



# Estimating tropical tree diversity indices from forestry surveys: A method to integrate taxonomic uncertainty



Stéphane Guitet<sup>a,b,\*</sup>, Daniel Sabatier<sup>c</sup>, Olivier Brunaux<sup>a</sup>, Bruno Hérault<sup>d</sup>, Méline Aubry-Kientz<sup>d</sup>, Jean-François Molino<sup>c</sup>, Christopher Baraloto<sup>e</sup>

<sup>a</sup> ONF, Direction Régionale de Guyane, 97307 Cayenne, French Guiana

<sup>b</sup> INRA, UMR Amap, TA A51/PS2, 34398 Montpellier Cedex 5, France

<sup>c</sup> IRD, UMR Amap, TA A51/PS2, 34398 Montpellier Cedex 5, France

<sup>d</sup> CIRAD, UMR EcoFoG, 97387 Kourou Cedex, French Guiana

<sup>e</sup> INRA, UMR EcoFoG, 97387 Kourou Cedex, French Guiana

## ARTICLE INFO

### Article history:

Received 19 March 2014

Received in revised form 21 May 2014

Accepted 24 May 2014

### Keywords:

Diversity

French Guiana

Tree community

Rainforest

Monte-Carlo process

Landscape scale

## ABSTRACT

Analyses of tree diversity and community composition in tropical rain forests are usually based either on general herbarium data or on a restricted number of botanical plots. Despite their high taxonomic accuracy, both types of data are difficult to extrapolate to landscape scales. Meanwhile, forestry surveys provide quantitative occurrence data on large areas, and are thus increasingly used for landscape-scale analyses of tree diversity. However, the reliability of these approaches has been challenged because of the ambiguity of the common (vernacular) names used by foresters and the complexity of tree taxonomy in those hyper-diverse communities.

We developed and tested a novel approach to evaluate taxonomic reliability of forestry surveys and to propagate the resulting uncertainty in the estimates of several diversity indicators (alpha and beta entropy, Fisher-alpha and Sørensen similarity). Our approach is based on Monte-Carlo processes that simulate communities by taking into account the expected accuracy and reliability of common names. We tested this method in French Guiana, on 9 one-hectare plots (4279 trees – DBH  $\geq$  10 cm) for which both common names and standardized taxonomic determinations were available. We then applied our method of community simulation on large forestry inventories (560 ha) at the landscape scale and compared the diversity indices obtained for 10 sites with those computed from precise botanical determination situated at the same localities.

We found that taxonomic reliability of forestry inventories varied from 22% (species level) to 83% (family level) in this Amazonian region. Indices computed directly with raw forestry data resulted in incorrect values, except for Gini–Simpson beta-diversity. On the contrary, our correction method provides more accurate diversity estimates, highly correlated with botanical measurements, for almost all diversity indices at both regional and local scales. We obtained a robust ranking of sites consistent with those shown by botanical inventories.

These results show that (i) forestry inventories represent a significant part of taxonomic information, (ii) the relative diversity of regional sites can be successfully ranked using forestry inventory data using our method and (iii) forestry inventories can valuably contribute to the detection of large-scale diversity patterns when biases are well-controlled and corrected.

The tools we developed as R-functions are available in supplementary material and can be adapted with local parameters to be used for forest management and conservation issues in other regional contexts.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

The conservation of biodiversity is one of the major challenges in tropical rainforests (Koh and Sodhi, 2010). To progress in this direction, policy makers and forest managers need practical tools

\* Corresponding author at: INRA, UMR Amap, TA A51/PS2, 34398 Montpellier Cedex 5, France. Tel.: +33 467615800.

E-mail address: [stephane.guitet@cirad.fr](mailto:stephane.guitet@cirad.fr) (S. Guitet).

to highlight future protected-areas where conservation issues should have priority. Even though genetic diversity represents a fundamental component of biodiversity (Moritz and Faith, 1998), estimating species diversity, i.e., detecting richest or very original (i.e., dissimilar) communities, remains the most tractable method to achieve this objective (Myers et al., 2000).

Clearly, a complete census of biodiversity, including different plants and animals groups, is usually impossible for almost any operational project especially in mega-diverse regions (Landeiro et al., 2012). Instead, practical biodiversity indicators have often been proposed, based on a single group or on a reduced subset of surrogate study groups (Kessler et al., 2011) or on a simplified para-taxonomic assessment (Basset et al., 2004). Among these indicator-groups, tree composition data are widely collected in tropical forests and as a consequence are frequently used for large scale diagnostics (e.g., ter Steege et al., 2006, 2013; Stropp et al., 2009). Moreover, even if tree species communities probably do not provide the most efficient surrogate for overall biodiversity (Kessler et al., 2011), most of assessments about biodiversity in tropical forest are already made with tree species communities, given their clear linkages with at least some other taxonomic groups and with important ecosystem functions such as primary production. ... Nevertheless, even for tree communities, installing botanical plots and collecting vouchers in tropical forest requires fastidious and expensive fieldwork, in addition to the contributions of a limited number of taxonomists (Baraloto et al., 2013). Therefore, as well as for the other biological groups, tree communities data are mostly limited to forest sites intensively monitored for research purposes and rarely replicated or applied on operational scales for forest management, i.e., hundreds or thousands of hectares (Kessler et al., 2011). As a result, practitioners and managers have to find alternative methods like “rapid assessments” (Higgins and Ruokolainen, 2004), to evaluate tree diversity and changes in community composition at larger scales.

Large scale forestry surveys based on common names from vernacular sources represent a promising source of data to estimate trees' diversity because of their large coverage and low cost (ter Steege et al., 2000). Forestry surveys have already been used in various contexts to investigate community originality (beta-diversity) and local diversity patterns (alpha-diversity) at both continental and regional scales (e.g., ter Steege, 1998; ter Steege et al., 2006; Rejou-Mechain et al., 2008; Emilio et al., 2010; Fayolle et al., 2012; Moscoso et al., 2013). However, the reliability of these approaches is rarely evaluated, despite the ambiguity of the common names used by foresters and the complexity of tree taxonomy in tropical forests (Lacerda and Nimmo, 2010; Gomes et al., 2013). For example, in Central Africa, the reliability of commercial inventories was estimated to be lower than 62% for species rank and 76% for family rank (Rejou-Mechain et al., 2011). In Amazonia, to evaluate tree diversity and distribution patterns, some authors have assumed that more than 95% of trees in forestry surveys are correctly identified to genus rank (ter Steege, 1998; ter Steege et al., 2006), whereas other authors measured that less than 80% of trees can correctly be identified at this rank (Hawes et al., 2012).

Throughout the tropics, common names used in forestry surveys mix different local languages and dialects. Whatever their origin, these vernacular names are generally related to traditional uses and to physiognomy but rarely to taxonomical criteria (Oldeman, 1968; Rollet, 1980; Kvist et al., 1995; Wilkie and Saridan, 1999). Despite this intrinsic heterogeneity, the common nomenclatures have been empirically and gradually normalized by local forest services to refine commercial inventories and especially for national scale surveys facilitated by FAO since the 1970s (Clément, 1978). Modern forestry surveys therefore include valuable taxonomic information to address production issues (commercial volumes, biomass) and to examine floristic composition

(Couteron et al., 2003; Rejou-Mechain et al., 2011). Nevertheless, errors and biases need to be better understood in order to control uncertainties associated with these kinds of data. This is especially true in Amazonia where species are often lumped or, on the contrary, split in parataxonomic nomenclatures used by communities and foresters. For example in Acre only 50% of the common names are unique to a single taxonomic species (Baraloto et al., 2008). A simple method that could correct these errors and these biases would be very useful to improve biodiversity measurement by integrating forestry inventory data with true taxonomic surveys.

French Guiana is a particularly favorable context to study this issue for Amazonia and to develop this kind of approach (e.g., Aubry-Kientz et al., 2013). The tree flora is quite well-known due to a high collecting effort (Haripersaud et al., 2010; ter Steege et al., 2013). Despite this effort, the geographical coverage of existing botanical plots network remains insufficient to perform a clear practical mapping of floristic composition. On the other hand, several broad scale forestry surveys are available to complete this network (Brunaux and Demenois, 2003). Moreover, forest nomenclature in French Guiana shares part of its origins with other Amazonian nomenclatures (Hammond et al., 1996) and its study can therefore provide a basis for a larger regional standardization effort.

Here we integrate large scale forestry and botanical inventory data with a novel modeling approach to address three questions: (1) what is the taxonomic reliability of the information contained in the common nomenclature used by foresters in this Amazonian region, (2) what are the resulting uncertainty and bias in the estimates of several diversity indicators using forestry surveys at different scales and (3) can we correct estimates of tree alpha and beta diversity from forestry surveys using appropriate statistical methods?

## 2. Material and method

### 2.1. Study area

French Guiana covers 85,000 km<sup>2</sup> in the East of the Guiana Shield, between the Brazilian state of Amapa and Suriname. Altitude generally ranges between 0 and 200 m a.s.l (mean 140 m) with few mountainous peaks exceeding 800 m. Climate is equatorial with annual rainfall ranging from 4000 mm in the Northeast to 2000 mm in South and West, and a mean annual temperature of about 26 °C. The number of consecutive months with less than 100 mm precipitation (dry season) varies from 2 in the North to 3 in the South with high inter-annual variations. Evergreen rainforest covers more than 90% of the country. The last checklist of trees species (including palms) of French Guiana reports 1581 species from 391 genus and 80 families (Molino et al., 2009). Alpha-diversity of tree communities in northern French Guiana is reputed to be among the highest of the Amazonian region with 10% of all Amazonian species occurring in the 1.3% of surface area represented, and a Fisher's alpha value about 180, far above other eastern Amazonian regions (ter Steege et al., 2003; Saatchi et al., 2008).

