Predicting the potential distribution of *Taxus sumatrana* using Maximum Entropy Model

M H Saputra and A D Sunandar

Balai Penelitian dan Pengembangan Lingkungan Hidup dan Kehutanan Aek Nauli Jalan Raya Parapat KM 10,5 Simalungun - Sumatera Utara
e-mail address: mhadis.ms@gmail.com

Abstract. *Taxus sumatrana* as the medicinal plant is noted as endangered species. As a potential medicinal plant, conservation act is important. This study aims to identify the environmental variables that contribute to *T. sumatrana* distribution and predict the potential areas for *T. sumatrana* conservation area. The Maximum Entropy (Maxent) was applied to model the species distribution. The presence of *T. sumatrana* was collected from 77 locations across the Sumatra. Climate and biophysical variables were used to determine the potential distribution of species. The results showed that a minimum temperature of coldest months has the highest permutation importance by 66.3% while the highest contribution variable was altitude by 48.1%. Soil type has an important value that is not presented in other variables. Potential regions for *T. sumatrana* distribution were: North Sumatra (Karo, Dairi, Toba Samosir, and Simalungun District), West Sumatra (Agam, Tanah Datar, Sawahlunto and Solok District), Jambi (Bungo, Merangin and Kerinci District), Bengkulu (Bengkulu Utara), and South Sumatra (Lahat District). The potential environmental condition for *T. sumatrana* is 1,500m above sea level and upper, minimum temperature of the coldest month below 10°C with precipitation in the driest month between 50-70 mm/month. The species is potentially found in Humid Andosol and Vitric Andosol soil types.

1. Introduction

One of genus in Taxaceae family which has potential as medicinal plant is Taxus. *Taxus sumatrana*, like many other Taxus also produces paclitaxel which is recognized as cancer inhibitors especially for breast and ovarium cancer [1], breast cancer, lymph node cancer and several other types of cancer [2]. Beside as cancer inhibitor, Taxus also has potential as anticonvulsant, antipyretic, analgesic [3], and also antifungal [4].

Taxus genus is spread generatively by seed[5] and found mostly in temperate region (Europe and North America). In Asia, Taxus spread from the continent in the north to Indonesia in the south and in Indonesia, *T. sumatrana* was found in Sumatra and Sulawesi islands [6], [7], [8]. However, the spread of Taxus in Sumatra was in spotted pattern, only in certain small area. Thus, in IUCN red list, *T. sumatrana* is categorized as endangered species.
Unfortunately, very limited information is available due to the limited existence of the species in its natural habitat. Small number of trees were found in Karo regencies in North Sumatra province while larger number of trees were found in Kerinci regencies, Jambi Province. Low growth rate and low reproduction capability of the species [9] have become internal threat and decreased the distribution of the species in the wild. This condition also happened in Andalusia as described by [10].

The method for estimating the potential geographic distribution of species is the Species Distribution Models (SDM). SDM characterizes the environmental conditions that are suitable for the species and identifies where suitable environments are distributed in space[11]. One of the appropriate models for this study is Maximum Entropy (Maxent). Maxent addresses the distribution of species using the presence-only data with small number of sample and was proven to be advantageous over the SDM methods[12]. The model offers quantitative approach to understand the correlation between environmental variables and species distribution in geographic form. Maxent result gives the insight of the potential area of *T. sumatrana* to survive based on the environmental variables that suitable with the species requirement. This study aims to determine the most influential factors in the spread of *T. sumatrana* and identifies suitable areas for the development of the species.

2. Materials and Method

2.1. Study site
The *T. sumatrana* is distributed across the Sumatera island from Aceh to Lampung. The Sumatra island is part of Indonesia with latitude 6°N to -6°S and Longitude 95° to 110°E. The area consists of 9 provinces i.e. Aceh, North Sumatra, West Sumatra, Riau, Jambi, Bengkulu, South Sumatra, Bangka Belitung, and Lampung. Generally, the climate in Sumatra has a heavy rainfall for a year [13]. Sumatra land use is dominantly covered by forest and crossed by the mountainous such as Pegunungan Seribu in the western part [14], [15]. The elevation ranged from 0 to 3,800m above the sea level where the peak of the island is Mount Kerinci. The flora biodiversity is very high with more than 820 species richness in Sumatra. Most of the flora type is Malesiana related to the Asian country [16].

2.2. Data collection of species occurrence
The 77 spots of *T. sumatrana* location were collected. The presence of *T. sumatrana* location was gained from primary and secondary data. The primary data were collected by land survey in North Sumatra and Jambi region. From this area, 74 (8 from North Sumatra and 66 from Mount Kerinci) spots were collected. Meanwhile, 3 location data were noted from the literature review. The location was found in Mount Dempo in South Sumatra [17]. All *T. sumatrana* location was given a latitude and longitude coordinates. The data were digitized in ascii file format. The result was input in computerized geographic information system (GIS) using PC ArcGIS software. The locations of *T. sumatrana* are presented in Figure 1.

