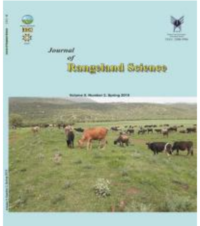
Contents available at ISC and SID

Journal homepage: www.rangeland.ir



Research and Full Length Article:

Comparing Different Modeling Techniques for Predicting Presence-absence of Some Dominant Plant Species in Mountain Rangelands, Mazandaran Province

Mansoureh Kargar[A], Davoud Akhzari[B]*, Amir Saadatfar[C]

[A] Ph.D. Rangeland Sciences,Watershed and Natural Resources Administration of Alborz Province, Karaj, Iran.

[B]Associate Professor, Department of Range and Watershed Management, Malayer University, Iran *(Corresponding author), Email: d_akhzari@yahoo.com

[C] Department of Medicinal Plant, Research and Technology Institute of Plant Production (RTIPP), Shahid Bahonar University of Kerman, Kerman, Iran

Received on: 25/04/2018
Accepted on: 27/10/2018

Abstract. In applied studies, the investigation of the relationship between a plant species and environmental variables is essential to manage ecological problems and rangeland ecosystems. This research was conducted in summer 2016. The aim of this study was to compare the predictive power of a number of Species Distribution Models (SDMs) and to evaluate the importance of a range of environmental variables as predictors in the context of rangeland vegetation. In this study, Aflah rangelands with 5721 ha were selected. In this research, predictor variables included climatic, topographic and edaphic parameters. The sampling method was equal random-classification for vegetation and soil. Topographic factors including slope, elevation and aspect were determined in Arc GIS software. In each sample unit, 10 plots were established (total 350 plots) and the lists of the species, their number, their presence or absence were recorded. The efficacy of five different modelling techniques to predict the distribution of five dominant rangeland plant species (*Agropyron repens*, *Festuca ovina*, *Leucopoa sclerophylla*, *Stachys lavandulifolia* and *Tragopogon graminifolius*) was evaluated. The models were Generalized Linear Model (GLM), Classification and Regression Trees (CART), Boosted Regression Trees (BRT), Generalized Additive Models (GAM), and Random Forest (RF). Data analysis was done using the R software, version 3.1.1. The results showed that GAM model demonstrated most consistently high predictive power over the species in the rangeland context investigated here. GAM had higher Area Under the Curve (AUC). The AUC (0.67, 0.77, 0.69, 0.64 and 0.60 and Kappa values (0.10, 0.10, 0.19, 0.01 and 0.11) were obtained for *Agropyron repens*, *Festuca ovina*, *Leucopoa sclerophylla*, *Stachys lavandulifolia* and *Tragopogon graminifolius*, respectively. GAM model exhibited the most predictive power. The importance analysis of the environmental variables showed that N, pH and aspect were the most important variables in the GAM model. Overall, N, P and C/N soil (0.452, 0.437 and 0.389) were the most important environmental variables.

Keywords: Vegetation, Soil, Topographic, Random Forest, Aflah rangelands

Introduction

Rangeland modeling provides valuable information about rangeland ecosystems (Piri Sahragard and Zare Chahouki, 2016). Species Distribution Models (SDMs) have been applied to a great variety of organisms from viruses (Machado, 2012) and phytoplankton (Hallstan *et al.*, 2012) to vascular plants (Meier *et al.*, 2010; Pellissier *et al.*, 2010; Engler *et al.*, 2011) and lichens (Bergamini *et al.*, 2007) to insects (Maggini *et al.*, 2002; Lütolf *et al.*, 2006; Marmion *et al.*, 2009), birds (Wisz *et al.*, 2007), fishes (Sundblad *et al.*, 2009; Jones *et al.*, 2012) and mammals (Boitani *et al.*, 2007; Rondinini *et al.*, 2011). In addition, quantifying the environmental niche of species (Guisan *et al.*, 1998; Rondinini *et al.*, 2011) is used in SDMs to test ecological (Petitpierre *et al.*, 2012) or evolutionary hypotheses (Vega *et al.*, 2010; Schorr *et al.*, 2012), and/or land use changes (Dirnböck *et al.*, 2003; Vicente *et al.*, 2011) on SDMs are defined similar to statistical analytical algorithms according to field observations and geographical distribution of species range maps of the environment which can be determined by statistical methods (Hengl *et al.*, 2009) based on factors such as the distribution of the dependent variable or response (Poisson, and binomial), type or independent variables (Dubuis, 2013). The emergence of any plant is affected by environmental factors and the relationships between one or more species. If the most effective factors for each plant species can be determined and its behavior with environmental variables and the associated species can be studied, it will be possible to obtain the distribution of species forecasting models (Guisan and Zimmerman, 2000). Data often have a non-constant variance distribution and show many explanatory variables in a collinear manner. Consequently, linear regression cannot be appropriate or may lead to a high unexplained variation (Guisan *et al.*, 2002). Non-parametric and machine learning techniques may be better able to fit the identified problems of linear regression. For plant species, many studies have been carried out to evaluate the predictive performance of habitat models (Garzon *et al.*, 2006; Tarkesh, 2012; Piri Sahragard and Zarechahocki, 2015; Cao *et al.*, 2016; Jafarian and Kargar, 2016; Sor *et al.*, 2017), and they concluded that there was no best modelling technique, but depending on the scope and goal of the study, some techniques will be better suited than others in particular situations (Dubuis, 2013). In the domain of forest site quality assessment, Mckenney and Pedlar (2003) successfully used Classification and Regression Trees (CART) to model site index from environmental variables for two tree species in Canada. The performances of non-parametric techniques such as CART, Generalized Additive Models (GAM) and Artificial Neural Networks (ANN) were compared to parametric techniques by Moisen and Frescino (2002) for the prediction of several independent forest species characteristics in USA. Comparing and ranking different modelling techniques for prediction of site index in Mediterranean mountain forests site, Aertsena *et al.* (2010) showed that BRT is a good alternative in case the ecological interpretability of the technique is of higher importance. When user-friendliness is more important, MLR and CART are the preferred alternatives. Despite its good predictive performance, ANN is penalized for its complex, non-transparent models and big training effort. Tarkesh (2012) compared the performance of six predictive vegetation models. Kappa statistical coefficient was calculated using the receiver operating characteristic and area under the curve. The results showed that MARS and MAXENT had the first and second highest precision, respectively. Many studies have already concluded that there is no general best modelling technique, but depending on the scope and goal of the study, some techniques will be better suited than others in particular situations (Dubuis, 2013). This study can be a good guide to obtain the strengths of different techniques in modelling species. Because the majority of the country's total area including rangeland ecosystems are not satisfactory ecosystems, every action and research that will lead to the better management and planning of the ecosystems are worthwhile. The specific objectives of this study are to compare the modelling techniques with respect to their predictive performance and to investigate the importance of environmental variables as predictors of SDMs to assess which one can improve single species prediction and which one has the highest performance as predictor in the models under study.

