

## Dissimilarity measurements and the size structure of ecological communities

Miquel De Cáceres<sup>1,2\*</sup>, Pierre Legendre<sup>3</sup> and Fangliang He<sup>4,5</sup>

<sup>1</sup>CTFC (Forest Science Center of Catalonia), Ctra. St. Llorenç de Morunys km 2, E-25280 Solsona, Catalonia, Spain; <sup>2</sup>CREAF (Centre for Ecological Research and Applied Forestries), Autonomous University of Barcelona, Bellaterra, Catalonia, Spain; <sup>3</sup>Département de Sciences Biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, QC H3C 3J7, Canada; <sup>4</sup>State Key Laboratory of Biocontrol, SYSU-Alberta Joint Lab for Biodiversity Conservation and School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, China; and <sup>5</sup>Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2H1, Canada

### Summary

1. Measurements of community resemblance in ecology are often based on species composition, and the starting point for calculations is usually a site-by-species data table. However, resemblance measurements may not be sufficiently accurate when communities are described using species composition only. Characteristics such as the size of their constituting organisms are also important to understand community organization.

2. Here, we provide a framework that generalizes conventional resemblance measurements by incorporating the size structure of the compared communities. We first introduce the concept of cumulative abundance profile, which generalizes traditional species abundance values, and describe how to calculate it. We then explain our approach to compare cumulative abundance profiles in community resemblance measurements and use a small simulation study to determine which resemblance coefficients appropriately deal with compositional and structural differences. After that, we present an illustrative example where we study the structural and compositional variation between and within six Douglas-fir forest plots in British Columbia, Canada.

3. According to our investigations, the generalizations we suggest for the percentage difference (alias Bray–Curtis dissimilarity) and the Ružička coefficients are appropriate to measure community resemblance in terms of size structure, species composition or both.

4. Our framework allows community resemblance to be measured in terms of either size structure or species composition, or both. A broad range of applications is expected. In the case of terrestrial plant communities, potential applications include analyses of community dynamics and classification of vegetation.

**Key-words:** beta diversity, community ecology, cumulative abundance profile, dissimilarity coefficients, forest dynamics, size structure

### Introduction

The notion of community resemblance is central in ecology. Ecologists determine routinely the similarity or dissimilarity between pairs of communities with the aim of quantifying the amount of community change along time (e.g. before and after a disturbance; Philippi, Dixon & Taylor 1998), across space (e.g. to estimate beta diversity; Anderson *et al.* 2011; Legendre & De Cáceres 2013) or due to experimental treatments. Since the beginning of the twentieth century, many dissimilarity/similarity coefficients have been proposed to measure the resemblance between communities (Koleff, Gaston & Lennon 2003; Legendre & Legendre 2012). Community resemblance is almost always assessed on the basis of species composition data in the form of a site-by-species data table. In some cases, this table simply contains binary data describing species incidence (i.e. presence or absence), whereas in other cases, it

contains species abundance values (e.g. counts, cover, biomass or some other measure of relative or absolute importance). Although data on species composition are fundamentally important for describing communities, composition alone may be insufficient because communities that are similar in composition may differ in other characteristics such as the size of organisms and *vice versa*. As a result, the organization of communities may be oversimplified if represented by species composition alone. To more accurately describe community organization and the variation of communities across space or along time, it is necessary to generalize the conventional approach to community resemblance by incorporating structural data describing the size of constituent organisms in addition to compositional data. Incorporating structural data in resemblance assessments would allow the analyst to exploit valuable information obtained during field surveys (e.g. Holopainen & Kalliovirta 2006). If available, there is no reason to ignore the wealth of information regarding size when measuring community resemblance.

\*Correspondence author. E-mail: miquelcaceres@gmail.com

In this paper, we present a general framework to measure community resemblance in terms of composition, size structure or taking both elements into account. In the next sections, we first introduce the concept of cumulative abundance profile (CAP), which generalizes traditional species abundance values and allows one to describe the structural component of community organization. After explaining our approach to compare CAPs, we present six dissimilarity coefficients that can deal with CAPs and are generalizations of well-known compositional indices. We then use synthetic community data to evaluate the ability of these coefficients to appropriately measure community resemblance along simulated compositional and structural gradients. An example with real data is then presented where we study the structural and compositional variation between and within Douglas-fir forest plots in British Columbia, Canada. Finally, we discuss the advantages and limitations of the proposed framework and suggest potential applications.

### The cumulative abundance profile

Ecological communities can have similar *species composition* (i.e. similar species abundance values) and differ at the same time in *size structure* (i.e. distribution of individual sizes). One might consider that a proper way to accommodate the size of organisms into resemblance indices is to define abundance, so that the value of each species is an indication of its overall biomass in the sampled community. However, this strategy would not indicate whether the biomass value for a given species is the biomass value of a single very large individual or the sum of biomass values of several small ones. A vector where each species has a single abundance value is not sufficient to describe a community in terms of both its size structure and species composition. The size structure for the population of each species in the community will be available if the species identity and the size of each individual were determined during survey. In sessile communities, such as coral reefs and forests, the same information will also be available if abundance values (e.g. cover) are estimated for a set of predefined vertical strata. Although in this second case, the individuals constituting the community may not be identified, and a (simplified) description of the size structure is also obtained.

To determine the distribution of sizes one needs first to choose which structural variable is used to represent the size of organisms. For example, in plant communities, the most natural choice for structural variable is plant height, but other structural variables, such as the trunk diameter, may be used instead. In the description of our approach, we will use the term *size* for the sake of generality, but readers may envision particular variables. The only restriction we impose to the structural variable is that it cannot be negative. Similarly, we do not constrain the definition of *abundance*, as long as its value is non-negative. With the previous considerations in mind, we define the *cumulative abundance profile* (CAP) to be a function that takes a value of *size* as input and returns the *cumulative abundance of organisms whose size is equal to or larger than the input value*. For example, if height is chosen to

be the structural variable, the function will return the cumulative abundance of organisms as tall as or taller than the given height. In other words, an organism of height  $h$  will contribute to the cumulative abundance values at  $h$  and also at lower heights. The cumulative abundance profile is maximal for value zero of the structural variable and is a non-increasing continuous function.

The actual calculation of CAP values varies depending on the format of the community data. Let us consider three cases, which illustrate how different ways of describing the size structure can be accommodated to the CAP framework:

#### CALCULATION OF CAP FROM STRATIFIED COMMUNITY DATA

Consider the case where the size of organisms has been simplified into  $s$ -ordered size classes (or vertical strata), and the community has been sampled by assessing the abundance of each species within each size class  $t$  ( $t = 1, \dots, s$ ). The class index can be readily used as the structural variable, provided that classes have been ordered from small to large. In this case, the CAP can be represented as a vector of  $s$  values and the value for a given class  $t$  is the sum of abundances of the target species in all classes  $u \geq t$ :

$$\text{CAP}(t) = \sum_{u=t}^s x(u) \quad \text{eqn 1}$$

where  $x(u)$  is the recorded abundance of the target species in size class  $u$ .

