Mohammad Ali ZARE CHAHOUKI*, Lyla KHALASI AHVAZI, Hossein AZARNIVAND

Department of Rehabilitation of Arid and Mountainous Regions, University of Tehran, Iran, P.O. Box: 31585-4314, *e-mail: mazare@ut.ac.ir (corresponding author)

COMPARISON OF THREE MODELING APPROACHES FOR PREDICTING PLANT SPECIES DISTRIBUTION IN MOUNTAINOUS SCRUB VEGETATION (SEMNAN RANGELANDS, IRAN)

ABSTRACT: The predictive modeling of plant species distribution has wide applications in vegetation studies. This study attempts to assess three modeling approaches to predict the plant distribution in the dry (precipitation 128–275 mm) mountainous (altitude 1129–2260 m a.s.l.) scrub vegetation on the example of the rangelands of northeastern Semnan, Iran. The vegetation of the study area belongs to the communities of Artemisia, Astragalus, Erytia and other scrub species. The main objective of this study is to compare the predictive ability of three habitat models, and to find the most effective environmental factors for predicting the plant species occurrence. The Canonical Correspondence Analysis (CCA), Logistic Regression (LR), and Artificial Neural Network (ANN) models were chosen to model the spatial distribution pattern of vegetation communities. Plant density and cover, soil texture, available moisture, pH, electrical conductivity (EC), organic matter, lime, gravel and gypsum contents and topography (elevation, slope and aspect) are those variables that have been sampled using the randomized systematic method. Within each vegetation type, the samples were collected using 15 quadrates placed at an interval of 50 m along three 750 m transects. As a necessary step, the maps of all factors affecting the predictive capability of the models were generated. The results showed that the predictive models using the LR and ANN methods are more suitable to predict the distribution of individual species. In opposite, the CCA method is more suitable to predict the distribution of the all studied species together. Using the finalized models, maps of individual species (for different species) or for all the species were generated in the GIS environment. To evaluate the predictive ability of the models, the accuracy of the predicted maps was compared against real-world vegetation maps using the Kappa statistic. The Kappa (κ) statistic was also used to evaluate the adequacy of vegetation mapping. The comparison between the vegetation cover of a map generated using the CCA application and its corresponding actual map showed a good agreement (i.e. κ= 0.58). The results also revealed that maps generated using the LR and ANN models for Astragalus spp., Halocnemum strobilaceum, Zygophyllum eurypterum and Seidlitzia rosmarinus species have a high accordance with their corresponding actual maps of the study area. Due to the high level of adaptability of Artemisia sieberi, allowing this specie to grow in most parts of the study area with relatively different habitat conditions, a predictive model for this species could not be fixed. In such cases, a set of predictive models may be used to formulate the environment-vegetation relationship. Finally, the predictive ability of the LR and ANN models for mapping Astragalus spp. was determined as κ = 0.86 and κ = 0.91 respectively, implying a very good agreement between predictions and observations.

It is concluded that the combination of modelling of the local species distribution constitutes
a promising future research area, which has the potentiality to enhance assessments and conservation planning of vegetation (like rangelands) based on predictive species models.

**KEY WORDS:** predictive model, Canonical Correspondence Analysis, geostatistical method, logistic regression, artificial neural networks, Kappa statistic.

1. INTRODUCTION

With the rise of new powerful statistical techniques and GIS tools, the development of predictive plant species distribution models has rapidly increased in vegetation science. The predictive models of plant species distribution on the basis of environmental variables are used to evaluate the sensitivity of vegetation to hypothesized environmental influences (Cairns 2001). Plant cover data may be used to classify the studied plant community into a vegetation type and to test different ecological hypothesis concerning the plant abundance and distribution. Plant cover data may also be used in gradient studies in which the effects of different environmental gradients on the abundance of specific plant species are investigated (Austin 2007). Such gradient studies may be used in the prognostic modeling of the effects of global warming and nitrogen deposition on plant community dynamics and, consequently, to predict the fate of specific ecosystems in a changing environment (Guisan and Thuiller 2005).

The habitat is the core of the hierarchical patch dynamic theory (Allen and Starr 1982, Wu and Huffer 1997), and is defined as a physical place at a particular spatio-temporal scale where and when an organism either actually or potentially lives (Kearney and Porter 2006). The modelling of the different types of habitat requires to identify a set of environmental variables related to the spatial distribution of vegetation at each scale. One of the current challenges in ecology is to forecast accurately the implications of future environmental changes on species distribution (Morin and Lechowicz 2008; Thuiller et al. 2008).

