Check for updates

Received: 6 July 2022 | Revised: 9 January 2023 | Accepted: 11 January 2023

DOI: 10.1111/jvs.13175

Journal of Vegetation Science IAVS

RESEARCH ARTICLE

Probabilistic and preferential sampling approaches offer integrated perspectives of Italian forest diversity

Nicola Alessi[1,2,3,4] | Gianmaria Bonari[5] | Piero Zannini[1,2,3] | Borja Jiménez-Alfaro[6] | Emiliano Agrillo[4] | Fabio Attorre[7] | Roberto Canullo[8] | Laura Casella[4] | Marco Cervellini[1,8] | Stefano Chelli[8] | Michele Di Musciano[1,9] | Riccardo Guarino[10] | Stefano Martellos[3,11] | Marco Massimi[7] | Roberto Venanzoni[12] | Stefan Zerbe[5] | Alessandro Chiarucci[1,3]

[1]BIOME Lab, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

[2]LifeWatch, Lecce, Italy

[3]Plant Data Interuniversity Research Centre for Plant Biodiversity and Big Data, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

[4]Italian Institute for Environmental Protection and Research, Rome, Italy

[5]Faculty of Science and Technology, Free University of Bozen-Bolzano, Bolzano, Italy

[6]Research Unit of Biodiversity (CSUC/UO/PA), University of Oviedo, Mieres, Spain

[7]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy

[8]School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

[9]Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy

[10]Department of Biological, Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Palermo, Italy

[11]Department di Life Sciences, University of Trieste, Trieste, Italy

[12]Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy

Correspondence
Piero Zannini, BIOME Lab, Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy.
Email: piero.zannini2@unibo.it

Funding information
CONplant (TN201H); LifeWatchPLUS (CIR-01_00028)

Co-ordinating Editor: Zoltán Botta-Dukát

Abstract

Aim: Assessing the performances of different sampling approaches for documenting community diversity may help to identify optimal sampling efforts and strategies, and to enhance conservation and monitoring planning. Here, we used two data sets based on probabilistic and preferential sampling schemes of Italian forest vegetation to analyze the multifaceted performances of the two approaches across three major forest types at a large scale.

Location: Italy.

Methods: We pooled 804 probabilistic and 16,259 preferential forest plots as samples of vascular plant diversity across the country. We balanced the two data sets in terms of sizes, plot size, geographical position, and vegetation types. For each of the two data sets, 1000 subsets of 201 random plots were compared by calculating the shared and exclusive indicator species, their overlap in the multivariate space, and the

Nicola Alessi and Gianmaria Bonari share first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Journal of Vegetation Science* published by John Wiley & Sons Ltd on behalf of International Association for Vegetation Science.

areas encompassed by spatially-constrained rarefaction curves. We then calculated an index of performance using the ratio between the additional and total information collected by each sampling approach. The performances were tested and evaluated across the three major forest types.

Results: The probabilistic approach performed better in estimating species richness and diversity of species assemblages, but did not detect other components of the regional diversity, such as azonal forests. The preferential approach outperformed the probabilistic approach in detecting forest-specialist species and plant diversity hotspots.

Conclusions: Using a novel workflow based on vegetation-plot exclusivities and commonalities, our study suggests probabilistic and preferential sampling approaches are to be used in combination for better conservation and monitor planning purposes to detect multiple aspects of plant community diversity. Our findings can assist the implementation of national conservation planning and large-scale monitoring of biodiversity.

KEYWORDS

biodiversity, co-occurrence data, detrended correspondence analysis, indicator species analysis, regional survey, spatially constrained rarefaction curve, temperate forests, vegetation database, zonal vegetation

1 | INTRODUCTION

Human-induced environmental changes are affecting the distribution, structure, and functioning of ecosystems, resulting in a global biodiversity crisis with evident impact on our society (Cardinale et al., 2012; Pecl et al., 2017). While international conservation programs strengthen protection goals (COM, 2020), the scientific community provides data, measurements, and evaluations of the biodiversity crisis from local to global scales (IPBES, 2019). Large collections of field observations at different spatio-temporal scales have thus become an essential tool to monitor, model, and predict the impact of global changes on natural ecosystems (Schmeller et al., 2015; Staude et al., 2020). Monitoring agencies need cost-effective sampling approaches to accomplish national conservation strategies and programs. In this context, the performance of a sampling approach can be evaluated based on the efficiency in collecting information. Useful approaches should detect multiple aspects of the spatio-temporal patterns of biodiversity (Chiarucci et al., 2011; Mihoub et al., 2017; Schmidt-Traub, 2021).

Currently, the geographical extent of plant diversity databases ranges from the regional to the global scale, including different types of diversity observations, such as species records and co-occurrence data (Chytrý et al., 2016; Sabatini, Lenoir, et al., 2021). One advantage of this latter type of data is to allow accurate estimations of local diversity due to the recording of complete — or almost complete — species lists within sampling units, e.g. a vegetation plot (Franklin et al., 2017). Co-occurrence data can also be transformed to single-species records, while species assemblages derived from aggregates of for example herbarium specimens, could lead to spurious results (Bottin et al., 2020). Herbarium specimens have shown a bias towards rare but colorful and charismatic species (Troudet et al., 2017; Adamo et al., 2021) when compared with aggregates of vegetation plot databases (Bottin et al., 2020). Estimates of beta diversity across large areas obtained by assembled plot data are similar to those obtained by species lists (Chiarucci et al., 2021), suggesting that the standardization of large vegetation plot databases allows sufficient representation of vegetation conditions, modeling and predicting biodiversity patterns at different spatio-temporal scales (Staude et al., 2020; Laughlin et al., 2021; Testolin et al., 2021).

Notwithstanding the amount of aggregated historical data, biodiversity monitoring requires continuous and expensive sampling efforts to detect changes in species diversity. The long tradition of vegetation surveys in Europe has allowed the implementation of different sampling approaches across the continent. Traditionally, preferential (opportunistic) sampling has been widely employed in Europe. This approach collects vegetation plots at environmentally homogeneous sites selected on the basis of expert selection and using variable numbers and grain sizes of plots to characterize plant communities (Braun-Blanquet, 1964). Despite some limitations in the use of preferentially collected data for inferential purposes (Chiarucci, 2007; Roleček et al., 2007), this approach is suitable for the assessment of total species richness of a given study area, as well as to detect rare vegetation types characterized by habitat specialist or alien species (Michalcová et al., 2011; Speak et al., 2018). Other studies have suggested the advantages

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

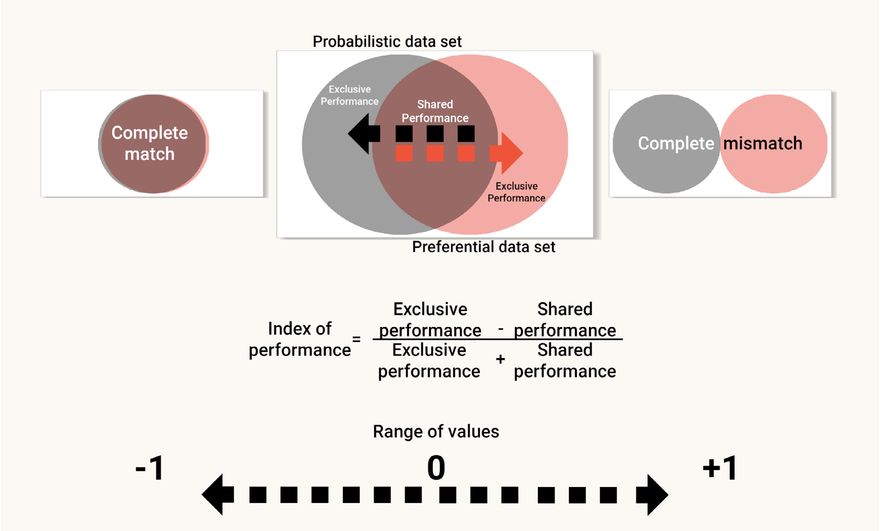

FIGURE 1 Graphical conceptualization of the methodology used to measure the performance of the two sampling approaches. The shared and exclusive biodiversity information calculated as percentages and emerging from the data sets highlights similarities, differences, and the overall performance of the two sampling approaches. The index of performance evaluates the additional information with respect to the common information collected by each sampling approach, weighted by their sum.

of probabilistic approaches, in which plots are placed according to a survey design to produce robust inferences on the abundance and distribution of species and vegetation types (Michalcová et al., 2011; Swacha et al., 2017).

The combination of probabilistic and preferential sampling approaches may detect different facets of plant community diversity, revealing both common and rare species distributions and abundances (Roleček et al., 2007). However, despite the urgent need for improving sampling schemes for plant diversity monitoring, the two approaches have been compared only at the landscape scale (Michalcová et al., 2011; Swacha et al., 2017; Speak et al., 2018), thus neglecting environmental and biogeographical factors which drive plant community diversity patterns. Since probabilistic data sets at the regional scale are difficult to retrieve, extensive diversity data sets are usually obtained by aggregating local data sets based on different sampling schemes. In turn, aggregated data sets suffer from biases in data distribution with respect to the most frequent vegetation types in a defined geographic area (Roleček et al., 2007). To efficiently monitor plant diversity and improve surveys at broad spatial scales, standardized measurements of performances of data sets could shed light on how to efficiently integrate both probabilistic and preferential data.

