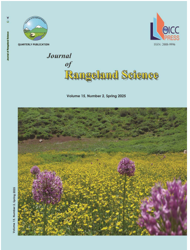

Volume 15, Issue 2, 152512 (1-10)

Journal of Rangeland Science (JRS)

https://doi.org/10.57647/j.jrs.2025.1502.12

Identification of suitable areas for the growth of *Teucrium polium* species using machine learning models (a case study of Khalil Abad and Kashmar Counties, Iran)

Javad Momeni Damaneh[1], Jalil Ahmadi[2,*], Ali Akbar Safdari[2], Zahra Jafar pour Chekab[3], Sajad Shams Beyranvand[3]

[1]*Department Natural Resources, Faculty of Agriculture and Natural Resources, University of Hormozgan, Bandar Abbas, Iran.*
[2]*Department Rehabilitation of Arid and Mountainous Regions, University of Tehran, Tehran, Iran.*
[3]*Department Range and Watershed Management Ferdowsi, University of Mashhad, Mashhad, Iran.*

[*]**Corresponding author:** jalilahmadi@ut.ac.ir

Original Research

Received:
10 July 2023
Revised:
1 May 2024
Accepted:
4 June 2024
Published online:
20 April 2025

Abstract:

With the advancement in computer technologies, the prediction of ecological niches for various plant species has become possible. The impact of climate change on the distribution of plants could be investigated using species distribution models. This study aimed to identify the climatic and environmental factors influencing the distribution of *Teucrium polium* species and determine its geographic range in Kashmar and Khalilabad counties, Khorasan Razavi province, Iran. To achieve this, 75 bioclimatic variables encompassing soil, topography, climate, and geology factors were initially analyzed for correlation, and variables with correlation coefficients higher than 0.80 were eliminated. The data of 57 GPS-recorded presence points were collected from two areas during 2021 – 2022. The environmental data and presence points were processed and predicted using the BIOMOD2 package within the R software, which encompasses 11 models. The models were evaluated using Cohen's kappa coefficient (KAPPA), True Skill Statistic (TSS), and Receiver Operating Characteristic (ROC) indices. Model accuracy assessment revealed that the Random Forest (RF) model achieved 99.7% accuracy while the Ensemble model achieved 99.4% accuracy, indicating excellent modeling precision. The relative importance of different variables in the selected models were as follows: in the RF model, silt at a depth of 30 – 60 and 15 – 30 cm, topographic humidity index, annual mean temperature, and daily temperature range; and for Ensemble model, nitrogen at a depth of 15 – 30 cm, topographic humidity index, soil bulk density at a depth of 5 – 15 cm, silt at a depth of 30 – 60 cm, and nitrogen at a depth of 15 – 5 cm, highlighting the influence of soil factors on the species distribution. The obtained results could be utilized for the conservation, management, and expansion of *Teucrium polium* habitats in similar areas.

Keywords: Habitat suitability; *Teucrium polium*; Soil properties; Machine Learning; BIOMOD2

Introduction

Teucrium polium, known as Kalpoureh here in Iran, is an important medicinal plant that grows in some regions worldwide (Koocheki et al., 2008). It belongs to the Lamiaceae family and is an evergreen herbaceous plant. *T. polium* is predominantly found in temperate and tropical regions of Southern Europe, North Africa, and Western Asia (El Atki et al., 2019). It has tall woody stems and dark rectangular leaves; it has beautiful pink and light purple flowers arranged in clustered spikes. In traditional medicine, *T. polium* is known for its therapeutic properties in treating digestive disorders, rheumatism as an anti-inflammatory and antimicrobial agent (Falah-Hosseini et al., 2004; Cozzani et al., 2005; Majdoub et al., 2022).

T. polium is an herbaceous plant that grows in warm and dry temperate regions. It grows in areas with temperatures ranging from 10 to 30 ^{o}C and at altitudes below 2500 m above sea level. *T. polium* is commonly found in sunny locations, as well as areas with sandy soil and moist, fertile soils. Its adaptable nature allows it to grow as an herbaceous and mountainous plant in areas with sparse vegetation, plains, road edges, and even on rocks (Mozaffarian, 2012; Salmaki et al., 2016).

Considering the ecological importance and the protective role of *T. polium* in soil conservation and medicinal uses, identifying suitable areas for its cultivation and growth could optimize its cultivation, enhance its productivity, and contribute to its conservation efforts. In this regard, the utilization of machine learning models has gained attention as a powerful tool for predicting suitable areas for the cultivation and growth of *T. polium.* These models can easily determine favorable habitats for its growth, aiding future management planning for the preservation and development of this species. Machine learning models have been extensively used in various studies, highlighting their effectiveness in predicting species distribution. Despite differences in scale, scope, and methodology, these models have proven to be valuable and cost-effective tools for natural resource managers, enhancing their awareness and decision-making capabilities regarding the impact of climate change on different species.

Nowadays, with the development of remote sensing systems and various prediction models, it is possible to predict the ecological niches of different plant and animal species. Habitat prediction using data collection from natural habitats and various environmental information is feasible (Kargar et al., 2018; Guo et al., 2017; Rahmanian et al., 2022). Some predictions have been examined through species distribution models (**Guisan2003**; Zare Chahuki and Abbasi, 2018). In the past decades, changes in humans' lifestyles and industrialization of production processes have led to increased air pollution and severe climate changes. (Barnes and Harrison, 1982). These changes not only had an impact on human health and life, but also posed serious threats to natural ecosystems and biodiversity.

Some factors have destructive impact on plant diversity including increased temperature, and alterations in precipitation patterns including shifts in the temporal and spatial distribution of rainfall. However, the effects of these changes are not uniform across the globe. Arid and semi-arid regions are particularly vulnerable to climate change, and the plant species within these regions face significant challenges for survival under such stress conditions (Bellard et al., 2012; Diaz-Varela et al., 2010; Ernakovich et al., 2014; Feeley et al., 2011; Sproull et al., 2015). Machine learning models that can process large volumes of biological data and detect complex patterns within this data could be utilized for predicting the habitat suitability of plants (Chen and Jiang, 2021). These models combine machine learning algorithms and statistical analysis to model both quantitative and qualitative information related to factors influencing habitat suitability for a specific species, producing accurate predictions. Therefore, these models can serve as powerful tools for sustainable management of natural resources and biodiversity conservation (Elith et al., 2008; Moradi et al., 2019).