#### CALCULATION OF CAP FROM INDIVIDUAL DATA

Cumulative abundance profiles may be precisely calculated if the value of the structural variable is available for every individual  $i$  of the target species. The CAP value for the target species at a given size  $h$  is the cumulative abundance of all individuals of that species with size  $h$  or larger:

$$\text{CAP}(h) = \sum_i a_i \cdot I(h_i \geq h) \quad \text{eqn 2}$$

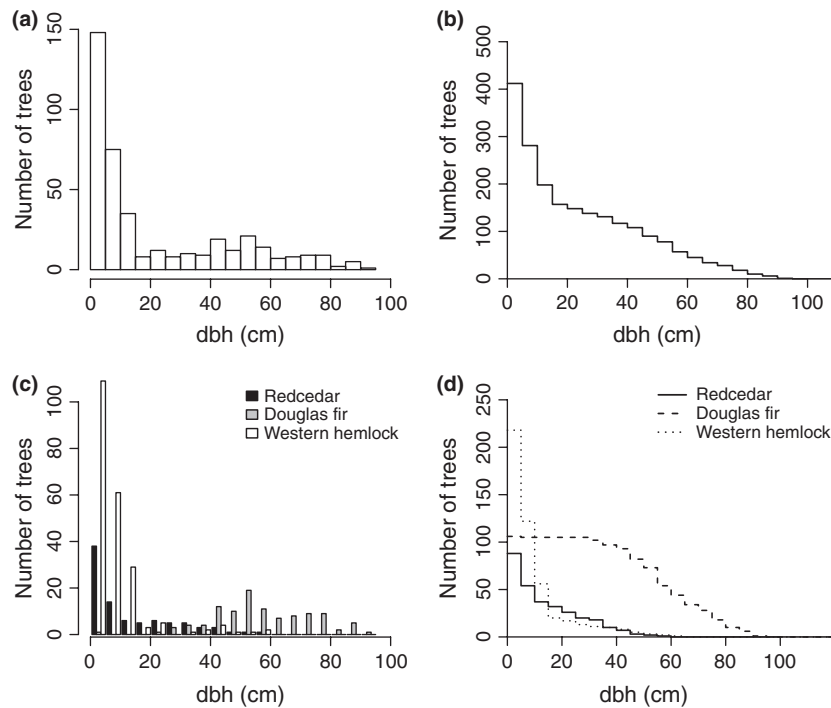
where  $I(h_i \geq h)$  is an indicator variable equal to one for individuals of size  $h$  or larger and zero otherwise. The interpretation of  $a_i$  values will depend on the definition chosen for *abundance*. If abundance is defined as ‘number of individuals’, then  $a_i = 1$  for all individuals. On the other hand, if abundance is defined as ‘cover’, then  $a_i$  will be the contribution of individual  $i$  (e.g. its crown cover) to the cumulative cover value.

#### CALCULATION OF CAP FROM A CONTINUOUS SIZE DISTRIBUTION

There is no need to limit the calculation of CAP to discrete data; it can be computed using a continuous size distribution for  $h$ :

$$\text{CAP}(h) = K \cdot \int_{l=h}^{\infty} f(l) \, dl \quad \text{eqn 3}$$

Here,  $f(l)$  is the probability density function of individuals of size  $l$ , and  $K$  is a constant used to translate cumulative



**Fig. 1.** Example of cumulative abundance profiles calculated for an old-growth Douglas-fir forest plot in British Columbia, Canada (see *Example of application* section): (a) histogram depicting the distribution of diameters at breast height (dbh) in 5 cm classes; (b) cumulative abundance profile (CAP) using dbh as the structural variable: for each diameter class, we count the number of individuals with that diameter or larger; (c) dbh histogram for each tree species separately; (d) CAP calculated for each tree species separately.

probability values into abundance units (e.g.  $K$  could be the total number of individuals in the community).

One can calculate CAPs for each species separately or for the entire community without taking species identity into account, that is, after pooling all species. As an example of calculation, we show in Fig. 1 the distribution of diameter at breast height (dbh) in 5-cm classes for trees in a Douglas-fir forest plot and the resulting CAPs calculated with and without species identity.

## Measuring community resemblance in terms of composition and structure

### COMPARISON OF CUMULATIVE ABUNDANCE PROFILES

The basis of our community resemblance framework consists in replacing species abundance values by CAPs. More specifically, we suggest that the comparison of two communities should involve the comparison of pairs of CAPs (one pair per species) instead of comparing pairs of species abundance values. Moreover, the comparison of the two CAPs for a given species  $j$  –  $CAP_{1j}$  and  $CAP_{2j}$  – should be done by integrating the comparison of cumulative abundance values along the values of the structural variable. This approach leads to distinguishing between the following:  $A_j$ , the area(s) where the two profiles overlap;  $B_j$ , the area(s) where  $CAP_{1j}$  exceeds  $CAP_{2j}$ ; and  $C_j$ , the area(s) where  $CAP_{2j}$  exceeds  $CAP_{1j}$  (see Fig. 2a). More formally:

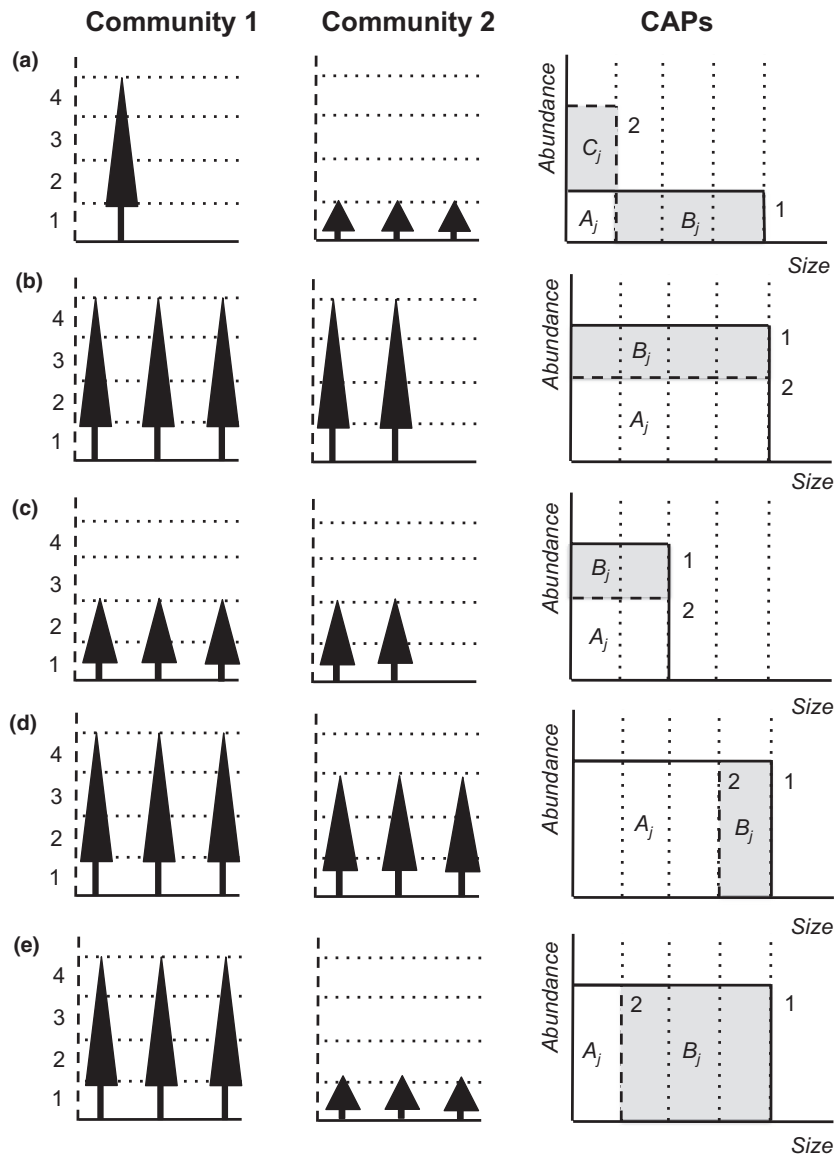
$$\begin{aligned} A_j &= \int_{h=0}^{\infty} \min(CAP_{1j}(h), CAP_{2j}(h)) \cdot dh \\ B_j &= \int_{h=0}^{\infty} (CAP_{1j}(h) - \min(CAP_{1j}(h), CAP_{2j}(h))) \cdot dh \\ C_j &= \int_{h=0}^{\infty} (CAP_{2j}(h) - \min(CAP_{1j}(h), CAP_{2j}(h))) \cdot dh \end{aligned} \quad \text{eqn 4}$$

