JOURNAL OF FOREST SCIENCE, *62*, 2016 (12): 562–570

doi: 10.17221/97/2016-JFS

Investigation of a possibility of spatial modelling of tree diversity using environmental and data mining algorithms

A. Abdollahnejad, D. Panagiotidis, P. Surový

Department of Forest Management, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

ABSTRACT*:* Biological diversity is the basis for a wide array of goods and services provided by forests. The variety of forest trees and shrubs plays a vital role in the daily life of forest communities. The purpose of this study is to investigate the possibility of modelling the diversity of tree species by characteristics of topography, soil and climate, using data mining algorithms *k*-NN, RF and SVM in Dr. Bahramnia forestry plan in the north of Iran. Based on the basal area factor for each species in a total of 518 sample plots, diversity indices such as species richness, evenness and heterogeneity were calculated for each plot. Topographic maps of primary and secondary properties were prepared using the digital elevation model. Categories of the soil and climate maps database of Dr. Bahramnia forestry plan were extracted. Modelling rates of tree and shrub species diversity using data mining algorithms, with 80% of the sampling plots were taken. Assessment of the model accuracy, using 20% of samples and evaluation criteria, was conducted. Results showed that topographic features, especially elevation, had the highest impact on the species diversity index. The modelling results also showed that Camargo evenness index had lowest root mean square error (RMSE) (0.14) and RMSE% (24.35), compared to other indicators of diversity. In addition, the results of the comparison between the algorithms showed that the random forest algorithms were more accurate in modelling the diversity.

Keywords: topographic characteristics; suborder soil; climate; non-parametric algorithms; richness; evenness indicators

Diversity or taxonomy is the middle level of hierarchical biodiversity classification and its purpose is to assess the diversity of plant and animal species within certain areas (van der Maarel 2005). The main concern is to compare taxonomic groups in different geographical areas. Diversity is a significant part of biodiversity and can be divided into two categories: the first is richness and the second is evenness (Ludwing, Reynolds 1988; Maguran 1996; Krebs 1999). Richness is one of the basic indicators for measuring diversity in terms of the region and it has a direct, scientific effect on diversity, meaning that the higher the number of species, the higher the diversity. On the other hand, evenness is an indicator that shows the distribution of trees in different classes of species (Ejtehadi et al. 2010). When comparing two different populations that have the same richness, the population that has higher evenness has higher diversity. On the contrary, when comparing two different populations that have the same evenness but not the same richness, the population with higher richness has higher diversity. In places where these two components (richness and evenness) are different, identifying areas with higher diversity becomes a difficult task. In order to solve this problem, non-parametric methods are used through the combination of richness and evenness components (Ardestani et al. 2010).

Forest diversity in the northern part of Iran is one of the richest forest ecosystems of the temperate forests. In order to achieve sustainable devel-

Supported by the Czech University of Life Sciences Prague, Project No. B07/15.

opment, further study of ecological and environmental factors which affect this ecosystem and the diversity of this forest is crucial (Fallahchay, Marvie Mohajer 2005; Marvie Mohajer 2006). Some of the ecological factors which can affect biodiversity are elevation, aspect, slope, climate, and human activities (Ejtehadi et al. 2010). So far there are different kinds of studies that have been conducted in different territories, trying to predict or investigate diversity distribution related to topographic factors (Pourbabaee 1998; Marvie Mohajer 2005; Gracia et al. 2007; Ismailzadeh, Hosseini 2007; Ghanbari 2008; Saatchi et al. 2008; Kymasi 2012), edaphic (Qomioghli et al. 2006; Ejtehadi et al. 2010; Kymasi 2012) and climatic factors (Mehdinya et al. 2006; Parmentier 2011; Gixhari et al. 2012). One of the main purposes of modelling research is to clarify the most appropriate method, regarding the spatial prediction of forest characteristics, based on sampling methods (Kint et al. 2003).

Spatial analyses belong among the non-classical methods for data processing in order to estimate the information on unmeasured areas (Whittaker 1977). The principle of these models is based on the hypothesis that environmental factors can control the spatial distribution of plants (Ghanbari et al. 2011). Different kinds of models such as remote sensing techniques, geostatistics, generalized linear regression, neural networks, nearest neighbours, decision trees and their variants, for example random forest, have been used to predict the biological characteristics of forests. According to Franklin (1998) and Shataee et al. (2012), nearest neighbours, support vector machine and random forest, which can do both classification and regression, are the most common algorithms among data mining algorithms. Many researchers used these algorithms to model the quantity of forest characteristics (Ismail, Mutango 2010; O'Sullivan et al. 2010; Parmentier 2011; Yazdani 2011; Shataee et al. 2012). Moreover, the advantages of non-parametric algorithms rely on the fact that they are not sensitive to a high number of independent variables as inputs for modelling.

The objective of this study is to investigate the possibility of modelling the spatial distribution of tree diversity using topographic, edaphic and climatic factors through three different types of algorithms.

