



Predicting current and future suitable habitat and productivity for Atlantic populations of maritime pine (*Pinus pinaster* Aiton) in Spain

Marcos Barrio-Anta¹ · Fernando Castedo-Dorado² · Asunción Cámara-Obregón¹ · Carlos A. López-Sánchez¹

Received: 7 November 2019 / Accepted: 12 March 2020 / Published online: 17 April 2020
© INRAE and Springer-Verlag France SAS, part of Springer Nature 2020

Abstract

• **Key message** By combining inventory data and spatially-continuous environmental information, we were able to develop models for Atlantic populations of maritime pine (*Pinus pinaster* Aiton) in Spain in order to predict suitable habitat and site index at a spatial resolution of 250 × 250 m.

• **Context** Currently available, spatially continuous environmental information was used to make reliable predictions about suitable habitat and forest productivity.

• **Aims** To develop raster-based distribution and productivity models for Atlantic populations of maritime pine in Spain to predict current and future suitable habitat and productivity.

• **Methods** Occurrence data and site index values were obtained from the Third Spanish National Forest Inventory and research plots, respectively. After testing different algorithms, random forest were selected for modelling the relationships between maritime pine occurrence, site index and spatially continuous environmental variables.

• **Results** The overall accuracy of the suitable habitat model was 73%, and climate (mainly thermal properties) and soil physical properties were the most important variables. The site index model explained 60% of the observed variability, and lithological properties were the most important variables. A slight increase in site index (0.46–0.51%) and a large increase in suitable habitat (50–66%) are expected for 2070 under the most pessimistic climate change scenario.

• **Conclusion** The currently available spatial continuous information enables the development of accurate raster data models for predicting suitable habitat and site productivity without the need for fieldwork. Climate change is expected to increase the potentially suitable habitat of Atlantic maritime pine populations in Spain in the coming decades.

Keywords *Pinus pinaster* Aiton · Site index · Species distribution model · Environmental variables · Random forest · Climate change

Handling Editor: Andreas Bolte

Contributions of the co-authors MB-A and CAL-S: designed the research, analysed and interpreted the data and wrote the manuscript. FC-D and AC-O: designed the experimental plot, collected data and reviewed drafts of the manuscript.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13595-020-00941-5>) contains supplementary material, which is available to authorized users.

✉ Carlos A. López-Sánchez
lopezscarlos@uniovi.es

¹ Department of Biology of Organisms and Systems, Mieres Polytechnic School, University of Oviedo, Campus Universitario de Mieres, C/Gonzalo Gutiérrez Quirós S/N, 33600 Mieres, Spain

² Department of Engineering and Agricultural Sciences, School of Agricultural and Forestry Engineering, University of León, Campus de Ponferrada, Avda. de Astorga S/N, 24400 Ponferrada, Spain

1 Introduction

Maritime pine occupies an environmentally diverse area in the Western Mediterranean basin and displays a remarkable degree of genetic variation, with multiple subspecies recognized. It is therefore considered an ecologically versatile species (Abad Viñas et al. 2016). In the Iberian Peninsula, the species have traditionally been classified into two groups (Atlantic maritime pine and Mediterranean maritime pine), although there is no consensus about its taxonomic status (Alía et al. 1996). These populations differ greatly in growth performance: Atlantic maritime pine is more productive and mainly found in Atlantic climate areas, whereas Mediterranean maritime pine is less productive and found in Mediterranean climate areas, characterised by lower, irregular precipitation and

higher temperatures (Alía et al. 1997; Bravo-Oviedo et al. 2011). Atlantic maritime pine populations expanded throughout Galicia in the seventeenth century (probably from Portugal) (Ruíz Zorrilla 1980) and are now widely distributed in NW Spain (in the Autonomous Communities of Galicia and Asturias and the province of León). The species occupies an area of 433,754 ha, which represents 18.2% of the total forest area (MAPA 2019), and it is the most important coniferous tree species in NW Spain in terms of both surface cover and wood production. The average volume harvested in the period 2005–2015 reached 2,649,249 m³ year⁻¹, which represents 31.9% of the total volume harvested annually in the region and 19.7% of the same in Spain (MAPA 2019). The species grows in even-aged stands, derived from plantations or natural regeneration after clear cutting or wildfire, and it has traditionally been managed for sawn timber or pulp production. The rotation age is usually between 30 and 40 years, although older stands are common as a result of a lack of adequate management. The area occupied by these maritime pine populations has increased in recent decades because the species has been widely used in afforestation programs (Álvarez-Álvarez et al. 2011).

Because of the economic importance of maritime pine, forest managers and practitioners must be able to identify the best quality sites for afforestation. It is widely accepted that the growth and yield of even-aged stands are largely determined by the productive capacity of the growing site, including many variables that collectively determine the site quality (e.g. Clutter et al. 1983; Gadov and Bredenkamp 1992; Vanclay 1994). Evaluation of site quality is desirable for volume production. However, it is seldom feasible to use direct measures of volume productivity because the volume attained by a stand at any given age may be strongly affected by factors other than site quality (e.g. stand density, cultural practices, pests and diseases), and historical records of yields from forested sites are limited (Clutter et al. 1983; Burkhart and Tomé 2012). Early empirical evidence showed that, for most species, good quality sites correspond to areas where height growth rates are high (Clutter et al. 1983; Skovsgaard and Vanclay 2008), and therefore height growth of dominant trees has commonly been used as an indirect measure of site quality (Clutter et al. 1983). Moreover, height growth is closely correlated with volume and is independent of stocking over quite a wide range of stand densities (Burkhart and Tomé 2012). In this regard, site index, defined as the average height of the dominant portion of the stand at a reference stand age, is one of the most common indirect measures of forest site productivity used worldwide for even-aged stands (Weiskittel et al. 2011a). Although site index enables determination of maximum of the mean annual increment (MAImax) in volume and rotation age (Barrio Anta and Diéguez-Aranda 2005), conversion of the site index measure for direct estimation of these variables requires the use of growth and yield

models. The maximum mean annual volume increment per hectare is considered the most appropriate direct measure of productivity because it is directly related to the wood volume that can be obtained in a given site (Skovsgaard and Vanclay 2008; Latta et al. 2009).

Site index can be estimated with reasonable accuracy at local scale, although this is usually expensive (stand dominant height and stand age must be determined) and requires the species to be present. These shortcomings can be resolved by constructing methods of estimating this index based only on environmental variables. Many studies have attempted to relate site index to environmental factors by using parametric approaches (e.g. Fontes et al. 2003; Romanyà and Vallejo 2004), non-parametric approaches (e.g. McKenney and Pedlar 2003; Albert and Schmidt 2010) or both (e.g. Wang et al. 2005; Aertsens et al. 2010; Álvarez-Álvarez et al. 2011). However, many of these published studies included physical or chemical soil-related variables that can only be determined by using expensive analytical techniques, and the models developed are therefore of little practical use. The current availability of remote sensing data and spatially interpolated gridded climate and soil data enable us to obtain spatially continuous environmental information, thus allowing site index to be estimated at landscape scale (Bontemps and Bouriaud 2014) without the need for fieldwork. Moreover, this spatially explicit information can be used in further research (Parresol et al. 2017). These types of raster models should be used to integrate the Spanish National Forest Inventory data and the first Nationwide Airborne Laser Scanning data to carry out yield predictions, as recently noted by Novo-Fernández et al. (2019).

At this point, the remaining challenge is to establish an adequate framework for depicting or mapping site index or site productivity for particular species throughout an area. Suitable habitat provides an appropriate reference framework for depicting relevant ecological information about a species. Suitable habitat can be defined as the area within a region where a species actually or potentially occurs. This area is determined by the complex relationships between numerous environmental variables, for which ecological researchers have commonly used empirical models relating species occurrence to environmental predictors. These models are known as Species Distribution Models (Guisan and Zimmermann 2000).

Climate-related variables are often key factors driving species distribution and productivity, and shifts in these variables are therefore expected to strongly affect species distribution, abundance and growth. Accelerated climate change is already affecting Southern Europe, and more dramatic implications are expected in the future with a predicted increase in mean annual temperature of 4–6 °C and a decrease in mean annual precipitation of 15–20% for 2080 in the worst possible scenario (EEA 2017). This change will contribute to restricting or expanding the geographic distribution of particular species

(e.g. Monzón et al. 2011; Bellard et al. 2012), and the impact will depend on the adaptability of different species to climate change (Venetier et al. 2007). Forests are particularly sensitive to this type of change, because the long lifespan of trees does not allow for rapid adaptation to environmental changes (Davis et al. 2005).

Species distribution models that include climatic variables can be used to predict where the species might move to, or be able to persist under suitable conditions, given a particular climate change scenario. In the same way, productivity models that include climate variables can predict future shifts in site productivity. Projecting species distribution and productivity models for climate change scenarios can thus provide information that will help decision-makers develop and implement appropriate afforestation programmes.

The main aim of this study was to develop raster-based distribution and productivity models for Atlantic populations of maritime pine in Spain in order to predict the current and future suitable habitat and productivity. For this purpose, the following secondary goals were identified as follows: (i) to investigate the environmental factors determining the current suitable habitat and site productivity, (ii) to develop species distribution and productivity models based on current environmental variables, (iii) to generate spatially continuous maps of these environmental characteristics, and (iv) to make future projections of the models and maps based on different climate change scenarios.

2 Materials and methods

2.1 Study area

The study area comprises region of provenance 1 (north-west) for maritime pine in Spain (Alía et al. 2009), including the Autonomous Community of Galicia and parts of Asturias and the province of León (Fig. 1). In this region of provenance, the climate is characterized as Atlantic, and precipitation is fairly uniformly distributed throughout the year. In coastal areas, precipitation reaches an annual average of 1342 mm, with an average summer precipitation of 132 mm. Summer drought, when it occurs, lasts for less than 1 month. The thermal regime is characterized by mild temperatures with an annual average of 12.9 °C and a low average daily thermal oscillation (10.8 °C). In inland areas, precipitation is lower (1031 mm of annual average and 100 mm of summer period) and daily thermal oscillation (14.5 °C) is higher. Summer drought lasts between 1 and 2 months in these areas (Alía et al. 2009). From a geological point of view, granitic rocks and sediments from the lower Palaeozoic, with lithofacies of sandstones, quartzites and slates, are predominant. The soils have a loamy to sandy loam texture, a low retention capacity and are strongly acidic (Alía et al. 1996).

2.2 Data collection

Four different types of data were considered in this study: (i) site index data obtained from research plots, (ii) maritime pine occurrence data obtained from the Third Spanish National Forest Inventory (SNFI3) and from research plots established for this purpose, (iii) data on current environmental variables, and (iv) future climatic data projections for different climate change scenarios.

2.2.1 Site index data

Site index data were obtained from a network of 85 research plots established by our research team in pure, even-aged stands of maritime pine. The plots were established throughout the current distribution area of the populations of Atlantic maritime pine in north-western Spain, with the aim of covering a wide range of ages, stand densities and site qualities. The plot size ranged from 700 to 1200 m², in order to include a minimum of 30 trees. In each plot, diameter at breast height and total height of all trees were measured. Stand age was obtained from the plantation date (if available) or as the average age of six representative selected trees. Dominant height was determined as the average height of the 100 largest-diameter trees per hectare. Site index (dominant height at a reference age of 20 years) was estimated from the algebraic difference form equations developed for Galicia (Álvarez-González et al. 2005) and Asturias (Álvarez-Álvarez et al. 2011). The summary statistics of the main stand variables of the research plots used for modelling site index from environmental variables are shown in Table 1.

2.2.2 Occurrence data

Information on maritime pine occurrence was drawn from the Third Spanish National Forest Inventory (SNFI3) (DGCN 2006). SNFI plots are placed at the nodes of a 1 km UTM square grid, and a total of 7293 SNFI3 plots were located within the study area. These plots with Atlantic populations of maritime pine presence/absence data were imported to a GIS database (ArcGIS 9.3, ESRI, Redlands, CA, USA), with presence defined as the occurrence of at least one live maritime pine tree within each SNFI3 plot.