In this research, the capability of three methods for predicting vegetation type is compared.

The **Canonical Correspondence Analysis** (CCA) model is one of them. Almost all models that use ordination techniques for predicting the distribution of species or communities are based on CCA (Zhang et al. 2005, Baruch 2005, He et al. 2007, Guoqing et al. 2008). CCA expresses the relationship between species as a linear combination of environmental variables (Zhang et al. 2005). CCA is also used to study spatial variation of species communities, and to assess which part of the variation can be explained by associated environmental variables (Bach and Michael 2006). CCA has found widespread use in ecological sciences.

The CCA method combines a multivariate ordination of species occurrence data with a constrained regression maximising the correlation between the species ordination axes and selected environmental variables. The species are assumed to have unimodal responses to the underlying environmental gradients as specified by the ordination axes. Significance of species environment correlation was tested by the distribution of the free Monte Carlo test (1000 permutations). In the Monte Carlo approach, the distribution of the test statistics under the null hypothesis is generated by random permutations of cases in the environmental data. On the other hand, CCA chooses the best weights for the environmental variables (Chahouki et al. 2010).

**Logistic Regression** (LR) is a kind of Generalized Linear Models (GLM) which is a suitable way to analyse a binary response variable. Logistic Regression (LR) uses a logic link to describe the relationship between the response and the linear sum of the predictor variables (Miller and Franklin 2002). This is accomplished by applying diverse regression equations, in which presence/absence of an object is transformed into a continuous probability ranging from 0 to 1. Values close to 1 represent high probability of presence, whereas values close to 0 represent high probability of absence. The probability of occurrence of each plant species is calculated with respect to the combined effect of site conditions.

The **Artificial Neural Networks** (ANN) is the third model. The recourse to ANN as used by Fukuda (2011), Willems et al. (2008), Tan et al. (2006), Watts et al. (2011),
and Scrinzi et al. (2007), is a promising area of predictive modeling of plant species distribution. However, most published applications of ANN in ecology are connected with the field of remote sensing (see Gutiérrez et al. 2008) and (non-spatial) assessment of environmental changes (Liu et al. 2011). Neural network models do not make any attempt to model the physical, chemical, or biological causes of ecological phenomena. Instead, they learn a direct mapping from input features to output predictions.

In ANN model the biophysical descriptors for predicting the spatial distribution of species or communities are used. Tan et al. (2006) suggest that GLM and ANN models are the most suitable and robust models for studying ecosystems with time-dependent dynamics and periodicities whose frequency are possibly less than the time scale of the data considered.

The objective of this research is to examine the spatial distribution pattern of vegetation types, analyse the ecological relationships between these communities and their environment, and finally compare the capability of three models (i.e. CCA, LR and ANN) in predicting plant species distribution. To do

Fig. 1. Location of study area and the distribution of the vegetation types.
so, key factors, affecting vegetation development, and other ecological parameters of the study area were identified. A further objective of this research was to establish a monitoring system that may serve to identify and predict future changes of vegetation cover and assess the impacts of conservation and management practices.

2. STUDY AREA

The study area (Fig.1) is located at the northeast of the Semnan province, central Iran (35°53’N, 54°24’E to 35°50’N, 53°43’E). The maximum elevation of the region is 2260 m a.s.l. and the minimum elevation is 1129 m a.s.l. Average annual precipitation of the study area ranges between 275 mm in the mountains and 128 mm in the saline lowlands. Minimum temperature occurs in December (around −6°C) while the highest temperature reaches +45°C in June.

Vegetation classification using dominant species and vegetation map of the area was conducted and 8 vegetation types were identified (Fig. 1).

A preliminary survey on the rangeland plant diversity in the flora of the Semnan province of Iran has been made with six families. The important families are respectively, Compositae, Gramineae, Chenopodiaceae, Papilionaceae, Labiatae and Cruciferae. This community covers an area of 74,000 ha.

