, and the conservation values are all taken into account along with the relative abundance of the species. If the species have the same abundance in each site, which means that $p_{i,k} = p_i$, for each site k and each species i , the corresponding β -diversity is equal to zero, reflecting the fact that in such a case there is no dissimilarity between sites. If we divide (16) by the upper bound B_5 from the inequality (B.4) of Appendix B, taking the weights (12), we obtain the relative β -diversity, which has the advantage of always being a number between 0 and 1.

At the same time, let us notice that Rao's index R_D given by (8) is a linear function of the joint distribution of the relative abundance of the pairs of species $\Theta = [p_i p_j]$, and, consequently, the corresponding β -diversity induced by the pairs of species is equal to zero, for an arbitrary distance matrix D and an arbitrary distribution of the relative abundance of individual species. Therefore, Rao's index is not suitable for use in the standard additive partitioning of diversity induced by pairs of species when an arbitrary dissimilarity distance between species is taken into account. Attempts were made [24, 25] to either find particular cases of distance matrices or define new kinds of nonstandard α -divergence for which the partitioning of diversity can still be performed.

2.6. The Multiplicative Partitioning of Biodiversity with Respect to the Pairs of Species. The weighted Gini-Simpson quadratic index GS_w given by (9), which can be used in the additive partitioning of diversity induced by pairs of species, may be transformed into the measure of diversity [21]:

$$\frac{1}{\left[\sum_{i,j} w_{ij} \pi_{ij} - GS_w(\Theta)\right]} = \left(\sum_{i,j} w_{ij} \pi_{ij}^2\right)^{-1}, \quad (17)$$

which can be used in the multiplicative partitioning of diversity induced by the pairs of species. This measure of biodiversity may be viewed as being the weighted version for pairs of species of the classic Hill number of first degree from [10]. Using the notations from the previous Section 2.5, the corresponding multiplicative γ -diversity, α -diversity, and β -diversity are

$$\gamma = \left[\sum_{i,j} w_{ij} \left(\sum_k \lambda_k \pi_{ij,k} \right)^2 \right]^{-1}, \quad \alpha = \left[\sum_k \lambda_k \sum_{i,j} w_{ij} \pi_{ij,k}^2 \right]^{-1},$$

$$\beta = \frac{\gamma}{\alpha}. \quad (18)$$

Due to the convexity of the function $\sum_{i,j} w_{ij} \pi_{ij}^2$, as a function of the joint distribution $\Theta = [\pi_{ij}]$, the γ -diversity cannot be smaller than the α -diversity, as it should be, and, consequently, the multiplicative β -diversity satisfies the inequality: $\beta = \gamma/\alpha \geq 1$. In the additive partitioning of diversity, the γ -diversity, α -diversity, and β -diversity are entities of the same kind and may be expressed in the same units. In the multiplicative partitioning of diversity, the β -diversity is simply a ratio between the total, regional biodiversity γ and the average within-site biodiversity α , a numerical indicator showing to what extent the regional biodiversity, as a whole, exceeds the average biodiversity of the sites of the respective region.

Let $\theta_k = (p_{1,k}, \dots, p_{n,k})$ be the vector whose components are the relative abundances of the individual species at site k . If the species are independent, $\pi_{ij,k} = p_{i,k} p_{j,k}$. Let also $D = [d_{ij}]$ be the matrix of the distances between species. Then, for the weights (12), the corresponding β -diversity from (18) is

$$\beta = \frac{\left[\sum_k \lambda_k \sum_{i < j} (v_i + v_j) d_{ij} p_{i,k}^2 p_{j,k}^2 \right]}{\left[\sum_{i < j} (v_i + v_j) d_{ij} \left(\sum_k \lambda_k p_{i,k} p_{j,k} \right)^2 \right]} \geq 1, \quad (19)$$

measuring the ratio between the regional biodiversity and the average biodiversity of the sites with respect to the pairs of species. Obviously, if the sites have the same species and the same abundance of these species, which means that $p_{i,k} = p_i$, for each species i , then $\beta = 1$.

