CHAPTER 6

Compositional similarity and β (beta) diversity

Lou Jost, Anne Chao, and Robin L. Chazdon

6.1 Introduction

Spatial variation in species composition is one of the most fundamental and conspicuous features of the natural world. Measures of compositional differentiation and similarity quantify this variation. Conservation biologists apply these measures when setting conservation priorities or evaluating regional conservation plans, and ecologists use them to study the homogenizing or diversifying effects of human activities, natural disturbances, or spatial variability of environmental conditions (McKinney and Lockwood 1999; Olden 2006; Vellend et al. 2007). Measures of similarity and differentiation are also essential tools for evaluating the effects of isolation by distance or geographic barriers, and for describing changes in species composition along environmental gradients.

Ecologists aim not merely to describe spatial variation in community composition, but ultimately to understand the causal factors that produce it. Is spatial heterogeneity caused or maintained by disturbances? Do species assemblages converge or diverge in composition over time during succession (Terborgh et al. 1996; Vandermeer et al. 2004)? How do environmental gradients affect the distributions of species (Bray and Curtis 1957; Whittaker 1960, 1972; Tuomisto et al. 1995)? To what extent do compositional differences reflect neutral variation in species abundance due solely to dispersal limitation (Hubbell 2001)? Carefully chosen measures of similarity and differentiation can be connected to appropriate null models to enable hypothesis testing.

In ecology, a general approach to assessing compositional differentiation partitions regional diversity (γ) into within- (α) and between- (β) group components, and derives similarity and differentiation measures from these components. A wide range of similarity, differentiation, and β diversity measures are in common use, reflecting the many ways that assemblages can be said to differ in composition (reviewed by Legendre and Legendre 1998; Vellend 2001; Koleff et al. 2003; Jurasinski et al. 2009; Tuomisto 2010).

Here, we provide a conceptual overview of approaches to quantifying compositional similarity and differentiation. We then discuss the fundamental connection between diversity and compositional differentiation, and show how some of the most important and useful similarity and differentiation measures are directly connected to diversity measures through the concept of β diversity.

6.2 State of the field

Before we can discuss the important approaches in this field, we need to standardize some terms and introduce some mathematical concepts. Our basic unit of analysis is the *assemblage*, the set of individuals exposed to our sampling efforts in a defined area or point. Ecologists generally try to delimit assemblages so that they are spatially uniform in expected species composition. Assemblages can be delimited by uniform habitats, uniform soil types, constant elevational ranges, or host species, and may range in size from a few cubic centimetres of soil to millions of hectares of forest, depending on the biology of the taxa being examined, as well as the question being asked. If an assemblage really is spatially

uniform in this sense, ecologists can make many samples throughout the assemblage, and each sample will be a replicate drawn from the same assemblage. When this assumption holds, incidence or abundance data from the replicate samples can be used to estimate the relative abundances of each species in the assemblage (Chapter 4). If the assemblage is not homogeneous in the statistical sense, it can still be characterized by its average properties.

Assemblages are said to have identical compositions if the *relative* abundance of each species in one assemblage is the same as the *relative* abundance of that species in the other assemblages (or, when ignoring abundances, they are identical if the lists of species in the assemblages are the same). Thus assemblages can be compositionally identical even if they differ greatly in density (number of individuals per unit area or volume) or area. This definition separates these logically distinct factors. At the other extreme, assemblages are maximally dissimilar if they share no species. A measure of compositional similarity between assemblages compares the relative abundance of each species in each of the assemblages and returns a summary measure of the closeness of these relative abundances.

A measure of *relative* compositional similarity ranges from 0 to 1, assigning a value of 0 to a set of assemblages that share no species, and a value of 1 to a set of compositionally identical assemblages. Compositional *differentiation* is, loosely speaking, the opposite of compositional similarity. A measure of *relative* compositional differentiation ranges from 0 to 1, assigning a value of 0 to a set of compositionally identical assemblages, and a value of 1 to a set of assemblages that share no species. Measures of compositional similarity and differentiation are really just measures of the divergence between the species probability distributions of each assemblage. Divergence measures play an important role in statistics and information theory, and their properties are well studied (Pardo 2006).

Ecologists use similarity measures to make inferences about their assemblages. These inferences usually conceal implicit assumptions about the mathematical properties of the similarity measure used. If the similarity measure does not have the required properties, the ecological inferences will be invalid. At least the following three basic properties should be present in any ecologically useful measure of similarity.

The most basic property is that the measure must truly reflect similarity. Similarity is a multifaceted concept, so it is not possible to define it with complete precision. Different similarity measures may legitimately differ in the way they rank sets of assemblages. However, the concept of similarity does have an unambiguous core. It is possible to create a sequence of assemblages whose compositional similarities unambiguously decrease. Suppose we have a set of equally large, perfectly even assemblages (all species equally common), and suppose each assemblage is identical in composition. If we add unique new species to each assemblage, at the same abundances as the pre-existing species, the similarity of these assemblages must decrease. This is an essential part of the meaning of 'compositional similarity', and this core concept will be referred to as 'monotonicity' with respect to unambiguously decreasing similarity. If a measure *increases* from one set of assemblages to the next in this sequence, it is not measuring similarity. Some commonly used similarity measures fail this most basic test. The reverse of this test applies to differentiation measures.

A measure of compositional similarity should not be sensitive to the raw abundances of the species in each assemblage, but only to their relative abundances. This property may be called 'density invariance'. A measure which lacks this property is not measuring compositional similarity, although it may have other legitimate interpretations.

A third property essential in relative similarity and differentiation measures (those that range between 0 and 1) is 'replication invariance'. Ecologists often compare the similarity of a subset of the data (perhaps a particular genus, or age class, or trophic guild) to the similarity of the whole dataset. For example, we may want to know if large-bodied moths are less spatially differentiated in a forest landscape than the moths as a whole. If these part-to-whole comparisons are to make sense, the measure of relative similarity or differentiation must be invariant with respect to pooling

of identical subsets. Suppose, for example, that we are studying large, medium, and small moths from two sites. Suppose each set of moths has exactly the same number of species and exactly the same abundances. Each set will show exactly the same degree of differentiation, no matter what measure is used. If ecologists made a part-to-whole comparison, looking at the relative differentiation of small moths compared to that of moths in general, they would have to conclude that the small moths are not exceptional in their degree of differentiation. If we want a differentiation measure to lead to this sensible conclusion, the measure must have the property known in economics as 'replication invariance'. If we have N identical subsets of abundances, and no species are shared between subsets, then the community as a whole should have the same degree of similarity as the individual subsets. Imagine what would happen if this were not the case. Taken to an extreme, every genus in the dataset might show the same high site-to-site relative differentiation, while the community as a whole could be assigned a low site-to-site relative differentiation. This actually happens with some common measures of similarity and differentiation. (Tables 6.1 and 6.2)

6.2.1 Measures of relative compositional similarity and differentiation

Incidence-based measures of relative similarity

Incidence-based measures of relative compositional similarity are based only on the presence or absence of species. These have always been very popular with ecologists. The first incidence-based similarity index in ecology was published at the turn of the last century by Jaccard (1900, 1901), and new ones have been proposed continuously since then. Today, a bewildering number of incidence-based similarity measures are available to ecologists. Hubalek (1982) listed 43 and Koleff et al. (2003) listed 24 incidence-type similarity indices. These papers, along with those of Gower (1985) and Legendre and Legendre (1998), provide comprehensive reviews of incidence-based indices. Table 6.1 presents some important incidence-based similarity measures that are monotonic with decreasing similarity, density-invariant, and replication-invariant. Both the original Jaccard measure and the classic Sørensen measure (Sørensen 1948) are widely used. Both were originally designed to compare two assemblages. They differ only in their perspective. The Jaccard index compares the number of shared species

Table 6.1 A class of incidence-based similarity indices for comparing two assemblages and their corresponding abundance-based versions.

