Forest tree species distribution for Europe 2000-2020: mapping potential and realized distributions using spatiotemporal Machine Learning

Carmelo Bonannella (carmelo.bonannella@opengeohub.org)
OpenGeoHub https://orcid.org/0000-0002-5391-8427
Tomislav Hengl
OpenGeoHub https://orcid.org/0000-0002-9921-5129
Johannes Heisig
University of Münster https://orcid.org/0000-0003-3586-3001
Leandro Parente
OpenGeoHub https://orcid.org/0000-0003-1589-0467
Marvin N Wright
University of Bremen https://orcid.org/0000-0002-8542-6291
Martin Herold
Wageningen University & Research https://orcid.org/0000-0003-0246-6886
Sytze de Bruin
Wageningen University & Research https://orcid.org/0000-0002-6884-2832

Research Article

Keywords: Species distribution model, Ecological niche, Ensemble modeling, Machine learning, Imbalanced data, Presence-absence, Spatiotemporal modeling, Stacked generalization, High resolution, Tree species

Posted Date: January 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1252972/v1

Forest tree species distribution for Europe 2000–2020: mapping potential and realized distributions using spatiotemporal Machine Learning

Carmelo Bonannella[1], Tomislav Hengl[1], Johannes Heisig[2], Leandro Parente[1], Marvin N Wright[3, 4], Martin Herold[5, 6], and Sytze de Bruin[6]

[1]**OpenGeoHub, Wageningen, The Netherlands**

[2]**Institute for Geoinformatics, University of Münster, Münster, Germany**

[3]**Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany**

[4]**University of Bremen, Bremen, Germany**

[5]**GFZ German Research Centre for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Telegrafenberg, Potsdam, 14473, Germany**

[6]**Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Wageningen, The Netherlands**

Corresponding author:

Carmelo Bonannella[1]

Email address: carmelo.bonannella@opengeohub.org

ABSTRACT

Paper describes a data-driven framework based on spatio-temporal ensemble machine learning to produce distribution maps for 16 forest tree species (*Abies alba* Mill., *Castanea sativa* Mill., *Corylus avellana* L., *Fagus sylvatica* L., *Olea europaea* L., *Picea abies* L. H. Karst., *Pinus halepensis Mill.*, *Pinus nigra J. F. Arnold*, *Pinus pinea* L., *Pinus sylvestris* L., *Prunus avium* L., *Quercus cerris* L., *Quercus ilex* L., *Quercus robur* L., *Quercus suber* L. and *Salix caprea* L.) at high spatial resolution (30 m). Tree occurrence data for a total of 3 million of points was used to train different Machine Learning (ML) algorithms: random forest, gradient-boosted trees, generalized linear models, k-nearest neighbors, CART and an artificial neural network. A stack of 585 coarse and high resolution covariates representing spectral reflectance (Landsat bands, spectral indices; time-series of seasonal composites), different biophysical conditions (i.e. temperature, precipitation, elevation, lithology) and biotic competition (other species distribution maps) was used as predictors for realized distributions, while potential distribution was modelled with environmental predictors only. Logloss and computing time were used to select the three best algorithms to train an ensemble model based on stacking with a logistic regressor as a meta-learner for each species. High resolution (30 m) probability and model uncertainty maps of realized distribution were produced for each species using a time window of 4 years for a total of 6 distribution maps per species for the studied period, while for potential distributions only one map per species was produced. Results of spatial cross validation show that *Olea europaea* and *Quercus suber* achieved the best performances in both potential and realized distribution, while *Pinus sylvestris* and *Salix caprea* achieved the worst. Further analysis shows that fine-resolution models consistently outperformed coarse resolution models (250 m) for realized distribution (average decrease in logloss: +53%). Realized distribution models achieved higher predictive performances than potential distribution ones. Importance of predictor variables differed across species and models, with the green band for summer and the NDWI and NDVI for fall for realized distribution and the diffuse irradiation and precipitation of the driest quarter being the most important and frequent for potential distribution. The ensemble model outperformed or performed as good as the best individual model in all potential species distributions, while for ten species it performed worse than the best individual model in modeling realized distributions. The framework shows how combining continuous and consistent EO time series data with state of the art ML can be used to derive dynamic distribution maps. The produced time-series occurrence predictions can be used to quantify temporal trends and detect potential forest degradation.

Submitted to PeerJ on: 7th of January 2022;

1 INTRODUCTION

Reforestation and restauration are considered key strategies for tackling global warming and enhancing CO_2 sequestration (Lefebvre et al., 2021; Domke et al., 2020; Nave et al., 2019). Under the European Green Deal and the EU biodiversity strategy for 2030, the EU has committed to planting at least 3 billion additional trees in the EU by 2030 (https://ec.europa.eu/environment/3-billion-trees_en). At the same time in Germany, trees in more than 2.5% of the country's total forest area have died because of beetles and drought fueled by a warming climate (Popkin, 2021). Obtaining reliable information on forest tree species distribution in space and time has become paramount for stakeholders and decision-makers to anticipate and minimize climate change impacts (Keenan, 2015).

The production of species distribution maps has been dominated by Species Distribution Modeling

(SDM) approaches (Franklin, 2010). Maps of species ecological niches are made by associating values of different predictors to known locations of the target species and then used to predict distribution in geographic space where no field data for the target species is available. Correlative SDM (C-SDM) in particular has recently become very popular due to the increase in advanced statistical techniques (i.e. Machine Learning) and geospatial data (satellites, drones etc.) availability (Gobeyn et al., 2019). This is also reflected in an increase of SDM-related publications in the last decade (Booth, 2018), as well as reviews (Booth et al., 2014).

From the late 2000s, the focus in SDM has shifted from traditional statistical models to machine-learning (ML) algorithms (Elith et al., 2008). While statistical models start by assuming a particular structural model whose parameters are estimated from the data, ML tries to learn the relationship between the response and the predictors through the observation of dominant patterns (Breiman, 2001b). There are no ecological assumptions explicitly embedded in the algorithms (e.g. Random Forest, Artificial Neural Networks) and most of the time it is difficult to interpret their outputs, which is also referred to as the *"black box"* problem of ML (Molnar, 2020). However, the exponential increase in computing power (Gorelick et al., 2017), predictor variables availability (Zhu et al., 2019) and ecological *"big data"* gathered by multiple sources (i.e. sensors, cameras etc.) (Hampton et al., 2013) make ML increasingly suitable tool for SDM. Possibly it is even the only practical means, to gain insights from such an amount of information and to map, monitor or forecast changes on multiple geographical scales (Gobeyn et al., 2019).

There is a broad variety of algorithms used for SDM (Gobeyn et al., 2019; Franklin, 2010): among the most used presence-absence we find Generalized Linear Models (GLM) (Nelder and Wedderburn, 1972), Classification and Regression Trees (CART) (Quinlan, 1986), gradient-boosted Trees (GBT) (Friedman, 2002), Random Forest (RF) (Breiman, 2001a) and Artificial Neural Networks (ANN) (McCulloch and Pitts, 1943). In literature, many examples of tree species distribution maps created with regression-based and ML SDM can be found: Hill et al. (2017) used GLM to model the distribution of *Acer platanoides* and *Fraxinus excelsior* for Great Britain at 1 km resolution, Marchi and Ducci (2018) compared the predictive performances of nine different algorithms, including RF, ANN, GLM and GBT, to model the current distribution of *Abies alba* and *Fagus sylvatica* for Italy at 1 km resolution using National Forest Inventory (NFI) data; they also evaluated the potential change in the distribution of these species in 2050 according to two of the four Representative Concentration Pathway (RCP) scenarios adopted by IPCC. Prasad et al. (2006) used NFI data as well and were among the first to include topographical and land cover variables to predict the current distribution and future climate scenarios of *Acer saccharum, Fagus grandifolia, Pinus taeda* and *Quercus alba* in the eastern United States using RF and CART at 1 km resolution. Following this trend, Cord et al. (2009) added another layer of complexity: to model the distribution of tree species of the genera *Pinus spp.* and *Quercus spp.* in Mexico, they included multi-temporal remotely sensed data at medium resolution (250 m) as predictors and compared the influence of Earth Observation (EO) data on the predictive performance of the model.

Meier et al. (2010) measured how strongly the influence of biotic interactions affects the prediction of species distribution in Swiss broadleaf forests. Biotic interactions are usually included in the form of distribution maps due to data availability: in this study they were included as predictors in the form

of relative abundance compared to the target species, relative abundance of large individuals and total shade by large individuals. Results showed how biotic variables considerably improved the predictive performances; their effect were independent from abiotic variables and helped in identifying aspects not easily captured by abiotic predictors only.

Brus et al. (2012) mapped 23 species groups over Europe at 1 km resolution using NFI data, bioclimatic and topographical variables. Contrary to most SDM studies, where each species is mapped individually and one map per species is provided, Brus et al. (2012) use multinomial regression to assign a probability value in the interval [0–1] to each species, with probabilities adding up to 1. This means that each pixel shows the prevalent species, but not if multiple species may coexist in the same area. Another state-of-the-art data source on forest species distribution in Europe is the European Atlas of Forest Tree species. This is among the largest data sources with information on forest tree species for Europe (San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, 2016). It describes in detail the auto-ecology of 76 different forest tree species and provides geographical information on each species in the form of chorological maps, probability of presence maps and maximum habitat suitability maps. While these predictions are useful to ascertain the species composition of European forests, there is now a need for spatio-temporal predictions of potential and realized distribution of forest tree species building on new methods that potentially better suited to deal with the increasing resolution and availability of data for both training and prediction.

Spatio-temporal modeling, Earth Observation data and specifically the use of high spatial resolution data have only recently started to be explored for SDM (Gelfand and Shirota, 2021; Pérez Chaves et al., 2018; Hefley and Hooten, 2016). Commonly, SDM still relies on climatic or bioclimatic factors at coarse spatial resolution ($\geq$1 km) while in the temporal dimension long time averages (30–50 years) are commonly used (Iturbide et al., 2018). The potential impact of differences in resolution of the input variables on the results is often ignored (Porfirio et al., 2014) despite the fact the forest spatial patterns in Europe are often linked to management decisions happening at local scales. Furthermore, previous studies comparing predictive performances at different spatial resolutions mention distribution maps with high spatial resolution ($<$ 100 m) and slightly lower prediction accuracy can still be more useful for invasive species management than coarser ($>$ 250 m) but more accurate maps (Manzoor et al., 2018; Guisan et al., 2013; Gottschalk et al., 2011; Prates-Clark et al., 2008).

In the last decade, ecologists have conducted hundreds of studies to determine which methods best suit the needs of SDM. Some general findings are that presence-absence approaches are usually better than presence-only and better predictive power comes at the cost of transferability of the model (Valavi et al., 2021; Pecchi et al., 2019; Guisan et al., 2017) but no consensus has been reached yet. For example, previous studies have shown that distribution maps derived from the same dataset but using different models can lead to quite opposite conclusions (Araújo and New, 2007; Pearson et al., 2006). Inter-model variability in projections has been tackled using ensemble modeling, where numerous independent models are fit using a range of methods applied to the same input data while the outputs of the individual models are aggregated into the final prediction. Hao et al. (2019) provide a good general review in this sense, focusing only on SDM studies that use ensemble methods: they reviewed a total of 224 papers, limited, however, to ensemble methods implemented in the BIOMOD software. Strong conclusions could not

be drawn due to lack of information on performances of ensemble models versus individual models. On top of that, only few ensemble strategies have been investigated: mean, median and weighted average. These strategies are intuitively simple and involve, in the first two cases, just taking the mean or median of the predictions of the individual models as a final prediction. The weighted average does the same but scales the predictions by weights assigned based on cross validation predictive performances of the models. A robust ensemble technique that, to our knowledge, has not been tested yet for SDM is stacking or stacked generalization. Outputs made by the individual models are the input of a meta-learner which then produces the final prediction (Wolpert, 1992). This approach will be used in this study.

To enable more insight into tree species dynamics over EU we have set the following objectives:

1. To develop a framework for modeling species distribution in space-time with state-of-the-art Machine Learning.
2. To assess the importance of various sources of Earth Observation data for mapping tree species distributions.
3. To explore and quantify the importance of high resolution data on model predictive performances.

We first review the theory for space-time modeling of forest tree species using presence and absence data. We then show results of our experimental design to build a spatio-temporal ensemble model for SDM based on ML algorithms, variable importance and accuracy assessment for selection of tree species (16) based on 5–fold spatial cross-validation. Finally, we discuss variable importance of various EO data sources and visualize final predictions for sample areas and whole of EU.

2 MATERIALS AND METHODS

2.1 Species distribution model theory

SDM, as any modeling tool, tries to provide a simplified and understandable conceptual representation of a complex phenomenon in the real world, in this case the geographical distribution of a species; it is therefore important to clarify, before any modeling attempt, what kind of distribution a particular study aims to quantify, i.e. what is the ecological meaning of the response variable predicted. Conceptually speaking, the applied procedure is the same: based on the similarity between the values of the predictors associated with the occurrence data, build a function that can be used to assign probability of presence of the target species to areas with unavailable field data. However, based on what is actually modelled, the choice of predictors, training data (presence-only or presence-absence) and modeling techniques may vary (Soberón, 2010).

Hutchinson (1957) defined the distribution of a species as the n-dimensional hypervolume in a vectorial space where each component is a variable that positively influences the growth of the species and on which the species has no influence. These variables are usually abiotic/environmental variables, such as temperature, precipitation or elevation (Soberón, 2010; Soberón and Peterson, 2005). By considering the species physiological response to the combined effect of all these variables in the vectorial space, it is possible to capture the fundamental niche of a species, i.e. all the possible conditions that a species can

occupy in the environment. The realized niche is instead defined as a narrower region of this vectorial space constrained by biotic interactions and dispersal. The relationship between abiotic factors, biotic interactions and the regions accessible by the species through dispersion was formalized by Soberón and Peterson (2005) in the BAM diagram.

By definition, it is not possible for SDMs to capture the entire fundamental niche because occurrence data gathered through field campaigns is already constrained by biotic interactions. While it is possible to analyze the fundamental niche of a species using mechanistic approaches, SDMs can only investigate the realm of the realized niche (Guisan and Thuiller, 2005). On top of the realized niche of a species, in this study we also model the potential niche. The potential niche is defined as that portion of the fundamental niche that exists in the study area at the time of the study (Peterson et al., 2011). The concept is relevant as it assesses whether the conditions as defined in the fundamental niche actually exist (Guisan et al., 2017). Guisan et al. (2014) suggested that the traditional approach of SDM to capture just a snapshot of the realized niche may be too restrictive. In the context of climate change, assessing the impact of changing conditions on the distribution of a species requires making predictions in space and time. For space-time modeling, the realized niche has to be built simultaneously considering: (a) different time periods, and (b) different geographical areas (e.g. native area or exotic areas). (Martinez-Minaya et al., 2018; Miller et al., 2007). The differences in definitions here discussed are then reflected in the way the two niches are modelled:

1. Potential distribution is modelled using variables that capture the ecological gradients of the study area and that are in no way influenced by other species (i.e. no biotic interactions),

2. Remotely sensed, reflectance-based data (i.e. land cover, spectral indices etc.) are not used to model potential distribution as vegetation species distribution may be influenced owing to human alteration of the landscape, underestimating the potential distribution, or capturing disturbed areas that are just temporary not suitable for the species (Bradley et al., 2012). The same is true for other human altered of environmental factors such as the Urban Heat Island (UHI) effect (Bechtel and Schmidt, 2011; Pigeon et al., 2007),

3. The realized distribution is modelled using a combination of environmental variables (temperature, precipitation, topography), biotic variables and reflectance-based variables.

Occurrence data used in SDM is either presence-only (i.e. records of locations where the species was observed) or presence-absence (i.e. locations of both presence and absence of a species have been recorded in the sampling design): ML algorithms for SDM treat the suitability task as a binary classification problem, where the response variable can assume a discrete value of 1 or 0 (i.e. binary response for respectively, presence or absence) or a continuous value from 0 to 100 (i.e. probability of presence), with the latter being the most used (Guisan et al., 2017). Species occurrence datasets are hard to obtain due to the high costs of the field surveys. Presence and true absence data usually suffer from preferential sampling, meaning that a) selection of the sampling locations is not independent from the response variable (i.e. species occurrence) and b) the locations in the study area are not equally likely to be sampled (Martinez-Minaya et al., 2018). True absence data is usually not available but replaced with background

data (Phillips et al., 2009), which does not try to infer where the species is absent but rather to characterize the environmental background of the study area. Alternatively, absence data is artificially generated and referred to as pseudo-absence. Generally, pseudo-absence indicates locations that researchers assume are potentially unsuitable for the target species, and points can be generated in several ways (Iturbide et al., 2015). Different kinds of absence data have to be used depending on the distribution modelled: potential distribution cannot be modelled with absence data coming from non environmental factors (i.e. a species being absent due to competition/biotic interactions) (Jiménez-Valverde et al., 2008).