MATERIAL AND METHODS

Study area. The study area is located in the southwest from Gorgan city in Golestan province in Iran (36°43'–36°46'N, 54°21'–54°24'E). The total area is

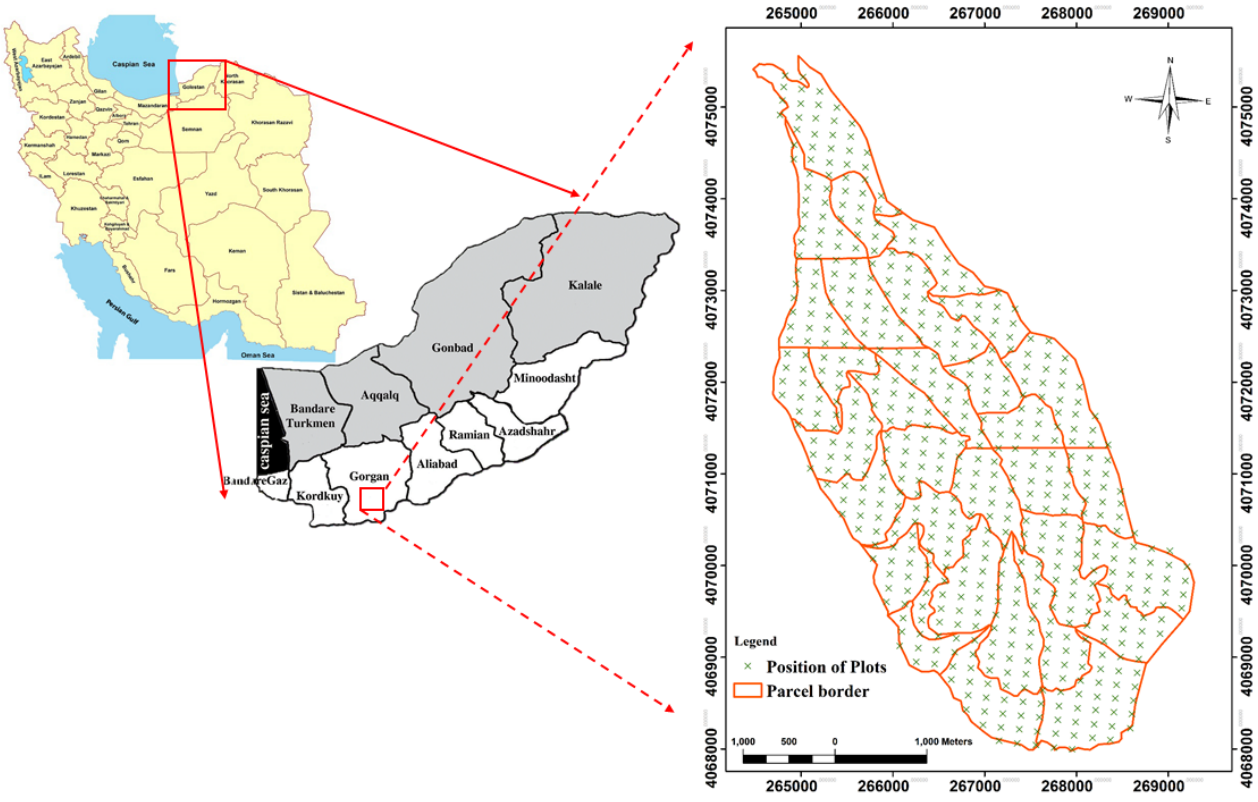

Fig. 1. Location of plots in Dr. Bahramnia's forestry district, Golestan province, northern Iran

Table 1. Primary topographic attributes that were computed by terrain analysis from digital elevation model data in this study (Beven, Kirkby 1979; Moore et al. 1993; Wilson, Gallant 2000)

Characteristic	Definition	Importance
Altitude	elevation	vegetation, climate conditions, solar energy
Slope	gradient	flow rate, precipitation, vegetation, flow velocity, soil conditions
Aspect	slope azimuth	evapotranspiration, species distribution (fauna and flora), solar energy
Specific catchment area	used to estimate saturation excess overland flow	runoff volume and rate, soil characteristics, water viscosity, geomorphological conditions
Profile curvature	slope profile curvature	runoff acceleration, erosion/deposition percentage, geomorphological conditions
Tangential curvature	plan curvature multiplied by slope	an alternative measure of local flow conditions and divergence
Plan curvature	contour curvature	soil and water content, soil characteristics

1,714 ha (Fig. 1). The elevation of the study area ranges from 220 to 1,012 m a.s.l. and the slope range is between 0 and 80% and the soil type is brown and grey-brown. The average precipitation is 649 mm.

In regard to the aspect, 45% of the total area is facing the west, 42% the north, 10% the east and 3% the south. Concerning the forest type, the majority of the area is covered by *Carpinus-Zelkova*. There were also other types of forest such as: *Zelkova-Carpinus*, *Fagus-Carpinus* and *Fagus*.

