Modeling and mapping potential distribution of Crimean juniper (Juniperus excelsa Bieb.) using correlative approaches

Kürşad ÖZKAN1*, Özdemir ŞENTÜRK2, Ahmet MERT3 and Mehmet Güvenç NEGİZ3

1 Suleyman Demirel University, Faculty of Forestry, Department of Forest Engineering, 32260, Isparta, Turkey
2 Mehmet Akif Ersoy University, Gölhisar Vocational School, 15400, Gölhisar, Burdur, Turkey
3 Suleyman Demirel University, Sutculer Vocational School, 32950, Sutculer, Isparta, Turkey

*Corresponding Authors Email: kursadozkan@gmail.com

Abstract

Modeling and mapping potential distribution of living organisms has become an important component of conservation planning and ecosystem management in recent years. Various correlative and mechanistic methods can be applied to build predictive distributions of living organisms in terrestrial and marine ecosystems. Correlative methods used to predict species’ potential distribution have been described as either group discrimination techniques or profile techniques. We attempted to determine whether group discrimination techniques could perform as well as profile techniques for predicting species potential distributions, using elevation (ELVN), parent material (ROCK), slope (SLOP), radiation index (RI) and topographic position index (TPI) as explanatory variables. We compared potential distribution predictions made for Crimean juniper (Juniperus excelsa Bieb.) in the Yukarı gökdere forest district of the Mediterranean region, Turkey, applying four group discrimination techniques (discriminate analysis (DA), logistic regression analysis (LR), generalized additive model (GAM) and classification tree technique (CT)) and two profile techniques (a maximum entropy approach to species distribution modeling (MAXENT), the genetic algorithm for rule-set prediction (GARP)). Visual assessments of the potential distribution probability of the applied models for Crimean juniper were performed by using geographical information systems (GIS). Receiver-operating characteristic (ROC) curves were used to objectively assess model performance. The results suggested that group discrimination techniques are better than profile techniques and, among the group discrimination techniques, GAM indicated the best performance.

Key words

Crimean juniper, Habitat selection models, Model selection, Potential distribution, Predictive accuracy, Predictive mapping

Introduction

The purpose of ecological restoration is to initiate the recovery of degraded ecosystems by creating conditions that will prompt the ecosystem to follow what is believed to be its natural pathway (Jordan et al., 1996; Anand and Desrochers, 2004).

Rather than actual distribution maps of species, their potential distribution maps give much more opportunities for developing strategies, making accurate policies and realization of suitable implementations in restoration of the ecosystems. Predictive distribution models have been increasingly used to form potential distribution maps of the species. Such models have been used for wild life animals (Johnson and Gillingham, 2005; Saqib et al., 2006; López-López et al., 2007), bird species (Martinez et al., 2003; Peterson et al., 2007; Ortega – Huerta and Peterson 2008), insect species (Ganeshaiah et al., 2003; Gallego et al., 2004; Jiménez-Valverde et al., 2008; Tognelli et al., 2009) and plant species (Vetaas and Grytnes, 2002; Marmion et al., 2009) in terrestrial ecosystems; fish (Sousa et al., 2006; Mounton et al., 2008), Jell fish (Drake and Bossenbroek, 2004; Bentlage et al., 2009) and mammal species (Kaschner et al., 2006) in aquatic or marine ecosystems. The models can also allow mapping potential distribution of forest types or vegetation communities as...
well (Brzeziecki et al., 1993; Zimmermann et al., 1999; Liu et al., 2009; Felicísimo et al., 2002; Miller and Franklin, 2002). Besides, they have been used to assess the potential impact of climate change on species distributions (Guisan and Theurillat, 2000; Thomas et al., 2004).

Predictive distribution models are mainly divided into 2 groups, namely mechanistic (static) models and correlative (dynamic) models. To apply the mechanistic models, detailed ecophysiological information of the target species should be provided. Unfortunately, it is not easy to provide such information for all species. On the contrary, such information is not required for application of correlative models. That is why, correlative models have been much more frequently used in prediction of species' distribution. According to the type of a response variable, correlative models are described as group discrimination techniques and profile techniques. Presence and absence of data (binary data) for a given species is required for the application of group discrimination techniques, while only presence data is evaluated using profile techniques.

In the present study, Crimean juniper (Juniperus excelsa Bieb.) was selected as focal species because it is the most common native tree species after pines and oaks throughout the Taurus Mountains of Turkey. Also, it has economical functions due to its hardwood which is highly durable and has aromatic properties (Gültekin and Gültekin, 2007; Özkán et al., 2010). That is why it has been subjected to overgrazing and individual selection for a long time in Turkey. Modeling and mapping potential distribution of Crimean juniper is particularly important to determine where new population of species could be established. This information is crucial to enable its sustainability.

In light of given information above, the performance of correlative techniques (4 group discrimination techniques (DA, LR, GAM, CART) and 2 profile techniques (MAXENT and GARP)) using geo-referenced data of Crimean juniper, and a set of environmental variables in the Yukarıgökdere forest district of the Mediterranean region, Turkey was assessed.

