Applied Vegetation Science

Modelling Buxus balearica distribution in Southern Spain.

Journal:	*Applied Vegetation Science*
Manuscript ID:	Draft
Manuscript Category:	Ordinary article
Date Submitted by the Author:	
Complete List of Authors:	Navarro, Rafael; University of Córdoba, Forestry Hernandez-Bermejo, Esteban; University of Córdoba, Agronomy Hernandez-Clemente, Rocío; University of Córdoba, Forestry
Keywords:	Environmental gradient, Habitat distribution model, Mediterranean shrubs, Maxent, Conservation

Modelling *Buxus balearica* distribution in Southern Spain.

Navarro-Cerrillo, R. M[1*], Hernendez-Bermejo, E[2]; Hernandez-Clemente, R[1],

[1] Forestry Department, School of Agriculture and Forestry, University of Córdoba, Edf. Leonardo da Vinci, Campus de Rabanales s/n, Mail Box 3048 (14071, Córdoba, Spain);

[2] Agricultural and Forestry Science Department, School of Agriculture and Forestry, University of Cordoba, Avda 10 Edificio Mutis (C4 2ª), Campus de Rabanales (14071, Cordoba, Spain).

*Corresponding author: e-mail: rmnavarro@uco.es; telephone: 34-957-218657 fax: 34-957-218563.

Abstract

Question: A comparison between predictive models to study the habitat distribution of *Buxus balearica* Lam. at a local scale.

Location: The distribution area of *Buxus balearica* studied was the Málaga and Granada populations (Southern Spain), which includes almost its total Iberian Peninsula distribution (36º and 38º N). The study area was restricted to basic soils, mainly marls and chalky ones, which occupied a surface of 38,180 km^2.

Methods: The prediction models were tested based on 17 environmental variables. Six methods were compared: Multivariate adaptative regression spline (MARS), Maximum entropy approach to modelling species' distributions (Maxent), two generic algorithms based on environmental metrics dissimilarity (Bioclim and Domain), Algorithm for Rule-set Prediction (GARP), and supervised learning methods based on generalized linear classifiers (Support vector machines-SVMs). As a test of the predictive power of the models we used *Kappa* index.

Results: Maxent showed the better predictive accuracy, following by GARP models. All the other models tested obtained lower accuracy values. By comparing the predictive power of Models, climate variables showed the highest contributions of the environmental variables studied. The variables with lowest contributions were the isolation models. A test of sensitivity to the reduction in the number of variables obtained an accuracy of over 0.90 by applying just 3 climatic variables (spring rainfall, mean temperature of warmest month, and mean temperature of coldest month). All algorithms produced maps that coincided well with the known distribution of the species.

Conclusions: Maxent has showed the better habitat-specific predictive mapping model of the *Buxus balearica* communities in southern Spain.

Keywords: Environmental gradient; Habitat distribution model; Mediterranean shrubs.

1. Introduction

The genus *Buxus* is represented in Spain by two species, *B. sempervirens* L. and *B. balearica* Lam. *B. balearica* is an evergreen shrub, rarely small trees, monoecious, more robust and bigger than *Buxus sempervirens*, up to 5 m. high. It has glabrous stems and entire, coriaceous and opposite (9-18 x 25-40 mm) leaves, dark green on their adaxial side and light green on their back, with short petiole. Small flowers in inflorescence c. 10 mm in some leaf axils, where only one is female, sessile in the centre of the clusters, and the rest are male, shortly pedicelled with four yellowish pieces and four stamens; the female ones have three persistent styles. Flowering is IV-VI, and fruits are ovoid, loculicide capsules, trilocule, up to 7 mm. widely oblong, with its styles as long as its capsule. Seeds are black, with 2 in each loculus.

Chorological studies show that *B. balearica* is a relict of Tertiary flora, which was much more abundant before but is nowadays relegated to calcareous soils at sea level or close to the coast. *B. balearica* is confined to the Western Mediterranean area, concretely in the Balearic Islands, Sardinia, and North of Morocco and also in sub-coastal mountains in the South East of the Iberian Peninsula (Lázaro et al. 2006). Two other populations are also known in Turkey, initially considered a different species. Its current distribution in Andalusía consists of a few fragmented and isolated populations distributed throughout it from Malaga to Almeria (Blanca 1999). Its populations have drastically decreased and a significantly reduced number of individuals are found due to both climate change in the Mediterranean region during the Holocene (Yll et al. 1997) and human activity during the last centuries (Yll et al. 1997, Lázaro et al. 2006). It has been catalogued as ''vulnerable'' by the UICN, and in Spain, it is considered to be at risk of extinction in Andalusia (Blanca 1999).

All the evidence shows that *B. balearica* in southern Spain finds acceptable conditions for flowering; fructification, seed production, dispersion and germination (Lázaro et al. 2006).

However, summer drought, the most crucial factor affecting the regeneration process in a number of Mediterranean species promotes an extreme environmental restriction for seedling establishment. Populations are thus doomed to disappear in the long term. The situation is even tougher due to the effect of wild fires and human pressure, which accelerate the population's fragmentation process and vanishment. This situation is creating a restrictive scenario for the species' conservation programs, demanding a higher precision in repopulation activities.

Ecological niche modelling is important for a variety of applications in ecology and conservation (Graham and Hijmans 2006). For example, they attempt to provide detailed distribution cartography related to the spread of species (Thuiller et al. 2004), impacts on climate change (Thomas et al. 2004, Matsui et al. 2009), and spatial patterns of species diversity (Feldemeyer-Christe et al. 2007). As such, ecological niche modelling has been the target for an impressive growth in attention in recent years (Guisan & Zimmermann 2000, Thuiller et al. 2004, Thomas et al. 2004, Araújo et al. 2005, Soberón & Peterson 2005, Graham and Hijmans 2006; Elith et al. 2006, Matsui et al. 2009) placing this technique among emerging new approaches relevant to ecology, biogeography, and conservation biology.

The basic approach of these models is to combine a set of known occurrences together with prediction variables such as topographic, climatic, edaphic, biogeography and remotely sensed ones. Accurate occurrence data (presence and absence) are rarely available, especially for rare species or inaccessible site locations. Correlative models using species presence and absence locations records for habitat predictions have been referred to as discrimination techniques, while those using only species presence records have been referred to as profile techniques (Jeschke & Strayer, 2008). Examples of discrimination techniques include those models based on discriminant analysis (Rogers et al. 1996), general linear model regression (GLM) (Cumming 2000), generalized additive models (GAM) (Leathwick & Whitehead

2001), Multivariate Adaptive Regression Spline (MARS) (Muñoz & Felicísimo 2004) and decision-tree based methods (Araújo & Williams 2000). On the other hand, climate envelope techniques (e.g., ANUCLIM, BIOCLIM, DOMAIN, FEM, HABITAT, and Mahalanobis distance) used a classic bioclimatic modelling approach (Hirzel et al. 2002, Anderson et al. 2003). Comparisons of modelling techniques are a current issue of research (Jeschke & Strayer 2008).