If CAPs are defined over  $s$  discrete size classes, each of width  $w(t)$ , quantities  $A_j$ ,  $B_j$  and  $C_j$  can be obtained using:

$$\begin{aligned} A_j &= \sum_{t=1}^s w(t) \cdot \min(CAP_{1j}(t), CAP_{2j}(t)) \\ B_j &= \sum_{t=1}^s w(t) \cdot (CAP_{1j}(t) - \min(CAP_{1j}(t), CAP_{2j}(t))) \\ C_j &= \sum_{t=1}^s w(t) \cdot (CAP_{2j}(t) - \min(CAP_{1j}(t), CAP_{2j}(t))) \end{aligned} \quad \text{eqn 5}$$

This decomposition of community resemblance into agreement ( $A_j$ ) and disagreement ( $B_j$  and  $C_j$ ) areas for a given species is analogous to the decomposition proposed by Tamás, Podani & Csontos (2001) for compositional resemblance coefficients.

Why should we compare *cumulative* abundance profiles? Our main motivation for comparing CAPs was that larger individuals often have larger impact on the organization, function and dynamics of communities than smaller ones. Accordingly, in our approach, the amount of agreement and disagreement is greater when a given change in abundance concerns large individuals compared to when it concerns small individuals (compare areas  $A_j$  and  $B_j$  in Fig. 2b–c). On the



**Fig. 2.** Pairs of hypothetical communities (a–e) used to illustrate the three areas issued from the comparison of two cumulative abundance profiles,  $CAP_{1j}$  (continuous line) and  $CAP_{2j}$  (dashed line), for a given species  $j$ :  $A_j$  is the area where the two profiles overlap, whereas  $B_j$  and  $C_j$  are the areas where  $CAP_{1j}$  exceeds  $CAP_{2j}$  and  $CAP_{2j}$  exceeds  $CAP_{1j}$ , respectively.

other hand, if the two communities compared have the same abundance value for a given species but differ in the size of organisms, our approach will yield a larger disagreement if the differences in size are larger (compare  $B_j$  in Fig. 2d–e).

#### RESEMBLANCE COEFFICIENTS FOR CAP COMPARISON

Before choosing a coefficient of compositional resemblance, it is important to discuss which mathematical and ecological properties are deemed essential for the question at hand (Hajdu 1981; Faith, Minchin & Belbin 1987; Tamás, Podani & Csontos 2001; Jost, Chao & Chazdon 2011; Beck, Holloway & Schwanghart 2013; Legendre & De Cáceres 2013). In line with this, we require here that a given coefficient needs to be function of areas  $A_j$ ,  $B_j$  and  $C_j$  for every species  $j$  – that is,  $A_1, A_2, \dots, A_p, B_1, B_2, \dots, B_p, C_1, C_2, \dots, C_p$  – (Property P1) to be

considered appropriate for measuring the community resemblance in terms of both structure and composition. In addition, if CAPs are defined for discrete size classes (or vertical strata), we require that *resemblance values should not change with the subdivision of classes into subclasses, provided that the CAPs do not change* (Property P2). Otherwise, there could be an artificial increase or decrease in resemblance derived from arbitrary decisions about the resolution of size classes. This second property is analogous to a property of compositional resemblance coefficients called *species replication invariance* (Jost, Chao & Chazdon 2011; Legendre & De Cáceres 2013).

Following the above requirements, we present in Table 1 six dissimilarity coefficients of compositional resemblance for which we could derive a suitable generalization to measure dissimilarity in both composition and structure: Whittaker's index of association (also known as relativized Manhattan)

(Whittaker 1952; Faith, Minchin & Belbin 1987), the modified Canberra metric (Lance & Willams 1967; Stephenson, Williams & Cook 1972), the percentage difference (Odum 1950) [alias Bray-Curtis (see historical note in Legendre & Legendre (2012, index, p. 311) about this)], the Ružička (1958) index (a generalization of Jaccard's binary similarity coefficient), the quantitative symmetric index of Kulczynski (1928) and a generalization of Ochiai's (1957) binary similarity coefficient to quantitative data that, to our knowledge, has never been explicitly described before, but follows the generalization scheme of Tamás, Podani & Csontos (2001). This generalization of Ochiai's index is different from other generalizations like the one suggested by Chao *et al.* (2006) or the chord and Hellinger distances (Orlói 1967; Rao 1995), which were deemed unsuitable for the current purpose. Appendix S1 provides the mathematical proofs that these six coefficients have properties P1 and P2.

#### TRANSFORMATIONS OF CAPS

When measuring the compositional resemblance between communities, transformations such as the square root or the logarithm are commonly employed to reduce the weight of abundance with respect to species presence (Van der Maarel 1979; Legendre & Legendre 2012). Another kind of transformation involves the transformation of abundance values using community level statistics, normally with the aim of excluding differences in total abundance from the resemblance measurement (Faith, Minchin & Belbin 1987; Legendre & Gallagher 2001). Three kinds of transformations can be applied on CAPs. We describe these transformations in Appendix S2.

#### Simulation study

In order to evaluate the six generalized dissimilarity coefficients under different situations, we conducted a simulation study where we created synthetic ecological communities differing in:

- *Size structure* – We used the Gamma distribution to stochastically model the size of individuals, and we varied the shape parameter (scale was 1 in all cases) to obtain differences in size structure (treatments labelled 'a' to 'e'; Fig. 3).
- *Species composition* – We used the multinomial distribution to stochastically model species identity, and we defined a compositional gradient by setting different proportions of five species (treatments labelled 'A' to 'E'; Fig. 3).
- *Community size* – We considered the following numbers of individuals per community: 25, 50, 100, 200 and 400 (treatments labelled '1' to '5').

We designed three experiments by crossing two of the three gradients each time: *Experiment 1* – gradients in composition and size structure (and 100 individuals); *Experiment 2* – gradients in composition and number of individuals (and structural treatment 'c'); *Experiment 3* – gradients in size structure and number of individuals (and compositional treatment 'C'). For each experiment, we generated  $5 \times 5 = 25$  communities, corresponding to all combinations of gradient positions, and we labelled each community with

its combination of treatments. CAPs were calculated for each species and community using the number of individuals as abundance measure and their size as structural variable. We then calculated the dissimilarity between each pair of communities using each of the six coefficients in Table 1. Finally, we displayed each of the resulting dissimilarity matrices using non-metric multidimensional scaling (nMDS) in two dimensions. We evaluated the performance of coefficients visually by determining whether the set of communities formed a two-dimensional grid in a plane, with axes corresponding to the two simulated gradients, as in Minchin (1987).