**Materials and Methods**

**Study area**: Yukarıgökdere forest district is located in the Lakes subregion of the Mediterranean region and, situated between north latitudes of 41°6′31″ and 41°8′99″ m and east longitudes of 30°32′12″ and 31°12′22″ m with an area of about 8000 ha (Fig. 1).

Limestone is dominant parent material. Locally conglomerates and ophiolite melanges are also present. Elevation ranges between 800 to 2000 m asl. In the district, a translation climate prevails between Mediterranean climate and continental climate with an average annual rainfall of 751 mm and an annual average temperature of 13.03 °C. The forest is mainly composed of Pinus brutia (Brutian pine), Pinus nigra (Crimean pine), Juniperus excelsa (Crimean juniper), Cedrus libani (Lebanon cedar) and Quercus (Oaks) species. The district is rich in endemic species with 61 endemic plant taxa (Fakir et al., 2009; Özkán and Negiz, 2011).

**Data and statistical evaluation**: The data was collected from one hundred and nineteen sampling plots. At each sampling plots, presence and absence data of Crimean juniper were recorded (Fig. 1). To provide distribution models of the species, by using presence and absence (binary) data of Crimean juniper, discriminate analysis (DA), logistic regression analysis (LR), generalised additive model (GAM) and classification tree technique (CT) were applied. Besides, two profile techniques called maximum entropy approach to species distribution modeling (MAXENT) and genetic algorithm for rule-set prediction (GARP) were applied using only presence data of the species. Environmental layers (explanatory variables) used in the study were elevation (ELVN), parent material (ROCK), slope (SLOP), radiation index (RI) and topographic position index (TPI).

DA, a supervised classification, involve the determination of a linear equation like regression that would predict which group the case belonged to. DA assume a multivariate normal distribution of the predictor variables and a common within group co-variance of the variables for all points defining presence and absence of a vector (Rogers et al., 1996). Like DA, LR, a generalized linear model (GLM) used for binomial regression, was used to distinguish between two or more groups. While either technique can be used for such applications, LR was generally preferred when the response variable was dichotomous (presence/absence) (Felicísimo et al., 2004). GAM is a non-parametric version of GLM, which is produced from multiple linear regression analysis (MLR) (Moisen et al., 2006). GAM uses transformation techniques that are independent for each predictor variable, which are counted together to calculate the response variable (Guisan and Zimmermann, 2000). This allows exploration of shapes of species response curves to environmental gradients, and allows the fitting of statistical models in better agreement with ecological theory (Austin, 2002; Lehmann et al., 2002). CT is a nonparametric tree building technique. The essential purpose of CT is to partition the main data into homogeneous subgroups. In this way, the data is represented by a tree structure (Breiman et al., 1984; Navarrate et al., 2011).

As mentioned before, MAXENT and GARP are profile techniques in order to model potential distribution of species by using only presence data. MAXENT focuses on fitting a probability distribution for occurrence of the species in question to the set of pixels across the study region, based on the idea that the best explanation to unknown phenomena will maximize the entropy of the probability distribution, subject to the appropriate constraints. In case of modeling ecological niches of species, these constraints consist of the values of those pixels at which the species were detected (Phillips et al., 2006). GARP produces sets
of rules that delineate ecological niches in an artificial-intelligence-based approach. Occurrence points are resampled randomly to create test and training data sets and the algorithm works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection (Tsoar et al., 2007).

To identify the best model, the Area Under the Curve (AUC) of the Relative Operating Characteristic (ROC) was calculated for each model (Fielding and Bell, 1997).

Results and Discussion

A DEM of the Yukarigökdere district was used for calculating four digital terrain models (i.e. ELEV, SLOP, RI and TPI) as raster layers with a spatial resolution of 1 ha (100x100 meter). Parent material map composed of 3 main bedrock types (i.e. limestone group including neritic limestone, pelagic limestone and charty limestone with Halobia (a), conglomerate (b) and ophiolitic melange (c)) was also constructed by transforming to the same size grid. As mentioned before, response data was obtained from 119 sampling points and the presence of Crimean juniper at 91 sampling points were detected (Fig. 1).

DA was performed with a stepwise procedure. The obtained function having a Wilks’ Lambda value of 86.4% was found highly significant (p< 0.000). Instead of canonical discriminate function, fisher linear discriminate functions (\(Y0=-17.092+0.025*\text{ELEV}-0.025*\text{TPI} \) and, \(Y1=-20.220+0.028*\text{ELEV}-0.009*\text{TPI}\)) were used to generate a potential distribution map due to fact that variance-co-variance matrices of the function were found equivalent (Box’s M=0.717, p=0.874). To this end, the standardized values (\(Y_i=x_i / (x_{\text{max}}-x_{\text{min}})\)) of the differences (\(x_i\)) between the estimated values obtained from \(Y0\) and \(Y1\) equations were calculated for each cell.