These models utilize various algorithms to predict the occurrence of a specific event. The differences between algorithms in machine learning models have always posed a challenge in selecting the best algorithm for a particular region (Renner and Warton, 2013). To overcome these challenges, the available algorithms in the BIOMOD package can be employed, which provides a suitable platform for examining the geographical distribution of plant species (Elith et al., 2008). BIOMOD 2 is a popular and practical package in the field of statistics and bioinformatics, designed for the R programming language. This package includes various machine-learning algorithms for processing biological data. Some of the algorithms available in the BIOMOD 2 package include the Generalized Linear Model (GLM), Generalized Boosting Method (GBM), Generalized Additive Model (GAM), Classification Tree Analysis (CTA), Artificial Neural Network (ANN), Surface Range Envelope (SRE), Flexible Denotative Analysis (FDA), Multivariate Adaptive Regression Spline (MARS), Random Forest (RF), Maximum entropy model (MaxEnt), and Eensembles (ESMs). Additionally, the package provides useful functions and tools such as statistical functions, data visualization, data preprocessing, and feature selection (Zarkami et al., 2021; Abdollahi and Naderi, 2012; Phillips et al., 2006).

Materials and methods

Study area

The present study was conducted in Khalilabad and Kashmar counties, Khorasan Razavi province, Iran, with a total area of 2,271 km^2. The geographical coordinates of the study area are (34$^{\mathrm{o}}$ 54′19″ to 35$^{\mathrm{o}}$23′54″ N latitude and 57$^{\mathrm{o}}$58′20″ to 58$^{\mathrm{o}}$47′59″ E longitude) (figure 1). The study area is located in the Irano-Turanian phytogeographic region, and it exhibits a wide range of ecological characteristics. Based on climatic classification, this area is placed in the cold semi-arid class. The annual average precipitation was 209.8 mm in the period from 1989 to 2012. However, precipitation wasn′t evenly distributed across the province. The rainfall declined from north to south. Recorded data showed a range of 116.2 to 312.8 mm for annual precipitation. The annual average temperature of the province ranged from 12.2 $^{\mathrm{o}}$C, to 18.2 $^{\mathrm{o}}$C during the same period, with an average annual temperature of 15.6 $^{\mathrm{o}}$C (Damaneh et al., 2022).

Determining presence points

Presence data for the study areas were obtained using 1:25,000 topographic maps, which included features such as roads, ridgelines, and rivers. After visiting the field and surveying the region, useful and exploitable habitat areas were identified. These areas were defined as locations that in addition to being dominated by the target species, covered a minimum area of 100 m^2. Subsequently, presence points were sampled during field visits conducted between 2021 and 2022. Circular plots with a radius of 5.64 m (covering an area of 100 m^2) were used for sampling. A total of 57 presence points were recorded using a GPS from two different regions (Fig. 1). To prevent spatial autocorrelation and reduce sampling errors, the useful areas were converted into 1000 × 1000 m grids using ArcGIS 10.8 software (Diaz-Varela et al., 2010). One presence point was obtained from each grid cell, and these points were then used in the habitat

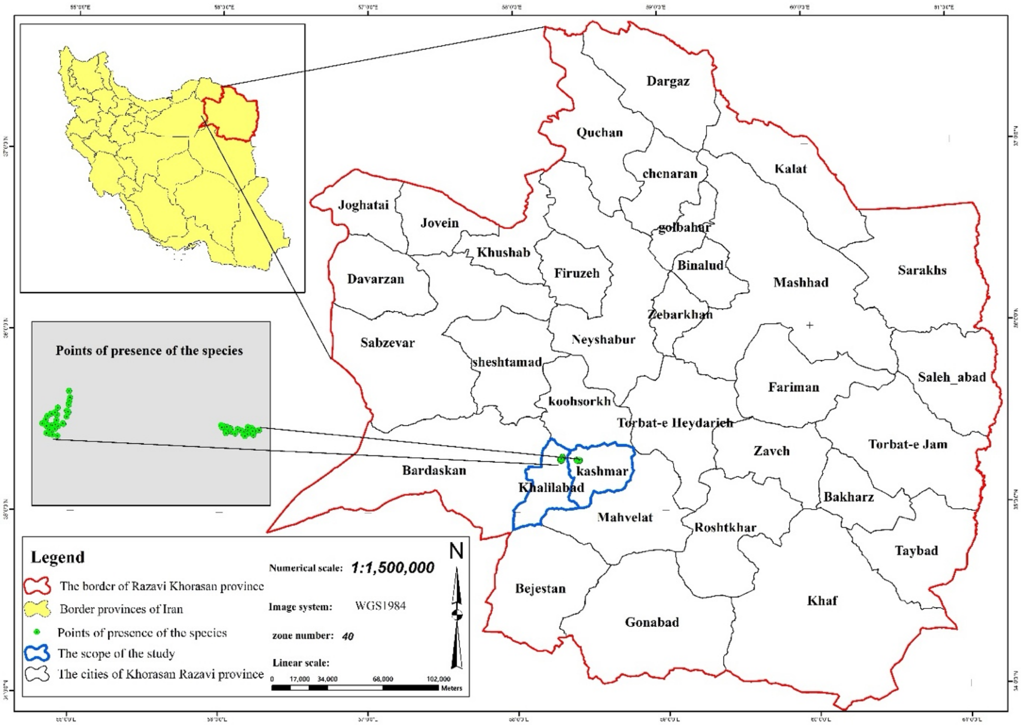

Figure 1. The map of location of the study area and the distribution of species presence points in Khorasan Razavi Province, Iran.

suitability modeling.

