



# A general method for the classification of forest stands using species composition and vertical and horizontal structure

Miquel De Cáceres<sup>1,2</sup> · Santiago Martín-Alcón<sup>3</sup> · Jose Ramón González-Olabarria<sup>1</sup> · Lluís Coll<sup>1,2,4</sup>

Received: 6 July 2018 / Accepted: 10 March 2019 / Published online: 12 April 2019  
© INRA and Springer-Verlag France SAS, part of Springer Nature 2019

## Abstract

• **Key message** We present a novel approach to define pure- and mixed-forest typologies from the comparison of pairs of forest plots in terms of species identity, diameter, and height of their trees.

• **Context** Forest typologies are useful for many purposes, including forest mapping, assessing habitat quality, studying forest dynamics, or defining sustainable management strategies. Quantitative typologies meant for forestry applications normally focus on horizontal and vertical structure of forest plots as main classification criteria, with species composition often playing a secondary role. The selection of relevant variables is often idiosyncratic and influenced by a priori expectations of the forest types to be distinguished.

• **Aims** We present a general framework to define forest typologies where the dissimilarity between forest stands is assessed using coefficients that integrate the information of species composition with the univariate distribution of tree diameters or heights or the bivariate distribution of tree diameters and heights.

• **Methods** We illustrate our proposal with the classification of forest inventory plots in Catalonia (NE Spain), comparing the results obtained using the bivariate distribution of diameters and heights to those obtained using either tree heights or tree diameters only.

• **Results** The number of subtypes obtained using the tree diameter distribution for the calculation of dissimilarity was often the same as those obtained from the tree height distribution or to those using the bivariate distribution. However, classifications obtained using the three approaches were often different in terms of forest plot membership.

• **Conclusion** The proposed classification framework is particularly suited to define forest typologies from forest inventory data and allows taking advantage of the bivariate distribution of diameters and heights if both variables are measured. It can provide support to the development of typologies in situations where fine-scale variability of topographic, climatic, and legacy management factors leads to fine-scale variation in forest structure and composition, including uneven-aged and mixed stands.

**Keywords** Dissimilarity coefficients · Forest plot · Forest typology · Mixed forests · Stand structure

---

**Handling Editor:** Laurent Bergès

**Contributions of co-authors:** MDC and LC conceived the idea. MDC developed the methods and conducted data analyses. SMA and JRG contributed to interpretation of results. All co-authors contributed to manuscript writing.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13595-019-0824-0>) contains supplementary material, which is available to authorized users.

✉ Miquel De Cáceres  
miquelcaceres@gmail.com

<sup>1</sup> Forest Science and Technology Center of Catalonia (CTFC),  
25280 Solsona, Spain

<sup>2</sup> Centre for Ecological Research and Forestry Applications (CREAF),  
08193 Cerdanyola del Valles, Spain

<sup>3</sup> Agresta S. Coop., Calle Duque de Fernán Núñez, 2,  
28012 Madrid, Spain

<sup>4</sup> Department of Agriculture and Forest Engineering (EAGROF),  
University of Lleida, 25198 Lleida, Spain

## 1 Introduction

Vegetation classifications are powerful tools to describe, summarize, and represent the variation of vegetation across space and time (De Cáceres et al. 2015). The development and use of typologies in forestry can be traced back to the birth of forest science, because forest types have traditionally helped foresters to summarize site and stand quality, a prerequisite for predicting potential growth or timber yield (Cajander 1949). Nowadays, forest typologies are important tools for assessing and monitoring the state of forest ecosystems at several scales (Corona 2016). They are useful for many purposes, such as assessing habitat quality or the vulnerability to disturbances, anticipating forest dynamics, or mapping forest resources for the definition of sustainable stand-oriented management strategies (e.g., James and Wamer 1982; Bebi et al. 2001; Bruciamacchie 2001; Barbati et al. 2007, 2014). In addition, combining detailed forest typologies with three-dimensional remote sensing data has recently shown its potential for mapping variations on developmental stages, post-disturbance regeneration patterns, or even the change of management activities on forested landscapes (Falkowski et al. 2009; Bottalico et al. 2014; Martín-Alcón et al. 2015; Valbuena et al. 2016, 2017).

Forest typologies can be rather informal qualitative descriptions, sometimes complemented with profile diagrams to make them easier to communicate (O'Hara et al. 1996; Larsen and Nielsen 2007). Alternatively, they can be defined quantitatively and the resulting forest classes be differentiated using formal assignment rules (e.g. classification trees) on the basis of chosen thresholds for relevant variables (e.g., Bebi et al. 2001; Bruciamacchie 2001; Giannetti et al. 2018). Quantitative assignment rules can be defined by expert knowledge, after preliminary inspection of the main axes of variation of potential variables (e.g., Aunós et al. 2007), but a more data-driven approach involves two steps: (1) to define homogeneous groups of forest plots using unsupervised classification methods (i.e., hierarchical or non-hierarchical clustering); (2) to define assignment rules by means of supervised classification (e.g., discriminant analysis, classification trees, or neural networks), using the classification produced in the first step as training data (Reque and Bravo 2008; Martín-Alcón et al. 2012; De Cáceres and Wiser 2012). The first step of the data-driven approach requires deciding how to measure the *resemblance* between forest plots, i.e., deciding which structural and compositional variables are relevant and how to combine their information in a resemblance (dissimilarity or similarity) coefficient.

Fine-grained quantitative forest typologies normally focus on structure as main classification criteria (e.g., basal area, canopy cover, the distribution of tree diameters and heights, stocking density, or wood volume) (Faith et al. 1985; McElhinny et al. 2005). Very often, the chosen variables are expressed in different units (e.g., basal area vs. dominant

height) and/or are partially correlated (e.g., basal area vs. mean quadratic diameter), which requires employing variable standardization or factor analysis prior to the calculation of distances between forest stands (e.g., Reque and Bravo 2008; Martín-Alcón et al. 2012; Casals et al. 2015). Besides the need to represent both horizontal and vertical structure, there are no standard choices of relevant variables and procedures to define typologies, which results in variable choices being rather idiosyncratic to each classification analysis. In addition, the scope of typologies is often restricted to forests where a single target species is dominating or has minimum abundance (in terms of basal area or number of individuals). For example, in Spain, quantitative forest typologies have focused on silver fir (*Abies alba* Mill.), beech (*Fagus sylvatica* L.), sessile oak (*Quercus petraea* Matt.), or mountain pine (*Pinus uncinata* Ram.) forests (Aunós et al. 2007; Reque and Bravo 2008; Martín-Alcón et al. 2012). The role of other tree species than the target ones is only accounted for implicitly, by considering variables such as the percentage of the total basal area corresponding to the target species or the ratio of the average size for individuals of the target species compared to others (Aunós et al. 2007; Reque and Bravo 2008). Although coarse forest typologies exist simultaneously addressing variation in structure and composition (e.g., Godinho-Ferreira et al. 2005; Barbati et al. 2014), quantitative classifications addressing in detail the compositional and structural variation of mixed-forests are rare (Ngo Bieng et al. 2006). This may seem surprising, since mixed forests occupy almost one-fourth (23%) of the European's forested land (Forest Europe, UNECE, FAO 2011) and their importance is progressively increasing in response to the growing complexity of societal demands (Bravo-Oviedo et al. 2014). Adopting similarity or dissimilarity coefficients designed to account for multiple structural and compositional attributes would allow defining fine forest typologies in different contexts (even-aged vs. uneven-aged, pure vs. mixed) in a more consistent way.

Ecologists have long compared the composition of plant or animal communities using species presence-absence or abundance data (i.e., cover or density values) and multivariate resemblance coefficients (Legendre and Legendre 2012). However, the usefulness of this type of coefficients in forestry is rather limited, due to the inability to simultaneously consider the vertical and horizontal structure of stands. At most, compositional coefficients have been used for the classification of stands in terms of the composition within canopy and understory layers separately (Youngblood 1993). De Cáceres et al. (2013) presented a general framework to compare ecological communities that takes into account the size structure and composition simultaneously and illustrated their approach for the comparison of forest stands in terms of diameter distribution and tree species identity. While this framework has been successfully applied to describe and typify the structure and composition of Anatolian black pine forests (Yılmaz et al.

2018) and to assess the amount structural and compositional variation within forests (Yao et al. 2019), its full potential for the definition of forest typologies remains to be explored. Moreover, when characterizing and/or differentiating forest stands from a forestry perspective, both the distribution of tree diameters and tree heights may be relevant, because the provision of several forest goods (e.g., timber quantity and quality) and services (e.g., soil protection or habitat quality) depends on the distribution of these two variables. Although the two are often correlated, general frameworks to measure the resemblance between forest stands could take into account the joint bivariate distribution of diameters and heights.

The objectives of this article are twofold. First, we introduce an extension of the De Cáceres et al. (2013) framework to allow considering two size variables (namely height and diameter), instead of only one, when calculating the resemblance between forest stands. Second, we illustrate the usefulness of this approach to define typologies of different kinds, including pure vs. mixed stands or even-aged vs. uneven-aged stands. In the following section, we first review the conceptual and mathematical aspects of the framework of De Cáceres et al. (2013), describe its extension to consider two size variables, and present a small simulation study that illustrates the behavior of the proposed dissimilarity coefficients. Then, we use our dissimilarity framework and unsupervised clustering techniques to derive pure- and mixed-forest typologies from forest inventory data. When doing that, we compare the results obtained using the bivariate distribution of tree diameters and heights with those obtained using one of the two marginal distributions only. Finally, we discuss the advantages and limitations of the resemblance framework, including potential applications beyond the definition of typologies.

## 2 Resemblance between forest stands in terms of structure and composition

### 2.1 Original framework

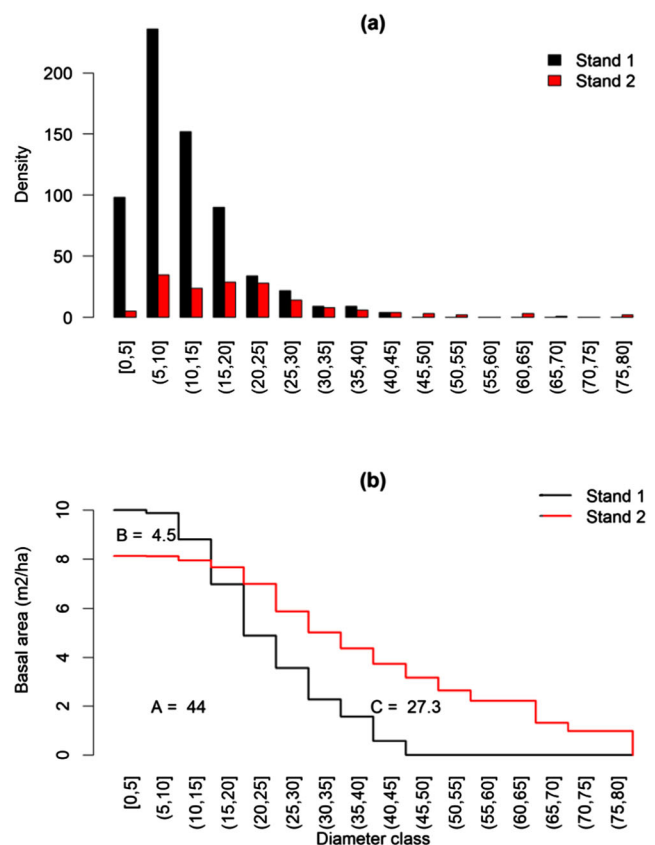
The framework of De Cáceres et al. (2013) allows defining and calculating the dissimilarity between forest stands in a flexible way, depending on the choice of variables made by the user regarding three aspects of stand organization (Pommerening 2002):

- Abundance** (i.e., horizontal structure): A non-negative variable to represent the space occupied by plants, such as density, (projected) crown cover or basal area.
- Size** (i.e. vertical structure): A (continuous or ordinal) variable to represent plant size, such as tree diameter or plant height.

