Plant Ecol (2009) 203:303–315
DOI 10.1007/s11258-008-9551-6

Can we reliably estimate species richness with large plots? an assessment through calibration training

Frédéric Archaux · Sylvaine Camaret · Jean-Luc Dupouey · Erwin Ulrich · Emmanuel Corcket · Laurence Bourjot · Alain Brêthes · Richard Chevalier · Jean-Francois Dobremez · Yann Dumas · Gérard Dumé · Marie Forêt · Françoise Forgeard · Myriam Lebret Gallet · Jean-François Picard · Franck Richard · Jean-Marie Savoie · Laurent Seytre · Jean Timbal · Jean Touffet

Received: 14 January 2008 / Accepted: 29 October 2008 / Published online: 2 December 2008

Abstract The number of species (species richness) is certainly the most widely used descriptor of plant diversity. However, estimating richness is a difficult task because plant censuses are prone to overlooking and identification errors that may lead to spurious interpretations. We used calibration data from the French ICP-level II plots (RENECOFOR) to assess the magnitude of the two kinds of errors in large forest plots. Eleven teams of professional botanists recorded all plants on the same eight 100-m^2 plots in 2004 (four plots, eights teams) and 2005 (four plots, nine teams including six from 2004), first independently and then consensually. On average, 15.5% of the shrubs and trees above 2 m were overlooked and 2.3% not identified at the species level or misidentified. On average, 19.2% of the plant species below 2 m in

Electronic supplementary material The online version of this article (doi:10.1007/s11258-008-9551-6) contains supplementary material, which is available to authorized users.

F. Archaux (✉)
CEMAGREF Domaine des Barres,
F-45290 Nogent sur Vernisson, France
e-mail: frederic.archaux@cemagref.fr

S. Camaret
Laboratoire d'Ecologie Alpine (LECA),
Université de Savoie, UFR CISM,
F-73376 Le Bourget-du-lac Cedex, France
e-mail: camaret@univ-savoie.fr

J.-L. Dupouey
INRA Nancy, Forest Ecology and Ecophysiology Unit,
F-54280 Champenoux, France
e-mail: dupouey@nancy.inra.fr

E. Ulrich
Département Recherche, Office National des Forêts,
Boulevard de Constance, F-77300 Fontainebleau,
France
e-mail: erwin.ulrich@onf.fr

E. Corcket
UMR1202 BioGeCo, Université Bordeaux 1, Avenue des Facultés, F-33405 Talence, France

L. Bourjot
Bourjot Environnement, B.P. 300, Savoie Technolac,
F-73375 Le Bouget du Lac, France
e-mail: bourjot@univ-savoie.fr

A. Brêthes
Direction Territoriale de l'Office Nationale des Forêts,
Parc Technologique Orléans-Charbonnière,
100, boulevard de la Salle, B.P. 18,
F-45760 Boigny sur Bionne, France
e-mail: alain.brethes@onf.fr

R. Chevalier · Y. Dumas
Cemagref, Domaine des Barres,
F-45290 Nogent sur Vernisson, France
e-mail: richard.chevalier@cemagref.fr

Y. Dumas
e-mail: yann.dumas@cemagref.fr

J.-F. Dobremez
Université de Savoie, Dynamique des Ecosystèmes d'Altitude, F-73376 Le Bourget du Lac Cedex,
France
e-mail: dobremez@univ-savoie.fr

height were overlooked and 5.3% were misidentified and 1.3% were misidentified at the genus level (especially bryophytes). The overlooking rate also varied with plant species, morphological type, plot and team. It was higher when only one botanist made the census. It rapidly decreased with species cover and increased with plot species richness, the recording time of the census in the tree layer and the number of the censuses carried out during the day in the ground layer. Familiarity of the team with the local flora reduced the risk of overlooking and identification errors, whereas training had little impact. Differences in species richness (over space or time) in large plots should be cautiously interpreted, especially when several botanists participate in the survey. In particular, the quality of the data needs to be evaluated using calibration training and, if necessary, may be improved by involving more experienced botanists working in teams and by fixing a minimum recording time.

Keywords Calibration · Data quality · Long-term monitoring · Observer effect · Plant survey

Introduction

Plant communities may be investigated in forest ecosystems using large plots, in order to assess the impact of different land uses (e.g. Vellend et al. 2007), the temporal changes in the environment (e.g. van Tol et al. 1998) or the plant diversity itself (e.g. Thimonier et al. 1994). If the number of species (species richness) is certainly the most widely used descriptor of plant diversity, it has rarely been stated that the observed number of species might be a systematically biased, underestimate of true species richness because some species are unavoidably missed during the censuses (overlooking errors).

Non-exhaustiveness is a problem for any plant study; it is still more problematic in the case of biogeographical and monitoring studies as many botanists typically carry out the censuses over space and/or over time. Thus, differences in skill level among botanists may cause spurious spatial or temporal trends or may hide true trends. For instance, resampling of old relevés often shows an increase in species richness over time (e.g. Grabherr et al. 1994; Thimonier et al. 1994): this increase may simply result from a change in botanist, possibly because the botanists involved in the resampling searched for species more carefully than the former botanists (furthermore, they may have already known the list of past species, while the former botanists were not aware that their plots would be resampled).