MARS has been applied to Mediterranean species in the past few years for many authors dealing with habitat prediction (Muñoz & Felicísimo 2004; Navarro-Cerrillo et al. 2006) reporting accurate results for tree species. Classification algorithms commonly used are CART or TREE NET, often less accurate than MARS (Muñoz & Felicisimo, 2004; Segurado & Araujo 2004). Another group of techniques are generics algorithms based on environmental dissimilarity metrics (BIOCLIM, HABITAT, DOMAIN o ENFA) (Beaumont et al., 2005). More recently, a machine learning approach to predictive modelling has applied new algorithms such as GARP (Genetic Algorithm for Rule-Set Prediction) (Stockwell & Peterson 2002, Anderson et al. 2003, Pape & Gaubert 2007) or MAXENT (Maximum Entropy) (Phillips & Dubik 2008), which has proven especially successful in predicting species potential distributions under a wide variety of situations.

Modelling prediction methods results in different geographic distributions depending mainly on species ecology specialization and number of samples (Stockwell & Peters 2002; Randin et al. 2006, Pearson 2006), so it is important to compare the performance of the many algorithms and approaches to project distributions within the same regions used to train the models (Elith et al. 2006, Guisan et al. 2007). Most studies have focused on species with a wide geographical distribution and large number of presence data (Beaumont et al. 2005; Phillips & Dubik 2008). On the contrary, spatial distribution modelling at a local scale has been less tested and experience obtained with habitat prediction of spread species cannot be easily generalized.

The aim of this contribution is to provide a comparison between modelling algorithms to study the habitat distribution of such a fragmented species as *Buxus balearica* Lam. by evaluating the prediction capability of different spatial distribution models at a local scale.

2. Methods

2.1. Study area

Buxus balearica was chosen as the species for analysis based on the criteria of ecological relevance and fairly broad geographic distributions, giving sample sizes sufficient for analysis. The distribution area of *Buxus balearica* studied in this work was the Málaga and Granada populations, which includes almost its total Iberian Peninsula distribution with the exception of the Rágol locality (Almeria province) (36º and 38º north) (Figure 1). This area extends along the southern slope of the Sierras de Almijara, Cázulas and Los Guájares. Part of these mountains is nowadays protected by the Andalusian Government, the Natural Park of the Sierras de Tejeda, Almijara and Alhama, with 40,600 ha. The study area was restricted to basic soils, mainly marls and chalky ones, which occupied a surface of 38,180 km^2.

Those mountain ranges form a geographical barrier between the provinces of Málaga and Granada and constitute a continuous alignment of rocky escarpments reaching the most prominent height in the South of Spain and Portuguese coast (Maroma summit, 2,068 m). It acts as a corridor, connecting the Sierras of Málaga with Sierra Nevada, due to its longitudinal distribution. The predominant geology is a consistent limestone, although in some areas there are also Eocene loamy sandstones, limestones, and dolomite marbles. The growing season is about 8-months (March to October), and annual precipitation averages from 450 or 500 mm in coastal areas (Almuñécar, Vélez-Málaga) to more than 1000 mm in some mountainous areas of the northern slope (Alfarnate, Zafarraya), the average values being 600-700 mm in medium height mountainous areas facing the Mediterranean Sea (Cómpeta, Canillas). Dominant climate is a semi-arid Mediterranean (sensu Quézel) below 500 m above sea level and sub-humid Mediterranean (in mountains above 600 and 700 m). Monthly average temperatures oscillate between 11ºC and 26ºC. In spite of its Mediterranean climate, with a long dry period during the summer, Western warm and humid winds from the Strait of Gibraltar reach these mountains. These sea breezes condense at a certain altitude (generally in the evening) and compensate water deficit.

Forest species composition includes many myrtled-shape leaf species well adapted to Mediterranean conditions, such as *Pistacia lentiscus* L., *P. terebinthus* L., *Rhamnus alaternus* L., *Buxus balearica*, *Coriaria myrtifolia* L., *Osyris quadripartite* Salzman *ex* Decaine, *Cneorum tricoccon* L. and *Quercus coccifera* L. among others. *B. balearica* can only be found in the termo and meso-Mediterranean belts.

2.2. Data processing

The prediction models were tested based on 17 environmental variables (Environmental Information Network of Andalusia-Consejerıa de Medio Ambiente) (Table 1). All data sets were resampled to 10 m resolution for analysis to reflect the spatial resolution of the occurrence data.

The topographical data – aspect (ASP), slope (SLP), Euclidean distance to the closest drainage (EDW), and isolation models (MPI) – were derived from a Land Digital Model (10 m resolution) (CMA, 1998). Altitude was not considered because the species appears indistinctly distributed from approx. 0 to 800 m. (Elevation range registers within the study area). Isolation was calculated with SHORTWAVC aml application (Felicísimo et al. 2002). Eight isolation models were calculated, one every 45 days throughout the year, on which total radiation received at the surface of the earth over a period of time was estimated (Table 2).

Finally, taking into account the preference of *B. balearica* for areas where ambient humidity may accumulate, Euclidean distance to the closest drainage area was calculated.

Meteorological data – winter rainfall (TP_1), spring rainfall (TP_2), summer rainfall (TP_3), autumn rainfall (TP_4), mean temperature of warmest month (ATWM), mean maximum temperature of warmest month (AMTWM), mean temperature of coldest month (ATCM), mean minimum temperature of coldest month (AMTCM), – were obtained from a 28 weather station network (National Institute of Meteorology). The time serial data were 1956–2000. The methodology applied to obtain a 100 m grid of meteorological variables was the interpolation of data from each station. This interpolation was based on multiple regression analysis taking into account the position of each point (coordinates x, y), the altitude (obtained from Land Digital Model), the distance to the sea and orientation weight (maximum value of 4 to SW orientation and minimum value of 0 to NE). The litology (LITO) was obtained form a digital lithological map of the study area (scale: 1:100,000).

Current locations of *Buxus balearica* were obtained by field surveys using a GPS (GPSmap® 60CSx), from herbarium specimens (Botanical Gardens-University of Córdoba) and from historical citations. A value of presence (1) of *B. balearica* was assigned to 220 points and it was considered when the species presented itself, independently of population density.