Determining environmental variables

By reviewing previous studies and considering the target species and available baseline information within the study area, a total of 75 important and influential variables affecting the distribution of the *T. polium* species were identified. Layers of these variables were prepared from various sources. Using the available information, 75 environmental variables were considered, including 10 physiographic variables, 19 climatic variables, 45 soil variables, and one geological variable for the modeling process (Table 1). Physiographic and geomorphological variables were obtained from the U.S. Geological Survey website (www.earthexplorer.usgs.gov). Climatic variables were obtained from the WorldClim (www.worldclim.org) database at a resolution of 2.5 minutes, which interpolated weather data from 1950 to 2000. Soil variables were obtained from the Soil Grids (www.soilgrids.org) database at a resolution of 250 m. Geological maps at a scale of 1:250,000 from the Geological Survey of Iran were used. To ensure that all input layers had the same coordinate system, scale, and extent, data preparation and initial processing of the layers were performed using Idrisi Selva software. The information layers in Table 1 were prepared, and the layers were standardized to a pixel size of 1000×1000 m in Idrisi Selva by applying a Pearson correlation coefficient threshold of 0.80. Variables with correlation coefficients below 0.80 Momeni Damaneh et al. (2022b) were selected (Fig. 2). Ultimately, 27 habitat variables were selected as predictor variables (Table 1) and called in the R software along with the presence points of the target species for modeling. The BIOMOD 2 package including GLM, GBM, GAM, CTA, ANN, SRE, FDA, MARS, RF, MaxEnt, and ESMs models was used to determine the suitable areas for the *T. polium* species in the present time. The accuracy of the models was evaluated using Kappa, TSS (True Skill Statistic), and ROC (Receiver Operating Characteristic) indices, which are commonly used indices for determining and identifying areas of equal potential. It should be noted that the preparation of information layers was performed using Idrisi Selva software (Table 1).

Species distribution modeling

In the present study, ten algorithms available in the BIOMOD software package Thuiller et al. (2009) and Phillips et al. (2006) were used for modeling the distribution of *T. polium* species (Table. 2). Additionally, the BIOMOD software was utilized to generate the points of absence. The points of absence were randomly generated, with 300 points of absence and an equal number of presence points (57 points). In the modeling process, 70% of the presence points were used to build the models while the remaining 30% were used for model evaluation. To enhance modeling accuracy, the process was repeated five times.

Model evaluation

The accuracy of the models was assessed using three statistical coefficients. The first method employed was the evaluation of the Receiver Operating Characteristic (ROC) curve. The ROC curve is a graphical representation of a model′s ability to predict species presence and absence based on relevant environmental variables (Fielding and Bell, 1997). The second method involved calculating the True Skill Statistic (TSS), which is applicable when presence-absence models are used. TSS can be considered as evidence for interpreting

Table 1. Environmental variables used in *T. polium* habitat modeling.

Category	Variable name and Description	Abbreviation	Units	*T. polium*
Bioclimatic	The monthly average temperature	BIO 1	(°C)	Accept
	Average monthly (max temp – min temp)	BIO 2	(°C)	Reject
	Isothermality = (BIO2/BIO7) (×100)	BIO 3	Percent	Reject
	Mean annual temp. variation calculated from SD of monthly temp.×100	BIO 4	(°C)	Accept
	The highest monthly temperature that has been recorded in a certain year	BIO 5	(°C)	Reject
	The occurrence of the lowest monthly temperature in a given year	BIO 6	(°C)	Reject
	Temperature variation over a given period (BIO5-BIO6)	BIO 7	(°C)	Reject
	The average temperatures experienced during the wettest quarter	BIO 8	(°C)	Reject
	The average temperatures experienced during the driest quarter	BIO 9	(°C)	Reject
	The average temperatures experienced during the hottest quarter	BIO 10	(°C)	Reject
	The average temperatures in the coldest quarter	BIO 11	(°C)	Reject
	This is the sum of all total monthly precipitation values	BIO 12	mm	Accept
	The total amount of precipitation experienced in the wettest month	BIO 13	mm	Reject
	The total amount of precipitation experienced in the driest month	BIO 14	mm	Reject
	The monthly total precipitation SD from the monthly total precipitation	BIO 15	Percent	Reject
	The total amount of precipitation experienced during the wettest quarter	BIO 16	mm	Reject
	The overall amount of precipitation experienced during the driest quarter	BIO 17	mm	Reject
	The total amount of precipitation that falls during the hottest	BIO 18	mm	Reject
	The total amount of precipitation experienced during the coldest quarter	BIO 19	mm	Reject
Edaphic				
Physical properties	Bulk density in depth 0-5 cm	Bulk 0-5 cm	cg/cm^3	Accept
	Bulk density in depth 5-15 cm	Bulk 5-15 cm	cg/cm^3	Accept
	Bulk density in depth 15-30 cm	Bulk 15-30 cm	cg/cm^3	Accept
	Bulk density in depth 30-60 cm	Bulk 30-60 cm	cg/cm^3	Accept
	Bulk density in depth 60-100 cm	Bulk 60-100 cm	cg/cm^3	Accept
	Sand in depth 0-5 cm	Sand 0-5 cm	g/kg	Reject
	Sand in depth 5-15 cm	Sand 5-15 cm	g/kg	Reject
	Sand in depth 15-30 cm	Sand 15-30 cm	g/kg	Accept
	Sand in depth 30-60 cm	Sand 30-60 cm	g/kg	Reject
	Sand in depth 30-60 cm	Sand 60-100 cm	g/kg	Reject
	Silt in depth 0-5 cm	Silt 0-5 cm	g/kg	Reject
	Silt in depth 5-15 cm	Silt 5-15 cm	g/kg	Reject
	Silt in depth 15-30 cm	Silt 15-30 cm	g/kg	Accept
	Silt in depth 30-60 cm	Silt 30-560 cm	g/kg	Accept
	Silt in depth 60-100 cm	Silt 60-100 cm	g/kg	Reject
	Clay content in depth 0-5 cm	Clay 0-5 cm	g/kg	Accept
	Clay content in depth 5-15 cm	Clay 5-15 cm	g/kg	Reject
	Clay content in depth 15-30 cm	Clay 15-30 cm	g/kg	Reject
	Clay content in depth 30-60 cm	Clay 30-60 cm	g/kg	Reject
	Clay content in depth 60-100 cm	Clay 60-100 cm	g/kg	Reject
	Coarse fragments in depth 0-5 cm	Coarse 0-5 cm	cm^3/dm^3	Accept
	Coarse fragments in depth 5-15 cm	Coarse 5-15 cm	cm^3/dm^3	Reject
	Coarse fragments in depth 15-30 cm	Coarse 15-30 cm	cm^3/dm^3	Accept
	Coarse fragments in depth 30-60 cm	Coarse 30-60 cm	cm^3/dm^3	Reject
	Coarse fragments in depth 60-100 cm	Coarse 60-100 cm	cm^3/dm^3	Reject
Edaphic				
Chemical	Nitrogen in depth 0-5 cm	N 0-5 cm	cg/kg	Reject

Continued of Table 1.

