Ecological Engineering 51 (2013) 83–87

Contents lists available at SciVerse ScienceDirect

Ecological Engineering

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

Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills

Xue-Qing Yang[a,1], S.P.S. Kushwaha[b,*], S. Saran[b,2], Jianchu Xu[c,3], P.S. Roy[b,1]

[a] Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China
[b] Indian Institute of Remote Sensing, Indian Space Research Organization, Dehradun 248001, U.K., India
[c] World Agroforestry Centre, China and East Asia Office, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, Yunnan, China

ARTICLE INFO

Article history:
Received 19 March 2012
Received in revised form 23 August 2012
Accepted 1 December 2012
Available online 3 January 2013

Keywords:
AUC
J. adhatoda
Jackknife
Maxent
Species distribution

ABSTRACT

The population of the medicinal plant, Malabar nut (*Justicia adhatoda* L.) is shrinking in Dun valley due to habitat fragmentation, invasion by *Lantana camara*, over-exploitation, and an ever-increasing human population – the most important being the increasing demand on land for agriculture, industries and the urbanization. Predicting potential geographic distribution of the species is important from species and habitat restoration point of view. This paper reports the results of a study carried out in the Lesser Himalayan foothills in India (Dun valley) on potential distribution modeling for Malabar nut using Maxent model. The *Worldclim* bioclimatic variables, slope, aspect, elevation, and the land use/land cover (based on IRS LISS-III) data and 46 spatially well-dispersed species occurrence points were used to predict the potential distribution of *J. adhatoda* in ca. 1877 km^2 study area. Jackknife test was used to evaluate the importance of the environmental variables for predictive modeling. Maxent model was highly accurate with a statistically significant AUC value of 92.3. The approach could be promising in predicting the potential distribution of medicinal plant species and thus, can be an effective tool in species restoration and conservation planning.

1. Introduction

Habitat degradation and fragmentation, invasion by alien species, over-exploitation, and an ever-increasing human population are some of the important factors responsible for the species loss throughout the world (Barnosky et al., 2011), bringing about one-fifth of the plant species at the risk of extinction (Brummitt and Bachman, 2010). Species habitat restoration is one of the successful ecological engineering measures for the species rehabilitation and habitat conservation (Polak and Saltz, 2011). A detailed knowledge on the current distribution of species is often a pre-requisite to rehabilitate the species in any ecosystem (Franklin, 2009; Barik and Adhikari, 2011). The analysis of species–environment relationship has always been a central issue in ecology and biogeography (Guisan and Zimmerman, 2000). Remote sensing provides several useful input variables such as vegetation type and density, biome, landscape or eco-region maps (Turner et al., 2003; Kushwaha, 2011) while the Geographic Information System (GIS) happens to be an ideal tool for geospatial database creation, data integration and modeling (e.g. Irfan-Ullah et al., 2006).

Models predicting the potential geographic distribution of species are important for a variety of applications in conservation biology (Ferrier, 2002; Graham et al., 2004). A large number of statistical models are currently in use to simulate the spatial distribution of plant species (Kumar and Stohlgren, 2009; Adhikari et al., 2012), spread of invasive species (Peterson et al., 2003; Thuiller et al., 2005), spatial patterns of species diversity (Graham et al., 2006), or impact of climate change (Thomas et al., 2004; Saran et al., 2010). Maximum entropy (Maxent) model (Phillips et al., 2004) is a species distribution model (SDM) originating from the statistical mechanics (Jaynes, 1957). It is a general purpose environmental model for predicting the potential distribution of species. The method has several advantages; it requires only species presence (or occurrence) data and environmental information (Elith et al., 2011). Presence-only modeling methods simply require a set of known occurrences together with predictor variables such as topography, climate, soil, biogeography etc. (Phillips and Dudik, 2008). It can make use of both continuous and categorical data and incorporate the interactions between the variables (Phillips et al.,

* Corresponding author. Tel.: +91 0135 2524170; fax: +91 0135 2741987.
E-mail address: spskushwaha@gmail.com (S.P.S. Kushwaha).
[1] Tel.: +91 0135 2524170; fax: +91 0135 2741987.
[2] Tel.: +91 0135 2744583; fax: +91 0135 2741987.
[3] Tel.: +86 0871 5223014; fax: +86 0871 5216350.