### 2.2. Floristic data

We used four floristic datasets: (i) 30 forestry surveys with floristic determination according to common names localized on the whole French Guiana as shown in Fig. 1 and detailed in supplementary information (Table S11), (ii) a taxa-abundances table provided with the checklist of tree species (Molino et al., 2009) and derived from a large phyto-ecological network scattered over French Guiana, (iii) a subset of 29 one-hectare botanical plots (100 m × 100 m) with taxonomic determination localized close to our forestry surveys

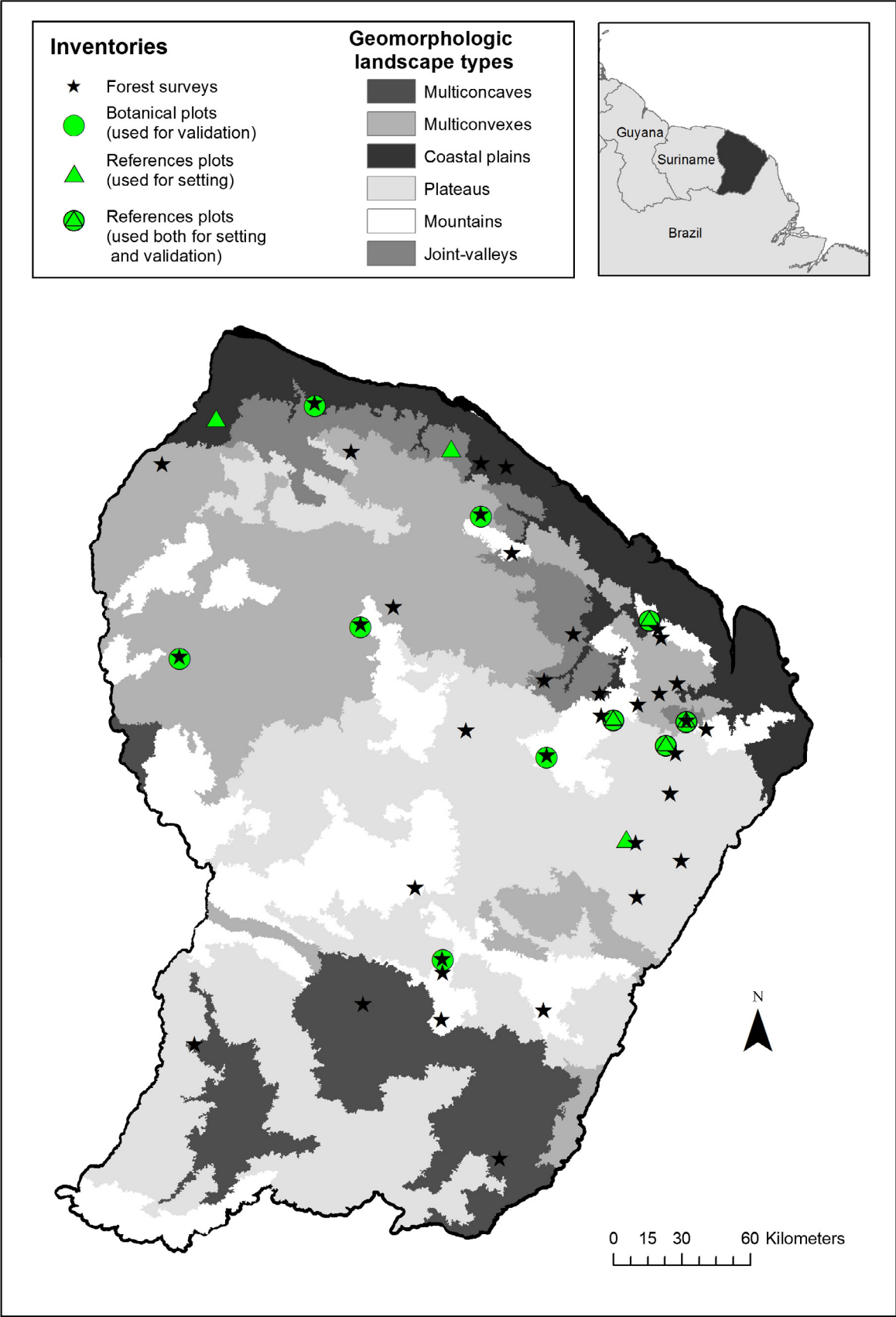


Fig. 1. Forestry surveys and botanical plots location in the main Guianese landscapes.

(Fig. 1) and (iv) a subset of 9 one-hectare reference plots (100 m × 100 m) with both common names and taxonomic determination for all trees (Fig. 1).

1. Forestry surveys have been collected during 2006–2012 by ONF's (French National Agency) fieldworkers whose main activities are pre-harvesting commercial inventories. Ninety-eight line-transects (2.5–3 km long per 20 m large) were inventoried in 30 sites which are homogeneous from a landscape point of view (Guitet et al., 2013) and representative of environmental gradients and vegetation diversity, from low-statured white-sand forests to high-statured terra Firme forests. All trees (including palms) with diameter at breast height (1.3 m; hereafter DBH) above 17.5 cm have been measured and floristic determinations have been assigned using common names (detailed in [Supplementary Information – Table S12](#)). The dataset encompasses 108,650 trees from a total area of 560 ha and is used to estimate diversity indicators at a regional scale (i.e., one site represents about 30 km<sup>2</sup>). The common nomenclature used by foresters in French Guiana gathers more than 500 different vernacular names that derived from two main languages: “Nengue tongo” and “Créole” vernacular (Oldeman, 1968). The correspondences between names and taxonomy were first studied fifty years ago, and revealed many synonyms (Bena, 1960). During the 1970s, nomenclature has been simplified and normalized to analyze large-scale commercial inventories that required homogeneous data for analyses (Valeix and Mauperin, 1989). As a result, many rarely used names were lumped into more generic and common names. Since this time, common nomenclature continues to change with empirical practices and botanical observations, sometimes adding new names relating to species that are deemed easy to recognize, or pointing out fuzzy and inconsistent relationship especially in very large and complex families (e.g., Lauraceae, Sapotaceae and Annonaceae). The current list contains 243 common names.
2. The taxa-abundances table was derived from a phyto-ecological plot network gathering 75,000 inventoried trees, partly included in the Amazon Tree Diversity Network (Molino et al., 2009). As it proved to be one of the most intensive botanical network in the region and spread over almost the whole territory (ter Steege et al., 2013), we assumed this abundance table is representative of our regional meta-community. However, as 493 very rare species of the checklist had no record ( $N = 0$ , i.e., species known from French Guiana but not recorded in the phyto-ecological plots), we added them a residual weight ( $w < 1$ ).
3. The 29 one-hectare dataset was obtained by selecting among botanical plots (100 m × 100 m and DBH ≥ 10 cm) included in ATDN for French Guiana, those situated close to our forestry surveys (i.e., <12 km) and on the same landscape type ([Supplementary Information – Table S11](#)). This subset provided *true* diversity measurements to compare with the diversity estimates from forestry surveys at operational scales on 10 of our study sites.
4. The 9 reference plots dataset come from one-hectare plots (100 m × 100 m) on seven sites laying in the main forest landscapes ([Supplementary Information – Table S11](#)) but localized in the more accessible Northern region of French Guiana. This dataset gathers both common and botanical names for 4279 trees with DBH ≥ 10 cm in 100 m × 100 m plots, determined independently during two separate field-missions, except in one case (plot PARM). It is used to test the taxonomic reliability of common names. Five of these plots are located close to large forestry surveys and are also included in the 29 botanical plots dataset.

The family and genus classification follows the Angiosperm Phylogeny Group III findings (APG III, 2009).

### 2.3. Process to estimate diversity with forestry surveys

Our method is based on the simulation of species occurrence tables. Using a Monte-Carlo scheme we simulate different types of determination errors in forestry surveys, taking into account the taxonomic accuracy of common names. Fig. 2 summarizes the computation process in four steps: (1) formatting nomenclature used in forestry surveys to define the expected taxonomic accuracy of common names, (2) computing taxonomic reliability of the common names taking into account the expected taxonomic accuracy, (3) propagating uncertainty in diversity estimates through simulated communities and (4) computing different diversity indices (i.e., alpha and beta entropy, Fisher-alpha and Sørensen index) and validating the estimates with botanical data.

#### Step one: formatting nomenclature used in forestry surveys

We based our approach on the most recent reference list (ONF, 2004) updated with recent observations (see [Supplementary Information – Table S12](#)). This list contained 243 common names corresponding to one or several botanical species in the French Guiana checklist. Then we classified common names into four groups ( $g$ ) of accuracy, according to their correspondence to botanical taxa of the French Guiana checklist: 138 names ( $g = 1$ ) corresponding to one species, 92 ( $g = 2$ ) corresponding to multiple species within a single genus, 12 ( $g = 3$ ) corresponding to multiple species from different genus in the same family. Other names without any precise taxonomic determination ( $g = 4$ ) are considered as “unknown”.

