Climate change may threaten the southernmost \textit{Pinus nigra} subsp. \textit{salzmannii} (Dunal) Franco populations: an ensemble niche-based approach

Rafael M Navarro-Cerrillo (1), Joaquín Duque-Lazo (1), Rubén D Manzanedo (2), Raúl Sánchez-Salgueiro (3), Guillermo Palacios-Rodríguez (1)

We used Species Distribution Modeling to predict the probability of Iberian pine (\textit{Pinus nigra} subsp. \textit{salzmannii} [Dunal] Franco) occurrences in southern Spain in response to environmental variables and to forecast the effects of climate change on their predicted geographical distribution. The ensemble modeling approach “biomod2” was used, together with present Iberian pine data, to predict the current potential distribution based on bioclimatic explanatory variables (200 m resolution) and to forecast future suitability by studying three periods (2040, 2070, and 2100), considering the Global Circulation Models BCM2, CNCM3, and ECHAM5, and the regional model EGAMAM, for different scenarios (SRAB1, SRA2, SRB1). Model evaluation was performed using Kappa, True Skills Statistic (TSS), and Area Under the Curve (AUC) values. The biomod2 approach highlighted the average number of days with a minimum temperature equal to or below 0°C, annual precipitation, and aridity index as the most important variables to describe the \textit{P. nigra} occurrence probability. Model performances were generally satisfactory and the highest AUC values and high stability of the results were given by GAM and GLM, but MaxEnt and the SRE model were scarcely accurate according to all our statistics. The ensemble Species Distribution Modeling of \textit{P. nigra} in Andalusia predicted the highest probability of species occurrence in the eastern areas, Sierra de Cazorla being the area with the highest occurrence of \textit{P. nigra} in Andalusia. In the future habitat, the general probability of \textit{P. nigra} occurrence in Andalusia will decrease widely; the species is expected to lose habitat suitability at moderate altitudes and its occurrence probability will have decreased by nearly 70% on average by 2100, affected by the selected scenario. Populations in Sierra de Cazorla are those most suitable for \textit{P. nigra} growth, even under the most pessimistic scenarios. It is likely that the natural southern populations of \textit{P. nigra} will be very sensitive to changes in climate.

Keywords: Species Distribution Modeling, Climate Change, Ensemble Modeling, Iberian Pine, Mediterranean Relict Forests

Introduction

Tree species distribution is mainly determined by climate and soil (i.e., ecological drivers), despite the effect of historical factors such as the presence of glacial refuges (Svenning & Skov 2007) and the naturalization of species beyond their former range (Pyšek et al. 2014). Projected trends in the context of global change suggest an increase in the mean annual temperatures, drought, and frequency of extreme events in the Mediterranean Basin. Understanding the effects of climate on tree distribution is particularly relevant in relict drought-prone forests; for example, those of southern Iberia (Benito Garzón et al. 2008), which are considered to be highly exposed to climate change, in which both warming and aridification trends have been observed (Giorgi 2006). These recent trends had negative effects on the tree growth (Sánchez-Salgueiro et al. 2012), forest productivity (Madrigal-González & Zavala 2014), and distribution of the southernmost Iberian coniferous forests (Benito Garzón et al. 2008). An improved understanding of the responses to climate of the distribution of relict circum-Mediterranean pine populations provides a valuable system for investigating the role played by ongoing warming at the southern distribution limit of these species, where they usually face severe drought stress.

The Iberian pine (\textit{Pinus nigra} Arn. subsp. \textit{salzmannii} [Dunal] Franco – hereafter abbreviated as \textit{P. nigra} subsp. \textit{salzmannii}) is widely distributed in the western Mediter-
and a few have modeled the changes in habitat suitability of *P. nigra* subsp. *salzmannii* and its future distribution in the Iberian Peninsula (Falcitono et al. 2011, López-Tirado & Hidalgo 2014), but the use of ensemble SDM approaches might improve the accuracy of biodiversity conservation and management policies in relict forests (Porfírio et al. 2014). For *P. nigra* subsp. *salzmannii*, it is at the edge of its distribution and in the highest mountains where we can expect the strongest response to purely climatic conditions. Likewise, for *P. nigra* subsp. *salzmannii* the populations at the southernmost edge of the current distribution represent an important genetic reservoir for the future conservation of the species under increasing drought and aridity (Soto et al. 2010). The genetic characteristics of *P. nigra* subsp. *salzmannii* might help to conserve the species at higher latitudes, even in areas where it already occurs but is not genetically adapted to the forecasted change in climate conditions (Herrero et al. 2013, López-Tirado & Hidalgo 2014). Moreover, in situ conservation and biodiversity management require information on plant species distribution at a local scale (Van Gils et al. 2012).

The modeling of the distribution of endangered species, like *P. nigra* subsp. *salzmannii*, is especially challenging because of the lack of data and the high level of human disturbance in some of the populations (Elith & Franklin 2013). In this paper, we have used ensemble SDM of habitat to explore current and future impacts of climate change on relict populations of Iberian pine in the southern Iberian Peninsula. We used the biomod2 R-package (Thuiller et al. 2012), which combines 10 different model algorithms and calculates a final ensemble prediction. We aimed to identify the major environmental variables explaining black pine climate suitability in the southernmost Iberian forests, and then to explore the potential responses of the species to predicted changes in climate using three intercontinental Global Circulation Models and one regional model, for three different scenarios, at a very fine scale (200 meters resolution). In addition, we discuss whether the factors driving the occurrence of this tree species in the southern Iberian Peninsula are similar to those reported previously for other Iberian relict regions.

