

COST Action FP1202

“Strengthening conservation: a key issue for adaptation of marginal/peripheral populations of forest trees to climate change in Europe (MaP-FGR)”

Guest Editors: Fulvio Ducci, Kevin Donnelly

## Some refinements on species distribution models using tree-level National Forest Inventories for supporting forest management and marginal forest population detection

Maurizio Marchi,  
Fulvio Ducci

Spatial modelling is a fundamental tool to support forest management strategies. National Forest Inventories (NFIs) provide extensive and detailed data for spatial analysis. In this study, the most recent Italian NFI (INFC2005) was used to evaluate possible refinements on species distribution model (SDM) techniques and to derive the future scenarios for two target species (*Fagus sylvatica* L. and *Abies alba* Mill.) sharing a similar ecological environment and geographic range. A weighted SDM and a provenance distribution model (PDM) were tested, based on tree-level selection of NFI plots using species basal area as a filter. Two climate projections were analysed for 2050s according to the IPCC 5<sup>th</sup> Assessment Report (AR5). The results were evaluated as possible guidelines for management of the Italian region of the EUFGIS network, where many marginal forest populations (MaPs) are currently included as genetic conservation units (GCUs). The uncertainty of coordinates of inventory points did not affect the results of SDM. No statistical differences were found when comparing the niche realization for the two model species (ANOVA  $p > 0.05$ ) mainly due to spatial autocorrelation between the environmental predictors. Based on the classic SDM evaluation method (True Skill Statistic - TSS) little improvements in predictions were observed when weighting each presence/absence records, possibly due to the lack of adequate ancillary data but also to the evaluation method. A higher accuracy of predictions (TSS  $> 0.85$ ) was obtained when different “provenances” were modelled separately, due to the reduction in the “background noise”. We showed that for classical SDM, the prevalence of certain ecological features of some locations may drive algorithms to produce coarse averaged predictions. Provenance distribution modelling may represent a valuable step forward in spatial analysis, particularly for the detection of marginal peripheral populations. The exact spatial co-ordinates of plots and additional information on site quality (e.g., stand age, site index, etc.) in NFI data could greatly help in better weighting presence/absence data and properly test the new evaluation methods.

**Keywords:** INFC2005, European Beech, Silver Fir, Modelling Uncertainties, Provenance Modelling, Climate Change, Mediterranean Area

□ CREA - Research Centre for Forestry and Wood, v.le S. Margherita 80, I-52100 Arezzo (Italy)

@ Maurizio Marchi  
([maurizio.marchi@crea.gov.it](mailto:maurizio.marchi@crea.gov.it))

Received: Mar 23, 2017 - Accepted: Feb 06, 2018

**Citation:** Marchi M, 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* 11: 291-299. - doi: [10.3832/ifor2441-011](https://doi.org/10.3832/ifor2441-011) [online 2018-04-13]

Communicated by: Piermaria Corona

### Introduction

The Mediterranean region is nowadays acknowledged as one of the main “biodiversity hotspots” for endemic plants and trees (Petit et al. 2003, Noce et al. 2016) and recognized as a biogeographic zone whose biodiversity is significantly threatened (Myers et al. 2000). Knowledge of the existing genetic and adaptive potential of forest genetic resources (FGR) is needed to safeguard tree species of the Mediterranean region and help forest managers mitigating the impact of climate change on forest ecosystems (Williams & Dumroese 2013, Orlovic et al. 2014). Genetic diversity will play a key role in protecting forests under climate change, as information on species richness and genetic diversity is considered essential for the choice of ro-

bust silvicultural systems (Fady et al. 2016). The maintenance of a high genetic diversity of forests can contrast with other objectives such as timber production (Kingsolver & Diamond 2011), however the trade-off between productivity and biodiversity conservation is a key requirement of forest policy.

Endangered gene pools may require transfer measures aimed at maintaining genetic diversity through assisted gene flow and assisted migration protocols based on robust data and models (Hamann et al. 2015), as well as on the knowledge of adaptive traits and the spatial distribution of genetic diversity (Kramer et al. 2017, Piotti et al. 2017). Recently, forest research has focused on the impact of climate change on forest resources and their spatial distribu-

tion (Isaac-Renton et al. 2014, Marchi et al. 2016), and on changes in the provision of ecosystem goods and services (Castaldi et al. 2017, Ray et al. 2017), in order to support forest management strategies under climate change. For example, the “European Atlas of Forest Tree Species” of the European Commission (San-Miguel-Ayanz et al. 2016) is a key resource aimed to support forest management with a comprehensive coverage at a European scale.

Species distribution modelling (SDM) is a promising technique commonly applied to predict the impact of climate change on species' biodiversity and populations (Thuiller et al. 2015) and to support forest management and silvicultural practices (Rehfeldt et al. 2014). The goal of such technique (often referred to as “ecological niche modelling”) is to characterize the ecological requirements of a target (plant or animal) species using statistical models. Various algorithms are used (e.g., Generalized Linear Models, Generalized Additive Models, Neural Networks, Regression Trees, etc.), assuming the current spatial distribution of a species to be a proxy of its ecological niche. Environmental predictors such as temperature, precipitation and several bioclimatic indices are generally derived from wide datasets by means of statistical tests (e.g., co-linearity), or using an expert knowledge on the target species (Naimi et al. 2011, Vacchiano & Motta 2015). New opportunities may be offered from using National Forest Inventory (NFI) ancillary data, such as timber volume and species abundance. Many freely available or restricted spatial European datasets and empirical data of forest species have been used in SDM studies, such as the EUFOR-

GEN (<http://www.euforgen.org/species/>) or EFI distribution maps (Brus et al. 2012), but local databases such as forest types are generally thought to provide a higher level of accuracy. All NFIs are characterized by: (i) high spatial precision (metres); (ii) detection of tree species and their relative abundance; (iii) availability of quantitative data (such as basal area, site index, age, average height, etc.); (iv) national scale and total coverage of the territory; and (v) standardized methods and outputs. Consequently, a NFI provides an important country-level data source, due to high-quality, statistically designed geo-referenced data (Fattorini 2015). In many countries, a NFI survey is regularly performed (e.g., every 5 years) and often repeated on the same sampling points (Fridman et al. 2014) with the aim of collecting temporal series and assessing local growth trends. In addition to using robust and highly-performing algorithms and validated datasets, the assessment of uncertainty is an important step in such modelling techniques (Petr et al. 2014). This includes uncertainties in climate data (and future projections) as well as positional errors in presence/absence data. Such step is especially needed when small populations are studied, such as Marginal and Peripheral forest populations (MaP), where a reduced part of the whole distribution of a forest species must be carefully represented by using its realised niche (Marchi et al. 2016).

In this paper, an effort to improve the SDM approach using national forest inventory data is proposed, with the aim of supporting active adaptive management strategies for maintaining endangered forest gene pools and genetic diversity under pro-

jected climate change. The study focuses on the ecological niche of two target species (*Fagus sylvatica* L. and *Abies alba* Mill.) derived from the spatial distribution of inventory points. Five different approaches were evaluated to forecast possible effects of climate change on a selected group of populations included into the EUFGIS database (<http://www.eufgis.org/>). Two different climate change projections (RCP4.5 and RCP8.5 of AR5, IPCC – <http://www.ipcc.ch/report/ar5/syr/>) were used to explore the expected shifts in the fundamental niche space of the study species.

