Comparative Analysis of Species Distribution Modeling of *Daphne papyracea* in Dabka Watershed Nainital District, Uttarakhand

Nancy Raina[1] Y. S. C. Khuman[2] K. S. Rao[1] S.Sreekesh[3]
1.Botany Department, University of Delhi, Delhi
2. School of Agriculture, Indira Gandhi National Open University, New Delhi
3.CSRD, Jawaharlal Nehru University, Delhi
E-mail of corresponding author: nraina19@gmail.com

Abstract
Species distribution modeling with different algorithms is used to understand the ecological and geographical distribution of the species in the Himalayan region of Nainital district using various climatic and topographic variables. In this paper, various algorithms were compared using different measures of performance evaluation (threshold-dependent and threshold-independent) for *Daphne papyracea*, and the modelling results were assessed using indices: sensitivity, specificity, overall prediction success and Cohen's kappa statistic, NMI, AUC, CCRate, MCRate. These indices were used to identify the best model for the species in the study area and best performance was observed for Environmental Distance Algorithm. Such studies therefore, will try to define the suitable management strategies (conservation and protection) for Daphne papyracea that is an economically important plant having medical value.
Keywords: Species distribution modeling, *Daphne papyracea* , Himalaya, Environmental Distance Algorithm, performance evaluation

1. Introduction

With increasing transformation in the land use due to various anthropogenic activities like urbanization, deforestation that has resulted into habitat fragmentation and its loss thus affecting economically important plant species. Therefore species distribution modeling can be used for ecological niche modeling identifying the potential geographical areas using environmental variables (i.e abiotic conditions such as rainfall rate, topography, temperature) in addition to the geographic position of the plant species (Mendoza-Gonzalez et al., 2013). However, even based on limited field data species distribution patterns can also be predicted (Austin, 1998). Such models can also be used for extrapolating species distribution with different environmental variables inside and outside areas of species niche producing native and invaded geographic distribution (Peterson, 2003). These ecological distributional models is an important component of conservation biology that uses known distribution records of a species, as well as environmental and spatial explanatory variables, for interpolating species' distributions across region (Guisan and Zimmermann, 2000). Models may also extrapolate species' distributions to sets of environmental conditions outside those used to build the models (Peterson, 2003).

The Himalayas are repository to large amount of faunal and floral diversity. It is considered to be one of the hotspots. It covers a total of 18% of India's geographical area, and accounts for more than 50% of India's forest cover and 40% of the species endemic to the Indian subcontinent (Semwal et al., 2004). The region is geo-dynamically unstable, ecologically fragile and underdeveloped. With the rapid increase in population over the years, there has been increase in the demand for food, fuel, fodder putting an increased pressure on various natural resources (Tiwari, 2000). In present scenario, sustainable development of natural resources of the Himalayas has acquired added importance in order to maintain the ecological balance in the mountains. The primary step is to identify sensitive areas, and derive the conservation plan for such area. Thus the geographic location of plant species is the basis for planning any management strategy. Since the distribution pattern of the plant species is scarce in the area, the study was an attempt for finding the distribution pattern of ***Daphne papyracea*** using various abiotic variables with different algorithms models.

2. Material & Methods

2.1 Study Area

The study region of Dabka watershed falls in the Himalayan zone, Nainital district of Kumaon region in Uttarakhand with the total area of the watershed is 69.06 sq. km (Figure 1). Its geographical location lies between 79° 17' 53" and 79° 25' 38" longitude, 29° 30' 19" and 29° 24' 09" latitude. The elevation ranges from an altitude of as high as 2455 m asl and as low as 750 m asl bounded by steep slope, topography.

It has subtropical climate and is characterized by abundant and seasonally intense rainfall. The annual rainfall is 1865 mm with the maximum rainfall in the months of July and August together showing rainfall of more than 1000mm.The temperature increases to as high as 36^0 C in summers (May- June) in lower elevation zones whereas at higher altitude of the watershed the temperature lowers to 0^0 C in winter season (Dec- Feb). Humidity varies from 98% (Max) in rainy season (July- September), 3.5% (Min) during winters (Dec- Feb).