The percentage difference (alias Bray-Curtis) and the Ružička index yielded good results under the three experiments (Fig. 4). The case of Whittaker's index of association can also be considered satisfactory because this index excludes differences in total abundance among the compared communities. The Kulczynski and Ochiai indices yielded curvilinear distortions but the results were reasonably acceptable. Finally, the Canberra index yielded strong distortions in two experiments.

#### Example of application

In order to illustrate our framework with real data, we analyse in this section the variation between and within six plots in a Douglas-fir forest located in the Greater Victoria Watershed District in southern Vancouver Island, British Columbia, Canada. These data were obtained during a chronosequence survey made by the Canadian Forest Service to study the changes caused by converting old-growth coastal temperate forests to managed forests (He & Duncan 2000; Getzin *et al.* 2006). The advantage of using our framework here is that it allows measuring community changes derived from management by focusing on compositional and structural aspects either separately or simultaneously.

The dominant species in this coastal forest are the shade-intolerant Douglas-fir [*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco], the shade-tolerant western hemlock [*Tsuga heterophylla* (Raf.) Sarg.] and the western redcedar (*Thuja plicata* Donn ex D. Don). Three of the six plots are located in the northern part of the Victoria Watershed District, whereas the other three are located in the southern part. Within each of the two areas, a distinct plot was established in *immature* (25–45 years since last management), *mature* (65–85 years) and *old-growth* (>200 years) forest stands. For our illustration, we considered that differences in topography or soil among stands (Trofymow *et al.* 1997) were of lower importance than stand age. All trees were georeferenced within the boundaries of each plot, and the diameter at breast height (dbh) of each live tree was measured. Because some of the six plots were larger than others, for our analysis, we only used the data corresponding to subplots of  $60 \times 60$  m delineated from the south-west corner of each plot. We excluded seedlings and saplings (<0.5 cm in diameter) and discarded all individuals that did not belong to the three dominant species mentioned above.

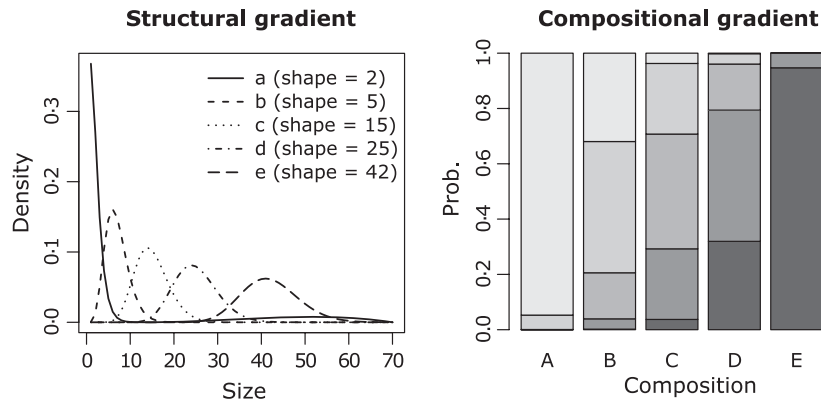
We first calculated species composition in all plots, using the number of individuals as the measure of abundance (see Table



**Table 1.** Six compositional dissimilarity coefficients generalized here to measure the dissimilarity in composition and structure. The second column presents the original formulation of the coefficients for compositional resemblance, using data from site-by-species data table  $\mathbf{X} = [x_{ij}]$  with  $n$  sites and  $p$  species ( $x_{1+}$  and  $x_{2+}$  indicate the sum of values for the corresponding rows of  $\mathbf{X}$ ;  $pp$  is the number of species that are present in both communities). The third column presents the generalized coefficients for the discrete case (where  $y_{1j}(t)$  and  $y_{2j}(t)$  indicate the CAP values for species  $j$  in size class  $t$ ;  $w(t)$  is the width of size class  $t$ ;  $y_{1+}$  and  $y_{2+}$  stand for the weighted sum of CAP values across species using  $w(t)$  as weights) and the continuous case (where  $y_{1j}(h)$  and  $y_{2j}(h)$  indicate the CAP values for species  $j$  and size  $h$ ;  $y_{1+}$  and  $y_{2+}$  stand for the integral of the corresponding CAP over size). Finally, the fourth column presents the same indices using the  $A_j B_j C_j$  notation (see text and Appendix S1)