2.2.3 Data on current environmental variables

Environmental variables related to climate, topography and soil characteristics were included as potential predictors in the species distribution and productivity models. In total, a set of 50 spatial environmental variables were available for analysis (Table 2). Topographic variables were based on a 30 m resolution digital elevation model provided by the Spanish National Plan for Aerial Orthophotography (PNOA;

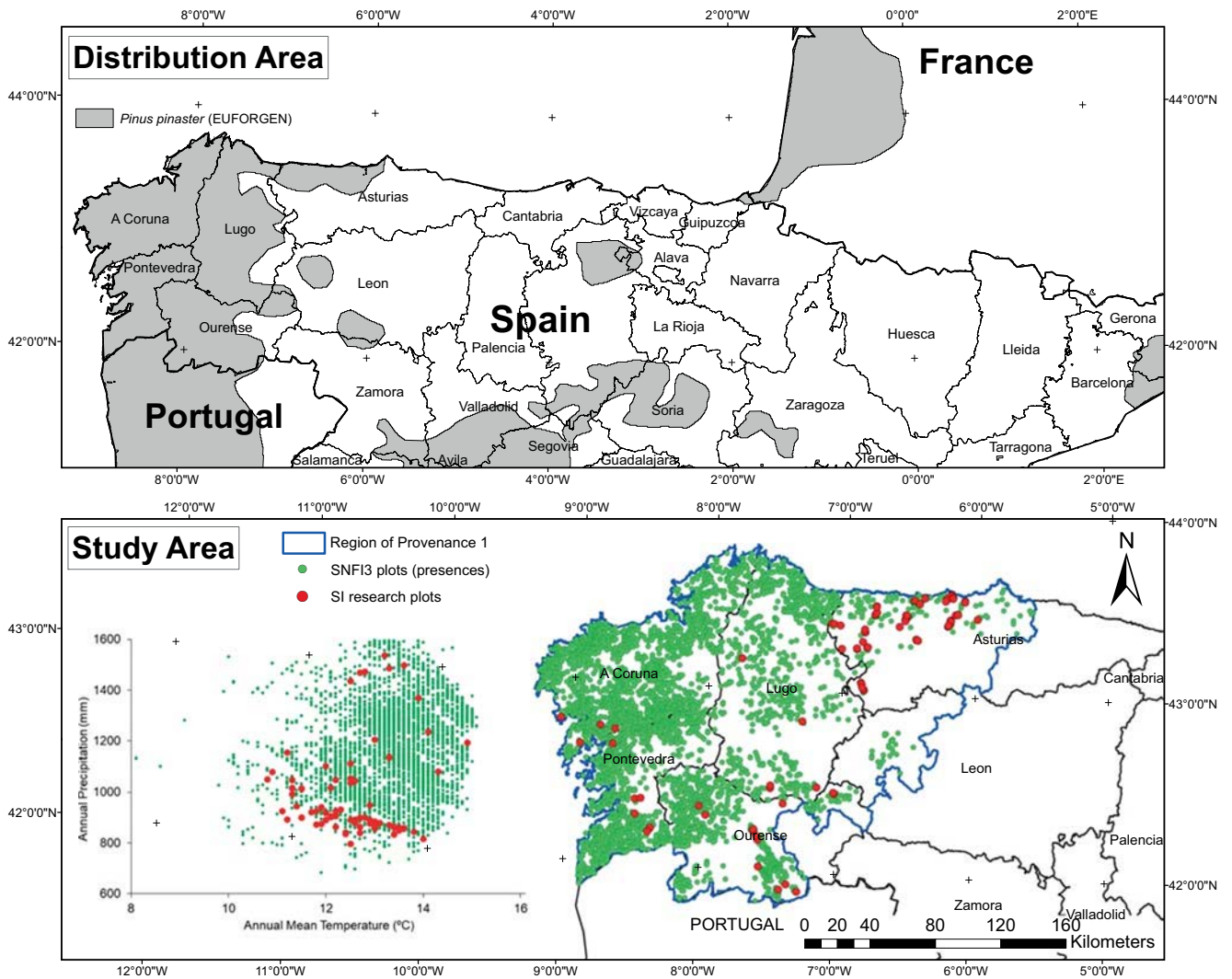


Fig. 1 Location of the study area (region of provenance 1) within the maritime pine distribution area in Northern Iberian Peninsula and Southern France. Plots used in the Third Spanish National Forest Inventory (SNFI3) where the presence of maritime pine was detected

(green dots) and site index (SI) research plots used in the present study (red dots). Below (left) scatter plot of showing the plots used in this study superimposed on the bivariate plot of annual mean temperature and total annual precipitation

<http://www.pnoa.ign.es>). The System for Automated Geoscientific Analyses (SAGA; Conrad et al. 2015) and Geographical Information System (GIS) software (version 3.0.0) were used to obtain eight topographic variables. Elevation was ruled out as an independent variable as it is

strongly correlated with climatic variables such as mean temperature and annual precipitation.

Gridded data for 19 climate variables with a 30-arc sec resolution (approximate 800 m) were obtained from WorldClim (Hijmans et al. 2005) by means of interpolations

Table 1 Summary statistics for stand variables in the 85 research plots used to develop the productivity model

	Number of trees (ha ⁻¹)	Basal area (m ² ha ⁻¹)	Quadratic mean diameter (cm)	Dominant height (m)	Age (years)	¹ Site index (m)
Minimum	377.78	12.77	10.32	7.25	12.00	7.80
Maximum	2580.00	76.20	41.50	26.96	61.50	19.66
Mean	1183.10	37.46	21.34	14.82	25.85	12.80
Standard deviation	485.65	14.79	7.61	4.94	13.14	2.44

¹ Dominant height at a reference age of 20 years

Table 2 Environmental variables considered potential predictors for the distribution and productivity models

Type	Abbreviation	Description	Unit	Source	
Terrain	SLP	Slope based on a digital elevation model	%	PNOA Lidar	
	ASP	Aspect based on a digital elevation model	°		
	CU	Curvature			
	PLC	Plan curvature			
	PRC	Profile curvature			
	TSI	Terrain shape index			
Climate	WI	Wetness index		WorldClim	
	DHN	Euclidean distance to nearest hydrographic network (m)	m		
	BIO1	Annual mean temperature	°C		
	BIO2	Mean diurnal range (Mean of monthly (max temp - min temp))	°C		
	BIO3	Isothermality (BIO2/ BIO7) (*100)	%		
	BIO4	Temperature seasonality (standard deviation *100)	%		
	BIO5	Max temperature of warmest month	°C		
	BIO6	Min temperature of coldest month	°C		
	BIO7	Temperature annual range (BIO5- BIO6)	°C		
	BIO8	Mean temperature of wettest quarter	°C		
	BIO9	Mean temperature of driest quarter	°C		
	BIO10	Mean temperature of warmest quarter	°C		
	BIO11	Mean temperature of coldest quarter	°C		
	BIO12	Annual precipitation	mm		
	BIO13	Precipitation of wettest month	mm		
	BIO14	Precipitation of driest month	mm		
	BIO15	Precipitation seasonality (coef. of variation)	%		
	BIO16	Precipitation of wettest quarter	mm		
	BIO17	Precipitation of driest quarter	mm		
	BIO18	Precipitation of warmest quarter	mm		
BIO19	Precipitation of coldest quarter	mm			
Soil	Sr_ss	Solar radiation in summer solstice	$\text{kJ m}^2 \text{ year}^{-1}$	PNOA Lidar	
	Sr_eq	Solar radiation in equinox	$\text{kJ m}^2 \text{ year}^{-1}$		
	Sr_ws	Solar radiation in winter solstice	$\text{kJ m}^2 \text{ year}^{-1}$		
	SC	Soil organic carbon content	mG/ha	SoilGrids250	
	pH_H2O	Soil pH in H ₂ O solution			
	pH_KCl	Soil pH in KCl solution			
	BD	Bulk density of fine earth fraction (< 2 mm)	kg m^{-3}		
	CLAY	Percentage of clay in soil	Weight %		
	SAND	Percentage of sand in soil	Weight %		
	SILT	Percentage of silt in soil	Weight %		
	CF	Coarse fragments	Volumetric %		
	CEC	Cation-exchange capacity	$\text{cmol} + \text{kg}^{-1}$		
	DB	Absolute depth to bedrock	cm		
	DB200	Depth to bedrock (R horizon) up to 200 cm	cm		
	R	Probability of occurrence of R horizon within 200 cm	%		
	SC_fef	Soil organic carbon content (fine earth fraction)	g		
	Geo_units	Geochronological units			SGM
	Geo_lit_units	Lithological units			
LIT_dco	Hydrogeology		SSM		
LIT_dlo	Lithostratigraphy				
LIT_per	Lithostratigraphy permeability				
WRB-FULL	Full soil code of the Soil typological units from the World Reference Base (WRB) for soil resources		ESDB		
WRB-LEV1	Soil reference group of the soil typological units from the World Reference Base (WRB) for Soil Resources				

SGM Spanish geological map (1:1000,000), SSM Spanish stratigraphic map (1:200,000), ESDB European soil database v2.0 (1:1000,000)

methods using the ANUSPLIN package (Booth et al. 2014). In addition, 12 soil variables were compiled from the SoilGrids250m (Hengl et al. 2017), which provides a collection of updatable global predictions for soil properties at 250 m spatial resolution based on machine learning algorithms. Soil type and group were compiled from the European soil database

v2.0 scale 1:1000,000. Lithostratigraphic type and permeability were obtained from the Spanish Stratigraphic Map scale 1:200,000, and Geology from the Spanish Geological Map scale 1:1000,000 (IGME 2015a, b). All climate, soil and topography variable raster grids were resampled at 250 m resolution.

2.2.4 Climate data projections

Climate data projections are required for predicting future suitable habitat and site productivity under different climate change scenarios. We used the global climate model for the years 2050 and 2070 based on the CMIP5 model of the 5th assessment report of the Intergovernmental Panel on Climate Change (<http://www.worldclim.org/CMIP5>). Bioclimatic predictions for two opposing scenarios of representative concentration pathways (RCP) were considered. The first “moderate scenario” (RCP 4.5) assumes a CO₂ concentration of 650 ppm and a 1.0–2.6 °C increase by 2100, whereas the second “pessimistic scenario” (RCP 8.5) considers a CO₂ concentration of 1350 ppm and a 2.6–4.8 °C increase by 2100 (van Vuuren et al. 2011; IPCC 2013; Harris et al. 2014; Dyderski et al. 2017).

2.3 Modelling species distribution and site productivity from environmental variables

2.3.1 Modelling techniques

To date, parametric regression techniques such as multiple linear regression and generalized linear modelling have been widely used to explore critical factors related to species occurrence (e.g. Roces-Díaz et al. 2015; Shirk et al. 2018) and forest productivity (e.g. Sánchez-Rodríguez et al. 2002; Aertsens et al. 2010; Sharma et al. 2012). However, traditional parametric regression techniques may not be adequate for analysis involving a potentially large number of predictors with complex interactions (Guisan and Zimmermann 2000; Prasad et al. 2006; Jiang et al. 2015). Advances in computer-assisted statistical analysis techniques simplify the implementation of non-parametric statistical models (Weiskittel et al. 2011b; Sabatia and Burkhart 2014; Jiang et al. 2015; Huang et al. 2017), generalized additive models (Aertsens et al. 2010; Brandl et al. 2014; Bjelanovic et al. 2018) and artificial neural networks (Drummond et al. 2003; Williams et al. 2009; Cosenza et al. 2015). We first carried out a preliminary analysis testing several algorithms implemented in the freely available BIOMOD2 R package (Thuiller et al. 2016) for fitting the species distribution model, and we used WEKA open source software (Hall et al. 2009) for fitting the site index model. Once the best performing algorithms were selected, they were fitted by WEKA for further assessment, evaluation and implementation. To select the potentially most important regressor variables, a wrapper method was used to select the subsample of variables as this usually produces the best results (Zhiwei and Xinghua 2010). The method selects the subsample of variables by using a learning algorithm as part of the evaluation function.