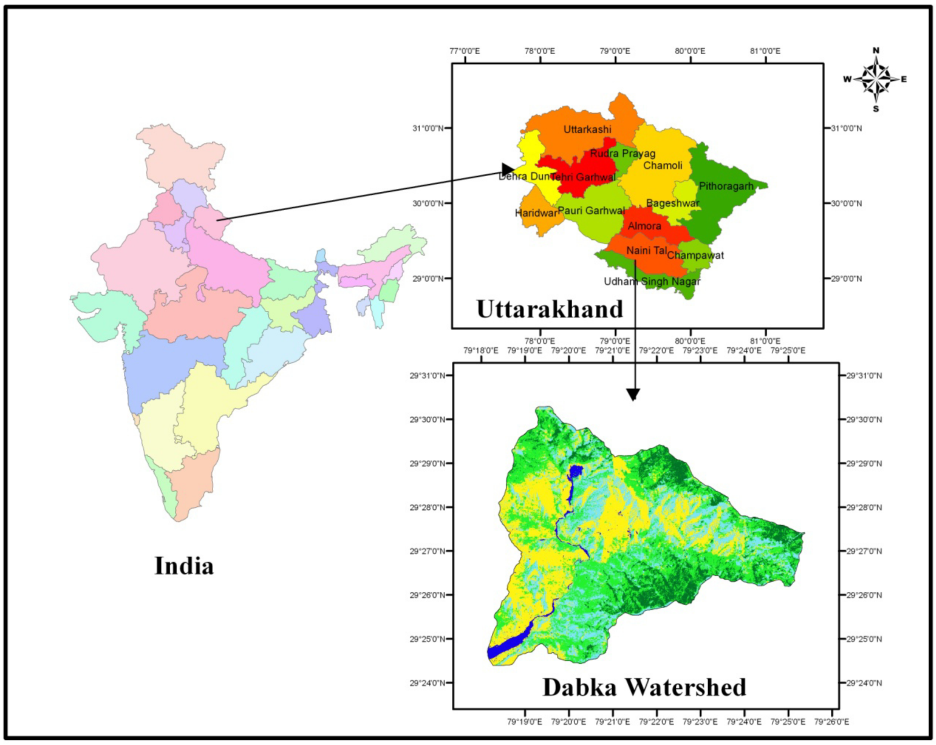


Figure 1: Study area

2.2 Species Distribution Modelling

Daphne papyracea Wall. Ex Steud. (Thymelaeaceae) is an evergreen shrub of height and grows in the Himalayan region in the elevation zone of 1600-2500m. The bark/fruits are used as dye-yielding plant. It has medicinal value as it used as medicine for curing high blood pressure. (Kumari et al., 2011).

Modelling was performed using Open-Modeller (Version 1.0.8) with different algorithms (Table 1). Open Modeller was chosen due to its wide application to modeling of ecological niches and geographical distribution of species (Stockwell and Peters, 1999). The 19 environmental data layers described in (Table 2) with a resolution of 30 arc-second resolution grid ("1 km^2") resolution along with field absence-presence data of *Daphne papyracea* was collected from 221 sample points across the entire study area and latitude and longitude with GPS (Garmin GPS-76). 136 points were used for performing training and the rest for the testing. Based on different algorithm in the computing system, results were obtained in the form of maps (Figure 2).

Table 1: Algorithms used for predictive distribution modelling of *Daphne papyracea* (openModeller)

Algorithm	Study	Description
Bioclim	Nix, (1986); Brereton *et al*. (1995); Beaumont and Hughes, 2002	Implements the Bioclimatic Envelope Algorithm. Each variable has its own envelope represented by the interval [**m - c*s, m + c*s**], where '**m**' is the mean; '**c**' is the cutoff input parameter; and's' is the standard deviation. Besides the envelope, each environmental variable has additional upper and lower limits taken from the maximum and minimum values related to the set of occurrence points.
Climate Space Model		CSM is a principle components based algorithm. The component selection process in this algorithm implementation is based on the Broken-Stick cutoff where any component with an eigenvalue less than n standard deviation above a randomized sample is discarded.
Envelope Score	Piñeiro, R.,*et al*., (2007)	For each given environmental variable the algorithm finds the minimum and maximum at all occurrence sites. The probability of occurrences is determined as: p = layers within min-max threshold / number of layers
Environmental Distance	Carpenter *et al*., (1993)	Generic algorithm based on environmental dissimilarity metrics. When used with the Gower metric and maximum distance 1, this algorithm should produce the same result of the algorithm known as DOMAIN.
GARP *Genetic algorithm for rule-set prediction*	Stockwell, (1999); Stockwell and Noble, (1992)	GARP is a genetic algorithm that creates ecological niche models for species. The models describe environmental conditions under which the species should be able to maintain populations. For input, GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive.
GARP (single run) - new openModeller implementation	Stockwell (1999); Stockwell and Noble, (1992)	Gene values changed from integers (between 1 and 253) to floating point numbers (between -1.0 and 1.0). This avoids precision problems in environment values during projection (for example, if an environment variable has the value 2.56 in some raster cell and 2.76 in another one, DesktopGarp rounds them off to 3).
GARP with Best Subsets - new openModeller implementation	Peterson (2001); Anderson et al., (2002a; 2002b); Anderson et al., (2003)	GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive. This algorithm applies the Best Subsets procedure using the new openModeller implementation in each GARP run.
GARP with best subsets - DesktopGARP implementation	Peterson (2001); Anderson et al., (2002a; 2002b); Anderson et al. (2003).	GARP uses a set of point localities where the species is known to occur and a set of geographic layers representing the environmental parameters that might limit the species' capabilities to survive. This algorithm applies the Best Subsets procedure using the DesktopGARP implementation in each GARP run.
SVM (Support Vector Machines)	Vapnik, (1995); Schölkopf, et al.,(2000); Schölkopf, (2001). Cristianini and Shawe-Taylor (2000)	Support vector machines map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data. The separating hyperplane is the hyperplane that maximises the distance between the two parallel hyperplanes. An assumption is made that the larger the margin or distance between these parallel hyperplanes the better the generalisation error of the classifier will be.
Artificial Neural Network	Thuiller (2003,2004); Araújo et al. (2005)	Neural networks are non-linear statistical data modeling or decision making tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.