Ground Data. To calculate the diversity indicators, we used the information on measured trees from 518 permanently visible sample plots with a radius of 17.5 m that were used in a systematic network of 150 × 200 m grids. The geographic position of plot centres and forest attributes such as diameter, height, crown diameter, name of species and tree health status were recorded on inventory forms. We included all trees with diameters greater than 10 cm for measuring the diversity indicators. Based on the basal area factor for each species in each plot, we were able to extract and calculate the amount of diversity indicators using the ecological methodology software by Krebs (1999).

We then constructed a digital elevation model – DEM (cell size = 30 m) for the study area using the topographic map (1:25000 scale and with 10 m contour interval). To check the DEM quality, we first used the hillshade tool to create a shaded relief from DEM, by considering the illumination source angle and shadows in order to be able to visually identify large errors (noise and sudden change in values). Then, we used the contour tool to recreate the contours from DEM (15 m contour interval), and finally we compared the original contours with new contours which came from interpolation.

Using a variety of software such as ArcGIS (Version 9.3, 2008) and TAS (Version 1.0, 2003), we used DEM to construct primary and secondary topographic characteristic maps (Tables 1 and 2). Climatic maps such as: average annual precipitation (Eq. 1), average annual temperature (Eq. 2) and average annual evaporation (Eq. 3) were made using DEM in ArcGIS software and the formula for the Ghare Sou area based on information of metrological stations for our study area during two decades (Byroodyan 1990):

$$Y = \frac{282X^2 - 285{,}000X + 18.10^7}{X^2 - 1{,}000X + 45.10^4} \quad (1)$$

where:
Y – average annual precipitation (mm),
X – elevation.

$$T = -0.006X + 17.75 \quad (2)$$

where:
T – average annual temperature (°C).

$$ETP = 651 - 0.092X \quad (3)$$

where:
ETP – average annual evaporation (mm).

The zonal statistics algorithm was used for extracting the suborder soil factor from Kardgar (2012) and other layers such as topographic and climatic factors using a buffer layer around the centre of the plots (17.5 m radius).

Diversity indicators. There are two different methods for measuring diversity: (*i*) numerical indicators, (*ii*) non-numerical or parametrical indicators. Numerical indicators present one number as a result. These types of indicators using richness component or evenness component or both together, can be divided into three categories: (*i*) richness indicators, (*ii*) evenness indicators, (*iii*) heterogeneity indicators. In this study we choose five dif-

Table 2. Secondary topographic attributes that were computed by terrain analysis from digital elevation model data in this study (Beven, Kirkby 1979; Moore et al. 1993; Wilson, Gallant 2000)

Characteristics	Definition	Importance
	$SPI = A_s \tan \beta_R$ where: A_s – specific catchment area, β_R – local slope angle.	It is a measure of erosive power of flowing water, predicts tangential concavity and net deposition in areas of profile concavity and net erosion in areas of profile convexity.
Stream power indices (*SPI*)	$LS = (M + 1)\left(\frac{A_s}{22.13}\right)^m \left(\frac{\sin \beta}{0.0896}\right)^n$ where: LS – length-slope factor, M = 0, A_s – specific catchment area ($m^2 \cdot m^{-1}$), m = 0.4, β (^{0}) – slope gradient, n = 1.3.	It is the Revised Universal Soil Loss Equation in certain circumstances, predicts locations of net erosion and net deposition areas.
Topographic wetness index	$\ln \frac{a}{\tan b}$ where: a – local upslope area draining through a certain point per unit contour length, b – local slope in radians.	For steady-state flow conditions, it describes the spatial distribution of the saturation zone for runoff generation, soil transition, slope gradient.
Radiation indices	$R_{ne} = (1 - \alpha)\, R^{\downarrow} + \sigma\, (\varepsilon_a\, T_a^4 - \varepsilon_s\, T_s^4)$ where: R_{ne} – estimated net radiation ($W \cdot m^{-2}$), α – albedo (dimensionless), $R^{\downarrow}$ – incoming short wave solar radiation ($W \cdot m^{-2}$), σ – Stefan-Boltzmann constant (5.67×10^{-8} $W \cdot m^{-2} \cdot {}^{\circ}K^{-4}$), ε_a – atmospheric emissivity (dimensionless), determined according to equation $\varepsilon_a = \phi(e_a / T_a)1/7$ (Brutsaert 1975), ϕ – empirical coefficient, e_a – air vapor pressure (kPa), T_a – air temperature (°K), ε_s – surface emissivity (dimensionless), T_s – surface temperature (°K).	The three main terms account for direct-beam, diffuse, and reflected irradiance. A variety of methods are used by different authors to calculate these individual components. The methods vary tremendously in terms of sophistication, input data, and accuracy.

ferent kinds of indicators for measuring diversity based on previous research, study of changing the tree and shrub diversity in different environmental conditions which was conducted by Abdollahnejad and Shataee (2014).