Dissimilarity coefficient and references	Formulation for compositional data	Generalization for $s$ discrete size classes and for continuous CAP definition	Formulation using $A_j B_j C_j$ notation (see text)
Whittaker's index of association (Whittaker 1952)	$\frac{1}{pp} \sum_{j=1}^p \left  \frac{x_{1j} - x_{2j}}{x_{1+}} - \frac{x_{2j}}{x_{2+}} \right $	$\frac{1}{2} \sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \left  \frac{y_{1j}(t)}{y_{1+}} - \frac{y_{2j}(t)}{y_{2+}} \right  \frac{1}{2} \sum_{j=1}^p \sum_{t=1}^s \left  \frac{y_{1j}(h)}{y_{1+}} - \frac{y_{2j}(h)}{y_{2+}} \right  dh$	$\sum_{j=1}^p (B_j + C_j)^{\dagger}$
Canberra metric (Lance & Williams 1967; Stephenson, Williams & Cook 1972)	$\frac{1}{pp} \sum_{j=1}^p \frac{ x_{1j} - x_{2j} }{(x_{1j} + x_{2j})}$	$\frac{1}{pp} \sum_{j=1}^p \left[ \frac{\sum_{t=1}^s w(t) \cdot  y_{1j}(t) - y_{2j}(t) }{\sum_{t=1}^s w(t) \cdot (y_{1j}(t) + y_{2j}(t))} \right] \frac{1}{pp} \sum_{j=1}^p \left[ \frac{\int_{h=0}^{\infty}  y_{1j}(h) - y_{2j}(h)  dh}{\int_{h=0}^{\infty} (y_{1j}(h) + y_{2j}(h)) dh} \right]$	$\frac{1}{pp} \sum_{j=1}^p \left[ \frac{B_j + C_j}{2A_j + B_j + C_j} \right]$
Percentage difference ( <i>alias</i> Bray-Curtis) (Odum 1950)	$1 - \left[ \frac{2 \sum_{j=1}^p \min(x_{1j}, x_{2j})}{x_{1+} + x_{2+}} \right]$	$1 - \left[ \frac{2 \sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \min(y_{1j}(t), y_{2j}(t))}{y_{1+} + y_{2+}} \right] 1 - \left[ \frac{2 \sum_{j=1}^p \int_{h=0}^{\infty} \min(y_{1j}(h), y_{2j}(h)) dh}{y_{1+} + y_{2+}} \right]$	$1 - \left[ \frac{\sum_{j=1}^p 2A_j}{\sum_{j=1}^p (2A_j + B_j + C_j)} \right]$
Ruzicka index (Ruzicka 1958)	$1 - \left[ \frac{\sum_{j=1}^p \min(x_{1j}, x_{2j})}{\sum_{j=1}^p \max(x_{1j}, x_{2j})} \right]$	$1 - \left[ \frac{\sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \min(y_{1j}(t), y_{2j}(t))}{\sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \max(y_{1j}(t), y_{2j}(t))} \right] 1 - \left[ \frac{\sum_{j=1}^p \int_{h=0}^{\infty} \min(y_{1j}(h), y_{2j}(h)) dh}{\sum_{j=1}^p \int_{h=0}^{\infty} \max(y_{1j}(h), y_{2j}(h)) dh} \right]$	$1 - \left[ \frac{\sum_{j=1}^p A_j}{\sum_{j=1}^p (A_j + B_j + C_j)} \right]$
Kulezynski index (Kulezynski 1928)	$1 - \left[ \frac{\sum_{j=1}^p \min(x_{1j}, x_{2j})}{y_{1+}} + \frac{\sum_{j=1}^p \min(x_{1j}, x_{2j})}{y_{2+}} \right]$	$1 - \left[ \frac{\sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \min(y_{1j}(t), y_{2j}(t))}{y_{1+}} + \frac{\sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \min(y_{1j}(t), y_{2j}(t))}{y_{2+}} \right]$ $1 - \left[ \frac{\sum_{j=1}^p \int_{h=0}^{\infty} \min(y_{1j}(h), y_{2j}(h)) dh}{y_{1+}} + \frac{\sum_{j=1}^p \int_{h=0}^{\infty} \min(y_{1j}(h), y_{2j}(h)) dh}{y_{2+}} \right]$	$1 - \frac{1}{2} \left[ \frac{\sum_{j=1}^p A_j}{\sum_{j=1}^p (A_j + B_j)} + \frac{\sum_{j=1}^p A_j}{\sum_{j=1}^p (A_j + C_j)} \right]$
Generalized Ochiai index	$1 - \left[ \frac{\sum_{j=1}^p \min(x_{1j}, x_{2j})}{\sqrt{x_{1+} \cdot x_{2+}}} \right]$	$1 - \left[ \frac{\sum_{j=1}^p \sum_{t=1}^s w(t) \cdot \min(y_{1j}(t), y_{2j}(t))}{\sqrt{y_{1+} \cdot y_{2+}}} \right] 1 - \left[ \frac{\sum_{j=1}^p \int_{h=0}^{\infty} \min(y_{1j}(h), y_{2j}(h)) dh}{\sqrt{y_{1+} \cdot y_{2+}}} \right]$	$1 - \left[ \frac{\sum_{j=1}^p A_j}{\sqrt{\left( \sum_{j=1}^p (A_j + B_j) \right) \left( \sum_{j=1}^p (A_j + C_j) \right)}} \right]$

<sup>†</sup>This formula is correct only for the CAPs obtained after dividing by  $y_{1+}$  and  $y_{2+}$ , respectively (see Appendices S1 and S2).



**Fig. 3.** Size structure treatments (a–e in the left panel; shape indicates the parameter of a gamma distribution) and species composition treatments (A–E in the right panel) used in the three experiments of our study with simulated community data.

S3.1 in Appendix S3). Abundance values were log-transformed to decrease the importance of large numbers of individuals, producing matrix  $\mathbf{X}^{\text{COMP}}$ . We then calculated CAPs either disregarding species identity (matrix  $\mathbf{Y}^{\text{STR}}$ ) or considering it (matrix  $\mathbf{Y}^{\text{COMP-STR}}$ ). In both cases, we also used the number of individuals as the measure of abundance, and we used 5-cm-diameter classes as the structural variable. Cumulative abundance values were also log-transformed (see Figs S3.1 and S3.2 in Appendix S3).

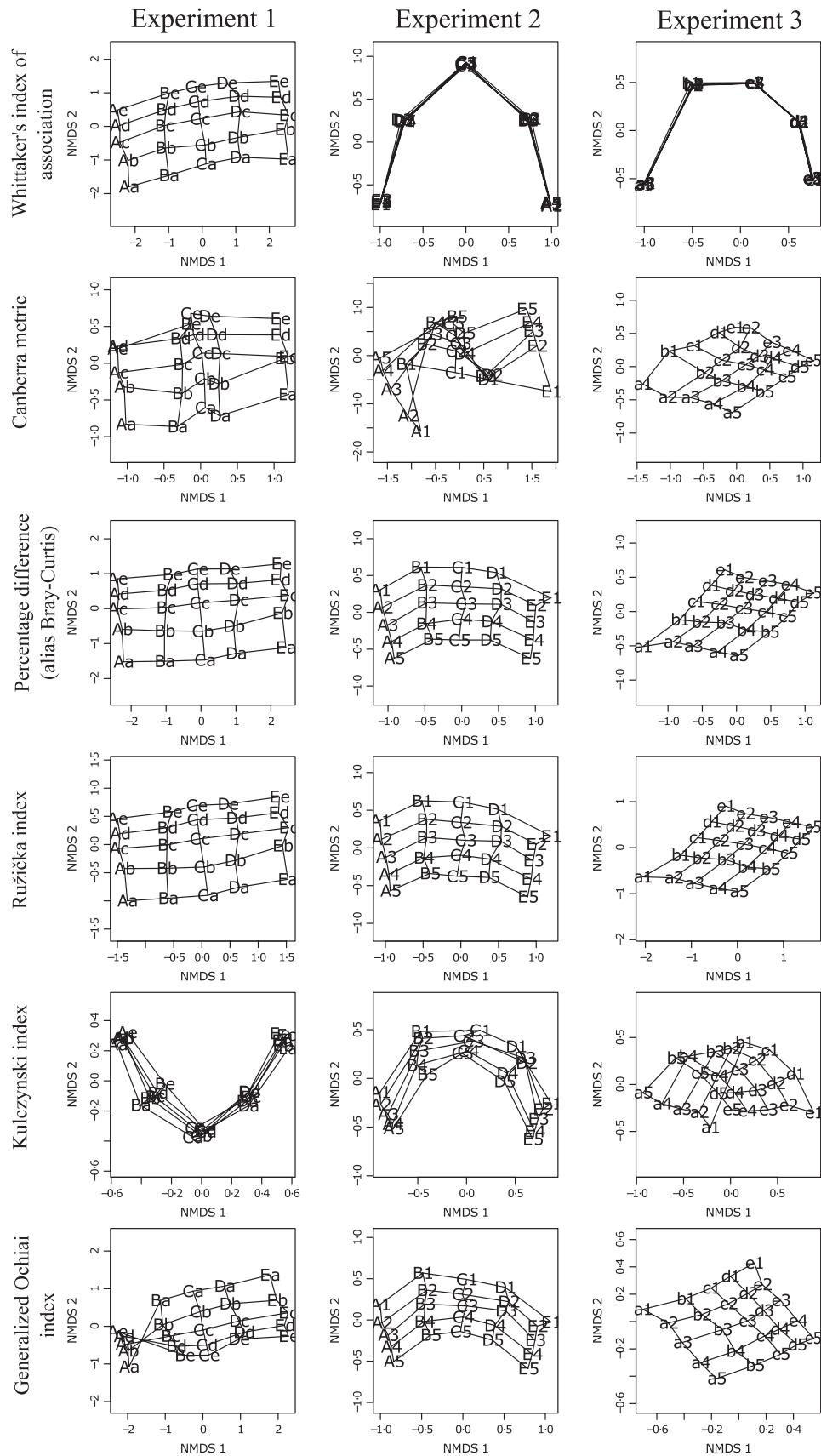
In each of the two sampling areas (north and south) separately, we determined the dissimilarity between immature, mature and old-growth stands by calculating the percentage difference (*alias* Bray–Curtis) coefficient on matrices  $\mathbf{X}^{\text{COMP}}$ ,  $\mathbf{Y}^{\text{STR}}$  and  $\mathbf{Y}^{\text{COMP-STR}}$ . The dissimilarity values obtained in each case are shown in Table 2. Compositional dissimilarity values were larger between plots in the north than between plots in the south (Table 2a). Dissimilarities assessing differences in structure but not in composition were again larger between plots of the northern area than between plots of the southern area (Table 2b). Nevertheless, for plots in the southern area, structural dissimilarities were comparatively larger than the corresponding dissimilarities in composition. When accounting for differences in composition and structure simultaneously, dissimilarity values were also larger between plots of the northern area than between plots of the southern area (Table 2c). Dissimilarity values in composition and structure were always larger than the corresponding dissimilarities in either composition or structure alone.