Table 2: Bioclimatic variables used for the predictive modelling

Code	Variable	Unit
BIO1	Annual Mean Temperature	^{0}C
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	^{0}C
BIO3	Isothermality (BIO2/BIO7) (* 100)	-
BIO4	Temperature Seasonality (standard deviation *100)	C of V
BIO5	Max Temperature of Warmest Month	^{0}C
BIO6	Min Temperature of Coldest Month	^{0}C
BIO7	Temperature Annual Range (BIO5-BIO6)	^{0}C
BIO8	Mean Temperature of Wettest Quarter	^{0}C
BIO9	Mean Temperature of Driest Quarter	^{0}C
BIO10	Mean Temperature of Warmest Quarter	^{0}C
BIO11	Mean Temperature of Coldest Quarter	^{0}C
BIO12	Annual Precipitation	mm
BIO13	Precipitation of Wettest Month	mm
BIO14	Precipitation of Driest Month	mm
BIO15	Precipitation Seasonality (Coefficient of Variation)	C of V
BIO16	Precipitation of Wettest Quarter	mm
BIO17	Precipitation of Driest Quarter	mm
BIO18	Precipitation of Warmest Quarter	mm
BIO19	Precipitation of Coldest Quarter	mm
SLP	Slope	^{0}
ASP	Aspect	^{0}
ELEV	Elevation	m

2.3 Performance criteria

Several performance criteria are from the values in confusion matrix, (Table 3) various performance measures including overall prediction success (matching coefficient; Buckland and Elston, 1993), sensitivity, specificity, NMI (Forbes, 1995) Area under Curve (AUC; Fielding and Bell, 1997), Misclassification Rate and Kappa (Cohen, 1960) were calculated.

Table 3: Confusion Matrix (Fielding & Bell 1997)

	True Presence	True Absence
Predicted Presence	a	b
Predicted Absence	c	d

Sensitivity (Sn) is defined as the percentage of true positives predicted correctly whereas Specificity (Sp) is defined as percentage of true negatives correctly predicted. The values range from 0 to 1 with higher value indicating higher accuracy in the model (Mouton et al., 2010).

Cohen's kappa, a derived statistic that measures the proportion of all possible cases of presence or absence that are predicted correctly by a model after accounting for chance. For kappa, values of 0.0-0.4 are considered to indicate slight to fair model performance, values of 0.4-0.6 moderate, 0.6-0.8 substantial and 0.8- 1.0 almost perfect (after Landis and Koch, 1977).

Except AUC all other variables used are threshold dependent. It also include plots based on receiver-operating characteristics (ROC plots), that indicates model performance independently of threshold values that are required in presence-absence models. They assess the performance of model output at all possible probability thresholds at which presence might be accepted (i.e. $p > 0$ to $p < 1$). The curve is obtained by plotting sensitivity vs. (1 - specificity) for varying probability thresholds. Good model performance is characterized by a curve that maximizes sensitivity for low values of (1- specificity) (Robertson et al., 1983). High performance models are indicated by large areas under the ROC curves (i.e. large areas under the curve; AUC). The values ranges from 0 to 1, where according to Thullier et al., (2005) value ranging in between 0.9-1.0 indicates perfect discrimination and very good model, values between 0.9-0.95 indicates a good model, 0.8-0.9 indicates model with fair performance and values below 0.8 shows a poor model. It is considered to be best effective measure having discriminatory ability for assessing the accuracy of presence/absence distribution models (Thuiller et al., 2003; Rushton et al., 2004; Austin, 2007).

NMI make full use of the information contained in the confusion matrix (Fielding and Bell, 1997) and quantifies the information included in the model predictions compared to that included in the observations (Mouton et al., 2010). The values range from 0 (completely inaccurate), to 1 where presence–absence is perfectly predicted (Forbes 1995).

