

Alpha diversity of vascular plants in European forests

Running title: Alpha diversity in European forests

Martin Večeřa^{1*}, Jan Divíšek^{1,2}, Jonathan Lenoir³, Borja Jiménez-Alfaro⁴, Idoia Biurrun⁵, Ilona Knollová¹, Emiliano Agrillo⁶, Juan Antonio Campos⁵, Andraž Čarni⁷, Guillermo Crespo Jiménez⁸, Mirjana Čuk⁹, Panayotis Dimopoulos¹⁰, Jörg Ewald¹¹, Federico Fernández-González⁸, Jean-Claude Gégout¹², Adrian Indreica¹³, Ute Jandt^{14,15}, Florian Jansen¹⁶, Zygmunt Kącki¹⁷, Valerijus Rašomavičius¹⁸, Marcela Řežníčková¹, John S. Rodwell¹⁹, Joop H. J. Schaminée²⁰, Urban Šilc⁷, Jens-Christian Svenning^{21,22}, Grzegorz Swacha¹⁷, Kiril Vassilev²³, Roberto Venanzoni²⁴, Wolfgang Willner^{25,26}, Thomas Wohlgemuth²⁷ & Milan Chytrý¹

* Corresponding author: martinvec@seznam.cz

¹ Department of Botany and Zoology, Masaryk University, Brno, Czech Republic

² Department of Geography, Masaryk University, Brno, Czech Republic

³ UR 'Ecologie et Dynamique des Systèmes Anthropisés' (EDYSAN, UMR 7058 CNRS-UPJV), Jules Verne University of Picardie, Amiens, France

⁴ Research Unit of Biodiversity (CSIC/UO/PA), University of Oviedo, Mieres, Asturias, Spain

⁵ Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU, Bilbao, Spain

⁶ Department of Environmental Biology, Sapienza University of Rome, Rome, Italy

⁷ Jovan Hadži Institute of Biology, Research Centre of the Slovenian Academy of Sciences and Arts, Ljubljana, Slovenia

⁸ Institute of Environmental Sciences, University of Castilla-La Mancha, Toledo, Spain

⁹ Department of Biology and Ecology, University of Novi Sad, Novi Sad, Serbia

¹⁰ Department of Biology, Laboratory of Botany, University of Patras, Patras, Greece

¹¹ Faculty of Forestry, Weihenstephan-Triesdorf University of Applied Sciences, Freising, Germany

¹² Université de Lorraine, AgroParisTech, INRA, Silva, Nancy, France

¹³ Department of Silviculture, Transilvania University of Braşov, Braşov, Romania

¹⁴ Department of Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Halle, Germany

¹⁵ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

¹⁶ Faculty of Agricultural and Environmental Sciences, University of Rostock, Rostock, Germany

¹⁷ Botanical Garden, University of Wrocław, Wrocław, Poland

¹⁸ Laboratory of Flora and Geobotany, Institute of Botany, Nature Research Centre, Vilnius, Lithuania

¹⁹ Lancaster, UK

²⁰ Wageningen Environmental Research, Wageningen, the Netherlands

²¹ Center for Biodiversity Dynamics in a Changing World, Department of Bioscience, Aarhus University, Aarhus, Denmark

²² Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Aarhus, Denmark

²³ Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

²⁴ Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy

²⁵ Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria

²⁶ Vienna Institute for Nature Conservation and Analyses, Vienna, Austria

²⁷ Forest Dynamics Research Unit, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Acknowledgements

We thank Pavel Daněk, Jürgen Dengler, Pavel Novák, Jan Roleček, David Zelený, and Dominik Zukal for helpful comments on the methods and interpretation of the results; František Kuda, Petr Vybrál, and Lubomír Tichý for their help with GIS and JUICE analyses; Els De Bie, Henry Brisse, Laura Casella, Úna FitzPatrick, Xavier Font, Thilo Heinken, Tomáš Peterka, Solvita Rūsiņa, Željko Škvorc, Ioannis Tsiripidis, Milan Valachovič, and Institut national de l'information géographique et forestière (<https://inventaire-forestier.ign.fr>) for significant data contributions from the vegetation-plot databases they manage. Our special thanks belong to all the botanists and foresters who sampled the vegetation plots used in our analyses and to the three anonymous reviewers for their helpful comments on previous versions of the manuscript.

Funding

M.V., J.D., I.K., M.Ř., and M.C. were supported by the Czech Science Foundation (Centre of Excellence Pladias; project no. 14–36079G). I.B. and J.A.C. were supported by the Basque Government (IT936-16). B.J.-A. was supported by the Marie Curie Clarín-COFUND program of the Principate of Asturias and the European Union (ACB17-26). J.-C.S. considers this work a contribution to his VILLUM Investigator project “Biodiversity Dynamics in a Changing World” funded by VILLUM FONDEN (grant 16549) and his Danish Council for Independent Research | Natural Sciences TREECHANGE project (grant 6108-00078B).

Abstract

Aim

The former continental-scale studies modelled coarse-grained plant species-richness patterns (gamma diversity). Here we aim to refine this information for European forests by (i) modelling the number of vascular plant species that co-occur in local communities (alpha diversity) within spatial units of 400 m²; and (ii) assessing the factors likely determining the observed spatial patterns in alpha diversity.

Location

Europe roughly within 12°W–30°E and 35–60°N.

Taxon

Vascular plants.

Methods

The numbers of co-occurring vascular plant species were counted in 73,134 georeferenced vegetation plots. Each plot was classified by an expert system into deciduous broadleaf, coniferous, or

1
2
3 sclerophyllous forest. Random Forest models were used to map and explain spatial patterns in alpha
4 diversity for each forest type separately using 19 environmental, land-use, and historical variables.

6 **Results**

7
8 Our models explained from 51.0 to 70.9% of the variation in forest alpha diversity. The modelled alpha-
9 diversity pattern was dominated by a marked gradient from species-poor North-western to species-rich
10 South-eastern Europe. The most prominent richness hotspots were identified in the Calcareous Alps and
11 adjacent north-western Dinarides, the Carpathian foothills in Romania, and the Western Carpathians in
12 Slovakia. Energy-related factors, bedrock types, and terrain ruggedness were identified as the main
13 variables underlying the observed richness patterns. Alpha diversity increases especially with
14 temperature seasonality in deciduous broadleaf forests, on limestone bedrock in coniferous forests, and
15 in areas with low annual actual evapotranspiration in sclerophyllous forests.

20 **Main conclusions**

21
22 We provide the first predictive maps and analyses of environmental factors driving the alpha diversity
23 of vascular plants across European forests. Such information is important for the general understanding
24 of European biodiversity. This study also demonstrates a high potential of vegetation-plot databases as
25 sources for robust estimation of the number of vascular plant species that co-occur at fine spatial grains
26 across large areas.

30 **Keywords**

31 Diversity, Europe, European Vegetation Archive (EVA), Forest vegetation, Plant community, Predictive
32 modelling, Random Forests, Species-richness patterns, Vascular plants, Vegetation-plot database

38 **Introduction**

39
40 Describing and understanding patterns of plant species richness have been among the overarching goals
41 of biogeography since its beginnings. Even now, plant diversity remains one of the key topics of
42 community ecology and macroecology, as it summarizes the outcome of many ecosystem processes
43 (Grace, 1999). New incentives for studying plant diversity patterns have arisen in recent decades in the
44 face of rapid human-induced environmental changes and subsequent biodiversity losses at the global
45 scale (e.g., Pereira et al., 2010). Given the fact that plants contribute, by far, the largest proportion of
46 terrestrial biomass on Earth (Bar-On, Phillips, & Milo, 2018), understanding fine-grain plant diversity
47 patterns across large spatial extents and their drivers is of primary interest in both science and
48 conservation practice.

49
50 Vascular plant diversity has been explored and mapped in numerous studies performed all around
51 the world. It has been addressed in various geographical and environmental settings ranging from (i)
52 regional extents (e.g., Albuquerque, Olalla-Tárraga, Montoya, & Rodríguez, 2011; Divíšek & Chytrý,
53 2018; Wohlgemuth, 1998), sometimes focused on specific biomes (e.g., Crespo-Mendes, Laurent,
54 Bruun, & Hauschild, 2019); (ii) through continental extents (e.g., Araújo, Thuiller, Williams, &

1
2
3 Reginster, 2005; Lahti & Lampinen, 1999; Ronk, Szava-Kovats, & Pärtel, 2015); (iii) up to the global
4 extent (e.g., Barthlott et al., 2007; Kreft & Jetz, 2007; Wulff, 1935). Despite considerable progress in
5 mapping plant species-richness patterns across large areas during the past decades (Barthlott et al., 2007;
6 Kier et al., 2005; Kreft & Jetz, 2007; Ulloa Ulloa et al., 2017), most studies to date have focused on
7 vascular plant species richness assessed at coarse spatial resolutions, using data from grid atlases (e.g.,
8 Araújo et al., 2005), expert-drawn range maps (e.g., Currie & Paquin, 1987), or floras and regional
9 checklists (e.g., Ulloa Ulloa et al., 2017). Such studies described the whole diversity of relatively large
10 areas, i.e., gamma diversity (Whittaker, 1972). If such studies analysed factors affecting plant species
11 richness across broad spatial extents (e.g., Francis & Currie, 2003; Kreft & Jetz, 2007; Qian, White, &
12 Song, 2007), they have mostly focused on the role of large-scale environmental drivers, especially
13 current climate (Hawkins et al., 2003). Very few studies have considered factors related to habitat
14 heterogeneity and environmental history and combined them into a single model of species richness
15 (e.g., Jiménez-Alfaro et al., 2018).

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Despite the wealth of knowledge on vascular plants as a taxonomic group, we still lack information on the global or continental distribution of their species richness at fine spatial resolution (Beck et al., 2012), also known as alpha diversity (Whittaker, 1972), i.e., the grain at which plant species potentially interact. This stems from the fact that continental and global vegetation-plot databases had been missing until very recently (Bruehlheide et al., 2019; Chytrý et al., 2016). Yet, only vegetation-plot data providing complete lists of species co-occurring locally can be used to model the spatial distribution of alpha diversity and to understand its determinants. This kind of data contains relatively precise information on species richness at georeferenced sites (Chytrý et al., 2016) as well as information on species composition that makes it possible to differentiate vegetation types. By producing maps of vascular plant species richness at the level of individual plots (16 and 400 m² for grasslands and forests, respectively), Divíšek & Chytrý (2018) demonstrated that vegetation-plot databases are a valuable source of data for high-resolution mapping of plant species richness of different vegetation types and species groups because each of them can exhibit a different diversity pattern. This and other recent studies (Jiménez-Alfaro et al., 2018; Lenoir et al., 2010) also showed that considering just one habitat type in a broad spatial extent may provide essential insights into the driving mechanisms of species richness, often related to ecology and history of a specific habitat.

Here, we focus on the alpha diversity of forest vegetation across a large part of the European continent at the level of all forest types together, and three main forest types separately – deciduous broadleaf, coniferous, and sclerophyllous. These forest types correspond to the main functional groups of European trees, which are driven by different biogeographical histories and ecological requirements (e.g., Barbero, Loisel, & Quézel, 1992; Bohn et al., 2003; Tzedakis, Emerson, & Hewitt, 2013). Thus we expect potential differences in their alpha-diversity patterns and related explanatory factors. Forests represent a major component of Europe's terrestrial ecosystems (Bengtsson, Nilsson, Franc, & Menozzi, 2000). Therefore we hypothesize that their alpha diversity displays similar spatial pattern as that

1
2
3 documented by previous maps of plant gamma diversity (e.g., Krefl & Jetz, 2007), i.e., the gradient from
4 the species-poor north to the species-rich south of the continent. The highest gamma diversity, as well
5 as the largest site-specific species pools of vascular plants in Europe, can be found in mountainous areas
6 in the southern and central parts of the continent (Araújo et al., 2005; Ronk et al. 2015). It is these areas
7 where we expect also the location of the highest alpha diversity in European forests (hereafter referred
8 to as 'richness hotspots').
9

10
11
12 North-south gradients in species richness may result partly from current climate conditions. It is
13 generally believed that species-richness patterns at extratropical latitudes with relatively cooler climates
14 can be explained by water-energy interactions, in which ambient heat is the main limiting factor for both
15 plants and animals (Hawkins et al., 2003; Jiménez-Alfaro, Chytrý, Mucina, Grace, & Rejmánek, 2016).
16 Nevertheless, there are other factors operating on regional to local scale that can play an important role.
17 Habitat heterogeneity, including land-cover composition, topography, and bedrock-chemistry gradients
18 in the surroundings of local communities, influences the species-pool size for the focal habitat, mass
19 effects, and dispersal limitation of specialized species (Divíšek & Chytrý, 2018; Janišová, Michalcová,
20 Bacaro, & Ghisla, 2014). Further, environmental or biogeographical history, especially past climate, has
21 been proposed as a strong factor controlling current richness patterns (e.g., Lenoir et al., 2010; Svenning,
22 Normand, & Skov, 2009). The European flora has undergone considerable changes in distribution
23 ranges as well as massive diversity losses during the Pleistocene climatic oscillations, which left
24 significant imprints in current floristic composition and richness patterns (Ewald, 2003; Svenning et al.,
25 2009).
26

27
28 On a local scale, numerous environmental factors have been suggested to affect species richness
29 of the herb layer in European forests, e.g., herb-layer productivity (e.g., Axmanová et al., 2012), light
30 availability (e.g., Ewald, 2008), soil pH (e.g., Zerbe, Schmidt, & Betzin, 2007), soil moisture (e.g.,
31 Härdtle, von Oheimb, & Westphal, 2003), soil nutrients (e.g., Pausas, 1994), tree layer species
32 composition (e.g., Kooijman & Cammeraat, 2010), fine-scale terrain configuration, or forest
33 management (e.g., Horvat, Biurrun, & García-Mijangos, 2017). The effects of these factors on species
34 richness are often interrelated, and their relative importance, as well as the direction of their effects, may
35 vary among individual forest community types, groups of species, or regional environmental contexts
36 (Axmanová et al., 2012; Härdtle et al., 2003). Given the absence of continent-wide data on such fine-
37 scale, spatially highly variable factors, we cannot directly include them in the modelling procedure.
38

39
40 As the focus of our study is the pattern of alpha diversity across a continent, i.e., regional
41 comparisons, we hypothesize that broad-scale environmental factors, particularly energy-related
42 climatic variables, play the main role here. Past climate and other historical factors can also have
43 influenced regional differences in alpha diversity. However, we also expect important effects of more
44 fine-grained factors (e.g., topographical heterogeneity, bedrock chemistry, and land cover) on alpha-
45 diversity patterns.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 To examine these hypotheses, we followed a predictive modelling approach combining a large set
4 of georeferenced vegetation plots covering most of the European forest types and spatial grids describing
5 the current and past environment, including the factors operating both on the coarse and fine scale. The
6 predictive modelling enabled us to join two perspectives on alpha-diversity patterns – mapping and
7 understanding the drivers – in a single project. The aims of this paper are: (i) to map the expected number
8 of vascular plant species which may co-occur locally within European deciduous broadleaf, coniferous,
9 and sclerophyllous forests; and (ii) to identify drivers that might determine the observed spatial patterns
10 in alpha diversity.
11
12
13
14
15
16

17 **Materials and methods**

18 *Study area*

19 Our study area spans from Portugal in the west to the European part of Turkey in the east, and from
20 southern Scandinavia in the north to Sicily and Crete in the south. We omitted the territories of Russia,
21 Belarus, Ukraine, Moldova, Iceland, Malta, Finland, and large parts of Norway and Sweden because of
22 a low density of available forest vegetation plots with complete species records.
23
24
25
26
27

28 *Data on species richness*

29 Georeferenced vegetation-plot records were obtained mainly from the European Vegetation Archive
30 (EVA; Chytrý et al., 2016; accessed on 18 January 2016), which integrates dozens of national and
31 regional vegetation databases covering most of Europe and contains more than 1.5 million vegetation
32 plots. For France, we additionally downloaded data from the French National Forest Inventory database
33 (Robert et al., 2012; <https://inventaire-forestier.ign.fr/>) to improve the spatial coverage. To identify
34 forest vegetation in the databases, we used an expert system for classification of European forests based
35 on their species composition (Schaminée et al., 2014), run in the JUICE program (Tichý, 2002). We
36 selected those plots which were classified by this expert system into forest habitat types according to
37 the modified EUNIS classification as used in the project of the European Red List of Habitats (Janssen
38 et al., 2016). Subsequently we merged these forest types into three main types – deciduous broadleaf,
39 coniferous, and sclerophyllous forests. We used plots with a surface area ranging between 100 and
40 1000 m², which were the most common plot sizes used for sampling. We initially selected 93,484 forest
41 plots. An overview of contributing databases and numbers of plots assigned to particular EUNIS forest
42 habitat types is available in Appendix S1, Tables S1.1 and S1.2 in Supporting Information.
43
44
45
46
47
48
49
50
51

52 In each plot, we merged multiple records of the same species occurring in more than one vegetation
53 layer (e.g., tree species occurring in the herb, shrub, and tree layers), removed non-vascular plants, and
54 counted the number of species. If the sampling date was available (82.1% of plots), we removed plots
55 older than 1970 (13.2%) in order to avoid potential bias related to long-term vegetation changes. A
56 common deficiency of vegetation-plot databases is that some plots are not accurately georeferenced,
57 i.e., there is a certain spatial difference between the actual sampling site and geographical coordinates
58
59
60

1
2
3 provided in the database (hereafter referred to as 'location uncertainty'). From the set of plots with
4 information on location uncertainty (56.3% of the remaining plots), we removed plots with an indicated
5 uncertainty larger than 2821 m (i.e., the radius of a circular buffer zone of 25 km²) and plots
6 georeferenced into grid cells of 10 × 10 km or larger (altogether 7.7%). By doing so, we reduced the
7 spatial mismatch between data on species richness and data on environmental conditions extracted
8 mostly from coarse-grain spatial grids. However, we retained plots with no information on location
9 uncertainty because some countries were represented only by this sort of data. Generally, the information
10 on location uncertainty was often not available in the database even for relatively recent plots
11 presumably geo-referenced with GPS devices.
12
13

14
15
16
17 Based on the selection criteria and assignment of plots to certain forest types, we prepared four
18 different datasets: (i) 'all forests' containing 73,134 plots (Figure 1); (ii) 'deciduous broadleaf forests'
19 containing 54,521 plots; (iii) 'coniferous forests' containing 15,978 plots; and (iv) 'sclerophyllous
20 forests' containing 2635 plots (Appendix S1, Figure S1.1).
21
22
23

24 25 *Environmental variables*

26
27 Predictor variables related to environmental and historical drivers of alpha-diversity patterns were
28 extracted from vector maps and raster grids using ArcGIS (ESRI, 2011) and QGIS (QGIS Development
29 Team, 2015). First, we delimited a circular buffer zone (~ 25 km²) around each vegetation plot to
30 calculate values of predictor variables that were then assigned to the respective plot. For quantitative
31 variables (e.g., mean annual temperature), we calculated the mean value within the buffer zone. For
32 categorical variables (e.g., area of forest), we calculated the proportion of the buffer zone area occupied
33 by each (e.g., land-cover) category (Appendix S1, Figure S1.3). We used this approach because the
34 coordinates of some vegetation plots are inaccurate and because environmental and historical data differ
35 in terms of spatial resolution. The buffer approach smoothed potentially extreme values of
36 environmental variables from point extractions and reduced potential spatial autocorrelation in model
37 residuals. The area of circular buffers around each vegetation plot corresponds to the size of the UTM
38 grid cells (5 × 5 km) that we used for visualization of the predicted alpha-diversity values across the
39 European continent.
40
41
42
43
44
45
46

47 The predictor variables can be divided into three broad groups: (i) current environment; (ii)
48 environmental history; and (iii) land cover (Table 1). The first group is represented by factors related to
49 topography, substrate (geology and soil), and climate. The second group contains variables related to
50 glacial history and palaeoclimate. The third group contains proportions of aggregated land-cover types
51 in the surroundings of each vegetation plot. A detailed description of the set of predictor variables used
52 as well as correlation matrices of variables for all forests and for each particular forest type are provided
53 in Appendix S1. In addition, we used plot size as a covariate in all of our models to account for the
54 potential effect of plot size variation (i.e., species-area relationships).
55
56
57
58
59
60

Modelling and mapping alpha-diversity patterns

We followed the procedure used by Divišek & Chytrý (2018) for modelling and mapping plant species richness at the local scale using the Random Forest algorithm (Breiman, 2001). This machine-learning method combining predictions of individual regression trees grown on a large number of bootstrap samples is considered very powerful for modelling complex relationships (Prasad, Iverson, & Liaw, 2006) that exist between species richness and the underlying predictor variables. Details of model fitting are described in Appendix S1.

The 19 variables listed in Table 1 as well as plot size were used as predictors and their relative importance within the models was expressed by ‘Mean decrease in accuracy’ and ‘Mean decrease in node impurity’. The former quantifies non-randomness of the effect of the particular predictor in the model, while the latter quantifies the predictive power of the predictor (see Appendix S1 for details on these measures). In order to consider both aspects of predictor importance, we rescaled both the measures to values ranging from 0 to 1 and calculated their average. Thus, we obtained a single measure of ‘integrated’ relative importance.

The models were calculated using the R software (version 3.2.5; R Core Team, 2016) with the ‘randomForest’ package (Liaw & Wiener, 2002). Relationships between species richness and individual predictors were visualized using partial dependence plots (Hastie, Tibshirani, & Friedman, 2009), which depict the effect of a given predictor on species richness while averaging effects of the other predictors in the model. Spatial autocorrelation in residuals of each model was checked using spatial correlograms showing Moran’s *I* statistics for distance classes defined using Sturges’ rule. As we were interested in spatial autocorrelation at short spatial distances, we randomly placed 500 rectangular windows (with a side length of 200 km) across sampled regions within the study area and calculated a spatial correlogram for a subset of plots in each of these windows. Resulting correlograms (i.e., Moran’s *I* values corresponding to distance classes defined in each window) were then summarized using local polynomial regression fitting with a span of 0.75.

Alpha-diversity values were predicted using the Random Forest models, based on the values of environmental variables calculated within UTM grid cells of 5×5 km, i.e., the same area as that of the circular buffer zones used to assign values of environmental variables to vegetation plots. The inclusion of plot size to the models enabled us to set a fixed plot size of 400 m² for the model predictions. This plot size was the one most frequently occurring in our dataset. Thus the predicted alpha diversity is the number of species expected to be found in a plot of 400 m² in a respective forest type within a given UTM grid cell of 5×5 km, which was used to visualize the predicted alpha-diversity value on maps. The cell size of 5×5 km used for visualization of our results was selected as a compromise between sufficiently high resolution and low computational demands. However, we also provide zoomed windows with the predicted alpha diversity for all forests visualized at a finer spatial resolution of 1×1 km, based on values of environmental variables calculated within UTM grid cells of this size (Appendix S2, Figure S2.8).

1
2
3 Richness hotspots and the top hotspots within them were identified as grid cells with predicted
4 alpha-diversity values in or above the 95th percentile (Prendergast, Quinn, Lawton, Eversham, &
5 Gibbons, 1993) and the 99th percentile, respectively.
6
7

8 To estimate the uncertainty of the predicted alpha diversity, we calculated and mapped standard
9 deviation of predictions of individual regression trees ($n = 500$) within each Random Forest model. We
10 also checked the spatial distribution of grid cells whose environmental conditions are out of range of
11 conditions covered by the vegetation-plot data, to find out areas where the models are extrapolating in
12 terms of environment.
13
14
15

16 17 *Stability of the predicted alpha-diversity patterns*

18 Vegetation-plot data were originally collected by botanists, foresters, nature conservationists, and
19 people from related fields for various purposes (Divíšek & Chytrý, 2018). This implies several
20 peculiarities of such kind of data, two of which are especially relevant for modelling species-richness
21 patterns. First, it is an uneven sampling effort (Knollová, Chytrý, Tichý, & Hájek, 2005), with large
22 variability in spatial coverage and density of vegetation-plot records across Europe in our data. Some
23 regions are undersampled, while in others, plots tend to be spatially clustered. This implies a risk of bias
24 in model results. A specific case of clustering is that groups of several plots share the same geographical
25 coordinates (35.7% of the plots used). These groups originate from repeated surveys over time (i.e., time
26 series from permanent or quasi-permanent plots) or from surveys of small areas, e.g., nature reserves,
27 in which all individual plots were assigned the same geographical coordinates. The second important
28 issue is preferential sampling, which is traditionally applied within vegetation surveys following the
29 European phytosociological school (Westhoff & van der Maarel, 1978). Preferential sampling may
30 result in species-rich communities being sampled more evenly across a study area than species-poor
31 communities (Chytrý, 2001), thus again potentially introducing a bias to estimates of species richness.
32
33
34
35
36
37
38
39
40

41 To test whether the predicted patterns of alpha diversity might be affected by these issues, we
42 applied three different resampling strategies (following Divíšek & Chytrý, 2018) and compared the
43 resulting models with the model based on the original dataset. The first strategy preferred plots from
44 undersampled regions: We calculated the density of plots across the study area using the Kernel Density
45 tool in ArcGIS and assigned a density value to each plot. Then we applied weighted random sampling
46 to create a subset containing two-thirds of plots from the original dataset with a preference of plots from
47 the poorly sampled regions. The other two strategies reduced local oversampling using criteria of spatial
48 distance among plots and their similarity in species composition: If two plots were sampled close to
49 each other and they were, at the same time, similar in species composition, then just one plot from such
50 a pair was selected (see Appendix S1 for a detailed description of the algorithm). We did two variants
51 of this resampling: (i) selecting the richer plot, and (ii) selecting a random plot from each pair. The
52 former approach reflected the assumption that species-rich plots might cover the study area more
53
54
55
56
57
58
59
60

1
2
3 representatively than species-poor ones, while the latter is close to the standard way of spatial
4 resampling. The resampling procedures were performed on the dataset of all forests.

5
6 A Random Forest model was fitted to each subset based on one resampling strategy. Predictions
7 of alpha diversity from the models were then correlated with each other and with the model based on
8 the original non-resampled dataset. We also checked spatial autocorrelation in the residuals of each
9 model (following the same procedure as explained in the section *Modelling and mapping alpha-diversity*
10 *patterns*).
11
12
13
14
15
16

17 **Results**

18
19 The Random Forest models explained between 51.0% and 70.9% of the variation in alpha diversity
20 (Table 2), with the best fit achieved for coniferous forests. Moran's *I* values at the 1st lag distance were
21 negative and close to zero for all the considered forest types, indicating no obvious spatial patterns in
22 the residuals (Appendix S2, Figures S2.9–S2.12).
23
24

25 We detected a marked geographical gradient in vascular plant alpha diversity from relatively
26 species-poor forests in North-western Europe to relatively species-rich forests in South-eastern Europe
27 (Figure 2a), i.e., from oceanic lowlands to more continental mountainous and hilly areas. The most
28 prominent top richness hotspots of all forests (on average 47 ± 4 species predicted to co-occur within a
29 400-m² plot) were modelled for both the Northern and Southern Calcareous Alps and adjacent north-
30 western Dinarides, the Carpathian foothills in Romania, and the Western Carpathians (Figure 2b).
31 Alpha-diversity patterns were similar for both deciduous broadleaf and coniferous forests, with species-
32 rich areas concentrated mostly in Central and South-eastern Europe (Figure 2c, d). The top richness
33 hotspots for deciduous broadleaf forests (on average 45 ± 4 species) were more scattered, including
34 mostly lower parts of mountain ranges from the western Pyrenees to the Carpathian foothills in Romania,
35 but also lowland forests of the hemiboreal zone in Lithuania (Appendix S2, Figure S2.13a). For
36 coniferous forests, the top richness hotspots (on average 48 ± 3 species) were limited to mountainous
37 areas, especially to the Calcareous Alps, north-western Dinarides, and Western Carpathians (Appendix
38 S2, Figure S2.13b). In sclerophyllous forests, high alpha diversity is predicted mainly in southern parts
39 of the three Southern European peninsulas (Figure 2e), with the top richness hotspots (on average $43 \pm$
40 4 species) in the eastern-central Iberian Peninsula, the central parts of Sicily, the Peloponnese, and along
41 the Ionian coast of Greece (Appendix S2, Figure S2.13c).
42
43
44
45
46
47
48
49
50
51

52 Predictors contributing the most to the total explained variation in all forests were temperature
53 seasonality, the proportion of limestone bedrock, and the proportion of forest areas (Figures 3a and 4a).
54 In deciduous broadleaf forests, the most important predictors were temperature seasonality, annual
55 precipitation, and the proportion of forest area (Figure 3b). High alpha diversity was related to higher
56 values of both the climatic variables, while the proportion of forest area showed no obvious trend with
57
58
59
60

1
2
3 only slightly higher alpha diversity related to less forested landscapes (Figure 4b). In coniferous forests,
4 alpha diversity was best explained by the proportion of limestone bedrock, terrain ruggedness, and
5 temperature seasonality (Figure 3c). All the predictors showed a positive effect here, with high alpha
6 diversity related to medium to high proportions of limestone bedrock, more rugged terrains, and higher
7 values of temperature seasonality (Figure 4c). In sclerophyllous forests, alpha diversity responded
8 primarily to climatic variables (Figure 3d), being high particularly in areas with low annual actual
9 evapotranspiration, low annual precipitation, and high annual potential evapotranspiration (Figure 4d).
10 The effects of the lower-ranked predictors on alpha diversity are summarized in Appendix S2, Figures
11 S2.15–S2.18.

12
13
14
15
16
17 The mean prediction uncertainty of all the models, measured as the standard deviation of
18 predictions of individual regression trees, was ± 12 species. Higher uncertainties were found generally
19 in species-rich areas such as the Western Alps and in some areas with deficient data, e.g., the northern
20 part of the Romanian Carpathians (Appendix S2, Figure S2.19). The range of environmental conditions
21 covered by the input vegetation-plot data was well captured within almost the whole study area
22 (Appendix S2, Figure S2.20). Relatively larger areas where values of one predictor variable are out of
23 the range covered by the vegetation-plot data were found only in small parts of Sweden, the Baltic
24 countries, and Poland. In these cases, there were higher values of temperature change since the Last
25 Glacial Maximum (LGM) and current temperature seasonality than covered by the plot data.

26
27
28
29
30
31
32 In all cases, the models based on different resampling strategies predicted almost the same patterns
33 of alpha diversity of all forests as the model based on the original non-resampled dataset (Spearman's
34 correlation coefficient between models ranged from 0.96 to 0.98). The latter model had a slightly better
35 fit to the data (56% of explained variation) than the models based on resampling (47–51%). There was
36 a negligible slightly negative spatial autocorrelation in residuals from all the models (see *Tests of*
37 *stability of the predicted alpha-diversity patterns* in Appendix S2). These results indicate that the
38 predicted alpha-diversity patterns are very robust, not depending on whether a resampling procedure is
39 applied or not.