Probably as a result of botanists' generally low awareness of this problem and/or willingness to

G. Dumé · M. Forêt
Inventaire Forestier National, Château des Barres,
F-45290 Nogent sur Vernisson, France
e-mail: gdume@ifn.fr

F. Forgeard · M. Lebret Gallet · J. Touffet
Laboratoire d'Ecologie Végétale,
Université de Rennes I, Complexe Scientifique
de Beaulieu, 35042 Rennes Cedex,
France
e-mail: francoise.forgeard@univ-rennes1.fr

M. Lebret Gallet
e-mail: myriam.lebret@univ-rennes1.fr

J. Touffet
e-mail: ecoveg@univ-rennes1.fr

J.-F. Picard
Institut National de la Recherche Agronomique,
Equipe phytoécologie, F-542820 Champenoux,
France
e-mail: dupouey@nancy.inra.fr

F. Richard
Office National des Forêts, Pont de l'Orta,
20250, Corte, Corsica, France
e-mail: richard.fran@wanadoo.fr

J.-M. Savoie
Ecole Supérieure d'Agriculture de Purpan,
75, voie de TOEC, 31076 Toulouse Cedex 3,
France
e-mail: jm.savoie@esa-purpan.fr

L. Seytre
Conservatoire botanique national du Massif central,
Le Bourg, 43230 Chavaniac-Lafayette, France
e-mail: conservatoire.siege@cbnmc.fr

J. Timbal
Institut National de la Recherche Agronomique,
Centre de Bordeaux-Cestas, Equipe Ecophysiologie
et Nutrition, Unité de Recherche Forestière,
Domaine INRA de l'Hermitage, Pierroton,
F-33610 Cestas, France
e-mail: timbal@zouk.pierroton.inra.fr

accept this fact, relatively few studies have investigated the quality of plant censuses; however, all of them showed high levels of overlooking errors and significant differences between botanists (Nilsson and Nilsson 1985; Lepš and Hadincová 1992; Klimeš et al. 2001; Scott and Hallam 2002; Kercher et al. 2003; Archaux et al. 2006). Only three studies quantified the misidentification rate in plant censuses: Klimeš et al. (2001) found that misidentification was more important than overlooking, whereas Scott and Hallam (2002) and Archaux et al. (2006) found the contrary. However, the procedures used in the three studies to estimate identification errors are questionable since the botanists did not go back to the plots they had sampled to agree together on an accurate final list of species. Some factors influencing overlooking and misidentification have been evidenced, such as plant cover, morphological type and experience of the botanist (Klimeš et al. 2001; Scott and Hallam 2002; Archaux et al. 2006). However, some questions remain. Scott and Hallam (2002) apparently included a non-botanist manager and a student in their analyses, but we do not know whether differences between observers still persist among experienced botanists. Furthermore, many factors potentially impacting the census quality have never been investigated such as vertical layer, the number, training and fatigue of the botanists, although most botanists would acknowledge that these factors influence the quality of the plant censuses.

To reduce overlooking as much as possible, some monitoring programmes have been based on small plots, subdivided into subplots (e.g. Økland 1995 with 1-m^2 plots subdivided in 16 subplots). However, in temperate forests, only a very few species are usually found per m^2, so that many monitoring programmes, including the only pan-European one, the ICP Forests Level II programme (international co-operative programme on assessment and monitoring of air pollution effects on forests), as well as many resampling studies and floristic databases (e.g. Gégout et al. 2005) are based on larger plots. For instance, 577 of the 708 ICP Forests Level II plots are 100 m^2 or more (de Vries et al. 2003).

In this context, our study aimed at:

1. Estimating the exhaustiveness of plant censuses carried out on large (100-m^2) plots to assess whether differences in species numbers (over space or time) can be reliably interpreted using large plots,
2. Investigating the factors that may impact the quality of the censuses, such as plant cover-abundance, plant morphology, vertical layer, number of botanists, their familiarity with the flora, fatigue and training. The identification of impacting factors may help improve sampling protocols.

Methods

Sampling design and relevés

We used the calibration training data gathered by the Quality Assurance (QA) procedure of the RENECOFOR programme (Camaret et al. 2004). The RENECOFOR programme is the French part of the ICP Forests Level II programme and includes 102 permanent plots sampled every 5 years by professional botanists. Each plot is composed of eight 2 × 50-m subplots (four being fenced to exclude large herbivores). One of the objectives of the calibrations is to estimate the overlooking and misidentification rates for the teams involved in the monitoring of the permanent plots (11 teams in 1995, 10 in 2000 and 2005, 16 in total).

Calibrations were organised during two successive years in June. In each year, four 2 × 50 m subplots were chosen so that they were as heterogeneous as possible in terms of species richness and composition: species numbers in the strata below 2 m in height ranged from 15 to 44 among the four subplots sampled in 2004 and from 37 to 67 in 2005 (Table 1). In 2004, calibration training took place near Issoudun (central France), and involved eight teams. In 2005, the calibration was located near Nancy (north-eastern France) and involved nine teams, including six teams who had participated in the calibration near Issoudun. The least experienced botanist had sampled plant communities since 1995, whereas most botanists had more than 15 years of experience (up to 30 years for three botanists).

In each team, the number of botanists varied from one to three (Table 1). On the first day of the calibration training, each team visited each subplot according to a random sequence. They walked

Table 1 Location, sampling year, main tree species and species richness in the ground layer (in italics, species richness in the tree and shrub layers) of the eight subplots, ground-layer vegetation overlooking and misidentification (in italics) rates (%) for the 11 teams (in brackets the number of botanists in the team; only for team five, botanist number was one in 2004 but two in 2005)

