# Maxent Modeling of *Myrica esculenta* for Estimating Geographical Distribution in Kumaun Himalayas, Uttarakhand

**Rajeev Shankhwar, M.S. Bhandari, R.K. Meena and H.S. Ginwal**
Division of Genetics & Tree Propagation, FRI, Dehardun - 248195, Uttarakhand, India

**Abstract**

The forest of Uttarakhand has species of medicinal and economic importance known as *Myrica esculenta* (*M. esculenta*) (Vern. *Kafal*). The evergreen small tree found in mid-Himalayas with an elevation ranged from 1500-2100 m amsl. The enormous anthropogenic pressure along with habitat degradation and habitat fragmentation resulted in the declining population of this species. Under Forest Genetics Resource (FGR) conservation and management prospective, the geographical mapping of *M. esculenta* becomes utmost priority in conservation genetics.

The survey was conducted during 2016-2017 and geo-coordinates were recorded through Global Position System (GPS) in different forest area of Almora, Bageshwar, Champawat, Nainital and Pithoragarh districts of Kumaun Himalayas. The Maximum Entropy Modeling with a set of 19-Bioclimatic variables [1] was used for estimating the distribution of *M. esculenta*.

We use 60 % well distributed geo-coordinates in the Maxent model for predicting distribution of *M. esculenta.* Remaining 40 % geo-coordinates were used for validation of model. It has been found that 78.13 % of geo-coordinates were overlaid on the actual area of occurrence of *M. esculenta.* According to Jackknife test environmental variables: Temperature Seasonality (18.4%) and Altitude (17.6%) showed highest contribution for Maxent modeling to predict the distribution of *M. esculenta* in Kumaun Himalayas. It could be concluded that Maxent model is highly suitable for habitat distribution and prediction of *M. esculenta* in Kumaun Himalayas.

Keywords: Forest Genetics Resource, Kumaun Himalayas, Maximum Entropy Modeling

## 1. Introduction

Forest Genetics Resources (FGRs) spread in the form of trees, scrubs and other non-woody plant species, are extremely important to environmental prospective, have economic value besides fulfilling social needs. FGRs play an important role in environmental processes, like climate balancing, reduce soil erosion, conserving water in catchments for tributaries & rivulets, and protect wildlife, ensuring food and health security for rapidly increasing human population and livestock. Sustainability of all resources depended on the availability of rich diversity maintained between and within the species. The high level of genetic diversity ensures that forest trees can survive, adapt and evolve under changing climatic conditions [1].

In the last 6-7 decades, human's population rise at exponential rate and developed industrial resources which rapidly destruct forests. The exploitation of FGRs creates more pressure on forest and many biological species have been disappearing, and rest comes under threat zone. According to The State of the World's Forest Genetic Resources (FAO, 2014), roughly half of the forest species were found to be threatened or subject to genetic erosion [2].

The utmost priority is to conserve and protect the species under FGRs. The distributions of FGRs need to be tapped and recorded of continuously. The Remote Sensing and Geographical Information System (RS & GIS) based technology is fully utilized for species distribution mapping. The technology is very reliable and authentic which could help in programming priority areas of conservation and help in achieving Sustainable Developmental Goals (SDG) from FGRs. *M. esculenta* is one such species that need conservation and which is commonly found in sub-temperate to sub-tropical region of Kumaun Himalayas. The *M. esculenta* occurs under canopy covers of *Pinus roxburghii* (Chir pine), *Quercus leucotrichophora* (Banj Oak), *Rhododendron arboreum* (Burans), *Lyonia ovalifolia* (Ayaar) in a scattered manner. Due to canopy covers, spectral reflectance of *M. esculenta* is not easy to measure using traditional multispectral satellite data. The satellite imagery from LISS-3, LISS-4, SPOT and LANDSAT data series is not august well to geographically map species apart from Hyperspectral RS. The species wise mapping need high spatial resolution with more spectral bands and such combination is not present in current data series. Therefore, Maxent model is used in the present study to know the actual area, estimated area and habitat suitability of species under different agro-climatic zones of Kumaun Himalayas.