In this study, we evaluated the performance of probabilistic and preferential sampling approaches for estimating different facets of forest diversity at the country scale. Using a novel workflow based on vegetation plot exclusivities and commonalities, we compared the two approaches in terms of representing: (i) habitat specialist composition, (ii) diversity of species assemblages, and (iii) species diversity estimates. We thus evaluated the performance of the two sampling approaches based on the additional information with respect to the shared information collected by each sampling approach, weighted by their sum. By combining vegetation data sets from across Italy, we aim to discuss the importance of collecting and combining spatial observations to develop biodiversity monitoring programs for national conservation planning (Hochkirch et al., 2021; Schmidt-Traub, 2021).

2 | METHODS

2.1 | Study area

Italian forests cover 90,851 km^2 (Gasparini et al., 2022). The high variation of the study area in latitude (from 35° to 47°), elevation (from 0 m to 4809 m a.s.l., with forest vegetation up to ~1700 to 1900 m a.s.l.), geomorphological heterogeneity (Fredi & Palmieri Lupia, 2017), and climatic conditions (from subtropical to cold-temperate climate; Fratianni & Acquaotta, 2017) is mirrored by a high diversity of forest types (Chiarucci et al., 2019; Agrillo et al., 2021). The main vegetation forest types in the study area are broad-leaved evergreen and deciduous forests of warm-temperate climate, broad-leaved deciduous forests of cool-temperate climate, and needle-leaved forests of cold-temperate climate (Dinerstein et al., 2017). According to the Italian forest inventory (INFC, 2015), "forests with high trees" cover 89,567 km^2, of which 17% is represented by high-elevation coniferous forests, 67% by broad-leaved deciduous forests, 13% by Mediterranean evergreen forests, and 3% by riparian forests.

2.2 | General workflow

We compared a single vegetation plot data set (hereafter, "probabilistic data set") with a larger aggregate of vegetation plot data sets (hereafter, "preferential data set"), both representing vascular plant diversity of Italian forests across the whole country (see the next paragraph). The vegetation plots of the probabilistic data set were collected according to a formal and reproducible scheme, while the vegetation plots of the preferential data set were obtained by aggregating preferentially collected data. We developed a novel workflow based on the partitioning of the performance into shared and exclusive information emerging from each data set (Figure 1). While the shared information was defined as the portion collected by both data sets, the exclusive information differentiated a data set

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with respect to the others. The sum of these two components corresponded to the overall, or joint, information of the probabilistic and preferential approaches — that is, the total size or area of information retained. A heuristic index measuring the performance of each sampling approach was then calculated as the ratio of the difference between exclusive and shared information divided by their sum (Figure 1). Since we measured shared and exclusive information as proportions, the index ranges between −1 and 1. The index was calculated on randomly re-sampled and balanced subsets of the probabilistic and preferential data sets (Figure 2), as described in the following subsections. We evaluated multiple aspects of plant diversity retained by each data set using three different ecological analyses (Figure 2). We then applied this approach to three zonal forest types for evaluating widely distributed and rare plant communities (Figure 3).

2.3 | Probabilistic data set

The probabilistic data set of Italian forest vegetation consisted of plots collected in the framework of the BIOSOIL project (Hiederer & Durrant, 2010). It was obtained by extracting a probabilistic sample of plant communities based on a 16 km × 16 km grid superimposed on the whole Italian country (Level I network; Lorenz et al., 2002; Forests ICP, 2016; Chiarucci et al., 2019). Grid corners were selected if a forest patch larger than 0.01 km^2 occurred therein. Hence, a statistically representative sample of 261 sites was considered for field observations of forest ecosystems (Petriccione & Cindolo, 2006). Overall, 60 sites were excluded by ground surveys because they were not forests, inaccessible, or extremely disturbed (e.g. recent tree harvesting, cattle rest areas, ski slopes). This resulted in a final sample of 201 circular sampling sites (radius 25.24 m; sampled area 2000 m^2), in which four 10 m × 10 m plots were located at random distances from the center, along the main cardinal directions. In each plot, plant species identities and their relative cover were recorded (Canullo et al., 2013; Ferretti et al., 2013). Coordinates of sampling site centers were extracted from GPS devices with a positional uncertainty below 10 m. The field campaign was carried out in spring–summer 2007, employing 10 teams of two surveyors each, after a common training and calibration exercise following Quality Assurance guidelines (Allegrini et al., 2009; Canullo et al., 2016). Taxa identified at the genus level were excluded from the data set. Thus, a final data set containing plant cover values for 1,099 species observed in 804 plots distributed over 201 sites was obtained. Taxonomy was standardized according to the Italian flora (Pignatti et al., 2017–2019).

2.4 | Preferential data set

The preferential data set of Italian forest vegetation consisted of plot observations aggregated from four databases (see Appendix S1). The data set underwent a filtering process on an initial set of 51,529 plots. We selected plots with: (i) an estimated positional uncertainty

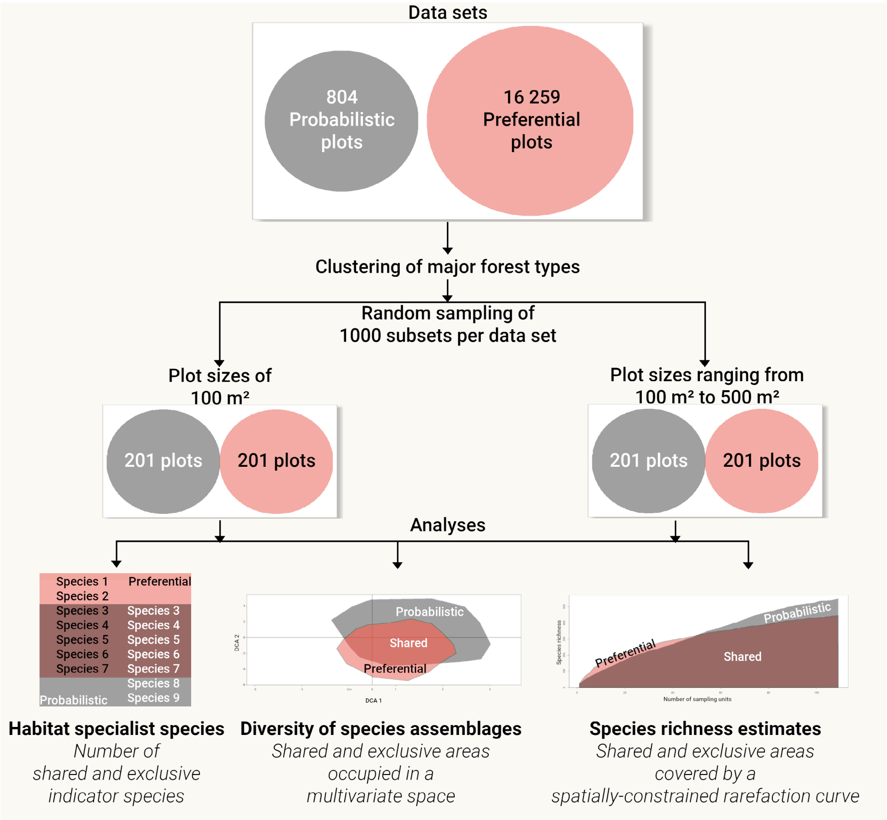
Data sets
804 Probabilistic plots
16 259 Preferential plots
Clustering of major forest types
Random sampling of 1000 subsets per data set
Plot sizes of 100 m²
Plot sizes ranging from 100 m² to 500 m²
201 plots
201 plots
201 plots
201 plots
Analyses
Species 1
Species 2
Species 3
Species 4
Species 5
Species 6
Species 7
Preferential
Species 3
Species 4
Species 5
Species 6
Species 7
Species 8
Species 9
Probabilistic
Probabilistic
Shared
Preferential
DCA 2
DCA 1
Probabilistic
Preferential
Shared
Species richness
Number of sampling units
Habitat specialist species
Number of shared and exclusive indicator species
Diversity of species assemblages
Shared and exclusive areas occupied in a multivariate space
Species richness estimates
Shared and exclusive areas covered by a spatially-constrained rarefaction curve

FIGURE 2 Graphical conceptualization of the workflow adopted to balance and analyze the probabilistic and preferential data sets. To balance data set sizes, we randomly re-sampled plots considering sizes, plot sizes, geographic distribution and vegetation types. To consider different aspects of plant community diversity, we applied the conceptual model to three ecological analyses: Indicator Species Analysis, Detrended Correspondence Analysis, and spatially-constrained rarefaction curves.