Location and year	Main tree species	Species richness	Team's overlooking and *misidentification* rates										
			1 (1)	2 (2)	3 (1)	4 (2)	5 (1/2)	6 (2)	7 (3)	8 (2)	9 (1)	10 (1)	11 (1)
Issoudun 2004	*Quercus petraea*, *Pinus pinaster*	15	26.7	–	–	33.3	40	20	13.3	20	26.7	–	53.3
		4	*0*			*20*	*22.2*	*0*	*0*	*8.3*	*0*		*0*
	Q. petraea	28	17.9	–	–	25	21.4	17.9	21.4	7.1	17.9	–	25
		4	*0*			*9.5*	*0*	*4.3*	*9.1*	*0*	*0*		*0*
	Q. robur, *Carpinus betulus*, *Sorbus torminalis*	44	11.4	–	–	20.5	11.4	13.6	22.7	2.3	20.5	–	25
		12	*7.7*			*8.6*	*12.8*	*10.5*	*14.7*	*4.7*	*5.7*		*9.1*
	Q. petraea	41	24.4	–	–	19.5	17.1	19.5	17.1	14.6	26.8	–	24.4
		8	*3.2*			*6.1*	*8.8*	*3*	*14.7*	*2.9*	*3.3*		*6.5*
Nancy 2005	*Tilia cordata*, *C. betulus*, *Acer campestre*	64	–	14.1	28.1	25	20.3	7.8	–	17.2	18.8	23.4	–
		7		*3.6*	*6.5*	*8.3*	*0*	*3.4*		*1.9*	*0*	*2*	
	Q. petraea, *C. betulus*, *S. aria*	50	–	10	20	18	24	14	–	6	10	26	–
		11		*0*	*10*	*22*	*2.6*	*0*		*2.1*	*6.7*	*2.7*	
	Q. petraea, *C. betulus*, *A. campestre*	67	14.9	11.9	28.4	20.9	11.9	7.5	–	6	11.9	13.4	–
		9	*3.5*	*0*	*8.3*	*13.2*	*1.7*	*0*		*1.6*	*6.8*	*12.5*	
	Fagus sylvatica, *Q. petraea*, *C. betulus*	37	29.7	10.8	32.4	24.3	24.3	10.8	–	10.8	13.5	21.6	–
		7	*3.8*	*6.1*	*0*	*17.9*	*0*	*0*		*3*	*12.5*	*3.4*	

We included in the misidentifications the cases where a team correctly identified a species but also noted a second, wrong name from small/atypical specimens belonging to the same species

outside the subplots to avoid damaging the vegetation as much as possible but entered the subplots whenever necessary for identification. They recorded all vascular plants and terricolous bryophytes and estimated their cover-abundance in four vertical strata using the Braun-Blanquet semi-quantitative scale (*r*: cover less than 5%, only one individual; +: cover less than 5% and rare; 1: cover less than 5% and abundant; 2: 5–25% cover; 3: 25–50% cover; 4: 50–75% cover; 5: 75–100% cover). If plant identification (at the species or genus level) was doubtful, the teams added the *confere* (*cf.*) Latin prefix before the species or genus name. Time spent doing relevés was not controlled but was recorded. On the second day, to produce a consensual list of species in the subplots, all the teams surveyed the four plots again and together re-identified all the species they had individually recorded the day before. During the second day, very few new species were found (one in 2004 and two in 2005). Some plants could not be identified with certainty to the species level (small vegetative specimens). Teams noted the cause of the discrepancies between the consensual list and their own list (e.g. species considered outside the subplot by the team or the consensus, species overlooked, species misidentified, inversion of strata). In 2005, the teams also consensually estimated the plant cover during the second day. To get a similar consensual cover estimate for 2004 plant data, for each plant, we calculated the median of the reported Braun-Blanquet cover values for all the teams that recorded it. In the few cases where an equal number of teams recorded two different cover classes, we kept the most likely one. We weighted each observation by the probability that the cover reported by the team was correct. This probability was calculated for a given team and a given Braun-Blanquet class using 2005 data as the ratio between the number of times the team reported this Braun-Blanquet class correctly (i.e. the consensus agreed upon this cover class for the same plant) and the total number of times the team reported this cover class (correctly or not). The nomenclature is Flora Europaea (Tutin et al. 1968–1980, 1993).

Overlooking and misidentification rates

We considered four kinds of error: overlooking, misidentification at the species level, misidentification at the genus level and complex misidentification. The overlooking rate corresponds to the probability of missing a species during the census. The misidentification rate at the species level is the probability that a plant is misidentified (the species name is wrong); or identified at the genus only (the species name is lacking). Following Scott and Hallam (2002), we considered identifications at the genus level (incomplete identifications) as misidentifications at the species level (when the genus was correct), rather than analysing them separately, because there is a continuum between incomplete identifications and true misidentifications (i.e. the species name is wrong): facing the same seedling of *Quercus robur*, different teams could note *Quercus sp.*, *Quercus cf. petraea* or *Quercus petraea* depending naturally on their level of experience, but also to some extent on their willingness to note incomplete names. None of the misidentifications at the species level for the tree layer were incomplete identifications, whereas incomplete identifications represented 48% of the misidentifications at the species level for the ground layer. The misidentification rate at the genus level is the probability that a plant is given a wrong genus name. Complex misidentifications correspond to cases where a team distinguishes two closely related species (for instance *Tilia cordata* and *Tilia platyphyllos*) when only one is present (e.g. *Tilia cordata*); complex misidentification rate is the probability of committing such an error.

The overlooking rate for a given team in a given plot was calculated as the ratio between the number of species overlooked by this team in this plot and the number of species consensually agreed by all teams in the same plot. The misidentification rate for a given team in a given plot was the ratio between the number of misidentifications made by this team in this plot and the number of species the team had recorded.

The number of cases of misidentification at the species or genus level in the tree layer (respectively, 10 and 3 out of 442 records) and of complex misidentification in the ground layer (18 out of 2,357 records) was too low to explore the factors explaining these errors (thus, only mean values are reported). Therefore, factors affecting misidentification rates at the species and genus level were only analysed for ground vegetation.

Data selection

We analysed the error rates in two vegetation layers (ground and tree layers, respectively, below and above 2 m in height). As data initially comprised four vegetation layers (plus the bryophyte layer), we merged records of vascular species found in the two ground vegetation strata below 2 m and the two tree strata above 2 m and we kept the highest Braun-Blanquet cover value in each. The consensual lists of species retained in 2004 and 2005 are given in Appendix 1 in Supplementary Material (65% of the 148 species recorded were overlooked at least once and 40.5% misidentified at least once).

In the consensus list for each plot, we kept only the taxa that were identified at the species level and deleted those identified only at the genus level or with a *cf.* in the taxon name (on average 3.6 plants per plot for ground vegetation, representing 9.6% of the total number of taxa recorded; with a range over the eight plots of 0–6 plants, 0–21% of the total); thus, all species kept for analyses could be identified at the species level. Alternatively, we could have applied this data selection for the calculation of the misidentification rate only and kept all records for the overlooking rate but we preferred to use the same data set in all the analyses. We considered that the observations with a *cf.* in the taxon name corresponded to correct identifications if the species name recorded by the team was the one agreed upon by the consensus (e.g. team 1 noted *Viola cf. riviniana* and the consensus agreed on the presence of *Viola riviniana*).