2.3.2 Model assessment and evaluation

The *k*-fold cross-validation approach was used to test the accuracy of the algorithms. In this process, the data set is divided into *k* subsets and each time the model is applied, one of the subsets is used as the test set and the other *k*-1 subsets form the training data set. This provides a good indication of how well the classifier will perform on unseen data. We used *k* = 10 and applied the algorithm several times and computed various standard metrics to assess model performance. In order to assess the accuracy of the species distribution model predictions, we used a confusion matrix that reflects the four possible ways that a sample point can be classified. Values of this matrix were used to calculate several commonly used metrics (Shirk et al. 2018): (i) the overall accuracy, (ii) sensitivity, (iii) specificity, (iv) the true skill statistic, (v) Cohen’s kappa; and (vi) the area under the ROC curve. The algorithm reports a probability of presence (PoP) of maritime pine as an output variable. A binary model is needed in order to calculate Cohen’s kappa and overall accuracy, and a threshold PoP was therefore selected to convert PoP to binary presence–absence outputs. To select this threshold (PoP_{threshold}), we used the average value of the PoP that maximized the sum of sensitivity and specificity and the PoP that minimized the difference between the absolute values of sensitivity and specificity.

To evaluate site index predictions by the fitted algorithm, the pseudo-coefficient of determination (R^2) (Ryan 1997), the mean absolute error (MAE) and the root mean squared error (RMSE) were used. For implementation of machine learning algorithms, WEKA has an embedded feature ranking technique called the variable importance measure (VIM), which we used as a tool to guide selection of predictors for inclusion in the final model. To ensure that values of variable importance were expressed on comparable scales, the VIM values were normalized so that they added up to a unitary value (normalized importance), and they were also expressed in relative terms: relative importance = (VIM – VIM_{min}) / (VIM_{max} – VIM_{min}). We then constructed the marginal response curves in order to explore the relationships between the response and each of important predictor variables. These curves represent the predicted PoP of the species or the site index prediction value (*y*-axis) as a function of a single environmental variable (*x*-axis), when all other explanatory variables were held constant at their mean values.

2.3.3 Making predictions of current and future suitable habitat and site productivity

The models finally selected were applied to the current environmental spatial variables (resampled to a 250 m × 250 m resolution) to generate spatially continuous maps. As previously pointed out, the maximum mean annual volume

increment per hectare is a practical direct measure of productivity. To convert site index to this direct measure of productivity in the absence of field data, the plantation density and the density regulation regime must be established. For the former, we considered plantation densities of 900 to 1400 stems ha^{-1} , as 1100 stems ha^{-1} is a common value in the area (Rodríguez Soalleiro and Madrigal Collazo, 2008), and for the latter we consider that only the competition-related mortality and random events (and not thinning management) will drive the density evolution of the stand. Under these assumptions, we carried out the following steps to determine MAI_{max} : (i) we generated a database of site index values from the minimum and maximum observed values in the plot data at 0.5 m intervals; (ii) we applied dynamic growth and yield models (Arias-Rodil et al. 2016; Diéguez-Aranda et al. 2009) to this database considering initial plantation densities of between 900 and 1400 stems ha^{-1} at 50 stems ha^{-1} intervals; (iii) we determined the MAI_{max} and the associated rotation age for each simulation; and (iv) we developed a parametric model for estimating MAI_{max} as a function of site index and plantation density. We also fitted an additional model for estimating the optimal rotation age as a function of MAI_{max} .

In addition, we also projected species distribution and productivity models onto spatial projections at 250 m resolution of the environmental variables reflecting the two climate change scenarios, i.e. moderate (RCP 4.5) and pessimistic (RCP 8.5).

For both the current and future scenarios, we used FRAGSTATS 4.2 (McGarigal et al. 2016) to quantify the area of habitat and degree of habitat fragmentation based on the binary model. We use three indicators to quantify suitable habitat surface: (i) total area, (ii) mean patch area and (iii) largest patch index (the percentage of the landscape encompassed by the largest patch). The fragmentation was assessed with the aggregation index, which equals 0 when suitable habitat is maximally disaggregated into single grid cell patches disconnected from all other patches and increases to 1 as suitable habitat is increasingly aggregated into a single, compact patch. We also quantified the degree of change for each future scenario relative to the current situation, classifying habitat as either gained, maintained or lost. The main methodological steps are graphically summarized in Fig. 2.

3 Results

3.1 Modelling technique selected

As a result of the preliminary analysis of different modelling techniques (Table 3), the non-parametric random forest ensemble learning method was selected as the best option for fitting species distribution and productivity models. Random forest is a widely used non-parametric classification and regression approach that consists of constructing hundreds of

decision trees from randomized subsets of predicted and predictor variables (Breiman 2001). The success of the technique is based on the use of numerous trees, developed with different independent variables that are randomly selected from the complete original set of features (e.g. Deschamps et al. 2012; Wang et al. 2016).

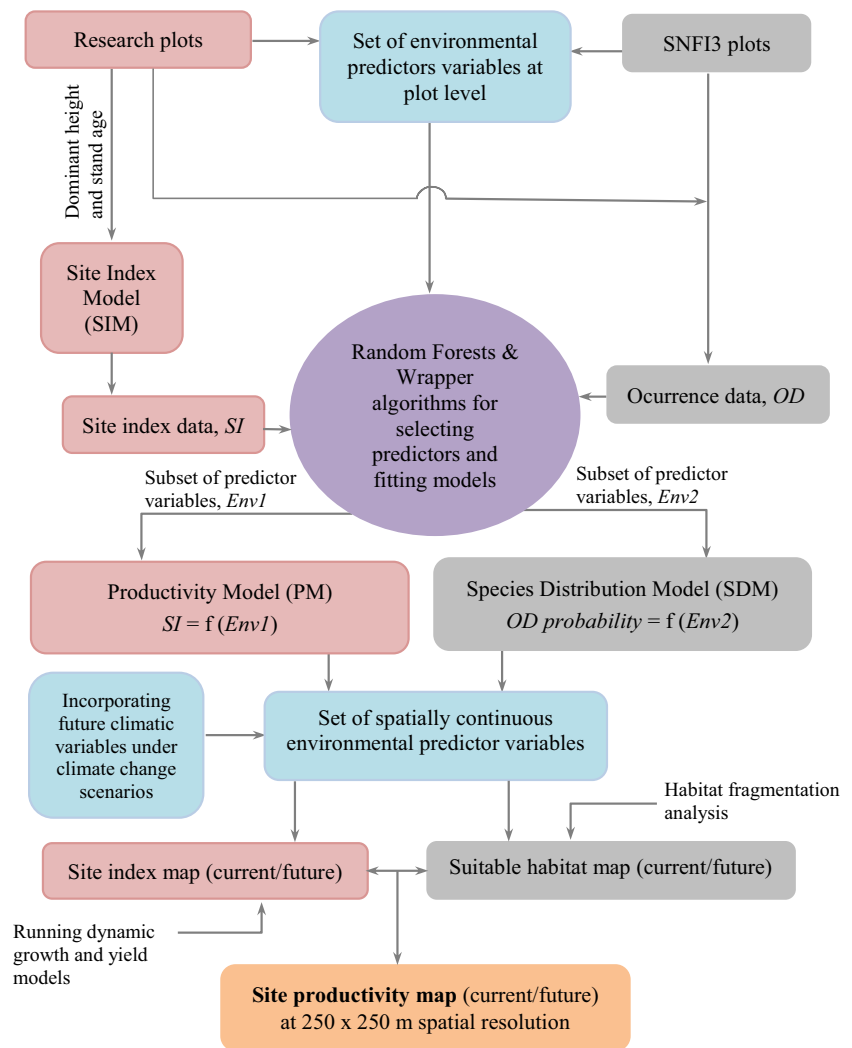
3.2 Species distribution model: current and future suitable habitat

Of the 7378 sites surveyed, maritime pine was present at 3760 sites and absent from 3618. The elevations of the sites where the trees were present ranged from 0 m to 1327 m (mean elevation = 315 m) and the latitudinal distribution ranged from 43.73 to 41.82 degrees north (mean latitude = 42.82 degrees north). As a result of the feature selection process, 11 of the 50 variables were retained as an optimal subset size for the random forest method, showing that the current distribution of the species is driven by many interrelated environmental variables (Table 4). According to the evaluation data set, the model performance was rather good (Table 5): area under the ROC curve = 0.81, overall accuracy = 0.73, true skill statistic = 0.46, kappa = 0.46. The true positive rate (sensitivity) was 0.76 and the true negative rate (specificity) was 0.69.

According to the normalized importance score, five climate variables together contributed most (52.23%) to explaining the suitable habitat for Atlantic populations of maritime pine, with thermal variables contributing 33.39% and pluviometry variables the remaining 18.84%. Five soil variables contributing 39.82% to the final model were also retained; physically related variables made up 32.31% of this contribution and a soil type classification variable accounted for the remaining 7.51%.

The functional forms of the marginal response plots of the six variables of greatest relative importance were increasing or decreasing curves (Fig. 3). The annual mean temperature (BIO1) was the most important variable, with the probability of presence increasing from 50% (mean annual temperature of 10 °C) to almost 100% at a mean annual temperature of 16 °C. The highest mean annual temperatures are reached in south-west coastal areas, in the basin of the main rivers and in the interior plane areas. The next two most important variables (BIO3 and BIO4) refer to the variability in temperature and both indicate that this species prefers rather stable temperatures, as in the coastal areas and in lower altitude river basins and plates. The fourth and fifth most important variables were the precipitation in the wettest quarter (BIO16) and seasonality of precipitation (BIO15). The former indicates that Atlantic populations of maritime pine prefer areas with a winter precipitation of around 500–550 mm, and the latter indicates preference for areas with high monthly precipitation variation, which corresponds to the south-west part of the study region where precipitation is lowest in the summer period. The sixth

Fig. 2 Workflow adopted in this study for modelling and mapping the current and future suitable habitat and site productivity for Atlantic maritime pine populations



most important variable was the soil bulk density, indicating that the presence of maritime pine was more likely when soil bulk density increases from 1150 and 1500 kg/m³. In accordance with the results obtained with the species distribution model, the total surface area of the current suitable habitat is 1.86 million hectares (Table 6).

Figure 4 shows the distribution of the five most important climatic variables (for an accumulated normalized importance value of up to 60%) for the current conditions and the future climatic scenarios for the 2050 and 2070 horizons. The future projections reveal that the main climatic variables will shift under both climatic scenarios, but that the greatest change will occur under the more pessimistic scenario (RCP 8.5) regardless of the temporal horizon. Two temperature variables, i.e. BIO1 (mean annual temperature) and BIO4 (temperature seasonality), will clearly shift towards higher values in the future. However, isothermality (BIO3) will not change in the optimistic scenario but will decrease in the pessimistic scenario. Precipitation in the wettest quarter (winter) (BIO16) will decrease slightly in the future, whereas the monthly variation in

precipitation (precipitation seasonality, BIO15) will increase notably by 2050. Changes in temperature ranges rather than changes in precipitation will apparently have a greater impact on the suitable habitat for the Atlantic populations of maritime pine.