Fig. 1: The study area and the sampling plots occupied and unoccupied by Crimean juniper
Likewise, ELVN and TPI were found best predictors as a result of applied LR with a forward stepwise procedure. Nagelkerke R Square showed that about 21.7% of the variation in response variable was explained by the logistic model. The obtained LR model was: The occurrence of Crimean juniper tree probability ($p$) =\[\exp\left(-3.544 + 0.004 \times \text{ELVN} + 0.019 \times \text{TPI}\right) / (1+\exp\left(-3.544 + 0.004 \times \text{ELVN} + 0.019 \times \text{TPI}\right)).\]

10 fold cross-validations were run to obtain optimal CT model. This tree model produced 3 nodes and 2 splits. The tree model was built by ELVN and TPI (Fig. 2). According to tree model, the most suitable sites for Crimean juniper seemed to be the areas approximately above 1250 m (Fig 2).

Before performing GAM, a quasi-binomial model was chosen. The best model was obtained by means of same variables found in the distribution models of other group discrimination techniques. The stepwise selection of predictors for Crimean juniper selected the following model: The occurrence of Crimean juniper probability= $s(\text{ELVN}, 4) + S(\text{TPI}, 4)$ (explained deviance: 41.5%).

Where $s$ is spline smoother, 4 is d.f. for the spline smoother. The predictor’s partial response is shown in Fig. 3, where most suitable sites for Crimean juniper were characterized by ELEV values approximately 1500 m asl and TPI values between 0 and 50.

As for profile techniques, the most important factors contributing MAXENT distribution model were found to be TPI and RL (Fig 4). Since GARP used a random approach for the rule sets and the predictions varied between runs, ten models were generated using all model approaches preferring 50 % training and 50% test option. A final model, containing summary of ten models provided the potential distribution in percentage of each species. ArcInfo GIS was used to create the potential suitability models as maps and given in Fig. 5. According to the results of the applied ROC analysis, GAM was found the best model with a AUC value of 0.869. The AUC value of CT was found 0.778, while those values for DA and LR were 0.762 and 0.764, respectively (Fig. 4). MAXENT and GARP had lowest ROC values of 0.686 and 0.712, respectively.

In the present study, expected results were achieved from the performances of statistical approaches point of view. Namely, the model results of MAXENT and GARP were failure because those techniques had a disadvantage since the beginning of the statistical process due to fact that they created models by using only presence data of the species as response data (Pereson et al., 2007; Wang et al., 2010).

With regard to comparisons of group discriminate techniques with each other, DA is a linear statistical method that falls behind to explain nonlinear relationships. To use DA, the explanatory variables should be normally distributed as well (Blackard and Dean, 1999; Manel et al., 1999).

However, explanatory variables (environmental variables) are generally non-normal. Because of this reason, this method is rarely used for spatial distribution modelling. Unlike DA, LR can be used when the explanatory variables are non-normal (Pohar et al., 2004).

The relationship between the predictor and response variables is not a linear function in logistic regression; instead, the logistic regression function is used, which is the logit transformation of a probability of occurrence ($p$). In other words,
the logistic or logit function was used to transform an ‘S’-shaped curve into an approximately straight line and to change the range of proportion from 0–1 to -∞ to +∞ (Ozkan and Şentürk, 2012).

However, even though LR was used for a long time as a predictive distribution model, it was inadequate for curvilinear or bell-shaped nonlinear relationships. To explain such relationships, the best method was CT or GAM. Those methods have been, therefore, much more commonly used than LR to predict species distribution. In comparison to GAM with CT, GAM gives generally more interpretive results than CT. Because the number of prediction values of CT are based on the number of terminal nodes (Ozkan, 2013) whereas GAM model predicts the values referring the values of explanatory variables (Guisan et al., 2002).

In addition to the properties of distribution models, it should be noted that there generally existed a bell shaped relationship between distribution of organisms and environmental conditions.
factors in the mountain forest ecosystems in particular karstic mountain forest districts like Yukarıgökdere forest district. That was probably the main reason that GAM was found to be the best predictive model in this study.

Predictive distribution models are important tools to generate potential distribution or habitat suitability maps of plants and animals. In comparison with profile techniques, group discriminate techniques seem more reliable according to the results obtained from this study. The reason of this is probably due to fact that real absence-presence data of the species is used by group discriminate techniques.

Properties of the focal species, size of the study area, qualities and quantities of environmental layers and accuracy and sufficiency of collected inventory data of the species also play important roles to obtain an accurate distribution model as well as the selection of the best one of the modeling techniques.

The Mediterranean region has a vast and rich forest potential. Majority of Mediterranean forests have been degraded. That is why mapping the potential distributions of focal species is crucial to prepare essential management and conservational plans for the Mediterranean ecosystems.

Even though numerous studies were already focused on the relationships between plant distribution and environmental factors, few considerable studies have been carried out about the spatial distribution modelling of plant species in the Mediterranean region so far (Özkan and Şentürk, 2012; Şentürk, 2012; Özkan, 2013). Therefore the present study is important as an essential reference for future studies on species distribution modelling of the species in Turkey, in particular in the Mediterranean region.

References


Modeling and mapping potential distribution of Crimean Juniper


Özkan, K. and M.G. Negiz: Woddy vegetation classification and mapping by using hierarchical methods in Isparta Yukangökdere district. SDU Faculty of Forestry, J., 12, 27-33 (2011).