44 45 46 **Discussion**

47 *Modelling and mapping alpha diversity across European forests*

48
49 Based on a set of predictors capturing current environmental conditions and land cover as well as past
50 climate changes, the proportion of the variation in vascular plant alpha diversity of European forests
51 explained in our models was on average 59.5%. Other recent studies using vegetation plot data to model
52 species-richness patterns in forest communities explained 47% (for all forests in the Czech Republic;
53 Divišek & Chytrý, 2018) and 31% of the variation (for European beech forests; Jiménez-Alfaro et al.,
54 2018), respectively. However, these comparisons have to be taken with caution, as the amount of
55 explained variation is always somehow influenced by the modelling algorithm and the number of
56 predictor variables used.
57
58
59
60

1
2
3 We provide the first maps of the number of vascular plant species expected to co-occur locally
4 (i.e., alpha diversity) for the major forest types across a considerable part of the European continent. We
5 reveal broad-scale patterns in alpha diversity that are somewhat similar to former work depicting
6 European plant species-richness patterns at a much coarser grain (i.e., gamma diversity; Araújo et al.,
7 2005; Lahti & Lampinen, 1999; Ronk et al., 2015; Svenning et al., 2009), with a spatial gradient ranging
8 from relatively species-poor oceanic lowlands of North-western Europe to relatively species-rich
9 continental and mountainous areas of South-eastern Europe. This similarity between alpha diversity of
10 European forests and gamma diversity of the whole flora might be a result of the species-pool effect
11 (e.g., Cornell & Harrison, 2014), assuming that gamma diversity well reflects the species-pool sizes of
12 forest habitats. However, the congruence between gamma diversity and species pool may strongly vary
13 region by region, and not even a significant correlation between the size of the species pool and richness
14 per unit area necessarily demonstrates the effect of species pool on the richness of local communities
15 (Herben, 2000). Nevertheless, such a comparison was not the purpose of this study.

16
17
18
19
20
21
22
23
24 Although we found forest richness hotspots mainly in Central and South-eastern European
25 mountain ranges, thus supporting former findings from the studies of gamma diversity, our maps also
26 suggest important and ‘unexpected’ richness hotspots in Eastern Europe, within the lowland forests of
27 the hemiboreal zone in Lithuania, as well as along the Carpathian foothills in Romania. This trend is
28 consistent with the observation of very high forest species richness in continental areas within the forest-
29 steppe zone (Chytrý et al., 2012). We suggest that these continental areas should receive more attention
30 from both the scientific and nature conservation perspectives in the pan-European context.

31
32
33
34
35 Species-richness patterns across large spatial extents have been traditionally understood as a
36 function of the current environment and especially climate conditions (e.g., Currie & Paquin, 1987;
37 Francis & Currie, 2003). This is valid also for alpha diversity of vascular plants in European forests.
38 Although the three forest types differ in the proposed main driving mechanisms of alpha diversity,
39 energy-related climatic factors are among the three most important predictors of each of the types
40 (Figure 3). The importance of habitat heterogeneity and landscape context for fine-grained species-
41 richness patterns (Divíšek & Chytrý, 2018; Janišová et al., 2014) was also corroborated by high ranking
42 of limestone bedrock, terrain ruggedness, and land cover of the surrounding landscape such as
43 proportions of forest area and mosaic land area with non-forest semi-natural habitats. The past climate
44 has been documented to influence especially gamma diversity (e.g., Svenning et al., 2009), while its
45 effects on alpha-diversity patterns are less obvious (Lenoir et al., 2010). We accounted for such historical
46 effects by including predictor variables related to (i) climate change (e.g., Sandel et al., 2011), assuming
47 that more stable climatic conditions between the late Pleistocene and Holocene may have favoured
48 refugial effects, and thus higher species richness; and (ii) LGM glaciation, assuming that current forest
49 communities existing in the areas recolonized after deglaciation are poor in species due to dispersal lag
50 (e.g., Svenning & Skov, 2004). The climate change predictors (temperature and precipitation change
51 since the LGM) were ranked relatively high in the case of coniferous and deciduous broadleaf forests.
52
53
54
55
56
57
58
59
60

1
2
3 However, the presence of LGM glaciers had the lowest ranking in our models (Figure 3), suggesting
4 that this predictor was not able to capture the potentially significant role of postglacial dispersal from
5 refugia of forest alpha diversity. Since the only refugia of forest species in Europe with relatively well-
6 known locations are those of beech forests (Magri, 2008), developing a predictor of forest species
7 richness accounting for postglacial dispersal events remains a challenge for further research (Jiménez-
8 Alfaro et al., 2018).

9
10
11
12 The proportion of unexplained variation in the data may be attributed to local-scale environmental
13 conditions (e.g., Axmanová et al., 2012; Ewald, 2008), forest management (e.g., Horvat et al., 2017),
14 and historical factors not accounted for by our predictor variables, as suggested by the negative (although
15 negligible) spatial autocorrelation in the model residuals (Table 2). A crucial point here may be the role
16 of forest management and its history, as it largely influences local-scale environmental conditions,
17 especially the availability of light. Almost all forests in Europe (>99%) are subject to some kind of
18 management (Parviainen, 2005), and the management intensity is highly variable in space and time.
19 Only the most intensive management practices (particularly clear-cutting and changes in tree species
20 composition) have been proven to have a significant negative effect on vascular plant species richness
21 (Paillet et al., 2010). In contrast, low-intensity or ‘traditional’ forest management practices such as
22 coppicing can promote species richness (e.g., Hédler, Kopecký, & Komárek, 2010). The vegetation-plot
23 data we used come mostly from botanical (phytosociological) field surveys that traditionally focus on
24 forests with near-natural species composition, therefore they should not be markedly affected by
25 intensive forest management. We rather suppose that locally positive effects of management on alpha
26 diversity may be reflected in our results. For instance, in some of the predicted richness hotspots, low-
27 intensity management practices such as forest grazing (e.g., in the Alps or Jura Mountains; Ewald, 2000;
28 Peringer et al., 2013), selective logging (e.g., in Slovenia; Veselič et al., 2008), or coppicing in the past
29 (e.g., along the Carpathian foothills in Romania and in Poland; Szymura, 2012) are supposed to support
30 or maintain high species richness. These traditional practices are also known to favour an open forest
31 structure (e.g., Decoq et al., 2004), creating benign conditions for forest specialists in shady patches as
32 well as canopy openings suitable for more light-demanding and generalist species coming from the
33 surrounding landscape. Our data did not directly include such drivers, even though we used proxies
34 potentially capturing some of their local variation (e.g., Jiménez-Alfaro et al., 2018). Nevertheless, non-
35 measured variables and stochastic processes such as patch dynamics and spatial mass effects may play
36 a role in shaping spatial patterns in alpha diversity.

37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 *Deciduous broadleaf forests*

Our results suggest a strong relationship between the alpha diversity of deciduous broadleaf forests and energy-related variables such as temperature seasonality and annual precipitation. High alpha diversity in deciduous broadleaf forests is predicted in areas with pronounced temperature seasonality, i.e., more

1
2
3 continental climates, and also in more oceanic precipitation-rich areas. These patterns may be considered
4 from the perspective of two different spatial scales: (i) the continental scale, and (ii) the regional scale.

5
6 On the continental scale, these two climatic modes generally do not overlap, and the richness
7 hotspots associated with them thus require more or less separate explanations. In more continental areas,
8 alpha diversity can be positively affected by relatively high energy inputs during the warm and long
9 growing seasons. This may apply to both deciduous broadleaf and coniferous forests. In addition, there
10 can be indirect effects of pronounced temperature seasonality – in particular, the existence of forest-
11 steppe in more continental areas of South-eastern Europe enables mixing of forest and light-demanding
12 species. The high alpha diversity of deciduous broadleaf forests in precipitation-rich areas on the pan-
13 European scale is found especially along the mountain fringes, where rainfall, large areas of forests, and
14 high topographical heterogeneity coincide, e.g., in the Western Alps, Southern Calcareous Alps, and
15 northern Dinarides. The adjacent areas in the southeast and southwest are considered to have served as
16 refugia of some forest species during the Pleistocene glaciations (Magri, 2008; Tzedakis et al., 2013).
17 Postglacial migration from the nearby refugia has been suggested to be among the key factors shaping
18 current high species richness of beech forests in the Alps (Jiménez-Alfaro et al., 2018; Ujházyová et al.,
19 2016). The survival of forest species in the refugia may have been facilitated, among other factors, by
20 relatively humid conditions due to orographic precipitation. From this point of view, current high
21 precipitation may not have a direct effect on current species richness in that area, but it may correlate
22 with relatively high precipitation in the full glacial, which may have been crucial.

23
24 On the regional scale, locally specific combinations of precipitation and temperature seasonality
25 may occur. It has been shown that slightly higher continentality in a relatively precipitation-rich area or
26 slightly higher precipitation in a relatively continental area may cause locally or regionally increased
27 species richness (e.g., Horvat et al., 2017; Wohlgemuth, 1998). This may be, for instance, the case of
28 high forest alpha diversity in some areas of the western Balkan Peninsula.

29
30 A low proportion of forest area in the surroundings of the sampled site is related to slightly higher
31 alpha diversity for deciduous broadleaf forests, however, the trend is not very pronounced (Figure 4b).
32 The same equivocal pattern was found by Divíšek & Chytrý (2018) at a regional scale within Czech
33 forests. This can likely be attributed to two contrasting tendencies, both affecting forest species richness
34 via enlarging/maintaining species pools: (i) spatial mass effects resulting from habitat heterogeneity in
35 the landscapes with smaller forest areas, more ecotones, and consequently an even higher chance for
36 generalist species to occur in a forest; and (ii) within-forest habitat heterogeneity, connectivity, long-
37 term continuity and stability of forest environments in more forested landscapes (Honnay et al., 2002).

38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 *Coniferous forests*

All the top richness hotspots for coniferous forests are in areas dominated by limestone or other
calcareous bedrocks. This pattern is likely related to the evolutionary and glacial history of Central
Europe. Existing systems of indicator values suggest that a vast majority of Central European vascular

1
2
3 plant species prefer base-rich and calcareous soils (e.g., Ellenberg et al., 1991), which is also valid for
4 the forest flora (Ewald, 2003; Ujházyová et al., 2016). A positive relationship between species richness
5 and soil pH may be mediated through species-pool effects, namely the legacy of selection towards a
6 base-rich flora within biogeographical regions dominated by calcareous bedrocks (Ewald, 2003; Pärtel,
7 2002).

8
9
10
11 Montane coniferous forests, often growing on shallow soil and in rugged terrain, tend to have a
12 relatively sparse canopy which enables occurrence of light-demanding species with an optimum in non-
13 forest vegetation. This is true for both calcareous and non-calcareous soils, but the species pool on
14 calcareous bedrock is much larger. Another factor increasing local species richness of montane
15 coniferous forests on limestones is an admixture of acidophilous species in local patches of acid forest
16 floor arising from poorly decomposed needle litter and old deadwood structures, particularly at higher
17 elevations (Ewald, 2008; Ujházyová et al., 2016).

18
19
20
21
22 Terrain ruggedness was the second most important driver of alpha-diversity patterns in European
23 coniferous forests. More rugged terrains tend to harbour more species-rich forests than flat or gently
24 undulating landscapes. This partly reflects diverse environmental conditions and partly higher isolation
25 of habitats within topographically heterogeneous landscapes. These effects result in various processes
26 increasing species richness, in particular: (i) habitat heterogeneity and spatial mass effects (e.g., Zelený,
27 Li, & Chytrý, 2010); (ii) locally buffered climate change and related refugial effects during the periods
28 of macroclimatic oscillations (e.g., Sandel et al., 2011); and (iii) taxonomic diversification (e.g.,
29 Rosenzweig, 1995). Ruggedness may also be a proxy for poor accessibility, low management intensity,
30 and high stand age of mountain forests, which favours vascular plant understorey richness (Ewald,
31 2008).

32 33 34 35 36 37 38 39 *Sclerophyllous forests*

40
41 In sclerophyllous forests, the highest numbers of species are predicted for areas with relatively low
42 annual actual evapotranspiration and annual precipitation, and high annual potential evapotranspiration.
43 Similarly, Naveh and Whittaker (1979) found increasing alpha diversity with decreasing rainfall in
44 Mediterranean shrublands and woodlands in Israel. We assume that low precipitation combined with
45 high energy inputs (i.e., high potential evapotranspiration) determines openness of forest stands in the
46 Mediterranean, enabling the co-existence of forest specialists and light-demanding species with optima
47 outside forests. The climatic causes of the modelled patterns are further supported by low summer
48 rainfall amounts in some richness hotspots, e.g., in Sicily and south-western Greece. This suggests that
49 high alpha diversity in sclerophyllous forests is related to areas with a more pronounced Mediterranean
50 climate with its warm dry summers and mild wet winters. In warmer and drier areas, there is also a
51 larger habitat-specific pool of sclerophyllous forest species, including a higher diversity of
52 sclerophyllous trees and shrubs (Rivas-Martínez, 2011).

1
2
3 The Mediterranean Basin and other Mediterranean-climate ecosystems have been identified as
4 global biodiversity hotspots (Cowling, Rundel, Lamont, Arroyo, & Arianoutsou, 1996; Kreft & Jetz,
5 2007; Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000). The high plant alpha diversity in
6 these ecosystems is supposed to be related to open habitats, including long-term grazed shrublands and
7 woodlands, whereas denser shrublands and woodlands tend to have lower diversity (Cowling et al.,
8 1996). A large pool of open-habitat species, some of which are able to disperse and survive in forests,
9 is partly due to the long-term adaptation of these species to natural disturbances, especially to fire and
10 intensive grazing that have maintained open landscapes for millennia (Naveh & Whittaker, 1979). More
11 recently, the regime of frequent disturbances has continued with some natural processes being
12 substituted or even intensified due to human activities (e.g., deforestation; Bohn et al., 2003), largely
13 favouring open-habitat species. Species-rich sclerophyllous forests may also be found at higher
14 elevations with different climatic conditions than the Mediterranean lowlands, e.g., higher summer
15 rainfall (Appendix S2, Figure S2.18). In such conditions, their communities are enriched with species
16 of deciduous forests, which may be the case in some parts of the eastern Iberian Peninsula. However, a
17 thorough analysis of the ecological preferences of species making up high alpha diversity in European
18 sclerophyllous forests remains a task for further research.

30 *Methodological aspects of modelling species-richness patterns using vegetation-plot data*

31 In this study, we followed the recent trend of using vegetation-plot data to analyse broad-scale
32 biogeographical phenomena (e.g., Jiménez-Alfaro et al., 2018; Wagner et al., 2017), integrating
33 community ecology and macroecological modelling. This approach opens new perspectives in
34 understanding relationships between species-richness patterns and spatially structured environmental
35 factors (Divíšek & Chytrý, 2018). However, it implies also several issues which must be carefully
36 considered, namely: (i) uneven geographical coverage of vegetation-plot data, (ii) preferential sampling,
37 and (iii) scale mismatch between plot-derived species richness and environmental variables.

38 The data used to infer alpha-diversity patterns must be sufficiently representative in terms of both
39 geographical coverage and potential bias in richness estimates. We excluded marginal areas of the
40 continent with scarce data and assumed that species-richness patterns might be well interpolated for the
41 data gaps in non-marginal areas by our models, since (i) environmental conditions of such gaps were
42 well represented in the data from more densely sampled areas, and (ii) all biogeographical regions and
43 all main forest types within the study area were covered. The extent of the UTM grid used for prediction
44 of alpha diversity almost fully fitted the range of environmental conditions covered by the vegetation-
45 plot data, with minor exceptions in the very northeast and southeast of the study area. That means model
46 predictions were very rarely extrapolated in terms of environmental conditions. High prediction
47 uncertainties were typical mostly of species-rich areas. Such a pattern is partly related to the fact that
48 there are species-rich and species-poor communities often occurring and sampled at nearby sites, which
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 might make the prediction output ambiguous. In contrast, areas dominated by species-poor communities
4 had very low prediction uncertainties.
5

6
7 However, there were still parts of the geographic range of some forest types where vegetation-plot
8 data were lacking, which was partly reflected in the maps of prediction uncertainty. Regarding the
9 European sclerophyllous forest range, for instance, we had fewer plots from the Iberian Peninsula than
10 from the Italian and Balkan Peninsulas. Also, the predicted richness hotspot on the Peloponnese, though
11 corroborated by an expert evaluation, should be better supported by plot data. Thus, the predicted
12 species-richness pattern may change slightly in the future with more data from the current gaps added.
13 Another related inaccuracy might be in the predicted pattern for coniferous forests. Their range was
14 estimated from the CORINE Land Cover database which does not distinguish native forest stands from
15 conifer tree plantations. Therefore the predicted species richness might be overestimated in some areas
16 with extensive plantations (e.g., the United Kingdom, northwestern Iberian Peninsula, and the Czech
17 Republic).
18
19

20
21
22
23
24 The use of different resampling strategies that either explicitly preferred plots from undersampled
25 areas or diminished the role of densely sampled areas in the models did not lead to notable differences
26 in the resulting predictions (Appendix S2, Figures S2.21 and S2.25). This finding suggests that the
27 species-richness pattern in European forests captured by the plot data is robust. In the end, we used the
28 model based on the original (non-resampled) dataset, as it explained most of the variation and had the
29 lowest mean of squared residuals, although its statistical parameters may be inflated due to high spatial
30 autocorrelation at very short distances in some areas. The preferential selection of sampling sites is
31 inherent to vegetation sampling in the European phytosociological school (Westhoff & van der Maarel,
32 1978), and thus it probably affects most of the databases we used. We examined the potential effect of
33 preferential sampling on predicted spatial patterns in alpha diversity by applying data resampling
34 strategies that selected either richer or random plots from plot clusters in oversampled areas. However,
35 the results were almost the same as those based on all data. Divišek and Chytrý (2018) also explicitly
36 tested how predicted species-richness patterns may change if they are based on 50% of plots from the
37 database with higher or lower species richness. In both cases, there was almost no difference in predicted
38 spatial patterns compared with those based on non-resampled data, although the average predicted
39 species richness might be slightly increased or decreased. Therefore, we assume that in the case of
40 European forests preferential sampling does not lead to any strong bias in predicted spatial patterns of
41 alpha diversity.
42
43
44
45
46
47
48
49
50
51

52
53 Our methodological approach involved a scale mismatch between species-richness data from small
54 plots and rather coarse grain of environmental variables used. As our study focused on inter-regional
55 differences in alpha diversity, which are primarily influenced by broad-scale environmental variables,
56 we do not assume that such mismatch would have unduly affected our results. Given that the vegetation
57 plots are not accurately georeferenced in some cases and that GIS layers of predictor variables differ in
58 spatial resolution, we used buffer zones to calculate values of predictor variables within the surroundings
59
60

1
2
3 of the plot coordinates provided in the database. The model prediction for a fixed plot size of 400 m²
4 was then projected into 5 × 5 km grid cells based on the values of explanatory variables calculated within
5 them, i.e., spatial units of the same size as those used as the model input. We consider this approach to
6 be appropriate for integrating (i) fine-scale information on alpha diversity with more coarse-scale
7 information on the environment, and (ii) different spatial resolutions of the explanatory variables which
8 are unavoidable in this kind of broad-scale analysis.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

List of table and figure legends

Table 1: Summary statistics of the predictor variables used in modelling vascular plant alpha diversity across European forests, for all forests and for each forest type separately.

Table 2: Results of Random Forest models and spatial autocorrelation in model residuals for vascular plant alpha diversity across European forests.

Figure 1: Spatial distribution of forest vegetation plots, ranging from 100 to 1000 m² in surface area, used as an input dataset for modelling alpha-diversity patterns (n = 73,134). The colours indicate the total number of vascular plant species recorded per plot (i.e., alpha diversity). For the purpose of visualization, alpha-diversity values displayed on this map were rescaled to 400 m² (the most frequent plot size within the dataset) using the species-area relationship obtained from the data (Appendix S1, Figure S1.2a). The richest plots are displayed on the top. Analogous maps for the subsets of deciduous broadleaf, coniferous, and sclerophyllous forests are in Appendix S1, Figure S1.1.

Figure 2: Predicted alpha diversity of vascular plants across European forests based on Random Forest models for (a) all forests (n = 73,134); (c) deciduous broadleaf forests (n = 54,521); (d) coniferous forests (n = 15,978); and (e) sclerophyllous forests (n = 2635). All maps are provided in a UTM 5 × 5 km grid, with species numbers predicted for a plot size of 400 m². Grid cells with forest area <10% (according to the CORINE Land Cover database; Bossard, Feranec, & O'ahel', 2000) are left white – for (a) the 10% threshold was applied to the area of all CORINE forest types; for (c) to broadleaf and mixed-forest area; for (d) to coniferous forest area; and for (e) to the broadleaf and mixed-forest area within the range of sclerophyllous forests according to Bohn et al. (2003).

(b) richness hotspots (light brown) and the top richness hotspots (black) of vascular plants in European forests for all forests – grid cells with predicted alpha-diversity values equal to or greater than the 95th and 99th percentile values, respectively. Analogous maps of richness hotspots for the subsets of deciduous broadleaf, coniferous, and sclerophyllous forests are in Appendix S2, Figure S2.13.

Figure 3: Integrated relative importance of predictor variables for explaining alpha-diversity patterns for all forests and for each forest type separately. The black part of the bars indicates the contribution of the 'Mean decrease in node impurity' to the integrated relative importance, the grey part indicates the contribution of the 'Mean decrease in accuracy' (see Appendix S1 for details on these measures, and Appendix S2 and Figure S2.14 for original values of these measures). LGM – Last Glacial Maximum.

Figure 4: Effects of the three top-ranked predictor variables on alpha diversity for all forests and for each forest type separately (partial dependence plots). The integrated importance scores are shown in parentheses. AET – actual evapotranspiration; PET – potential evapotranspiration; VRM – vector ruggedness measure. The effects of the lower-ranked predictors on alpha diversity are summarized in Appendix S2, Figures S2.15–S2.18.

Table 1: Summary statistics of the predictor variables used in modelling vascular plant alpha diversity across European forests, for all forests and for each forest type separately.

	All forests			Deciduous broadleaf			Coniferous			Sclerophyllous		
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
Current environment												
Elevation [m a.s.l.]	-5	529	2583	-5	473	2552	-1	745	2583	1	390	1595
Terrain ruggedness [VRM × 10 ⁴]	0	75	696	0	68	696	0	94	629	0	99	374
Heat load [index × 10 ⁴]	5354	5978	6896	5368	5974	6693	5354	5993	6896	5366	5992	6403
Topsoil pH	3	6	8	3	6	8	4	6	8	5	7	8
Limestone area [%]	0	17	100	0	16	100	0	19	100	0	25	100
Mean annual temperature [°C]	-2	9	19	-1	9	19	-2	7	18	7	14	19
Temperature seasonality [SD × 100]	302	638	881	302	635	881	333	659	879	366	564	736
Annual precipitation [mm/yr]	290	835	2051	290	813	1985	348	925	2051	359	738	1487
Precipitation seasonality [coeff. of var.]	7	26	101	8	25	101	7	27	99	15	41	88
Summer rainfall [%]	8	54	73	8	54	73	8	57	72	10	34	57
Annual potential evapotranspiration [mm/yr]	374	783	1345	375	796	1345	374	712	1234	633	937	1305
Annual actual evapotranspiration [mm/yr]	252	590	810	252	590	810	297	594	794	309	566	783
Environmental history												
Presence/absence of an LGM glacier	<i>binary variable</i>											
Temperature change since the LGM [°C]	1	4	13	1	4	13	2	5	13	2	3	4
Precipitation change since the LGM [mm]	1	43	567	1	42	567	1	46	293	16	63	443
Land cover												
Forest area [%]	0	47	100	0	46	100	0	55	100	0	35	99
Mosaic land area [%]	0	22	100	0	23	100	0	20	98	0	15	80
Shrubland area [%]	0	6	94	0	5	88	0	7	91	0	19	94
Intensively cultivated land area [%]	0	16	100	0	18	100	0	8	93	0	15	97

LGM – Last Glacial Maximum; SD – standard deviation of mean monthly temperatures; VRM – vector ruggedness measure.

Table 2: Results of Random Forest models and spatial autocorrelation in model residuals for vascular plant alpha diversity across European forests.

	All forests	Deciduous broadleaf	Coniferous	Sclerophyllous
No. of cases (vegetation plots)	73,134	54,521	15,978	2635
Explained variation [%]	56.3	51.0	70.9	59.8
Mean of squared residuals	80.2	80.9	73.9	42.8
Moran's I at 1 st lag distance (\pm SD)	-0.004 \pm 0.04	-0.012 \pm 0.04	-0.002 \pm 0.07	-0.027 \pm 0.11

For Peer Review

Data accessibility

Vegetation-plot data used to model the alpha-diversity patterns in European forests are stored in the European Vegetation Archive (EVA) – <http://euroveg.org/eva-database>. The particular dataset used in this study can be obtained from EVA upon request. Shapefiles containing the modelled alpha-diversity values, the identified richness hotspots as well as the maps of prediction uncertainty can be downloaded from the Figshare repository (<https://doi.org/10.6084/m9.figshare.8005703.v2>). All the other data used in this study are freely available from the cited publicly accessible databases.

References

- Albuquerque, F. S., Olalla-Tárraga, M. Á., Montoya, D., & Rodríguez, M. Á. (2011). Environmental determinants of woody and herb plant species richness patterns in Great Britain. *Ecoscience*, 18, 394–401.
- Araújo, M. B., Thuiller, W., Williams, P. H., & Reginster, I. (2005). Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography*, 14, 17–30.
- Axmanová, I., Chytrý, M., Zelený, D., Li, C.-F., Vymazalová, M., Danihelka, J., ... Diekmann, M. (2012). The species richness–productivity relationship in the herb layer of European deciduous forests. *Global Ecology and Biogeography*, 21, 657–667.
- Barbero, M., Loisel, R., & Quézel, P. (1992). Biogeography, ecology and history of Mediterranean *Quercus ilex* ecosystems. *Vegetatio*, 99/100, 19–34.
- Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the USA*, 115, 6506–6511.
- Barthlott, W., Hostert, A., Kier, G., Küper, W., Kreft, H., Mutke, J., ... Sommer, J. H. (2007). Geographic patterns of vascular plant diversity at continental to global scales. *Erdkunde*, 61, 305–315.
- Beck, J., Ballesteros-Mejia, L., Buchmann, C. M., Dengler, J., Fritz, S. A., Gruber, B., ... Dormann, C. F. (2012). What's on the horizon for macroecology? *Ecography*, 35, 673–683.
- Bengtsson, J., Nilsson, S. G., Franc, A., & Menozzi, P. (2000). Biodiversity, disturbances, ecosystem function and management of European forests. *Forest Ecology and Management*, 132, 39–50.
- Bohn, U., Neuhausl, R., Gollub, G., Hettwer, C., Neuhauslová, Z., Raus, T., ... Weber, H. (2003). *Map of the Natural Vegetation of Europe, Scale 1 : 2,500,000: Explanatory text with CD, Legend, Maps*. Münster, DE: Landwirtschaftsverlag.
- Bossard, M., Feranec, J., & Otáhel, J. (2000). *CORINE Land Cover technical guide – addendum 2000*. Copenhagen, DK: European Environment Agency.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.