(1) Richness indicators. These indicators are the simplest and oldest method for measuring the diversity and they are based on the number of species (s) and total number of individuals in the sample (N), one of these indicators is so called Menhinick's index – D_{Mn} (Whittaker 1977), as Eq. 4:

$$D_{Mn} = \frac{s}{\sqrt{N}} \tag{4}$$

(2) Evenness indicators. For evenness indicators we used two different types of index:

(*i*) Camargo index (E'). Richness and scarce species cannot influence the Camargo index (Camargo 1992), as Eq. 5:

$$E' = 1 - \left(\sum_{i=1}^{S} \sum_{j=i+1}^{S} \left[\frac{|P_i - P_j|}{S}\right]\right) \tag{5}$$

where:

P_i – number of i species,

P_j – number of j species,

S – total number of species.

(*ii*) Smith and Wilson index (E_{var}). This index, suggested by Smith and Bastow Wilson (1996), is based on the variance of species frequency, as Eq. 6:

$$E_{var} = 1 - \left[\frac{2}{\pi \arctan\left\{\sum_{i=1}^{S} (\log_e(n_i) - \sum_{j=1}^{S} \log_e(n_i)/S)^2 / S\right\}}\right] \tag{6}$$

where:

n_i – number of i species,

n_j – number of j species,

S – total number of species in all the samples.

(3) Heterogeneity indicators.

(*i*) Shannon-Wiener index (*H'*). This index is the most common index and it can be calculated using Shannon and Weaver (1949), as Eq. 7:

$$H' = -\sum_{i=1}^{s} P_i \ln P_i = -\sum_{i}^{s} (P_i)(\log_2 P_i) \quad (7)$$

where:

P_i – percentage of *i* species,

s – number of species.

(*ii*) Simpson index (*D*). This is the most popular and the first non-parametric index for diversity and it is more sensitive to evenness than to richness. It can be calculated using Simpson (1949), as Eq. 8:

$$1 - D = 1 - \sum_{i=1}^{s} P_i^2 \quad (8)$$

where:

P_i – percentage of *i* species,

s – number of species.

Calculation of diversity indicators. For each plot, we calculated the cross-sectional area of each tree using the diameter, and then according to the tree species we were able to calculate the total basal area of each species in a sample plot using the ecological methodology software by Krebs (1999).

Applied algorithms. The prediction of tree diversity was based on 80% of the samples (414 plots) with different independent variables, such as topography, climate, edaphic factors and several different combinations of these variables.

Three different kinds of data mining algorithms were used to predict the diversity of tree species in the whole study area (all measured plots), and for our calculations we used the Statistica software (Version 7.0.61, 2006).

k-Nearest Neighbour (*k*-NN) is the most common algorithm based on training samples. The hypothesis of this algorithm is that all the samples are located in a space with *n* dimensions and it specifies the neighbours, based on the standard Euclidean distance.

The meaning of *k* is the number of nearest neighbours. In order to find the optimal number of *k*, we used the cross-validation method with *k* ranging from 1 to 50. To measuring the metric distance between the known (neighbour plots) and unknown plots (estimated plots using neighbours data), we used weighted Euclidean distance as the most appropriate option, providing comparison with other options that the software has.

Another algorithm that we used for estimating the diversity of species was Random Forest (RF) algorithm. To apply RF algorithm, we used 400 decision trees and we considered 5 as the minimum and 100 as the maximum number of nodes for each decision tree.

The final algorithm that we used was the so-called Support Vector Machine (SVM). The option that we used considers two kernel functions where one is the so-called type-1 regression and the other is the Radial Basis Function (RBF). Moreover, for improving the accuracy of the model prediction, we used the cross-validation method (Yazdani 2011; Shataee et al. 2012).

Assessing the accuracy of modelling. The purpose of this assessment was mainly to investigate the ability of the models that we used for estimating diversity using training samples. We used relative root mean square error (RMSE), RMSE%, BIAS and BIAS% in order to evaluate the accuracy of the algorithm results.

RESULTS

The results of the inventory plots showed that the majority of the populations comprised 8 different species. In addition, the number of species as described in Table 3 had a high variance between the plots. As an example, based on inventory data we observed that in some plots there was only one spe-

Table 3. Statistical table displays the species distribution based on inventory data