## 2. Material and Methods

### 2.1. Study Area

The study was carried out in five districts (Almora, Bageshwar, Champawat, Nainital and Pithoragarh) of Kumaun Forest Division, Uttarakhand (Figure 1). The study area lies between 29°03'20.71" N to 30°29'21.76" N latitude and 79°03'59.53" E to 81°01'14.41" E longitude with an elevation range from 250 to 7415 m amsl. The total geographical area of the five districts is 18,492 $Km^2$. The forest cover of Almora is 1,583 $km^2$; Bageshwar is 1,363 $Km^2$; Champawat is 1,184 $Km^2$; Nainital is 3,004 $Km^2$ & Pithoragarh is 2,102 $Km^2$ (FSI- ISFR-2015).

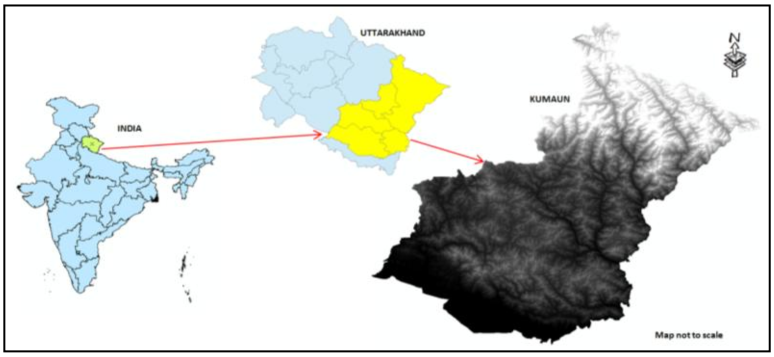



Figure 1. Geographical location of study area

### 2.2. Geo-spatial Data for *M. esculenta*

Geo-coordinates were recorded for *M. esculenta* populations with the help of Geographical Positioning System (GPS) Garmin Oregon 650 model. A total 400 coordinates of *M. esculenta* population were taken in five districts. Geo-coordinates were recorded in degree decimal format with a positional accuracy of 5 to 8 m. The 300 m distance was maintained between two GPS coordinates. Geo-coordinates were also collected for scattered trees all along the motor roads, in addition to trekking in dense to moderate forests. The field survey provides different slope, aspect and elevation range of species. The variation in topological factors is helpful for species distribution modeling. Sixty percent well distributed geo-coordinates were used in the Maxent model for estimating and predicting distribution of *M. esculenta*. The remaining forty percent geo-coordinates were used for validation of model.

## 2.3. Environmental variables

Different bioclimatic variables are derived from the monthly temperature and rainfall values (Table 1). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation), seasonality (e.g., annual range in temperature and precipitation), and extreme or limiting environmental factors (e.g., temperature of the coldest & warmest month and precipitation of the wet and dry quarters). A quarter is a period of three months (1/4$^{th}$ of the year). In the present study, bioclimatic variables (Hijmans *et al*., 2005), 30 arc-seconds resolution World-Clim data version 1.4: Current condition (1960-1990) had been used [3]. To generate aspect, elevation and slope; GIS based data layers, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Global Digital Elevation Model (GDEM) with 30 m resolutions was used [4].