Materials and Methods

Study area

The study area is located in the north of Iran, Aflah Rangelands (52°5′–52°11′ N; 35°46′–35°49′ E) (Fig.1). It covers 5000 ha and has an elevation ranging from 2500 to 3910 m form sea level. The climate is semi-arid and cool. The mean annual temperature and precipitation are 10°C and 518 mm, respectively. Climate data were provided from eight climatology stations around the area. Data were used for drawing a bar graph for a 15-year period (1995-2010) (Fig. 2) (Moghari *et al.*, 2014).

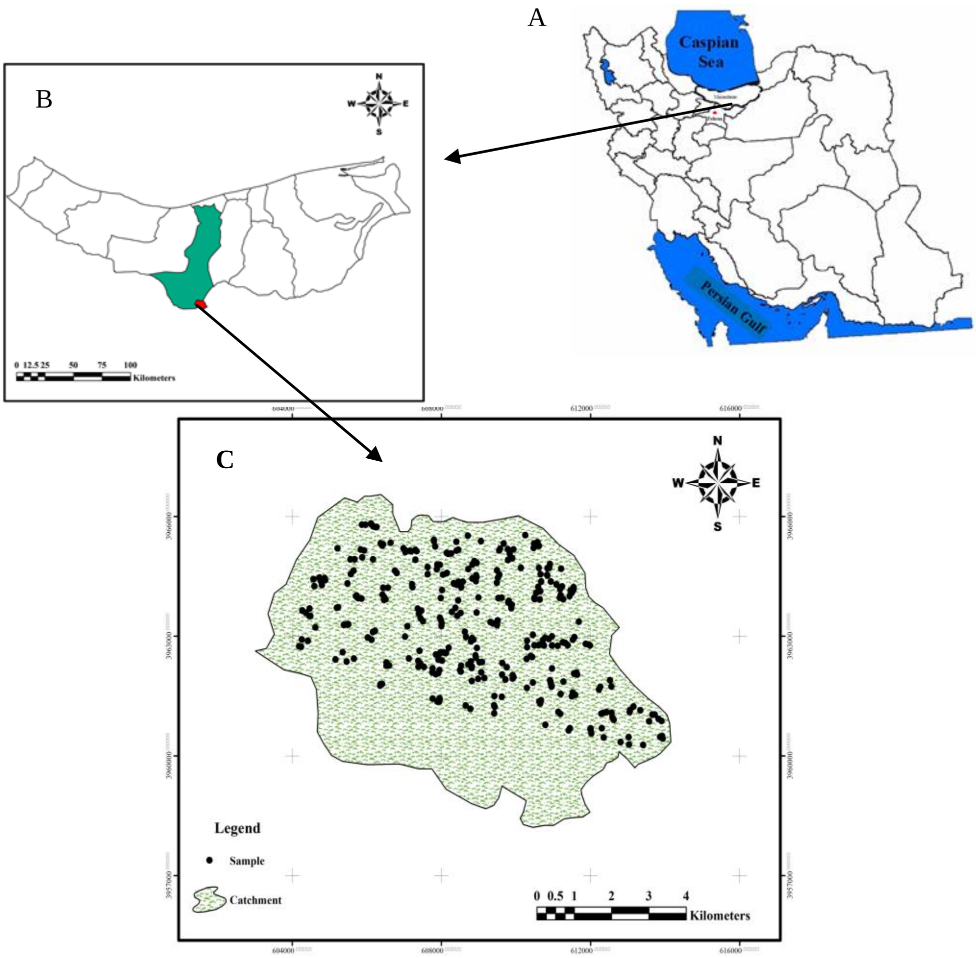


Fig. 1. (A and B) The location of the study area in Iran and Mazandaran province, (C) The presence and absence sites of the species

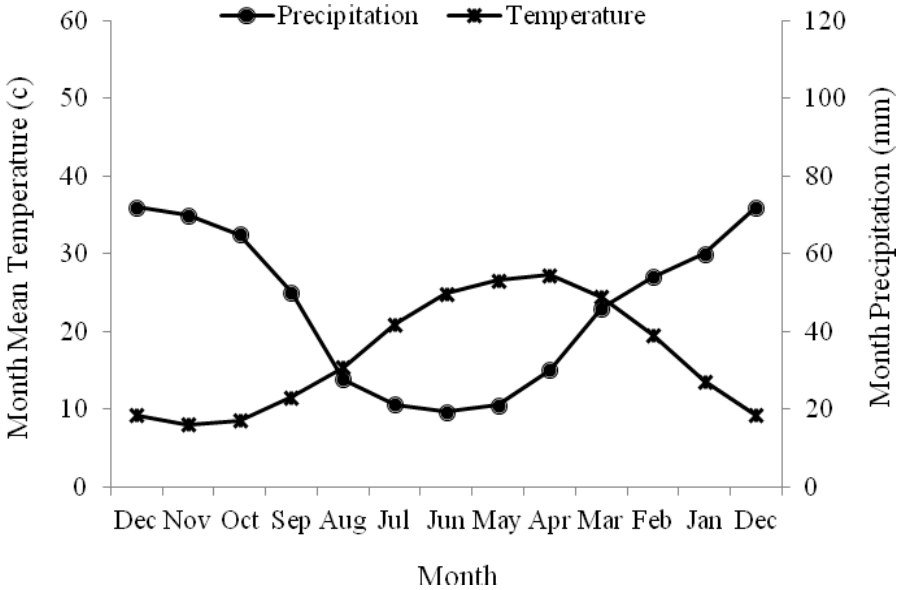


Fig.2. Amberotermic curve of Amol (2001-2017) (References: weather station of Amol)

Data Collection

During the summer of 2016, 350 sampling plots of 1 m^2 were established using a random stratified sampling procedure (Hirzel and Guisan, 2002) based on elevation, slope and aspect, and were extensively inventoried (Pottier *et al.,* 2012). Each sampling point was separated from the others by a minimum distance of 100 m as it has been shown that from this distance onwards, there is no autocorrelation between the plots in the study area. The analyses were limited to the five dominant species. An equal-stratified sampling procedure was used for data collection. The plan provided appropriate distribution and random sampling to facilitate reliable statistical analyses for the area. Primarily three stratifying variables were selected: slope, aspect, elevation. A Digital Elevation Model (DEM) was produced using Arc GIS 10.3 software and slope, aspect, and elevation maps were created. The stratified map was superimposed on the existing species distribution map (scale=1:25000).Then, each was split into several classes. The study area was partitioned: this was done on a map combining the classes with geology to generate 35 homogenous units. Ten 1m^2 quadrates were randomly distributed in each sampling site. In total, 350 plots were established at the study area. Presence and absence of six dominant species in the study area were selected. Soil samples were collected from 0-30cm depth from each sampling site. The samples were air-dried and passed through a 2 mm sieve to be prepared for the tests. The methods used for testing were as follows: The Bouyoucos hydrometer method for soil texture (Beretta *et al.*, 2014), the Kjeldahl method for total nitrogen (Jafari haghighi, 2003), and the modified Walkley-Black wet oxidation procedure for organic carbon content (Pellissier *et al.*, 2010). The pH in the soil/water ratio was 1:1, the total phosphorus was determined colorimetrically from wet digestion with H2SO4+HClO4 (Jafari haghighi, 2003), potassium was determined after extraction by 1N ammonium acetate adjusted with pH 7 (Dubuis, 2013). Soil Maps of slope, aspect, elevation and sample points were overlaid to extract physiographic data at each sampling point.