- Composition**: A categorization of plants, such as species composition, growth forms, or other functional distinction.

Choices for (a) and (b) imply specifying a distribution of the total abundance into size classes (e.g., overall density or basal area into tree diameter or height classes). Instead of using this size distribution directly, however, the framework operates on the basis of cumulative values. The *cumulative abundance profile* (CAP) is a function taking a size value as input and returning the cumulative abundance of plants whose size is equal to or larger than the input value. For example, if tree diameter is chosen as the size variable and basal area as the abundance variable, the CAP value is the cumulative basal area corresponding to trees as thick as or thicker than the input diameter (see Fig. 1). CAP is a non-increasing function whose value is always maximal for the smallest value of the size variable (e.g., the cumulative basal area corresponding to the smallest diameter class is equal to the basal area of the stand).

The procedure to calculate the resemblance between a pair of forest stands can be summarized in three steps: (1)



**Fig. 1** Example of calculation of resemblance between a pair of monospecific stands using the CAP framework. **a** Distribution of trees in diameter 5-cm classes in the two stands. **b** CAPs for the two stands, using diameter as size variable and basal area as abundance variable. Values of  $A_k$ ,  $B_k$ , and  $C_k$  are indicated in each of three areas resulting from the intersection of CAPs

calculating the CAP for each tree species (or compositional class) in each of the two stands, (2) comparing the CAPs of the two stands for each species, and (3) pooling the result of CAP comparisons across species with an appropriate dissimilarity coefficient. The details of the three steps are as follows (De Cáceres et al. 2013):

**Step 1** Calculation of CAP, following the definition given above, is done for each  $k$  species and for each of the two stands separately. The importance accorded to differences in stand horizontal structure (a), vertical structure (b), and composition (c) will depend on the variables and/or transformations chosen to represent these aspects when building CAPs. For example, dividing trees into angiosperm and gymnosperms, instead of using species composition, will decrease the importance of compositional vs. structural differences. Similarly, the importance of vertical structure can be modulated by changing the resolution of diameter or height bins (Yao et al. 2019). Unequal widths of diameter or height bins can be used to modulate the importance accorded to differences in abundance across tree size (e.g., using diameter quadratic bins will increase the importance of differences in saplings). Finally, the role of horizontal structure can be modulated by transforming CAPs prior to their comparison. For example, a square root transformation applied to cumulative abundance values will decrease the importance of differences in horizontal structure with respect to the other aspects (De Cáceres et al. 2013).

**Step 2** The comparison of the two CAPs for a given species  $k$ — $CAP_{1k}$  and  $CAP_{2k}$ —is done by integrating the comparison of cumulative abundance values along the values of the structural variable. This comparison leads to distinguishing between three regions in the diagram that shows the two CAPs:  $A_k$ , the areas where the two profiles overlap;  $B_k$ , the areas where  $CAP_{1k}$  exceeds  $CAP_{2k}$ ; and  $C_k$ , the areas where  $CAP_{2k}$  exceeds  $CAP_{1k}$  (Fig. 1b). If  $\theta$  is the size variable, the three quantities are mathematically calculated using

$$\begin{aligned} A_k &= \int_{\theta=0}^{\infty} \min(CAP_{1k}(\theta), CAP_{2k}(\theta)) \cdot d\theta \\ B_k &= \int_{\theta=0}^{\infty} (CAP_{1k}(\theta) - \min(CAP_{1k}(\theta), CAP_{2k}(\theta))) \cdot d\theta \\ C_k &= \int_{s=0}^{\infty} (CAP_{2k}(\theta) - \min(CAP_{1k}(\theta), CAP_{2k}(\theta))) \cdot d\theta \end{aligned} \quad (1)$$

While eq. 1 assumes a continuous size variable, if one employs a discrete size variable (i.e., height or diameter bins) to define CAPs, then  $A_k$ ,  $B_k$ , and  $C_k$  are calculated replacing integrals in eq. 1 by summations across size classes.

**Step 3** The species-wise comparison of pairs of CAPs results in values  $A_k$ ,  $B_k$ , and  $C_k$  for each species  $k$ . At this point, compositional resemblance coefficients typical of community ecology can be calculated, provided they can be decomposed into quantities analogous to  $A_k$ ,  $B_k$ , and  $C_k$  (Tamás et al. 2001). Although several coefficients were presented in De Cáceres et al. (2013), we focus here on  $D_{man}$  and  $D_{bray}$ , which are generalizations of the Manhattan (or city-block) distance and the percentage difference (alias Bray-Curtis) dissimilarity (Odum 1950; Bray and Curtis 1957), respectively. The Manhattan distance calculated across all species (from  $k=1$  to  $k=p$  species or compositional classes) is

$$D_{man} = \sum_{k=1}^p (B_k + C_k) \quad (2)$$

and the corresponding generalization of the Bray-Curtis dissimilarity is

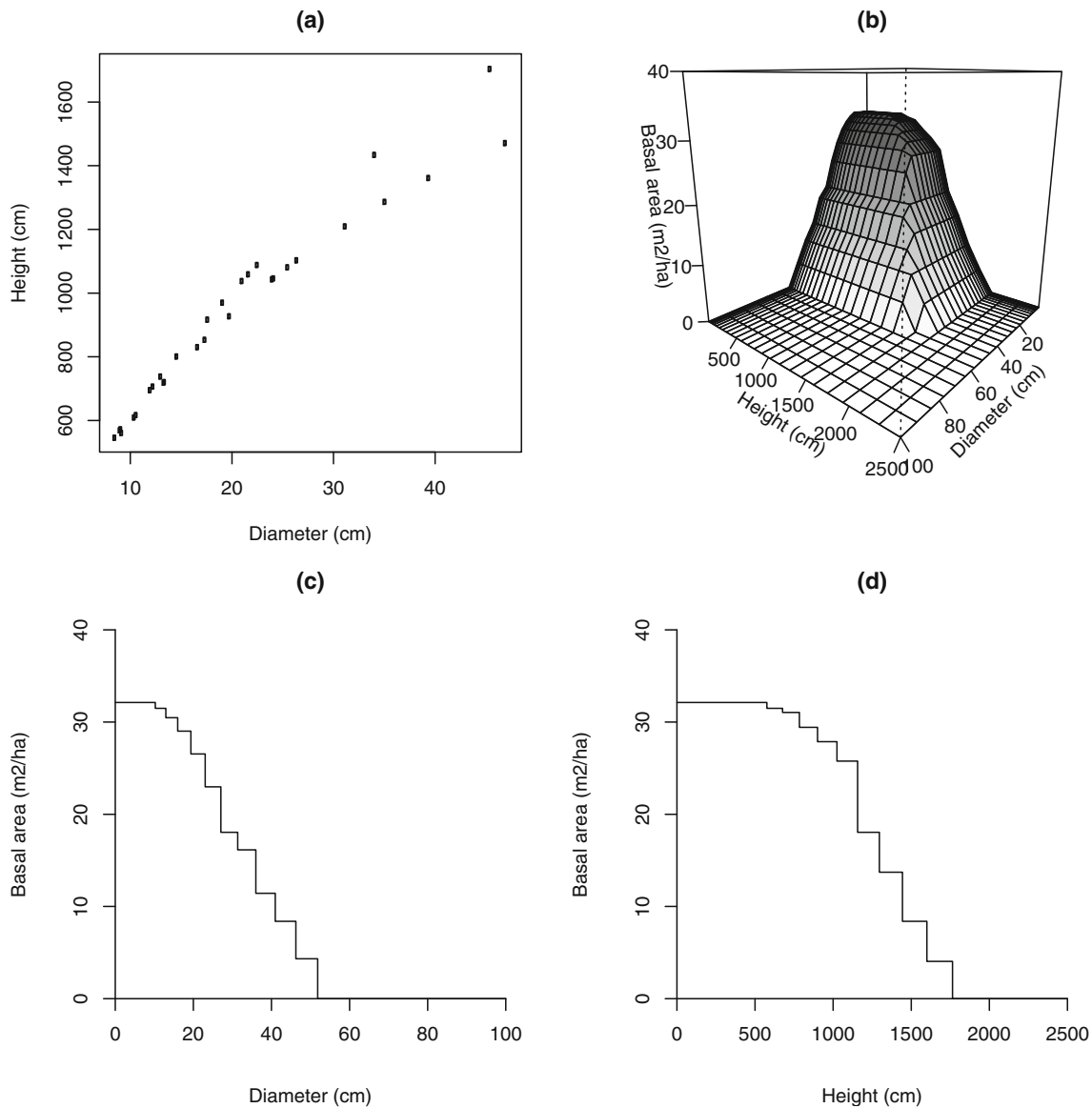
$$D_{bray} = \frac{\sum_{k=1}^p (B_k + C_k)}{\sum_{k=1}^p (2A_k + B_k + C_k)} \quad (3)$$

The main difference between  $D_{man}$  and  $D_{bray}$  is that the latter involves a normalizing factor that relativizes the dissimilarity between the two stands to the [0–1] interval. For example, in the single-species case of the CAPs represented in Fig. 1b,  $D_{man} = 31.8$  and  $D_{bray} = 0.265$ .

## 2.2 Extension of the framework

With the aim of comparing forest stands in terms of their observed bivariate distribution of tree heights and diameters, we define here the *cumulative abundance surface* (CAS) as a function taking a *pair* of values of size variables and returning the cumulative abundance of individuals whose sizes are equal to or larger than the values given. If one takes height and diameter as size variables and basal area as abundance measure, the CAS function will return the basal area of trees as tall as or taller than the input height value and, at the same time, whose diameter is as thick as or thicker than the input diameter value (Fig. 2a, b). Like the marginal distributions of any bivariate joint distribution, CAS can be marginalized into the CAPs of its two size variables (Fig. 2c, d).

The calculation of resemblance between stands is analogous to the original procedure but replacing the role of CAPs by CASs. In step 1, one calculates the CAS for each species in each stand. The comparison of  $CAS_{1k}$  and  $CAS_{2k}$



**Fig. 2** Example of the construction of a CAS and its marginal CAPs for diameters and heights from individual tree data. **a** A scatterplot of diameter and height values in a 20 × 20 m forest plot. **b** The cumulative

abundance surface (CAS) calculated after deciding the limits of bins for heights and diameters. **c, d** The CAPs that result from marginalizing the CAS

for a given species  $k$  (step 2) is done by integrating the comparison of cumulative abundance values across the plane defined by the two size variables. In this case, the intersection of the two surfaces defines three volumes:  $A_k$ —the volume common to the two cumulative surfaces;  $B_k$ —the volume that occurs in  $CAS_{1k}$  but not in  $CAS_{2k}$ ; and  $C_k$ —the volume that occurs in  $CAS_{2k}$  but not in stand  $CAS_{1k}$ . If  $\theta$  and  $\varphi$  are the two size variables, the three quantities are

$$\begin{aligned}
 A_k &= \int_{\theta=0}^{\infty} \int_{\varphi=0}^{\infty} \min(CAS_{1k}(\theta, \varphi), CAS_{2k}(\theta, \varphi)) \cdot d\theta \cdot d\varphi \\
 B_k &= \int_{\theta=0}^{\infty} \int_{\varphi=0}^{\infty} (CAS_{1k}(\theta, \varphi) - \min(CAS_{1k}(\theta, \varphi), CAS_{2k}(\theta, \varphi))) \cdot d\theta \cdot d\varphi \\
 C_k &= \int_{\theta=0}^{\infty} \int_{\varphi=0}^{\infty} (CAS_{2k}(\theta, \varphi) - \min(CAS_{1k}(\theta, \varphi), CAS_{2k}(\theta, \varphi))) \cdot d\theta \cdot d\varphi
 \end{aligned} \tag{4}$$

As before, we require the resemblance indices (step 3) to be a function of  $A_k$ ,  $B_k$ , and  $C_k$  for each species. In particular,  $D_{man}$  and  $D_{bray}$  (eqs. 2 and 3) can be calculated without modification. Functions to build CASs and to calculate the resemblance coefficients  $D_{man}$  and  $D_{bray}$  have been included in the R package ‘vegclust’ (De Cáceres et al. 2010).