Table 1. Nineteen Bioclimatic variables used for species distribution modeling

| Label | Variable | Scaling Factor | Units |
|---|---|---|---|
| Bio 1 | Annual Mean Temperature | 10 | Degrees Celsius |
| Bio 2 | Annual Mean Diurnal Range | 10 | Degrees Celsius |
| Bio 3 | Iso-thermality | 100 | Degrees Celsius |
| Bio 4s | Temperature Seasonality (Standard Deviation) | 100 | Degrees Celsius |
| Bio 5 | Max Temperature of Warmest Month | 10 | Degrees Celsius |
| Bio 6 | Min Temperature of Coldest Month | 10 | Degrees Celsius |
| Bio 7 | Annual Temperature Range | 10 | Degrees Celsius |
| Bio 8 | Mean Temperature of Wettest Quarter | 10 | Degrees Celsius |
| Bio 9 | Mean Temperature of Driest Quarter | 10 | Degrees Celsius |
| Bio 10 | Mean Temperature of Warmest Quarter | 10 | Degrees Celsius |
| Bio 11 | Mean Temperature of Coldest Quarter | 10 | Degrees Celsius |
| Bio 12 | Annual Precipitation | 1 | Millimeters |
| Bio 13 | Precipitation of Wettest Month | 1 | Millimeters |
| Bio 14 | Precipitation of Driest Month | 1 | Millimeters |
| Bio 15 | Precipitation Seasonality (CV) | 100 | Millimeters |
| Bio 16 | Precipitation of Wettest Quarter | 1 | Millimeters |
| Bio 17 | Precipitation of Driest Quarter | 1 | Millimeters |
| Bio 18 | Precipitation of Warmest Quarter | 1 | Millimeters |
| Bio 19 | Precipitation of Coldest Quarter | 1 | Millimeters |

## 2.4. Categorical Data

### 2.4.1. Forest Cover Map (FCM)

The forest cover density thematic map generated using LANDSAT-8 satellite image (path–row: 145- 40 and 145-39) with 30 m spatial resolution. The image was classified into 2 major classes i.e., forest and non-forest (Agriculture, Barren land, Settlements, scrub land and water body).

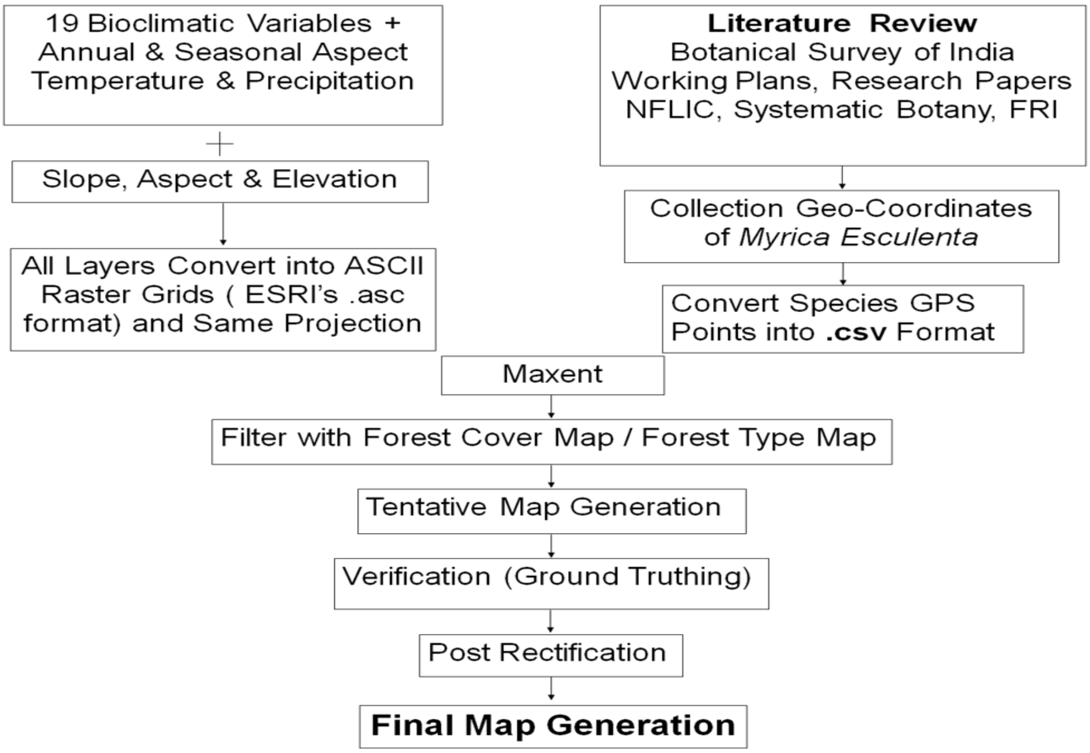
### 2.4.2. Forest Type Map (FTM)

Forest Type Map (FTM) of India Atlas 2011 prepared by Forest Survey of India was used. Forest Thematic Mapper (FTM) available at 1:50000 scale in which 178 forest types were mapped out of 200 Forest Types (FT) described by Champion & Seth (1968). Further, categorical data layers resample (using nearest neighbor approach) into 1 Km spatial resolution to match bioclimatic variable resolution.