3. METHODS

3.1. Data collection

A one-year vegetation quantitative survey was carried out in 2009. Sampling was conducted in homogeneous units (vegetation patches) resulted from overlaying of hypsometric, aspect, slope and geologic maps. Within each unit, the samples were collected using 15 quadrates placed at an interval of 50 m along three 750 m transects. Floristic list, density and canopy cover percentage were determined and recorded for each quadrant. The quadrant size was determined using the minimal area method (Green and Ostling 2003). The randomized systematic method was used for sampling. Soil samples were taken from 0–20 cm and 20–80 cm in starting and ending points of each transect. Measured soil factors included texture (determined by Bouyoucos hydrometer), available moisture (weighting method), soil organic matter (determined using Walkely and Black rapid titration; Black 1979), pH in the saturation extract (determined by pH meter), electrical conductivity (EC) (determined by conductivity meter), and lime (determined using 1 N HCl; Jackson 1967). The elevation and slope (using GPS) and slope direction were determined at the location of each quadrate.

The collected information of all environmental variables are summarized in APPENDIX I including twenty two soil physical and chemical parameters and two geographical factors.

3.2. Methods of data analysis

For plant predictive mapping it is necessary to prepare the maps of all effective factors used in the models (an example Fig. 2). Topographic data (elevation, slope and aspect) were derived from Digital Elevation Models (DEM) with a resolution of 10 m characteristics. Spatial statistical methods were used to map soil characteristics. The methods consist of Block Kriging and Inverse Distance Weighting by GS* (a comprehensive geostatistics software program that is fast, efficient, and easy to use) and GIS software to predict soil factor. The cross validation was performed using the statistical parameters of Mean Absolute Errors (MAE) and Mean Biase Errors (MBE).

CCA was performed using PCORD, Ver. 4.17 program (McCune and Mefford 1999). In order to enable the allocation of the plant habitats into a range of values, the Z score was calculated. These values were then used to predict the location of the community on a prepared Z surface.

The probability of occurrence of each plant species according to Linear Regression was calculated with respect to the combined effect of site conditions using the following equation:

$$Y = \frac{\exp(LP)}{1+\exp(LP)} = \frac{\exp(b_0 + b_1x_1 + \ldots + b_nx_n)}{1 + \exp(b_0 + b_1x_1 + \ldots + b_nx_n)}$$

(1)
where: $b_j$ is the constant and $\exp$ is an exponential function and where $b_1$, $b_2$, ..., $b_n$ are the logic coefficients of $x_1$, $x_2$, ..., and $x_n$ variables respectively, in which presence/absence of an object is transformed into a continuous probability ranging from 0 to 1.

The models were calculated with individual selected variables and their combination using SPSS, 15.0. The best model is selected using two criteria: (i) approximate variance explained (Nagelkerka $R$ square) and (ii) goodness of fit (Hosmer and Lemeshow test statistics).

ANN models are capable of classifying data into more than two categories. Therefore, all vegetation types were included in the formulation of the ANN model and all 24 environmental variables were included in the procedure. The back propagation neural network (Cairns 2001) was used to generate the ANN model with one input, one hidden and one output layer. In the ANN approach, initial values of the weights are randomly assigned and for the model there are multiple possible solutions. This led that twenty-five ANN models were created for each number of nodes (7 and 10). Any of the ANN models with the highest accuracy value was then used to classify the data into two validation data sets. The mean accuracy based on two validation data sets for each ANN model (7–10 nodes) was compared to determine the ANN model which best classified the data. The best measure of agreement between observed and predicted presence-absence is Kappa ($\kappa$) statistic (Cohen 1960, Monserud and Leemans 1992, Fiedling and Bell 1997, Guisan and Ziemerlmann 2000, Moisen and Frescino 2002, Robertson et al. 2003, Liu et al. 2005). Kappa is used as the main measure to evaluate the models in this study. Monserud and Leemans (1992) suggested the following ranges of agreement for the $\kappa$ statistic: no agreement <0.05; very poor 0.05±0.20; poor 0.20±0.40; fair 0.40±0.55; good 0.55±0.70; very good 0.70±0.85; excellent 0.85±0.99; and perfect 0.99±1.00. Negative values indicate extremely poor agreement. The accuracy of predicted maps and adequacy of vegetation types mapping were evaluated using the Kappa statistic.

