Ecological Modelling 309–310 (2015) 64–71

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

# An evaluation of predictive habitat models performance of plant species in Hoze soltan rangelands of Qom province

H.P. Sahragard [a], M.A. Zare Chahouki [b,*]

[a] Range and Watershed Department, University of Zabol, Iran
[b] Department of Rehabilitation of Arid and Mountainous Regions, Natural Resources Faculty, University of Tehran, Iran

ARTICLE INFO

*Article history:*
Received 17 December 2014
Received in revised form 9 March 2015
Accepted 7 April 2015
Available online 15 May 2015

*Keywords:*
Modeling distribution of plant species
Qom rangelands
Logistic regression
Maximum entropy
Artificial neural network
Geostatistics

ABSTRACT

The current study addresses an assessment of performance of modeling techniques including logistic regression (LR), maximum entropy modeling technique (MaxEnt) and artificial neural network (ANN) to predict habitat distribution of plant species in Qom Province rangelands of Iran. After determination of homogeneous units, vegetation sampling was carried out using random systematic method. Depending on the plant species, the plot size was determined using Minimal Area method from 2 to 25 $m^2$. Sample size was also determined to be 60 plots with respect to vegetation cover variations using statistical method. In order to sample the soil at each habitat, eight holes was drilled and samples were taken from 0 to 30 and 30 to 80 cm depths. Plant distribution modeling was conducted using LR, the MaxEnt and ANN. After implementation of the model, to evaluate and predict the actual maps conformity, Kappa coefficient and true skill statistic (TSS) were measured. On the basis of Kappa and TSS values calculated, prediction accuracy of the methods used varies for different habitats. Results indicate that LR model is capable to predict habitats distribution of *Halocnemum strobilaceum*, which has limited ecological niche at very good level ($\kappa = 0.71$). The model obtained from the MaxEnt could predict habitat distribution of *Artemisia sieberi* at very good level. However, the prediction maps derived from ANN models for all studied habitats were obtained to be at good and very good level. Results indicate a strong relationship between model performance and the kinds of species distributions being modeled. Some methods performed generally better, but no method was superior in all circumstances. Based on these results it can be said that in order to choose the optimal approach of habitats distribution modeling in addition to the statistical considerations, purpose and expected accuracy, data available types, ecological niche range of species and be interpreted of method in terms of ecological concepts also should be considered.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Vegetation cover of each region is the resultant of its environmental gradients. Therefore, it can be said that a combination of ecological factors such as climate, soil and physiography affect the establishment of plant species. Predictive vegetation modeling (PVM) can be defined as predicting the distribution of vegetation across a landscape based on the relationship between the spatial distribution of vegetation and certain environmental variables (Franklin, 1995; Guisan and Zimmermann, 2000). Predictive vegetation modeling has become an important tool in ecology because of its ability to investigate the species–environment relationships (Segurado and Araújo, 2004; Austin, 2007). To do so, PVM requires digital maps of the environmental variables, spatial information on the vegetation attribute of interest (e.g. species, type, abundance), usually from a sample of locations, and an appropriate statistical model. Results obtained from these analyses can be used to produce habitat-suitability maps, which can infer the potential distribution of a species. Such predictions are useful for ecosystem studies (Ferrier, 2002; McNally et al., 2003), conservation planning (Polasky and Solow, 2001), and studying biogeographically large-scale issues such as contraction of geographic ranges (Donald and Greenwood, 2001) and management of invasive species (Peterson and Shaw, 2003).

Predictive models of distributions can generally be classified into one of the two categories based on the type of data they require (Guisan and Zimmermann, 2000). The first category requires the presence and absence of a species data (presence/absence models), while the second requires only species presence data (presence-only models). A number of studies have demonstrated that different modeling approaches have the potential to yield substantially different predictions (e.g. Brotons et al., 2004; Elith et al., 2006; Loiselle

* Corresponding author. Tel.: +98 2632249313.
*E-mail address:* mazare@ut.ac.ir (M.A.Z. Chahouki).

http://dx.doi.org/10.1016/j.ecolmodel.2015.04.005
0304-3800/© 2015 Elsevier B.V. All rights reserved.

et al., 2003; Segurado and Araújo, 2004; Thuiller, 2004). So the choice of the right statistical method in a specific modeling context is an important issue and now supported by many published comparisons (e.g. Moisen and Frescino, 2002; Segurado and Araújo, 2004). The most comprehensive set of model comparison to date was provided by Elith et al. (2006). The authors compared 16 modeling methods using 226 species across six regions in the world. Through comparison of 16 modeling techniques, these authors concluded that boosted regression trees (BRT) and MaxEnt are among the best performing methods.

In practice, model selection will be influenced by factors including whether (1) observed absence data are available, (2) data on some of the environmental variables are categorical, and (3) evaluation of the influence of different variables on the model prediction is important or not. Assessing the accuracy of a model's predictions is commonly termed 'validation' or 'evaluation' which is a vital step in model development. Application of the model will be less effective if we have not assessed the accuracy of its predictions. Validation thus enables us to determine the suitability of a model for a specific application and to compare different modeling methods (Pearce and Ferrier, 2000). In general, models classified as 'best' were those that were able to identify complex relationships existed in the data, including interactions among environmental variables.