## 2.5 Maxent Model

Maximum Entropy approach for estimating distribution and habitat suitability of a species based on model run by Maxent software uses a set of 19 Bioclimatic variables. Species Geo- coordinates, ASTER- GDEM, categorical variables (FCM) and certain parameters like replicates run 100 (Flory *et al*., 2012) allows model to run multiple times and averaging the results from all models. For reducing model over fitting regularization, multiplier value set be 0.1 (Phillips *et al*., 2004), number of iterations set be 5000 (Young *et al.*, 2011), as it allows model to have sufficient time for convergence to avoid model over-prediction. Other values were kept as default. On basis of different variables Maxent predicts the high value for grid cell in the data which is suitable for species estimated distribution. After model output came, masking tool applied on output using FCM. Masking was essential because model prediction came over non-forest area also, which decreases the results accuracy. By applying masking; we eliminate the over-predicted area and increase the distribution accuracy of a species. The mapping scheme of *M. esculenta* through Maxent Modeling is shown in Figure 2.

Figure 2. Geographical Distribution Mapping of *Myrica esculenta* through Maxent Modeling




## 3. Results and Discussions

### 3.1. Analysis of Receiver Operating Characteristic (ROC) Curves

Sensitivity (1-Omission rate) vs 1–Specificity is area under the ROC Curve or AUC (Figure 3). The graph is formulated with sensitivity on the (vertical) axis and 1–specificity on the x (horizontal) axis. The area in the uppermost left region provides the most useful discrimination in terms of accuracy of training set modeled for the prediction. The average test AUC for the replicate runs is 0.952. The high value of AUC means that all possible observations used in the modelling are measured very accurately. The ROC curve is steep (exponential) and above the diagonal line. The categorized models with values >0.9 are highly accurate for prediction modeling.

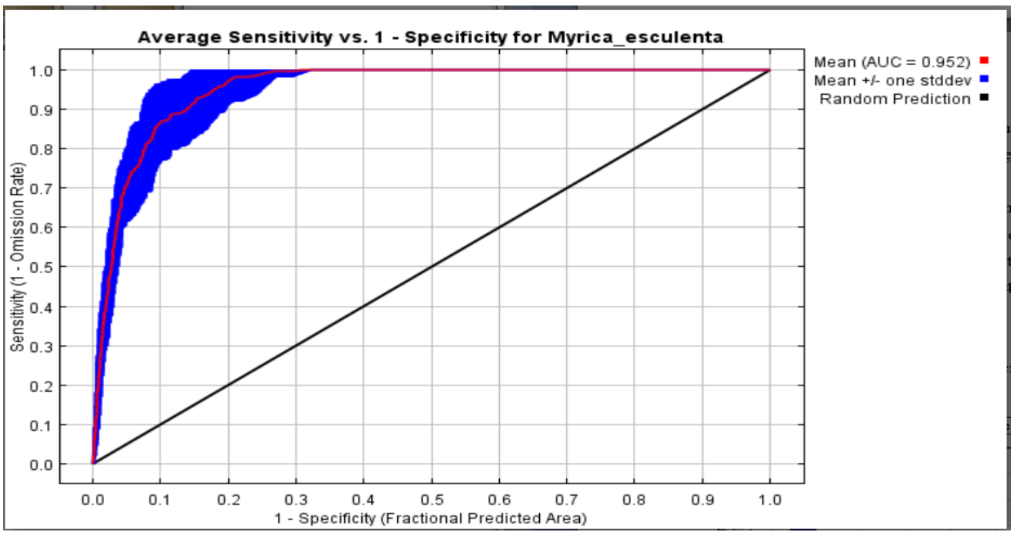



Figure 3. ROC curve shows, the average test AUC for the replicate runs is 0.952

### 3.2. Analysis of Variable Contributions

The estimates of relative contributions of each environmental variable to the Maxent model were shown in Table 2. Higher the contribution, the more impact that particular variable has on predicting the occurrence of *M. esculenta*. According to Jackknife test environmental variables: Temperature Seasonality (18.4 %) and Altitude (17.6 %) showed highest contribution.