The data were selected on three climatic variables including average annual precipitation, average annual temperature, and average annual relative humidity. Interpolation of these climatic factors relative to elevation was used to extract the climatic data of the sampling points. To investigate the relative importance of predictor variables in the models, we built a final set of models for each plant species. For each species, the importance of each variable in the models was assessed in BIOMOD by randomizing each variable individually and then recalibrating the model with the randomized variable while keeping the other variables unchanged (R Development Core Team, 2012) (Fig. 3).

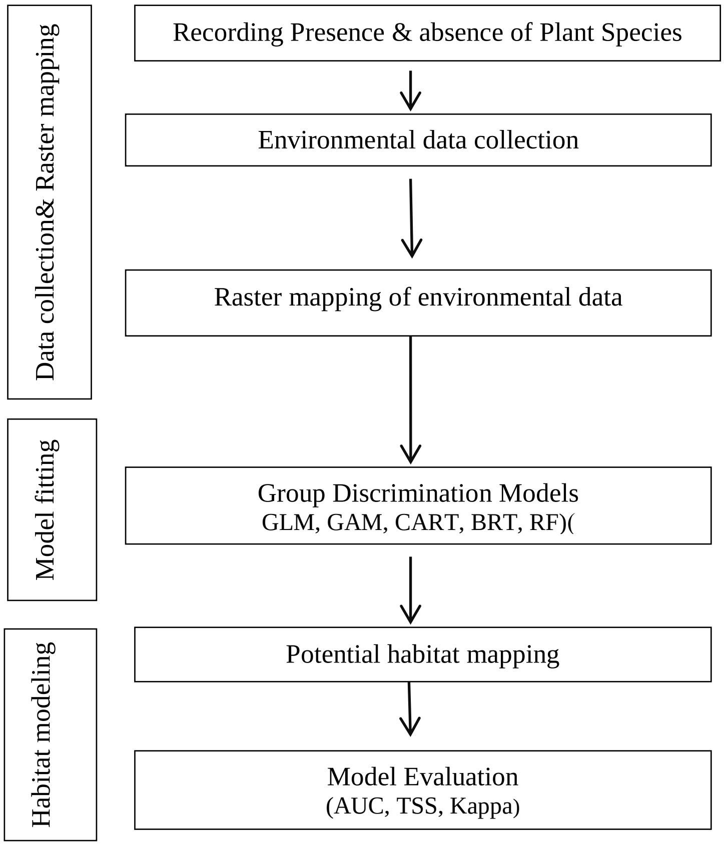


Fig. 3. Flowchart of species distribution modeling using GLM. GAM, RF, BRT, CART

Modelling Techniques

Generalized Linear Model (GLM)

This model is a parametric model developed from linear models. In this model, the formula, the relationship between explanatory variables, and the response provided by the estimated regression parameters in addition to measuring confidence intervals are obtained. GLM is developed for a situation when observations are not normally distributed and when other methods are not suitable regression models (Khalasi Ahvazi *et al*., 2012).

Generalized Additive Models (GAM)

Similar to a generalized linear model, a generalized additive model consists of three steps: 1) Selecting the distribution for the response variable 2) Defining the systematic explanatory variables 3) Identifying the link between the expected values and the systematic response variable. The generalized additive model is not unlike the linear regression model equation (Hastie and Tibshirani, 1990).

Classification and Regression Trees (CART)

CART encompasses a non-parametric regression technique that 'grows' a decision tree based on a binary partitioning algorithm which recursively splits the data until groups are either homogeneous or contain no fewer observations than a user-defined threshold. The predicted value of a 'terminal' node is the average of the response values in that node (Breiman *et al.*, 1984). CART is a popular technique because it represents information in a way that is intuitive and easy to visualize. Preparation of candidate predictors has been simplified because predictor variables can be of any type (numeric, binary, categorical, etc.), and model outcomes are unaffected by monotone transformations and differing scales of measurement among predictors. Regression trees are insensitive to outlets and can accommodate missing data in predictor variables using surrogates (Breiman *et al.*, 1984). The hierarchical structure of a regression tree means that the response to one input variable depends on the values of inputs higher in the tree so interactions between predictors are automatically modelled.

Boosted Regression Trees (BRT)

BRT is a combination of statistical and machine learning techniques. This technique is one of several techniques that use a combination of multiple models to help improve the performance of a single model. BRT uses a combination of two algorithms: CART regression model and composition series of models (Boosting). Boosting overcomes the biggest weakness of a single decision tree that has a relatively weak fitness. The benefits of BRT are as follows: 1) dealing with different types of predictive variables 2) correcting lost data 3) no need to convert or eliminate data outputs 4) fitting complex nonlinear relationships 5) automatically controlling the interaction between the variables. The regression model has been strengthened by R 3.0.1 software and can be used by GBM (Elith *et al.*, 2008).

Random Forest (RF)

Random forest is a new and powerful method that presents considerable developments in data mining technology. Nevertheless, it is relatively unknown in ecological studies. Random forest approach is based on combining data in new ways where a large number of decision trees have been created and then, all the trees are combined together to predict (Cutler *et al.*, 2006). When the predictive variables are identified and targeted, random forests grow a tree in the CART model (Breiman, 2001).

Model Evaluation

Kappa is the maximum one. This means that there is a complete agreement between the actual values and the prediction. Zero values and the prediction of the probability of random or non-random real values are negative, showing that the model is unrealistic. Kappa coefficient matrix is used to calculate the error.

$$\text{Kappa statistic} = \frac{\left(\frac{a+d}{n}\right) - \frac{(a+b)+(a+c)+(c+d)(d+b)}{n^2}}{1 - \frac{(a+b)(a+c)+(c+d)(d+b)}{n^2}}$$

Where:

a is the number of findings that are rated as negative by both raters, b and c are the numbers of findings rated as positive by one rater but negative by the other, and d is the number of findings rated as positive by both raters. There are $a+d$ concordant pairs of ratings and $b+c$ discordant pairs among n pairs of observations.