As a source of homogeneously distributed true absence data for both potential and realized distribution, in this study we will use the Land Use/Cover Area Survey (LUCAS) (EUROSTAT, 2017) dataset: in-situ observations of land use and land cover distributed on a 2 x 2 km grid covering the whole European Union (see d'Andrimont et al. (2021) for more information and https://ec.europa.eu/eurostat/web/lucas/data/lucas-grid for the official grid).

2.2 General workflow

Potential and realized distribution maps were produced for continental Europe for the time period January 2000 – December 2020 using a spatio-temporal modeling approach. Final prediction maps show the probability of presence (0–100%) of at least one individual of the target species in the area covered by a 30 m pixel. Probability of presence is relative to the mapped target species, irrespective of the potential co-occurrence of other species in the same 30 m pixel and should not be confused with the absolute abundance or proportion of each species in the pixel area. The sum of the presence probabilities of different species in the same pixel can thus exceed 100 %. With each probability map comes a model uncertainty map, where the user can check the reliability of probability map values per pixel. We produced one potential distribution map and six realized distribution maps for each species: the assumption is that the conditions in the study area that determine the potential distribution of the species did not change over the time period analyzed; this does not hold for the realized distribution. We split the time period analyzed in six time windows according to the following scheme:

- 2000–2002,
- 2002–2006,
- 2006–2010,
- 2010–2014,
- 2014–2018,
- 2018–2020.

One realized distribution map was produced for each time period. The general workflow used to derive the distribution maps is shown in Fig. 1: two datasets for each species (potential and realized dataset) were produced and overlaid with a set of both static and dynamic predictor variables, matching both the point location in space (x, y) and time (t). Predictor variables used for potential and realized distribution follow the rules listed in the previous section: reflectance-based predictor variables and species distribution maps

were used to model only the realized distribution. The classification matrix produced was used to fit a spatio-temporal ensemble model based on stacking. The workflow was implemented in the Python and R programming languages.

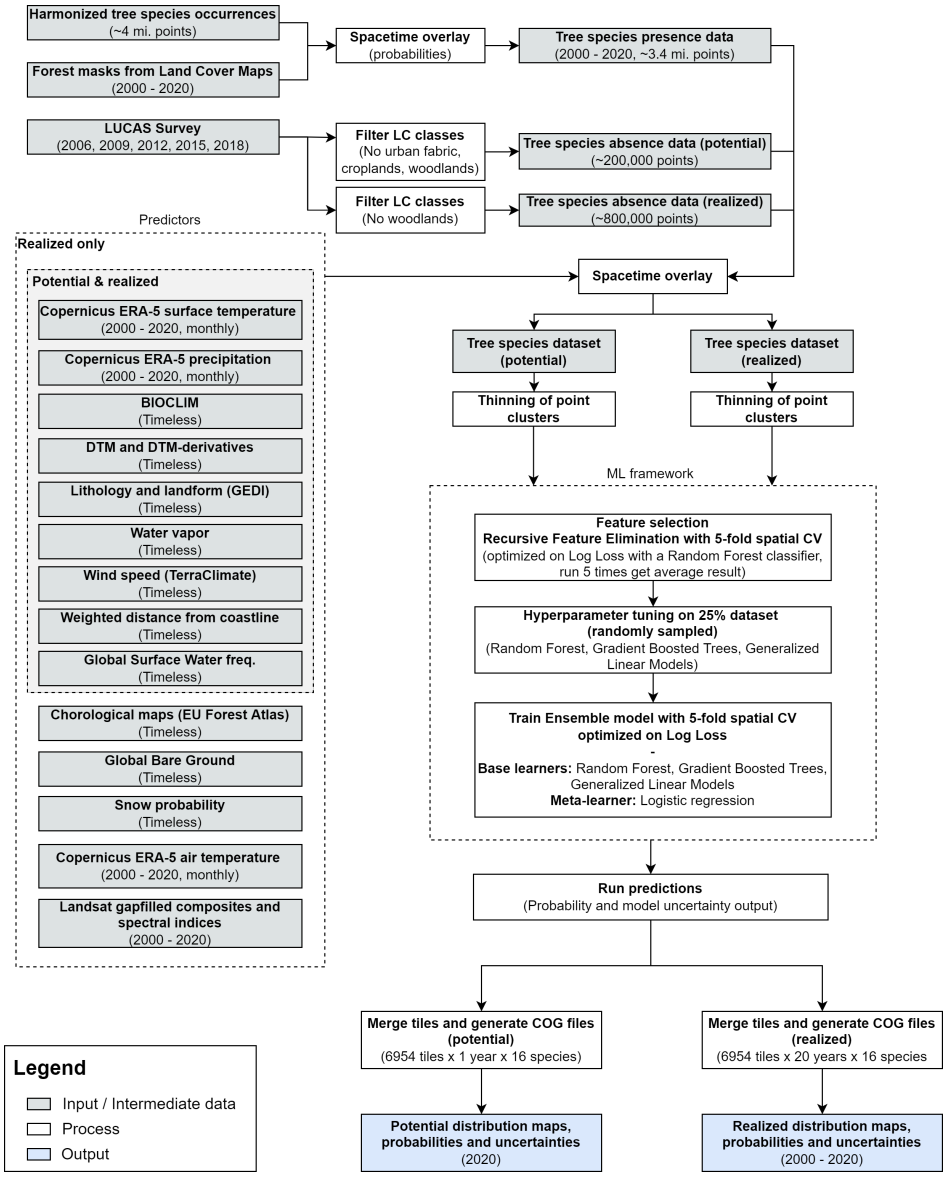


Figure 1. General workflow illustrating the preparation of the point data, the predictor variables used, model building (feature selection — hyperparameter optimization — training) and preparation of distribution maps.

2.3 Study area

The study area covers the European continent, that is all countries included in the Corine Land Cover (CLC) database (Büttner et al., 1998) except Turkey. European forests cover 33% of the continent's land area. Owing to the variety of climatic conditions across both latitudinal and longitudinal gradients, twelve out of the 20 FAO Forest Ecological Zones are represented in European forests (de Rigo et al., 2016b).

The European Atlas of Forest Tree Species (San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, 2016) reports detailed information for a total of 76 forest tree species. From those, the following 16 were chosen and modelled in this study:

1. Silver fir (*Abies alba* Mill.),
2. Sweets chestnut (*Castanea sativa* Mill.),
3. Common hazel (*Corylus avellana* L.),
4. European beech (*Fagus sylvatica* L.),
5. Olive tree (*Olea europaea* L.),
6. Norway spruce (*Picea abies* L. H. Karst.),
7. Aleppo pine (*Pinus halepensis* Mill.),
8. Austrian pine (*Pinus nigra* J. F. Arnold),
9. Stone pine (*Pinus pinea* L.),
10. Scots pine (*Pinus sylvestris* L.),
11. Sweet cherry (*Prunus avium* L.),
12. Turkey oak (*Quercus cerris* L.),
13. Holm oak (*Quercus ilex* L.),
14. Common oak (*Quercus robur* L.),
15. Cork oak (*Quercus suber* L.),
16. Goat willow (*Salix caprea* L.).

2.4 Training points

2.4.1 Preparing and combining legacy occurrence points

A total of 2,454,997 tree species occurrence locations from three different sources was gathered. The majority of points (71%) comes from the Global Biodiversity Information Facility (GBIF) which is an open database fed by field observations from individual researchers. National forest inventory data from multiple EU member states published by Mauri et al. (2017) forms another 23% of the dataset. The remaining 6% comes from the LUCAS dataset.

All sources were filtered, harmonized and supported by additional information. Entries were filtered for species included in the European Atlas of Forest Tree Species (San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, 2016). Occurrences with a taxonomy rank other than species or genus were omitted. Same applies to points which had flags indicating serious location issues (i.e. missing coordinates). Geometries were re-projected to coordinate reference system ETRS89 / LAEA Europe (EPSG: 3035). A high resolution land mask for Europe (Hengl et al., 2020) was applied to further

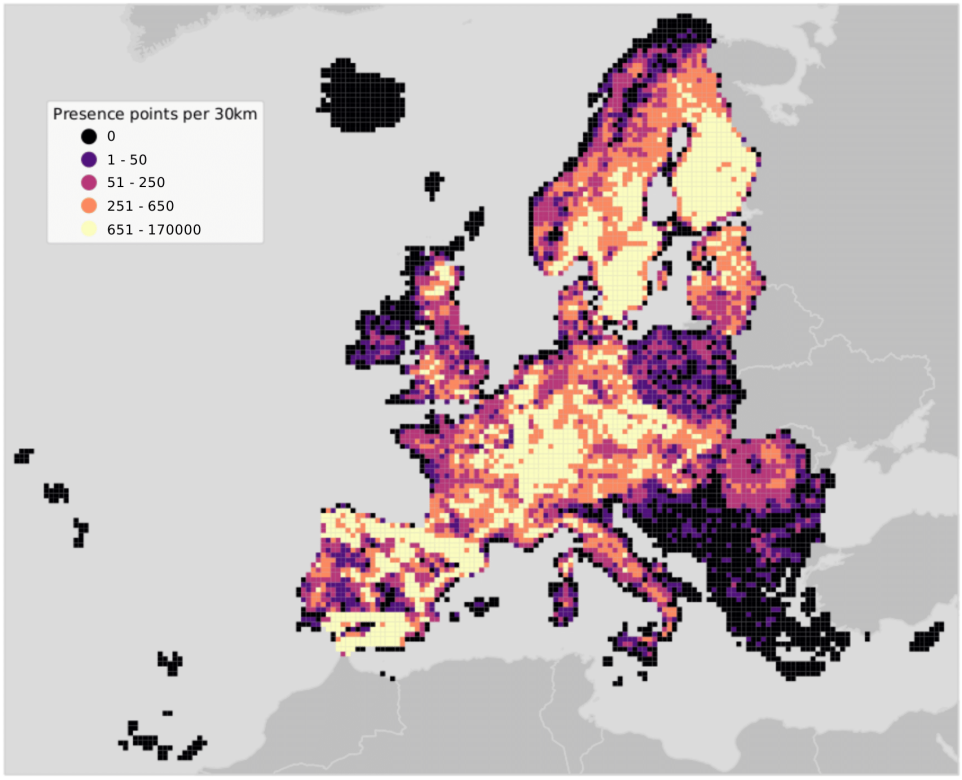


Figure 2. Map of the study area overlaid with a grid of 30 km tiles used for spatial 5–fold cross validation. Pixel value shows the number of presence points per tile.

exclude misplaced occurrence points. GBIF taxon and genus keys were derived for the other two data sources. Quality flag variables for location accuracy and date were established from existing metadata to indicate potentially problematic entries. The harmonized point dataset has information on species and genus (including respective GBIF keys), year of observation, country, original data source, citation, and license among other auxiliary variables. Its generation is reproducible. The dataset was published separately and is openly available along with the code and individual GBIF dataset citations (Heisig and Hengl, 2020).

We used yearly forest masks derived from Witjes et al. (2021) to decide upon including point data lacking the year of observation. Witjes et al. (2021) provides yearly probability maps at 30 m for the period 2000–2020 for 43 land cover classes according to the CLC level 3 legend. We overlaid the points with the probability maps for the classes:

- 311: Broad-leaved forest,
- 312: Coniferous forest,
- 313: Mixed forest,
- 323: Sclerophyllous forest,

- 324: Transitional woodland-shrub,
- 333: Sparsely vegetated area.

Points were used only if the probability value extracted for at least one of the classes was $\geq$50% for all the years considered. Each unique combination of longitude, latitude and year was then considered as an independent sample. An additional quality flag was added to distinguish points coming from this operation and the points with original year of observation coming from source datasets.

2.4.2 Preparing non-occurrence points

A total of 883,630 land cover points was gathered from the LUCAS database as provided by Eurostat and used as absence data. All LUCAS survey data (2006, 2009, 2012, 2015 and 2019) was used: each survey was first downloaded individually and then aggregated. As for the occurrence points, each unique combination of longitude, latitude and year was considered as an independent sample. The survey assigns each location as belonging to one of the following 8 main categories:

- A: Artificial land,
- B: Cropland,
- C: Woodland,
- D: Shrubland,
- E: Grassland,
- F: Bareland,
- G: Water,
- H: Wetlands.

The above classes were used for selecting observations for the absence dataset. Points in class C were excluded as absence data as that class already served the selection of presence data. For modelling the actual distribution, all remaining points were included in the first selection, while for the potential distribution points coming from classes A and B were excluded. For modelling the potential distribution the selection of absence points was more restrictive. Points were overlaid with a rasterized chorological map for each of the target species downloaded from the European Atlas of Forest Tree Species portal (see subsection: 2.1). Only points falling outside the area indicated by the chorological map were used as absence for the potential distribution.

2.5 Predictor variables

A total of 585 harmonized variables covering continental Europe (Hengl et al., 2020) at different spatial resolution were used as predictors to model the realized distribution of the species. In this study we included both dynamic (i.e. time-series of data of different temporal resolution) variables covering the time period January 2000 – December 2020 and static (i.e. variables not expected to change during the

modelled time period) variables. A subset of only 318 variables were used instead to model the potential distribution (see Fig. 1). All data was reprojected in the Projected coordinate system for Europe (EPSG code: 3035) before the analysis.

2.5.1 Dynamic data

We used a reprocessed version of Landsat ARD data provided by GLAD (Potapov et al., 2020): time series used in this study covers the period 1999–2020. Cloud and cloud shadow pixels were removed from the images, maintaining only the quality assessment-QA values labeled as clear-sky. Afterwards, individual images were averaged by season according to three different quantiles (25th, 50th and 75th) and the following calendar dates for all period:

- Winter: December 2 of previous year until March 20 of current year,
- Spring: March 21 until June 24 of current year,
- Summer: June 25 until September 12 of current year,
- Fall: September 13 until December 1 of current year.

84 images (3 quantiles × 4 seasons × 7 Landsat bands) were produced for each year. Missing values were imputed using the *Temporal Moving Window Median* algorithm. For more details on the preprocessing of Landsat data for this study see Witjes et al. (2021). 7 different spectral indices were computed for each year and season using the 50th quantile only, for a total of 7 x 4 = 28 spectral indices variables per year.

Table 1. Table with Landsat-derived spectral indices used in this study.

Spectral Index	Abbreviation	Formula	Reference
Enhanced Vegetation Index	EVI	$2.5 \times \frac{NIR - RED}{NIR + 6 \times RED - 7.5 \times BLUE + 1}$	(Huete et al., 2002)
Enhanced Vegetation Index 2	EVI2	$2.5 \times \frac{NIR - RED}{NIR + 2.4 \times RED + 1}$	(Jiang et al., 2008)
Modified Soild Adjusted Vegetation Index	MSAVI	$\frac{(2 \times NIR + 1) - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - RED)}}{2}$	(Qi et al., 1994)
Normalized Burned Ratio	NBR	$\frac{NIR - SWIR2}{NIR + SWIR2}$	(Key and Benson, 1999)
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - RED}{NIR + RED}$	(Tucker, 1979)
Normalized Difference Wetness Index	NDWI	$\frac{NIR - SWIR1}{NIR + SWIR1}$	(Gao, 1996)
Soil Adjusted Vegetation Index	SAVI	$(1 + 0.5) \times \frac{NIR - RED}{(NIR + RED + 0.5)}$	(Huete, 1988)

A reprocessing of the ERA5 Land hourly dataset has been used to have monthly aggregates of air temperature (2 meters above ground), surface temperature and precipitation. Original ERA5 data was aggregated to daily data, and subsequently to monthly data, with increased resolution (1 km) using CHELSA data (Karger et al., 2020): in this way the general spatial and temporal pattern of ERA5 Land dataset was kept while using the fine spatial detail coming from the CHELSA dataset. The following steps were used for temperature data:

1. aggregate CHELSA to ERA5 spatial resolution,

2. calculate difference between ERA5 Land and aggregated CHELSA,
3. interpolate differences with a Gaussian filter to 30 arc seconds,
4. add the interpolated differences to CHELSA.

A different approach was used for precipitation, with proportions instead of differences: using proportions ensures that areas without recorded precipitation remain areas without precipitation; only in the case of actual precipitation in a given area, precipitation was redistributed according to the spatial detail of CHELSA:

1. aggregate CHELSA to ERA5 spatial resolution,
2. calculate proportion between ERA5 Land and aggregated CHELSA,
3. interpolate proportion with a Gaussian filter to 30 arc seconds,
4. multiply the interpolated proportion with CHELSA.

For air and surface temperature we obtained the monthly minimum, mean and maximum, while for precipitation the monthly sum. We then computed the standard deviation of each of these variables and added these as additional predictor variables for modeling. To simulate the cumulative effect of temperature and precipitation regimes on a short term scale (5 years), we also calculated the monthly average of the previous 5 years for each variable, for a total of 336 climatic time series layers.