**Material and methods**

**Study areas**

The study areas are located in Andalusia, southern Spain, a characteristic region of the Mediterranean Basin located in the south of the Iberian Peninsula (Fig. S1 in Supplementary material). The climate in this region is typically Mediterranean, with hot and dry summers and precipitation concentrated in spring and autumn. Forests of *P. nigra* subsp. *salzmannii* are located in the highest forested range of the Baetic Mountains, characterized by great topographical variability and a high degree of natural disturbances, including wild fires and, often, non-sustainable forest management. Andalusia contains most of the natural stands of Iberian pine in its southernmost populations, with the exception of those located in Morocco and Algeria (Camarero et al. 2013). The natural *P. nigra* subsp. *salzmannii* forests in eastern Andalusia are dominant in north-oriented and high-elevation sites (1500-2700 m a.s.l.) and are highly diverse (ca. 107,000 ha). There are also extensive plantations (ca. 40,000 ha) of *P. nigra* subsp. *salzmannii*, which were mostly planted in mid to high-elevation sites (1200-2510 m a.s.l.) between the 1960s and 1970s. We did not consider these plantations in this study because they may be established outside their bioclimatic range (López-Tirado & Hidalgo 2014).

**Location data and environmental variables**

Information on species occurrences and bioclimatic variables was compiled from the Andalusian Regional Government database (REDIAM 2010). We used 977 presence records (points) of natural *P. nigra* subsp. *salzmannii* forests from the third National Spanish Forest Inventory (Villanueva 2005). We selected the presence records within natural populations because natural *P. nigra* subsp. *salzmannii* stands presented a more plastic response to climate change and more severe decline have been associated to planted trees (Sanchez-Salgueiro et al. 2013), which suggest that some of *P. nigra* subsp. *salzmannii* plantations might be outside of its bioclimatic range. An equal number of pseudo-absences (Prevalence equal to 0.50) with 10 replicates (Barbet-Massin et al. 2012) within the study area were generated using the biomod2 R package (Thuiller et al. 2013).

The climate data consisted of 29 environmental variables (Tab. S1 in Supplementary material) averaged for the period 1960-2000 and the future predictions were carried out for three periods: 2011-2040; 2041-2070, and 2071-2099, following regional scenarios (REDIAM 2010). For each period, we considered three intercontinental Global Circulation Models (BCCM2, CNCM3, and ECHAM5) and one regional model (EGHAM), developed in the project Climate Change Scenarios updated in the 4th IPCC report, as well as three climate change scenarios (A1B, A2, and B1 – IPCC 2014), to take into account a high variability in future projections. The complete dataset had a spatial resolution of 200-meter grids. In all our analyses, we used average period values.

**Variable reduction and importance**

The initial 29 environmental variables were reduced by the variable inflation factor (VIF<10 – Quinn & Keough 2002), principal component analysis (PCA), and the relative quality of the statistical models mea-
asured by Akaike's Information Criterion (AIC) values. We took the non-collinear variables as the means of the first two principal components (Franklin 2010), and from those we selected the final variables based on their importance in predicting the climate suitability, as judged by AIC.

Variable importance was quantified using the Variable Importance function within biomod2, which is implemented by calculating the Pearson’s correlation (r) between model predictions including all remaining variables (a “full model”) and predictions excluding the variable being tested (a “reduced model”). If a variable has a low contribution to a model, the two outputs would be similar, and (1-r) would be low, while the opposite would be the case for important variables (Thullier et al. 2013). This variable importance procedure does not allow the comparison of importance between models. Therefore, we used the ranking system described in Syphard & Franklin (2009) to compare variable selection between model types.

**Statistical models**

To deal with model selection, we used ensemble models from 10 SDM algorithms (Thullier et al. 2013), employing the biomod2 package (R Core Development Team 2014). This package allows one to perform ensemble SDM using presence-absence data that includes: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Flexible Discriminate Analysis (FDA), Artificial Neural Networks (ANN), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy Modelling (MaxEnt), Random Forest (RF) and Boosted Regression Trees (BRT) and one similar to BIOCLIM, Surface Range Envelop (SRE).

We used ensemble models calculated using the mean, median, coefficient of variation, upper and lower confidence interval, committee averaging and probability mean weight decay of the single model predictions (Thullier et al. 2013). The consensus methods employed in this study make up a representative sample of the most commonly used techniques. Four methods (Median, Mean, Coefficient of Variation, and Confidence Intervals) are based on their statistical functions, whereas the Probability Mean Weight Decay (WD) and Committee Averaging (CA) methods preselect the single models based on certain predefined criteria. In WD, half of the single-model outputs are preselected on the basis of the True Skills Statistic (TSS) values. The selected single models are combined using an averaging function. In CA the probabilities of the selected models are first converted into binary, according to the maximal TSS threshold, and the CA score is calculated as the average of the binary predictions (Thullier et al. 2013).