0925-8574/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoleng.2012.12.004

2006). There is no dearth of successful cases of this model's application in predicting species distribution (Reiss et al., 2011; Fuller et al., 2012).

Malabar nut, *Justicia adhatoda* (syn. *Adhatoda zeylanica*, *A. vasica*; Family: Acanthaceae), is a 1–6 m tall, evergreen, perennial shrub. It is distributed in the open/sparse tree canopy habitats in tropical to sub-tropical areas up to 1450 m altitude. The plant has played an important role not only in the ancient Indian system of medicine – the Ayurveda (Das et al., 2009), but also offers important inputs to modern pharmaceutical and cosmetics industries. Over time, the habitat of Malabar nut has shrunk due to an ever-increasing demand on land for cultivation, industry, human habitation, and other economic development activities. It is, therefore, of considerable interest if information on the potential habitats of this medicinally important plant is known for reintroduction and restoration.

2. Study area

The study area, Dun valley is located in the Lesser Himalayan foothills falling in Dehradun district of Uttarakhand, India. The Valley (30°00′–30°30′N and 77°36′–78°18′E), with an area of 1877 km^2, is bounded by Shivalik hills in the south and Lesser Himalayas in the north. It receives a mean annual rainfall of 2051 mm. The temperature ranges from 2 °C in winter to 42 °C in summer. Dun valley is a forested landscape with forests, agriculture, settlements, orchards and tea gardens. The principal forest types are: (i) *Moist Bhabar-Dun Sal Forest* (3C/C_2b_i), (ii) *Lower Himalayan Moist Temperate Forest* (12C_1), (iii) *Himalayan Subtropical Pine Forest* (9/C_1), and (iv) *Northern Dry Mixed Deciduous Forest* (5B/C_2) (Champion and Seth, 1968). *Shroea robusta*, *Terminalia tomentosa*, *Anogeissus latifolia*, *Mallotus philippensis*, *Dalbergia sissoo*, and *Acacia catechu* are some of the important tree species in the Valley. After becoming capital city of the Uttarakhand province in the year 2000, Dehradun district, in general, and Dun valley in particular have come under tremendous anthropogenic pressure that has taken a heavy toll on the several important plant species including Malabar nut.

3. Materials and methods

3.1. Land use/land cover (LULC) mapping

The Indian Remote Sensing satellite (IRS P6) LISS-III images (path-row: 96-49 and 96-50) with 23.5 m spatial resolution were mosaiced to extract the False Color Composite (FCC) image of the study area. Landsat 7 ETM+ image was used as reference image and the LISS-III images were geometrically corrected with sub-pixel accuracy using uniformly distributed ground control points (GCPs). First order polynomial transformation and nearest neighbor re-sampling method were chosen to preserve original brightness values in the image. The WGS 84 datum and Universal Transverse Mercator (UTM) projection (Zone – 44 N) were used throughout. Two-season data (of December 10, 2008 and March 16, 2009) were used for LULC ("categorical" data) mapping using on-screen visual interpretation. The LULC classes were field validated for mapping accuracy assessment.

3.2. Environmental variables and species occurrence data

Nineteen bioclimatic variables (Hijmans et al., 2005) with 30 seconds (ca. 1 km) spatial resolution, downloaded from *WorldClim* dataset (www.worldclim.org), were used to find out the most influential variables associated with *J. adhatoda* distribution. The Shuttle Radar Topographic Mission (SRTM) Digital Terrain Model (DTM) with 90 m resolution, downloaded from the www.srtm.usgs.gov website, was used to generate the slope, aspect and elevation data layers (Appendix-Table A.1). The multicollinearity test was conducted by using Pearson Correlation Coefficient (r) to examine the cross-correlation and the variables with cross-correlation coefficient value of >±0.8 were excluded (Appendix-Table A.2). A total of 102 species occurrences were recorded randomly in the study area in February and March, 2011 (during flowering season of *J. adhatoda*). A hand-held multi-channel *Trimble-SB* Global Positioning System (GPS) receiver with ±5 m positional accuracy was used to acquire the species occurrence geo-coordinates. Fifty six spatially correlated species occurrence points were excluded from the analysis and remaining 46 were used in modeling.