2.5.2 Static covariate datasets

As additional static covariates, we used the following datasets:

- 19 bioclimatic variables (Hijmans et al., 2005) for the period 1979 - 2013 to provide a baseline of the actual state of the climate; we however decided use the bioclimatic variables from the CHELSA dataset (1 km spatial resolution) since CHELSA has been claimed to better match data from meteorological stations than WorldClim (Karger et al., 2017). Layers were freely downloaded from https://chelsa-climate.org/bioclim/,
- 50 different chorological maps downloaded from the Europena Atlas of Forest Tree Species web portal. Chorological maps provide a qualitative overview of the spatial distribution of a species over an area, differentiating between native and introduced. We considered both the native and introduced areas as the potential distribution of a species for the time period covered by the study. The maps are harmonized products derived from different historical biobliographic sources or from the EUFORGEN website (http://www.euforgen.org); they are provided as vectorial layers and the minimum mapping unit may vary across the species. To include these layers in the models, we first rasterized the maps on a 10 km grid, assigning a value of 100 or 0, respectively, where the species could potentially be present or not. We then resampled the maps at a spatial resolution of 30 m using the cubic spline method to create a smoothed, continuous area around the reported range,

- Global bare ground cover from Hansen et al. (2013). The layer is derived from annual composites from 2000 to 2012 and provides information on bare ground cover on a percent (1–100) scale at 30 m resolution; it was downloaded from https://glad.umd.edu/dataset/global-2010-bare-ground-30-m,
- Solar direct and diffuse irradiation images at 1 km resolution were downloaded from https://globalsolaratlas.info/download,
- 13 cloud fraction layers (monthly averages and annual average) derived from MODIS (Wilson and Jetz, 2016) at 1 km resolution obtained from https://www.earthenv.org/cloud,
- Digital terrain model (DTM) for Europe (Hengl et al., 2020) and DTM-derived (slope, hillshade) variables at 30 m resolution, available along with its reproducible code at https://doi.org/10.5281/zenodo.4724549,
- Easterness, northness derived in GRASS GIS (Olaya, 2009), and positive and negative openness derived using SAGA GIS and available via https://doi.org/10.5281/zenodo.4486135,
- Probability of surface water occurrence at 30 m resolution derived from Landsat time series (Pekel et al., 2016), obtained from and freely available at https://global-surface-water.appspot.com/,
- Height above nearest drainage (HAND) and flow accumulation area at 90 m resolution from the MERIT Hydro global hydrography datasets, freely available at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/,
- Long-term flood hazard map calculated on a 500 years time period at 1 km resolution (Dottori et al., 2016),
- Continental Europe surface lithology based on the European Geological Data Infrastructure (EGDI) at 1:1M scale produced by GEOZS, Slovenia, and available at https://doi.org/10.5281/zenodo.4787631,
- Water vapor pressure (kPa) based on the WorldClim2.1 dataset (Fick and Hijmans, 2017); obtained from http://www.worldclim.com/version2,
- Long-term snow probability (2000–2012) at 500 m resolution based on the MODIS (MOD10A2) and available at https://doi.org/10.5281/zenodo.5774953,
- Monthly wind speed (1998–2018) at 5 km resolution from TerraClimate available at https://www.climatologylab.org/terraclimate.html.

2.6 Species datasets

Points were overlaid with the predictor variables and two presence-absence datasets were produced for each species, one to be used for potential distribution and one for realized distribution. The final dataset contains predictor variables plus 8 columns with metadata used to uniquely identify the points:

- id: unique point identifier,
- year: year of observation,
- postprocess: quality flag to identify if the temporal reference of an observation comes from the original dataset or is the result of spatiotemporal overlay with forest masks,
- tile id: contains the tile id from the 30 km grid,
- easting: longitude coordinates in Coordinate Reference System ETRS89 / LAEA Europe (= EPSG code 3035),
- northing: latitude coordinates in Coordinate Reference System ETRS89 / LAEA Europe (= EPSG code 3035),
- atlas class: name of the tree species according to the European Atlas of Forest Tree Species or NULL in case of absence point,
- lc1: contains original LUCAS land cover class or NULL if it's a presence point.

While absence points are homogeneously distributed in the study area, the same is not true for presence points. Fig. 2 shows the number of presence points for all the species selected in this study aggregated using a 30 km grid: clusters of points can be observed in Scandinavia, France and Spain, while some countries (i.e. Cyprus, Iceland) have no points at all.

To obtain a spatially balanced presence-absence dataset for each species, we empirically defined a fixed amount of points to be selected from each tile: we first overlaid the points with a 30 km grid and counted the amount of points per tile. We selected the highest number scored by each species and then calculated the median of the distribution made by all the highest scores. Results of this thinning operation can be seen in Table 2 and Fig. 3.

2.7 Feature selection

Features for potential and realized distribution for each species were selected using the Recursive Feature Eliminitation (RFE) strategy, implemented in the Python programming language and publicly available in the scikit-learn library. For each combination of species and modelled distribution we trained a random forest classifier (num.trees = 50, default values were used for the other parameters): RFE fits the model and removes the weakest feature (or features) until a specified number of features is reached, then ranks the importance of the features based on the model's coefficients (for regression-based models) or feature importance (for random forest).

The minimum number of features was not known before hand: to select this number, we ran the Recursive Feature Elimination with a spatial 5–fold Cross Validation (RFECV), using the logarithmic loss, or logloss, as a scoring estimator. Logloss is one of the most robust performance metric when it comes to imbalanced datasets (Ferri et al., 2009). Logloss is indicative of how close the predicted probability for

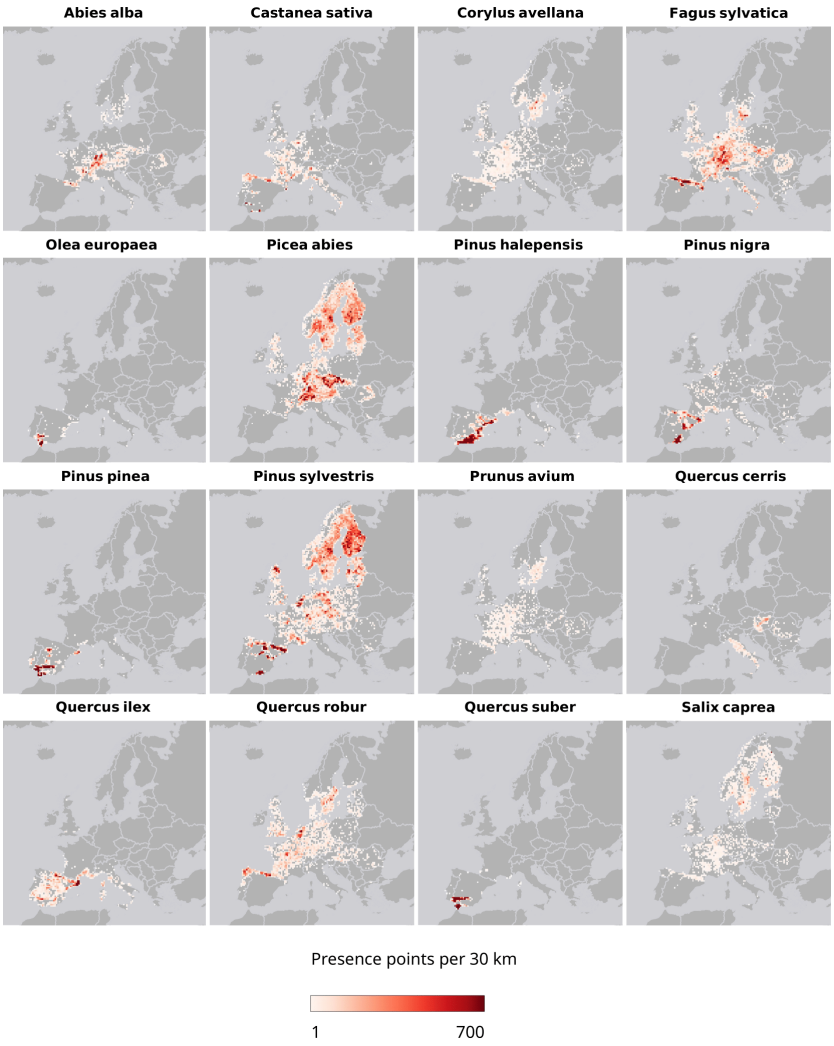


Figure 3. Distribution of presence points per species after thinning. Absence points are omitted for visualization purposes.

an observation i is to the corresponding label y. For binary classification with label $y \in 0, 1$ the overall logloss was calculated as:

$$f^{*}_{\text{Logistic}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \ln\left[p(y_i)\right] + (1 - y_i) \cdot \ln\left[1 - p(y_i)\right] \quad (1)$$

where N is the total number of observations and $p(y_i)$is the predicted probability for an observation with $y_i = 1$. It follows that values close to 0 indicate high prediction performances, with logloss = 0 being a perfect match, and values that are positive to infinite are progressively worse scores. For comparison, the value of logloss for random assignment depends on the number of classes (a) and the prevalence of the classes (b): for binary classification and a balanced (50:50) dataset with $N = 10$ observations, the equation (1) gives a value of 0.69.

We ran the RFECV on a 25% random subsample for each species and modelled distribution; this

Table 2. Number of presence and absence points used to model potential and realized distribution for each species before and after the thinning operation.

Distribution	Species	Presence	Absence	Presence thinned	Absence thinned	Prevalence	Prevalence thinned
Potential	*Abies alba*	45,495	108,603	45,410	108,445	0.42	0.42
Potential	*Castanea sativa*	77,382	107,667	49,269	107,363	0.72	0.46
Potential	*Corylus avellana*	32,321	64,135	32,141	63,955	0.50	0.51
Potential	*Fagus sylvatica*	197,240	55,595	180,817	55,266	3.50	3.30
Potential	*Olea europaea*	50,656	191,830	12,761	191,769	0.26	0.07
Potential	*Picea abies*	360,271	172,879	351,063	172,769	2.10	2.00
Potential	*Pinus halepensis*	233,964	180,795	71,258	179,931	1.30	0.40
Potential	*Pinus nigra*	139,516	225,382	54,739	224,752	0.62	0.24
Potential	*Pinus pinea*	239,254	221,263	41,188	221,050	1.10	0.19
Potential	*Pinus sylvestris*	507,681	44,652	415,859	44,077	11.00	9.40
Potential	*Prunus avium*	22,978	88,517	22,862	88,401	0.26	0.26
Potential	*Quercus cerris*	13,856	104,782	13,774	104,700	0.13	0.13
Potential	*Quercus ilex*	57,690	203,273	52,213	203,207	0.28	0.26
Potential	*Quercus robur*	113,044	52,493	111,399	52,190	2.20	2.10
Potential	*Quercus suber*	419,975	211,403	26,361	211,253	2.00	0.12
Potential	*Salix caprea*	45,967	76,263	45,789	76,082	0.60	0.60
Realized	*Abies alba*	45,495	558,564	44,882	558,493	0.08	0.08
Realized	*Castanea sativa*	77,382	558,564	48,969	554,503	0.14	0.09
Realized	*Corylus avellana*	32,321	558,564	32,141	554,781	0.06	0.06
Realized	*Fagus sylvatica*	197,240	558,564	179,108	557,490	0.35	0.32
Realized	*Olea europaea*	50,656	558,564	12,483	542,354	0.09	0.02
Realized	*Picea abies*	360,271	558,564	348,682	557,571	0.64	0.63
Realized	*Pinus halepensis*	233,964	558,564	68,400	553,831	0.42	0.12
Realized	*Pinus nigra*	139,516	558,564	53,717	557,050	0.25	0.10
Realized	*Pinus pinea*	239,254	558,564	39,766	555,684	0.43	0.07
Realized	*Pinus sylvestris*	507,681	558,564	412,502	555,730	0.91	0.74
Realized	*Prunus avium*	22,978	558,564	22,862	558,556	0.04	0.04
Realized	*Quercus cerris*	13,856	558,564	13,774	558,559	0.03	0.03
Realized	*Quercus ilex*	57,690	558,564	52,213	558,280	0.10	0.09
Realized	*Quercus robur*	113,044	558,564	110,748	558,267	0.20	0.20
Realized	*Quercus suber*	419,975	558,564	25,617	556,598	0.75	0.05
Realized	*Salix caprea*	45,967	558,564	45,769	558,558	0.08	0.08

operation was replicated 5 times. For each iteration we selected the minimum of the function showed in Fig. 4 and the averaged result was then used as the minimum number of features for the RFE.

2.8 Model building and evaluation

2.8.1 Modeling methods

To build an ensemble model, we decided to compare predictive performances and computing time (hyperparameter tuning — cross validation — prediction time) of different machine learning algorithms on a random 25% subset of observations for both potential and realized distribution datasets. A detailed workflow of this process is shown in Fig. 5. We decided to conduct this test on seven different species: choice of the species was based on the spatial distribution of the training points and the ratio between presence and absence points. In this way, algorithms performances can be tested on different ecological conditions (latitudinal and longitudinal gradient) and imbalance of classes. The species selected were: *Abies alba*, *Castanea sativa*, *Fagus sylvatica*, *Picea abies*, *Pinus halepensis* and *Pinus sylvestris*.
The following algorithms were compared:

- Random Forest (RF) (Breiman, 2001a),
- Gradient-boosted trees (GBT) (Friedman, 2002),
- Classification trees (CART) (Therneau and Atkinson, 2011),

- Generalized Linear Models (Nelder and Wedderburn, 1972) with Lasso regularization (Tibshirani, 1996) (just GLM from now on),
- C5.0 (Quinlan, 1986),
- K-nearest neighbor (KNN) (Fix and Hodges, 1989),
- Artificial Neural Network (ANN) (Ripley and Venables, 2017),

Analyses were conducted using the *mlr* (Machine Learning for R) package (Bischl et al., 2016): the package provides a standardized framework that wrap different machine learning algorithms already implemented in R in other independent packages. It also provides hyperparameter tuning with different optimization strategies, resampling functionalities (bootstrapping, cross validation), benchmarking and visualizations through the *ggplot2* package. For each algorithm, an hyperparameter space was defined: combinations of hyperparameters were generated per model based on a grid search of 5 steps per hyperparameter. Due to computational constraints, we set the *num.trees* parameter for Random forest to 85. The tested activation functions for the neural network were *sigmoid* and *tanh*, while for the output we tested both *sigmoid* and *softmax*. For GLM we used the automatically generated λ sequence and selected the λ_{min}. The rest of the hyperparameter ranges are shown in Table 3.

2.8.2 Selecting component models

We evaluated each combination of hyperparameters by comparing logarithmic loss values during a 5–fold spatial cross validation replicated 5 times: we used spatial cross validation instead of normal cross validation for hyperparameter tuning because it educes overoptimistic performance results in the presence of strong data clustering (Schratz et al., 2019). We used the tile ID produced in the tiling system for

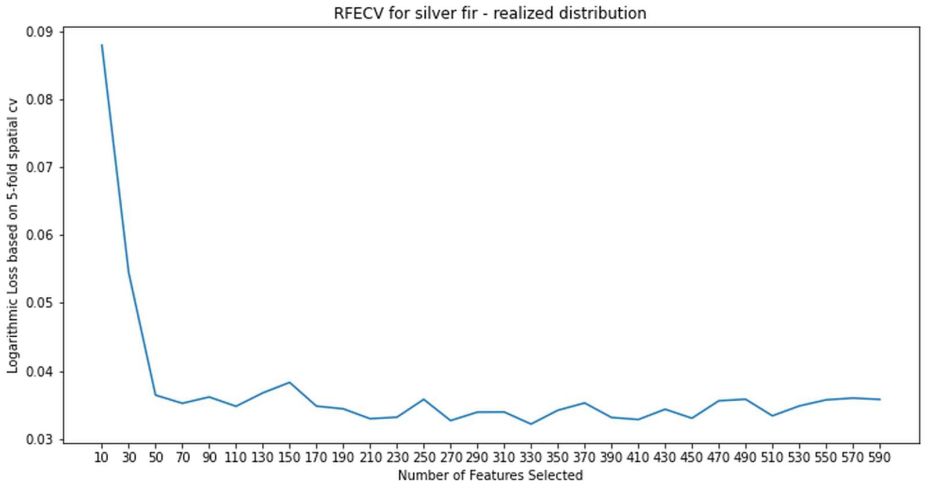


Figure 4. Log Loss performances by number of selected features. In this iteration ran for the realized distribution of the silver fir, 330 is the minimum of the function.