4. RESULTS

4.1. CCA - based modeling

Significance of the species-environment correlation was tested using the distribution-free Monte Carlo test (1000 permutations). In the Monte Carlo test, the distribution of the test statistics under the null hypothesis is generated by random permutations of cases in the environmental data. According to Table 1, first and second axes (Eigenvalue = 0.869 and Eigenvalue = 0.182) accounted for 98.7% and 4.4% variation in environmental data. The correlation between the first axis and species-environmental variables was 0.99 and the Monte Carlo permutation test for the first axis was highly significant ($P = 0.01$). Similarly, the correlation between the second axis and species-environmental variables was 0.92 and the Monte Carlo test for the second axis was highly significant ($P = 0.02$).

The canonical coefficients ($Z$ scores) were obtained from the CCA method according to the following equation (2):

$$Z = -0.2589 \text{gravel}_1 + 0.2691 \text{silt}_2 - 0.2437 \text{sand}_3 - 0.2356 \text{sand}_4 + 0.244 \text{available moisture}_5 + 0.2662 \text{gypsum}_6 + 0.2662 \text{gypsum}_7 + 0.2662 \text{EC}_8 + 0.2653 \text{EC}_9$$

Table 1. Canonical correspondence analysis for environmental data.

<table>
<thead>
<tr>
<th></th>
<th>Axis 1</th>
<th>Axis 2</th>
<th>Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>0.869</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>Variance in species data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of variance explained</td>
<td>98.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Cumulative % explained</td>
<td>98.7</td>
<td>99.1</td>
<td>99.4</td>
</tr>
<tr>
<td>Pearson Correlation, Spp-Envt*</td>
<td>0.998</td>
<td>0.920</td>
<td>0.959</td>
</tr>
<tr>
<td>Kendall (Rank) Corr., Spp-Envt</td>
<td>0.481</td>
<td>0.706</td>
<td>0.584</td>
</tr>
<tr>
<td>Species-Environment Correlation</td>
<td>0.998</td>
<td>0.920</td>
<td>0.959</td>
</tr>
<tr>
<td>$P$ value of Mont Carlo test</td>
<td>0.0100</td>
<td>0.0200</td>
<td>0.0100</td>
</tr>
</tbody>
</table>

* correlation between sample scores for an axis derived from the species data and the sample scores that are linear combinations of the environmental variables. Set to 0.000 if axis is not canonical.
Fig. 2. Spatial distribution of lime content in 0–20 cm depth performed by geostatistical method of Block Kriging.

Fig. 3. Predicted map of vegetation types using the CCA model. The vegetation was classified into 5 different groups or vegetation classes. Class 1 is the biggest one and most of vegetation types are located here; CCA model cannot separate these types, because they are similar. Classes 2, 3 and 4 contain the Zygophyllum eurypterum, Seidlitzia rosmarinus and Artemisia sieberi communities respectively. Class 5: the class is separated very well because of significant soil differences in Halocnemum strobilaceum community.
They were used to weight the importance of a specific environmental variable in determining vegetation type. Based on the predictive model obtained using the CCA method, predictive vegetation maps were generated in the GIS environment (Fig. 3). The results indicate that the CCA method is associated with a broader set of species.

4.2. Logistic Regression based modeling

The predicted occurrence probability of vegetation species is expressed via the equations (1–7) given in APPENDIX II. Regarding equation (1), the occurrence of *Artemisia aicheri* is dependent to the content of clay, lime and available moisture in the soil depth of 0–20 (W 1)

Based on the predictive model obtained using the LR method, predictive vegetation maps were generated in the GIS environment. Fig 4 shows the predicted map of *Astragalus* spp. using the logistic regression model.

4.3. ANN based modeling

The accuracy of the ANN models was variable and depended on the number of nodes existed in the hidden layer of the model. The ANN models with 7 and 10 nodes had the highest accuracy (i.e. 0.56). Since both the 7- and 10-node models produced the same accuracy (using the average of the two validation data sets), we chose the 7-node model as the best one because it performed better on the training data set than did the 10 node model. The average accuracy for the 7-node ANN model was 0.56 and $\kappa = 0.51$.

Fig 5 shows the predicted map of *Astragalus* spp. generated using the ANN model.

5. DISCUSSION AND CONCLUSIONS

There is good agreement between the CCA model and actual vegetation map ($\kappa = 0.679$). The comparison between the LR and ANN models are presented in Table 2. The results suggest that ANN models can predict as well or better than other models. The predictions obtained using the three modelling approaches provide different insights into the potential distribution and biology of the target organism, implying that each of the methods may be appropriate to be used in different situations. The choice of model is likely to be influenced by the aims of the study, the biology of the target organism, the level of knowledge on the target organism’s biology, and data quality.