Category	Variable name and Description	Abbreviation	Units	*T. polium*
properties	Nitrogen in depth 5-15 cm	N 5-15 cm	cg/kg	Accept
	Nitrogen in depth 15-30 cm	N 15-30 cm	cg/kg	Accept
	Nitrogen in depth 30-60 cm	N 30-60 cm	cg/kg	Accept
	Nitrogen in depth 60-100 cm	N 60-100 cm	cg/kg	Reject
	Soil organic carbon in depth 0-5 cm	OC 0-5 cm	dg/kg	Accept
	Soil organic carbon in depth 5-15 cm	OC 5-15 cm	dg/kg	Accept
	Soil organic carbon in depth 15-30 cm	OC 15-30 cm	dg/kg	Accept
	Soil organic carbon in depth 30-60 cm	OC 30-60 cm	dg/kg	Accept
	Soil organic carbon in depth 60-100 cm	OC 60-100 cm	dg/kg	Reject
	Soil pH water in depth 0-5 cm	pHw 0-5 cm	pH $\times$10	Accept
	Soil pH water in depth 5-15 cm	pHw 5-15 cm	pH $\times$10	Reject
	Soil pH water in depth 15-30 cm	pHw 15-30 cm	pH $\times$10	Reject
	Soil pH water in depth 30-60 cm	pHw 30-60 cm	pH $\times$10	Reject
	Soil pH water in depth 60-100 cm	pHw 60-100 cm	pH $\times$10	Reject
	Cation exchange capacity (at pH 7) in depth 0-5 cm	CEC 0-5 cm	mmol(c)/kg	Accept
	Cation exchange capacity (at pH 7) in depth 5-15 cm	CEC 5-15 cm	mmol(c)/kg	Accept
	Cation exchange capacity (at pH 7) in depth 15-30 cm	CEC 15-30 cm	mmol(c)/kg	Reject
	Cation exchange capacity (at pH 7) in depth 30-60 cm	CEC 30-60 cm	mmol(c)/kg	Reject
	Cation exchange capacity (at pH 7) in depth 60-100 cm	CEC 60-100 cm	mmol(c)/kg	Reject
Topographic				
Altitude	Altitude absl. (Obtained from optical sensors ASTER satellite, 90 m)	DEM	m	Accept
TWI	Topographic wetness index (Produced from DEM)	TWI	Unit less	Accept
TRI	Terrain Ruggedness Index	TRI	Unit less	Reject
Slope	Percent change in that elevation over a certain distance	Slope	Degree	Reject
Geology	Geology	Geology	Unit less	Accept

real ecological phenomena. Studies have shown a high correlation between ROC and TSS values. In studies where the results are presented as presence-absence maps, TSS can serve as a suitable alternative to ROC (Walther et al., 2002). The Kappa coefficient measures the agreement between two evaluators who have independently ranked N cases into C mutually exclusive categories. The first use of a statistic similar to Kappa is attributed to Galton and Smeeton (Galton, 1892; Smeeton, 1985). ROC, Kappa, and TSS values serve as indicators of model performance. Scores below 0.5 signify poor performance while those between 0.5 and 0.6 suggest very weak agreement. Values from 0.6 to 0.7 indicate weak agreement, 0.7 to 0.8 signify moderate agreement, and 0.8 to 0.9 point to good agreement. Scores ranging from 0.9 to 1 represent high, desirable agreement in modeling (Swets, 1988). ROC, Kappa, and TSS scores below 0.5 suggest poor model performance, while those between 0.5 and 0.6 indicate very weak agreement. Scores from 0.6 to 0.7 represent weak agreement, 0.7 to 0.8 suggest moderate agreement, and 0.8 to 0.9 indicate good agreement. A score between 0.9 and 1 signifies high, desirable agreement in modeling (Swets, 1988).
Furthermore, to gain a geographical perspective on the areas with suitable climatic and environmental conditions for the studied species, maps of potential habitats were created for the present and future scenarios (Fig. 3). The potential habitat maps were derived using habitat suitability models, which range from 0 to 1000, with 0 representing the lowest probability and 1000 representing the highest probability. To enhance the understanding of the distribution, the maps were classified into four categories using the Natural Breaks or Jenks algorithm in ArcGIS 10.8: Unsuitable habitats ranging from 0 to 250, habitats with low suitability ranging from 250 to 500, habitats with moderate suitability ranging from 500 to 750, and Unsuitable habitats ranging from 750 to 1000 (Table 4).

Results

The values of the Kappa, TSS, and ROC indices, which are prominent and widely used indices for determining and identifying areas of similar potential, are presented in Table 2. The modeling process in most of the employed models in this study resulted in high (desirable) and excellent levels of accuracy. The Random Forest (RF) model exhibited the highest accuracy with values of 91.6%, 99.7%, and 99.8% for the Kappa, TSS, and ROC parameters, respectively. Therefore, based on the results and values in Table 2, the Random Forest (RF) model is selected as the preferred model for all the examined scenarios of the *T.*