Europe (Coordinate Reference System ETRS89 / LAEA Europe, epsg:3035) as the blocking parameter in the training function in *mlr*. All the compared algorithms were used in *"probability"* mode, that is, predicting for each observation in the dataset a probability value for presence (class 1) and absence (class 0). Besides the performance achieved in the logloss metric, computing time for the hyperparameter tuning, a 5–fold spatial cross validation and prediction time for a 30 km tile were also considered as additional criteria: we calculated these two metrics only for the species that had the highest computational costs (*Pinus sylvestris*). This gave us an estimate of how long the process of training each component model could take during the building of the ensemble model. We used logloss performance as the first criteria to choose the component models: only in the case of two or more methods performing within one standard deviation from the average performance, we chose the computationally fastest one across all areas (hyperparameter tuning — cross validation — prediction) and tasks (potential — realized).

2.8.3 Training ensemble model using stacking

Stacked generalization involves combining predictions made by level 0 models and using them as training data for a level 1 model (or meta-model or meta-learner from now on) (Wolpert, 1992). To avoid or limit overfitting, there are currently two approaches used to prepare the training data: via hold-out method, which splits the original dataset in training and test, or via k–fold cross validation. We used the latter in this study, via a 5–fold spatial cross validation: the out-of-fold predictions were used to build a level 1

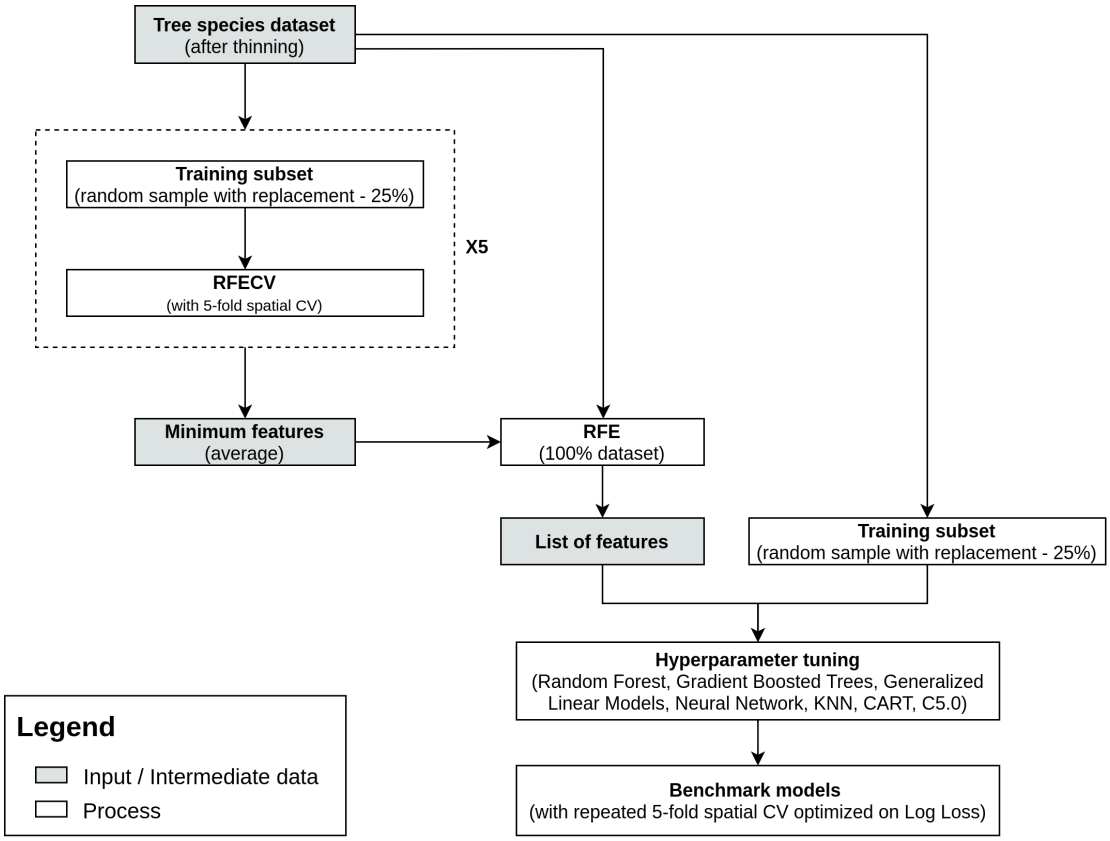


Figure 5. Example workflow illustrating the feature selection and benchmarking process for one species

Table 3. Hyperparameter space for the analyzed algorithms. In light gray the name of the R package used to implement the algorithm is reported, in brackets the name of the algorithm. p refers to the number of predictor variables, while columns *Lower* and *Upper* indicate the bounds of the regions in the hyperparameter space.

Algorithm	Hyperparameter	Type	Lower	Upper
C5.0				
(Classification trees)	minCases	integer	0	10
	CF	numeric	0	0.5
kknn				
(k-nearest neighbor)	k	integer	1	50
deepnet				
(Artificial neural network)	learning rate	numeric	0.0001	0.00001
	numepochs	integer	10	20
	batchsize	integer	50	150
	hidden_dropout	numeric	0.1	0.3
	activationfunction	discrete	-	-
	output	discrete	-	-
	momentum	numeric	0	0.05
	number.of.layers	integer	2	4
	units	integer	32	64
ranger				
(Random forest)	mtry	integer	$\sqrt{p}/3$	p
rpart				
(CART)	minsplit	integer	20	25
	minbucket	integer	5	10
	cp	numeric	0.01	0.1
	maxcompete	integer	3	4
	maxsurrogate	integer	4	5
	usesurrogate	discrete	-	-
	surrogatestyle	discrete	-	-
	maxdepth	integer	5	15
xgboost				
(gradient-boosted trees)	nrounds	integer	10	20
	max_depth	integer	3	5
	eta	numeric	0.01	0.1
	subsample	numeric	0.5	0.9
	min_child_weight	integer	10	20
	colsample_bytree	numeric	0.5	0.9

training dataset for the meta-learner. We used logistic regression with Lasso regularization (Tibshirani, 1996) as a meta-learner, which is usually the most used model for classification problems (Gomes et al., 2012). Final predictions are delivered as probability maps (0–100%) for presence together with model uncertainty maps: we consider as model uncertainty the standard deviation of the predicted values of the base learners. The principle is that the higher the standard deviation the more uncertain the model is towards the right value to assign to the pixel (Brown et al., 2020).

2.8.4 Variable importance assessment

To assess to what extent the three level 0 models used different parts of the available feature space and the agreement between these models, we compared the variable importance when possible. For Random forest and CART we used Gini importance, for C5.0 the *"percentage of training set samples that fall into*

all the terminal nodes after the split"(Quinlan, 1986), for Gradient boosted trees the gain metric (Shi et al., 2019) and for Generalized linear models the coefficients for the minimum fitted value of λ (Hastie et al., 2016).

2.8.5 Model evaluation

Predictive performances of the ensemble model was assessed through spatial 5–fold cross-validation repeated 5 times (Roberts et al., 2017) with logloss as performance metric. To investigate if the ensemble model outperformed the component models, we compared results of the spatial cross validation of the ensemble with the results of the component models. The area under the ROC curve (AUC) is a commonly used metric to evaluate SDMs predictive performances due to being threshold independent (Shabani et al., 2018); however, it is also more sensitive to prevalence than logloss (Ferri et al., 2009), hence our choice of logloss as a performance metric to compare different models coming from different training datasets.

3 RESULTS

3.1 Spatio-temporal ensemble machine learning framework

Table 4 shows the logloss performances based on the spatial 5–fold cross-validation repeated 5 times. This clearly indicates that GLM and Random forest have the highest predictive performances for all species. Both algorithms scored the lowest logloss among the other algorithms in 7 cases out of 14, with GLM excelling in potential distribution tasks and Random forest having better performances in the realized distribution tasks.

Table 4. Average logloss and standard deviation for the compared algorithms and for the subset of seven target species. In bold are highlighted the best performing learners for each task.

Distribution	Species	ANN	C5.0	GBT	GLM	KNN	RF	CART
Potential	*Abies alba*	0.170±0.014	0.042±0.003	0.089±0.001	**0.009±0.001**	0.057±0.012	0.023±0.001	0.057±0.005
Potential	*Castanea sativa*	0.253±0.018	0.055±0.004	0.100±0.003	**0.015±0.002**	0.066±0.016	0.028±0.003	0.082±0.011
Potential	*Fagus sylvatica*	0.423±0.014	0.069±0.005	0.115±0.002	**0.021±0.001**	0.076±0.009	0.036±0.002	0.170±0.008
Potential	*Picea abies*	0.450±0.012	0.070±0.002	0.120±0.001	**0.032±0.001**	0.091±0.008	0.040±0.002	0.142±0.008
Potential	*Pinus halepensis*	0.341±0.033	0.035±0.001	0.087±0.002	**0.008±0.001**	0.049±0.010	0.016±0.003	0.076±0.017
Potential	*Pinus sylvestris*	0.451±0.016	0.080±0.004	0.134±0.001	0.050±0.001	0.107±0.005	**0.043±0.003**	0.212±0.006
Potential	*Quercus robur*	0.401±0.018	0.074±0.006	0.120±0.002	**0.023±0.001**	0.070±0.009	0.040±0.003	0.156±0.006
Realized	*Abies alba*	0.628±0.004	0.143±0.002	0.138±0.007	0.074±0.007	0.206±0.039	**0.069±0.007**	0.189±0.016
Realized	*Castanea sativa*	0.565±0.019	0.247±0.074	0.187±0.020	0.161±0.018	0.372±0.068	**0.107±0.021**	0.271±0.035
Realized	*Fagus sylvatica*	0.544±0.013	0.094±0.017	0.113±0.003	0.050±0.005	0.145±0.034	**0.048±0.006**	0.127±0.016
Realized	*Picea abies*	0.650±0.058	0.122±0.009	0.168±0.003	0.093±0.001	0.244±0.022	**0.079±0.005**	0.244±0.015
Realized	*Pinus halepensis*	0.595±0.028	0.125±0.034	0.127±0.004	**0.039±0.004**	0.112±0.033	0.053±0.009	0.143±0.027
Realized	*Pinus sylvestris*	0.316±0.016	0.072±0.005	0.114±0.003	0.056±0.006	0.103±0.017	**0.042±0.006**	0.111±0.011
Realized	*Quercus robur*	0.627±0.039	0.183±0.024	0.180±0.008	0.107±0.009	0.249±0.032	**0.098±0.007**	0.248±0.024

Overall, GLM shows best performance in sense of achieving the lowest values of standard deviation. Random forest, however, scored close to GLM hence the difference does not seem to be significant. The absolute difference between values scored by GLM and Random forest is lower (0.089) than when Random forest had the advantage over GLM (0.105). This indicates a high reliability of Random forest performances even when other models outperform it. The neural network scored the highest logloss values in all tasks, so it was immediately excluded from the pool of level 0 models to choose from. It was time consuming to find a common hyperparameter range well suited for different tasks, since neural networks are often extremely situation-dependent. After a preliminary selection, we used the range shown in Table

3: despite that, our results remained inferior to those obtained with the other learners. On top of that, the *mlr* implementation of neural networks, based on the *deepnet* package, doesn't allow the use of ReLU (rectified linear activation function) as an activation function, which would have been beneficial for our purposes. Based on logloss performances, we selected RF and GLM as the first two components of the ensemble. Based on similar values of logloss (within one standard deviation of the average performance) scored by C5.0, GBT, KNN and CART, we used computational costs to choose the third component model. Table 5 shows the required time by each algorithm to run the hyperparameter tuning, 5–fold spatial cross validation and to predict a probability map on one tile. KNN was excluded due to computing time values being from one to two order of magnitude higher than the ones scored by the other models.

Table 5. Hyperparameter tuning, cross validation and prediction time for each model and distribution task. Time values are reported in seconds. Tests were conducted in a parallel computing setup on a CPU server running 2 x Intel(R) Xeon(R) Gold 6248R - 3.00GHz (96 threads) with 504 GB RAM.

Distribution	Process	ANN	C5.0	GBT	GLM	KNN	RF	Rpart
Potential	Tuning	661.2	220.7	680.1	345.9	2334.6	318.3	1019.3
Potential	Cross validation	54.8	45.5	34.4	305.7	582.5	426.1	14.5
Potential	Prediction	24.12	231.1	24.8	14.8	19272.9	53.4	14.1
Realized	Tuning	1772.3	944.5	851.1	427.2	16198.4	1873.8	3298.1
Realized	Cross validation	16.2	184.8	20.5	344.1	2919.9	2347.5	39.9
Realized	Prediction	19.3	389.3	32.6	17.8	$>$ 1 day	78.4	16.4
	Total	**2547.92**	**2015.9**	**1643.5**	**1455.5**	$>$ **1 day**	**5097.5**	**4402.3**

Even though the CART scores very low value in cross validation and prediction time in both potential and realized tasks, tuning time is the second highest, just behind KNN. C5.0 is faster than GBT in the whole potential workflow (497.3 seconds against 739.3) but slower in the realized workflow (1518.6 seconds against 904.2). Considering both workflows, GBT proved to be faster and more consistent in cross validation and prediction time, showing an increase in tuning time of just 25% with double the amount of training data (see Table 2). After the training of the component models and the meta-learner, this is an example of final fitted ensemble model:

```
Species: Pinus_sylvestris
Distribution: Potential
Logloss: 0.0255

Call:
stats::glm(formula = f, family = "binomial", data = getTaskData(.task,
    .subset), weights = .weights, model = FALSE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5683   0.0586   0.0587   0.0595   3.4006

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -5.78660    0.05221 -110.83   <2e-16 ***
classif.ranger  8.12492    0.10953   74.18   <2e-16 ***
```

```
classif.xgboost  1.49208    0.08233   18.12   <2e-16 ***
classif.glmnet   2.53462    0.07296   34.74   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 290519  on 459919  degrees of freedom
Residual deviance:  23472  on 459916  degrees of freedom
AIC: 23480

Number of Fisher Scoring iterations: 8
```

The output shows that all algorithms are significant predictors for the response variable (column $Pr(>|z|)$), while from the coefficient estimate values we infer that predictions coming from Random forest have, for this specific case, the highest weight in the final predictions of the meta-learner, followed by GLM with Lasso and finally GBT.

3.2 Variable importance

We generated the variable importance for the three component models for all species and distributions: for each of these combinations we computed the relative variable importance and then took the top–20 variables from each set. This resulted in a total of 146 unique variables for potential distributions and 178 for realized distributions; of those, 54 variables are present in both potential and realized set of important variables. To better analyze the results and identify driving factors of species distributions, we aggregated the whole set of 270 variables in 7 macro-classes:

- Climate (i.e. precipitation, wind speed, water vapor, snow probability etc.),
- Temperature (i.e. time series of recorded temperatures for the observed time period),
- Bioclim (i.e. bioclimatic variables from CHELSA),
- Topography (i.e. DTM and DTM-derivative variables),
- Landsat band (i.e. all percentiles, all seasons),
- Distribution (i.e. species distribution maps from European Atlas of Forest Tree Species),
- Spectral index (i.e. spectral indices derived from Landsat bands).

Results are presented in Fig. 6. Each plot can be divided in four quadrants, from the top left clockwise: variables with high relative importance but low frequency (i.e. important for one or few species), variables with high importance and high frequency (i.e. important for all species), variables with low importance and high frequency (i.e. they occured often but were not important) and variables with low importance and low frequency.