2.3. Environmental variables
Nineteen bioclimatic variables [18] with 30-second (1 km) spatial resolution were obtained from the WorldClim dataset (http://www.worldclim.com). The permutation importance of bioclimatic variable into the model was used to detect the most influential variables associated with the distribution of *T. sumatrana*. The variables used include soil type, altitude, aspect and slope. Altitude was downloaded from DIVA GIS website that provides near global 90meter resolution elevation data (https://www.diva-gis.org/Data). Slope (as degree of an angle) and aspect (as an eight-class categorical variable: N, NE, E, SE, S, SW, W, NW) were derived using ArcGIS 10’s Spatial Analyst. The Sumatra soil unit referred to Food and Agriculture Organization was defined into 43 soil units[19]. The data for environmental variable consist of elevation, aspect, slope, climate and soil type are shown in Table 1.
Table 1. Data source of Environmental Variable

| No | Data     | source                                          | Type | Extraction Method                     |
|----|----------|-------------------------------------------------|------|---------------------------------------|
| 1  | Elevation| www.diva-gis.org/Data                            | .tif | Spatial Analysis                      |
| 2  | Aspect   | spatial analysis from elevation data             | .tif | Analysis of the aspect of the surface topography |
| 3  | Slope    | spatial analysis from elevation data             | .tif | Analysis of the slope of the surface topography |
| 4  | Climate  | www.worldclime.org                              | .bil | Spatial Analysis                      |
| 5  | Soil Type| www.fao.org/soils-portal                         | .asc | Spatial Analysis                      |

Figure 1. T. sumatrana natural distribution Sumatra
2.4. Species distribution model (SDM)
SDMs have been developed regarding the improvement in GIS and spatial data [11]. The information of occurrence data and the variables as the input increased the possibility of using machine learning to predict the SDM. One of the promising approached in SDMs is Maxent. The advantages of Maxent is it uses the presence data only rather than complete information [20].

Maxent uses the Maximum entropy algorithm to predict the potential distribution of the species due to the presence location of the species. This model shows the probability of the potential areas that is suitable based on the average value of environmental variables in presence location where the species occur. Maxent idea is to estimate a target probability distribution by finding the probability distribution of maximum entropy, subject to a set of constraints that represent our incomplete information about the target distribution [12].

Maxent model relies on Baye’s rule [21]

\[
P(y = 1 | x) = \pi(x) P(y = 1 | X)
\]

where \(P(y=1 \mid x)\) is the probability that the species is present at the site \(x\) (\(y\) ranges from 0-1), \(\pi(x)\) is the present observation or realized distribution at \(x\) area, \(P(y=1)\) is the probability of presence, and \(|X|\) is the all-area probability from the \(X\) site.

Maxent uses the Gibbs distribution from the sets of features \(f\) and weights \(\lambda\) rather than estimating the Baye’s estimation rule. Gibbs distribution calculates the realized distribution (\(\pi(x)\)) from the Baye’s rule. The formula was defined by [21]

\[
q_\lambda(x) = \frac{\exp \left( \sum_{j=1}^{n} \lambda_j f_j(x) \right)}{Z_\lambda}
\]

where \(q_\lambda(x)\) is the estimated probability of presence with the maximum entropy of \(\pi(x)\), \(\exp \left( \sum_{j=1}^{n} \lambda_j f_j(x) \right) / Z_\lambda\) is an exponential distribution parameterized by a vector of feature (\(f\)) weights (\(\lambda\)), and \(Z_\lambda\) is a normalization constant ensuring the \(q_\lambda(x)\) sum to one over the area.

After obtaining an estimate of \(q_\lambda\), sufficient information is obtained to get the probability distribution \(P(y=1|x)\), as indicated

\[
P(y = 1 | x) = \frac{(e^{Hq_\lambda})}{(1+e^{Hq_\lambda(x)})}
\]

where \(H\) is the entropy of \(q_\lambda\).

2.5. Method
The presence of \(T.\ sumatrana\) location was used as the input in Maxent together with the environmental variables. The Maxent program uses the Comma Separated Value (CSV) format for presence location. Furthermore, the environmental data were collected and selected for the Sumatra region. Maxent requires the raster format for the environmental layer that consists of elevation, slope, aspect, soil type, temperature and precipitation with the same resolution, extent and geographical coordinate system. The resolution is 30 meters and the extents for environmental layers are 5.90635 (top), 95.00927 (left), 108.91843 (right), and -6.16531 (bottom). The environmental layer will be extracted into Action Script Communication (ASC) file to be applied in Maxent program. All spatial data used WGS 1984 as the spatial reference.

The environmental variables used were soil unit, altitude, slope, aspect, and 8 bioclimatic factors. Those climatic factors are bio2 (annual mean temperature diurnal range), bio3 (isothermality), bio4 (temperature seasonality), bio6 (minimum temperature of coldest month), bio14 (precipitation of driest month), bio15 (precipitation seasonality), bio18 (precipitation of warmest quarter), bio19 (precipitation of coldest quarter) [22]. The bioclimate factors were chosen based on permutation importance in the model. