A limited number of species is sufficient to assign a vegetation plot to a forest vegetation unit

Lise Maciejewski, Paulina E. Pinto, Stéphanie Wurpillot, Jacques Drapier, Serge Cadet, Serge Muller, Pierre Agou, Benoît Renaux, Jean-claude Gégout

MS. LISE MACIEJEWSKI (Orcid ID : 0000-0002-9142-0507)

DR. PAULINA E. PINTO (Orcid ID : 0000-0002-1628-1826)

Article type : Methods in Vegetation Science

Jörg Ewald

A limited number of species is sufficient to assign a vegetation plot to a forest vegetation unit

Running title: few species to identify a vegetation unit

Lise Maciejewsk[1, 2], Paulina Elcira Pinto[1], Stephanie Wurpillot[3], Jacques Drapier[4], Serge Cadet[5], Serge Muller[6], Pierre Agou[7], Benoît Renaux[8], Jean-Claude Gégout[1]

ORCID:

Lise Maciejewski https://orcid.org/0000-0002-9142-0507

Paulina Elcira Pinto https://orcid.org/0000-0002-1628-1826

Serge Muller https://orcid.org/0000-0002-4545-9427

Jean-Claude Gégout https://orcid.org/0000-0002-5760-9920

Authors' institutional affiliations

[1] Université de Lorraine, AgroParisTech, INRAE, Silva, 54000, Nancy, France.

[2] OFB, MNHN, CNRS, UMS 2006 PatriNat, CP 41, 36 rue Geoffroy-Saint-Hilaire F-75005 Paris, France.

[3] IGN (Institut National de l'Information Géographique et Forestière), 73 Avenue de Paris, 94160 Saint-Mandé, France.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/AVSC.12650

[4] IGN (Institut National de l'Information Géographique et Forestière), Direction interrégionale nord-est, 1 rue des Blanches Terres, 54390 Champigneulles, France.

[5] ONF (Office national des forêts), Réseau Habitats-Flore, Bureau d'études territorial 13/84, 46 av. Cézanne, 13098 Aix-en-Provence, France.

[6] MNHN, CNRS, UPMC, EPHE, UMR 7205 ISYEB, F-75005 Paris, France.

[7] Biotope, Centre Bourgogne, 122-124 Faubourg Bannier, 45000 Orléans, France.

[8] Conservatoire botanique national du Massif Central, Le Bourg, 43230 Chavaniac-Lafayette, France.

Correspondance

Paulina Elcira Pinto, Université de Lorraine, AgroParisTech, INRAE, Silva, 14 rue Girardet F-54000 Nancy, France. Email: paulina.pinto@agroparistech.fr

Funding information

This study was supported by the French National Institute for Agricultural, Food and Environmental Research (Forest Grassland and Freshwater Ecology Department, EFPA) through the ONF-INRAE Interface Grant "Station, distribution, croissance et choix des essences dans un contexte environnemental changeant", by the French National Research Agency (ANR) through the Laboratory of Excellence ARBRE (ANR- 12-LABXARBRE-01), and by the Regional Council of Grand Est through the project "Contribution à l'identification et à la cartographie à fine résolution des zones humides forestières à l'aide du caractère hygrophile des plantes".

ABSTRACT

Aims

Inventorying the habitats composing Natura 2000 sites is mandatory in the European Union and is necessary to implement relevant conservation measures. Vegetation plots, recording the presence or abundance of all plant species co-occurring within a plot, are currently used to identify terrestrial Natura 2000 habitat types, whose descriptions are mainly based on phytosociological units. However, vegetation plots are time-consuming and frequently restricted to the growing season. Moreover, no vegetation plots can be regarded as exhaustive, and significant inter-observer variation has been highlighted. We studied whether reducing the number of recorded species and the time spent carrying out a vegetation plot had an impact on vegetation unit assignment using species presence. We also studied if vegetation plots recorded on winter could be used for vegetation unit assignment.

Location: mainland France.

Methods

We used 273 vegetation plots covering French temperate and mountainous forests. The time at which species were sighted was recorded. We also estimated whether a species was recognisable in winter. We used a classification program to compare assignments based on complete and incomplete vegetation plots.

Results

Ten species and five minutes were sufficient to assign a plot to an association, and seven species and four minutes to an alliance. Vegetation unit assignment proved feasible in winter, especially at the alliance level.

Conclusions

We confirmed that a limited number of species is sufficient to assign vegetation plots to vegetation units. However, mapping habitats requires habitat identification and delimitation. This study confirms current field habits, particularly when creating a habitat map, usually based on a limited number of recorded species. Lastly, it confirms that the

use of vegetation plots coming from a great variety of sources is relevant to create habitat time series, crucial tools for monitoring habitats at a national scale.

Keywords: vegetation classification, automatic classification program, expert system, habitat, forest ecosystem, phytosociology, Natura 2000, vegetation typology, incomplete vegetation records.

1. INTRODUCTION

The natural and semi-natural habitat types of Community interest listed in Annex I of the Habitats Directive ('Natura 2000 habitat types' hereafter) are rare, in danger of disappearance, or present outstanding examples of typical characteristics of one of the biogeographical regions in Europe (EC Council, 1992). Phytosociology has long been used for plant community description in Europe. As a consequence, phytosociological units were used to describe the terrestrial habitats of Community interest (Evans, 2010). The Habitats Directive states that all European Union countries must protect and restore the species and habitats of Community interest (EC Council, 1992). At a national level, each European Union country must monitor, assess and report to the European Commission the conservation status of these threatened species and habitats (Art. 11 and Art. 17). Within their territory, they have to implement a coherent ecological network of Special Areas of Conservation (SAC) set up under the title Natura 2000 (Art. 3) along with the special protection areas created by the Birds Directive (EC Council, 2010). In this context, important means are deployed at the European, national, and local levels to identify and monitor the targeted ecosystems.

Mapping habitats requires habitat identification and delimitation. Therefore, carrying out an inventory of the habitat types occurring in the SAC of the Natura 2000 network is necessary to create maps of endangered ecosystems, which are key for implementing relevant management actions. To carry out such an inventory, site managers have to assign vegetation observations to vegetation units, whether they record the presence of plant species co-occurring in vegetation plots or simply spot several plant species without recording them. In practice as well as traditionally in phytosociology, assignments are performed manually in the field on the basis of species presence and expert human knowledge. That is how vegetation plots, currently provide the data used to identify Natura 2000 habitats. Vegetation plots are often restricted to the growing season, require botanical and phytosociological skills, and are time-consuming. For example, completing a vegetation plot takes one hour in the protocol of Archaux et al. (2006). This time effort can be a limiting factor in the face of the large number of vegetation plots necessary to carry out habitat inventories within Natura 2000 sites.

Vegetation plots involving observers are characterized by an inherent degree of error consisting of overlooking error (not observing actually present species), and misidentification error (incorrectly identifying species) (Morrison, 2016). Ten to 30 % of the species remain undetected by single observers (Archaux et al., 2006; Morrison, 2016; Vittoz et al., 2010; Vittoz and Guisan, 2007). Misidentification error generally appears to be of a smaller magnitude than overlooking error, and unsurprisingly smaller at the genus level than at the species level (Morrison 2016). Only considering overlooking error (as we do in this study), two different observers will inevitably produce two different inventories of species for a given plot. Thus, no vegetation plot can be regarded as exhaustive.

Another source of potential incompleteness of vegetation plots is the variability of plot size according to the observer and the sampling method (Chytrý and Otýpková, 2003; Chytrý and Tichý, 2018). These data are frequently missing for part of the plots in national or international databases (Chytrý et al., 2016). While there is no clear evidence of a relationship between plot size and error in species lists (Morrison 2016), plot size theoretically affects the species richness of a vegetation plot positively. To deal with this issue, the traditional phytosociology method recommends to collect data on a "minimal area", whose value is given by the asymptotically horizontal part of the species-area curve (Braun-Blanquet et al., 1932). But the numerous methods proposed to determine ''minimal area'' (e.g. Hopkins, 1957) ignore the nature of species-area curves (Preston, 1962), which usually cannot exhibit an asymptote (Williamson et al. 2001; Dengler et al. 2009). As a result, species richness in a same place could substantially differ among vegetation plots based on different plot sizes.

Vegetation plots are often carried out during the growing season and are time-consuming. Moreover, no species inventory of vegetation plots can be considered exhaustive because of systematic overlooking errors, and because plot size can vary depending on observers and protocols and thereby affect species richness. We do not know yet if these findings affect the vegetation unit assignment process, with potential impacts on the inventory of Natura 2000 habitats within sites, and their monitoring at a national level. This is why we assessed the influence of the number of species recorded in vegetation plots on vegetation unit assignment. More precisely, we investigated whether using incomplete vegetation plots – which implies recording a lower number of species and spending less time carrying out the vegetation plot– had an impact on vegetation unit assignment using species presence.

2. MATERIALS AND METHODS

We used a large number of vegetation plots well distributed across mainland French temperate and mountain forests to perform a large-scale study encompassing a wide range of forest ecosystems. The time at which plant species were sighted since the beginning of the inventory was recorded for each vegetation plot. We created incomplete vegetation plots, by truncating the complete floristic inventories of the plots. Then, we used an automatic classification program for vegetation unit assignment based on both complete and incomplete vegetation plots, and we compared their efficiency.

2.1 Sampled sites and vegetation plots

Six teams of field agents of the French National Forest Inventory (NFI) carried out 273 vegetation plots between May and September 2013 across temperate and mountainous forests of mainland France (Fig. 1) (see Hervé 2016 for a presentation of the French NFI method). Each plot had been randomly localized before the fieldwork phase. Vascular plant species found in the understory layer and terricolous bryophytes were recorded across each circular 700-m² plot (15-m radius). The floristic inventory included all herbaceous and shrub species, and tree species below 7.5 cm in diameter at 1.30 m height. The time, and thus the order in which species were sighted was recorded. The cover of each species in the plot was assessed visually using the Braun-Blanquet approach (Braun-Blanquet et al., 1932), after completion of the timed vegetation inventory (see Pinto et al., 2016 for a more detailed description of the 273 timed vegetation plots). A coefficient of 1 was applied for the species whose cover was missing

(15%). The cover of tree species above 7.5 cm in diameter was assessed visually, using 10 % classes, at the end of the understory layer inventory.

2.2 Incomplete vegetation plots

2.2.1 Incomplete vegetation plots truncated according to the number of species

The mean species richness of vegetation plots was 33 species (STD=16), with a minimum of four species, and a maximum of 83 species. We created incomplete vegetation plots by selecting an increasing number of species per vegetation plot: one, two, three, until the total species richness of each plot was reached. Four methods of selection of the subset of species were used: (i) random selection (100 repetitions), (ii) according to the inventory order of appearance of the species in the vegetation plots ('increasing order' hereafter), (iii) according to the decreasing order of appearance of the species in the vegetation plot ('decreasing order' hereafter), and (iv) the decreasing abundance of the species in the vegetation plots. For this last method of selection, we first randomly selected species among those with the highest cover observed in the vegetation plot. Then, we randomly selected the next species in the second highest cover observed in the vegetation plot, and so on until the total richness of the plot was reached. In the case of a species present in both the understory and tree layers, we used its highest cover.

2.2.2 Incomplete vegetation plots truncated according to the time spent in the field

The surveyors spent 5 to 62 minutes to carry out each vegetation plot in the field, which represented the range of the time effort needed to carry out a complete floristic inventory. Using the time when species were recorded, we created incomplete vegetation plots of increasing duration.

2.2.3 Incomplete vegetation plots truncated according to the species potentially recognisable in winter

The vegetation plots were carried out between May and September. To assess the feasibility of vegetation unit assignment in winter, we created incomplete vegetation plots (called 'winter plots' hereafter) with species that we considered recognisable during the winter season: phanerophytes, chamaephytes, woody lianas, mosses, and species with leaves still present in winter and evergreen leaves. The species richness of the winter plots was equal or less than the species richness of the complete vegetation plots. To avoid bias linked to richness differences between 'winter plots' and complete vegetation plots, we equalize the species richness by randomly choosing a subset of species (called 'random winter plots' hereafter) among the complete vegetation plot. In order to average the variability linked to random selection, 100 series of 'random winter plots' were created for each plot.

2.3 Automatic classification program

To study whether reducing the number of species and time spent carrying out a vegetation plot had an impact on vegetation unit assignment, we tested the assignment of each incomplete vegetation plot to a vegetation unit. To this end, we used a typology based on the units published at the association and alliance levels (228 and 43 phytosociological units, respectively) (Bardat et al., 2004; Bioret et al., 2014; Gégout et al., 2009; Maciejewski et al., 2020). To repeat the assignments as many times as necessary, based on a previous study we chose the 'Phi-program' (Gégout and Coudun, 2012; Maciejewski et al., 2020), an automatic classification program which performs similarly as individual experts in vegetation unit assignments (Maciejewski et al., 2020). This automatic classification program uses species fidelity indexes to classify vegetation plots. It works with a baseline of 9,827 vegetation plots (containing 1,648 species) classified by experts at the association level within the chosen phytosociological system. These vegetation plots were used to compute a fidelity index φ (the Phi coefficient) (Chytrý et al., 2002) of each of the 1,648 species to each of the 228 associations. As a result, we obtained the matrix M, with the Phi coefficient of each species to each association, with 1,648 rows and 228 columns. The Phi-program tested the assignment of a given vegetation plot to each association by extracting the n species present in the vegetation plot from the matrix M and calculating the average value of the fidelity indexes of these n species, all vegetation layers taken together, for each association. The program assigned the vegetation plot to the association whose average fidelity index was the highest.

The objective was to compare the efficiency of the Phi-program for vegetation unit assignment based on complete and incomplete vegetation plots. The efficiency of automatic classification programs has usually been assessed based on one expert judgment considered as a “true assignment” (e.g. Černá and Chytrý 2005; van Tongeren et al. 2008; De Cáceres et al. 2009), which could be merely regarded as a validation. But strong and significant inter-expert variation has already been observed (Eriksen et al., 2019; Hearn et al., 2011; Stevens et al., 2004; Ullerud et al., 2018). Consequently, the choice of the reference is of crucial importance, and the use of a unique reference can be contested. This is why we assessed the efficiency of the Phi-program in a previous study not by comparing it with one expert, but by comparing it with the efficiency of five expert organisations (the National Botanical Conservatory of the Massif Central; Biotope, an ecological consultancy; the botanist’s network of the National Forestry Office of France; the National Museum of Natural History of France; the French National Forest Inventory). To do so, we used an index of consistency called ‘agreement ratio’ (Maciejewski et al., 2020). A lower agreement ratio than those of the expert organisations would have meant a poor ability of the program to reproduce expert judgments. A higher agreement ratio would have meant that the program could agree with several experts more often than one expert with several others, suggesting that it could provide a more consensual assignment.