Table 4: List of variables used for performance measurement (Manel *et al.* 2001 & openModeller)

Performance measure	Definition	Formula
Overall prediction success or Correct Classification Rate	Percentage of all correctly predicted (S)	a + d/n
Sensitivity	Percentage of true positives correctly predicted (Sn)	a/(a +c)
Specificity	Percentage of true negatives correctly predicted (Sp)	d/(b + d)
Odds ratio	Ratio of correctly assigned cases to incorrectly assigned cases	ad/cb
NMI Normalized Mutual Information statistic		$= 1 - \frac{a X \ln(a) - bX\ln(b) - cX\ln(c) - dX\ln(d) + (a+B)X\ln(a+b)X\,ln(a+b) + (c+d)X\,ln(c+d)}{n\,X\ln(n) - ((a+c)X\,ln(a+c) + (b+d)X\,ln\,(b+d)}$
Kappa	Proportion of specific agreement	[(a+d)-((a+c)(a+b)+(b+d)(c+d))/n)]/[n-(((a+c)x(a+b))+(b+d)(c+d))/n)]
Misclassification Rate (MC Rate)		(b+c)/n

3. Results

The maps obtained through different algorithm of models (Figure 2) shows the potential distribution of *Daphne papyracea*. At 50% threshold, from the training data the average AUC for *Daphne papyracea* was found to be 1.00 and highest in the model "Environmental Distance" with the next highest value to be around 0.847 in GARP with best subsets- DesktopGARP implementation". The lowest value of 0.545 was found to be for the model SVM One Class (Table 5).

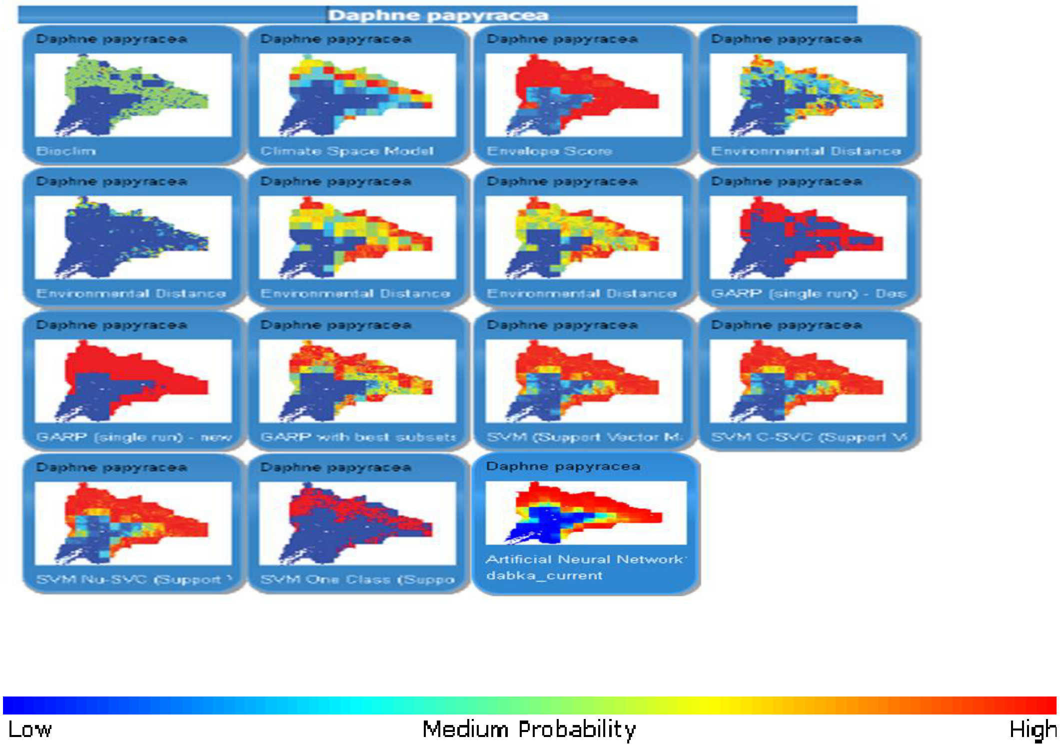


Figure 2: Potential distribution map areas of Daphne papyracea

Table 5. Statistics results of confusion matrix performed with different algorithms

Algorithm	**Avg. AUC**	**Avg. Accuracy (%)**	**Avg. Omission (%)**	**Avg. Commission (%)**
Bioclim	0.786	83.929	0.000	42.857
Climate Space Model	0.717	58.929	47.143	30.952
Envelope Score	0.786	79.464	0.000	54.762
Environmental Distance	1.000	83.929	0.000	42.857
Environmental Distance - Chebyshev	1.000	92.857	0.000	19.048
Environmental Distance - Mahalanobis	1.000	79.464	0.000	54.762
Environmental Distance - Manhattan	1.000	80.357	0.000	52.381
GARP (single run) - DesktopGARP	0.745	79.464	5.714	45.238
GARP (single run) – new openModeller	0.760	81.250	2.857	45.238
GARP with best subsets - DesktopGARP implementation	0.805	80.357	5.714	45.238
SVM (Support Vector Machines)	0.787	79.464	2.857	50.000
SVM C- SVC (Support Vector Machines)	0.787	80.357	2.857	47.619
SVM Nu-SVC (Support Vector Machines)	0.789	79.464	2.857	50.000
SVM One Class(Support Vector Machines)	0.595	57.143	50	30.952
Artificial Neural Network	0.79	78.877	23.913	52.381

Index for AUC:
Swets 1988: AUC values of 0.5-0.7 = low accuracy, values of 0.7-0.9 = useful applications and values of >0.9 = high accuracy
(Elith et al., 2002): 0.90–1.00 _ excellent; 0.80–0.90 _ good; 0.70–0.80 _ fair; 0.60–0.70 _ poor; 0.50–0.60 _ fail. AUC values of less than 0.5 indicate that the model tends to predict presence at sites at which the species is, in fact, absent.
(Thuiller et al. 2005): Poor (AUC < 0.8), fair (0.8 < AUC < 0.9), good (0.9 < AUC < 0.95) and very good (0.95 < AUC < 1.0).