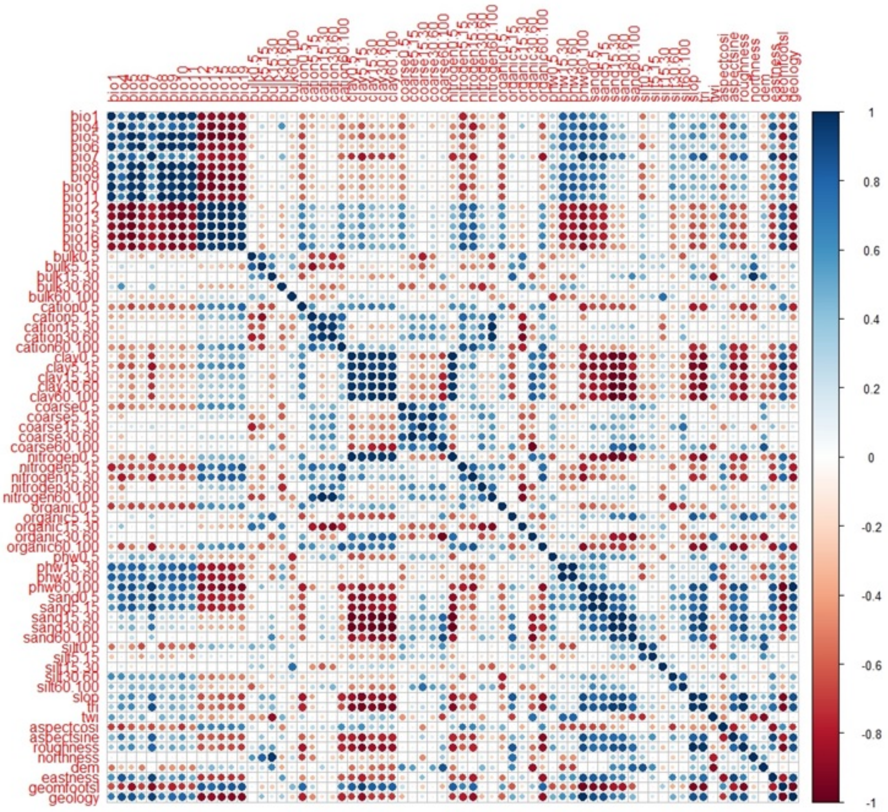

Figure 2. Pearson correlation test of 75 predictors. Negative correlations are shown in red, and positive correlations are shown in blue. The intensity of color and the size of the circle represent the magnitude of the correlation coefficients.

polium species, serving as the basis for subsequent calculations. The highest accuracy values and influential variables in the species distribution are specified in Tables 2 and 3. Overall, the ensemble model and the selected model will ultimately be used to present the results.

The percentage of the relative importance of environmental variables in modeling the distribution of suitable areas for *T. polium* indicates that the highest impact of environmental factors using the Random Forest (RF) model includes silt at a depth of 30 – 60 cm, silt at a depth of 15 – 30 cm, Topographic Wetness Index (TWI), mean annual temperature (BIO1), and mean diurnal temperature range (BIO2). In the ensemble model, the variables with the highest impact on the distribution of *T. polium* are nitrogen at a depth of 15 – 30 cm, Topographic Wetness Index (TWI), bulk density at a depth of 5 – 15 cm, silt at a depth of 30 – 60 cm, and nitrogen at a depth of 15 – 5 cm (Fig.3).

The assessment of the relative importance of all environmen-

Table 2. Accuracy evaluation in modeling suitable locations for *T. polium* using three Current time modeling.

Accuracy parameter	Abbreviation	Current time modeling		
		KAPPA	TSS	ROC
Artificial Neural Network	ANN	0.803	0.987	0.993
Classification Tree Analysis	CTA	0.891	0.993	0.996
Techniques and their ensembles	ESMs	0.909	0.994	0.998
Flexible Denotative Analysis	FDA	0.916	0.995	0.998
Generalized Boosting Method	GBM	0.916	0.995	0.998
Generalized Liner Model	GLM	0.916	0.995	0.997
Multivariate Adaptive Regression Spline	MARS	0.916	0.995	0.997
Maximum entropy model	MAXENT	0.891	0.993	0.996
Random Forest	RF	0.916	0.995	0.998
Surface Range Envelope	SRE	0.912	0.993	0.997
Generalized Additive Model	GAM	0.861	0.987	0.993

KAPPA = Cohen's kappa coefficient, TSS = True Skill Statistic, ROC = Receiver Operating Characteristic.

Table 3. Relative importance percentage of environmental variables in two modeling of Random Forest (RF) and Techniques and their ensembles (Esms) suitable locations for *T. polium*.

Environmental/ soil variables used in the models	RF	Esms
Bio 1 (Monthly average temperature)	0.599	0.234
Bio 4 (Seasonal temperature (SD × 100))	0.535	0.1558
Bio 12 (Monthly precipitation)	0.566	0.1799
Bulk density in 0-5 cm depth	0.292	0.1026
Bulk density in 5-15 cm depth	0.525	0.3263
Bulk density in 15-30 cm depth	0.231	0.0759
Bulk density in 30-60 cm depth	0.349	0.0805
Bulk density in 60-100 cm depth	0.247	0.0671
Cation exchange capacity (CEC) in 0-5 cm depth	0.236	0.0437
Cation exchange capacity (CEC) in 5-15 cm depth	0.229	0.0366
Sand in 15-30 cm depth	0.562	0.1443
Silt in 15-30 cm depth	0.691	0.2971
Silt in 30-60 cm depth	0.713	0.3238
Clay in 0-5 cm depth	0.297	0.2438
Coarse fragments in 0-5 cm depth	0.458	0.1096
Coarse fragments in 15-30 cm depth	0.439	0.0653
Nitrogen in 5-15 cm depth	0.307	0.3203
Nitrogen in 15-30 cm depth	0.279	0.3623
Nitrogen in 30-60 cm depth	0.229	0.1056
Soil organic carbon in 0-5 cm depth	0.465	0.1275
Soil organic carbon in 5-15 cm depth	0.258	0.0672
Soil organic carbon in 15-30 cm depth	0.293	0.0624
Soil organic carbon in 30-60 cm depth	0.105	0.0606
Soil pH in 0-5 cm depth	0.065	0.0254
Topographic wetness index (TWI)	0.641	0.3435
Dem	0.543	0.1628
Geology	0.528	0.296

tal parameters indicated that soil factors have significant importance compared to topographic and climatic factors in the distribution of *T. polium* in the Kashmar and Khalil Abad regions (figure 3). According to Table 4, the results of this section showed that in the RF model with an area of 1713 km^2, equivalent to 62.4%, and based on the ensemble model with an area of 11111 km^2, equivalent to 72.5%, the studied areas had a moderate to high potential for the growth and exploitation of *T. polium*. Overall, by examining the outputs of the selected Random Forest and ensemble models and comparing the suitable areas for *T. polium*, the highest distribution suitable for the utilization of this species is scattered in the northwestern to northeastern regions, including the central parts of Khalilabad, which include the outskirts district, and the central part of Kashmar, which includes the Paein Valayat district and the Farah Dasht Kashmar district, which includes the Razagh Abad and Qaleh Bala districts. It can be stated that the habitat has shown a consistent expansion (figure 3). Table 3 shows the relative importance ratios of the most important environmental input variables in the Random Forest and ensemble models.