Index	Incidence-based In terms of a, b, c	Incidence-based In terms of S_1, S_2, S_{12}	Abundance-based (See text for details)
Jaccard	$\frac{a}{a+b+c}$	$\frac{S_{12}}{S_1+S_2-S_{12}}$	$\frac{UV}{U+V-UV}$
Sørensen; Dice	$\frac{2a}{(2a+b+c)}$	$\frac{2S_{12}}{S_1+S_2}$	$\frac{2UV}{U+V}$
Ochiai	$\frac{a}{[(a+b)(a+c)]^{1/2}}$	$\frac{S_{12}}{(S_1\ S_2)^{1/2}}$	$(UV)^{1/2}$
Anderberg	$\frac{a}{a+2(b+c)}$	$\frac{S_{12}}{2S_1+2S_2-3S_{12}}$	$\frac{UV}{2U+2V-3UV}$
Kulczynski	$\frac{a}{b+c}$	$\frac{S_{12}}{S_1+S_2-2S_{12}}$	$\frac{UV}{U+V-2UV}$
Kulczynski; Cody	$\frac{a}{2(a+b)} + \frac{a}{2(a+c)}$	$\frac{1}{2}\left(\frac{S_{12}}{S_1} + \frac{S_{12}}{S_2}\right)$	$\frac{1}{2}(U+V)$
Lennon et al. (2001) (conditional Sørensen)	$\frac{a}{a+\min(b,c)}$	$\frac{S_{12}}{\min(S_1, S_2)}$	$\frac{UV}{\min(U,V)}$
General	A function of a, b, c satisfying some conditions (see text)	Replace a, b, c by S_{12}, $S_1 - S_{12}$, $S_2 - S_{12}$	Replace a, b, c by UV, $U(1-V)$, $V(1-U)$

This table is an expanded version of Table 2 of Chao et al. (2006). Symbols: a, the number of shared species; b, the number of unique species in the first assemblage; c, the number of unique species in the second assemblage; S_1 the number of species in Assemblage 1; S_2 the number of species in Assemblage 2; S_{12} the number of shared species; U and V: see text.

Table 6.2 Some similarity indices for two and multiple assemblages and their properties (density invariance, replication invariance, and monotonicity, given in Section 6.2),

Index	Two-assemblage	Multiple-assemblage	Density Invariance (6.2)	Replication Invariance (6.2)	Monotonicity (6.2)
Incidence-based					
(1) Jaccard	Jaccard (1900)	Koch (1957)	Yes	Yes	Yes
(2) Sørensen	Sørensen (1948)	Diserud and Ødegaard (2007)	Yes	Yes	Yes
Abundance-based					
(1) Horn index	Horn (1966)	Chao et al. (2008)	Yes	Yes	Yes
(2) Morisita–Horn index	Morisita (1959)	Jost (2006)	Yes	Yes	Yes
(3) Additive (based on Gini–Simpson)	Lande (1996)	Lande (1996)	Yes	No	No
(4) Additive (based on entropy)	Lande (1996)	Lande (1996)	Yes	No	No
(5) Bray–Curtis	Bray–Curtis (1957)	Use the average of	No	Yes	Yes
(6) Percentage Similarity	Renkonen (1938)	pairwise similarities or	Yes	Yes	Yes
(7) Standardized Gower/Euclidean/ Minkowski measure	Gower (1971, 1985)	the average distance from an individual assemblage to the centroid	No	Yes	Yes
(8) Canberra index	Lance and Williams (1967)		No	No	Yes
(9) Correlation coefficient	See, for example, Krebs (1999)		Yes	Yes	No
(10) Normalized expected species shared (NESS)	Grassle and Smith (1976)	Chao et al. (2008)	Yes	No	Yes
(11) Chao–Jaccard and Chao-Sørensen	Chao et al. (2005)	Not available	Yes	Yes	Yes

Formulas

Assume that there are S_i species in the ith assemblage and S species in the combined assemblage. Let $\overline{S}$ denote the average number of species, M_{ir} denote the abundance of the ith species in the rth assemblage, and p_{ir} denote the relative abundance, $i = 1, 2, \ldots, S$ and $r = 1, 2, \ldots, N$. Thus, we have N sets of species abundance $\{(M_{1r}, M_{2r}, \ldots, M_{Sr});\ r = 1, \ldots, N\}$ and N sets of the relative abundance $\{(p_{1r}, p_{2r}, \ldots, p_{Sr});\ r = 1, \ldots, N\}$.

Incidence-based indices

(1) Incidence-based Jaccard for two assemblages = $S_{12}/(S_1 + S_2 - S_{12})$
Incidence-based Jaccard for multiple assemblages = $(\overline{S}/S - 1/N)/(1 - 1/N)$
(2) Incidence-based Sørensen for two assemblages = $2S_{12}/(S_1 + S_2)$
Incidence-based Sørensen for multiple assemblages = $(N - S/\overline{S})/(N - 1)$

Abundance-based indices

(1) Horn overlap index for two assemblages = equation 6.1
Horn overlap index for N assemblages = equation 6.4
(2) Morisita–Horn index for two assemblages = equation 6.2
Morisita–Horn index for N assemblages = equation 6.5
(3) Additive (based on Gini–Simpson index) for N assemblages = H_α/H_γ, where

$$H_\alpha = (1/N)\sum_{j=1}^{N}(1 - \sum_{i=1}^{S} {}_i\, p_{ij}^2), \quad H_\gamma = 1 - \sum_{i=1}^{S}\overline{p}_i^2 \text{ and } \overline{p}_i = \sum_{j=1}^{N} p_{ij}/N$$

Table 6.2 *Continued.*

(4) Additive (based on entropy) for N assemblages = H_α / H_γ, where

$$H_\alpha = -\frac{1}{N}\sum_{i=1}^{S}\sum_{j=1}^{N} p_{ij}\log(p_{ij}),\ H_\gamma = -\sum_{i=1}^{S}\overline{p}_i\log(\overline{p}_i) \text{ and } \overline{p}_i \text{ is defined in (3)}$$

(5) Bray–Curtis index for two assemblages = equation 6.6

(6) Percentage similarity for two assemblages = equation 6.7

(7) Gower/Euclidean/Minkowski similarity for two assemblages = equation 6.8

(8) Canberra index for two assemblages = $1 - \frac{1}{N}\sum_{i=1}^{S}\frac{|M_{i1}-M_{i2}|}{M_{i1}+M_{i2}}$

(9) Correlation coefficient = correlation of the two sets of abundances ($M_{11}, M_{21}, \ldots, M_{S1}$) and ($M_{12}, M_{22}, \ldots, M_{S2}$). This measure takes value between -1 and 1.

(10) Normalized expected species shared (NESS(m)) for two assemblages:

$$\frac{2\sum_{i=1}^{S}\mu_{i1}(m)\mu_{i2}(m)}{\sum_{i=1}^{S}[\mu_{i1}(m)]^2 + \sum_{i=1}^{S}[\mu_{i2}(m)]^2}$$

where $\mu_{ij}(m) = 1 - (1 - p_{ij})^m$.
NESS(m) for N assemblages:

$$\frac{2}{(N-1)}\frac{\sum_{i=1}^{S}\sum_{j<k}\mu_{ij}(m)\mu_{ik}(m)}{\sum_{i=1}^{S}\sum_{j=1}^{N}[\mu_{ij}(m)]^2}$$

(11) Chao–Jaccard for two assemblages = $UV/(U + V - UV)$
Chao-Sørensen for two assemblages = $2UV/(U + V)$

where U(V) denotes the total relative abundance of the shared species in assemblage 1 (2).

to the total number of species in the combined assemblages, while the Sørensen index compares the number of shared species to the mean number of species in a single assemblage. The Jaccard index takes a global view while the Sørensen index takes a local view. Table 6.3 gives some examples. The Sørensen index is a true overlap measure (Wolda 1981): if the two assemblages have the same number of species, the Sørensen index gives the proportion of shared species in each assemblage (i.e. the proportion of an assemblage's species list that overlaps with the species list of the other assemblage).

The Sørensen index is the harmonic mean of the proportion of shared species in the first assemblage and the proportion of shared species in the second assemblage. The Ochiai index is the geometric mean and the Kulczynski–Cody index (Kulczynski 1928) is the arithmetic mean of the two proportions. Because of the relationship between these means, it is always true that Sørensen index $\leq$ Ochiai index $\leq$ Kulczynski–Cody index.

When the richness of one assemblage is much greater than the richness of the other, both the Jaccard and Sørensen indices are always small. This is an accurate reflection of the difference in species compositions of the two assemblages, but for some applications it can be useful to normalize the measures so that they take the value unity when overlap is as large as it can be, given the respective richnesses of the two assemblages. Lennon et al. (2001) proposed this modification to the Sørensen index, which may be called the conditional Sørensen index. The formula for the conditional Sørensen index is given in Table 6.1. When there are no unique species in one of the assemblages, the conditional Sørensen index always yields a value of 1 no matter how many unique species there are in the other assemblage. Hence, this measure is not informative when applied to two nested assemblages

Table 6.3 Artificial examples with four sets (A, B, C, D) of assemblages illustrating various similarity indices.