**Model calibration and validation**

We randomly divided our dataset into two subsets, which we used to train the model (70% of the data) and to evaluate it independently (the remaining 30%). This yielded 100 different fits per model, each with its own accuracy indicator, mean, and quantiles of model accuracy calculated. Models with higher mean values and smaller variations were considered as being the most accurate ones (Duque-Lazo et al. 2016).

Model performance was evaluated by the AUC of the Receiver Operating Characteristic plot, Kappa, and TSS. The AUC represents the models’ discriminative capacity with regard to the data and is obtained by plotting the commission error (1-specificity; false positives) on the horizontal axis, versus the omission error (sensitivity; actual positives which are correctly identified as such) on the vertical axis, for numerous thresholds. The AUC ranges between 0 (opposite prediction) and 1 (perfect prediction) with 0.5 = random, where 1 represents a perfect discrimination between presence and absence and 0.5 represents a random fit. As an evaluation metric, AUC has the advantage of being prevalence and threshold independent (Franklin 2010).

Cohen’s Kappa (k) was derived from the 2 x 2 confusion matrix to measure the rate of agreement between actual and predicted values in the spatial space for categorical Kappa values; however, the matrix depends on the defined threshold for presence. Values of k near to 0.5 indicate no discrimination capacity (random agreement), whereas a value of 1 represents the perfect discrimination model (Cohen 1960). The TSS is concerned with omission and commission errors and is also prevalence independent. It ranges from -1 to +1, where +1 indicates perfect agreement and 0 indicates a random performance. The TSS measures the difference between the actual agreement and the randomly expected agreement; it is particularly useful for the modeling of rare species with limited point locations and it can be used to compare different modeling techniques. The TSS is defined as (Allouche et al. 2006 – eqn. 1):

\[
TSS = \text{sensitivity} \times \text{specificity} - 1
\]

Finally, Kappa and TSS were calculated and evaluated, considering a threshold equal to prevalence (Vale et al. 2014).

**Distribution maps**

We generated continuous probabilistic maps with values of between 0 and 1 for each grid point, for the present and future habitat distributions of *P. nigra* subsp. *salzmannii* in Andalusia. In order to assist the visual interpretation of the model predictions, the probability values were classified into four categories (0-25%; 25-50%; 50-75%; 75-100%). To calculate the shifts in the range of the surface area from the present to the future distributions, we reclassified the probabilities as 0 (unsuitable habitat) and 1 (suitable habitat); the methods used to estimate the threshold were the same as those employed to calculate TSS and...
Kappa. Finally, we focused on six different relict areas (Andalusia, Sierra Cazorla, Sierra María, Sierra de Baza, Sierra Nevada, and Sierra de las Nieves – Fig. 1) to evaluate the effect of climate change on each population.

Results

Variable reduction

The results of the variable selection highlighted that 17 predictors out of 29 were affected by collinearity problems (VIF>10), of which eight were selected with high correlations with the first two principal components of the PCA and six were highlighted by AIC model selection. The final set of the six variables selected included: the average reference evapotranspiration (ETO); aridity index (IAR); average number of days with a minimum temperature equal to or below 0 °C (NDF); annual precipitation (PRC); aridity index (IAR); annual sum of the positive differences between precipitation and reference evapotranspiration (SSUP); and average snow precipitation (SNOW – Tab. 1). Moreover, the biomd2 ranked classification of variable importance (Tab. 2) highlighted NDF as the most important variable, followed by PRC and IAR. The single scorings of variable importance from the separate model techniques generally agreed with this classification, reinforcing our variable selection (Tab. 2).

Our results show that the probability of *P. nigra* subsp. *salzmannii* occurrence increases with PRC and NDF, while it decreases with increases in ETO and SSUP. We also observed slight responses to aridity and snow cover (Fig. 1). Locations with low annual precipitation (PRC < 915 mm) or high transpiration losses (ETO > 908 mm or SSUP > 719 mm) were highly unsuitable for *P. nigra* subsp. *salzmannii* growth. Surprisingly, we observed a weak response to aridity; low aridity values maximized the occurrence probability but the response was otherwise rather flat. Similarly, snow cover did not significantly affect the occurrence probability (Fig. 1).

Model selection and validation

The single model predictions were compared by their accuracy as given by TSS, Kappa, and AUC; overall, these showed good model accuracy (Fig. 2). The model performances were generally satisfactory and the highest AUC values and lowest variances were shown by GAM and GLM, although they had lower TSS and Kappa values than ANN, BRT, and RF, which were also highly accurate. Although CART, FDA, and MARS possessed low predictive power and stability, they exhibited high (CART) or moderate (FDA and MARS) correlation with other models. However, the MaxEnt and SRE models were scarcely accurate according to all our statistics (Fig. 2). The bivariate Pearson correlation analysis highlighted the low correlation of the MaxEnt and SRE predictions with those of the rest of the model techniques used. On the