After comparing plots to each other, we wondered which forest stands were more structurally and/or compositionally heterogeneous and whether within-plot variability changed with stand age. To address these questions, we repeated our dissimilarity calculations after dividing each of the initial  $60 \times 60$  m plots into nine  $20 \times 20$  m subplots. Note that using smaller sampling units may artificially increase the amount of spatial variation because fewer plants are used to describe the community (Bellehumeur, Legendre & Marcotte 1997). Nevertheless, variability comparisons are still valid among sets of sampling units of the same size. As before, we calculated percentage difference dissimilarity with emphasis on compositional data,

structural data and using both attributes. Figs. 5a–c show the corresponding ordination diagrams, obtained using principal coordinate analysis (PCoA, Gower 1966) computed on the square roots of the dissimilarities to avoid the production of negative eigenvalues. In order to facilitate the interpretation of these plots, we added either compositional and/or structural variables as arrows. We then calculated the amount of spatial variation (i.e. non-directional beta diversity) found within each plot as the sum of the dissimilarity values between the nine subplots divided by  $72 (= n(n-1) \text{ for } n = 9)$  (Legendre, Borcard & Peres-Neto 2005; Legendre & De Cáceres 2013); actually, the beta diversity calculation method uses squared dissimilarities, but these had been square-rooted to make the dissimilarity matrices Euclidean. The results showed that the amount of variation in either composition or structure generally tended to increase with stand stage (see ‘Var’ values in Fig. 5), but other patterns were less clear. We tested for homogeneity of variances using the permutational test developed by Anderson (2006) available in function ‘betadisper’ of the R package ‘vegan’ (Oksanen *et al.* 2012). The three forest plots in the southern area did not differ in the amount of internal variation for composition ( $F = 2.747$ ;  $P\text{-value} = 0.0844$ ), structure ( $F = 2.264$ ;  $P\text{-value} = 0.1171$ ) or composition and structure ( $F = 1.950$ ;  $P\text{-value} = 0.1622$ ). In contrast, differences in amount of internal variation turned out to be significant for the three plots in the northern area (composition  $F = 6.331$ ,  $P\text{-value} = 0.0044$ ; structure  $F = 3.767$ ,  $P\text{-value} = 0.0121$ ; composition and structure  $F = 9.047$ ,  $P\text{-value} < 0.0001$ ), because the immature stand was much more internally homogeneous in all aspects compared to the other two.

## Discussion

Ecological communities have long been studied both in terms of species composition and size structure, although the two components are usually analysed separately (e.g. Lee *et al.* 2002; Fang *et al.* 2012). The inherently multivariate nature of species assemblages has led ecologists to embrace multivariate statistical methods for the description and analysis of community compositional patterns (Legendre & Legendre 2012). In



**Fig. 4.** Non-metric multidimensional scaling (nMDS) ordinations of the dissimilarity matrices obtained by computing each of the six dissimilarity coefficients on the 25 synthetic community matrices created in each experiment. Communities are labelled using the combination of treatments that were varied in the experiment.



contrast, analyses of the structural component of communities are not normally carried out using multivariate statistics. For example, variables such as the mean tree diameter, height or basal area are usually calculated separately to describe forests in terms of their size structure (e.g. Fang *et al.* 2012). If more detailed structural information is needed, plant size (height or diameter) frequency distributions are also calculated and compared (e.g. Davies, Palmiotto & Ashton 1998; Lee *et al.* 2002).

A few studies have been published in the past proposing to measure community resemblance in terms of size structure (see Faith *et al.* 1985 and references therein). However, we are the first to present a general framework to determine community resemblance integrating differences in size structure and species composition in a single measurement. The framework is grounded in the concept of cumulative abundance profile, which, if abundance is defined as number of individuals, is directly obtained from the empirical distribution of the chosen structural variable. Our framework allows users to focus on

either species composition or size structure if desired. On the one hand, species composition can be disregarded if species are merged into a single entity. On the other hand, size structure can be disregarded if all organisms are assumed to have the same size. If both species composition and size structure were disregarded, then resemblance measurements would be based on the overall abundance in the community (e.g. the total number of individuals or total biomass). In fact, the overall abundance is a feature of community organization underlying both compositional (i.e. abundance divided among species) and structural (i.e. abundance divided among size classes) attributes. Although we used terrestrial plant communities in most of our examples, our proposals can be readily applied to freshwater or marine benthic communities such as aquatic macrophytes, coral reefs, periphyton species in multilayer mats (e.g. Tall *et al.* 2006) and other types of plant or animal communities where the size of organisms is considered relevant for community organization.

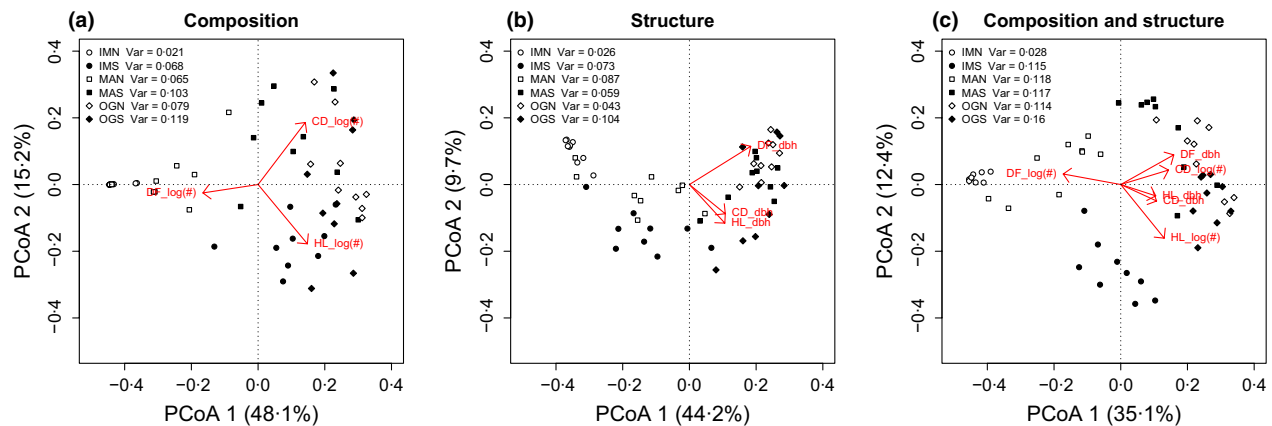
**Table 2.** Percentage difference (*alias* Bray-Curtis dissimilarity) values calculated between the immature (IM), mature (MA) and old-growth (OG) Douglas-fir forest plots on Vancouver Island (Canada). Dissimilarities were calculated using a different approach for each of the three cases (a–c) described in the text