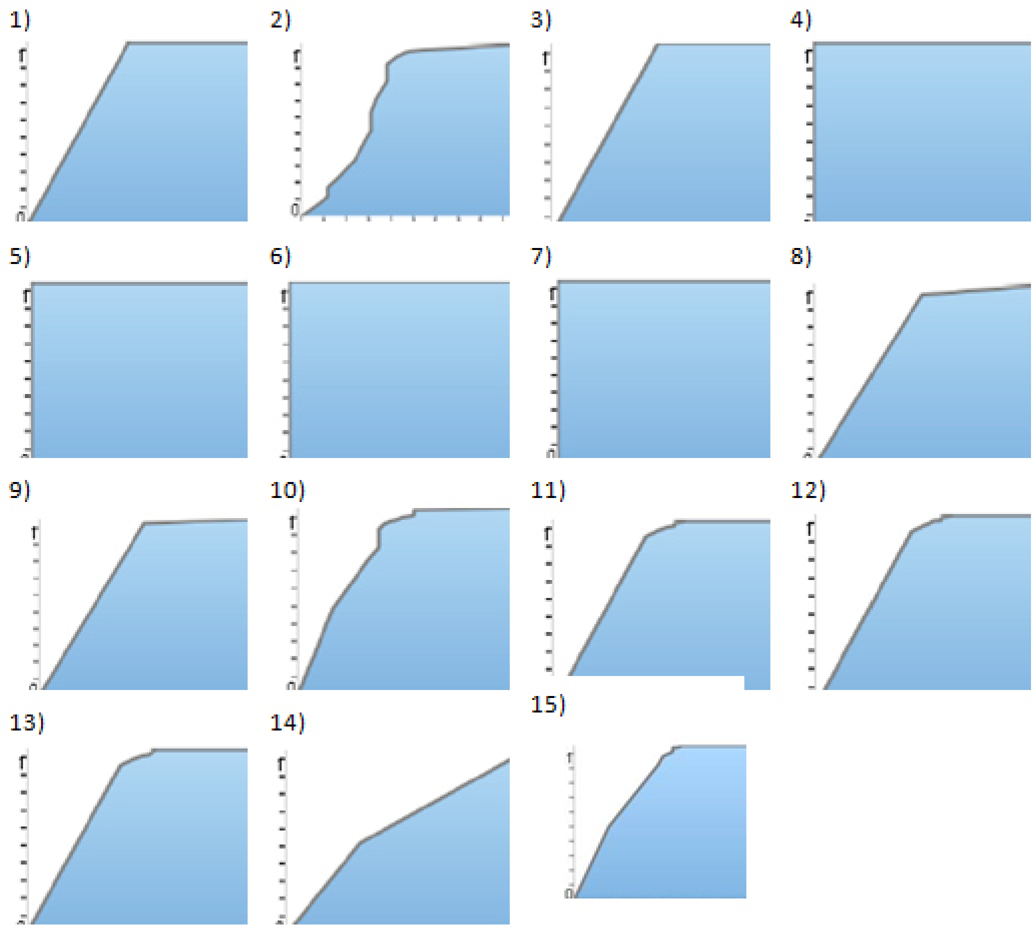
Correct Classification Rate was highest for Environment Distance with the value of 1 and the minimum was found in model of SVM One Class vector with the value of 0.598 (Table 5). Kappa coefficient was highest for all the modified versions of Environmental Distance with value of 1 and lowest was for SVM One Class (Support Vector Machines) with the value of 0.221. Sensitivity value of 1 was found in Bioclim, CSM, ES, ED, ED-C, ED-Mh, ED-Mn, GARP best subsets, SVM, SVM C-SVC, SVM C-SVC, SVM Nu-SVC and the lowest level of 0.529 was found in SVM 1-CL. Specificity was lowest in CSM with the value of 0.500 (Table 6).

A ROC plot is obtained by plotting all sensitivity values (true positive fraction) on the y axis against their equivalent (1 - specificity) values (false positive fraction) for all available thresholds. Environmental distance and its modified versions maximized sensitivity for low values of (1-specificity) and indicated large areas under ROC curves (Figure 3).Since in the study data the confusion matrix consists of 0 values, it cannot be applied due to its dependence on logarithmic data (Manel et al., 2001). The highest value of NMI was also shown by Environment Distance and its modified versions algorithms (Table 6).