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

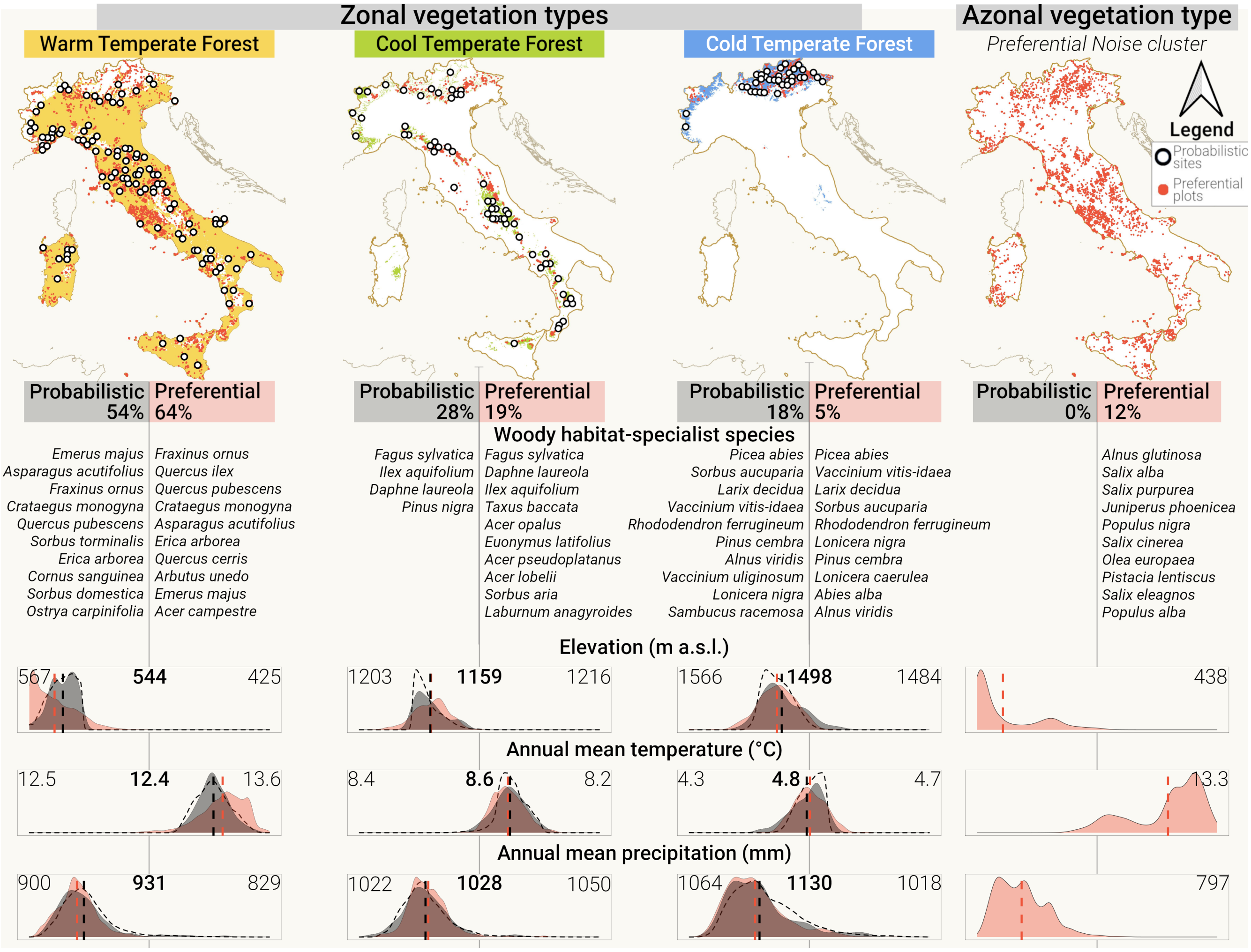

FIGURE 3 Graphical summary of the differences between probabilistic and preferential data sets in characterizing Italian forest types. We obtained forest vegetation types using a multivariate regression tree on the probabilistic data set and assignment of the preferential plots with a noise clustering technique. The azonal vegetation type occurs only in the preferential data set. We show the geographical distribution and the frequency of clustered co-occurrence data sets (804 probabilistic and 16,259 preferential plots), the first 10 significant woody indicator species ($p < 0.01$) ranked by their association values (phi coefficient), the density plots and mean values of environmental variables for the two data sets. The dashed density plot and the bold character in the mean values represent the Italian forests.

below 1000 m; (ii) cumulative tree species cover above 30%; and (iii) all taxa identified at the species level. Duplicated plots were removed. Taxonomy was standardized according to the Italian flora (Pignatti et al., 2017–2019). The final data set included 16,259 plots containing plant cover values for 2,948 species, including plots of different sizes (17% of the plots had no information about plot size). The preferential data set resulted in a total of 946 geographically distinct locations identified in a 16 km × 16 km cell grid.

2.5 | Environmental variables

To characterize each plot in terms of environmental variables, we extracted elevation (European Union, 2021) and all the 19 bioclimatic variables of the Chelsa data set at 1 ArcSec resolution (~1000 m at the equator; Karger et al., 2017).

2.6 | Forest vegetation types

For the sole purpose of defining major vegetation types in Italian forests, we used the probabilistic data set at the site level (aggregation of four 100-m^2 plots) by square-root-transforming species cover values and performing a multivariate regression tree with environmental variables (De'ath & Fabricius, 2002). This technique identifies the most probable vegetation type given a certain climate, by concurrently accounting for species co-occurrences and environmental variables (Borcard et al., 2011). Clustering the probabilistic data set at the site level instead of the plot level (201 sites × 4 plots) allowed the assignment of the four plots forming sites to a unique vegetation type. Because of the heterogeneous sources of the preferential data set, we preferred to use the probabilistic data set for the definition of vegetation types due to its statistical representativeness of the distribution of forest vegetation

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the study area. After checking for multicollinearity among environmental variables with Variation Inflation Factor analysis (VIF, Zuur et al., 2010) using a threshold of 10, we used eight predictors in the model — i.e. elevation, minimum temperatures of the driest and wettest quarter, temperature seasonality (the standard deviation of the monthly temperatures), isothermality (the ratio of diurnal variation to annual variation in temperatures), precipitation seasonality (the standard deviation of the monthly precipitation estimates expressed as a percentage of the mean of those estimates), and precipitations of the wettest month and of the coldest quarter (Karger et al., 2017). We characterized and named the three obtained clusters using their geographical and environmental distributions along with their list of indicator species resulting from the Indicator Species Analysis (De Cáceres et al., 2010). The three clusters were considered as mean ecological prototypes representative of Italian zonal forest types: (i) the warm-temperate forest, dominated by evergreen and deciduous broad-leaved trees (109 sites, 54%); (ii) the cool-temperate forest, dominated by deciduous broad-leaved trees (56 sites, 28%); and (iii) the cold-temperate forest, dominated by needle-leaved trees (36 sites, 18%). Then, chord-transformed preferential plots were assigned to zonal forest types based on their chord distance from prototype centroids expressed as species composition in a multivariate space, namely by using noise clustering (De Cáceres et al., 2017). By setting a threshold distance for plot assignment, we excluded outlier plots from the probabilistic prototypes — meaning 12% of the data set, composed mainly of azonal forest stands such as riparian types and coastal areas. Most of the preferential plots were assigned to a zonal vegetation type (64% warm-temperate; 19% cool-temperate; 5% cold-temperate). A detailed description of the methodology is reported in Appendix S2, while details on the characterization of forest types are graphically summarized in Figure 3. The environmental characterization of the two data sets was compared with the whole distribution of Italian forests provided by Copernicus Land Monitoring Service products upscaled at 1000-m spatial resolution (Figure 3; European Union, 2021). The area occupied by each forest type was obtained as a prediction of the multivariate regression tree model performed on environmental variables. A complete list of indicator species sorted by life forms and ordered by fidelity values to forest vegetation types is provided in Appendix S3. Specifically, we calculated indicator values for preferential plots: (i) regardless of plot size; (ii) with a plot size of 100 m^2 except for cold-temperate forest (plots with sizes ranging between 100 and 300 m^2); and (iii) with plot sizes ranging between 100 and 500 m^2.

2.7 | Performance measurement

To compare the performance of probabilistic and preferential data sets, considering their discrepancy in plot numbers, we sampled 1000 different subsets of 201 plots for each data set. Plots were randomly sampled at the site level. While sites for the probabilistic data set were defined in the sampling design as random forested corners of a 16 km × 16 km grid, for the preferential data set we simulated a similar re-sampling design selecting random plots located within cells of the 16 km × 16 km grid. Thus, the probabilistic subsets were aggregated selecting a random plot for each of the 201 sites. In the preferential plot subsets, we maintained proportions among vegetation types observed with the probabilistic data set — meaning 109 plots for the warm-temperate forest, 56 plots for the cool-temperate forest, and 36 plots for the cold-temperate forest. To standardize plot sizes between the two data sets, we selected only preferential 100-m^2 plots, except for cold-temperate forest for which we selected plots with sizes ranging between 100 and 300 m^2. To evaluate the effect of plot size on the performance measurements, we repeated the analyses also using 1000 preferential subsets of plots with sizes ranging between 100 and 500 m^2.