Table 2. Estimates of relative contribution of environmental variables to the Maxent model

| Label | Variable | Percent Contribution | Permutation Importance |
|---|---|---|---|
| Bio 4 | Temperature Seasonality (Std. Deviation) | 18.4 | 0.6 |
| Alt | Altitude | 17.6 | 1.4 |
| Bio 6 | Min Temperature of Coldest Month | 8.2 | 1.4 |
| Bio 15 | Precipitation Seasonality (CV) | 7.7 | 15.2 |
| Bio 16 | Precipitation of Wettest Quarter | 7.3 | 11.1 |
| Bio 7 | Annual Temperature Range | 6.6 | 0.1 |
| Bio 2 | Annual Mean Diurnal Range | 6.2 | 3.6 |
| Bio 13 | Precipitation of Wettest Month | 6 | 26.6 |
| Bio 18 | Precipitation of Warmest Quarter | 5.3 | 10.8 |
| Asp | Aspect | 4.5 | 0.2 |
| Bio 14 | Precipitation of Driest Month | 3.3 | 2.2 |
| Bio 19 | Precipitation of Coldest Quarter | 3.1 | 6.3 |
| Bio 17 | Precipitation of Driest Quarter | 2.6 | 5.7 |
| Bio 11 | Mean Temperature of Coldest Quarter | 0.8 | 0.4 |
| Bio 3 | Iso-thermality | 0.8 | 0.2 |
| Bio 12 | Annual Precipitation | 0.7 | 1.7 |
| Bio 9 | Mean Temperature of Driest Quarter | 0.6 | 0.6 |
| Bio 5 | Max Temperature of Warmest Month | 0.2 | 11.6 |

### 3.3. Jackknife Test of Variable Importance

The Jackknife shows the training gain of each Bioclimatic variable and compares it to the training gain with all the variables (Figure 4). The environmental variable with the highest gain when used in isolation is Altitude, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain most, when it is omitted is bio14 and therefore appears to have the most information which is not present in the other variables. This environmental variable allows us a reasonably good fit to the training data. We see that if Maxent uses only asp, it achieves almost no gain, so that variable is not (by itself) useful for estimating the distribution of *M. esculenta*.