Data analysis

Each observation was a binary variable (species recorded or overlooked for the overlooking rate; species identified or not for the misidentification rate) corresponding to a given species in a given subplot sampled by a given team. To relate these variables to sets of explanatory random and fixed factors, we applied generalised linear mixed-effect models (GLMEM) with a binomial logit link function using the lmer function from the R-package lme4

Table 2 Mixed-effect models relating the error rates to random and fixed explanatory variables depending on the strata (Ground: below 2 m, Tree: above 2 m)

Strata	Ground			Tree
Variables	Overlooking	Misid. species	Misid. genus	Overlooking
Random effects				
Plant species	1.45 ± 0.14	1.12 ± 0.04	1.91 ± 0.07	1.30 ± 0.24
Team	0.46 ± 0.12	0.48 ± 0.12	0.36 ± 0.02	0.00 ± 0.00
Subplot	0.27 ± 0.03	0.09 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Fixed effects				
Plant cover	−2.31 ± 0.17***	−0.33 ± 0.24	−0.88 ± 0.68	−3.92 ± 0.62***
Bryophytes	*0.95 ± 0.54*	0.07 ± 0.75	2.91 ± 1.69	
Forbs	−1.27 ± 0.42**	0.31 ± 0.46	1.24 ± 1.41	
Trees/Shrubs	−0.93 ± 0.43*	0.16 ± 0.52	1.11 ± 1.58	
Number of botanists	*−0.45 ± 0.27*	−0.14 ± 0.35	−0.80 ± 0.85	−1.07 ± 0.34**
Order of the census	*0.14 ± 0.07*	−0.03 ± 0.13	0.12 ± 0.33	0.22 ± 0.17
Recording time	−0.01 ± 0.01	−0.00 ± 0.01	−0.03 ± 0.03	−0.03 ± 0.01*
Familiarity	−0.35 ± 0.10***	−0.74 ± 0.14***	−1.02 ± 0.34**	−0.43 ± 0.50
Training	−0.03 ± 0.03	0.14 ± 0.05**	0.13 ± 0.11	0.08 ± 0.05
Species richness	*0.03 ± 0.02*	0.01 ± 0.02	−0.02 ± 0.05	−0.08 ± 0.02**
Understory cover	0.00 ± 0.01	−0.01 ± 0.01	0.02 ± 0.02	
Year/Plot	−0.85 ± 0.56	−1.26 ± 0.68*	−0.21 ± 1.76	1.06 ± 0.69
Diff. identification	−0.37 ± 0.34	1.09 ± 0.37**	1.08 ± 0.98	
Freq. detection		−0.05 ± 0.84	−0.12 ± 2.02	

Coefficients for random effects are standard deviation and their standard deviation calculated from Monte Carlo Markov Chain sampling. Coefficients for fixed effects are estimates and their standard error. Italics: $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. See Method section for explanation about the error rates and explanatory variables

(Bates et al. 2008; R Development Core Team 2008). GLMEM estimates *P*-values for fixed factors but not for random factors. To get an idea of the precision of the estimates and thus the magnitude of the random factors, we generated 1,000 Markov Chain Monte Carlo random effect estimates from the posterior distribution of the fitted GLMEM parameters using R function mcmcsamp; we then calculated the standard deviation of these 1,000 samples.

Explanatory variables included in the models were: plant species, morphological type and Braun-Blanquet class; subplot, subplot species richness, percentage cover of the ground vegetation in the subplot and plot/year (both being confounded); identity of the team (variable hereafter called "team"), recording time, order of the subplot survey in the day, training level of the team and familiarity with the plant species, difficulty to identify the plant, proportion of times the species was overlooked in the subplot. The last variable was naturally not included to model the overlooking rate. Variables used for the four error rates studied can be found in Table 2. Subplot, team and plant species were considered as random factors, all others as fixed factors. This allowed us to simultaneously consider (1) a random team effect and a fixed effect of the number of people in the team, and (2) a random subplot effect and fixed effects of the subplot species richness and percentage cover of the subplot ground vegetation. Plot/Year was considered to be a fixed factor to test the hypothesis that error rates were smaller in 2005 thanks to the 2004 calibration training (an informal test as plot and year were confounded variables).

Morphological type was defined as a four-class variable: bryophyte, graminoid (grasses and grass-like plants), other herbaceous species (forbs), ligneous species (lianas, shrubs and trees) and was used only for ground vegetation (because only ligneous species were recorded above 2 m). A preliminary inspection of the data revealed that species whose

Braun-Blanquet cover was 2 or more were never overlooked. Therefore, we merged the Braun-Blanquet cover classes from 2 to 5 into a single 2+ class (see Fig. 1 for ground vegetation). This ordinal variable was coded from 1 to 4 in the models (1 for the r class, 2 for the +, 3 for the 1 and 4 for the classes over 1). We used the same transformation of Braun-Blanquet cover for the analysis of the misidentification rate since very few plants with Braun-Blanquet cover 2 or more were misidentified (Fig. 2), suggesting that the risk of misidentification for a plant is roughly the same, the plant Braun-Blanquet cover being 2, 3, 4 or 5.