### Tab. 1 - Variable importance ranking and parametric characterization of *P. nigra* subsp. *salzmannii* for 1000 runs and 10 habitat prediction model techniques. Variable abbreviation is indicated in parentheses. (Qtl): quartile.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variable</th>
<th>Min</th>
<th>1st Qt</th>
<th>Mean</th>
<th>3rd Qt</th>
<th>Max</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average number of days with a minimum temperature equal to or below 0 °C (NDF)</td>
<td>35.30</td>
<td>65.40</td>
<td>76.30</td>
<td>90.00</td>
<td>158.90</td>
<td>days</td>
</tr>
<tr>
<td>2</td>
<td>Annual precipitation (PRC)</td>
<td>306.00</td>
<td>558.00</td>
<td>760.00</td>
<td>919.00</td>
<td>1513.00</td>
<td>mm</td>
</tr>
<tr>
<td>3</td>
<td>Aridity index (IAR)</td>
<td>35.00</td>
<td>94.00</td>
<td>116.00</td>
<td>160.00</td>
<td>296.00</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Annual sum of the positive differences between precipitation and reference evapotranspiration (SSUP)</td>
<td>1.00</td>
<td>142.00</td>
<td>312.00</td>
<td>453.00</td>
<td>1046.00</td>
<td>mm</td>
</tr>
<tr>
<td>5</td>
<td>Average reference evapotranspiration (ETO)</td>
<td>92.00</td>
<td>840.00</td>
<td>891.00</td>
<td>934.00</td>
<td>1087.00</td>
<td>mm</td>
</tr>
<tr>
<td>6</td>
<td>Average snow precipitation (SNOW)</td>
<td>27.00</td>
<td>134.00</td>
<td>266.00</td>
<td>446.00</td>
<td>1170.00</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>- Aspect (ASPECT)</td>
<td>-453.00</td>
<td>-27.00</td>
<td>8.00</td>
<td>49.00</td>
<td>333.00</td>
<td>degree</td>
</tr>
<tr>
<td></td>
<td>- Sum of water balances at the end of each month (BH)</td>
<td>1.00</td>
<td>685.00</td>
<td>1894.00</td>
<td>3011.00</td>
<td>7707.00</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>- Digital elevation model (DEM)</td>
<td>7.00</td>
<td>1164.00</td>
<td>1423.00</td>
<td>1705.00</td>
<td>2995.00</td>
<td>m</td>
</tr>
<tr>
<td></td>
<td>- Average net primary production (DF)</td>
<td>182.00</td>
<td>1223.00</td>
<td>1928.00</td>
<td>2471.00</td>
<td>3371.00</td>
<td>hours</td>
</tr>
<tr>
<td></td>
<td>- Average number of days with a maximum temperature equal to or above 35 °C (NDC)</td>
<td>0.10</td>
<td>5.70</td>
<td>10.20</td>
<td>16.20</td>
<td>44.40</td>
<td>days</td>
</tr>
<tr>
<td></td>
<td>- Annual radiation (RN)</td>
<td>26.00</td>
<td>95.00</td>
<td>103.00</td>
<td>107.00</td>
<td>126.00</td>
<td>Julian m²</td>
</tr>
<tr>
<td></td>
<td>- Annual sum of the negative differences between precipitation and reference evapotranspiration (SDEF)</td>
<td>84.00</td>
<td>394.00</td>
<td>440.00</td>
<td>506.00</td>
<td>657.00</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>- Slope (SLOPE)</td>
<td>1.00</td>
<td>11.00</td>
<td>17.00</td>
<td>24.00</td>
<td>100.00</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>- Average maximum temperature (T_MAX)</td>
<td>12.20</td>
<td>16.60</td>
<td>17.60</td>
<td>18.60</td>
<td>22.20</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Average mean temperature (T_MED)</td>
<td>4.10</td>
<td>10.20</td>
<td>11.20</td>
<td>12.20</td>
<td>15.60</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Average minimum temperature (T_MIN)</td>
<td>-4.50</td>
<td>3.70</td>
<td>5.00</td>
<td>6.10</td>
<td>9.50</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Maximum of the monthly average maximum temperatures (TMAX)</td>
<td>24.40</td>
<td>28.80</td>
<td>30.00</td>
<td>31.10</td>
<td>34.70</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Average maximum temperature of all months (TMC)</td>
<td>14.30</td>
<td>20.20</td>
<td>21.20</td>
<td>22.50</td>
<td>26.00</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Average minimum temperature of all months (TMIN)</td>
<td>-1.80</td>
<td>3.10</td>
<td>4.00</td>
<td>4.90</td>
<td>8.50</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>- Minimum of the monthly average minimum temperatures (TMINF)</td>
<td>-8.40</td>
<td>-2.10</td>
<td>-1.00</td>
<td>-0.10</td>
<td>3.30</td>
<td>°C</td>
</tr>
</tbody>
</table>