For potential distribution, diffuse irradiation, precipitation of the driest quarter (BIO17) and precipitation of the driest month (BIO14) were the most important and most frequent predictors across all models

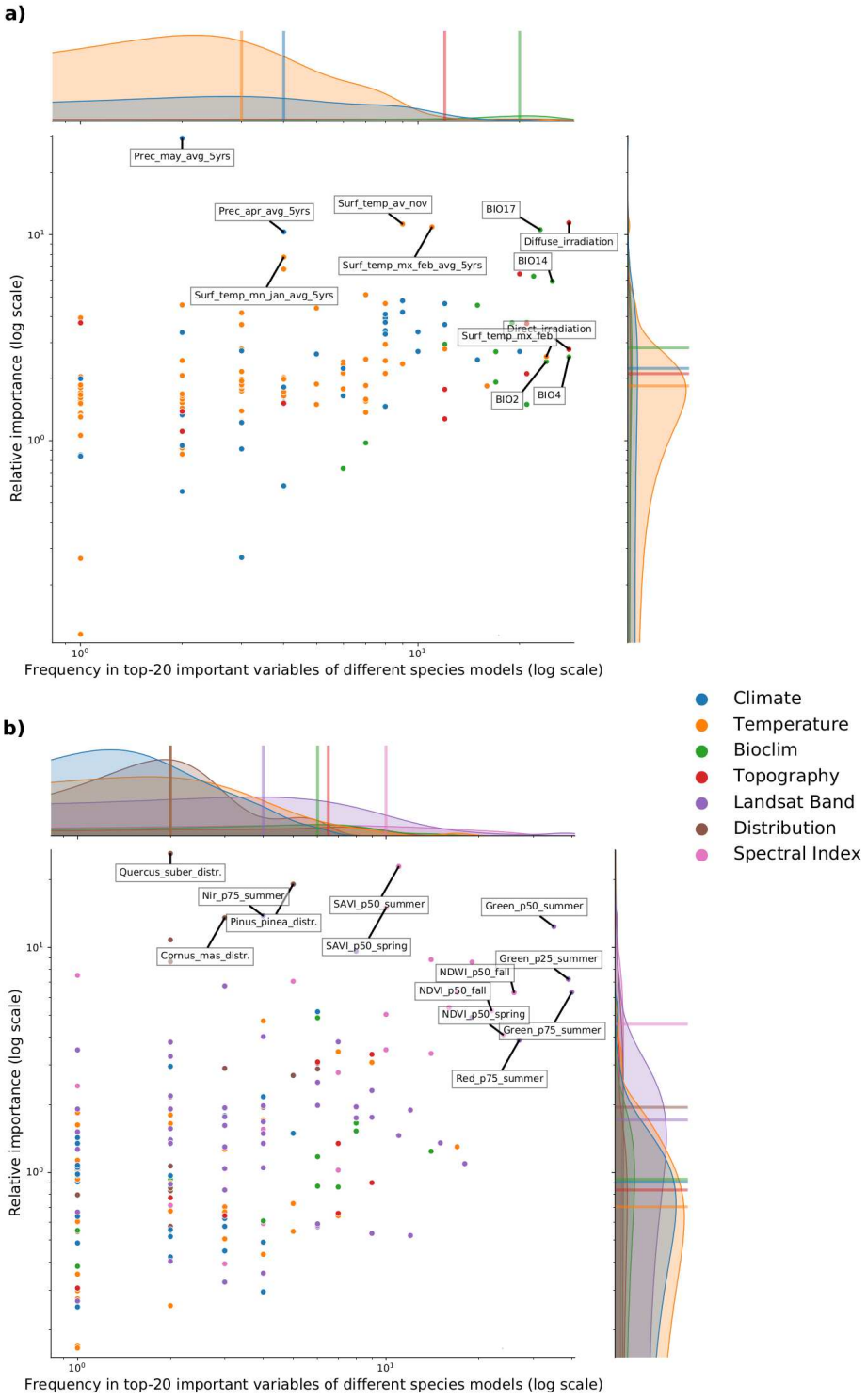


Figure 6. Relative variable importance vs frequency of the variables of the top–20 most important across the component models and all species for potential **(a)** and realized **(b)** distribution. Variables were aggregated in seven different classes, colors are shown in legend. Labeled dots are variables that recorded high values of relative variable importance or frequency: each plot shows in the top right corner the most important and most frequent variables across all models and species.

and species. The density distributions per macro-class help understanding how the Bioclim macro-class was the one with on average both most important and most frequent variables. Other variables are more species-specific: the cumulative precipitation of May computed on a 5 years time window records the highest absolute value in relative importance but it was important for just two species (*Abies alba* and *Castanea sativa*, see additional statistical outputs at https://zenodo.org/record/5821865). The Temperature macro-class accounts the highest numbers of predictors, but the values recorded in both variable importance and frequency are the lowest among all the macro-classes. The Climate macro-class had the largest variety in predictors and variables in this class are homogeneously spread out across all the species in both variable importance and frequency.

For realized distribution, the summer aggregates of Landsat green (25th and 50th quantiles) were the two most important and most frequent variables across all models and species, closely followed by the fall aggregates of NDVI and NDWI and summer and spring aggregates of SAVI. Reflectance-based macro-classes clearly outperformed the other ones in this case. The distribution maps scored the highest values for variable importance (distribution of the *Quercus suber* and the *Pinus pinea*) but they were species-specific (i.e. important for just three species, see additional statistical outputs at https://zenodo.org/record/5821865). Despite the green Landsat band scoring the best values across all the other predictors, the Spectral index macro-class scored on average as the most important in both variable importance and frequency.

Overall, the component models show more differences in variable importance in the potential distribution models than in the realized ones. On average, Random forest and Gradient Boosted Trees selected the same variables in the top–10 but not always in the same order, while GLM tended to choose completely different variables. This suggests how the ensemble models tend to use a wider proportion of the feature space than single models. This tendency is most apparent in the potential distribution models. In the realized distribution models, the component models agree in selecting the top-10 most important variables predictors from Landsat bands or Spectral indices. Random forest and Gradient Boosted Trees considered on average the Landsat bands as the most important, while GLM selected the spectral indices more often.

3.3 Accuracy assessment

We ran a 5–fold spatial cross validation repeated 5 times with the tuned component models and the ensemble model to provide conservative estimates of predictive performance. Predictions on the left-out folds were aggregated to assess model performance. Table 6 shows logloss and standard deviation values for component models and ensemble model across all species and distributions compared with the value of logloss for the intercept-only model. All models were better than random assignment. In general, the models for realized distribution achieved better predictive performances than those for potential distribution. This does not hold for three species: *Fagus sylvatica, Olea europaea* and *Quercus suber*. The ensemble model:

- **outperformed** the component models in **15** cases of the potential distribution tasks but only in **one** case of the realized distribution tasks,
- performed **as good as** the best component model in **one** case of the potential distribution tasks and

in **five** cases of the realized distribution tasks,

- performed **worse** than the best component model in **ten** cases of the realized distribution tasks, with small differences.

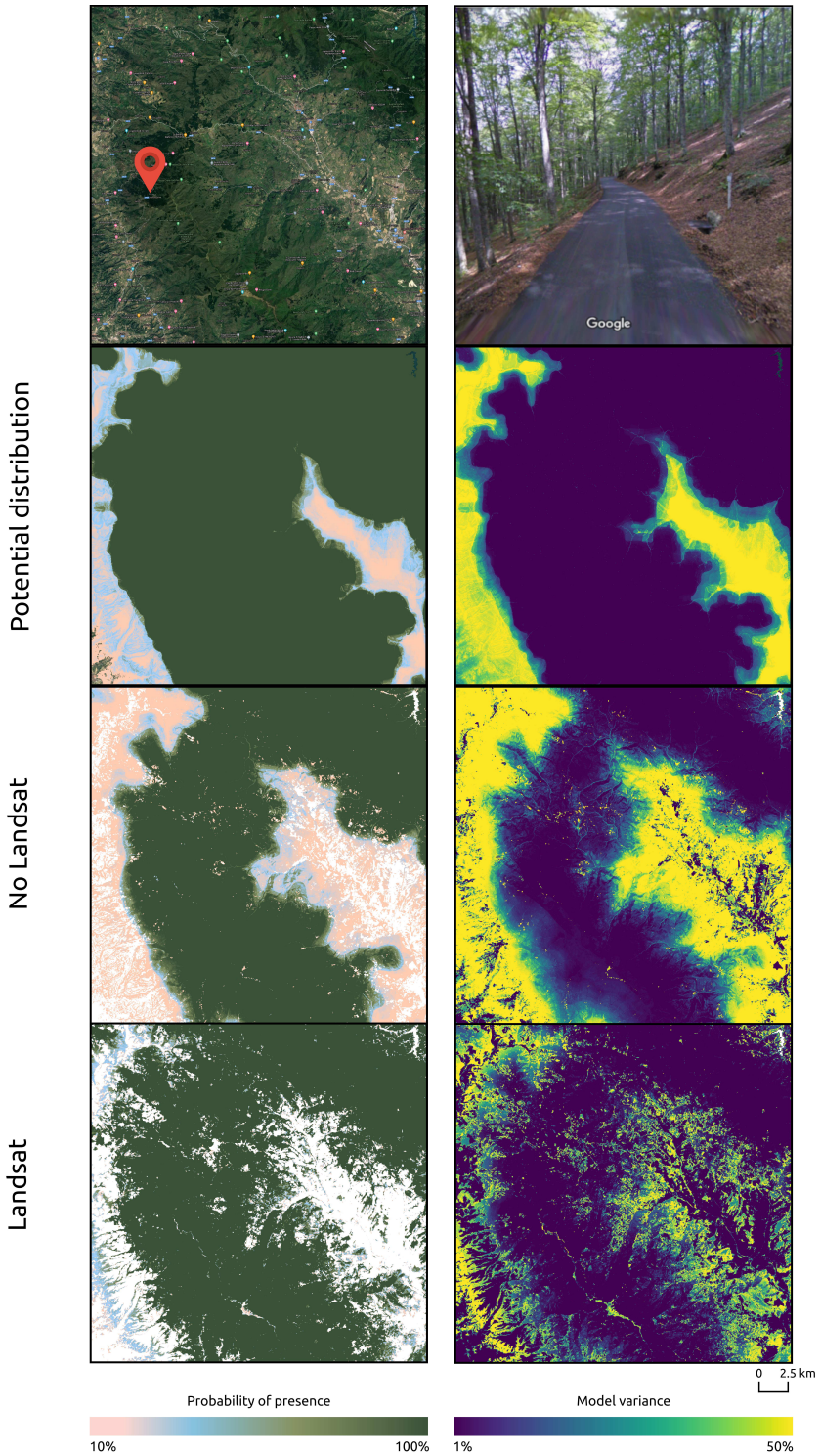
Table 6. Average logloss and standard deviation for the component learners and the ensemble model. In bold are highlighted the best performing learners for each task. Random logloss values are dependent on presence-absence ratio in the species dataset and are here used as a baseline for predictive performances comparison.

Distribution	Species	GBT	GLM	RF	Ensemble	Random
Potential	*Abies alba*	0.062 ±0.008	0.071 ±0.005	0.071 ±0.011	**0.046 ±0.005**	0.606
Potential	*Castanea sativa*	0.133 ±0.052	0.148 ±0.016	0.119 ±0.027	**0.080 ±0.013**	0.622
Potential	*Corylus avellana*	0.109 ±0.018	0.125 ±0.008	0.118 ±0.027	**0.080 ±0.009**	0.638
Potential	*Fagus sylvatica*	0.032 ±0.004	0.074 ±0.007	0.048 ±0.008	**0.029 ±0.003**	0.544
Potential	*Olea europaea*	0.007 ±0.004	**0.005 ±0.002**	0.008 ±0.004	**0.005 ±0.002**	0.233
Potential	*Picea abies*	0.069 ±0.004	0.089 ±0.005	0.072 ±0.005	**0.067 ±0.004**	0.634
Potential	*Pinus halepensis*	0.048 ±0.012	0.037 ±0.005	0.057 ±0.011	**0.033 ±0.006**	0.596
Potential	*Pinus nigra*	0.091 ±0.011	0.101 ±0.008	0.117 ±0.012	**0.077 ±0.005**	0.494
Potential	*Pinus pinea*	0.033 ±0.007	0.035 ±0.004	0.047 ±0.012	**0.023 ±0.004**	0.434
Potential	*Pinus sylvestris*	0.031 ±0.003	0.055 ±0.003	0.035 ±0.004	**0.026 ±0.002**	0.315
Potential	*Prunus avium*	0.096 ±0.010	0.115 ±0.007	0.103 ±0.019	**0.071 ±0.006**	0.509
Potential	*Quercus cerris*	0.063 ±0.017	0.037 ±0.009	0.049 ±0.011	**0.024 ±0.005**	0.360
Potential	*Quercus ilex*	0.071 ±0.008	0.094 ±0.007	0.071 ±0.015	**0.058 ±0.007**	0.508
Potential	*Quercus robur*	0.075 ±0.007	0.113 ±0.010	0.084 ±0.006	**0.069 ±0.007**	0.626
Potential	*Quercus suber*	0.016 ±0.012	0.011 ±0.004	0.021 ±0.012	**0.009 ±0.003**	0.348
Potential	*Salix caprea*	0.106 ±0.012	0.138 ±0.010	0.117 ±0.011	**0.091 ±0.008**	0.662
Realized	*Abies alba*	0.017 ±0.002	**0.014 ±0.002**	0.022 ±0.002	0.016 ±0.002	0.264
Realized	*Castanea sativa*	**0.024 ±0.004**	0.026 ±0.003	0.031 ±0.004	0.025 ±0.002	0.281
Realized	*Corylus avellana*	0.024 ±0.003	**0.02 ±0.002**	0.031 ±0.002	0.022 ±0.002	0.212
Realized	*Fagus sylvatica*	**0.030 ±0.002**	0.031 ±0.001	0.043 ±0.001	**0.030 ±0.002**	0.554
Realized	*Olea europaea*	0.012 ±0.004	**0.006 ±0.001**	0.012 ±0.004	0.008 ±0.002	0.107
Realized	*Picea abies*	**0.037 ±0.002**	0.043 ±0.002	0.049 ±0.001	0.040 ±0.002	0.666
Realized	*Pinus halepensis*	0.013 ±0.002	**0.011 ±0.001**	0.020 ±0.003	**0.011 ±0.001**	0.346
Realized	*Pinus nigra*	0.018 ±0.002	**0.016 ±0.002**	0.027 ±0.004	**0.016 ±0.002**	0.297
Realized	*Pinus pinea*	0.013 ±0.005	**0.009 ±0.002**	0.019 ±0.005	0.010 ±0.002	0.245
Realized	*Pinus sylvestris*	**0.043 ±0.001**	0.053 ±0.001	0.058 ±0.001	0.045 ±0.002	0.682
Realized	*Prunus avium*	0.022 ±0.002	**0.020 ±0.002**	0.030 ±0.002	**0.020 ±0.002**	0.165
Realized	*Quercus cerris*	0.009 ±0.002	**0.006 ±0.001**	0.013 ±0.003	0.007 ±0.002	0.113
Realized	*Quercus ilex*	0.023 ±0.004	0.019 ±0.002	0.027 ±0.004	**0.018 ±0.002**	0.291
Realized	*Quercus robur*	0.041 ±0.002	**0.036 ±0.002**	0.045 ±0.003	**0.036 ±0.001**	0.448
Realized	*Quercus suber*	0.012 ±0.008	**0.005 ±0.001**	0.011 ±0.004	0.006 ±0.002	0.180
Realized	*Salix caprea*	**0.028 ±0.002**	0.031 ±0.002	0.035 ±0.002	0.030 ±0.002	0.268

Among the best performing models, the ensemble for potential distribution of *Olea europaea* and GLM for realized distribution of *Quercus suber* achieved both the absolute and relative (i.e. scaled with the random value) lowest values of logloss. The ensemble for potential distribution of *Salix caprea* achieved the highest absolute value, while the highest relative value was achieved by the ensemble for potential distribution of *Pinus nigra*.