Species	Cross section area (cm^2)			Basal area (cm^2)	Frequency	Species occurrence by plot
	minimum	maximum	average			
Fagus orientalis Lipsky	78.50	18,859.62	2,514.16	3,539,937.87	1,408	243
Carpinus betulus Linnaeus	78.50	20,096	1,568.18	5,087,192.50	3,244	440
Quercus castaneifolia C.A. Meyer	78.50	9,498.5	689.99	234,597.25	340	38
Alnus subcordata C.A. Meyer	78.50	18,859.62	795.95	506,226.87	636	102
Acer velutinum Boissier	176.60	1,256	330.35	11,892.75	733	247
Zelkova carpinifolia (von Pallas) C. Koch	78.50	5,671.62	330.20	200,430.12	36	6
Diospyros lotus Linnaeus	78.50	5,671.62	330.20	200,430.12	607	194
Parrotia persica (de Candolle) C.A. Meyer	78.50	13,266.50	703.41	3,014,812.12	4,286	426
Other species	176.62	3,316.62	488.11	83,955.75	172	34

Table 4. Results of the application of data mining algorithm in a case study (richness-evenness indicators)

Independent variables	Algorithms	Index											
		Menhinick's				Camargo				Smith and Wilson			
		RMSE	RMSE%	BIAS	BIAS%	RMSE	RMSE%	BIAS	BIAS%	RMSE	RMSE%	BIAS	BIAS%
Topography	*k*-NN	0.24	31.12	–0.18	–23.72	0.24	43.05	–0.10	–17.5	0.30	68.73	–0.205	–45.53
	SVM	0.26	33.19	–0.11	–13.43	0.24	42.45	–0.11	–19.63	0.28	64.42	–0.205	–49.7
	RF	0.25	31.12	–0.16	–20.3	0.14	24.35	–0.13	–23.63	0.24	51.8	–0.16	–35.88
Soil	*k*-NN	0.38	47.99	–0.06	–6.93	0.37	65.19	–0.29	–81.25	0.47	90.18	–0.207	–139.50
	SVM	0.30	37.92	–0.18	–24.02	0.24	43.22	–0.14	–26.53	0.31	72.30	–0.205	–41.55
	RF	0.25	32.37	–0.17	–21.018	0.23	39.13	–0.13	–23.86	0.27	60.03	–0.14	–30.53
Climate	*k*-NN	0.25	31.99	–0.17	–21.3	0.24	42.65	–0.11	–19.90	0.30	70.62	–0.205	–46.21
	SVM	0.25	31.74	–0.18	–23.59	0.24	42.82	–0.11	–18.59	0.28	64.81	–0.205	–49.46
	RF	0.26	32.90	–0.17	–21.76	0.17	30.01	–0.12	–21.77	0.28	59.50	–0.164	–34.88
All variables	*k*-NN	0.24	30.63	–0.18	–23.59	0.24	43.06	–0.10	–18.19	0.27	63.32	–0.205	–45.53
	SVM	0.27	33.92	–0.11	–12.43	0.25	44.09	–0.11	–20.30	0.30	69.11	–0.205	–47.81
	RF	0.28	34.33	–0.21	–26.85	0.16	24.74	–0.13	–23.65	0.26	54.67	–0.16	–35.46
10 affecting layers	*k*-NN	0.24	30.96	0.17	–21.67	0.24	42.89	–0.10	–18.04	0.28	65.82	–0.205	–44.60
	SVM	0.27	34.29	–0.11	–12.51	0.24	42.53	–0.11	–19.47	0.30	68.75	–0.205	–47.99
	RF	0.26	33.28	–0.01	–22.19	0.15	26.32	–0.14	–24.15	0.25	54.12	–0.162	–34.25

k-NN – *k*-Nearest Neighbour algorithm, SVM – Support Vector Machine algorithm, RF – Random Forest algorithm, RMSE – root mean square error

cies (low richness) while in other plots we found up to seven different species (high richness).

Using different combinations of independent variables: (*i*) topographic, (*ii*) edaphic, (*iii*) climatic ones, we created continuous maps for the whole study area where the results of accuracy evaluation by using 20% of the plots (testing plots) showed that the RF algorithm had the highest accuracy ~99% of indicators, compared to the other two algorithms. More specifically, the results of RF showed RMSE% = 24.35 for Camargo index, RMSE% = 51.81 for Smith and Wilson index, RMSE% = 34.02 for Simpson index and finally RMSE% = 34.69 for Shannon-Wiener index. Menhinick's index had the highest accuracy using the *k*-NN algorithm with RMSE% = 30.63 (Tables 4 and 5).

Table 5. Results of the application of data mining algorithm in a case study (heterogeneity indicators)

