Marginality indices for biodiversity conservation in forest trees

Nicolas Picard [a], Maurizio Marchi [b,*], Maria Jesus Serra-Varela [c], Marjana Westergren [d], Stephen Cavers [e], Eduardo Notivol [f], Andrea Piotti [b], Paraskevi Alizoti [g], Michele Bozzano [h], Santiago C. González-Martínez [i], Delphine Grivet [c,j], Filippos A. Aravanopoulos [g], Giovanni Giuseppe Vendramin [b], Fulvio Ducci [k], Bruno Fady [l], Ricardo Alía [c,j]

[a] GIP ECOFOR, Paris, France
[b] CNR - Institute of Biosciences and BioResources, Florence Research Area, Italy
[c] Mixed Unit INIA-UVA on Sustainable Forest Management, Palencia, Spain
[d] Slovenian Forestry Institute, Ljubljana, Slovenia
[e] UK Centre for Ecology & Hydrology, Bush Estate, Penicuik, United Kingdom
[f] Departamento de Sistemas Agrícolas, Forestales y Medio Ambiente, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Zaragoza, Spain
[g] School of Forestry and Natural Environment, Aristotle University of Thessaloniki, Thessaloniki, Greece
[h] European Forest Institute, Barcelona, Spain
[i] INRAE, Univ. Bordeaux, BIOGECO, Cestas, France
[j] INIA-CSIC, Dpt. Genetics and Ecophysiology, Madrid, Spain
[k] CREA – Research centre for Forestry and Wood, Arezzo, Italy
[l] NRAE, URFM, Ecology of Mediterranean Forests, Avignon, France

ARTICLE INFO

ABSTRACT

Marginal and peripheral populations are important for biodiversity conservation. Their original situation in a species' geographic and ecological space often confers them genetic diversity and traits of high adaptive value. Yet theoretical hypotheses related to marginality are difficult to test because of confounding factors that influence marginality, namely environment, geography, and history. There is an urgent need to develop metrics to disentangle these confounding factors. We designed nine quantitative indices of marginality and peripherality that define where margins lie within species distributions, from a geographical, an environmental and a historical perspective. Using the distribution maps of eight European forest tree species, we assessed whether these indices were idiosyncratic or whether they conveyed redundant information. Using a database on marginal and peripheral populations based on expert knowledge, we assessed the capacity of the indices to predict the marginality status of a population. There was no consistent pattern of correlation between indices across species, confirming that the indices conveyed different information related to the specific geometry of the species distributions. Contrasting with this heterogeneity of correlation patterns across species, the relative importance of the indices to predict the marginality status of populations was consistent across species. However, there was still a significant country effect in the marginality status, showing a variation in expert opinion of marginality vis-á-vis the species distribution. The marginality indices that we developed are entirely based on distribution maps and can be used for any species. They pave the way for testing hypotheses related to marginality and peripherality, with important implications in quantitative ecology, genetics, and biodiversity conservation.

1. Introduction

Mapping and characterising marginal populations is essential to numerous research and applied fields in ecology, especially for biodiversity conservation (Lesica and Allendorf, 1992; Araújo and Williams, 2001; Abeli et al., 2018). Referring to Soulé (1973), we here define a marginal population of a species as "a population exposed to an extreme of one or more relevant variables" across the species distribution. Sticking to this definition, marginality is a direct result of the statistical distribution of these variables across the species' range. Environmental marginality follows from environmental variables. Peripherality (a synonym for geographical marginality) follows from geographical variables (i.e. spatial coordinates in the geographical space) (Soulé, 1973). It has been frequently argued that marginal populations deserve priority and specific management for their protection (Rehm et al., 2015; Fady et al., 2016). However, too much focus on these populations can also bias conservation (Araújo and Williams, 2001).

Because marginality is not defined by species characteristics such as abundance, genetics or demography, there is great interest in testing whether marginality is related to these characteristics. Several theories and hypotheses predict such relationships. According to the abundant-centre hypothesis, a species' abundance is typically higher at the centre of its geographical range and lower toward its edges (Sagarin and Gaines, 2002). In accordance with this hypothesis, marginal and peripheral populations may exhibit different traits from central populations in terms of distribution, demography, size, adaptive value, genetic makeup, etc. (Soulé, 1973; Sagarin and Gaines, 2002; Hampe and Petit, 2005; Eckert et al., 2008; Pironon et al., 2017).

Peripheral populations are expected to have lower genetic diversity due to genetic drift and inbreeding (Frankham, 1996). They are also expected to have lower genetic fitness than core populations if local adaptation is constrained by gene flow from more central populations. However, gene flow from central populations could also increase genetic diversity, and consequently the probability of successful adaptation (Kremer et al., 2012). Geographic distance reduces gene flow. Therefore, close populations are expected to be more genetically similar than separated ones ('isolation-by-distance' pattern; Wright, 1943). Similarly, gene flow is more likely to be successful between similar environments (Sexton et al., 2014). Therefore, populations in similar environments are expected to have more genetic resemblance than those in different environments ('isolation-by-environment' or 'isolation-by-adaptation'; Nosil et al., 2008). It has been shown that ecological marginality gradients explain variation in species' demographic performance better than geographical gradients (Pironon et al., 2017).

Genetic structure in peripheral populations can also be influenced by both historical and contemporary demographic changes. Leading and rear edges of range expansions that took place after the last glaciation differ in time of establishment, demographic stochasticity, within and among population diversity and regional diversity (Hampe and Petit, 2005).

Theoretical expectations for marginal/peripheral populations remain largely unresolved because of contradictory results brought by empirical studies (Sagarin and Gaines, 2002; Eckert et al., 2008; Pironon et al., 2017). Confounding effects can bias the comparisons between central and peripheral populations. It is currently not clear to what extent geographical and ecological marginality gradients impact the main assumption of the centre-periphery hypothesis (CPH) (Pironon et al., 2017). Therefore, any test of the CPH should disentangle the effects of the different factors (distribution, ecology, geography and history) on the demography and genetics of populations (Pironon et al., 2017).

The difficulty to define a marginal population (Soulé, 1973) is a major reason for the lack of conclusive results from empirical studies testing the CPH hypothesis, or the ones related. Indeed, most of these studies use a relatively arbitrary categorical definition of marginality (Eckert et al., 2008). Only few studies use a quantitative metric of marginality. Channell and Lomolino, 2000 used centrality indices (distance or area to the core). Schwartz et al., 2003 used the area necessary to fit a given number of home ranges. Araújo and Williams, 2001 used aspects related to the density of the distributions. Yet metrics to disentangle confounding factors implicit in the CPH are lacking. Therefore, the development of such metrics to assess the marginality of populations and their importance in conservation programs is very urgent (Lesica and Allendorf, 1992; Hardie and Hutchings, 2010).

Our objective here is to provide metrics for identifying where margins lie within species distributions, from a geographical, an environmental and a historical perspective. We defined nine quantitative indices of marginality and peripherality using data from eight ecologically and economically important European forest tree species. In a first step, we examined correlations between indices for each species. This analysis allowed us to identify indices that were idiosyncratic and those showing the same relationship regardless of the species considered. In a second step, we relied on an expert database composed of 1,252 marginal and peripheral gene conservation candidate populations. We assessed the capacity of the indices to predict the marginality status of these populations.

2. Materials and methods

2.1. Species and data

We selected eight widely distributed forest tree species which are part of gene conservation activities under the auspices of EUFORGEN (de Vries et al., 2015), namely *Abies alba* Mill., *Fagus sylvatica* L., *Picea abies* (L.) H.Karst., *Pinus halepensis* Mill., *Pinus nigra* J.F.Arnold, *Pinus pinaster* Aiton, *Pinus pinea* L. and *Pinus sylvestris* L. These species are interesting models to investigate the still unknown relationship between marginality and genetic diversity (Fady et al., 2022). We considered only the European range of the species, limited by the Ural Mountains that often represent a natural limit for subspecies or gene pools (Tsuda et al., 2016).