This research showed three different modeling approaches for predicting plant species distribution. In this research, using the CCA, LR and ANN approaches, the regression equations were constructed utilizing topographic and soil characteristics from one side and vegetation features from the other side. Using the equations, a series of species predictive maps were then generated. The

<table>
<thead>
<tr>
<th>No</th>
<th>Vegetation type</th>
<th>Model</th>
<th>Kappa ($\kappa$)</th>
<th>levels of agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Artemisia aicheri</em>– Bromous tomentellus</td>
<td>LR</td>
<td>0.43</td>
<td>fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.52</td>
<td>fair</td>
</tr>
<tr>
<td>2</td>
<td><em>Astragalus</em> spp.</td>
<td>LR</td>
<td>0.86</td>
<td>excellent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.91</td>
<td>excellent</td>
</tr>
<tr>
<td>3</td>
<td><em>Eurotia ceratoides</em></td>
<td>LR</td>
<td>0.56</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.55</td>
<td>good</td>
</tr>
<tr>
<td>4</td>
<td><em>Artemisia sieberi</em></td>
<td>LR</td>
<td>0.33</td>
<td>poor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.3</td>
<td>poor</td>
</tr>
<tr>
<td>5</td>
<td><em>Seidlitzia rosmarinus</em></td>
<td>LR</td>
<td>0.6</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.72</td>
<td>very good</td>
</tr>
<tr>
<td>6</td>
<td><em>Zygophyllum eurypterum</em></td>
<td>LR</td>
<td>0.58</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.62</td>
<td>good</td>
</tr>
<tr>
<td>7</td>
<td><em>Halocnenum strobilaceum</em></td>
<td>LR</td>
<td>0.46</td>
<td>fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANN</td>
<td>0.53</td>
<td>fair</td>
</tr>
</tbody>
</table>
Fig. 4. Predicted map of *Astragalus* spp. using the logistic regression model (see APPENDIX II).

Fig. 5. Predicted map of *Astragalus* spp. using the ANN model.
results of the CCA analysis adequately well showed relative positions of species and sites along the most important ecological gradients. The CCA method is the preferable method for mapping rare species and where a low number of sample plots exists. In this method, it is not necessary to use quantification attributes like density, frequency, biomass and canopy cover. These properties are severely affected by the sampling method, shape, size and number of quadrate, as well as precipitation, whereas absence-presence is not dependent to above-mentioned factors.

The advantage of the CCA method is that here a model is constructed for all sampled species of a taxon at once, using only few model relevant environmental parameters. This means that every species is explained by the same parameters, and, more importantly, the ordination of species can be judged where there is a low number of observations. In fact, the small number of sampling plots in this study makes CCA a preferable option (Guisan et al. 1999). CCA appears easier than GLMs to implement into Geographical Information Systems (GIS). A drawback of this approach is that predictions are not probabilistic but are expressed as a distance from the centroid of each species.

Comparing the different modelling techniques showed that logistic regressions and artificial neural networks are acceptable choices for frequent (occurrence at many sample plots) species where there are a sufficient number of samples. It was also shown that the distribution of species could be better defined if the logistic regression model would be used. The ANN method demonstrated that it could predict the spatial distribution of species with an acceptable level of accuracy. There is no preference between the later two methods in modeling the distribution of species in general. The decision for a modeling method should always be guided by the questions to be answered by the model results (Guisan et al. 1999, Tan et al. 2006).

The results of Mi et al. (2010) showed that the ANN model was generally more accurate than the LR model. They suggested that artificial neural network models are more applicable than regression models. Also Özesmi et al. (2006) suggested that ANN models make good definitions of a study system but are too specific to generalize well to other ecologically complex systems unless input variable distributions are very similar.

Within the same group, the ANN method and the nonlinear logistic regression-based classifier were more accurate than the CCA model. The obtained results showed that predicted probability provides more valuable information about the performance of a predictive model than accuracy. Future research should be carried out to evaluate the applicability of the approaches by a wide variety of species and taxonomic groups and examine the necessary data requirements for both occurrence records and environmental coverage (Stockwell and Peterson 2002, Anderson 2003).

Similar research works will be necessary to apply dynamic models to simulate the relationship between changing environment space and the potential for species to disperse through fragmented landscapes, and to further our understanding of the complex dynamics of model systems consisting of multiple interacting species.