2.4 Comparing the assignments of the automatic classification program based on complete and incomplete vegetation plots

The agreement ratio of the automatic classification program run with complete vegetation plots was 0.25 for association assignment, slightly below 0.27, the average agreement ratio of the five expert organisations (Fig. 2a). It was 0.48 for alliance assignment, similar to the average agreement ratio of the five expert organisations (Fig. 2b). Thus, the agreement ratios of the automatic classification program were quite similar to the agreement ratios of the expert organisations. These ratios were defined as the reference agreement ratio (RAR hereafter, Fig 2) to evaluate the efficiency of incomplete vegetation plots to provide assignments similar to those of experts.

For each number of species or time spent in the field studied, the program assigned a vegetation unit (an association and an alliance) to each of the 325 vegetation plots. We compared these assignments with the ones provided by the five expert organisations and calculated an agreement ratio. Then, we calculated the percentage of this agreement ratio against the RAR to interpret the results more easily, and we assessed the number of species, or the time, needed to reach 100 % of the RAR. The closer the agreement ratios drew to the RAR, the more relevant the species or group of species in the incomplete vegetation plots were, and brought substantial information for vegetation unit assignment.

Slightly lower accuracy can be accepted in some projects, or depending on the available means. In these contexts, reaching 90% of the RAR obtained with complete vegetation plots could be considered as sufficient. Furthermore, the method also allowed exceeding 100 %, i.e. a higher agreement ratio than the one obtained with the complete vegetation plot if the last species recorded in the field had a confounding effect on the assignment of a relevant vegetation unit to the vegetation plot.

We calculated the average, minimum and maximum values of the agreement ratios of the 100 series of incomplete vegetation plots of randomly chosen species order. We also ran a Pearson correlation test to investigate the relation between the order and the cover of each species, to know whether the first species to be observed were the most abundant ones.

We calculated the agreement ratio and the percentage of the RAR reached with the winter plots. Then, we investigated whether the decrease in species richness in winter or the types of removed species most impacted the agreement ratios. To do so, we calculated the median, minimum and maximum values of the agreement ratios of the 100 series of randomly selected winter plots.

3. RESULTS

3.1 Number of species required for vegetation unit assignment

The percent of the reference agreement ratio (RAR) reached by the agreement ratio of the Phi-program using incomplete vegetation plots increased faster when selecting species according to their decreasing abundance or increasing order than when selecting them according to a random selection or a decreasing order (Fig. 3). The most abundant species or the first sighted species were more relevant for vegetation unit assignment that the last ones. The correlation between the order and the cover of the species was significant ($p<0.001$), but the order only explained 8 to 11% of the variability of the cover.

Based on the incomplete vegetation plots of increasing order, which reflected an on-the-ground reality, the first 21 species sighted and recorded in the field were sufficient for the automatic classification program to reach nearly the same agreement ratio with incomplete vegetation plots as with complete vegetation plots for association and alliance assignment (100 % and 99 % of the RAR, respectively) (Fig. 3).

To reach 90% of the RAR, only the first ten species of the vegetation plot were needed to assign an association, and only the first seven species were needed to assign an alliance (Fig. 3). Beyond these numbers, the new added species did not bring significant extra information for vegetation unit assignment.

3.2 Time spent on the field for vegetation unit assignment

In the first minute of the vegetation plot, 74 % of the RAR was reached regarding association assignment, and 86 % regarding alliance assignment (Fig. 4). The first five minutes and first four minutes were enough to record species reaching > 90 % of the RAR for association and alliance assignment, respectively. Twenty-five minutes were needed to reach 100 % of the RAR for both assignments (Fig. 4). After the 25th minute, the agreement ratios stabilized around 100% of the RAR.

3.3 Influence of the season on vegetation unit assignment

The mean species richness of the winter plots was 22 (STD=9), i.e. an average decrease of slightly more than 30% in comparison with the complete vegetation plots. The agreement ratio of the Phi-program based on winter plots was 0.21 for association assignment (84 % of the RAR), and 0.47 for alliance assignment (97 % of the RAR) (Fig. 5).

Using the random winter plots, the mean agreement ratio was 0.23 for association assignment (min. 0.21; max. 0.26), and 0.47 for alliance assignment (min. 0.44; max. 0.48) (Fig. 5). Therefore, the species considered as recognisable in winter were the least relevant ones for association assignment, but were quite relevant for alliance assignment.

4. DISCUSSION

A limited number of species or a limited field time for flora inventory proved sufficient to assign vegetation plots to vegetation units. In addition, a subset of species recognisable in winter proved sufficient for vegetation unit assignment, particularly at the alliance level. We performed a large-scale study involving a great number of vegetation units and covering a wide range of temperate and mountainous forest ecosystems. As the definition of terrestrial Natura 2000 habitats is mainly based on phytosociological units, these results can likely be extended not only to Natura 2000 forest habitats, but to other habitats of Community interest, especially those with definitions focused on species. As to possible vegetation unit assignments in winter, the result can be extended to Natura 2000 forest habitats, but maybe not to other habitat types, e.g. grassland habitats composed of a majority of hemicryptophytes barely recognisable after the growing season.

4.1 Extra information on the possible use of incomplete vegetation plots

No vegetation plot can be regarded as exhaustive, and significant inter-observer variation exists among specialists carrying out vegetation plots in the field (Archaux et al., 2006; Morrison, 2016; Vittoz et al., 2010; Vittoz and Guisan, 2007). However, this study shows that the incompleteness of vegetation plots has only a slight impact on vegetation unit assignment. Some studies advise caution when using incomplete vegetation plots; for example, relative species diversity metrics are influenced by undersampling but are more robust to inventory incompleteness than absolute measures (Azovsky, 2018). Furthermore, a trade-off between robustness to undersampling and sensitivity to rare species has been highlighted (Beck et al., 2013). But a possible use of incomplete vegetation plots has been shown in other studies to detect species temporal trends (Bruelheide et al., 2020) or predict species distributions (Braunisch and Suchant, 2010). Moreover, strong robustness to vegetation plot incompleteness has been proved when bioindicating forest soil acidity, nitrogen and mineral nutrition (Pinto et al., 2016). Therefore, our study complements previous results by showing that vegetation unit assignment is very robust to the use of vegetation plots considered as incomplete because of overlooked species.

4.2 Plot vegetation species are not all equally informative in vegetation unit assignments

Using the same number of species, we obtained different agreement ratios according to the species kept in the incomplete vegetation plots. More specifically, the most abundant species or the first species to be observed were the most relevant ones for assigning a vegetation unit. A significant correlation between the order in which species were recorded in the field and the cover of the species was highlighted in the studied dataset. Therefore, the first species to be observed on the field were usually the most abundant ones, as already observed (Pinto et al., 2016), and the most informative ones regarding vegetation unit assignment.