2.7. The Weighted Shannon Entropy. The weighted Shannon entropy was introduced in [26]. If we have n species such that the distribution of the relative abundance of these species is $\theta = (p_1, \dots, p_n)$ and the nonnegative weights assigned to the species are $w = (w_1, \dots, w_n)$, then the weighted entropy is

the nonnegative, concave function $H_w(\theta) = -\sum_i w_i p_i \ln p_i$. Similarly, if $W = [w_{ij}]$ is a matrix of nonnegative weights and $\Theta = [\pi_{ij}]$ a joint probability distribution assigned to the pairs of species, the joint weighted entropy is $H_W(\Theta) = -\sum_{i,j} w_{ij} \pi_{ij} \ln \pi_{ij}$. It is possible, in principle, to remake the analysis from Sections 2.1–2.6 using the weighted Shannon entropies $H_w(\theta)$ and $H_W(\Theta)$ instead of the weighted Gini-Simpson indices $GS_w(\theta)$ and $GS_W(\Theta)$. However, the Shannon entropy is actually a measure of uncertainty and we cannot justify its use as a measure of diversity, as we did for the Gini-Simpson index at the beginning of Section 2.1. Also, since the Shannon entropy is a logarithmic function, it is much more difficult to obtain simple analytical formulas for its maximum values subject to given constraints. The weighted Gini-Simpson index is a simpler and more effective tool in measuring biodiversity.

3. Discussion

It seems to be much easier to discuss the significance of the concepts introduced in Section 2 by showing a representative numerical example. Let us assume that in a certain region there are three sites ($m = 3$) and three species ($n = 3$). If A_{ik} denotes the absolute abundance (number of individuals) of species i within site k , let us assume that

$$\begin{aligned} A_{11} &= 2, & A_{21} &= 24, & A_{31} &= 14; \\ A_{12} &= 32, & A_{22} &= 4, & A_{32} &= 14; \\ A_{13} &= 24, & A_{23} &= 36, & A_{33} &= 20. \end{aligned} \quad (20)$$

The corresponding relative abundance is

$$\begin{aligned} p_{1,1} &= 0.05, & p_{2,1} &= 0.60, & p_{3,1} &= 0.35; \\ p_{1,2} &= 0.64, & p_{2,2} &= 0.08, & p_{3,2} &= 0.28; \\ p_{1,3} &= 0.30, & p_{2,3} &= 0.45, & p_{3,3} &= 0.25. \end{aligned} \quad (21)$$

Thus, in this example, $\theta_1 = (0.05, 0.60, 0.35)$, $\theta_2 = (0.64, 0.08, 0.28)$, and $\theta_3 = (0.30, 0.45, 0.25)$.

3.1. Biodiversity with Respect to the Individual Species. Using the Rich-Gini-Simpson index $GS_n(\theta)$, given by (1) with the weights $w_i = n$, to calculate the amount of diversity with respect to the individual species, in each site, we obtain $GS_3(\theta_1) = 1.5450$, $GS_3(\theta_2) = 1.5168$, and $GS_3(\theta_3) = 1.9350$. The maximum biodiversity in this case would be $n - 1 = 2$. We can see that the first two sites have almost the same biodiversity, both a little smaller than the biodiversity of the third site which is close to the maximum value, when only the richness and the abundance of species are taken into account.