It is concluded that the combination of modelling of the local species distribution constitutes a promising future research area, which has the potentiality to enhance assessments and conservation planning of vegetation (like range-lands) based on predictive species models.

And choosing an appropriate population model for investigating plants’ distribution could lead to the identification of suitable habitat within a specified distance of an invasion front. It could be closely monitored for seedling occurrence, or corridors of unsuitable habitat that could be maintained between infested areas and vulnerable areas. In fact, the modelling and prediction of vegetation changes is essential in the environmental impacts assessment and decision making.

6. REFERENCES


Austin M. 2007 – Species distribution models and ecological theory: a critical assessment


Received after revision December 2011
## APPENDIX I. Environmental factors and SD (±) of mean values in the study area (for different vegetation types).

<table>
<thead>
<tr>
<th></th>
<th>Depth (cm)</th>
<th>Gravel (%)</th>
<th>Clay (%)</th>
<th>Silt (%)</th>
<th>Sand (%)</th>
<th>Lime (%)</th>
<th>Organic matter (%)</th>
<th>Available P (%)</th>
<th>Gypsum (%)</th>
<th>EC (ds/m)</th>
<th>pH (- log H⁺)</th>
<th>Elevation (m)</th>
<th>Slope (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artemisia aucheri-Eurotia ceratooides</td>
<td>0-20</td>
<td>28.22±2.7</td>
<td>22.17±1.1</td>
<td>20.50±0.72</td>
<td>57.33±1.31</td>
<td>13.89±1.13</td>
<td>0.60±0.11</td>
<td>6.39±0.70</td>
<td>0.00±0.00</td>
<td>0.255±0.03</td>
<td>8.10±0.04</td>
<td>1756±1.00</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>45.63±3.48</td>
<td>21.00±2.40</td>
<td>18.33±2.03</td>
<td>60.67±3.60</td>
<td>14.93±0.77</td>
<td>0.63±0.26</td>
<td>5.03±1.71</td>
<td>0.00±0.00</td>
<td>0.277±0.05</td>
<td>8.20±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halocnemum strobilaceum</td>
<td>0-20</td>
<td>0.00±0.00</td>
<td>26.8±4.23</td>
<td>37.33±5.38</td>
<td>35.83±8.71</td>
<td>21.70±2.12</td>
<td>0.50±0.14</td>
<td>17.30±2.16</td>
<td>14.33±4.75</td>
<td>8.85±1.5</td>
<td>8.10±0.11</td>
<td>1132±2.02</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>2.84±1.74</td>
<td>29.33±3.45</td>
<td>35.33±5.51</td>
<td>35.33±7.63</td>
<td>22.26±0.45</td>
<td>0.43±0.12</td>
<td>17.17±2.39</td>
<td>13.91±2.93</td>
<td>7.20±1.71</td>
<td>8.1±0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artemisia sieberi-Zygodylum eurypterum</td>
<td>0-20</td>
<td>35.52±2.83</td>
<td>17.50±1.03</td>
<td>20.33±2.49</td>
<td>62.17±3.04</td>
<td>22.43±0.91</td>
<td>0.53±0.09</td>
<td>7.58±0.58</td>
<td>0.02±0.00</td>
<td>0.27±0.02</td>
<td>8.2±0.03</td>
<td>1442±1.05</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>50.03±7.91</td>
<td>16.00±1.13</td>
<td>16.17±3.29</td>
<td>67.83±3.90</td>
<td>20.24±2.06</td>
<td>0.44±0.11</td>
<td>7.24±0.77</td>
<td>0.00±0.00</td>
<td>0.28±0.03</td>
<td>8.3±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zygodylum eurypterum-Artemisia sieberi</td>
<td>0-20</td>
<td>27.59±2.12</td>
<td>16.67±1.17</td>
<td>25.00±3.08</td>
<td>58.33±3.21</td>
<td>20.28±0.80</td>
<td>0.34±0.07</td>
<td>8.47±0.93</td>
<td>0.00±0.00</td>
<td>0.22±0.01</td>
<td>8.3±0.02</td>
<td>1516±2.34</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>36.44±5.15</td>
<td>23.67±0.84</td>
<td>17.83±1.97</td>
<td>58.50±2.19</td>
<td>17.95±0.56</td>
<td>0.38±0.08</td>
<td>6.91±0.35</td>
<td>0.00±0.00</td>
<td>0.22±0.01</td>
<td>8.3±0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artemisia aucheri-Astragalus spp-Bromus tomentellus</td>
<td>0-20</td>
<td>28.48±3.17</td>
<td>26.45±1.58</td>
<td>25.50±0.89</td>
<td>48.00±1.77</td>
<td>11.22±1.06</td>
<td>0.93±0.08</td>
<td>12.68±1.30</td>
<td>0.00±0.00</td>
<td>0.21±0.01</td>
<td>8.2±0.02</td>
<td>1760±110.12</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>47.64±5.99</td>
<td>33.17±2.04</td>
<td>22.50±2.16</td>
<td>44.33±3.82</td>
<td>13.89±2.01</td>
<td>0.62±0.07</td>
<td>12.79±1.30</td>
<td>0.00±0.00</td>
<td>0.17±0.01</td>
<td>8.2±0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seidlitzia rosmarinus</td>
<td>0-20</td>
<td>28.15±2.64</td>
<td>22.83±1.17</td>
<td>30.00±2.07</td>
<td>47.17±2.79</td>
<td>15.80±0.55</td>
<td>0.70±0.12</td>
<td>11.91±0.87</td>
<td>0.02±0.00</td>
<td>0.31±0.05</td>
<td>8.2±0.05</td>
<td>1826±4.00</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>20-80</td>
<td>37.48±5.25</td>
<td>20.67±3.10</td>
<td>21.50±3.05</td>
<td>54.00±5.70</td>
<td>15.31±2.00</td>
<td>0.60±0.13</td>
<td>11.52±1.06</td>
<td>0.02±0.00</td>
<td>0.18±0.04</td>
<td>8.3±0.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX II. The Linear Regression equations for different species of study area. Description of variables see APPENDIX I, W1 – soil depth 0–20 cm, W2 – soil depth 20–80 cm.