4.3 Forest habitats can be mapped with limited means and in winter

We consider that a large number of Natura 2000 site managers have limited means available, and sometimes limited botanical and phytosociological skills, for inventorying natural habitat types within their site. Our results confirm that a vegetation plot can be assigned to a vegetation unit when observing only some of the species present in the plot. Considering that mapping requires identifying and delineating habitats, our study is the first to confirm that incomplete vegetation plots can be used to create forest habitat maps and that the observation of abundant species during vegetation mapping is an efficient field approach for habitat mapping. Such a practice has already been recommended to make acceptable savings (Rodwell, 2006). Our study also confirms that little time can be spent in the field for vegetation unit assignment, hence a limited cost.

Vegetation plots are frequently restricted to the growing season because the sampling date and the species phenology might affect the creation of vegetation units in a classification (Jenackovic et al., 2019, 2016; Vymazalova et al., 2016, 2014). Yet, we showed that assigning a vegetation plot to a forest vegetation unit was feasible in winter, especially at the alliance level, even if the species considered recognisable in winter are not the most relevant ones for vegetation unit assignment.

Our results provide a general overview of forest habitats. The low number of plots we used and the sampling approach did not provide specific results for each habitat type. A further study with an *ad hoc* design will be needed to assess whether we need a different number of species and time spent in the field according to each habitat type, which could lead, for example, to less time for species inventory and fewer species inventoried in species-poor habitats.

Nonetheless, we should keep in mind that this potential optimisation of the time spent on vegetation plots and the extended period of vegetation plot feasibility only concerns field work dedicated to vegetation unit assignment. Other studies may require complete vegetation plots performed at a particular time of the year, especially for the initial description of habitats when creating a typology, and for projects such as biodiversity monitoring or environmental assessment.

4.4 Choosing the hierarchical level in the typology

The most common level for habitat of Community interest is the alliance. A slightly lower number of species and less time was needed to reach the RAR when assigning a vegetation plot to an alliance than when assigning it to an association. Therefore, creating a forest habitat map at the alliance level could take less time, and thus be cheaper than creating a forest habitat map at the association level. But even if finding the best cost-effectiveness ratio is a key issue for sampling (Mateo et al., 2018), the needs for detailed information on the vegetation units should guide the choice of the hierarchical level (Eriksen et al., 2019).

4.5 An automatic classification program can create standardised habitat assignment data from a great variety of data sources

A large number of databases and vegetation plots are available at the European level (see Chytrý et al., 2020 for multiple examples). These databases contain vegetation plots inventoried by multiple observers (see e.g. Chytrý et al., 2016; Jaskova et al., 2020). Quite inevitably, different skills and experience levels influence the quality and exhaustiveness of the vegetation plots, different plot sizes are used (Chytrý and Otýpková, 2003; Chytrý and Tichý, 2018), and this can theoretically affect the species richness associated to a vegetation plot. We presently show that vegetation unit assignment is robust to the use of vegetation plots considered as incomplete, and thereby to inter-observer variation. We also prove that an automatic classification program can create standardised habitat assignment data, even with a great variety of data sources, and then be used to attune databases merged from different organisations. This is an interesting perspective for creating habitat time series, which are crucial for the monitoring and the conservation of endangered habitats. In particular, this could help to implement the mandatory monitoring of Natura 2000 habitats across Europe, as demanded by the Habitats Directive (EC Council 2006).

ACKNOWLEDGEMENTS

We thank the providers of the NFI vegetation plots, as well as the people who collected floristic data and were involved in the assignment of the vegetation plots to the vegetation units at the NFI (Andre G., Bernard S., Bircker L., Blond W., Boithias B., Bourrinet F., Cano R., Daubigney L., Daviaud J.-F., De Taxis Du Poet T., Delayat J.-M., Delhaye S., Delquaire P., Hirsch B., Hugerot Y., Leclaire G., Ledeme G., Magnette F., Malemanche L., Michel C., Paque G., Payen D., Pedrot L., Pihou O., Rives J.-F., Vaillot J.-B.) and at the national forestry office (Blin M., Blondel F., Darnis T., Fallour-Rubio D., Gattus J.-C., Holveck P., Hum P., Keller J., Lathuillière L., Loustalot-Forest F., Ritz F., Rollier C., Témoin J.-L.). We also thank Anne Gégout-Petit for her help with statistics.

Author contributions

L.M. performed the analyses and wrote the first draft of the manuscript with contributions from J-C.G and P.E.P.; L.M., J-C.G., P.E.P. designed the study and methodology; S.W., J-C.G., P.E.P., provided the data set; J.D, S.C, S.M., P. A., B.R. contributed to the plots classification in vegetation units; J-C.G. supervised the work; all authors provided critical feedback and helped shape the research and the manuscript.

Data accessibility statement:

The data and the code used in this study are available from the first author (LM) upon request.

REFERENCES

Archaux, F., Gosselin, F., Bergès & L., Chevalier, R. (2006). Effects of sampling time, species richness and observer on the exhaustiveness of plant censuses. *Journal of Vegetation Science* 17, 299–306. https://doi.org/10.1111/j.1654-1103.2006.tb02449.x

Azovsky, A.I. (2018). Robustness of relative diversity metrics to inventory incompleteness: Could we estimate the near imponderable? *Global Ecology and Biogeography* 27, 1200–1212. https://doi.org/10.1111/geb.12777

Bardat, J., Bioret, F., Botineau, M., Boullet, V., Delpech, R., Géhu, J.-M., Haury, J., Lacoste, A., Rameau, J.-C., Royer, J.-M., Roux, G. & Touffet, J. (2004). *Prodrome des végétations de France*, Muséum national d'Histoire naturelle. ed, Publication scientifique du Muséum national d'Histoire naturelle. Paris.

Beck, J., Holloway, J.D. & Schwanghart, W. (2013). Undersampling and the measurement of beta diversity. *Methods in Ecology and Evolution* 4, 370–382. https://doi.org/10.1111/2041-210x.12023

Bioret, F., Bensettiti, F. & Royer, J.-M. (2014). Presentation of the Prodrome of French vegetation, from the first version (PVF1) to the second, declension at plant community level (PVF2). *Acta Botanica Gallica* 161, 395–402. https://doi.org/10.1080/12538078.2014.972978

Braun-Blanquet, J., Fuller, G.D. & Conard, H.S. (1932). *Plant sociology: the study of plant communities*. New York and London, McGraw-Hill book company, inc.