2.3. Modelling methods

Several approaches have been used to approximate the ecological niches of species (Elith et al. 2006). In this study, six methods were compared:

a) Multivariate adaptative regression spline, MARS, proposed by Friedman (1991) provides an alternative regression-based method for fitting non-linear responses, using piecewise linear fits rather than smooth functions. MARS uses a stepwise addition/deletion strategy with linear splines and it adopts the generalized cross-validation criterion to add and select basic functions.

b) Maximum entropy approach to modelling species' distributions (Maxent) (Phillips et al. 2006, Phillips 2006). Maxent estimates species probability distribution by finding the probability distribution of the maximum entropy, subject to a set of constraints that represent our incomplete information about the target distribution. Mathematical approach of Maxent is based on the estimation of probability function for a species satisfying all the bioclimatic or environmental limits associated while maintaining its distribution and maximum entropy. The Maxent algorithm calculates the distance between the given environmental conditions to each occurrence point and selects the closest distance. In terms of potential distribution, the probability of the presence of a species can be interpreted as the probability of not finding any biophysical limitations to the species' existence; therefore, the greater the entropy of a system, the greater the probability of encountering sites without such limitations (Phillips & Dubik 2008).

c) Climate envelope techniques.- The climate envelope modelling approach has its foundations in ecological niche theory. Climate envelopes can be defined as constituting the climatic component of the fundamental ecological niche, or the 'climatic niche'. Some bioclimatic models are based on empirical relationships between observed species distributions and environmental variables (Peterson 2006). Some examples are ANUCLIM, BIOCLIM, DOMAIN, FEM, HABITAT, and Mahalanobis distance. They fit a minimal envelope in a multidimensional climate space and use presence-only instead of presence/absence data. On this study, two models were selected: BIOCLIM and DOMAIN.

d) Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell & Noble 1992) is a genetic algorithm that creates ecological niche models for species. The models describe environmental conditions under which the species should be able to maintain populations. For input, GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capability to survive. GARP works in an iterative process of rule selection, testing, and incorporation or rejection: first, a method is chosen from a set of possibilities (logistic regression, bioclimatic rules, range rules, negated range rules), and then is applied to the training data and a rule developed to maximize predictivity (Anderson et al. 2003).

e) Support vector machines (SVMs) are a group of supervised learning methods that belong to a family of generalized linear classifiers. They can also be considered a special case of Tikhonov regularization. SVMs can perform binary classification (pattern recognition) and real valued function approximation (regression estimation). A special property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers. Support vector machines map input vectors to a higher dimensional space, where a maximal separating hyper plane is constructed. This hyper plane will attempt to split the input data into two classes. The separating hyper plane is that which maximises the distance between the two classes. The model only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Intuitively, this would make the classification correct for testing data that was near, but not identical to the training data.

2.4. Model evaluation

To assess the agreement between the presence-absence and the predictor’s records three statistics were used: the area under curve (AUC), the correlation coefficient (COR) and the maximum Kappa (κ). The AUC ranges between 0 and 1. Cohen’s Kappa (κ) is one measurement that can be derived from the confusion matrix. As a validation tool (i.e. when the ‘truth’ is known), it states the overall accuracy of a prediction once the element of chance has been removed (Liu et al. 2005). Otherwise, Kappa can serve as a tool to assess reliability of prediction in terms of relative agreement. A value near 0 indicates no discrimination (agreement by chance); a value of 1 represents perfect discrimination (agreement); a value of > 0.6 is considered ‘good’ and >0.8 as ‘excellent’ (Graham & Hijmans, 2006). Kappa is relatively tolerant to zeros in the confusion matrix and considers both omission and commission errors in one parameter. However, Kappa is unimodally dependent on prevalence, i.e. on the proportion of all presences in the full validation dataset (Allouche et al. 2006). This would have been a problem in this project, because the presences of known species were few (training: ~ 150; test: ~70) compared to the large number of random background points (>1000), which were deemed necessary for a reliable confusion matrix given the large study area. Hence, the TSS (Allouche et al. 2006) was computed instead of Kappa. It is defined as TSS = sensitivity + specificity – 1. In contrast to Kappa, TSS values can be used to compare prediction performance regardless of both the validation dataset size and the prevalence contained therein, while still featuring the same strengths of Kappa: full consideration of sensitivity, specificity and chance (Allouche et al. 2006). Finally, the concordance among models was evaluated through a correlation analysis.

3. Results

The area under curve (AUC) scores for these models showed a value of 0.73 for Maxent, compared with 0.60 for GARP models, suggesting that GARP models are less predictive than Maxent models (Figure 2). The same results were obtained with the kappa (κ) coefficient and the correlation coefficient (COR) (Figure 2). All the other models tested obtained lower accuracy values. The lowest accuracy was ranked by the Environmental Distance model. The correlation matrix of the different models shows the spatial prediction concordance between models (Table 2). The highest correlations were obtained in two separate groups; first, between the generic algorithms based on environmental dissimilarity metrics. The highest correlations were ranked by the Environmental Distance and the SVM and Bioclime. The second was between Maxent and MARS, obtaining the highest significance of the correlation

Maxent and MARS are different algorithms and follow different computational routines although both produced similar potential distribution models on this study. The relative contributions of the environmental variables were similar as well. Figure 3 shows a heuristic estimate of the relative contributions of the environmental variables to the Maxent and MARS model. To determine the estimate in each iteration of the training algorithm, increase in regularized gain was added to the contribution of the corresponding variable. The highest contributions fell on the climate variables: spring rainfall (TP_2), mean temperature of warmest month (ATWM), and mean temperature of coldest month (ATCM) with proportional weights. The variables with lowest contributions were the terrain isolation models. Figure 4 shows the results of the jackknife test of a variable importance. The environmental variable with the highest gain when used in isolation is ATCM, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is ATCM, which therefore appears to have the largest amount of information that was not present in the other variables.

Once Maxent was identified as the most accurate model, a test of sensitivity to the reduction in the number of variables and number of presence input data was performed. Figure 5 shows the curve described by the AUC depending on the number of variables introduced in the analysis. The number of variables introduced in each model is explained in Table 3. Results show a decreasing tendency of the AUC with the reduction of the number of variables. However, the model obtains an accuracy of over 0.90 by applying just 3 variables (TP_2, ATWM, and ATCM). In contrast, the reduction in the number of presence input data shows a greater sensitivity (Figure 6). The maximum number of points taken into the analysis reach high values in the accuracy assessment of both Maxent and MARS. A reduction of 50% in the number of presence data shows a reduction in kappa coefficient around 14.5% for MARS and 6.7% for Maxent. Comparing both models, MARS is more sensitive to the reduction in the number of presence input data.

All algorithms produced maps that coincided well with the known distribution of the species (Figure 7), although the GARP prediction tended to be overly extensive, and the Maxent model tended to be somewhat underpredicted. GARP models continued to reconstruct many of the species' known distributions. Maxent models, on the other hand, produced an odd pattern coincident with the input data set at higher probability values reconstructing the on-diagonal quadrants or the off-diagonal quadrants, depending on which were used to train the models (Figure 7).

4. Discussion

4.1 Model evaluation.

Differences in prediction performance between modelling methods at a local scale is shown on this study. In general terms, results agree with those obtained from the application of the same models to the species distribution at a regional and global scale (Elith et al. 2006, Austin 2007). The generics algorithms based on environmental dissimilarity metrics obtain poor results compared with those obtained by the machine learning approach. The accuracy of Maxent was significantly higher than all the prediction models evaluated. The computational efficiency of each model should be evaluated considering three main aspects: pre- processing required for the input data, processing efficiency of the model itself, post-processing required to acquire statistical and cartographical prediction. Generic algorithm models based as BIOCLIM, DOMAIN, SVC, or GARP, are of a great computational efficiency. Implemented in common free software like the open modeller (http://openmodeller.sourceforge.net), and identical input data are used for all of them. The user-friendly interface and the fast computation of the algorithm may prove the usefulness of these models at a regional and global scale. Nevertheless, at a local scale, the results show low accuracy values and poor prediction maps. Maxent and MARS significantly improves all the predictions. Although Maxent share all the computational advantages of generic algorithms, MARS presents some drawbacks. The input data have to be introduced in a specific format (.sav) with the environmental data information extracted from the presence/absence data. This process involves the necessity to process all the information with Arc/info. Once the functions are generated, it is necessary to apply a specific routine programmed in AML to obtain the raster prediction with ArcInfo.