Table 6. Statistics results of tests performed with different algorithms

Algorithm	CCRate	Sensitivity	Specificity	Kappa	MC Rate	NMI
Bioclim	0.839	1.000	0.571	0.625	0.161	0.659
Climate Space Model	0.813	1.000	0.500	0.556	0.188	0.624
Envelope Score	0.839	1.000	0.571	0.625	0.161	0.659
Environmental Distance	1.000	1.000	1.000	1.000	0.000	1
Environmental Distance – Chebyshev	1.000	1.000	1.000	1.000	0.000	1
Environmental Distance – Mahalanobis	1.000	1.000	1.000	1.000	0.000	1
Environmental Distance – Manhattan	1.000	1.000	1.000	1.000	0.000	1
GARP (single run) – DesktopGARP	0.777	0.886	0.595	0.502	0.223	0.545
GARP (single run) – new openModeller	0.813	0.986	0.524	0.560	0.188	0.605
GARP with best subsets – DesktopGARP implementation	0.804	1.000	0.476	0.532	0.196	0.613
SVM (Support Vector Machines)	0.813	1.000	0.500	0.556	0.188	0.624
SVM C- SVC (Support Vector Machines)	0.813	1.000	0.500	0.556	0.188	0.624
SVM Nu-SVC (Support Vector Machines)	0.813	1.000	0.500	0.556	0.188	0.624
SVM One Class(Support Vector Machines)	0.598	0.529	0.714	0.221	0.402	0.457
Artificial Neural Network	0.687	0.778	0.500	0.281	0.313	0.455

Index for Kappa coefficient:
Landis and Koch (1977): 0.81–1.00 = almost perfect; 0.61–0.80 = substantial; 0.41–0.60 = moderate; 00.21–0.40 = fair; 0.00–0.20 = fail.



1. Bioclim
2. Climate Space Model
3. Envelope Score
4. Environmental Distance
5. Environmental Distance - Chebyshev
6. Environmental Distance - Mahalanobis
7. Environmental Distance - Manhattan
8. GARP (single run) - DesktopGARP
9. GARP (single run) – new openModeller
10. GARP with best subsets -DesktopGARP implementation
11. SVM (Support Vector Machines)
12. SVM C- SVC (Support Vector Machines)
13. SVM Nu-SVC (Support Vector Machines)
14. SVM One Class(Support Vector Machines)
15. Artificial Neural Network

Figure 3: Receiver Operating Characteristic curve of *Daphne papyracea* for different algorithm

4. Discussion

Predictive Distribution models provide the ways to develop the set of environmental variables affecting the distribution at different geographical scales (Manel et al., 2001). Since the number of threatened species is increasing, the species distribution models are been increasingly used to derive the effective and accurate models for conservation management plan for the species.

A threshold level of 50% is taken as it is widely used threshold level in ecology (Manel et al., 2001; Luck, 2002; Stockwell and Peterson, 2002; Bailey et al., 2002; Woolf et al., 2002). For assessment of prediction of species distributions many indices can be used including sensitivity, specificity, correct classification rate, kappa, AUC. The area under the ROC function (AUC) has been used extensively for species distribution modeling by various authors and is considered to be an important index because it measures the accuracy without depending upon the threshold and indicates the usefulness of a model for prioritizing areas for a particular species conservation (Elith et al., 2006). Reviewing the AUC range values for prediction of model (Swets 1988; Elith et al., 2002; Thullier et al., 2005), it ranges from 0 to 1. In our study area the average AUC for all the models studied for *Daphne papyracea*, were above 0.7 except for SVM-One Class. Maximum mean scores of the AUC, do not exceed 0.82 and is around 0.70 for most species for different models (Elith et al., 2006). According to Thuiller et al, (2005) the AUC values were categorized as: poor (AUC < 0.8), fair (0.8 < AUC < 0.9), good (0.9 < AUC < 0.95) and very good (0.95 < AUC < 1.0). Environmental Distance and its variants showed the best performance with a score of 1.

Kappa is chance dependent measurement and is used to distinguish between predicted presence and predicted absence in ecological studies having presence-absence data. It measures the correct classification rate (proportion of correctly classified presences and absences) that considers both commission and omission errors (Elith et al. 2006). Landis and Koch (1977) proposed a scale on Kappa with categories: almost perfect (0.81–1.00) ; substantial (0.61–0.80) ; moderate (0.41–0.60); fair (00.21–0.40); and fail (0.00–0.20). In the different models studied SVM One Class and ANN showed a fair performance and Environmental distance and its modified versions showed perfect performance in the *Daphne papyracea*. From the testing data conducted the average AUC values of models of Environmental distance and its modified versions also showed the perfect score of 1 and the lowest was for SVM One class.