3.3. Spatial modeling

The *ENVI* 4.3 and *ArcGIS* 9.1 were used to create the spatial data layers. The categorical data were re-sampled to 1 km spatial resolution using nearest neighbor re-sampling technique. Maxent model is a maximum entropy-based machine learning program that estimates the probability distribution for a species' occurrence based on the environmental constraints (http://www.cs.princeton.edu/~schapire/maxent/). The principle of maximum entropy approach is to ensure that approximation satisfies any constraints on the unknown sites, meaning that the

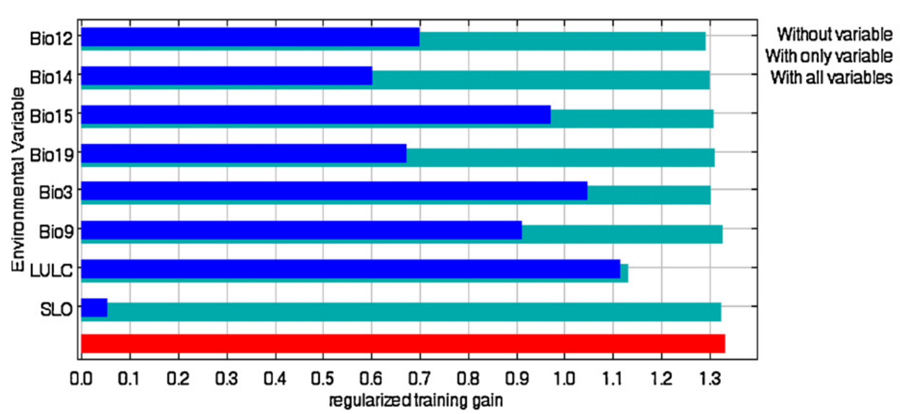

Fig. 1. The Jackknife test for evaluating the relative importance of environmental variables for *J. adhatoda* in Dun valley. (*Note*: "Bio12" is annual precipitation; "Bio14" is precipitation of driest period; "Bio15" is precipitation seasonality; "Bio19" is precipitation of coldest quarter; "Bio3" is isothermality; "Bio9" is mean temperature of driest quarter; "LULC" is land use/land cover map; "SLO" is slope.)

estimated probability of unknown distribution involves less number of constraints but more choices (Jaynes, 1957).

In order to avoid over-fitting of the test data, we set the regularization multiplier value as 0.1 (Phillips et al., 2004). The maximum number of background points was 10,000. Linear, quadratic and hinge features were used. We selected 70% data for training and the rest 30% for testing. A total of 100 runs were set for model building (Flory et al., 2012). Other values were kept as default. The Area Under the Receiving Operator Curve (AUC) was used to evaluate model's goodness-of-fit and model with highest AUC value was considered as the best performer. The *Jackknife* procedure was used to assess the importance of the variables. The final potential species distribution map had a range of values from 0 to 1 which were regrouped in to four classes of potential habitats *viz*., 'high potential' (>0.6), 'good potential' (0.4–0.6), 'moderate potential' (0.2–0.4) and 'least potential' (<0.2).

4. Results and discussion

The mapping accuracy of the LULC map was 91.8%. The Jackknife evaluation results indicated isothermality and LULC as main factors influencing *J. adhatoda* distribution (Fig. 1). The percent contribution values given in Appendix-Table A.1 are only heuristically defined; they depend on the particular path that the Maxent code uses to get to optimal solution. These results were consistent with the Jackknife evaluation. The model output provided satisfactory results with the given set of training and test data, the final model

Table A.1
Environmental variables used in the study and their percentage contribution.

Code	Environmental variables	Unit	% Contribution
Bio1	Annual mean temperature	°C	
Bio2	Mean diurnal range (mean of monthly max. and min. temp.)	°C	
Bio3	**Isothermality** ((Bio2/Bio7) × 100)	–	72.2
Bio4	Temperature seasonality (standard deviation × 100)	C of V	
Bio5	Maximum temperature of warmest month	°C	
Bio6	Minimum temperature of coldest month	°C	
Bio7	Temperature annual range (Bio5–Bio6)	°C	
Bio8	Mean temperature of wettest quarter	°C	
Bio9	**Mean temperature of driest quarter**	°C	0.7
Bio10	Mean temperature of warmest quarter	°C	
Bio11	Mean temperature of coldest quarter	°C	
Bio12	**Annual precipitation**	mm	1.3
Bio13	Precipitation of wettest period	mm	
Bio14	**Precipitation of driest period**	mm	1.3
Bio15	**Precipitation seasonality (CV)**	C of V	1.0
Bio16	Precipitation of wettest quarter	mm	
Bio17	Precipitation of driest quarter	mm	
Bio18	Precipitation of warmest quarter	mm	
Bio19	**Precipitation of coldest quarter**	mm	0.8
LULC	**Land use and land cover**	15 types	21.9
ELE	Elevation	m	
SLO	**Slope**	°	0.8
ASP	Aspect	°	