\[
P(\text{Artemisia aucheri}) = \frac{\text{Exp}(4.881 \text{clay} - 13.369 \text{lime} + 7.759 \text{Available moisture} - 42.688)}{1 + \text{Exp}(4.881 \text{clay} - 13.369 \text{lime} + 7.759 \text{Available moisture} - 42.688)} \quad (1)
\]

\[
P(\text{Bromus tomentellus}) = \frac{\text{Exp}(4.881 \text{clay} - 13.369 \text{lime} + 7.759 \text{Available moisture} - 42.688)}{1 + \text{Exp}(4.881 \text{clay} - 13.369 \text{lime} + 7.759 \text{Available moisture} - 42.688)} \quad (2)
\]

\[
P(\text{Astragalus}) = \frac{\text{Exp}(37.659 \text{slope} + 1.253 \text{elevation} + 66.024 \text{Available moisture} - 3138.027)}{1 + \text{Exp}(37.659 \text{slope} + 1.253 \text{elevation} + 66.024 \text{Available moisture} - 3138.027)} \quad (3)
\]

\[
P(\text{Artemisia sieberi}) = \frac{\text{Exp}(36.64 \text{lime} - 86.576 \text{Available moisture} + 83.14)}{1 + \text{Exp}(36.64 \text{lime} - 18.287 \text{Available moisture} + 83.14)} \quad (4)
\]

\[
P(\text{Eurotia ceratoides}) = \frac{\text{Exp}(0.118 \text{slope} - 2.864)}{1 + \text{Exp}(0.118 \text{slope} - 2.864)} \quad (5)
\]

\[
P(\text{Zygoiphyllum eurypterum}) = \frac{\text{Exp}(-0.854 \text{clay} - 18.287 \text{gypsum} + 16.97)}{1 + \text{Exp}(0.854 \text{clay} - 18.287 \text{gypsum} + 16.97)} \quad (6)
\]

\[
P(\text{Halocnemum strobilacum}) = \frac{\text{Exp}(9.85 E1 - 23.631)}{1 + \text{Exp}(9.85 E1 - 23.631)} \quad (7)
\]

\[
P(\text{Seidlitzia rozmarinus}) = \frac{\text{Exp}(0.211 \text{elevation} - 12.194 \text{slope} - 242.79)}{1 + \text{Exp}(0.211 \text{elevation} - 12.194 \text{slope} - 242.79)} \quad (8)
\]