In contrast, Maxent not only exhibits a high computational efficiency, it also produces an extensive statistical report of the model. In order to test the sensitivity of the model to different analysis were performed. The first one consisted of the reduction of the number of variables. The high accuracy obtained with just three climate variables shows the strong dependency of the species studied on the climatology conditions. This last result is a very important one from the point of view of the study and assessment of the distribution of *Buxus balearica*. The last test performed on the Maxent model was the reduction in the number of presence input data. Compared to the MARS model, the reduction in accuracy was significantly lower with the Maxent model. This last result is also very important insofar as the input data of presence are relatively inaccessible. Apart from this difference in accuracy and computational efficiency, the prediction maps agree between Maxent and MARS, as well as the importance of the variables in the model.

4.2. Potential distribution model of Buxus balearica.

Within the meso-Mediterranean and supra-Mediterranean climate types, *B. balearica* was observed at 220 locations. The digital model developed for predicting the distribution of *B. balearica* is a probabilistic raster map, shown in Figure 7. Kappa index results achieved

0.84 accuracy rating in predicting both presence and absence of the species. Response curve diagrams produced along with the model output indicate the effect of individual variables on the Maxent prediction. Environmental variables based on meteorological data made the highest contribution to the potential distribution. The single variable making the greatest contribution to defining the potential distribution of the species was mean temperature of warmest month (ATWM). The species presents an alarming sensitivity to thermic variations. *B. balearica* limits itself to areas with average temperatures ranging from 24.6 to 25.4ºC during the warmest month. Results of the mean maximum temperature of warmest month (AMTWM) indicate a species tolerance within a range between 31ºC and 32ºC. As a consequence, a climate change scenario of solely one degree could greatly diminish the survival possibilities of this species. It is also clear that the low tolerance of *B. balearica* to frost, disappearing as a result of competition with other species in areas where mean temperature of coldest month (ATCM) is below 7ºC and mean minimum temperature of coldest month (AMTCM) is below 3ºC.

Seasonal precipitation also showed itself to have a great impact on the distribution model, measured by total trimester precipitation. Spring rainfall (TP_2) showed yet again the species's sensitivity to climatic and irregularity during the spring-summer season. It could be said that in areas with the presence of *B. balearica* annual precipitation is above 500 mm, with spring precipitations above 110 mm and summer rainfall (TP_3) at least 25 or 30 mm. Consequently, *B. balearica* requires certain soil moisture. However, this interpretation may be incomplete, considering the lack of data concerning crypto-precipitations. This added condensation is continually scarcer due to the climate variations observed in the last decades within the region.

The response to orientation models MDI5 (12.5º) and MDI4 (22.5º) with respect to MDI2 (0º) showed a significant contrast. During the summer solstice (sun at its maximum height), the majority of locations with presence of the species receive a great amount of energy while during the winter solstice (sun at its lowest position) solar radiation is at its lowest. According to model results, *B. balearica* distribution is more affected by solar isolation during the period between the summer solstice and fall equinox (MDI5-12.5º). The distribution of the species, on the other hand, was indifferent to aspect, even though a slight preference was shown for a southerly orientation. These results agree with field observations, in which *B. balearica* was seen to be distributed with various orientations. Slope was also no significant in determining the presence or absence of the species, as it appears in slopes of a varying steepness.

Euclidean distance to nearest drainage, based on a watershed model, was an influential factor in determining the presence of the species. *B. balearica* specimens appear to take refuge in precipices, streams and rivers, and certainly not so much because of subsoil moisture, but more likely because it is in such areas that the species is subject to less incident radiation. Even more so because in these areas it is able to avoid forest fires, which here do not reach great intensity, and conditions allow the plant resprouting. Moreover, seedling mortality due to summer drought can be overcome in the riparian zone (brooks and streams of seasonal flows).

5. Conclusion

An evaluation of model performance to predict the distribution of *Buxus balearica* is described in this paper. It is important to note that many applications of species distribution models depend on predicting potential distributions, rather than realized distributions. Facing increasing climate change and human pressure, *B. balearica* populations will tend to shelter in areas with microclimate conditions that are less unfavourable and which limit evapotranspiration, regardless of altitude or topography. Average performance of different habitat distribution models for current versions showed that Maxent, evaluated on independent presence/absence test, may be applied to predict potential distributions of *B. balearica*.

Acknowledgements

This study was supported by the coordinated Spanish MEC project DINAMED (CGL2005-05830-C03-02/BOS), INTERBOS (CGL2008-04503-CO3-02) and FEDER funding. We thank the Córdoba Botanical Garden (University of Córdoba, Spain) for providing the herbarium data. We thank Antonio Pulido Pastor for their facilities for data collection in the Parque Natural Sierra de Tejeda, Almijara y Alhama (Málaga-Granada, Spain).

References

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223–1232.

Anderson, R.P., Lew, D. & Peterson, A.T.2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162: 211-232.

Araújo, M.B., Pearson, R. G., Thuiller, W. & Erhard, M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology* 11: 1504-1513.

Araújo, M. B. & Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96: 331- 345.

Austin M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 1-2: 1-19.

Beaumont, L.J., Hughes, L. & Poulsen, M. 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecological Modelling* 2: 251-270

Blanca, G. (Ed.) 1999. *Libro Rojo de la flora silvestre amenazada de Andalucía. Tomo I: especies en peligro de extinción.* Consejería de Medio Ambiente, Junta de Andalucía, Sevilla.

Cumming, G.S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography* 27: 441–455.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. & Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29: 129-151.

Feldmeyer-Christe, E.; Ecker, K.; Küchler, M.; Graf, U. & Waser, L. 2007. Improving predictive mapping in Swiss mire ecosystems through re-calibration of indicator values. *Applied Vegetation Science* 10: 183-192.

Felicísimo, A.M., Francés, E., Fernández, J.M., González-Díez, A. & Varas, J. 2002. Modeling the potential distribution of forests with a GIS. *Photogrammetric Engineering & Remote Sensing* 5: 455-461.

Friedman, J.H., 1991. Multivariate adaptive regression splines. *Annals of Statistics* 19: 1-141.

Graham, C.H. & Hijmans, R.J. 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography* 15: 578-587.

Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.

Guisan, A., Zimmermann, N.E, Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. 2007. What matters for predicting spatial distributions of trees: techniques, data, or species' characteristics?. *Ecological Monographs* 4: 615-630.

Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. 2002. Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data?, *Ecology* 83: 2027-2036.

Jeschke, J.M. & Strayer, D.L. 2008. Usefulness of Bioclimatic Models for Studying Climate Change and Invasive Species. *Annals of the New York Academy of Sciences* 1134: 1–24.