### 2.3 Simulation study

We simulated tree data using Johnson’s Sbb distribution (Johnson 1949; Schreuder and Hafley 1977) to study whether  $D_{man}$  and  $D_{bray}$  can reflect differences in terms of total basal area, tree diameter distribution, and tree height

distribution (see details in Annex 1). We did not include differences in composition because they were already studied in De Cáceres et al. (2013). In a first experiment, we simulated 25 different stands by crossing five treatments in stand basal area (5, 10, 15, 20, and 25 m<sup>2</sup>/ha; labeled '1' to '5') with five treatments in size distribution (labeled 'a' to 'e'), assuming that when stands had trees larger in diameter, those trees were also taller (i.e., in treatment 'a' mean dbh/height was 5 cm/4 m; in 'b' 12 cm/5 m; up to 80 cm/20 m in 'e'). In a second experiment, we simulated stands having all the same basal area (25 m<sup>2</sup>/ha) but differing in height distribution, diameter distribution, or both. In this case, we crossed five treatments in diameter distribution (labeled 'a' to 'e') with five treatments in tree height distribution (labeled '1' to '5'). Thus, stands could be composed of short and thin trees, tall and thin trees, short and thick trees, or tall and thick trees; with intermediates situations being also covered. For every stand, we simulated tree individuals until the target basal area was met.

The CAS for each stand was built taking basal area as abundance variable and tree diameter and tree height as size variables. We then calculated the resemblance between all pairs of the 25 stands using either  $D_{man}$  or  $D_{bray}$ . Kruskal's non-metric multidimensional scaling was used to display each of the resulting dissimilarity matrices in a two-dimensional scatter plot (see Annex 1, Figs. 9 and 11). Both coefficients were responsive to variations in stand basal area, tree height distribution and tree diameter distribution, and the resulting resemblance matrices satisfactorily preserved the ordering of stands along the simulated gradients of the two experiments. Nevertheless, the behavior of  $D_{man}$  and  $D_{bray}$  differed in the magnitude of dissimilarity values for stands having low basal area. CASs enclose a larger volume when either the basal area of the stand increases or when trees are larger (see Annex 1, Figs. 8 and 10), which leads to larger values of  $A_k$ ,  $B_k$ , and  $C_k$  when comparing CASs in those situations. Since  $D_{man}$  does not have a normalizing factor in its formula, the resulting distance value was affected by differences in CAS volume. The choice of a dissimilarity index should be made in relation to the problem at hand (Legendre and De Cáceres 2013). After inspecting the results of the simulation study, we concluded that  $D_{man}$  is more appropriate than  $D_{bray}$  for forestry applications because it better represents the differences in the distribution of wood volume across species or tree sizes.

### 3 Definition of forest types in Catalonia using forest inventory data

Here, we illustrate the usefulness of the framework presented in the previous section to define forest typologies on the basis of differences in stand composition and structure. Starting from forest inventory data, we generated a classification

system for forests in Catalonia (NE of Spain) with two classification levels. Forest types of the first (coarser) level were defined using species dominance as single classification criterion whereas the definition of subtypes of the second (finer) level included both composition and structure as classification criteria. To evaluate the effect of considering the bivariate distribution of diameters and heights vs. considering one of the two marginal distributions alone, we compare the typologies obtained using the extended (CAS) framework with the typologies obtained using the original (CAP) framework.

#### 3.1 Data preparation

We compiled 10,546 plot records of the Third Spanish Forest Inventory (NFI3) in Catalonia (DGCN 2005). NFI3 plots consist of four nested circular subplots (of radius 5, 10, 15, and 25 m). For each subplot, species identity, height (cm), and diameter at breast height (dbh, cm) of a living tree are recorded only if its diameter is larger than a threshold (7.5, 12.5, 22.5, and 42.5 cm, respectively). For example, between 15 m and 25 m radius, only trees larger than 42.5 cm are recorded. In subplots of 5 m radius, the number of saplings ( $2.5 \text{ cm} \leq \text{dbh} \leq 7.5 \text{ cm}$ ) per species and their average height is also recorded. We calculated the basal area per hectare (m<sup>2</sup>/ha) corresponding to each measured tree by multiplying the normal section of the tree by a density factor that depended on the subplot indicated by its dbh. We focused on forests dominated by the 12 most prevalent tree species in the region: Aleppo pine (*Pinus halepensis* Mill.), Black pine (*P. nigra* Arnold), Scots pine (*Pinus sylvestris* L.), Mountain pine (*Pinus uncinata* Ram.), Stone pine (*Pinus pinea* L.), Maritime pine (*Pinus pinaster* Aiton), European fir (*Abies alba* Mill.), Holm oak (*Quercus ilex* L.), Cork oak (*Q. suber* L.), Downy oak (*Quercus humilis* Mill.), Portuguese oak (*Quercus faginea* Lam.), and European beech (*Fagus sylvatica* L.). We discarded from the data set a total of 3168 forest plots where tree species other than those mentioned accounted for more than 20% of basal area or more than 20% of all individuals.

#### 3.2 Level 1–dominance-based forest types

With the remaining 7378 plots, we first defined forest types depending on species dominance. A given plot was considered to be dominated by one of the twelve tree species if that species alone accounted simultaneously for more than 80% of basal area and more than 80% of tree density. Similarly, the plot was considered as being co-dominated by two species if each of them alone accounted for between 20 and 80% of basal area or density and altogether they accounted for more than 80% of both variables. Forest types defined in this way were only considered valid if they had at least 10 plots assigned to them, to ensure a minimum number of class members. Among the 7378 plots evaluated, 4337 were dominated

by a single species whereas 2034 were co-dominated by two species (see Table 1). The remaining 1007 plots were considered as showing intermediate dominance patterns and were discarded for the definition of forest subtypes. We calculated the non-parametric (Spearman's  $\rho$ ) correlation between tree diameters and tree heights of each species in each forest plot. Average correlations were rather high for forests dominated by *Abies alba* but much lower for those dominated by oak species like *Quercus ilex* or *Q. faginea* (Table 1).

### 3.3 Level 2–forest subtypes

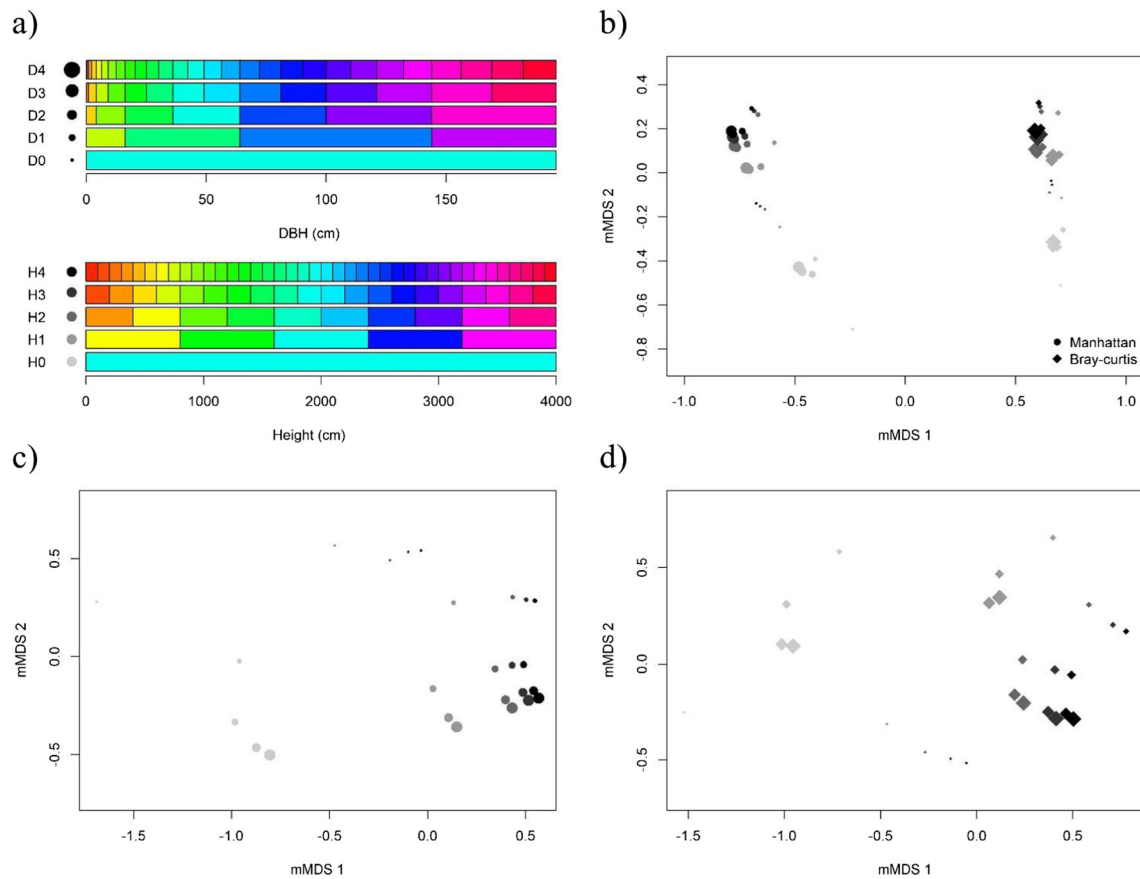
For each dominance-based forest type, we considered the definition of subtypes according to both composition and structure. We generated subtypes using either the extended (CAS) framework or the original (CAP) framework. CASs for each plot and species were built using basal area as abundance variable and taking tree diameter and height as size variables. In order to increase the sensitivity of dissimilarities with respect to recruitment stages, diameter bins were finer to distinguish thin trees than to distinguish thick trees: dbh limits (in cm) for class  $r$  were defined as  $4 + [(0.5 \cdot (r-1))^2, (0.5 \cdot r)^2]$  cm. The resulting bins were 4–6.25 cm, 6.25–9 cm, 9–12.25 cm, 12.25–16 cm, 16–20.25 cm, etc. Height classes were defined using 100-cm bins. CAPs were

built in the same way as for CASs, taking either tree diameter or tree height as (single) size variable (these will be referred to as CAP(d) or CAP(h), respectively). We evaluated the dissimilarity between pairs of plots using  $D_{man}$  (eq. 2), with  $B_k$  and  $C_k$  calculated from either CAS pairs (eq. 4), CAP(d) pairs, or CAP(h) pairs (eq. 1). The correlation between the three resulting dissimilarity matrices was almost always above 0.85 (see Table S1.1 in SEM1), indicating a large degree of agreement in the three ways of assessing resemblance between forest stands (but see classification results below). To illustrate how choices regarding the dissimilarity metric and the definition of diameter and height bins may affect the dissimilarity between forest plots, we include in Fig. 3 the results of a small sensitivity analysis using the 434 forest plots dominated by *Pinus nigra*. They indicate that the dissimilarity metric ( $D_{man}$  or  $D_{bray}$ ) is the most critical choice (Fig. 3b). After that, including or not the height or diameter distributions (i.e., splitting abundance data into height or diameter bins) leads to markedly different dissimilarity values. Whether bins are more or less finely defined is of lower importance, especially beyond a certain level of resolution (D3–D4 or H3–H4 in Fig. 3c).