Braunisch, V. & Suchant, R. (2010). Predicting species distributions based on incomplete survey data: the trade-off between precision and scale. *Ecography* 33, 826–840. https://doi.org/10.1111/j.1600-0587.2009.05891.x

Bruelheide, H., Jansen, F., Jandt, U., Bernhardt-Römermann, M., Bonn, A., Bowler, D., Dengler, J., Eichenberg, D., Grescho, V., Harter, D., Jugelt, M., Kellner, S., Ludwig, M., Wesche, K. & Lütt, S. (2020). Using incomplete floristic monitoring data from habitat mapping programmes to detect species trends. *Diversity and Distributions* 26, 782-794. https://doi.org/10.1111/ddi.13058

Černá, L. & Chytrý, M. (2005). Supervised classification of plant communities with artificial neural networks. *Journal of Vegetation Science* 16, 407–414. https://doi.org/10.1111/j.1654-1103.2005.tb02380.x

Chytrý, M., Hennekens, S.M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J.H.J., Aćić, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., Campos, J.A., Carlón, L., Čarni, A., Casella, L., Csiky, J., Ćušterevska, R., Stevanović, Z.D., Danihelka, J., Bie, E.D., Ruffray, P. de, Sanctis, M.D., Dickoré, W.B., Dimopoulos, P., Dubyna, D., Dziuba, T., Ejrnæs, R., Ermakov, N., Ewald, J., Fanelli, G., Fernández-González, F., FitzPatrick, Ú., Font, X., García-Mijangos, I., Gavilán, R.G., Golub, V., Guarino, R., Haveman, R., Indreica, A., Gürsoy, D.I., Jandt, U., Janssen, J.A.M., Jiroušek, M., Kącki, Z., Kavgacı, A., Kleikamp, M., Kolomiychuk, V., Ćuk, M.K., Krstonošić, D., Kuzemko, A., Lenoir, J., Lysenko, T., Marcenò, C., Martynenko, V., Michalcová, D., Moeslund, J.E., Onyshchenko, V., Pedashenko, H., Pérez-Haase, A., Peterka, T., Prokhorov, V., Rašomavičius, V., Rodríguez-Rojo, M.P., Rodwell, J.S., Rogova, T., Ruprecht, E., Rūsiņa, S., Seidler, G., Šibík, J., Šilc, U., Škvorc, Ž., Sopotlieva, D., Stančić, Z., Svenning, J.-C., Swacha, G., Tsiripidis, I., Turtureanu, P.D., Uğurlu, E., Uogintas, D., Valachovič, M., Vashenyak, Y., Vassilev, K., Venanzoni, R., Virtanen, R., Weekes, L., Willner, W., Wohlgemuth, T. & Yamalov, S. (2016). European Vegetation Archive (EVA): an integrated database of European vegetation plots. *Applied Vegetation Science* 19, 173–180. https://doi.org/10.1111/avsc.12191

Chytrý, M. & Otýpková, Z. (2003). Plot sizes used for phytosociological sampling of European vegetation. *Journal of Vegetation Science* 14, 563–570. https://doi.org/10.1111/j.1654-1103.2003.tb02183.x

Chytrý, M. & Tichý, L. (2018). National vegetation classification of the Czech Republic: a summary of the approach. *Phytocoenologia* 48, 121–131. https://doi.org/10.1127/phyto/2017/0184

Chytrý, M., Tichý, L., Hennekens, S.M., Knollová, I., Janssen, J.A.M., Rodwell, J.S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., Hájek, M., Dengler, J., Novák, P., Zukal, D., Jiménez-Alfaro, B., Mucina, L., Abdulhak, S., Aćić, S., Agrillo, E., Attorre, F., Bergmeier, E., Biurrun, I., Boch, S., Bölöni, J., Bonari, G., Braslavskaya, T., Bruelheide, H., Campos, J.A., Čarni, A., Casella, L., Ćuk, M., Ćušterevska, R., Bie, E.D., Delbosc, P., Demina, O., Didukh, Y., Dítě, D., Dziuba, T., Ewald, J., Gavilán, R.G., Gégout, J.-C., Galdo, G.P.G. del, Golub, V., Goncharova, N., Goral, F., Graf, U., Indreica, A., Isermann, M., Jandt, U., Jansen, F., Jansen, J., Jašková, A., Jiroušek, M., Kącki, Z., Kalníková, V., Kavgacı, A., Khanina, L., Korolyuk, A.Y., Kozhevnikova, M., Kuzemko, A., Küzmič, F., Kuznetsov, O.L., Laiviņš, M., Lavrinenko, I., Lavrinenko, O., Lebedeva, M., Lososová, Z., Lysenko, T., Maciejewski, L., Mardari, C., Marinšek, A., Napreenko, M.G., Onyshchenko, V., Pérez-Haase, A., Pielech, R., Prokhorov, V., Rašomavičius, V., Rojo, M.P.R., Rūsiņa, S., Schrautzer, J., Šibík, J., Šilc, U., Škvorc, Ž., Smagin, V.A., Stančić, Z., Stanisci, A., Tikhonova, E., Tonteri, T., Uogintas, D., Valachovič, M., Vassilev, K., Vynokurov, D., Willner, W., Yamalov, S., Evans, D., Lund, M.P., Spyropoulou, R., Tryfon, E. & Schaminée, J.H.J. (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science* 23, 648–675. https://doi.org/10.1111/avsc.12519

Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. (2002). Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13, 79–90. https://doi.org/10.1111/j.1654-1103.2002.tb02025.x

De Cáceres, M., Font, X., Vicente, P. & Oliva, F. (2009). Numerical reproduction of traditional classifications and automatic vegetation identification. J*ournal of Vegetation Science* 20, 620–628. https://doi.org/10.1111/j.1654-1103.2009.01081.x

Dengler, J. (2009). Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography* 36, 728–744. https://doi.org/10.1111/j.1365-2699.2008.02038.x

Dengler, J., Löbel, S. & Dolnik, C. (2009). Species constancy depends on plot size – a problem for vegetation classification and how it can be solved. *Journal of Vegetation Science* 20, 754–766. https://doi.org/10.1111/j.1654-1103.2009.01073.x

EC Council (2010). *Directive 2009/147/CE du Parlement européen et du Conseil du 30 novembre 2009 concernant la conservation des oiseaux sauvages*, 020.

EC Council (1992). *Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora.*

Eriksen, E.L., Ullerud, H.A., Halvorsen, R., Aune, S., Bratli, H., Horvath, P., Volden, I.K., Wollan, A.K. & Bryn, A. (2019). Point of view: error estimation in field assignment of land-cover types. *Phytocoenologia* 49, 135–148. https://doi.org/10.1127/phyto/2018/0293

Evans, D. (2010). Interpreting the habitats of Annex I: past, present and future. *Acta Botanica Gallica* 157, 677–686. https://doi.org/10.1080/12538078.2010.10516241

Gégout, J.-C. & Coudun, C. (2012). The right relevé in the right vegetation unit: a new typicality index to reproduce expert judgement with an automatic classification programme. *Journal of Vegetation Science* 23, 24–32. https://doi.org/10.1111/j.1654-1103.2011.01337.x

Gégout, J.-C., Rameau, J.-C., Renaux, B., Jabiol, B. & Bar, M. (2009). *Les habitats forestiers de la France tempérée ; typologie et caractérisation phytoécologique*, AgroParisTech-ENGREF. ed. Nancy.

Hearn, S.M., Healey, J.R., McDonald, M.A., Turner, A.J., Wong, J.L.G. & Stewart, G.B. (2011). The repeatability of vegetation classification and mapping. *Journal of Environmental Management* 92, 1174–1184. https://doi.org/10.1016/j.jenvman.2010.11.021

Hervé, J.-C. (2016). France, in: *National Forest Inventories - Assessment of Wood Availability.* Springer International Publishing, pp. 385–404.

