POTENTIAL DISTRIBUTION MODELLING AND MAPPING OF DOG ROSE (Rosa canina L.) IN THE NUR MOUNTAINS OF GAZİANTEP DISTRICT, TURKEY

KARAKAYA, T.1* – YÜCEL, E.2

1Republic of Turkey General Directorate of Forestry, Ankara, Turkey
2Department of Biology, Faculty of Science, Eskişehir Technical University, Eskişehir, Turkey
(phone: + 90-222-321-3550; fax: +90-320-4910)

*Corresponding author
e-mail: turgay26k@gmail.com; phone: +90-312-296-4265; fax: +90-312-296-4203

(Received 13th Jan 2021; accepted 9th Apr 2021)

Abstract. This study was carried out to make potential distribution modeling of dog rose (Rosa canina L.) species found naturally in the Nur Mountains of Gaziantep District, Turkey. In this study data obtained from 79 sample sites was used. Interspecific correlation analysis (ICA) was applied to define indicator species of dog rose. The results of the applied ICA showed that Abies cilicica Carr., Pinus nigra Arnold, and Rubus caesius L. were positively associated with dog rose whereas Pinus brutia Ten. had a negative association, becoming an indicator species. Classification and regression tree technique (CART) and maximum entropy approach (MAXENT) have been used to obtain the distribution models of the species. The optimum model obtained from CART was built by altitude, aspect and heat index. The variables found in the MAXENT model were altitude, slope % and topographical position index. According to the results of ten-fold cross validations, AUC (Area Under Curve) values of the models were found to be higher than 80%. The obtained model-based maps pointed out that habitat as the most suitable sites of dog rose are the lower and middle parts of the valleys found in the upper zone of the Nur Mountains.

Keywords: habitat suitability, modelling, non-wood forest products, profile techniques, discrimination methods, dog rose

Introduction

Plant species diversity of Turkey is mostly found in forest ecosystems. Non-timber forest products make up the majority of biological diversity in forest ecosystems. These products are ecologically, biologically as well as economically important and are widely used in pharmaceutical industry, food, chemistry, and cosmetic sectors. For this reason, the importance given to non-timber forest products in Turkey is increasing. However, due to reasons such as unplanned, faulty and excessive utilization, some species are endangered. In order to meet the demand for non-timber forest products of high economical value and maintain sustainable use, it is necessary to conserve and use them in a planned manner and expand their expansion areas via plantation. The number of studies regarding these non-timber forest products of high economic value has increased in Turkey.

Located at the intersection of Iran-Turanian and Mediterranean phytogeographical regions, The Nur Mountains in Gaziantep province are rich in forest resources and non-timber forest products due to their climate and topographic features. One of these non-timber forest products is Dog rose (Rosa canina), which is ecologically and economically important and is native to The Nur Mountains. Rosa canina can easily adapt to extreme climate conditions and therefore is most widely distributed in Turkey and in the world (Nilsson, 1972). Some studies have been carried out in Turkey...
regarding habitat requirements and geographical distribution of *Rosa canina*, but both international and national studies regarding *Rosa canina* fruit have largely focused on its taxonomic, genotypic, phenological, pomological, palynological, and morphological features (Uggla, 1991; Kazankaya et al., 2001; Ercişli and Eşitken, 2004; Güneş and Dölek, 2010; Yıldız and Çelik, 2011; Ekincialp and Kazankaya, 2012).

There are many studies about the relationship between ecological requirements and habitat factors of plant species. However, these studies have been carried out on woody plants in forest ecosystems. Recent years have also seen studies about determining ecological features in terms of vegetation and environmental relationships (Eser, 2014; Özkcan, 2004; Negiz, 2013; Karatepe, 2004; Fontaine et al., 2007). These studies have focused on the effects of site factors on the distribution and growing of plant species, determining optimum site conditions, and site factor classification. Moreover, it has also been found out that different site factors have positive or negative effects on plant growth, productivity, expansion and diversity. Plant species are distributed in areas appropriate to their habitat requirements and these areas are potential distribution areas for these species; nonetheless, due to factors such as unplanned, faulty, and excessive utilization; plant diseases and fires, actual expansion areas of some species do not represent potential distribution areas. It is necessary to carry out studies regarding model-based potential distribution and habitat suitability mapping of species to conserve target species in today’s climate conditions, expand their distribution areas, determine potential geographical area changes in case of future climate changes, and take precautions for sustained utilization of species in the light of this information (Özdemir et al., 2020).

Plans for non-timber forest products require determining the ecological features of species. Studies regarding these issues have gained importance within the last years. Together with the development of geographical information systems and statistical methods and software, scientific studies based on analytical methods have been carried out in many parts of the world regarding distribution modeling of trees, bushes, and grass species (Vayssieres et al., 2000; Lehmann et al., 2002; Vetaas and Grytnes, 2002; Thuiller et al., 2003; Dobrowski et al., 2006; Wang et al., 2006; Vaclavik and Ortega, 2008; Özdemir, 2018). In scientific research carried out in Turkey, geographical distribution models of species have been created and in some studies, model-based potential distribution maps of plant species and plant species diversity have been developed (Şentürk, 2012; Şentürk et al., 2010, 2014; Gülsoy, 2011; Özkcan and Şentürk, 2012; Özkcan, 2013a; Gülsoy et al., 2014; Özdemir and Özkcan, 2016).