- 1
2
3 Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., ...
4 Zverev, A. (2019). sPlot – a new tool for global vegetation analyses. *Journal of Vegetation Science*,
5 30, 161–186.
6
7
8 Chytrý, M. (2001). Phytosociological data give biased estimates of species richness. *Journal of*
9 *Vegetation Science*, 12, 441–444.
10
11 Chytrý M., Ermakov N., Danihelka J., Hájek M., Hájková P., Horsák M., ... Zelený D. (2012). High
12 species richness in hemiboreal forests of the northern Russian Altai, southern Siberia. *Journal of*
13 *Vegetation Science*, 23, 605–616.
14
15 Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., ... Yamalov
16 S. (2016). European Vegetation Archive (EVA): an integrated database of European vegetation
17 plots. *Applied Vegetation Science*, 19, 173–180.
18
19 Cornell, H. V. & Harrison, S. P. (2014). What are species pools and when are they important? *Annual*
20 *Review of Ecology, Evolution, and Systematics*, 45, 45–67.
21
22 Cowling, R. M., Rundel, P. W., Lamont, B. B., Arroyo, M. K., & Arianoutsou, M. (1996). Plant
23 diversity in mediterranean-climate regions. *Trends in Ecology and Evolution*, 11, 362–366.
24
25 Crespo-Mendes, N., Laurent, A., Bruun, H. H., & Hauschild, M. Z. (2019). Relationships between
26 plant species richness and soil pH at the level of biome and ecoregion in Brazil. *Ecological*
27 *Indicators*, 98, 266–275.
28
29 Currie, D. J. & Paquin, V. (1987). Large-scale biogeographical patterns of species richness of trees.
30 *Nature*, 329, 326–327.
31
32 Decocq, G., Aubert, M., Dupont, F., Alard, D., Saguez, R., Wattez-Franger, A., ... Bardat, J. (2004).
33 Plant diversity in a managed temperate deciduous forest: understorey response to two silvicultural
34 systems. *Journal of Applied Ecology*, 41, 1065–1079.
35
36 Divišek, J. & Chytrý, M. (2018). High-resolution and large-extent mapping of plant species richness
37 using vegetation-plot databases. *Ecological Indicators*, 89, 840–851.
38
39 Ellenberg, H., Weber, H. E., Düll, R., Wirth, V., Werner, W., & Paulissen D. (1991). Zeigerwerte von
40 Pflanzen in Mitteleuropa. *Scripta Geobotanica*, 18, 1–248.
41
42 ESRI (2011). *ArcGIS Desktop*. Redlands, CA: Environmental Systems Research Institute.
43
44 Ewald, J. (2000). Long-term impact of forest pasture on the understorey of mountain forests in the
45 Tegernsee Alps (Bavaria). *Zeitschrift für Ökologie und Naturschutz*, 9, 161–171.
46
47 Ewald, J. (2003). The calcareous riddle: Why are there so many calciphilous species in the Central
48 European flora? *Folia Geobotanica*, 38, 357–366.
49
50 Ewald, J. (2008). Plant species richness in mountain forests of the Bavarian Alps. *Plant Biosystems*,
51 142, 594–603.
52
53 Francis, A. P., & Currie, D. J. (2003). A globally consistent richness-climate relationship for
54 angiosperms. *The American Naturalist*, 161, 523–536.
55
56
57
58
59
60

- 1
2
3 Grace, J. B. (1999). The factors controlling species density in herbaceous plant communities: an
4 assessment. *Perspectives in Plant Ecology, Evolution and Systematics*, 2, 1–28.
- 5
6 Härdtle, W., von Oheimb, G., & Westphal, C. (2003). The effect of light and soil conditions on the
7 species richness of the ground vegetation of deciduous forests in northern Germany (Schleswig-
8 Holstein). *Forest Ecology and Management*, 182, 327–338.
- 9
10
11 Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining,*
12 *inference, and prediction* (2nd ed.). New York, NY: Springer-Verlag.
- 13
14 Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guégan, J.-F., Kaufman, D. M., ... Turner, J.
15 R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, 84,
16 3105–3117.
- 17
18
19 Hédl, R., Kopecký, M., & Komárek, J. (2010). Half a century of succession in a temperate oakwood:
20 from species-rich community to mesic forest. *Diversity and Distributions*, 16, 267–276.
- 21
22 Herben, T. (2000). Correlation between richness per unit area and the species pool cannot be used to
23 demonstrate the species pool effect. *Journal of Vegetation Science*, 11, 123–126.
- 24
25 Honnay, O., Verheyen, K., Butaye, J., Jacquemyn, H., Bossuyt, B., & Hermy, M. (2002). Possible
26 effects of habitat fragmentation and climate change on the range of forest plant species. *Ecology*
27 *Letters*, 5, 525–580.
- 28
29
30 Horvat, V., Biurrun, I., & García-Mijangos, I. (2017). Herb layer in silver fir-beech forests in the western
31 Pyrenees: Does management affect species diversity? *Forest Ecology and Management*, 385, 87–96.
- 32
33 Janišová, M., Michalcová, D., Bacaro, G., & Ghisla, A. (2014). Landscape effects on diversity of
34 semi-natural grasslands. *Agriculture, Ecosystems & Environment*, 182, 47–58.
- 35
36 Janssen, J. A. M., Rodwell, J. S., García Criado, M., Gubbay, S., Haynes, T., Nieto, A., ... Valachovič
37 M. (2016). *European Red List of Habitats. Part 2. Terrestrial and freshwater habitats.*
38 Luxembourg City, LU: Publications Office of the European Union.
- 39
40
41 Jiménez-Alfaro, B., Chytrý, M., Mucina, L., Grace, J. B., & Rejmánek, M. (2016). Disentangling
42 vegetation diversity from climate–energy and habitat heterogeneity for explaining animal
43 geographic patterns. *Ecology and Evolution*, 6, 1515–1526
- 44
45
46 Jiménez-Alfaro, B., Girardello, M., Chytrý, M., Svenning, J.-C., Willner, W., Gégout, J.-C., ...
47 Wohlgemuth, T. (2018). History and environment shape species pools and community diversity in
48 European beech forests. *Nature Ecology and Evolution*, 2, 483–490.
- 49
50
51 Kier, G., Mutke, J., Dinerstein, E., Ricketts, T. H., Küper, W., Kreft, H., & Barthlott, W. (2005).
52 Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, 32, 1107–
53 1116.
- 54
55
56 Knollová, I., Chytrý, M., Tichý, L., & Hájek, O. (2005). Stratified resampling of phytosociological
57 databases: some strategies for obtaining more representative data sets for classification studies.
58 *Journal of Vegetation Science*, 16, 479–486.
- 59
60

- 1
2
3 Kooijman, A. M., Cammeraat, E. (2010). Biological control of beech and hornbeam affects species
4 richness via changes in the organic layer, pH and soil moisture characteristics. *Functional Ecology*,
5 24, 469–477.
6
7
8 Kreft, H. & Jetz, W. (2007). Global patterns and determinants of vascular plant diversity. *Proceedings*
9 *of the National Academy of Sciences of the USA*, 104, 5925–5930.
10
11 Lahti, T. & Lampinen, R. (1999). From dot maps to bitmaps: Atlas Florae Europaeae goes digital.
12 *Acta Botanica Fennica*, 162, 5–9.
13
14 Lenoir, J., Gégout, J.-C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E., ... Svenning,
15 J.-C. (2010). Cross-scale analysis of the region effect on vascular plant species diversity in
16 southern and northern European mountain ranges. *PLoS ONE*, 5, e15734.
17
18 Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22.
19
20 Magri, D. (2008). Patterns of post-glacial spread and the extent of glacial refugia of European beech
21 (*Fagus sylvatica*). *Journal of Biogeography*, 35, 450–463.
22
23 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000).
24 Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.
25
26 Naveh, Z. & Whittaker, R. H. (1979). Structural and floristic diversity of shrublands and woodlands in
27 northern Israel and other Mediterranean areas. *Vegetatio*, 41, 171–190.
28
29 Paillet, Y., Bergès, L., Hjältén, J., Ódor, P., Avon, C., Bernhardt-Römermann, M., ... Virtanen, R.
30 (2010). Biodiversity differences between managed and unmanaged forests: meta-analysis of
31 species richness in Europe. *Conservation Biology*, 24, 101–112.
32
33 Pärtel, M. (2002). Local plant diversity patterns and evolutionary history at the regional scale.
34 *Ecology*, 83, 2361–2366.
35
36 Parviainen, J. (2005). Virgin and natural forests in the temperate zone of Europe. *Forest Snow and*
37 *Landscape Research*, 79, 9–18.
38
39 Pausas, J. G. (1994). Species richness patterns in the understorey of Pyrenean *Pinus sylvestris* forest.
40 *Journal of Vegetation Science*, 5, 517–524.
41
42 Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-
43 Manjarrés, J. F., ... Walpole, M. (2010). Scenarios for global biodiversity in the 21st century.
44 *Science*, 330, 1496–1501.
45
46 Peringer, A., Siehoff, S., Chetelat, J., Spiegelberger, T., Buttler, A., & Gillet, F. (2013). Past and
47 future landscape dynamics in pasture-woodlands of the Swiss Jura Mountains under climate
48 change. *Ecology and Society*, 18, 11.
49
50 Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques:
51 bagging and Random Forests for ecological prediction. *Ecosystems*, 9, 181–199.
52
53 Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C., & Gibbons, D. W. (1993). Rare
54 species, the coincidence of diversity hotspots and conservation strategies. *Nature*, 365, 335–337.
55
56
57
58
59
60

- 1
2
3 Qian, H., White, P. S., & Song, J.-S. (2007). Effects of regional vs. ecological factors on plant species
4 richness: an intercontinental analysis. *Ecology*, 88, 1440–1453.
5
6 QGIS Development Team (2015). *QGIS Geographic Information System*. Open source geospatial
7 foundation project.
8
9 R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, AT: R
10 Foundation for Statistical Computing.
11
12 Rivas-Martínez, S. (Ed.) (2011). Map of series, geoserries and geopermaseries of vegetation in Spain.
13 *Itinera Geobotanica*, 18, 5–424 (in Spanish).
14
15 Robert, N., Vidal, C., Colin, A., Hervé, J.-C., Hamza, N., & Cluzeau, C. (2012). France. In E.
16 Tomppo, T. Gschwantner, M. Lawrence & R. E. McRoberts (Eds.), *National Forest Inventories.*
17 *Pathways for Common Reporting* (pp. 207–221). London, UK: Springer.
18
19 Ronk, A., Szava-Kovats, R., & Pärtel, M. (2015). Applying the dark diversity concept to plants at the
20 European scale. *Ecography*, 38, 1015–1025.
21
22 Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge, UK: Cambridge
23 University Press.
24
25 Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C.
26 (2011). The influence of late Quaternary climate-change velocity on species endemism. *Science*,
27 334, 660–664.
28
29 Schaminée, J. H. J., Chytrý, M., Hennekens, S. M., Janssen, J. A. M., Jiménez-Alfaro, B., Knollová, I.,
30 ... Data-providers (2014). *Vegetation analysis and distribution maps for EUNIS habitats. Report*
31 *EEA/NSV/14/006*. Copenhagen, DK: European Environment Agency.
32
33 Svenning, J.-C. & Skov, F. (2004). Limited filling of the potential range in European tree species.
34 *Ecology Letters*, 7, 565–573.
35
36 Svenning, J.-C., Normand, S., & Skov, F. (2009). Plio-Pleistocene climate change and geographic
37 heterogeneity in plant diversity-environment relationships. *Ecography*, 32, 13–21.
38
39 Szymura, T. (2012). How does recent vegetation reflect previous systems of forest management?
40 *Polish Journal of Ecology*, 60, 859–886.
41
42 Tichý, L. (2002). JUICE, software for vegetation classification. *Journal of Vegetation Science*, 13,
43 451–453.
44
45 Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in
46 northern Europe. *Trends in Ecology and Evolution*, 28, 696–704.
47
48 Ujházyová, M., Ujházy, K., Chytrý, M., Willner, W., Čiliak, M., Máliš, F., & Slezák, M. (2016).
49 Diversity of beech forest vegetation in the Eastern Alps, Bohemian Massif and the Western
50 Carpathians. *Preslia*, 88, 435–457.
51
52 Ulloa Ulloa, C., Acevedo-Rodríguez, P., Beck, S., Belgrano, M. J., Bernal, R., Berry, P. E. ...
53 Jørgensen, P. M. (2017). An integrated assessment of the vascular plant species of the Americas.
54 *Science*, 358, 1614–1617.
55
56
57
58
59
60

- 1
2
3 Veselič, Ž. (Ed.), Breznikar, A., Grecc, Z., Habič, Š., Hostnik, R., Diaci, J., & Golob, A. (2008).
4 *Forest management by mimicking nature: close-to-nature forest management in Slovenia: how to*
5 *conserve forests by using them.* Ljubljana, SI: Slovenia Forest Service.
6
7
8 Wagner, V., Chytrý, M., Jiménez-Alfaro, B., Pergl, J., Hennekens, S., Biurrun, I., ... Pyšek, P. (2017).
9 Alien plant invasions in European woodlands. *Diversity and Distributions*, 23, 969–981.
10
11 Westhoff, V. & van der Maarel, E. (1978). The Braun-Blanquet approach. 2nd ed. In R. H. Whittaker
12 (Ed.), *Classification of plant communities* (2nd ed., pp. 619–729). The Hague, NL: Junk.
13
14 Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, 21, 213–251.
15
16 Wohlgenuth, T. (1998). Modelling floristic species richness on a regional scale: a case study in
17 Switzerland. *Biodiversity and Conservation*, 7, 159–177.
18
19 Wulff, E. W. (1935). Versuch einer Einteilung der Vegetation der Erde in pflanzengeographische
20 Gebiete auf Grund der Artenzahl. *Repertorium Specierum Novarum Regni Vegetabilis*, 12, 57–83.
21
22 Zelený, D., Li, C.-F., & Chytrý, M. (2010). Pattern of local plant species richness along a gradient of
23 landscape topographical heterogeneity: result of spatial mass effect or environmental shift?
24 *Ecography*, 33, 578–589.
25
26
27 Zerbe, S., Schmidt, I., & Betzin, J. (2007). Indicators for plant species richness in pine (*Pinus*
28 *sylvestris* L.) forests of Germany. *Biodiversity and Conservation*, 16, 3301–3316.
29
30
31
32

33 Biosketch

34 Martin Večeřa is a PhD student interested in biogeography and ecology, with a particular focus on
35 patterns of species diversity. His research integrates broad-scale mapping and spatial modelling with
36 community ecology.
37

38 Author contributions: M.V., J.D., J.L., B.J.-A. & M.C. conceived the ideas; M.V., J.D., J.L. & M.C.
39 designed methodology; J.L., I.B., E.A., J.A.C., A.Č., M.Ć., P.D., J.E., F.F.-G., J.-C.G., A.I., U.J., F.J.,
40 Z.K., V.R., J.S.R., J.H.J.S., U.Š., G.S., K.V., R.V., W.W., T.W. & M.C. significantly contributed the
41 data; M.V., B.J.-A., I.K., M.Ř. & M.C. prepared and processed the dataset of vegetation plots; M.V. &
42 J.D. analysed the data; M.V., J.D., J.L., B.J.-A., I.B., E.A., J.A.C., A.Č., G.C.J., P.D., J.E., F.F.-G., J.-
43 C.G., A.I., U.J., F.J., Z.K., V.R., U.Š., J.-C.S., G.S., W.W., T.W. & M.C. interpreted the results and/or
44 regional specifics in alpha-diversity patterns; M.V. & M.C. led the writing of the manuscript. All authors
45 commented on the manuscript drafts.
46
47
48
49
50
51
52
53
54

55 Appendices

56 Appendix S1 – SUPPLEMENTARY MATERIALS AND METHODS

57 Appendix S2 – SUPPLEMENTARY RESULTS
58
59
60

1
2
3
4
5
6
7
8
9
10

SUPPORTING INFORMATION

Alpha diversity of vascular plants in European forests

Journal of Biogeography

11 Martin Večeřa, Jan Divíšek, Jonathan Lenoir, Borja Jiménez-Alfaro, Idoia Biurrun, Ilona Knollová, Emiliano
12 Agrillo, Juan Antonio Campos, Andraž Čarni, Guillermo Crespo Jiménez, Mirjana Ćuk, Panayotis Dimopoulos,
13 Jörg Ewald, Federico Fernández-González, Jean-Claude Gégout, Adrian Indreica, Ute Jandt, Florian Jansen,
14 Zygmunt Kački, Valerijus Rašomavičius, Marcela Řezníčková, John S. Rodwell, Joop H. J. Schaminée,
15 Urban Šilc, Jens-Christian Svenning, Grzegorz Swacha, Kiril Vassilev, Roberto Venanzoni, Wolfgang Willner,
16 Thomas Wohlgemuth & Milan Chytrý
17
18
19
20
21
22

CONTENTS

Appendix S1 – SUPPLEMENTARY MATERIALS AND METHODS

Overview of input vegetation-plot data used	2
Species-area relationships for visualization of alpha diversity of vegetation plots	6
Description of environmental variables used as predictors	8
Correlation matrices of predictors	10
Random forests – model fitting and relative importance measures	14
Resampling procedure to reduce local oversampling	14

Appendix S2 – SUPPLEMENTARY RESULTS

Examples of fine-scale alpha-diversity maps in zoomed windows	15
Spatial correlograms	16
Supplement to the richness hotspots	20
Supplement to the relative importance of predictors	21
Maps of prediction uncertainty	27
Range of environmental conditions in plot data as covered by the grid used for prediction	28
Tests of stability of the predicted alpha-diversity patterns	29

Supplementary references	34
---------------------------------	-----------

56
57
58
59
60

Appendix S1 – SUPPLEMENTARY MATERIALS AND METHODS

Overview of input vegetation-plot data used

Table S1.1 – Contributing vegetation-plot databases: number (n) and percentage (%) of plots used in the final dataset, N = 73,134.

Name	GIVD*	Custodian	Deputy custodian	EUNIS-classified plots	
				n	%
Austrian Vegetation Database	EU-AT-001	Wolfgang Willner	-	458	0.6
Balkan Vegetation Database	EU-00-019	Kiril Vassilev	Hristo Pedashenko	1434	2.0
Croatian Vegetation Database	EU-HR-002	Željko Škvorc	Daniel Krstonošić	2201	3.0
Czech National Phytosociological Database	EU-CZ-001	Milan Chytrý	Ilona Knollová	9247	12.6
Database of Forest Vegetation in Republic of Serbia & Vegetation Database of Northern Part of Serbia (AP Vojvodina)	EU-RS-003 & EU-RS-004	Mirjana Ćuk	-	227	0.3
Dutch National Vegetation Database	EU-NL-001	Joop Schaminée	Stephan Hennekens	3659	5.0
EcoPlant	EU-FR-005	Jean-Claude Gégout	Ingrid Seynave	2665	3.6
European Mire Vegetation Database	EU-00-022	Tomáš Peterka	Martin Jiroušek	9	0.01
French National Forest Inventory**	-	-	-	9235	12.6
GVRD Vegetation Reference Database Halle	EU-DE-014	Ute Jandt	Helge Bruelheide	3450	4.7
Hellenic Natura 2000 Database	EU-GR-005	Panayotis Dimopoulos	Ioannis Tsiripidis	1770	2.4
Hellenic Woodland Database & Hellenic Beech Forests Database	EU-GR-006 & EU-GR-007	Ioannis Tsiripidis	-	467	0.6
Iberian and Macaronesian Vegetation Information System (SIVIM)	EU-00-004	Xavier Font	-	772	1.1
INBOVEG	EU-BE-002	Els De Bie	-	1286	1.8
Irish Vegetation Database	EU-IE-001	Úna FitzPatrick	Lynda Weekes	745	1.0
Lithuanian Vegetation Database	EU-LT-001	Valerijus Rašomavičius	Domas Uogintas	473	0.6
Polish Vegetation Database	EU-PL-001	Zygmunt Kącki	Grzegorz Swacha	5868	8.0
Romanian Forest Database	EU-RO-007	Adrian Indreica	Pavel Dan Turtureanu	2176	3.0
SE Europe Forest Database	EU-00-021	Andraž Čarni	-	1609	2.2
Semi-natural Grassland Vegetation Database of Latvia	EU-LV-001	Solvita Rūsiņa	-	44	0.1
SIVIM Catalonia	EU-00-004	Xavier Font	-	19	0.03

SIVIM Deciduous forests	EU-00-023	Juan Antonio Campos	Xavier Font	1213	1.7
SIVIM Floodplain forests	EU-00-024	Idoia Biurrun	Xavier Font	1052	1.4
SIVIM Sclerophyllous forests	EU-00-004	Federico Fernández-González	Xavier Font	406	0.6
Slovak Vegetation Database	EU-SK-001	Milan Valachovič	Jozef Šibík	3426	4.7
SOPHY	EU-FR-003	Henry Brisse	Patrice de Ruffray	1649	2.3
Swiss Forest Vegetation Database	EU-CH-005	Thomas Wohlgemuth	-	4281	5.9
The Nordic Vegetation Database	EU-00-018	Jonathan Lenoir	Jens-Christian Svenning	520	0.7
UK National Vegetation Classification Database	EU-GB-001	John Rodwell	-	313	0.4
Vegetation Database of Slovenia	EU-SI-001	Urban Šilc	-	1605	2.2
Vegetation Database of Habitats in the Italian Alps	EU-IT-010	Laura Casella	Pierangela Angelini	1182	1.6
Vegetation Plot Database – Sapienza University of Rome	EU-IT-011	Emiliano Agrillo	Fabio Attorre	4160	5.7
Vegetation-Plot Database of the University of the Basque Country	EU-00-011	Idoia Biurrun	Itziar García-Mijangos	1389	1.9
VegetWeb	EU-DE-013	Florian Jansen	Jörg Ewald	1969	2.7
VegItaly, the Italian Vegetation database	EU-IT-001	Roberto Venanzoni	Flavia Landucci	764	1.0
VegMV	EU-DE-001	Florian Jansen	Christian Berg	1391	1.9

* Global Index of Vegetation-plot Databases ID

** From the French National Forest Inventory database, we used only plots sampled during the growing season (from April to September) and excluded plots that could be clearly identified as plantations of non-native trees or stands heavily affected by natural disturbances (e.g., windstorm). The data we used were sampled between 2009 and 2014.

Table S1.2 – Numbers of plots (n) in the final dataset (N = 73,134) belonging to particular forest habitat types of the revised EUNIS classification as used in the European Red List of Habitats (G1 ~ Deciduous broadleaf forests, G2 ~ Sclerophyllous forests, G3 ~ Coniferous forests).

Code	Name	n
G1.1	Temperate and boreal softwood riparian woodland	1356
G1.2a	<i>Alnus</i> woodland on riparian and upland soils	3060
G1.2b	Temperate and boreal hardwood riparian woodland	1034
G1.3	Mediterranean and Macaronesian riparian woodland	661
G1.4	Broadleaved swamp woodland on non-acid peat	1885
G1.5	Broadleaved swamp woodland on acid peat	800
G1.6a	<i>Fagus</i> woodland on non-acid soils	11853
G1.6b	<i>Fagus</i> woodland on acid soils	4331
G1.7a	Temperate and submediterranean thermophilous deciduous woodland	6654
G1.7b	Mediterranean thermophilous deciduous woodland	41
G1.8	Acidophilous <i>Quercus</i> woodland	4394
G1.9a	Temperate and boreal mountain <i>Betula</i> and <i>Populus tremula</i> woodland on mineral soils	50
G1.9b	Mediterranean mountain <i>Betula</i> and <i>Populus tremula</i> woodland on mineral soils	8396
G1.Aa	<i>Carpinus</i> and <i>Quercus</i> mesic deciduous woodland	5303
G1.Ab	Ravine woodland	4677
G1.Ba	<i>Alnus cordata</i> woodland	26
G2.1	Mediterranean evergreen <i>Quercus</i> woodland	2608
G2.2	Mainland laurophyllous woodland	24
G2.5a	South-Aegean <i>Phoenix</i> grove	3
G3.1a	Temperate mountain <i>Picea</i> woodland	5254
G3.1b	Temperate mountain <i>Abies</i> woodland	2791
G3.1c	Mediterranean mountain <i>Abies</i> woodland	13
G3.2	Temperate subalpine <i>Larix</i> , <i>Pinus cembra</i> and <i>Pinus uncinata</i> woodland	565
G3.4a	Temperate continental <i>Pinus sylvestris</i> woodland	4704
G3.4b	Temperate and submediterranean montane <i>Pinus sylvestris</i> - <i>Pinus nigra</i> woodland	1026
G3.4c	Mediterranean montane <i>Pinus sylvestris</i> - <i>Pinus nigra</i> woodland	49
G3.6	Mediterranean and Balkan subalpine <i>Pinus heldreichii</i> - <i>Pinus peuce</i> woodland	44
G3.7	Mediterranean lowland to submontane <i>Pinus</i> woodland	470
G3.9a	<i>Taxus baccata</i> woodland	29
G3.9b	Mediterranean <i>Cupressaceae</i> woodland	122
G3.A	<i>Picea taiga</i> woodland	6
G3.B	<i>Pinus sylvestris</i> taiga woodland	2
G3.Da	<i>Pinus mire</i> woodland	323
G3.Db	<i>Picea mire</i> woodland	580

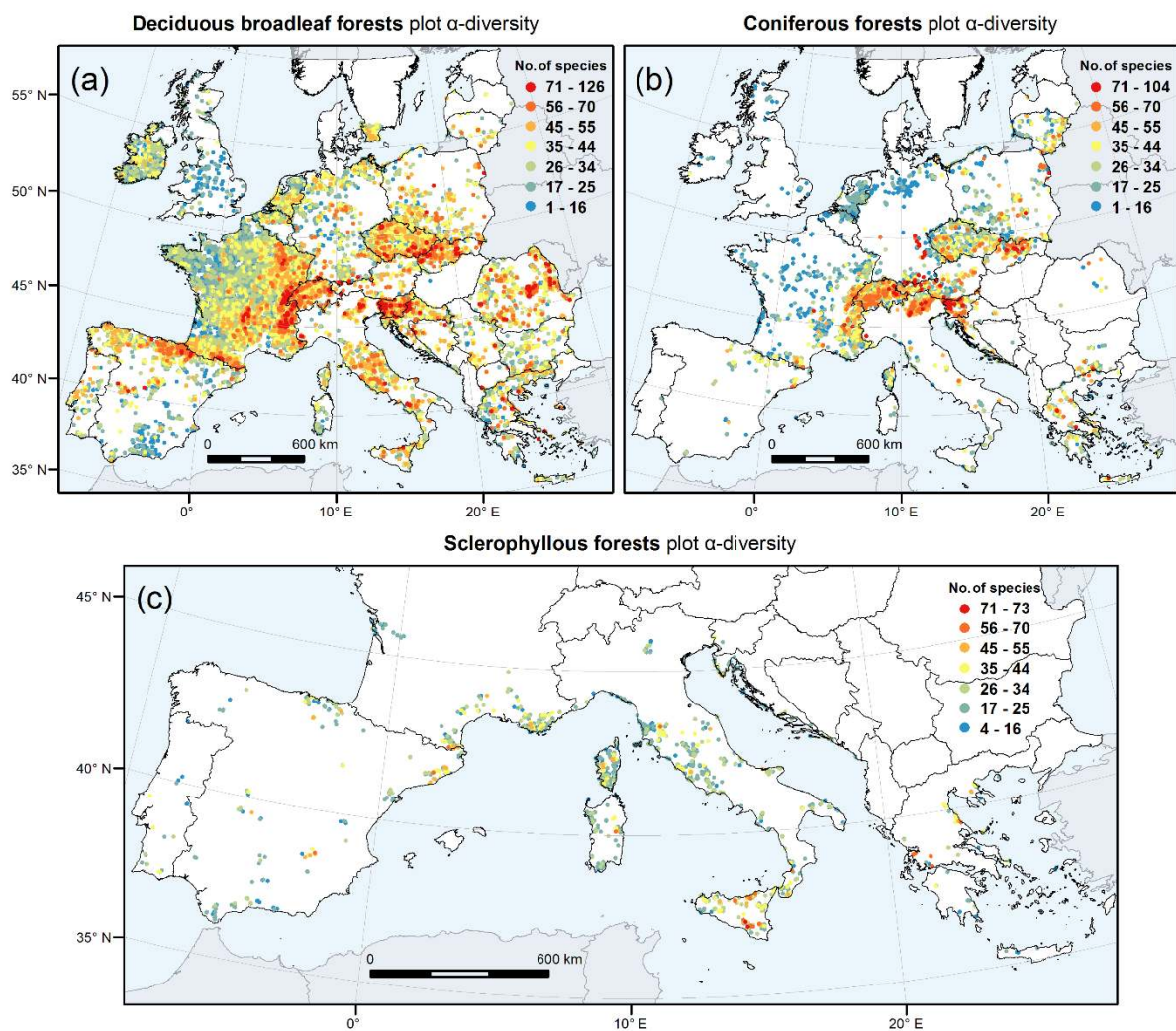
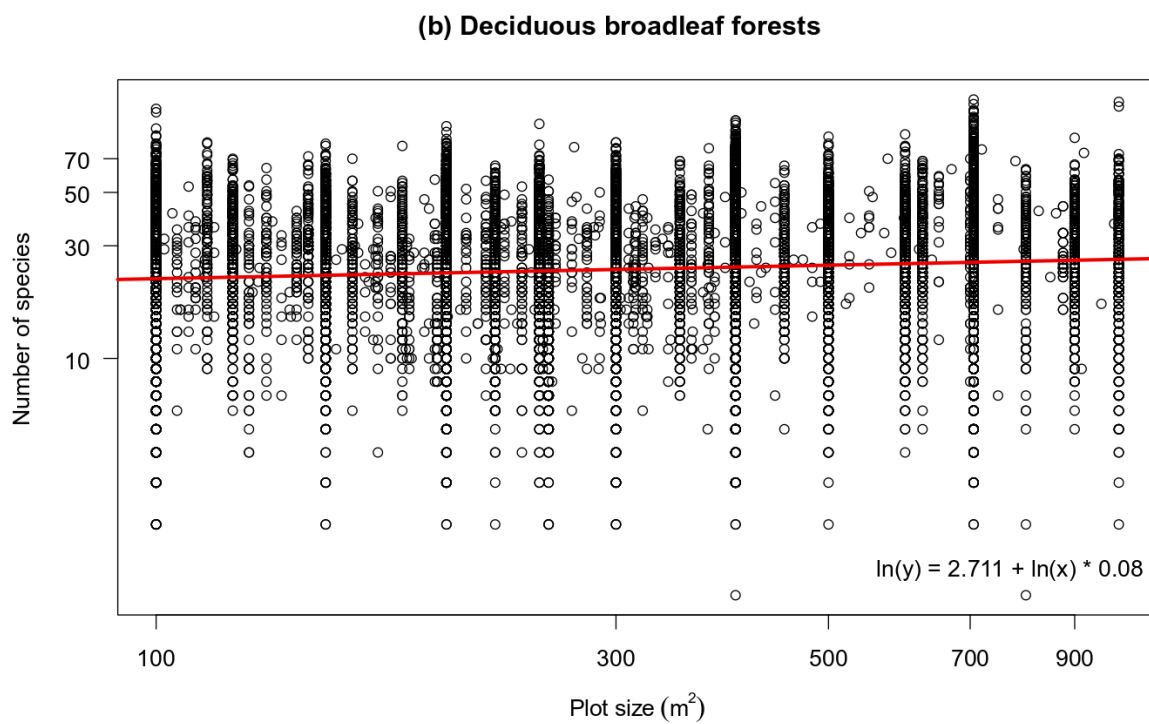
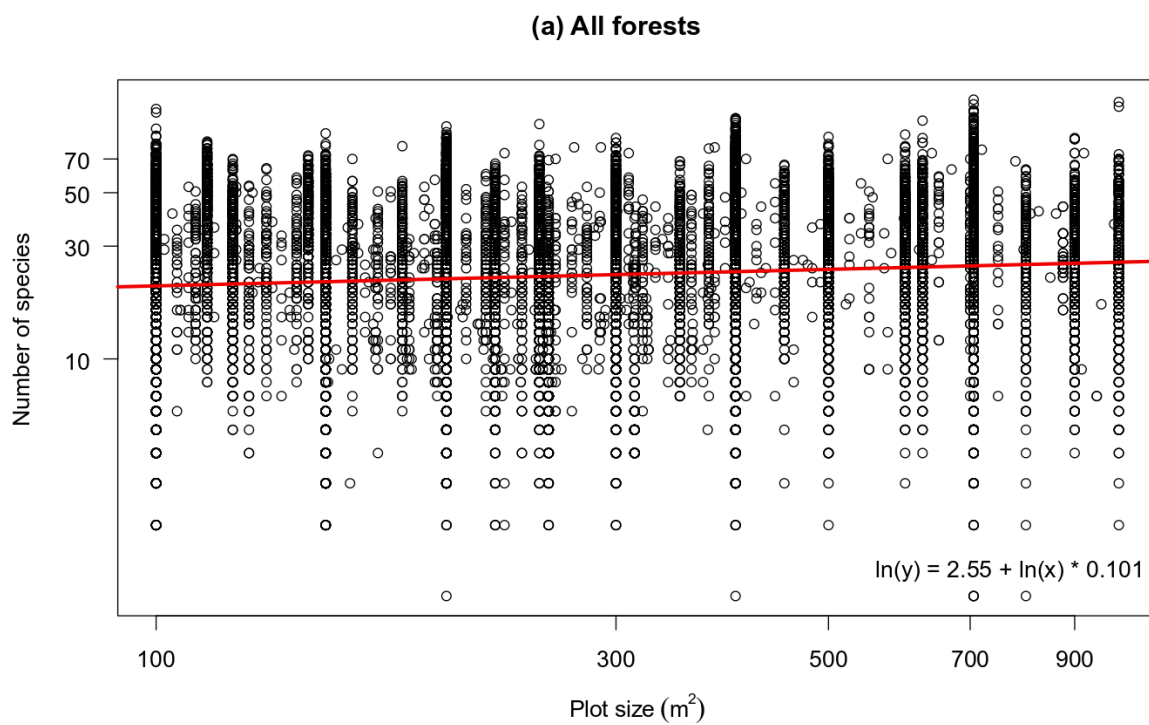
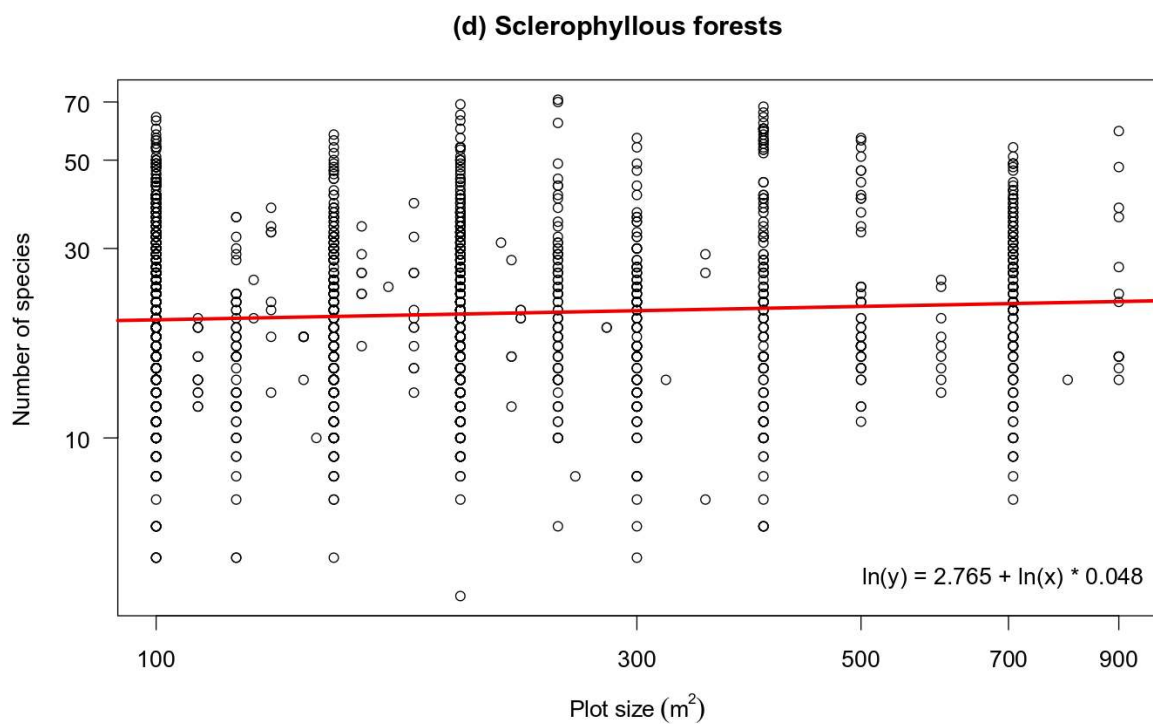
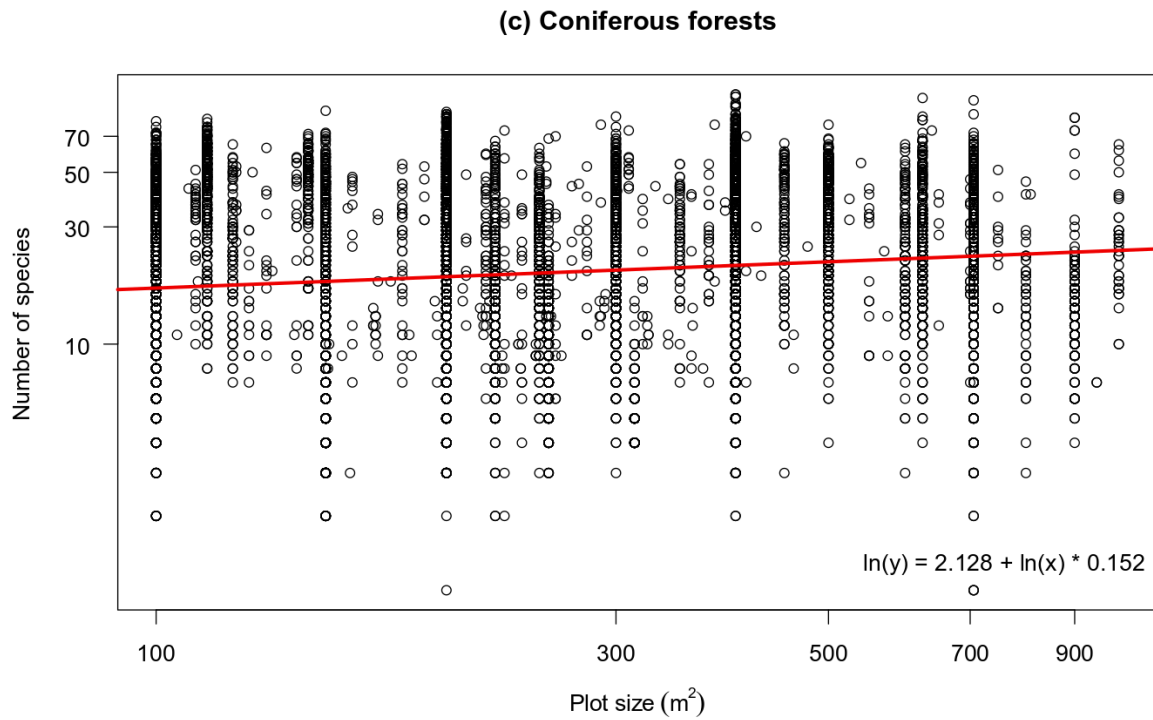