Discussion

According to the results of the current study, the machine learning models demonstrated high performance in predicting potential habitats for *T. polium* based on the accuracy metrics of KAPPA, TSS, and ROC. These models can effectively capture the environmental characteristics of its habitat. Considering that machine learning models are widely used in species distribution studies, this outcome was expected. However, given the specific characteristics of each plant species and the lack of research on *T. polium* distribution, an investigation of its distribution was necessary. The results indicate that in the Random Forest (RF) model, factors such as silt at depths of $30-60$ cm and $15-30$ cm, topographic wetness index (TWI), mean annual temperature

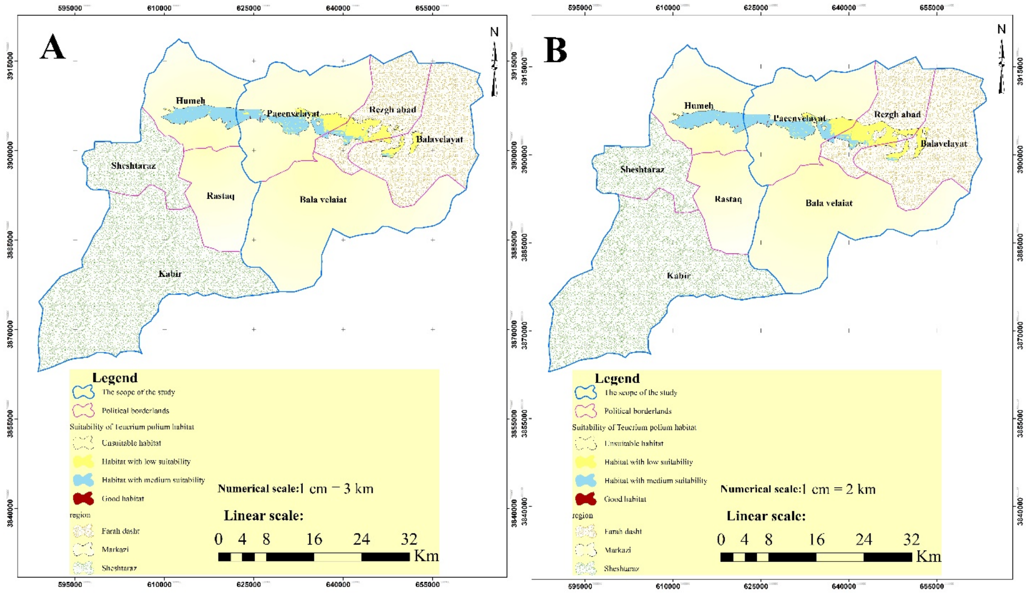

Figure 3. Habitat Suitability Map for *T. polium* under current conditions using the Random Forest (A) and Ensemble (B) model.

(BIO1), and mean diurnal temperature range (BIO2) play a significant role in shaping the distribution patterns of plant species and various vegetation characteristics (distribution, diversity, density, presence, etc.). In the aggregated model, parameters such as nitrogen at depths of 15 – 30 cm, TWI, bulk density at depths of 5 – 15 cm, silt at a depth of 30 – 60 cm, and nitrogen at depths of 15 – 5 cm also have a major influence on regulating the distribution patterns of plant species and vegetation attributes. Furthermore, the accuracy assessment of the models revealed that the RF model performed exceptionally well with an accuracy of over 99.7%, making it the preferred model for this study. Previous studies by Poyan et al. (2022), Cheng et al. (2012), Haidarian Aghakhani et al. (2017), Momeni Damaneh et al. (2022a), Momeni Damaneh et al. (2022b), Damaneh et al. (2022), Momeni Damaneh et al. (2020), and MomeniDamaneh et al. (2021) and others have also highlighted the effectiveness of the Random Forest model in species prediction modeling. The evaluation of habitat suitability for *T. polium* in the output maps generated by the RF model indicates that the majority of the study area is unsuitable for the establishment of this species, with the highest concentration of suitable habitats forming a strip in the northern to northeastern regions of Kashmar and Khalil Abad counties.
According to the findings of the current research, the ability of machine learning models to predict potential habitats for *T. polium*, as evaluated by the accuracy metrics of KAPPA, TSS, and ROC, was assessed as very good. These models can effectively capture the environmental characteristics of its habitat. These results are consistent with the studies conducted by Haidarian Aghakhani et al. (2017), Pirisahragard and Pahlavan-Rad (2020), Momeni Damaneh et al. (2022a), Momeni Damaneh et al. (2022b), and Zarabi et al. (2017) and Damaneh et al. (2022) which identified climate, edaphic, and physiographic factors as the most influential factors in plant species distribution. Climate factors, particularly temperature and precipitation, are considered the most important factors in the distribution of plant species (Sarhangzadeh, 2020). However, in the study area of the present research, with an average precipitation of 170 mm and its uniform distribution in different locations, and considering that the region falls within dry and semi-arid climate, the ecological constraints related to temperature are not significant. Therefore, the distribution of plant species in this area is primarily affected by moisture-related factors, with precipitation and temperature playing a secondary and tertiary role, respectively. Researchers have found that the expansion of plants towards higher altitudes and their establishment in elevated areas is an example of species migration in response to climate change, creating more favorable conditions for survival (Momeni Damaneh et al., 2022b; Walther et al., 2002). Thuiller et al. (2009)

Table 4. Area and percentage of suitable regions for *T. polium* in modeling.