Lazaro, A., Traveset, A. & Castillo, A. 2006. Spatial concordance at a regional scale in the regeneration process of a circum-Mediterranean relict (*Buxus balearica*): connecting seed dispersal to seedling establishment. *Ecography* 29: 683-696.

Liu, C., Berry, P.M., Dawson T.P. & Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species. *Ecography* 28: 385-393.

Lehmann, A., McC Overton, J. & Leathwick, J.R. 2006. GRASP: generalized regression analysis and spatial prediction. *Journal of Applied Ecology* 43, 413–423.

Leathwick J.R. & Whitehead D. 2001. Soil and atmospheric water deficits, and the distributions of New Zealand's indigenous tree species. *Functional Ecology* 15**:** 233–242.

Matsui, T.; Takahashi, K.; Tanaka, N.; Hijioka, Y.; Horikawa, M.; Yagihashi, T. & Harasawa, H. 2009. Evaluation of habitat sustainability and vulnerability for beech (*Fagus crenata*) forests under 110 hypothetical climatic change scenarios in Japan. *Applied Vegetation Science* 12: 328–339.

Muñoz, J. & Felicísimo, A.M. 2004. Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15: 285-292.

Navarro Cerrillo, R.M., Lara Fernández, A.; Blanco Oyonarte, P.; Calzado Martínez, C.; López Quintanilla, J. Fernández Cancio, A., Guzmán Álvarez, J.R. & Sánchez Salguero, R. 2006. Aproximación a la definición del hábitat fisiográfico del *Abies pinsapo* Boiss. en Andalucía. *Investigación Agraria. Sistemas y Recursos Forestales.* Fuera de serie, 137-152

Pape, M. & Gaubert, P. 2007. Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and Distributions* 13: 890–902.

Peterson, A. T. 2006. Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3: 59-72.

Phillips, S.J., & Dubik, M. 2008. Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-175.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Phillips, S. J. 2006. Maxent software for species habitat modeling, ver. 2.2.0. Bhttp://www.cs.princeton.edu/_schapire/maxent/

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E.; Zappa, M. & Guisan, A. 2006. Are niche-based species distribution models transferable in space?. *Journal of Biogeography* 33: 1689–1703.

Rogers D.J., hay, S.I. & Packer, M.J. 1996. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. *Annals of Tropical Medicine and Parasitology* 3: 25-241.

Segurado, P. & Araujo, M. 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31: 1555–1568.

Soberon, J. & Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2: 1-10.

Stockwell, D.R.B. and Noble, I. R. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of analysis. *Mathematics and Computers in Simulation* 33: 385-390.

Stockwell, D.R.B. & Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1-13.

Thomas, C.D., Williams, S.E., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., De Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T. & Phillips, O.L. 2004. Biodiversity conservation: Uncertainty in predictions of extinction risk/Effects of changes in climate and land use/Climate change and extinction risk. *Nature* 430: 6995.

Thuiller, W., Araújo, M.B. & Lavorel, S. 2004. Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography* 31: 353–361.

Yll, E.I., Perez-Obiol, R., Pantaleon-Cano, J. & Roure, J.M. 1997. Palynological evidence of climatic change and human activity during the Holocene on Minorca (Balearic Islands). *Quaternary Research* 3: 339-347.

Table 1. Environmental variables resampled to 10 m resolution used on the prediction models.

Variable	COD	Source
Aspect	ASP	Generated from DEM
Slope	SLO	Generated from DEM
Euclidean distance to watershed	EDW	Generated from DEM
Model of Potential Insolation , Solar Declination 22.5°	MPI+22.5	Generated from DEM
Model of Potential Insolation, Solar Declination 12.5°	MPI+12.5	Generated from DEM
Model of Potential Insolation, Solar Declination 0°	MPI 0	Generated from DEM
Model of Potential Insolation, Solar Declination -12.5°	MPI-12.5	Generated from DEM
Model of Potential Insolation, Solar Declination -22.5°	MPI-22.5	Generated from DEM
Total Precipitation (January-March)	TP1	National Institute of metherology (INM) (www.mapya.es)
Total Precipitation (April-June)	TP2	National Institute of metherology (INM) (www.mapya.es)
Total Precipitation (July-September)	TP3	National Institute of metherology (INM) (www.mapya.es)
Total Precipitation (October-December)	TP4	National Institute of metherology (INM) (www.mapya.es)
Average temperature of the warmest month	ATWM	National Institute of metherology (INM) (www.mapya.es)
Average maximum temperature of the warmest month	AMTWM	National Institute of metherology (INM) (www.mapya.es)
Average temperature of the coldest month	ATCM	National Institute of metherology (INM) (www.mapya.es)
Average minimum temperature of the coldest month	AMTCM	National Institute of metherology (INM) (www.mapya.es)
Lithology	LITO	Department of Environment of the Government of Andalusia

Table 2. Bivariate Pearson Correlation matrix of the concordance of the different spatial distribution models analysed.

	Bioclim	Maxent	MARS	GARP	SVM	ENV_DIST
Bioclim	1.000					
Maxent	-0.040	1.000				
MARS	-0.180	0.807***	1.000			
GARP	0.769**	0.303*	0.240	1.000		
SVM	0.856**	0.290	-0.110	0.798**	1.000	
ENV_DIST	0.856**	0.290	-0.110	0.798**	1.000**	1.000

**. Correlation is significant at the 0.01 level (2-tailed)

***. Correlation is significant at the 0.001 level (2-tailed)

Table 3. Sensitivity analysis of Maxent model. The table shows from right to left, the variables introduced, the code assigned to the model and the relative importance of the last variable introduced. Variables on table 1.

Model	Cod	Imp.
ATWM	Mod_01	24
ATWM, ATCM	Mod_02	12.9
ATWM, ATCM, TP2	Mod_03	12.2
ATWM, ATCM, TP2, AMTCM	Mod_04	7.8
ATWM, ATCM, TP2, AMTCM, MPI-22.5	Mod_05	5.8
ATWM, ATCM, TP2, AMTCM, MPI-22.5	Mod_06	5.6
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3	Mod_07	5.3
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3,	Mod_08	5.2
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4	Mod_09	5.1
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP	Mod_10	4.7
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM	Mod_11	3.8
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW	Mod_12	3.5
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1	Mod_13	2.1
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1, MPI0	Mod_14	0.7
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1, MPI0, MPI-12.5	Mod_15	0.7
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1, MPI0, MPI-12.5, ASP	Mod_16	0.4
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1, MPI0, MPI-12.5, ASP, MPI+12.5	Mod_17	0.1
ATWM, ATCM, TP2, AMTCM, MPI-22.5, TP3, TP4, TP, AMTWM, EDW, TP1, MPI0, MPI-12.5, ASP, MPI+12.5, MPI+22.5	Mod_18	0.1

