
Jyotishman Deka1, Sanjeeb Bharali1, Anup Kr Das1, Om Prakash Tripathi1 and Mohamed Latif Khan1

1Department of Forestry, North Eastern Regional Institute of Science and Technology, Itanagar, 791109, Arunachal Pradesh, India.

Abstract

Pinus merkusii is a vulnerable gymnosperm and occurs mainly in the eastern part of Anjaw district of Arunachal Pradesh, northeastern India. However, rapid extraction from its natural habitat had led to its population decline and hence considered vulnerable. Documentation and distribution of the species in this part of the country is lacking till date. As such in order to overcome the constrictions of the region’s remoteness and rugged terrain for field surveys, the present study utilized the species occurrence information and remotely sensed environmental variables to generate the distribution map of P. merkusii and its validation. The AUC of 0.979 and 0.972 observed for the training and test data, respectively, indicated higher success rates. This study revealed that the Enhanced Vegetation Index (EVI) proposed by the MODIS play an important role in providing dependable spatial and temporal information of vegetation cover and can be suitably applied to species distribution modeling. However, multicollinearity among the predictor variables which is commonly associated with remotely sensed data sets should be emphasized. Of the total 10 predictor variables used to build our model, evi_081 appeared to be the strongest predictors. The study also focuses on the potentiality of Maximum entropy (Maxent) modeling for identifying species distributions. The highly distributed P. merkusii area class was found to be 130 km²; medium 354 km² and low 847 km². It was further hoped that the results drawn in the current study will acts an effective tool for monitoring as well in preparation of effective conservation planning of the species in this particular region.

Key words: Enhanced vegetation index, MODIS, Pinus merkusii, Species distribution, Vulnerable.

1. Introduction

Pinus merkusii Jungh et de Vries is the only surviving member of the genus that naturally extends its distribution across the south equator. In India, it occurs mainly in the eastern part of Anjau district of Arunachal Pradesh which is neighboring to the India–Myanmar Pine Forests Eco-region. It is the dominant tree species in this region. However, in recent year’s unsustainable extraction for timber, deforestation, clear felling, logging etc. have built up the pressure on the habitat of the species and as a result population is decreasing significantly (Bharali et al., 2012). With the increasing awareness towards the importance of Himalayan biodiversity and alarming rate at which they are being exploited from natural habitats, there is a need to initiate various conservation and management strategies to mitigate such uncontrolled resource exploitation (Ray et al., 2011). As such the various key features such as species geographic distribution, its habitat and ecological characteristic, must be studied properly for the species protection / restoration activities.

Locating populations of rare, endangered, and threatened species is an important consideration in biodiversity conservation and for developing effective strategies for both *in situ* and *ex situ* conservation (Menon *et al*., 2010). With the advancement of remote sensing and GIS technologies, use of ecological niche models (ENM) to generate potential geographic distributions of species has rapidly increased in ecology, conservation and evolutionary biology (Terribile *et al*., 2010). The most common strategy for estimating the actual or potential geographic distribution of a species is to characterize the environmental conditions that are suitable for the species, and identify where suitable environments are distributed in space for any particular species (Pearson, 2007). ENM establishes relationships between occurrences of species with biophysical and environmental conditions (Soberon and Peterson, 2005; Kumar and Stohlgren, 2009). A large number of species distribution modeling methods are now available to predict potential distribution for a species but the basic constrain with most of the model is that they are sensitive to sample size. In such case with limited sample size, Maxent species distribution models were proven to be advantageous over the methods (Philips *et al*., 2006). Maxent is a bioclimatic model which deals with presence only data and has a number of other aspects that make it well-suited for species distribution modeling (Chitale and Behera, 2012). As such a large number of researchers worldwide have implemented Maxent for predicting the distribution of endangered and threatened species (Elith *et al*., 2007; Philips *et al*., 2006; Yost *et al*., 2008; Kumar and Stohlgren, 2009). Further free availability of high to low resolution satellite imageries, environmental variables databases and interpolated spatial datasets on climate and vegetation from public domain has enhanced the accuracy of prediction of these models manifold (Adhakari *et al*., 2012).

In the current study, we used species occurrence records, GIS, environmental layer (Enhanced vegetation index) and the maximum entropy distribution modeling approach Phillips *et al*., (2008) to predict the potential distribution for this vulnerable gymnosperm in order to develop the strategic planning for its conservation.