Let us assume now that the three species have the following conservation values: $v_1 = 6$, $v_2 = 3$, and $v_3 = 3$. These conservation values $v = (6, 3, 3)$ contribute to the diversity of the three sites. Taking the weights $w_i = nv_i$, we have $w_1 = 18$, $w_2 = 9$, $w_3 = 9$. Therefore, $w = (18, 9, 9)$. Using the weighted Gini-Simpson index $GS_w(\theta)$ given by (1), we obtain the following values of the biodiversity of each site: $GS_w(\theta_1) = 5.0625$, $GS_w(\theta_2) = 6.624$, and $GS_w(\theta_3) = 7.695$. When the species have these conservation values, the

biodiversity of the second and third sites are closer and higher than the biodiversity of the first site. But in order to have a better understanding of these numbers, we have to compare them with the bounds B_1 and B_2 from the inequalities (2) and (A.1), respectively. For the weights $w = (18, 9, 9)$, the loose upper bound B_1 for GS_w , which takes into account only the number of species $n = 3$ and the maximum weight $\max_i w_i = 18$, has the value 12. For the much better upper bound B_2 for GS_w from (A.1), mentioned in Appendix A, which takes into account the number of species $n = 3$ and all the weights $w = (18, 9, 9)$, we get the value 8.1. Therefore, we can see that the bound B_2 is obviously better than B_1 . With respect to B_2 , the second and third sites have 81.78% and 95% of the maximum biodiversity for the given weights, whereas the first site has only 62.5%. If we do not discriminate among sites with respect to size, altitude, or any other factor, then the parameters assigned to the three sites are $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. In such a case, we have

$$\begin{aligned} \sum_k \lambda_k \theta_k &= \frac{1}{3}(0.05, 0.60, 0.35) + \frac{1}{3}(0.64, 0.08, 0.28) \\ &+ \frac{1}{3}(0.30, 0.45, 0.25) \\ &= (0.3300, 0.3767, 0.2933) \\ &= (q_1, q_2, q_3). \end{aligned} \quad (22)$$

According to (4), the γ -diversity and α -diversity, with respect to the single species, are

$$\begin{aligned} \gamma &= GS_w \left(\sum_k \lambda_k \theta_k \right) = \sum_i w_i q_i (1 - q_i) = 7.9584, \\ \alpha &= \sum_k \lambda_k GS_w(\theta_k) = \frac{1}{3}(5.0625 + 6.624 + 7.695) = 6.4605. \end{aligned} \quad (23)$$

Thus, in the additive partitioning of diversity, the β -diversity is $\beta = \gamma - \alpha = 1.4979$. For the weights $w = (18, 9, 9)$ and $n = 3$, according to the formula (A.1) from Appendix A, the maximum value of GS_w is $B_2 = 8.1$. Therefore, the biodiversity γ of the entire region is 98.25% of the maximum and the average within-site biodiversity α is 79.76%. The value of the between-site diversity β shows the average differentiation between sites corresponding to a difference of 18.49% between the values of γ and α . We note that for identical sites, the value of β would be equal to zero, as could be seen from (5). The advantage of the use of the additive partitioning of biodiversity is that the values of α , β , and γ are expressed on the same scale of values.

Doing the multiplicative partitioning of biodiversity for $\lambda_i = 1/3$, ($i = 1, 2, 3$), and $w = nv = (18, 9, 9)$, from (7) we get $\gamma = 0.2493$ and $\alpha = 0.1815$. Consequently, $\beta = \gamma/\alpha = 1.3736$.

3.2. Biodiversity with Respect to the Pairs of Species. Let us assume that we have the matrix of the phylogenetic distances between the three species $D = [d_{ij}]$, where $d_{12} = 3$, $d_{13} = 2$, and $d_{23} = 2$. If we assume that within each site the species are supposed to be independent from the point of view of

their relative abundance, then the relative abundance of the pair of species (i, j) , in this order, is the product of the relative abundance of the corresponding individual species, namely, $p_{i,k}p_{j,k}$, within every site k . Therefore, the matrices $\Theta_k = [p_{i,k}p_{j,k}]$ are:

$$\begin{aligned} \Theta_1 &= \begin{bmatrix} 0.0025 & 0.0300 & 0.0175 \\ 0.0300 & 0.3600 & 0.2100 \\ 0.0175 & 0.2100 & 0.1225 \end{bmatrix}, \\ \Theta_2 &= \begin{bmatrix} 0.4096 & 0.0512 & 0.1792 \\ 0.0512 & 0.0064 & 0.0224 \\ 0.1792 & 0.0224 & 0.0784 \end{bmatrix}, \\ \Theta_3 &= \begin{bmatrix} 0.0900 & 0.1350 & 0.0750 \\ 0.1350 & 0.2025 & 0.1125 \\ 0.0750 & 0.1125 & 0.0625 \end{bmatrix}. \end{aligned} \quad (24)$$