The recording time increased with plot richness according to a seemingly log-linear relationship

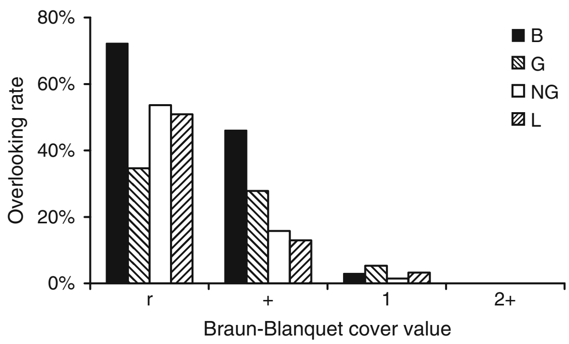

Fig. 1 Relationship between the overlooking rate (%) and the plant Braun-Blanquet cover value in the ground-layer vegetation, according to plant morphology (*B* Bryophyte, *G* Graminoid, *NG* Non-Graminoid herbaceous, *L* Ligneous species)

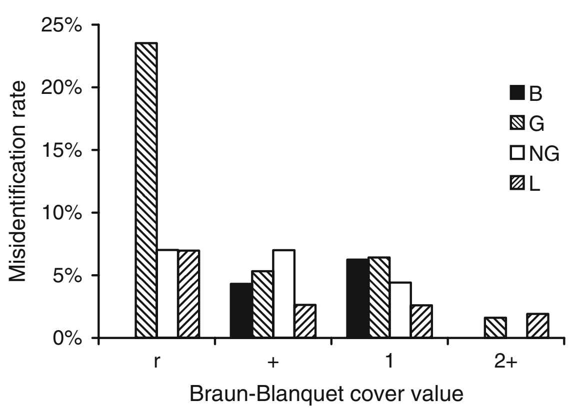

Fig. 2 Relationship between the misidentification rate at the species level (%) and the plant Braun-Blanquet cover value in the ground-layer vegetation, according to plant morphology (*B* Bryophytes, *G* Graminoids, *NG* Non-Graminoid herbaceous, *L* ligneous plants)

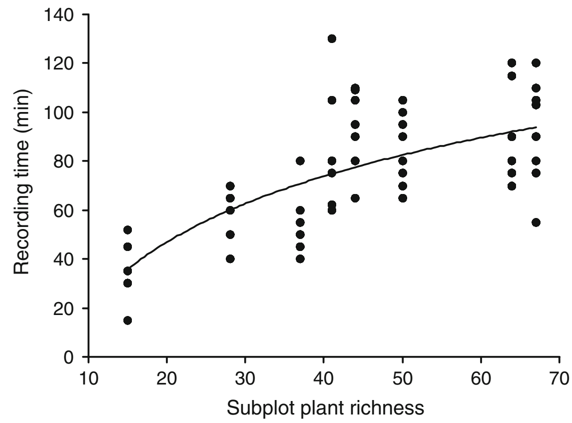

Fig. 3 Relationship between recording time and subplot species richness (eight subplots, 8–9 teams per plot, regression line: time = 40.12 ln(richness) − 74.28, $R^2 = 0.51$)

($R^2 = 0.51$, Fig. 3). In order to distinguish the effect of plot richness from the effect of recording time on the overlooking and misidentification rates, we had to remove this relationship. Therefore, the recording time was entered into the models as the residual of the regression of the recording time on the logarithm of plot plant richness. By using this transformed variable, we effectively tested whether a team who spent more time than expected given the plot richness, detected and identified a greater proportion of the plants.

The variable "order" corresponded to the rank of visit, by each team, of the four plots during the first day of the calibration training (thus the order varied from 1 to 4).

The training of the team was defined as the number of RENECOFOR plots the team had sampled during the 3 months preceding the calibration exercise.

The familiarity of the team with plant species was defined as a four-level ordered class variable: (1) species never encountered before the calibration training; (2) species encountered very occasionally; (3) species that the team records in ca 1% of its censuses (not restricted to the censuses done for the RENECOFOR monitoring); (4) species that the team records in ca 5% or more of its censuses. We considered this variable to be continuous (varying from 1 for unknown species to 4 for well-known species). This information was obtained through a questionnaire filled in by the teams shortly after the 2005 calibration and was lacking for one team (team 5).

We defined the difficulty to identify a species in a plot as a two-state variable depending on whether at least one team (or no team) used the *cf.* prefix for the species in the plot, thus showing they were not sure they had correctly identified it. As a result, a given plant species may be "difficult" in some plots and not in others.

Results

Tree layer

The mean overlooking rate (over the 66 censuses) was 15.5 ± 2.3% SE. The mean misidentification rates were 2.3 ± 0.9% SE at the species level and 0.9 ± 0.5% SE at the genus level. The model indicated high random variation among species and little variation between teams and subplots. The overlooking rate decreased with the plant cover, the number of botanists, the recording time and, surprisingly, the subplot species richness.

Ground vegetation layer

The mean overlooking rate over the 66 censuses was 19.2 ± 1.1% SE. The mean overall misidentification rate at the species level was 5.3 ± 0.7% (including complex misidentifications even if the teams reported the correct species name). Simple confusions—a species being given a single (wrong or incomplete) name—accounted for 102/120 = 85% of the misidentifications. The mean overall misidentification rate at the genus level was 1.3 ± 0.4% SE.

The overlooking rate decreased with plant cover (less rapidly than for species in the tree layer), morphological type (bryophytes being more often overlooked than graminoids, graminoids being more often overlooked than other vascular plants) and familiarity with the species. More marginally, the overlooking rate was smaller (1) in plots with fewer plant species, (2) when teams included at least two people and (3) for censuses done early in the day. Important residual random variation was found among plant species, teams and subplots.

The probabilities of misidentification at the species and genus level were best explained by familiarity of the team with the species. Species for which at least one team reported a *cf.* in the name were more often misidentified at the species level. Unexpectedly, the training increased the risk of misidentification at the species level. The overall misidentification rate at the species level tended to be smaller in 2005 than in 2004, while bryophytes tended to be more often misidentified at the genus level than vascular plants.

Discussion

Sampling errors: mainly overlooking or mainly identification errors?

Our results show that about one out of five ground-layer plant species was missed on average. Although lianas, shrubs and trees in the tree layer were less often overlooked, about one out of six of these species was nonetheless overlooked. Such high overlooking rate values may question either the way we calculated them, the experience of the professional botanists who participated in the RENECOFOR programme and/or the method of survey (large plots, Braun-Blanquet relevés).