### 2. Methodology

#### 2.1 Study site

The study was carried out in Anjaw district of Arunachal Pradesh. The district is located between 26°55' and 28°40' N latitudes and between 92°40' and 94°21' E longitudes and with an altitudinal range of 1000-2000 m asl (Figure 1). It is surrounded by China on north, Myanmar is on the eastern part of the district, and Lower Dirang Valley district is in the west and Lohit district in the South. The major rivers are Lohit, Delei and Tellu. The vegetation of the district falls under four broad climatic categories and classified in five broad forest types with a sixth type of secondary forests. These are tropical forests, sub-tropical forests, pine forests, temperate forests and alpine forests (Kaul and Haridasan, 1987). The climate is sub-tropical, wet and highly humid in the lower elevations and the valleys; and cold in the higher elevations. Annual rainfall in the district varies from 3500 to 5500 mm. Soil type was generally sandy and acidic in nature.

#### 2.2 Selected species and species records

*P. merkusii*, is the only surviving member of the genus that naturally extends its distribution across the south equator (Figure 2). It is a medium to large size tree, grows up to 70 m height making it one of the world’s tallest pines, with a trunk diameter up to 1 m and pyramidal to conical crown on young age, and flatter and spreading on old age. The needles are
very slender, rigid, straight 15-25 cm long and less than 1 mm thick, green to yellow color and are found in a pair of two. It flowers during March-June and the seeds ripen during October-November of the following year. P. merkusii has a large altitudinal range from a few meters to over 1800 m elevation (Cooling, 1968).

We obtained 30 occurrence records of P. merkusii tree species from field visit. These records were considered as the total known distribution for the species. All the coordinated were collected in degree decimal and converted to point vector file for modeling the distribution of the species.

2.3 Environmental variables and distribution modeling

In the present study, twenty three layers of 16 day composite MODIS EVI images (MOD13Q1) with a spatial resolution of 250 m were downloaded from Oak Ridge National Laboratory Distributed Active Archive Centre (http://daac.ornl.gov/MODIS/modis.html) and were used to characterize environments across the region. The images were downloaded in Geo-tiff format and converted to ASCII raster grids. These layers correspond to the year 2011 during which the field survey was conducted. EVI (Enhanced vegetation index) has been considered as the modified NDVI with improved sensitivity to high biomass regions and improved vegetation monitoring capability (Huete et al., 1999). In order to avoid cloud cover and other spurious effects the MODIS EVI products are generated by compositing daily data every 16 days, resulting in 23 composites per year (Huete et al., 2002). We did not use elevation layers as their influence on species distribution as P. merkusii has a large altitudinal range from a few meters to over 1800 m elevation (Cooling, 1968). As MODIS EVI dataset has been proved to be quite sensitive to the variations in topographic conditions topographic variables such as slope and aspect was also not included in the model (Matsushita et al., 2007; Adhakari et al., 2012). Further multicollinearity among the predictor variables which is commonly associated with remotely sensed data sets can be a source of inaccurate results and hence should be avoided (Dormann et al., 2007). In order to identify the multicollinearity among of layers (r > 0.9), all the 23 EVI layers were then subjected to correlation test. From each of the highly correlated pair of layers one of the EVI layers was then excluded from the ENM analysis. Finally, 10 environmental data layers were selected for habitat distribution modelling.

2.4 Model validation

ENMs were built using the maximum entropy niche modeling approach implemented in the software Maxent v3.3.3k (http://www.cs.princeton.edu/~schapire/maxent) as it estimates the probability distribution for a species occurrence based on environmental constraints and is applicable with presence-only data (Philips et al., 2006). Of the total records, seventy five percent were used for model training and twenty five percent for testing. To provide a test of model predictability, jackknife-based approach presented by Pearson et al. (2007) was applied, which is designed explicitly for situations of small sample size (Menon et al., 2010). The area under the receiver operating characteristic curve (AUC) value was evaluated and the model was graded 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) following Thuiller et al. (2005).

3. Results and Discussion

Maxent model was run with the following configurations: Hinge features, perform jackknife tests, logistic output format with a threshold rule of 10 percentile training presence. As the program is already calibrated on a wide range of species datasets, other parameters were set to default. Hinge features can effectively replace quadratic, product and threshold features and was found suitable with any presence-only dataset and are used by default in the software whenever there are at least 15 presence sites (Phillips and Dudík, 2008).

The model calibration test for P. merkusii generated acceptable results, AUC train = 0.979 and AUC test = 0.972 with standard deviation of 0.007 (Figure 3). The environmental variable with highest gain when used in isolation is evi_081 and contributed 23.4% to the MaxEnt mode which therefore appears to have the most useful information. The environmental variable that decreases the gain the most when it is omitted is evi_273, which therefore appears to have the most information that isn't present in the other variables. It was found from the literature and also from the field survey that the layers (evi_081 and evi_273) correspond to the period of flowering and fruiting phage of the species. Sensitivity of EVI layers towards flowering and fruiting phage of the species Ilex khasiana Purk., was also

Figure 3. (a) Jackknife of regularized training gain for *P. merkusii* (b) Sensitivity vs 1- Specificity for *P. merkusii*.