Note: The highlighted variables, selected through multi-collinearity test, were used in modeling.

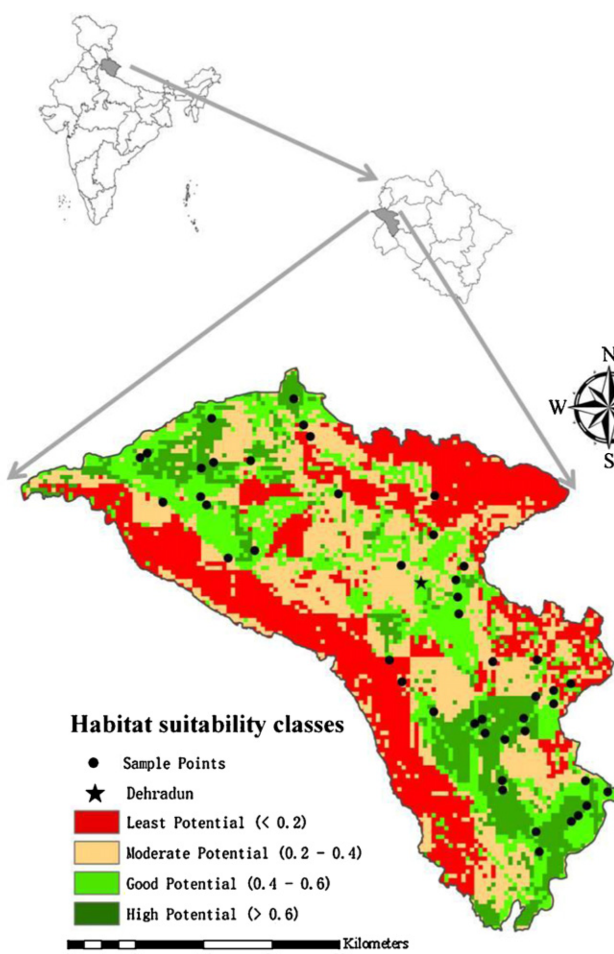

Fig. 2. Predicted potential distribution of *J. adhatoda* within Dun valley.

Table A.2
Multi-collinearity test by using cross-correlations (Pearson correlation coefficients, *r*) among environmental variables.