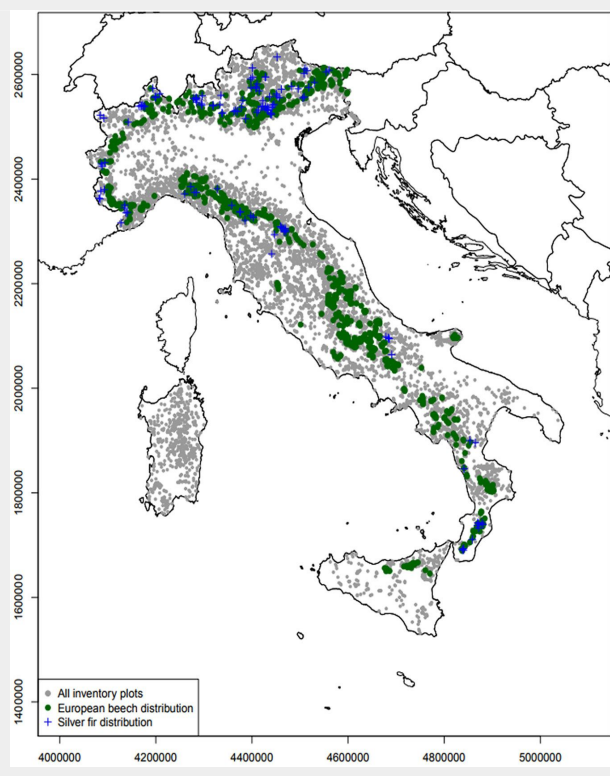
## Materials and methods

### Presence/absence datasets: the Italian National Forest Inventory

The last Italian NFI (INFC2005) began in 2003 and raw data has been recently released (Borghetti & Chirici 2016). The sampling procedure was developed following three main phases in a “stratified sampling approach” and 7,272 geo-referenced inventory points were available in the WGS84 reference system, grouped into 23 forest categories. From the 263 inventoried species, we selected European beech (*Fagus sylvatica* L.) and silver fir (*Abies alba* Mill.), whose Italian ranges are both characterised by the presence of widely disjoint meta-populations. Due to their valuable economic and ecological importance, both species are well studied at the European scale in terms of distribution of genetic diversity (Rose et al. 2009, Piotti et al. 2017), adaptive traits (Kerr et al. 2015, Kramer et al. 2017), and impacts of climate change on their spatial distribution (Schueler et al. 2013, Tinner et al. 2013). In addition, many marginal forest populations (MaPs) of both species do exist in Italy as possible target of FGR management.

In Italy beech forests cover 1,035,103 ha corresponding to 9.89% of the total forest cover. Almost 89% of these are pure beech stands and just 11% is a mixture between beech and other conifers (mainly *Pinus nigra* or *Abies alba*). Concerning silver fir, 68,460 ha are dominated by this species (0.65% of the Italian forests) with 75.5% as pure or beech-fir mixed forest. Both species are mainly distributed along the Apennines chain (Fig. 1), but also in parts of the Alps. While the beech range is almost continuous along the whole Italian peninsula, silver fir is scattered and concentrated in Calabria, Abruzzo, Tuscany, Piedmont and North-eastern Alps. Both species are absent in Sardinia while some beech populations are located in Sicily. In addition to their shared ecological features, beech and silver fir were selected as examples of managed forest tree species representing: (i) a dominant species which forms pure stands (*Fagus sylvatica*) with an almost stable fundamental niche; (ii) a shade-tolerant species (*Abies alba*) often planted outside its native range and able to grow under a close canopy cover in mixed forests with

**Fig. 1** - The Italian National Forest Inventory (INFC 2005) points (grey dots), the spatial distribution of European beech (green dots) and silver fir (blue crosses).



an expanded niche.

Spatial co-ordinates in the Italian NFI dataset are subject to privacy restriction, and presence/absence data are provided with the co-ordinates of the south-western corner of a 1-km grid. To deal with this, a preparatory analysis was performed to test possible biases in niche estimation.

#### Spatial climate data pre-processing and coordinate uncertainty test

The choice of environmental predictors in spatial modelling needs care, as the exclusion of real drivers may lead to unreliable results, whereas the use of a wide spectrum of highly correlated (co-linear) variables can affect the importance of predictors and lengthen the computational time. The WorldClim database (Fick & Hijmans 2017) has often been used for climatic data and indices in SDM applications (Molyneux et al. 2014, Vacchiano & Motta 2015). These data are provided as a global raster with a maximum resolution of 30 arc-seconds (approximately 1 km at the equator). However, the use of WorldClim in local studies has been often criticised due to its global coverage (Bedia et al. 2013). Therefore, we used the ClimateEU software, an online tool freely available at the website of the University of Alberta (Alberta, Canada). ClimateEU provides 36 climatic variables suitable for predicting local climate conditions in dendrochronological studies and simulations in Europe and Italy (Isaac-Renton et al. 2014, Marchi et al. 2015). Fifty monthly, seasonal, and annual variables are available using this software, including many economically or biologically relevant variables such as growing and chilling degree days, heating and cooling degree days, Hargreave's moisture deficit and reference evaporation. Future projections are based on 15 Global Circulation Models referenced in the IPCC Fifth Assessment Report for two emission scenarios (RCP4.5 and RCP8.5) and three standard time-slices (2020s, 2050s, 2080s). The use of multiple custom queries in the ClimateEU software allowed us to create 36 raster maps of the main climatic variables and indices over the Italian territory with spatial resolution of 250 m. For this study, the current average normal climate period 1981-2010 was used as reference. Two emission scenarios were also computed (RCP4.5 and RCP8.5) for the 2050s time-slice, using the IPSL-CM5A-MR model from the European Union implemented into ClimateEU. The full list of climatic variables calculated by ClimateEU is reported in Tab. S1, Tab. S2 and Tab. S3 (Supplementary material). In addition, a soil map of Italy (Costantini et al. 2014) was included in the dataset.

Principal Component Analysis (PCA) was carried out on a total of 37 factors. To avoid the problem of different scales among different variables (temperatures, precipitation, potential evapo-transpiration indices, soil categories etc.), the dataset was centered and scaled. Calculated com-

ponents up to 99% of the cumulative explained variance were used as predictors.