reported by Adhikari *et al.* (2012). Menon *et al.* (2010) also used 11 EVI layers to characterize environments variables for predicting the distribution of *Gymnocladus assamicus* in Tawang and West Kameng districts of Arunachal Pradesh. Waring *et al.* (2006) reported that EVI has more advantages over other vegetation index in that it is sensitive to site-specific variations and may be more sensitive to local variation in canopy leaf area and chlorophyll concentrations. This shows that the EVI proposed by the MODIS play an important role in providing dependable spatial and temporal information of vegetation cover and can be suitably applied to species distribution modeling. Other layers of EVI together contributed 76.6% to the distribution model of *P. merkusii* of which evi_273, evi_193 and evi_017 contributed 23.8%, 21.4% and 15.9% respectively. Approximations of percent contributions and permutation importance of different environmental variables are given in Table 1.

The Maxent model predicted potential distribution for *P. merkusii* with high success rates and are well agreed with the known distribution of the species (Figure 4). Most suitable habitat for *P. merkusii* was predicted in the north eastern parts of the district i.e., Kibithoo and Walong, forming almost a continuous belt and few other scattered areas towards Hawaii, Goiliang and Chaglagam. Using arbitrarily defined scale we grouped the probability values into four classes as high (0.6-1.0), medium (0.3-0.6), small (0.1-0.3) and no distribution (0.0-0.1). The highly distributed *P. merkusii* area class was found to be 130 km$^2$; medium 354 km$^2$ and low 847 km$^2$. Thus study showed that ENMs can be used to predict the potential distribution of species in areas with relatively few available data (Stockwell and Peterson, (2002); Pearson, 2007; Babar *et al.*, 2012).

This study makes available the first predicted potential distribution map for a plant species (*P. merkusii*) in India.
Table 1. Percent contributions and permutation importance of different environmental variables.

<table>
<thead>
<tr>
<th>Environmental Variable</th>
<th>Percent contribution</th>
<th>Permutation importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>evi_081 (22 March - 6 April)</td>
<td>23.4</td>
<td>30.6</td>
</tr>
<tr>
<td>evi_273 (30 Sept - 15 Oct)</td>
<td>23.8</td>
<td>24.1</td>
</tr>
<tr>
<td>evi_017 (17 Jan - 1 Feb)</td>
<td>15.9</td>
<td>16.6</td>
</tr>
<tr>
<td>evi_193 (12 July - 27 July)</td>
<td>21.4</td>
<td>16.1</td>
</tr>
<tr>
<td>evi_241 (29 Aug - 13 Sept)</td>
<td>5.5</td>
<td>9.9</td>
</tr>
<tr>
<td>evi_257 (14 Sept - 29 Sept)</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>evi_161 (10 June - 25 June)</td>
<td>3.4</td>
<td>0.8</td>
</tr>
<tr>
<td>evi_145 (25 June - 9 June)</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>evi_225 (14 Aug - 28 Aug)</td>
<td>5.6</td>
<td>0.3</td>
</tr>
<tr>
<td>evi_097 (7 April - 22 April)</td>
<td>0.3</td>
<td>0</td>
</tr>
</tbody>
</table>

The study showed that with the use of small number of occurrence records and environmental variables derived from remotely sensed data; potential habitat as well as distribution patterns of plant species such as *P. merkusii* can be suitably modeled using Maxent. Since the species is confined to a very small geographical area, results of this study will play a major role in building strategic planning for the conservation of species. The potential habitat distribution map for *P. merkusii* can be helpful in land use management planning around its existing populations, identify conservation sites and also in selecting new sites for reintroduction of the species in its natural habitat. More research in the near future is necessary in order to safeguard the species and its habitat.

4. Conclusions

The present study utilized the species occurrence information and remotely sensed environmental variables to generate the distribution map of the target species and its validation. The AUC of 0.979 and 0.972 observed for the training and test data respectively, indicated higher success rates. Of the 10 predictor variables used to build our model, evi_081 appeared as the strongest predictors. Hence it has also been observed that environmental variables like MODIS enhanced vegetation index images derived from remotely sensed data can plays a key role in determining the distribution of potential habitats of *P. merkusii*. The result further indicates that species distribution models offer new understandings on the geographical bionetwork of these species. Mapping and modeling the geographic range of a species grips its potentiality in conservation biology. It was further hoped that the predicted potential distribution habitat of the species in this study would help in developing conservation strategies for monitoring and managing the species in its native range.
Acknowledgement

Firstly the authors are grateful to Oak Ridge National Laboratory Distributed Active Archive Centre for free distribution the MODIS data. Thanks are due to Biranjoy Basumatary and Bijit Basumatary for their help during the field study.

References


Chitale, V.S. and Behera, M.D., 2012, Can the distribution of sal (Shorea robusta Gaertn. f.) shift in the northeastern direction in India due to changing climate?. Current Science, 2, 1126–1135.