Species	**Set A**		**Set B**		**Set C**		**Set D**		
	Assemblage		Assemblage		Assemblage		Assemblage		
	1	2	3	4	5	6	7	8	9
a	0.125	0.25	0.6	0.55	0.8	0.10	0.75	0.20	0.10
b	0.125	0.25	0.1	0.05	0.1	0.55	0.01	0.15	0.20
c	0.125	0.25	0.1	0.03	0.1	0	0.02	0.05	0.50
d	0.125	0.25	0.1	0.02	0	0.33	0.03	0	0
e	0.125	0	0.05	0	0	0.01	0.19	0	0
f	0.125	0	0.05	0	0	0.01	0	0.4	0
g	0.125	0	0	0.20	0	0	0	0.2	0
h	0.125	0	0	0.15	0	0	0	0	0.08
i	0	0	0	0	0	0	0	0	0.12
Incidence based									
Jaccard	0.50		0.50		0.33		0.33 (See text)		
Sørensen (C_{0N})	**0.67**		**0.67**		**0.50**		**0.60 (See text)**		
Lennon et al.	1.00		0.67		0.67		0.50 (Baselga et al. 2007)		
Abundance based									
Bray–Curtis	0.33		0.53		0.20		0.23 (pairwise mean)		
Percentage	0.50		0.65		0.20		0.22 (pairwise mean)		
Horn entropy (C_{1N})	**0.69**		**0.73**		**0.43**		**0.41 (equation 6.4)**		
Morisita–Horn (C_{2N})	**0.67**		**0.89**		**0.25**		**0.27 (equation 6.5)**		
Chao–Jaccard	0.50		0.61		0.61		Not available		
Chao–Sørensen	0.67		0.75		0.75		Not available		

For each species in the assemblage, the relative abundance is indicated. (For computing the Bray–Curtis index, the total number of individuals is assumed to be 100 in Assemblages 1, 3, 5, 7, and 9, and 200 for Assemblages 2, 4, 6, and 8.)

(see Set A of Table 6.3). See Legendre and Legendre (1998) and Koleff et al. (2003) for the interpretation of the other measures listed in Table 6.1.

Most of these measures were originally designed to compare two assemblages. When more than two assemblages are compared, ecologists have often averaged their pairwise similarities (e.g. Lennon et al. 2001; Vellend 2001). However, the pairwise similarities tend to be correlated, which will cause an inference problem (Diserud & Ødegaard 2007). Aside from the correlation problem, pairwise similarities cannot fully characterize multiple-assemblage similarity when some species are shared across two, three, or more assemblages. A simple example is given by Chao et al. (2008). These problems have motivated the search for multiple-assemblage generalizations of similarity measures that can take global similarity into account.

Although there were measures in the literature that can be regarded as multiple-site Sørensen or related measures, a 'direct' extension of the Sørensen index to simultaneously compare multiple assemblages was first derived by Diserud & Ødegaard (2007). It has the form $(N - S/\overline{S})/(N - 1)$, where N denotes the number of assemblages, S denotes the total number of species in the combined assemblage, and $\overline{S}$ denotes the average number of species per assemblage. This measure is a monotonic transformation of Whittaker's β diversity (see below for details). It is identical to the complement of the 'turnover' measure by Harrison et al. (1992). For multiple assemblages, this generalized Sørensen index provides the overall proportion of shared species from a local perspective. Consider the special case in which each of the N assemblages has S species. Suppose that exactly R species are shared by all assemblages, and the remaining

species are unique to their assemblage. This measure then gives the proportion of each assemblage's species that overlap with the other assemblages, R/S The conditional Sørensen index has also been generalized to multiple assemblages (Baselga et al. 2007).

A multiple-site Jaccard index was presented by Koch (1957). It is expressed as $(\overline{S}/S - 1/N)/(1 - 1/N)$, which measures the overall proportion of the shared species in the combined assemblage. In the example of the preceding paragraph, this measure reduces to $R/$(total number species in the combined assemblage). Like the multiple-assemblage generalization of the Sørensen index, it is a monotonic transformation of Whittaker's multiplicative β diversity (Jost 2007). This measure is less useful than the multiple-assemblage Sørensen index because it tends to decline as more assemblages are compared, even if overlap between assemblages is kept constant.

In Table 6.3, Set D includes three assemblages with completely different non-shared species. There are nine species in the whole assemblage, and in each assemblage there are five species, three of which are shared by all assemblages. The proportion of shared species relative to the combined assemblages (multiple-assemblage Jaccard index) is 1:3, whereas the average proportion of shared species in each assemblage (multiple-assemblage Sørensen index) is 3:5. The values for the conditional Sørensen index are also given in Table 6.3.

Each of these measures provides different information about compositional similarity. When choosing a measure, ecologists should think carefully about whether their perspective is local or global. If it is local, the multiple-assemblage Sørensen index is an informative and widely used index, facilitating comparisons across studies. The conditional version of the Sørensen index gives useful additional information, so if diversities differ greatly among assemblages this should also be presented. If the perspective is global, the multiple-assemblage Jaccard index and its conditional generalization could be used, but if the number of assemblages is high, both should be interpreted with caution (see above). When there are many assemblages, it is often easier to convey an intuitive picture of similarity through the Sørensen index rather than the Jaccard index, regardless of whether the perspective is local or global.

A more basic decision is whether or not to use incidence-based similarity measures at all. They are attractive because of the apparent ease of fieldwork, since only presence or absence needs to be noted. When nearly all species in each assemblage can be sampled, they are easily interpretable tools for comparing species lists. In addition, incidence-based measures may be the only alternative available when it is hard to define or count individuals of each species, as in 'uncountable' microbes, colonies of insects, or plant and invertebrate taxa with clonal growth.

In species-rich assemblages, however, intensive sampling efforts are needed to make complete or nearly complete lists of the species in a given assemblage. In practice, similarity measures must be estimated from sample data, so the species richness and shared species richness in the formulas of Table 6.1 are replaced by the observed counts. It is well known (e.g. Wolda, 1981, 1983; Magurran, 2004, p. 175) that all incidence-based indices are biased when sampling is incomplete, and the biases are likely to be substantial for assemblages with high species richness and a large fraction of rare species. The Jaccard and Sørensen indices are often biased downwards (Chao et al. 2006), but could be biased upwards too. The biases exist even in the simplest case when all species are equally abundant. The bias cannot be reduced or removed by using equal sampling fractions, equal sample sizes, or equal effort. Not only the bias but also the variance depends on species abundances (not incidence alone); thus, it is impossible to correct for the bias or to estimate errors without using abundance data. Even when abundance data are available, correcting the bias and assessing variance is not easy. See Chao et al. (2005, 2006) for relevant discussions.

One may be inclined to estimate species richness and shared species richness using non-parametric, low-bias estimators (Chao 2005), plugging these into the formulas in Table 6.1. However, replacing observed values of species richness and shared species with non-parametric estimates is problematic for assessing assemblage similarity. Combining

these estimates in the formulas of Table 6.1 unavoidably inflates the variance and often renders the resulting estimate useless.

Finally, and perhaps most importantly, incidence-based similarity measures greatly oversimplify the relationships between assemblages. All such measures regard a maple forest with a few scattered pine trees as identical in composition to a pine forest with one or two maples, even though the two forests are ecologically very different. The relative abundances of species are ecologically important quantities. They should not be ignored, unless there is no choice.

Abundance-based similarity and distance measures

An ecologically more meaningful and informative approach to assessing similarity focuses on the differences in species frequencies among assemblages. This approach is also statistically more accurate and precise than incidence-based approaches. Some authors have adopted a parametric approach to abundance-based similarity measures by assuming that the relative abundances follows a parametric distribution (e.g. Smith, Solow & Preston 1996; Plotkin & Muller-Landau 2002). However, it is almost impossible to test whether a given sample really came from a particular parametric distribution (Chao 2005). We therefore focus on nonparametric indices that do not assume a particular kind of species abundance distribution.

In all the formulas that follow, we assume there are N assemblages, with S total species in the combined assemblages Let M_{ir} denote the absolute abundance of the ith species in the rth assemblage, and p_{ir} denote its relative abundance, with $i = 1, 2, \ldots, S$ and $r = 1, 2, \ldots, N$. Thus, we have N sets of species absolute abundances $\{(M_{1r}, M_{2r}, \ldots, M_{Sr}); r = 1, \ldots, N\}$ and N sets of the relative abundances $\{(p_{1r}, p_{2r}, \ldots, p_{Sr}); r = 1, \ldots, N\}$

Horn and Morisita–Horn overlap measures

The incidence-based Sørensen overlap index has abundance-based relatives. One of the most useful is the Horn (1966) overlap measure, based on Shannon's entropy (Chapter 4). In its original form it compares two assemblages, species by species. To make this clearer, it can be expressed as

$$S_H = \frac{1}{\log 2}\sum_{i=1}^{S}\left[\frac{p_{i1}}{2}\log\left(1+\frac{p_{i2}}{p_{i1}}\right)+\frac{p_{i2}}{2}\log\left(1+\frac{p_{i1}}{p_{i2}}\right)\right]. \tag{6.1}$$

This index equals unity if and only if $p_{i1} = p_{i2}$ for all i. When the two assemblages have equal numbers of species and consist entirely of equally common species, the Horn index is equal to the Sørensen index and gives the proportion of shared species in an assemblage.