Hopkins, B. (1957). The Concept of Minimal Area. *Journal of Ecology* 45, 441–449. https://doi.org/10.2307/2256927

Jaskova, A., Braslayskaya, T.Y., Tikhonova, E., Paal, J., Rusina, S., Laivins, M., Kucherov, I.B., Genikova, N., Knollova, I., Chernenkova, T., Churakova, E.Y., Diekmann, M., Halvorsen, R., Kirichok, E., Korotkov, V.N., Kryshen, A.M., Lugovaya, D.L., Morozova, O., Potapov, P., Prokazina, T.S., Schei, F.H., Semenishchenkov, Y.A., Shevchenko, N.E., Sidorova, O., Smirnov, N.S., Smirnova, O., Tsvirko, R., Turubanova & S.A., Chytry, M. (2020). European Boreal Forest Vegetation Database. *Phytocoenologia* 50, 79–92. https://doi.org/10.1127/phyto/2019/0336

Jenackovic, D.D., Lakusic, D., Zlatkovic, I., Juskovic, M. & Randelovic, N.V. (2019). Emergent wetland vegetation data recording: Does an optimal period exist? *Applied Vegetation Science* 22, 200–212. https://doi.org/10.1111/avsc.12419

Jenackovic, D.D., Zlatkovic, I.D., Lakusic, D.V. & Randelovic, V.N. (2016). The assessment of seasonal variability in emergent macrophyte communities. *Biologia* 71, 287–297. https://doi.org/10.1515/biolog-2016-0033

Maciejewski, L., Pinto, P.E., Wurpillot, S., Drapier, J., Cadet, S., Muller, S., Agou, P., Renaux, B. & Gégout, J.-C. (2020). Vegetation unit assignments: phytosociology experts and classification programs show similar performance but low convergence. *Applied Vegetation Science* 23, 698–709. https://doi.org/10.1111/avsc.12516

Mateo, R.G., Gaston, A., Jose Aroca-Fernandez, M., Saura, S. & Ignacio Garcia-Vinas, J. (2018). Optimization of forest sampling strategies for woody plant species distribution modelling at the landscape scale. *Forest Ecology and Management* 410, 104–113. https://doi.org/10.1016/j.foreco.2017.12.046

Morrison, L.W. (2016). Observer error in vegetation surveys: a review. *Journal of Plant Ecology* 9, 367–379. https://doi.org/10.1093/jpe/rtv077

Pinto, P.E., Dupouey, J.-L., Hervé, J.-C., Legay, M., Wurpillot, S., Montpied, P. & Gégout, J.-C. (2016). Optimizing the bioindication of forest soil acidity, nitrogen and mineral nutrition using plant species. *Ecological Indicators* 71, 359–367. https://doi.org/10.1016/j.ecolind.2016.05.047

Preston, F.W. (1962). The Canonical Distribution of Commonness and Rarity: Part I. *Ecology* 43, 185–215. https://doi.org/10.2307/1931976

Rodwell, J.S. (2006). *National Vegetation Classification: Users' Handbook, Joint Nature Conservation Committee*, ed. Elsevier, Peterborough.

Stevens, J.P., Blackstock, T.H., Howe, E.A. & Stevens, D.P. (2004). Repeatability of Phase I habitat survey. *Journal of Environmental Management* 73, 53–59. https://doi.org/10.1016/j.jenvman.2004.05.009

Ullerud, H.A., Bryn, A., Halvorsen, R. & Hemsing, L.Ø. (2018). Consistency in land-cover mapping: Influence of field workers, spatial scale and classification system. *Applied Vegetation Science* 21, 278–288. https://doi.org/10.1111/avsc.12368

van Tongeren, O., Gremmen, N. & Hennekens, S. (2008). Assignment of releves to pre-defined classes by supervised clustering of plant communities using a new composite index. *Journal of Vegetation Science* 19, 525–536. https://doi.org/10.3170/2008-8-18402

Vittoz, P., Bayfield, N., Brooker, R., Elston, D.A., Duff, E.I., Theurillat, J.-P. & Guisan, A. (2010). Reproducibility of species lists, visual cover estimates and frequency methods for recording high-mountain vegetation. *Journal of Vegetation Science* 21, 1035–1047. https://doi.org/10.1111/j.1654-1103.2010.01216.x

Vittoz, P. & Guisan, A. (2007). How reliable is the monitoring of permanent vegetation plots? A test with multiple observers. *Journal of Vegetation Science* 18, 413–422. https://doi.org/1http://10.0.4.87/j.1654-1103.2007.tb02553.x

Vymazalova, M., Tichy, L. & Axmanova, I. (2016). The role of vernal species in vegetation classification: a case study on deciduous forests and dry grasslands of Central Europe. *Phytocoenologia* 46, 9–20. https://doi.org/10.1127/phyto/2016/0034

Vymazalova, M., Tichy, L. & Axmanova, I. (2014). How does vegetation sampling in different parts of the growing season influence classification results and analyses of beta diversity? *Applied Vegetation Science* 17, 556–566. https://doi.org/10.1111/avsc.12087

Williamson, M., Gaston, K.J. & Lonsdale, W.M. (2001). The species–area relationship does not have an asymptote! *Journal of Biogeography* 28, 827–830. https://doi.org/10.1046/j.1365-2699.2001.00603.x

FIGURES WITH LEGENDS

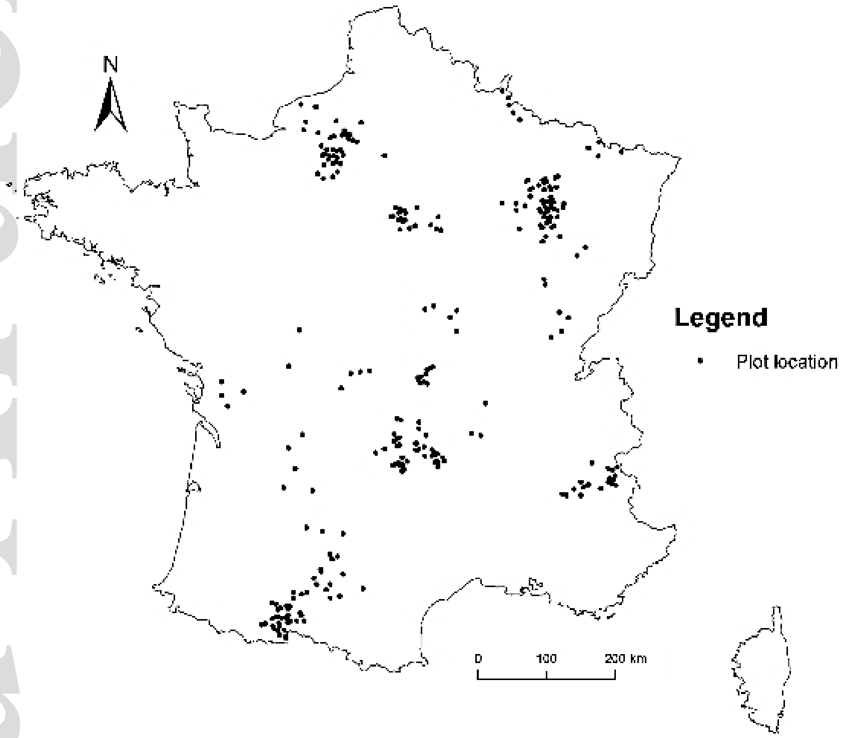

Figure 1: Map of the 273 vegetation plots surveyed between May and September 2013 by six teams of the French National Forest Inventory (NFI) across the forests of mainland France.