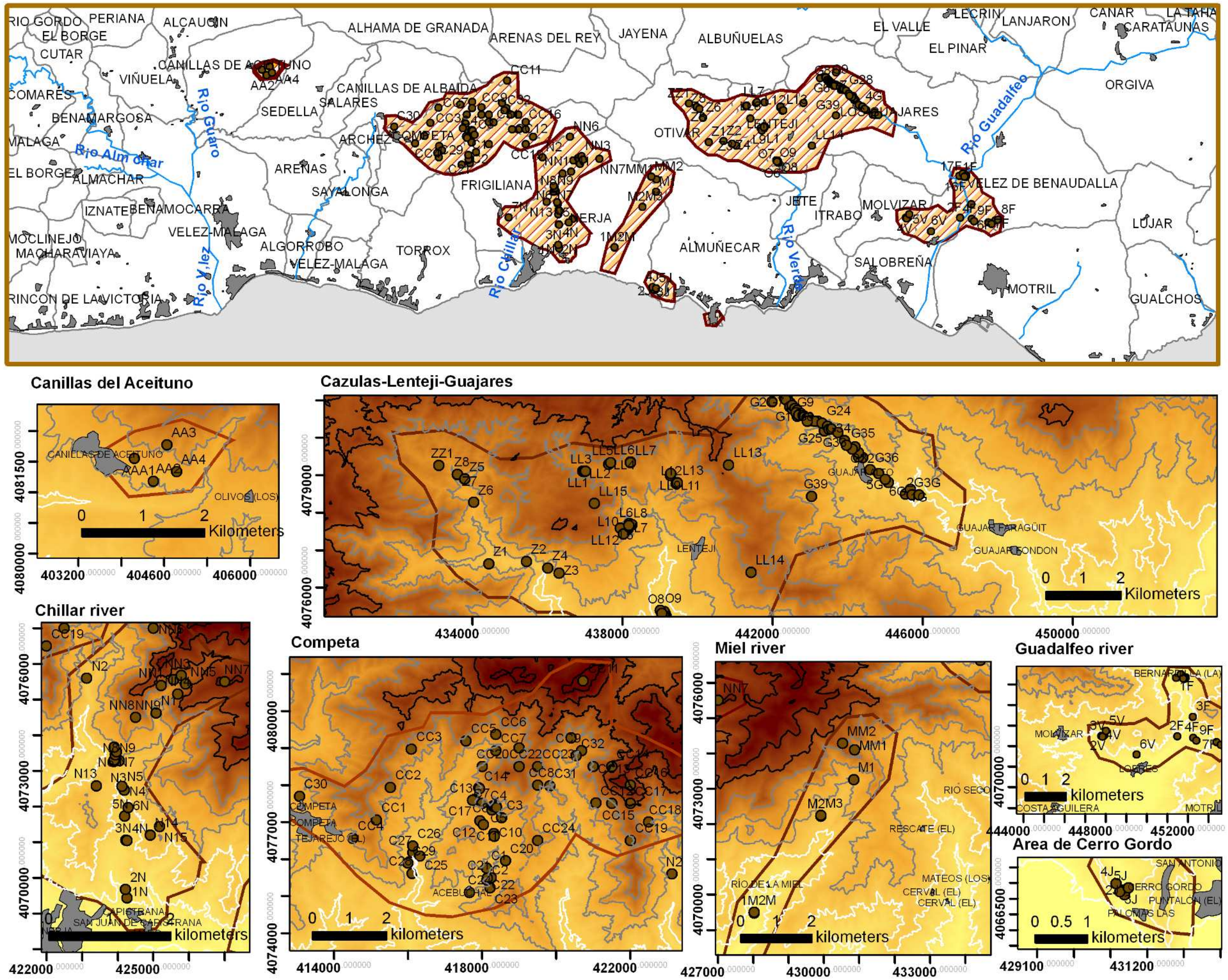


Figure 1. Distribution of *Buxus balearica* on the study locations area (Southern Spain).

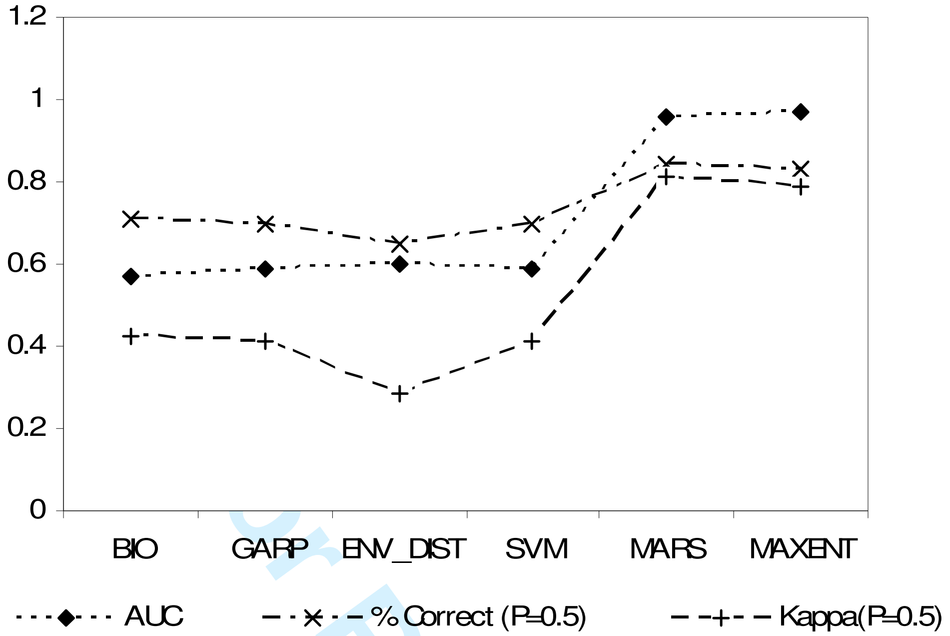
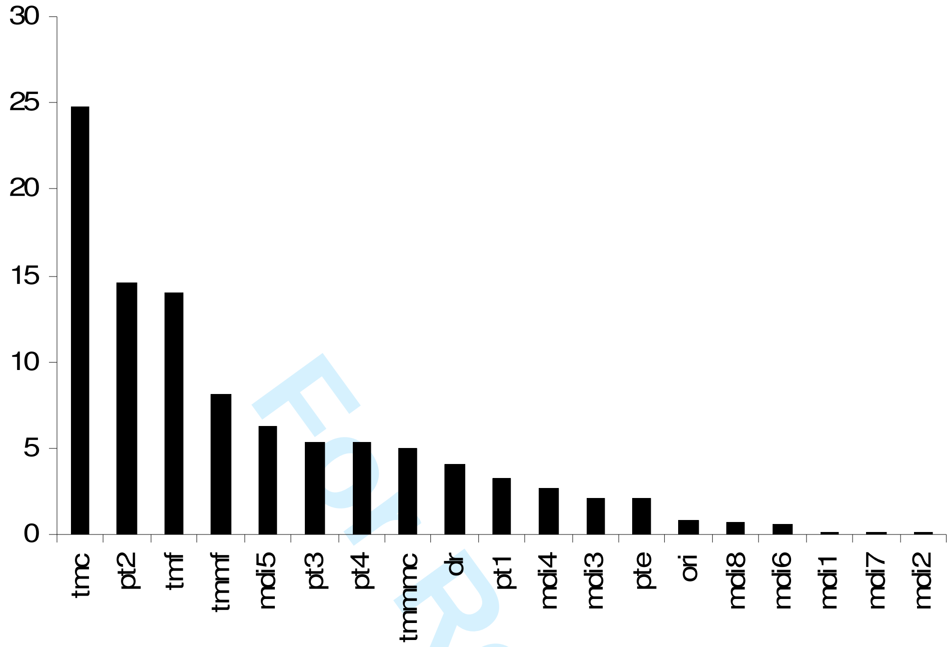