Another popular overlap measure is the Morisita Horn index, derived by Morisita (1959) and Horn (1966):

$$S_{MH} = \frac{2\sum_{i=1}^{S} p_{i1}p_{i2}}{\left[\sum_{i=1}^{S} p_{i1}^2 + \sum_{i=1}^{S} p_{i2}^2\right]} = 1 - \frac{\sum_{i=1}^{S}(p_{i1}-p_{i2})^2}{\sum_{i=1}^{S} p_{i1}^2 + \sum_{i=1}^{S} p_{i2}^2} \tag{6.2}$$

When the two assemblages are equally diverse and consist entirely of equally common species, the Morisita–Horn index is equal to the Horn index and the Sørensen index, and all give the proportion of shared species in an assemblage. The right side of equation 6.2 shows that the Morisita–Horn measure is based on the squared differences of the relative abundances of each species in the two assemblages. Because of this squared distance, the Morisita–Horn index is dominated by the most abundant species, while the relatively rare species have little effect (even if there are many of them). This makes the measure resistant to under-sampling because the influential abundant species are always sampled relatively accurately. Since ecological processes are often most strongly influenced by the dominant species, this measure is a good measure when looking for functional differences between ecosystems. It may not be appropriate when rare species are important, as in conservation applications. The Horn overlap measure would be more useful in those applications.

C_{qN}, a general multiple-assemblage abundance-based overlap measure

Surprisingly, all the overlap measures we have discussed—the Sørensen index, the Horn index, and the Morisita–Horn index—are special cases of a single general multiple-assemblage overlap measure C_{qN} (Chao et al. 2008):

$$C_{qN} = \frac{\frac{1}{(N^q - N)} \sum_{i=1}^{s} [(p_{i1} + p_{i2} + \ldots + p_{iN})^q - (p_{i1}^q + p_{i2}^q + \ldots p_{iN}^q)]}{\frac{1}{N} \sum_{i=1}^{S} (p_{i1}^q + p_{i2}^q + \ldots p_{iN}^q)} \tag{6.3}$$

Here q is a parameter that determines the measure's sensitivity to species' relative abundances, and N is the number of assemblages. When $N = 2$, setting $q = 0$ yields the Sørensen index, taking the limit as q approaches unity yields the Horn index, and setting $q = 2$ yields the Morisita–Horn index. Setting $N > 2$ yields the corresponding multiple-assemblage generalizations. The generalized Sørensen index is $C_{0N} = (N - S/\overline{S})/(N - 1)$ and the generalized Horn index can be expressed as

$$C_{1N} = \frac{1}{\log N} \sum_{i=1}^{S} \sum_{j=1}^{N} \left[\frac{p_{ij}}{N} \log \left(1 + \sum_{k \neq j} p_{ik} / p_{ij} \right) \right] \tag{6.4}$$

The multiple-assemblage Morisita–Horn index can be simplified as

$$C_{2N} = \frac{2 \sum_{i=1}^{S} \sum_{j<k} p_{ij} p_{ik}}{(N-1) \sum_{i=1}^{S} \sum_{j=1}^{N} p_{ij}^2} \tag{6.5}$$

For integer values of q between 2 and N, overlap measures C_{qN} have a simple statistical interpretation as the ratio of two probabilities, ${}^qG_D/{}^qG_S$. The numerator is the probability that q randomly sampled individuals belong to the same species, given that they did not all come from the same assemblage. The denominator is the probability that q randomly sampled individuals belong to the same species given that they are all drawn from the same assemblage (Chao et al. 2008). This interpretation shows that the value of q determines the 'depth' of the measure: when $q = 2$, only pairwise similarity is considered, but when $q = 3$ the measure also takes into account species that are shared by three assemblages. This measure is density and replication invariant, and monotonic with respect to unambiguously decreasing similarity. Like the multiple-assemblage generalizations of the Sørensen index, C_{qN} gives the true overlap R/S for all orders of q when each assemblage consists of S equally common species, with R species contained in all assemblages and the remainder unique to single assemblages. In the section on using transformations of β to measure compositional similarity (below) we give an example of the use and interpretation of these measures.

The expected value of C_{2N} (the Morisita–Horn index and its multiple-assemblage generalization) can be predicted under Hubbell's (2001) neutral model of biodiversity, which makes it an important theoretical tool. In Hubbell's neutral model, the community structure depends only on the number of communities N, the number of individuals in each community n (all communities are assumed to have the same size), the migration rate between communities m, and the speciation rate v. After many generations it reaches an equilibrium, and the structure at equilibrium can be predicted from the model parameters. This model is mathematically identical to Wright's finite-island model in population genetics. The expected equilibrium value of the complement of the Morisita–Horn index has been derived for that model (Jost 2008); under the approximation that $1 \gg m \gg v$, the formula for the expected equilibrium value of the Morisita–Horn index itself is $S_{MH} = m/[v(N-1) + m]$. If similarity between assemblages is much different from this, some causal factors are implicated. C_{1N} (the Horn measure of similarity and its multiple-assemblage generalization) can also be connected to this model under certain special conditions (Sherwin et al. 2006), although a simple analytical expression in terms of model parameters still eludes us.

Bray–Curtis and Renkonen similarity measures

One of the most frequently used abundance-based similarity measures is the Bray–Curtis or 'quantitative Sørensen' index (Bray & Curtis 1957), developed by Bray and Curtis during their pioneering work on plant community ordination. Beals (1984) gives a detailed review of the Bray–Curtis ordination and related techniques. The Bray–Curtis similarity is expressed as

$$S_{BC} = \frac{2 \sum_{i=1}^{S} \min(M_{i1}, M_{i2})}{\sum_{i=1}^{S} (M_{i1} + M_{i2})} = 1 - \frac{\sum_{i=1}^{S} |M_{i1} - M_{i2}|}{\sum_{i=1}^{S} (M_{i1} + M_{i2})} \tag{6.6}$$

where $\min(M_{i1}, M_{i2})$ denotes the smaller of the two numbers, M_{i1} or M_{i2}. This index reduces to the

Sørensen index if all species in each assemblage are equally abundant. The expression on the right in equation 6.6 shows that it takes the maximum value of 1 if and only if the two sets of absolute abundances are identical, that is, $M_{1i} = M_{2i}$ for all species. This shows that the index confounds density with compositional similarity, so it cannot be considered a measure of compositional similarity. It approaches 0 when sample sizes are very different, whether the assemblages have the same compositions or completely different compositions. For example, suppose every species in a young secondary forest has exactly the same relative abundance as in a primary forest. Suppose the secondary forest has four times as many stems per hectare as the primary forest. Then the Bray–Curtis index will be approximately $1-\frac{3}{5} = 0.4$ rather than unity. If the secondary forest had 10 times the density of the primary forest, the Bray–Curtis index would be about 0.18 instead of unity. If it had 100 times the density of the primary forest, the Bray–Curtis index would be close to 0.02.

In addition to these conceptual problems, the Bray–Curtis index also has some statistical problems when applied to samples. The *observed* absolute abundance for any species depends on the sampling fraction (the ratio of sample size to the total number of individuals in the assemblage), so this index becomes meaningless and performs erratically when the sampling fraction is unequal in the two assemblages (Chao et al. 2006). It also generally has a very large bias in this case (Chao et al. 2006). For these reasons, from both statistical and conceptual perspectives, the Bray–Curtis index cannot be recommended unless sampling fractions are known to be equal. Given the unlikely prospect of establishing such conditions for field data, the Bray–Curtis index seems rarely to be an acceptable choice for such data. Note that equalizing the *number of individuals* (sample sizes) in all samples by rarefaction before calculating the Bray–Curtis index, as suggested by Horner-Devine et al. (2004), does not equalize sampling fractions unless the assemblages themselves can be reasonably assumed to have the same total number of individuals susceptible to sampling.

It follows from equation 6.6 that the Bray–Curtis measure can be interpreted as a normalized Manhattan distance (absolute difference). When there are multiple assemblages, as in the conventional distance-based approach, an average of pairwise Bray–Curtis values is generally used as an overall similarity measure. This approach, unfortunately, ignores information about the species shared among three or more assemblages (Chao et al. 2008).