This study aims to reveal ecological features and the relationships between habitat features and geographical distribution, to determine ecological indicators be considered when determining potential geographical distribution, and to create an analytical model using these indicators and a potential distribution map of *Rosa canina* in The Nur Mountains.

**Materials and methods**

**Study area**

The research material of this study is *Rosa canina* and the study area is The Nur Mountains located in the west of Gaziantep province. The Nur Mountains are geographically located at the intersection of the Mediterranean Region, South-east Anatolia Region, and at the intersection of Iran-Turanian and Mediterranean phytogeographical regions (Fig. 1).
The climate of this province has unique features thanks to its geographical location. The climate features of the region were determined using the average 30-year data taken from Islahiye Meteorological Station, the closest to the study area (TSMS, 2012). According to the obtained data, the annual average heat is 16.8°C. The coldest month is January (5.3°C), whereas the hottest month is August (28.2°C). North and west winds are the dominant winds in the study area. The dominant wind direction is the north-west. The annual average relative humidity is 60% and the annual average precipitation is 791.2 mm, 50% of which falls in the winter. Elevation, aspect, orographic features, and air masses affect the amount of precipitation, and thus, the amount of annual precipitation is higher in the Nur Mountains since humid air masses of the Mediterranean meet the Amanos Mountains. Air masses get colder while passing over these mountains and cause orographic rains. All these show that the Mediterranean precipitation regime is dominant in the region and the study area is located in the climate zones of the Mediterranean climate and continental climate (Özuslu and Tel, 2011).

Flora in the Nur Mountains are as follows: scrub vegetation consisting of bushes, deciduous broadleaf tree communities of the Mediterranean mountain zone, evergreen coniferous trees, and grasses. Bush formation in the area distributes at an elevation of between 450 and 1000 m. Bush species consisting of both deciduous and evergreen scrub vegetation affected by continuous forest destruction distributes at these elevations. Forest formation is found at elevations of between 550 and 2000 m. The dominant tree species in this formation is Pinus brutia (Özuslu and Tel, 2011). Relict forest communities belonging to Colchic-Euxine flora region that inhabits the north-facing slopes and local areas and exist under humid climate conditions are found in the study area. There are also oriental beech trees (Fagus orientalis), which normally belong to the Euxine flora region, in the north-facing slopes at 1100-1500 m (Topaloğlu et al., 2016).

Data collection

Before doing the fieldwork, digital topographic map was used and expeditions were made into the region in order to determine possible sample locations. Because choosing
the sample locations to depend on grid net cells obtained from the climate data, bio-climate data was downloaded from the website http://www.worldclim.org (WorldClim 1.4 climate data for 1960-1990.). The maps downloaded were cut in the analysis area scale in the Global Mapper 14 software, converted into vector format, and a grid net was created. This grid net was transferred to GPS and inventory studies were done within the related cell limits. Then 79 sample locations were identified via the random distribution method (Fig. 1). Slope, aspect, and elevation variables of the area were taken into consideration. Moreover, it was important for the sample locations to represent the study area and to be distributed homogeneously while choosing the sample locations. After this process, field inventory study was commenced. The studies were carried out in sample locations sized 20 x 20 m. and suitable for ecological research (Fontaine et al., 2007; Özkan and Gülsoy, 2009; Yücel and Güner, 2010). It was determined whether Rosa canina is found in sample locations or not and habitat factors were recorded in the inventory. Furthermore, herbaceous plant species and woody plant species were recorded on a table (present-absent 1-0) and vegetation data matrix was created. All the obtained data was transferred to Microsoft Office Excel program.

Creating digital layers of ecological variables

Geographical information systems (GIS), software, various parameters, algorithms, digital layers, and data sets are used to determine potential distribution areas of species. In order to determine ecological features of a species, obtain potential distribution models, and map them, it is necessary to create or obtain digital layers of all the related variables in the model (Mert et al., 2013). Ecological variables used in potential distribution modelling of the target species are elevation, aspect, slope; topographic position index, landform index, aspect suitability index, radiation index, and heat index and they are in the cell value of 100 × 100 m.

Latitude and longitude, elevation, aspect, slope, topographic position index, and landform index maps were created using digital elevation model (DEM), while aspect suitability index, radiation index, and heat index maps were created using ArcGIS 10.2 package program (ESRI, 2011).

In order to create digital layers of ecological variables, topographic maps with a scale of 1/25000 were obtained and digital elevation model map was created using ArcGIS 10.2 package program with these maps. Aspect and slope maps were obtained using digital elevation model map. Topography tools extension of a software developed by Jenness (2006) was used to create topographic position index and landform index.

Topographic position index enables to classify the area by surface shapes and slope values using cell values of elevation steps. It compares average elevation values of neighbouring cells according to elevation values of each cell. Positive values show high landforms like hills or mountains, while negative values represent landforms such as valleys and canyons. Values approximate to zero show plains, ridges, or stable slope lands (Jenness, 2006; Weiss, 2001). Landform index makes classification by topographic position index model in at least two scales. TPIN models of different scales create various land variability types (Jenness, 2006; Tagil and Jenness, 2008). Aspect suitability index (ASI) (Eq. 1), radiation index (RI) (Eq. 2), and heat index (HI) (Eq. 3) values were calculated using different equations.

The equation for creating aspect suitability map is:

$$ASI = \cos(A_{max} - A) + 1$$  \hspace{1cm} (Eq.1)
Amax 202.5°, A represents aspect of which values are radians. The values of this equation range between 0 and + 2 (Ewald, 2000; Vanderpuye et al., 2002).