Distribution model projections for the two different emissions scenarios reveal important shifts, with the suitable habitat for maritime pine increasing over time horizon for the most pessimistic emissions scenario. Major changes are already predicted by 2050 and imply an increase in the surface of suitable habitat of between 0.85 and 1.1 million hectares, representing a respective increase of between 45.86 and 60.75% over the current surface. In addition, projections for 2070 reveal a further increase in suitable habitat of between 4.01 and 5.38% over that predicted for 2050. Although a substantial global increase in suitable habitat is predicted, mainly reached by filling gaps between the limits of the current area, a loss of suitable habitat is predicted for north-west Asturias and León for 2070 (Fig. 7). There is no clear displacement in latitude or longitude, but there is an increase in the average

Table 3 Values of the comparative statistics (after the 10-fold cross-validation approach) of the different algorithms tested in a preliminary analysis of species distribution and productivity models

Species distribution model		CTA	FDA	GAM	GBM	GLM	MARS	RF	SRE
AUC	ANN 0.7866 (0.0201)	0.7589 (0.0175)	0.7812 (0.0160)	0.7902 (0.0188)	0.7917 (0.0162)	0.782 (0.0176)	0.785 (0.0183)	0.807 (0.0206)	0.620 (0.0164)
Productivity model		kNN	MARS	RT	RF				
R ²	MLR 0.4135 (0.2298)	0.4433 (0.2585)	0.5204 (0.1638)	0.4728 (0.1980)	0.5954 (0.2309)				

Comparative statistics: area under the ROC curve (AUC) and pseudo-coefficient of determination (R^2). Values in brackets are the standard deviations for the 10 predictions

ANN artificial neural networks, CTA classification tree analysis, FDA flexible discriminant analysis, GAM generalized additive models, GBM generalized boosted regression models, GLM generalized linear models, MARS multivariate adaptive regression splines, RF random forest, SRE rectilinear surface range envelop (equivalent to BIOCLIM), MLR multiple linear regression, KNN nearest neighbour, regression trees using M5P algorithm

elevation from 362.68 m to 486.21 m a.s.l. in the worst scenario for 2070. Considering the area occupied and the degree of fragmentation of suitable habitat for Atlantic populations of maritime pine, the mean patch area will be 170–400% higher, the large path area index will increase by 26–37% and aggregation index will increase by 6–7% (Table 6).

3.3 Productivity model: current and future site productivity

The feature selection process in the random forest method retained only 7 of 50 of current environmental variables as the optimal subset size for site index prediction. Only two variables contributed 75.37% of the importance score to the model, with lithostratigraphy being the most important (alone contributing 51.24%) and the geochronology (information about the era and period of the geological material) the second most important (contributing 24.13%). Globally, geological and soil-related variables represented 90.28% of the importance score, and climate-related variables represented the remaining 9.71% (Table 7). The marginal response plots of the two most influential variables revealed different groups or units where height growth rates of maritime pine are different. The highest productivity levels are reached on acid plutonic rocks (granites, graniodorites and quartzodiorites, code 2 in Table S1), distributed over 38.73% of the study area, with a mean site index of 15.68 m. The lithostratigraphic groups ranked second are schists, graphitic schists, phyllites, amplites and lidites (code 16 in Table S1), distributed over 12.25% of the study area, with a mean site index of 14.50 m. The group comprising alluvial deposits of gravel, sand and silt (code 56 in Table S1), distributed over 4.6% of region, yielded a mean site index of 13.81 m. Moderate productivity values were achieved in three representative lithostratigraphic groups formed by schist, mica schist, quartzitic schist, paragneiss, orthogneiss, migmatites and gravel, sand and silt deposits (alluvial sites, valley bottoms and low river terraces) (codes 1, 18 and 17 in Table S1) distributed over 22.62% of the study area, with maritime pine stands reaching an average site index value of 13.45 m. The lowest site index values were reached by the lithostratigraphic group formed by slates, sandstones and quartzites, with an average site index value of 11.37 (codes 22 and 23 in Table S1), distributed over 6.16% of the study area.

Concerning geochronological information, stands with the highest site index values (averaged value of 15.59 m) are located on geological substrates from the Palaeozoic (Devonian-Carboniferous-Permian) (code 15 in Table S2), distributed over 40.58% of the study area. Another two representative groups from the Palaeozoic, i.e. Cambrian-Ordovician (code 12 in Table S2) and Silurian-Devonian (code 22 in Table S2), yielded the next highest values of site index (average values of 14.27 m) and are distributed over

Table 4 Variables included in the SDM including their type and variable importance

Type	Variable	Normalized importance	Relative importance	Summarized values for suitable habitat			
				Mean	Max.	Min.	SD
Climate	BIO1	11.57	100.00	12.96	15.10	8.00	1.00
Climate	BIO03	11.22	93.49	3.94	4.40	3.60	0.14
Climate	BIO04	10.60	81.81	413.34	565.40	307.00	60.44
Climate	BIO16	9.44	59.84	447.94	662.00	241.00	77.91
Climate	BIO15	9.40	59.09	42.26	53.00	27.00	4.77
Soil	BD	9.28	56.85	1274.66	1506.00	1042.00	59.12
Soil	SILT	8.63	44.69	32.42	44.00	20.00	3.85
Soil	R	8.14	35.36	16.02	43.00	1.00	5.48
Terrain	TSI	7.96	31.95	0.00	0.32	-0.26	0.03
Soil	WRB-LEV ¹	7.51	23.52	–	–	–	–
Soil	DB200	6.26	0.00	199.82	200.00	161.00	1.37

To ensure values of variable importance were expressed on a comparable scale for each of the response variable, the scores of all the predictors selected were normalized so that they added up to a unitary value (normalized importance) or were expressed as relative values. Relative importance = $(VIM - VIM_{min}) / (VIM_{max} - VIM_{min})$

¹ Qualitative variable

18.73% of the study area. The lowest site index values (average value of 11.98 m) correspond to sedimentary material from the Cambrian and Ordovician periods (codes 11, 18 and 27 in Table S2) and distributed over the 12.03% of the study area.

The third most important variable (6.31% of normalized importance score) was the probability of occurrence of R horizon within 200 cm, which can be interpreted as a surrogate for effective soil depth. As this probability increases, the site index rapidly decreases (Fig. 5). The fourth most important variable (6.17% of normalized importance score) was isothermality. The expected site index values decrease as the isothermality increases.

The predictive performance of the site index from current environmental conditions was satisfactory (60.01% of the explained variability), with no tendency for underestimation or overestimation (Fig. 6). Regarding the residual metrics, the root mean square error was 1.54 m and the mean average error 1.24 m, representing 12.04% and 9.66% of the observed mean value, respectively.

Conversion of site index to maximum mean annual increment in volume per hectare and its associated optimal

rotation age was accomplished by the following two linear relationships:

$$MAI_{max} = 5.3243 + 1.094 \cdot SI + 0.0058 \cdot SI^2 - 2897.1760 \cdot (1/N) \quad (1)$$

$$Ra = 110.5400 - 5.3243 \cdot MAI_{max} + 0.0931 \cdot MAI_{max}^2 \quad (2)$$

where MAI_{max} is the maximum mean annual increment in volume ($m^3 ha^{-1} year^{-1}$), SI is the site index (m), N is density of plantation (stems ha^{-1}) and Ra (years), the optimal rotation age (age at which the MAI_{max} is reached). The coefficients of determination for Eq. (1) and Eq. (2) were 0.9991 and 0.9953, respectively.

Figure 7 shows the raster map reflecting the spatial and temporal suitable habitat and forest productivity under the two different climate change scenarios. Figure 8 shows the projections of site index model under climate change scenarios, revealing that mean site index will only increase very slightly (between 0.45 and 0.51%) relative to the current predicted mean value (14.5 m). Moreover, in all the climate change scenarios, the surface of stands with site index lower than 16 m will increase substantially. Note that in order to determine the only

Table 5 Model fit metrics for the species distribution model (SDM)

Model	Data set	AUC	OA	TSS	Kappa	Sensitivity	Specificity	PoP _{threshold}
SDM	Test	0.81	0.73	0.46	0.46	0.76	0.69	0.40
	Train	1.00	1.00	1.00	1.00	1.00	1.00	0.40

Model fit metrics include the area under the receiver operator curve (AUC), overall accuracy (OA), true skill statistic (TSS), Cohen's kappa, sensitivity and specificity. PoP is the probability of presence. Model fitting and model evaluation metrics were obtained from the training and test datasets, respectively. All the values represent the mean 10-fold cross-validation

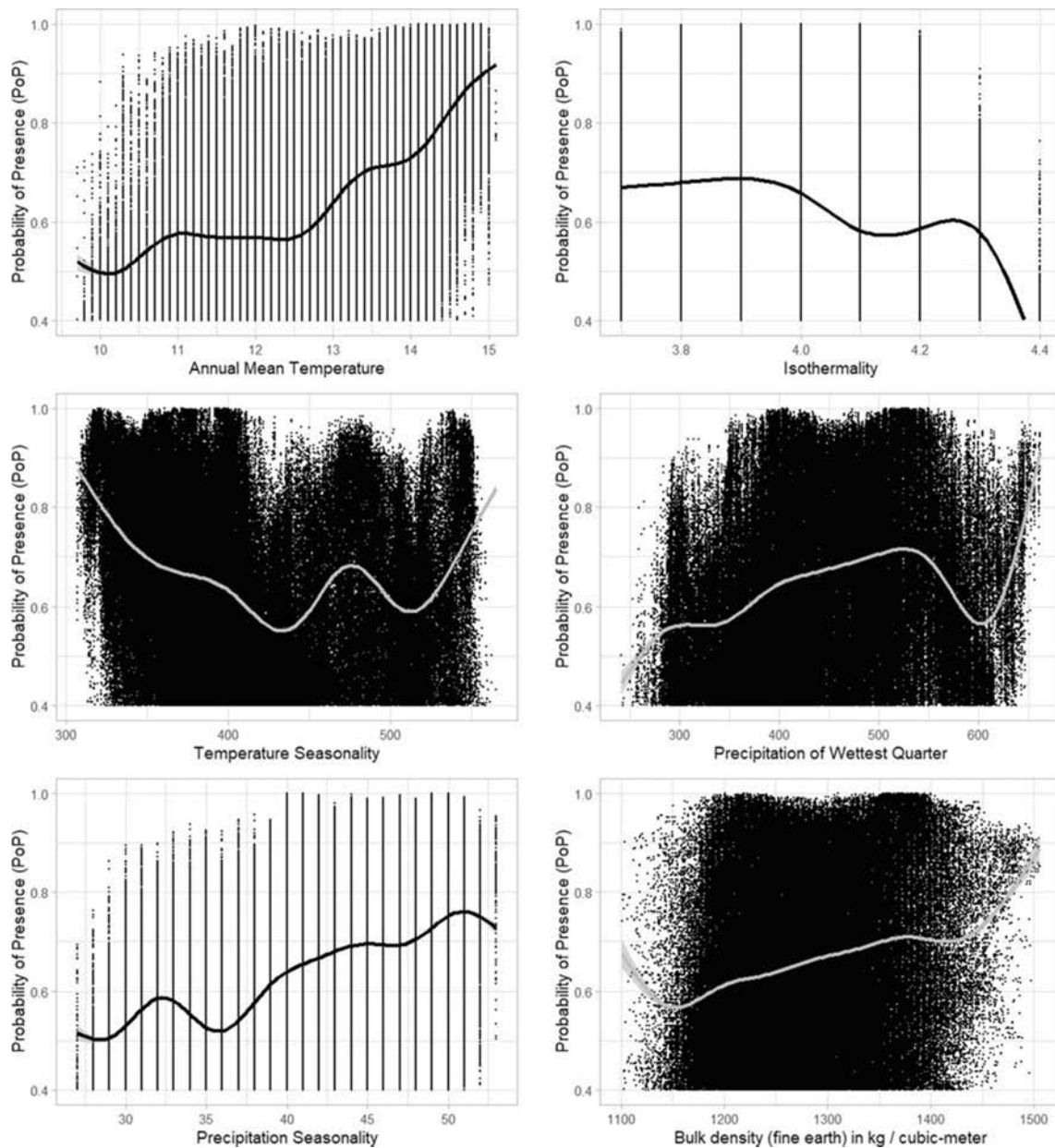


Fig. 3 Marginal response curves for the most important six variables included in the species distribution model for current environmental conditions. The variables are ordered in relation to their contribution to the model (importance score), annual mean temperature (BIO1, °C),

isothermality (BIO3, °C), temperature seasonality (BIO4, %), precipitation of wettest quarter (BIO16, mm), precipitation seasonality (BIO15, %) and bulk density of fine earth fraction (kg m^{-3}). The mean level (black or grey lines) and standard deviation (grey area) of the occurrence probability

Table 6 Changes in the distribution (mean elevation and latitude), area (total area) and fragmentation (mean patch area, largest patch index and aggregation index) of the suitable habitat of the Atlantic populations of maritime pine for current and four future climate change scenarios

Climate scenario	Elevation (m a.s.l.)	Latitude (degree)	Total area (km^2)	Largest patch index (%)	Mean patch area (ha)	Aggregation index (%)
Current	382.68	42.79	18,615.31	45.04	899.29	90.87
2050 RCP 4.5	440.82	42.81	27,152.94	71.10	2426.54	96.97
2050 RCP 8.5	471.18	42.82	29,924.31	79.28	3797.50	97.81
2070 RCP 4.5	449.74	42.81	27,899.19	73.21	2687.78	97.25
2070 RCP 8.5	486.21	42.82	30,926.19	82.52	4495.09	98.27

effect on site index of climate change; this variable was projected onto current suitable habitat, to prevent the additional effect on site index of the increase in surface area.