If we do not discriminate among sites with respect to size, altitude, or any other factor, then the parameters assigned to the three sites are $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. In such a case, we have

$$\lambda_1 \Theta_1 + \lambda_2 \Theta_2 + \lambda_3 \Theta_3 = \begin{bmatrix} 0.1674 & 0.0721 & 0.0906 \\ 0.0721 & 0.1896 & 0.1150 \\ 0.0906 & 0.1150 & 0.0878 \end{bmatrix}. \quad (25)$$

Let us use Rao's index (8) for doing the additive partitioning of diversity with respect to the pairs of species. Successively, we obtain $R_D(\Theta_1) = 1.0900$, $R_D(\Theta_2) = 1.1136$, and $R_D(\Theta_3) = 1.5600$. The corresponding α -diversity is $\alpha = \lambda_1 R_D(\Theta_1) + \lambda_2 R_D(\Theta_2) + \lambda_3 R_D(\Theta_3) = 1.255$, and the γ -diversity is $\gamma = R_D(\lambda_1 \Theta_1 + \lambda_2 \Theta_2 + \lambda_3 \Theta_3) = 1.255$. Consequently, the β -diversity is $\beta = \gamma - \alpha = 0$, which is not surprising because Rao's index is a linear function of the joint distribution of the pairs of species.

If we use the weighted Gini-Simpson index (11) with the weights $w_{ij} = d_{ij}$, we obtain

$$\begin{aligned} GS_D(\Theta_1) &= 0.9070, & GS_D(\Theta_2) &= 0.9674, \\ GS_D(\Theta_3) &= 1.3775, \end{aligned} \quad (26)$$

and the corresponding α -diversity is

$$\alpha = \lambda_1 GS_D(\Theta_1) + \lambda_2 GS_D(\Theta_2) + \lambda_3 GS_D(\Theta_3) = 1.0840, \quad (27)$$

the γ -diversity is $\gamma = GS_D(\lambda_1 \Theta_1 + \lambda_2 \Theta_2 + \lambda_3 \Theta_3) = 1.1381$, and the β -diversity is $\beta = \gamma - \alpha = 0.0541$. Calculating the upper bound B_5 of GS_w given in the inequality (B.4) from Appendix B, for the weights $w_{ij} = d_{ij}$, which means $W = D$, we obtain $\max GS_D = 2.75$. Compared to this maximum value, $GS_D(\Theta_1)$ represents 32.98%; $GS_D(\Theta_2) = 35.18\%$; $GS_D(\Theta_3) = 50.09\%$; $\gamma = 41.39\%$; $\alpha = 39.42\%$; $\beta = 1.97\%$.

We take now into account the number of species $n = 3$, the parameters assigned to the sites $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, the phylogenetic distances between species $d_{12} = 3$, $d_{13} = 2$, $d_{23} = 2$, and the conservation values of the species $v_1 = 6$, $v_2 = 3$, $v_3 = 3$. The computation of the weighted Gini-Simpson index given by (13), with the weights $w_{ij} = (n(n-1)/2)(1/2)(v_i + v_j)d_{ij}$, gives

$$\begin{aligned} GS_{n,v,D}(\Theta_1) &= 9.2580, & GS_{n,v,D}(\Theta_2) &= 12.6659, \\ GS_{n,v,D}(\Theta_3) &= 16.7994, \end{aligned} \quad (28)$$