Some of the plants may have been missed during both the independent and consensual visits, so that the overlooking rate may actually be slightly higher than the value we estimated. However, very few species were found during the consensual visits done by eight or nine teams suggesting that this effect is likely to be small. A second potential source of bias in the data comes from the fact that some of the plants recorded by some teams could not be found again during the consensual visits, and were thus considered to be misidentifications, instead of overlooking errors. Considering the low level of misidentification errors, it is also unlikely that this phenomenon caused a major bias in the two rates. In addition, most of the botanists who participated in our study are professional botanists who have been recognised in their field for many years in France. Some previous plant studies have quantified observer effect but using a variety of indices, such as the pseudo-turnover rate PT (Nilsson and Nilsson 1985; Lepš and Hadincová 1992; Klimeš et al. 2001; Kercher et al. 2003), the Sørensen similarity index SI (Gray and Azuma 2005) and the agreement rate AR (Kirby et al. 1986; Scott and Hallam 2002. All these indices are closely related (indeed: $PT = 1 - SI = (1 - AR)/(1 + AR)$), so that the

Table 3 Sørensen Similarity Index (SI = 2 * number of paired records/total number of records) reported in and calculated from the literature

Study	Vegetation type	Number of botanists	Quadrat size	Number of quadrats	SI (%)
Nilsson and Nilsson 1985	Swedish forested islands	2	0.03–2.19 ha	41	88.6 (80.6–95.8)
Kirby et al. 1986	British forests	2	0.20 ha	36	70.7 (55–79.5)
Lepš and Hadincová 1992	Central European open land	2	ca 0.025	40	87
Klimeš et al. 2001	Central European grassland	5	10 cm^2–4 m^2	7	(60–88)
Scott and Hallam 2002	Range of British vegetation types	2	0.16 m^2	10 * 10	76.5
Gray and Azuma 2003	Range of North American vegetation types	2	168 m^2	48 * 4	66.6
Kercher et al. 2003	North American herbaceous wetlands	2	1 m^2	12 * 10	80.9 (69.5–90.7)
This study	French forests	11	100 m^2	2 * 4	89.1 (70–97.8)

In brackets, range of SI values between observers and plots. For our study, we calculated the SI by comparing paired raw lists of species recorded by two teams in the same plot

results from these studies can be compared to our study (e.g. using Sørensen similarity index, Table 3). Although conditions varied greatly among studies (in terms of geographic region, vegetation types, number of teams/botanists, quadrat and sample sizes), the SI values are remarkably consistent among studies (incl. our study). It seems that higher values were found for larger plots or equivalently for forest plots (since larger plots are used in forests).

Our results are slightly lower than those reported by Klimeš et al. (2001) (ca 30%) and Archaux et al. (2006) (20–30%), possibly because:

1 Plant species may be more easily detected in the forest (mainly herbs/forbs) than in grasslands (mainly grasses),
2 Mean sampling time was 74 min (for 100 m^2 quadrats), versus 60 min for 400-m^2 quadrats in Archaux et al. (2006),
3 We deleted all records not consensually identified at the species level. These records often referred to low-covering plants (often bryophytes). As low-covering plants are more often missed, our overlooking rate values are probably slightly optimistic.

About 5% of the ground layer species and 2% of the tree layer shrubs and trees were misidentified at the species level (ca 1% at the genus level for the two categories). These figures are very close to the ones reported in Scott and Hallam (2002) in various vegetation types where on average 2.7% of the plants were misidentified at the species level by experts and 0.2% at the genus level. Archaux et al. (2006) found a mean misidentification rate of 6.8% for 400-m^2 forest plots. The slightly higher value reported in the latter study may come from the fact that experts were less familiar with the local flora and/or potential misidentifications were inferred by comparing the team lists without confirmation in the field.

As found by Scott and Hallam (2002), Archaux et al. (2006) and suggested by Kirby et al. (1986), the misidentification rate at the species level was much lower than the overlooking rate. However, Klimeš et al. (2001) suggested that the observer discrepancies observed in small grassland quadrats resulted primarily from the misidentification of small specimens, rather than from overlooking errors. Although this might be true for very small quadrats where often only very few specimens per species are available for identification and/or although the risk of overlooking a species in small quadrats may be lower, this may not be true in general. Furthermore, it should be remembered that misidentifications are conditional on detection, so that we cannot estimate the misidentification rate of overlooked species, had they been detected; we may expect the misidentification rate to be higher for overlooked species, because these mostly include low-covering, infertile specimens. However, the risk of misidentification did not increase with the number of teams who overlooked the species in our study. Similarly, plants presenting a misidentification risk were not more often overlooked. Thus, the (non-estimatable) misidentification rate of overlooked species is probably close to the misidentification rate of detected plants.

Plant and quadrat factors affecting the quality of the data

Plant cover was the main factor influencing the overlooking rate (Fig. 1 and 2), a result repeatedly found in plant studies (Lepš and Hadincová 1992; Klimeš et al. 2001; Archaux et al. 2006). On the contrary, the risk of misidentification was not significantly related to the cover of the plant, although Fig. 2 suggests this may hold for small, isolated graminoids. Scott and Hallam (2002) found that both the overlooking and misidentification rates varied greatly between species morphological types. In our study, the probability of missing a species was effectively greater for bryophytes than for graminoids, and greater for graminoids than for forbs and ligneous species. Bryophytes also tended to be misidentified at the genus level more often than vascular plants.

More marginally, species in the ground layer had a greater probability of being missed in rich plots (but not of being misidentified), but surprisingly, the reverse was found for trees and shrubs: botanists seem to more carefully screen the shrub and tree layers when the ground layer is species rich. This last result needs to be confirmed. Nonetheless, a bias of census exhaustiveness between rich and poor plots in favour of poor plots, similar to the one we found in the ground layer, had already been evidenced in temperate floristic relevés; this bias tended to vanish with longer censuses (Archaux et al. 2006).