3.4 Influence of high resolution on predictive performances

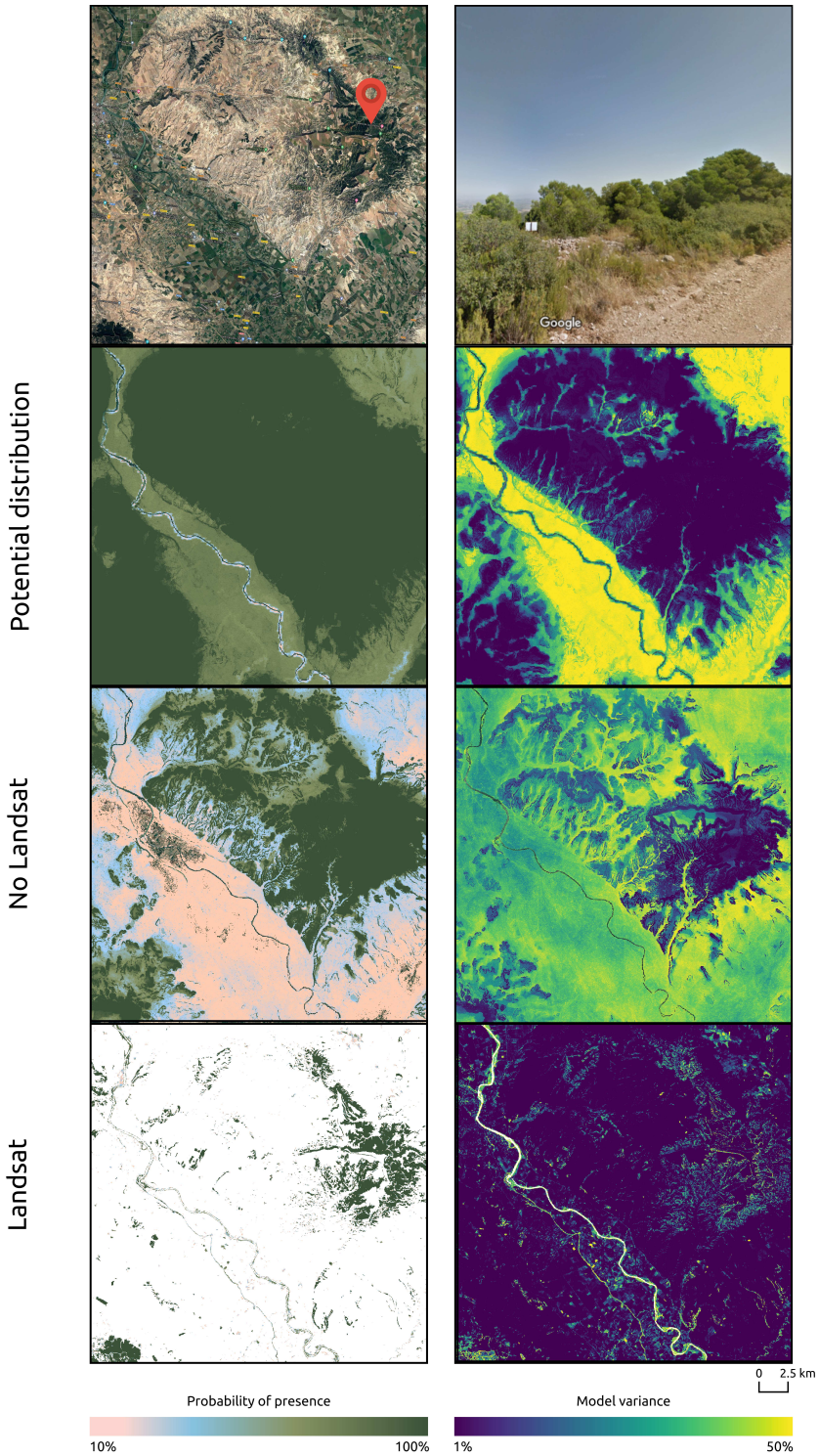
To assess the effect of high resolution products on predictive performances, we excluded Landsat bands and Landsat-derived spectral indices from the list of predictors used for realized distribution. We then



649

Figure 7. Comparison of potential distribution with realized distribution for the period 2018–2020 for *Fagus sylvatica* on a 30 km tile located in Tuscany, Italy. This area was chosen due to the presence of a national park (Foreste Casentinesi National Park) and a Natura 2000 area (Vallombrosa and Sant'Antonio forest) well known for *Fagus sylvatica* stands. Satellite view provided by Google Maps (left) and ground view provided by Google Street View (right) are shown in Row 1. Row 2 to row 4 show probability (left) and uncertainty (right) maps. Two realized distributions are shown: excluding and including Landsat data among the predictor variables.

650



651

Figure 8. Comparison of potential distribution with realized distribution for the period 2018–2020 for *Pinus halepensis* on a 30 km tile located in the Ebro Basin, Spain. The species was introduced in this area in the Mesolithic period. Satellite view provided by Google Maps (left) and ground view provided by Google Street View (right) are shown in Row 1. Row 2 to row 4 show probability (left) and uncertainty (right) maps. Two realized distributions are shown: excluding and including Landsat data among the predictor variables

652

applied our spatio-temporal ensemble machine learning framework (feature selection — hyperparameter tuning — ensemble model training) on each species and we ran a 5–fold spatial cross validation repeated 5 times to evaluate model performances. For the ensemble model we used the same component models (Random forest, GBT and penalized GLM) and meta-learner (penalized logistic regression). Results of this analysis were then compared with the performances achieved by the ensemble models using Landsat data. Comparison is shown in Table 7: for all species, ensemble models including Landsat data consistently overperformed models of the same species without the Landsat data. In all cases, models including Landsat data show better predictive performances (lower values of logloss) and less uncertainty (lower values of standard deviation). Fig. 7 and 8 show a visual comparison of the models for two species (*Fagus sylvatica* and *Pinus halepensis*) on sample areas.

Table 7. Average logloss for modelling realized distribution with and without the Landsat bands and spectral indices. Random logloss values are shown as a baseline for predictive performances.

Species	Landsat	No Landsat	Random
Abies alba	0.016 ±0.002	0.036 ±0.003	0.264
Castanea sativa	0.025 ±0.002	0.065 ±0.007	0.281
Corylus avellana	0.022 ±0.002	0.045 ±0.003	0.212
Fagus sylvatica	0.030 ±0.002	0.088 ±0.003	0.554
Olea europaea	0.008 ±0.002	0.010 ±0.003	0.107
Picea abies	0.040 ±0.002	0.093 ±0.002	0.666
Pinus halepensis	0.011 ±0.001	0.041 ±0.004	0.346
Pinus nigra	0.016 ±0.002	0.039 ±0.004	0.297
Pinus pinea	0.010 ±0.002	0.018 ±0.003	0.245
Pinus sylvestris	0.045 ±0.002	0.104 ±0.003	0.682
Prunus avium	0.020 ±0.002	0.041 ±0.003	0.165
Quercus cerris	0.007 ±0.002	0.018 ±0.003	0.113
Quercus ilex	0.018 ±0.002	0.032 ±0.003	0.291
Quercus robur	0.036 ±0.001	0.108 ±0.006	0.448
Quercus suber	0.006 ±0.002	0.010 ±0.003	0.180
Salix caprea	0.030 ±0.002	0.051 ±0.003	0.268

4 DISCUSSION

4.1 Summary findings

In this paper we have developed, tested and reported a methodological framework for predicting the potential and realized distributions of 16 forest tree species by Ensemble Machine Learning and analysis-ready Earth Observation data. In general, our ensemble model achieved better predictive performances than individual models when modeling the potential distribution, while performing slightly worse than the best component model for ten species when modeling the realized distribution.

Even at high resolution (i.e. on a local scale), climate proved to be the key driver of vegetation potential distribution across Europe mainly through temperature and precipitation. These results are consistent with findings of Hutchinson (1957) and Mather and Yoshioka (1968). Reflectance-based covariates were the most important predictors of the realized distributions. Overall, potential distribution proved to be significantly more complex to map accurately than realized distribution. In general, the ensemble and component models achieved better predictive performances for the realized distributions

than for the potential distributions as judged from the cross-validation logloss estimates (see Table 6 and Fig. 9).

Our results indicate a consistent increase in predictive performances for realized distribution when adding high resolution data, especially Landsat data at 30 m resolution and vegetation indices to the list of predictors (see Table 7). Significant findings of our work include:

1. Predictive modeling of forest tree species can be efficiently automated to the level of full automation, but this assumes high quality / artifact free training points with a homogenous distribution of occurrence and absence points whenever possible (Fig. 3)

2. Complexity of Ensemble Machine Learning methods can be significantly reduced by implementing the following two generic processes: (1) hyperparameter tuning on data subsets, (2) running efficient feature selection (Fig. 4).

3. Analysis-ready Landsat time-series are maybe cumbersome to prepare and gap-fill for clouds and artifacts (Witjes et al., 2021), but overall come as the most important inputs for mapping forest tree species, hence the Landsat project stands out as the most important data source for continuous and consistent time-series forest mapping.

Compared to the previous application of Machine Learning methods for mapping forest tree species over the whole European continent described in Hengl et al. (2018), we have identified the following methodological improvements:

- A framework to face the problem of preferential sampling and absence data unavailability to create a presence-absence datasets mainly by using LUCAS points (EUROSTAT, 2017). In contrast, Hengl et al. (2018) and de Rigo et al. (2016a) used only occurrence points, which probably introduced bias (i.e. overestimation or underestimations) in the predictions.

- A framework for spatio-temporal predictive mapping. We built a single model per forest tree species to predict through the whole spatio-temporal datacube cube and produce time-series of predictions. Maps can then be used as input data for further analysis / work, i.e. to detect species geographical range contraction or expansion, areas of forest degradation / restoration, species composition changes.

- Methodological steps to help with reduction of the model complexity (Fig. 4): these have shown to be beneficial for reducing the production costs and enabled us to map more forest tree species.

4.2 Modeling framework

Combining models using the ensemble approach is thought to reduce model uncertainty and increase its robustness in modelling species distributions (Araújo and New, 2007). We used ensemble with stacked generalization as ensemble approach, which has not been tested yet for species distribution modelling. We also trained the models in a spatio-temporal framework, expecting the models to generalize better when predicting in a temporal window not included in the training data. Part of the intent of the paper is to provide a reproducible framework to model species distributions, so we compare our results with

those available from previous publications that use ensemble modeling for either potential or realized distributions. Our study, however:

- models both potential and realized distributions,
- uses only one functional group (trees) on one continent (Europe),
- adds several layer of complexity (ensemble based on stacking, spatio-temporal framework),
- uses a large (585) set of predictor variables,
- uses logloss as a performance measure,
- uses spatial cross validation,
- does not test model transferability,

so not all results from other studies can be directly compared. Hao et al. (2020) used a similar methodological framework to the one used in this study. They modelled the distribution of 13 species of the genus *Eucalyptus* in South Australia and tested performances of ensemble model against individual models; they used mean and weighted average as ensemble strategies. They also tested cross validation versus spatial cross validation for model performances. The study doesn't specify which type of distribution was modelled: according to the definition provided in this study and the list of predictor variables used by Hao et al. (2020), we can classify the task as potential distribution.

Their results show how spatial cross validation performances were more conservative than cross validation ones when compared with performances on independent validation sets. This supports and reinforces our use of spatial cross validation as a validation strategy for the modeling framework. Ensemble models performed well but were outperformed by not tuned individual models and by a tuned GBT. There was also no clear advantage in predictive performances when using different ensemble strategies. This is in contrast with our results, where the ensemble based on stacking outperformed even tuned component models in 15 cases of the potential distribution tasks and performing as good as the best component model in the last case. This may be an indication of stacking being a better ensemble strategy when modeling species distribution.

In our case, a tuned GBT outperformed the ensemble only in realized distribution tasks. By comparing Table 4 and 6 we can also see that Random forest and GLM are clearly the best component models to map both potential and realized distributions when trained on a data sample, but GBT often outperforms Random forest or even the ensemble when tuned and trained on the whole dataset. In general, differences in predictive performances between the ensemble and the component models are also higher in potential distribution than in realized distribution. The list of variable importance per component model, species and task (see additional statistical outputs at `https://zenodo.org/record/5821865`) may give an insight to this: in the potential tasks, the component models use different parts of the feature space before the predictions are combined by the meta-learner. All the models select as most important variables for the task different predictors. For realized distribution tasks, the models all agree in selecting either Landsat

bands or spectral indices as most important variables, resulting in predictions that are highly correlated and with less variance between the models.

Ensemble modeling is known to perform best when there is a high diversity between the base models (a) and no or negative correlation between their outputs (b) (Zhou, 2019). The introduction of Landsat bands and spectral indices in general greatly increased the predictive performances of the models for realized distribution (see Table 7) compared to potential distribution models (see Table 6). However, it also homogenized predictions, which makes the (b) condition reported above not valid. We separately compared the repeated spatial cross validation performances of ensemble and component models excluding the Landsat bands and spectral indices (see additional statistical outputs at https://zenodo.org/record/5821865). In this case, the ensemble performed worse than the best component model only in four cases instead of ten. While most of ensemble SDM studies use no more than 15–20 predictor variables (Kaky et al., 2020; Hao et al., 2019; Kindt, 2018), even after feature selection we ended up using more than 300 variables to model the realized distribution of some species. A stricter feature selection could maybe remove the problem in future works. In general, if the ensemble provides predictive performances as good as or worse than the best component model, the best component model must be preferred (Zhang and Ma, 2012). However, ensemble models can still provide more advantages than individual models since they reduce model uncertainty and are more robust towards extrapolation (Mehra et al., 2019).

4.3 Species distributions

Our cross-validation accuracy assessment results indicate high predictive performances for all species, in both potential and realized distributions. In the case of mapping potential distribution, diffuse irradiation and precipitation of the driest quarter (BIO17) come as overall most important covariate layers. By design, potential distribution maps produced in this study are indicators of where a species is not likely to survive (see section 2.1). Our interest was to map those areas deemed not suitable by the model and identify which of the considered predictor variables would have been the limiting factors. Which environmental variables and their relative importance as limiting factors are still unclear for many tree species and few peer reviewed studies focused on investigating the importance of different environmental predictors on potential distributions are available. We found that, on average, each component model considers two or more predictors from the Bioclim macro-class among the top-10 most important variables to predict the potential distribution (see additional statistical outputs at https://zenodo.org/record/5821865). Previous findings in literature have shown the importance of bioclimatic variables when modeling species distributions (Fourcade et al., 2018), but this may also be a consequence of bioclimatic variables and elevation being the most employed, if not the only, predictors in numerous SDM studies (Fois et al., 2018). Bucklin et al. (2015) compared the influence of different sets of environmental predictors on model performances, but the list of predictors used in the study included human influenced factors, so their results cannot be used to assess the driving factors for potential distributions. Pearson and Dawson (2003) argued that environmental predictors only are not able to describe all the processes that limit species ranges, but this claim is directed at realized distributions only. Even if our results show the bioclimatic variables as the most important predictors for potential distributions, further studies in this direction may

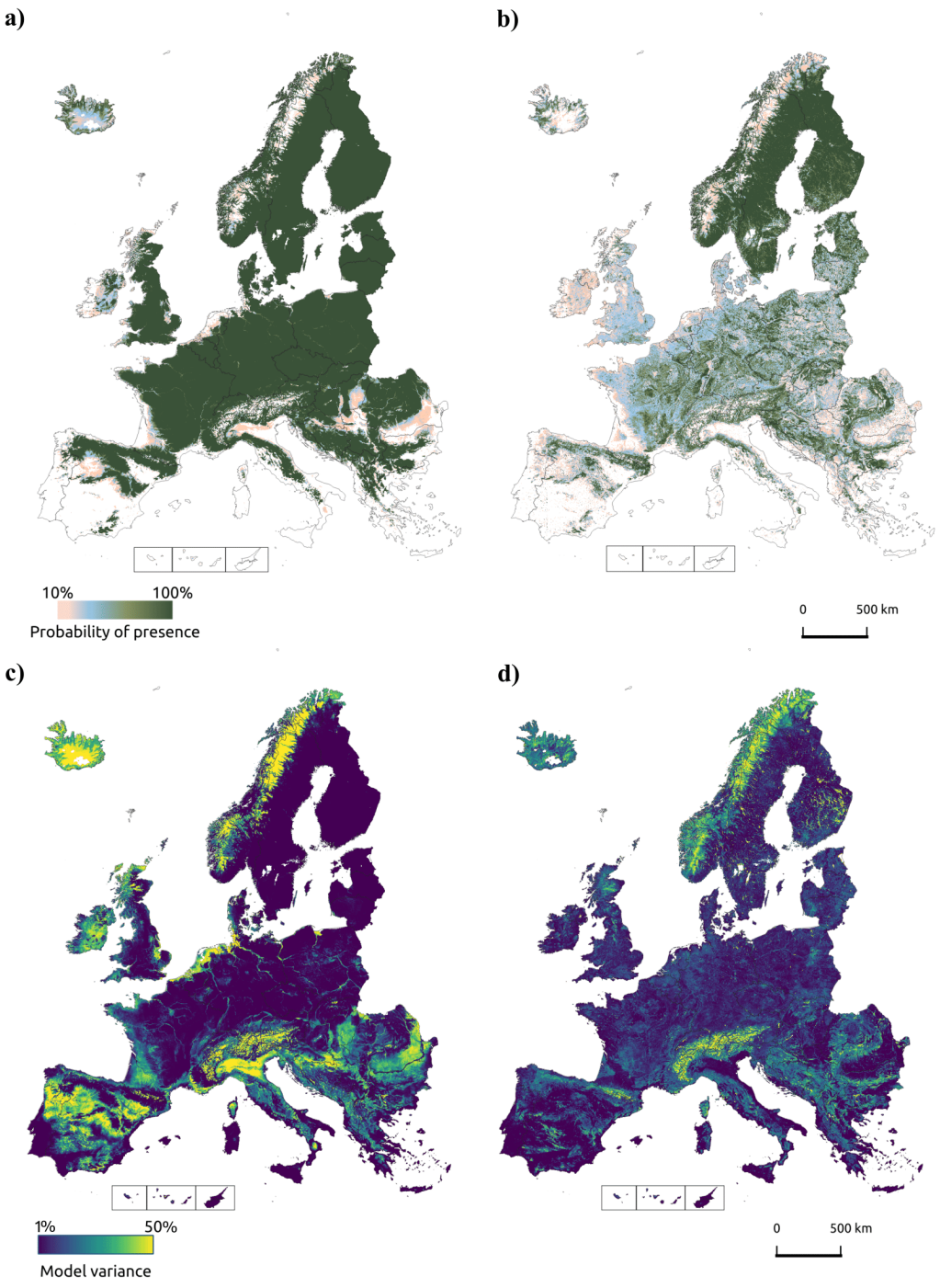


Figure 9. Potential and realized distribution for *Pinus sylvestris* for the period 2018–2020. Clockwise, from the top left corner: probability map for potential distribution **(a)**, probability map for realized distribution **(b)**, uncertainty map for potential distribution **(c)** and uncertainty map for realized distribution **(d)**.

be needed. The scale of the study may affect the importance of predictor variables: on a large scale, distribution may be influenced by macro environmental factors, while at a local scale, other environmental factors may limit mangrove distribution more significantly. Walthert and Meier (2017) and Weigel et al. (2019) proved that soil properties are more important than either bioclimatic or only climatic variables when modeling potential tree species distribution at, respectively, country and regional scale.