Once climatic surfaces were prepared, a set of “artificial niches” were simulated to test for possible biases due to co-ordinates uncertainty. Given that the co-ordinates of the NFI dataset were those of the south-western corner of 1-km grid, all NFI points were firstly virtually modified, moving each point in the centre of each NFI location cell (+500 m to both X and Y co-ordinates). This layer was then spatially overlaid onto four climatic raster layers and the values of the climatic variables at each modified NFI point were extracted. The original 250-m rasters were then upscaled to 500 m, 750 m and 1 km, and a simulation was performed. A random sequence of spatial co-ordinates was generated across all the 1-km cells covered by each species in the NFI dataset. The total number of randomly-generated co-ordinates was equivalent to the real NFI points for each species (1316 for beech, 348 for silver fir). Such random extraction was repeated 10,000 times, and the results averaged across extractions. Using this approach, a set of “artificial” ecological niches was created and compared with the basic niche corresponding to the current position of the NFI plot (south-western edge) by non-parametric ANOVA (Kruskal & Wallis 1952), in order to test whether the variance among artificial ecological niches was higher than the variance within them.

#### SDM algorithms and modelling procedure

It is well known that different modelling techniques which use similar inputs could lead to divergent predictions of climate or species distributions, mainly in cases of local studies or reduced datasets (Araújo & New 2007). SDMs have been often criticised as suitable for modelling the probability of detection, whilst providing poorly predictions of species performance (Isaac-Renton et al. 2014). In this study, the availability in the Italian NFI dataset of tree-level data (e.g., diameter at breast height and total height of each tree in each plot) allowed us to test for five different modelling approaches (in terms of presence/ab-

sence extraction) characterized by an expectedly increasing degree of complexity:

- (1) The ecological niche of the target species was derived and modelled using all the records belonging to the most suitable NFI forest category (hereafter defined as forest category approach – FCA) included into the INFC2005 database (e.g., silver fir forests to model *Abies alba*, beech forests to model *Fagus sylvatica*).
- (2) A tree-level selection of inventory points was performed using a basal area approach (BAA), thereby including all the NFI points where the species contributed to basal area by more than 1%, irrespective of the NFI forest category the corresponding point is included.
- (3) A novel provenance distribution model approach (PDM) was devised. First, NFI points detected for each species using BAA were grouped based on environmental characteristics using hierarchical clustering prior to modelling. Then an adequate number of clusters (groups) was selected and modelled separately.
- (4) A weighted BAA and PDM were computed (wBAA and wPDM), weighting each presence point differently according to the proportion of basal area of the target species in each.

Differences among methods were then evaluated by non-parametric ANOVA and post-hoc tests (Friedman 1937, Kruskal & Wallis 1952) applied to True Skill Statistic (TSS) values. The aim was not to assess the best modelling procedure for each species but to test for differences among methods in modelling the distribution of two species which share a similar ecological environment.

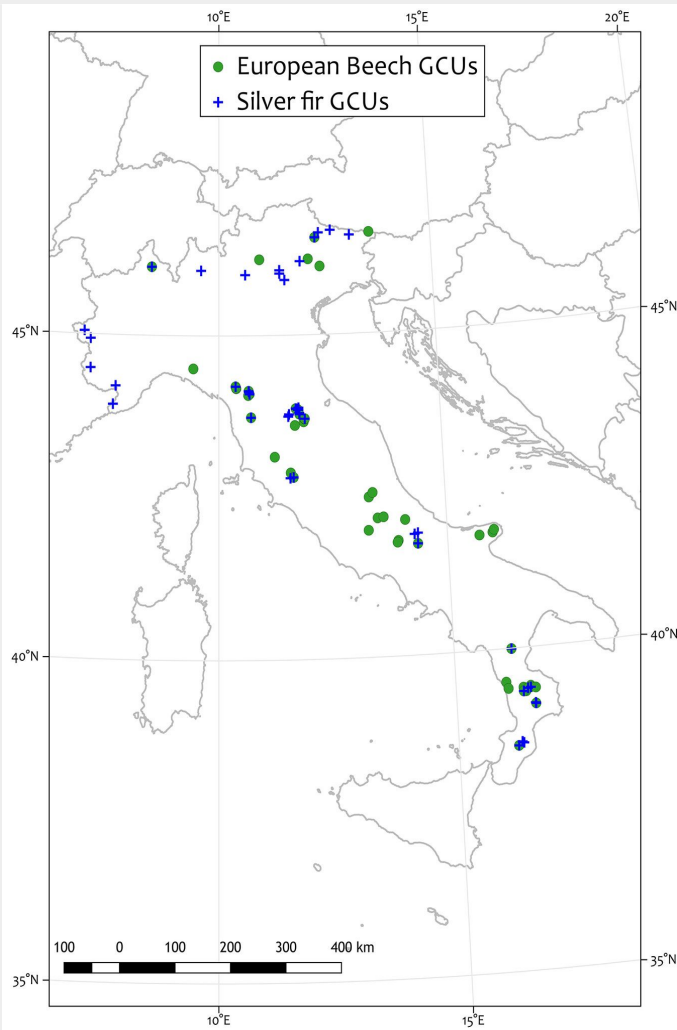
The five approaches of presence/absence selection and the number of NFI points used in each approach are presented in Tab. 1. Due to the different numeric consistency (i.e., absence points were always almost six times more abundant than the presence points), the prevalence parameter was always set to 0.5 to balance the predictions (Allouche et al. 2006). A set of nine different algorithms from the literature were tested and averaged to build a consensus map using the software pack-

**Tab. 1** - Modelled species and inventory (NFI) points considered for the tested methodology. (FCA): Forest Category Approach; (BAA): Basal Area Approach; (wBAA): weighted Basal Area Approach; (PDM): Provenance Distribution Modelling; (wPDM): weighted Provenance Distribution Modelling; (GCUs): number of genetic Conservation Units available from the EUFGIS database.

Modelled species	Method	Presence points in NFI	Proportion of NFI points	Absence points in NFI	GCUs
European beech	FCA	796	10.95%	6,476	47
	BAA & wBAA	1316	18.10%	5,956	
	PDM & wPDM	1316 (6 groups)	Variable	Variable	
Silver fir	FCA	131	1.80%	7,141	48
	BAA & wBAA	348	4.79%	6,924	
	PDM & wPDM	348 (6 groups)	Variable	Variable	



**Fig. 2** - Spatial distribution of the Italian Genetic Conservation Units (GCUs) for the two studied species (European beech and silver fir).



age “biomod2” (Thuiller et al. 2016) in R ver. 3.3.3 (R Core Team 2017). The set was comprised of: Generalized Linear Model (GLM), Generalized Boosted-regression Model (GBM), Generalized Additive Model (GAM), Classification Trees Analysis (CTA), Artificial Neural Networks (ANN), Flexible Discriminant Analysis (FDA), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF) and the classical version of MAXimum ENTropy (MAXENT).

To evaluate the quality of predictions delivered by each model, the dataset was randomly split into a training subset (75% of observations) and a testing subset (25% of observations). The splitting procedure was repeated 50 times and the average value and the standard deviation of the TSS index were calculated for each model (Alouche et al. 2006).

We used the best set of modelling procedure (ensemble) to map European beech and silver fir in Italy. A consensus model was calculated using the full-model predictions (i.e., using 100% of the available NFI points) and no TSS threshold was introduced to filter model before averaging. This analysis produced three distribution maps for each target species, one for the current baseline climate and conditions, and two maps showing the predicted distribution for future climate projections for the period centred on the 2050s for RCP4.5 and RCP8.5, respectively (Fig. 2). Finally, the consensus maps were included in a GIS and indications for the management of the Italian Genetic Conservation Units (GCUs) were derived for the two studied species included into the EUFGIS database (<http://www.eufgis.org/>).