For each dominance-based forest type and each of the three approaches, partitions of different numbers of clusters (from 2 to 12) were obtained as follows. Starting from

**Table 1** Forest types (12 single-species forests and 26 two-species forests) defined by dominance criteria, number of NFI3 plots assigned to them

Forest type	<i>N</i>	Average $\rho$ (H, DBH)	Forest type	<i>N</i>	Average $\rho$ (H, DBH)
Single-species forests (12)			Two-species forests (cont.)		
<i>Pinus halepensis</i>	1297	0.666	<i>P. nigra</i> – <i>Q. ilex</i>	110	0.662/0.481
<i>Pinus nigra</i>	434	0.761	<i>P. nigra</i> – <i>Q. humilis</i>	90	0.795/0.588
<i>Pinus sylvestris</i>	679	0.694	<i>P. nigra</i> – <i>Q. faginea</i>	53	0.712/0.693
<i>Pinus uncinata</i>	431	0.741	<i>P. sylvestris</i> – <i>P. uncinata</i>	117	0.681/0.736
<i>Pinus pinea</i>	60	0.622	<i>P. sylvestris</i> – <i>A. alba</i>	19	0.616/0.890
<i>Pinus pinaster</i>	40	0.622	<i>P. sylvestris</i> – <i>Q. ilex</i>	143	0.635/0.612
<i>Abies alba</i>	42	0.831	<i>P. sylvestris</i> – <i>Q. humilis</i>	134	0.668/0.551
<i>Quercus ilex</i>	798	0.569	<i>P. sylvestris</i> – <i>Q. faginea</i>	24	0.685/0.624
<i>Quercus suber</i>	178	0.701	<i>P. sylvestris</i> – <i>F. sylvatica</i>	31	0.609/0.615
<i>Quercus humilis</i>	189	0.671	<i>P. uncinata</i> – <i>A. alba</i>	38	0.722/0.769
<i>Quercus faginea</i>	95	0.612	<i>P. pinea</i> – <i>P. pinaster</i>	14	0.457/0.712
<i>Fagus sylvatica</i>	94	0.740	<i>P. pinea</i> – <i>Q. ilex</i>	69	0.525/0.640
Total	4337		<i>P. pinea</i> – <i>Q. suber</i>	44	0.616/0.650
Two-species forests (26)			<i>P. pinaster</i> – <i>Q. ilex</i>	12	0.653/0.048
<i>P. halepensis</i> – <i>P. nigra</i>	94	0.714/0.792	<i>P. pinaster</i> – <i>Q. suber</i>	33	0.565/0.646
<i>P. halepensis</i> – <i>P. pinea</i>	74	0.622/0.665	<i>A. alba</i> – <i>F. sylvatica</i>	30	0.855/0.612
<i>P. halepensis</i> – <i>Q. ilex</i>	341	0.670/0.597	<i>Q. ilex</i> – <i>Q. suber</i>	113	0.618/0.668
<i>P. halepensis</i> – <i>Q. humilis</i>	29	0.610/0.638	<i>Q. ilex</i> – <i>Q. humilis</i>	152	0.608/0.565
<i>P. halepensis</i> – <i>Q. faginea</i>	41	0.665/0.440	<i>Q. ilex</i> – <i>Q. faginea</i>	45	0.601/0.610
<i>P. nigra</i> – <i>P. sylvestris</i>	174	0.741/0.665	<i>Q. suber</i> – <i>Q. humilis</i>	10	0.711/0.865
			Total	2034	



**Fig. 3** Sensitivity analysis regarding the effect of dissimilarity metric and the resolution of diameter and height bins on the dissimilarities between plots. Dissimilarities between the 434 forest plots dominated by *Pinus nigra* were calculated using 50 combinations of dissimilarity metric (either  $D_{man}$  or  $D_{bray}$ ; filled circles and diamonds, respectively) and five degrees of resolution in the definition of diameter bins (D0 to D4) and height bins (H0 to H5). **a** Bin definitions graphically, along with the

symbol sizes (for diameters) and gray tones (for heights) used in the remaining panels (**b–d**). **b** A metric MDS representation in two dimensions (stress = 0.068) of the correlation between the 50 different dissimilarity matrices. **c**, **d** The MDS representation when focusing on the 25 combinations corresponding to  $D_{man}$  (stress = 0.099) and when focusing on those corresponding to  $D_{bray}$  (stress = 0.131), respectively

the square matrix containing distances between pairs of plots, an initial partition was generated by cutting the dendrogram produced by Ward's hierarchical clustering (Ward 1963). Then, we used the clustering algorithm Partitioning Around Medoids (Kaufman and Rousseeuw 1990) to refine this initial partition. A cluster medoid is defined as the object (i.e., a plot) having the minimal sum of distances to all the other objects in the cluster. We evaluated the final partitions corresponding to different numbers of clusters using Silhouette analysis (Kaufman and Rousseeuw 1990). For each forest type, we selected the partition with the maximum number of clusters (subtypes) as long as the average cluster Silhouette was larger than 0.20 and all clusters had at least 10 plots assigned to them. Subtypes were not defined if no partition fulfilled these requirements.

The number of subtypes recognized for each dominance-based forest type under each of the three

approaches is indicated in Table 2. In general, the number of subtypes depended on the total number of plots included in the dominance-based forest type. For example, pure *P. halepensis* forests, the most frequent type, were subdivided into the largest number of subtypes, whereas only two subtypes were recognized for forests dominated by *A. alba* or *P. pinaster*, both species with low prevalence in the study area. Forests co-dominated by two species followed the same rule. While six CAS-based subtypes were recognized for the common *P. halepensis*–*Q. ilex* forests, several other mixed forests were represented by too few plots to be divided into subtypes. Clearly, those dominance-based forest types less represented in the study area were described in less detail.

For 19 out of 36 dominance-based types, the number of subtypes distinguished using the CAS framework was equal to the number of subtypes distinguished using either CAP(d) or CAP(h) (Table 2). However, the overall number of subtypes



**Table 2** Number of forest subtypes (clusters) in classifications based on CAP(d) (i.e., using diameter distribution alone), classifications based on CAP(h) (i.e., using height distribution alone) and classifications based on CAS (i.e., taking into account the bivariate distribution of diameters and heights). The adjusted Rand (Hubert and Arabie 1985) index evaluates the degree of agreement between the membership matrices issued from each classification approach. Adjusted Rand values corresponding to line ‘Total’ were calculated after pooling the membership matrices corresponding to the different dominance-based types

(a) Single-species forests (12) Forest type	Number of subtypes			Adjusted Rand index		
	CAP(d)	CAP(h)	CAS	CAP(d) vs. CAP(h)	CAP(d) vs. CAS	CAP(h) vs. CAS
<i>Pinus halepensis</i>	7	12	12	0.545	0.460	0.464
<i>Pinus nigra</i>	5	6	5	0.631	0.513	0.621
<i>Pinus sylvestris</i>	3	5	6	0.326	0.374	0.529
<i>Pinus uncinata</i>	4	5	10	0.571	0.319	0.398
<i>Pinus pinea</i>	3	3	3	0.616	0.886	0.704
<i>Pinus pinaster</i>	2	2	2	0.900	0.900	1.000
<i>Abies alba</i>	3	2	2	0.602	0.472	0.647
<i>Quercus ilex</i>	3	8	11	0.387	0.324	0.498
<i>Quercus suber</i>	4	4	7	0.436	0.424	0.487
<i>Quercus humilis</i>	3	3	3	0.781	0.391	0.369
<i>Quercus faginea</i>	2	2	2	0.873	0.791	0.751
<i>Fagus sylvatica</i>	2	3	3	0.467	0.378	0.648
Total	41	55	68	0.579	0.509	0.567
(b) Two-species forests (26) Forest type	Number of subtypes			Adjusted Rand index		
	CAP(d)	CAP(h)	CAS	CAP(d) vs. CAP(h)	CAP(d) vs. CAS	CAP(h) vs. CAS
<i>P. halepensis–P. nigra</i>	3	3	3	0.624	0.481	0.672
<i>P. halepensis–P. pinea</i>	3	4	4	0.573	0.519	0.809
<i>P. halepensis–Q. ilex</i>	4	5	6	0.455	0.470	0.439
<i>P. halepensis–Q. humilis</i>	2	2	1	0.508		
<i>P. halepensis–Q. faginea</i>	3	3	2	0.565	0.455	0.455
<i>P. nigra–P. sylvestris</i>	3	6	5	0.509	0.477	0.563
<i>P. nigra–Q. ilex</i>	3	3	4	0.890	0.774	0.753
<i>P. nigra–Q. humilis</i>	3	4	3	0.465	0.521	0.556
<i>P. nigra–Q. faginea</i>	2	2	2	0.533	0.716	0.533
<i>P. sylvestris–P. uncinata</i>	4	4	4	0.463	0.760	0.517
<i>P. sylvestris–A. alba</i>	1	1	1			
<i>P. sylvestris–Q. ilex</i>	2	3	3	0.378	0.480	0.601
<i>P. sylvestris–Q. humilis</i>	2	5	2	0.318	0.854	0.339
<i>P. sylvestris–Q. faginea</i>	2	2	2	0.681	0.681	1.000
<i>P. sylvestris–F. sylvatica</i>	2	1	1			
<i>P. uncinata–A. alba</i>	2	2	2	1.000	1.000	1.000
<i>P. pinea–P. pinaster</i>	1	1	1			
<i>P. pinea–Q. ilex</i>	2	3	4	0.477	0.543	0.583
<i>P. pinea–Q. suber</i>	2	2	2	0.821	0.909	0.738
<i>P. pinaster–Q. ilex</i>	1	1	1			
<i>P. pinaster–Q. suber</i>	2	2	1	0.765		
<i>A. alba–F. sylvatica</i>	1	2	2			0.425
<i>Q. ilex–Q. suber</i>	2	2	3	0.796	0.201	0.209
<i>Q. ilex–Q. humilis</i>	3	3	4	0.824	0.519	0.560
<i>Q. ilex–Q. faginea</i>	3	2	1	0.489		
<i>Q. suber–Q. humilis</i>	1	1	1			
Total	59	69	65	0.702	0.708	0.673

was 133 using CAS, 124 using CAP(h), and 100 using CAP(d), indicating a slightly greater diversity of bivariate distributions and height distributions, compared to the diversity of diameter distributions. We calculated the Rand Index (Hubert and Arabie 1985) to compare the classifications into subtypes obtained under each approach. This index allows comparing classifications with different numbers of groups and is ranged between 0 (indicating a match between the two classifications no greater than would be expected by chance) and 1 (indicating a perfect match). Surprisingly, the agreement between approaches was rather low for several forest types (Table 2). A high level of agreement was obtained for forests dominated by *Pinus pinea* or *Pinus pinaster*, probably due to a simplicity of forest structures derived from management, but in the remaining cases the agreement was moderate or even low. In particular, the subtypes obtained using either CAS, CAP(d), or CAP(h) were rather different for forests dominated by *Pinus uncinata*, *Pinus sylvestris*, *Quercus ilex*, *Q. suber*, and *Q. humilis*, indicating a high degree of complexity in forest structures.