Confusion matrix-based measure should meet four requirements and obey six additional constraints measuring agreement and not association (Forbes, 1995). Kappa is a proportion of specific agreement but is sensitive to sample. NMI on other hand obeys all the requirements and constraints while being the most conservative of measures tested. It bears difficulty when there are zero values in any category of confusion matrix (Manel et al., 2001).

Since no model can excel on all the aspects of model performance, it depends upon the preferences of the modelers (Mouton et al., 2010). Rather than using single modeling technique, different models for each species should be performed and the most accurate technique should be selected (Heikkinen et al., 2006). It is thus concluded that the Environment Distances and its modified versions showed best values in all the indices, for D. papyracea that was further validated by test of AUC showing value of 1. The model of Environment Distance and its modified versions could be used for species distribution prediction in the Dabka watershed or regions with similar environmental characteristics. Since the plant studied has the economic value in form of its dye yielding properties (Gaur, 2008) and medicinal properties (Uniyal et al. 2002), it is important to conserve the area of its presence. Since the choice of model and variables influences the identification of areas for conservation depending upon the species locations or variables used, Conservation biologist could choose a model based on the ecology of the species and the availability of requisite data (Johnson and Gillingham, 2005). Thus our studies can help in defining the management strategies (conservation and protection) and designing conservation map by locating the potential areas of *Daphne papyracea* in the watershed.

References

Anderson, R. P., Gomez-Laverde M. & Peterson, A.T. (2002a), “Geographical distributions of spiny pocket mice in South America: insights from predictive models”, *Global Ecology Biogeography* 11, 131–141.

Anderson, R.P., Lewc, D. & Peterson, A.T. (2003). “Evaluating predictive models of species’ distributions: criteria for selecting optimal models”, *Ecological Modelling* 162, 211–232.

Anderson, R.P., Peterson, A. T. & Gomez-Laverde M. (2002b).”Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice”, *Oikos* 98, 3–16.

Araujo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005). “Validation of species–climate impact models under climate change”, *Global Change Biology* 11, 1504–1513.

Austin, M. (2007), “Species distribution models and ecological theory: a critical assessment and some possible new approaches”, *Ecological Modeling* 200, 1-19.

Austin, M.P. (1998), “An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests”, *Annals of the Missouri Botanical Garden* 85, 2-17.

Bailey, S.A., Haines-Young, R.H. & Watkins, C. (2002), “Species presence in fragmented landscapes: modeling of species requirements at the national level”, *Biological Conservation* 108, 307- 316.

Beaumont, L.J. & Hughes, L. (2002), “Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change”, *Global Change Biology* 8, 954–971.

Brereton, R., Bennett, S. & Mansergh, I. (1995), “Enhanced greenhouseclimate change and its potential effect on selected fauna of SouthEastern Australia: a trend analysis”, *Biological Conservation* 72, 339–354.

Buckland, S.T. & Elston, D.A., (1993) “Empirical models for the spatial distribution of wildlife”, *Journal of Applied Ecology* 30, 478-495.

Carpenter, G., Gillison, A.N. & Winter, J. (1993), “DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals” *Biodiversity and Conservation* 2, 667–680.

Cohen J. (1960), “A coefficient of agreement for nominal scales”, *Educational and Psychological Measurements* **20**(1), 37-46

Cristianini, N. & Taylor, J.S. (2000), "An introduction to Support Vector Machines and other Kernel based Learning Methods", Cambridge University Press.

Elith, J., Burgman, M.A. & Regan, H.M. (2002), "Mapping epistemic uncertainties and vague concepts in predictions of species distributions", *Ecological Modelling* 157, 313–29.

Elith, J., Graham C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S., Richardson, K., Schachetti Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M. & Zimmermann, N.E. (2006), "Novel methods improve predictions of species' distributions from occurrence data", *Ecography* 29(2), 129–151.

Fielding, A.H. & Bell, J. F. (1997), "A review of methods for assessment of prediction errors in conservation presence/absence models", *Environment Conservation* 24, 38-49.

Forbes, A.D. (1995), "Classification-algorithm evaluation: five performance measures based on confusion matrices", *Journal of Clinical Monitoring* 11, 189–206.

Gaur, R. (2008), "Traditional dye yielding plants of Uttarakhand, India", *Natural Product Radiance* **7**(2), 154-165.

Guisan, A. & Zimmermann, N.E. (2000), "Predictive habitat distribution models in ecology", *Ecological Modelling* 135, 147–186.

Heikkinen, R. K., Luoto, M., Araujo, M. B., Virkkala, R., Thuiller W.& Sykes M. T. (2006), "Methods and uncertainties in bioclimatic envelope modelling under climate change", *Progress in Physical Geography* **30**(6), 751-777.

Johnson, C.J. & Gillingham, M.P. (2005), "An evaluation of mapped species distribution models used for conservation Planning", *Environmental Conservation* **32** (2), 117–128.

Kumari, P., Joshi, G.C. & Tewari, L.M. (2011), "Diversity and status of ethno-medicinal plants of Almora district in Uttarakhand, India", *International Journal of Biodiversity and Conservation* **3**(7), 298-326.