Habitat	RF		ESMs	
	The predicted area (km^2)	(%)	The predicted area (km^2)	(%)
Unsuitable	2166.45	95.38	2157.51	94.99
Low suitability	45.09	1.99	52.08	2.29
Moderate suitability	59.68	2.63	61.64	2.71
Suitable	0.12	0.01	0.12	0.01

also identified the limited distribution range shift of plant species as one of the most significant impacts of climate change and rising temperatures. It should be noted that according to Thuiller et al. (2009), ecosystem responses to climate change vary across different ecosystems, and each ecosystem needs to be examined using appropriate methods. As mentioned earlier, the importance of rainfall amount and its temporal and spatial distribution in dry and semi-arid climates for the survival of various animal and plant species and biodiversity is well recognized by ecologists although the study of this parameter may vary depending on the scale, objective, precision, time, cost, etc. Nonetheless, there are still many hidden aspects of its impact on vegetation cover that require further investigation. It should be noted that each plant species establishes a relationship with environmental characteristics based on its habitat preferences, ecological needs, and tolerance range (Abdollahi et al., 2012; Kumari et al., 2021). Therefore, caution should be exercised in generalizing the obtained results to similar regions, taking into account regional-specific features and ecological requirements.

Conclusion

Human destructive activities such as overgrazing and unsustainable exploitation of *T. polium* along with climate change have put the current habitats of this species in Kashmar and Khalil Abad severely at risk. These unsustainable practices, without considering the environmental capacities in natural resource management, pose a significant challenge to the natural resources of Khorasan province and the country, leading to the degradation of water, soil, and vegetation in the region. Although this study only focused on examining the current climate and soil factors to delineate the potential areas of *T. polium* presence, to gain a deeper understanding and better knowledge for the restoration of damaged areas, conservation of endangered areas, and improvement of the predictive capability of ecological models for potential habitats of plant species, it is essential to consider other parameters (Barnes and Harrison, 1982). These parameters include human factors, various types of exploitation, grazing practices, wildlife, socioeconomic status, and many other factors that directly and indirectly influence the distribution of this species. Numerous studies have been conducted on various plant species, which despite differences in scale, extent, and methodology, introduce machine learning methods as effective approaches. In this research, the aim was to evaluate various machine learning models for species distribution, and the Random Forest model was selected as the most suitable model. Species distribution models serve as valuable and cost-effective tools for natural resource managers, enhancing their knowledge and decision-making abilities regarding the effects of climate change on species (Haidarian Aghakhani et al., 2017). The generated maps in the Random Forest model can identify sensitive and suitable areas for the growth of *T. polium* in the present time. This information can be useful in arid and semi-arid regions for conservation, development, and corrective strategies, assisting range and ecosystem managers in ensuring the future presence of *T. polium.*

Acknowledgment
We would like to express our utmost appreciation and gratitude to all the personnel of the Khorasan Razavi Department of Natural Resources and Watershed Management and the Natural Resources Departments of the respective counties for their assistance and support throughout the various stages of this research.

Authors contributions
Authors have contributed equally in preparing and writing the manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdollahi J., Arzani H., Naderi H., Arab Zade M. R. (2012) Effect of precipitation and high-temperature variability on forage production of some plant species in the Yazd steppe rangelands during the period of 2000-2008 (Case study: Ernan region). *Journal of Arid Biome* 2 (1): 58–69.

Abdollahi J., Naderi H. (2012) The effect of topographic variables and soil physicochemical properties on the performance of factors affecting the growth of Artemisia sieberi in the steppe rangelands of Nedushan, Yazd. *Journal of Watershed Management Research (Research and Development).* 25 (4): 62–52.

Barnes P. W., Harrison A. T. (1982) Species distribution and community organization in a Nebraska sandhill mixed prairie as influenced by plant/soil-water relationships. *Oecologia* 52:192–201.

Bellard C., Bertelsmeier C., Leadley P., Thuiller W., Courchamp F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters* 15 (4): 365–377.

Chen J., Jiang J. (2021) Application of machine learning models in plant habitat suitability prediction: A review. *Ecological Modeling* 438:109395.

Cheng Y. Y., Chan P. P., Qiu Z. W. (2012) Random forest-based ensemble system for short term load forecasting. *In 2012 international conference on machine learning and cybernetics IEEE.* 1:52–56.

Cozzani S., Muselli A., Desjobert J. M., Bernardini A. F., Tomi F., Casanova J. (2005) Chemical composition of essential oil of Teucrium polium subsp. capitatum (L.) from Corsica. *Flavour and Fragrance Journal* 20 (4): 436–441.

Damaneh J. M., Ahmadi J., Rahmanian S., Sadeghi S. M. M., Nasiri V., Borz S. A. (2022) Prediction of wild pistachio ecological niche using machine learning models. *Ecological Informatics* 72:101907.

Diaz-Varela R. A., Colombo R., Meroni M., Calvo-Iglesias M. S., Buffoni A., Tagliaferri A. (2010) Spatio-temporal analysis of alpine ecotones: A spatial explicit model targeting altitudinal vegetation shifts. *Ecological Modeling* 221 (4): 621–633.

El Atki Y., Aouam I., Taroq A., Lyoussi B., Taleb M., Abdellaoui A. (2019) Total phenolic and flavonoid contents and antioxidant activities of extracts from Teucrium polium growing wild in Morocco. *Materials Today: Proceedings* 13:777–783.

Elith J., Leathwick J.R., Hastie T. (2008) A working guide to boosted regression trees. *Journal of animal ecology* 77 (4): 802–813.

Ernakovich J. G., Hopping K. A., Berdanier A. B., Simpson R. T., Kachergis E. J., Steltzer H., Wallenstein M. D. (2014) Predicted responses of arctic and alpine ecosystems to altered seasonality under climate change. *Global Change Biology* 20 (10): 3256–3269.

Falah-Hosseini H., Larijani B., Fakhraee H., Rajabipour B., Akhoondzadeh S., Toolit T., Hashemzadeh R. (2004) Clinical trial on the effect of the extract of the seed of Silybum marianum (silibinin) on blood lipid in patients with type 2 diabetes with hyperlipidemia. *Iranian Journal of Diabetes and Metabolism* 3 (2): 201–206.

Feeley K. J., Silman M. R., Bush M. B., Farfan W., Cabrera K. G., Malhi Y., Meir P., Revilla N. S., Quisiyupanqui M. N. R., Saatchi S. (2011) Upslope migration of Andean trees. *Journal of Biogeography* 38 (4): 783–791.