### 3.4 Interpretation of CAS forest subtypes

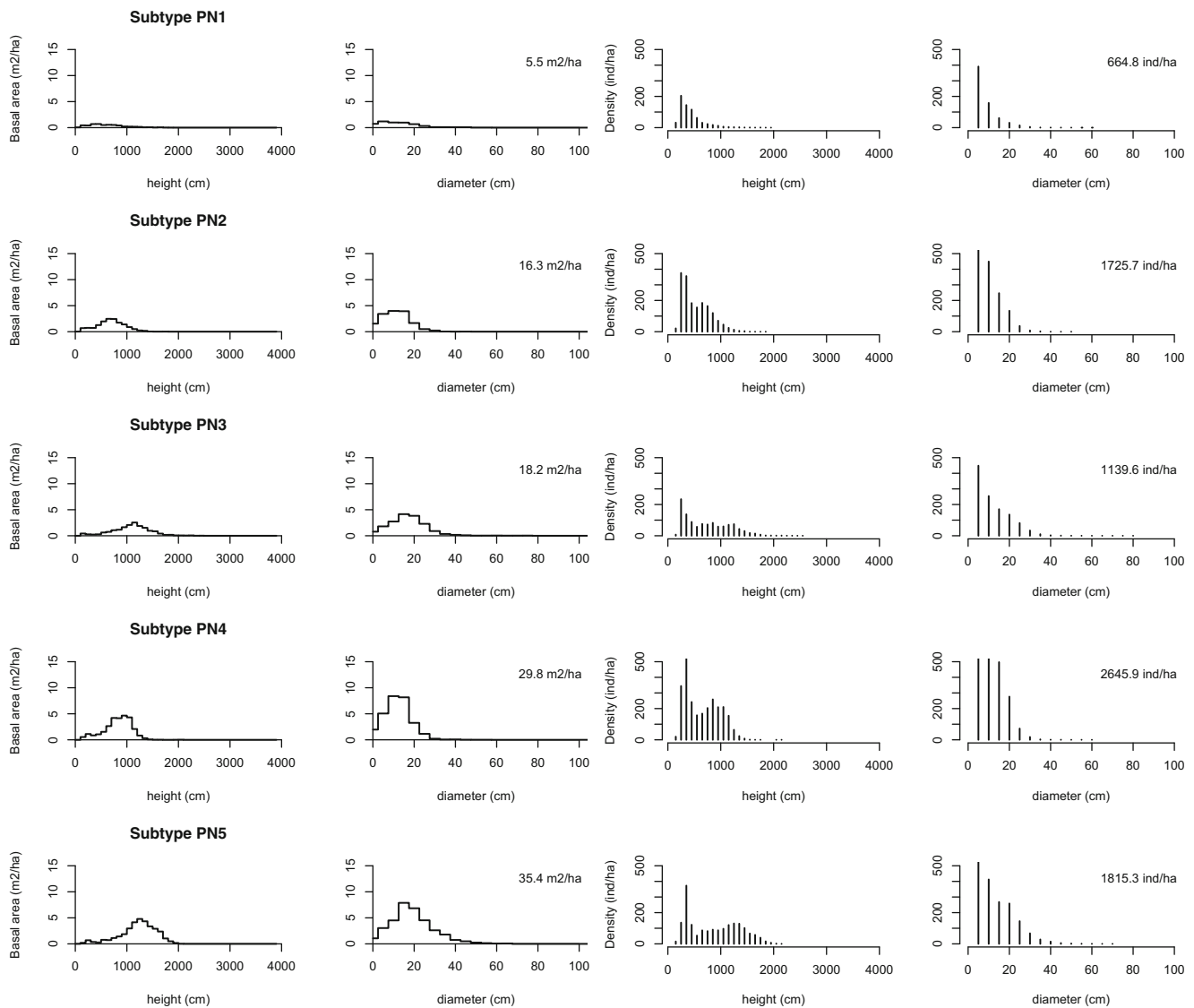
The characterization of all CAS-based subtypes is given in tables (structural statistics) and figures (density and basal area distributions across diameter and height classes) of SEM 2 and 3. Examining the distribution of density along tree size classes allowed us to interpret subtypes from silvicultural and ecological perspectives. The distinction between subtypes of single-species forests could arise due to a higher density only, without substantial differences in tree diameter or height distribution; or be due to differences in the distribution of one or both size variables. For instance, among the forest stands dominated by *P. nigra* (Fig. 4 and Table S2.2), four subtypes (PN1, PN2, PN4, and PN5) had diameter and height distributions that indicated different stages of development of even-aged forests, whereas a fifth subtype (PN3) included greater horizontal and vertical irregularity (i.e., uneven-aged stands). Among the even-aged subtypes, PN2 and PN4 reflected similar developmental stages but were noticeably different with regard to stand density. For its part, subtype PN5 not only involved the most mature stand development stage among even-aged forests, but could also be described as a mature and densely stocked uneven-aged stand.

Forests co-dominated by two species with similar degree of shade-tolerance included subtypes where both species had similar contribution to total basal area and subtypes when either one species or the other was more abundant than the other, but often with the two species having similar tree size distributions. For example, the interpretation of subtypes co-dominated by *P. sylvestris* and *P. uncinata* (Fig. 5 and

Table S3.10) led to the identification of one young and one more mature stand subtypes, both with similar occupancy of the two species (PSPU1 and PSPU2), and other two mature subtypes, one dominated by *P. sylvestris* (PSPU3), the other by *P. uncinata* (PSPU4). Similar interpretations could be made for pine-pine mixed forests (e.g., *P. nigra*–*P. sylvestris*, *P. halepensis*–*P. pinea*, or *P. halepensis*–*P. nigra*) and oak-oak mixed forests (e.g., *Q. ilex*–*Q. humilis*). In forests co-dominated by two species showing different degrees of shade-tolerance, trees of the shade-intolerant species were generally of larger size than trees of the shade-tolerant species. However, situations where the shade-tolerant species dominated the stand were also identified (see SEM3). For example, the six subtypes co-dominated by *P. halepensis* and *Q. ilex* (Fig. 6 and Table S3.3) reflected structural and compositional differences that in some cases indicated different successional stages between the shade-intolerant and the shade-tolerant species. A first subtype (PHQI1) matched with young and open stands co-dominated by the two species. Four of other subtypes (PHQI2, PHQI3, PHQI5, and PHQI6) showed a dominant layer of *P. halepensis* over a non-negligible layer of *Q. ilex*. These subtypes were differentiated by the size and density of pines. The remaining subtype (PHQI4) suggested a more advanced successional stage in which *Q. ilex*, although growing in a lower stratum, had already become dominant in terms of basal area. We reached similar interpretations for several pine-oak mixed forests (i.e., combinations between *P. halepensis*, *P. nigra*, *P. sylvestris*, or *P. pinea*, on one side, and *Q. ilex*, *Q. suber*, *Q. humilis*, or *Q. faginea*, on the other), but also for other kinds of mixed forests (e.g., *P. sylvestris*–*F. sylvatica* or *P. uncinata*–*A. alba*).

## 4 Discussion

The approach presented here attempts to simplify the task of determining the resemblance between forest stands. Instead of requiring the selection of multiple variables, often having different units and therefore being difficult to consider simultaneously, our approach allows the analyst to compare the vertical and horizontal structure of forest plots by performing a small set of choices (one or two size variables, an abundance variable and a compositional resolution). Similarly to other approaches based on the Lorenz curves (Valbuena et al. 2013), our approach tries to exploit all the information in the uni- or bivariate distribution of diameters and heights. Thanks to the flexibility of dissimilarity indices based on CAPs and CASs, our approach allows comparing and classifying forest stands of very different kinds. Therefore, even if its spatial grain is the forest plot, the same framework can be applied consistently for large areas if combined with data from large-scale forest surveys such as national forest inventories. Thus,



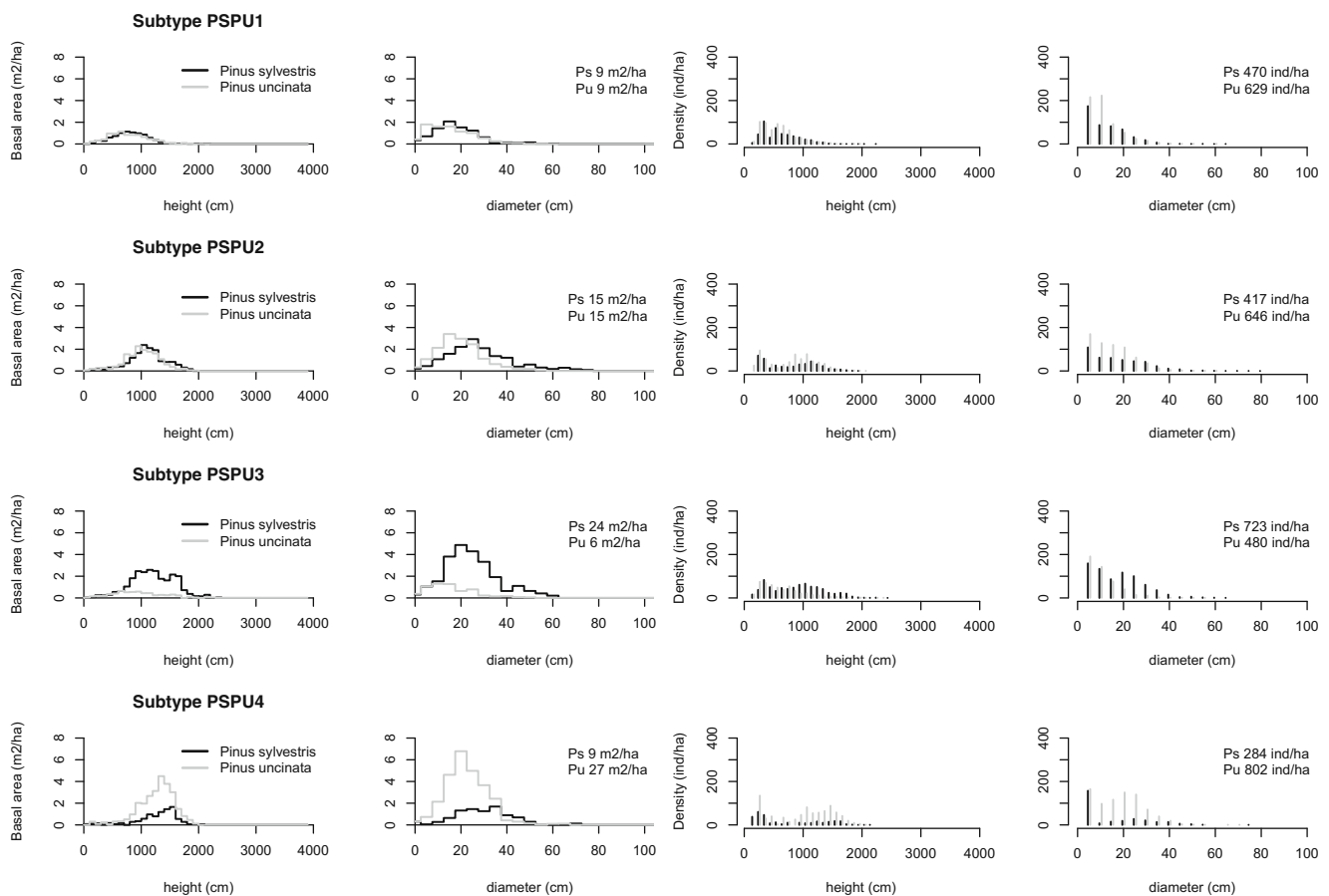
**Fig. 4** Five subtypes of forests dominated by *Pinus nigra* (PN), ordered by increasing stand's basal area (see Table S2.2). For each subtype, we show the distribution of basal area (m<sup>2</sup>/ha) among 1-m height classes and

5-cm diameter classes (first and second columns, respectively) and the distribution of density (ind./ha) among height and dbh classes (third and fourth columns, respectively)

it could be used to complement coarse forest typologies (Barbati et al. 2014) with finer classification levels.