**Results**

The results of PCA on climate variables showed that the 99% of the environmental variability could be summarized into seven uncorrelated PCA axes components (Tab. 2). The first component (PC1) accounted for most of the variance (77.82%) with a well balanced contribution of 37 environmental predictors. As expected, no single environmental variable prevailed on the others. On PC1, the average (absolute) eigenvector was 0.160 with a standard deviation of 0.039, demonstrating a high degree of co-linearity between the environmental predictors. The range (in absolute values) was between 0.017 (winter precipitations) and 0.185 (average temperature of the autumn) with a median value of 0.177. The remaining proportion of variance was distributed between six other components, with more than 95% included in the first four principal components. As a consequence, we used seven predictors in the biomod2 calculation (see below).

Concerning niche simulations, no statistical differences were observed between the ecological niche used for modelling (i.e., using the original co-ordinates provided by INFC) and any of the “artificial niches” we simulated (Tab. 3) with p-values >0.05.

**Tab. 2** - Summary statistics of the PCA eigenvectors for the 37 predictive variables considered.

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
MIN Eigenvector	0.016	0.001	0.003	0.001	0.001	0.002	0.002
MEDIAN Eigenvector	0.177	0.118	0.087	0.061	0.085	0.067	0.093
MEAN Eigenvector	0.160	0.134	0.112	0.113	0.119	0.109	0.122
Eigenvectors SD	0.039	0.097	0.122	0.121	0.115	0.125	0.112
MAX Eigenvector	0.185	0.356	0.560	0.545	0.478	0.703	0.519
Explained standard deviation	5.365	1.822	1.602	0.947	0.611	0.548	0.467
Proportion of variance	0.778	0.090	0.069	0.024	0.011	0.008	0.006
Cumulative proportion of variance (%)	77.82	86.79	93.73	96.16	97.17	97.98	99.3

**Tab. 3** - Statistical differences between covered niches.

Group A	Group B	ANOVA p-value
Raster 250 m	Average 500 m	>>0.05
Raster 250 m	Average 750 m	>>0.05
Raster 250 m	Average 1 km	>>0.05
Raster 250 m	Random 250 km	Average >>0.05 ± 0.003
Raster 250 m	Average 500 m	>>0.05
Raster 250 m	Average 750 m	>>0.05
Raster 250 m	Average 1 km	>>0.05
Raster 250 m	Random 250 km	Average >>0.05 ± 0.001

Regarding the different modelling approaches, the FCA showed low predictive performances, with a mean TSS always lower than 0.7 for all algorithms (a threshold generally accepted as satisfactory for environmental models – Swets 1988), with an average TSS value of  $0.642 \pm 0.058$  (Tab. 4) and a coefficient of variation always lower than 9.06%. The Generalized Boosted-regression Model (GBM) was the best performing algorithm (TSS = 0.679), being statistically different from the other algorithm applied.

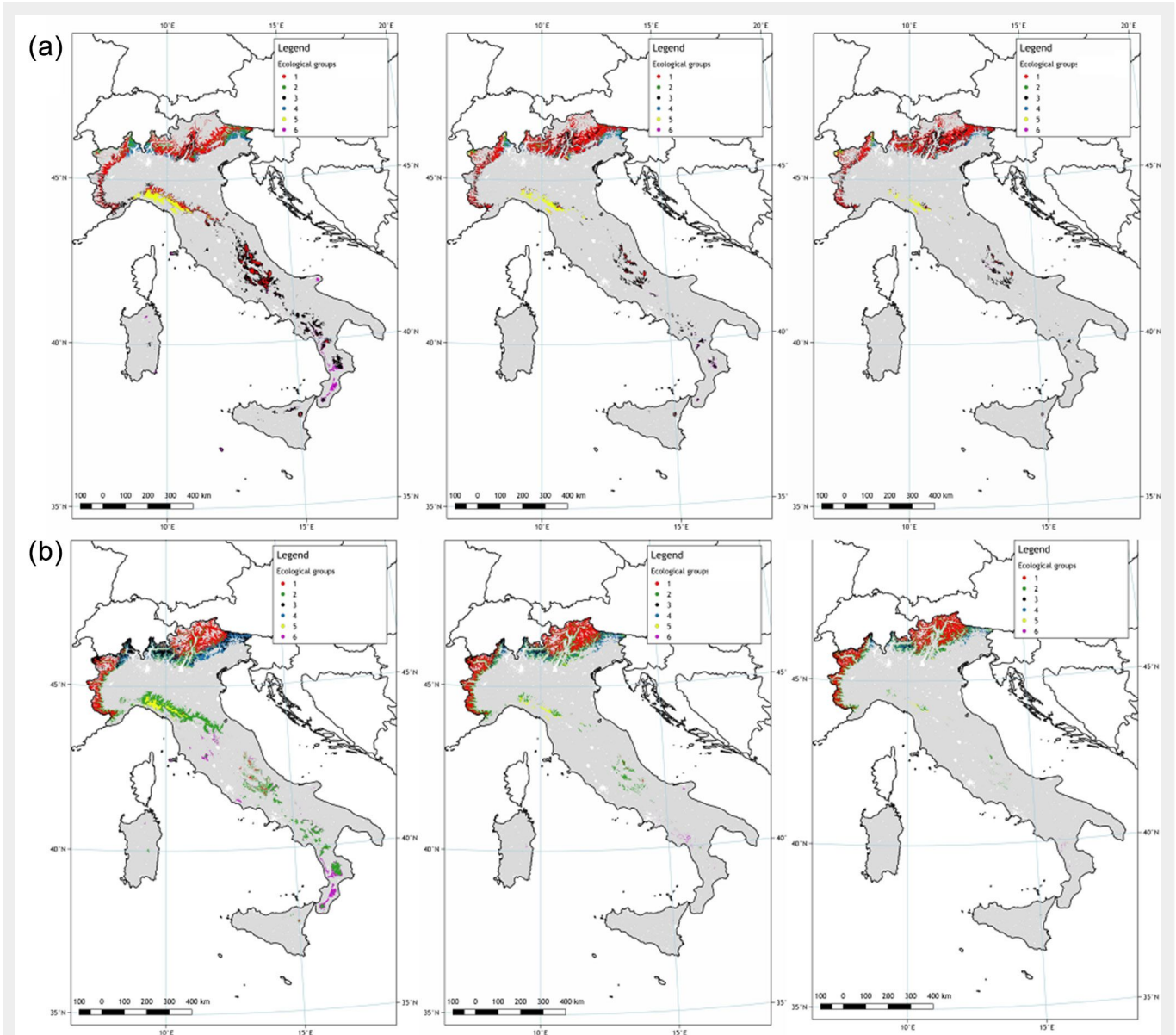
No significant improvements were observed in the prediction of species distribution using the BAA approach. However, a lower variability was detected among runs and algorithms (the global coefficient of variation was 7.15%), indicating a more stable prediction among runs. Using BAA, RF was the best performing model (TSS = 0.661), being statistically different from the

**Tab. 4** - Average TSS values for each extraction method and statistical groups according to non-parametric ANOVA. (\*): Average value of the 6 ecological provenances for each species.