### Tab. 2 - Mean variable importance values for 100 runs, for each selected variable and the habitat prediction model techniques. Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Flexible Discriminate Analysis (FDA), Artificial Neural Networks (ANN), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy Modeling (MaxEnt), Random Forest (RF) and Boosted Regression Trees (BRT) and one similar to BIOClim, Surface Range Envelop (SRE). Variable description in Tab. 1. (*): P < 0.05; (**): P < 0.01.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ANN</th>
<th>BRT</th>
<th>CART</th>
<th>FDA</th>
<th>GAM</th>
<th>GLM</th>
<th>MARS</th>
<th>MaxEnt</th>
<th>RF</th>
<th>SRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETO</td>
<td>0.19</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.30</td>
<td>0.18</td>
<td>0.28</td>
<td>0.01</td>
<td>0.03</td>
<td>0.31</td>
</tr>
<tr>
<td>IAR</td>
<td>0.07</td>
<td>0.01</td>
<td>0.03</td>
<td>0.06</td>
<td>0.12</td>
<td>0.04</td>
<td>0.13</td>
<td>0.01</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>NDF</td>
<td>0.73**</td>
<td>0.83**</td>
<td>0.90**</td>
<td>0.96**</td>
<td>0.67*</td>
<td>0.82**</td>
<td>0.73*</td>
<td>0.75*</td>
<td>0.58*</td>
<td>0.57*</td>
</tr>
<tr>
<td>PRC</td>
<td>0.56*</td>
<td>0.00</td>
<td>0.03</td>
<td>0.98**</td>
<td>0.99**</td>
<td>0.80**</td>
<td>0.97**</td>
<td>0.01</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td>SNOW</td>
<td>0.24</td>
<td>0.00</td>
<td>0.05</td>
<td>0.38</td>
<td>0.39</td>
<td>0.32</td>
<td>0.39</td>
<td>0.00</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>SSUP</td>
<td>0.07</td>
<td>0.00</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.07</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>
other hand, it is interesting to note that the models, in general, presented higher correlations (Tab. 3). The classification (BRT, CART, and RF) and regression method (GAM and GLM) predictions were highly correlated with each other. Moreover, ANN, FDA, and MARS presented acceptable correlations with the rest of the models (Tab. 4).

All ensemble models presented TSS>0.77, Kappa>0.90, and AUC>0.98, the ensemble models committee averaging (CA) and probability mean weight decay (WD) being the best ones (Tab. 4). In this table, we have excluded the results corresponding to the ensemble models in accordance with the coefficient of variance result appearing in Fig. 3, with variability between all model predictions.

Fig. 2 - Adjustment values obtained for 10 different distribution models in Andalusia, including Artificial Neural Networks (ANN), Boosted Regression Trees (BRT), Classification and Regression Trees (CART), Flexible Discriminate Analysis (FDA), Generalized Additive Models (GAM), Generalized Linear Models (GLM), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MaxEnt), Random Forest (RF), and Surface Range Envelop (SRE).

Tab. 3 - Pearson’s correlation matrix of the concordance of the different spatial distribution models analyzed: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Flexible Discriminate Analysis (FDA), Artificial Neural Networks (ANN), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy Modelling (MaxEnt), Random Forest (RF) and Boosted Regression Trees (BRT) and one similar to BIOCLIM, Surface Range Envelop (SRE). All correlation were significant with p<0.001 (2-tailed).

<table>
<thead>
<tr>
<th>Model</th>
<th>ANN</th>
<th>BRT</th>
<th>CART</th>
<th>FDA</th>
<th>GAM</th>
<th>GLM</th>
<th>MARS</th>
<th>MaxEnt</th>
<th>RF</th>
<th>SRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BRT</td>
<td>0.885</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CART</td>
<td>0.843</td>
<td>0.966</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FDA</td>
<td>0.890</td>
<td>0.888</td>
<td>0.865</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GAM</td>
<td>0.920</td>
<td>0.832</td>
<td>0.782</td>
<td>0.856</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GLM</td>
<td>0.891</td>
<td>0.878</td>
<td>0.841</td>
<td>0.862</td>
<td>0.971</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MARS</td>
<td>0.831</td>
<td>0.876</td>
<td>0.828</td>
<td>0.844</td>
<td>0.873</td>
<td>0.875</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.821</td>
<td>0.881</td>
<td>0.856</td>
<td>0.751</td>
<td>0.762</td>
<td>0.799</td>
<td>0.872</td>
<td>1.000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RF</td>
<td>0.921</td>
<td>0.950</td>
<td>0.912</td>
<td>0.867</td>
<td>0.851</td>
<td>0.889</td>
<td>0.889</td>
<td>0.801</td>
<td>1.000</td>
<td>-</td>
</tr>
<tr>
<td>SRE</td>
<td>0.603</td>
<td>0.666</td>
<td>0.651</td>
<td>0.623</td>
<td>0.588</td>
<td>0.601</td>
<td>0.620</td>
<td>0.685</td>
<td>0.642</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Tab. 4 - Adjustment values obtained with the ensemble models of Pinus nigra habitat prediction in Andalusia: Cohen’s Kappa, Area Under the Curve (AUC), True Skills Statistic (TSS), sensitivity (true positive rate), and specificity (true negative rate).