Figure S1.1 – Spatial distribution of vegetation plots, ranging from 100 to 1000 m² in surface area, used as input datasets for modelling alpha-diversity patterns in (a) deciduous broadleaf forests, $n = 54,521$; (b) coniferous forests, $n = 15,978$; and (c) sclerophyllous forests, $n = 2635$. The colours indicate the total number of vascular plant species recorded per plot (i.e., alpha diversity). For the purpose of visualization, alpha diversity values displayed on these maps were rescaled to 400 m² (the most frequent plot size within the dataset) using the species-area relationship obtained from the data (Appendix S1, Figures S1.2b, c, d). The richest plots are displayed on the top.

Species-area relationships for visualization of alpha diversity of vegetation plots



55
56
57
58
59
60

Figure S1.2 – Species-area relationships used for the visualization of alpha diversity of vegetation plots rescaled to 400 m² for all forests and for each forest type separately (see also Figure S1.1), the axes are in a logarithmic scale.

Description of environmental variables used as predictors

(i) Current environment

Topographic variables were all derived from the global digital elevation model with a spatial resolution of ~90 m at the equator (Jarvis, Reuter, Nelson, & Guevara, 2008): *elevation*; *terrain ruggedness* (calculated as a vector ruggedness measure (VRM) according to Sappington, Longshore, & Thompson, 2007); and *heat load index* (McCune & Keon, 2002; Equation 3). Terrain ruggedness was used as a proxy to capture the environmental heterogeneity occurring within a 25 km² buffer zone. Substrate variables are *topsoil pH* (measured in water) with a spatial resolution of ~4.3 km (Wieder, Boehnert, Bonan, & Langseth, 2014) and *area of limestone bedrock* derived from the European Soil Database (CEC, 2004; Panagos, Van Liedekerke, Jones, & Montanarella, 2012). Regarding the climatic variables, we extracted data from the WorldClim database (version 1.4.; Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and the Consortium for Spatial Information (CGIAR-CSI) database (<http://www.cgiar-csi.org/data>) with a spatial resolution of ~1 km at the equator: *mean annual temperature*; *temperature seasonality*; *annual precipitation*; *precipitation seasonality*; *rainfall during the vegetation period* (hereafter *summer rainfall*, the ratio of precipitation in April–September to annual precipitation – the lower values, the more Mediterranean climate); *annual potential evapotranspiration* (Zomer, Trabucco, Bossio, van Straaten, & Verchot, 2008); and *annual actual evapotranspiration* (Trabucco & Zomer, 2010).

(ii) Environmental history

This group contains the following variables: *presence/absence of a Last Glacial Maximum (LGM) glacier* (Ehlers, Gibbard, & Hughes, 2011) within at least 50% of the buffer zone surrounding each plot; *temperature change since the LGM* calculated as a mean of the differences in mean annual temperature between (i) “LGM” and “Mid-Holocene”, and (ii) “Mid-Holocene” and “Present” conditions; and *precipitation change since the LGM* calculated as a mean of the differences in annual precipitation between (i) “LGM” and “Mid-Holocene”, and (ii) “Mid-Holocene” and “Present” conditions. Past climatic conditions were available at a spatial resolution of ~1 km at the equator based on the CCSM4 model available in the WorldClim database (version 1.4.; Hijmans et al., 2005).

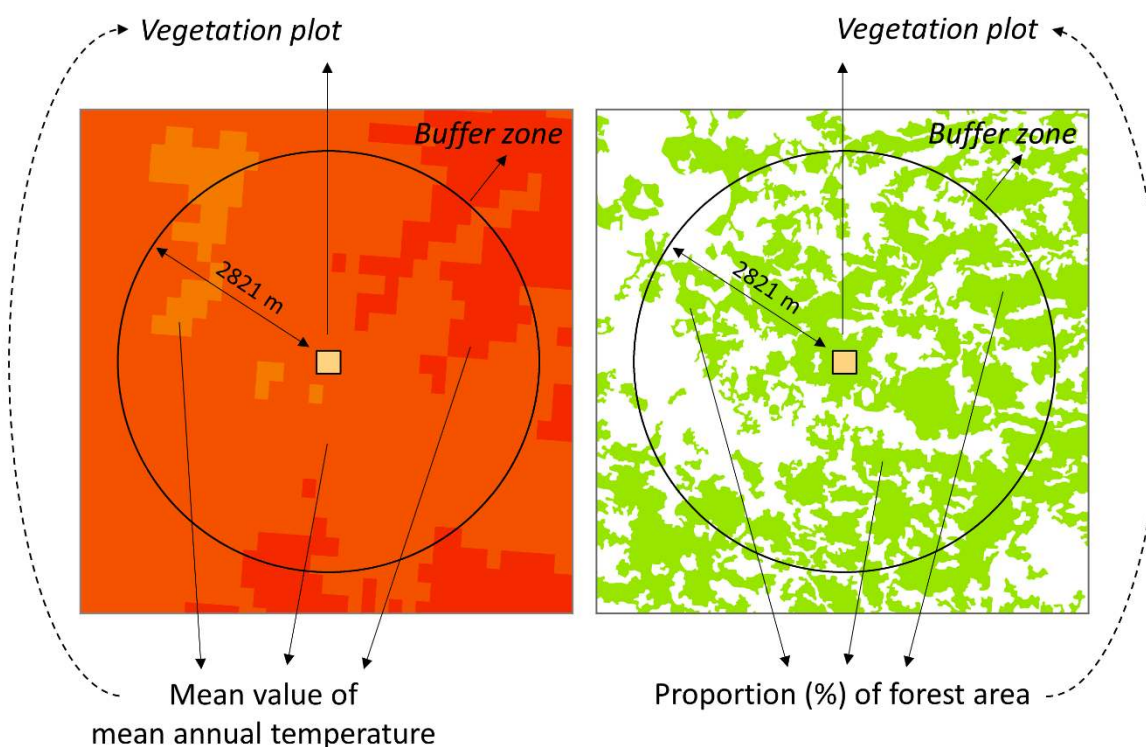


Figure S1.3 – Buffer zone (25 km²) around a vegetation plot used to calculate mean values of quantitative predictor variables and the proportion of occupancy for categorical variables. The calculated values were then assigned to the respective plot.

(iii) Land cover

Aggregated land-cover types were extracted from the CORINE 2000 Land Cover database (Bossard, Feranec, & Ouhahel, 2000): *forests* (311 broadleaf, 312 coniferous, and 313 mixed forests); *mosaic lands* (231 pastures, 321 natural grasslands, 242 complex cultivation patterns, 243 cultivated land with significant areas of natural vegetation, and 244 agro-forestry areas); *shrublands* (322 moors and heathland, 323 sclerophyllous vegetation, and 324 transitional woodland shrub); and *intensively cultivated lands* (211 non-irrigated arable land, 212 permanently irrigated land, 213 rice fields, 221 vineyards, 222 fruit trees and berry plantations, 223 olive groves, and 241 annual crops associated with permanent crops). The proportion of occupancy of these four land-cover categories within the 25 km² buffer zones was used as a predictor.

Correlation matrices of predictors

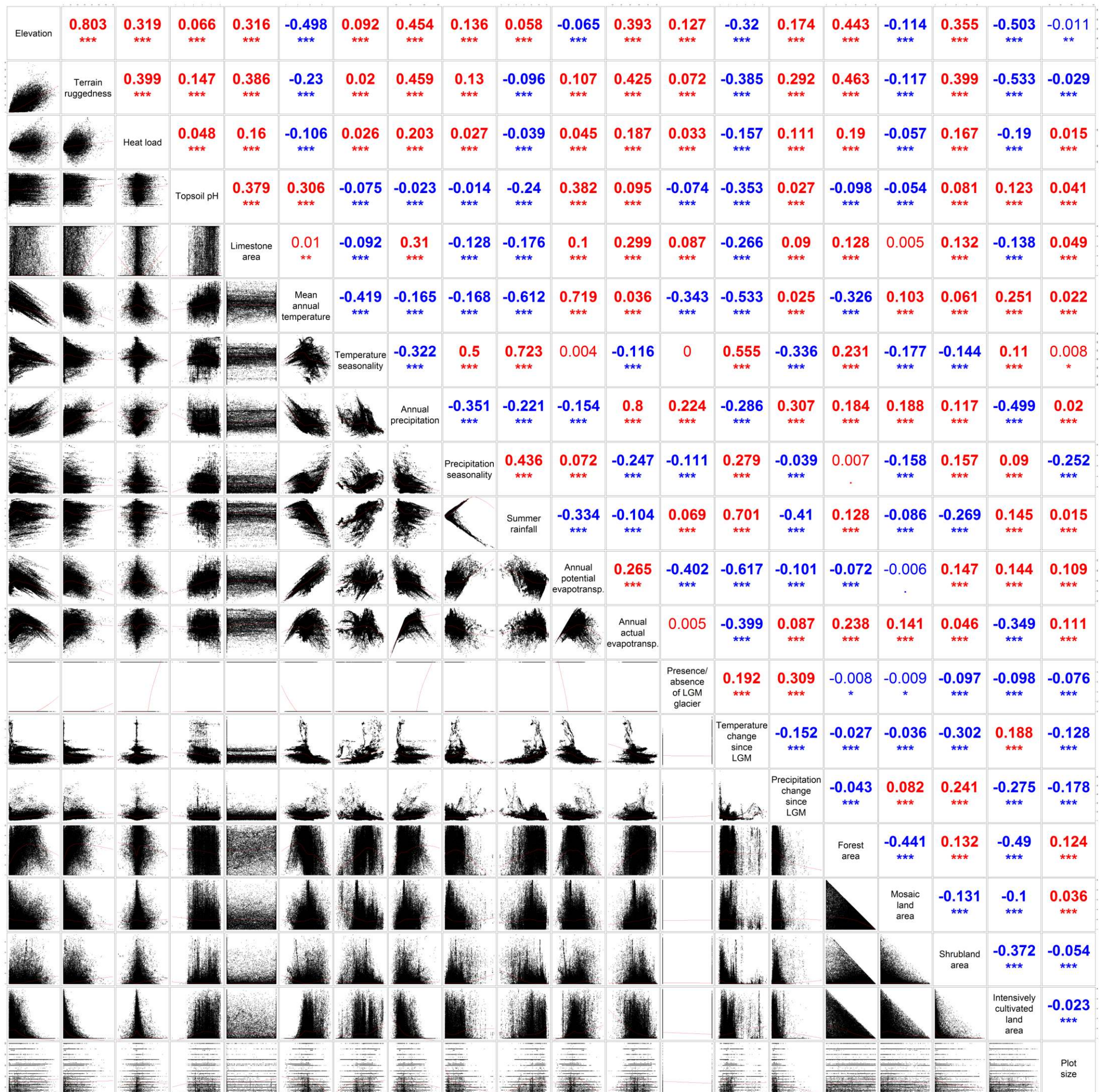


Figure S1.4 – Matrix of correlations between predictors in the dataset of all forests. Upper triangle: Spearman's correlation coefficients and their significance (p-value ‘.’ ≥ 0.05 and < 0.1 ; ‘*’ ≥ 0.01 and < 0.05 ; ‘**’ ≥ 0.001 and < 0.01 ; ‘***’ < 0.001); lower triangle: scatter plots for pairs of predictors.

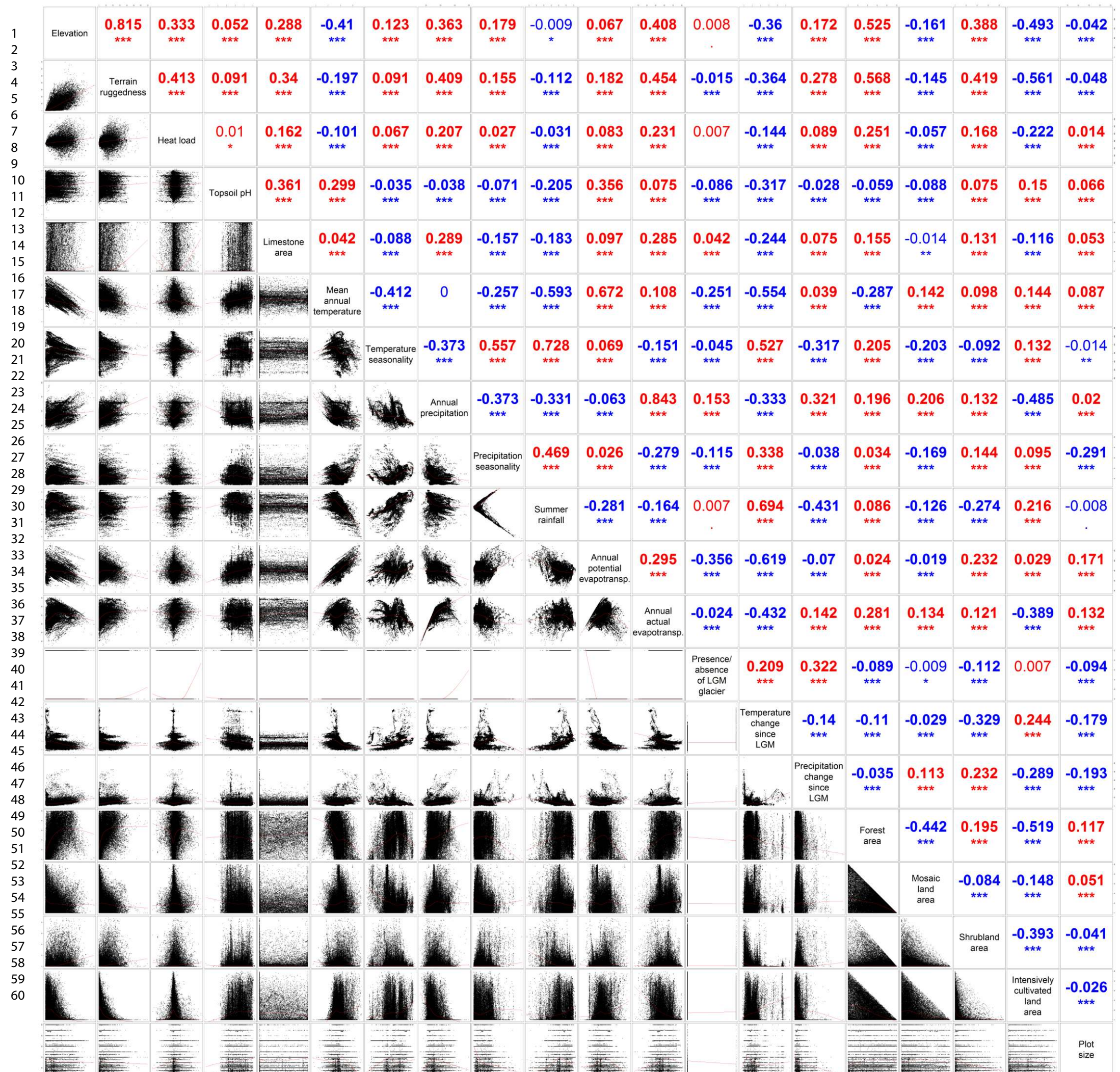


Figure S1.5 – Matrix of correlations between predictors in the dataset of deciduous broadleaf forests. Upper triangle: Spearman's correlation coefficients and their significance (p-value ‘.’ ≥ 0.05 and < 0.1 ; ‘*’ ≥ 0.01 and < 0.05 ; ‘**’ ≥ 0.001 and < 0.01 ; ‘***’ < 0.001); lower triangle: scatter plots for pairs of predictors.

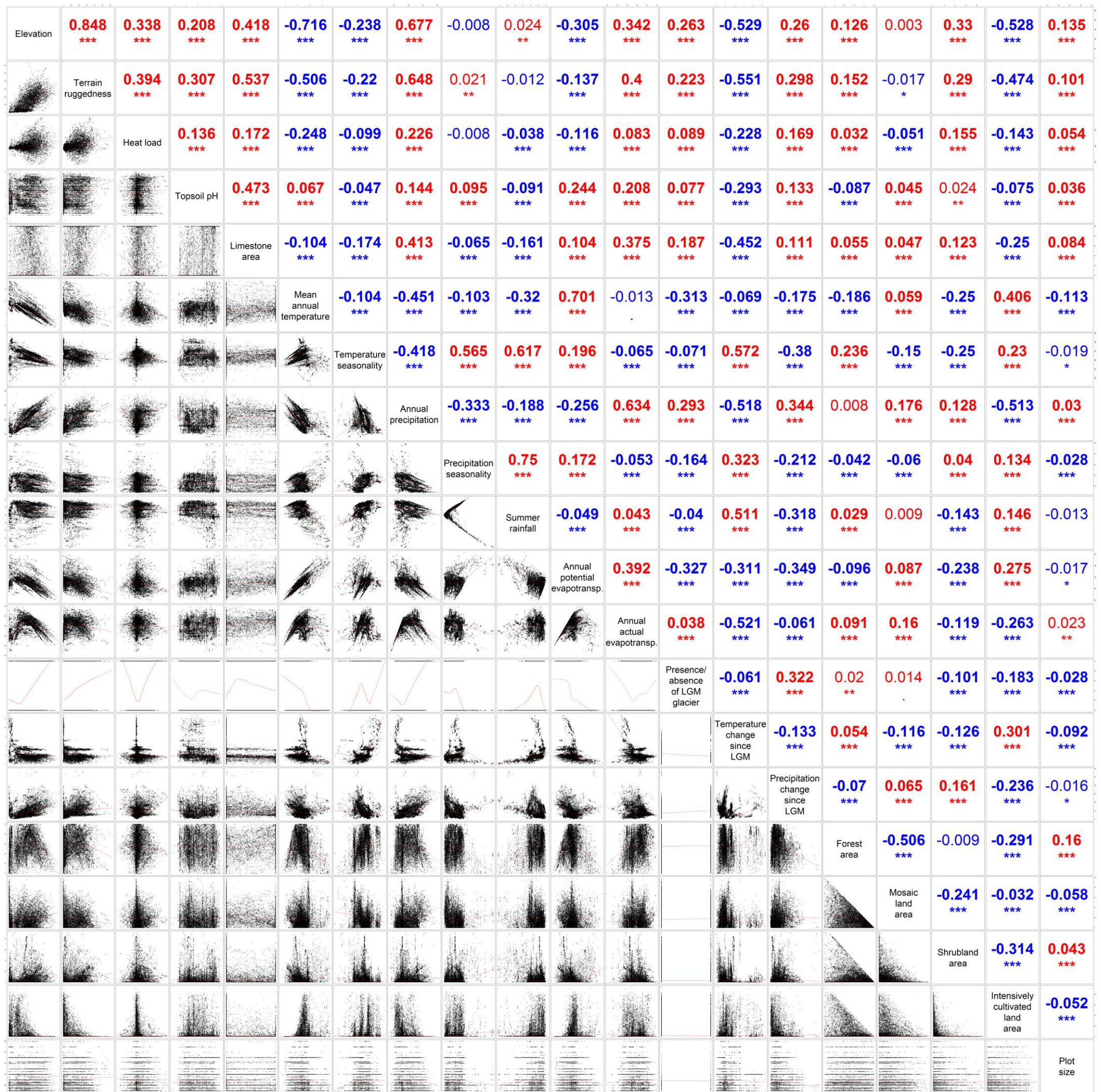


Figure S1.6 – Matrix of correlations between predictors in the dataset of coniferous forests. Upper triangle: Spearman's correlation coefficients and their significance (p-value ‘.’ ≥ 0.05 and < 0.1 ; ‘*’ ≥ 0.01 and < 0.05 ; ‘**’ ≥ 0.001 and < 0.01 ; ‘***’ < 0.001); lower triangle: scatter plots for pairs of predictors.

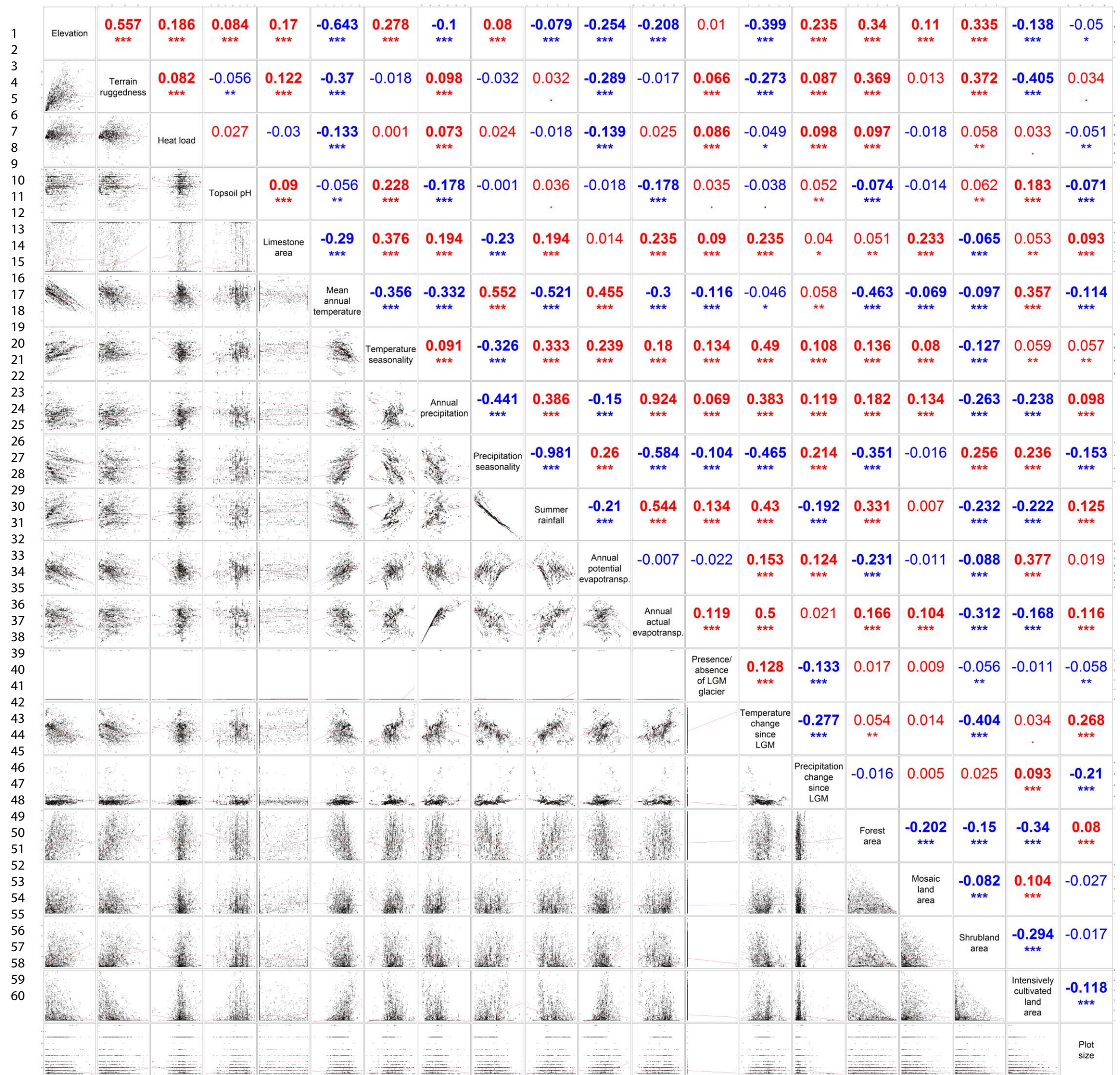


Figure S1.7 – Matrix of correlations between predictors in the dataset of sclerophyllous forests. Upper triangle: Spearman's correlation coefficients and their significance (p-value ‘.’ ≥ 0.05 and < 0.1 ; ‘*’ ≥ 0.01 and < 0.05 ; ‘**’ ≥ 0.001 and < 0.01 ; ‘***’ < 0.001); lower triangle: scatter plots for pairs of predictors.

Random forests – model fitting and relative importance measures

While fitting a Random Forest model, 500 bootstrap samples of the data, each containing 2/3 of plots in the dataset, were selected, and regression trees were grown on them. The remaining 1/3 constituted the out-of-bag (OOB) portion of the data, which were used to test the predictive power of the tree. Each regression tree was grown using m randomly selected variables to avoid high correlation among trees. To select the most appropriate m value, we tested the predictive power of the models with different settings of m (from 1 to 20), running each model ten times. These tests confirmed that default settings ($m = 1/3$ of explanatory variables) performed best in most cases. Therefore, we used the default setting of $m = 6$. The relative importance of predictors within Random Forest models is expressed by two measures. The first, called ‘Mean decrease in accuracy’, quantifies how much randomization of the target explanatory variable decreases the accuracy of the model, i.e., increases the mean of squared prediction errors (MSE) estimated from OOB data. The more randomization decreases the accuracy of the model, the more important the variable. We used 10 permutations of each variable in each model run. The second measure, called ‘Mean decrease in node impurity’, quantifies the strength of a particular predictor in terms of the decrease in residual sum of squares (RSS) between the non-split and split bulk of observations, which is averaged over all node splits and all regression trees. The relative importance of predictors was measured in 10 runs and then averaged.