Figure 2. Comparison of the validation coefficients obtained by the application of six different spatial distribution models: BIOCLIM (BIO), GARP, DOMAIN (ENV_DIST), SVM, MARS and MAXENT.

a)



b)

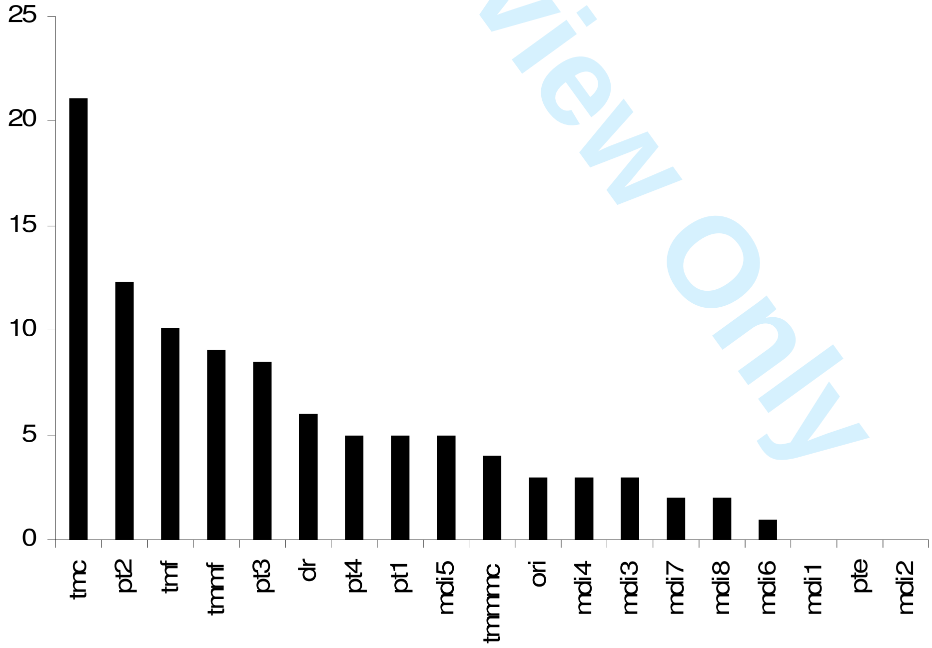


Figure 3.- Relative contributions of the environmental variables to the Maxent model (a) and to the MARS model (b).

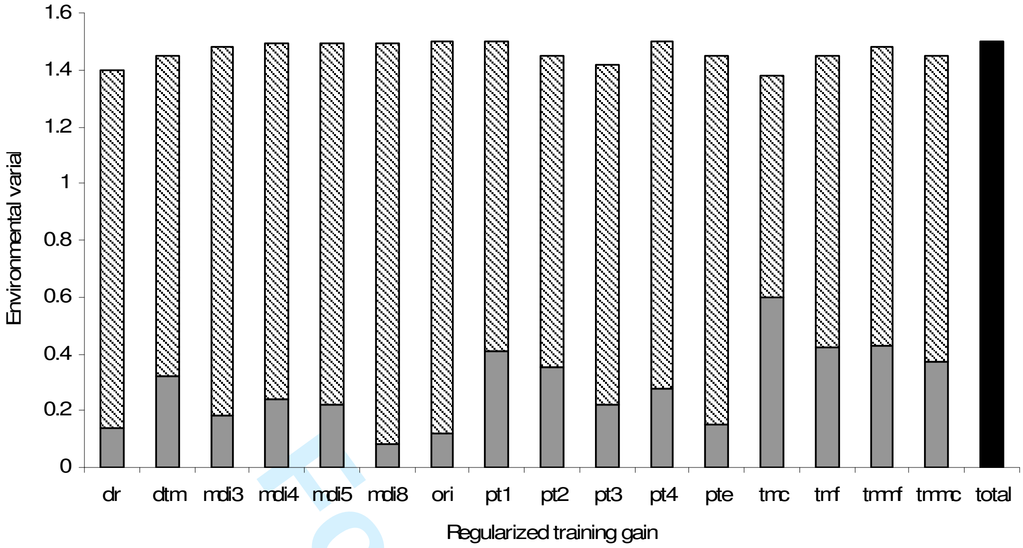


Figure 4.- Jack-knife test of variable importance to the Maxent model.

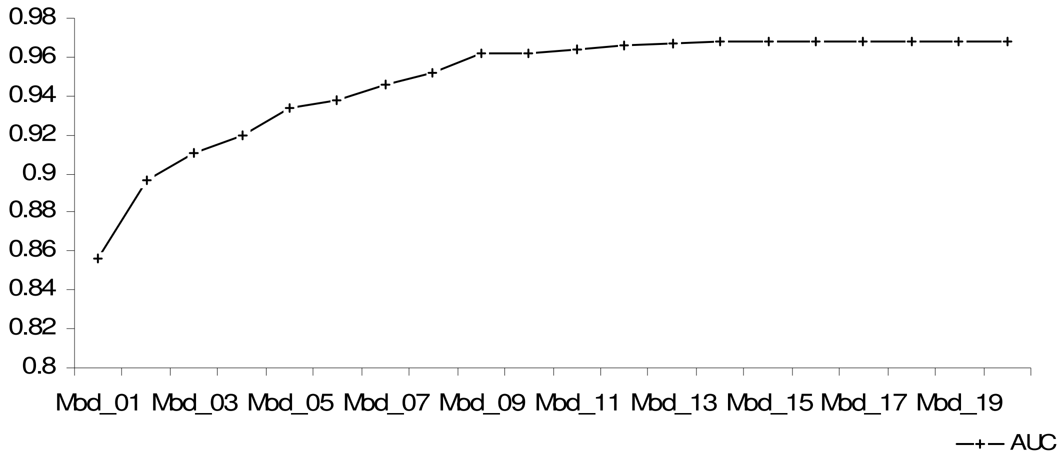
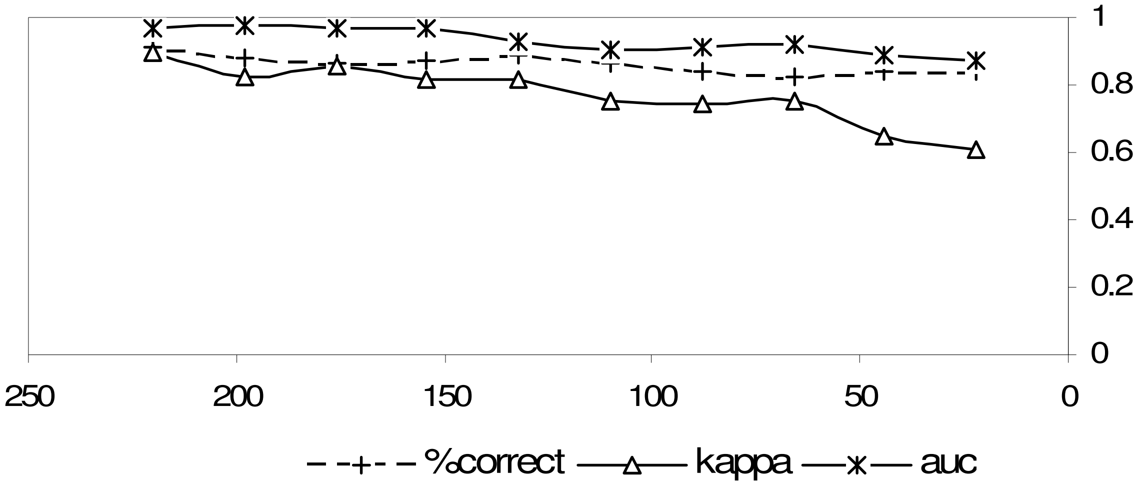