<table>
<thead>
<tr>
<th>Ensemble model</th>
<th>Kappa</th>
<th>TSS</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.786</td>
<td>0.904</td>
<td>0.986</td>
<td>0.9725</td>
<td>0.9313</td>
<td>0.869</td>
</tr>
<tr>
<td>Lower Confident interval</td>
<td>0.786</td>
<td>0.904</td>
<td>0.986</td>
<td>0.9714</td>
<td>0.9332</td>
<td>0.846</td>
</tr>
<tr>
<td>Upper Confident interval</td>
<td>0.784</td>
<td>0.904</td>
<td>0.986</td>
<td>0.9735</td>
<td>0.9311</td>
<td>0.881</td>
</tr>
<tr>
<td>Median</td>
<td>0.774</td>
<td>0.903</td>
<td>0.983</td>
<td>0.9724</td>
<td>0.9307</td>
<td>0.913</td>
</tr>
<tr>
<td>Committee averaging</td>
<td>0.802</td>
<td>0.905</td>
<td>0.987</td>
<td>0.9756</td>
<td>0.9298</td>
<td>0.800</td>
</tr>
<tr>
<td>Probability mean weight decay</td>
<td>0.786</td>
<td>0.904</td>
<td>0.986</td>
<td>0.9725</td>
<td>0.9319</td>
<td>0.806</td>
</tr>
</tbody>
</table>
predictions of below 25%, which means that, mostly, all the models agree with the predictions. Pixels with a high coefficient of variance are recognized as being the uncertain pixels predicted.

**Black pine present and future habitat suitability**

The ensemble SDM of *P. nigra* subsp. *salzmannii* in Andalusia predicted the highest probability of occurrence over the eastern areas (Sierra Cazorla, Sierra María, Sierra de Baza, and Sierra Nevada – Fig. 3), with a lesser presence in the mid-south (Sierra de las Nieves), which largely agrees with the current distribution of the species. The three ensemble models selected (CA, WD, and mean) showed a high degree of agreement in their predictions (Fig. 4). Sierra de Cazorla appears as the area with the highest *P. nigra* subsp. *salzmannii* occurrence within Andalusia, with a total surface area coverage of 50% (Fig. 3, see also Tab. S2 in Supplementary material) in its current prediction.

In the forecasted habitat suitability, the general probability of *P. nigra* subsp. *salz-
mannii occurrence in Andalusia widely decreases (Fig. 4, see also Tab. S2, Fig. S2 in Supplementary material), and the species is expected to lose habitat suitability at moderate altitudes, probably extending to higher ones as climate change proceeds (Fig. 4, Fig. S2). The southern areas (Sierra de las Nieves) disappear and the probability of species occurrence decreases, compared to the present predictions, for the eastern populations. Overall, all the populations showed a strong decrease in habitat suitability across the different scenarios (Fig. 4, Tab. S2 and Fig. S2 in Supplementary material). The predictions were significantly affected by the scenario, especially for longer periods of time (Fig. S2). The probability occurrence of P. nigra subsp. salzmannii will have diminished by nearly 70% on average by 2099 (Fig. 4, Tab. S2 and Fig. S2). It is worth mentioning that, potentially, Sierra de Cazorla will remain the most suitable area for P. nigra subsp. salzmanni growth, even under the most pessimistic scenarios.

The future predictions forecast a reduction in the potential habitat of P. nigra subsp. salzmannii, of 50% in 2040 and 99% in 2099 (Tab. S2 and Fig. S2 in Supplementary material). These results varied depending on the GCM used and the scenario, while the ensemble modeling technique used had a minor effect. The more optimistic results predict a habitat suitability reduction close to 40% (BCM2, SRB1) by 2099, while others predict a loss of over 99% (EGMAM, SRA2) in the same period. In general, the regional GCM (EGMAM) presented predictions more pessimistic than those of the intercontinental ones (Tab. S2 and Fig. S2). The most optimistic results were given by CCM3 and BCM2; the potential area remaining exceeded 20% in some cases.

Discussion

Species distribution models are a simple yet efficient statistics-based method used to map the spatial range of species and to forecast climate change impacts on species ranges (Franklin 2010, Elith & Franklin 2015). They have been used for many different purposes, including: conservation management, theory testing in biogeography and ecology, species management, and climate change impact assessment (Franklin 2010). In this work, we used an integrated model approach to assess the spatiotemporal patterns of suitable habitats for P. nigra subsp. salzmannii in the southern Iberian Peninsula, under four climate models. Biomod2 has shown the simplicity involved in the modelling, which enables its use for model distribution shifts resulting from climate change. In particular, our study constitutes a new approach with respect to other models previously used for predicting the distribution of the species (Attorre et al. 2011, López-Tirado & Hidalgo 2014, Marchi et al. 2016), due to its focus on a relevant regional-scale design and on the populations most sensitive to climate change. One of the remarkable characteristics of our approach is the applicability of the results in biodiversity conservation and forest management (Hampe & Petit 2005).