For each of the 1000 subsets of the two data sets, the overall information on habitat specialist species was quantified by summing the relative number of shared and exclusive indicator species for each data set (De Cáceres et al., 2010). We used the “multipatt” function of the *indicspecies* R package (De Cáceres et al., 2020) with 999 permutations and counted those species with a significant phi coefficient (p value <0.01). This analysis allowed us to compare clusters of unequal sizes (Tichý & Chytrý, 2006).

Overall information on species assemblage diversity was calculated by summing the shared and exclusive occupied areas of each data set over the two first axes of Detrended Correspondence Analysis (DCA). We used the “decorana” function of the *vegan* R package (Oksanen et al., 2020). DCA axes were used to estimate species turnover and summarize patterns of variation among plant assemblages (Eilertsen et al., 1990).

Overall information on species diversity estimates among plots was calculated by summing the relative shared and exclusive areas encompassed by spatially-constrained rarefaction curves for each data set (Chiarucci et al., 2009). Spatially-constrained rarefaction allowed accounting for the spatial arrangement of plots to calculate rarefaction curves by moving toward geographically close plots (Chiarucci et al., 2009). We thus used the “rare_alpha” function of the *Rarefy*R package (Thouverai et al., 2021). For the estimation of the exclusive information of each data set, we excluded the relative shared part of information.

The index of performance for each data set was then calculated based on the obtained percentage measurements of the exclusive and shared information in the three ecological analyses. Positive values of the index correspond to a good performance of the sampling approach which collects more exclusive than shared information. On the other hand, negative values mean that the shared information collected by the approach is larger than the exclusive information. To compare the performances of the probabilistic and preferential approaches, the index was tested for significance using the non-parametric Mann–Whitney test.

All the analyses were performed using QGIS version 3.16 (QGIS Development Team, 2020) and R version 4.1.2 (R Core Team, 2022).

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | RESULTS

The probabilistic and preferential data sets showed a different species composition and environmental characterization of the three zonal forest types, especially in the warm forest type (Figure 3, Appendix S3). The probabilistic data set showed an environmental characterization closer to the whole forest distribution provided by Copernicus Land Monitoring Service (European Union, 2021) in comparison with the preferential data set. The preferential data set showed more plots at low elevation in the warm-temperate and cold-temperate forests. The preferential data set also showed a high number of woody habitat specialist species in the cool-temperate forest type.

Regarding the three ecological analyses we performed, namely the Indicator Species Analysis, the DCA, and the spatially-constrained rarefaction curves, we found significantly different performances for the two data sets across the zonal forest types. We found concordant results for the heterogeneous and larger plot-size aggregated data set (Appendix S4).

The probabilistic data set showed the lowest values for the overall and exclusive information in the habitat specialist species (Figure 4a). The preferential data set outperformed the probabilistic data set for the whole data set, and for the warm- and cold-temperate forests types (Figure 5a). For the cool-temperate forest, the two data sets performed similarly instead.

The probabilistic data set showed the highest values for the overall and exclusive information on the diversity of species assemblages (Figure 4b). The probabilistic data set outperformed the preferential data set for both the cool- and cold-temperate forests, but it showed negative performance for the whole data set and warm-temperate forest (Figure 5b).

The probabilistic data set showed the lowest values for the overall and exclusive information on species richness estimates in the whole data set and warm-temperate forests, but the highest values in the cool- and cold-temperate forests (Figure 4c). We found lower performances for the probabilistic data set in the whole data set and in the warm-temperate forest (Figure 5c). To the contrary, the probabilistic data set outperformed the preferential data set in the cool- and cold-temperate forests.

4 | DISCUSSION

We provide a comparison of multifaceted performance between probabilistic and preferential sampling approaches in evaluating plant community diversity at a large scale. We confirmed our findings considering both similar and larger plot sizes for preferential plots with respect to probabilistic plots, suggesting plot size as a weak factor driving the analyzed patterns. The performance was assessed among three zonal forest types obtained by numerical clustering on the probabilistic data set and subsequent assignment of the forest types to preferential plots based on the species composition. This clustering approach allowed grouping of the most frequent forest communities, which resulted to be the zonal forest types. To the contrary, basing the clustering on a large and heterogeneous aggregate of plots, such as the preferential data set, might emphasize uncommon vegetation types which would make our comparison unstable. In general, the probabilistic approach failed in detecting the regional (gamma) diversity, by neglecting azonal forests — meaning riparian and coastal forest types. The data sets analyzed here were differentiated by habitat-specialist species and environmental distribution, with the preferential data set having more plots at warmer sites. The preferential approach also outperformed the probabilistic approach in detecting assemblages rich in habitat specialists. To the contrary, the probabilistic data set showed the higher performance in detecting diversity of species assemblages and spatially assembled regional species richness estimates. Notwithstanding this, in the given zonal forest types, the two sampling approaches deviate from this general finding.

4.1 | The probabilistic approach

Given an equal sampling effort, a systematic approach applied to forest areas represented by the probabilistic data set performed better in detecting richness of species and diversity of species assemblages at the national scale compared to a heterogeneous aggregate of preferential vegetation plot data sets. According to Botta-Dukát et al. (2007), this result may be due to the sampling of degraded forest stands in which species typical of species-rich open habitats tend to occur. Degraded sites are often avoided during preferential vegetation sampling surveys because of their mixed species composition which may hardly be assigned to targeted habitat types (e.g. Chytrý et al., 2020). The exception to this finding is represented by the warm-temperate forests in which the two approaches performed similarly. This is probably due to a bias of the preferential data set toward sites with warm-temperate forest types occurring at lower elevations with higher temperatures, in which evergreen forest stands occur. This bias is confirmed by the occurrence of evergreen species as woody habitat specialist species in the preferential data set — for example *Arbutus unedo* or *Quercus ilex*. These forests have traditionally attracted the interest of botanists and vegetation ecologists because of their relatively easy accessibility combined with typical species composition. In turn, this particular attention could have enriched the preferential data set in terms of species diversity. The evergreen warm-temperate forests may be difficult to detect by random or systematic sampling if we consider their limited distribution in the study area with respect to the deciduous warm-temperate forests (Agrillo et al., 2021). The low performance of the probabilistic data set in detecting habitat specialist species is in line with the results of Swacha et al. (2017) and suggests the difficulty of including undisturbed and characteristic forest patches in the data collection because of their limited geographical distribution in the study area. Thus, the data collection in transitional zones rich in non-specialist species constrained by the probabilistic sampling design could have increased the observed diversity of species and