and the corresponding α -diversity is

$$\alpha = \lambda_1 GS_{n,v,D}(\Theta_1) + \lambda_2 GS_{n,v,D}(\Theta_2) + \lambda_3 GS_{n,v,D}(\Theta_3) = 12.9078, \quad (29)$$

while the γ -diversity is $\gamma = GS_{n,v,D}(\lambda_1 \Theta_1 + \lambda_2 \Theta_2 + \lambda_3 \Theta_3) = 13.5321$, which gives the β -diversity: $\beta = \gamma - \alpha = 0.6243$. Calculating the upper bound B_5 of $GS_{n,v,D}$ given in the inequality (B.4) from Appendix B, for the weights $w_{ij} = (n(n-1)/2)(1/2)(v_i + v_j)d_{ij}$, we obtain $\max GS_{n,v,D} = 34.2237$. Compared to this maximum value, $GS_{n,v,D}(\Theta_1)$ represents 27.05%; $GS_{n,v,D}(\Theta_2) = 37.01\%$; $GS_{n,v,D}(\Theta_3) = 49.09\%$; $\gamma = 39.54\%$; $\alpha = 37.72\%$; $\beta = 1.82\%$.

Doing the multiplicative partitioning of biodiversity for $\lambda_i = 1/3$, ($i = 1, 2, 3$), and $w_{ij} = d_{ij}$, from (18) and (19), we get $\gamma = 8.56$ and $\alpha = 5.86$. Consequently, $\beta = \gamma/\alpha = 1.46$. Doing the multiplicative partitioning of biodiversity for the site parameters $\lambda_i = 1/3$, ($i = 1, 2, 3$), and the weights $w_{ij} = (n(n-1)/2)(1/2)(v_i + v_j)d_{ij}$, from (18), we get $\gamma = 0.75$ and $\alpha = 0.51$. Consequently, $\beta = \gamma/\alpha = 1.47$.

4. Conclusion

Using a measure of biodiversity, as a mathematical tool, the distribution of biodiversity at multiple sites of a region has been traditionally investigated through the 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. According to Whittaker [8], 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 [5], MacArthur and Wilson [7], and Lande [17], β -diversity is the difference between γ -diversity and α -diversity. This is the additive partitioning of diversity. All these diversities, namely, α -diversity, β -diversity, and γ -diversity, should be nonnegative numbers. In general, a measure of biodiversity ought to be nonnegative, in which case the corresponding α -diversity and γ -diversity, calculated by using such a measure, are nonnegative as well, as they should be. But the corresponding β -diversity is also nonnegative, in the additive partitioning of the biodiversity, or larger than 1, in the multiplicative partitioning of biodiversity, if the measure of biodiversity used is a concave function of the distribution of the relative abundance of species.

The best known measures of biodiversity are Shannon's entropy and the Gini-Simpson index. Both of them measure the biodiversity taking into account only the relative abundance of species. The widely used Rao's index measures the dissimilarity between species taking into account not only the relative abundance of species but also a distance between species, such as the phylogenetic distance, for instance. Both Shannon's entropy and the classic Gini-Simpson index satisfy the mathematical properties (nonnegativity and concavity) that allow them to be successfully used in the additive partitioning of biodiversity. Unfortunately, as was pointed out recently [12, 13], these two measures do not give good results when the number of species is very large. On the other hand, Rao's index of dissimilarity is not a concave function