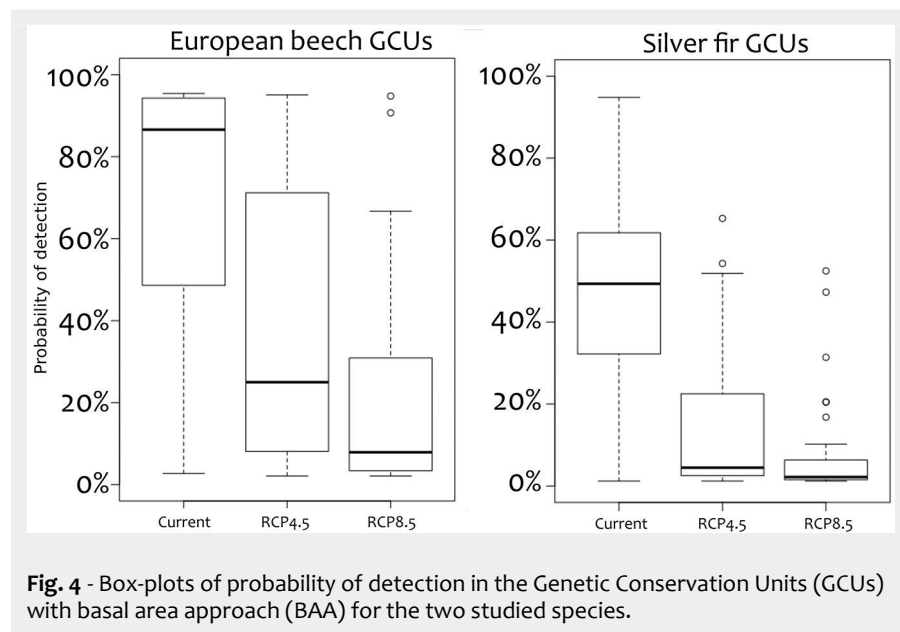
Method	Average TSS	Standard Deviation	Groups
PDM	0.830*	0.184	a
wPDM	0.643	0.244	b
FCA	0.642	0.058	b
BAA	0.636	0.046	b
wBAA	0.557	0.123	c

others. The weighted version of this approach (wBAA) did not show any improvement in the predictions. On the contrary, TSS values calculated for wBAA were much lower and statistically different from the others (average TSS = 0.557), indicating that wBAA was the worst model. The best model performances were ob-

tained using the PDM approach. This method was based on the result of a hierarchical clustering where six groups were selected for both species as fair trade-off between environmental variability and spatial fragmentation. Hierarchical clustering allowed for the whole ecological niche to be split into several clearly-defined sec-



**Fig. 3** - Provenance distribution model (PDM) for (a) European beech and (b) silver fir for 2050 climate projections driven by the two selected green-house gas emissions scenarios RCP4.5 and RCP8.5.



**Fig. 4** - Box-plots of probability of detection in the Genetic Conservation Units (GCUs) with basal area approach (BAA) for the two studied species.

tions with a strong “ecological signal”. Using this procedure, models’ calculations were fitted on a smaller and more homogeneous part of the realized ecological niche. The predictions obtained by this approach were significantly improved, with an average TSS of 0.830 and maximum values of about 0.9 using GLM, GBM, RF and MARS algorithms. However, very low and unexpected values were found using MAXENT, with TSS close to 0.3. Similarly, the TSS for wPDM also decreased as compared with PDM, though the observed average values were not statistically different from using either the FCA and BAA approaches.

The results of the comparisons (ANOVA) among all models and algorithm combinations are reported in Tab. S4 (Supplementary material).

An ensemble model was built for each provenance of each species (Fig. 3) and future scenarios using the software biomod2. A consensus map was implemented in GIS showing the probability of detection of European beech and European silver fir for each Italian GCU of the EUFGIS database, *i.e.*, the probability of detection of each ecological provenance in each GCU. Given the structure of PDM, six different suitability values were available in each GCUs for both species. To deal with this, the best performing provenance in each GCUs was extracted in each analysed time slice. The results of this analysis are summarised in Fig. 4 which shows the boxplots of the probability of detection (*i.e.*, survival) for each species in each climate projection. Based on model predictions, a reduction of suitable habitats is expected under future climate projections in all GCUs, as well as a range shrinking for both beech and silver fir.

## Discussion

Despite a possible bias in niche estimation stemming from positional uncertainties of the NFI dataset, the use of ancillary

tree-level data derived from the Italian NFI and the use of more sophisticated models did not significantly improve SDM performances in this study. Both the applied weighting procedures (wBAA and wPDM) showed a reduction in performance compared to unweighted methods (BAA and PDM). Although model complexity and performances have often been demonstrated to be uncorrelated in ecological analysis (Merow et al. 2014, Marchi et al. 2017), such differences between weighted and unweighted models may be explained by pure modelling aspects and evaluation methods. Indeed, the True Skill Statistic (TSS) was developed to improve Cohen’s kappa dependence on prevalence (Allouche et al. 2006) and is the most widely used measure for SDM performances and for models generating presence/absence predictions in general (Barbet-Massin et al. 2012). However, results demonstrated this index might be unreliable when a weighted SDM is performed. The TSS is typically used for binomial response variables (1/0), involving the calculation of sensitivity and specificity. In this study, weighted models were applied to calculate a smoothed value, corresponding to the performance of the species in any environmental (*i.e.*, geographic) location. All presence points are evaluated from biomod2 input (*i.e.*, presence=1, absence=0), and values equal to 1 are likely to be obtained for the ecological optimum in predicting species performance. This could likely explain the lower TSS values obtained using wBAA and wPDM, as compared with classical unweighted models.

With BAA modelling no differences in SDM performance were detected when compared to the more general FCA method. The use of a hierarchical clustering before the biomod2 analysis in the PDM approach highly improved the SDM results, leading to a more precise niche estimation. When many points are used for modelling

(as in the case for presence/absence datasets), a strong “background noise” is included in the calculation, and a “coarse” ecological niche is obtained, thereby losing some species-specific ecological features, such as those held in marginal and peripheral forest populations, which are characterized by a limited number of NFI points (Eckert et al. 2008). This aspect was obvious in our study where a TSS value of 0.83 was calculated using the PDM approach, which was much higher than that obtained using FCA and BAA (0.642 and 0.636, respectively). The average value for PDM was highly influenced by results of MAXENT, whose TSS value was surprisingly low ( $\approx 0.3$ ). A TSS value of approximately 0.9 was calculated for PDM when averaging all the algorithms except MAXENT, whose prediction was probably affected by the use of a low-memory version of the algorithm. MAXENT is widely used and well-referenced in the literature (Merow et al. 2013), but the evaluation of its performance was out of the scope of this study.