Renkonen (1938) proposed the following measure of similarity:

$$S_P = \sum_{i=1}^{S} \min(p_{i1}, p_{i2}) = 1 - \frac{1}{2}\sum_{i=1}^{S} |p_{i1} - p_{i2}| \tag{6.7}$$

This measure is based on the Manhattan distance of the relative abundances of the species in the assemblages, instead of the absolute abundances as in the Bray–Curtis index. It is therefore density invariant, so it is a valid measure of compositional similarity. It attains a minimum value of 0 when two assemblages are completely distinct (no shared species) and it attains a maximum value of 1 if and only if the two sets of relative abundances are identical. Following the work of Renkonen (1938), Whittaker (1952, 1972) and Wolda (1981), this index has become one of the most commonly used measures among ecologists. Gregorius (1987) also suggested its use for a proper genetic distance among populations.

Smith and Zaret (1982) found that in their simulation trials this measure tends to be biased when estimated from sample data, but Wolda (1981) recommended its use if a logarithm transform is applied to species frequencies to reduce the dominance of the most abundant species. This measure suffers from the same disadvantage as the Bray–Curtis index, namely that the analysis for multiple sites can only be based on pairwise comparisons, and thus information about species shared between three or more sites is ignored.

Distance measures

We have seen that the two-assemblage Bray–Curtis and Renkonen similarity indices are complements of some types of normalized Manhattan distance, and the two-assemblage Morisita–Horn index is the complement of a type of normalized Euclidean distance. This use of normalized distance

between assemblages is a natural way to form an index of similarity or differentiation between two sets of quantitative variables. There are many distances, and some common approaches include Euclidean distance, Manhattan distance, Lance and Williams' (1967) Canberra distance, chi-square distance, Minkowski distance, Orlóci's chord distance, and Hellinger distance (Legendre & Legendre 1998; Gower 1985). As indicated by Gregorius (1996), a distinction between 'distance' measures and 'differentiation' measures is that 'distance' refers to a pair of assemblages and obeys the triangle inequality whereas 'differentiation' refers to any number of assemblages but need not satisfy the triangle inequality. A unified class of distance-based similarity measures is the one-complement of the normalized Minkowski distance:

$$1 - \frac{1}{S}\sum_{i=1}^{S}\frac{|M_{i1} - M_{i2}|^{p}}{R_i^{p}} \qquad (6.8)$$

where $p \geq 1$, and R_i denotes the range (i.e. the difference of the maximum and minimum) of the ith species abundances. The special case $p = 1$ corresponds to the Gower (1971) similarity index. The special case $p = 2$ corresponds to the similarity index based on the Euclidean distance.

Due to the restriction that 'distance' is only calculated for a pair of assemblages, this approach cannot be directly extended to the case of multiple assemblages. Thus, the information shared by at least two assemblages is ignored in the traditional approach using the average of pairwise distances as the overall dissimilarity. Anderson et al. (2006) modified this approach by considering a multivariate dispersion as the 'β' diversity or a dissimilarity measure. A multivariate dispersion is defined as the average of the distance from an individual assemblage to the 'centroid.' This approach does simultaneously compare the whole set of assemblages because the information shared by multiple assemblages can be incorporated in the 'centroid'.

Indices based on total abundance of shared species

The abundance-based indices discussed above match species relative abundances, *species by species*. That is, the typical similarity indices assess a normalized probability that two randomly chosen individuals, one from each assemblage, belong to *the same* species. Chao et al. (2005, 2006) derived a class of measures which look at a different kind of similarity, based on assessing the normalized probability that two randomly chosen individuals, one from each assemblage, belong to *shared* species (not necessarily *the same* species). It is simple to convert any incidence-based measure that is replication invariant to its shared-abundance version. Let U denote the total relative abundances associated with the *shared* species in Assemblage 1 and let V denote the total relative abundances of the *shared* species in Assemblage 2. By replacing a, b and c in Table 6.1 with UV, $U(1-V)$ and $V(1-U)$, we obtain for each incidence-based index its shared-abundance version, given in the last column of Table 6.1. These indices attain the maximum value of 1 if and only if $U = V = 1$ (even if the two sets of abundances may be different). These two shared-abundance indices are called the Chao–Jaccard and Chao–Sørensen abundance indices in the literature and software packages (e.g. EstimateS by Colwell (2006) and SONS by Schloss and Handelsman (2006a)).

Unlike the ordinary similarity indices that match species-by-species abundances, these measures match the total relative abundances of species shared between two assemblages. In Sets B and C considered in Table 6.3, the total shared species abundances are the same for the first assemblage in each set ($U = 0.9$) and also the same for the second assemblage in each set ($V = 0.65$), yielding the same Chao–Sørensen and Chao–Jaccard shared abundance indices, even though the patterns of abundance for species a and b are drastically different for the two sets. Set C in particular shows that these two indices are quite different from species-by-species abundance-based similarity measures.

One main advantage of these measures is that the under-sampling bias due to unseen shared species can be evaluated and corrected. This class of measures can also be extended to deal with replicated incidence data. The extension of this approach to assessing similarity of multiple assemblages, although conceptually simple, is statistically complicated and is still under development. These measures are designed to be sensitive to rare shared species while still taking abundance into account, so they may make large jumps as more shared species are discovered. They should be used only if their

concept of similarity is relevant to the question of interest. For examples of their use and interpretation, see Chao et al. (2005, 2006), Anderson et al. (2006) and Schloss and Handelsman (2006a).

6.2.2 Diversity and compositional similarity

There is an intimate connection between species diversity and compositional differentiation and similarity. If assemblages are identical in composition, pooling them will leave their diversity unchanged. If assemblages are very different in composition, pooling them together with equal weights should result in a total diversity considerably greater than the diversity of any of the assemblages individually. Comparing the diversities of the individual assemblages to the diversity of the pooled assemblages will provide information about their compositional similarity, as long as the pooling gives equal weight to each assemblage. Many of the similarity and differentiation measures discussed above are based on different ways of comparing the diversities of the individual assemblages to the diversity of the pooled assemblages.

Similarity ratio based on diversity measures

Ecologists often use the ratio of mean within-group diversity (the 'α diversity') to total pooled diversity ('γ diversity') as a measure of the compositional similarity of the groups (Lande 1996). If this ratio is close to unity, ecologists infer that the groups must be similar in composition. If the ratio is low, they infer that the groups are highly differentiated in composition.

While this reasoning seems logical, its validity actually depends on the measure used. When the most common measures of complexity are used, such as Shannon entropy (see Chapters 3 and 5)

$$H_S = -\sum_{i=1}^{S} p_i \log p_i$$

or the Gini–Simpson index

$$H_{GS} = 1 - \sum_{i=1}^{S} p_i^2$$

the ratio of mean within-group 'diversity' to total pooled 'diversity' always approaches unity when within-group diversity is high (MacArthur 1965; Whittaker 1972; Jost 2006, 2007; Tuomisto 2010) even if all the groups are completely distinct (no species in common). This problem arises frequently in practice and can lead to serious misinterpretations (Jost et al. 2010). These 'similarity' measures based on the Gini–Simpson index or Shannon entropy are also not monotonic with respect to unambiguously increasing similarity, so they do not rank datasets correctly in terms of their similarity (Jost et al. 2010). They should not be used to infer compositional similarity across space or time. Because of their non-linearity with respect to pooling of equally diverse groups, they are also dangerously misleading in conservation applications (Jost 2009).

The great mathematical ecologist Robert MacArthur was the first to notice the problems with computing similarity ratios based directly on Shannon entropy or the Gini–Simpson index (MacArthur 1965). He solved the problem by converting these measures to effective number of species. The effective number of species is the number of equally common species needed to produce the observed value of an index. For example, to find the effective number of species for a Shannon entropy of 2.4 (assuming this value of 2.4 was calculated using natural logarithms), we need to find out how many equally common species are needed to obtain a Shannon entropy of 2.4. The answer is 11 species. Shannon entropy is converted to effective number of species by taking its exponential:

$$^1D = \exp(H_S) = \exp\left(-\sum_{i=1}^{S} p_i \log p_i\right)$$

$$= \text{'exponential of Shannon entropy'}$$

while the Gini–Simpson index or heterozygosity is converted by the following formula:

$$^2D = 1/(1 - H_{GS}) = 1/\sum_{i=1}^{S} p_i^2$$

$$= \text{'inverse Simpson concentration'}$$

MacArthur (1965) observed that when the numerator and denominator of the 'similarity' ratio described in the preceding paragraphs is converted to effective number of species, the ratio truly did

reflect the compositional similarity of the assemblages, as long as both the within-group mean and the pooled diversity are calculated using equal weights for each assemblage. Pooling with equal weights means not pooling samples directly, but pooling them in equal proportions.