Variables	Bio1	Bio2	Bio3	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19	LULC	ELE	SLO	ASP
Bio1	1																						
Bio2	≈1	1																					
Bio3	.94	.94	1																				
Bio4	.98	.98	.90	1																			
Bio5	≈1	≈1	.94	.98	1																		
Bio6	≈1	.99	.94	.97	≈1	1																	
Bio7	≈1	≈1	.93	.99	≈1	.99	1																
Bio8	≈1	≈1	.93	.99	≈1	≈1	≈1	1															
Bio9	.82	.83	.80	.84	.83	.82	.83	.83	1														
Bio10	≈1	≈1	.93	.99	≈1	.99	≈1	1	.83	1													
Bio11	≈1	≈1	.94	.97	≈1	.99	≈1	≈1	.81	≈1	1												
Bio12	−.69	−.69	−.62	−.73	−.70	−.69	−.70	−.70	−.64	−.70	−.70	1											
Bio13	−.51	−.51	−.44	−.56	−.52	−.51	−.52	−.52	−.44	−.51	−.50	.95	1										
Bio14	−.53	−.54	−.50	−.62	−.54	−.53	−.55	−.55	−.70	−.55	−.50	.78	.64	1									
Bio15	.45	.45	.51	.42	−.45	.48	.44	.45	.58	.44	.45	−.14	.11	−.46	1								
Bio16	−.63	−.63	−.55	−.68	−.64	−.63	−.64	−.64	−.57	−.64	−.61	.99	.97	.73	.98	1							
Bio17	−.74	−.75	−.73	−.78	−.75	−.75	−.75	−.75	−.77	−.74	−.72	.88	.74	.88	−.53	.81	1						
Bio18	−.74	−.74	−.68	−.78	−.75	−.74	−.75	−.75	−.70	−.74	−.72	.99	.90	.78	−.18	.97	.88	1					
Bio19	−.72	−.74	−.72	−.67	−.73	−.73	−.73	−.71	−.63	−.71	−.72	.78	.67	.59	−.49	.70	.85	.77	1				
LULC	−.66	−.66	−.66	−.66	−.66	−.66	−.66	−.66	−.63	−.66	−.66	.65	.65	.64	−.66	.65	.64	.64	.65	1			
ELE	−1	≈−1	≈−1	≈−1	≈−1	≈−1	≈−1	≈−1	−.99	≈−1	≈−1	.97	.97	.94	−.99	.97	.94	.96	.96	.66	1		
SLO	−.46	−.46	−.52	−.49	−.47	−.40	−.48	−.46	−.41	−.47	−.44	.61	.59	.62	−.53	.60	.64	.63	.62	.28	.46	1	
ASP	−.84	−.84	−.85	−.85	−.84	−.83	−.84	−.84	−.82	−.84	−.84	.85	.85	.84	−.85	.85	.84	.85	.85	.60	.84	.47	1

Note: If two variables had >±0.8, only one of them was selected in the same model. Correlations were significant at $\alpha = 0.05$ (calculated by statistic software, *SPSS 16.0*).

had high accuracy with an AUC value of 92.3. It is important to note that AUC values tend to be lower for species that have broad distribution scope (Mcpherson and Jetz, 2007; Evangelista et al., 2008). As evident from the Fig. 2, the eastern part of Dun valley, in general, has higher potential than the western part for *J. adhatoda*, which could primarily be attributed to the LULC types in this part of the Valley, i.e. open forest, scrub, agricultural land, forest plantations, and the non-forest areas. The model results showed that an area of 638.80 km^2 (34.03%) has high to good potential. Approximately 574.38 km^2 (30.6%) area showed medium potential while 663.82 km^2 (35.37%) area showed least potential. The least potential area is mostly covered by dense forest, which is an unsuitable habitat for light loving *J. adhatoda*.

Our results support the statement that the predicted potential distribution areas through Maxent modeling almost always appear as over-estimated compared to realized niche of the species, i.e. the habitat. Since Maxent model considers only niche-based presence data, it predicts the species fundamental niche rather than realized niche (Pearson, 2007; Kumar and Stohlgren, 2009). In reality, a species might have failed to disperse due to geographic barriers, human disturbance or associated competitive species. Noxious weed, *L. camara* happens to be a major constraint for the growth and distribution of *J. adhatoda* in Dun valley. The areas predicted suitable but currently not occupied by the species are the candidate areas to be considered for conservation prioritization and propagation of this species. The method is certainly promising in predicting the potential distribution of other medicinal plant species and can be a valuable tool in species conservation planning and climate change-species distribution studies.

Acknowledgements

Xueqing Yang gratefully acknowledges the funding and support from the Center for Space Science and Technology Education in Asia and the Pacific, Dehradun (affiliated to United Nations) and CGIAR Research Program-6 on Forests, Trees and Agroforestry. The authors wish to thank the reviewers for their constructive comments on the early version of the manuscript.

Appendix A.

See Tables A.1 and A.2.

References

Adhikari, D., Barik, S.K., Upadhaya, K., 2012. Habitat distribution modelling for reintroduction of Ilex khasiana Purk., a critically endangered tree species of northeastern India. Ecol. Eng. 40, 37–43.

Barik, S.K., Adhikari, D., 2011. Predicting geographic distribution of an invasive species *Chromolaena odorata* L. (King) & H.E. Robins. In: Bhatt, J.R., Singh, J.S., Tripathi, R.S., Singh, S.P., Kohli, R.K. (Eds.), Invasive Alien Plants – An Ecological Appraisal for the Indian Sub-continent. CABI, Oxfordshire, pp. 77–88.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the earth's sixth mass extinction already arrived? Nature 471, 51–57.