Resampling procedure to reduce local oversampling

The resampling procedure uses criteria combining spatial distance among plots and their similarity in species composition. For this purpose, species nomenclature across the dataset was unified according to the SynBioSys species list (Chytrý et al., 2016) and the expert system for classification of European forests (Schaminée et al., 2014). The similarity in species composition between each pair of plots was measured using the β_{sim} index, which is independent of differences in species richness (Koleff, Gaston, & Lennon, 2003). The following criteria were used: if two plots were situated closer than 5642 m (diameter of circular buffer zones used for calculation of environmental variables), and their compositional similarity was 0.8 or higher, then one plot from the pair was selected. If more than two plots were close to each other, the algorithm started from the pair that was most similar in species composition.

We did two variants of the resampling: (i) selecting the richer plot, and (ii) selecting a random plot, both from pairs of plots similar in species composition and located close to each other.

1
2 **Appendix 2 – SUPPLEMENTARY RESULTS**
3

4 *Examples of fine-scale alpha-diversity maps in zoomed windows*
5

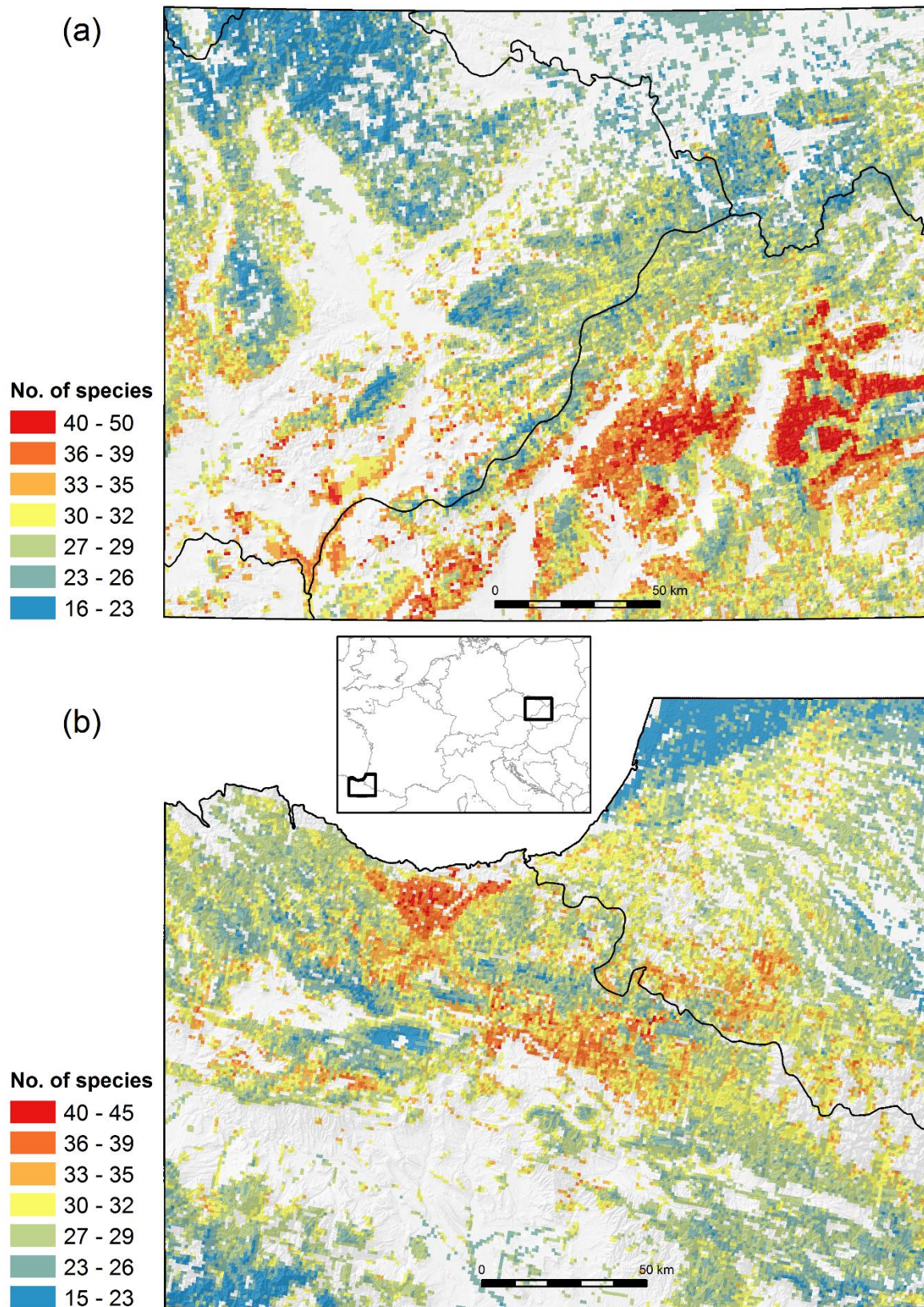
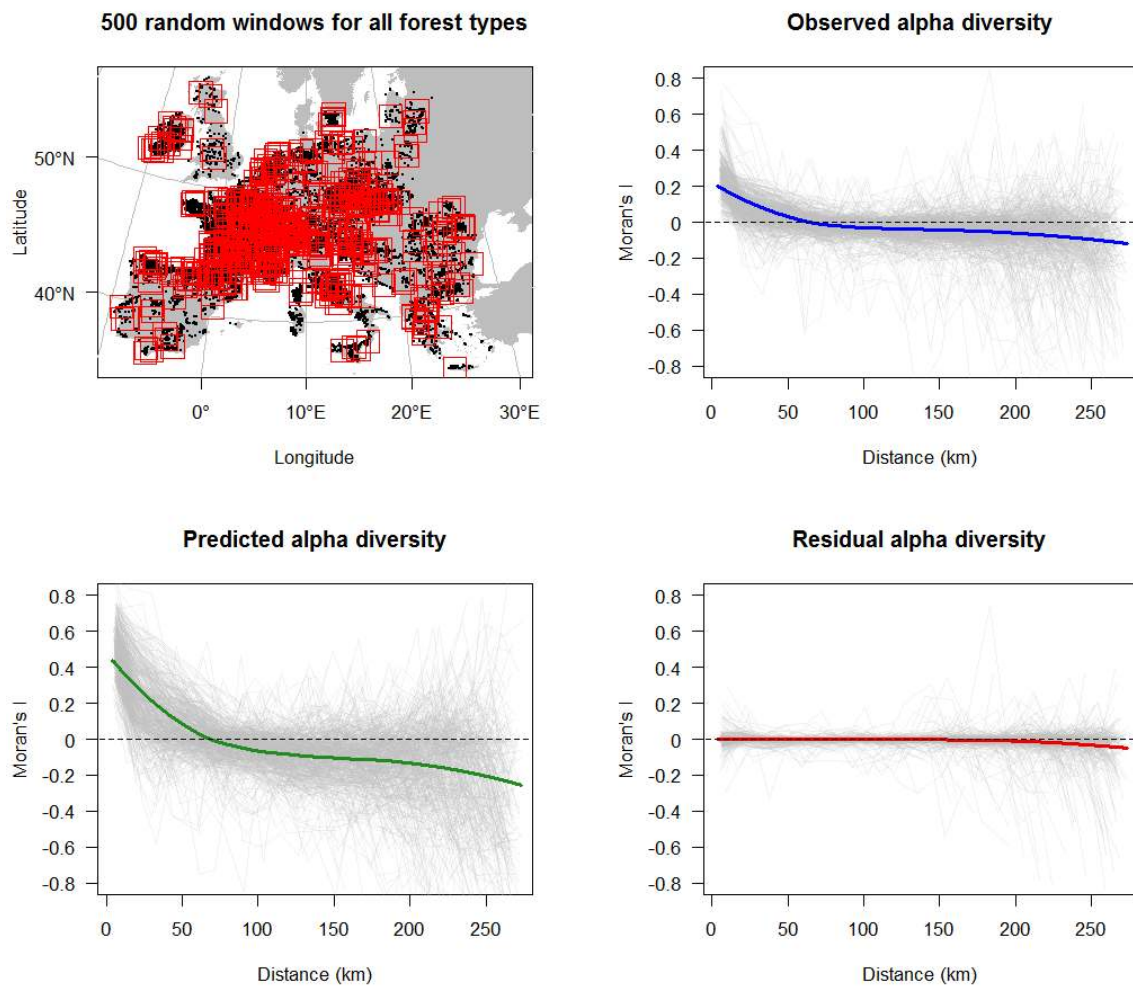


Figure S2.8 – Predicted alpha diversity of vascular plants in forests (for a plot size of 400 m²) in a spatial resolution of 1 × 1 km based on the pan-European Random Forest model for all forests displayed for two selected areas: (a) borderlands of the Czech Republic and Slovakia, and (b) north-western borderlands of Spain and France.

1
2
3 *Spatial correlograms*
4
5
6
7



40 **Figure S2.9** – Spatial correlograms of the residual, observed, and predicted alpha diversity for all forests,
41 calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled
42 regions in Europe.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

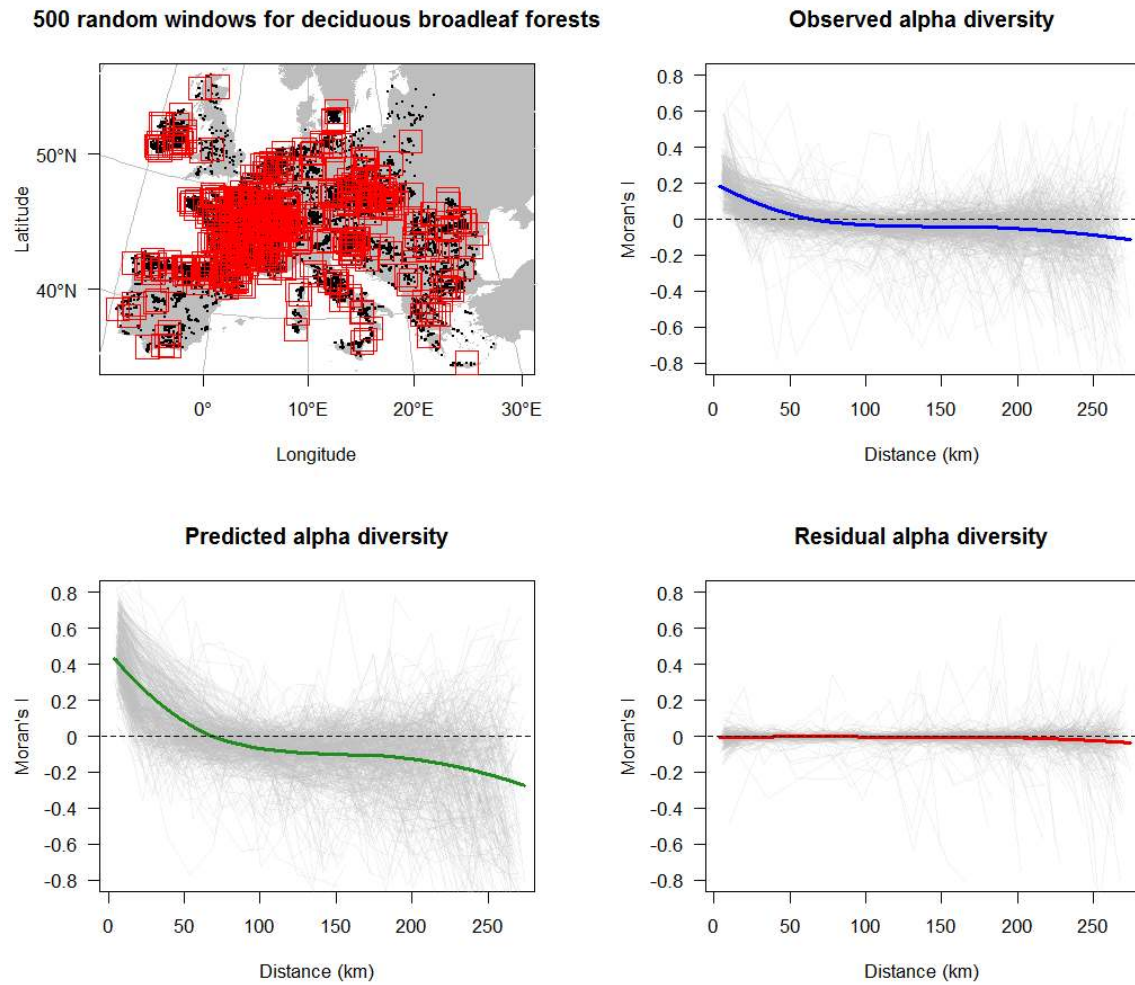


Figure S2.10 – Spatial correlograms of the residual, observed, and predicted alpha diversity for deciduous broadleaf forests, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

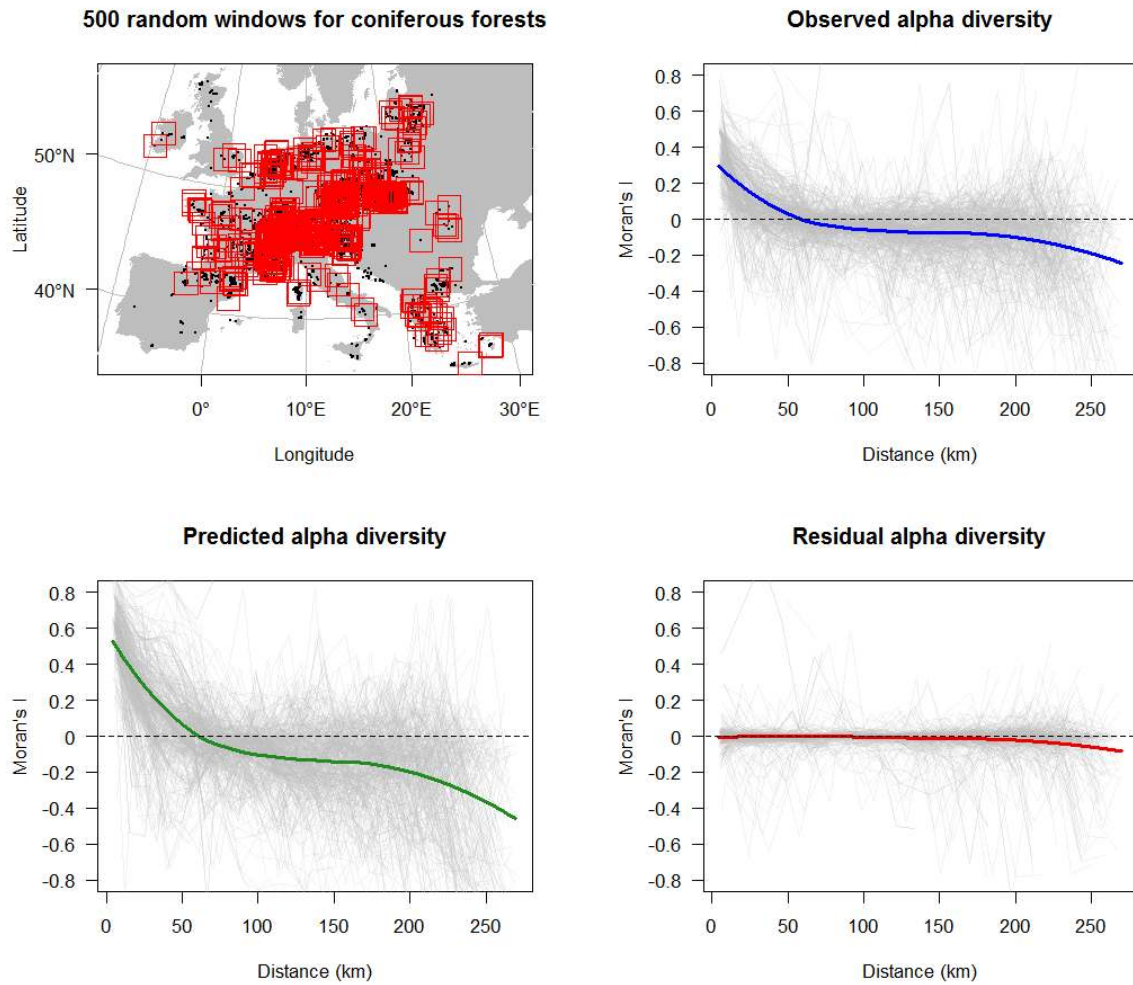


Figure S2.11 – Spatial correlograms of the residual, observed, and predicted alpha diversity for coniferous forests, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

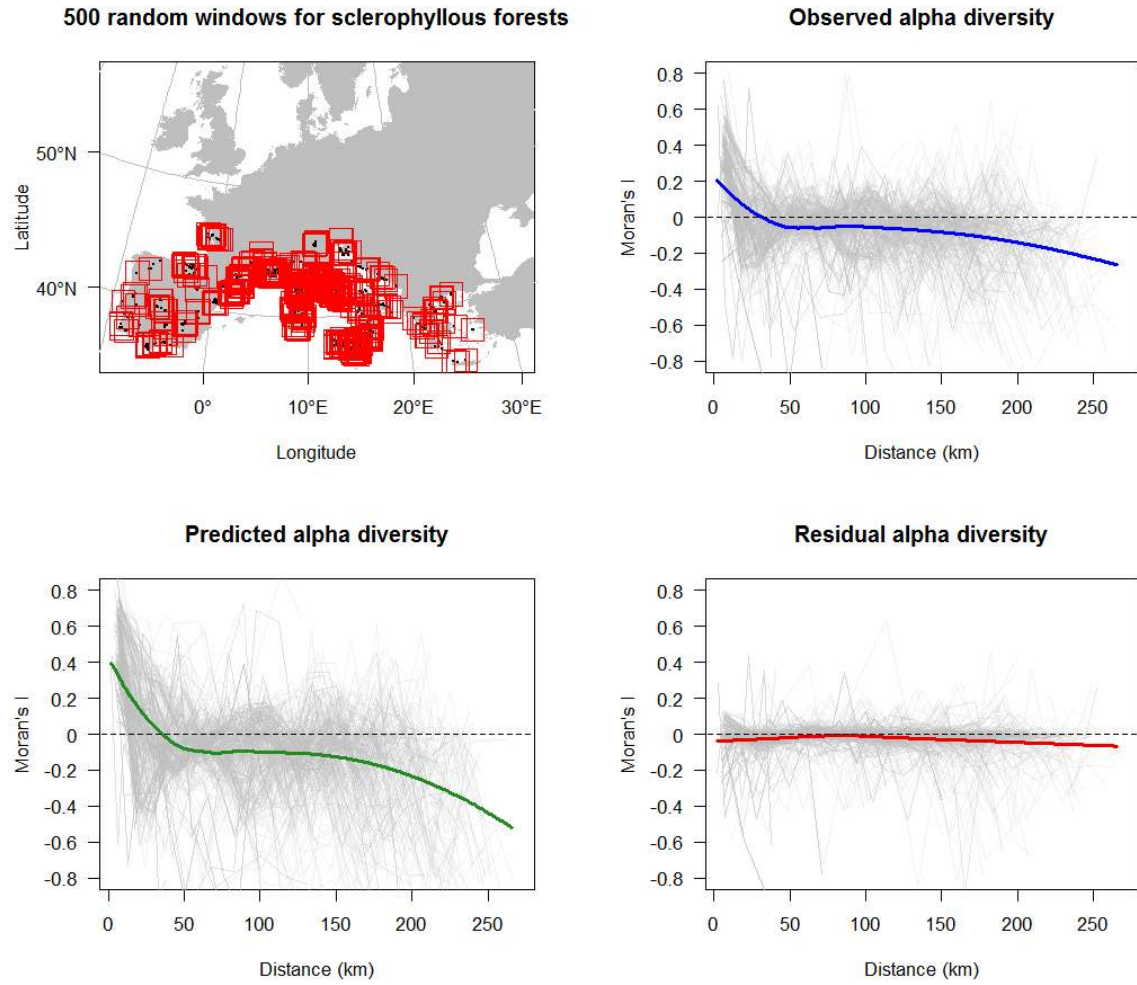


Figure S2.12 – Spatial correlograms of the residual, observed, and predicted alpha diversity for sclerophyllous forests, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

Supplement to the richness hotspots

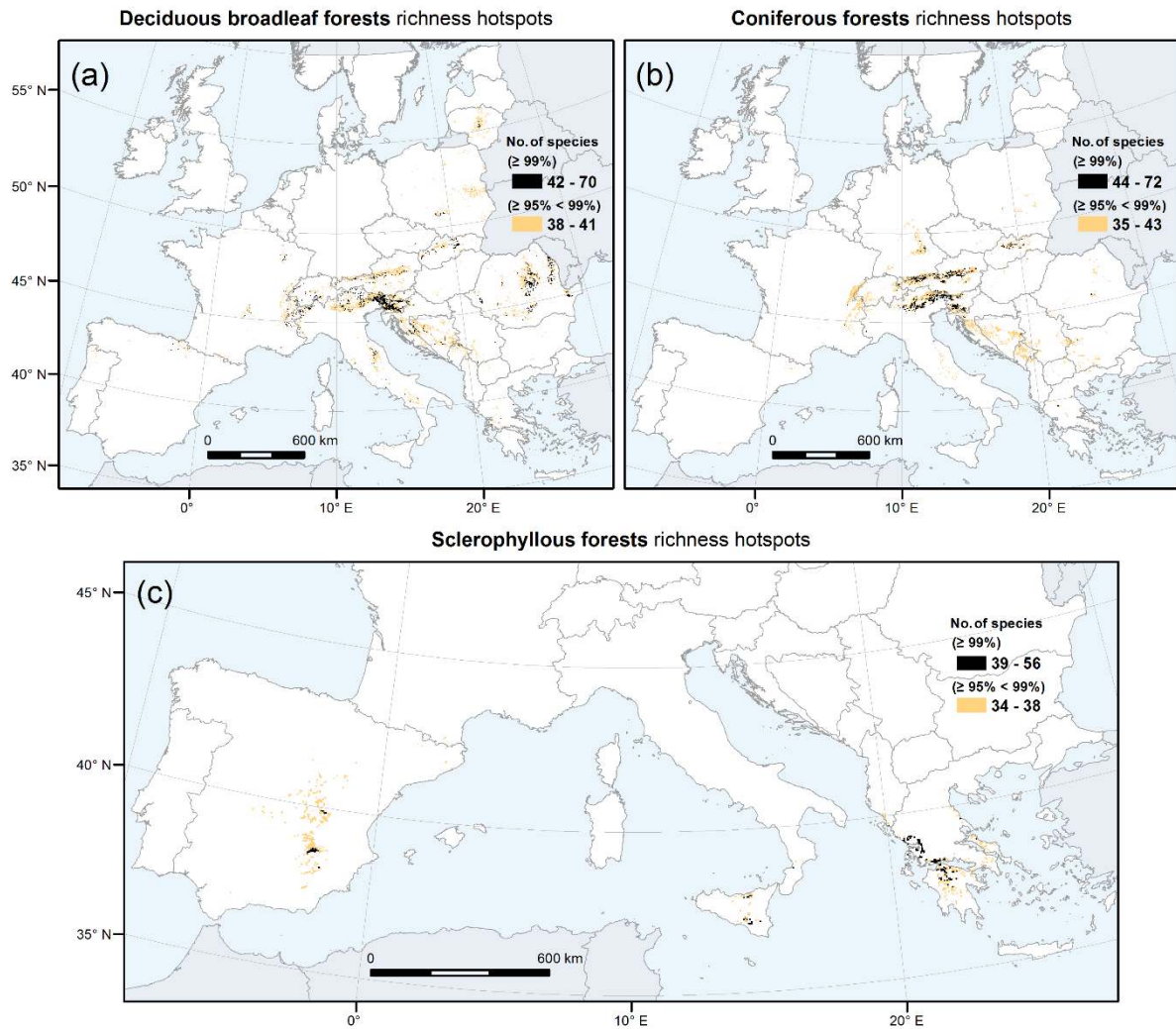
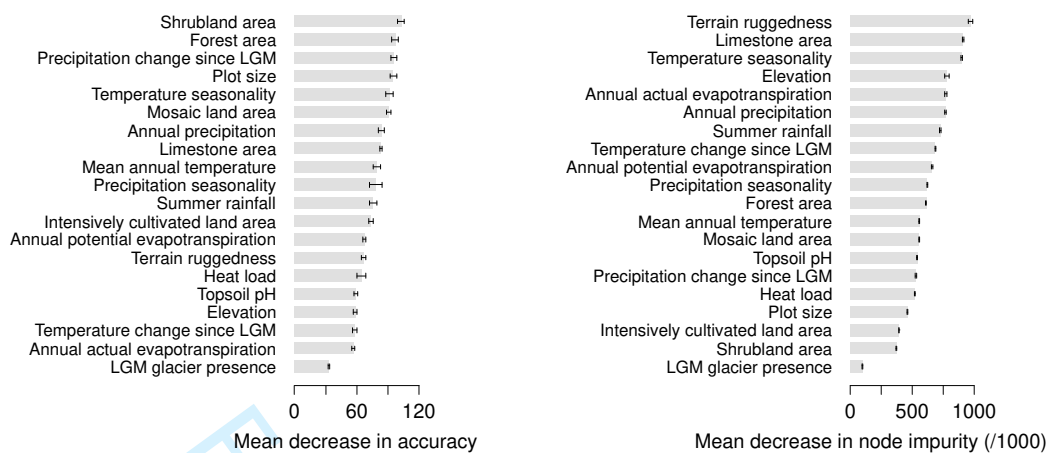


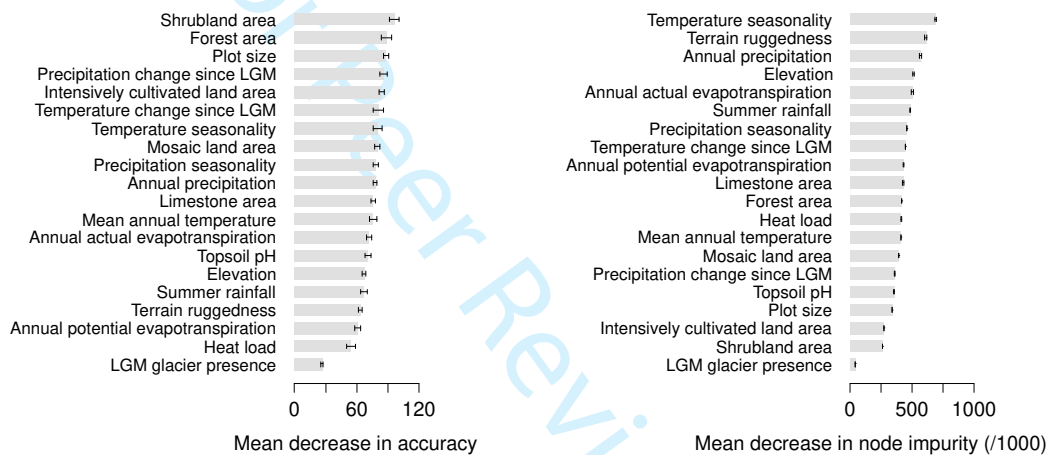
Figure S2.13 – Richness hotspots (light brown) and the top richness hotspots (black) of vascular plant alpha diversity in (a) deciduous broadleaf forests; (b) coniferous forests; and (c) sclerophyllous forests in Europe – grid cells with modelled richness values equal to or greater than the 95th and 99th percentile values, respectively.

Supplement to the relative importance of predictors

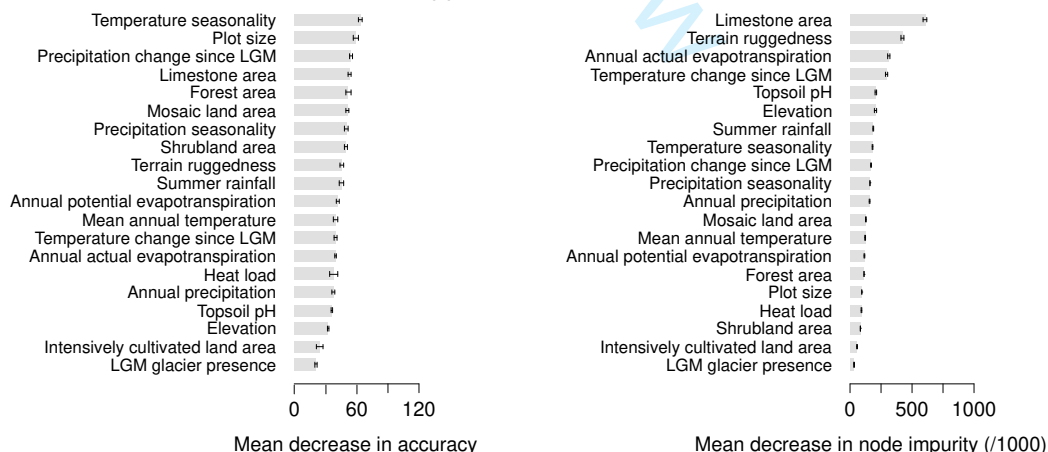
(a) All forests



(b) Deciduous broadleaf



(c) Coniferous



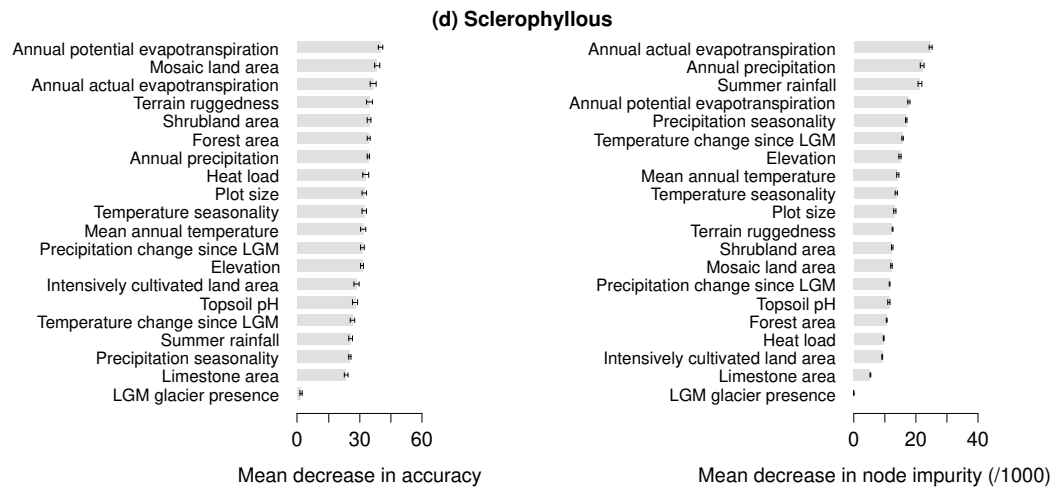


Figure S2.14 – Relative importance of predictor variables for explaining alpha-diversity pattern, measured by ‘Mean decrease in accuracy’ and ‘Mean decrease in node impurity’, for all forests and for each forest type separately. LGM – Last Glacial Maximum.