The particular mathematical property which makes some measures behave properly in similarity ratios while others fail is known in ecology as the doubling property or in economics as the replication principle (Dalton 1920; Hannah and Kay 1977). Suppose we make N equally large copies of an ecosystem, each with identical species abundance distributions but with completely different species. The replication principle states that the diversity of the N pooled ecosystems (which share no species) must be N times the diversity of the original ecosystem. This replication principle seems to be implicit in many forms of reasoning about diversity (Hill 1973; Jost 2009). (Replication *invariance*, discussed above, refers to measures that do not change when assemblages are replicated N times.)

Hill (1973) derived a parametric family of diversity measures that obey the replication principle:

$$ {}^{q}D = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)} $$

This formula gives the effective number of species of all standard complexity and diversity measures (Jost 2007). As discussed earlier, the value of q determines the sensitivity of the measure to species relative abundances; ecologists can either emphasize ($q > 1$) or de-emphasize ($q < 1$) the relative abundances. When $q = 0$, this formula gives species richness. When $q = 2$, it gives the inverse Simpson concentration (Simpson 1949). When $q = 1$ (the value that weighs all species exactly by their frequency), the formula is undefined (because of a division by zero in the exponent), but the limit as q approaches unity exists and equals the exponential of Shannon entropy. Shannon entropy thus can be derived directly from the mathematics of diversity and does not need to be borrowed from other fields. In order to avoid invalid inferences about diversity, these measures should always be used in preference to the raw Shannon entropy or Gini–Simpson index (MacArthur 1965; Whittaker 1972; Routledge 1979; Peet 1974; Jost 2006, 2007, 2009). Measures of compositional complexity that obey the replication principle are the 'true diversities' of Jost (2007, 2009) and Tuomisto (2010).

When a complexity measure that obeys the replication principle is used in the similarity ratio, with all assemblages weighted equally and with the mean defined in equation 6.9a below, the similarity ratio ranges from unity (complete similarity) to $1/N$ (complete dissimilarity). This range does not depend on their diversities and is easily normalized onto the interval from 0 to 1.

Partitioning diversity into α and β components

As mentioned in the preceding section, α diversity is a mean (not necessarily the arithmetic mean) of the diversities of a set of assemblages, whereas γ diversity is the diversity of the pooled assemblages. γ diversity is at least as great as α diversity, and will exceed α diversity when the assemblages are differentiated. γ diversity therefore has two contributions, one from α diversity and the other from differentiation between assemblages. The contribution of differentiation to the total or γ diversity is called β diversity.

What is the relationship between α, β, and γ? Whittaker (1972) proposed that β be defined through a multiplicative relationship:

$$ \alpha \times \beta = \gamma $$

He applied this relationship to effective numbers of species (species richness, the exponential of Shannon entropy, and the inverse Simpson concentration). He restricted his discussion to the case of equally weighted assemblages, which is the appropriate choice when the goal is to measure compositional similarity.

Lande (1996) proposed instead that β be defined through an additive relationship with α:

$$ \alpha + \beta = \gamma $$

This additive approach is meant to apply to concave measures of diversity and complexity (see Lande (1996) for a definition of concave), such as species richness (see Chapter 4), Shannon entropy (see Chapters 3 and 5), and the Gini–Simpson index. In this approach, α is the weighted arithmetic mean of the diversity or complexity measure of the

individual assemblages. β is found by subtracting α from γ. The concavity property ensures that γ is never less than α, so β defined in this way is always non-negative. This approach is superficially similar to the additive partitioning of variances.

These approaches are often respectively called multiplicative partitioning and additive partitioning of γ into α and β components. However, 'partitioning' usually means a *complete* separation of the two components, α (the within-group component) and β (the between-group component). A genuine partitioning of γ into α and β components would place all the within-group influence into the α component, and all the between-group influence into the β component. If the partitioning is complete, the α component provides no information about the value of the β component, and vice versa; they are completely independent components of γ diversity or complexity measure.

When a complete partitioning of γ exists, its form is unique for any given measure. It exists for all standard diversity and complexity measures when assemblages are given equal weights. For species richness ($q = 0$), the complete partitioning is multiplicative: $\alpha \times \beta = \gamma$. For Shannon entropy ($q = 1$), it is additive: $\alpha + \beta = \gamma$. For the Gini–Simpson index, it is neither additive nor multiplicative: $\alpha + \beta - \alpha \times \beta = \gamma$. For all effective number of species, or Hill numbers of any order, which are true diversities in the sense of Jost (2007), it is multiplicative: $\alpha \times \beta = \gamma$ (Jost 2007). The components are:

$$ {}^{q}D_{\alpha} = \left(\frac{1}{N} \sum_{i=1}^{S} p_{i1}^{q} + \frac{1}{N} \sum_{i=1}^{S} p_{i2}^{q} + \ldots + \frac{1}{N} \sum_{i=1}^{S} p_{iN}^{q} \right)^{1/(1-q)} \quad (6.9a) $$

$$ {}^{q}D_{\gamma} = \left\{ \sum_{i=1}^{S} \left[\frac{1}{N} (p_{i1} + p_{i2} + \ldots + p_{iN}) \right]^{q} \right\}^{1/(1-q)} \quad (6.9b) $$

$$ {}^{q}D_{\beta} = {}^{q}D_{\gamma} / {}^{q}D_{\alpha} \quad (6.9c) $$

β for true diversities is just the reciprocal of the similarity ratio of the previous section. It ranges from 1 to N, and gives the effective number of completely distinct assemblages; it is a divergence measure between the species probability distributions of the assemblages. It is a 'true β diversity' in the senses of Jost (2007) and Tuomisto (2010). In the special case of $q = 0$, equation 6.9c reduces to Whittaker's β diversity $S/\overline{S}$ based on species richness. Investigators using these measures should at least report the values of β for $q = 0$, 1, and 2, in order to give a picture of how differentiation varies between all species ($q = 0$), the typical species ($q = 1$), and the dominant species ($q = 2$). A more complete picture is conveyed by a graph of β diversity from $q = 0$ to $q = 4$ or 5 (beyond this it usually does not change much). This is the β diversity profile. This can be easier to interpret when normalized onto the unit interval to give a similarity profile, as described in the next section (see Box 6.1 for an example).

When assemblage area or size needs to be taken into account, as when measuring regional heterogeneity (Horn 1966) instead of compositional differentiation, a complete partitioning into a measure of mean within-site diversity and a measure of between-site differentiation only exists for measures based on Shannon entropy and species richness (Jost 2007). For the exponential of Shannon entropy, the components are:

$$ {}^{1}D_{\alpha} = \exp\left(-w_1 \sum_{i=1}^{S} p_{i1} \log p_{i1} - w_2 \sum_{i=1}^{S} p_{i2} \log p_{i2} \ldots - w_N \sum_{i=1}^{S} p_{iN} \log p_{iN} \right) \quad (6.10a) $$

$$ {}^{1}D_{\gamma} = \exp\left[-\sum_{i=1}^{S} (w_1 p_{i1} + \ldots + w_N\, p_{iN}) \log(w_1 p_{i1} + \ldots + w_N\, p_{iN}) \right] \quad (6.10b) $$

$$ {}^{1}D_{\beta} = {}^{1}D_{\gamma} / {}^{1}D_{\alpha} \quad (6.10c) $$

Here the weights w_j are usually taken to be the relative sizes (either in terms of area or total population) of the assemblages. This β ranges from 1 to the weight diversity, which is the exponential of the Shannon entropy of the weights themselves. It takes the latter value when the assemblages are completely different in composition (no shared species). When other measures of diversity are used with unequal weights, it may still be possible to extract information about assemblage differentiation by comparing them to more complicated reference diversities; see Gregorius (2010) for discussion.

Box 6.1 Numerical example: assessing the compositional similarity of trees and seedlings across six sites

We illustrate our multiple-assemblage similarity indices by using the frequency data collected in six forests in north-eastern Costa Rica 2006 by Chazdon and colleagues; see Norden et al. (2009) for details of the six sites. The six 1-ha plots vary in land-use history and in protection from hunting. Two plots are young secondary forests: El Bejuco (EB, 12 years old in 2006) and Juan Enriquez (JE, 12 years old in 2006). Two plots are 'intermediate' secondary forests: Lindero Sur (LSUR, 21 years old in 2006) and Lindero El Peje secondary (LEPS, 29 years old in 2006). Two plots are in old growth forests: Lindero El Peje primary (LEPP, over 200 years) and Selva Verde (SV, over 200 years). The four secondary forests were all cleared, burned, and used for pasture for several years and were subsequently abandoned. Three sites are located within La Selva Biological Station (LSUR, LEPS, and LEPP), and three sites are located 7 km to the west in Chilamate.