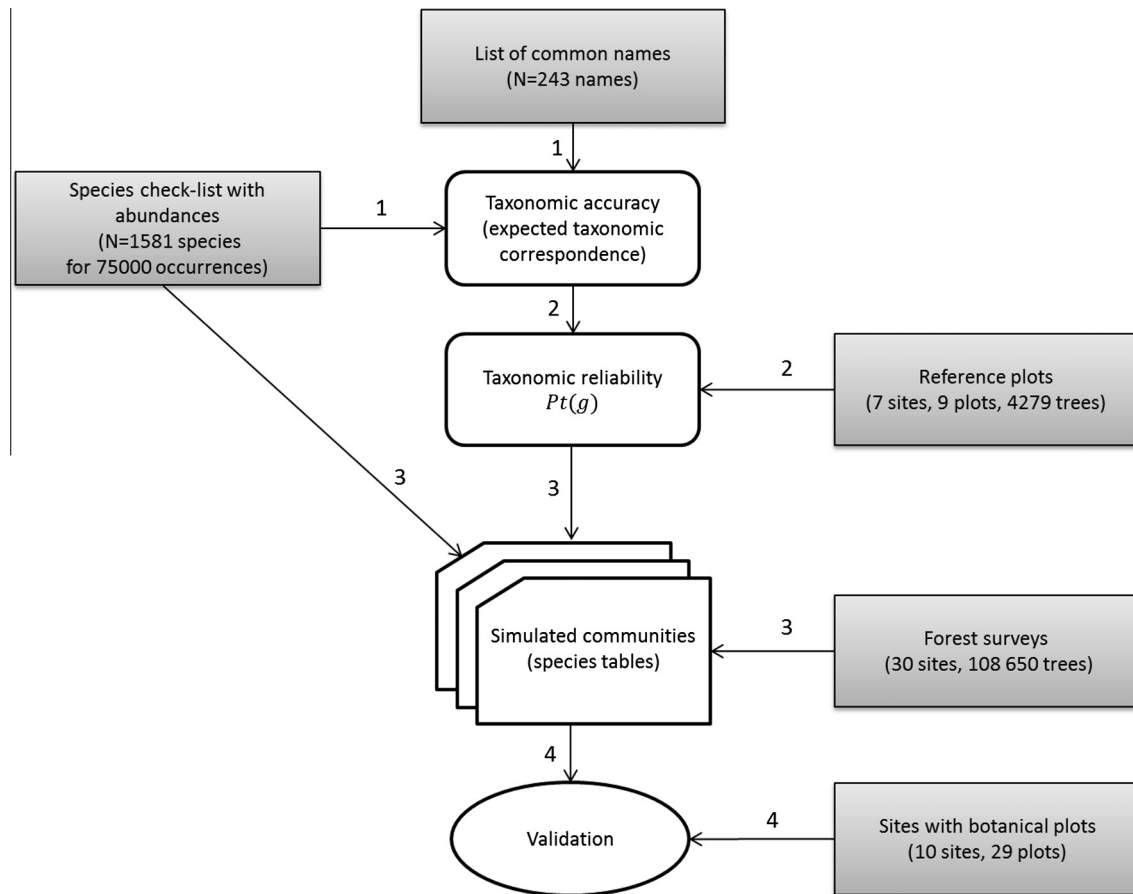
#### Step two: computing the taxonomic reliability of common names

We used the 4279 trees sampled on the 9 reference plots dataset to compute the *taxonomic reliability* of the common names, i.e., the probabilities that the species, genus and family expected from the common name given by foresters are the ones indicated by the botanists. We noticed  $P_t(g)$  the probabilities of a *determination error* at a given taxonomic rank  $t$  (1: for species, 2: genus and 3: family) for each group of common names  $g$  and approximated these probabilities by the corresponding frequencies of each event in our dataset. This yielded four triplets ( $P_1(g), P_2(g), P_3(g)$ ) for species, genus and family ranks respectively, with  $P_1 > P_2 > P_3$ . The *taxonomic reliability* can then be defined as  $1 - P_t(g)$  according to taxonomic ranks and accuracy groups.

To highlight factors that could influence the error values,  $P_t(g)$  has also been calculated for different sub-categories of trees considering their accuracy (groups), their location (plots) and their DBH class. Differences between sub-categories were investigated using linear regression models and ANOVA after verifying assumptions of normality using Bartlett test.

#### Step three: propagating uncertainty in diversity estimates through simulated communities

At this stage, we simulated community samples using the forestry surveys as input and applied a Monte-Carlo process taking into account the expected accuracy and reliability of common names ( $P_t(g)$ ). First, for each tree  $i$  of the forestry survey, with a common name included a group  $g_i$ , we sample  $p_i$  in a uniform law  $U[0, 1]$  and compare it to the  $P_t(g_i)$  triplet to simulate a *determination error*. Then, once the type of error was determined, we sampled a random species within the corresponding subset of the taxa-abundances table, using the relative abundance of each species (i.e.,  $N_{\text{specie}} / \sum N_{\text{subset}}$  in taxa-abundances table) as species probability-sampling, as follows:



**Fig. 2.** Methodological framework used to estimate diversity indices by simulating species occurrence tables from forestry surveys: grey rectangular boxes indicate original tables and white rounded boxes indicate deduced parameters used for Monte-Carlo method. Numbers indicate the different steps detailed in the text.

- If  $p_i \leq P_3(g_i)$  (i.e., error extra-family), we sampled in the restricted list of species from a different family from what was expected for the common name of tree  $i$ ; obviously, in the particular case when name of tree  $i$  was included in group 4 (“unknown”), we sampled in all the checklist.
- If  $P_3(g_i) < p_i \leq P_2(g_i)$  (i.e., true family but false genus), we sampled in the restricted list of species from the same family but a different genus from what was expected for the common name of tree  $i$ .
- If  $P_2(g_i) < p_i \leq P_1(g_i)$  (i.e., true genus but false species), we sampled in the restricted list of species that were different from what was expected for the common name of tree  $i$  but in the same genus.
- If  $P_1(g_i) < p_i$  (i.e., true species) then we used the expected species for the common name of tree  $i$ .

In some particular cases, intra-genus error and intra-family error were actually not computable (species alone in its genus or genus alone in its family). To simplify the model we decided in these cases to sample the correct species or the correct genus with some risk of underestimating the species diversity. We also tested other more sophisticated models that introduced different types of extra-family errors (confusion with common and well-known species vs. confusion with rare and unknown species), but as final results were highly correlated with previous ones we chose to apply the simplest one presented here (Supplementary Information Table S13 and Fig. S12).

#### 2.4. Validation phase: test of robustness and accuracy at different scales

This process, yielding the taxa abundances table of a simulated sample, was iterated 999 times to estimate the diversity indicators and their uncertainty. We computed for each simulated table several diversity indices (Table 2) including Fisher’s alpha and the Sørensen similarity, respectively noted  $\alpha$  and  $\beta$ , which are long and widely-used indices for practical measurements of diversity (Chao et al., 2005). We also calculated the entropy noted  $H_s$  (with the order  $n = 0, 1$  or  $2$  and the part  $s = \alpha$  or  $\beta$ ) which provide unbiased diversity measures and allow partitioning diversity in a coherent framework (Marcon et al., 2014). The R script we used for all the process is available in supplementary material (Supplementary Information S15).

We first simulated 999 values and verified that means and variances were stabilized after 99 simulations using bootstrapping in the 999 simulations. Then we set the number of iterations to 99 in the final script (Supplementary Information – Fig. S11).

We also tested the robustness of our method by degrading the quality of the two sets of parameters used in the Monte-Carlo method: the table of errors  $P_t(g)$  and the taxa-abundance table. First, instead of computing the mean errors probabilities table ( $P_{\text{mean}}$ ) with the nine reference-plots, we computed  $P_t(g)$  with only one plot (with the worst error probabilities – noted  $P_{\text{max}}$ ). Secondly, instead of basing our species probability-sampling on the precise abundances table of the species checklist



**Table 1**

Proportion of each groups regarding levels of taxonomic accuracy in each dataset: in the list of common names (number of names and percentage in bracket), in the reference plots and forestry surveys (number of trees and percentage in bracket).

Taxonomic accuracy	Species	Genus	Family	Unknown
Reference list (common names)	138 (57%)	92 (38%)	12 (5%)	1 <sup>a</sup> (0%)
Reference plots (trees)	1293 (30%)	2557 (60%)	385 (9%)	44 (1%)
Large forestry surveys (trees)	31755 (29%)	69372 (64%)	6801 (6%)	725 (1%)

<sup>a</sup> All names without any well-known taxonomic determination are gathered into “unknown”(see also Table SI2).

**Table 2**

Diversity indicators tested in this study – order indicate the type of metrics used in this indices (0 = based on presence, 1 = based on abundance, 2 = based on concentration).

Diversity indices	Type of diversity	Interpretation and characteristics	References
Fisher's alpha ( $\alpha$ )	Alpha	Rarefaction indice widely used and theoretically corrected from sampling bias	Fisher et al. (1943)
Sørensen ( $\beta$ )	Beta	Similarity indice widely used but sensible to sampling bias	Sørensen (1948)
Entropy ( $^nH_s$ )	Alpha or beta	Order 0 = species richness Order 1 = equivalent Shannon diversity Order 2 = equivalent Gini–Simpson diversity	Marcon et al. (2012, 2014), Marcon and Hérault (2013)

(i.e.,  $N_{\text{specie}}/\Sigma N_{\text{subset}}$  in taxa-abundances table noted  $N_{\text{prec}}$ ), we used rough abundances classes (i.e.,  $N[w, 5^0, 5^1, 5^2, 5^3, 5^4]$  noted  $N_{\text{class}}$ ) to simulate the case where species abundances are not available and have to be estimated by an expert (i.e.,  $N_{\text{specie}} = 625$  for the very abundant species,  $N_{\text{specie}} = 125$  for abundant ones,  $N_{\text{specie}} = 25$  for common ones,  $N_{\text{specie}} = 5$  for uncommon ones,  $N_{\text{specie}} = 1$  for rare ones and  $N_{\text{specie}} = w$  for very rare). Last of all, we tested different values for  $w$ , the residual weight of the very rare species, to simulate different degrees of completeness of the botanical data: a low  $w$  value, close to 0, assumes a very complete and representative database while a high value, close to 1, assumes a quite uncompleted database where scarce species are under-sampled.