	IM vs. MA	MA vs. OG	IM vs. OG
a) Composition (matrix $\mathbf{X}^{\text{COMP}}$ )			
Southern area	0.095	0.054	0.098
Northern area	0.209	0.259	0.482
b) Structure (matrix $\mathbf{Y}^{\text{STR}}$ )			
Southern area	0.191	0.111	0.273
Northern area	0.210	0.216	0.410
c) Composition and structure (matrix $\mathbf{Y}^{\text{COMP-STR}}$ )			
Southern area	0.263	0.167	0.277
Northern area	0.349	0.356	0.649

#### ABOUT THE CHOICES FOR 'ABUNDANCE' AND 'STRUCTURAL VARIABLE'

Because we wanted to keep our framework as general as possible, we did not constrain the definition of *abundance* and *structural variable*, leaving to users the choices regarding specific measurements (or possible transformations). We discuss here the range of options within each of these two general variables.

Although we did not require abundance to be defined using any specific metric, our cumulative abundance profiles are best suited to measurements such as number of individuals, percentage cover or basal area, because all these can broadly be related to the occupation of space; hence, integration of cumulative abundance values across the domain of the structural variable provides measurements that can be interpreted as 'volumes' or 'biomass'. Abundance measurements that relate



**Fig. 5.** Ordination diagrams obtained using principal coordinate analysis (Gower 1966) on percentage difference dissimilarity values (computed after taking the square root of dissimilarities, to avoid negative eigenvalues): (a) composition, (b) structure, (c) composition and structure. Compositional variables (the logarithm of the number of individuals for each species (CD\_log(#), HL\_log(#), DF\_log(#), where CD = western redcedar, HL = western hemlock and DF = Douglas fir) and structural variables (the average diameter for each species: CD\_dbh, HL\_dbh, DF\_dbh) were added as arrows in the ordination diagrams to facilitate interpretation. Forest subplots are drawn using different symbols depending on the forest plot they belong to. We also indicate the amount of variance (Var, non-directional beta diversity) found within each forest plot, following Legendre & De Cáceres (2013). Plot identifiers: IM = immature, MA = mature, OG = old growth; N = north, S = south.

to both the occupation of space and size of individuals, such as biomass, are less suited for our framework unless they are measured separately for different strata. We did not require a specific definition of the structural variable either, because we do not think that applications of our framework should be restricted to a single type of structural attribute. As much as ecologists choose to measure abundance using different metrics depending on the application, different attributes may be used to represent the size of organisms. Sometimes one may have a specific attribute in mind but use another attribute as a surrogate (e.g. dbh instead of height). One could even use the age of the organisms as the structural variable instead of their size. Clearly, further exploration is required to assess the effect of all those decisions on the resemblance values.

#### ABOUT DISSIMILARITY COEFFICIENTS AND ORDINATION ANALYSES

Our purpose when evaluating dissimilarity coefficients was not to show the benefits of one specific index over others. However, we used synthetic communities to study the ability of the different coefficients to capture structural and compositional differences. Five of the dissimilarity coefficients (namely Whittaker's, the percentage difference, Kulczynski, Ružička and the new generalization of Ochiai) can be considered CAP-based generalizations of coefficients that have been recommended in the past for measurements of compositional resemblance either in their presence-absence form (Bloom 1981; Janson & Vegelius 1981) or in their abundance-based form (Hajdu 1981; Faith, Minchin & Belbin 1987). Other indices, such as the chord/Hellinger distances, have also been recommended to measure compositional resemblance (Faith, Minchin & Belbin 1987; Legendre & Gallagher 2001; Legendre & De Cáceres 2013) but were deemed less suitable for the CAP framework because they could not be expressed in  $A_j-B_j-C_j$  notation (Property P1). Although we believe our choices are valid, other coefficients may be developed in the future to better measure the resemblance in composition and size structure. Other properties may also be required depending on the type of communities or the purpose of the application.

Because none of the recommended dissimilarity measures can be emulated by calculating the Euclidean distance on transformed rectangular matrices (Legendre & Gallagher 2001; Legendre & De Cáceres 2013), adopting the proposed resemblance framework has implications for classification and ordination analyses. For unconstrained ordinations, the present approach can only be used in combination with metric or non-metric multidimensional scaling techniques. For this reason, one cannot draw structural or compositional variables directly in ordination biplots, although they can be added *a posteriori*, as we did in Fig. 5. Regarding constrained ordination, our framework does not directly accommodate the analysis of species-environmental relationships using methods such as redundancy analysis or canonical correspondence analysis. However, distance-based redundancy analysis (Legendre & Anderson 1999; Legendre & Legendre 2012, Section 11.1.5) can still be used to display and test the relationship between

compositional and/or structural changes and potential explanatory factors.

#### POTENTIAL APPLICATIONS

We believe that incorporating structure in community resemblance measurements will be particularly useful for community ecology studies conducted at relatively fine scales. Despite scale limitations, our framework for dissimilarity measurements is useful in a broad range of studies, in fundamental or applied ecological research. Clear examples are studies focusing on plant community dynamics (e.g. Christensen 1977; Davies, Palmiotto & Ashton 1998; Harcombe *et al.* 2002) because our framework allows users to describe temporal patterns in structure and composition jointly. Grouping species according to their functional or structural attributes prior to calculation of cumulative abundance profiles can increase the potential of our framework. For example, fire management decisions usually benefit from the classification of forest stands in terms of their flammability. Measuring the dissimilarity between stands using information about the vertical arrangement and composition of fuel types may be useful to obtain such a classification.

Ecologists are nowadays interested in different components of the spatial variation of communities (i.e. beta diversity) other than the one issued from species (taxonomic) composition (Graham & Fine 2008; Swenson, Anglada-Cordero & Barone 2011). By allowing the incorporation of size structure into the analysis of community resemblance, we are opening the door to the quantification of the structural component of beta diversity. Structural beta diversity may be quantified independently or in combination with species composition, as we did in this paper. Whether structure should be combined with other components of beta diversity remains to be explored.

#### Software availability

Functions to calculate, transform and plot cumulative abundance profiles from either stratified community data or individual data have now been included in the 'vegclust' R package (De Cáceres, Font & Oliva 2010). The package also includes a function to calculate the six dissimilarity coefficients studied in this paper for discrete size classes.

#### Acknowledgements

The work was initiated when the authors visited Sun Yat-sen University in July 2012. The authors would like to thank Albert Petit for useful discussions in the field and Pau Vericat for providing interesting ideas on potential future applications of this framework. This research was supported by respective NSERC grants to P. Legendre (no. 7738) and F. He. M. De Cáceres was supported by research projects BIONOVEL (CGL2011-29539/BOS) and MONTES (CSD2008-00040) funded by the Spanish Ministry of Education and Science. The authors are thankful to Sun Yat-sen University for providing a stimulating collaborative environment.