4 Discussion

4.1 Species distribution model and effects of climate change on suitable habitat

The distribution of the Atlantic populations of maritime pine has been expanding for more than 300 years, mainly due to human input. The question of the extent to which expansion has been human-induced is important in developing distribution models for species. We assume that because of its high economic value, the species will currently occupy adequate habitat for growth. Therefore, starting from this premise, and considering the large surface area occupied by the species in north-western Spain, we consider there is enough information to describe and model the suitable habitat.

Maritime pine is a thermophilic species that grows in areas characterised by mean annual temperatures higher than 10 °C and very variable annual rainfall (range 500 to 1300 mm) (Alía et al. 2009). Our findings have shown that temperature-related variables are the environmental characteristics that most strongly affect the distribution of the Atlantic populations of maritime pine in Spain. The probability of presence of the species is greatest in areas where the mean annual temperature is highest and the variability is lowest (BIO3 and BIO4 variables), in accordance with the autoecological data published for the species (Abad Viñas et al. 2016; Nicolás and Gandullo 1967; Gandullo and Sánchez Palomares 1994).

The minor effect of the precipitation-related variables is consistent with the fact that in north-western Spain annual rainfall is between 1000 and 1300 mm, which is clearly higher than the minimum of around 400–500 mm required by the species (Abad Viñas et al. 2016). The increase in the presence of this species in areas with seasonal precipitation, which is highest in winter, can be spatially translated into lower altitude areas near the Galician coast. In Galicia, rainfall decreases with distance from the sea, as a consequence of the general atmospheric circulation pattern with fronts arriving from the west (De Uña Álvarez 2001).

Regarding soil characteristics, soil bulk density values may be related to chemical and physical soil properties. This physical soil variable is usually negatively and closely related to organic carbon content or soil organic matter (e.g. Alexander 1989; Grigal and Vance 2000; Périé and Ouimet 2008; Sakin 2012) and strongly positively correlated with the sand content (Chaudhari et al. 2013). Bulk density can therefore be viewed as a surrogate for soil properties such as soil organic matter or sand content, among others, suggesting that maritime pine prefers soils with low organic matter content and of sandy

texture. The former statement is consistent with the influence of the mean annual temperature on distribution of the Atlantic populations of maritime pine, as higher rates of organic matter mineralisation are expected in areas with higher temperatures (Álvarez-Álvarez et al. 2011). Concerning the latter statement, sandy textured soils have long been recognized as most appropriate for the species (Nicolás and Gandullo 1967; Bará and Toval 1983; Gandullo and Sánchez Palomares 1994; Abad Viñas et al. 2016).

The spatial analysis predicted an increase in suitable habitat surface by 2050 of between 46 and 61% of the current area. These findings are consistent with those of broader scale analyses. According to the Forest Focus European dataset, an important degree of expansion of the suitable habitat area for north-western Spain is expected in the next few decades (<http://data.europa.eu/88u/dataset/74bb6627-4c19-4ac7-ac59-73e302bfc117>), whereas Bede-Fazekas and Levente Horváth (2014) indicated a slight expansion. However, loss of suitable habitat is predicted for areas near the coast in western Asturias and the north-west of León for 2070. This is a consequence of current precipitation in the wettest quarter (BIO16) in the area (close to 300 mm) and the expected shifts in this variable towards lower values in the future (Fig. 4).

As Atlantic populations of maritime pine are almost all managed, the predicted gains in suitable habitat should be interpreted as areas where new plantations of this species can be established. Nevertheless, colonization simulations in Mediterranean populations have shown that maritime pine, mainly dispersed by wind, has a notable colonization capacity that allows extension of the distribution area and occupation of the new available land (Juez et al. 2014). In addition, the high level of standing variation and phenotypic plasticity of the Atlantic populations of maritime pine should enhance its capacity to adapt to the future climate (Serra-Varela et al. 2017).

The predicted high expansion of suitable habitat for Atlantic maritime pine stands is reflected in the expected metrics of landscape fragmentation, with more continuous stands of the species being possible in the future.

4.2 Productivity model and effects of climate change on productivity

As the productivity model enables site index to be predicted as a function of environmental variables (including several climatic variables), the model is capable of predicting changes in site index under non-constant climate. Lithological variables (lithostratigraphy and the age of the geological formations) were clearly the prominent features explaining the site index variability in the Atlantic maritime pine stands. This is not surprising as the geological formation has previously been reported to be an important factor affecting the potential growth of maritime pine (e.g. Bravo-Oviedo et al. 2011;

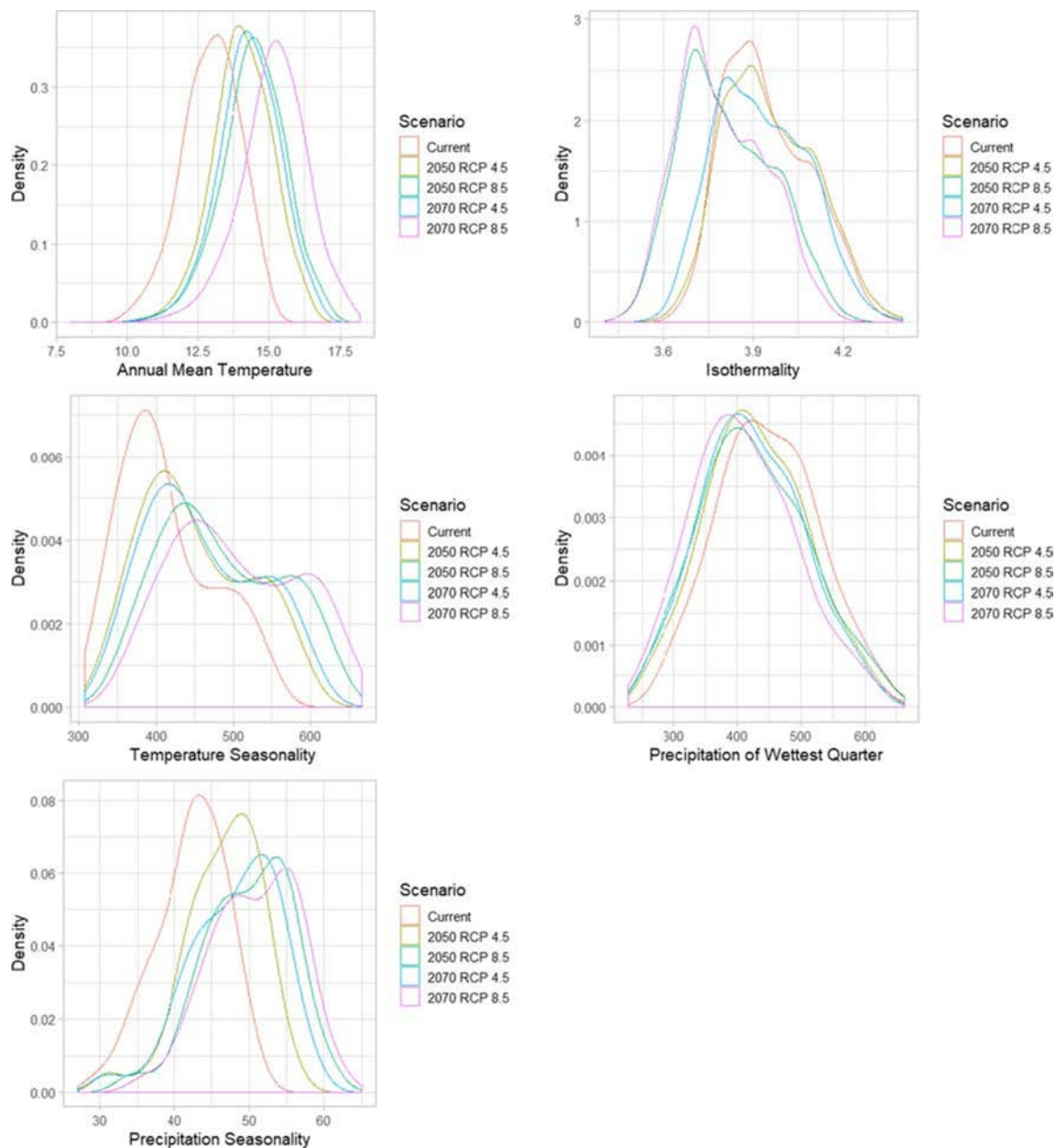


Fig. 4 Distribution of the climatic variables that contributed to the model explaining the distribution under five scenarios: (1) the current reference period; (2) 2050 under the RCP 4.5 emissions scenario; (3) 2050 under the RCP 8.5 emissions scenario; (4) 2070 under the RCP 4.5 emissions

scenario; and (5) 2070 under the RCP 8.5 emissions scenario. The variables shown presented an accumulated normalized importance of up to 52.23%

Eimil-Fraga et al. 2014). According to these authors (*op. cit.*), lithological variables are strongly related to soil depth, availability of some soil nutrients, elevation, slope and temperature as result of its distribution and geomorphology over the region. The high site index of Atlantic maritime pine stands established in soil over plutonic rocks (granites, granodiorites and quartz diorites) may be due to the fact that these stands are located at lower elevations and that the soil may be deep owing to the higher degree of tectonization in areas with abundant faults and fractures (Eimil-Fraga et al. 2014). By contrast, the lowest site index reached on quartzites, slates and

sandstones between the Cambrian and Ordovician may be related to the high resistance of these rocks to weathering, which frequently results in shallow soils (Eimil-Fraga et al. 2014).