The predictive power was estimated by the Area Under the Curve (AUC). The AUC of a receiver is the operating characteristic plot (Fielding and Bell, 1997). The AUC values range from 0.5 for models with random predictions to 1 for models perfectly fitting the data. A model is rated as fair if its AUC is higher than 0.7 (Sweets, 1988). The TSS values vary between 0 for a random model and 1 for a model showing perfect agreement.

TSS is calculated by the sum of sensitivity and specificity minus 1 and ranges from −1 to +1 where +1 indicates perfect agreement and zero or less values indicate performance no better than random.

TSS=Sensitivity+specificity-1

Model predictive performance

In this study, 10-fold cross-validation was used to assess the predictive performance of the model. In 10-fold cross-validation, the data are divided into 10 subsets of equal size. The regression technique is then applied 10 times, each time leaving out one of the subsets and using that subset to compute the prediction accuracy. Predictive performance is quantified by calculating model evaluation measures on the predicted values for cross-validation (Maggini *et al.,* 2006).

Results

In overall, 25 SI-models were built using five modelling techniques for each of the five species. All models were critically investigated for environmental factors. The five studied species clearly differ in site needs as expressed by different models. Model predictive performance indicated by higher AUC, Kappa and TSS values revealed that GAM had the highest values of AUC for two species *A. repens* (0.69), and *L. sclerophylla* (0.78) while for RF had high values of AUC for two species *S. lavandulifolia* (0.74), *T. graminifolius* (0.64), and the highest TSS for species (*A. repens*). While for the single species (*F. ovina*), CART had high values of AUC (0.79). BRT was always a poorer predictor than the other models according to all indicators. For *F. ovina*, most models (CART, GAM and RF) have an AUC higher than 0.70, which has been well-performance (Table 1).

Table 1. Performance indices of all SI-models for the five species and five modelling techniques: Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Boosted Regression Trees (BRT), and Random Forest (RF)

Statistical index	GLM	GAM	CART	BRT	RF
Agropyron repens					
AUC	0.67	0.69	0.61	0.67	0.63
KAPPA	0.10	0.17	0.27	0.10	0.11
TSS	0.26	0.22	0.19	0.15	0.12
Festuca ovina					
AUC	0.57	0.77	0.79	0.64	0.76
KAPPA	0.10	0.11	0.18	0.09	0.05
TSS	0.07	0.10	0.10	0.15	0.09
Leucopoa sclerophylla					
AUC	0.69	0.78	0.66	0.51	0.67
KAPPA	0.19	0.12	0.20	0.03	0.02
TSS	0.23	0.18	0.12	0.07	0.23
Stachys lavandulifolia					
AUC	0.64	0.71	0.65	0.68	0.74
KAPPA	0.01	0.11	0.12	0.19	0.12
TSS	0.2	0.18	0.23	0.14	0.19
Tragopogon graminifolius					
AUC	0.60	0.54	0.62	0.51	0.64
KAPPA	0.11	0.16	0.57	0.09	0.03
TSS	0.07	0.09	0.16	0.011	0.03

The analysis of environmental variables for different models was done in BIOMOD. The results of this analysis showed that in the GLM model, elevation and potassium were the most important environmental variables. In the GAM model, nitrogen and pH had more importance. Also, the variables of BRT model are nitrogen, elevation, and slope. In the RF model, carbon to nitrogen ratio and aspect had more importance. Also, in the CART model, organic matter and nitrogen were more important (Table 2).

Table 2. The analysis studied the importance of environmental variables in the package BIOMOD

Variables	RF	BRT	CART	GAM	GLM	Mean
Slope	0.171	0.369	0.231	0.371	0.431	0.321
Aspect	0.429	0.226	0.301	0.438	0.196	0.327
Elevation	0.130	0.367	0.151	0.350	0.205	0.263
Moisture	0.364	0.128	0.110	0.216	0.520	0.307
Precipitation	0.259	0.365	0.295	0.339	0.414	0.344
pH	0.342	0.211	0.218	0.512	0.362	0.356
C/N	0.582	0.522	0.162	0.317	0.390	0.452
OM	0.240	0.351	0.432	0.161	0.084	0.209
P	0.381	0.275	0.118	0.431	0.471	0.389
N	0.370	0.379	0.381	0.764	0.236	0.437
K	0.296	0.196	0.312	0.280	0.538	0.321
Silt	0.320	0.261	0.210	0.338	0.321	0.310
Saturation	0.159	0.310	0.113	0.116	0.328	0.228
P.W.P	0.130	0.177	0.164	0.129	0.202	0.159
Avail. mois	0.259	0.290	0.187	0.251	0.301	0.275
Bulk density	0.164	0.312	0.119	0.317	0.477	0.320

Accordingly, soil properties N, P and pH content, climatic factors Precipitation and topographic factors aspect seem to be common predictors for all species. The five studied species clearly differed in ecological needs as evident from the different models (Table 3). According to the RF model, *Tragopogon graminifolius* responded most to C/N, Elevation, N, and OM, *Stachys lavandulifolia* to Field capacity, Elevation and Moisture, the ANN model *Stachys lavandulifolia* to P content and sand, the GAM model *Agropyron repens* to Silt, and Slope, the CART model *Festuca ovina* to Elevation, P, N and *Leucopoa sclerophylla* to P (Table 3). Among soil properties, N, P and pH were more important than the other soil variables. Among topographic properties, aspect and climatic properties of the precipitation were more important. Also, climatic- topographic properties in *Stachys lavandulifolia* important to other plant species (Table 3).

Table 3. Overview of the predictor variables selected by site index models developed by five modelling techniques

Rangeland species	Modelling technique	Variable(s) selected by the model
Agropyron repens	GLM	pH, P, Field capacity, Moisture, Elevation
	GAM	Silt, Slope
	CART	C/N, Aspect
	BRT	Available moisture Saturation, Field capacity, N, Silt
	RF	pH, Precipitation
Festuca ovina	GLM	pH, C/N
	GAM	P, Aspect, Avail.mois
	CART	Elevation, P, N
	BRT	Available moisture**,** Aspect, Silt, P
	RF	C/N, Elevation, N
Leucopoa sclerophylla	GLM	OM, Aspect, Slope, Field capacity, Bulk density, N
	GAM	P
	CART	P, Silt, Field capacity
	BRT	Saturation, pH, Aspect, Slope
	RF	Precipitation**,** Elevation
Stachys lavandulifolia	GLM	K, Saturation, C/N, Silt
	GAM	Available moisture, C/N, Aspect, Slope, N, P, K, pH, Silt, Precipitation
	CART	C/N**,** Slope, Aspect
	BRT	Aspect, N
	RF	Field capacity, Elevation, Moisture
	GLM	Saturation, moisture, P, N, Precipitation
Tragopogon graminifolius	GAM	P, pH, Precipitation
	CART	Avail.mois, Saturation, N
	BRT	Saturation, Available moisture, PH**,** Aspect, Slope
	RF	C/N, Elevation, N, OM

The variables used by RF were sorted according to a decreasing degree of importance in the modelling: field capacity, saturation and slope. The generalized additive model was clearly more accurate, followed by Random forest model, the regression and classification tree models (Fig.4 and Fig.5). Predicted maps representing the probability of the occurrence of *Festuca ovina* are available (Fig. 6).