The equation for creating radiation index is:

$$RI = \frac{1-\cos\left(\pi/180 \cdot (0-30)\right)}{2}$$

(Eq.2)

Q represents aspect value. The values of this equation range between 0 and 2. Values in the north and northeast directions approximate to 0, while in the hotter and drier south and south-east directions, these values approximate to 2. (Moisen and Frescino, 2002; Aertsen et al., 2010; Wei et al., 2010; Brown and Ahl, 2011).

The equation for creating heat index map is:

$$HI = \cos(alfa_1) \times \tan(alfa_2) = (\cos(A_{max} - A) + 1) \times \tan(Slope)$$

(Eq.3)

Amax represents 202.5° and A represents aspect. 202.5 ° is assumed to be the highest heat load on the slopes facing southwest. alfa2 represents slope (Austrheim et al., 1999; Zelený and Chytry, 2007; Pal Axel et al., 2009). Heat index values range between 0 and 2.

19 different pieces of bio-climate data created by Hijmans et al. (2005) were downloaded from the website http://www.worldclim.org to create maps for climatic variables and cut according to the analysis area scale on Global Mapper 14 package program. They were converted into vector forms and grid net of the analysis area was created.

Ecological variables

ArcGIS 10.2 program was used for modeling and mapping studies. The data matrix was created with digital values of climate and land variables in the sample locations from digital layers. Ecological variables and codes were presented in Table 1.

Statistical assessment method

Interspecific correlation analysis

Interspecific correlation analysis was suggested by Cole (1949) and started to be used in ecology (Özkan, 2002). In this study, interspecific correlation analysis was done using SPSS 26.0 software package (IBM, 2019) with vegetation data matrix (created using herbaceous and woody plant species found in sample locations) and bedrock formation data in the sample locations. The analysis was carried out in order to determine positive and negative indicator plant species and indicator bedrock formations according to presence-absence data of Rosa canina (dependent variable) in sample locations (Özkan, 2002).

Factor analysis and Pearson correlation analysis

Climatic variables used in the study are highly correlated with each other. Furthermore, in studies carried out on distribution modeling of species in Turkey, it was stated that climatic variables might be highly correlated with elevation variables, which might cause multicollinearity problem (Eser, 2014; Şentürk, 2012; Süel, 2014). In order to prevent this, to obtain clearer analysis results and models that are more reliable, and
to eliminate multicollinearity problem, factor analysis among climatic variables was performed using SPSS 26.0 package program.

With the factor analysis, variables that can represent each other among many bio-climate variables were reduced. Component matrix belonging to bio-climate variables were obtained with variance and variance participation rate values that appeared according to the components. Bio-climate variables with the highest coefficient were determined as a result of the factor analysis. Following this, Pearson correlation analysis was carried out to identify the correlation between the bio-climate variables with the highest coefficients and elevation variables. This bio-climate variable was taken as the basis for modeling and mapping studies to be done after this stage.

Table 1. Codes and explanations of ecological variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation (m)</td>
<td>ELVT</td>
<td>Lowest heat of the coldest month</td>
<td>BIO6</td>
</tr>
<tr>
<td>Aspect (°)</td>
<td>ASPECT</td>
<td>Annual heat</td>
<td>BIO7</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>SLOPE</td>
<td>Average heat of the most humid first 3 months</td>
<td>BIO8</td>
</tr>
<tr>
<td>Deep Canyon</td>
<td>DEECAN</td>
<td>Average heat of the driest first 3 months</td>
<td>BIO9</td>
</tr>
<tr>
<td>Shallow Valley</td>
<td>SHAVAL</td>
<td>Average heat of the warmest first 3 months</td>
<td>BIO10</td>
</tr>
<tr>
<td>Dry Stream</td>
<td>DRYSTR</td>
<td>Average heat of the coldest first 3 months</td>
<td>BIO11</td>
</tr>
<tr>
<td>U-shaped Valley</td>
<td>USHAPE</td>
<td>Annual precipitation</td>
<td>BIO12</td>
</tr>
<tr>
<td>Lowland-Plain</td>
<td>LOWPLA</td>
<td>Precipitation of the most humid month</td>
<td>BIO13</td>
</tr>
<tr>
<td>Open Slopes</td>
<td>OPSESLO</td>
<td>Precipitation of the driest month</td>
<td>BIO14</td>
</tr>
<tr>
<td>Upper Slopes</td>
<td>UPPSLO</td>
<td>Seasonal Precipitation</td>
<td>BIO15</td>
</tr>
<tr>
<td>Hills in Valley</td>
<td>HILVAL</td>
<td>Precipitation of the most humid first 3 months</td>
<td>BIO16</td>
</tr>
<tr>
<td>Middle Slopes</td>
<td>MIDSLO</td>
<td>Precipitation of the driest first 3 months</td>
<td>BIO17</td>
</tr>
<tr>
<td>Summit - Back</td>
<td>SUMMIT</td>
<td>Precipitation of the warmest first 3 months</td>
<td>BIO18</td>
</tr>
<tr>
<td>Topographic Position Index</td>
<td>TPI</td>
<td>Precipitation of the coldest first 3 months</td>
<td>BIO19</td>
</tr>
<tr>
<td>Aspect Suitability Index</td>
<td>ASI</td>
<td>Alluvium</td>
<td>ALVYN</td>
</tr>
<tr>
<td>Radiation Index</td>
<td>RI</td>
<td>Basalt</td>
<td>BAZLT</td>
</tr>
<tr>
<td>Heat Index</td>
<td>HI</td>
<td>Basic rock</td>
<td>BZKYC</td>
</tr>
<tr>
<td>Annual Average Heat</td>
<td>BIO1</td>
<td>Carbonates</td>
<td>KRBNT</td>
</tr>
<tr>
<td>Day Class Average</td>
<td>BIO2</td>
<td>Limestone</td>
<td>KIREC</td>
</tr>
<tr>
<td>Isotherm</td>
<td>BIO3</td>
<td>Quaternary</td>
<td>KVTRN</td>
</tr>
<tr>
<td>Seasonal Heat</td>
<td>BIO4</td>
<td>Cuttings</td>
<td>KIRINT</td>
</tr>
<tr>
<td>Highest Heat of the Hottest Month</td>
<td>BIO5</td>
<td>Ophiolitic Melange</td>
<td>OFLML</td>
</tr>
</tbody>
</table>