The next most important variable, the probability of occurrence of R horizon within 200 cm, can be viewed as a surrogate for exploitable soil volume or rootable depth. Other authors have found that maritime pine site productivity is primarily associated with soil depth (e.g. Bará and Toval 1983; Oliveira et al. 2000; Álvarez-Álvarez et al. 2011; Eimil-Fraga et al. 2014). The observation that the only variable directly

Table 7 Variables included in the productivity model, including their type and variable importance

Type	Variable	Normalized importance	Relative importance	Summarized values for suitable habitat			
				Mean	Max	Min	SD
Lithostratigraphy	LIT_dlo ¹	51.24	100.00	—	—	—	—
Geology	Geo_es ¹	24.13	43.16	—	—	—	—
Soil	R	6.31	5.80	16.02	43.00	1.00	5.48
Climate	BIO03	6.17	5.52	3.94	4.40	3.60	0.14
Soil	CEC	4.71	2.45	25.04	66.00	15.00	2.82
Soil	WRB-LEV1 ¹	3.89	0.72	—	—	—	—
Climate	BIO08	3.54	0.00	8.59	12.50	2.60	1.64

To ensure values of variable importance were expressed on a comparable scale for each of the response variable, the scores of all the predictors selected were normalized so that they added up to a unitary value (normalized importance) or were expressed as relative values. Relative importance = $(VIM - VIM_{min}) / (VIM_{max} - VIM_{min})$

¹ Qualitative variables

related to the nutrient soil content (cation-exchange capacity) is of little importance in estimating site index indicates the

frugality of the species. Frugality is a characteristic of pioneer species and usually is associated with the presence of fungi

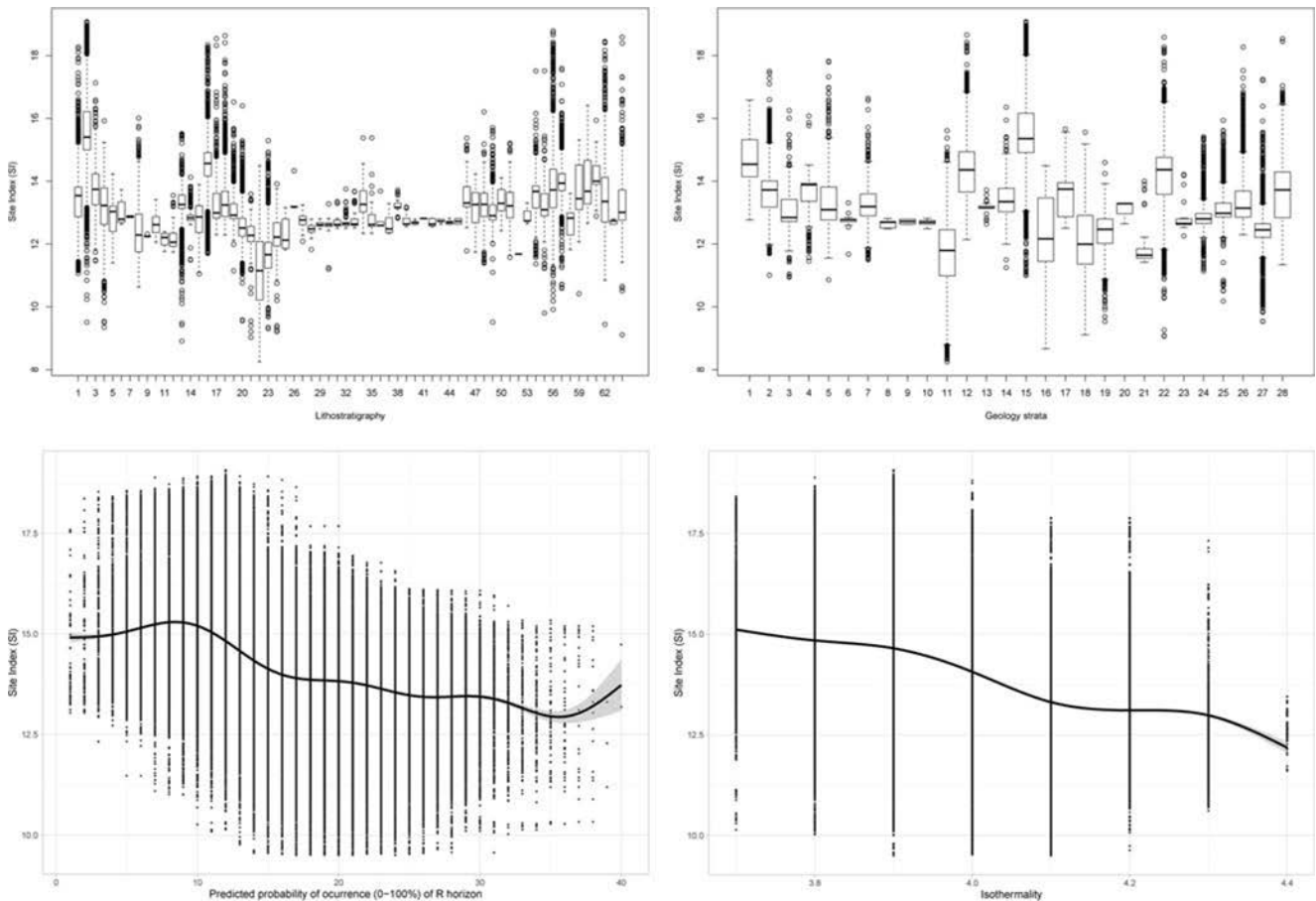


Fig. 5 Marginal response curves for the four variables included in the productivity model for current environmental conditions. Variables are ordered by their contribution to the model (importance score). LIT_dlo, lithostratigraphy; Geo_units, geology strata; R, probability occurrence of R horizon within 0–200 cm (%), isothermality (BIO3, °C). The mean (black line) and standard deviation (grey area) of the probability presence.

The prediction value of site index is shown as a function of each variable while all other variables are held at their median values for locations where Atlantic populations of maritime pine is present. Note: supplementary Tables S1 and S2 include the significance of the numerical codes shown in the two first graphs

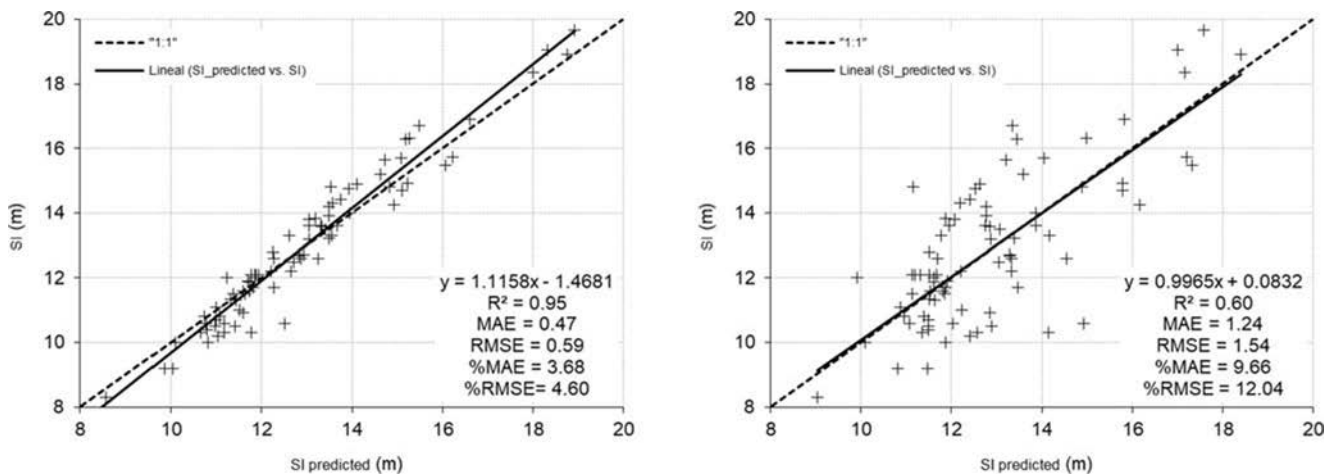


Fig. 6 Field measures vs. predicted values of site index (SI) for training (left) and evaluation (right). Solid lines indicate the regression fits ($n = 85$, 10-fold-CV) and dashed lines. The 1:1 identity line

that commonly form the mycorrhizae (Bará and Toval 1983; Maugé 1987). The negligible influence of chemical soil variables on site index has long been observed (Bará and Toval 1983; Álvarez-Álvarez et al. 2011; Eimil-Fraga et al. 2014) and explains the remarkable character of maritime pine as a “site-demanding” rather than a “nutrient demanding” species.

The first climatic variable contributing to explaining site index (isothermality, BIO3) may indicate that productivity is more closely related to low variations in temperature than to the mean annual temperature, although other authors have found that the latter variable is also positively correlated with site index in NW Spain (Eimil-Fraga et al. 2014). By contrast, in maritime pine stands in Mediterranean areas of Spain, mean annual temperature is only positively correlated with site index in areas characterised by high precipitation (Bravo-Oviedo et al. 2011). The high rainfall in NW Spain may

explain why precipitation variables do not significantly explain site index variation, unlike in the Mediterranean area.

The total variance explained by the fitted site index model (60%) can be considered suitable within the framework of site index estimation from environmental variables (Aertsen et al. 2010; Bontemps and Bouriaud 2014). A few studies have reported better performance of site index models developed for other species, mainly conifers: e.g. Fontes et al. (2003) for Douglas fir, Sharma et al. (2012) for Norway spruce and Scots pine, Brandl et al. (2014) for Norway spruce and Bjelanovic et al. (2018) for trembling aspen, lodgepole pine and white spruce. Many studies have reported similar or less accurate results; e.g. Bergès et al. (2005) for Sessile oak, Albert and Schmidt (2010) for beech and Norway spruce, Aertsen et al. (2010) for *Pinus brutia*, *Pinus nigra* and *Cedrus libani*, Sabatia and Burkhart (2014) for loblolly pine and Watt et al.

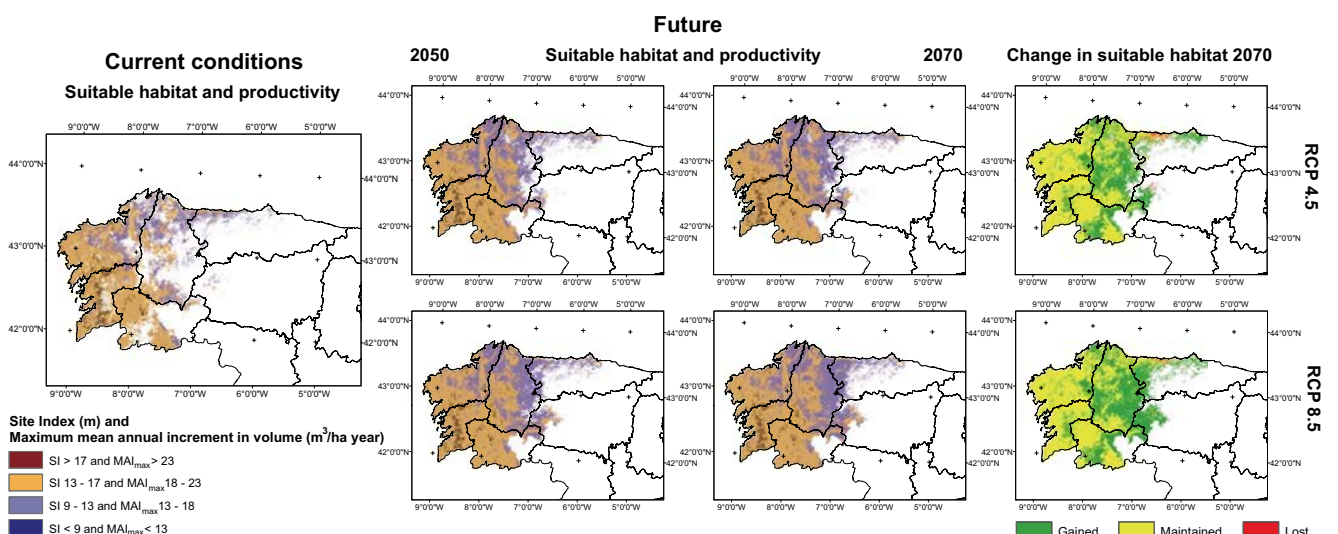
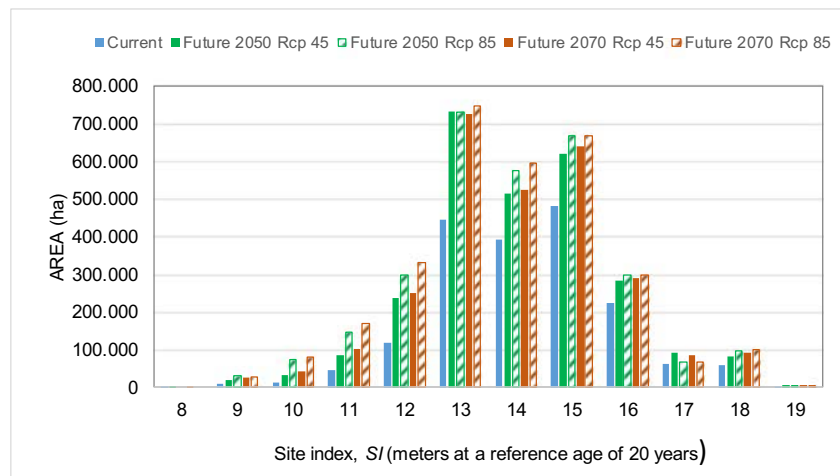