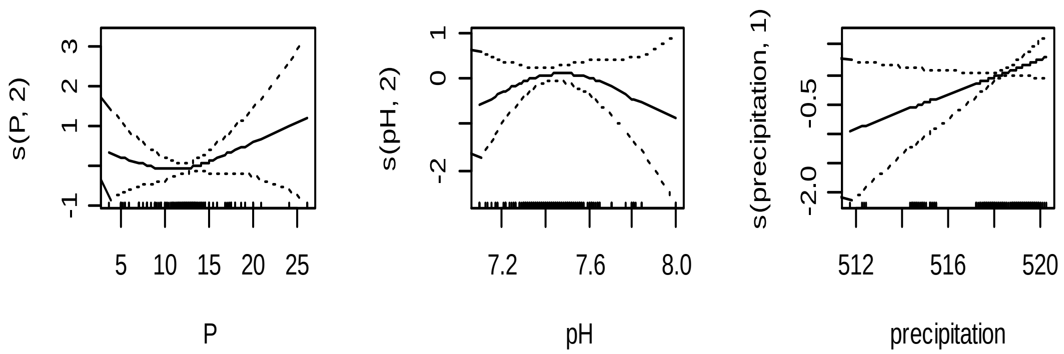


Fig.4. Partial dependence plots of the three predictor variables in the GAM-model for predicting the presence of *Festuca ovina* (full line). Dashed lines represent upper and lower twice-standard-error curves. Rug plots at inside bottom of graphs show distribution of sample sites along that variable.

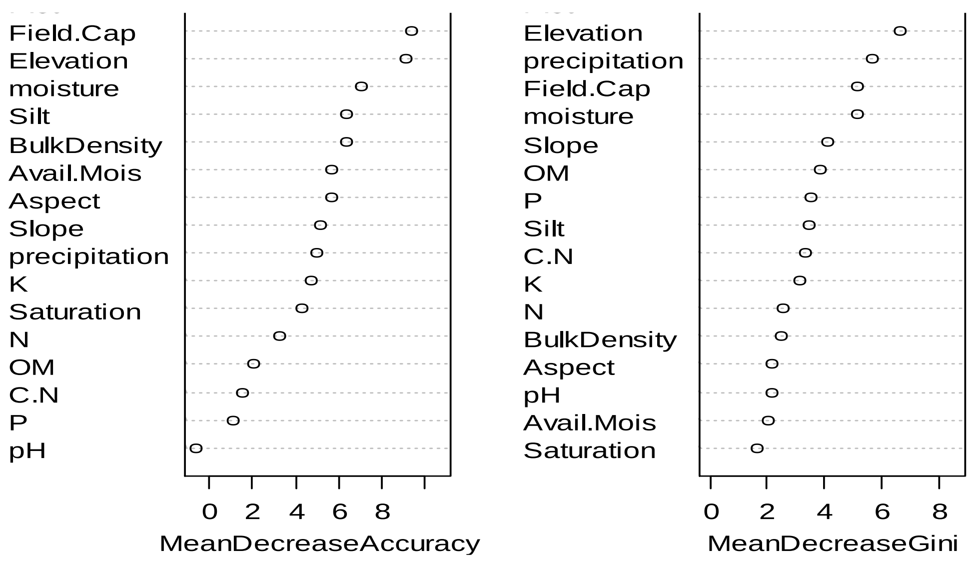


Fig. 5. Variable importance plot generated by random forest algorithm for *Stachys lavandulifolia*

This plot shows the importance of the measured variables as Mean Decrease Accuracy (MDA) and also as Mean Decrease Gini (MDG). The variables are shown by their full names in Table 1**.**

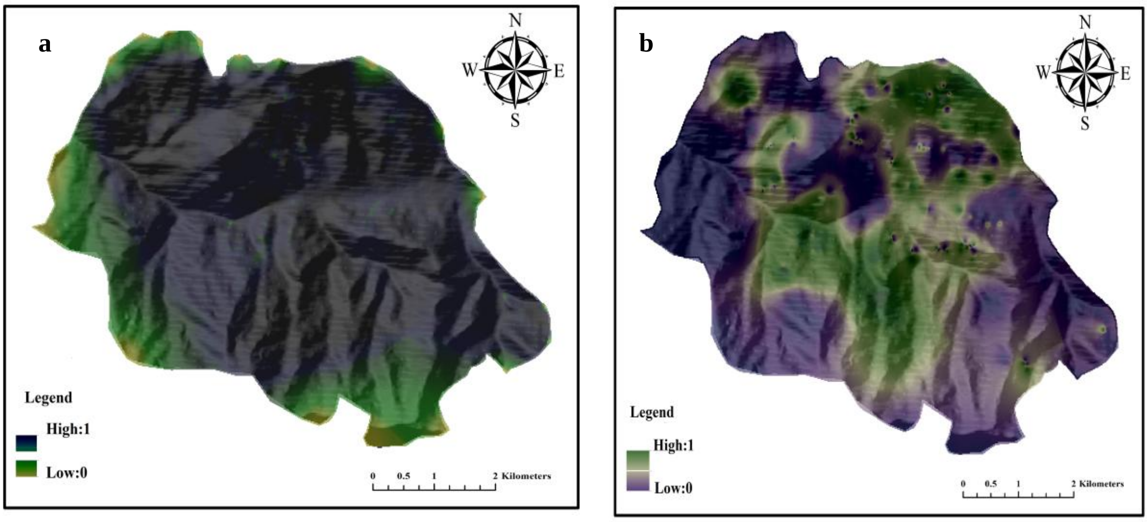


Fig. 6. Predicted maps representing the probability of the occurrence of *Festuca ovina* a) Generalized Additive Model (b) Random Forest Model

Discussion

Modelling the environmental needs of the species in protected areas clearly shows the critical areas where protected the species. Soil properties determine plant species which are in turn effective in the nitrogen cycle and soil properties. Soil texture has a high impact on the control of soil moisture and provides the plants easy access to soil nutrients (Piri sahargard and Zarechahocki, 2015; Cao *et al*., 2016; Asadian *et al*., 2017). The results showed that the most effective soil properties in plant species in rangelands in Iran were soil texture and K. In a specific climatic zone, soil texture will have a more effect on the growth and regeneration of plants compared to chemical fertility (Dubuis, 2013).