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

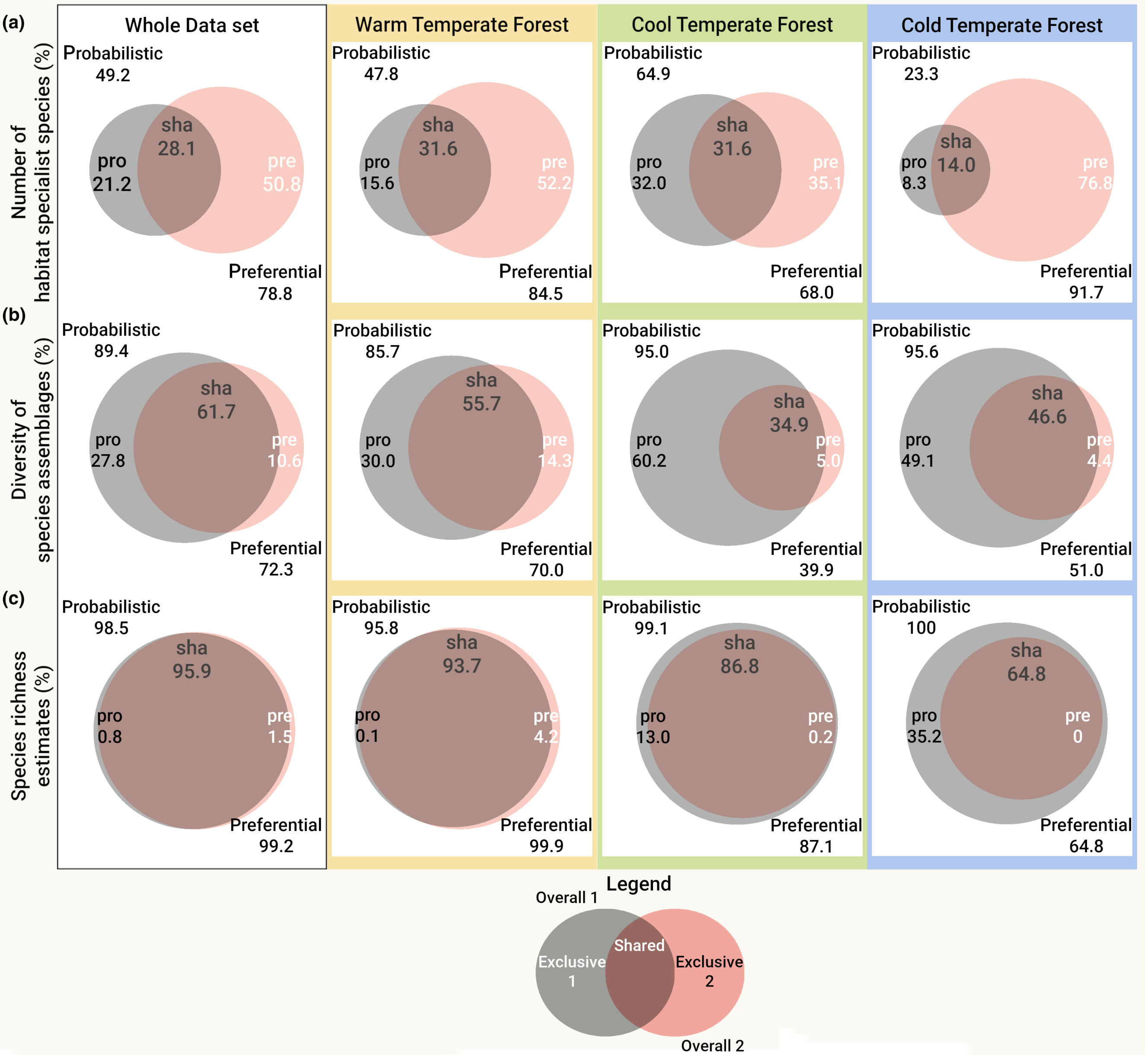

FIGURE 4 Overall, shared and exclusive information emerging from the probabilistic and preferential data sets for: (a) the habitat specialist species; (b) diversity of species assemblages; and (c) species richness estimates. Proportional Venn diagrams with median values of their components: the exclusive components ("pro" and "pre"), the shared component ("sha") and the overall information ("Probabilistic" and "Preferential"). The components were calculated for the indicator species, the areas occupied by the data sets in a Detrended Correspondence Analysis bi-plot, and the areas encompassed by spatially-constrained rarefaction curves of the two data sets.

species assemblages. Moreover, the constrained sampling adopted in the probabilistic approach prevents the detection of scattered patches and linear elements of, for example, remnant riparian forests, which frequently occur in peculiar landscape configurations detectable by fine topographic variables only (Douda, 2010).

4.2 | The preferential approach

The preferential approach performed better with forest stands rich in specialist species with a higher number of relatively rare species, e.g. *Acer lobelii*, *Taxus baccata* or *Abies alba* in the cool- and cold-temperate forests, respectively. Chytrý (2001) suggested preferential data sets of species-poor vegetation may be biased toward higher species richness because of the surveyors' tendency to increase the plot size to include indicator species. We found low performance of the preferential sampling for detecting species richness and diversity of species assemblages in the typically species-poorer cool- and cold-temperate forests, which is in line with Botta-Dukát et al.'s (2007) findings. The focus of the preferential approach with respect to undisturbed forest stands amplifies the sampling of rare specialist species but reduces the sampling of ecotonal species connected to forest dynamics, thus decreasing the

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

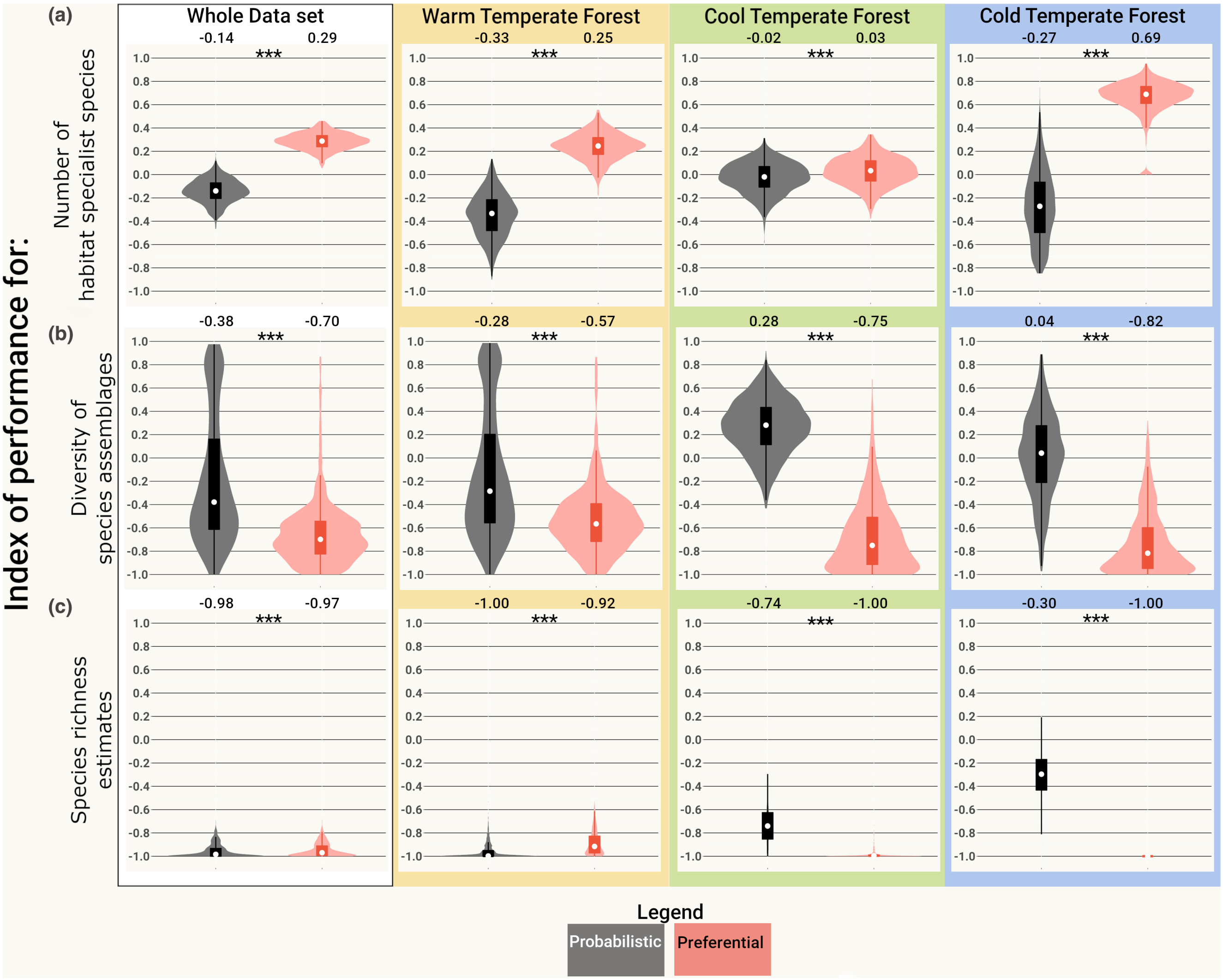

FIGURE 5 Index of performance for the probabilistic and preferential data sets calculated for three types of diversity analyses. The index is calculated as the ratio between the shared performance subtracted from the exclusive performance and their sum. Performances are calculated as percentages and the derived index ranges between −1 and 1. We calculated the index for the whole data sets and three vegetation types. Median values and significance of *p* values obtained from the Mann–Whitney test are shown (***, $p < 0.001$).

sampled species richness. By contrast, this approach allows sampling similar or higher species richness estimates and assemblage diversity than the probabilistic approach in complex vegetation types such as the warm-temperate forest. Because of ecological, biogeographical, evolutionary, and historical factors, the warm-temperate forests have a particularly rich species pool (Box, 2015; Rundel et al., 2016; Večeřa et al., 2019), especially in annual species (Večeřa et al., 2021). Interestingly, this high diversity is also reflected in a high number of vegetation types (Preislerová et al., 2022). Moreover, when considering patterns of forest area gains and losses, in combination with levels of protection and population density, low-elevation forest stands of the Italian country have been affected by a high degree of human impact during the last century (Zannini et al., 2022). This could result in more undisturbed conditions characterized by rare forest specialists at remote and high-elevation sites and a *continuum* between rural landscapes and secondary forest stands at low-elevation sites. Thus, the exclusive species richness combined with a complex forest landscape affected by millennia of anthropogenic impacts (Sadori et al., 2011) exacerbates the diversity of species assemblages occurring in limited areas (Agrillo et al., 2021). This complexity may be better detected with the support of a preferential approach because of the localization of sampling units positively conditioned by the knowledge of expert botanists.

4.3 | Combining probabilistic and preferential sampling

The complementary perspective of the probabilistic and the preferential approaches for detecting multiple facets of plant community diversity suggests the high potential of combining both sampling approaches. The large availability of preferential plots can be used for explorative analyses and for obtaining descriptive statistics, whereas the probabilistic data set is essential for hypothesis testing (Botta-Dukát et al., 2007). While the probabilistic data set can unbiasedly

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

represent the ecological status of forests at the country scale by evaluating the most frequent species richness or composition (Roleček et al., 2007), the preferential data set focuses on stands rich in undisturbed and specialist species as well as peculiar forest types, attaining additional information that is crucial for estimating regional diversity. Accordingly, the mean composition of Italian forests assessed with the probabilistic data set resulted in lower occurrences of habitat specialist species with respect to the species composition provided by an aggregate of preferential data sets. However, the indicative value of the preferential approach highlights its potential to detect conservation-relevant vegetation types (Chytrý et al., 2020). The scattered and rare distribution of undisturbed forest remnants rich in specialist species (e.g. old-growth forest stands; Barredo et al., 2021) could be better detected through targeted sampling. This holds true also for hotspots of plant diversity driven by topographic or biogeographic factors — for example, the azonal and Mediterranean evergreen forests (Naiman & Décamps, 1997; Rundel et al., 2016). A combination of both approaches in field surveys is thus recommended as it allows efficient and comprehensive evaluation of ecosystem diversity and status.