One of the main advantages of our approach is its flexibility with regard to the kind of input data that it accepts and the number of ways to analyze it. Multiple choices can be made for size (i.e., vertical structure) and abundance (i.e., horizontal structure) variables, as well as for the definition of compositional classes. Here, we used height and diameter, which seem natural choices of size variables when comparing forest plots using individual tree data. Comparisons between forest stands can be done using both variables and cumulative abundance surfaces (CASs), if both variables have been measured, or using

cumulative abundance profiles (CAPs) if only one of them is available. If, for a given forest data set, the vertical structure is defined as vegetation strata, one can adapt the method by defining the highest stratum reached by plants as the size variable. Our approach is also flexible with respect to the variable chosen to represent horizontal structure (e.g., basal area, projected cover or number of individuals). For example, if forest data consists of records from Airborne LiDAR, one could classify them into forest types choosing height classes as size variable and the number of first LiDAR returns corresponding to each height class as abundance (i.e., horizontal structure) variable. With regard to compositional classes, species



**Fig. 5** Four subtypes of forests co-dominated by *P. sylvestris* (PS) and *P. uncinata* (PU) (both shade-intolerant species), ordered by increasing basal area (see Table S3.10). For each subtype and species, we show the distribution of basal area ( $\text{m}^2/\text{ha}$ ) among 1-m height classes and 5-cm

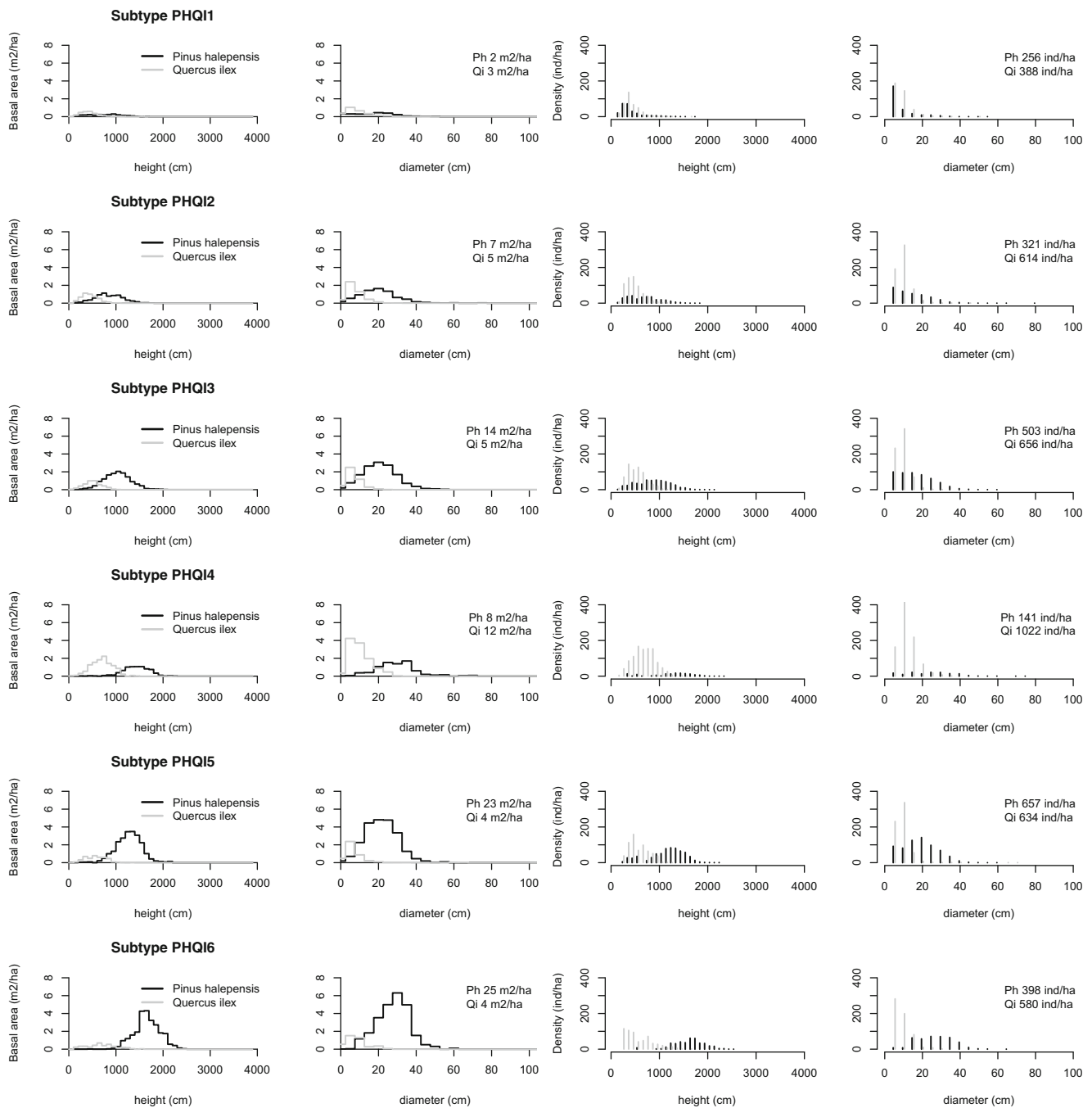
diameter classes (first and second columns, respectively) and the distribution of density ( $\text{ind}/\text{ha}$ ) among height and dbh classes (third and fourth columns, respectively)

taxonomic identity seems a natural choice, but other classes such as functional groups could be used instead. If the presence of dead trees, such as standing snags or fallen logs is deemed relevant (e.g., if forest types need to represent successional stages or habitat quality for wildlife; McComb and Lindenmayer 1999), one can add compositional classes to represent them (diameter and height/length of the dead tree could still be used as size variables). Another source of flexibility comes from the possibility of altering the scale of the size variables or the abundance variable. For example, in our application with forest inventory data we defined quadratic diameter bins, but linear or logarithmic bins would also be possible.

We found that tree diameters did not strongly correlate (monotonically) with heights in our forest inventory data. Further, classifications obtained using the CAS framework were often not very similar to classifications obtained using the original framework (CAP), even if the number of subtypes recognized was often the same. Taking into account the bivariate distribution of diameters and heights may better

accommodate to the complexity of forest structures than univariate distributions. However, the observed higher overall number of subtypes could be an artifact derived from inaccuracies in the estimation of tree sizes. Although we focused here on how dissimilarities are defined, other important choices for the definition of typologies are those of the unsupervised classification algorithm and its parameters. With the chosen values for average Silhouette (0.2) and minimum membership (10 plots), the number of subtypes of some forests, like those of *Pinus halepensis*, is probably too high for practical use. Larger values of these thresholds would lead to fewer subtypes in these cases, but could compromise the recognition of subtypes less represented in the forest inventory data. Practitioners are welcome to search for parameter combinations that produce interpretable and operational typologies.

While general and flexible, our approach has limitations in the kind of structural variables that can be included (McElhinny et al. 2005). Because it focuses on the structure provided by plants (i.e., trees, shrubs, and even



**Fig. 6** Six subtypes of forests co-dominated by *P. halepensis* (PH; shade-intolerant) and *Q. ilex* (QI; shade-tolerant), ordered by increasing basal area (see Table S3.3). For each subtype and species, we show the distribution of

basal area ( $m^2/ha$ ) among 1-m height classes and 5-cm diameter classes (first and second columns, respectively) and the distribution of density (ind./ha) among height and dbh classes (third and fourth columns, respectively)

herbs), it does not accommodate easily to other structural features such as litter cover or ground cover. Species diversity and variations in plant dimensions are fully represented in our approach, but spacing between plants is only represented in terms of density (Pommerening 2002). Unlike in Ngo Bieng et al. (2006), patterns of spatial clumping or segregation of

plants are not taken into account for the definition of forest types. Another limitation of the classification approach employed here is that it does not produce determination keys easy to apply for field sampling or assignment rules for mapping with remote sensing data (Valbuena et al. 2013). Supervised classification might be needed to produce

membership rules for mapping or keys for field determination. Finally, our approach relies on a sufficiently accurate estimation of the uni- or bivariate distribution of tree diameters and heights in the input data. Limitations derived from field sampling protocols, such as a small plot size or the nested plot structure of our forest inventory data, may be a source of error in the estimation of dissimilarities and the resulting forest classes (Nanos and de Luna 2017).

In our opinion, the approach presented here allows the description and classification of forest stands (i.e., through forest plot data) at a sufficient level of detail to facilitate forest planning and decision-making in biogeographic contexts where fine-scale variability of topographic, climatic, and land-use factors leads to large variability in forest structure and composition (e.g., Bruciamacchie 2001). In particular, it allows classifying mixed-forests according to structural attributes over large areas, which remained a difficult and challenging task until now. The compositional and structural typologies obtained using our approach can provide forest managers with critical information for the efficient organization of silvicultural operations in time and space. For example, they could be used to develop type-based management guidelines aiming to either maintain a certain stock and size distribution of trees (e.g., Gove 2004) or emulate natural succession (e.g., Attiwill 1994). If necessary, the structural and compositional dissimilarity between the managed stand and specific forest types could be used to accurately monitor deviations from planned sequences. Finally, our approach could also be used in the study of forest dynamics (De Cáceres et al. 2019). The definition of a forest typology from consecutive surveys of permanent forest plots could provide critical information on the prevalence of regressive and successional series across a study area, and allow the evaluation of the most important factors driving such changes, including the role of disturbances (Sánchez-Pinillos et al. 2019).

**Acknowledgements** The authors would like to thank Mario Beltrán (CTFC) for useful discussions around the method and its potential applications.

**Funding** The study was supported by projects 979S/2013 (Autonomous Agency of National Parks, Spanish Ministry of Agriculture Food and Environment) and a Spanish “Ramon y Cajal” fellowship to M.D.C (RYC-2012-11109).

**Data availability** The Spanish National Forest Inventory data are publicly available on the following website: <http://www.mapama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/>.

## Compliance with ethical standards

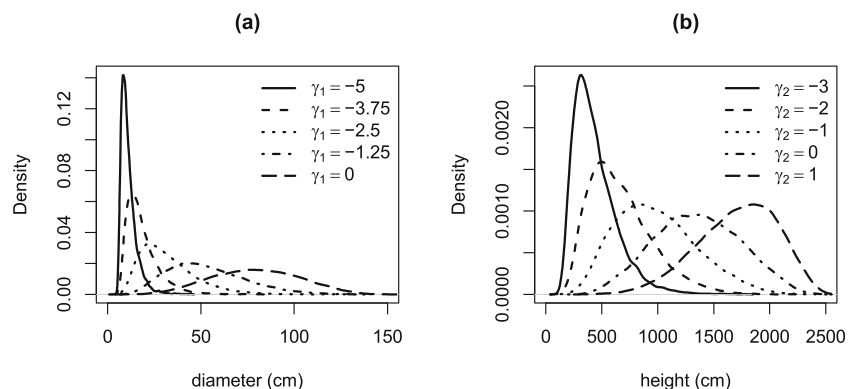
**Conflicts of interest** The authors declare that they have no conflict of interest.

## Annex 1. Simulation study

### Simulated treatments

We simulated forest tree data using Johnson’s Sbb distribution (Johnson 1949; Schreuder and Hafley 1977) to study the appropriateness of  $D_{man}$  and  $D_{bray}$  to reflect differences in terms of total basal area, tree diameter distribution and tree height distribution. For simplicity, we did not include differences in composition; these were considered in De Cáceres et al. (2013). In all simulations some parameters of the Sbb distribution were fixed ( $\rho = 0.8$ ,  $\delta_1 = \delta_2 = 1.5$ ). We changed parameters  $\gamma_1$  and  $\gamma_2$  of the Sbb distribution to simulate variations in the distribution of tree diameter and tree height. Five treatments were considered for each size variable (see Fig. 7). Five treatments of stand basal area were also considered: 5, 10, 15, 20, and 25 m<sup>2</sup>/ha. For each simulated stand we generated tree individuals until to the stand’s target basal area was met.