Maximum entropy (Maxent) approach

Maximum entropy (Maxent) approach was employed to obtain potential distribution modeling of *Rosa canina*. The modeling was done using logistic regression option in Maxent 3.3.3k package program and also the data matrix belonging to the ecological variables were used (Maxent, 2016). Presence data in the sample locations of *Rosa canina* were taken as dependent variable in distribution models of the species. Maximum entropy model presents suitability model assessing the data of sample locations where the target species exist and areas with similar features (Baldwin, 2009). Maxent is a model based on presence data in the sample locations and enables to obtain more realistic distribution and suitability models with less data (Wisz et al., 2008; Hernandez et al., 2006; Mert and Kiraç, 2017; Oruç et al., 2017). Maxent approach
makes it possible to examine correlations among ecological variables that might have an effect on the distribution of the target species, and thus we can eliminate the negative effects of ecological variables with high correlations on potential distribution models. The main principle of maximum entropy approach is that it explains that a random variable and a related uncertainty measure exist (Elith et al., 2010). Entropy approach gives a measure of how many options might be included in the selection process (Shannon, 1948). Maxent method shows the distribution of the likelihood of the presence of target species in each grid in the analysis area over the whole analysis area (Yost et al., 2008). Maxent software begins with a uniform distribution and performs multiple iterations that increase the probability of the species sample locations.

Maxent program supports three different models: raw, logarithmic, and logistic formats. Logistic format was used in this study since it shows probability values in the range of 0 and 1, can be interpreted more correctly, and gives the most realistic values (Baldwin, 2009; Phillips, 2008, 2010). Maxent software enables to iterate the model as many times as desired and therefore results can be compared every time the software is run and average values of the model and sufficient information can be obtained. In our study, model was iterated 10 times and model output were compared and evaluated. Suitability models that constitute the model reveal one probability for 0.01 km2-cells in the area. It is possible to find out to what extent the area is suitable for the target species by joining these probabilities. Cells are graded from suitable to unsuitable for the distribution of the target species in the study location and as a result, distribution map can be obtained (Freeman and Moisen, 2008; Phillips, 2010).

After the models were created using Maxent software and effects and relations of ecological variables that affect the models were evaluated using Jackknife statistics and decision curves analysis (Phillips, 2010). In this method, one ecological variable is excluded in each iteration; the remaining variables are used and the final gain is calculated. Then the model calculates a second gain using the variable of which effect is wanted to be calculated. Finally, all the variables are added to the analysis and total gain for the model is calculated. The gain is used to measure the contribution of the variables to the model. If total gain decreases significantly in the absence of a variable, this variable has an effect on the distribution of the target species. If the gain of one variable is low when used alone compared to the total gain, this variable does not have an effect on the distribution of the species. The effects of the ecological variables that constitute the model and the relation between the variable and the target species are understood from the marginal response curves, which are explained through the direction and degrees of these relations (Süel, 2014).

**Accuracy and selection of models**

In order to examine the accuracy of the models obtained through Maxent method, curves obtained through ROC (Receiver Operating Characteristic) analysis were used. ROC curves are composed of sensitivity and specificity indices. Sensitivity index shows how accurately the data that constitutes the model output predicts the probability layer that specifies which area the species exists as a result of the model. Specificity index is the accuracy measure of the prediction that the species does not exist in the area. According to the classification made by the values that are under the curve obtained through ROC analysis, it is seen that if the AUC (Area Under Curve) value is higher than 0.90 the model has a perfect explanation; if the value ranges between 0.90 and 0.80, it is good; if the value ranges between 0.80 and 0.70, it is suitable; if the value
ranges between 0.70 and 0.60, it is weak and if the value is lower than 0.60, the model is not successful (Araújo et al., 2005).

The reliability of model results depend on the quality of the data and any deviation in the data causes deviation in the model results. It is important to carefully examine the data sources and choose the data in an unbiased manner. The methods and criteria used for the reliability of the model results, the nature of the data source, and choosing and using the data source affect the power of the model. The power of a model depends on whether multicollinearity problem among independent variables exists or not. For this reason, there should not be a significant deviation problem between prediction values and real values. Explain-ability of a model should be great. It is important that the model is meaningful to the test set. The actual confirmation of the models obtained are done using test data set. The area under the ROC curve should be high. Therefore, the difference between validity test results for both training set and test data set should be small. If the validity value for the test data set of the model is significantly higher than the training data set, a problem arises in the model suitability (Özkan, 2012).