4.4 | Spatial and temporal baselines for conservation planning

We have presented here a workflow to test performances and detect unbalanced data distributions in large vegetation plot databases. The combination of existing large vegetation plot data sets has been shown to be a reliable reference system to extrapolate spatial and temporal baselines for biodiversity conservation planning (Franklin et al., 2017; Chytrý et al., 2020). Our study underlines the importance of studying diversity patterns while considering a well-designed integration of different sampling schemes to provide a description of multiple facets of plant community diversity. These sampling schemes should use standards to consistently apply statistical assumptions but also evaluate local and regional species diversity considering heterogeneous landscapes as a result of biogeographical and land-use history (Canullo et al., 2013; Speak et al., 2018). Using significant environmental strata to select sampling sites, probabilistic approaches will result in diversity measurements with known uncertainty values. On the other hand, the flexibility of preferential approaches identifies species-rich areas supporting the development and implementation of conservation planning and targeted actions — for example identifying habitat types (Chytrý et al., 2020), vegetation types (Bonari et al., 2021), old-growth forests (Sabatini, Bluhm, et al., 2021), refugia (Jiménez-Alfaro et al., 2018; Alessi et al., 2019), and riparian forest remnants (Douda et al., 2016). Geographical and biogeographical gaps in diversity monitoring data could be filled in a step-by-step procedure based on two (or more) sampling approaches which include: (i) tracing the spatial and temporal diversity baselines using existing large electronic archives; (ii) evaluating data deficiencies and performances; and (iii) planning efficient monitoring surveys based on historical data. In this procedure, adaptative sampling strategies may be effective for monitoring highly diverse and rare species or habitats (Fattorini et al., 2022). In adaptive sampling strategies an additional sampling effort is allocated to areas where the ecological *phenomenon* was observed in the earlier sampling surveys (Pacifici et al., 2016). Aggregated archives and preferential surveys could play an important role as baseline for monitoring regional diversity. This workflow should generate standardized diversity work and data flows between the scientific community and environmental agencies to implement conservation planning at the national scale (Mihoub et al., 2017; Hillebrand et al., 2018; Schmidt-Traub, 2021). Our combined approach thus encompasses a comprehensive integrated view that will eventually result in an optimized tool for assessing plant diversity in natural and semi-natural ecosystems.

AUTHOR CONTRIBUTIONS

Nicola Alessi, Gianmaria Bonari, Piero Zannini, Borja Jiménez-Alfaro, and Alessandro Chiarucci conceived the idea. Nicola Alessi, Gianmaria Bonari, and Piero Zannini designed the methodology. Nicola Alessi performed the analyses. Nicola Alessi, Gianmaria Bonari, and Piero Zannini wrote the paper, with major inputs from Borja Jiménez-Alfaro and Alessandro Chiarucci. All authors contributed to the acquisition of data, critically commented on the draft, and gave their final approval.

ACKNOWLEDGEMENTS

We thank the ICP Forests' Italian Focal Centre (CUFA, Comando per la Tutela della Biodiversità e dei Parchi – Ufficio Studi e Progetti, Roma) and the Forestry Department of the Autonomous Province of Bolzano-South Tyrol for providing the probabilistic data set. We thank Silvia Guglielmi for useful suggestions on graphics and figures. Open Access Funding provided by Università degli Studi di Bologna within the CRUI-CARE agreement. Research project implemented under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union - Next Generation EU. Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP J33C22001190001, Project title "National Biodiversity Future Center - NBFC". Open Access Funding provided by Università degli Studi di Bologna within the CRUI-CARE Agreement.

FUNDING INFORMATION

Nicola Alessi and Piero Zannini were supported by *LifeWatch Italy* through the project LifeWatchPLUS (CIR-01_00028). Gianmaria Bonari and Stefan Zerbe were supported by the Free University of Bozen-Bolzano through the CONplant project (TN201H).

DATA AVAILABILITY STATEMENT

Data sets were obtained from the VPD-Sapienza University of Rome (http://www.givd.info/ID/EU-IT-021; Agrillo et al., 2017), AMS-VegBank (http://www.givd.info/ID/EU-IT-021; Alessi et al., 2022), Vegetation database of Habitats in the Italian Alps – HabItAlp (http://www.givd.info/ID/EU-IT-010), and CircumMed Forest database

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(http://www.givd.info/ID/EU-00-026; Bonari et al., 2019). The list of the 16,259 preferential plots selected for the analyses is reported in Appendix S4.

ORCID

Nicola Alessi https://orcid.org/0000-0002-4479-950X
Gianmaria Bonari https://orcid.org/0000-0002-5574-6067
Piero Zannini https://orcid.org/0000-0003-2466-4402
Borja Jiménez-Alfaro https://orcid.org/0000-0001-6601-9597
Emiliano Agrillo https://orcid.org/0000-0003-2346-8346
Fabio Attorre https://orcid.org/0000-0002-7744-2195
Roberto Canullo https://orcid.org/0000-0002-9913-6981
Laura Casella https://orcid.org/0000-0003-2550-3010
Marco Cervellini https://orcid.org/0000-0002-0853-2330
Stefano Chelli https://orcid.org/0000-0001-7184-8242
Michele Di Musciano https://orcid.org/0000-0002-3130-7270
Riccardo Guarino https://orcid.org/0000-0003-0106-9416
Stefano Martellos https://orcid.org/0000-0001-5201-8948
Marco Massimi https://orcid.org/0000-0003-2137-8160
Roberto Venanzoni https://orcid.org/0000-0002-7768-0468
Stefan Zerbe https://orcid.org/0000-0002-9426-1441
Alessandro Chiarucci https://orcid.org/0000-0003-1160-235X

REFERENCES

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K. & Mammola, S. (2021) Plant scientists' research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nature Plants*, 7(5), 574–578. Available from: https://doi.org/10.1038/s41477-021-00912-2

Agrillo, E., Alessi, N., Massimi, M., Spada, F., De Sanctis, M., Francesconi, F. et al. (2017) Nationwide vegetation-plots database – Sapienza University of Rome: state of the art, basic figures and future perspectives. *Phytocoenologia*, 47, 221–229. Available from: https://doi.org/10.1127/phyto/2017/0139

Agrillo, E., Filipponi, F., Pezzarossa, A., Casella, L., Smiraglia, D., Orasi, A. et al. (2021) Earth observation and biodiversity big data for forest habitat types classification and mapping. *Remote Sensing*, 13(7), 1231. Available from: https://doi.org/10.3390/rs13071231

Alessi, N., Bruzzaniti, V., Buldrini, F., Centomo, E., Cervellini, M., Enea, M. et al. (2022) AMS-VegBank: a new database of vegetation plots for the Italian territory. *Vegetation Classification and Survey*, 3, 177–185. Available from: https://doi.org/10.3897/VCS.85083

Alessi, N., Těšitel, J., Zerbe, S., Spada, F., Agrillo, E. & Wellstein, C. (2019) Ancient refugia and present-day habitat suitability of native laurophylls in Italy. *Journal of Vegetation Science*, 30(3), 564–574. Available from: https://doi.org/10.1111/jvs.12743

Allegrini, M.C., Canullo, R. & Campetella, G. (2009) ICP-Forests (international co-operative programme on assessment and monitoring of air pollution effects on forests): quality assurance procedure in plant diversity monitoring. *Journal of Environmental Monitoring*, 11(4), 782–787. Available from: https://doi.org/10.1039/b818170p

Barredo, J.I., Brailescu, C., Teller, A., Sabatini, F.M., Mauri, A. & Janouskova, K. (2021) *Mapping and assessment of primary and old-growth forests in Europe*. Luxemburg: Publications Office of the European Union. Available from: https://doi.org/10.2760/13239

Bonari, G., Fernández-González, F., Çoban, S., Monteiro-Henriques, T., Bergmeier, E., Didukh, Y.P. et al. (2021) Classification of the Mediterranean lowland to submontane pine forest vegetation. *Applied Vegetation Science*, 24(1), e12544. Available from: https://doi.org/10.1111/avsc.12544

Bonari, G., Knollová, I., Vlčková, P., Chytrý, M., Xystrakis, F., Çoban, S. et al. (2019) CircumMed pine Forest database: an electronic archive for Mediterranean and Submediterranean pine forest vegetation data. *Phytocoenologia*, 49(3), 311–318. Available from: https://doi.org/10.1127/phyto/2019/0311

Borcard, D., Gillet, F. & Legendre, P. (2011) *Numerical ecology with R*. New York: Springer.