#### References

- Anderson, M.J. (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, **62**, 245–253.

- Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L. *et al.* (2011) Navigating the multiple meanings of  $\beta$  diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.
- Beck, J., Holloway, J.D. & Schwanghart, W. (2013) Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution*, **4**, 370–382.
- Bellehumeur, C., Legendre, P. & Marcotte, D. (1997) Variance and spatial scales in a tropical rain forest: changing the size of sampling units. *Plant Ecology*, **130**, 89–98.
- Bloom, S. (1981) Similarity indices in community studies: potential pitfalls. *Marine Ecology Progress Series*, **5**, 125–128.
- Chao, A., Chazdon, R.L., Colwell, R.K. & Shen, T.-J. (2006) Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics*, **62**, 361–371.
- Christensen, N.L. (1977) Changes in structure, pattern and diversity associated with climax forest maturation in Piedmont, North Carolina. *American Midland Naturalist*, **97**, 176–188.
- Davies, S., Palmiotto, P. & Ashton, P. (1998) Comparative ecology of 11 sympatric species of Macaranga in Borneo: tree distribution in relation to horizontal and vertical resource heterogeneity. *Journal of Ecology*, **86**, 662–673.
- De Cáceres, M., Font, X. & Oliva, F. (2010) The management of vegetation classifications with fuzzy clustering. *Journal of Vegetation Science*, **21**, 1138–1151.
- Faith, D.P., Minchin, P.R. & Belbin, L. (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, **69**, 57–68.
- Faith, D., Austin, M., Belbin, L. & Margules, C. (1985) Numerical classification of profile attributes in environmental studies. *Journal of Environmental Management*, **20**, 73–85.
- Fang, J., Shen, Z., Tang, Z., Wang, X., Wang, Z., Feng, J. *et al.* (2012) Forest community survey and the structural characteristics of forests in China. *Ecography*, **35**, 1059–1071.
- Getzin, S., Dean, C., He, F., Trofymow, J.A., Wiegand, K. & Wiegand, T. (2006) Spatial patterns and competition of tree species in a Douglas-fir chronosequence on Vancouver Island. *Ecography*, **29**, 671–682.
- Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325–338.
- Graham, C.H. & Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.
- Hajdu, L.J. (1981) Graphical comparison of resemblance measures in phytosociology. *Vegetatio*, **48**, 47–59.
- Harcombe, P.A., Bill, C.J., Fulton, M., Glitzenstein, J.S., Marks, P.L. & Elsik, I.S. (2002) Stand dynamics over 18 years in a southern mixed hardwood forest, Texas, USA. *Journal of Ecology*, **90**, 947–957.
- He, F. & Duncan, R.P. (2000) Density-dependent effects on tree survival in an old-growth Douglas-fir forest. *Journal of Ecology*, **88**, 676–688.
- Holopainen, M. & Kalliovirta, J. (2006) Modern data acquisition for forest inventories. *Forest Inventory – Methodology and Applications* (eds A. Kangas & M. Maltamo), pp. 343–362. Springer, Netherlands.
- Janson, S. & Vegelius, J. (1981) Measures of Ecological Association. *Oecologia*, **49**, 371–376.
- Jost, L., Chao, A. & Chazdon, R.L. (2011) Compositional similarity and  $\beta$  (beta) diversity. *Biological Diversity: Frontiers in Measurement and Assessment*, (eds A.E. Magurran & B.J. McGill), pp. 66–84. Oxford University Press, New York.
- Koleff, P., Gaston, K.J. & Lennon, J.J. (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367–382.
- Kulczynski, S. (1928) Die pflanzenassoziationen der pieinenen. *Bulletin International de L'Academie Polonaise des Sciences et des Lettres, Classe des Sciences Mathematiques et Naturelles, B*, Suppl. II (1927), 57–203.
- Lance, G.N. & Williams, W.T. (1967) Mixed-data classificatory programs. I. Agglomerative systems. *Australian Computational Journal*, **1**, 15–20.
- Lee, H., Davies, S., LaFrankie, J. & Tan, S. (2002) Floristic and structural diversity of mixed dipterocarp forest in lambir hills national park, Sarawak, Malaysia. *Journal of Tropical Forest*, **14**, 379–400.
- Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1–24.
- Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.
- Legendre, P. & De Cáceres, M. (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.
- Legendre, P. & Gallagher, E. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.
- Legendre, P. & Legendre, L. (2012) *Numerical Ecology*, 3rd edn. Elsevier Science BV, Amsterdam, NL.
- Minchin, P.R. (1987) An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, **69**, 89–107.
- Ochiai, A. (1957) Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. *Bulletin of the Japanese Society of Scientific Fisheries*, **22**, 526–530.
- Odum, E.P. (1950) Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology*, **31**, 587–605.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2012) *vegan: Community Ecology Package*. R package version 2.0-5. <http://CRAN.R-project.org/package=vegan>.
- Orlói, L. (1967) An agglomerative method for classification of plant communities. *Journal of Ecology*, **55**, 193–206.
- Philippi, T., Dixon, P. & Taylor, B. (1998) Detecting trends in species composition. *Ecological Applications*, **8**, 300–308.
- Rao, C.R. (1995) A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestió (Quaderns D'Estadística i Investigació Operativa)*, **19**, 23–63.
- Ružička, M. (1958) Anwendung mathematisch-statistischer Methoden in Geobotanik (Synthetische Bearbeitung von Aufnahmen). *Biologia (Bratisl.)*, **13**, 647–661.
- Stephenson, W., Williams, W. & Cook, S. (1972) Computer Analyses of Petersen's Original Data on Bottom Communities. *Ecological Monographs*, **42**, 387–415.
- Swenson, N.G., Anglada-Cordero, P. & Barone, J.A. (2011) Deterministic tropical tree community turnover: evidence from patterns of functional beta diversity along an elevational gradient. *Proceedings of the Royal Society B (Biological Sciences)*, **278**, 877–884.
- Tall, L., Cattaneo, A., Cloutier, L., Dray, S. & Legendre, P. (2006) Resource partitioning in a grazer guild feeding on a multilayer diatom mat. *Journal of the North American Benthological Society*, **25**, 800–810.
- Tamás, J., Podani, J. & Csontos, P. (2001) An extension of presence/absence coefficients to abundance data: a new look at absence. *Journal of Vegetation Science*, **12**, 401–410.
- Trofymow, J.A., Porter, G.L., Blackwell, B.A., Arksey, R., Marshall, V. & Pollard, D. (1997) *Chronosequences for Research Into the Effects of Converting Coastal British Columbia old-Growth Forests to Managed Forests: An Establishment Report*. Canadian Forest Service, Forest Ecosystem Processes Network, Victoria, British Columbia, Canada.
- Van der Maarel, E. (1979) Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio*, **39**, 97–114.
- Whittaker, R. (1952) A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs*, **22**, 1–44.

Received 11 March 2013; accepted 30 August 2013

Handling Editor: Daniel Faith

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1** Mathematical proofs for the properties of dissimilarity coefficients

**Appendix S2** Transformations of cumulative abundance profiles

**Appendix S3** Species composition and CAPs of the six Douglas-fir forest plots in British Columbia, Canada