Landis, J.R. & Koch, G.G. (1977), « The measurements of observer agreement for categorical data", *Biometrics* 33, 159–174.

Luck, G.W. (2002), "The habitat requirements of the rufous treecreeper (*Climacteris rufa*). 2. Validating predictive habitat models", *Biological Conservation* 105, 395-403.

Manel, S., Williams, H.C. & Ormerod, S.J. (2001), "Evaluating presence-absence models in ecology: the need to account for prevalence", *Journal of Applied Ecology* 38, 921 -931.

Mendoza-Gonzalez, G., Martinez, M.L., Rojas-Soto, O.R., Vazquez, G. & Gallego-Fernandez, J.B. (2013), "Ecological niche modeling of coastal dune plants and future potential distribution in response to climate change and sea level rise", *Global Change Biology* 19, 2524-2535.

Mouton, A.M., De Baets, B. & Goethals, P.L.M. (2010), "Ecological relevance of performance criteria for specific distribution models", *Ecological Modelling* 221, 1995-2002.

Nix, H.A. (1986), "A biogeographic analysis of Australian elapid snakes. InAtlas of elapid snakes of Australia", (ed.) Longmore,R.Australian Government Publishing Service, Canberra.

Peterson, A.T. (2001), "Predicting species geographic distributions based on ecological niche modeling", *Condor* 103, 599–605.

Peterson, A.T. (2003), "Predicting the geography of species invasions via ecological niche modeling", *Quarterly Review of Biology* 78, 419-433.

Piñeiro, R., Aguilar, J. F., Munt, D.D. & Feliner, G. N. (2007), "Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub *Armeriapungens* (Plumbaginaceae)", *Molecular Ecology* 16, 2155-2171.

Robertson, E.A., Zweig, M.H. & Van Steirteghem, M.D. (1983), "Evaluating the clinical accuracy of laboratory tests", *American Journal of Clinical Pathology* 79, 78–86.

Rushton , S.P., Oremerod, S.J. & Kerby, G. (2004), "New paradigms for modeling species distribution?" *Journal of Applied Ecology* 41, 193-200.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola A.J. & Williamson, R.C. (2001), "Estimating the support of a high-dimensional distribution", *Neural Computation* 13, 1443-1471.

Schölkopf, B., Smola, A., Williamson, R. & Bartlett, P.L. (2000), "New support vector algorithms", *Neural Computation* 12, 1207-1245.

Semwal, R.L., Nautiyal, S., Sen, K.K., Rana, U., Maikhuri, R.K., Rao, K.S. & Saxena, K.G. (2004), "Patterns and ecological implications of agricultural land-use changes: a case study from Central Himalaya, India", *Agriculture Ecosystem and Environment* 102, 81–92.

Stockwell, D.R.B. (1999), "Genetic algorithms II. In: Machine learning methods for ecological applications", (ed.) A. H. Fielding.Kluwer Academic Publishers, Boston.

Stockwell, D.R.B. & Noble, I.R. (1992), "Induction of sets of rules from animal distribution data: A robust and informative method of analysis", *Mathematics and Computers in Simulation* 33, 385-390.

Stockwell, D. & Peters, D. (1999), "The GARP modelling system: problems and solutions to automated spatial prediction", *International Journal of Geographical Information Science* 13, 143–158.

Stockwell, D.R.B. & Peterson, A.T. (2002), "Effects of sample size on accuracy of species distribution models", *Ecological Modelling* 148, 1-13.

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

Thuiller, W. (2003), "BIOMOD: optimising predictions ofspecies distributions and projecting potential future shiftsunder global change", *Global Change Biology* 9, 1353–1362.

Thuiller, W. (2004), "Patterns and uncertainties of species' rangeshifts under climate change", *Global Change Biology* 10, 2020–2027.

Thuiller, W., Richardson D.M., Pysek, 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 Biology* 11, 2234-2250.

Thuiller, W., Araujo, M.B. & Lavorel, S. (2003), « Generalized models vs. Classification tree analysis: predicting spatial distributions of plant species at different scales", *Journal of Vegetation Science* 14, 669-680.

Tiwari, P.C. (2000), "Land-use changes in Himalaya and their impact on the plains ecosystem: need for sustainable land use", *Land Use Policy* **17**(2),101-111.

Uniyal, S.K., Awasthi, A. & Rawat, G.S., (2002), "Traditional and ethnobotanical uses of plants in Bhagirathi Valley (Western Himalayas)", *Indian Journal of Taditional Knowledge* **1**(1), 7-19.

Vapnik, V. (1995), "The Nature of Statistical Learning Theory", Springer. New York.

Woolf, A., Nielsen, C.K ., Weber, T. & Gibbs-Kieninger, T.J. (2002). "Statewide modeling of bobcat, *Lynx rufus*, habitat in Illinois, USA", *Biological Conservation* 104, 191-198.