Then, we compared the different estimates obtained with common names on the 9 reference plots ( $\alpha$ ,  $\beta$ ,  $^nH_s$ ) with actual indices computed with botanical data ( $\alpha$ ,  $\beta$ ,  $^nH_s$ ) in order to test the efficiency of our approach for ranking and estimating. Finally, we applied the tool on 10 validation sites with large forestry inventories to evaluate the method at the operational scale. In order to compare results (rank and estimates) with actual measures and considering the low effective of our reference plots that did not allow testing the normality, we used Spearman rank tests to statistically compare the ranking and regression analysis only to evaluate bias and error of estimates. All analyses were performed in the R environment (R Development Core Team, 2009) and used packages *vegan* (Oksanen et al., 2013) and *entpart* (Marcon and Hérault, 2013).

### 3. Results

#### 3.1. Reliability of the taxonomic information contained in forestry surveys

Table 3 shows the reliability and determination errors computed on the nine reference plots for groups of common names, considering expected accuracy; and for taxonomic ranks, for all trees and for trees with DBH  $\geq 17.5$  cm (minimum DBH in our forestry surveys). Reliabilities per site and per DBH class are shown in Fig. 3.

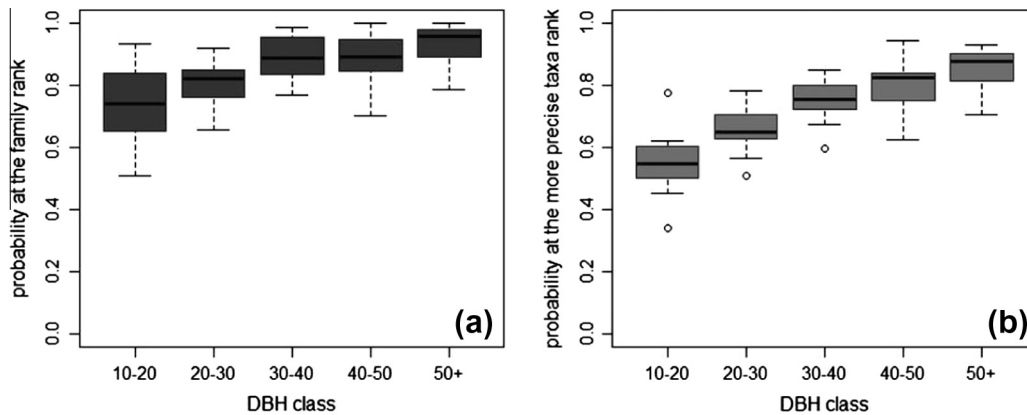
The reliability at the most precise taxonomic rank ( $1 - P_t(g)$  with  $t = g$ ) varies between 65% and 84% for all DBH classes and between 72% and 87% for DBH  $\geq 17.5$  cm (Table 2). In fact, the reliability gradually increases with DBH (Fig. 3): from 55% for DBH  $< 20$  cm to 85% for DBH  $\geq 50$  cm at the most precise taxonomic rank (taxon level) and, in the same way, from 74% to 93% at the family level.

We tested the DBH class effect with ANOVA using mean probability per groups and considering plots as replicates. We selected data with more than 20 trees to assure the normality of the observed probabilities and verifying the normality of residues. At the taxon level and for trees with DBH  $\geq 17.5$  cm, the results of ANOVA tests show strong influence of the DBH ( $df = 3$ ,  $F = 21.951$ ,  $P < 0.001$ ) no groups effect ( $df = 2$ ,  $F = 0.694$ ,  $P = 0.507$ ) but a significant plot-effect ( $df = 8$ ,  $F = 3.985$ ,  $P = 0.002$ ) mainly due to the plot where foresters and botanists collected the information at the

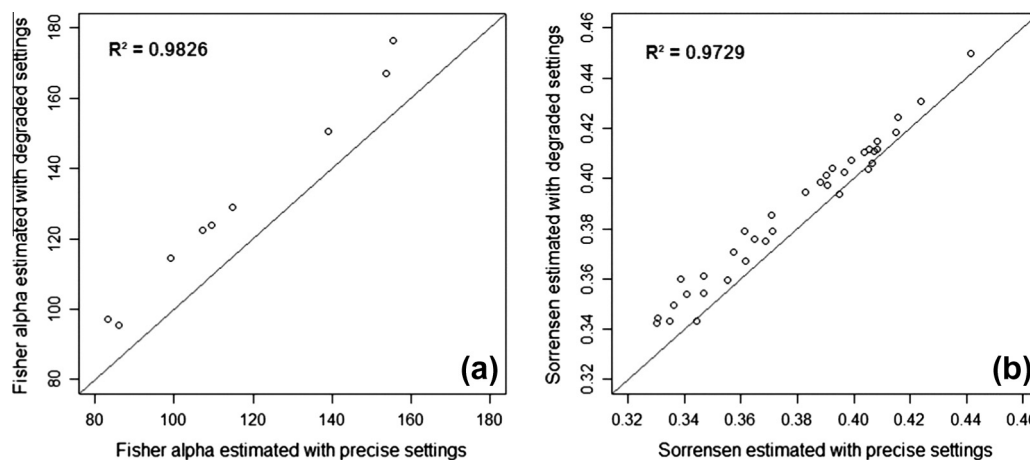
**Table 3**

Probabilities of determination error  $P_t(g)$  and taxonomic reliability ( $1 - P_t(g)$ ) computed for the three taxonomic ranks ( $t$ ) per group of common names' expected taxonomic accuracy ( $g$ ).

Selected trees	Taxonomic accuracy group ( $g$ )	$N$ trees	$P_1(g)$ (error of species)	$P_2(g)$ (error of genus)	$P_3(g)$ (error of family)	$1 - P_t(g = t)$ (reliability at the most precise rank)
All	Species (1)	1293	0.35	0.25	0.14	0.65
	Genus (2)	2557	1	0.34	0.23	0.66
	Family (3)	385	1	1	0.16	0.84
	None (4)	44	1	1	1	0
$\geq 17.5$ cm	Species (1)	762	0.24	0.14	0.08	0.76
	Genus (2)	1235	1	0.28	0.18	0.72
	Family (3)	209	1	1	0.13	0.87
	None (4)	19	1	1	1	0
$\geq 17.5$ cm Without plot PARM	Species (1)	662	0.26	0.15	0.08	0.74
	Genus (2)	1072	1	0.30	0.20	0.70
	Family (3)	184	1	1	0.14	0.86
	None (4)	18	1	1	1	0



**Fig. 3.** Taxonomic reliability of forestry surveys using common names (a) at the family rank (reliability is defined as  $1 - P_3(g)$ ) and (b) at the most precise taxonomic level (reliability is defined as  $1 - P_t(g)$  with  $g = t$ ) per DBH class and for the 9 reference plots.



**Fig. 4.** Correlation between estimates obtained with precise settings ( $w = 0.0001$ ,  $P = P_{\text{mean}}$  and  $N = N_{\text{prec}}$ ) and with degraded settings ( $w = 0.5$ ,  $P = P_{\text{max}}$  and  $N = N_{\text{class}}$ ) – examples for Fisher-alpha (a) and for Sørensen per pairs of plots (b). Solid line indicate 1:1 relation.

same time (plot PARM –  $df = 7$ ,  $F = 2.928$ ,  $P = 0.0221$  when this plot is excluded). The same tests performed at the family rank show significant effects for all factors (respectively  $df = 2$ ,  $F = 8.79$ ,  $P < 0.001$  for groups,  $df = 3$ ,  $F = 9.195$ ,  $P < 0.001$  for DBH and  $df = 8$ ,  $F = 5.121$ ,  $P < 0.001$  for plots) with still lower plot-effect when PARM is excluded from analysis ( $df = 7$ ,  $F = 4.257$ ,  $P = 0.0032$ ).

Thereafter, we applied the probabilities computed without the site PARM (Table 2 – last rows), when applying to the large forestry surveys. As a result, the mean reliability estimated for forestry surveys taking into account the relative abundances of the groups of accuracy (Table 1 last row) and computed with these statistics is estimated at 83% for family level; 70% for genus level; 22% at the species level; and 74% at the most precise expected taxonomic level.

### 3.2. Test at the local scale on the one-hectare reference plots

The estimated diversity values obtained with the different settings ( $P_{\text{mean}}/P_{\text{max}}$ ,  $N_{\text{prec}}/N_{\text{class}}$ ,  $w$ ) showed high and significant correlation in all cases ( $r^2 > 0.90$  –  $p < 0.001$  for all possible pair of results). Estimated ranks are thus well-conserved whatever the parameters we used. However, the most degraded settings ( $P_{\text{max}}$ ,  $N_{\text{class}}$ ,  $w = 0.5$ ) led to increased estimated values (i.e., a higher alpha diversity for all plots and a higher similarity between plots – Fig. 4).

Accordingly, we used in all further analyses the parameters  $w = 0.00001$ ,  $P = P_{\text{mean}}$  and  $N = N_{\text{prec}}$ . The correlation tests between the simulated estimates and actual values are detailed in Table 4.

The table also shows the results that would have been obtained without our correction method (i.e., directly with the common names).

In all cases, the correlation tests with unadjusted estimates were weak ( $p < 0.05$ ) or not significant except for  $^2H_\beta$  ( $p < 0.001$ ). Most of the correlation tests were greatly improved after using the correction method:

- Spearman rank-test was still not significant for  $^1H_\beta$ , weakly significant ( $p < 0.05$ ) for  $\alpha$  and  $^0H_\beta$ , but highly significant for all other indices ( $p < 0.001$ ) such that ranks are well estimated for a majority of diversity measures after correction.
- $R^2$  are high ( $> 0.80$ ) and highly significant ( $p < 0.001$  or  $p < 0.01$ ) for all  $H_\alpha$  and  $^2H_\beta$  whereas they are lower ( $< 0.60$ ) and less significant ( $p < 0.05$  or  $p < 0.01$ ) for  $^0H_\beta$ ,  $\alpha$ ,  $\beta$ .