Botta-Dukát, Z., Kovács-Láng, E., Rédei, T., Kertész, M. & Garadnai, J. (2007) Statistical and biological consequences of preferential sampling in phytosociology: Theoretical considerations and a case study. *Folia Geobotanica*, 42, 141–152. Available from: https://doi.org/10.1007/BF02893880

Bottin, M., Peyre, G., Vargas, C., Raz, L., Richardson, J.E. & Sanchez, A. (2020) Phytosociological data and herbarium collections show congruent large-scale patterns but differ in their local descriptions of community composition. *Journal of Vegetation Science*, 31(1), 208–219. Available from: https://doi.org/10.1111/jvs.12825

Box, E.O. (2015) *Warm-temperate deciduous forests around the northern hemisphere*. London New York: Springer.

Braun-Blanquet, J. (1964) Pflanzensoziologie. In: *Grundzüge der Vegetationskunde*, 3rd edition. Wien: Springer.

Canullo, R., Starlinger, F. & Giordani, F. (2013) Diversity and composition of plant and lichen species. In: Ferretti, M. & Fischer, R. (Eds.) *Forest monitoring: methods for terrestrial investigations in Europe with an overview of North America and Asia. Developments in environmental science*, Vol. 12. Oxford: Elsevier, pp. 237–250. Available from: https://doi.org/10.1016/B978-0-08-098222-9.00013-3

Canullo, R., Starlinger, F., Granke, O., Fischer, R. & Aamlid, D. (2016) Assessment of ground vegetation. In: UNECE ICP Forests Programme Co-ordinating Centre (Ed.) *Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forest*. Hamburg: Thünen Institute of Forest Ecosystems, pp. 1–18.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P. et al. (2012) Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59–67. Available from: https://doi.org/10.1038/nature11148

Chiarucci, A. (2007) To sample or not to sample? That is the question… For the vegetation scientist. *Folia Geobotanica*, 42(2), 209–216. Available from: https://doi.org/10.1007/BF02893887

Chiarucci, A., Bacaro, G., Rocchini, D., Ricotta, C., Palmer, M. & Scheiner, S. (2009) Spatially constrained rarefaction: incorporating the autocorrelated structure of biological communities into sample-based rarefaction. *Community Ecology*, 10(2), 209–214. Available from: https://doi.org/10.1556/ComEc.10.2009.2.11

Chiarucci, A., Bacaro, G. & Scheiner, S.M. (2011) Old and new challenges in using species diversity for assessing biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1576), 2426–2437. Available from: https://doi.org/10.1098/rstb.2011.0065

Chiarucci, A., Buldrini, F., Cervellini, M., Guarino, R., Caccianiga, M., Foggi, B. et al. (2021) Habitat type and Island identity as drivers of community assembly in an archipelago. *Journal of Vegetation Science*, 32(1), 1–14. Available from: https://doi.org/10.1111/jvs.12953

Chiarucci, A., Nascimbene, J., Campetella, G., Chelli, S., Dainese, M., Giorgini, D. et al. (2019) Exploring patterns of beta diversity to test the consistency of biogeographical boundaries: A case study across forest plant communities of Italy. *Ecology and Evolution*, 9(20), 11716–11723. Available from: https://doi.org/10.1002/ece3.5669

Chytrý, M. (2001) Phytosociological data give biased estimates of species richness. *Journal of Vegetation Science*, 12, 439–444. Available from: https://doi.org/10.1111/j.1654-1103.2001.tb00190.x

Chytrý, M., Hennekens, S.M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F. et al. (2016) European Vegetation Archive (EVA):

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

an integrated database of European vegetation plots. *Applied Vegetation Science*, 19(1), 173–180. Available from: https://doi.org/10.1111/avsc.12191

Chytrý, M., Tichý, L., Hennekens, S.M., Knollová, I., Janssen, J.A.M., Rodwell, J.S. et al. (2020) EUNIS Habitat Classification: expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, 23(4), 648–675. Available from: https://doi.org/10.1111/avsc.12519

COM(2020) Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. EU biodiversity strategy for 2030 - bringing nature back into our lives. No. 380 final. Brussels: Europe.

De Cáceres, M., Font, X. & Oliva, F. (2017) The management of vegetation classifications with fuzzy clustering. *Journal of Vegetation Science*, 21(6), 1138–1151. Available from: https://doi.org/10.1111/j.1654-1103.2010.01211.x

De Cáceres, M., Jansen, F. & Dell, N. (2020) "Indicspecies": relationship between species and group of sites. R Package Version 1.7.9. https://cran.r-project.org/web/packages/indicspecies/indicspecies.pdf

De Cáceres, M., Legendre, P. & Moretti, M. (2010) Improving indicator species analysis by combining groups of sites. *Oikos*, 119(10), 1674–1684. Available from: https://doi.org/10.1111/j.1600-0706.2010.18334.x

De'ath, G. & Fabricius, K.E. (2002) Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology*, 83(4), 1105–1117. Available from: https://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E. et al. (2017) An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience*, 67(6), 534–545. Available from: https://doi.org/10.1093/biosci/bix014

Douda, J. (2010) The role of landscape configuration in plant composition of floodplain forests across different physiographic areas. *Journal of Vegetation Science*, 21(6), 1110–1124. Available from: https://doi.org/10.1111/j.1654-1103.2010.01213.x

Douda, J., Boublík, K., Slezák, M., Biurrun, I., Nociar, J., Havrdová, A. et al. (2016) Vegetation classification and biogeography of European floodplain forests and alder carrs. *Applied Vegetation Science*, 19(1), 147–163. Available from: https://doi.org/10.1111/avsc.12201

Eilertsen, O., Okland, H.R., Okland, T. & Pederson, O. (1990) Data manipulation and gradient length estimation in DCA ordination. *Journal of Vegetation Science*, 1(2), 261–270. Available from: https://doi.org/10.2307/3235663

European Union. (2021) *Copernicus land monitoring service*. European Environment Agency (EEA).

Fattorini, L., Cervellini, M., Franceschi, S., Di Musciano, M., Zannini, P. & Chiarucci, A. (2022) A sampling strategy for assessing habitat coverage at a broad spatial scale. *Ecological Indicators*, 143, 109352. Available from: https://doi.org/10.1016/j.ecolind.2022.109352

Ferretti, M., Beuker, E., Calatayud, V., Canullo, R., Dobbertin, M., Eichhorn, J. et al. (2013) Data quality in field surveys. Methods and results for tree condition, phenology, growth, plant diversity and foliar injury due to ozone. *Developments in Environmental Science*, 12, 397–414. Available from: https://doi.org/10.1016/B978-0-08-098222-9.00021-2

Forests ICP(2016) MANUALS. Retrieved from http://icp-forests.net/page/icp-forests-manual

Franklin, J., Serra-Diaz, J.M., Syphard, A.D. & Regan, H.M. (2017) Big data for forecasting the impacts of global change on plant communities. *Global Ecology and Biogeography*, 26(1), 6–17. Available from: https://doi.org/10.1111/geb.12501

Fratianni, S. & Acquaotta, F. (2017) The climate of Italy. In: Soldati, M. & Marchetti, M. (Eds.) *Landscapes and landforms of Italy*. World Geomorphological Landscapes. Cham: Springer, pp. 29–38. Available from: https://doi.org/10.1007/978-3-319-26194-2_4

Fredi, P. & Palmieri Lupia, E. (2017) Morphological regions of Italy. In: Soldati, M. & Marchetti, M. (Eds.) *Landscapes and landforms of Italy*. World Geomorphological Landscapes. Cham: Springer, pp. 39–74. Available from: https://doi.org/10.1007/978-3-319-26194-2_5

Gasparini, P., Di Cosmo, L., Floris, A. & De Laurentis, D. (Eds.). (2022) *Italian National Forest Inventory—Methods and Results of the Third Survey: Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio—Metodi e Risultati della Terza Indagine*. Cham: Springer Nature.

Hiederer, R. & Durrant, T. (2010) *Evaluation of BioSoil demonstration project – Preliminary data analysis*. Luxembourg: Office for Official Publications of the European Communities, 126.

Hillebrand, H., Blasius, B., Borer, E.T., Chase, J.M., Downing, J.A., Eriksson, B.K. et al. (2018) Biodiversity change is uncoupled from species richness trends: consequences for conservation and monitoring. *Journal of Applied Ecology*, 55(1), 169–184. Available from: https://doi.org/10.1111/1365-2664.12959

Hochkirch, A., Samways, M.J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P. et al. (2021) A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, 35(2), 502–509. Available from: https://doi.org/10.1111/cobi.13589

INFC (2015) *Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio*. Arma dei Carabinieri – Comando Unità Forestali Ambientali e Agroalimentari & CREA – Centro di ricerca Foreste e Legno. Avilable fromwww.inventarioforestale.org/it/node/50 [Accessed 31 January 2021].