While choosing the most suitable model for the target species among the models obtained via Maxent method, firstly training and test values, ROC analysis iteration values, their standard deviation, and averages were calculated. Training and test sets of the models that are in the reliability range of 95% were included. The model with the highest AUC (Area Under Curve) value in the training set and test set and with the closest values was selected as the best model.

Results

Indicator plant species of Rosa canina

68 woody and herbaceous plant species showing a large distribution in the 79 sample locations have been found in the study location carried out in The Nur Mountains of Gaziantep Province. These species were recorded as present or absent and 19 plant species having a frequency value over 10% have been included in the statistical analyses.

*Rosa canina* species were found in 17 sample locations and had a frequency value of 21.5%. In order to determine indicator plant species belonging to *Rosa canina* species, interspecific correlation analysis was performed and it was found out whether relations of significant variables were negative or positive, and correlation coefficient values were calculated using Interspecific correlation analysis. The results of interspecific correlation analysis were shown in Table 2.

According to interspecific correlation analysis, *Pinus nigra* Arnold. (p: 0.002; ICA coefficient: 0.314), *Rubus caesius* L. (p: 0.004; ICA coefficient: 0.271) and *Abies cilicica* Carr. (p: 0.009; ICA coefficient: 0.261) species were found to be statistically meaningful and positive, whereas *Pinus brutia* Ten. (p: 0.016; ICA coefficient: -0.367) was found to be the negative indicator plant species.

Selection of climatic variables

Bio-climate variables used in the study were highly inter-correlated. In order to obtain clearer results from the statistical analyses, in order for models to be more reliable, and in order to eliminate multicollinearity problem, factor analysis was done on
bio-climate variables using SPSS 26.0 software package. Variables that could represent each other were reduced. Correlation coefficients among the main components were shown in Table 3.

According to the factor analysis results, two bio-climate variables with a variance higher than 1 and a variance participation rate higher than 10% were obtained. Variance participation rate of the first variable is 76.05%, while the second variable had a participation rate of 20.49%. Total variance explain-ability rate of these two variables was found to be 96.54%.

### Table 2. Positive and negative indicator plant species of Rosa canina

<table>
<thead>
<tr>
<th>Species</th>
<th>Chi square</th>
<th>P</th>
<th>ICA Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns</td>
<td>Pinus brutia Ten.</td>
<td>5.849</td>
<td>0.016</td>
</tr>
<tr>
<td>ns</td>
<td>Abies cilicica Carr.</td>
<td>6.796</td>
<td>0.009</td>
</tr>
<tr>
<td>ns</td>
<td>Arbutus andrachne L.</td>
<td>2.441</td>
<td>0.118</td>
</tr>
<tr>
<td>ns</td>
<td>Cedrus libani Rich.</td>
<td>1.527</td>
<td>0.217</td>
</tr>
<tr>
<td>ns</td>
<td>Cistus salviifolius L.</td>
<td>0.091</td>
<td>0.763</td>
</tr>
<tr>
<td>ns</td>
<td>Crataegus monogyna L.</td>
<td>9.639</td>
<td>7.002</td>
</tr>
<tr>
<td>ns</td>
<td>Crataegus orientalis PALLAS EX BIEB. var or.</td>
<td>0.026</td>
<td>0.871</td>
</tr>
<tr>
<td>ns</td>
<td>Echinops onopordum P.H.Davis</td>
<td>2.315</td>
<td>0.128</td>
</tr>
<tr>
<td>ns</td>
<td>Fagus orientalis L.</td>
<td>1.346</td>
<td>0.246</td>
</tr>
<tr>
<td>ns</td>
<td>Juniperus oxycedrus L.</td>
<td>1.964</td>
<td>0.161</td>
</tr>
<tr>
<td>ns</td>
<td>Paliurus spina-christi Mill.</td>
<td>0.064</td>
<td>0.800</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns</td>
<td>Pinus nigra Arnold.</td>
<td>9.629</td>
<td>0.002</td>
</tr>
<tr>
<td>ns</td>
<td>Quercus cerris L.</td>
<td>2.202</td>
<td>0.138</td>
</tr>
<tr>
<td>ns</td>
<td>Quercus coccifera L.</td>
<td>0.247</td>
<td>0.619</td>
</tr>
<tr>
<td>ns</td>
<td>Quercus infectoria Olivier.</td>
<td>3.034</td>
<td>0.082</td>
</tr>
<tr>
<td>ns</td>
<td>Rhus coriaria L.</td>
<td>2.135</td>
<td>0.128</td>
</tr>
<tr>
<td>Positive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ns</td>
<td>Rubus caesius L.</td>
<td>8.254</td>
<td>0.004</td>
</tr>
<tr>
<td>ns</td>
<td>Spartium junceum L.</td>
<td>0.788</td>
<td>0.375</td>
</tr>
<tr>
<td>ns</td>
<td>Sytrax officinalis L.</td>
<td>0.998</td>
<td>0.318</td>
</tr>
</tbody>
</table>

ns: statistically non-significant. Species determined to be significant as a result of the interspecific correlation analysis (p < 0.05) are expressed in bold.