Figure 5.- Sensitivity analysis of Maxent model based on the reduction in the number of variables.

a)



b)

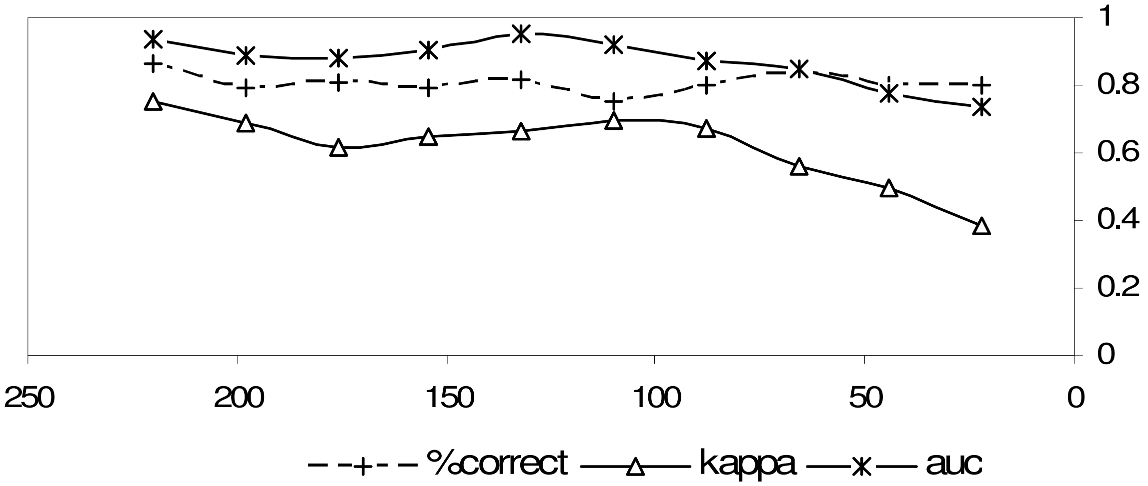


Figure 6.- Accuracy functions of Maxent model and MARS depending on the number of presence input data.

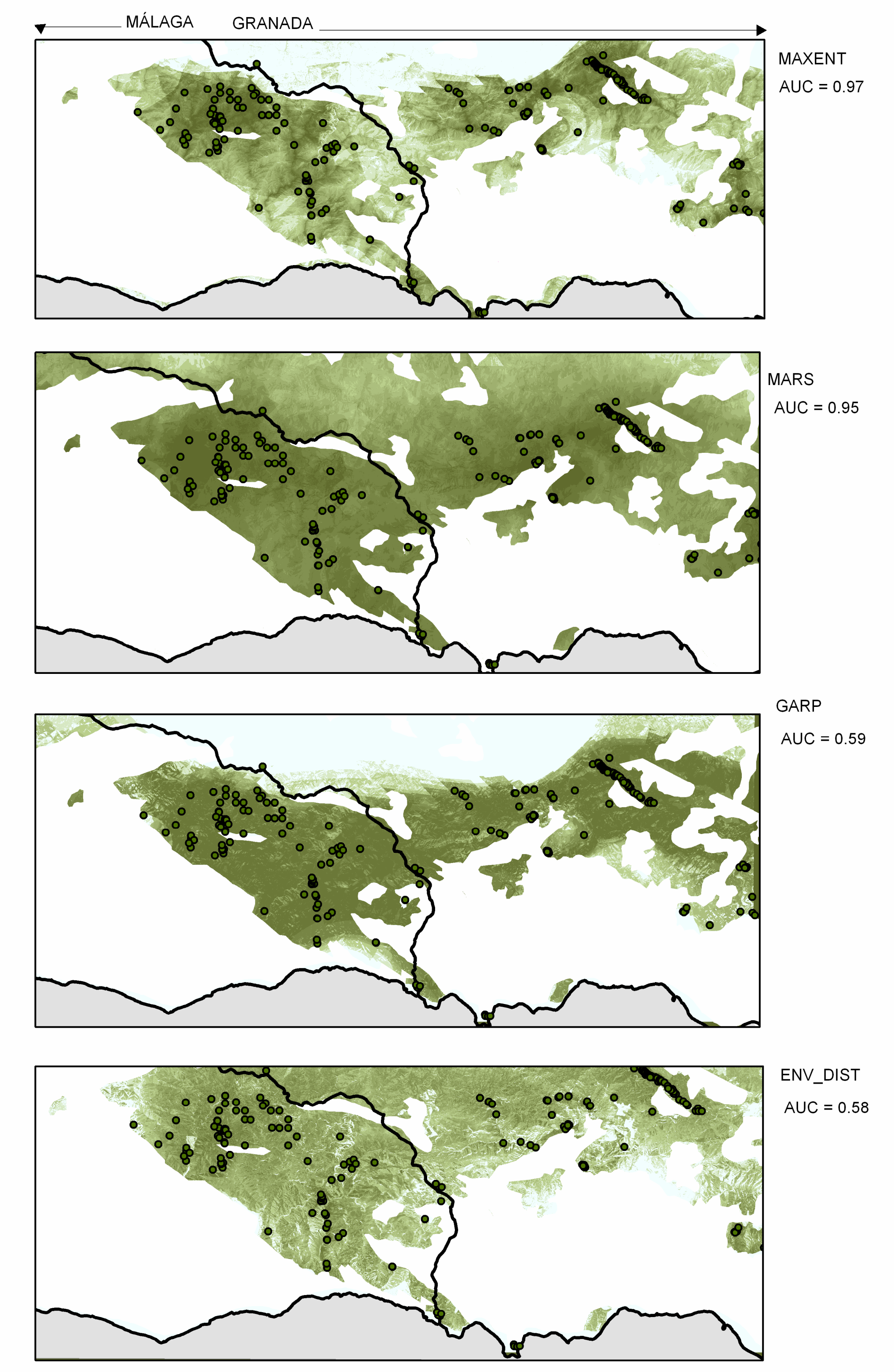


Figure 7.- Potential distributions of *Buxus balearica* obtained by the application of MAXENT, MARS, GARP and ENVIRONMENTAL DISTANCE (ENV_DIST).