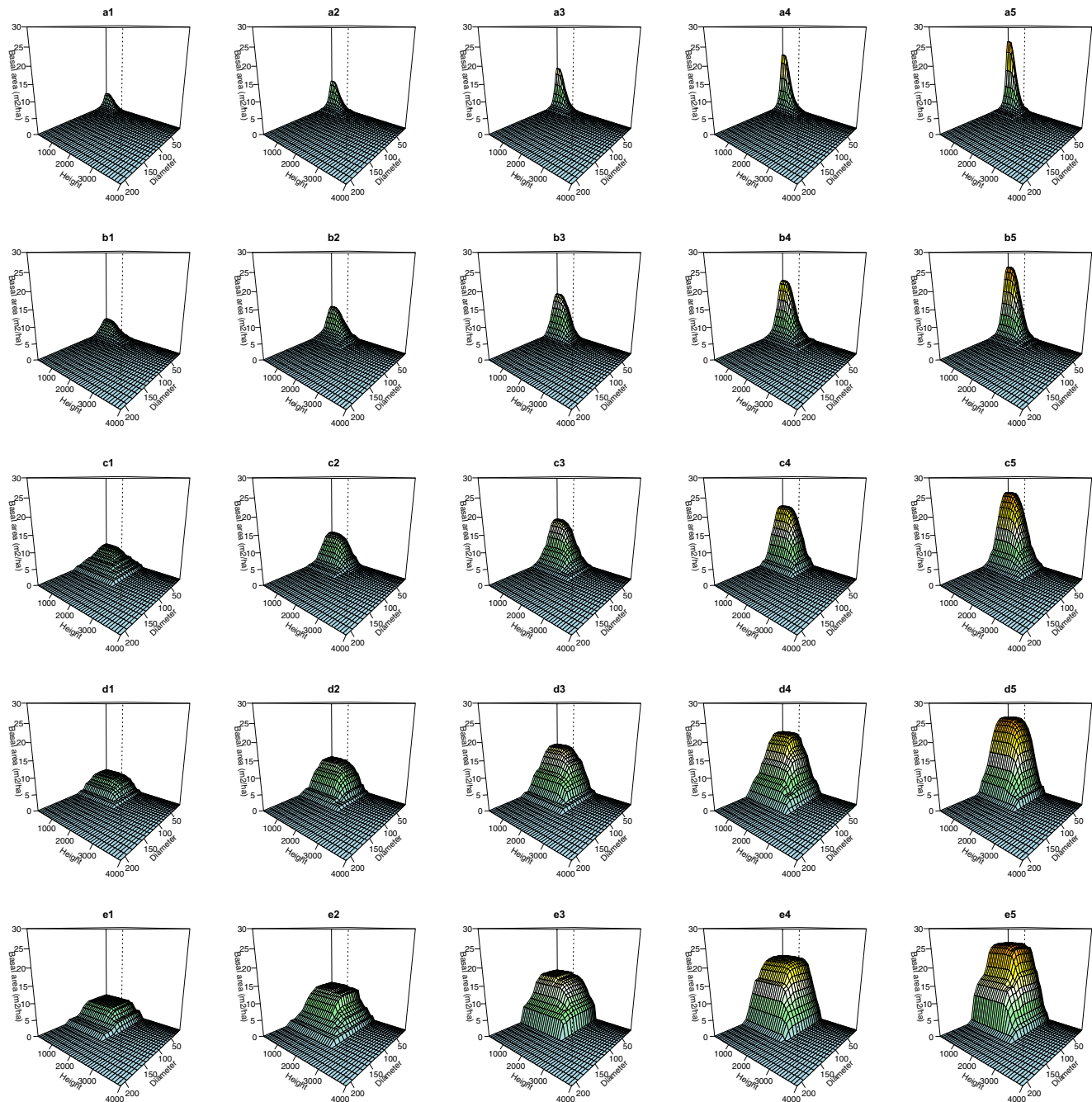
**Fig. 7** Marginal probability density functions for tree diameters and tree heights obtained by setting different values for Sbb parameters  $\gamma_1$  and  $\gamma_2$ , respectively



## First simulation experiment

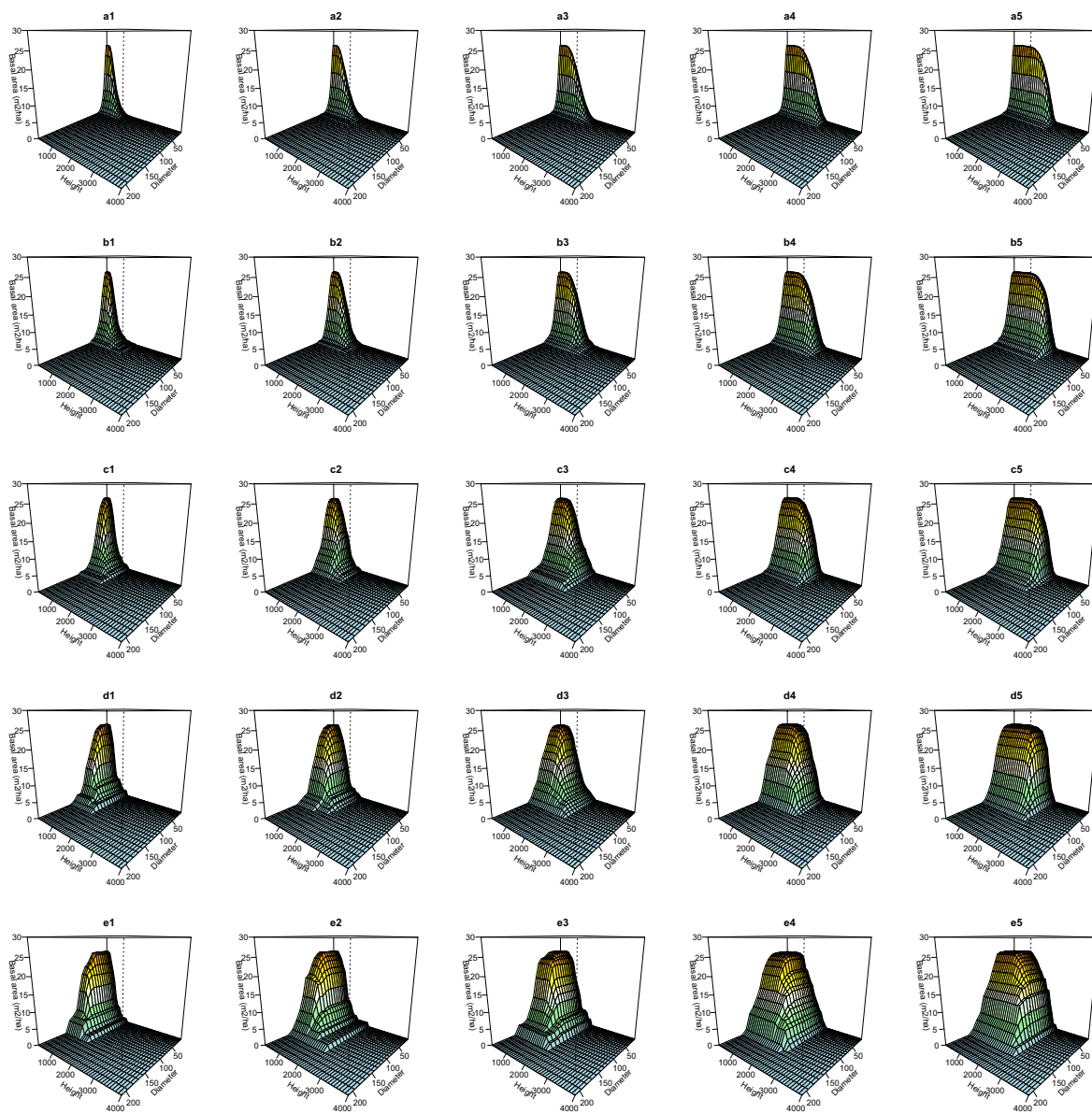
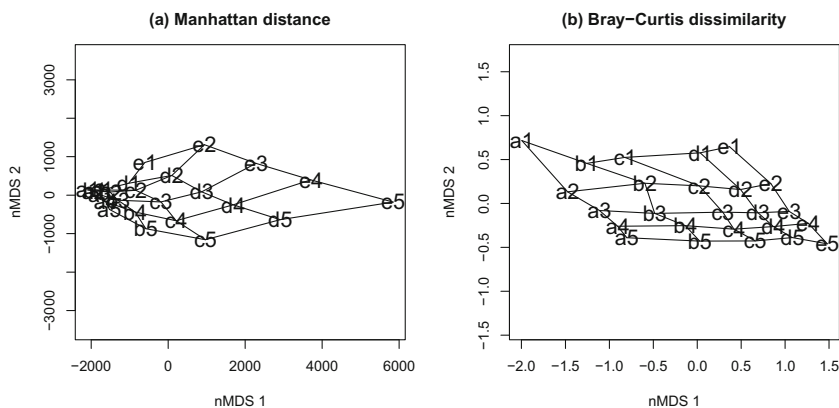
In a first experiment, we generated 25 different stands by crossing the five treatments in stand basal area (labeled ‘1’ to ‘5’) with five treatments in tree size distribution (labeled ‘a’ to ‘e’), assuming that when trees were larger in diameter they were also larger in height (i.e., in treatment ‘a’  $\gamma_1 = -5$  and  $\gamma_2 = -3$ ; in ‘b’  $\gamma_1 = -3.75$  and  $\gamma_2 = -2$ ; etc.). Thus, simulated stands were composed of few small individuals, many small individuals, few large individuals, or many large individuals; with intermediate situations

being also covered. The CAS for each plot was built using basal area as abundance variable and tree diameter and tree height as size variables. We defined quadratic diameter bins, meaning that more classes were defined to distinguish thin trees than to distinguish thick trees. Specifically, diameter limits (in cm) for class  $r$  were defined as  $4 + [(0.5 \cdot (r-1))^2, (0.5 \cdot r)^2]$  cm. The resulting classes were: 4–6.25 cm, 6.25–9 cm, 9–12.25 cm, 12.25–16 cm, 16–20.25 cm etc. Height classes were defined linearly using 100-cm bins. Figure 8 shows the CAS corresponding to each of the 25 simulated stands.



**Fig. 8** Cumulative abundance surfaces (CASs) corresponding to the 25 simulated forest stands of the first experiment. Each of the five rows corresponds to the bivariate distributions of diameter and height presented in Fig. 7. Columns differ in the stand’s basal area simulated (5, 10, 15, 20, and 25 m<sup>2</sup>/ha)

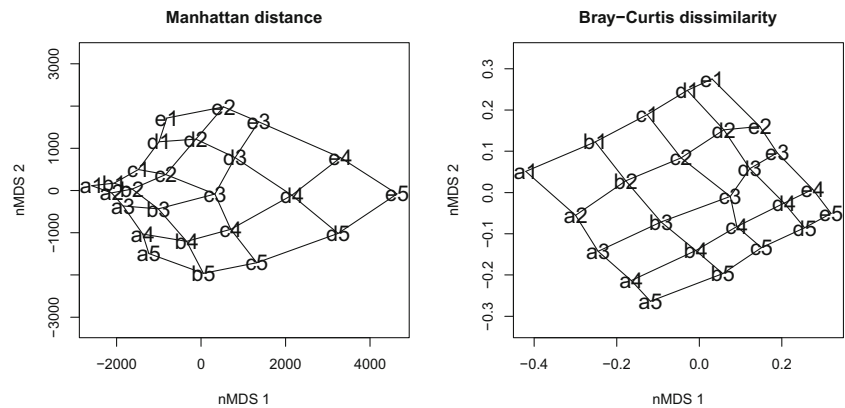
**Fig. 9** Non-metric multidimensional scaling (nMDS) of the dissimilarity matrix obtained using either  $D_{man}$  or  $D_{bray}$ . Edges indicate stands that are contiguous along treatment sequences



**Fig. 10** Cumulative abundance surfaces (CASs) of the 25 simulated forest stands of the second experiment. Each panel corresponds to a combination of diameter distribution (rows) and height distribution (columns) among those presented in Fig. 7. All stands have a basal area of 25 m<sup>2</sup>/ha



**Fig. 11** Non-metric multidimensional scaling (nMDS) of the dissimilarity matrix obtained using either  $D_{man}$  or  $D_{bray}$ . Labels indicate forest stands simulated using combinations of diameter distribution (a–e) and height distribution (1–5). Edges indicate stands that are contiguous along treatment sequences



then calculated the resemblance between all pairs of simulated stands using either  $D_{man}$  or  $D_{bray}$ . We used Kruskal's non-metric multidimensional scaling (nMDS) to display each of the resulting dissimilarity matrices in a two-dimensional scatter plot (Fig. 9). Both coefficients were responsive to changes in total basal area and tree size. However, in the dissimilarity matrix obtained using  $D_{man}$  the distances due to differences in tree size were smaller between stands with small basal area than between stands with large basal area (i.e., compare sequences a1–e1 and a5–e5 in Fig. 9). Similarly, the distances due to differences in basal area were smaller between stands composed of small individuals than between stands composed of large individuals (i.e., compare sequences a1–a5 and e1–e5 in Fig. 9a). These patterns did not occur in the case of  $D_{bray}$ . Contrastingly, with this coefficient the dissimilarity between a pair stands differing in basal area was somewhat larger when the two stands had 5 and 10 m<sup>2</sup>/ha, respectively (i.e., a1 vs. a2, b1 vs. b2, ...; Fig. 9b), than when they had 20 and 25 m<sup>2</sup>/ha, respectively (i.e., a4 vs. a5, b4 vs. b5, ...; Fig. 9b). Similarly, the dissimilarity between pairs of plots differing in tree distribution was larger between small tree treatments (i.e., a1 vs. b1, a2 vs. b2, ...; Fig. 9b) than between large tree treatments (i.e., d1 vs. e1, d2 vs. e2, ...; Fig. 9b).