However, the absolute values for  $H_\alpha$ ,  $\alpha$ ,  $\beta$  are still over-estimated (i.e., positive biases) whereas they are under-estimated for all  $H_\beta$  (i.e., negative biases). The relative mean errors are particularly important for  $\alpha$ ,  $\beta$  (respectively 55% and 48%). They show opposite patterns between  $H_\beta$  and  $H_\alpha$ : higher but decreasing with the order for  $H_\beta$  (from 38% at order 2 to 15% at order 0), and lower but increasing with the order for  $H_\alpha$  (from 1% at order 2 to 21% at order 0). The error is particularly strong for one plot (TORT1 – see Fig. 5), for which the common names are the least accurate (we have only family information for 17% of the trees vs. less than 10% on all other plots). The correlations are largely improved for

**Table 4**

Correlation tests between diversity values estimated from common names and measured from botanical data in the one-hectare reference plots, using Spearman's rank-tests for ranking and linear regression tests for predictions. The first two columns give results for unadjusted estimates (directly with inventory tables) and the last ones give result after using our correction method:

Indice	DF	Without correction		After correction			
		Spearman's rank-test	Adjusted $R^2$ with $F$ test	Spearman's rank-test	Adjusted $R^2$ with $F$ test	Bias	Mean absolute% error
$^0H_\alpha$	7	$S = 24^*$	0.44 <sup>*</sup>	$S = 23^{**}$	0.86 <sup>***</sup>	22.668	21
$^1H_\alpha$	7	$S = 24^*$	0.55 <sup>*</sup>	$S = 14^{**}$	0.88 <sup>***</sup>	0.269	6
$^2H_\alpha$	7	$S = 22^*$	0.65 <sup>***</sup>	$S = 4^{***}$	0.85 <sup>***</sup>	0.007	1
$^0H_\beta$	7	$S = 62^{ns}$	0.13 <sup>ns</sup>	$S = 32^*$	0.49 <sup>*</sup>	−53.611	15
$^1H_\beta$	7	$S = 70^{ns}$	0.28 <sup>ns</sup>	$S = 56^{ns}$	0.81 <sup>***</sup>	−0.301	26
$^2H_\beta$	7	$S = 8^{***}$	0.79 <sup>***</sup>	$S = 2^{***}$	0.82 <sup>***</sup>	−0.007	38
$\alpha$	7	$S = 70^{ns}$	0.32 <sup>ns</sup>	$S = 28^*$	0.55 <sup>*</sup>	41.325	55
$\beta$	34	$S = 4741^*$	0.17 <sup>*</sup>	$S = 3321^{***}$	0.24 <sup>**</sup>	0.111	48

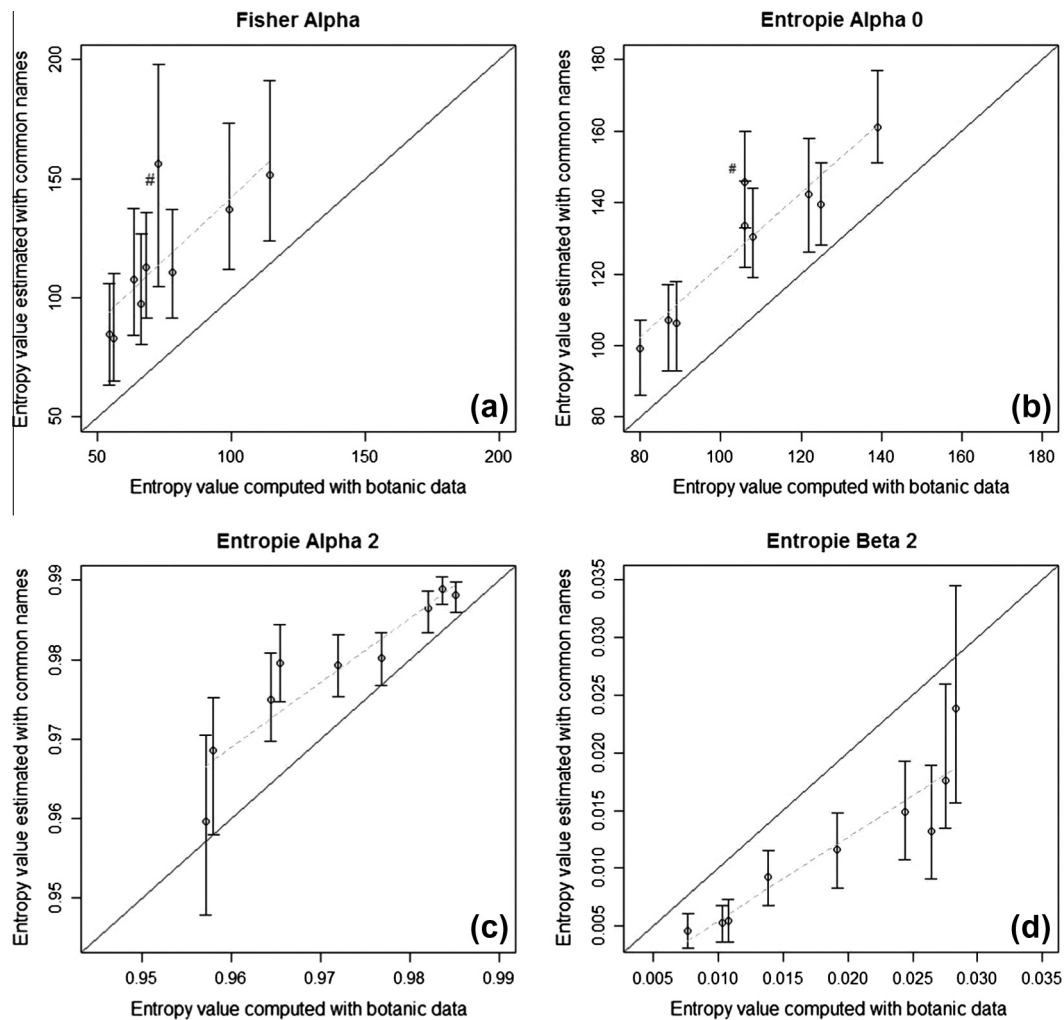
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , <sup>ns</sup> Not significant.

$^0H_\beta$  and  $\alpha$  when this outsider is removed (respectively: adjusted  $R^2 = 0.77$ – $p < 0.01$ ; adjusted  $R^2 = 0.92$ – $p < 0.001$ ).

### 3.3. Validation at the operational scale on 10 sites with large forestry surveys and multiple botanical plots

Because botanical and forestry surveys have different extent and sampling effort at this scale, we only considered the correlation tests and did not pay attention to biases and absolute

error (Table 5). Despite the sampling differences, all diversity indices, except  $^0H$ , showed strong correlation between the botanical measures and the values estimated from forestry surveys after applying the correction method, whereas only  $^1H_\beta$ ,  $^2H_\beta$  and  $\beta$  were successful without correction. Similarly to the plot scale, the Spearman rank-test was not significant for  $^1H_\beta$ , and weakly significant for  $\alpha$ , whereas highly significant for all other indices, excepted  $^0H$ . As a consequence, site ranks were globally well preserved for a majority of estimated diversity indices after correction.  $R^2$  were



**Fig. 5.** Correlation between the different diversity indices computed from common names and actual values measured from botanical data on reference plots – solid line indicate 1:1 relation and dashed line indicate fitted regression: for  $\alpha$  (a) and  $^0H_\alpha$  (b) the correlation is greatly improved once the outlying plot (TORT1 marked with #) is removed. No outliers for  $^2H_\alpha$  (c) and  $^2H_\beta$  (d).



high ( $>0.75$ ) and significant ( $p < 0.001$ ) for all  $H_\alpha$  and  $H_\beta$  at order 1 and 2 whereas they were lower for  $\alpha$  ( $R^2 = 0.63$ ,  $p < 0.01$ ) and  $\beta$  ( $R^2 = 0.66$ ,  $p < 0.001$ ).

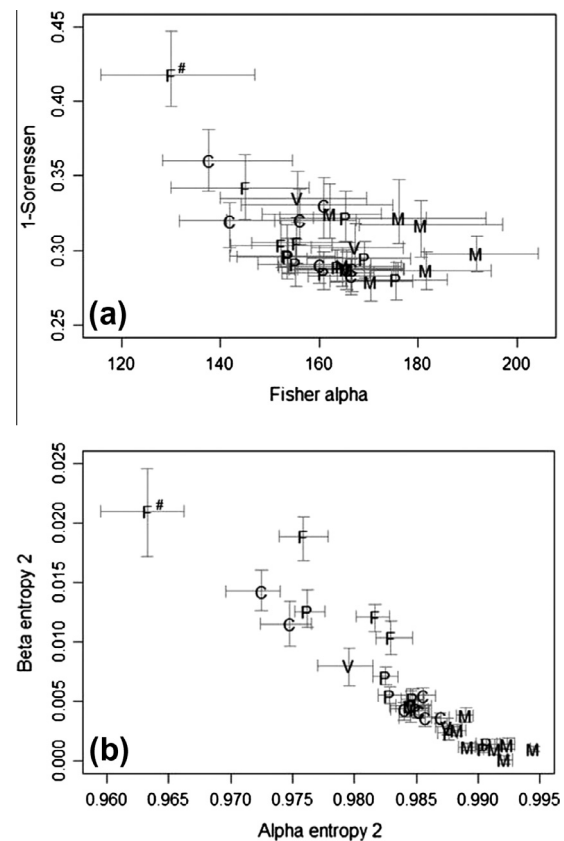
Finally, when the method was applied across the whole forestry surveys dataset (Fig. 5), significant differences among sites were detected, especially for alpha and beta entropies of order 2 which had smaller uncertainty (Fig. 6b) than for classical Fisher-alpha and Sørensen indices (Fig. 6a). In both cases, we observed similar patterns relatively to the landscapes classes: all mountainous sites (noted M), exhibited high alpha and low beta diversities, whereas flat lowland sites (noted F) exhibited a higher beta-diversity than other landscapes and a relatively lower alpha-diversity.