The PDM approach presented here has rarely been implemented in the literature. Researchers often use the whole distribution range of forest species (San-Miguel-Ayanz et al. 2016), and a similar approach was recently developed by Isaac-Renton et al. (2014) with Douglas-fir, where a SDM was used to clip a RandomForest classification of the European environment, based on the ecological features of Douglas-fir in the Pacific North West. New insights can be achieved using provenance distribution models, which could be integrated with the classical use of forecasting range shifts, such as replacing local provenances with better adapted material from lower latitudes or lower elevations (O’Brien et al. 2007, Marchi et al. 2016). Indeed, using the provenance distribution model indirect hypotheses can be made on the genetic diversity distribution of a species, assuming a connection between ecological and genetic diversity as a result of long term-evolutionary processes and local adaptation (Eckert et al. 2008, Lira-Noriega & Manthey 2014, Fady et al. 2016). However, forest genetic diversity is not directly included into NFIs data, and it needs spatially explicit information to infer genetic variation parameters.

## NFI data and the provenance distribution modelling method in a European context

At a wider scale, merging all the European NFIs could lead to the construction of an important dataset of species distribution. NFIs are often difficult to harmonized due to differences in national inventory definitions, sampling designs, plot size and density, measured variables, and measurement protocols (McRoberts et al. 2009). The use of NFI data at a European scale would have great impact on SDM results due to the possibility of excluding broad envelopes or statistical maps and correctly



detecting the true absence of species. With such data, the reliability of SDM and PDM for the whole Europe would be improved. Indeed, while many studies in species distribution modelling have focused on the importance of presence points (Manel et al. 2001), the reliability and availability of true absence points is rarely discussed. Presence points are fundamental to derive ecological niches and are treated as the most relevant issue in SDM approaches. However, a much higher degree of uncertainty lies in the absence of a given species at a particular location, whether is due to ecological or human-related reasons. Such information is fundamental to set an “exclusion threshold” of each species for better modelling its distribution. Many tools have been tested to describe and simulate absence data (e.g., pseudo-absences or background points), for example: random spatial selection, regular grids, ecological distance (also known as Surface Range Envelope method) and spatial distance from presence records (Barbet-Massin et al. 2012). The use of NFI datasets could reduce the problem of the lack of quality of absence data.

#### *PDM in the framework of marginal forest populations and climate change*

Robust models and tools are essentials to drive forest practice, forest management and the use and transfer of forest reproductive materials (Williams & Dumroese 2013). Even when gene pools are artificially moved or transferred (e.g., assisted gene flow or migration) the use of well-adapted seed and forest reproductive material is a fundamental issue under the projected climate change (Ivetić et al. 2016). In this context, the conservation of MaPs and endemic species should be seen as a priority (Fady et al. 2016). Novel SDM techniques and approaches in combination with monitoring efforts of European networks (e.g., EUFGIS or ICP-Forests) will help forest managers to develop mitigation and adaptation strategies. Novel methods to preserve and connect endangered and fragmented rear-edge meta-populations across the Mediterranean basin have to be developed.

The rear-edge of many Mediterranean forest species represented by fragmented habitats in southern Europe is a strategic area of research to detect marginal populations and their precise niche using SDM techniques. An effective approach should integrate range-wide analysis with data on adaptive traits and phenotypic plasticity of provenances from comparative field trials. In fact, combining the information on genetic variation and adaptive potential of species in NFI datasets would provide a very useful tool to detect marginality, possibly useful MaPs, and help their management. For instance, all forest populations lying close to the boundary of suitable/un-suitable niche space in suitability maps (e.g., 5%-10%-20% of probability of detec-

tion) may be studied as candidates for ecological marginality.

The Italian peninsula is widely recognized as a forest biodiversity hotspot (Petit et al. 2003, Noce et al. 2016) based on studies of the environment and climate conditions developed for PDM at a regional scale. The genetic resources of mesophilic species will be increasingly endangered under climate change, and the probability of their retreat or disappearance is high. Moreover, because of their adaptive potential and/or phenotypic plasticity, southern populations may be transferred to more northerly latitudes for conservation purposes and/or to improve the resilience of northern and central-European forests under climate change. In this context, a typical example is the European silver fir from Serra S. Bruno (Calabria, Italy – Kerr et al. 2015).

New interesting scenarios may be explored using the PDM approaches, which could help establish effective recommendations and guidelines to face climate change. On the one hand, local models (i.e., ecological groups) may be more appropriate to predict small endangered populations recognized as ecologically marginal. On the other hand, MaP conservation protocols may be studied, including gene pool improvement using material from different ecological zones that is predicted to grow and adapt well in the selected MaP. The possible introduction into endangered populations of new genotypes from different ecological regions already adapted to forest ecosystems in projected climates, will require additional information which is not provided in NFIs, but that can be inferred through the projected spatial distribution of genetic diversity (Lira-Noriega & Manthey 2014).

#### **Conclusions**

The PDM approach used in this study may represent a valuable step forward for spatial analysis, thanks to the possibility of modelling each group separately. Although statistically-designed and characterized by a wide range of ancillary data, the information from NFIs can improve the high-resolution species distribution models and the provenance suitability models, going beyond the simple concept of “ecological suitability”. All the unweighted methods presented in this study show a high probability of detection of genotypes within marginal forest populations. Exact spatial coordinates of plots and additional information on site quality such as stand age and or site index are required to weight presence/absence properly, and to test new evaluation methods. In our opinion, novel evaluation criteria should be developed to complement the classical true skill statistic (TSS), the area under the receiver operating characteristic curve (AUC, ROC or AUROC) or Cohen’s Kappa. For instance, wBAA and wPDM approaches could be validated using site index and average volume increment for each inventory plot. Fertility

and average growth can not be currently derived from the Italian National Forest Inventory (INFC2005) because age is currently missing and only current volume increment is available.

#### **Acknowledgements**

This work has been supported by the COST Action FP1202 MaP-FGR (<http://map-fgr.entecra.it/>) and TREES4FUTURE FP7 project (<http://www.trees4future.eu/>). We are grateful to all the reviewers for their useful suggestions which improved the early version of this manuscript.