Independent variables	Algorithms	Index							
		Shannon-Wiener				Simpson			
		RMSE	RMSE%	BIAS	BIAS%	RMSE	RMSE%	BIAS	BIAS%
Topography	*k*-NN	0.52	45.84	–0.904	–78.11	0.20	46.45	–0.07	–15.7
	SVM	0.51	45.17	–0.904	–80.99	0.19	44.84	–0.09	–20.69
	RF	0.46	38.07	–0.89	–77.14	0.17	37.05	–0.11	–25.66
Soil	*k*-NN	1.27	90.01	–0.904	–237.78	0.41	93.67	–0.32	–21.7
	SVM	0.56	49.57	–0.904	–73.42	0.21	49.23	–0.14	–35.65
	RF	0.49	41	–0.88	–73.62	0.19	40.30	–0.12	–26.73
Climate	*k*-NN	0.52	46.21	–0.909	–79.95	0.19	43.77	–0.085	–18.47
	SVM	0.50	44.57	–0.907	–76.50	0.19	84.43	–0.09	–20.92
	RF	0.44	36.54	–0.89	–74.83	0.17	36.28	–0.10	–23.26
All variables	*k*-NN	0.51	45.06	–0.904	–77.90	0.19	44.64	–0.08	–17.24
	SVM	0.52	46.25	–0.905	–82.57	0.21	47.28	–0.13	–32.37
	RF	0.43	34.69	–0.902	–80.28	0.16	34.02	–0.13	–15.42
10 affecting layers	*k*-NN	0.50	44.09	–0.904	–78.31	0.20	44.96	–0.07	–15.42
	SVM	0.51	44.95	–0.904	–81.11	0.20	46.91	–0.14	–35.86
	RF	0.48	39.17	–0.900	–77.76	0.17	37.44	–0.12	–26.69

k-NN – *k*-Nearest Neighbour algorithm, SVM – Support Vector Machine algorithm, RF – Random Forest algorithm, RMSE – root mean square error

Feature Selection and Variable Screening algorithms were used to find the 10 most important predictors that highly influence the variable of interest (Table 6). The results from estimation using these layers showed that the RF algorithm had the highest accuracy of modelling evenness and heterogeneity indicators and *k*-NN had the highest prediction accuracy of richness index (RMSE% = 30.96).

Table 6. The most significant predictors (independent variables) in modelling diversity indicators

Independent variables	*P*-value
Elevation (most affecting factor)	0.0009
Tangential curvature	0.0848
Slope percentage	0.0417
Solar radiation	0.1223
Suborder soil	0.1254
Temperature	0.1824
Evaporation	0.1824
Aspect	0.2584
Specific catchment area	0.3373
Precipitation	0.3463

DISCUSSION AND CONCLUSIONS

According to the results of Menhinick's index (richness index) and heterogeneity indicators, we found out that modelling based on all independent variables had the highest accuracy. These results also showed that evenness maps which were made by topographic variables had the highest accuracy. From the conclusions above it is evident that richness and heterogeneity indicators are more connected to environmental factors like soil and climate, compared to the evenness indicators. According to the results of Table 6 it is clear that topographic factors, specifically elevation, had the most significant influence on the distribution of tree diversity. Changes in elevation can cause different environmental conditions such as temperature, precipitation, moisture, solar radiation, air pressure etc. and eventually they can affect the spatial distribution of trees. Moreover, due to the difficulties of human activities (selective cutting) at higher elevations these areas are closer to natural patterns (different frequency of species). Many studies like Peffer et al. (2003), Marvie Mohajer (2006), Gracia et al. (2007), Guoyu (2011), Shirzad and Tabari (2011), Kymasi (2012), and Momeni Moghaddam et al. (2012) have shown that elevation is the most important environmental factor affecting the spatial distribution of species. This conclusion is also verified in our study. Additional research of Bale et al. (1998) and Ghanbari et al. (2011) showed that topographic factors like slope and aspect can also influence the spatial distribution of tree species. Different amounts of solar radiation, exposure, airflow pressure, water flow sources, density of cloud cover and fog in different aspects can cause different environmental conditions that can affect the distribution and combination of tree species.

In addition, a varying amount of soil drainage in different percentages of slope can affect soil conditions and create an altered habitat for tree species.

Solar radiation is one of the secondary topographic characteristics that cause a significant impact on the spatial distribution of trees. According to Table 6, solar radiation had the greatest affect among the secondary topographic characteristics (Peffer et al. 2003; Saatchi et al. 2008; Ghanbari et al. 2011). Overall, using topographic layers can yield reasonable outputs in the modelling of spatial distribution of plants and diversity (Zimmermann, Kienast 1999).

In our study, according to Table 6, soil characteristics such as soil nutrients, soil depth, soil humidity and types of soil were some of the most important factors in the distribution of tree species because they influence site conditions (McKenney, Pedlar 2003; Stephenson et al. 2006).

Regarding the algorithms that were used in this paper, the results comparing different algorithms showed that RF had the highest accuracy compared to both *k*-NN and SVM. The ability of RF in determining important coefficients, weighting the independent variables and its non-necessity of decision tree structure pruning are all factors that enhance the functionality and effectiveness of this algorithm. On the contrary, *k*-NN and SVM algorithms use the same proportions of weighting for all independent variables (Kernes, Ohmann 2004). Comparing this with other algorithms, data mining algorithms are easier to comprehend; they need little data preparation (no need to normalize data) and can handle numerical and categorical data, in addition to that, a large amount of data can be analysed by data mining algorithms in a reasonable time. As a result, this enables researchers to study the effects of different independent variables on one factor simultaneously. Also, data mining algorithms such as RF can compare the effectiveness of independent variables and consider different weights for independent layers (input data) in the modelling process. Data mining algorithms with this potential for classification and regression of forest attributes at a high level of accuracy and robustness can be used as a promising approach in a wide range of forest research.