Current potential habitat description

The six main environmental variables used, listed in order of importance, were: NDF, FRC, IAR, SSUP, ETO, and SNOW (Tab. 1). NDF had the strongest influence on P. nigra distribution, in agreement with other studies that stressed the key role of low temperatures to explain the geographic distribution of P. nigra in Mediterranean areas (Gandullo & Palomares 1994, Elena-Rosselló & Sánchez-Palomares 1991). P. nigra is benefited by an increase in the NDF, particularly in dry Mediterranean mountain areas, where it can compete with other co-existing coniferous species which are more sensitive to drought (P. sylvestris and P. pinaster) or to low temperatures (P. halepensis). This geographic distribution has traditionally been related to a relatively short vegetative period, high moisture, and a high degree of continentality (Elena-Rosselló & Sánchez-Palomares 1991), as well as to variables that are negatively related to the minimum average temperature (Sanchez-Salgueiro et al. 2013). This agrees with the experimental findings of Gandullo & Palomares (1994), who reported an optimal distribution for this species on northern slopes above 1000 m a.s.l., with mean annual temperatures ranging between 6.1 and 10.5°C. The dependence of P. nigra subsp. salzmannii on the annual precipitation shows its high sensitivity to annual and seasonal rainfall oscillations. The greatest potential suitability for the species has been reported in areas with annual rainfall of over 950 mm, with minimum summer precipitation of over 25-30 mm (Gandullo & Palomares 1994, López-Tirado & Hidalgo 2014). Consequently, the species requires a certain degree of soil moisture, although it is also one of the few tree species that resists the most severe winter conditions (SNOW) in southern Spain as well as very dry summers (IAR), frequently found in Mediterranean high mountains supporting a snowpack as an important compensation feature for dry conditions. Under such conditions (mountain climate at high altitudes), P. nigra subsp. salzmannii is highly competitive and is one of the main forest species in the Mediterranean region (Herrero et al. 2013).

Model selection

The biomod2 software has been used to predict the habitat of forest species of different forms and parameterizations (Duque-Lazo et al. 2016). In this work, we have demonstrated differences in prediction performance between modeling methods at a regional scale. In general terms, the results agree with those obtained from the application of biomod2 to the prediction of the distribution of forest species at a local and regional scale (López-Tirado & Hidalgo 2014). The linear methods (GLM and GAM) surpassed in accuracy the rest of the models used. Simple linear relationships using the selected climate variables and occurrence point locations could explain the highly specific habitat suitability of P. nigra subsp. salzmannii. The classification methods (BRT, CART, and RF), which are based on multiple data partitioning and increase the model’s complexity, also gave accurate relationships, supporting the linear relationship between the habitat suitability of P. nigra and the environmental variables. The methods ANN, FDA, and MARS yielded accurate responses, although with a high variance; such responses might be due to chance or proper set data through cross-validation. MaxEnt showed a surprisingly low predictive power and stability for the three statistics studied, despite being one of the most widely used modelling techniques and being able to build linear relationships between response and explanatory variables (Duque-Lazo et al. 2016). Finally, the power of SRE in the modelling of species distribution, in comparison with other methods, has been doubted. In this study, the used ensemble models differed in terms of future predictions, depending on the GCM studied. The ensemble model calculated by WD presented a higher over all total area for the current prediction than the ensemble modeling calculated by CA. Moreover, the percentage area loss was also higher for the ensemble model calculated by WD. These results could be due to an over-prediction of the current area, the difference being marked in the Sierra Nevada study area with 500 ha more in the ensemble model calculated by WD than for CA (Tab. 1, Tab. 2).

Future habitat projection

Recent studies have predicted a reduction in P. nigra subsp. salzmannii habitat suitability and tree growth in the next few decades in southern Europe (Sanchez-Salgueiro et al. 2012, López-Tirado & Hidalgo 2014, Bede-Fazekas et al. 2014, Marchi et al. 2016), related to an increase in temperatures. According to our results, future temperature increases could affect the distribution of P. nigra subsp. salzmannii in southern Spain (Camarero et al. 2013), and that shifts to higher altitudes rather than higher latitudes are expected in the future (Navarro-Cerrillo et al. 2014). Habitat models showed the sensitivity of this species to minimum temperatures and precipitation, in agreement with previous studies in the Mediterranean region indicating these as crucial factors determining, for example, Quercus ilex and Q. faginea forests (Feliciísimo et al. 2011).

Our results suggest a dramatic reduction in locations representing a suitable habitat for P. nigra subsp. salzmannii due to climatic transitions: by 24.8% (ECHAM5-SRA2) to 52.8% (EGMAM-SRA1B) of the extent of
their current distributions (as simulated for the present climate) in 2040, and by 55-5% (BCM2-SRB1) to <99% (ECHAMS-5RA2, ECMWF-MSC-RA2 and SBD1) in 2099 (Tab. 4). This trend, namely the loss of suitable habitats, was consistent among climate scenarios and populations.

Alternatively, the populations could extend northwards to central Spain in response to future conditions, and we assumed here that the species fully achieves its potential changes in distribution (meaning that we did not assume any dispersal limitation). However, and under a more restricted and optimistic assumption. The fact that P. nigra subsp. salzmannii is a mountain species whose populations are highly isolated, with scant mountain ecosystems connecting the northern and central Iberian Peninsula populations with the southernmost ones, suggests that this scenario would be unlikely (Candel-Pérez et al. 2012). Therefore, P. nigra subsp. salzmannii may only persist in the long term as residual populations derived from its present distribution. Also, there is a great threat of local extinction, especially for those populations with a more restricted distribution (Sierra de Baza, Sierra de La Almijara, and Sierra de María – Camarero et al. 2013, Navarro-Cerrillo et al. 2014). Regarding altitude migration, although there are well-documented examples of populations migrating to altitudes higher than those at which they occur today as an apparent response to the ongoing climate change, the speed of that migration is much slower than would be needed to track the changing climate (Le- noir et al. 2008). In the case of P. nigra subsp. salzmannii in southern Spain, an upward migration of 300-400 m would be required to compensate for the change in climate expected for the year 2040 as predicted, for instance, by the A2 scenario of the Canadian GCM (Candel-Pérez et al. 2012). Forest mortality may be an observable response to these changes, as suggested by previous studies where dendroclimatological and ecological data showed the critical and marginal situation of P. nigra subsp. salzmannii populations (Linares & Tisca 2010, Pietramellati et al. 2012, Sanchez-Salgueiro et al. 2013).