The common basis for the computation of all marginality indices is the species distribution maps. We here used those provided by Caudullo et al. (2020) that are characterized by very smooth contours and little detail at small distances. It implies that fine scale results ($\lesssim 50$ km) were not considered in the current study. The computation of the indices was based on the polygons that define continuous area of natural occupancy of the species range. Isolates (point features in Caudullo et al., 2020's shapefiles) and synanthropic occurrences outside the natural range were both disregarded.

A dataset on marginal and peripheral populations was collected for each species based on expert knowledge (Ducci and Donnelly, 2017). We here used the geographic coordinates of a total of 1,252 marginal populations identified by experts (Table 1).

Table 1
Number of populations identified as marginal by experts for each species (Ducci and Donnelly, 2017).

Species	Nb. of marginal populations	Area of the species distribution (millions ha)
Abies alba	331	73.4
Fagus sylvatica	221	227.1
Picea abies	332	375.2
Pinus halepensis	47	33.6
Pinus nigra	29	42.5
Pinus pinaster	35	29.0
Pinus pinea	88	5.0
Pinus sylvestris	169	611.6
Total	1252	1397.4

2.2. Marginality indices

We designed nine marginality indices: one environmental, six peripheral, and two historical (Table 2). The three kinds of indices correspond to three spatio-temporal dimensions of variability (environment, geography, and history). Each of these dimensions may have an effect on the species' traits and population features (see Fig. 8 in Pironon et al., 2017 for a conceptual framework). All indices were computed from the species distribution map. The environmental index is model-based. Its values are the predictions of a model trained on the distribution map using environmental variables as predictors. All other indices are data-driven and were constructed geometrically from the maps.

Analyses based on correlations (but not those based on rank statistics) require indices to have comparable magnitudes of variation. Therefore, some of the indices were transformed. When the maximal value of an index across the entire species distribution differed by several orders of magnitude from its minimal value, a log-transform was applied. To standardize the distribution of a marginality index across the species distribution, the quantile transform was used. It replaces an index value by its quantile according to the empirical distribution function of the index across the entire species distribution. Transformed values thus vary between zero and one. All computations of the marginality indices were performed in the R software (see Appendix S2 for details).

2.2.1. Environmental marginality index

The index was defined using a species distribution model based on climatic conditions (Pecchi et al., 2019; Tessarolo et al., 2021), thus resulting in an index of climate marginality. We used climatic variables from WorldClim portal for the 1961–1990 normal period (version 1.4, Hijmans et al., 2005; see Appendix S2 for more details). An ensemble modelling approach was used. Five different algorithms were used for model fitting: Generalised Linear Model, Generalised Additive Model, Artificial Neural Network, Multivariate Adaptive Regression Splines and Random Forest. The modelling steps were accomplished using the species distribution maps from Caudullo et al. (2020) as presence data and by sampling a random sample of absences. The same weight was given to presences and absences by extracting an equal number of each. To improve accuracy, 10 different pseudo-absences datasets were generated and modelling was repeated 15 times averaging the results. In total, $5 \times 10 \times 15 = 750$ models were fitted for each species. Finally, the ensemble model was the weighted mean of these fits, using single-model performance as weight. Model performance was assessed by means of True Skill Statistic and Area Under the receiver operating characteristic Curve (AUC; Fielding and Bell, 1997). Probabilistic model outputs based on climatic conditions were transformed to quantiles in order to provide the climate marginality index (see maps of the index in Appendix S1).

2.2.2. Peripherality indices

There is no univocal way to identify extreme values in the geographical coordinates of a species distribution. Peripherality may refer to eccentricity, proximity to the border of the distribution, or isolation with respect to the main bulk of the distribution (Fig. 1). We designed six peripherality indices that we called: area, gravity, centroid, edge, isolation and second-nearest-core. With the exception of the centroid index, all peripherality indices relied on a morphological spatial pattern analysis (MSPA). MSPA uses mathematical morphological operators to segment a binary image into different classes depending on their connectivity within the image (Soille and Vogt, 2009). In our case, the binary image corresponded to the species distribution map with the foreground given by the presence land cells and the background given by the absence land cells and water cells. MSPA was used to divide the species distribution map into three categories: (i) cores, i.e. continuous patches of presences excluding their perimeters; (ii) edges (i.e.
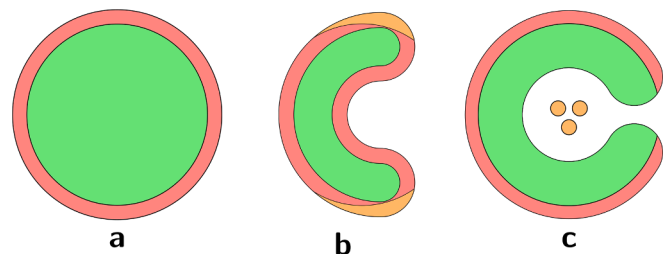
Table 2
List of the marginality indices used with their main characteristics. "Feature metric" indicates the feature that the index measures at each location. "Link with marginality" indicates how marginality (M) varies when the index (I) varies.

Index	Feature metric	Link with marginality	Construction
Environmental marginality indices			
Climatic	Climate suitability for the species	M decreasing with I	Model-based*
Peripherality indices			
Area	Size of the nearest population	M decreasing with I	Data-driven†
Gravity	Size of the nearest population	M decreasing with I	Data-driven
Centroid	Eccentricity with respect to the species distribution	M increasing with I	Data-driven
Edge	Distance to the border of the species distribution	M decreasing with I	Data-driven
Isolation	Isolation with respect to the species distribution	M increasing with I	Data-driven
Second-nearest-core	Isolation with respect to the species distribution	M increasing with I	Data-driven
Historical marginality indices			
North/south	Proximity to leading-edge or rear-edge along the latitudinal range	M U-shaped with I	Data-driven
East/west	Proximity to leading-edge or rear-edge along the longitudinal range	M U-shaped with I	Data-driven

* Model trained on the species distribution using climate covariates.
† Geometric construction based on the species distribution.

Fig. 1. Diagrams showing that extreme locations to identify peripheral populations may correspond to different geographical areas depending on the geometry of the distribution (= the coloured areas). (a) When the distribution is a disk, eccentric areas (= areas far from the centre of the distribution), areas adjacent to the border of the distribution, and areas isolated from the bulk (in green) of the distribution are the same (in red). (b) When the distribution is elongated, eccentric areas (in orange) differ from areas adjacent to the border (in red). (c) When the distribution is fragmented, eccentricity, borderness and isolation differ; here the three small disks (in orange) are neither eccentric nor adjacent to the border, but they are isolated from the bulk.

external core perimeters) and perforations (i.e. internal core perimeters), hereafter jointly designated as "edges"; and (iii) other MSPA classes (i.e. islets, loops, bridges and branches) that were not considered for the definition of the peripherality indices.

The area index was defined for any location as the area of the core that is the nearest to this location. Core area can be considered as a proxy for the effective size of the population that is the nearest to the location considered (Munguía-Rosas and Montiel, 2014). Yet other factors (e.g. habitat quality, connectivity) also determine this size. The area index has the advantage to provide straightforward interpretation but has the limitation to be spatially discontinuous (see maps in Appendix S1). Spatial discontinuity means that two populations can be as close to each other as desired and yet have very different index values. The area index was log-transformed for further analyses.

The gravity index was defined for any location as the weighted mean of the core areas with weights inversely proportional to the distance squared from the location to the cores:

$$\text{gravity index} = \frac{\sum_{i=1}^{n} A_i \Big/ d_i(\mathbf{x})^2}{\sum_{i=1}^{n} 1 \Big/ d_i(\mathbf{x})^2}$$

where n is the number of cores, A_i is the area of the ith core and $d_i(\mathbf{x})$ is the distance from location $\mathbf{x}$ to the ith core. This index is spatially continuous. We named it after gravity because its formula is similar to the gravity force if we assimilate the core area to a mass. The gravity index is log-transformed for further analyses. This index is related to the contagion index (Araújo and Williams, 2001). It was considered to solve the issue of the spatial discontinuity of the area index, while conveying an information similar to that index (compare maps in Appendix S1).