#### 4. Discussion

Even if the concept of habitat complementarity has replaced the sole richness-rankings in the current literature on conservation planning, this long-history approach is still widely used in practice to prioritize areas for conservation (e.g., Gotelli and Colwell 2001; Martini et al., 2007). In this paper, we propose a rigorous way to deal with taxonomic problems inherent to the use of forest inventory data for conservation planning. The script developed in the R environment, which is free and widely used in ecological analyses, is provided in supplementary information and contains all necessary functions to apply the correction method to improve the diversity estimates. This method proved to give relevant results compared with botanical data and has the potential to greatly improve the diversity mapping in tropical forest.

##### 4.1. Forestry surveys of tropical forests provide relevant taxonomic information

In French Guiana, common names provide genus-level information with taxonomic reliability of about 70%, which is quite comparable to other published studies. In Central Africa, Rejou-Mechain et al. (2011) found that agreement between forestry surveys and 1107 0.5-ha botanical plots varies between 32% and 67% at the genus level according to the DBH and site conditions. Using 17 0.1-ha plots in central Amazonia, Hawes et al. (2012) reported a similar expected taxonomic resolution as ours for local nomenclature (i.e., 18.4% for species and 59.8% for genus, to be compared to 29% and 64% respectively in our study), but obtained a better reliability for this small sample (i.e., 78.1% for genus level and 97.6% for family vs. 70% and 83% respectively in our study). At the species level, these performances can be compared to the morpho-species classification, commonly employed by botanists with sterile vouchers, whose reliability is evaluated between 52% and 67% (Gomes et al., 2013). Considering these results obtained with well-trained field-workers and precedent references (Baraloto



**Fig. 6.** Comparison of the estimated diversity values obtained from forestry surveys using correction method for alpha and beta entropy at order 2 (b) and for classical alpha-Fisher and Sørensen indices (a) – bars indicate uncertainties – labels indicate dominant landscape on each site (M = Mountains, C = Hills, F = Plains, P = Plateaus, V = Valleys). Mountains show higher alpha diversity values whereas plains express lower alpha diversity but higher beta diversity, especially the site on White sand marked by #.

et al. 2008), we think that the high reliability attributed to some Amazonian forestry surveys, i.e., 95% at the genus level and about 100% for family level (ter Steege, 1998; ter Steege et al., 2006), are overestimated and should be more precisely tested with several reference plots as in our study.

##### 4.2. Alpha and beta diversity can be accurately estimated from forestry surveys using our correction method

As common names often represent multiple species, diversity indices computed directly from forestry surveys are obviously

**Table 5**  
Correlation tests between diversity values estimated from common names on large forestry surveys using our correction method and actual values measured on multiple botanical plots at the same sites, using Spearman's rank-tests and adjusted  $R^2$ .

Indice	DF	Without correction		After correction	
		Spearman's rank-test	Adjusted $R^2$ with F test	Spearman's rank-test	Adjusted $R^2$ with F test
${}^0H_\alpha$	8	$S = 102^{ns}$	0.39 <sup>*</sup>	$S = 62^{ns}$	0.31 <sup>ns</sup>
${}^1H_\alpha$	8	$S = 92^{ns}$	0.31 <sup>ns</sup>	$S = 28^{**}$	0.77 <sup>***</sup>
${}^2H_\alpha$	8	$S = 98^{ns}$	0.06 <sup>ns</sup>	$S = 24^{**}$	0.75 <sup>***</sup>
${}^0H_\beta$	8	$S = 212^{ns}$	0.12 <sup>ns</sup>	$S = 162^{ns}$	0.08 <sup>ns</sup>
${}^1H_\beta$	8	$S = 48^{*}$	0.78 <sup>***</sup>	$S = 60^{ns}$	0.75 <sup>***</sup>
${}^2H_\beta$	8	$S = 74^{**}$	0.62 <sup>**</sup>	$S = 16^{***}$	0.84 <sup>**</sup>
$\alpha$	8	$S = 128^{ns}$	0.23 <sup>ns</sup>	$S = 28^{*}$	0.63 <sup>**</sup>
$\beta$	44	$S = 3376^{***}$	0.69 <sup>***</sup>	$S = 4020^{***}$	0.66 <sup>***</sup>

<sup>\*</sup>  $p < 0.05$ .

<sup>\*\*</sup>  $p < 0.01$ .

<sup>\*\*\*</sup>  $p < 0.001$ .

<sup>ns</sup> Not significant.

poorly correlated with actual diversity measures as shown by the correlation tests at the one-hectare plot scale (Table 4). However, raw forestry surveys data can directly provide quite correct estimates of Gini–Simpson beta-diversity at both local and regional scales. In fact abundant species, which have more weight in this order 2 metric (Pelissier and Couteron, 2007; Marcon et al., 2012), are also well-known by foresters. However using the proposed model to simulate communities improves the reliability at this order and allows for better estimates of alpha-diversities and a complete partitioning of diversity with a correct approach.

The proposed model also provides better diversity estimates and a robust ranking of sites, highly consistent with botanical measurements for almost all diversity indices at regional scale. Despite the mismatch between botanical plots and forestry surveys in extent and location (up to 12 km), we obtained strong correlations between true measures and estimated diversity values after correction (Table 5). The mean values obtained for  $\alpha$  at this scale (between 130 and 180) are in accordance with the previous references for French Guiana, i.e.,  $\alpha$  about 180 for DBH > 10 cm at the species rank (ter Steege et al., 2003; Saatchi et al., 2008) and about 80 at the genus rank for DBH > 30 cm (Stropp et al., 2009). Interestingly, corrected forestry survey estimates showed variation in diversity among sites, congruent with those shown by botanical inventories. For example the highest alpha-diversity values were found in sites situated on low mountains (up to 450 m) in accordance with the “mid-altitude bulge” effect suggested by several authors for rainforests (Eisenlohr et al., 2013; Sanchez et al., 2013). At the opposite, sites on lowlands present lower alpha-diversity values and higher dissimilarity values (beta-diversity). This is especially true for the SABL site on white sands. These highly resources-constrained soils constitute a strong environmental filter, resulting in low diversity and high endemism levels (Fine et al., 2010; Stropp et al., 2011).

Our results confirm that large forestry surveys are best suited to detect large-scale patterns of floristic variation, using metrics that give more weight to abundant species, i.e., Gini–Simpson and all other order 2 metrics (Rejou-Mechain et al., 2011). We also demonstrate that these data can provide valuable estimates, after correction, for other diversity indices which are widely used by ecologists and practitioners (Fig. 5). The same method could also be coupled with functional-traits databases (Ollivier et al., 2007) or with phylogenetic trees (Baraloto et al., 2012) available for French Guiana in order to get an initial estimate of the regional variability from a functional or phylogenetic perspective (ter Steege et al., 2006).

#### 4.3. Model implementation improvement and terms of use

The simple method presented here allows estimating forest tree community diversity on operational scales (i.e., areas of thousands hectares) using forestry surveys with reduced bias. Whereas 1 ha-plots require about 25 person days to be established and inventoried (Baraloto et al., 2013), considerably greater areas can be surveyed during the same duration with forestry workers. As a result, this rapid assessment method may be efficient for prioritizing conservation areas within which more comprehensive botanical sampling could be scheduled.

As common forestry nomenclatures share quite similar structures all over the tropics, we propose that this method could be used in other contexts. Scripts available in the supplementary information can be adapted with local parameters but should be used with caution as several conditions have to be controlled to assure the consistency of the results: (i) biogeographic conditions have to be relatively homogeneous and the fragmentation of the study area has to be limited, to ensure working on the same meta-community, (ii) a sufficiently complete checklist and

quantitative botanical inventories of the local flora should be available to determine the relative frequency of the most common species (at least with coarse abundance classes), (iii) foresters and botanists should have sufficient exchanges in the field to get a clear view on the usual relationship between common names and taxonomy and (iv) common reference plots must be surveyed independently by foresters and botanists in order to quantify the reliability of the forestry surveys regarding taxonomic information. We believe that these conditions exist in many areas, especially in Amazonia where RADAMBrasil surveys provide abundant and precise data that could complete the geographic coverage of tree diversity plots (ter Steege et al., 2006, 2013; Emilio et al., 2010; Moscoso et al., 2013). Such data also exist in forestry concessions adopting certification processes for reduced impact logging (RIL) that require the establishment of permanent monitoring plots (Pokorny et al., 2005). A notable example occurs in Central Africa where large forestry surveys are frequent in certified private concessions (Fayolle et al., 2012).

Our results confirm that forestry surveys can provide tremendous and valuable datasets for improving regional forest management plans and conservation priorities. Moreover it also underlines the complementary nature of forestry surveys and precise botanical sampling. Nevertheless, the absolute values of diversity estimated from forestry surveys are clearly biased, even when using our corrections, especially at the plot scale (i.e., over-estimation for alpha-diversity and under-estimation for beta-diversity indices). Indeed, we did not take into account the spatial auto-correlation due to limited dispersion which causes a local similarity in composition (Condit et al., 2000) at both tree neighborhood and local scales.

Therefore, we believe further improvements could be achieved in several ways. First, additional botanical inventories would not only augment species abundance data, but more interestingly would permit the definition of regional communities, instead of a single meta-community, to adapt the sampling process to each locality. Second, this increase in botanical effort could also provide aggregation indicators in order to reduce biases introduced by spatial auto-correlation. For example, the Monte-Carlo process could be modified using a maximum density indicator for each species, such that random species sampling depends on the precedent samplings. Third, additional reference plots could allow reliability estimates for each common name instead of groups of common names as we were restricted to implement. Continued sampling of tropical forest diversity remains an urgent priority (Feeley and Silman, 2011), and we call for continued collaboration between forestry professionals, botanists and ecologists in this endeavor.