However, projections of potential future distributions also need to be interpreted with caution. Even if the models presented in this study are quite accurate and are commonly used to assess the impact of global change (Thuiller 2014), we have not accounted here for biotic interactions that may modify habitat suitability. For example, positive interactions with mycorrhizal fungi may enable trees to withstand drier conditions without any significant growth reduction (Collins Johnson et al. 2010). Furthermore, trees are incredibly long-lived and resilient (Petit & Hampe 2006) and therefore we can expect individuals to persist under sub-optimal conditions for long periods of time (Sanchez-Salgueiro et al. 2012), while waiting for infrequent benign conditions to be reproduced. Therefore, an important aspect that must be considered is forest monitoring. Currently, only time and empirical data will give us a real idea about the reliability of such models (Kautz et al. 2017).

Our results have serious implications for species adaptation and forest management (Camarero et al. 2013, Sanchez-Salgueiro et al. 2015). If this is coupled with the loss of a critical stopover site, the results for P. nigra subsp. salzmannii populations in southern Spain might be catastrophic. Some populations, of course, may be able to migrate (Sierra de Cazorla) to higher latitudes or altitudes in the face of climate change; but other populations would seem to face local extinction. The Sierra de Cazorla populations showed the greatest overlap between the potential and future suitable distributions, highlighting this area as a key genetic resource for the in situ conservation of P. nigra forests in the southern Mediterranean region (Tiscar & Lucas-Borja 2016).

Conclusion
Given the accumulating evidence of recent climate changes, an evaluation of the current and future distributions of P. nigra subsp. salzmannii in Andalusia (southern Spain) is described in this paper. This information is necessary to ensure that the most threatened populations of such species will be afforded protection in response to climate change, by adjusting their protection status and forest management. Our results show an important contraction of the distribution area for all GCM models, with a minimum potential reduction of 40% (BCM2, SBD1) in 2099. The changing climate will inevitably result in impacts on biomes and community structures. Thus, its mitigation, as well as adaptation to potential future scenarios, is vital to the conservation of climate-sensitive species. Future research that combines bioclimatic niche modeling with a mechanistic disturbance, dispersal, and competition model will be likely to provide a greater insight into the potential range of P. nigra subsp. salzmannii in the face of climate change. Furthermore, it would supply information that influenced forest management options and restoration – that may include ecological restoration, selected thinning, or assisted migration. Some alternatives would ensure “fine-grained” landscapes with patches of a diverse range of semi-natural habitats – including woodlands, heathlands, and shrubs – that offered suitable habitat niches to sustain disturbance responses to changing environmental conditions. Additionally, the distribution of protected areas should be revised to include extensive areas of natural or semi-natural habitats of coniferous Mediterranean forests, in particular plantations such as artificial P. nigra subsp. salzmannii forests, which offer a diverse range of physical habitats.

Acknowledgements
We acknowledge the financial support of ESPECTRED (CGL2017-8616-R), Universidad de Córdoba – Campus de Excelencia CEIA, and AEMET (Agencia Estatal de Meteorología) for providing meteorological data. R. Sánchez-Salgueiro has been supported by a postdoctoral fellowship (FEDER - Programa de Formación en I+D+i de las Universidades 2014-2015 de la Junta de Andalucía) and the “CoMoReAdapt” Spanish project (CGL2013-48843-C2-1-R). The authors also thank Dr. Miguel B. Araujo for his valuable comments at the beginning of this work.

Authors’ contributions
RMN conceived the study, analyzed data, performed research, and wrote the paper; JDL analyzed data and contributed new methods; RDM conceived the study and performed research; RSS and GPR wrote the paper.

References


Supplementary Material

Tab. S1 - Environmental data used for the habitat prediction of P. nigra subsp. salzmannii occurrence in Andalusia (Spain).
Tab. S2 - Total (ha) and percentage (in parentheses) area loss, relative to the values for 1961-2000, of Pinus nigra future projections (2040, 2070, and 2099) with different scenarios (SRA2, SRA1B, and SRB1), three Global Circulation Models (CCM: BCM2, CNCM3, ECHAM5), and one Regional Circulation Model (EGMAM).

Fig. S1 - Iberian (a), Spanish (b), and Andalusian (c) distribution of P. nigra subsp. salzmannii.

Fig. S2 - Future P. nigra subsp. salzmannii probability of occurrence prediction obtained by probability mean weight decay ensemble modeling with the Global Circulation Model EGMAM, the scenarios SRA2, SRB1, and SRA1B, and the future projections 2040, 2070, and 2099.

Link: Navarro-Cerrillo_2588@suppl001.pdf