The centroid of a species distribution was defined as the location whose geographic coordinates are the averages of the geographic coordinates of all locations where the species is present. The centroid index was then defined for any location as the cost distance from the centroid to this location. Cost distances were determined using a cost path analysis and a conductance matrix (the inverse of a resistance matrix) that reflected the conductance of gene flow. Sea cells were assigned low conductance (10), land cells without the species intermediate conductance (50), and land cells where the species was present high conductance (100). The centroid index thus represents how far (in terms of gene flow) a location is from the centre of the species distribution.

The edge index was defined for any location as the distance from this location to the nearest core edge. To differentiate between locations inside and outside the cores, the distances of locations lying out of the core areas were considered as negative. The edge index thus indicates how far a location is from cores (if outside cores) or how deep inside the core it can be found.

The isolation index was defined for any location as the distance from this location to the nearest core greater than 100 ha and further than 50 km from the location. The isolation index can be considered as a measure of the level of broad-scale connectivity between cores. This measure of isolation has the advantage to match the isolation-by-distance model, where gene flow is most likely between neighbouring populations. The minimum core area of 100 ha roughly complies with the 50/500 rule of population viability (Franklin et al., 1980). However, this index has the limitation to be spatially discontinuous like the area index, and is moreover sensitive to the minimum size of the cores considered to be near.

The second-nearest-core index was defined for any location as the distance from this location to the second nearest core greater than 100 ha. The second-nearest-core index was designed as a spatially continuous alternative to the isolation index.

2.2.3. Historical marginality indices

Historical indices aimed to reflect the migration and demographic changes of the species, which can affect the CPH (Hampe and Petit, 2005). Such changes include the most common geographical trends known in forest species (Alberto et al., 2013), as well as those related to putative ice-age refugia.

The north/south index was defined for any location as the quantile of the latitude of the location according to the empirical distribution function of latitudes over the entire species distribution. It varies between zero when the location is further south than the southernmost limit of the distribution, and one when the location is further north than the northernmost limit (see maps in Appendix S1). Similarly, the east/west index was defined for any location as its longitude transformed to quantiles.

2.3. Analyses

MSPA is sensitive to the spatial grain of data and to the smoothness of map contours. Therefore, a sensitivity analysis was conducted for those marginality indices that depend on MSPA (i.e. the area, gravity, edge, isolation and second-nearest-core index). The analysis was based on random noise added to the species distribution. More details on this sensitivity analysis are provided in Appendix S3.

2.3.1. Relationship between indices

Relationships between indices were assessed to check whether the different indices conveyed different or redundant information on marginality. For a given species, the correlations between indices depend on the specific geometry of its spatial distribution. Because the different species distributions have different geometries, we expect the different indices to convey different information only if the pattern of correlations between indices differs from one species to another.

For each species s, correlations between marginality indices were computed in two different ways. First, linear relationships between indices were computed. We performed a principal component analysis (PCA) of the matrix giving the value of each index at each point of presence of the species. As a result, for each species s, a 9×2 matrix $\mathbf{C}_s$ giving the coordinates of the nine marginality indices along the first two principal components of the PCA was obtained. PCA was performed using the 'ade4' package of R.

Second, non-linear relationships between indices were computed based on marginal areas. For each index, we computed the percentiles of the index values in the species distribution. We then segmented the species distribution into marginal and non-marginal areas depending on how the index compared to its percentiles. Right-tail ($\geqslant 90\%$), left-tail ($\leqslant 10\%$) or two-tail ($\geqslant 95\%$ or $\leqslant 5\%$) percentiles were used depending on whether the indices increased, decreased or had a U-shaped relationship with marginality (Table 2). Fig. 2 shows for instance the resulting marginal and non-marginal areas of *Abies alba* according to each marginality index. The non-linear relationship between two marginality indices i and j was then assessed using the Jaccard coefficient J_{ij} between the marginal areas they delimited:

$$J_{ij} = \frac{|M_i \cap M_j|}{|M_i \cup M_j|}$$

where M_i is the marginal area delimited by index i and $|M_i|$ gives the area of a geographic zone. Given the definition of M_i, $|M_i|$ always equalled 10% of the area of the species distribution. The Jaccard coefficient varies between zero when there is no overlap between the marginal areas delimited by the two indices, and one when they completely overlap. For each species s, we finally used the classical metric multidimensional scaling to place the nine species on a bidimensional map. This operation reduced the 9×9 matrix of Jaccard coefficients into a 9×2 matrix $\mathbf{M}_s$ (p.31 Cox and Cox, 2001).

Matrix $\mathbf{C}_s$ for linear relationships and matrix $\mathbf{M}_s$ for non-linear relationships play a similar role: the closer two indices are in the space defined by these matrices, the greater the correlation between these two
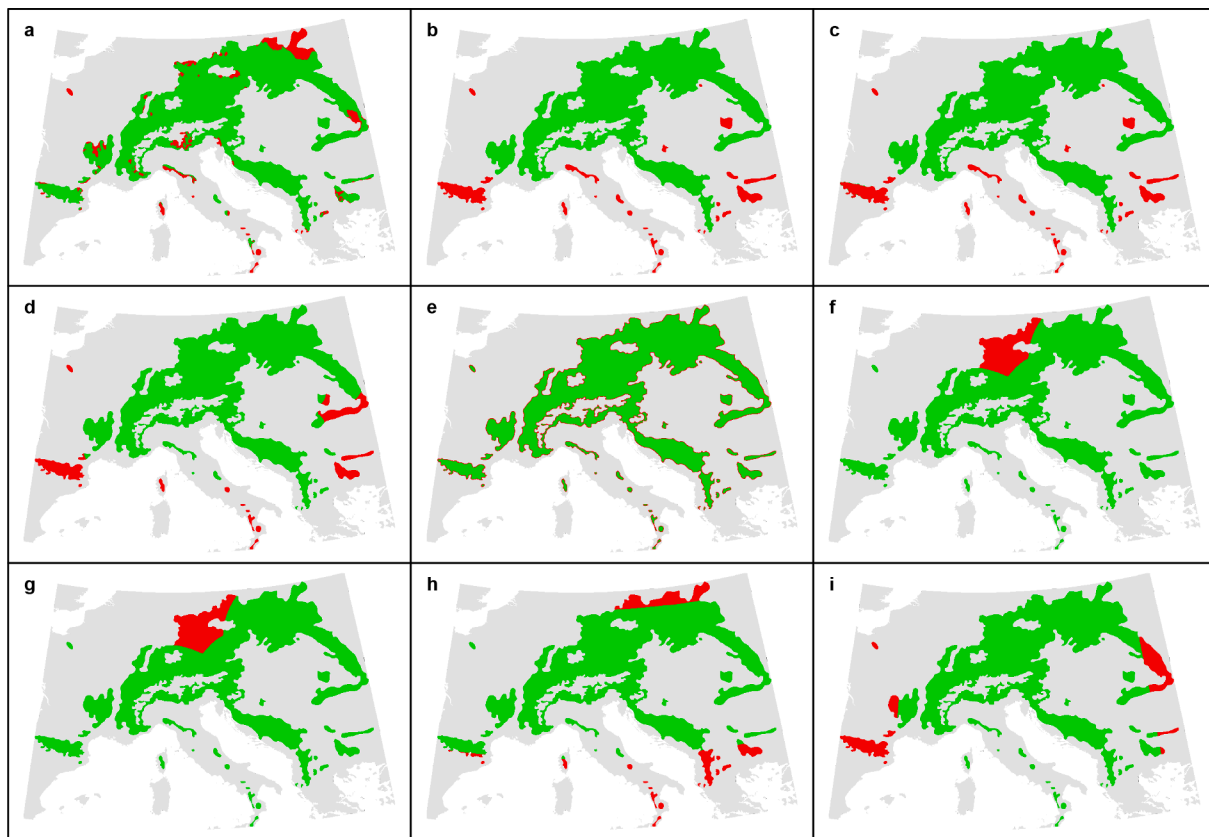
Fig. 2. Map of the marginal (red) and non-marginal (green) areas in the distribution of *Abies alba* according to the different marginality indices: (a) climatic, (b) area, (c) gravity, (d) centroid, (e) edge, (f) isolation, (g) second-nearest-core, (h) north/south, and (i) east/west. The delimitation between marginal and non-marginal areas is defined by one-tail or two-tail cut-off percentile(s) in the distribution of the index such that 10% of the locations be marginal.