In conclusion, the above forestry plan was under management for many years, which affected the structure and combination of tree species and disconnected the links between the current forests and environmental factors. In other words, this type of managed forest appears to be completely different from a natural forest and as a result there is no pure connection between this forest and environmental conditions. In future work, we suggest this type of research for unmanaged forests which are not influenced by human activities and are more connected to topographic and edaphic factors.

References

Abdollahnejad A., Shataee S.H. (2014): The study of tree and shrub species diversity changes in the parameters of a physiographic, soil and vegetation. District one of DR. Bahramnia forestry plan. Journal of Wood and Forest Science and Technology, 21: 61–84. (in Persian)

Ardestani E., Basir M., Torkesh M., Borhani M. (2010): Indicators for assessment of pasture species diversity in four places in Isfahan province. Journal of Rangeland, 4: 43–46. (in Persian)

Bale C.L., Williams J.B., Charley J.L. (1998): The impact of aspect on structure and floristics in some Eastern Australian site. Forest Ecology and Management, 110: 363–377.

Beven K.J., Kirkby M.J. (1979): A physically based, variable contributing area model of basin hydrology. Hydrolological Science Bulletin, 24: 43–69.

Brutsaert W. (1975): On a derivable formula for longwave radiation from clear skies. Water Resources Research, 11: 742–744.

Byroodyan M. (1990): Weather and Climatology (Ghare Sou River Watershed Studies). Gorgan, Agriculture Publication: 300. (in Persian)

Camargo J.A. (1992): New diversity index for assessing structural alterations in aquatic communities. Bulletin of Environmental Contamination and Toxicology, 48: 428–434.

Ejtehadi H., Sepehry A., Akafi H.R. (2010): Methods of Measuring Biodiversity. Mashhad, University of Mashhad: 288. (in Persian)

Fallahchay M.M., Marvie Mohajer M.R. (2005): Ecological role of elevation on tree diversity of Siahkal forest in the north of Iran. Iranian Journal of Natural Resources, 58: 89–101. (in Persian)

Franklin J. (1998): Predicting the distribution of shrub species in Southern California form climate and terrain-derived variables. Journal of Vegetable Science, 9: 733–748.

Ghanbari F. (2008): Predicting the spatial distribution of forest allometric growth properties using geostatistics and GIS. [MSc Thesis.] Gorgan, Gorgan University of Agricultural Sciences and Natural Resources: 160. (in Persian)

Ghanbari F., Shataee S.H., Mohseni A., Habashi H. (2011): Using a logistic regression model to predict the spatial characteristics of topography and forest type. Iranian Journal of Forest and Poplar Research, 19: 27–41. (in Persian)

Gixhari B., Ismaili H., Vrapi H., Elezi F., Dias S., Sulovari H. (2012): Geographic distribution and diversity of fruit tree species in Albania. International Journal of Ecosystems and Ecology Sciences, 2: 355–360.

Gracia M., Montané F., Piqué J., Retana J. (2007): Overstory structure and topographic gradient determining diversity and abundance of understory shrub species in temperature forest in central Pyrenees (NE Spain). Forest Ecology and Management, 242: 391–397.

Guoyu L. (2011): Topography related spatial distribution of dominant tree species in a tropical seasonal rain forest in China. Forest Ecology and Management, 262: 1507–1513.

Ismail R., Mutango O. (2010): Comparison of regression tree ensembles: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forest of KwaZulu-Natal, South Africa. International Journal of Applied Earth Observation and Geoformation, 12: 45–51.

Ismailzadeh A., Hosseini M. (2007): Relationship between ecological groups of plants with biodiversity indicators of plants in Afratakhteh cache for *Taxus bacata*. Journal of Ecology, 43: 21–30. (in Persian)

Kardgar N. (2012): Accuracy assessment of soil maps in Dr. Bahramnia forestry plan. [MSc Thesis.] Gorgan, Gorgan University of Agricultural Sciences and Natural Resources: 150. (in Persian)

Kernes B.K., Ohmann J.L. (2004): Evaluation and prediction of shrub cover in coastal Oregon forests (USA). Catena, 55: 341–365.

Kint V., van Meirvenne M., Nachtergale L., Geudens G., Lust N. (2003): Spatial methods for quantifying forest stand structures development: A comparison between nearest neighbour indices and variogram analysis. Forest Science, 49: 36–49.

Krebs C.J. (1999): Ecological Methodology. 2nd Ed. Menlo Park, Addison-Welsey Educational Publishers, Inc.: 620.