The results showed that the slope associated species *Agropyron repens* is consistent with the findings of Jafarian and Kargar (2012) in Polour rangelands in northern Iran. Jafarian *et al.* (2009) identified slope and aspect as the most effective topographical factors in the separation of plant habitats in the studied area. When the transition and movement of materials in the soil is associated with moisture, topographical factors and soil moisture affecting plant growth. In other researches that were conducted for modeling predictions of habitats of various plant species, logistic regression was used. Soil water movement occurs by the control of the effect of soil that is an important factor in the availability of nutrients and a potential factor in soil erosion. This impact can have a high ecological potential. Increase or decrease of the percentage of silt and clay along with increase or decrease of high levels of sand will limit the presence of plant species. Organic matter and humus are the main factors in the formation of soil structure and porosity and permeability and can thus increase the amount of soil. The generalized additive models were clearly more accurate. This was also observed by Jafarian and Kargar (2016). Due to its flexibility in determining the type and degree of communication and appropriate interpretability, generalized additive model has become a popular model and can be used for a wide range of data. For modelling data with different scales, additive models have been used in other sciences. Based on our data, non-parametric techniques outperform GAM for predicting the presence of species. For all species, GLM and BRT models performed worse than the other models in predicting forest characteristics. This has also been confirmed by Moisen and Frescino (2002). Leath wick *et al.* (2006) concluded from their study on modelling fish species richness that due to its capability for fitting interactions among predictor variables, RF appears to offer considerable performance gains over modelling techniques such as GAM. Also Moisen *et al.* (2006) predicted the basal area. Although the predictions were poor, BRT-like models performed better and obtained more stable results than GAM. Our study cannot confirm these findings. Based on most evaluation measures, GAM models perform better than BRT models. This may be due to the fact that BRT models together with CART tend to over-fit the data more strongly than other techniques. Nevertheless, the predictive success of the GAM model in terms of goodness-of-fit, i.e. AUC, Kappa and TSS are always the highest of all among modelling techniques making it at first sight the most suited technique for predicting species presence. There are many potential sources of error in the data sets used for modelling including measurement errors, sampling soil, limitations in field data collection, genetic variability, etc. These errors may affect the overall accuracy of the models (Moisen *et al.,* 2006; Dubuis, 2013; Sor *et al.,* 2017). Among the models used, GAM model had a better performance that is in accordance with the results of Dubuis (2013). One of the remarkable characteristics of CART model is its simplicity. It was, however, the least accurate predictive model in this study (AUC= 0.74). The importance of the measured variables in the random forest model indicated that C/N was the most influential variable in the modelling (Table 2). Thus, these results also show random forest as the most accurate of the three methods used. As summarized in Table 3, the random forest model is also the closest in the presence area to the actual distribution. To conclude, the modelling framework presented here provided good results with notably high and stable AUC values obtained by changing the tuning parameter achieved by the random forest learning method. These results indicate different qualities of the predictive models related to each species. The reason of this can be attributed to the special condition of each habitat, which other researchers have achieved similar results (Aertsena *et al.*, 2010). The small amount of TSS is due to the fact that species in the study area do not prefer a set of habitat conditions higher than the average environmental conditions of the region. This means that the studied plant species are specialist in the range of their habitat resources. In other words, the species has a relatively narrow ecological niche breath that is consistent with the findings (Barry and Elith, 2006). This means that whatever the species was adapted to and act the particular conditions of an environment, it will have less tolerance in dealing with the changing environmental conditions. These results warrant the consideration of alternative modelling techniques within production mapping environments. A final cautionary note is that statistical differences between the modelling techniques may not necessarily translate to relevant differences from a management perspective. Conversely, models that do not produce different global performance measures significantly may produce wildly different maps resulting in drastically different implications for management decisions. It should be noted that the determination of the main factors influencing the distribution of species and studying these environmental factors could be cost and time-saving.

Agropyron repens had a direct relationship with pH; Increasing CaCO3 causes a high pH and mineral rate at root environment. Organic carbon improves the soil's physical and biological properties that contribute to the plant's survival. The advantage of this method is that it uses the presence or absence of species.

Conclusion

It was concluding these methods can be used for mapping predictions of vegetation. Such maps along with obtained information similar to this research can facilitate the regeneration of rangelands, determine compatible species, and identify appropriate areas for seeding. Five modelling techniques were compared for predicting species presence for five tree species in Aflah rangelands. For predicting species presence, GAM had the highest values of AUC and Kappa for the majority of the species, while GAMs had the highest values for the majority of the species for sensitivity. Investigating the performance of the model in terms of management decisions seems necessary.

References

Asadian, GH., Javadi, S.A., Jafari, M., Arzani, H., Akbarzadeh, M., 2017.Relationships between Environmental Factors and Plant Communities in Enclosure Rangelands (Case study: Gonbad, Hamadan) J. of Range. Sci., 7(1): 20-34.

Aertsena, W., Kinta, V., Orshovena, J., Özkanb, K., and Muysa, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecolo Modell.,9 (221): 1119–1130.

Barry, S., and Elith, J., 2006. Error and uncertainty in habitat models. J. Appli Ecol., 10 (43): 413–423.

Beretta, A. N., Silbermann,A.V., Paladino, L., Torres,D., Bassahun,D., Musselli, R., and García-Lamohte, A., 2014. Soil texture analyses using a hydrometer: modification of the Bouyoucos method, Cien. Inv. Agr., 41(2):263-271.

Bergamini, A., Stofer, S., Bolliger J., and Scheidegger C., 2007. Evaluating macrolichens and environmental variables as predictors of the diversity of epiphytic microlichens. Lichenol, 4 (39): 475–489.

Boitani, L., Sinibaldi, I., Corsi, F., Biase, A., Carranza, I, d'Inzillo, Ravagli, M., Reggiani, G., Rondinini, C., and Trapanese, P., 2007. Distribution of medium- to large-sized African mammals based on habitat suitability models. Biodi and Conse., 3 (17): 605–621.

Breiman, L, Friedman, J.H, Olshen, R.A., Stone, C.J., 1984. Classification and Regression, Amazon publication, London, 354 pp.

Breiman, L., 2001. Random forests. Machine Learning, 45:15–32.

Cao, B., Chengke, B., Linlin, Z., Guishuang, L., Mingce, M., 2016. Modeling habitat distribution of *Cornus officinalis* with Maxent modeling and fuzzy logics in China. J. Plant Ecol., 9 (6): 742-51.

Cutler, A., and Stevens, J. R., 2006. Random forests for microarrays. Methods in Enzy., 411:422–432.

Dirnböck, T., Dullinger, S., Grabherr, G., and Dirnb, T., 2003. A regional impact assessment of climate and land-use change on alpine vegetation. J. of Biogeo., 5 (101): 401–417.

Dubuis, A., 2013. Predicting spatial patterns of plant biodiversity: from species to communities. Thesis Ph.D. University of Lausanne, 295 pp.

Elith, J., Leathwick, J. R., Hastie, T., 2008. A working guide to boosted regression trees. J Anim Ecol., 2 (77): 802–813.