Figures S2.15 – S2.18

Effects of predictor variables (ranked on 4th to 20th place according to the integrated importance score) on alpha diversity in European forests depicted with partial dependence plots. The integrated importance scores are shown in parentheses. AET – actual evapotranspiration; LGM – Last Glacial Maximum; PET – potential evapotranspiration; VRM – vector ruggedness measure.

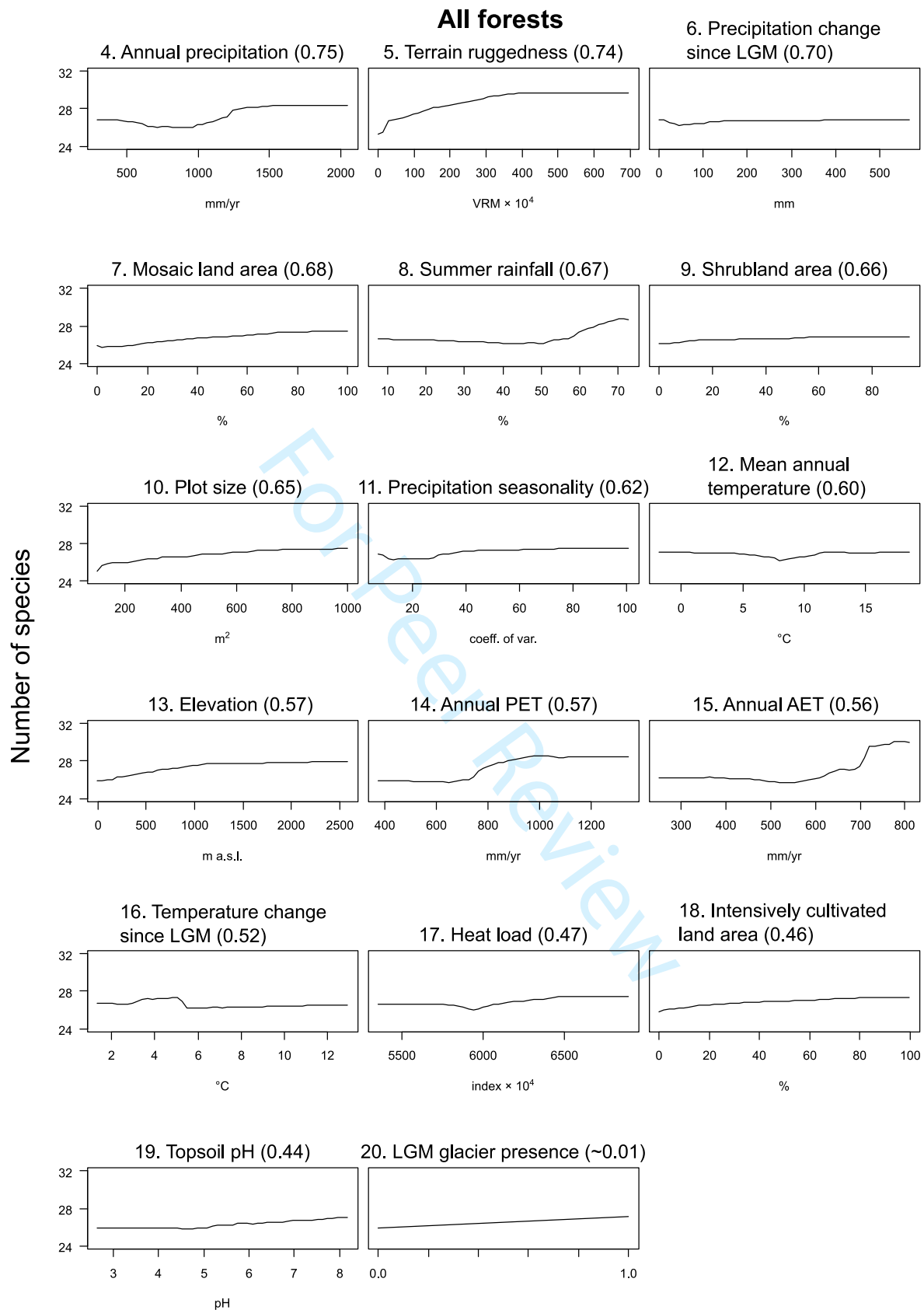


Figure S2.15 – Effects of predictor variables on alpha diversity modelled for all forests.

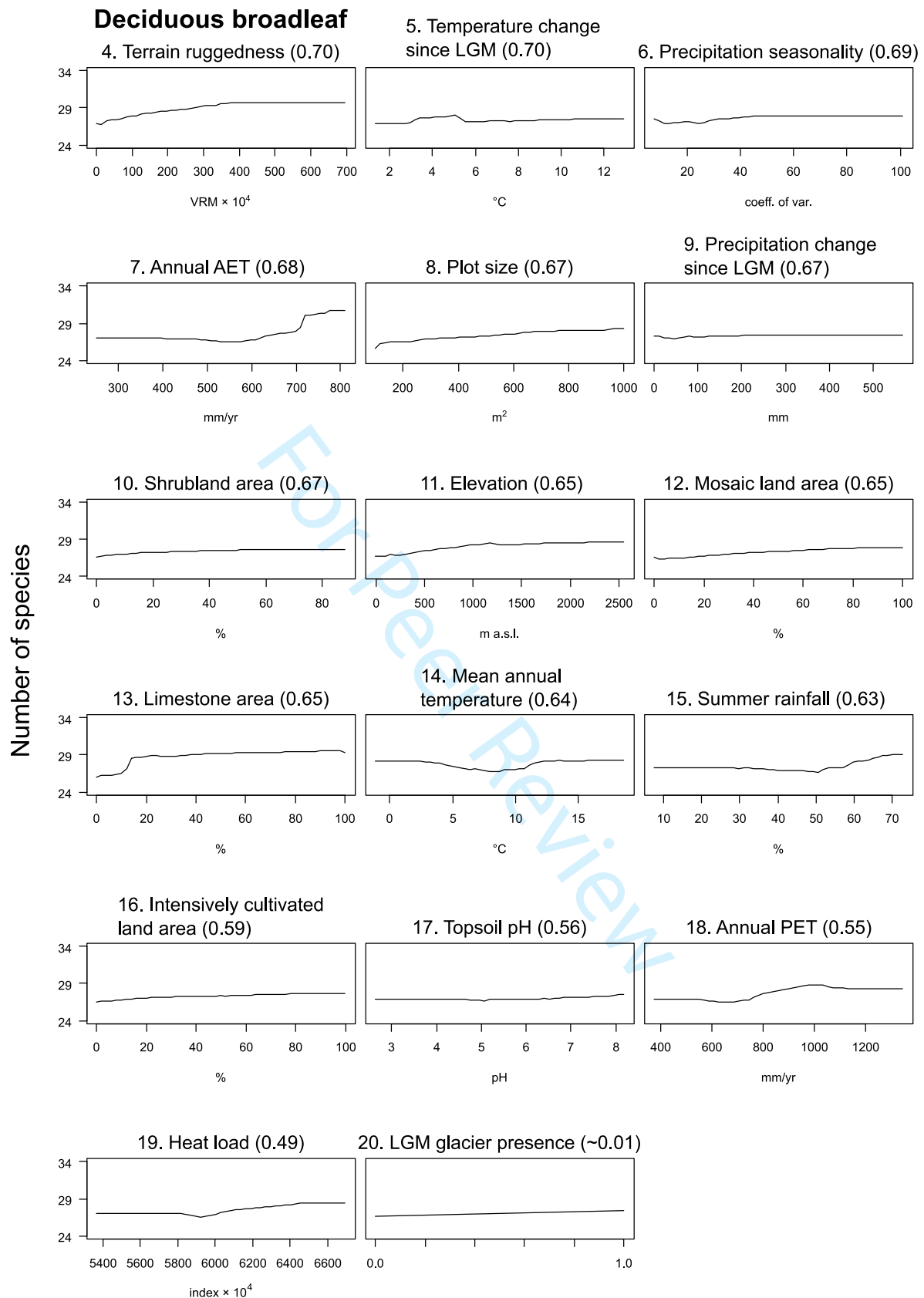


Figure S2.16 – Effects of predictor variables on alpha diversity modelled for deciduous broadleaf forests.

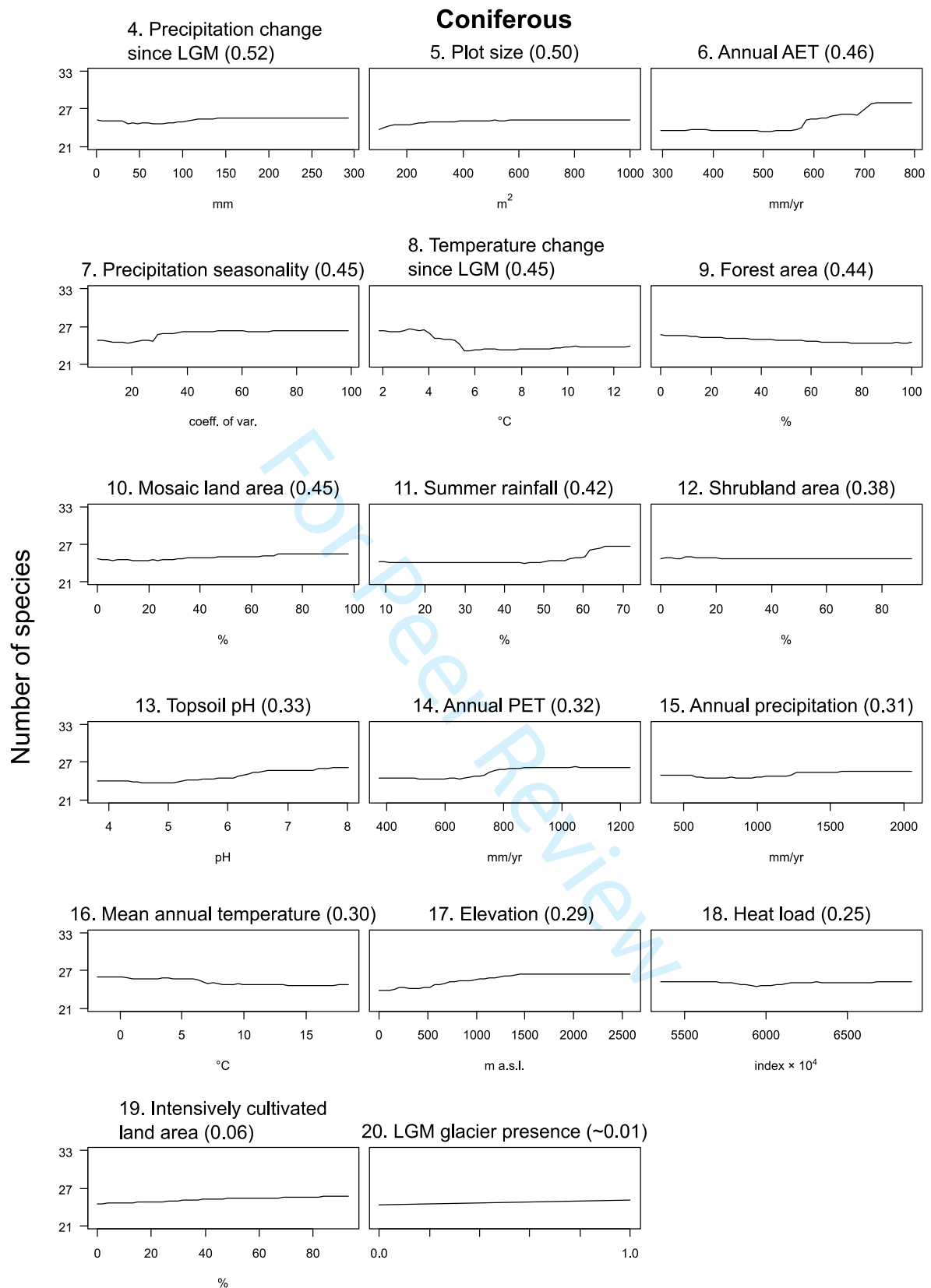


Figure S2.17 – Effects of predictor variables on alpha diversity modelled for coniferous forests.

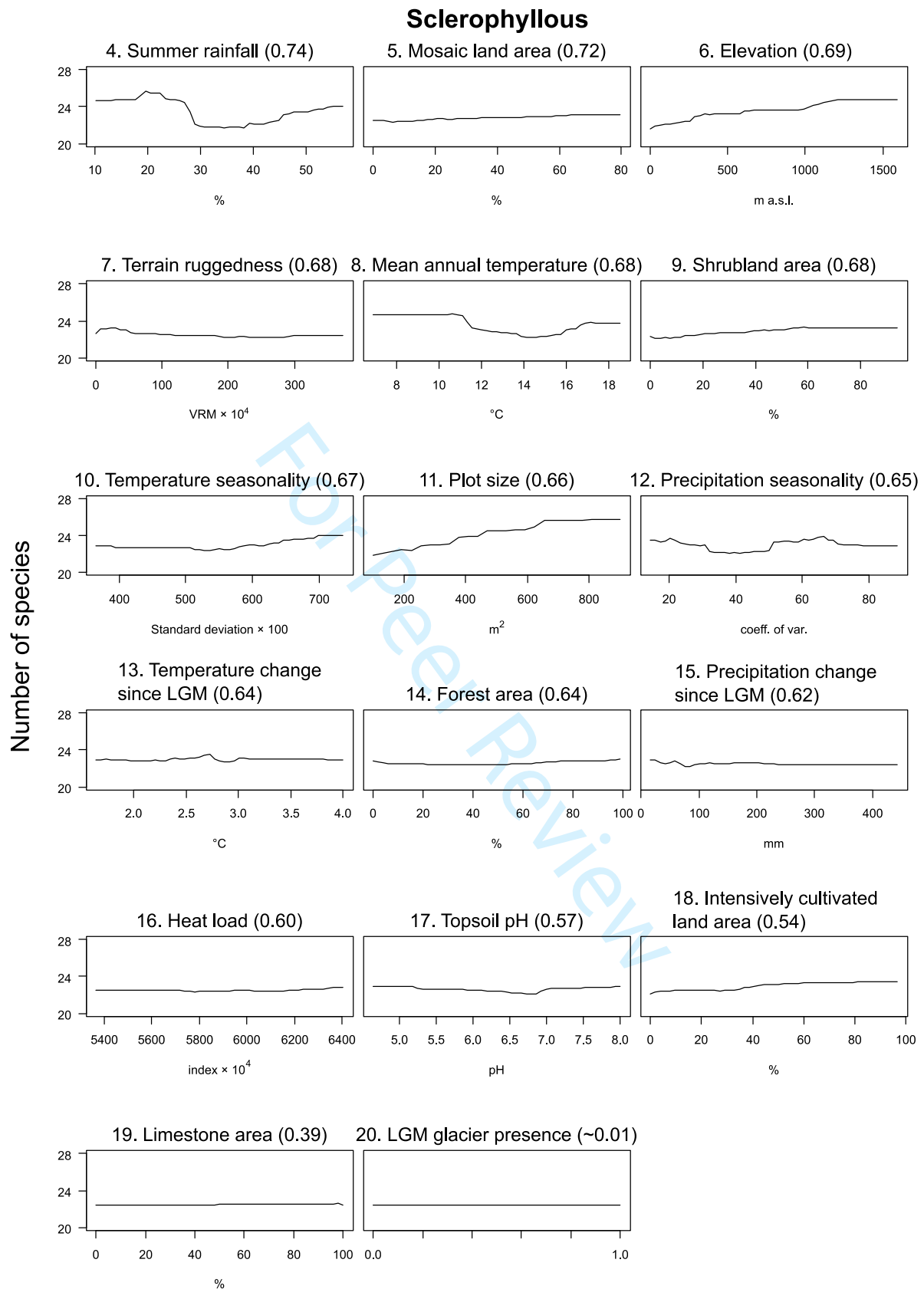
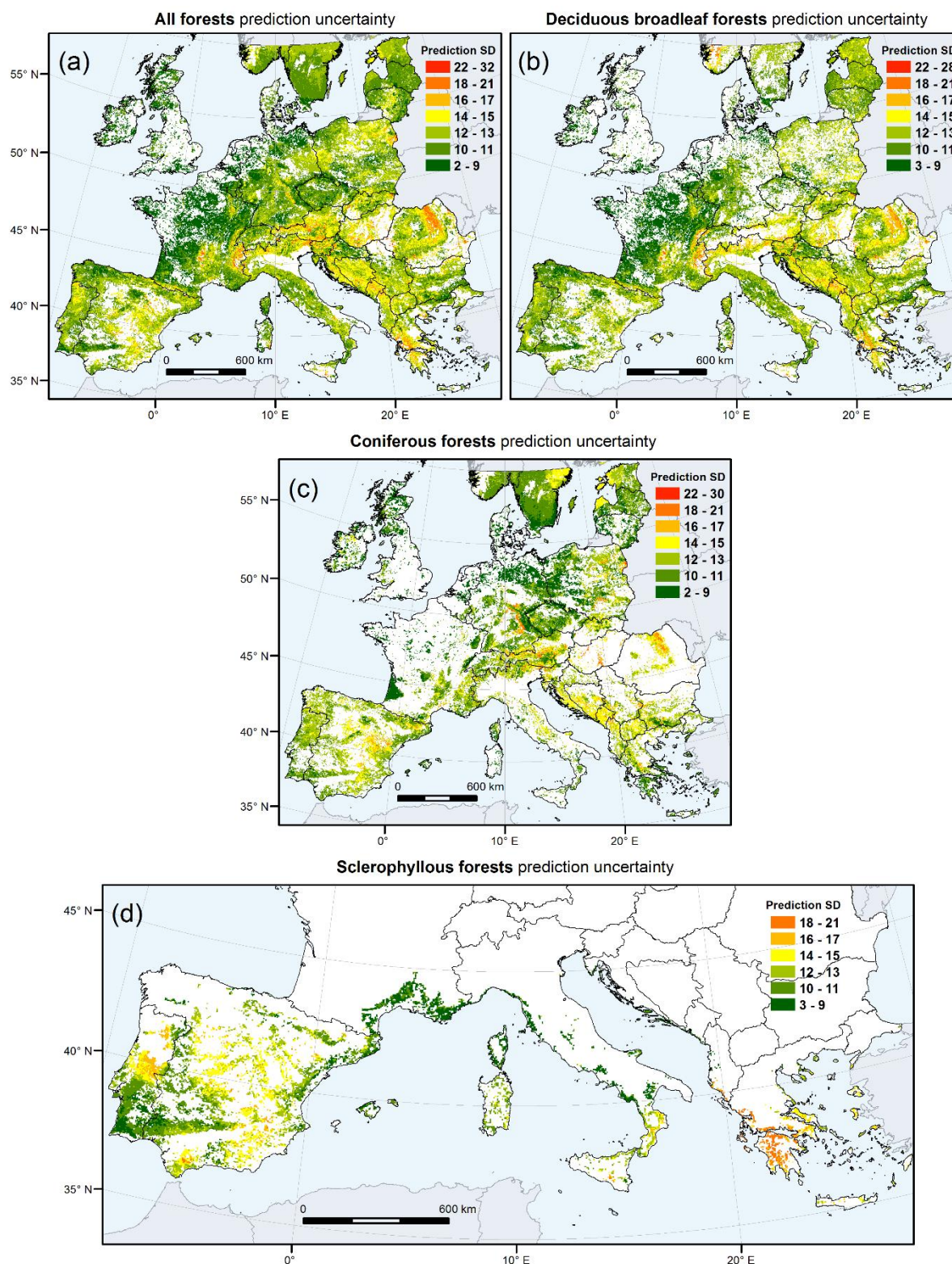


Figure S2.18 – Effects of predictor variables on alpha diversity modelled for sclerophyllous forests.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

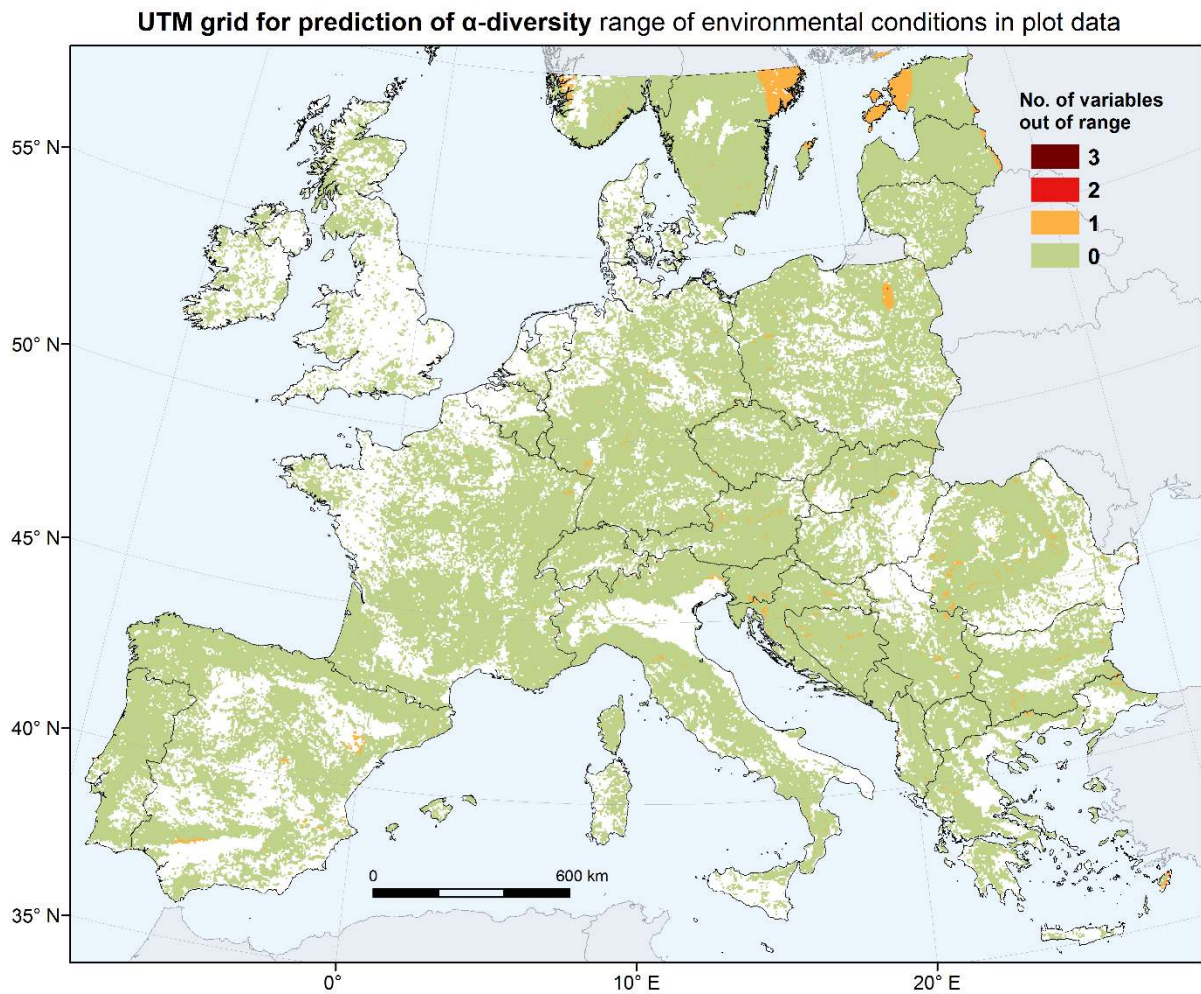
Maps of prediction uncertainty



55
56
57
58
59
60

Figure S2.19 – Uncertainty of the predicted alpha diversity in (a) all forests, (b) deciduous broadleaf forests, (c) coniferous forests, and (d) sclerophyllous forests depicted as the standard deviation of predictions of individual regression trees ($n = 500$) within respective Random Forest models. Grid cells with forest area $<10\%$ (according to the CORINE Land Cover database; Bossard et al., 2000) are left white. See Figure 2 for details.

1
2 *Range of environmental conditions in plot data as covered by the grid used for prediction*
3
4
5



38 **Figure S2.20** – UTM grid cells of 5×5 km used for the prediction of alpha diversity. Colours indicate the number
39 of explanatory variables of which values calculated in a particular cell are out of the range of conditions covered
40 by the vegetation-plot data used to model alpha-diversity patterns. Grid cells with forest area <10% (according
41 to the CORINE Land Cover database; Bossard et al., 2000) are left white.
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tests of stability of the predicted alpha-diversity patterns

Table S2.3 – Results of Random Forest models and spatial autocorrelation in model residuals for all forests based on four data subsets representing different resampling strategies: (a) non-resampled (original dataset), (b) resampled with the preference for richer plots, (c) resampled with randomly selected plots, and (d) undersampled. See also *Resampling procedure to reduce local oversampling* in Appendix S1.

	(a) <i>Non-resampled</i>	(b) <i>Resampled – richer</i>	(c) <i>Resampled – random</i>	(d) <i>Undersampled</i>
No. of cases (vegetation plots)	73,134	54,086	53,187	48,756
Explained variation [%]	56.3	48.5	47.4	51.3
Mean of squared residuals	80.2	89.5	89.6	84.8
Moran's <i>I</i> at 1 st lag distance (\pm SD)	-0.004 \pm 0.04	-0.005 \pm 0.05	-0.001 \pm 0.05	-0.006 \pm 0.04

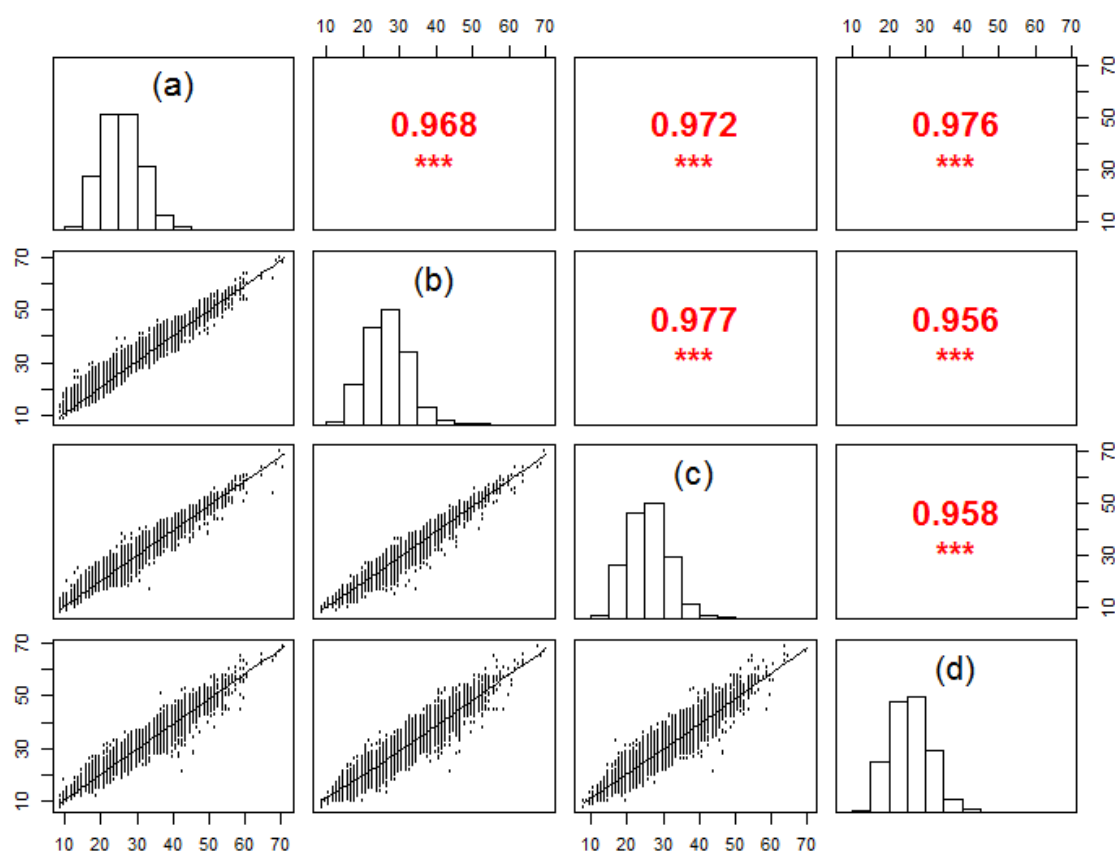
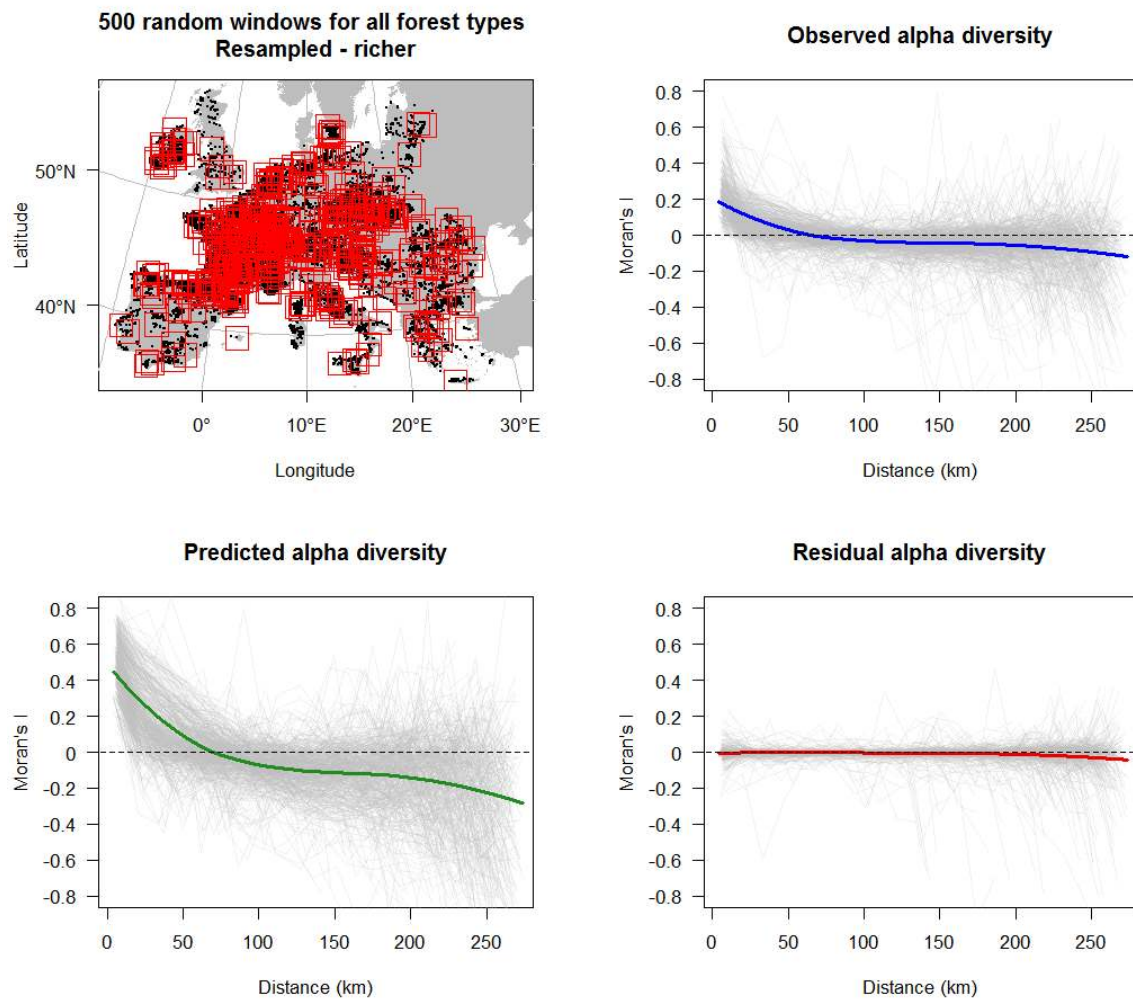


Figure S2.21 – Matrix of correlations between alpha-diversity patterns predicted by models based on four data subsets representing different resampling strategies: (a) non-resampled (original dataset), (b) resampled with the preference for richer plots, (c) resampled with randomly selected plots, and (d) undersampled. Upper triangle: Spearman's correlation coefficients and their significance (p-value ' ≥ 0.05 and < 0.1 '; '*' ≥ 0.01 and < 0.05 '; '**' ≥ 0.001 and < 0.01 '; '***' < 0.001); lower triangle: scatter plots for pairs of variables.