of the relative abundance of species for arbitrary distances between species and, consequently, can be used in the additive partitioning of biodiversity only for some particular distance matrices, but not in general. The main objective of this paper is to show that the weighted Gini-Simpson quadratic index GS_D given by (11), which is a generalization of the classic Gini-Simpson index GS to the pairs of species, is a suitable measure for use in the standard additive partitioning of biodiversity because, unlike the commonly used Rao's index of dissimilarity R , it is a concave function of the relative abundance of the pairs of species. Unlike the classic Gini-Simpson index GS , the weighted Gini-Simpson quadratic index $GS_{n,D}$ behaves very well when the number of species is very large. The index $GS_{n,D}$ may be generalized to get the diversity measure $GS_{n,v,D}$, given by (13), which takes into account not only the number of species, the relative abundance of the pairs of species, and the matrix D of the distances between species, but also a vector v of values assigned to the individual species, such as some conservation values for instance. The algebraic transformations (6) and (17) of the weighted Gini-Simpson quadratic indices GS_w , for single species, and GS_W , for pairs of species, given by (1) and (9), respectively, provide two measures of biodiversity which are suitable for use in the multiplicative partitioning of biodiversity. A detailed numerical example shows how the formulas should be implemented in applications.

From a practical point of view, the new weighted Gini-Simpson measure of biodiversity $GS_{n,v,D}$, which is a positive concave function of the relative abundance of the pairs of species, which essentially depends both on the matrix D of the distances between species and on the conservation values v of the species, is proposed as a suitable and improved replacement for the well-known Rao's index in the partitioning of biodiversity.

Appendices

A. An Upper Bound for $GS_w(\theta)$ for Individual Species

The weighted Gini-Simpson index $GS_w(\theta)$ is a nonnegative, concave, quadratic function of the distribution of the relative abundance of species $\theta = (p_1, \dots, p_n)$. We can apply the standard Lagrange multipliers technique from multivariate calculus in order to maximize $GS_w(\theta)$ subject to the constraint $\sum_i p_i = 1$. When the positive weights $w = (w_1, \dots, w_n)$ are given, the maximum value of the weighted Gini-Simpson index $GS_w(\theta)$, as a function of the weights, is

$$\max_{\theta} GS_w(\theta) \leq \frac{1}{4} \left[\sum_i w_i - (n-2)^2 \left(\sum_i w_i^{-1} \right)^{-1} \right]. \quad (A.1)$$

If the bound from the right-hand side of the inequality (A.1) is denoted by B_2 , the relative weighted biodiversity is $0 \leq GS_w(\theta)/B_2 \leq 1$.

B. An Upper Bound for $GS_W(\Theta)$ for the Pairs of Species

The weighted Gini-Simpson index $GS_W(\Theta)$ is a nonnegative, concave, quadratic function of the joint distribution assigned to the pairs of species $\Theta = [\pi_{ij}]$. We can apply the standard Lagrange multipliers technique from multivariate calculus in order to maximize $GS_W(\Theta)$ subject to the constraint $\sum_{i,j} \pi_{ij} = 1$. When the positive weights $W = [w_{ij}]$ are given, the maximum value of the weighted Gini-Simpson index $GS_W(\Theta)$, as a function of the weights, subject to the constraint $\sum_{i,j} \pi_{ij} = 1$, is

$$\max_{\Theta} GS_W(\Theta) \leq \frac{1}{4} \left[\sum_{i,j} w_{ij} - (n^2 - 2)^2 \left(\sum_{i,j} w_{ij}^{-1} \right)^{-1} \right]. \quad (B.1)$$

If the bound from the right-hand side of the inequality (B.1) is denoted by B_4 , the relative weighted biodiversity is $0 \leq GS_W(\Theta)/B_4 \leq 1$.

Let us note that if $\pi_{ij} = \pi_{ji}$, for the distinct pairs (i, j) , and $w_{ij} = w_{ji}$, $w_{ii} = 0$, which happens, for instance, in the important case when $w_{ij} = d_{ij}$, or when

$$w_{ij} = \left[\frac{n(n-1)}{2} \right] \left[\frac{(v_i + v_j)}{2} \right] d_{ij}, \quad (B.2)$$

where $v_i > 0$ is the conservation value of species i and d_{ij} is the distance between the distinct species (i, j) , then $GS_W(\Theta)$ may be written as