The matrix of bio-climate variables showed that variable with the highest coefficient in the first component was BIO17 (-0.982), whereas in the second component it was BIO16 (0.920). Bio-climate variables and elevation were highly correlated, which caused multicollinearity problem. In order to prevent this, Pearson correlation analysis was done between BIO16, BIO17, and elevation variables. Pearson correlation analysis showed that there was a correlation between BIO16-and elevation at a level of r: 0.025. The correlation between BIO17 and elevation was found to be r: 0.892. This showed that the correlation between BIO17 variable and elevation variable was above 85%. Thus, it was assumed that elevation variable could represent BIO17 (precipitation of the driest first three months) and it was decided that only BIO16 variable (precipitation of the most humid first 3 months) could be used, which would not cause multicollinearity problem.
Table 3. Correlation coefficient between bio-climate variables and main components

<table>
<thead>
<tr>
<th>Bio-climate variables</th>
<th>Components</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO1</td>
<td>Annual average heat</td>
<td>0.949</td>
<td>0.305</td>
</tr>
<tr>
<td>BIO2</td>
<td>Daily average range (average of daily maximum and minimum heat)</td>
<td>0.908</td>
<td>-0.359</td>
</tr>
<tr>
<td>BIO3</td>
<td>Isotherm</td>
<td>0.926</td>
<td>-0.048</td>
</tr>
<tr>
<td>BIO4</td>
<td>Seasonal heat</td>
<td>0.943</td>
<td>-0.209</td>
</tr>
<tr>
<td>BIO5</td>
<td>Lowest heat of the hottest month</td>
<td>0.979</td>
<td>0.192</td>
</tr>
<tr>
<td>BIO6</td>
<td>Average heat of the coldest month</td>
<td>0.925</td>
<td>0.368</td>
</tr>
<tr>
<td>BIO7</td>
<td>Annual heat range (BIO5-BIO6)</td>
<td>0.698</td>
<td>-0.620</td>
</tr>
<tr>
<td>BIO8</td>
<td>Average heat of the most humid first 3 months</td>
<td>0.938</td>
<td>0.334</td>
</tr>
<tr>
<td>BIO9</td>
<td>Average heat of the driest first 3 months</td>
<td>0.964</td>
<td>0.256</td>
</tr>
<tr>
<td>BIO10</td>
<td>Average heat of the hottest first 3 months</td>
<td>0.963</td>
<td>0.261</td>
</tr>
<tr>
<td>BIO11</td>
<td>Average heat of the coldest first 3 months</td>
<td>0.938</td>
<td>0.334</td>
</tr>
<tr>
<td>BIO12</td>
<td>Annual precipitation</td>
<td>-0.834</td>
<td>0.543</td>
</tr>
<tr>
<td>BIO13</td>
<td>Precipitation of the most humid month</td>
<td>-0.700</td>
<td>0.703</td>
</tr>
<tr>
<td>BIO14</td>
<td>Precipitation of the driest month</td>
<td>-0.953</td>
<td>-0.206</td>
</tr>
<tr>
<td>BIO15</td>
<td>Seasonality of precipitation</td>
<td>0.889</td>
<td>0.420</td>
</tr>
<tr>
<td>BIO16</td>
<td>Precipitation of the most humid first 3 months</td>
<td>-0.357</td>
<td><strong>0.920</strong></td>
</tr>
<tr>
<td>BIO17</td>
<td>Precipitation of the driest first 3 months</td>
<td><strong>-0.982</strong></td>
<td>0.009</td>
</tr>
<tr>
<td>BIO18</td>
<td>Precipitation of the hottest first 3 months</td>
<td>-0.978</td>
<td>-0.100</td>
</tr>
<tr>
<td>BIO19</td>
<td>Precipitation of the coldest first 3 months</td>
<td>-0.357</td>
<td>0.920</td>
</tr>
</tbody>
</table>

Selected representative variables are expressed in bold

Potential distribution modeling and mapping via maxent method

In order to create potential distribution model and map via Maxent method, all ecological variables were included in the analysis on Maxent 3.3.3.k package program and later, modeling was iterated excluding variables with the lowest contribution level according to Jackknife statistics results. 10-fold cross-validation test was iterated 10 times in each modeling stage and the area under the curve as a result of ROC analysis of training and test results were recorded. Data of the locations where target species existed and locations showing similarities with the analysis location were evaluated and more than one suitability model were created. After the modeling stage, effects and correlations of ecological variables that affect the models were evaluated using Jackknife statistics and marginal response curves method. Training and test values, iteration values of ROC analysis, and standard deviation and averages were calculated. Training and test data sets of models in the 95% reliability range were taken into consideration. The model with the highest AUC value in the training set and test set and with the closest values to each other was selected as the best model.

In the model obtained using Maxent method, elevation, slope, and topographic position index variables were the ecological variables establishing the model. Elevation variable made the most contribution, while topographic position index made the least contribution (Fig. 2). According to Jackknife statistics results, it was seen that contributions and effects of the variables on total gain were different. Thus, the model was found to produce the best result. Accordingly, the contribution of elevation variable
was more than the contributions of slope and topographic position index and the biggest lost in total gain occurred when the elevation variable was excluded from the model.

Figure 2. Jackknife statistics result of the model obtained via Maxent method

The effects of ecological variables that constitute the model and the relation between the variables and target species can be understood from the marginal response curves. These curves are explained through the direction and degrees of the relations. According to these results, there was a positive correlation between species choice and elevation, whereas the correlation between slope and topographic position index variables increased up to a level and then decreased (Fig. 3).