35
36
37
38
39

Figure S2.22 – Spatial correlograms of the residual, observed, and predicted alpha diversity for all forests based on the “Resampled – richer” data subset, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For spatial correlograms for the “Non-resampled” dataset, see Figure S2.9.

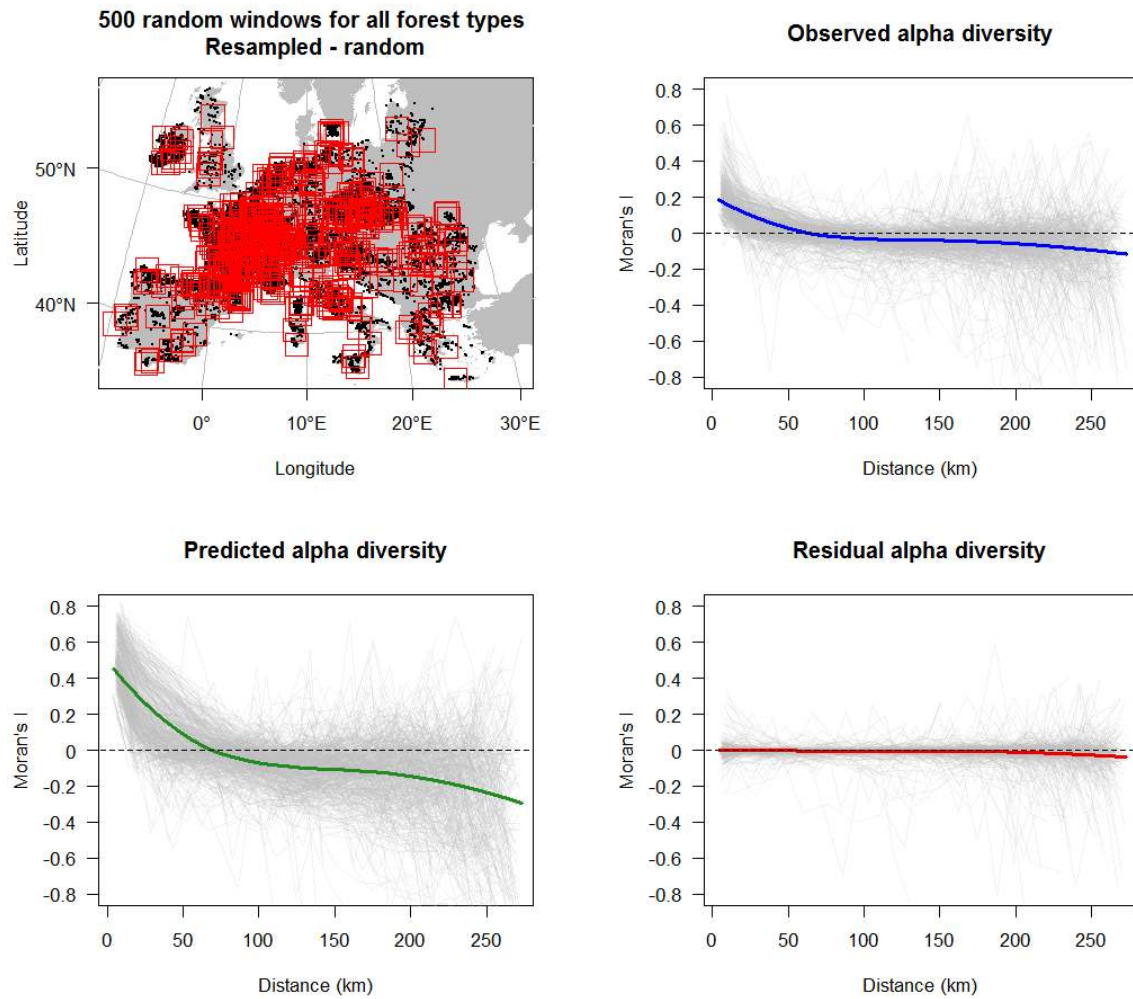


Figure S2.23 – Spatial correlograms of the residual, observed, and predicted alpha diversity for all forests based on the “Resampled – random” data subset, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

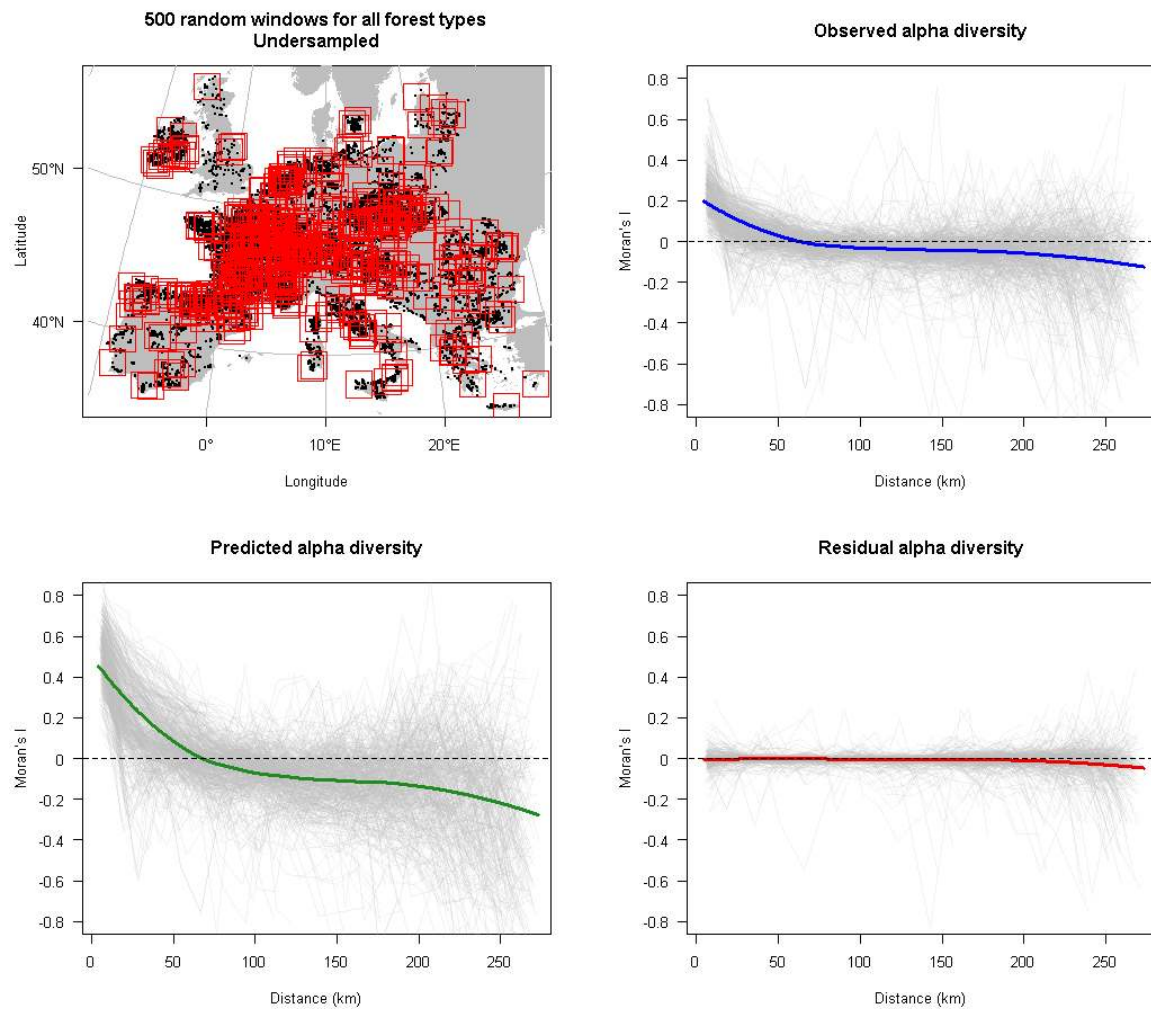


Figure S2.24 – Spatial correlograms of the residual, observed, and predicted alpha diversity for all forests based on the “Undersampled” data subset, calculated within 500 rectangular windows (a side length of 200 km), randomly distributed across the sampled regions in Europe.

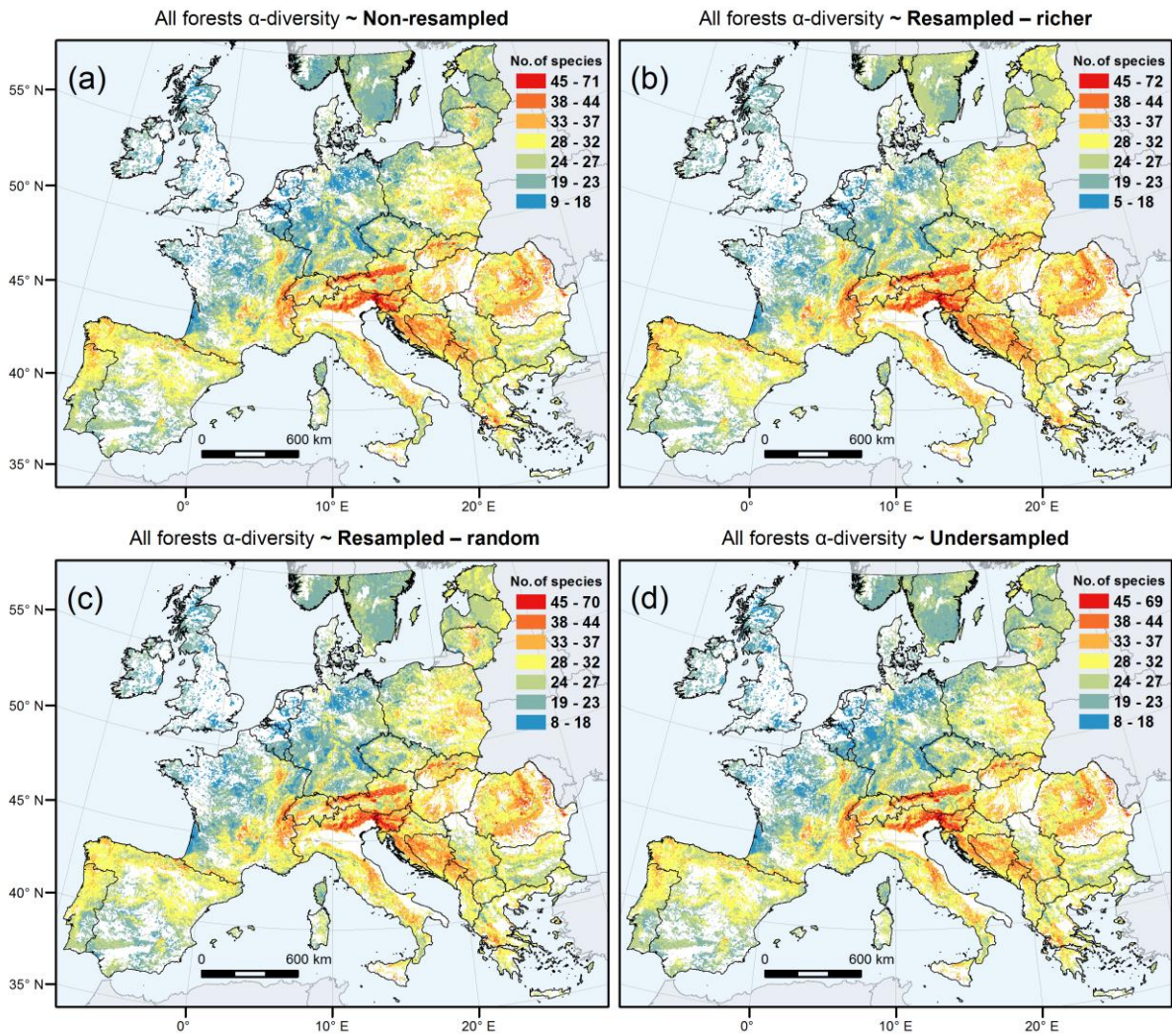
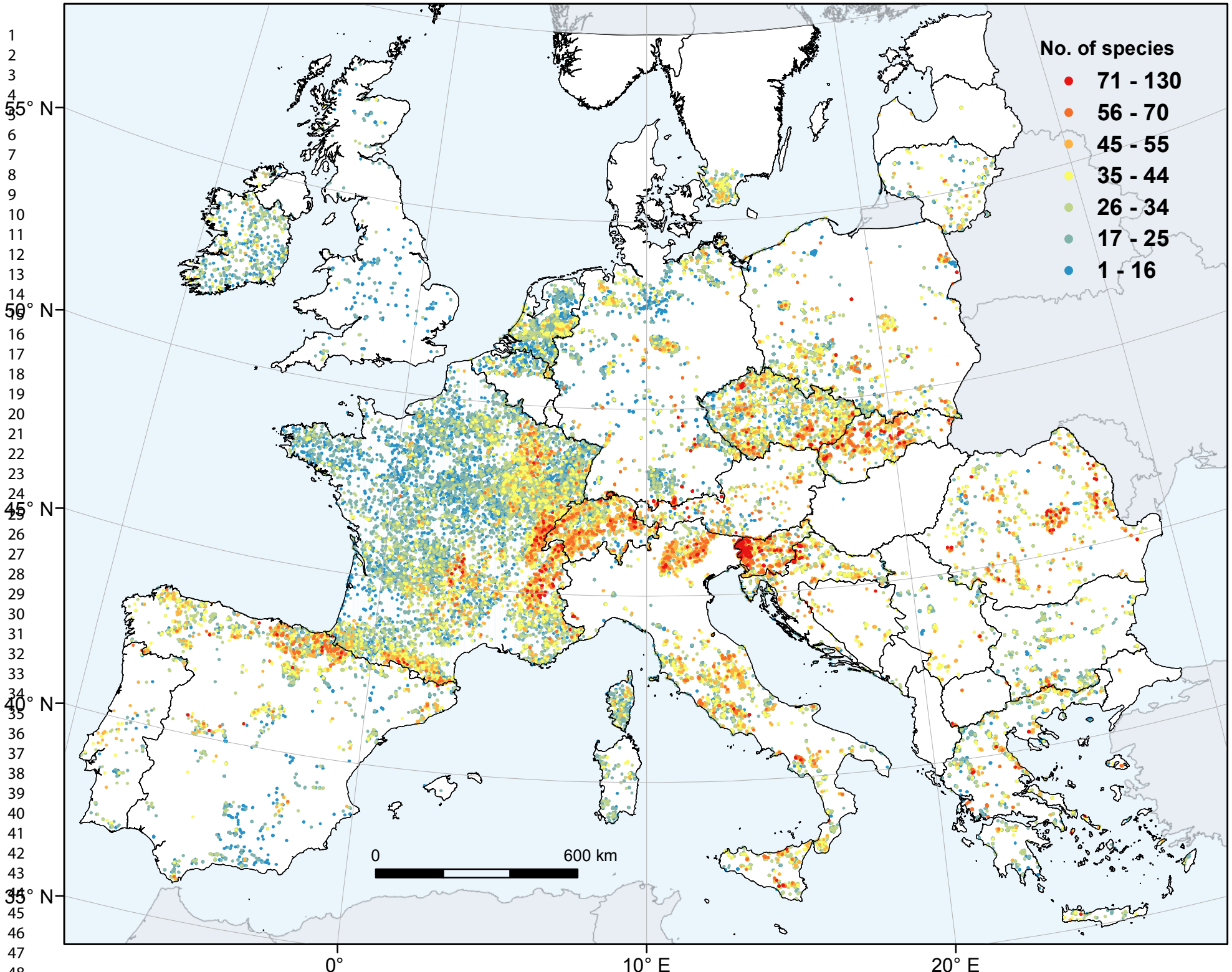


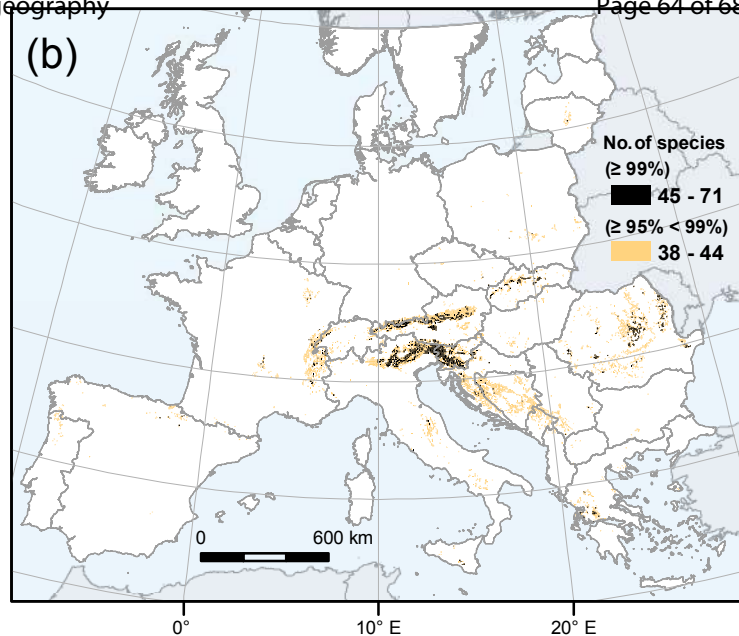
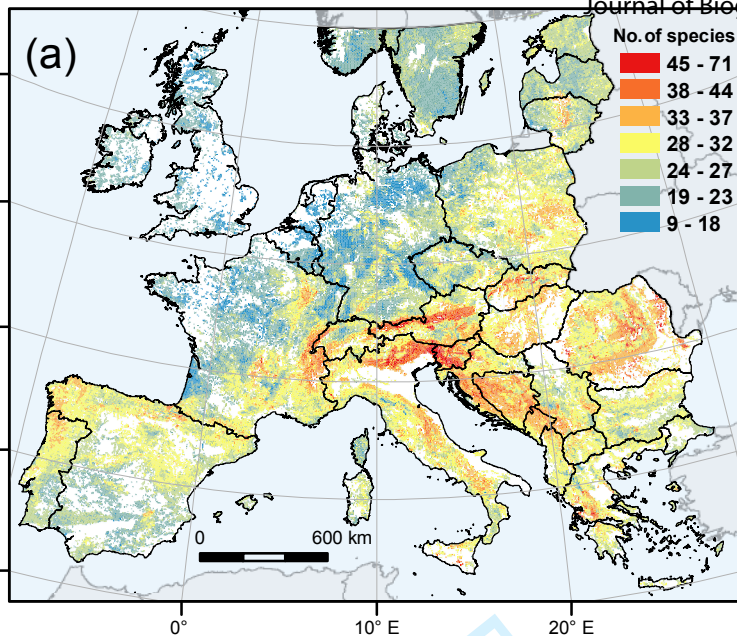
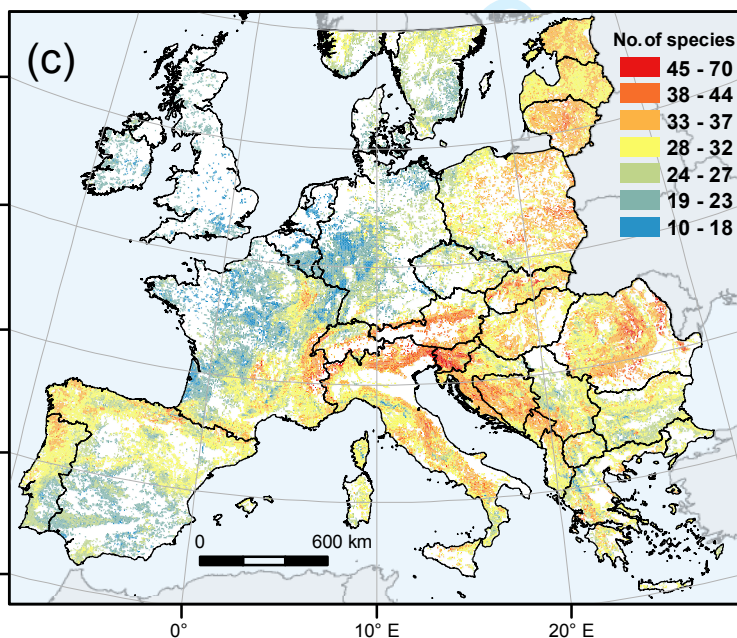
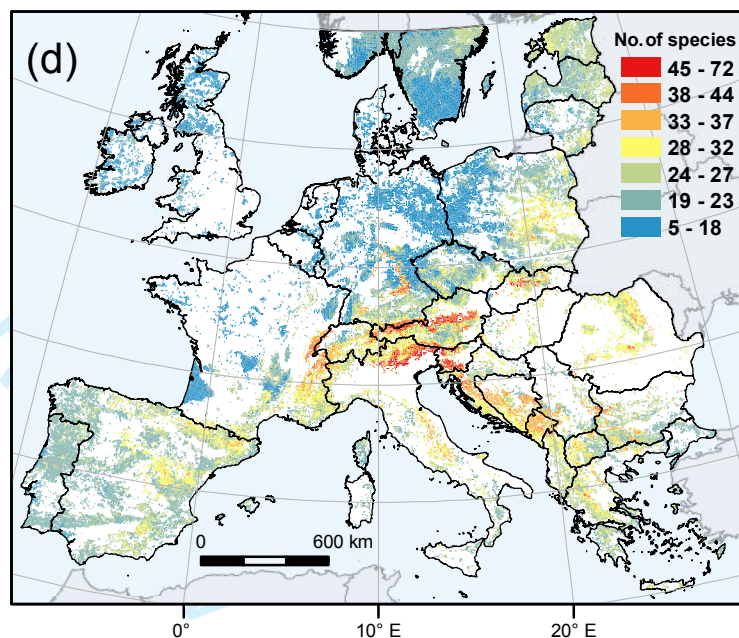
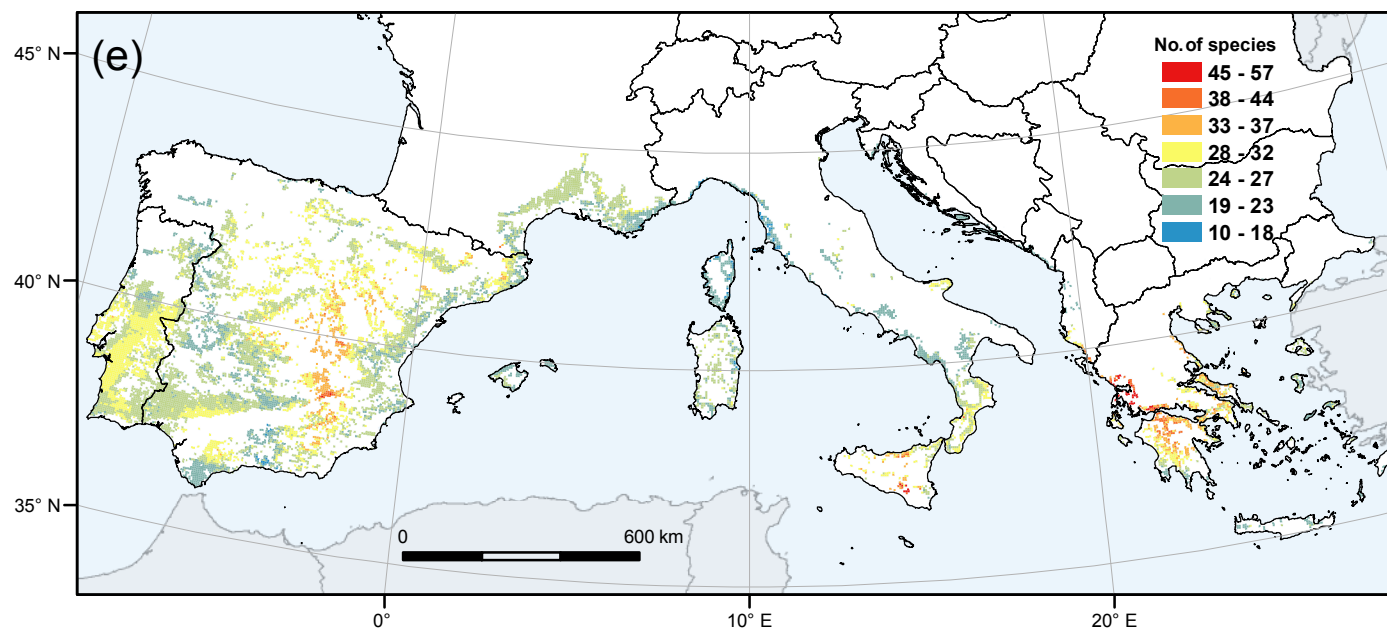
Figure S2.25 – Vascular plant alpha diversity of all forests in Europe – comparison of the predictive maps based on four data subsets representing different resampling strategies: (a) non-resampled (original dataset), (b) resampled with the preference for richer plots, (c) resampled with randomly selected plots, and (d) undersampled.