indices for species *s*. If the different marginality indices convey different information, we expect matrices $\mathbf{C}_s$ to differ between species. On the contrary, if the different indices convey redundant information, we expect matrices $\mathbf{C}_s$ to be similar for all species. The same holds for the set of matrices $\mathbf{M}_s$. For any pair of species (s, t), the similarity between matrices $\mathbf{C}_s$ and $\mathbf{C}_t$ (or between matrices $\mathbf{M}_s$ and $\mathbf{M}_t$) was assessed using the similarity index defined by Indahl et al. (2018). This similarity index is implemented in the R software by the function 'SMI' of the package 'MatrixCorrelation'. It varies between zero (when the two matrices remain dissimilar even after transforming them with Procrustes rotations) and a maximum of one (when the two matrices are equal to within Procrustes transformations). Significance of the similarity was estimated using the permutation test provided.

2.3.2. Predicting the marginality status of a population

Based on the expert database, we used models to assess the capacity of the marginality indices to predict the marginality status of a tree population. In addition to the marginality indices, the country where each marginal population is located was included as a categorical predictor in the model. This categorical predictor allowed us to test for a country effect in the expert assessment of marginality. The model must classify any tree population as being marginal or not, i.e. be a binary classifier. The training data consisted of populations that are known to be marginal according to expert knowledge, but we had no information on populations known to be non-marginal. Hence, the data consisted of presence-only data. Moreover, the predictors were mapped and the marginal populations were known by their geographic locations. The maximum entropy (Maxent) model is a binary classifier that is suitable in a geographic context with presence-only data (Phillips and Dudík, 2008; see Appendix S2 for more details).

The overall quality of fit of the Maxent model was estimated using

the AUC. The relative importance of a variable was assessed by computing the training gain (= the opposite of the minimized log loss, see Appendix S2) when using this variable as the only predictor. Another measure of the variable importance was the training loss when removing this variable from the full model with all predictors. Both the training gain and the training loss were expressed as a percentage of the minimized log loss of the full Maxent model with all predictors. In addition, the percent contribution of each variable was also reported (Phillips and Dudík, 2008). We used the "cloglog" output of Maxent that can be interpreted as an estimate between 0 and 1 of the probability of being marginal (Phillips and Dudík, 2008). Maxent models were fitted using the 'dismo' package in R with default settings.

3. Results

Indices that are not based on MSPA (i.e. the climatic index, the centroid index and the historical marginality indices) are hardly sensitive to noise in the distribution map. Therefore, results with these indices do not depend much on the quality and spatial grain of maps. Among the five indices that rely on MSPA, the edge index that does not depend on the cores found by the MSPA is moderately sensitive to noise. On the contrary, the four others are highly sensitive to noise (Appendix S3). Hence, for these four indices, results below are also conditional on the smoothness and large grain of Caudullo et al., 2020's maps.

3.1. Relationship between marginality indices

For all species, there was a positive correlation between the area index and the gravity index on one hand, and between the isolation index and the second-nearest-core index on the other. Apart from these two pairs of redundant indices, the pattern of correlations between the

marginality indices generally differed from one species to another (Fig. 3). For instance, the centroid index and the east/west index were largely independent for *Fagus sylvatica* (Fig. 3b), negatively correlated for *Picea abies* (Fig. 3c), and positively correlated for *Pinus halepensis* (Fig. 3d). One exception to this overall heterogeneity of correlation patterns was the similar correlation pattern shown by *Picea abies* and *Pinus sylvestris* (compare panels c and h in Fig. 3). These two species have distributions with a similar geometry extending far to the east.

Reflecting this heterogeneity of correlation patterns, the similarities between species based on their matrices **C**$_s$ were low overall (Table 3). Several pairs of species had very low similarities (e.g. *Fagus sylvatica* and *Pinus nigra*, or *Abies alba* and *Pinus pinaster*, or *Pinus nigra* and *Pinus pinaster*). In agreement with the similarity of their correlation patterns, *Picea abies* and *Pinus sylvestris* were an exception and showed a high similarity value. This value was the only one that did not correspond to significantly dissimilar matrices according to the permutation test.

Similar results were obtained when considering the Jaccard coefficients between marginal areas. Again, the area and the gravity indices delimited very similar marginal areas (compare for instance panels b and c in Fig. 2 for *Abies alba*). So did the isolation and the second-nearest-core indices (compare panels f and g in Fig. 2). As a consequence, when mapped using multidimensional scaling, the area

and the gravity indices were located close to each other for all species (Fig. 4). So were the isolation index and the second-nearest-core indices. Apart from these two pairs of redundant indices, the pattern of marginal areas delimited by the marginality indices generally differed from one species to another. One exception again was the pair *Picea abies* and *Pinus sylvestris* (compare panels c and h in Fig. 4). Accordingly, the similarities between species based on their matrices **M**$_s$ were low overall, with the exception of *Picea abies* and *Pinus sylvestris* that had a high similarity value (Table 4).

3.2. Predicting the marginality status from indices

Because the area index and the isolation indices were redundant with other indices, they were not considered to predict the marginality status of populations determined by experts. On the contrary, the country was considered a predictor. Overall, the relative importance of the different predictors was consistent across species (as shown by the length of the bars Fig. 5). The edge index was consistently the most relevant variable to predict the marginality status of tree populations, followed by the country where the population was located, and the climatic index. On the other hand, the centroid, the north/south and the second-nearest-core indices were overall not very relevant to predict the marginality
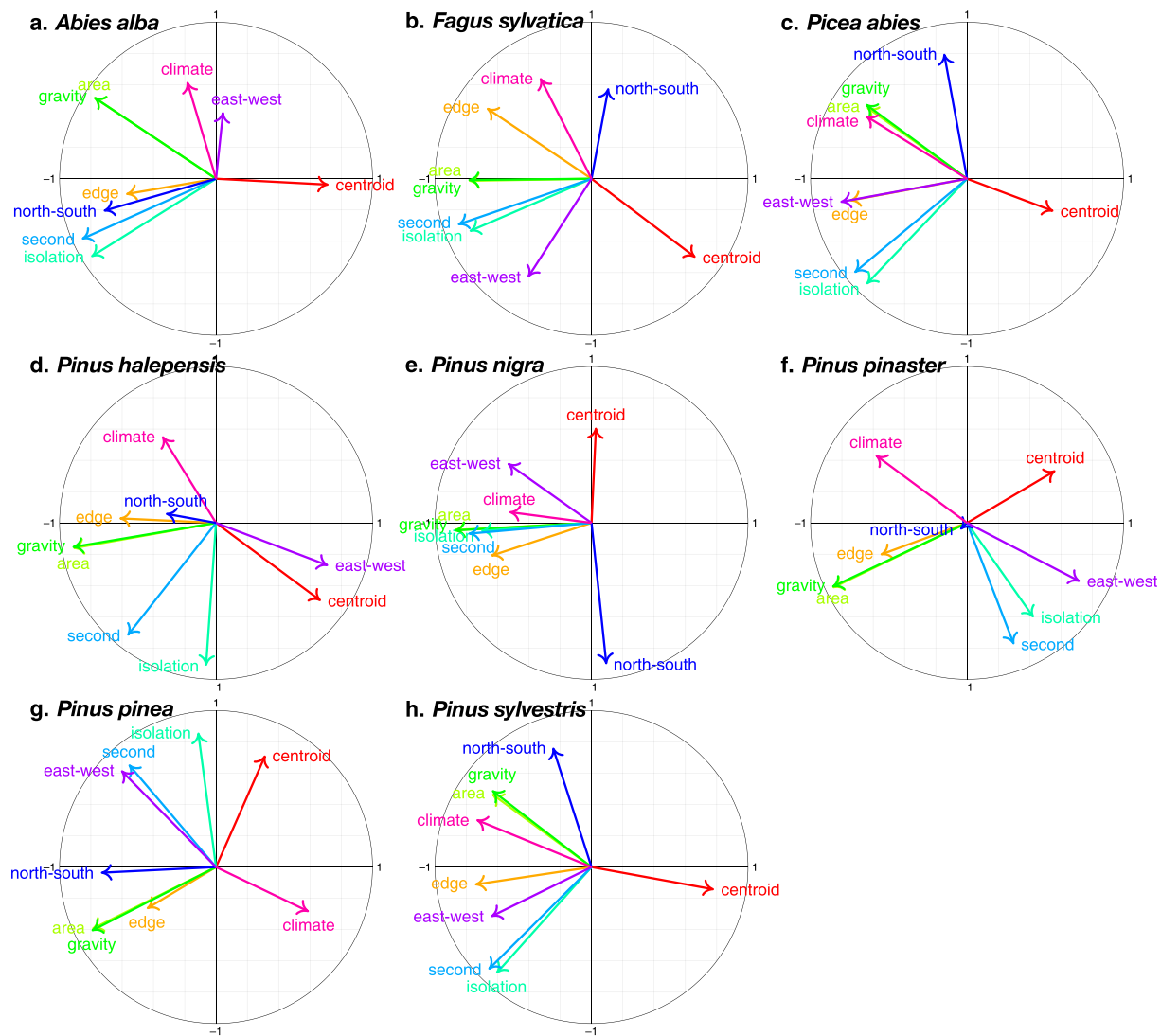