Figure 3. Marginal response curves that structure the model

AUC value of the training data set was found to be 0.84 and AUC value of the test data set was found to be 0.81 as a result of ROC analysis. The meaningfulness of AUC value of ROC curve is measured by the size of the area. AUC value ranges from 0.5 to 1. The model created with the proximity of the area under the curve to 1 is a successful
model. Elevation, slope, and topographic position index variables that structured the model obtained via Maxent method were used to create the potential suitability map for *Rosa canina* species and was shown in Figure 4.

![Figure 4. Potential suitability map obtained via Maxent method](image)

The potential suitability map of *Rosa canina* species shows locations belonging to elevation, slope, and topographic index values that meet the suitable ecological requirements affecting the species potential distribution. These areas are located mostly in the middle and bottom valleys in higher parts of The Nur Mountains. Blue zones in the map show the locations not suitable for ecological requirements and habitat conditions of *Rosa canina* species. Locations suitable for the ecological requirements of the species are shown in green, yellow, and red according to their suitability degree. Red shows the most suitable locations where the species might potentially exist in terms of ecological requirements. The Nur Mountains are shown with the rectangle-shaped line and when the model data is extended on the map, application outside the study location is also seen. This shows that the data obtained in the sample locations can be extended and used in areas near the analysis area in order to determine the potential distribution areas of the target species.

**Discussion**

As a result of vegetation detection studies carried out in 79 sample locations, it was understood that *Rosa canina* species exist in 17 sample locations and has a frequency value of 21.5%. In order to determine the indicator plant species of *Rosa canina* species, interspecific correlation analysis was done on 19 plant species over the frequency value of 10%. Whether correlations of significant variables were negative or positive was
determined and correlation coefficient values were calculated via interspecific correlation analysis. The results of interspecific correlation analysis show that *Abies cilicica* Carr., *Pinus nigra* Arnold., and *Rubus caesius* L. were positive indicator species, whereas *Pinus brutia* Ten was the negative indicator species. Accordingly, it is possible to say that in terms of ecological requirements and habitat factors, *Rosa canina* species can expand naturally in areas where black pine, blackberry, and cilician fir grow and *Rosa canina* species can most possibly exist in these areas. Habitats where Calabrian pine grow are less suitable for ecological requirements of *Rosa canina* species and thus it is less possible that *Rosa canina* can grow in these habitats. Indicator species are used to determine potentially suitable habitats for the target species. Many studies have been carried in Turkey to determine indicator plant species for potential distribution areas of many plant and animal species (Karatepe, 2006; Çelik et al., 2006; Özkan, 2002; Negiz et al., 2014; Öztürk and Güvenç, 2010; Gülsoy et al., 2013).

Similarly, there are studies carried out to determine indicator species of *Rosa canina* species and interspecific correlation analysis has been used in these studies. For instance, in the basin of Beyşehir Lake positive indicator species of *Rosa canina* species were *Pyrus elaeagnifolia* (Wild pear) *Prunus spinosa* (Blackthorn) and *Lonicera etrusca* (Honeysuckle), while *Cotoneaster nummularia* (Medlar) species was the negative indicator species (Özkan and Bilir, 2008). In one study carried out in the Central Anatolia Region, it was found that *Cistus laurifolius*, *Pyrus elaeagnifolia*, *Lonicera caucasica* subsp. orientalis, *Quercus vulcanica*, and *Amelanchier rotundifolia* subsp. *Rotundifolia* were positive indicator species (Güner et al., 2011a). In another study carried out in the same region, *Rosa canina* species was shown as the indicator species among woody indicator species in areas where Anatolian black pines could potentially be productive (Güner et al., 2011c). Furthermore, in a study carried out to determine indicator woody plant species for potential habitats of some wild mammals in Gölçük Nature Park, *Rosa canina* species was found to be the positive indicator species for Beech marten (*Martes foina*) (Oğurlu and Aksan, 2013).

19 bio-climate variables used in the study are highly correlated. For the statistical analyses to produce clearer results, for the models to be more reliable, and eliminate multicollinearity problem factor analysis was done on bio-climate variables. Thus, reduction of bio-climate variables was done to identify bio-climate variables to be used in potential distribution modelling and mapping studies of *Rosa canina* species. Variables that might represent each other among bio-climate variables were reduced to less variables via factor analysis. According to these results, two bio-climate variables with a variance of 1 and with a participation rate of more than 10% were obtained. Variance participation rate of the first variable is 76.05%, while the second variable had a participation rate of 20.49%. Total variance explanation rate of these two variables was found to be 96.54%. The matrix of bio-climate variables showed that variable with the highest coefficient in the first component was BIO17 (-0.982) (precipitation of the driest first three months) whereas in the second variable was BIO16 (0.920) (precipitation of the most humid first 3 months). Bio-climate variables and elevation were highly correlated, which might cause multicollinearity problem. To prevent this, Pearson correlation analysis was done between BIO16, BIO17, and elevation variables. Pearson correlation analysis showed that there was a correlation between BIO16-and elevation at a level of 0.025. The correlation between BIO17 and elevation was found to be 0892. This showed that the correlation between BIO17 variable and elevation variable was above 85%. Thus, it was assumed that elevation variable could represent...
BIO17 (precipitation of the driest first three months) and it was decided that only BIO16 variable (precipitation of the most humid first 3 months) could be used, which would not cause multicollinearity problem.