Engler, R., Randin, C.F., Thuiller,W., Dullinger, S., Zimmermann, N.E., Araujo, M.B., Pearman, P.B., Le Lay, G., Piedallu, C., Albert, C.H., Choler P., Coldea, G., De Lamo, X., Dirnbock T., Gegout, J.C., Gomez-GarcianD., Grytnes, J.-A.A., Heegaard E., Hoistad, F., Nogues-Bravo, D., Normand S., Puscas, M., Sebastia, M.T., Stanisci, A., Theurillat J.-P.P., Trivedi, M.R., Vittoz, P., Guisan, A., Araújo M.B., Dirnböck, T., Gégout, J.C., Gómez-García, D., Høistad, F., Nogués-Bravo, D., Puşcaş M., and Sebastià, M.T., 2011. 21st century climate change threatens mountain flora unequally across Europe. Glob Change Biol., 7 (17): 2330–2341.

Fielding, A.H., and Bell J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environ Conserve., 2 (24): 38–49.

Garzon, M.B., Blazek, R., Neteler, M., de Diosa, R., Ollero, H., and Furlanello, C., 2006. Predicting habitat suitability with machine learning models: The potential area of Pinus sylvestris L. in the Iberian Peninsula, Ecolog mode., 197:383–393.

Guisan, A., Edwards, T.C., and Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecolog Modell., 8 (157): 89–100.

Guisan, A., Theurillat, J.P., and Kienast, F., 1998. Predicting the potential distribution of plant species in an Alpine environment. J. Veget Scie., 3 (9): 65–74.

Guisan, A., and Zimmermann, N.E., 2000. Predictive Habitat Distribution Models in Ecology. Ecol Model., 135:147-186.

Hallstan, S., Johnson, R. K., Willén, E., and Grandin, U., 2012. Comparison of classification-then modelling and species-by-species modelling for predicting Lake Phytoplankton assemblages. Ecol Model., 2 (231):11–19.

Hastie, T., and Tibshirani, R., 1990. Non-parametric logistic and proportional odds regression, Appli statis., 260-276.

Hengl, T., Sierdsema, H., Radovi, A., and Dilo, A., 2009. Spatial prediction of species distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecolo Model., 6 (220): 3133-3222.

Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modeling? EcoloModell., 157: 331–341.

Jafarian, Z., Arzani, H., Zahedi, GH. And Azarnivand, H., 2009. Application of discriminate analysis for determination relationship between distribution of plant species with environmental factors and satellite data at Rineh rangeland in province of Mazandaran.watershed researches., 3 (88):64-71.

Jafarian, Z., and Kargar, M., 2012. Environmental factors affecting the ecological species groups using logistic regression at Polour rangeland, Mazndaran. J. Environ Scie., 10 (2): 107-118.

Jafarian, Z., and Kargar, M., 2016. Modelling plant protected and valued species distributions using GLM and GAM in Polour rangeland. Geo and develop., 15(46):117-132.

Jafari haghighi, M.,2003. Sampling method analysis methods, Neda Zoha publication, 272 pp.

Jones M.C, Dye S.R, Pinnegar J.K, Warren R, and Cheung W.W.L., 2012. Modelling commercial fish distributions: Prediction and assessment using different approaches. Ecolo Model., 4 (225):133–145.

Khalasi Ahvazi, L., Zare Chahouki, M.A., Ghorbannezhad, F., 2012. Comparing Discriminant Analysis, Ecological Niche Factor Analysis and Logistic Regression Methods for Geographic Distribution Modelling of *Eurotia ceratoides* (L.) C. A. Mey. J. Range Sci., 3(1): 45-57.

Leathwick, J.R., Elith, J., Francis, M.P., Hastie, T., and Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Mar. Ecolog –Progr., 12 (321): 267–281.

Lütolf, M., Kienast, F., Guisan, A., 2006. The ghost of past species occurrence: improving species Distribution models for presence-only data. J. Appli Ecol., 43: 802–815.

Machado, E.A., 2012. Empirical mapping of suitability to dengue fever in Mexico using species distribution modeling. Appl Geogr., 11 (33): 82–93.

Maggini, R., Guisan, A., and Cherix, D., 2002. A stratified approach for modeling the distribution of a threatened ant species in the Swiss National Park. Biodiv and Conser., 11:

2117–2141.

Maggini, R., Lehmann, A., Zimmermann, N.E., and Guisan, A., 2006. Improving generalized regression analysis for the spatial prediction of forest communities. J. Biogeo., 33: 1729–1749.

Marmion, M., Luoto, M., Heikkinen, R.K., and Thuiller, W., 2009. The performance of state-of-the-art modeling techniques depends on geographical distribution of species. Ecolo Modell., 220: 3512–3520.

Mckenney, D.W., Pedlar, J.H., 2003. Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. Forest Ecolo Manage., 3 (175): 497–507.

Meier, E.S., Kienast, F., Pearman, P.B, Svenning, J-C., Thuiller, W., Araújo, M.B., Guisan, A., and Zimmermann, N.E., 2010. Biotic and abiotic variables show little redundancy in explaining tree species distributions. Ecog, 12(33): 1038–1048.

Moghari, M., Arzani, H., Tavili, A., Azarnivand, H., Saravi, M., and Farahpoor, H. 2014. Identifying and Determining the Competency of Medicinal Herbs in the Pashang Haraz Basin Rangelands, Amol County, Mazandaran Province. Two Iranian. J. Medic. Arom. Plants., 30 (6): 898-914.

Moisen, G.G., Freeman, E.A, Blackard, J.A, Frescino, T.S, Zimmermann, N.E, and Edwards,T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol. Model., 8 (199): 176–187.

Moisen, G.G., and Frescino, T.S., 2002. Comparing five modelling techniques for predicting. Forest characteristics. Ecol Model, 4 (157): 209–225.

Pellissier, L., Bråthen, K.A., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E., Guisan, A., and Brathen, K.A., 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. Ecogr, 8 (33): 1004–1014.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C, and Guisan, A., 2012. Climatic niche shifts are rare among terrestrial plant invaders. Scie., 8 (335): 1344–8.

Piri Sahragard, H., and Zare Chahoki, M. A., 2015. An evaluation of predictive habitat models performance of plant species in Hoze soltan rangelands of Qom province. Ecol Model., 2 (309): 64-71.

Piri Sahragard, H., and Zare Chahouki, M.A., 2016. Modeling of *Artemisia sieberi* Besser Habitat Distribution Using Maximum Entropy Method in Desert Rangelands. Journal of Rangeland Science. 6(2): 93-101.

Pottier, J., Dubuis, A., and Pellissier, L., 2012. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Glob Ecolog Biogeo., 6 (22): 52–63.