Fig. 3. Plot of matrices **C**$_s$ for each species. Each panel shows the correlation circle of the principal component analysis of the nine marginality indices at all locations where the species is present. The x-axis gives the correlation with the first principal component. The y-axis gives the correlation with the second principal component. Each arrow represents a marginality index. The coordinates of the extremities of the arrows are given by matrices **C**$_s$.

Table 3

Matrix of similarity between the eight species, using the similarity index by Indahl et al. (2018) between matrices $\mathbf{C}_s$. For each species s, $\mathbf{C}_s$ gives the coordinates of the marginality indices along the first two principal components of the PCA. The lower triangular part of the similarity matrix gives the similarities when keeping the two pairs of redundant marginality indices. The upper triangular part gives the similarities when excluding redundant marginality indices (i.e. keeping the gravity index and the second-nearest-core index but dropping the area index and the isolation index). Shaded cells highlight similarity values with a p-value < 0.5.

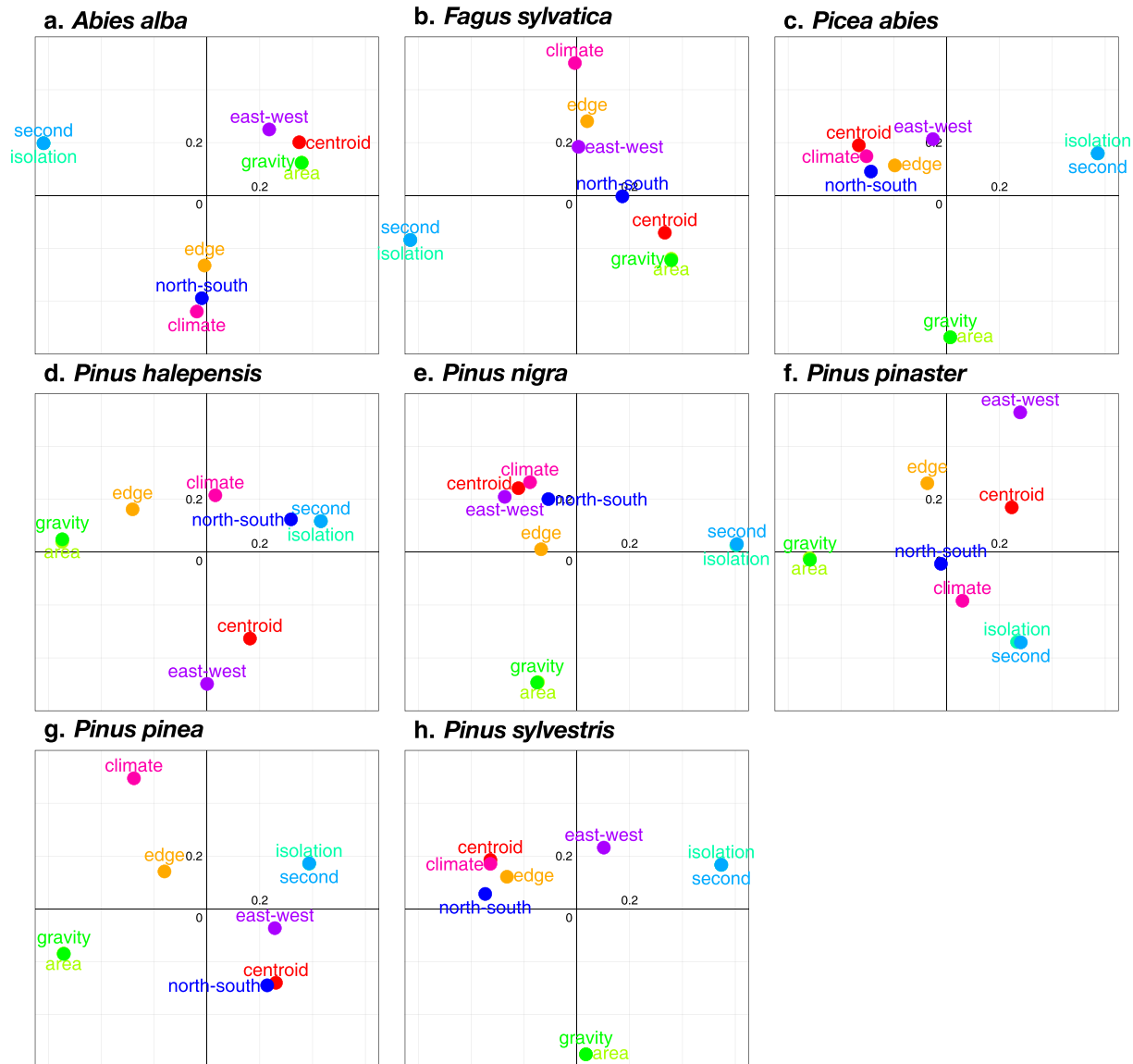|	A. alb.	F. syl.	P. abi.	P. hal.	P. nig.	P. pina.	P. pine.	P. syl.
A. alba	1	0.60	0.10	0.62	0.43	0.15	0.55	0.40
F. sylvatica	0.53	1	0.21	0.79	0.15	0.52	0.22	0.78
P. abies	0.04	0.10	1	0.02	0.14	0.01	0.03	0.51
P. halepensis	0.33	0.42	0.15	1	0.24	0.74	0.21	0.35
P. nigra	0.29	0.00	0.24	0.14	1	0.09	0.26	0.02
P. pinaster	0.05	0.18	0.32	0.81	0.04	1	0.04	0.16
P. pinea	0.15	0.23	0.33	0.64	0.16	0.61	1	0.22
P. sylvestris	0.03	0.09	**0.99**	0.18	0.23	0.37	0.36	1

Fig. 4. Plot of matrices $\mathbf{M}_s$ for each species. Each panel shows the multidimensional scaling map of the nine marginality indices based on the Jaccard coefficients of the marginal areas that these indices delimit. The two axes correspond to the two dimensions of the multidimensional scaling. Each dot represents a marginality index. The coordinates of dots are given by matrices $\mathbf{M}_s$.

status. The east/west index was a relevant predictor for the two species whose distribution extended farthest to the east (*Picea abies* and *Pinus sylvestris*), as well as for *Fagus sylvatica*. Otherwise, it was not a relevant predictor. Finally, the gravity index was a relevant predictor of

marginality for *Fagus sylvatica* and *Pinus pinaster* as well as, to a lesser extent, for *Pinus halepensis*.