Kymasi F. (2012): Spatial distribution of tree and shrub species diversity in forests in Golestan province using GIS. [MSc Thesis.] Gorgan, Gorgan University of Agricultural Sciences and Natural Resources: 170. (in Persian)

Ludwing J.A., Reynolds J.F. (1988): Statistical Ecology: A Primer on Methods and Computing. New York, John Wiley & Sons: 202.

Maguran A.E. (1996): Ecological Diversity and Its Measurement. Princeton, Chapman & Hall: 179.

Marvie Mohajer M.R. (2006): Silviculture. Tehran, Tehran University Press: 387. (in Persian)

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 Ecology and Management, 175: 497–507.

Mehdinya T., Ejtehadi H., Sepehri A. (2006): Physiographic variables and the correlation between rainfall and vegetation communities present in the watershed of the Babol, Mazandaran province using geographic information systems. Journal of Agricultural Sciences and Natural Resources, 13: 99–107. (in Persian)

Momeni Moghaddam T., Sagheb Talebi K.H., Akbarinia M., Akhavan M., Hosseini S.M. (2012): Impact of physiographic and edaphic factors on some of qualitative and quantitative characteristics of Juniperus trees. Case study: Layn region – Khorasan. Iranian Journal of Forest, 4: 143–156. (in Persian)

Moore I.D., Gessler P.E., Nielsen G.A., Petersen G.A. (1993): Terrain attributes: Estimation methods and scale effects. In: Jakeman A.J., Beck M.B., McAleer M. (eds): Modelling Change in Environmental Systems. London, John Wiley & Sons: 189–214.

O'Sullivan S., Keady E., Keane S., Irwin O'Halloran J. (2010): Data mining for biodiversity prediction in forests. In: Coelho H., Studer R., Wooldridge M. (eds): Proceedings of the 19th European Conference on Artificial Intelligence, Lisbon, Aug 16–20, 2010: 289–294.

Parmentier I. (2011): Predicting alpha diversity of African rainforests: Models based on climate and satellite-derived data do not perform better than a purely spatial model. Journal of Biogeography, 38: 1164–1176.

Peffer K., Pebesma E.J., Burrough P.A. (2003): Mapping alpine vegetation using vegetation observation and topographic attributes. Landscape Ecology, 18: 759–776.

Pourbabaee H. (1998): Biodiversity of wooden plants in the forests of Gillan. [Ph.D. Thesis.] Tehran, Tarbiat Modares University: 264. (in Persian)

Qomioghli A.S., Hosseini M., Mataji A., Jalali G.H. (2006): Biodiversity of wooden plants on different soil in two different plant communities. Journal of Biology, 20: 200–207. (in Persian)

Saatchi S., Buermann W., ter Steege H., Mori S.A., Smith T.B. (2008): Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sensing of Environment, 112: 2000–2017.

Shannon C.E., Weaver W. (1949): The Mathematical Theory of Communication. Urbana, University of Illinois Press: 163.

Shataee S.H., Kalb S., Fallah A., Pelz D. (2012): Forest attribute imputation using machine-learning methods and ASTER data: Comparison of *k*-NN, SVR and random forest regression algorithms. International Journal of Remote Sensing, 33: 6254–6280.

Shirzad M.A., Tabari M. (2011): Effect of some environmental factors on diversity of woody plants in *Junipeus excelsa* habitat of Hezarmasjed mountains. Iranian Journal of Biology, 24: 800–808. (in Persian)

Simpson E.H. (1949): Measurement of diversity. Nature, 163: 688.

Smith B., Bastow Wilson J. (1996): A consumer's guide to evenness indices. Oikos, 76: 70–82.

Stephenson C.M., MacKenzie M.L., Edwards C., Travis J.M.J. (2006): Modeling establishment probabilities of an exotic plant, *Rhododendron ponticum*, invading a heterogeneous woodland landscape using logistic regression with spatial autocorrelation Ecological Modeling, 193: 747–758.

van der Maarel E. (2005): Vegetation Ecology. London, Blackwell Publishing: 273.

Whittaker R.H. (1977): Evolution of species diversity in land communities. Evolutionary Biology, 10: 1–67.

Wilson J.P., Gallant J.C. (2000): Terrain Analysis: Principles and Applications. New York, John Wiley & Sons: 520.

Yazdani S. (2011): Quantitative estimation of forest characteristics using QuickBird images. [MSc Thesis.] Gorgan, Gorgan University of Agricultural Sciences and Natural Resources: 129. (in Persian)

Zimmermann N.E., Kienast F. (1999): Predictive mapping of alpine grassland in Switzerland: Species versus community approach. International Journal of Vegetable Science, 10: 469–482.

Received for publication August 1, 2016
Accepted after corrections October 18, 2016

Corresponding author:

Ing. Azadeh Abdollahnejad, Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences, Department of Forest Management, Kamýcká 1176, 165 21 Prague 6-Suchdol, Czech Republic;
e-mail: abdolahnejad@fld.czu.cz