Team-related factors affecting the quality of the data

We found strong, complex observer effects on both the overlooking and misidentification rates; these effects were related to the identity of the botanists, the familiarity with the local flora and, to a lesser extent, to the number of botanists in the team, training and fatigue. Observer effects were the main factor affecting the misidentification rates (together with the difficulty to identify the species for the probability of simple misidentification at the species level). Our results (and others) are contrary to the suggestion by Kirby et al. (1986)'s that overlooking is mainly a matter of chance and therefore, should be more or less constant across observers.

Fewer species were missed when teams were composed of at least two people, certainly because the plot area effectively sampled increases with the number of people surveying the plot. The teams with at least two observers often included one expert plus one or two less-experienced observer(s): this probably explains why the number of observers had little impact on misidentification rates. Klimeš et al. (2001) recommended that at least three observers participate in plant censuses to guarantee data quality.

Experience has been repeatedly pointed out as one of the main factors affecting the quality of vegetation censuses (Kirby et al. 1986; Klimeš et al. 2001; Scott and Hallam 2002; Archaux et al. 2006). Scott and Hallam (2002) give figures showing that experts misidentify fewer species than less competent observers (misidentification rate of 2.7% for experts against 4.6% for competent observers and 14.1% for less-experienced observers). It is not easy to evaluate the experience of an observer. In our case, we separated two aspects of experience: the familiarity with the local flora and the number of censuses carried out during the three months preceding the calibration exercises (training). Logically, the teams tended to more often misidentify the species they were less familiar with; a less expected result was that the teams also overlooked them more often: part of the overlooking errors concerning species teams are not familiar with are probably misidentifications in reality. This phenomenon probably explains why teams who misidentified a greater proportion of species also tended to miss more species (although the relationship is weak, Fig. 4). As a result, increasing the recording time would not necessarily reduce the differences in observed richness between teams; Archaux et al. (2006) effectively observed that differences in observed species richness between botanists do not decrease as the recording time increases.

Contrary to our expectation, training did not improve the quality of the censuses; on the contrary, we found that teams that carried out more censuses three months before the calibration trainings misidentified a greater proportion of plants in the ground layer. This would support the hypothesis that a kind of routine occurs in the course of a survey, i.e. trained teams may identify plants too quickly. Similarly, censuses done late in the day were generally less exhaustive in the ground layer than censuses done

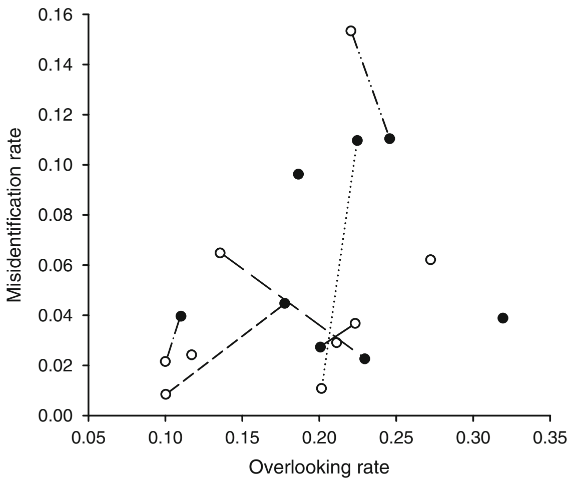

Fig. 4 Relationship between the overlooking rate and the misidentification rate for ground vegetation in 2004 (open symbols) and 2005 (closed ones) (Spearman's $r = 0.40$, $z = 1.58$, $P = 0.11$). Paired dots correspond to the six teams who were involved both years

early in the day. One explanation may be that the vegetation is progressively damaged by the visits of the teams, although the teams mainly stayed outside the subplots. A more likely explanation is that the quality of the censuses diminishes as teams become more tired.

Surprisingly, the recording time reduced only the overlooking rate in the tree layer, indicating that recording time essentially limits gross errors. The limited effect of recording time on the quality of census results is contrary to Archaux et al. (2006) but agrees with the statement by Klimeš et al. (2001) that some observers/teams of observers may be slower or faster at completing their censuses. However, neither our work, nor the study by Klimeš et al. (2001) investigated how recorded species accumulate during the censuses, contrary to Archaux et al. (2006); this difference may explain why conclusions differed between these studies.

Finally, plants for which certain teams expressed some doubt about identification were indeed more often misidentified than the average.

Recommendations for plant surveys

The comparison of our results with former studies shows that a significant and relatively constant proportion of plant species are missed or misidentified in vegetation surveys. To our knowledge, however, the study of the magnitude of the observer effect has been restricted to temperate and boreal ecosystems and has never been done in tropical ecosystems. In particular, the overlooking and misidentification rates in the tree layer are likely to be higher in tropical forests than in temperate and boreal forests, due to the far higher diversity in tree species. In temperate and boreal regions, our results question the use of large plots to track differences in species richness over space or time, especially if the expected differences in richness are small. For instance, plant communities usually evolve slowly during the forest succession (except during the youngest successional stages) (Aubert et al. 2003); thus long-term monitoring programmes such as ICP Forests may not allow us to reliably show such slow dynamics, unless plots are surveyed by a single team over time and team skills do not change over time (two options probably not realistic in the long term). Similarly, biogeographical studies based on large data sets gathered by different teams of botanists in different areas may suffer similar biases. For instance, in 2000, teams participating in RENECOFOR were asked to record the presence of browsing or fraying indices in the plots in addition to the name and abundance of the plant species: the spatial distribution of the records of browsing and fraying exactly matches the spatial distribution of the teams (Camaret et al. 2004). Since the teams survey groups of plots, entire regions, sometimes with high abundance of large herbivores, seem erroneously free from herbivore impact.