Each plot was 50×200 m (1 ha). For this example, we consider two size classes of canopy tree species: trees ($>$ 5 cm in DBH) including canopy palms, and canopy tree seedlings ($<$ 1 cm in DBH, but $>$ 20 cm in height). All trees were marked and measured for diameter within a 1-ha plot in each forest. Seedlings were sampled in five strips 2×200 m long, running every 10 m (0.2 ha). The number of species and individuals are summarized in Table 6.4. Note that SV old-growth forest has by far the highest seedling abundance; 72% are seedlings of a single canopy palm species, *Welfia regia*. Species richness of trees is significantly greater in the old-growth plots, but seedling species richness is similar (Norden et al. 2009; Table 6.4).

Table 6.4 The observed species richness in four secondary-growth and two old-growth sites (numbers in parentheses denote the number of individuals)

Site	Age in 2006	Seedlings	Trees
EB young	12	65 (3797)	56 (498)
JE young	12	58 (2258)	48 (546)
LSUR intermediate	21	42 (777)	51 (762)
LEPS intermediate	29	47 (1303)	79 (1093)
LEPP old-growth	$>$ 200	53 (1160)	89 (618)
SV old-growth	$>$ 200	54 (5411)	114 (832)
Average number of species		53.17	72.83
Number of species in the six plots		97	184

See Box 6.1 for explanation of EB, JE, LSUR, LEPS, LEPP, and SV.

Table 6.5 β diversities and multiple-assemblage similarity indices (with s.e. in parenthesis) across the six sites for trees and seedlings.

Index	Sizes	$q = 0$ (Presence/ absence)	$q = 1$ (Entropy based)	$q = 2$ (Simpson index based)
β diversity	Trees	2.526 (0.052)	1.982 (0.026)	1.626 (0.037)
	Seedlings	1.824 (0.034)	1.566 (0.011)	1.797 (0.019)
Multiple-assemblage similarity	Trees	0.695 (0.009)	0.618 (0.007)	0.535 (0.015)
C_{qN}	Seedlings	0.835 (0.007)	0.750 (0.004)	0.467 (0.008)

The β diversity formula is given in equation 6.9c and multiple-assemblage similarity in equation 6.3 or 6.12.

Although a specialized group of tree species colonize abandoned pastures and become dominant in early secondary forests, seedling species are more likely to be shared with old-growth forests, due to similar conditions within the shaded forest understory in all plots. We therefore predicted that multiple-assemblage similarity would be lower for trees than for seedlings across the six forest plots.

Here, we assess the compositional similarity indices of trees across the six plots. We then compare the similarity indices for seedlings across the six plots. In Table 6.5 we present the β diversities and multiple-assemblage similarity indices of orders 0, 1, and 2 along with their bootstrapped s.e. for trees and seedlings. We use the observed species and abundance to compute the measures for $q = 0$ and 1 because for these data the under-sampling bias is relatively low so that it would not affect the general 'relative' pattern. For the measures of $q = 2$, we use the nearly unbiased estimators (Chao et al. 2008).

Based on Table 6.5 and the multiple-assemblage similarity profile in Figure 6.1A, for $q = 0$ and 1, the cross-site similarity of seedlings is larger than the cross-site similarity of trees. However for $q \geq 2$, the result is reversed. This is because abundant species strongly influence the assemblage similarity index for $q \geq 2$. This pattern is due to a single species that is superabundant in the SV old-growth forest, *Welfia regia*. When this species is excluded in the comparison, it is clear (Figure 6.1B) the cross-site similarity of seedlings is always larger than the cross-site similarity of trees for all orders. The above

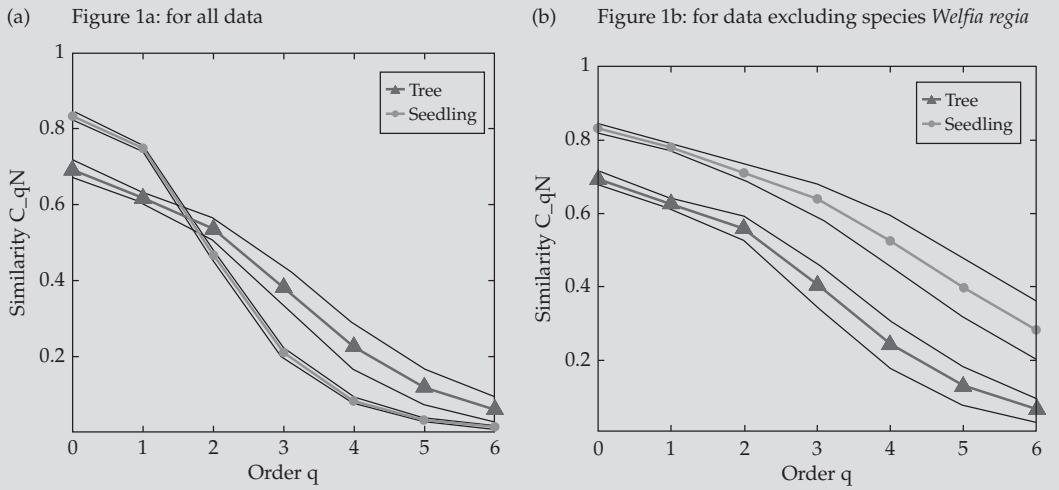

Figure 6.1 The multiple-assemblage similarity profile C_{qN} (for $q = 0 - 6$) and the associated 95% confidence interval: *x*-axis, order *q*; *y*-axis, similarity index C_{qN}. (a) For all data; (b) for data excluding species *Welfia regia*.

pattern is also generally valid for any pair of plots for the three orders.

The results show that, among these six plots, β diversity of trees is higher than for tree seedlings. This pattern reflects the ecological specialization of common trees in secondary vs old-growth forests. In contrast, tree seedlings established in both secondary and old-growth forest are more evenly mixed across the plots and do not show evidence of ecological segregation or differentiation. In the secondary forests, many of these seedlings are colonizing these forests for the first time (Norden et al. 2009). Gradually, these changes in species composition will lead to differentiation and reassembly of tree species composition as the shade-tolerant species become recruited in larger size classes (Chao et al. 2005). Since the under-sampling bias for the estimates of the measure of order 0 (only presence/absence data are used) generally cannot be corrected and a measure based on abundance data is more informative, it is suggested that the multiple-assemblage similarity index, particularly for measures of order $q = 1$ and $q = 2$ based on abundance data, provides a sensitive and statistically robust metric for comparing compositional similarity of tree and seedling assemblages in forests at different stages of succession (Chao et al. 2008).

The complete partitioning lets us evaluate the multiplicative and additive approaches. Whittaker's multiplicative approach gives the complete partitioning of all effective numbers of species. The additive approach gives the complete partitioning of Shannon entropy but not species richness or the Gini–Simpson index. The additive species richness β is still a useful measure, as shown below, but the additive Gini–Simpson β is confounded with α in a particularly inappropriate way. When α is high, this β necessarily approaches zero, even if all groups are completely differentiated (Jost 2006, 2007, 2008; Jost et al. 2010; Tuomisto 2010). One consequence of this is that additive β using the Gini–Simpson index is not monotonic with respect to unambiguously increasing differentiation (Jost et al. 2010). It cannot be directly interpreted as a measure of differentiation for a given set of assemblages. It needs to be properly normalized before it can be interpreted (Jost et al. 2010). Hierarchical additive partitioning of this measure gives misleading results as well, since α generally increases at higher hierarchical levels, limiting this β to ever-smaller values as we go up the hierarchical pyramid.

Using transformations of β to measure compositional similarity

β diversity based on complete partitioning is independent of α. However, we may also be interested in measures of the number of species (or effective

number of species) by which the assemblages differ; this kind of measure necessarily depends on α. For example, if assemblage weights are taken to be equal, the effective number of regional species not contained in a typical local assemblage is given by ${}^{q}D_{\gamma} - {}^{q}D_{\alpha} = {}^{q}D_{\alpha}({}^{q}D_{\beta} - 1)$. This is essentially additive partitioning of effective number of species (but taking care to use equal weights when q is different from 0 or 1, and making sure that ${}^{q}D_{\alpha}$ is calculated as above). Economo and Kiett (2008) used this partitioning to investigate the behaviour of Hubbell's neutral theory of biodiversity. Dividing this by $N - 1$ gives the effective number of species unique to a typical local assemblage. This is given by ${}^{q}D_{\alpha}({}^{q}D_{\beta} - 1)/(N - 1)$. These measures should be interpreted with care when comparing regions, since they depend on α diversity.