#### **References**

- Allouche O, Tsoar A, Kadmon R (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223-1232. - doi: [10.1111/j.1365-2664.2006.01214.x](https://doi.org/10.1111/j.1365-2664.2006.01214.x)
- Araújo MB, New M (2007). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42-7. - doi: [10.1016/j.tree.2006.09.010](https://doi.org/10.1016/j.tree.2006.09.010)
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3: 327-338. - doi: [10.1111/j.2041-210X.2011.00172.x](https://doi.org/10.1111/j.2041-210X.2011.00172.x)
- Bedia J, Herrera S, Gutiérrez JM (2013). Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global and Planetary Change* 107: 1-12. - doi: [10.1016/j.gloplacha.2013.04.005](https://doi.org/10.1016/j.gloplacha.2013.04.005)
- Borghetti M, Chirici G (2016). Raw data from the Italian National Forest Inventory are on-line and publicly available. *Forest@ - Journal of Silviculture and Forest Ecology* 13 (1): 33-34. [in Italian with English summary] - doi: [10.3832/efor0083-013](https://doi.org/10.3832/efor0083-013)
- Brus DJ, Hengeveld GM, Walvoort DJJ, Goedhart PW, Heidema AH, Nabuurs GJ, Gunia K (2012). Statistical mapping of tree species over Europe. *European Journal of Forest Research* 131: 145-157. - doi: [10.1007/s10342-011-0513-5](https://doi.org/10.1007/s10342-011-0513-5)
- Castaldi C, Vacchiano G, Marchi M, Corona P (2017). Projecting nonnative Douglas fir plantations in Southern Europe with the forest vegetation simulator. *Forest Science* 63: 101-110. - doi: [10.5849/forsci.16-048](https://doi.org/10.5849/forsci.16-048)
- Costantini E, Barbetti R, Fantappiè M, L’Abate G, Lorenzetti R, Napoli R, Marchetti A, Riviaccio R (2014). The soil map of Italy: a hierarchy of geodatabases, from soil regions to sub-systems. In: “GlobalSoilMap Basis of the global spatial soil information system” (Arrouays D, McKenzie N, Hempel J, Richer de Forges A, McBratney AB eds). CRC Press Taylor & Francis Group, London, UK, pp. 109-112.
- Eckert CG, Samis KE, Loughheed SC (2008). Genetic variation across species’ geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* 17: 1170-1188. - doi: [10.1111/j.1365-294X.2007.03659.x](https://doi.org/10.1111/j.1365-294X.2007.03659.x)
- Fady B, Aravanopoulos FA, Alizoti P, Mátyás C, Wühlisch G, Westergren M, Belletti P, Cvjetkovic B, Ducci F, Huber G, Kelleher CT, Khaldi A, Kharrat MBD, Kraigher H, Kramer K, Mühlenthaler U, Peric S, Perry A, Rousi M, Sbay H, Stojnic S, Tijardovic M, Tsvetkov I, Varela MC, Ven-