For all species, the Maxent model provided a good quality of fit (AUC ranging between 0.90 and 0.97, with an average value across species of

Table 4

Matrix of similarity between the eight species, using the similarity index by Indahl et al. (2018) between matrices **M**ₛ. For each species *s*, **M**ₛ gives the coordinates of the marginality indices in the multidimensional scaling map based on the Jaccard coefficients. The lower triangular part of the similarity matrix gives the similarities when keeping the two pairs of redundant marginality indices. The upper triangular part gives the similarities when excluding redundant marginality indices (i.e. keeping the gravity index and the second-nearest-core index but dropping the area index and the isolation index). Shaded cells highlight similarity values with a p-value < 0.5.

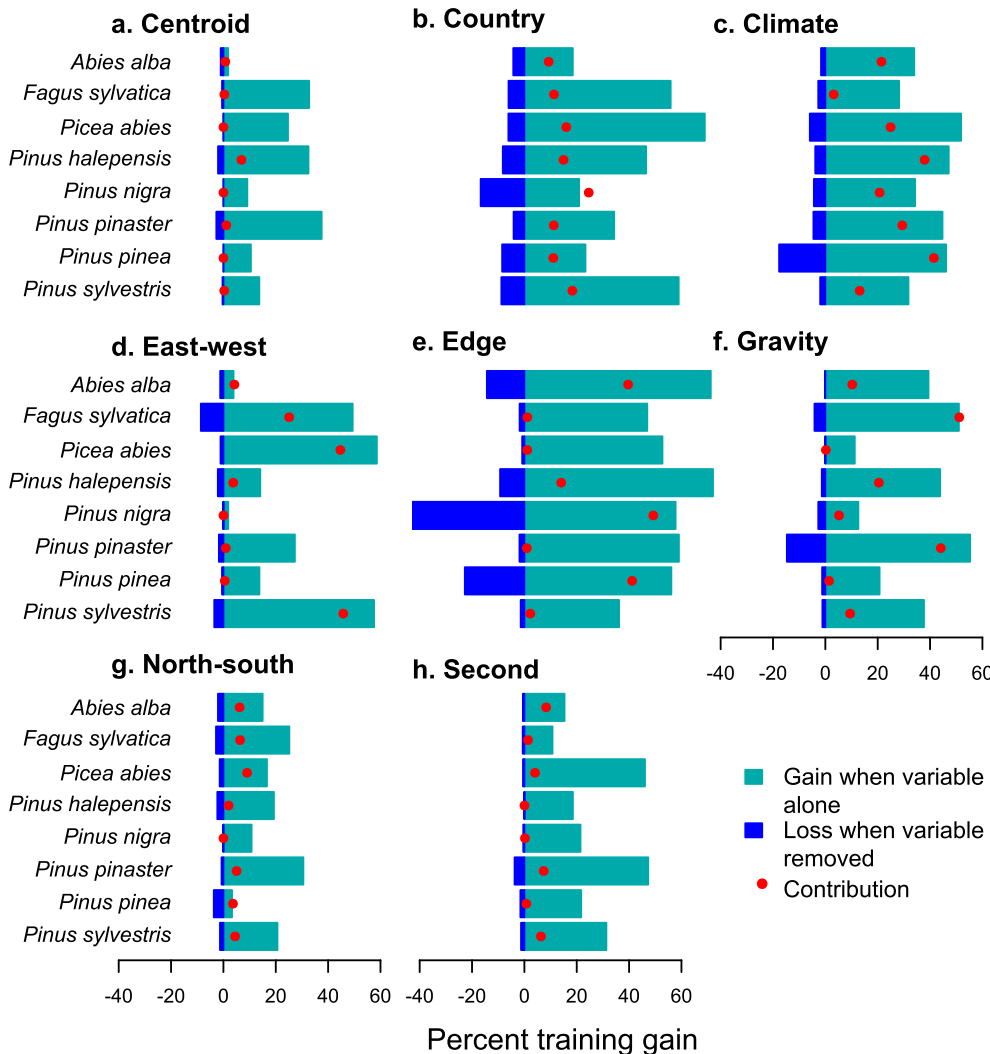|	A. alb.	F. syl.	P. abi.	P. hal.	P. nig.	P. pina.	P. pine.	P. syl.
A. alba	1	0.60	0.08	0.21	0.28	0.12	0.00	0.02
F. sylvatica	0.92	1	0.34	0.26	0.41	0.00	0.02	0.33
P. abies	0.55	0.66	1	0.00	0.85	0.12	0.00	**0.95**
P. halepensis	0.54	0.50	0.12	1	0.04	0.08	0.60	0.01
P. nigra	0.84	0.82	0.83	0.37	1	0.26	0.00	0.67
P. pinaster	0.31	0.41	0.05	0.75	0.16	1	0.30	0.03
P. pinea	0.35	0.40	0.12	0.84	0.27	0.83	1	0.02
P. sylvestris	0.43	0.58	**0.97**	0.09	0.70	0.06	0.12	1

Fig. 5. Relative importance of seven marginality indices (panels a, c–h) and country of location (panel b) to predict the marginality status of populations. Relative importance is measured as the percentage of training gain or loss provided by each predictor when fitting a Maxent model. Training gain (> 0) or loss (< 0) is shown on the x-axis. A separate model is fitted for each of the eight tree species shown on the y-axis. Greenish bars correspond to the percentage of training gain when the variable is the only predictor of the model. Blue bars correspond to the percentage of training loss when the variable is removed from the full model with all predictors. Red dots correspond to the percent contribution of the variable.

0.93). The maps of the predicted probability of being marginal are shown in Appendix S1. For instance, geographical areas where *Abies alba* populations had the highest probability to be marginal were the Pyrenees, Corsica, the southern part of the Apennines range, and the Carpathians. In contrast, populations located in the most central part of its distribution across Switzerland, Austria, Germany, Hungary and Poland were the less likely to be marginal.

For all species, there was a significant country effect in diagnosing the marginality status of a population. This effect was apparent in the map of the probabilities of being marginal with discontinuities at

country borders (e.g. Swiss or Austrian borders for *Abies alba*).

4. Discussion

We provided consistent measures of marginality and peripherality that are applicable to species with broad distribution range, using nine indices covering climatic, geographical and historical factors. Two indices were specifically designed to solve the issue of spatial discontinuity of two other indices. Apart from these two pairs of redundant indices, there was no consistent pattern of correlations between indices

across species. Hence, the indices conveyed different information related to the specific geometry of the species distributions. Contrasting with this heterogeneity of correlation patterns across species, the relative importance of the indices to predict the marginality status of populations was consistent across species. The edge index and the climatic index consistently had the greatest importance. Taken together, these two results concur to the idea that marginality is jointly defined by geographical marginality and ecological marginality that may or may not overlap (Abeli et al., 2014).

The evaluation of the different components of marginality and its outcomes are pivotal to avoid caveats in some studies trying to address the CPH or other related topics. It is of special importance in defining conservation policies, for instance within the EUFORGEN conservation programme. It will avoid biases already detected towards marginal populations (Araújo and Williams, 2001), especially if the classification is based on expert knowledge as tested in this work.

Even though the same indices consistently contributed to predicting the marginality status of populations across species, the marginality maps generated by the model substantially varied across species. For species with a large spatial distribution (e.g. *Abies alba*, *Fagus sylvatica*, *Picea abies*, *Pinus sylvestris*), areas with a high probability of being marginal clearly differed from the main bulk of the distribution (Appendix S1). In contrast, for species with a limited and fragmented spatial distribution (particularly *Pinus nigra* and *Pinus pinea*), areas with a high probability of being marginal largely coincided with the species distribution itself. In other words, basically all the populations of these latter species were considered as marginal. There are at least two different interpretations of this result: (1) the marginality indices are too much data-driven and lack robustness to capture the marginality status of populations; (2) the perception that experts have of the marginality of a population is influenced by the overall spatial pattern of the species distribution.