$$GS_W(\Theta) = 2 \sum_{i < j} w_{ij} \pi_{ij} (1 - \pi_{ij}). \quad (B.3)$$

Maximizing $GS_W(\Theta)$, which in this case depends only on $n(n-1)/2$ variables π_{ij} , ($i < j$), subject to the constraint: $2 \sum_{i < j} \pi_{ij} = c$, where $0 < c = 1 - \sum_i \pi_{ii} \leq 1$, we obtain

$$\max_{\Theta} GS_W(\Theta) \leq \frac{1}{2} \left[\sum_{i < j} w_{ij} - \left(\frac{n(n-1)}{2} - 1 \right)^2 \left(\sum_{i < j} w_{ij}^{-1} \right)^{-1} \right]. \quad (B.4)$$

If the bound from the right-hand side of the inequality (B.4) is denoted by B_5 , the relative weighted biodiversity is

$$0 \leq \frac{GS_W(\Theta)}{B_5} \leq 1. \quad (B.5)$$

C. Concavity of the Weighted Gini-Simpson Index GS_W for Individual Species

Using the notation from Section 2.2 and taking into account that

$$\begin{aligned} & -\lambda_k^2 p_{i,k}^2 + \lambda_k p_{i,k}^2 \\ & = \lambda_k (1 - \lambda_k) p_{i,k}^2 \\ & = \lambda_k (\lambda_1 + \dots + \lambda_{k-1} + \lambda_{k+1} + \dots + \lambda_m) p_{i,k}^2 \\ & = (\lambda_1 \lambda_k + \dots + \lambda_{k-1} \lambda_k + \lambda_k \lambda_{k+1} + \dots + \lambda_k \lambda_m) p_{i,k}^2, \\ & \text{for every } 1 \leq k \leq m, \end{aligned} \quad (C.1)$$

we get

$$\begin{aligned} \beta & = GS_W \left(\sum_k \lambda_k \theta_k \right) - \sum_k \lambda_k GS_W(\theta_k) \\ & = \sum_i w_i \left(\sum_k \lambda_k p_{i,k} \right) \left(1 - \sum_k \lambda_k p_{i,k} \right) \\ & \quad - \sum_k \lambda_k \sum_i w_i p_{i,k} (1 - p_{i,k}) \\ & = \sum_i w_i \left(\sum_k \lambda_k (1 - \lambda_k) p_{i,k}^2 - \sum_{k \neq r} \lambda_k \lambda_r p_{i,k} p_{i,r} \right) \\ & = \sum_i w_i \left[\sum_{k < r} \lambda_k \lambda_r (p_{i,k}^2 - 2 p_{i,k} p_{i,r} + p_{i,r}^2) \right] \\ & = \sum_i w_i \sum_{k < r} \lambda_k \lambda_r (p_{i,k} - p_{i,r})^2 \geq 0. \end{aligned} \quad (C.2)$$

Remark 1. In the paper [21], two corrections are needed: (a) on page 799, in the first column, the rows 12-13, the numerical values should be $\alpha = 0.723$, $\gamma = 0.745$, $\beta = 0.022$; (b) on page 798, (11) should be: $\gamma = \sum_{i,j} d_{ij} (\sum_k \lambda_k p_{i,k} p_{j,k}) (1 - \sum_k \lambda_k p_{i,k} p_{j,k})$.

Acknowledgments

The authors would like to thank the editor, Professor Jean-Guy Godin, and the two referees for their detailed and very helpful comments.