For realized distribution, variable importance confirms that Earth Observation layers such as the 25th and 50th quantile summer aggregates for the Landsat green band and the 50th quantile fall aggregates of NDVI and NDWI are overall the most important layers for mapping realized distribution of species (Fig. 6). The inclusion of Landsat data and derived spectral indices increases predictive performances (see Table 7) and contains more detailed information on species distribution ranges (see Fig. 7 and 8). Importance of NDVI is well known since it is one of the most used proxies in vegetation studies such as biodiversity estimation (Madonsela et al., 2017; He et al., 2009), net primary productivity (Schloss et al., 1999) and land degradation (Easdale et al., 2018), phenology (Fawcett et al., 2021) and species composition changes (Wang et al., 2021). NDVI incorporates information from the red and the near-infrared (NIR) portion of the electromagnetic spectrum (see Table 1). Vegetation's behavior in this portion of the spectrum has long been used in vegetation mapping to distinguish between coniferous and deciduous tree species (Hoffer, 1984). The green band, although usually less important than the red and NIR band, has already proved useful in vegetation mapping to classify forest types (Gao et al., 2015), predict forest variables (stem volume, diameter and tree height) at species level (Astola et al., 2019) and forest biomass at community level (Nandy et al., 2017).

Comparing our results with chorological maps from the European Atlas of Forest Tree Species (San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, 2016), we can see that in general both potential and realized distribution correctly capture the species ranges. Overall, potential distribution maps show homogeneous patterns of high probability values for all target species, while realized distribution maps show very fragmented patterns. Fig. 7 and 8 show sample areas for two different species where the potential distribution model indicates high probability for the target species. The realized distribution model helps discriminating the presence or absence of the species due to biotic or other external factors. A high geographical overlap between probability maps of realized distribution may reflect co-existence of different species within the same forest stands and could help with clearly defining forest communities.

4.4 High resolution contributions: is finer always better?

Bioclimatic variables available only at coarse spatial resolution were used as predictor variables in both potential and realized distribution. The Landsat bands and the spectral indices were not the only high resolution layers used in this study: terrain and terrain-derived predictors were also included at 30 m resolution. However, regardless of the terrain data high resolution, the tree species potential distribution patterns mostly reflect the original spatial resolution of the bioclimatic variables. Thus, climate influences species distribution on European scale. Even though this might indicate that mapping potential distributions at high resolution may not be necessary, it can still be useful for different case studies. For example, comparing the difference, and hence mapping the gap, between potential and

realized distribution at the same fine scale, as we show in Fig. 7 and 8, may prove to be an invaluable tool for both forest managers and conservation planners that work on the local level.

Potential distribution maps can be used to identify suitable areas for species in reforestation and restoration programs; realized distribution maps can inform the forest managers on the presence or absence of said species in those areas at a particular point in space and time (see Fig. 10). By removing the biotic factors that limit the presence of the species in a potential reforestation site, using multiple distribution maps and including expert knowledge on species synecology, structurally complex forest stands could be planned and developed in a much more informed way. The opposite approach could be used by conservation planners. Potential distribution is modelled by studying the relationship between a species and the environmental conditions found in its native range, where the species is at equilibrium (Jiménez-Valverde et al., 2011). Invasive species are usually more abundant and have greater performances in the introduced range than in their native ranges (Hierro et al., 2005). This is due to the absence of biotic factors that normally limit species distribution in their native range in the introduced range. Thus, a species that occupies only 10% of its potential distribution in its native range may end up occupying a bigger percentage of it in the introduced range. Estimation of potential distribution in the introduced range that depends only on environmental factors are conservative by definition, potential distribution maps may provide a good indication to conservation planners of how much the invasive species could spread in the introduced range.

For realized distribution, including high resolution predictor variables in the model not only increases predictive performances but also lowers overall (see Table 7) and local (see Fig. 7 and 8) values of uncertainty. For forest management purposes, a large, consistent, standardized, long-term and high resolution image collection such as the one provided by the Landsat program can help extending in space and time information on tree species presence, composition and abundance. A spatial resolution of 30 m is particularly well suited for national forest inventory applications: Strickland et al. (2020) derived probability maps of forest tree species for a 25 years time period (1985–2010) using yearly Landsat composites to extend missing information from the Canadian NFI and estimating changes in forest cover, species composition and forest disturbances. The increasing availability of even higher-spatial resolution satellite data from the European Copernicus program (i.e. Sentinel 1 and 2) and commercial providers (i.e. Planet) can potentially further enhance predictions by including more data and a better spatial matching of in-situ and satellite-derived information.

4.5 Technical limitations

The spatio-temporal Ensemble Machine Learning based on stacking is highly computational and requires significant investment in High Performance Computing. The total computing time required to conduct feature selection, hyperparameter tuning, model training, accuracy assessment and predictions of all probability and uncertainty maps described for just 1 species exceeds 25,000 CPU hours based on the Intel Xeon Gold chips, which within our infrastructure implied 19 days of non-stop computing with full capacity. This estimation doesn't include training data and predictor variables pre-processing and harmonization. We decreased processing costs by using Amazon S3 (Palankar et al., 2008) and full parallelization of all processes including spatio-temporal overlay (see `http://eumap.readthedocs.org/`), hyper-

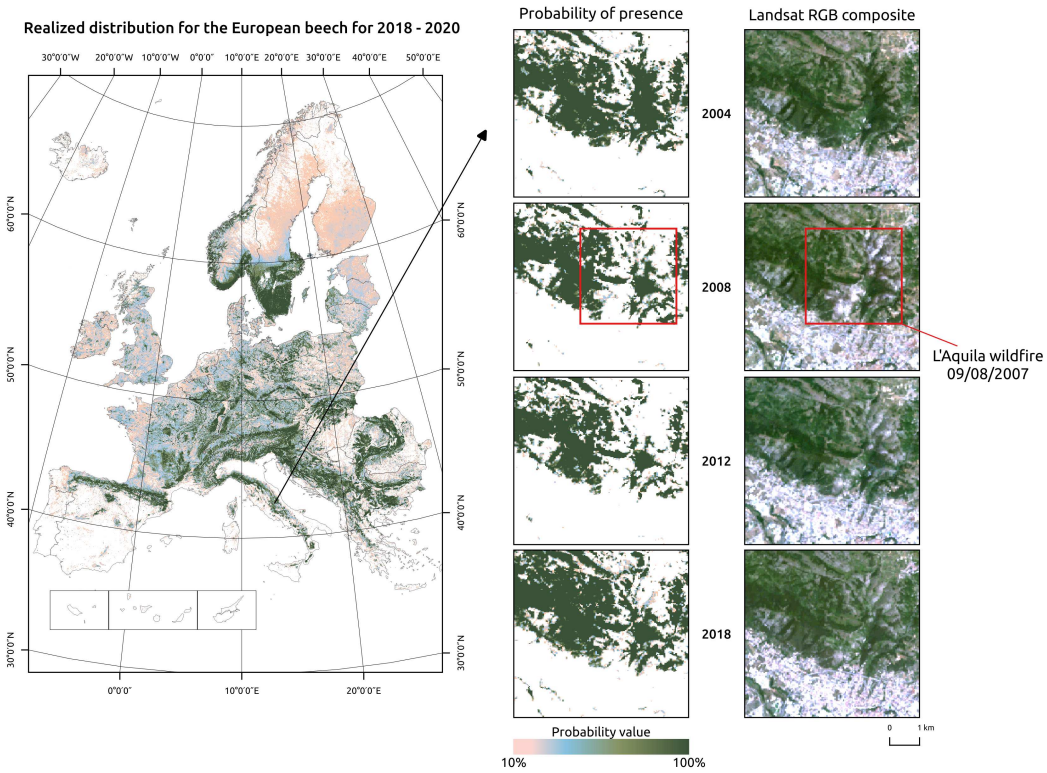


Figure 10. Realized distribution of *Fagus sylvatica* for the period 2018-2020. Detailed insets show a region around L'Aquila city, in Central Italy. The *Fagus sylvatica* forest on the northern outskirts of the city was affected by a serious wildfire in 2007. The realized distribution maps can be used to track changes through time.

parameter tuning and prediction, but the processing costs remain high. Due to the high production costs and data availability we have eventually decided to deviate from the original plan to map all forest tree species from the European Forest Atlas (76) to focus on a smaller subset (16).

Even though we achieved high values of predictive performances, this result was possible only thanks to a high quality and artifact free dataset. Density of observations in the dataset varies greatly across Europe and different species, leaving large areas either overrepresented or poorly covered (Fig. 2). Model uncertainty around predicted probability values depends strongly on both density and quality of observations, with areas scarcely covered or underrepresented in the datasets having higher uncertainty values or becoming source of extrapolation for the model. In our maps this can be seen in various locations in Norway (no absence data available) and Iceland (no presence or absence data available) across multiple species distribution maps, in both potential and realized distributions (Fig. 9).

Because our interest was to map distribution of forest species on a long time scale, we had to balance between using the best Earth Observation images and time-coverage (2000–2020). Eventually we have decided not to use any of the Sentinel-2 or similar products as they only span 2016–2020. In reality, Sentinel-1,2 images are more detailed (from 20 to 10 m resolution) and could probably help increase accuracy of the more recent years.

We have tried implementing predictions models that are holistic i.e. include inter-specific interactions

between target forest species and other species through species distribution maps; in practice this was not easy to implement because consistent, long-term distributions of species i.e. distribution maps from the European Atlas of Forest Tree Species (San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, 2016) are only available at relatively coarse spatial resolutions and not for all of the forest tree species. In addition, we have not considered other valuable information that can be classified among the biotic factors i.e. relative abundance/richness (Kass et al., 2020) or abundance of large individuals (Meier et al., 2010) due to data unavailability for both spatial and temporal scale of the study. The prediction errors are provided as standard deviation of the base learners. This approach is computationally acceptable, but could be further fine-tuned to increase accuracy of errors as in van den Hoogen et al. (2020).

4.6 Future work and directions

The predictions of forest tree species we have produced are possibly the most detailed and certainly the largest open datasets available to date. Although we had to innovate both Species Distribution Modelling methods and our computing and coding skills, we still recognize many future areas of improvements. These include:

- Fusing ALL Earth Observation data available e.g. Harmonized Landsat Sentinel-2 (HLS) and eventually Sentinel 1 data to help improve predictions.
- Adding and testing hyper-spectral images (e.g. from the future hyper-spectral missions such as ENMAP, see https://www.enmap.org/) for helping increase accuracy of predictions for especially species that grow under dominant species.
- Incorporating more species data, in particular those from NFI plots would further improve the species predictions at European and national scales.
- Adding more sophisticated and different Machine Learning algorithms such as Deep learning techniques (Lakshminarayanan et al., 2016) and similar.
- Modeling dynamics of forest tree species not as fields, but as objects through cellular automata or similar algorithms.
- Using spatio-temporal Machine Learning to predict future states of forest tree distributions using climate scenarios and similar.

We have released the maps and the code under open data / open source licenses to enable other similar research and to help speed up land restoration and reforestation projects in Europe. The code is available in our GitLab repository at https://gitlab.com/geoharmonizer_inea/spatial-layers/-/tree/master/veg_mapping. To suggest any improvement visit instead https://gitlab.com/geoharmonizer_inea/spatial-layers/-/issues. The predictions of tree species are available as Cloud-Optimized GeoTIFFs via https://maps.opendatascience.eu and can be displayed using compare tool in 2D and 3D. (Fig. 11).

European forest dynamics, even though some recent results indicate increased mortality in forest tree species e.g. Popkin (2021), is probably among the least troubling in comparison to other continents. Our

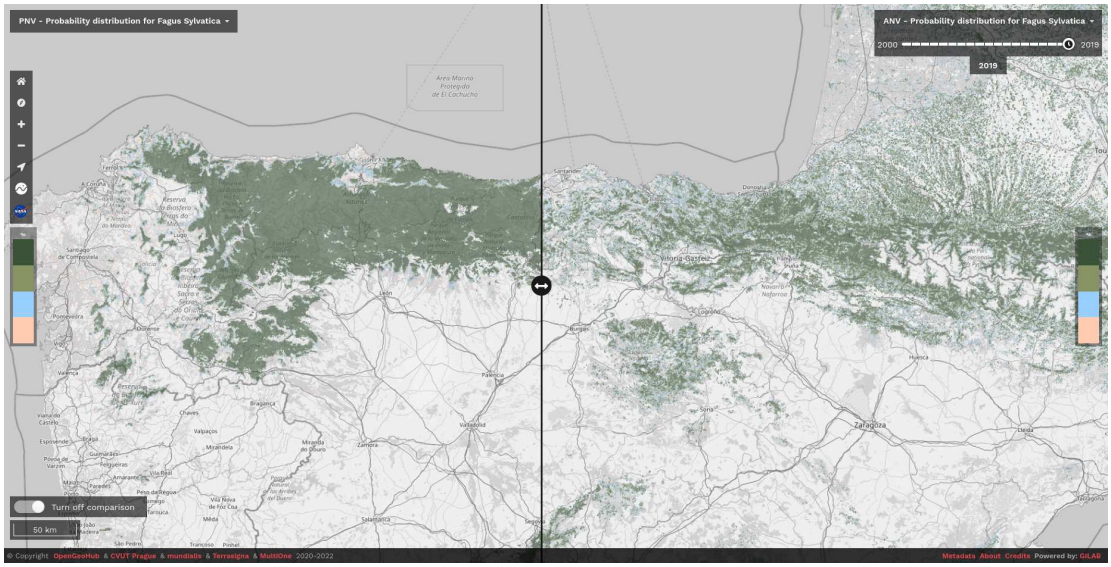


Figure 11. Difference between potential and realized distribution for *Fagus sylvatica* in Northern Spain for the period 2018–2020 visualized using slider in the Open Data Science Europe viewer (https://maps.opendatascience.eu).

methodological framework could thus potentially be implemented also at a global scale, and possibly through Google Earth Engine (van den Hoogen et al., 2021) or through the ESA's OpenEO platform (https://openeo.cloud/) to produce high resolution (10–30 m) predictions of global forest dynamics. Our estimate is that globally there are much more forest tree species important for forest management and monitoring i.e. at the scale of 500–1000 m. For example, in Brazil, it has been estimated that about 220 forest tree species cover most of the land and represent over 95 of the biomass i.e. so called *"hyper-dominant species"* (Draper et al., 2021). Scaling up approach described in this paper to help produce objective predictions and help monitor forest dynamics and support re-forestation efforts across globe is our next frontier.

ACKNOWLEDGEMENTS

This work is co-financed by the European Union CEF Telecom project 2018-EU-IA-0095. We are grateful to the GiLAB company from Belgrade, Serbia for their support with processing and publishing produced data via the opendatascience.eu data portal. We are also grateful to the Geo-harmonizer project partners CVUT Prague, mundialis, Terrasigna & MultiOne for helping with processing all LUCAS ground observations and quality control.

CONFLICT OF INTEREST

The authors confirm no conflict of interest.

REFERENCES

Araújo, M. B. and New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1):42–47.

Astola, H., Häme, T., Sirro, L., Molinier, M., and Kilpi, J. (2019). Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, 223:257–273.

Bechtel, B. and Schmidt, K. J. (2011). Floristic mapping data as a proxy for the mean urban heat island. *Climate Research*, 49(1):45–58.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., and Jones, Z. M. (2016). mlr: Machine learning in r. *Journal of Machine Learning Research*, 17(170):1–5.

Booth, T. H. (2018). Species distribution modelling tools and databases to assist managing forests under climate change. *Forest Ecology and Management*, 430(15):196–203.

Booth, T. H., Nix, H. A., Busby, J. R., and Hutchinson, M. F. (2014). BIOCLIM : the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, 20(1):1–9.

Bradley, B. A., Olsson, A. D., Wang, O., Dickson, B. G., Pelech, L., Sesnie, S. E., and Zachmann, L. J. (2012). Species detection vs. habitat suitability: Are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling*, 244:57–64.

Breiman, L. (2001a). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L. (2001b). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231.

Brown, K. E., Bhuiyan, F. A., and Talbert, D. A. (2020). Uncertainty quantification in multimodal ensembles of deep learners. In *The Thirty-Third International Flairs Conference*.

Brus, D., Hengeveld, G., Walvoort, D., Goedhart, P., Heidema, A., Nabuurs, G., and Gunia, K. (2012). Statistical mapping of tree species over europe. *European Journal of Forest Research*, 131(1):145–157.