Concurring with the former idea, those indices that are highly sensitive to noise in data also have little contribution to the Maxent model predictions. Concurring with the latter idea, the strong country effect in the Maxent model means that experts from different countries have different perception of marginality. Bias in expert knowledge on ecosystems is a known caveat that may require the use of elicitation techniques (Perera et al., 2012). Our results support the idea that marginality, although a concept that is intuitively easy to grasp, still lacks metrics and a theoretical background (Gómez-Sanz, 2019). Metrics are needed to clarify the concept and make it a useful tool for quantitative ecology. The indices defined in this study may help experts in their decision making, so as to converge towards a consensual rationale on how to assess marginality.

More importantly, the relevance of the marginality indices proposed here will depend on their predictive performance when trying to explain biological processes that are believed to depend on them. Such processes include population abundance, reproductive performance, genetic diversity, population dynamics, tree morphology, or phenotypic plasticity (Soulé, 1973; Abeli et al., 2014). The difficulty here is to simultaneously develop a theory that explains the effects of marginality in plants and develop marginality indices that are suitable for this still unknown theory.

The marginality indices defined here could be extended in several directions. For instance, environmental marginality was defined using climatic variables only. Other environmental dimensions of the ecological niche (e.g. soil or topography) could be used, either jointly with climatic variables, or separately from them. In the latter case, edaphic or topographic marginality indices would be thus obtained (Gómez-Sanz, 2019). The distance to the glacial refugia of species is another example of a historical marginality index that could complement the north/south and east/west indices (Hampe and Petit, 2005; Abeli et al., 2018). As for peripherality, there are many other metrics of isolation, fragmentation or connectivity that could potentially be derived from the species distributions (Saura et al., 2011).

Contrary to the other indices that were data-driven, the climatic index was model-based. The modelling approach allowed us to integrate all climate variables into a single index. However, it raises the question of the performance of the model used, especially in a context of climate change (Santini et al., 2021). On the one hand, the importance of the climatic index for predicting the marginality status of populations confirmed the relevance of the modelling approach. On the other hand, future work may explore different climatic marginality indices based on different climate variables.

While the radiative exploration of additional marginality indices would be worth continuing, we believe that future progress in quantifying marginality will come first from confronting these indices to marginality-related theories. One principle proposed for the subsequent development of both marginality indices and theories is that marginality is a concept operating at the coarse spatial scale of the whole species distribution. *P. abies* and *P. sylvestris* exemplify this principle. These species have similar marginality index values due to the similarity of their broad scale distribution. Yet they locally differ in their probabilities of being marginal. This principle contrasts for instance with the micro-scale approach followed by Gómez-Sanz (2019). This author compared the marginality level of pairs of declining and non-declining *Pinus halepensis* forest stands at different locations in a subset of the species distribution. This difference between the coarse scale and the micro scale in shaping plant-environment relationships was also invoked by Midolo et al. (2021) to explain why the fitness-periphery hypothesis (a sub-hypothesis of CPH) was rarely observed.

In conclusion, this study defined idiosyncratic indices of environmental and geographic marginality that aim to quantify population marginality in a standardized way. The marginality status of a population as determined by expert knowledge can be to a large extent predicted by these indices. However, it also strongly depends on the experts' perception of marginality, evidencing that marginality still is a concept open to the subjective perception of experts. Converging towards useful marginality metrics for quantitative ecology and genetics will require a back-and-forth between metrics and theory. Metrics will need to be evaluated for marginality-related theories (e.g. the CPH). In turn, theories will need to be developed based on updated metrics. To disentangle the effects of the data dependency of indices from the possible bias in expert perception of marginality, future research may use elicitations techniques. Future research may also develop marginality indices that are more robust to noise in data.

Credit authorship contribution statement

MW, SC, EN, AP, PA, MB, SCGM, DG, FAA, GGV, FD, BF and RA conceptualized and designed the study. MW, SC, EN, AP, PA, MB, SCGM, DG, FAA, GGV, FD and RA provided data. NP, MM and MJSV analysed data. NP, MM, MJSV and RA drafted the article. MB, DG, GGV and BF revised it critically. All authors contributed critically to the drafts and gave final approval for publication.

Data accessibility

Maps and R code available from Zenodo https://doi.org/10.5281/zenodo.4679159 (Picard et al., 2021). Database on marginal populations available from Zenodo https://doi.org/10.5281/zenodo.4767651 (Fady et al., 2021).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecolind.2022.109367.

References

Abeli, T., Gentili, R., Mondoni, A., Orsenigo, S., Rossi, G., 2014. Effects of marginality on plant population performance. J. Biogeogr. 41, 239–249. https://doi.org/10.1111/jbi.12215.

Abeli, T., Vamosi, J.C., Orsenigo, S., 2018. The importance of marginal population hotspots of cold-adapted species for research on climate change and conservation. J. Biogeogr. 45, 977–985. https://doi.org/10.1111/jbi.13196.

Alberto, F.J., Aitken, S.N., Alía, R., González-Martínez, S.C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., Savolainen, O., 2013. Potential for evolutionary responses to climate change–evidence from tree populations. Glob. Change Biol. 19, 1645–1661. https://doi.org/10.1111/gcb.12181.

Araújo, M.B., Williams, P.H., 2001. The bias of complementarity hotspots toward marginal populations. Conserv. Biol. 15, 1710–1720. https://doi.org/10.1046/j.1523-1739.2001.99450.x.

Caudullo, G., Welk, E., San-Miguel-Ayanz, J., 2020. Chorological data for the main European woody species. Mendeley Data. 10.17632/hr5h2hcgg4.12.

Channell, R., Lomolino, M.V., 2000. Trajectories to extinction: spatial dynamics of the contraction of geographical ranges. J. Biogeogr. 27, 169–179. https://doi.org/10.1046/j.1365-2699.2000.00382.x.

Cox, T.F., Cox, M.A.A., 2001. Multidimensional scaling. Number 88 in Monographs on Statistics and Probability, 2nd ed.,. Chapman and Hall/CRC, Boca Raton, FL. 10.1201/9780367801700.

Ducci, F., Donnelly, K., 2017. Forest tree marginal populations in Europe: report on the state of knowledge on forest tree marginal and peripheral populations in Europe. Ann. Silvicultural Res. 41, 1–12. https://doi.org/10.12899/asr-1586.

Eckert, C.G., Samis, K.E., Lougheed, S.C., 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Mol. Ecol. 17, 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x.

Fady, B., Aravanopoulos, F.A., Alizoti, P., Mátyás, C., Von Wühlisch, G., Westergren, M., Belletti, P., Cvjetkovic, B., Ducci, F., Huber, G., Kelleher, C.T., Khaldi, A., Kharrat, Bou Dagher, 2016. M., Kraigher, H., Kramer, K., Mühlethaler, U., Peric, S., Perry, A., Rousi, M., Sbay, H., Stojnic, S., Tijardovic, M., Tsvetkov, I., Varela, M.C., Vendramin, G.G., Zlatanov, T., 2016. Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. Forest Ecol. Manage. 375, 66–75. https://doi.org/10.1016/j.foreco.2016.05.015.

Fady, B., Božič, G., Ducousso, A., Berthold, H., Eliades, N.G., Sbay, H., Ballian, D., Gömöry, D., Kelleher, C., Ladier, J., Danusevicius, D., Ducci, F., Bou Dagher Kharrat, M., Markiewicz, P., Malvolti, M.E., David-Schwartz, R., Westergren, M., Vendramin, G.G., Myking, T., Alizoti, E., Medak, J., Kraigher, H., Aravanopoulos, F.A., Máchová, P., Perić, S., Mattioni, C., Villani, F., 2021. A dataset of 577 marginal and peripheral forest tree populations in Europe. Zenodo. 10.5281/zenodo.4767651.