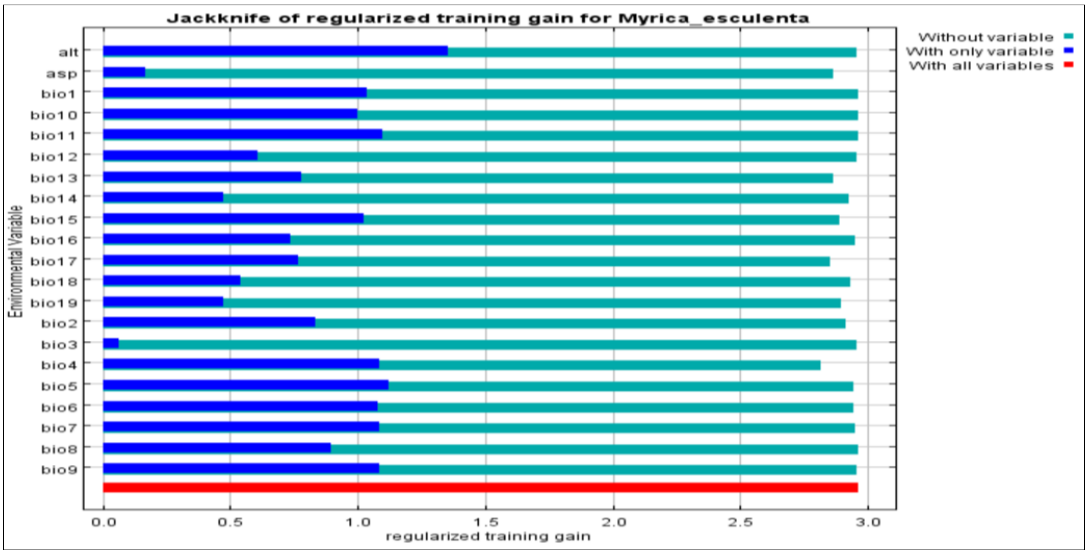



Figure 4. Jackknife of regularized training gain for *M. esculenta*

### 3.4. Distribution of *M. esculenta* in Kumaun Himalayas

On the basis of RS & GIS technology, the total estimated area of *M. esculenta* in Kumaun Himalayas is 2789.58 Km$^2$ (Table 3.). In all five districts of Kumaun Himalayas, Almora district showed highest distribution of *M. esculenta* (780.56 Km$^2$) followed by Nainital (637.90 km$^2$), Bageshwar (543.89 km$^2$), Pithoragarh (481.96 Km$^2$) and Champawat (345.27 km$^2$).

Table 3. Estimated area under *M. esculenta* in different districts of Kumaun Himalayas through Maxent Modeling.

| Districts | Geo-coordinates Recorded Areas | Geographical Area ($Km^2$) | Forest Cover ($Km^2$) | Estimated Area ($Km^2$) | Estimated area % in respect to geographical area | Estimated area % in respect to forest cover |
|---|---|---|---|---|---|---|
| Almora | Binsar Wildlife Sanctuary, Shitla-Khet, Ranikhet to Bhatroajkhan, Dunagiri, Kukchina | 3,139 | 1,583 | 780.56 | 24.87 | 49.31 |
| Bageshwar | Kausani | 2,246 | 1,363 | 543.89 | 24.22 | 39.90 |
| Champawat | Chirapani, Munch | 1,766 | 1,184 | 345.27 | 19.55 | 29.16 |
| Nainital | Bhowali | 4,251 | 3,004 | 637.90 | 15.01 | 21.24 |
| Pithoragarh | Kaamlekh Dev, Dhula | 7,090 | 2,102 | 481.96 | 6.80 | 22.93 |
| Grand Total | | **18,492** | **9,236** | **2789.58** | **15.09** | **30.20** |

(Source: Geographical Area and Forest Cover: Forest Survey of India Report, 2015)

More detailed *M. esculenta* analysis with different elevation zones and forest cover density classes [Very Dense Forest (VDF), Moderate Dense Forest (MDF) and Open Forest (OF)] were carried out and shown in Figure 5.