Brummitt, N., Bachman, S., 2010. Plants under pressure a global assessment. The first report of the IUCN Sampled Red List Index for Plants. Royal Botanic Gardens, Kew.

Champion, H.G., Seth, S.K., 1968. A Revised Survey of the Forest Types of India. Manager of Publications, Govt. of India, New Delhi.

Das, S., Khan, M.L., Rabha, A., Bhattacharya, D.K., 2009. Ethnomedicinal plants of Manas National Park, Assam, Northeast India. Indian J. Tradit. Know. 8, 514–517.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of Maxent for ecologists. Divers. Distrib. 17, 43–57.

Evangelista, P.H., Kumar, S., Stohlgren, T.J., Jarnevich, C.S., Crall, A.W., Norman, J.B., Barnett, D.T., 2008. Modelling invasion for a habitat generalist and a specialist-plant species. Divers. Distrib. 14, 808–817.

Franklin, J., 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, U.K, 338.

Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst. Biol. 51, 331–363.

Flory, A.R., Kumar, S., Stohlgren, T.J., Cryan, P.M., 2012. Environmental conditions associated with bat white nose syndrome mortality in the north-eastern United States. J. Appl. Ecol. 49, 680–689.
Fuller, D.O., Ahumada, M.L., Quiñones, M.L., Herrera, S., Beier, J.C., 2012. Near-present and future distribution of *Anopheles albimanus* in Mesoamerica and the Caribbean Basin modeled with climate and topographic data. Int. J. Health Geogr. 11, 13.
Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol. Evol. 19, 497–503.
Graham, C.H., Moritz, C., Williams, S.E., 2006. Habitat history improves prediction of biodiversity in a rainforest fauna. Proc. Natl. Acad. Sci. U.S.A. 103, 632–636.
Guisan, A., Zimmerman, N.E., 2000. Predicting habitat distribution models in ecology. Ecol. Model. 135, 147–186.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surface for global land areas. Int. J. Climatol. 25, 1965–2198.
Irfan-Ullah, M., Amarnath, G., Murthy, M.S.R., Peterson, A.T., 2006. Mapping the geographic distribution of *Aglaia bourdillonii* Gamble (Meliaceae), an endemic and threatened plant using ecological niche modeling. Biodivers. Conserv. 16, 1917–1925.
Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620–630.
Kumar, S., Stohlgren, T.J., 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. J. Ecol. Nat. Environ. 1, 94–98.
Kushwaha, S.P.S., 2011. Remote sensing of invasive alien plant species. In: Bhatt, J.R., Singh, J.S., Tripathi, R.S., Singh, S.P., Kohli, R.K. (Eds.), Invasive Alien Plants – An Ecological Appraisal For the Indian Sub-continent. CABI, Oxfordshire, pp. 131–138.
Mcpherson, J.M., Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. Ecography 30, 135–151.
Pearson, R.G., 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History, Available at: http://ncep.amnh.org
Peterson, A.T., Papes, M., Kluza, D.A., 2003. Predicting the potential invasive distributions of four alien plant species in North America. Weed Sci. 51, 863–868.
Phillips, S.J., Miroslav, D., Schapire, R.E., 2004. Maxent Software for Species Distribution Modeling. http://cs.princeton.edu/~schapire/Maxent/
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.
Phillips, S.J., Dudik, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.
Polak, T., Saltz, D., 2011. Reintroduction as an ecosystem restoration technique. Conserv. Biol. 25, 424–427.
Reiss, H., Cunze, S., König, K., Neumann, H., Kröncke, I., 2011. Species distribution modelling of marine benthos: a North Sea case study. Mar. Ecol. Prog. Ser. 442, 71–86.
Saran, S., Joshi, R., Sharma, S., Padalia, H., Dadhwal, V.K., 2010. Geospatial modelling of brown oak (*Quercus semecarpifolia* Sm.) habitats in the Kumaun Himalaya under climate change scenario. J. Indian Soc. Remote Sens. 38, 534–547.
Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Aarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 27, 145–147.
Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.F., Hughs, G.O., Rouget, M., 2005. Niche-based modeling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biol. 11, 2234–2250.
Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. Trends Ecol. Evol. 18, 306–314.