## Second simulation experiment

In a second experiment, we calculated the resemblance among stands having all the same basal area (25 m<sup>2</sup>/ha) but differing in height distribution, diameter distribution, or both. In this case, we generated stands by crossing the five treatments in diameter distribution (labeled 'a' to 'e') with the five treatments in tree size distribution (labeled '1' to '5'). Thus, stands could be composed of short and thin trees, tall and thin trees, short and thick trees, or tall

and thick trees; with intermediates situations being also covered. The CAS for each plot was built as before (Fig. 10).

As with the first experiment, we calculated the resemblance between all pairs of simulated stands using either  $D_{man}$  or  $D_{bray}$  and displayed the resulting dissimilarity matrices using non-metric multidimensional scaling (Fig. 11).

Both coefficients were responsive to changes in tree diameter and tree height. Although the effects were less obvious, the behavior of both indices was similar to the first experiment. With  $D_{man}$  the distances due to differences in diameter distribution were smaller between stands with short trees than between stands with tall trees and, similarly, the distances due to differences in tree height were smaller between stands composed of thin individuals than between stands composed of thick individuals (Fig. 11a). With  $D_{bray}$  the dissimilarity between stands differing in tree diameter was larger between thin tree treatments than between thick tree treatments and, similarly, the dissimilarity between plots differing in tree height was larger between short tree treatments than between tall tree treatments (Fig. 11b).

## References

- Attwill PM (1994) The disturbance of forest ecosystems: the ecological basis for conservative management. For Ecol Manag 63:247–300. [https://doi.org/10.1016/0378-1127\(94\)90114-7](https://doi.org/10.1016/0378-1127(94)90114-7)
- Aunós A, Martínez E, Blanco R (2007) Tipología selvícola para los abetales españoles de *Abies alba*. Mill Investig Agrar Sist Recur For 16:52–64
- Barbati A, Corona P, Marchetti M (2007) A forest typology for monitoring sustainable forest management: the case of European forest types. Plant Biosyst 141:93–103. <https://doi.org/10.1080/11263500601153842>
- Barbati A, Marchetti M, Chirici G, Corona P (2014) European Forest types and Forest Europe SFM indicators: tools for monitoring

- progress on forest biodiversity conservation. For *Ecol Manag* 321: 145–157. <https://doi.org/10.1016/j.foreco.2013.07.004>
- Bebi P, Kienast F, Schönenberger W (2001) Assessing structures in mountain forests as a basis for investigating the forests' dynamics and protective function. For *Ecol Manag* 145:3–14. [https://doi.org/10.1016/S0378-1127\(00\)00570-3](https://doi.org/10.1016/S0378-1127(00)00570-3)
- Bottalico F, Travaglini D, Chirici G, Marchetti M, Marchi E, Nocentini S, Corona P (2014) Classifying silvicultural systems (coppices vs. high forests) in mediterranean oak forests by airborne laser scanning data. *Eur J Remote Sens* 47:437–460. <https://doi.org/10.5721/EuJRS20144725>
- Bravo-Oviedo A, Pretzsch H, Ammer C, Andenmatten E, Barbati A, Barreiro S, Brang P, Bravo F, Coll L, Corona P, den Ouden J, Ducey MJ, Forrester DI, Giergiczy M, Jacobsen JB, Lesinski J, Löf M, Mason WL, Matovic B, Metslaid M, Morneau F, Motiejunaite J, O'Reilly C, Pach M, Ponette Q, del Rio M, Short I, Skovsgaard JP, Soliño M, Spathelf P, Sterba H, Stojanovic D, Strelcova K, Svoboda M, Verheyen K, von Lüpke N, Zlatanov T (2014) European mixed forests: definition and research perspectives. For *Syst* 23:518–533. <https://doi.org/10.5424/fs/2014233-06256>
- Bray RJ, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349
- Bruciamacchie M (2001) Les typologies de peuplements, 20 ans après. *Rev For Fr* III:449–458
- Cajander AK (1949) Forest types and their significance. *Acta For Fenn* 56:1–71
- Casals P, Camprodon J, Caritat A, Rios AI, Guixé D, Garcia-Marti X, Martín-Alcón S, Coll L (2015) Forest structure of mediterranean yew (*Taxus baccata* L.) populations and neighbor effects on juvenile yew performance in the NE Iberian Peninsula. For *Syst* 24:e042. <https://doi.org/10.5424/fs/2015243-07469>
- Corona P (2016) Consolidating new paradigms in large-scale monitoring and assessment of forest ecosystems. *Environ Res* 144:8–14. <https://doi.org/10.1016/j.envres.2015.10.017>
- De Cáceres M, Wiser SK (2012) Towards consistency in vegetation classification. *J Veg Sci* 23:387–393. <https://doi.org/10.1111/j.1654-1103.2011.01354.x>
- De Cáceres M, Font X, Oliva F (2010) The management of vegetation classifications with fuzzy clustering. *J Veg Sci* 21:1138–1151. <https://doi.org/10.1111/j.1654-1103.2010.01211.x>
- De Cáceres M, Legendre P, He F (2013) Dissimilarity measurements and the size structure of ecological communities. *Methods Ecol Evol* 4: 1167–1177. <https://doi.org/10.1111/2041-210X.12116>
- De Cáceres M, Chytrý M, Agrillo E et al (2015) A comparative framework for broad-scale plot-based vegetation classification. *Appl Veg Sci* 18:543–560. <https://doi.org/10.1111/avsc.12179>
- De Cáceres M, Coll L, Legendre P et al (2019) Trajectory analysis in community ecology. *Ecol Monogr*. <https://doi.org/10.1002/ecm.1350>
- DGCN (2005) Tercer Inventario Forestal Nacional (1997–2007): Catalunya. Dirección General de Conservación de la Naturaleza. Ministerio de Medio Ambiente, Madrid
- Faith D, Austin M, Belbin L, Margules C (1985) Numerical classification of profile attributes in environmental studies. *J Environ Manag* 20: 73–85
- Falkowski MJ, Evans JS, Martinuzzi S, Gessler PE, Hudak AT (2009) Characterizing forest succession with lidar data: an evaluation for the inland northwest, USA. *Remote Sens Environ* 113:946–956. <https://doi.org/10.1016/j.rse.2009.01.003>
- Forest Europe, UNECE and FAO (2011) State of Europe's Forests 2011. Status and trends in sustainable forest management in Europe. pp 344. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Giannetti F, Barbati A, Mancini LD, Travaglini D, Bastrup-Birk A, Canullo R, Nocentini S, Chirici G (2018) European Forest types: toward an automated classification. *Ann For Sci* 75:6. <https://doi.org/10.1007/s13595-017-0674-6>
- Godinho-Ferreira P, Azevedo A, Rego F (2005) Carta da Tipologia Florestal de Portugal Continental. *Silva Lusit* 13:1–34
- Gove JH (2004) Structural stocking guides: a new look at an old friend. *Can J For Res* 34:1044–1056. <https://doi.org/10.1139/x03-272>
- Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2:193–218
- James FC, Wamer NO (1982) Relationships between temperate forest bird communities and vegetation structure. *Ecology* 63:159–171
- Johnson NL (1949) Bivariate distributions based on simple translation system. *Biometrika* 36:297–304
- Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley & Sons Inc, New York
- Larsen JB, Nielsen AB (2007) Nature-based forest management—where are we going? Elaborating forest development types in and with practice. For *Ecol Manag* 238:107–117. <https://doi.org/10.1016/j.foreco.2006.09.087>
- Legendre P, De Cáceres M (2013) Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecol Lett* 16: 951–963. <https://doi.org/10.1111/ele.12141>
- Legendre P, Legendre L (2012) Numerical Ecology, 3rd Eng. Elsevier Science BV, Amsterdam
- Martín-Alcón S, Coll L, Aunós A (2012) A broad-scale analysis of the main factors determining the current structure and understory composition of Catalanian sub-alpine (*Pinus uncinata* ram.) forests. *Forestry* 85:225–236. <https://doi.org/10.1093/forestry/cpr067>
- Martín-Alcón S, Coll L, De Cáceres M et al (2015) Combining aerial LiDAR and multispectral imagery to assess postfire regeneration types in a Mediterranean forest. *Can J For Res* 45:856–866
- McComb W, Lindenmayer D (1999) Dying, dead, and down trees. In: Maintaining Biodiversity in Forest Ecosystems. Hunter, M. L. (ed.). Cambridge University Press, Cambridge, pp 335–372
- McElhinny C, Gibbons P, Brack C, Bauhus J (2005) Forest and woodland stand structural complexity: its definition and measurement. For *Ecol Manag* 218:1–24. <https://doi.org/10.1016/j.foreco.2005.08.034>
- Nanos N, de Luna SS (2017) Fitting diameter distribution models to data from forest inventories with concentric plot design. For *Syst* 26:1–13. <https://doi.org/10.5424/fs/2017262-10486>
- Ngo Bieng MA, Ginisty C, Goreaud F, Perot T (2006) First typology of oak and scots pine mixed stands in Orléans Forest (France), based on the canopy spatial structure. *New Zeal J For Sci* 36:325–346
- O'Hara KL, Latham PA, Hessburg P, Smith BG (1996) A structural classification for inland northwest forest vegetation. *West J Appl For* 11:97–102
- Odum EP (1950) Bird populations of the highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology* 31:587–605
- Pommerening A (2002) Approaches to quantifying forest structures. *Forestry* 75:305–324. <https://doi.org/10.1093/forestry/75.3.305>
- Reque JA, Bravo F (2008) Identifying forest structure types using National Forest Inventory Data: the case of sessile oak forest in the Cantabrian range. *Investig Agrar Sist Recur For* 17:105–113
- Sánchez-Pinillos M, Leduc A, Ameztegui A, Kneeshaw D, Lloret F, Coll L (2019) Resistance, resilience, or change: post-disturbance dynamics of boreal forests after insect outbreaks. *Ecosystems*. In Press
- Schreuder HT, Hafley WL (1977) Bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics* 33:471–478
- Tamás J, Podani J, Csontos P (2001) An extension of presence/absence coefficients to abundance data: a new look at absence. *J Veg Sci* 12: 401–410. <https://doi.org/10.2307/3236854>
- Valbuena R, Packalen P, Mehtätalo L, García-Abril A, Maltamo M (2013) Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. *Can J For Res* 43:1063–1074

- Valbuena R, Maltamo M, Packalen P (2016) Classification of multilayered forest development classes from low-density national airborne lidar datasets. *Forestry* 89:392–401. <https://doi.org/10.1093/forestry/cpw010>
- Valbuena R, Maltamo M, Mehtätalo L, Packalen P (2017) Key structural features of boreal forests may be detected directly using L-moments from airborne lidar data. *Remote Sens Environ* 194:437–446. <https://doi.org/10.1016/j.rse.2016.10.024>
- Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
- Yao J, Zhang C, De Cáceres M et al (2019) Variation in compositional and structural components of community assemblage and its determinants. *J Veg Sci*. <https://doi.org/10.1111/jvs.12708>
- Yılmaz OY, Yılmaz H, Akyüz YF (2018) Effects of the overstory on the diversity of the herb and shrub layers of Anatolian black pine forests. *Eur J For Res* 137:433–445. <https://doi.org/10.1007/s10342-018-1114-3>
- Youngblood A (1993) Community type classification of forest vegetation in young, mixed stands, interior Alaska. Res. Pap. PNW-RP-458. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland 49 p

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.