Fady, B., Esposito, E., Abulaila, K., Aleksic, J.M., Alia, R., Alizoti, P., Apostol, E.N., Aravanopoulos, P., Ballian, D., Bou Dagher Kharrat, M., Carrasquinho, I., Albassatneh, M.C., Curtu, A.L., David-Schwartz, R., de Dato, G., Douaihy, B., Eliades, N.G.H., Fresta, L., Gaouar, S.B.S., Illoul, M.H., Ivetic, V., Ivankovic, M., Kandemir, G., Khaldi, A., Khouja, M.L., Kraigher, H., Lefévre, F., Mahfoud, I., Marchi, M., Pérez Martín, F., Picard, N., Sabatti, M., Sbay, H., Scotti-Saintagne, C., Stevens, D.T., Vendramin, G.G., Vinceti, B., Westergren, M., 2022. Forest genetics research in the Mediterranean Basin: bibliometric analysis, knowledge gaps, and perspectives. Current Forestry Reports, in press. 10.1007/s40725-022-00169-8.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–39. https://doi.org/10.1017/S0376892997000088.

Frankham, R., 1996. Relationship of genetic variation to population size in wildlife. Conserv. Biol. 10, 1500–1508. https://doi.org/10.1046/j.1523-1739.1996.10061500.x.

Franklin, I.R., 1980. Evolutionary change in small populations. In: Soulé, M.E., Wilcox, B. A. (Eds.), Conservation biology – An evolutionary-ecological perspective. Sinauer Associates, Sunderland, MA, pp. 135–149 http://hdl.handle.net/102.100.100/293218?index=1.

Gómez-Sanz, V., 2019. Site-scale ecological marginality: Evaluation model and application to a case study. Ecol. Model. 408, 108739 https://doi.org/10.1016/j.ecolmodel.2019.108739.

Hampe, A., Petit, R.J., 2005. Conserving biodiversity under climate change: the rear edge matters. Ecol. Lett. 8, 461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x.

Hardie, D.C., Hutchings, J.A., 2010. Evolutionary ecology at the extremes of species' ranges. Environ. Rev. 18, 1–20. https://doi.org/10.1139/A09-014.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. https://doi.org/10.1002/joc.1276.

Indahl, U.G., Næs, T., Liland, K.H., 2018. A similarity index for comparing coupled matrices. J. Chemom. 32, e3049 https://doi.org/10.1002/cem.3049.

Kremer, A., Ronce, O., Robledo-Arnuncio, J.J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J.R., Gomulkiewicz, R., Klein, E.K., Ritland, K., Kuparinen, A., Gerber, S., Schueler, S., 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology letters 15, 378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x.

Lesica, P., Allendorf, F.W., 1992. Are small populations of plants worth preserving? Conserv. Biol. 6, 135–139. https://doi.org/10.1046/j.1523-1739.1992.610135.x.

Midolo, G., Wellstein, C., Faurby, S., 2021. Individual fitness is decoupled from coarse-scale probability of occurrence in North American trees. Ecography 44, 789–801. https://doi.org/10.1111/ecog.05446.

Munguía-Rosas, M.A., Montiel, S., 2014. Patch size and isolation predict plant species density in a naturally fragmented forest. PLoS ONE 9, e111742. https://doi.org/10.1371/journal.pone.0111742.

Nosil, P., Egan, S.P., Funk, D.J., 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation and multiple roles for divergent selection. Evol.: Int. J. Organic Evol. 62, 316–336. https://doi.org/10.1111/j.1558-5646.2007.00299.x.

Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., Chirici, G., 2019. Species distribution modelling to support forest management. A literature review. Ecolog. Modell. 411, 108817 https://doi.org/10.1016/j.ecolmodel.2019.108817.

Perera, A.H., Drew, C.A., Johnson, C.J., 2012. Expert knowledge and its application in landscape ecology. Springer, New York, NY, 10.1007/978-1-4614-1034-8.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Picard, N., Marchi, M., Serra-Varela, M.J., Westergren, M., Cavers, S., Notivol Paino, E., Piotti, A., Alizoti, P., Bozzano, M., González-Martínez, S., Grivet, D., Aravanopoulos, F.A., Vendramin, G.G., Ducci, F., Fady, B., Alía, R., 2021. Maps and R code from: Marginality indices for biodiversity conservation in forest trees. Zenodo. https://doi.org/10.5281/zenodo.4679159.

Pironon, S., Papuga, G., Villellas, J., Angert, A.L., García, M.B., Thompson, J.D., 2017. Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. Biol. Rev. 92, 1877–1909. https://doi.org/10.1111/brv.12313.

Rehm, E.M., Olivas, P., Stroud, J., Feeley, K.J., 2015. Losing your edge: climate change and the conservation value of range-edge populations. Ecology and Evolution 5, 4315–4326. https://doi.org/10.1002/ece3.1645.

Sagarin, R.D., Gaines, S.D., 2002. The 'abundant centre' distribution: to what extent is it a biogeographical rule? Ecol. Lett. 5, 137–147. https://doi.org/10.1046/j.1461-0248.2002.00297.x.

Santini, L., Benítez-López, A., Maiorano, L., Cengic, M., Huijbregts, M.A.J., 2021. Assessing the reliability of species distribution projections in climate change research. Divers. Distrib. 27, 1035–1050. https://doi.org/10.1111/ddi.13252.

Saura, S., Estreguil, C., Mouton, C., Rodríguez-Freire, M., 2011. Network analysis to assess landscape connectivity trends: application to European forests (1990–2000). Ecol. Ind. 11, 407–416. https://doi.org/10.1016/j.ecolind.2010.06.011.

Schwartz, M.K., Mills, L.S., Ortega, Y., Ruggiero, L.F., Allendorf, F.W., 2003. Landscape location affects genetic variation of Canada lynx (Lynx canadensis). Mol. Ecol. 12, 1807–1816. https://doi.org/10.1046/j.1365-294X.2003.01878.x.

Sexton, J.P., Hangartner, S.B., Hoffmann, A.A., 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68, 1–15. https://doi.org/10.1111/evo.12258.

Soille, P., Vogt, P., 2009. Morphological segmentation of binary patterns. Pattern Recogn. Letters 30, 456–459. https://doi.org/10.1016/j.patrec.2008.10.015.

Soulé, M.E., 1973. The epistasis cycle: a theory of marginal populations. Annu. Rev. Ecol. Syst. 4, 165–187. https://doi.org/10.1146/annurev.es.04.110173.001121.

Tessarolo, G., Lobo, J.M., Rangel, T.F., Hortal, J., 2021. High uncertainty in the effects of data characteristics on the performance of species distribution models. Ecol. Ind. 121, 107147 https://doi.org/10.1016/j.ecolind.2020.107147.

Tsuda, Y., Chen, J., Stocks, M., Källman, T., Sønstebø, J.H., Parducci, L., Semerikov, V., Sperisen, C., Politov, D., Ronkainen, T., Väliranta, M., Vendramin, G.G., Tollefsrud, M.M., Lascoux, M., 2016. The extent and meaning of hybridization and introgression between Siberian spruce (Picea obovata) and Norway spruce (Picea

abies): cryptic refugia as stepping stones to the west? Mol. Ecol. 25, 2773–2789. https://doi.org/10.1111/mec.13654.

de Vries, S.M.G., Alan, M., Bozzano, M., Burianek, V., Collin, E., Cottrell, J., Ivankovic, M., Kelleher, C.T., Koskela, J., Rotach, P., Vietto, L., Yrjänä, L., 2015. Pan-European strategy for genetic conservation of forest trees and establishment of a core network of dynamic conservation units. In: European Forest Genetic Resources Programme (EUFORGEN). Bioversity International, Rome, Italy.

Wright, S., 1943. Isolation by distance. Genetics 28, 114–138 https://www.ncbi.nlm.nih. gov/pmc/articles/PMC1209196/.