References

- [1] C. Gini, "Variabilità e mutabilità," in *Memorie di Metodologica Statistica*, E. Pizetti and T. Salvemini, Eds., Liberia Eredi Virgilio Veschi, Roma, Italy, 1912.
- [2] E. H. Simpson, "Measurement of diversity," *Nature*, vol. 163, p. 688, 1949.
- [3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, p. 379-423; 623-656, 1948.
- [4] L. Boltzmann, *Vorlesungen über Gastheorie*, J. A. Barth Akademische Druck-u, Leipzig, Germany, 1896.
- [5] R. H. MacArthur, "Patterns of species diversity," *Biological Review*, vol. 40, pp. 510-533, 1965.
- [6] E. C. Pielou, "Shannon's formula as a measure of specific diversity: its use and misuse," *American Naturalist*, vol. 100, pp. 463-465, 1966.
- [7] R. H. MacArthur and E. O. Wilson, *The Theory of Island Biogeography*, Princeton University Press, Princeton, NJ, USA, 1967.
- [8] R. H. Whittaker, "Evolution and measurement of species diversity," *Taxon*, vol. 21, pp. 213-251, 1972.
- [9] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed., pp. 547-561, University of California, Berkeley, Calif, USA, 1961.
- [10] M. Hill, "Diversity and evenness. A unifying notation and its consequences," *Ecology*, vol. 45, pp. 427-432, 1973.
- [11] L. Jost, "Entropy and diversity," *Oikos*, vol. 113, no. 2, pp. 363-375, 2006.

- [12] L. Jost, "Partitioning diversity into independent alpha and beta components," *Ecology*, vol. 88, no. 10, pp. 2427–2439, 2007.
- [13] L. Jost, "Mismeasuring biological diversity: response to Hoffmann and Hoffmann (2008)," *Ecological Economics*, vol. 68, pp. 925–928, 2009.
- [14] C. R. Rao, "Diversity and dissimilarity coefficients: a unified approach," *Theoretical Population Biology*, vol. 21, no. 1, pp. 24–43, 1982.
- [15] C. Ricotta, "Through the jungle of biological diversity," *Acta Biotheoretica*, vol. 53, no. 1, pp. 29–38, 2005.
- [16] S. Hoffmann and A. Hoffmann, "Is there a "true" diversity?" *Ecological Economics*, vol. 65, no. 2, pp. 213–215, 2008.
- [17] R. Lande, "Statistics and partitioning of species diversity, and similarity among multiple communities," *Oikos*, vol. 76, no. 1, pp. 5–13, 1996.
- [18] J. A. Veech, K. S. Summerville, T. O. Crist, and J. C. Gering, "The additive partitioning of species diversity: recent revival of an old idea," *Oikos*, vol. 99, no. 1, pp. 3–9, 2002.
- [19] L. Jost, P. Devries, T. Walla, H. Greeney, A. Chao, and C. Ricotta, "Partitioning diversity for conservation analyses," *Diversity and Distributions*, vol. 16, no. 1, pp. 65–76, 2010.
- [20] R. C. Guiasu and S. Guiasu, "New measures for comparing the species diversity found in two or more habitats," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 18, no. 6, pp. 691–720, 2010.
- [21] R. C. Guiasu and S. Guiasu, "The weighted quadratic index of biodiversity for pairs of species: a generalization of Rao's index," *Natural Science*, vol. 3, pp. 795–801, 2011.
- [22] R. C. Guiasu and S. Guiasu, "The Rich-Gini-Simpson quadratic index of biodiversity," *Natural Science*, vol. 2, pp. 1130–1137, 2010.
- [23] R. C. Guiasu and S. Guiasu, "Diversity measures and coarse-graining in data analysis with an application involving plant species on the Galápagos Islands," *Journal of Systemics, Cybernetics and Informatics*, vol. 8, no. 5, pp. 54–64, 2010.
- [24] C. Ricotta and L. Szeidl, "Diversity partitioning of Rao's quadratic entropy," *Theoretical Population Biology*, vol. 76, no. 4, pp. 299–302, 2009.
- [25] A. Chao, C. H. Chiu, and L. Jost, "Phylogenetic diversity measures based on Hill numbers," *Philosophical Transactions of the Royal Society B*, vol. 365, no. 1558, pp. 3599–3609, 2010.
- [26] M. Belis and S. Guiasu, "A quantitative-qualitative measure of information in cybernetic systems," *IEEE Transactions of Information Theory*, vol. 14, no. 4, pp. 593–594, 1968.