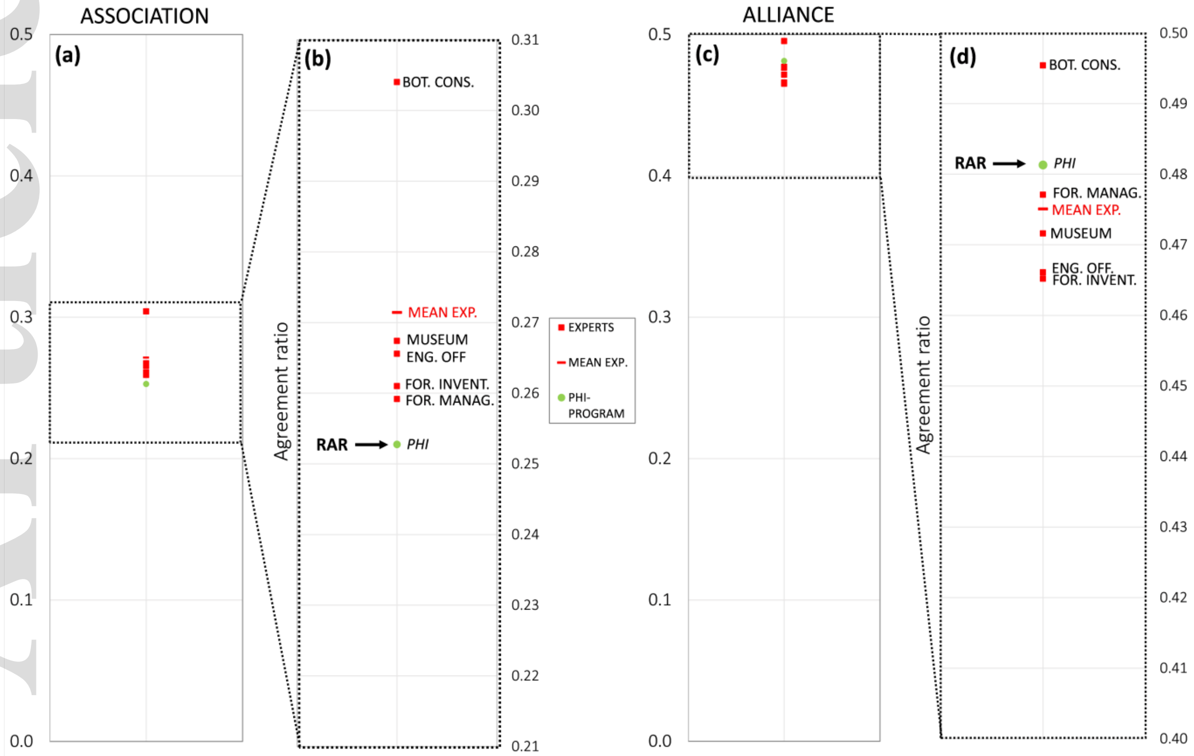

Figure 2: Agreement ratios, based on complete vegetation plots, calculated among expert organisations, and between the Phi-program and the expert organisations.

Experts: the National Botanical Conservatory of the Massif Central (Bot. Cons.), the National Forest Inventory of France (For. Invent.), the botanists' network of the National Forestry Office of France (For. Manag.), Biotope, an ecological consultancy (Eng. Off.), and the National Museum of Natural History of France (Museum).

Mean Exp.: average value of the five agreement ratios of the expert organisations.

PHI: automatic classification program called the Phi-program.

RAR: agreement Ratio of Reference, calculated based on complete vegetation plots between the Phi-program and the five the expert organisations.

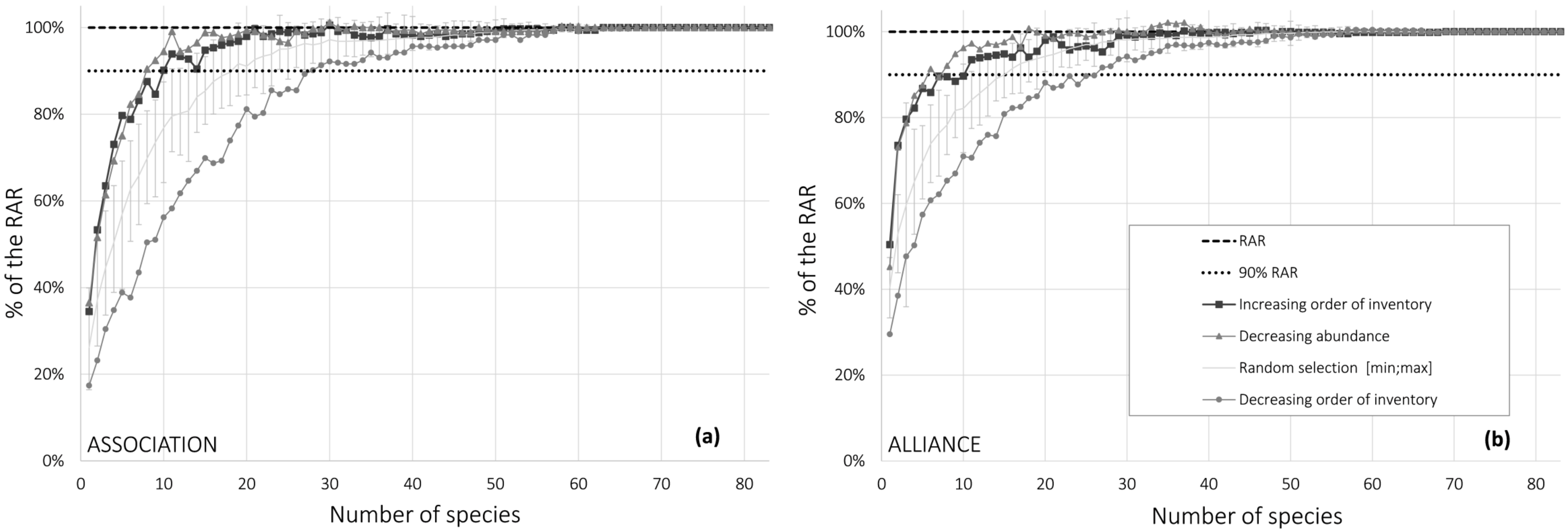

Figure 3: Percentage of the Agreement Ratio of Reference (RAR) reached by the agreement ratios of the Phi-program based on incomplete vegetation plots according to the number of species in the plot, regarding association (a) and alliance assignment (b). The number of species was selected using: (i) random selection (100 repetitions), (ii) increasing order of inventory, (iii) decreasing order of inventory, and (iv) decreasing abundance of the species in the vegetation plots.

RAR is the agreement ratio, calculated based on the complete vegetation plots, between the automatic classification program ('Phi-program') and the expert organisations (0.25 and 0.48 on 1, for association and alliances, respectively, see Fig. 2).