- dramin GG, Zlatanov T (2016). Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. *Forest Ecology and Management* 375: 66-75. - doi: [10.1016/j.foreco.2016.05.015](https://doi.org/10.1016/j.foreco.2016.05.015)
- Fattorini L (2015). Design-based methodological advances to support national forest inventories: a review of recent proposals. *iForest - Biogeosciences and Forestry* 8 (1): 6-11. - doi: [10.3832/ifer1239-007](https://doi.org/10.3832/ifer1239-007)
- Friedman M (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32: 675-701. - doi: [10.1080/01621459.1937.10503522](https://doi.org/10.1080/01621459.1937.10503522)
- Fick SE, Hijmans RJ (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302-4315. - doi: [10.1002/joc.5086](https://doi.org/10.1002/joc.5086)
- Fridman J, Holm S, Nilsson M, Nilsson P, Ringvall AH, Ståhl G (2014). Adapting National Forest Inventories to changing requirements - the case of the Swedish National Forest Inventory at the turn of the 20<sup>th</sup> century. *Silva Fennica* 48 (3): art-1095, pp. 29. - doi: [10.14214/sf.1095](https://doi.org/10.14214/sf.1095)
- Hamann A, Roberts DR, Barber QE, Carroll C, Nielsen SE (2015). Velocity of climate change algorithms for guiding conservation and management. *Global Change Biology* 21: 997-1004. - doi: [10.1111/gcb.12736](https://doi.org/10.1111/gcb.12736)
- Isaac-Renton MG, Roberts DR, Hamann A, Spiecker H (2014). Douglas-fir plantations in Europe: a retrospective test of assisted migration to address climate change. *Global Change Biology* 20: 2607-2617. - doi: [10.1111/gcb.12604](https://doi.org/10.1111/gcb.12604)
- Ivetic V, Devetaković J, Nonić M, Stanković D, Šijačić-Nikolić M (2016). Genetic diversity and forest reproductive material - from seed source selection to planting. *iForest - Biogeosciences and Forestry* 9 (5): 801-812. - doi: [10.3832/ifer1577-009](https://doi.org/10.3832/ifer1577-009)
- Kerr G, Stokes V, Peace A, Jinks R (2015). Effects of provenance on the survival, growth and stem form of European silver fir (*Abies alba* Mill.) in Britain. *European Journal of Forest Research* 134: 349-363. - doi: [10.1007/s10342-014-0856-9](https://doi.org/10.1007/s10342-014-0856-9)
- Kingsolver JG, Diamond SE (2011). Phenotypic selection in natural populations: what limits directional selection? *The American naturalist* 177: 346-57. - doi: [10.1086/658341](https://doi.org/10.1086/658341)
- Kramer K, Ducousso A, Gömöry D, Hansen JK, Ionita L, Liesebach M, LorenÅ A, Schüler S, Sulkowska M, De Vries S, Von Wühlisch G (2017). Chilling and forcing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between provenances and are phenotypically plastic. *Agricultural and Forest Meteorology* 234: 172-181. - doi: [10.1016/j.agrfor.2016.12.002](https://doi.org/10.1016/j.agrfor.2016.12.002)
- Kruskal W, Wallis W (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47: 583-621. - doi: [10.1080/01621459.1952.10483441](https://doi.org/10.1080/01621459.1952.10483441)
- Lira-Noriega A, Manthey JD (2014). Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution* 68: 1082-1093. - doi: [10.1111/evo.12343](https://doi.org/10.1111/evo.12343)
- Manel S, Williams HC, Ormerod SJ (2001). Evaluating presence absence models in ecology; the need to count for prevalence. *Journal of Applied Ecology* 38: 921-931. - doi: [10.1046/j.1365-2664.2001.00647.x](https://doi.org/10.1046/j.1365-2664.2001.00647.x)
- Marchi M, Castaldi C, Merlini P, Nocentini S, Ducci F (2015). Stand structure and influence of climate on growth trends of a marginal forest population of *Pinus nigra* spp. *nigra*. *Annals of Silvicultural Research* 39: 100-110. - doi: [10.12899/asr-1066](https://doi.org/10.12899/asr-1066)
- Marchi M, Nocentini S, Ducci F (2016). Future scenarios and conservation strategies for a rear-edge marginal population of *Pinus nigra* Arnold in Italian central Apennines. *Forest Systems* 25: e072. - doi: [10.5424/ifs/2016253-09476](https://doi.org/10.5424/ifs/2016253-09476)
- Marchi M, Chiavetta U, Castaldi C, Ducci F (2017). Does complex always mean powerful? A comparison of eight methods for interpolation of climatic data in Mediterranean area. *Italian Journal of Agrometeorology* 1: 59-72. - doi: [10.19199/2017.1.2038-5625.059](https://doi.org/10.19199/2017.1.2038-5625.059)
- McRoberts R, Tomppo E, Schadauer K, Vidal C, Ståhl G, Chirici G, Lanz A, Cienciala E, Winter S, Smith B (2009). Harmonizing national forest inventories. *Journal of Forestry* 58: 179-187. [online] URL: <http://www.researchgate.net/publication/274106424>
- Merow C, Smith MJ, Edwards TC, Guisan A, McMahon SM, Normand S, Thuiller W, Wüest RO, Zimmermann NE, Elith J (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography* 37: 1267-1281. - doi: [10.1111/ecog.00845](https://doi.org/10.1111/ecog.00845)
- Merow C, Smith MJ, Silander JA (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058-1069. - doi: [10.1111/j.1600-0587.2013.07872.x](https://doi.org/10.1111/j.1600-0587.2013.07872.x)
- Molyneux N, Soares I, Neto F (2014). Modeling current and future climates using WorldClim and DIVA software: case studies from Timor Leste and India. *Journal of Crop Improvement* 28: 619-640. - doi: [10.1080/15427528.2014.924369](https://doi.org/10.1080/15427528.2014.924369)
- Myers N, Mittermeier R, Fonseca G, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858. - doi: [10.1038/35002501](https://doi.org/10.1038/35002501)
- Naimi B, Skidmore AK, Groen TA, Hamm NAS (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modeling. *Journal of Biogeography* 38: 1497-1509. - doi: [10.1111/j.1365-2699.2011.02523.x](https://doi.org/10.1111/j.1365-2699.2011.02523.x)
- Noce S, Collalti A, Valentini R, Santini M (2016). Hot spot maps of forest presence in the Mediterranean basin. *iForest - Biogeosciences and Forestry* 9 (5): 766-774. - doi: [10.3832/ifer1802-009](https://doi.org/10.3832/ifer1802-009)
- O'Brien EK, Mazanec RA, Krauss SL (2007). Provenance variation of ecologically important traits of forest trees: implications for restoration. *Journal of Applied Ecology* 44: 583-593. - doi: [10.1111/j.1365-2664.2007.01313.x](https://doi.org/10.1111/j.1365-2664.2007.01313.x)
- Orlovic S, Ivankovic M, Andonoski V, Stojnic S, Isajev V (2014). Forest genetic resources to support global bioeconomy. *Annals of Silvicultural Research* 38: 51-61. - doi: [10.12899/asr-942](https://doi.org/10.12899/asr-942)
- Petit RJ, Aguinalde I, De Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563-1565. - doi: [10.1126/science.1083264](https://doi.org/10.1126/science.1083264)
- Petr M, Boerboom L, Ray D, Van Der Veen A (2014). An uncertainty assessment framework for forest planning adaptation to climate change. *Forest Policy and Economics* 41: 1-11. - doi: [10.1016/j.forpol.2013.12.002](https://doi.org/10.1016/j.forpol.2013.12.002)
- Piotti A, Leonarduzzi C, Postolache D, Bagnoli F, Spanu I, Brousseau L, Urbinati C, Leonardi S, Vendramin GG (2017). Unexpected scenarios from Mediterranean refugial areas: disentangling complex demographic dynamics along the Apennine distribution of silver fir. *Journal of Biogeography* 44 (7): 1547-1558. - doi: [10.1111/jbi.13011](https://doi.org/10.1111/jbi.13011)
- R Core Team (2017). R: a language and environment for statistical computing. R Foundation, Vienna, Austria. [online] URL: <http://www.r-project.org>
- Ray D, Petr M, Mullett M, Bathgate S, Marchi M, Beauchamp K (2017). A simulation-based approach to assess forest policy options under biotic and abiotic climate change impacts: A case study on Scotland's National Forest Estate. *Forest Policy and Economics* 89: 1-11. - doi: [10.1016/j.forpol.2017.10.010](https://doi.org/10.1016/j.forpol.2017.10.010)
- Rehfeldt GE, Worrall JJ, Marchetti SB, Crookston NL (2014). Adapting forest management to climate change using bioclimate models with topographic drivers. *Forestry* 88: 528-539. - doi: [10.1093/forestry/cpv019](https://doi.org/10.1093/forestry/cpv019)
- Rose L, Leuschner C, Köckemann B, Buschmann H (2009). Are marginal beech (*Fagus sylvatica* L.) provenances a source for drought tolerant ecotypes? *European Journal of Forest Research* 128: 335-343. - doi: [10.1007/s10342-009-0268-4](https://doi.org/10.1007/s10342-009-0268-4)
- San-Miguel-Ayanz J, De Rigo D, Caudullo G, Houston Durrant T, Mauri A (2016). European atlas of forest tree species. Publication Office of the European Union, Luxembourg. - doi: [10.2788/4251](https://doi.org/10.2788/4251)
- Schueler S, Kapeller S, Konrad H, Geburek T, Mengl M, Bozzano M, Koskela J, Lefèvre F, Hubert J, Kraigher H, Longauer R, Olrik DC (2013). Adaptive genetic diversity of trees for forest conservation in a future climate: A case study on Norway spruce in Austria. *Biodiversity and Conservation* 22: 1151-1166. - doi: [10.1007/s10531-012-0313-3](https://doi.org/10.1007/s10531-012-0313-3)
- Swets JA (1988). Measuring the accuracy of diagnostic systems. *Science* 240: 1285-1293. - doi: [10.1126/science.3287615](https://doi.org/10.1126/science.3287615)
- Thuiller W, Pollock LJ, Gueguen M, Münkemüller T (2015). From species distributions to meta-communities. *Ecology Letters* 18 (12): 1321-1328. - doi: [10.1111/ele.12526](https://doi.org/10.1111/ele.12526)
- Thuiller W, Georges D, Engler R (2016). biomod2: ensemble platform for species distribution modeling. R package version 3:3-7. [online] URL: <http://CRAN.R-project.org/package=biomod2>
- Tinner W, Colombaroli D, Heiri O, Henne PD, Steinacher M, Untenecker J, Vescovi E, Allen JRM, Carraro G, Conedera M, Joos F, Lotter AF, Luterbacher J, Samartin S, Valsecchi V (2013). The past ecology of *Abies alba* provides new perspectives on future responses of silver fir forests to global warming. *Ecological Monographs* 83: 419-439. - doi: [10.1890/12-2231.1](https://doi.org/10.1890/12-2231.1)
- Vacchiano G, Motta R (2015). An improved



species distribution model for Scots pine and downy oak under future climate change in the NW Italian Alps. *Annals of Forest Science* 72: 321-334. - doi: [10.1007/s13595-014-0439-4](https://doi.org/10.1007/s13595-014-0439-4)  
Williams MI, Dumroese RK (2013). Preparing for climate change: forestry and assisted migration. *Journal of Forestry* 111: 287-297. - doi: [10.5849/jof.13-016](https://doi.org/10.5849/jof.13-016)

### Supplementary Material

**Tab. S1** - Directly-calculated variables from the ClimateEU software and included into the PCA.

**Tab. S2** - Derived climatic indices from the ClimateEU software and included into the PCA.

**Tab. S3** - Derived seasonal variables from the ClimateEU software and included into the PCA.

**Tab. S4** - Non-parametric ANOVA on the combination between methods and algorithms.

**Link:** [Marchi\\_2441@suppl001.pdf](#)