Regarding the differences among models, selection of an appropriate algorithm is both difficult and crucial. In the other hand, identifying models that are generically 'best' is problematic since the approach used to assess predictive performance depends on the aim of the modeling. Optimal method selection of species distribution modeling will be achieved through making a comparison between them and discovering the best method of prediction considering different conditions. Despite all these, few studies have been conducted in order to compare different methods of modeling and specify each method capacity to anticipate in comparison with other methods in order to select the best modeling approach. It is clear that comparison of different modeling techniques leading to selection of more efficient methods can also lead to more accurate and reliable results while providing a strong base for correct management decision. On the other hand, it leads to saving time and cost for conducting future studies.

Considering the importance of the issues mentioned, it is necessary to conduct researches in order to evaluate the performance of different methods used in predictive modeling of distribution plant species to identify areas of highest fitness for the establishment of different species with the highest possible precision identified and to help rangeland management in making appropriate amendment decisions. In the current study, modeling was performed using three methods namely LR, MaxEnt and ANN, in addition, model evaluation were conducted using related indices. Furthermore, according to the research objective, the presences of optimal threshold levels were determined for different plant species in order to achieve the highest prediction accuracy.

## 2. Methods

### 2.1. The study area

The study area is located in the central part of Qom province and 50 km from the city of Qom in Geographic coordinates area 50′ 50° 30″–50′ 54° 30″ E and 34′ 59° 30″–35′ 03° 30″ N. This region is located in the west of Qom and covers an area of 3000 Ha. The study area is in plain area. Minimum and maximum altitudes in the study area are 796 and 1100 m above sea level, respectively. This region was chosen because of changes in vegetation cover in relation to soil changes. Clarity vegetation covers variations and easy separation of plant communities. Fig. 1 shows the general location of the study area in Iran and Qom province.

**Table 1**
List of variables in the data set.

| Variable | Code | Unit | Mean ± standard deviation |
|---|---|---|---|
| Elevation | abs | M | 790 ± 15 |
| Slope | Slope | % | 5 ± 0.2 |
| Gravel | gr | % | 9.85 ± 1.86 |
| Clay | Caly | % | 13.81 ± 7.02 |
| Silt | Silt | % | 32.34 ± 10.16 |
| Sand | Sand | % | 54.89 ± 10.25 |
| Saturation moisture | sm | % | 38.34 ± 5.09 |
| Available water | A.W. | % | 19.10 ± 5.26 |
| Gypsum | gy | % | 3.31 ± 1.34 |
| Organic matter | OM. | % | 0.57 ± 0.27 |
| Lime | Lime | % | 7.01 ± 0.48 |
| pH (acidity) | pH | – | 7.23 ± 0.17 |
| ECe | EC | ds/m | 97.49 ± 26.79 |
| Sodium ion ($Na^+$) | Na | meq/l | 647.46 ± 45.99 |
| Potassium ion ($K^+$) | K | meq/l | 4.98 ± 0.58 |
| Calcium ion ($Ca^{2+}$) | Ca | meq/l | 316.22 ± 11.46 |
| Magnesium ($Mg^{2+}$) | Mg | meq/l | 99.88 ± 13.67 |
| Chlorine ($Cl^-$) | Cl | meq/l | 831.75 ± 45.61 |
| Carbonate ($Co_3{}^{2-}$) | Co | meq/l | 1.4 ± 0.068 |
| Bicarbonate ($HCO_3{}^-$) | $HCO_3{}^-$ | meq/l | 9.35 ± 1.82 |
| Sulfate ($So_4{}^{2-}$) | $SO_4$ | meq/l | 235.46 ± 24.39 |

### 2.2. Environmental predictor variables

After determination of homogeneous units using basic maps of the study area (digital elevation, aspect, slope and geology maps, scale 1:25,000), considering condition of the area, in the homogeneous units, vegetation sampling was carried out using random systematic method via the plots established along four transect with 200–1000 m lengths. Depending on the plant species, the plot size was determined using Minimal Area method from 2 to 25 $m^2$. The sample size used was determined to be 60 plots with respect to vegetation cover variations using statistical method. Vegetation sampling was done in the key area of homogeneous units. Besides vegetation data (name of plant species and canopy cover percent), information related to the geographical boundaries of habitats, slope, aspect and altitude were also recorded. For soil sampling at each habitat eight holes were drilled and samples were taken from 0 to 30 and 30 to 80 cm depths. Since most of the root activities is within 0–30 cm depth range, so the 0–30 cm and 30–80 cm depth ranges were selected as the first and second layers, respectively. In order to have good distribution, soil profiles were spread in the study area. After sampling, soil characteristics consisting gravel percent, texture, saturation moisture, available water, lime, gypsum, organic matter, Acidity (pH), Electrical Conductivity (EC) and soluble solute ($Na^+$, $Ca^{2+}$, $Mg^{2+}$, $K^+$, $Cl^-$, $Co_3{}^{2-}$, $Hco_3{}^-$ and $So_4{}^{2-}$) were measured by routine methods (Table 1). Using geostatistical and kriging interpolation method with the same spatial resolution (pixel size 30 m × 30 m) soil digital layers were prepared and stored in GIS. Arc GIS 9.3 and $GS^+$ Version fifth software were used for mapping soil properties. Digital elevation map of the region 1:25,000 scale was used for mapping slope, aspect and elevation.

### 2.3. Model development

#### 2.3.1. LR

GLMs are a suite of parametric methods (McCullagh and Nelder, 1989) allowing more flexible relationships to be specified in the form of a number of link functions between the response and predictor variables than linear regression models. When response data is binary, the appropriate GLM is a logistic model using a logit link to describe the relationship between the response and the linear sum

Download English Version:

https://daneshyari.com/en/article/4375676

Download Persian Version:

https://daneshyari.com/article/4375676

Daneshyari.com