Bucklin, D. N., Basille, M., Benscoter, A. M., Brandt, L. A., Mazzotti, F. J., Romanach, S. S., Speroterra, C., and Watling, J. I. (2015). Comparing species distribution models constructed with different subsets of environmental predictors. *Diversity and distributions*, 21(1):23–35.

Büttner, G., Steenmans, C., Bossard, M., Feranec, J., and Kolár, J. (1998). The European CORINE land cover database. *International Archives of Photogrammetry and Remote Sensing*, 32:633–638.

Cord, A., Schmidt, M., and Dech, S. (2009). Potential and limitations of multi-temporal Earth observation data to improve model results of tree species distribution in Mexico. *Proceedings, 33rd International Symposium on Remote Sensing of Environment, ISRSE 2009*, pages 285–288.

d'Andrimont, R., Verhegghen, A., Meroni, M., Lemoine, G., Strobl, P., Eiselt, B., Yordanov, M., Martinez-Sanchez, L., and van der Velde, M. (2021). Lucas copernicus 2018: Earth-observation-relevant in situ data on land cover and use throughout the european union. *Earth System Science Data*, 13(3):1119–1133.

de Rigo, D., Caudullo, G., Houston Durrant, T., and San-Miguel-Ayanz, J. (2016a). The European Atlas of Forest Tree Species: modelling, data and information on forest tree species. *European Atlas of Forest Tree Species*, page e01aa69.

de Rigo, D., Houston Durrant, T., Caudullo, G., and Barredo, J. I. (2016b). European forests: an ecological overview. In *European Atlas of Forest Tree Species*, pages 24 – 31. Publication Office of the European Union, Luxembourg.

Domke, G. M., Oswalt, S. N., Walters, B. F., and Morin, R. S. (2020). Tree planting has the potential to increase carbon sequestration capacity of forests in the united states. *Proceedings of the National Academy of Sciences*, 117(40):24649–24651.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F. A., and Feyen, L. (2016). Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*, 94:87–102.

Draper, F. C., Costa, F. R., Arellano, G., Phillips, O. L., Duque, A., Macía, M. J., Ter Steege, H., Asner, G. P., Berenguer, E., Schietti, J., et al. (2021). Amazon tree dominance across forest strata. *Nature ecology & evolution*, 5(6):757–767.

Easdale, M. H., Bruzzone, O., Mapfumo, P., and Tittonell, P. (2018). Phases or regimes? R evisiting NDVI trends as proxies for land degradation. *Land Degradation & Development*, 29(3):433–445.

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4):802–813.

EUROSTAT (2017). Land Cover/Use Statistics (LUCAS) Database.

Fawcett, D., Bennie, J., and Anderson, K. (2021). Monitoring spring phenology of individual tree crowns using drone-acquired ndvi data. *Remote Sensing in Ecology and Conservation*, 7(2):227–244.

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognition Letters*, 30(1):27–38.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12):4302–4315.

Fix, E. and Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3):238.

Fois, M., Cuena-Lombraña, A., Fenu, G., and Bacchetta, G. (2018). Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions. *Ecological Modelling*, 385:124–132.

Fourcade, Y., Besnard, A. G., and Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, 27(2):245–256.

Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378. Nonlinear Methods and Data Mining.

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266.

Gao, T., Zhu, J., Zheng, X., Shang, G., Huang, L., and Wu, S. (2015). Mapping spatial distribution of larch plantations from multi-seasonal landsat-8 oli imagery and multi-scale textures using random forests. *Remote Sensing*, 7(2):1702–1720.

Gelfand, A. E. and Shirota, S. (2021). The role of odds ratios in joint species distribution modeling.

Environmental and Ecological Statistics, 28(2):287–302.

Gobeyn, S., Mouton, A. M., Cord, A. F., Kaim, A., Volk, M., and Goethals, P. L. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling*, 392(June 2018):179–195.

Gomes, C., Nocairi, H., Thomas, M., Ibanez, F., Collin, J.-F., and Saporta, G. (2012). Stacking prediction for a binary outcome. In *Compstat 2012*, pages 271–282, Limassol, Cyprus.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27.

Gottschalk, T. K., Aue, B., Hotes, S., and Ekschmitt, K. (2011). Influence of grain size on species-habitat models. *Ecological Modelling*, 222(18):3403–3412.

Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., and Kueffer, C. (2014). Unifying niche shift studies: Insights from biological invasions. *Trends in Ecology and Evolution*, 29(5):260–269.

Guisan, A. and Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9):993–1009.

Guisan, A., Thuiller, W., and Zimmermann, N. E. (2017). *Habitat Suitability and Distribution Models*. Cambridge University Press, Cambridge.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., Mcdonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P., and Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12):1424–1435.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S., and Porter, J. H. (2013). Big data and the future of ecology. *Frontiers in Ecology and the Environment*, 11(3):156–162.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853.

Hao, T., Elith, J., Guillera-Arroita, G., and Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, 25(5):839–852.

Hao, T., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, 43(4):549–558.

Hastie, T., Qian, J., and Tay, K. (2016). An introduction to glmnet.

He, K. S., Zhang, J., and Zhang, Q. (2009). Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *acta oecologica*, 35(1):14–21.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1(2):87–97.

Heisig, J. and Hengl, T. (2020). Harmonized Tree Species Occurrence Points for Europe. URL:

https://zenodo.org/record/4061816, Dataset Version: 0.2.

Hengl, T., Leal Parente, L., Krizan, J., and Bonannella, C. (2020). Continental Europe Digital Terrain Model at 30 m resolution based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and background layers.

Hengl, T., Walsh, M. G., Sanderman, J., Wheeler, I., Harrison, S. P., and Prentice, I. C. (2018). Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. *PeerJ*, 6:e5457.

Hierro, J. L., Maron, J. L., and Callaway, R. M. (2005). A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. *Journal of ecology*, 93(1):5–15.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15):1965–1978.

Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., and Brown, N. (2017). Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and evolution*, 7(4):1043–1056.

Hoffer, R. (1984). Remote sensing to measure the distribution and structure of vegetation. *The Role of Terrestrial Vegetation in the Global Carbon Cycle: Measurement by Remote Sensing*, pages 131–59.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the modis vegetation indices. *Remote sensing of environment*, 83(1-2):195–213.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3):295–309.

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22:415–427.

Iturbide, M., Bedia, J., and Gutiérrez, J. M. (2018). Background sampling and transferability of species distribution model ensembles under climate change. *Global and Planetary Change*, 166(March):19–29.

Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., and Gutiérrez, J. M. (2015). A framework for species distribution modelling with improved pseudo-absence generation. *Ecological Modelling*, 312:166–174.

Jiang, Z., Huete, A. R., Didan, K., and Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote sensing of Environment*, 112(10):3833–3845.

Jiménez-Valverde, A., Lobo, J. M., and Hortal, J. (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions*, 14(6):885–890.

Jiménez-Valverde, A., Peterson, A. T., Soberón, J., Overton, J., Aragón, P., and Lobo, J. M. (2011). Use of niche models in invasive species risk assessments. *Biological invasions*, 13(12):2785–2797.

Kaky, E., Nolan, V., Alatawi, A., and Gilbert, F. (2020). A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. *Ecological Informatics*, 60:101–150.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas.

Scientific data, 4(1):1–20.

Karger, D. N., Dabaghchian, B., Lange, S., Thuiller, W., Zimmermann, N. E., and Graham, C. H. (2020). High resolution climate data for europe.

Kass, J. M., Anderson, R. P., Espinosa-Lucas, A., Juárez-Jaimes, V., Martínez-Salas, E., Botello, F., Tavera, G., Flores-Martínez, J. J., and Sánchez-Cordero, V. (2020). Biotic predictors with phenological information improve range estimates for migrating monarch butterflies in mexico. *Ecography*, 43(3):341–352.

Keenan, R. J. (2015). Climate change impacts and adaptation in forest management: a review. *Annals of Forest Science*, 72(2):145–167.

Key, C. H. and Benson, N. C. (1999). The Normalized Burn Ratio (NBR): A Landsat TM radiometric measure of burn severity. *United States Geological Survey, Northern Rocky Mountain Science Center: Bozeman, MT, USA*.

Kindt, R. (2018). Ensemble species distribution modelling with transformed suitability values. *Environmental Modelling & Software*, 100:136–145.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*.

Lefebvre, D., Williams, A. G., Kirk, G. J., Burgess, J., Meersmans, J., Silman, M. R., Román-Dañobeytia, F., Farfan, J., Smith, P., et al. (2021). Assessing the carbon capture potential of a reforestation project. *Scientific reports*, 11(1):1–10.

Madonsela, S., Cho, M. A., Ramoelo, A., and Mutanga, O. (2017). Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS Journal of Photogrammetry and Remote Sensing*, 133:116–127.

Manzoor, S. A., Griffiths, G., and Lukac, M. (2018). Species distribution model transferability and model grain size-finer may not always be better. *Scientific Reports*, 8(1):1–9.

Marchi, M. and Ducci, F. (2018). Some refinements on species distribution models using tree-level National Forest Inventories for supporting forest management and marginal forest population detection. *iForest - Biogeosciences and Forestry*, 11(2):291–299.

Martinez-Minaya, J., Cameletti, M., Conesa, D., and Pennino, M. G. (2018). Species distribution modeling: a statistical review with focus in spatio-temporal issues. *Stochastic environmental research and risk assessment*, 32(11):3227–3244.

Mather, J. R. and Yoshioka, G. A. (1968). The role of climate in the distribution of vegetation. *Annals of the Association of American Geographers*, 58(1):29–41.

Mauri, A., Strona, G., and San-Miguel-Ayanz, J. (2017). EU-Forest, a high-resolution tree occurrence dataset for Europe. *Scientific Data*, 4(1):160123.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

Mehra, A., Tripathy, P., Faridi, A., and Chinmay, A. (2019). Ensemble learning approach to improve existing models. *International Journal of Innovative Science and Research Technology*, 4.

Meier, E. S., Kienast, F., Pearman, P. B., Svenning, J. C., Thuiller, W., Araújo, M. B., Guisan, A., and Zimmermann, N. E. (2010). Biotic and abiotic variables show little redundancy in explaining tree

species distributions. *Ecography*, 33(6):1038–1048.

Miller, J., Franklin, J., and Aspinall, R. (2007). Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling*, 202(3-4):225–242.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Nandy, S., Singh, R., Ghosh, S., Watham, T., Kushwaha, S. P. S., Kumar, A. S., and Dadhwal, V. K. (2017). Neural network-based modelling for forest biomass assessment. *Carbon Management*, 8(4):305–317.

Nave, L. E., Walters, B. F., Hofmeister, K., Perry, C. H., Mishra, U., Domke, G. M., and Swanston, C. (2019). The role of reforestation in carbon sequestration. *New Forests*, 50(1):115–137.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Olaya, V. (2009). Chapter 6 basic land-surface parameters. In Hengl, T. and Reuter, H. I., editors, *Geomorphometry*, volume 33 of *Developments in Soil Science*, pages 141–169. Elsevier.

Palankar, M. R., Iamnitchi, A., Ripeanu, M., and Garfinkel, S. (2008). Amazon S3 for science grids: a viable solution? In *Proceedings of the 2008 international workshop on Data-aware distributed computing*, pages 55–64.

Pearson, R. G. and Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography*, 12(5):361–371.

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., and Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of biogeography*, 33(10):1704–1711.

Pecchi, M., Marchi, M., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., and Chirici, G. (2019). Species distribution modelling to support forest management. A literature review. *Ecological Modelling*, 411(submitted):108817.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633):418–422.

Pérez Chaves, P., Ruokolainen, K., and Tuomisto, H. (2018). Using remote sensing to model tree species distribution in Peruvian lowland Amazonia. *Biotropica*, 50(5):758–767.

Peterson, A. T., Soberón, J. M., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., and Araújo, M. B. (2011). *Ecological niches and geographic distributions (MPB-49)*. Princeton University Press.

Phillips, S. J., Dudik, M., Elith, J., Graham, C. H., Leathwick, J., Ferrier, S., Applications, S. E., Jan, N., Phillips, J., Leathwick, J., and Elith, J. (2009). Sample Selection Bias and Presence-Only Distribution Models : Implications for Background and Pseudo-Absence Data Published by : Ecological Society of America Linked references are available on JSTOR for this article : Your use of the JSTOR archive indica. *Ecological Applications*, 19(1):181–197.

Pigeon, G., Legain, D., Durand, P., and Masson, V. (2007). Anthropogenic heat release in an old European agglomeration (Toulouse, France). *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(14):1969–1981.

Popkin, G. (2021). Germany's trees are dying. a fierce debate has broken out over how to respond. *Science*, 374.

Porfirio, L. L., Harris, R. M., Lefroy, E. C., Hugh, S., Gould, S. F., Lee, G., Bindoff, N. L., and Mackey, B. (2014). Improving the use of species distribution models in conservation planning and management under climate change. *PLoS ONE*, 9(11):1–21.

Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina, A., and Ying, Q. (2020). Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, 12(3):426.

Prasad, A. M., Iverson, L. R., and Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, 9(2):181–199.

Prates-Clark, C. D. C., Saatchi, S. S., and Agosti, D. (2008). Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. *Ecological Modelling*, 211(3-4):309–323.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment*, 48(2):119–126.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1):81–106.

Ripley, B. and Venables, W. (2017). *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. R package version 7.3-12.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.

San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. E. (2016). *European Atlas of Forest Tree Species*. Publication Office of the European Union.

Schloss, A., Kicklighter, D., Kaduk, J., Wittenberg, U., and Intercomparison, T. P. O. T. P. N. M. (1999). Comparing global models of terrestrial net primary productivity (NPP): comparison of NPP to climate and the Normalized Difference Vegetation Index (NDVI). *Global Change Biology*, 5(S1):25–34.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, 406:109–120.

Shabani, F., Kumar, L., and Ahmadi, M. (2018). Assessing accuracy methods of species distribution models: AUC, Specificity, Sensitivity and the True Skill Statistic. *Global Journal of Human Social Science*, 18(1):6–18.

Shi, X., Wong, Y. D., Li, M. Z.-F., Palanisamy, C., and Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accident Analysis & Prevention*, 129:170–179.

Soberón, J. M. (2010). Niche and area of distribution modeling: A population ecology perspective. *Ecography*, 33(1):159–167.

Soberón, J. M. and Peterson, A. T. (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, 2(10):1–10.

Strickland, G. E. I., Luther, J. E., White, J. C., Wulder, M. A., Strickland, G. E. I., Luther, J. E., White, J. C., and Wulder, M. A. (2020). Extending Estimates of Tree and Tree Species Presence-Absence through Space and Time Using Landsat Composites. *Canadian Journal of Remote Sensing*, 46(5):567–584.

Therneau, T. M. and Atkinson, E. J. (2011). An Introduction to Recursive Partitioning Using the RPART

Routines. *Mayo clinic*, 61:33.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., and Elith, J. (2021). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*, 0(0):1–27.

van den Hoogen, J., Geisen, S., Wall, D. H., Wardle, D. A., Traunspurger, W., de Goede, R. G., Adams, B. J., Ahmad, W., Ferris, H., Bardgett, R. D., et al. (2020). A global database of soil nematode abundance and functional group composition. *Scientific data*, 7(1):1–8.

van den Hoogen, J., Robmann, N., Routh, D., Lauber, T., van Tiel, N., Danylo, O., and Crowther, T. W. (2021). A geospatial mapping pipeline for ecologists. *BioRxiv*.

Walthert, L. and Meier, E. S. (2017). Tree species distribution in temperate forests is more influenced by soil than by climate. *Ecology and Evolution*, 7(22):9473–9484.

Wang, H., Liu, H., Huang, N., Bi, J., Ma, X., Ma, Z., Shangguan, Z., Zhao, H., Feng, Q., Liang, T., et al. (2021). Satellite-derived NDVI underestimates the advancement of alpine vegetation growth over the past three decades.

Weigel, R., Gilles, J., Klisz, M., Manthey, M., and Kreyling, J. (2019). Forest understory vegetation is more related to soil than to climate towards the cold distribution margin of european beech. *Journal of Vegetation Science*, 30(4):746–755.

Wilson, A. M. and Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS biology*, 14(3):e1002415.

Witjes, M., Parente, L., van Diemen, C. J., Hengl, T., Landa, M., Brodsky, L., Halounova, L., Krizan, J., Antonic, L., Ilie, C. M., et al. (2021). A spatiotemporal ensemble machine learning framework for generating land use/land cover time-series maps for Europe (2000–2019) based on LUCAS, CORINE and GLAD Landsat. *PeerJ*.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

Zhang, C. and Ma, Y. (2012). *Ensemble machine learning: methods and applications*. Springer.

Zhou, Z.-H. (2019). *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC.

Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., et al. (2019). Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224:382–385.