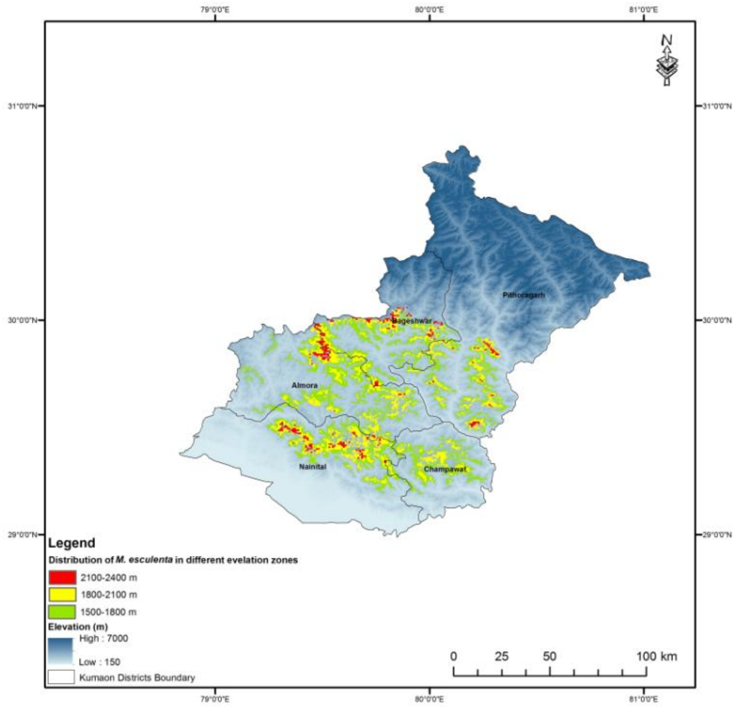



Figure 5. Estimated distribution of *M. esculenta* with different elevation zone in Kumaun Himalayas through Maxent Modeling

Estimated distribution of *M. esculenta* in Forest Covers Classes at different elevation zones were shown in Table 4. In our study, *M. esculenta* present in scattered form with an elevation ranged from 1500-1800 m amsl and covered 1576.06 $Km^2$ areas under the canopy cover of largely dominated *Pinus roxburghii* and *Quercus leucotrichophora* forests (Singh *et al*., 2016).

Elevation zone (1800-2100 m) largely covered by *Quercus leucotrichophora* (Singh *et al*., 2016), where our estimated distribution is 1037.40 $Km^2$ and under VDF is 289.77 $Km^2$, MDF is 597.93 $Km^2$ and OF is 149.70 $Km^2$, respectively.

In case of elevation zone (2100-2400 m), Maxent model predicted area of 176.12 $Km^2$ which is very low as compared to elevation zone ranged from 1500-2100 m amsl. This elevation zone is mixed with upper limit of *Quercus sp.* forest and broadleaved species (Singh *et al*., 2016). In this elevation range, *M. esculenta* found along with *Rhododendron arboreum and Lyonia ovalifolia*. The study suggested that with increasing altitude, the *M. esculenta* declines.

Table 4. Estimated distribution of *M. esculenta* in Very Dense, Moderately Dense and Open Forest classes in different elevation zones.

| **Elevation Zone (m)** | **Forest Classes** | **Area under Forest Classes ($Km^2$)** | **Total Area ($Km^2$)** |
|---|---|---|---|
| 1500-1800 | Very Dense Forest | 319.50 | **1576.06** |
| | Moderately Dense Forest | 957.32 | |
| | Open Forest | 299.24 | |
| 1800-2100 | Very Dense Forest | 289.77 | **1037.40** |
| | Moderately Dense Forest | 597.93 | |
| | Open Forest | 149.70 | |
| 2100-2400 | Very Dense Forest | 57.09 | **176.12** |
| | Moderately Dense Forest | 97.65 | |
| | Open Forest | 21.38 | |
| Total Area | | **2789.58** | **2789.58** |

(Source: ASTER GDEM)

**4. Conclusions**

Maxent model output provided acceptable result with AUC value of 95.2. 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). Geo-coordinates (Green in colour) used to run model and geo-coordinates (Pink in colour) used to validate model output. It has been found that, 78.13% geo-coordinates overlaid on the actual area of *M. esculenta* occurrence in Kumaun Himalayas. Similar to our study spatial data has been used in 70:30 ratio for training and testing in *Justicia adhatoda* (Yanga *et al*., 2013), 75:25 ratio for training and testing in *Ilex khasiana Purk* (Adhikari *et al*., 2012). The total geographical area of five districts is 18,492 $Km^2$. Estimated area through Maxent model is 2789.58 $Km^2$ which includes actual area of 309.08 $Km^2$ under *M. esculenta*. Highest area of *M. esculenta* occurs in the elevation zone from 1500-1800 m and decline sharply above 2200 m. Due to lopping and fire in *Pinus roxburghii* forest, *M. esculenta* spread in scattered manner and under grows in canopy cover of others trees. Our study provides the first eco-distribution map of *M. esculenta* in Kumaun Himalayas which will be helpful in better planning, conservation and management of this species. It could be concluded that Maxent model is very useful for predicting and estimating distribution of those species which are highly scattered, present in mixed forests and grows under canopy cover of other forest species.