#### Acknowledgements

We thank French Forest Agency (ONF), the National Park of French Guiana (PAG), ECOTROP and “Guyaspase” BGF program of French Minister for Environment (MEDD/MAAP) and the “Habitats” PO-FEDER program of European Union (UE) for funding this study. We also acknowledge all the field workers and technicians who participated to the forestry surveys and furnished the data, especially Atidong Nano now retired. Special thanks to Maxime Réjou-Méchain for his careful rereading and to two anonymous reviewers for their useful comments.

#### Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foreco.2014.05.045>.

## References

- App, I.L.I., 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161, 105–121. <http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x>.
- Aubry-Kientz, M., Hérault, B., Ayotte-Trépanier, C., Baraloto, C., Rossi, V., 2013. Toward trait-based mortality models for tropical forests. PLoS one 8, e63678. <http://dx.doi.org/10.1371/journal.pone.0063678>.
- Baraloto, C., Ferreira, E., Rockwell, C., Walthier, F., 2008. Limitations and applications of parataxonomy for community forest management in southwestern Amazonia. Ethnobotany Res. Appl. 5, 77–84. <http://www.ethnobotanyjournal.org/vol5/i1547-3465-05-077.pdf>.
- Baraloto, C., Hardy, O.J., Paine, C.E.T., Dexter, K.G., Cruaud, C., Dunning, L.T., Gonzalez, M.A., Molino, J.F., Sabatier, D., Savolainen, V., Chave, J., 2012. Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. J. Ecol. 100, 690–701. <http://dx.doi.org/10.1111/j.1365-2745.2012.01966.x>.
- Baraloto, C., Molto, Q., Rabaud, S., Hérault, B., Valencia, R., Blanc, L., Fine, P.V.A., Thompson, J., 2013. Rapid simultaneous estimation of aboveground biomass and tree diversity across Neotropical forests: a comparison of field inventory methods. Biotropica 45, 288–298. <http://dx.doi.org/10.1111/btp.12006>.
- Basset, Y., Novotny, V., Miller, S.E., Weiblen, G.D., Missa, O., Stewart, A.J.A., 2004. Conservation and biological monitoring of tropical forests: the role of parataxonomists. J. Appl. Ecol. 41, 163–174.
- Bena, P., 1960. Essences forestières de Guyane. Bureau Agricole et Forestier de l'Ouest Guyanais, Paris, p. 469.
- Brunaux, O., Demeo, J., 2003. Planning and logging in the tropical rain forests of French Guiana. Rev. Forest Fr. (Nancy) 55, 260–272. <http://dx.doi.org/10.4267/2042/5777>.
- Chao, A., Chazdon, R.L., Colwell, R.K., Shen, T.J., 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett. 8, 148–159. <http://dx.doi.org/10.1111/j.1461-0248.2004.00707.x>.
- Clément, J., 1978. Formulation, réalisation et utilisation des inventaires forestiers dans les pays tropicaux. Bois For. Trop. 55. [http://bft.cirad.fr/cd/BFT\\_182\\_55-71.pdf](http://bft.cirad.fr/cd/BFT_182_55-71.pdf).
- Condit, R., Ashton, P.S., Baker, P., Bunyavechewin, S., Gunatilleke, S., Gunatilleke, N., Hubbell, S.P., Foster, R.B., Itoh, A., LaFrankie, J.V., Lee, H.S., Losos, E., Manokaran, N., Sukumar, R., Yamakura, T., 2000. Spatial patterns in the distribution of tropical tree species. Science 288, 1414–1418. <http://dx.doi.org/10.1126/science.288.5470.1414>.
- Couteron, P., Pelissier, R., Mapaga, D., Molino, J.F., Teillier, L., 2003. Drawing ecological insights from a management-oriented forest inventory in French Guiana. Forest Ecol. Manage. 172, 89–108. [http://dx.doi.org/10.1016/S0378-1127\(02\)00310-9](http://dx.doi.org/10.1016/S0378-1127(02)00310-9).
- Eisenlohr, P.V., Alves, L.F., Bernacci, L.C., Padgurschi, M.C.G., Torres, R.B., Prata, E.M.B., dos Santos, F., Assis, M.A., Ramos, E., Rochelle, A.L.C., Martins, F.R., Campos, M.C.R., Pedroni, F., Sanchez, M., Pereira, L.S., Vieira, S.A., Gomes, J., Tamashiro, J.Y., Scaranello, M.A.S., Caron, C.J., Joly, C.A., 2013. Disturbances, elevation, topography and spatial proximity drive vegetation patterns along an altitudinal gradient of a top biodiversity hotspot. Biodiversity Conserv. 22, 2767–2783. <http://dx.doi.org/10.1007/s10531-013-0553-x>.
- Emilio, T., Nelson, B.W., Schietti, J., Desmouliere, S.J.M., Santo, H., Costa, F.R.C., 2010. Assessing the relationship between forest types and canopy tree beta diversity in Amazonia. Ecography 33, 738–747. <http://dx.doi.org/10.1111/j.1600-0587.2009.06139.x>.
- Fayolle, A., Engelbrecht, B., Freycon, V., Mortier, F., Swaine, M., Réjou-Méchain, M., Doucet, J.-L., Fauvet, N., Cornu, G., Gourlet-Fleury, S., 2012. Geological substrates shape tree species and trait distributions in African Moist Forests. PLoS one 7, e42381. <http://dx.doi.org/10.1371/journal.pone.0042381>.
- Feeley, K.J., Silman, M.R., 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity Distribution 17, 1132–1140. <http://dx.doi.org/10.1111/j.1472-4642.2011.00813.x>.
- Fine, P.V.A., Garcia-Villacorta, R., Pitman, N.C.A., Mesones, I., Kembel, S.W., 2010. A floristic study of the white-sand forests of Peru. Ann. Mo. Bot. Gard. 97, 283–305. <http://dx.doi.org/10.3417/2008068>.
- Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J. Anim. Ecol. 42–58.
- Gomes, A.C.S., Andrade, A., Barreto-Silva, J.S., Brenes-Arguedas, T., Lopez, D.C., de Freitas, C.C., Lang, C., de Oliveira, A.A., Perez, A.J., Perez, R., da Silva, J.B., Silveira, A.M.F., Vaz, M.C., Vendrami, J., Vicentini, A., 2013. Local plant species delimitation in a highly diverse Amazonian forest: do we all see the same species? J. Veg. Sci. 24, 70–79. <http://dx.doi.org/10.1111/j.1654-1103.2012.01441.x>.
- Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol. Lett. 4, 379–391.
- Guitet, S., Cornu, J.-F., Brunaux, O., Betbeder, J., Carozza, J.-M., Richard-Hansen, C., 2013. Landform and landscape mapping, French Guiana (South America). J. Maps 9 (3), 325–335. <http://dx.doi.org/10.1080/17445647.2013.785371>.
- Hammond, D.S., Gourlet-Fleury, S., Hout, P.V.D., ter Steege, H., Brown, V.K., 1996. A compilation of known Guianan timber trees and the significance of their dispersal mode, seed size and taxonomic affinity to tropical rain forest management. Forest Ecol. Manage. 83, 99–116. [http://dx.doi.org/10.1016/0378-1127\(95\)03680-6](http://dx.doi.org/10.1016/0378-1127(95)03680-6).
- Haripersaud, P., ter Steege, H., de Granville, J.J., Chevillotte, H., Hoff, M., 2010. Species abundance, distribution and diversity in time and space after centuries of botanical collecting in the Guianas. Taxon 59, 592–597.
- Hawes, J.E., Peres, C.A., Riley, L.B., Hess, L.L., 2012. Landscape-scale variation in structure and biomass of Amazonian seasonally flooded and unflooded forests. Forest Ecol. Manage. 281, 163–176. <http://dx.doi.org/10.1016/j.foreco.2012.06.023>.
- Higgins, M.A., Ruokolainen, K., 2004. Rapid tropical forest inventory: a comparison of techniques based on inventory data from western Amazonia. Conserv. Biol. 18, 799–811. <http://dx.doi.org/10.1111/j.1523-1739.2004.00235.x>.
- Kessler, M., Abrahamczyk, S., Bos, M., Buchori, D., Putra, D.D., Gradstein, S.R., Hohn, P., Kluge, J., Orend, F., Pitopang, R., Saleh, S., Schulze, C.H., Sporn, S.G., Steffan-Dewenter, I., Tjitrosoedirdjo, S.S., Tschardt, T., 2011. Cost-effectiveness of plant and animal biodiversity indicators in tropical forest and agroforest habitats. J. Appl. Ecol. 48, 330–339. <http://dx.doi.org/10.1111/j.1365-2664.2010.01932.x>.
- Koh, L.P., Sodhi, N.S., 2010. Conserving Southeast Asia's imperiled biodiversity: scientific, management, and policy challenges. Biodiversity Conserv. 19, 913–917. <http://dx.doi.org/10.1007/s10531-010-9818-9>.
- Kvist, L.P., Andersen, M.K., Hesselsoe, M., Vanclay, J.K., 1995. Estimating use-values and relative importance of Amazonian flood plain trees and forests to local inhabitants. Papers from the IUFRO Tropical Silviculture Subject Group, 51 07 00, IUFRO 20th 74, 293–300, 293. <http://www.cifor.org/nc/online-library/browse/view-publication/publication/27.html>.
- Lacerda, A.E.B.D., Nimmo, E.R., 2010. Can we really manage tropical forests without knowing the species within? getting back to the basics of forest management through taxonomy. Forest Ecol. Manage. 259, 995–1002. <http://dx.doi.org/10.1016/j.foreco.2009.12.005>.
- Landeiro, V.L., Bini, L.M., Costa, F.R.C., Franklin, E., Nogueira, A., de Souza, J.L.P., Moraes, J., Magnusson, W.E., 2012. How far can we go in simplifying biomonitoring assessments? an integrated analysis of taxonomic surrogacy, taxonomic sufficiency and numerical resolution in a megadiverse region. Ecol. Indic. 23, 366–373. <http://dx.doi.org/10.1016/j.ecolind.2012.04.023>.
- Marcon, E., Hérault, B., 2013. Entropart, an R package to partition diversity. <http://CRAN.R-project.org/package=entropart>.
- Marcon, E., Hérault, B., Baraloto, C., Lang, G., 2012. The decomposition of Shannon's entropy and a confidence interval for beta diversity. Oikos 121, 516–522. <http://dx.doi.org/10.1111/j.1600-0706.2011.19267.x>.
- Marcon, E., Scotti, I., Hérault, B., Rossi, V., Lang, G., 2014. Generalization of the partitioning of Shannon diversity. PLoS one 9. <http://dx.doi.org/10.1371/journal.pone.0090289>.
- Martini, A.M.Z., Fiaschi, P., Amorim, A.M., da Paixao, J.L., 2007. A hot-point within a hot-spot: a high diversity site in Brazil's Atlantic Forest. Biodiversity Conserv. 16, 3111–3128.
- Molino, J.F., Sabatier, D., Prévost, M.F., Frame, D., Gonzalez, S., Bilot-Guérin, V., 2009. Etablissement d'une liste des espèces d'arbres de la Guyane Française. IRD – French Ministry of Agriculture, internal report, Cayenne-Montpellier, p. 59. [http://agriculture.gouv.fr/IMG/pdf/7\\_ListeArbresGuyane-2.pdf](http://agriculture.gouv.fr/IMG/pdf/7_ListeArbresGuyane-2.pdf).
- Moritz, C., Faith, D.P., 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. Mol. Ecol. 7, 419–429. <http://dx.doi.org/10.1046/j.1365-294x.1998.00317.x>.
- Moscoso, V., Albernoz, A.L., Salomão, R.D.P., 2013. Niche modelling for twelve plant species (six timber species and six palm trees) in the Amazon region, using collection and field survey data. Forest Ecol. Manage. 310, 652–662. <http://dx.doi.org/10.1016/j.foreco.2013.08.064>.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. <http://dx.doi.org/10.1038/35002501>.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Vegan: community Ecology Package. Rpackage version 2.0-7. <http://CRAN.R-project.org/package=vegan>.
- Oldeman, R.A.A., 1968. Sur la valeur des noms vernaculaires des plantes en Guyane française. Bois For. Trop. 117, 17–23. [http://bft.cirad.fr/cd/BFT\\_117\\_17-23.pdf](http://bft.cirad.fr/cd/BFT_117_17-23.pdf).
- Ollivier, M., Baraloto, C., Marcon, E., 2007. A trait database for Guianan rain forest trees permits intra- and inter-specific contrasts. Ann. Forest Sci. 64, 781–786. <http://dx.doi.org/10.1051/forest:2007058>.
- ONF, 2004. Guide de reconnaissance des arbres de Guyane. Office National des Forêts, Cayenne, p. 374.
- Pelissier, R., Couteron, P., 2007. An operational, additive framework for species diversity partitioning and beta-diversity analysis. J. Ecol. 95, 294–300. <http://dx.doi.org/10.1111/j.1365-2745.2007.01211.x>.
- Pokorny, B., Sabogal, C., Silva, J.N.M., Bernardo, P., Souza, J., Zweede, J., 2005. Compliance with reduced-impact harvesting guidelines by timber enterprises in Terra Firme forests of the Brazilian Amazon. Int. Forest Rev. 7, 9–20.
- R Development Core Team, 2009. R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org/>.
- Rejou-Mechain, M., Fayolle, A., Nasi, R., Gourlet-Fleury, S., Doucet, J.L., Gally, M., Hubert, D., Pasquier, A., Billand, A., 2011. Detecting large-scale diversity patterns in tropical trees: can we trust commercial forest inventories? Forest Ecol. Manage. 261, 187–194.
- Rejou-Mechain, M., Pelissier, R., Gourlet-Fleury, S., Couteron, P., Nasi, R., Thompson, J.D., 2008. Regional variation in tropical forest tree species composition in the Central African Republic: an assessment based on inventories by forest