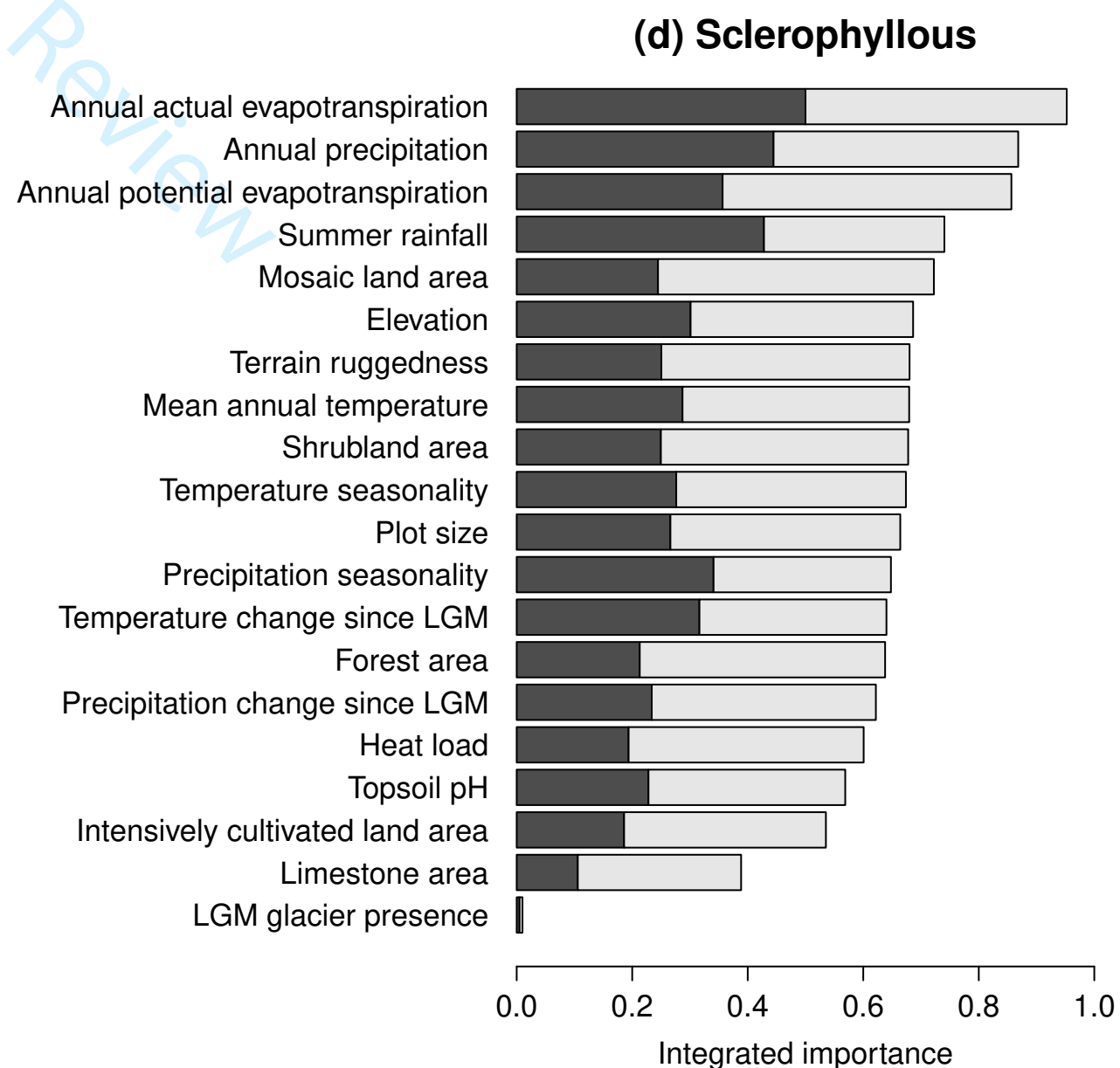
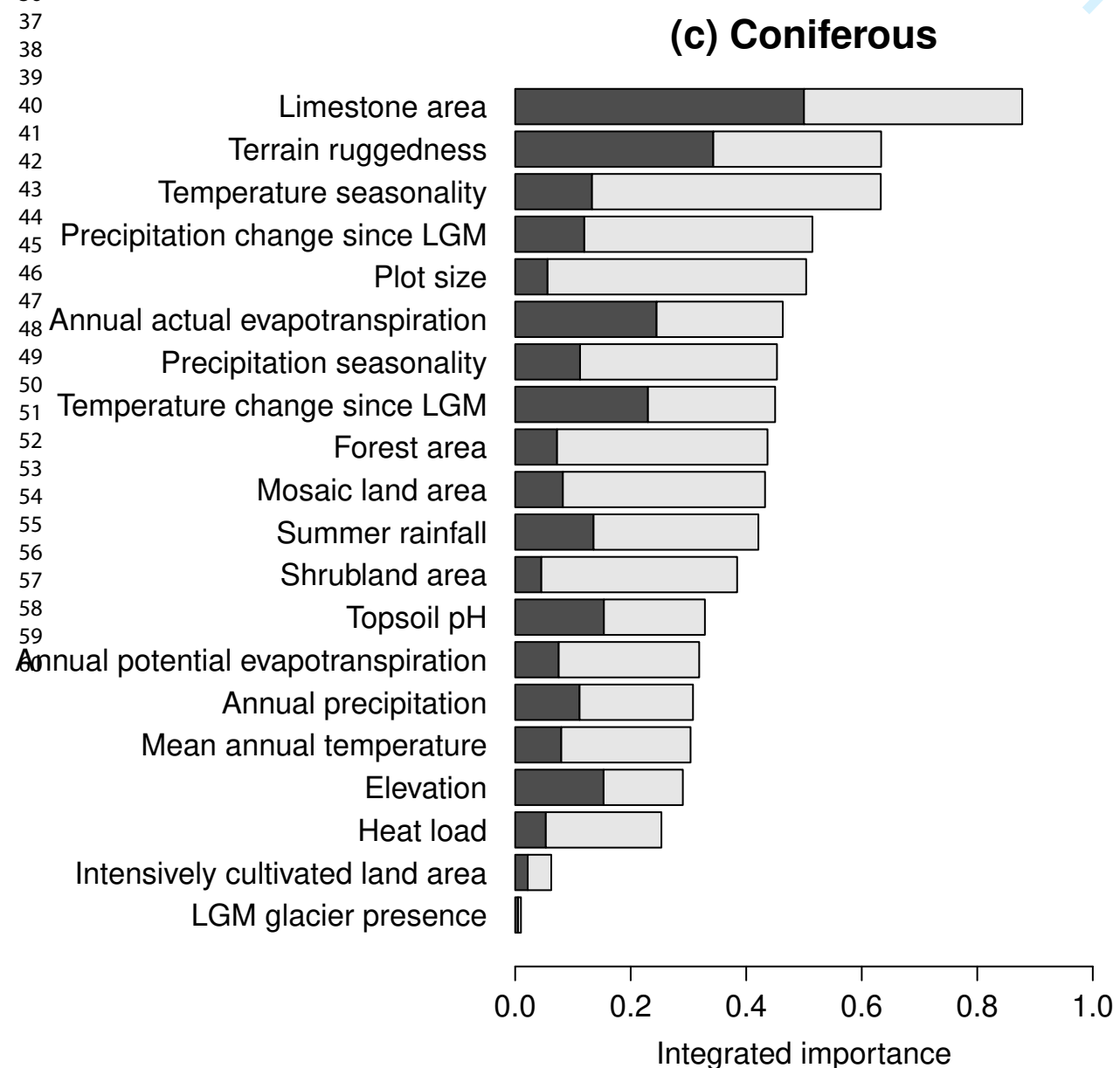
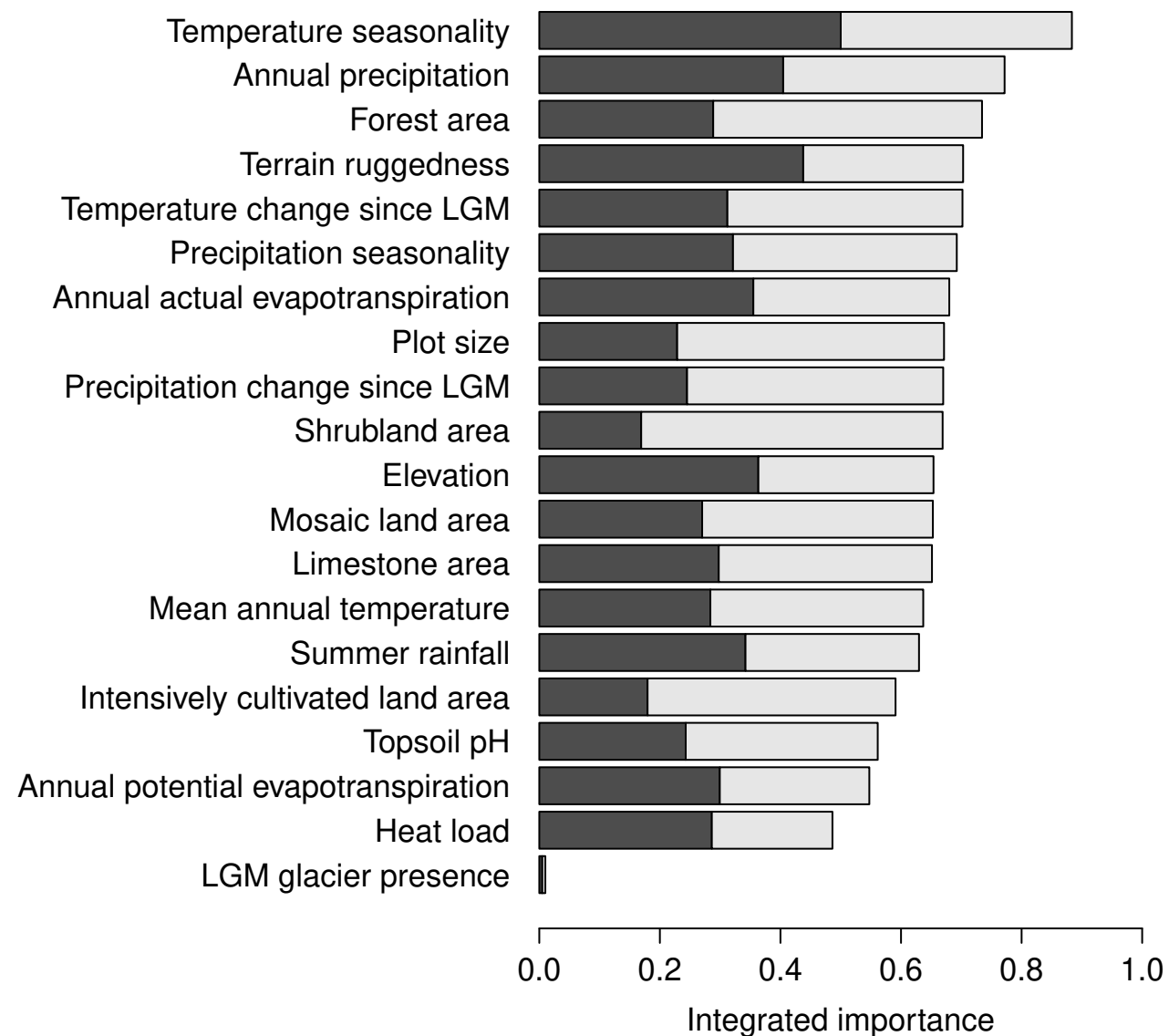
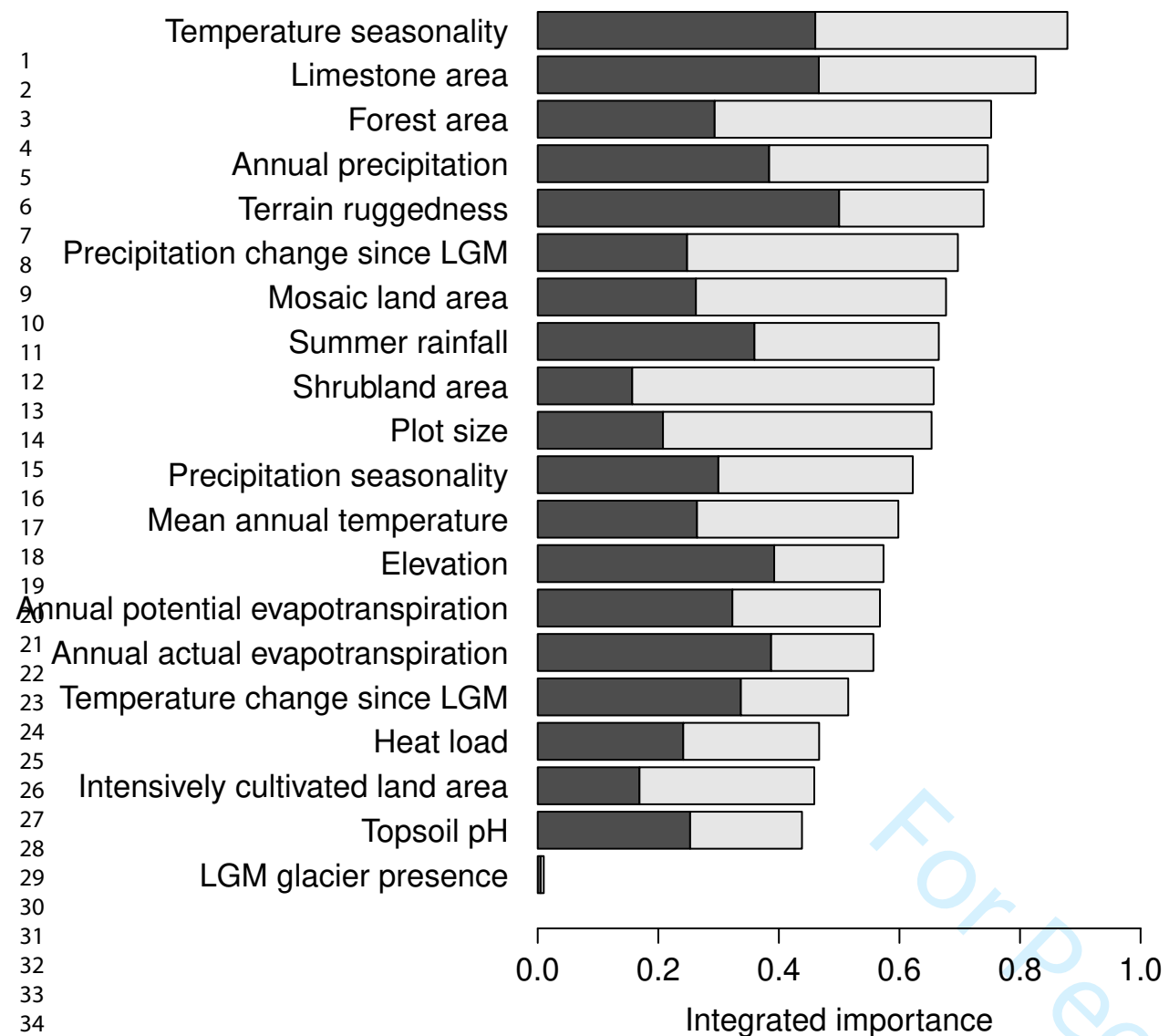
Supplementary references

- Bossard, M., Feranec, J., & O'ahel', J. (2000). *CORINE Land Cover technical guide – addendum 2000*. Copenhagen, DK: European Environment Agency.
- CEC (2004). *The European Soil Database distribution version 2.0*. European Commission and the European Soil Bureau Network (CD-ROM), EUR 19945 EN.
- Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., ... Yamalov, S. (2016). European Vegetation Archive (EVA): an integrated database of European vegetation plots. *Applied Vegetation Science*, 19, 173–180.
- Ehlers, J., Gibbard, P. L., & Hughes, P. D. (Eds). (2011). *Quaternary glaciations – extent and chronology: a closer look*. Oxford, UK: Elsevier.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.
- Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). *Hole-filled SRTM for the globe Version 4*. Available from the CGIAR-CSI SRTM 90m Database: <http://srtm.csi.cgiar.org/>.
- Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, 72, 367–382.
- McCune, B. & Keon, D. (2002). Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science*, 13, 603–606.
- Panagos, P., Van Liedekerke, M., Jones, A., & Montanarella, L. (2012). European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy*, 29, 329–338.
- Sappington, J. M., Longshore, K. M., & Thompson, D. B. (2007). Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *Journal of Wildlife Management*, 71, 1419–1426.
- Schaminée, J. H. J., Chytrý, M., Hennekens, S. M., Janssen, J. A. M., Jiménez-Alfaro, B., Knollová, I., ... Data-providers (2014). *Vegetation analysis and distribution maps for EUNIS habitats. Report EEA/NSV/14/006*. Copenhagen, DK: European Environment Agency.
- Trabucco, A. & Zomer, R.J. (2010). *Global Soil Water Balance Geospatial Database*. CGIAR Consortium for Spatial Information.
- Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). *Regridded Harmonized World Soil Database v1.2. Data set*. Available on-line [<http://daac.ornl.gov>] from Oak Ridge National Laboratory Distributed Active Archive Center. Oak Ridge, TN, USA.
- Zomer, R. J., Trabucco, A., Bossio, D. A., van Straaten, O., & Verchot, L.V. (2008). Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment*, 126, 67–80.

All forests plot diversity

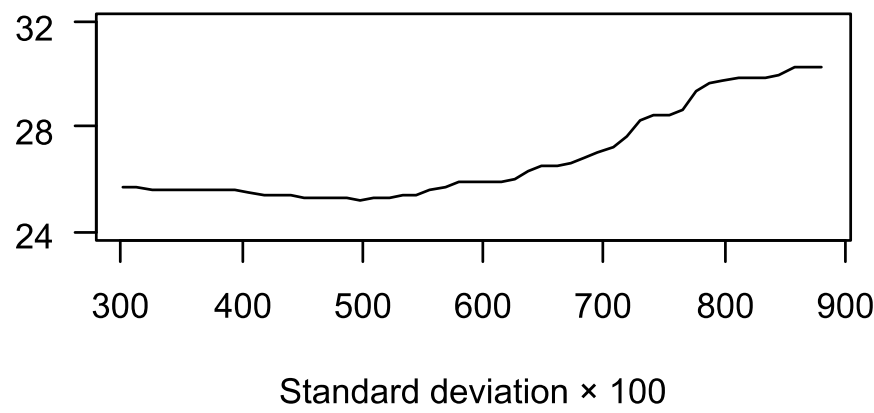


Deciduous broadleaf forests α -diversityConiferous forests α -diversitySclerophyllous forests α -diversity

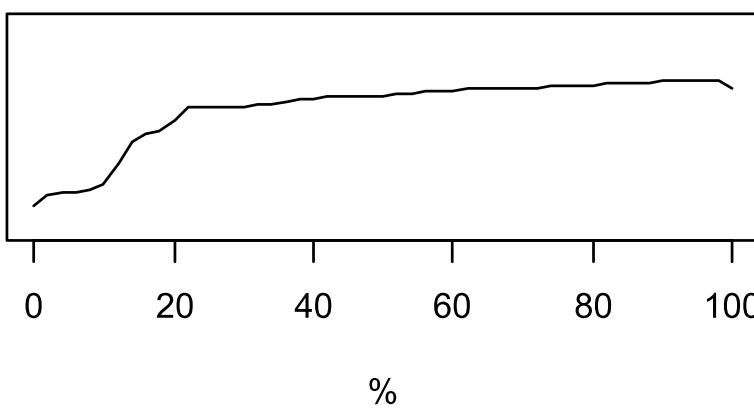


(a) All forests

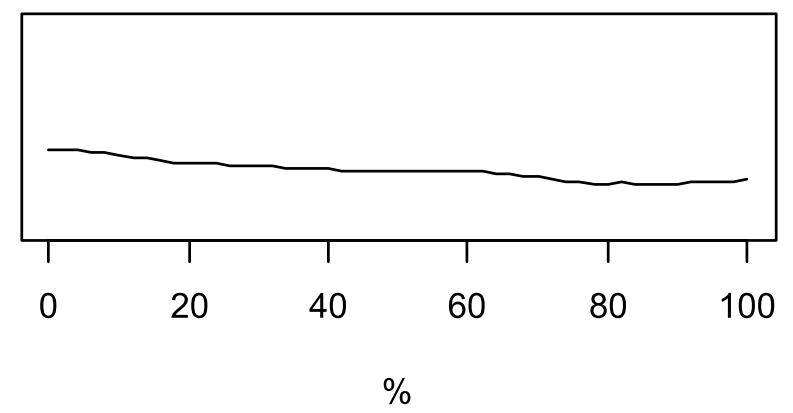
1. Temperature seasonality (0.88)



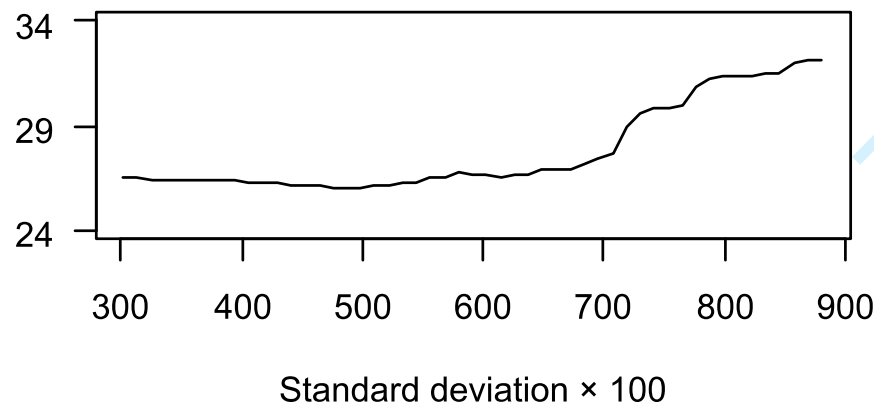
2. Limestone area (0.83)



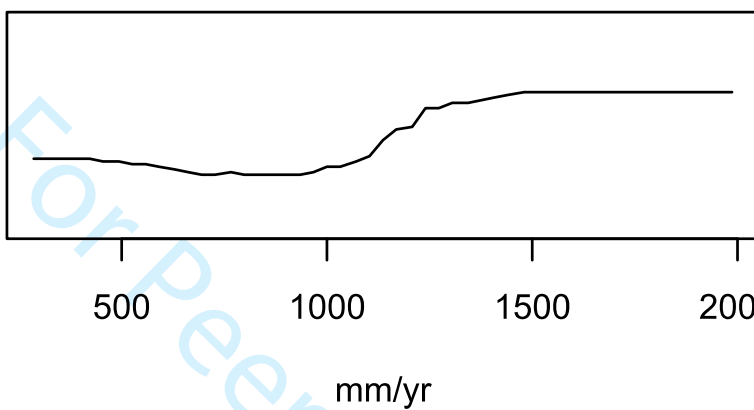
3. Forest area (0.75)

**(b) Deciduous broadleaf**

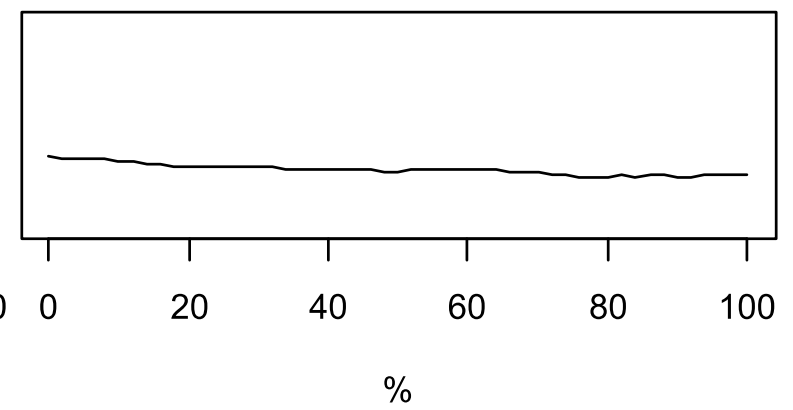
1. Temperature seasonality (0.88)



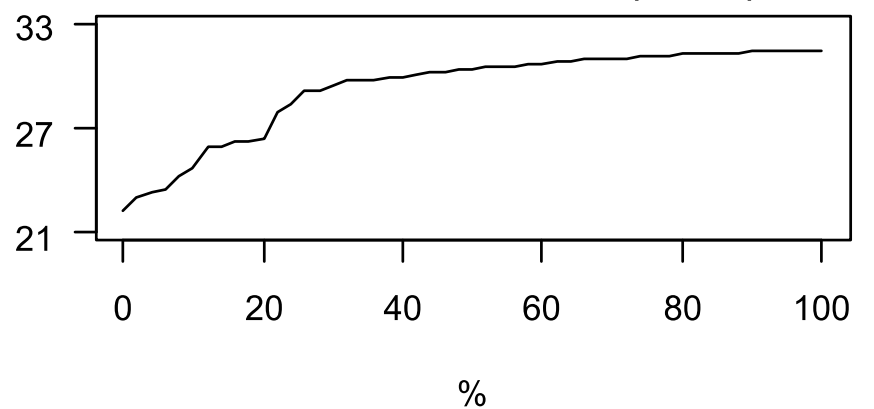
2. Annual precipitation (0.77)



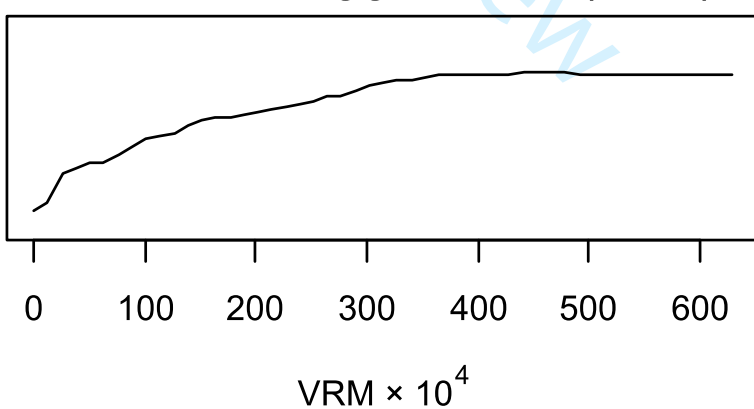
3. Forest area (0.73)

**(c) Coniferous**

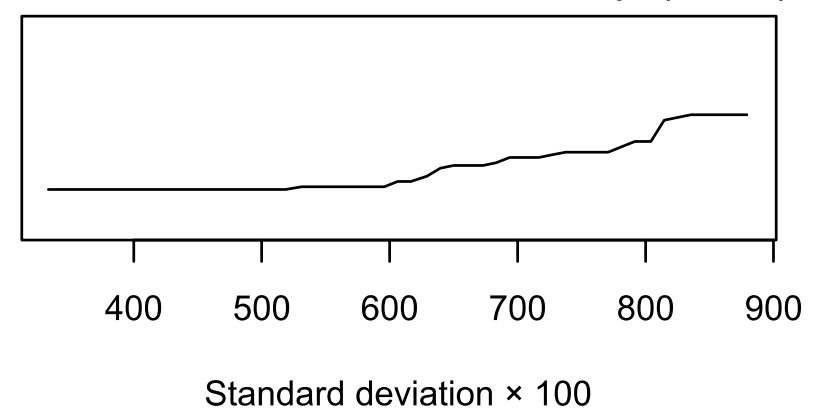
1. Limestone area (0.88)



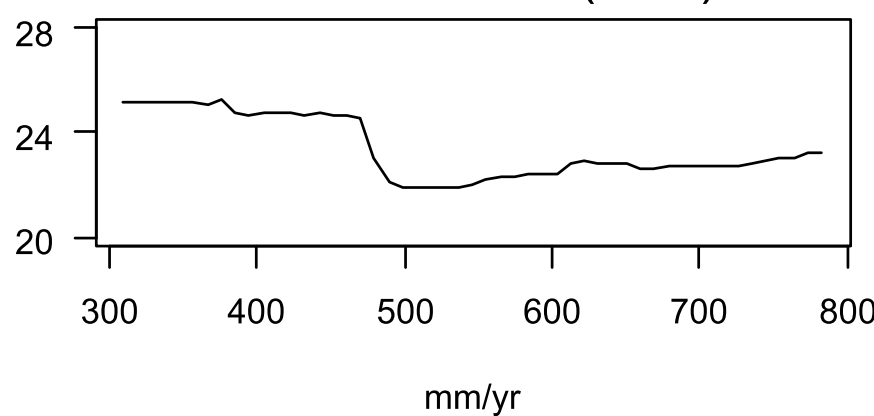
2. Terrain ruggedness (0.63)



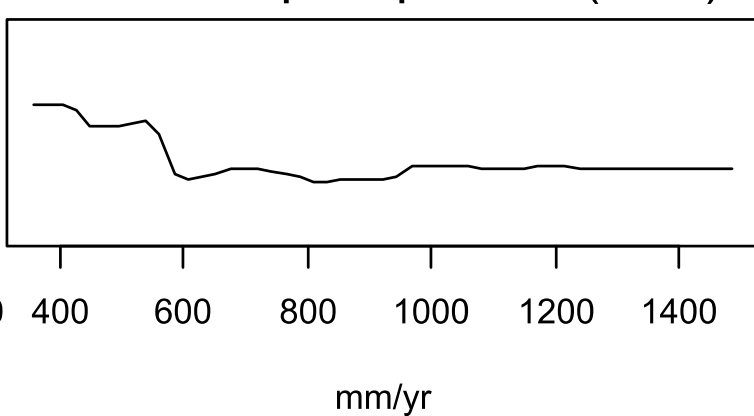
3. Temperature seasonality (0.63)

**(d) Sclerophyllous**

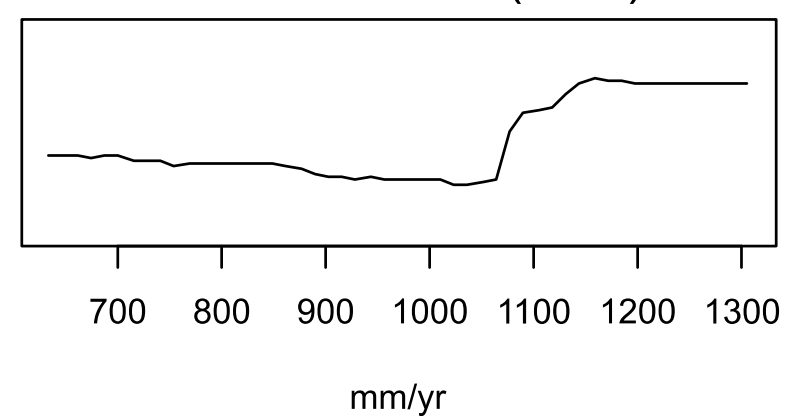
1. Annual AET (0.95)



2. Annual precipitation (0.87)



3. Annual PET (0.86)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 09-Apr-2019
4

5 Dear Dr. Večeřa,
6

7
8 Reference Number: JBI-18-0562.R1
9

10 Title: Alpha diversity of vascular plants in European forests
11

12 We have designated the above paper as on track for acceptance subject to final corrections being
13 carried out. Specifically we require that the data be made available so that the analyses can be
14 repeated. The handling editor, Holger Kreft, provides more specific s on what is required.
15
16

17 Please see below and see any attached file(s) for any editorial or reviewer comments that require
18 your attention.
19

20 We look forward to receiving your revised manuscript within 14 days.
21

22 Thank you again for submitting to the 'Journal of Biogeography'. We look forward to receiving the
23 revised version of your paper.
24
25

26 Yours sincerely,
27

28
29 Steven Higgins
30 Chief Editor, Journal of Biogeography
31
32

33
34
35
36 -----
37 EDITOR'S COMMENTS TO AUTHOR
38

39 Editor: Kreft, Holger

40 Comments to the Author:

41 Dear authors,
42
43

44 thanks for your careful revision. Your manuscript has now been seen by the two original referees and
45 they are generally very happy how you dealt with the issues raised and I agree with them.

46 There is one thing that you might want to consider that relates to Data accessibility. It is great that
47 you are planning to make available the modeled richness maps. However, I also strongly urge you to
48 consider publishing the full data frame that was used for the analyses. Otherwise, I am afraid that
49 your analyses will not be replicable.
50
51

52 Best regards,
53

54 Holger Kreft
55

56
57 Prof. Holger Kreft
58
59
60

1
2
3 Dear editors,
4
5

6
7 We are glad for the decision to accept our study.
8

9 Regarding your recommendation to publish full data that we used, it is not possible to make the
10 dataset freely available due to the European Vegetation Archive (EVA) policy. However, the dataset
11 we used is stored in EVA and can be obtained upon request. I have specified this issue in the Data
12 accessibility statement.
13
14

15
16 With best wishes,
17
18

19
20 Martin Večeřa
21

22 On behalf of all the co-authors
23
24
25
26 -----
27

28 REVIEWER COMMENTS TO AUTHOR
29

30
31 Referee: 1
32

33 Comments to the Author

34 Review Journal of Biogeography – JBI-18-0562.R1

35 Alpha-diversity of vascular plants in European forests
36
37

38 This is the revised version of a manuscript which I reviewed already in the original version. In general,
39 most of the remarks and comments of the reviewers and by the editor have been considered in this
40 revised version, mainly in focusing on mapping and understanding the drivers of species richness. I
41 am not completely satisfied considering the management or naturalness aspect on the selection
42 within the datasets, especially by the “phytosociological” or “natural forest habitat” criteria of
43 vegetation-plot sampling: maps for coniferous forests e.g. in Belgium, the Netherlands or Czech
44 Republic present spatial diversity patterns within areas which belongs not to landscapes where
45 coniferous forests will be close-to-nature.
46
47

48 Nevertheless, I recommend publishing the paper in this revised version.
49

50 Thank you for your comments and suggestions. We are aware of the mentioned issues, they are
51 discussed in the manuscript. We consider them to be a potential subject of improvement for future
52 maps of alpha diversity in European forests, once the data on non-native occurrence of coniferous as
53 well as other forest types are available at the pan-European extent.
54
55

56
57 Referee: 2
58

59 Comments to the Author
60

1
2
3 Dear authors,
4

5 Thank you for careful revision and detailed replies. I do appreciate the progress you have done with
6 the paper. Correlation plots are very helpful to understand the data used in the models. I still think
7 that percentages of different land-covers within a single model is not the best idea but I agree that
8 this does not change your results much. I am very happy to see the uncertainty maps.
9
10

11
12
13 Thank you for your comments and suggestions. Regarding the land-cover proportions, we will think
14 also about different solutions for similar projects in future.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review