Fig. 7 Model predictions of current and future suitable habitat and productivity for Atlantic maritime pine populations in Spain under two different climate change scenarios (RCP 4.5 and RCP 8.5). Note:

maximum mean annual volume increment (MAI_{max}) intervals correspond to a plantation density of $1100 \text{ stems ha}^{-1}$

Fig. 8 Distribution of area of suitable habitat for different site index classes under current environmental conditions and two different climatic scenarios (RCP 4.5 and RCP 8.5) for 2050 and 2070



(2015) for radiata pine. Concerning maritime pine, previous studies have obtained less accurate results than in the present study: Pacheco Marques (1991), Bravo-Oviedo et al. (2011) Álvarez-Álvarez et al. (2011) and Eimil-Fraga et al. (2014). In addition, these models include some soil chemical variables that are expensive to obtain, thus restricting or decreasing the practical application of the models. The model presented here can be considered operational and inexpensive, as fieldwork is not required for site index estimation.

Several statistical approaches, including multiple linear regression and artificial neural networks (Aertsen et al. 2010), have been used to develop models to predict site index from environmental variables. Non-parametric techniques are considered very flexible and robust for this purpose, although they have been less widely used than multiple linear regression (Jiang et al. 2015). This can be partly explained by the fact that the former techniques do not provide explicit equations that facilitate application by forest managers who are not specifically trained in machine learning techniques. To date, few studies have used random forest algorithms within the framework of the present study (but see e.g. Weiskittel et al. 2011b; Sabatia and Burkhart 2014; Jiang et al. 2015). Although the results obtained by these authors (op. cit.) were promising, it has been suggested that random forest algorithms could provide illogical site index predictions under extrapolation (e.g. Aertsen et al. 2010; Sabatia and Burkhart 2014). Considering current climate conditions, use of the widest possible range of observed data would prevent extrapolation-related problems. However, prediction of species distribution and productivity models for future climatic conditions may not be as reliable (Sabatia and Burkhart 2014).

The good performance of the linear model for the maximum mean annual volume increment as a function of site index was expected a priori because there is usually a strong relationship between maximum mean (annual or periodical) increment in volume and site index (Waring et al. 2006). In addition, the maximum mean annual volume increment data

was derived from simulations using dynamic growth models for the species, i.e. the data was smoothed and much of the variation removed (e.g. Hasenauer and Monserud 1997). The model used to estimate the optimal rotation age also performed very well and will serve practitioners as a support for considering different harvesting age alternatives.

Finally, two explanations are suggested for the minor effect of climate change in the expected productivity (increase < than 1%) of the Atlantic populations of maritime pine: (i) only two temperature variables, i.e. isothermality (BIO3) and mean temperature of the wettest quarter (winter -BIO8), are related to maritime pine productivity, and (ii) the joint contribution is low, accounting for only 9.71% of the normalized importance.

5 Conclusions

Two raster-based models of 250 m resolution were developed in order to predict the current and future suitable habitat and productivity for Atlantic populations of maritime pine in Spain under climate change scenarios. Both models were developed using the random forest machine learning technique and currently available spatially continuous environmental variables.

Climate variables drive the suitable habitat of the species, whereas site index variability is mainly explained by lithological or physical soil variables. Climate change is expected to lead to a large increase in the area of suitable habitat for Atlantic populations of maritime pine by 2070. Predicted gain in suitable habitat may lead to more favourable conditions for the species, making it preferable to radiata pine (also widely used in the region) for reforestation programmes. The proposed models may be useful as decision-making tools for forest managers and politicians planning sustainable use of the forest resources under changing climate. Further research aimed at obtaining a better understanding of the complex relationships between environmental variables and species

occurrence and productivity is advisable to enhance the climate-sensitive predictive models.

Funding information The research was supported by funding from the Spanish Ministry of Science and Innovation (projects AGL2008–02259/FOR) and from the local government of Asturias (projects CN-07-094 and SV-PA-13-ECOEMP-58).

Data availability The datasets generated and/or analysed during the current study are available in the Forest Inventory and Analysis Database. Ministerio de Agricultura, Pesca y Alimentación. Gobierno de España. https://www.miteco.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/infomacion-disponible/inf3_bbdd_descargas.htm.aspx

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

References

- Abad Viñas R, Caudullo G, Oliveira S, de Rigo D (2016) *Pinus pinaster* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, Houston Durrant T, Mauri A (eds) European Atlas of Forest Tree Species. Publ. Off. EU, Luxembourg, p e012d59+
- Aertsen W, Kint V, van Orshoven J, Ozkan K, Muys B (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol Model* 221:1119–1130. <https://doi.org/10.1016/j.ecolmodel.10.01.007>
- Albert M, Schmidt M (2010) Climate-sensitive modelling of site-productivity relationships for Norway spruce (*Picea abies* (L.) Karst.) and common beech (*Fagus sylvatica* L.). *Forest Ecol Manag* 259:739–749. <https://doi.org/10.1016/j.foreco.2009.04.039>
- Alexander EB (1989) Bulk density equations for southern Alaska soils. *Can J Soil Sci* 69:177–180
- Alía R, Martín S, De Miguel J, Galera R, Agúndez D, Gordo J, Salvador L, Catalán G, Gil A (1996) Regiones de procedencia *Pinus pinaster* Aiton. Dirección General de Conservación de la Naturaleza, Madrid
- Alía R, Moro J, Denis JB (1997) Performance of *Pinus pinaster* provenances in Spain: interpretation of the genotype by environment interaction. *Can J For Res* 27:1548–1559
- Alía R, García Del Barrio JM, Iglesias S, Mancha JA, De Miguel J, Nicolás JL, Pérez F, Sánchez De Ron D (2009) Regiones de Procedencia de especies forestales en España. Ministerio de Medio Ambiente y Medio Rural y Marino. Organismo Autónomo Parques nacionales, Madrid
- Álvarez-Álvarez P, Afif E, Cámara-Obregón A, Castedo-Dorado F, Barrio-Anta M (2011) Effects of foliar nutrients and environmental factors on site productivity in *Pinus pinaster* Ait. Stands in Asturias (NW Spain). *Ann For Sci* 68:497–509. <https://doi.org/10.1007/s13595-011-0047-5>
- Álvarez-González JG, Ruiz González AD, Rodríguez Soalleiro R, Barrio Anta M (2005) Ecoregional site index models for *Pinus pinaster* in Galicia (northwestern Spain). *Ann For Sci* 62:115–127. <https://doi.org/10.1051/forest:2005003>
- Arias-Rodil M, Barrio-Anta M, Diéguez-Aranda U (2016) Developing a dynamic growth model for maritime pine in Asturias (NW Spain): comparison with nearby regions. *Ann For Sci* 73(2):297–320. <https://doi.org/10.1007/s13595-015-0501-x>
- Bará S, Toval G (1983) Calidad de estación del *Pinus pinaster* Ait. en Galicia. INIA. Serie: Recursos Naturales, 24. Madrid
- Barrio Anta M, Diéguez-Aranda U (2005) Site quality of pedunculate oak (*Quercus robur* L.) stands in Galicia (northwest Spain). *Eur J Forest Res* 124:19–28. <https://doi.org/10.1007/s10342-004-0045-3>
- Bede-Fazekas A, Levente Horváth KM (2014) Impact of climate change on the potential distribution of Mediterranean pines. *Q J Hung Meteorol Serv* 118:41–52
- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecol Lett* 15: 365–377. <https://doi.org/10.1111/j.1461-0248.2011.01736.x>
- Bergès L, Chevalier R, Dumas Y, Franc A, Gilbert, JM (2005) Sessile oak (*Quercus petraea* Liebl.) site index variations in relation to climate, topography and soil in even-aged high-forest stands in northern France. *Ann For Sci* 62: 391–402. <https://doi.org/10.1051/forest:2005035>
- Bjelanovic I, Comeau PG, White B (2018) High resolution site index prediction in boreal forests using topographic and wet areas mapping attribute. *Forests* 9:113. <https://doi.org/10.3390/f9030113>
- Bontemps JD, Bouriaud O (2014) Predictive approaches to forest site productivity: recent trends, challenges and future perspectives. *Forestry* 87(1):109–128. <https://doi.org/10.1093/forestry/cpt034>
- Booth TH, Nix HA, Busby JR, Hutchinson MF (2014) Bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxENT studies. *Diversity Distrib* 20: 1–9
- Brandl S, Falk W, Klemmt HJ, Stricker G, Bender A, Rötzer T, Pretzsch H (2014) Possibilities and limitations of spatially explicit site index modelling for spruce based on National Forest Inventory Data and digital maps of soil and climate in Bavaria (SE Germany). *Forests* 5(11):2626–2646. <https://doi.org/10.3390/f5112626>
- Bravo-Oviedo A, Roig S, Bravo F, Montero G, Del-Río M (2011) Environmental variability and its relationship to site index in Mediterranean maritime pine. *Forest Syst* 20(1):50–64. <https://doi.org/10.5424/fs/2011201-9106>
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Burkhardt HE, Tomé M (2012) Modelling forest trees and stands. Springer, Berlin
- Chaudhari PR, Ahire DV, Ahire VD, Chkravarty M, Maity S (2013) Soil bulk density as related to soil texture, organic matter content and available total nutrients of Coimbatore soil. *Int J Sci Res* 3(2):1–8
- Clutter J, Fortson J, Pienaar L, Brister H, Bayley R (1983) Timber management: a quantitative approach. Wiley, New York
- Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for automated geoscientific analyses (SAGA) v. 3.0.0. *Geosci Model Dev* 8: 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cosenza DN, Leite HG, Marcatti GE, Binoti DHB, Alcántara AEM, de Rode R (2015) Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. *Scientia Forestalis* 43:955–963. <https://doi.org/10.18671/scifor.v43n108.19>
- Davis ME, Shaw RG, Etterson JR (2005) Evolutionary responses to climate change. *Ecology* 86:1704–1714. <https://doi.org/10.1890/03-0788>
- De Uña Álvarez E (2001) El clima, in: Atlas de Galicia. Tomo 1: Medio Natural, edited by: Precedo Ledo, A and Sancho Comíns, J, Sociedade para o Desenvolvimento Comarcal de Galicia, Xunta de Galicia
- Deschamps B, McNaim H, Shang J, Jiao X (2012) Towards operational radar-only crop type classification: comparison of a traditional decision tree with a random forest classifier. *Can J Remote Sens* 38:60–68. <https://doi.org/10.5589/m12-012>
- DGCN (2006) III Inventario Forestal Nacional (1997–2006). Principado de Asturias, Galicia y León. Dirección General de Conservación de la Naturaleza, Secretaría General de Medio Ambiente, Ministerio de Medio Ambiente. Madrid