Maxent method was used in the potential distribution modeling and mapping stage and it works with the presence data of the target species in the sample locations. Due to destruction of forest ecosystems, it is possible that the species might have existed previously in areas where absent data of the species was recorded. With this model, the most suitable models that show potential distribution areas in the analysis location were created. Accuracy of the models were examined using various methods. Then potential distribution map of the target species was created via geography information system. In the model obtained via Maxent method elevation, slope, and topographic position index variables were the ecological variables structuring the model and explaining the distribution of the species. Elevation variable made the most contribution, while topographic position index made the least contribution. Jackknife statistics results showed that contributions and effects of the variables on total gain were different. Thus, the model was found to produce the good results. Accordingly, the contribution of elevation variable was more than slope and topographic position index and the biggest lost in total gain occurred when the elevation variable was excluded from the model.

The effects of ecological variables that constitute the model and the relation between the variables and target species can be understood from the marginal response curves. These curves are explained through the direction and degrees of the relations. According to these results, there was a positive correlation between species’ choice and elevation, whereas the correlation between slope and topographic position index variables increased up to a level and then decreased.

AUC value of the training data set was found to be 0.84 and AUC value of the test data set was found to be 0.81 as a result of ROC analysis carried out to determine the accuracy of the chosen model. AUC value ranges from 0.5 to 1 and the model created with the proximity of the area under the curve to 1 is a successful model. According to the classification made by the values that are under the curve obtained through ROC analysis, it is seen that the AUC value ranges between 0.90 and 0.80. In some studies, in order to obtain a reliable model, training and data set values should be high and the difference between them should be small (Özkan, 2012).

The potential suitability map of *Rosa canina* species was created using elevation, slope, and topographic index variable that structure the model obtained via Maxent method. The map shows areas belonging to elevation, slope, and topographic index values that meet the suitable ecological requirements affecting the species’ potential distribution. These areas are located mostly in the middle and bottom valleys in the higher parts of the Nur Mountains. The greatest explanatory variable that structured the model was elevation. In many studies, elevation was found to be effective on biodiversity and distribution of plant species (Eser, 2014; Güner, 2011b; Karatepe, 2004; Fountaine et al., 2007; Özkan et al., 2013; Özkan and Gülsoy, 2009; Şentürk et al., 2013). Elevation factor has a significant effect on climatic variables. Amount of precipitation and heat in the habitat vary greatly depending on changes in elevation (Acar et al., 2001). Hence, this variability has an important effect on the diversity and distribution of plant species (Eser, 2014; Duran and Gürek, 2010; Günal, 2013).

Results of various other studies carried out in Turkey regarding ecological requirements and areas of *Rosa canina* species are similar to the results of our study. For example, in one study on *Rosa* L inhabiting Eskisehir city, it was found that 7 taxa
of the species were present in the analysis areas and their ecological features were examined. It was pointed out in this study that *Rosa canina*, one of these taxa, could exist in various habitats. Also, the species exist in heights between 240 and 1500 m in Eskisehir province, mostly in south aspects, in sandy, argillaceous, loamy, rocky, shallow- deep and humid-dry soils; in spaces under or within forests, along waysides, in anthropogenic step fields; in meadows, fields, or garden borders. They also prefer sunny and open spaces, sandy and dry soils, and south aspects (Yücel, 2001). In a study carried out in the western highlands of Buldan town on the factors that affect geographic distribution of *Rosa canina*, it was revealed that the species exist in heights between 1000 and 1400 m, in north and east aspects. It also had a positive correlation with early Pliocene sedimentary rocks and land shapes, whereas the correlation with heights of 200-1000 m, south aspect, and stony areas was negative. In one study carried out in the basin of Lake Egirdir, it was stated that *Rosa canina* exists up to 1750 m high in different bedrock types (Karatepe, 2006). In another study carried out in the Basin of Beyseshir Lake, the distribution of *Rosa canina* species and the relations between soil and topographic features were assessed. The study revealed that alluvial deposits, rocky land surfaces, dusty soil, and lands rich in organic materials influence the species distribution (Özkan and Bilir, 2008). Moreover, in one study carried out in the Central Anatolia Region, extended additive model method was employed and the variables creating the best model were found to be elevation, aspect, slope, and bedrock type. According to this model, the species preferred heights above 1440 m, shady aspects, mid-sharp or steep lands, and metamorphic bedrock types. Also, climate variables were the most effective factor in the expansion of the species (Güner et al., 2011a). Another study carried out in Sütçüler province on potential distribution modeling and ecological features of the species, it was found that the model was constituted by elevation, bedrock type, radiation index, and aspect suitability index (Kazaz, 2013)].

**Conclusions**

It is thought that the results obtained from this study can be used to evaluate the potential of non-timber forest products as target species and to develop policies about mentioned species. It is also predicted that it can be used to decide where Rosa canina species can be used in the province of Gaziantep. Furthermore, potential distribution map will enable to determination of suitable locations to grow the species for commercial purposes to create income for the locals. Such that model-based maps are the most effective tools that can be used for the mentioned purpose. These maps will also be used as main data sources to realize ecosystem-based administration plans for the sustainability of the regional ecosystems successfully. Model-based potential distribution maps of species are important to develop and implement correct policies for the protection, restoration, and sustainability of ecosystems. Consequently, all the results acquired show that this study is important both in terms of studies to be carried out for doge rose and in terms of future studies with similar methods.

**REFERENCES**


Karakaya - Yücel: Potential distribution modelling and mapping of dog rose (Rosa canina L.) in the Nur Mountains of Gaziantep district, Turkey
- 2758 -