Fielding A. H., Bell J. F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation* 24 (1): 38–49.

Galton F. (1892) Fingerprints. *Macmillan and Company*

Guo Y., Li X., Zhao Z., Wei H., Gao B., Gu W. (2017) Prediction of the potential geographic distribution of the ectomycorrhizal mushroom Tricholoma matsutake under multiple climate change scenarios. *Scientific reports* 7 (1): 46221.

Haidarian Aghakhani M., Tamartash R., Jafarian Z., Tarkesh Esfahani M., Tatian M. R. (2017) Forecasts of climate change effects on Amygdalus scoparia potential distribution by using ensemble modeling in Central Zagros *Journal of RS and GIS for Natural Resources* 8 (3): 1–14.

Kargar M., Jafarian Z., Tamartash R., Alavi S. J. (2018) Comparison of non-parametric and parametric species distribution models (SDM) in determining the habitat of dominant rangeland species (Case study: Khetteh Riz Rangelands) 25 (3): 512–523.

Koocheki A., Nasiri Mahalati M., Azizi G., Khazaee H. (2008) Study of agroecological requirements of Teucrium polium L. plants. *Iranian Journal of Crop Sciences* 6 (2): 395–404.

Kumari N., Srivastava A., Dumka U. C. (2021) A long-term spatiotemporal analysis of vegetation greenness over the Himalayan Region using Google Earth Engine. *Climate* 9 (7): 109.

Majdoub S., Dall'Acqua S., El Mokni R., Hammami S., Peron G. (2022) Chemical composition and antioxidant activity of essential oil from the aerial parts of Teucrium luteum (Mill.) Degen subsp. flavovirens (Batt.) Greuter & Burdet growing wild in Tunisia. *Applied Sciences* 12 (15): 7370.

Momeni Damaneh J., Esmaeili Pour Y., Gholami H., Farashi A. (2020) Application of maximum entropy model in determining the potential habitat of Astracantha gossypina (Fisch.) Podlech in northeastern Iran. *Plant Ecosystem Conservation* 19:236–217.

Momeni Damaneh J., Esmaeilpour Y., Gholami H., Farrashi A. (2022a) Predicting the geographical distribution of the genus Ferula (Ferula spp.) using habitat suitability modeling (Case study: Razavi and North Khorasan Provinces)' *Ecosystem Management* 2 (1): 25–35.

Momeni Damaneh J., Tajbakhsh S. M., Ahmadi J., Safdari A. (2022b) Comparison of species distribution models in determining the habitat suitability of wild pistachio (Pistacia vera L.) in Khorasan Razavi province. *Water and Soil Management and Modeling* 3 (4): 77–92.

MomeniDamaneh J., Esmaeilpour Y., Gholami H., Farashi A. (2021) Properly predict the growth of (Ferula assa-foetida L.) in northeastern Iran using the maximum entropy model'. *Iranian Journal of Range and Desert Research* 28 (3): 578–592.

Moradi H. R., Vazirizadeh A., Rafiee M. (2019) Habitat suitability modeling for plant species using machine learning algorithms. *Ecological Indicators* 101:440–448.

Mozaffarian V. (2012) Medicinal and Aromatic Plants of Iran. Tehran: *Farhang Moaser Publishers*

Phillips S. J., Anderson R. P., Schapire R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological modeling* 190 (3-4): 231–259.

Pirisahragard H., Pahlavan-Rad M. R. (2020) Prediction of soil properties using random forest with sparse data in a semi-active volcanic mountain. *Eurasian Soil Science* 53:1222–1233.

Poyan S., Rahmanian S., Amindin A., Pourghasemi H. R. (2022) Spatial and seasonal modeling of the land surface temperature using random forest. *In Computers in Earth and Environmental Sciences,* 221–234.

Rahmanian S., Pouyan S., Karami S., Pourghasemi H. R. (2022) Predictive habitat suitability models for Teucrium polium L. using boosted regression trees. *In Computers in Earth and Environmental Sciences,* 245–254.

Renner I. W., Warton D. I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics* 69 (1): 274–281.

Salmaki Y., Kattari S., Heubl G., Brauchler C. (2016) Phylogeny of non-monophyletic Teucrium (Lamiaceae: Ajugoideae): Implications for character evolution and taxonomy. *Taxon* 65 (4): 805–822.

Sarhangzadeh J. (2020) Modeling habitat suitability of Juniperus foetidissima Willd. in the Arasbaran Biosphere Reserve. *Forest Research and Development* 5 (1): 93–112.

Smeeton N. C. (1985) Early history of the kappa statistic. *Biometrics* 41:795.

Sproull G. J., Quigley M. F., Sher A., Gonzalez E. (2015) Long-term changes in composition, diversity and distribution patterns in four herbaceous plant communities along an elevational gradient. *Journal of Vegetation Science* 26 (3): 552–563.

Swets J. A. (1988) Measuring the accuracy of diagnostic systems. *Science* 240 (4857): 1285–1293.

Thuiller W., Lafourcade B., Engler R., Araujo M. B. (2009) BIOMOD–a platform for ensemble forecasting of species distributions. *Ecography* 32 (3): 369–373.

Walther G. R., Post E., Convey P., Menzel A., Parmesan C., Beebee T. J., Fromentin J. M., Hoegh-Guldberg O., Bairlein F. (2002) Ecological responses to recent climate change. *Nature* 416 (6879): 389–395.

Zarabi M., Haghdoosti R., Yousefi H. (2017) Modeling the habitat suitability of organic wild pistachio (Pistacia vera) using MaxEnt method in the forest area of Sarghas (Gonbadli sub-basin, Khorasan Razavi province). *Ecohydrology* 4 (3): 817–824.

Zare Chahuki M. A., Abbasi M. (2018) Modeling the habitat prediction of Rheum ribes L. using the maximum entropy (Maxent) model in the Chah Tash rangelands of Yazd province. *Journal of Rangeland Science and Watershed Management* 71 (2): 379–391.

Zarkami S., Ahmadi M., Abedini A. (2021) Modeling the habitat preference of the invasive plant Eichhornia crassipes in some wetlands of Gilan province. *Iranian Journal of Plant Biology (Iranian Journal of Biology)* 34 (2): 286–275.