- Diéguez-Aranda U, Rojo Alboreca A, Castedo-Dorado F, Álvarez-González JG, Barrio-Anta M, Crecente-Campo F, González-González JM, Pérez-Cruzado C, Rodríguez-Soalleiro R, López-Sánchez CA, Balboa-Murias MA, Gorgoso Varela, JJ, Sánchez-Rodríguez F (2009) Herramientas selvícolas para la gestión forestal sostenible en Galicia. Xunta de Galicia
- Drummond ST, Sudduth KA, Joshi A, Birrell SJ, Kitchen NR (2003) Statistical and neural methods for site-specific yield prediction. *Trans ASAE* 46:5–14. <https://doi.org/10.13031/2013.12541>
- Dyderski MK, Paz S, Frelich LE, Jagodzinski AM (2017) How much does climate change threaten European forest tree species distributions? *Glob Change Biol* 24:1150–1163. <https://doi.org/10.1111/gcb.13925>
- EEA (2017) Climate change, impacts and vulnerability in Europe 2016. An indicator-based report, European Environment Agency Report N° 1/2017
- Eimil-Fraga C, Rodríguez-Soalleiro R, Sánchez-Rodríguez F, Pérez-Cruzado C, Álvarez-Rodríguez E (2014) Significance of bedrock as a site factor determining nutritional status and growth of maritime pine. *Forest Ecol Manag* 331:19–24. <https://doi.org/10.1016/j.foreco.2014.07.024>
- Fontes L, Tomé M, Thompson F, Yeomans A, Sales LJ, Savill P (2003) Modelling the Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) site index from site factors in Portugal. *Forestry* 76:491–507. <https://doi.org/10.1093/forestry/76.5.491>
- Gadow K, Bredenkamp B (1992) Forest management. Academica, Pretoria
- Gandullo JM, Sánchez Palomares O (1994) Estaciones ecológicas de los pinares españoles. MAPA-ICONA, Madrid
- Grigal DF, Vance ED (2000) Influence of soil organic matter on forest productivity. *NZ J Forestry Sc* 30(1/2):169–205
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *SIGKDD Explorations* 11(1):10–18. <https://doi.org/10.1145/1656274.1656278>
- Harris RMB, Grose MR, Lee G, Bindoff NL, Porfirio LL, Fox-Hughes P (2014) Climate projections for ecologists. *WIREs Clim Change* 5: 621–637. <https://doi.org/10.1002/wcc.291>
- Hasenauer H, Monserud RA (1997) Biased predictions for tree height increment models developed from smoothed 'data'. *Ecol Model* 98:13–22. [https://doi.org/10.1016/S0304-3800\(96\)01933-3](https://doi.org/10.1016/S0304-3800(96)01933-3)
- Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotic A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12(2):e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- Hijmans RJ, Cameron SE, Parra JL, Jones P, 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>
- Huang S, Ramírez C, Conway S, Kennedy K, Kholer T, Liu J (2017) Mapping site index and volume increment from forest inventory, Landsat, and ecological variables in Tahoe National Forest, California. *USA Can J For Res* 47:113–124. <https://doi.org/10.1139/cjfr-2016-0209>
- IGME (2015a) Mapa Geológico de España a escala 1:200.000. Instituto Geológico y Minero de España, Ministerio de Ciencia, Innovación y Universidades Madrid
- IGME (2015b) Mapa Geológico de la Península Ibérica, Baleares y Canarias a escala 1:1.000.000. Instituto Geológico y Minero de España, Ministerio de Ciencia, Innovación y Universidades Madrid
- IPCC (2013) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel of climate change. Cambridge University Press, Cambridge
- Jiang H, Radtke PJ, Weiskittel AR, Coulston JW, Guertin PJ (2015) Climate- and soil-based models of site productivity in eastern US tree species. *Can J For Res* 45:325–342. <https://doi.org/10.1139/cjfr-2014-0054>
- Juez L, González-Martínez SC, Nanos N, de-Lucas AI, Ordóñez C, del Peso C, Bravo F (2014) Can seed production and restricted dispersal limit recruitment in *Pinus pinaster* Aiton from the Spanish northern plateau? *Forest Ecol Manag* 313:329–339. <https://doi.org/10.1016/j.foreco.2013.10.033>
- Latta G, Temesgen H, Barrett TM (2009) Mapping and imputing potential productivity of Pacific northwest forests using climate variables. *Can J For Res* 39:1197–1207. <https://doi.org/10.1139/X09-046>
- MAPA (2019) Anuario de Estadística Forestal 2019. Ministerio de Agricultura, Pesca y Alimentación. Madrid
- Maugé J (1987) Le pin maritime. Premier résineux de France. IDF, Paris, 192 pages
- McGarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat selection modeling: a review and outlook. *Landscape Ecol* 31:1161–1175. <https://doi.org/10.1007/s10980-016-0374-x>
- McKenney DW, Pedlar JH (2003) Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. *Forest Ecol Manag* 175:497–507. [https://doi.org/10.1016/S0378-1127\(02\)00186-X](https://doi.org/10.1016/S0378-1127(02)00186-X)
- Monzón J, Moyer-Horner L, Palamar MB (2011) Climate change and species range dynamics in protected areas. *BioScience* 61:752–761. <https://doi.org/10.1525/bio.2011.61.10.5>
- Nicolás A, Gandullo JM (1967) Ecología de los pinares españoles. I *Pinus pinaster* Ait. IFIE, Madrid, 311 p
- Novo-Fernández A, Barrio-Anta M, Recondo C, Cámara-Obregón A, López-Sánchez CA (2019) Integration of National Forest Inventory and nationwide airborne laser scanning data to improve forest yield predictions in North-Western Spain. *Remote Sens* 11(14):1693. <https://doi.org/10.3390/rs11141693>
- Oliveira AC, Pereira JS, Correia AV (2000). A silvicultura do pinheiro bravo. Centro Pinus, 102 pp. Portugal
- Pacheco Marques C (1991) Evaluating site quality of even-aged maritime pine stands in northern Portugal using direct and indirect methods. *For Ecol Manag* 41:193–204. [https://doi.org/10.1016/0378-1127\(91\)90103-3](https://doi.org/10.1016/0378-1127(91)90103-3)
- Parresol BR, Scott DA, Zarnoch SJ, Edwards LA, Blake JI (2017) Modelling forest site productivity using mapped geospatial attributes within a South Carolina landscape, USA. *For Ecol Manag* 406:196–207. <https://doi.org/10.1016/j.foreco.2017.10.006>
- Périé C, Ouimet R (2008) Organic carbon, organic matter and bulk density relationships in boreal forest soils. *Can J Soil Sci* 88(3):315–325
- Prasad A, Iverson L, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–199. <https://doi.org/10.1007/s10021-005-0054-1>
- Roces-Díaz JV, Jiménez-Alfaro B, Álvarez-Álvarez P, Álvarez-García MA (2015) Environmental niche and distribution of six deciduous tree species in the Spanish Atlantic region. *iForest* 8:224–231. <https://doi.org/10.3832/ifer1183-008>
- Rodríguez Soalleiro R, Madrigal Collazo, A (2008) Silvicultura de *Pinus pinaster* Ait. subsp. atlantica H. de Vill. In: Serrada R, Montero G, Reque JA (eds) Compendio de silvicultura aplicada en España. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Ministerio de Educación y Ciencia, Madrid, pp 367–398

- Romayà J, Vallejo VR (2004) Productivity of *Pinus radiata* plantation in Spain in response to climate and soil. For Ecol Manag 195:177–189. <https://doi.org/10.1016/j.foreco.2004.02.045>
- Ruiz Zorrilla P (1980) Notas para una historia del pino en Galicia. Ministerio de Cultura, Dirección General del Patrimonio Artístico, S.G. de Archivos, Madrid
- Ryan TP (1997) Modern regression methods. John Wiley & Sons, New York
- Sabatia CH, Burkhart HE (2014) Predicting site index of plantation loblolly pine from biophysical variables. For Ecol Manag 326:142–156. <https://doi.org/10.1016/j.foreco.2014.04.019>
- Sakin E (2012) Organic carbon organic matter and bulk density relationships in arid-semi arid soils in Southeast Anatolia region. Afr J Biotechnol 11(6):1373–1377. <https://doi.org/10.5897/AJB11.2297>
- Sánchez-Rodríguez F, Rodríguez-Soalleiro R, Español E, López CA, Merino A (2002) Influence of edaphic factors and tree nutritive status on the productivity of *Pinus radiata* D. Don plantations in Northwestern Spain. For Ecol Manag 171(1–2):181–189. [https://doi.org/10.1016/S0378-1127\(02\)00471-1](https://doi.org/10.1016/S0378-1127(02)00471-1)
- Serra-Varela MJ, Alía R, Ruiz Daniels R, Zimmermann NE, Gonzalo-Jiménez J, Grivet D (2017) Assessing vulnerability of two Mediterranean conifers to support genetic conservation management in the face of climate change. Divers Distrib 23:507–516. <https://doi.org/10.1111/ddi.12544>
- Sharma RP, Brunner A, Eid T (2012) Prediction from site and climate variables for Norway spruce and Scots pine in Norway. Scand J For Res 27:619–636. <https://doi.org/10.1080/02827581.2012.685749>
- Shirk AJ, Cushman SA, Waring KM, Wehenkel CA, Leal-Sáenz A, Toney C, Lopez-Sanchez CA (2018) Southwestern white pine (*Pinus strobiformis*) species distribution models project a large range shift and contraction due to regional climatic changes. Forest Ecol Manag 411:176–186. <https://doi.org/10.1016/j.foreco.2018.01.025>
- Skovsgaard JP, Vanclay JK (2008) Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry 81:13–31. <https://doi.org/10.1093/forestry/cpm041>
- Thuiller W, Georges D, Engler R, Breiner F (2016) ‘biomod2’: ensemble platform for species distribution modelling
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. Clim Chang 109:5–31. <https://doi.org/10.1007/s10584-011-0148-z>
- Vanclay JK (1994) Modelling forest growth and yield: applications to mixed tropical forests. Wallingford CAB International
- Vennetier M, Vilà B, Liang E, Guibal F, Taahbet A, Gadbin-Henry C (2007) Impact of climate change on pines forest productivity and on the shift of a bioclimatic limit in Mediterranean area. In: Leone V, Lovreglio R (eds), proceedings of the international workshop MEDPINE 3: conservation, regeneration and restoration of Mediterranean pines and their ecosystems. Options Méditerranéennes: Série a. Séminaires Méditerranéens n.75
- Wang Y, Raulier F, Ung CH (2005) Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods—a case study of lodgepole pine productivity. Forest Ecol Manag 214:201–211. <https://doi.org/10.1016/j.foreco.2005.04.025>
- Wang L, Zhou X, Zhu X, Dong X, Guo W (2016) Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J 4:212–219. <https://doi.org/10.1016/j.cj.2016.01.008>
- Waring RH, Milner KS, Jolly WM, Phillips L, McWethy D (2006) Assessment of site index and forest growth capacity across the Pacific and Inland Northwest U.S.A. with a MODIS satellite-derived vegetation index. Forest Ecol Manag 228:285–291. <https://doi.org/10.1016/j.foreco.2006.03.019>
- Watt SW, Dash JP, Bhandari S, Watt P (2015) Comparing parametric and non-parametric methods of predicting site index for radiata pine using combinations of data derived from environmental surfaces, satellite imagery and airborne laser scanning. Forest Ecol Manag 357:1–9. <https://doi.org/10.1016/j.foreco.2015.08.001>
- Weiskittel AR, Hann DW, Kershaw JA Jr, Vanclay JK (2011a) Forest growth and yield modeling. Wiley-Blackwell, Oxford
- Weiskittel AR, Crookston NL, Radtke PJ (2011b) Linking climate, gross primary productivity, and site index across forests of the western United States. Can J For Res 41(8):1710–1721. <https://doi.org/10.1139/x11-086>
- Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O’Brien JN, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. Divers Distrib 15:565–576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x>
- Zhiwei X, Xinghua W (2010) Research for information extraction based on wrapper model algorithm. 2010 Second international conference on computer research and development. Kuala Lumpur, Malaysia, pp. 652–655

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.