The β diversities for effective number of species ranges from 1 to N when weights are equal. They can be made into measures of relative compositional similarity by transforming them onto the unit interval. All such transformations inherit the important mathematical properties of this β (independence from α, density invariance, replication invariance, and others). Similarity measures that are transformations of β are not limited to two assemblages. There are many possible transformations of β onto the unit interval, each addressing a different aspect of compositional similarity (Jost 2007; Chao et al. 2008; Tuomisto 2010). A transformation that is linear in the proportion of regional diversity contained in the average assemblage is

$$ {}^{q}S = (1/{}^{q}D_{\beta} - 1/N)/(1 - 1/N) \tag{6.11} $$

(Jost 2007). When $q = 0$, this is the multiple-assemblage generalization of the Jaccard index. When $q = 2$ this is the multiple-community generalization of the Morisita–Horn index. The Jaccard and Morisita–Horn measures are members of the same family and are connected by a continuum of similarity measures which differ only in their sensitivity to species relative abundances.

The overlap measure C_{qN} discussed earlier is also a transformation of true β diversity:

$$ C_{qN} = [(1/{}^{q}D_{\beta})^{q-1} - (1/N)^{q-1}]/[1 - (1/N)^{q-1}] \tag{6.12} $$

This family of measures includes the multiple-assemblage generalizations of the Sørensen index, Horn index, and Morisita–Horn index. All three of these standard indices are measuring the same aspect of assemblage similarity, only differing in the weights they give to species frequencies. The Morisita–Horn index is unique in belonging to both families of similarity measures; it is the both linear in shared diversity and linear in the degree of overlap. It therefore has a wider range of interpretations than other similarity measures. It can be useful to transform the β in equation 6.10c to a normalized measure between 0 and 1, so that we can compare compositional similarity across several sets of assemblages. This transformation generalizes Horn's overlap measure (equation 6.4) to the unequal-weight case:

$$ S_{H\_w} = \frac{1}{-\sum_{j=1}^{N} w_j \log w_j} \sum_{j=1}^{N} \left\{ \sum_{i=1}^{S} (w_j p_{ij}) \log \left(1 + \sum_{k \neq j}^{N} w_k p_{ik} / w_j p_{ij} \right) \right\} \tag{6.13} $$

We recommend that investigators calculate at least C_{0N}, C_{1N}, and C_{2N} for their data (Box 6.1). They could also use equation 6.12 to graph a continuous similarity profile for a range of q (say, from $q = 0$ to $q = N$). As an example of the use of C_{qN}, consider the assemblage sets A–D in Table 6.3. In Set A, all species in each assemblage are equally abundant, thus the three similarity indices are about the same $C_{0N} = C_{2N} \approx C_{1N}$. In Set B, $C_{2N} > C_{1N}$, implying that the dominant species are more similar than the average species. In contrast, for Sets C and D, $C_{2N} < C_{1N}$, implying that the dominant species are less similar than the average species. C_{2N}, like the Simpson index it is based on, is more sensitive to the dominant species than C_{1N}. We can make comparisons across the four sets because all measures are normalized to the range of 0 and 1. The three similarity indices in both Sets A and B are consistently higher than those in Sets C and D. Based on the Horn index, Set D has the lowest similarity. The proposed indices C_{0N}, C_{1N}, and C_{2N} are featured in the Program SPADE (Chao & Shen 2003a), which can be freely downloaded from http://chao.stat.nthu.edu.tw/softwareCE.html.

6.2.3 Statistical estimation of assemblage differentiation and similarity

In practice, all differentiation and similarity measures need to be estimated from a relatively small sample taken from each of the study assemblages. Traditionally, species frequencies in the samples are assumed to give good estimates of the species frequencies in the assemblage itself. The sample species frequencies are therefore substituted for the assemblage species frequencies in the theoretical formulas for differentiation and similarity. The resulting estimate is called the maximum likelihood estimator (MLE). This approach is statistically valid only when sample size is relatively large compared to the number of species, so that almost all species are observed in samples. When sample size is relatively small, so that a large fraction of rare species are missed in samples, our simulations show that the MLEs for differentiation indices are generally biased upwards and thus biased downwards for similarity measures. Sampling limitations create challenges for making accurate estimates of differentiation and similarity measures. Moreover, any bias-reduction method valid for estimating α diversity in one assemblage cannot be directly extended to estimate γ diversity. This is because the abundance vectors cannot be simply pooled across assemblages unless the sampling efforts are equal in all assemblages, so that sample sizes reflect assemblage sizes.

The magnitude of the under-sampling bias depends on the diversity order q. For $q \geq 2$, a nearly unbiased estimator exists for any measure based on the Simpson index. Chao et al. (2008) provided such an estimator for the index C_{qN}. Similar estimators can be constructed for β diversity or differentiation measures. For $q = 1$, there exists no nearly unbiased estimator for any measure based on Shannon's entropy. The unseen species have moderate effect on the index. A two-sample jackknife method (Schechtman & Wang 2004) that removes most of the under-sampling bias of the MLE has been applied to Horn's heterogeneity and overlap indices (Norden et al. 2009; Goßner et al. 2009). A coverage-based Horvitz–Thompson approach, similar to that proposed by (Chao and Shen 2003b) for estimating entropy in one assemblage, is currently under development for multiple-assemblage measures. For $q = 0$, the under-sampling bias could be substantial, especially for highly diverse assemblages. Only an accurate lower bound for species richness (or shared species richness) is feasible (Chao 2005; Pan et al. 2009).

Regardless of how the similarity or differentiation is estimated, the result will always have some statistical uncertainty. Ecologists traditionally quantify this uncertainty by testing their results against the null hypothesis that there is no differentiation, and reporting a P value calculated by randomized resampling. This is rarely an appropriate approach, since the question of interest is not 'Is there any differentiation?'(there will always be some differentiation among natural populations) but rather 'How much differentiation is there?' The problem is one of parameter estimation rather than hypothesis testing. In parameter estimation, statistical uncertainty is expressed by means of a confidence interval. If sample size is sufficient, bootstrapping can be used to calculate an approximate confidence interval around the estimated value of the measure (Chao et al. 2008). If the confidence interval includes 0 (for relative differentiation) or 1 (for relative similarity), then the observations are insufficient to reject the null hypothesis of zero differentiation. If the interval does not include 0, not only can the investigator reject the null hypothesis but he or she will also know the range of magnitudes within which the true population value of the measure lies. This magnitude is the scientifically important aspect of differentiation or similarity. The magnitude of good differentiation and similarity measures can be understood by referring to simpler reference assemblages with the same value of the measure.

6.3 Prospectus

Exciting advances in the mathematics of diversity and differentiation have been made within the fields of genetics, economics, information science, and physics, but these advances are largely ignored outside of the disciplines that generated the advance. In addition the extensive statistical literature on divergence measures (see Pardo 2006)

has yet to be tapped by ecologists. Interdisciplinary synthesis of these developments holds great potential. It appears that all disciplines are independently converging on the same concepts, and we would get there faster if each science was aware of the progress made by the other sciences. Order 1 measures (those based on Shannon entropy) have many important mathematical properties not shared by other frequency-sensitive diversity measures, and come closest to the ecologists' intuitive idea of diversity and differentiation (Jost 2007; Jost et al. 2010). If ecologists or population geneticists can derive the expected behaviour of these measures under a neutral model, following the lead of Sherwin et al. (2006), these measures would become the standard in the field.

Standard similarity measures treat all species as equally distinct. Yet in many important ways, an assemblage of rats and an assemblage of mice are more similar than an assemblage of rats and an assemblage of whales. Some recently developed measures take into account functional, phylogenetic, or other kinds of species differences (Graham and Fine 2008; Allen et al. 2009; Cadotte et al. 2009; Pavoine et al. 2009). However, these measures are usually based on additive partitioning of a generalization of the Gini–Simpson index (e.g. Hardy and Senterre 2007). The resulting differentiation measure is not acceptable (Table 6.2, see also Jost (2008) and Hardy and Jost (2008)). Ricotta and Szeidl (2009) have recently extended the diversity of order 2 to take into account phylogenetic or functional differences using multiplicative partitioning. It is important to extend this measure to other orders, and resolve issues relating to the definition of α diversity.

6.4 Key points

1. The general overlap measure C_{qN} is the best overall measure of relative compositional similarity. It includes the Sørensen, Horn, and Morisita–Horn indices and their multiple-assemblage generalizations, which take global similarity into account.
2. The Morisita–Horn index is especially useful as an abundance-based similarity measure, since its expected value can be predicted from the parameters of Hubbell's neutral model.
3. Some common similarity and differentiation measures (traditional additive β and similarity measures, some distance measures, the Bray–Curtis measure, and between-group component of variance) do not behave as expected and do not necessarily reflect compositional similarity or differentiation.
4. Assemblage differentiation is intimately related to the concept of β diversity. Multiplicative partitioning of effective number of species best captures the notion of β diversity. It can be transformed into the Jaccard and C_{qN} similarity indices, among others.
5. Estimation of differentiation and similarity is statistically challenging, especially for measures based on species richness. Specialized estimators should be used.