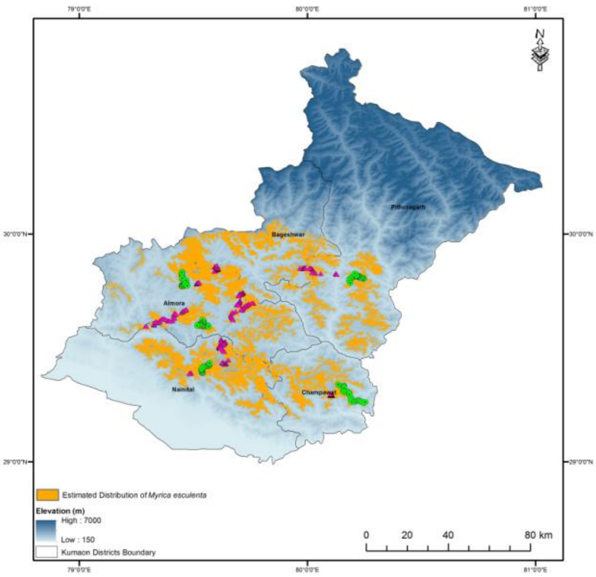



Figure 6. Geo-coordinates overlaid for validation in Maxent model output map.

**Acknowledgement**

The financial support by the Ministry of Environment, Forest and Climate change (MoEF&CC), Government of India, New Delhi is gratefully acknowledged.

**References**

Anonymous., 2011. Forest type Atlas, Forest Survey of India Dehradun, India.

Anonymous., 2015. India State Forest Report, Forest Survey of India Dehradun, India.

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. Ecological Engineering, 40, pp. 37 - 43.

Dwivedi, P., Kushwaha, S.P.S., Nandy, S., Rathore, C.S., 2006. Site suitability analysis for khair (*acacia catechu*) in part of Doon valley using Geoinformatics. Journal of the Indian Society of Remote Sensing, Vol. 34, No. 2.

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.

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. Journal of Applied Ecology, 49, pp. 680-689.

Giriraj, A., Irfan-Ullah, M., Ramesh, B.R., Karunakaran, P.V., Murthy, S.R., 2008. Mapping the potential distribution of *Rhododendron arboreum* Sm. spp. *nilagiricum* (Zenker) Tagg (*Ericacease*), an endemic plant using Ecological niche modeling. Current Science, Vol. 94 No. 12.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, pp. 1965– 1978.

Jane, E., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2010. A statistical explanation of MaxEnt for ecologists Diversity and Distributions.

Kumar, S., Stohlgren, T. J., 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree *Canacomyrica monticola* in New Caledonia. Journal of Ecology and Natural Environment, Vol. 1(4), pp. 094-098.

Mcpherson, J.M., Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. Ecography 30, 135–151.

Phillips, S.J., Miroslav, D., Schapire, R.E., 2004. Maxent Software for Species Distribution Modeling. http://cs.princeton.edu/~schapire/Maxent/.

Singh, G., Padalia, H., Rai, I.D., Bharti, R.R., Rawat, G.S., 2016. Spatial extent and conservation status of Banj Oak (*Quercus leucotrichophora* A. camus) forests in Uttarakhand, western Himalaya. Tropical Ecology, 57(2), pp. 255-262.

Yanga, X.Q., Kushwaha, S.P.S., Saran, S., Jianchu Xuc, Roy, P.S., 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. Ecological Engineering, 51, pp. 83– 87.

Young, N., Carter, L., Evangelista, P. 2011. A MaxEnt Model v3.3.3e Tutorial (ArcGIS v10).

1. URL. 2017 Download http://www.fao.org/forestry/fgr/en.
2. URL. 2017 Download http://www.euforgen.org/forest-genetic-resources/forest-management/
3. URL. 2017 Download http://ccafs-climate.org/downloads/docs/bioclimatic-variables.pdf
4. URL. 2017 Download https://reverb.echo.nasa.gov/reverb.