Rondinini, C., Di Marco, M., Chiozza, F., Santulli, G., Baisero, D., Visconti, P., Hoffmann, M., Schipper, J., Stuart, S.N., Tognelli M.F., Amori, G., Falcucci, A., Maiorano, L, and Boitani, L., 2011. Global habitat suitability models of terrestrial mammals. Philosophical transactions of the Royal Society of London. Series B, Bio scie., 366: 2633–41.

Schorr, G., Holstein, N., Pearman, P.B., Guisan, A., and Kadereit, J.W., 2012. Integrating species distribution models (SDMs) and phylogeography for two species of Alpine Primula. Ecol. Evol.,10 (2): 1260–77.

Sor, R., Park, Y.S., Boets, P., Goethals, L. M., Sovan, L., 2017. Effects of species prevalence on the performance of predictive models. Ecolog Mode, 23 (354): 11-19.

Sundblad, G., Härmä, M., Lappalainen, A., Urho, L., and, U., 2009. Transferability of predictive fish distribution models in two coastal systems. Estuarine, Coas. Shelf Science., 6 (83): 90–96.

Sweets, J. A., 1988. Measuring the accuracy of Diagnostic system. Science, 27 (240): 1285-1293.

Tarkesh, M., 2012. Comparison of six correlative models in predictive vegetation mapping on a local scale. Thesis Ph.D. University of Jena, 106 p.

Vega, R., Flojgaard, C., Lira-Noriega, A., Nakazawa, Y., Svenning, J.C., and Searle, J.B., 2010. Northern glacial refugia for the pygmy shrew Sorex minutus in Europe revealed by phylogeographic analyses and species distribution modelling. Ecogr, 10 (33): 260–271.

Vicente, J., Randin, C.F., Goncalves, J., Metzger, M.J., Lomba, A., Honrado, J., and Guisan, A., 2011. Where will conflicts between alien and rare species occur after climate and land-use change? A test with a novel combined modelling approach. Biolo Invas., 9 (13):1209–1227.

Wisz, M.S., Walther, B., and Rahbek, C., 2007. Using potential distributions to explore determinants of Western Palaearctic migratory songbird species richness in sub-Saharan Africa. J. Biogeogr., 12 (34): 828–841.

مقایسه تکنیک‌های مختلف مدل‌سازی برای پیش‌بینی حضور و عدم حضور برخی گونه‌های گیاهی غالب در مراتع کوهستانی استان مازندران

منصوره کارگر [الف]، داود اخضری [ب]*، امیر سعادتفر [ج]
[الف] دکتری علوم مرتع، اداره‌کل‌منابع طبیعی و آبخیزداری استان البرز
[ب] دانشیار گروه مرتع و آبخیزداری دانشگاه ملایر *(نگارنده مسئول)، پست الکترونیک: *d_akhzari@yahoo.com*
[ج] استادیار پژوهشی پژوهشکده فناوری تولیدات گیاهی، دانشگاه شهید باهنر کرمان
[د] دانشیار گروه مرتعداری، دانشگاه گنبدکاووس، گنبد

تاریخ دریافت: ۱۳۹۷/۰۲/۰۵
تاریخ پذیرش: ۱۳۹۷/۰۸/۰۵

چکیده. در مطالعات کاربردی، بررسی رابطه بین حضور و عدم حضور گونه گیاهی و متغیرهای محیطی برای مدیریت مشکلات اکولوژیکی و اکوسیستم‌های مرتعی ضروری است. این تحقیق در تابستان سال ۱۳۹۵ انجام شد. هدف از این مطالعه مقایسه قدرت پیش‌بینی تعدادی از مدل‌های پراکنش گونه (SDM) و ارزیابی اهمیت تعدادی از متغیرهای محیطی به عنوان پیش‌بینی‌کننده‌ها در ارتباط با پوشش گیاهان مرتعی بود. در این تحقیق، مراتع منطقه افلاح با وسعت ۵۷۲۱ هکتار به عنوان منطقه مطالعاتی انتخاب شد. متغیرهای پیش‌بینی کننده شامل فاکتورهای اقلیمی (بارندگی، دما و رطوبت)، توپوگرافی (شیب، جهت و ارتفاع) و عوامل ادافیکی بودند. روش نمونه‌گیری به صورت تصادفی- طبقه‌بندی شده برای پوشش گیاهی و خاک بود. عوامل توپوگرافی شامل شیب، ارتفاع و جهت در نرم افزار Arc GIS تعریف شدند. در هر واحد نمونه‌برداری، ۱۰ پلات یک متر مربعی (مجموع ۳۵۰ پلات) استقرار یافتند و لیست گونه‌ها، تعداد و حضور یا عدم حضور گونه‌های گیاهی در آنها ثبت شد. پنج روش مختلف مدل‌سازی برای پیش بینی حضور و عدم حضور پنج گونه گیاهی غالب (*Agropyron repens*، *Festuca ovina*، *Leucopoa sclerophylla*، *Stachys lavandulifolia* و *Tragopogon graminifolius*) مورد بررسی قرار گرفتند. مدل‌ها عبارت بودند از: مدل خطی تعمیم یافته (GLM)، طبقه بندی و رگرسیون درختی (CART)، درخت رگرسیون تقویت شده (BRT)، مدل جمعی تعمیم یافته (GAM) و جنگل تصادفی (RF).تجزیه و تحلیل داده‌ها با استفاده از نرم افزار R، انجام شد. نتایج نشان داد که مدل GAM توانایی پیش بینی بالایی برای گونه‌های موجود در مرتع مورد بررسی را داشته است. مدل GAM دارای ارزش عددی بالایی برای پارامتر سطح زیر منحنی (AUC) بود (به ترتیب ۰/۶۷، ۰/۷۷، ۰/۶۹، ۰/۶۴ و ۰/۶۰ برای گونه‌های *Agropyron repens*، *Festuca ovina*، *Leucopoa sclerophylla*، *Stachys lavandulifolia* و *Tragopogon graminifolius*. همچنین مقدار عددی شاخص Kappa برای گونه‌های *Agropyron repens*، *Festuca ovina*، *Leucopoa sclerophylla*، *Stachys lavandulifolia* و *Tragopogon graminifolius* به ترتیب برابر ۰/۱۰، ۰/۱۰، ۰/۱۹، ۰/۰۱ و ۰/۱۱ تعیین شد. تجزیه و تحلیل متغیرهای محیطی نشان داد که اسیدیته، نیتروژن و جهت مهمترین متغیرها در مدل GAM بودند. به طور کلی متغیرهای نیتروژن، نسبت کربن به نیتروژن و فسفر خاک به ترتیب با میزان ۰/۴۵۲، ۰/۴۳۷ و ۰/۳۸۹ از مهم‌ترین متغیرهای محیطی تاثیر گذار بر پراکنش گونه‌های مورد مطالعه بودند.

کلمات کلیدی: پوشش گیاهی، خاک، توپوگرافی، مدل جنگل تصادفی، مراتع افلاح