- companies. *J. Trop. Ecol.* 24, 663–674. <http://dx.doi.org/10.1016/j.foreco.2010.10.003>.
- Rollet, B., 1980. Intérêt de l'étude des écorces dans la détermination des arbres tropicaux sur pied. *Bois For. Trop.* 194, 3–128. <[http://bft.cirad.fr/cd/BFT\\_195\\_31-50.pdf](http://bft.cirad.fr/cd/BFT_195_31-50.pdf)>.
- Saatchi, S., Buermann, W., ter Steege, H., Mori, S., Smith, T.B., 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens. Environ.* 112, 2000–2017. <http://dx.doi.org/10.1016/j.rse.2008.01.008>.
- Sanchez, M., Pedroni, F., Eisenlohr, P.V., Oliveira, A.T., 2013. Changes in tree community composition and structure of Atlantic rain forest on a slope of the Serra do Mar range, southeastern Brazil, from near sea level to 1000 m of altitude. *Flora* 208, 184–196. <http://dx.doi.org/10.1016/j.flora.2013.03.002>.
- Sørensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr.* 5, 1–34.
- Stropp, J., ter Steege, H., Malhi, Y., 2009. Disentangling regional and local tree diversity in the Amazon. *Ecography* 32, 46–54.
- Stropp, J., Van der Sleen, P., Assunção, P.A., da Silva, A.L., ter Steege, H., 2011. Tree communities of white-sand and terra-firme forests of the upper Rio Negro. *Acta Amazonica* 41, 521–544. <http://dx.doi.org/10.1111/j.1600-0587.2009.05811.x>.
- ter Steege, H., 1998. The use of forest inventory data for a national protected area strategy in Guyana. *Biodiversity Conserv.* 7, 1457–1483.
- ter Steege, H., Nigel, C.A., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.F., Monteagudo, A., Vargas, P.N., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G.W., Marimon, B.S., Marimon, B.H., Vieira, I.C.G., Amaral, I.L., Brien, R., Castellanos, H., Lopez, D.C., Duivenvoorden, J.F., Mogollon, H.F., Matos, F.D.D., Davila, N., Garcia-Villacorta, R., Diaz, P.R.S., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J.D., Piedade, M.T.F., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V.A., Peres, C.A., Toledo, M., Gerardo, A.A.C., Baker, T.R., Ceron, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neill, D., Silveira, M., Paredes, M.R., Chave, J., Lima, D.D., Jorgensen, P.M., Fuentes, A., Schongart, J., Valverde, F.C., Di Fiore, A., Jimenez, E.M., Mora, M.C.P., Phillips, J.F., Rivas, G., van Andel, T.R., von Hildebrand, P., Hoffman, B., Zent, E.L., Malhi, Y., Prieto, A., Ruelas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A.A., Schutz, A.C., Gonzales, T., Nascimento, M.T., Ramirez-Angulo, H., Sierra, R., Tirado, M., Medina, M.N.U., van der Heijden, G., Vela, C.I.A., Torre, E.V., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Balslev, H., Ferreira, C., Mesones, I., Torres-Lezama, A., Giraldo, L.E.U., Zagt, R., Alexiades, M.N., Hernandez, L., Huamantupa-Chuquimaco, I., Milliken, W., Cuenca, W.P., Pauletto, D., Sandoval, E.V., Gamarra, L.V., Dexter, K.G., Feeley, K., Lopez-Gonzalez, G., Silman, M.R., 2013. Hyperdominance in the Amazonian tree flora. *Science* 342, 325. <http://dx.doi.org/10.1126/science.1243092>.
- ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Hout, P.V.D., Daly, D.C., Silveira, M., Phillips, O., Vasquez, R., Andel, T.V., Duivenvoorden, J., Oliveira, A.A., Ek, R., Lilwah, R., Thomas, R., Essen, J.V., Baider, C., Maas, P., Mori, S., Terborgh, J., Nunez Vargas, P., Mogollon, H., Morawetz, W., 2003. A spatial model of tree alpha-diversity and tree density for the Amazon. *Biodiversity Conserv.* 12, 2255–2277. <http://dx.doi.org/10.1023/A:1024593414624>.
- ter Steege, H., Pitman, N.C.A., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., Molino, J.F., Prevost, M.F., Spichiger, R., Castellanos, H., von Hildebrand, P., Vasquez, R., 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* 443, 444–447. <http://dx.doi.org/10.1038/nature05134>.
- ter Steege, H., Sabatier, D., Castellanos, H., Van Andel, T., Duivenvoorden, J., De Oliveira, A.A., Ek, R., Lilwah, R., Maas, P., Mori, S., 2000. An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *J. Trop. Ecol.* 16, 801–828. <http://dx.doi.org/10.1017/S0266467400001735>.
- Valeix, J., Mauperin, M., 1989. Cinq siècles de l'histoire d'une parcelle de forêt domaniale de la terre ferme d'Amérique du Sud. *Bois For. Trop.* 219, pp. 13–29. <[http://bft.cirad.fr/cd/BFT\\_219\\_13-29.pdf](http://bft.cirad.fr/cd/BFT_219_13-29.pdf)>.
- Wilkie, P., Saridan, A., 1999. The limitations of vernacular names in an inventory study, Central Kalimantan, Indonesia. *Biodiversity Conserv.* 8, 1457–1467. <http://dx.doi.org/10.1023/A:1008930012467>.