IPBES (2019) In: Brondízio, H.T., Settele, E.S., Díaz, J. & Ngo, S. (Eds.) *Global assessment report of the intergovernmental science-policy platform on biodiversity and ecosystem services*. Bonn, Germany: IPBES secretariat, 1148. Available from: https://doi.org/10.5281/zenodo.3831673

Jiménez-Alfaro, B., Girardello, M., Chytrý, M., Svenning, J.-C., Willner, W., Gégout, J.C. et al. (2018) History and environment shape species pools and community diversity in European beech forests. *Nature Ecology and Evolution*, 2, 483–490. Available from: https://doi.org/10.1038/s41559-017-0462-6

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W. et al. (2017) Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 1–20. Available from: https://doi.org/10.1038/sdata.2017.122

Laughlin, D.C., Mommer, L., Sabatini, F.M., Bruelheide, H., Kuyper, T.W., Mccormack, M.L. et al. (2021) Root traits explain plant species distributions along climatic gradients yet challenge the nature of ecological trade-offs. *Nature Ecology and Evolution*, 5, 1123–1134. Available from: https://doi.org/10.1038/s41559-021-01471-7

Lorenz, M., Mues, V., Becher, G., Seidling, W., Fischer, R., Langouche, D. et al. (2002) *Forest condition in Europe. Results of the 2001 Large-scale Survey*. Brussels, Geneva: Scheme UE and ICP Forests (UNECE-EC).

Michalcová, D., Lvončík, S., Chytrý, M. & Hájek, O. (2011) Bias in vegetation databases? A comparison of stratified-random and preferential sampling. *Journal of Vegetation Science*, 22(2), 281–291. Available from: https://doi.org/10.1111/j.1654-1103.2010.01249.x

Mihoub, J.B., Henle, K., Titeux, N., Brotons, L., Brummitt, N.A. & Schmeller, D.S. (2017) Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Scientific Reports*, 7(41591), 1–11. Available from: https://doi.org/10.1038/srep41591

Naiman, R.J. & Décamps, H. (1997) The ecology of interfaces: Riparian zones. *Annual Review of Ecology and Systematics*, 28(102), 621–658. Available from: https://doi.org/10.1146/annurev.ecolsys.28.1.621

Oksanen, A.J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D. et al. (2020) 'Vegan': Community ecology package. R Package Version 2.5-7.

Pacifici, K., Reich, B.J., Dorazio, R.M. & Conroy, M.J. (2016) Occupancy estimation for rare species using a spatially-adaptive sampling

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

design. *Methods in Ecology and Evolution*, 7, 285–293. Available from: https://doi.org/10.1111/2041-210X.12499

Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.C. et al. (2017) Biodiversity redistribution under climate change: impacts on ecosystems and human well-being. *Science*, 355(6332), 1333–1338. Available from: https://doi.org/10.1126/science.aai9214

Petriccione, B. & Cindolo, C. (2006) *Progetto BioSoil – biodiversity. Valutazione della biodiversità forestale sulla rete sistematica di livello I. Manuale Nazionale, Italia*. Roma: Corpo Forestale Dello Stato.

Pignatti, S., Guarino, R. & La Rosa, M. (2017–2019) *Flora d'Italia*, Vol. 1–4, Ed. 2 edition. Milano: Edagricole di New Business Media.

Preislerová, Z., Jiménez-Alfaro, B., Mucina, L., Berg, C., Bonari, G., Kuzemko, A. et al. (2022) Distribution maps of vegetation alliances in Europe. *Applied Vegetation Science*, 25(1), e12642. Available from: https://doi.org/10.1111/avsc.12642

QGIS Development Team(2020) QGIS geographic information system. Open Source Geospatial Fundation Project. http://qgis.osgeo.org

R Core Team(2022) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Roleček, J., Chytrý, M., Hájek, M., Lvončík, S. & Tichý, L. (2007) Sampling design in large-scale vegetation studies: Do not sacrifice ecological thinking to statistical purism! *Folia Geobotanica*, 42(2), 199–208. Available from: https://doi.org/10.1007/BF02893886

Rundel, P.W., Arroyo, M.T.K., Cowling, R.M., Keeley, J.E., Lamont, B.B. & Vargas, P. (2016) Mediterranean biomes: Evolution of their vegetation, floras, and climate. *Annual Review of Ecology, Evolution, and Systematics*, 47(1), 383–407. Available from: https://doi.org/10.1146/annurev-ecolsys-121415-032330

Sabatini, F.M., Bluhm, H., Kun, Z., Aksenov, D., Atauri, J.A., Buchwald, E. et al. (2021) European primary forest database. *Scientific Data*, 8(220), 1–14. Available from: https://doi.org/10.1038/s41597-021-00988-7

Sabatini, F.M., Lenoir, J., Hattab, T., Arnst, E.A., Chytrý, M., Dengler, J. et al. (2021) sPlotOpen – An environmentally balanced, open-access, global dataset of vegetation plots. *Global Ecology and Biogeography*, 30(9), 1740–1764. Available from: https://doi.org/10.1111/geb.13346

Sadori, L., Jahns, S. & Peyron, O. (2011) Mid-Holocene vegetation history of the Central Mediterranean. *Holocene*, 21(1), 117–129. Available from: https://doi.org/10.1177/0959683610377530

Schmeller, D.S., Julliard, R., Bellingham, P.J., Böhm, M., Brummitt, N., Chiarucci, A. et al. (2015) Towards a global terrestrial species monitoring program. *Journal for Nature Conservation*, 25, 51–57. Available from: https://doi.org/10.1016/j.jnc.2015.03.003

Schmidt-Traub, G. (2021) National climate and biodiversity strategies are hamstrung by a lack of maps. *Nature Ecology and Evolution*, 5, 1325–1327. Available from: https://doi.org/10.1038/s41559-021-01533-w

Speak, A., Escobedo, F.J., Russo, A. & Zerbe, S. (2018) Comparing convenience and probability sampling for urban ecology applications. *Journal of Applied Ecology*, 55(5), 2332–2342. Available from: https://doi.org/10.1111/1365-2664.13167

Staude, I.R., Waller, D.M., Bernhardt-Römermann, M., Bjorkman, A.D., Brunet, J., De Frenne, P. et al. (2020) Replacements of small- by large-ranged species scale up to diversity loss in Europe's temperate forest biome. *Nature Ecology and Evolution*, 4(6), 802–808. Available from: https://doi.org/10.1038/s41559-020-1176-8

Swacha, G., Botta-Dukát, Z., Kącki, Z., Pruchniewicz, D. & Żołnierz, L. (2017) A performance comparison of sampling methods in the assessment of species composition patterns and environment-vegetation relationships in species-rich grasslands. *Acta Societatis Botanicorum Poloniae*, 86(4), 1–15. Available from: https://doi.org/10.5586/asbp.3561

Testolin, R., Attorre, F., Borchardt, P., Brand, R.F., Bruelheide, H., Chytrý, M. et al. (2021) Global patterns and drivers of alpine plant species richness. *Global Ecology and Biogeography*, 30(6), 1218–1231. Available from: https://doi.org/10.1111/geb.13297

Thouverai, E., Pavoine, S., Tordoni, E., Rocchini, D., Ricotta, C., Chiarucci, A. et al. (2021) "Rarefy": Rarefaction methods. R Package Version 1.1. https://doi.org/10.1016/j.ecocom.2012.05.007

Tichý, L. & Chytrý, M. (2006) Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science*, 17(6), 809–818. Available from: https://doi.org/10.1111/j.1654-1103.2006.tb02504.x

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017) Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7(1), 1–14. Available from: https://doi.org/10.1038/s41598-017-09084-6

Večeřa, M., Axmanová, I., Padullés, C.J., Lososová, Z., Divíšek, J., Knollová, I. et al. (2021) Mapping species richness of plant families in European vegetation. *Journal of Vegetation Science*, 31(3), e13035. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/jvs.13035

Večeřa, M., Divíšek, J., Lenoir, J., Jiménez Alfaro, B., Biurrun, I., Knollová, I. et al. (2019) Alpha diversity of vascular plants in European forests. *Journal of Biogeography*, 46(9), 1919–1935. Available from: https://doi.org/10.1111/jbi.13624

Zannini, P., Frascaroli, F., Nascimbene, J., Halley, J.M., Stara, K., Cervellini, M. et al. (2022) Investigating sacred natural sites and protected areas for forest area changes in Italy. *Conservation Science and Practice*, 4(8), e12695. Available from: https://doi.org/10.1111/csp2.12695

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2010) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. Available from: https://doi.org/10.1111/j.2041-210X.2009.00001.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. List of selected preferential plots from four Italian databases.

Appendix S2. Summary of analyses and results on plots clustering to vegetation types.

Appendix S3. Lists of indicator species (p values <0.01) divided by life forms and ordered by decreasing fidelity value to Italian forest types calculated on probabilistic and preferential data sets with different plot sizes.

Appendix S4. Information emerging from each data set using both homogeneous and heterogeneous plot-size selection in the preferential data set.

How to cite this article: Alessi, N., Bonari, G., Zannini, P., Jiménez-Alfaro, B., Agrillo, E., Attorre, F. et al. (2023) Probabilistic and preferential sampling approaches offer integrated perspectives of Italian forest diversity. *Journal of Vegetation Science*, 34 e13175. Available from: https://doi.org/10.1111/jvs.13175

16541103, 2023, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jvs.13175 by Universita Di Camerino, Wiley Online Library on [23/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License