There are obvious ways to reduce these errors such as involving only botanists familiar with the local flora, making only a few censuses per day and limiting their number during the vegetation season, and regularly enhancing team motivation (as species are much more often missed than misidentified). In our study, most botanists were not experts in bryology, which may explain why many bryophytes were overlooked and/or misidentified: we would recommend sampling vascular plants and bryophytes separately because (1) searching for bryophytes on the ground distracts the attention of the botanist from higher vegetation layers and vice versa, and (2) botanists often have different identification skills for the two taxonomic groups.

The overlooking rate could also be reduced using smaller plots by focusing the attention of the botanists on a smaller area. However, this may not necessarily be the case. For instance, Archaux et al. (2007) found the repeatability of species richness to be similar for 2, 4 or 400-m^2 plots (see also Table 3). Increasing the recording time may be more effective than reducing the plot size. For instance, on 100-m^2 forest plots, it may be necessary to spend at least 2 or 3 hours to get more acceptable levels of exhaustiveness (it my also be sensible to subdivide the plots into smaller subplots, e.g. 50 1 m × 2 m subplots). Fixing such a minimum recording time may also limit the bias of exhaustiveness we observed between species-rich and species-poor quadrats (cf Archaux et al. 2006).

A second significant improvement of plant surveys would be the use of teams of observers, even if all of them are not experts, as recommended by Klimeš et al. (2001). Teams composed of one expert in vascular plants and one bryologist would probably be an ideal configuration for plant studies. As species whose identification teams doubted were indeed more often misidentified, we strongly encourage teams involved in monitoring programmes to note their doubt in the field as these records could be more easily related to previous or future records on the same plots.

Finally, even though calibration exercises apparently failed to significantly reduce observer effects, we do think they help convince botanists of the necessity to spend enough time on the plots to carefully check all specimens of species prone to misidentification. Calibrations also have some limits: bringing all the teams together is costly and team behaviour during training is likely to be different from that in the field. Control visits are better adapted to obtain more realistic estimates of observer errors: during the survey, each team samples a few plots shortly after these plots have been sampled by other teams; observer effects can be estimated from these pairs of independent censuses done by different teams on the same plots. However, control visits cannot be used to distinguish between overlooking and identification errors (unless the controlled team accompanies the control team during the control visit to establish a consensual list). Thus, we recommend associating calibrations and control visits in plant surveys in general and in monitoring programmes in particular.

Acknowledgements We thank Frédéric Gosselin for help in data analysis, Victoria Moore for carefully re-reading the manuscript and Leos Klimeš and two reviewers for constructive comments on the manuscript.

References

Archaux F, Bergès L, Chevalier R (2007) Are plant censuses carried out on small quadrats more reliable than on larger ones? Plant Ecol 188:179–190

Archaux F, Gosselin F, Bergès L, Chevalier R (2006) Effects of sampling time, quadrat richness and observer on exhaustiveness of plant censuses. J Veg Sci 17:299–306

Aubert M, Alard D, Bureau F (2003) Diversity of plant assemblages in managed temperate forests: a case study in Normandy (France). For Ecol Manage 175:321–337

Bates D, Maechler M, Dai B (2008) The lme4 Package. Available at: http://lme4.r-forge.r-project.org/

Camaret S, Bourjot L, Dobremez JF (2004) Suivi de la composition floristique des placettes du réseau (1994/95–2000) et élaboration d'un programme d'assurance qualité intensif. Office National des Forêts, Direction Technique, Fontainebleau

de Vries W, Reinds G, Posch M, Sanz MJ, Krause G, Calatayud V, Renaud J, Dupouey J, Sterba H, Vel E, Dobbertin M, Gundersen P, Voogd J (2003) Intensive monitoring of forest ecosystems in Europe, Technical Report 2003. EC-UN/ECE, Brussels, Geneva

Gégout JC, Coudun C, Bailly G, Jabiol B (2005) EcoPlant: a forest site database linking floristic data with soil and climate variables. J Veg Sci 16:257–260

Grabherr G, Gottfried M, Pauli H (1994) Climate effects on mountain plants. Nature 369:448

Gray AN, Azuma DL (2005) Repeatability and implementation of a forest vegetation indicator. Ecol Indic 5:57–71

Kercher SM, Frieswyk CB, Zedler JB (2003) Effects of sampling teams and estimation methods on the assessment of plant cover. J Veg Sci 14:899–906

Kirby KJ, Bines T, Burn A, Mackintosh J, Pitkin P, Smith I (1986) Seasonal and observer differences in vascular plant records from British woodlands. J Ecol 74:123–132

Klimeš L, Dancák M, Hájek M, Jongepierová I, Kucera T (2001) Scale-dependent biases in species counts in a grassland. J Veg Sci 12:699–704

Lepš J, Hadincová V (1992) How reliable are our vegetation analyses? J Veg Sci 3:119–124

Nilsson IN, Nilsson SG (1985) Experimental estimates of census efficiency and pseudoturnover on islands: error trend and between-observer variation when recording vascular plants. J Ecol 73:65–70

Økland RH (1995) Changes in the occurrence and abundance of plant species in a Norwegian boreal coniferous forest, 1988–1993. Nordic J Bot 15:415–438

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Scott WA, Hallam CJ (2002) Assessing species misidentification rates through quality assurance of vegetation monitoring. Plant Ecol 165:101–115

Thimonier A, Dupouey JL, Bost F, Becker M (1994) Simultaneous eutrophication and acidification of a forest ecosystem in North-East France. New Phytol 126: 533–539

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA (1968–1980, 1993) Flora Europaea. Cambridge University Press, 5 vols

van Tol G, van Dobben HF, Schmidt P, Klap JM (1998) Biodiversity of Dutch forest ecosystems as affected by receding groundwater levels and atmospheric deposition. Biodiv Cons 7:221–228

Vellend M, Verheyen K, Flinn KM, Jacquemyn H, Kolb A, van Calster H, Peterken G, Graae BJ, Bellemare J, Honnay O, Brunet J, Wulf M, Gerhardt F, Hermy M (2007) Homogenization of forest plant communities and weakening of species–environment relationships via agricultural land use. J Ecol 95:565–573