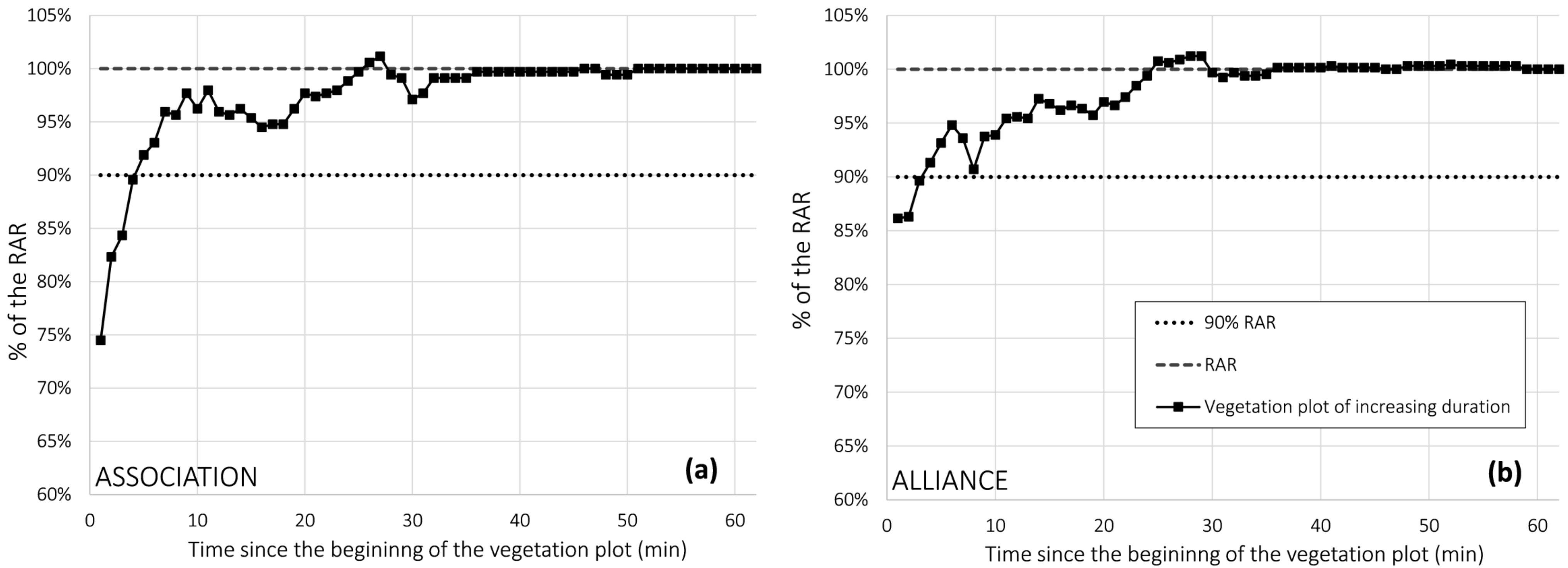

Figure 4: Percentage of the Agreement Ratio of Reference (RAR) reached by the agreement ratios of the program based on incomplete vegetation plots of increasing duration according to duration since the beginning of the vegetation plot, regarding association (a) and alliance assignment (b). RAR is the agreement ratio, based on the complete vegetation plots, between the automatic classification program ('Phi-program') and the expert organisations (0.25 and 0.48 on 1, for association and alliances, respectively, see Fig. 2).

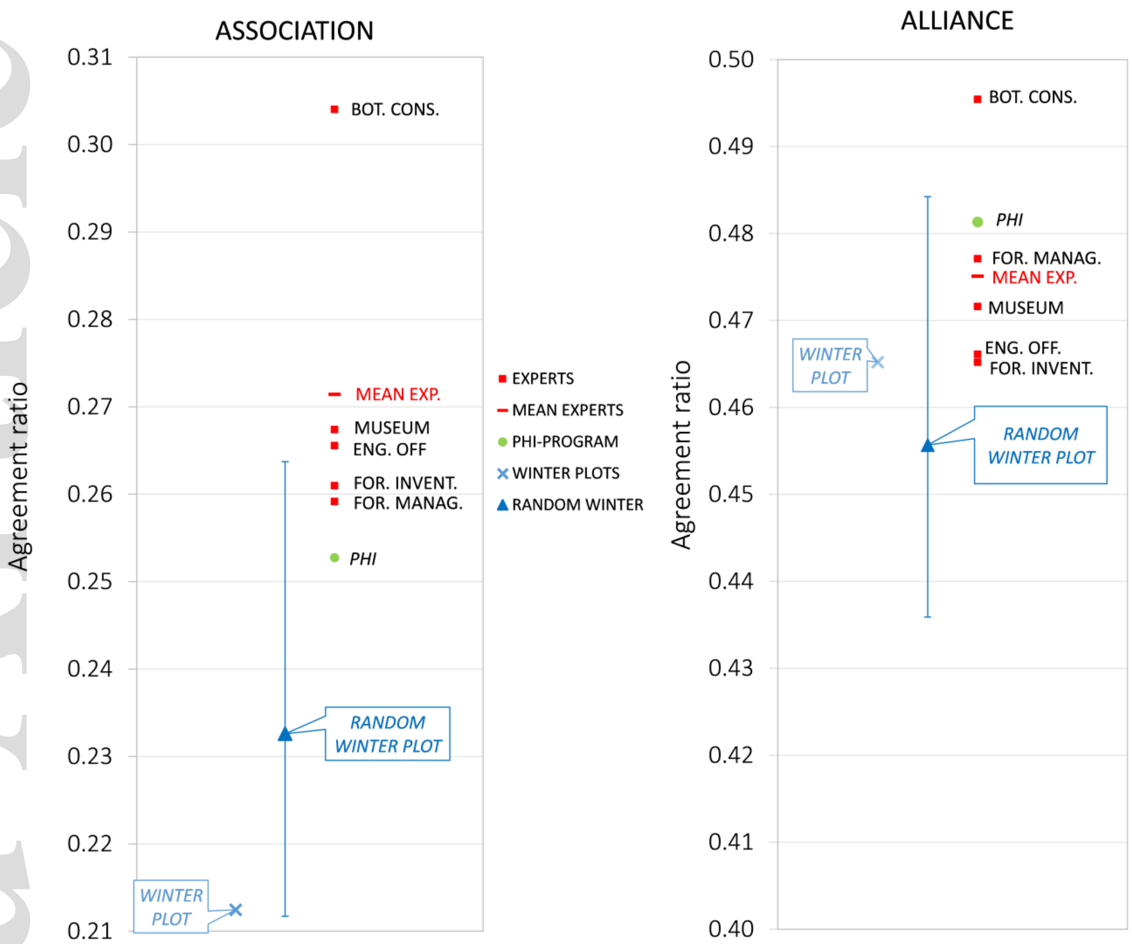

Figure 5: Comparison between agreement ratios calculated with winter plots and with complete vegetation plots, for association assignment (a) and alliance assignment (b).

Experts: defined in Fig.2

Mean Exp.: average value of the five agreement ratios of the expert organisations.

Phi: agreement ratio provided by the automatic classification program (Phi-program) using complete vegetation plots.

Winter plots: agreement ratio provided by the Phi-program using incomplete vegetation plots including species potentially recognisable in winter ('winter plots').

Random winter plots: agreement ratio provided by the Phi-program using incomplete vegetation plots created by randomly choosing species among the complete vegetation plots to equalise the species richness of the winter plots.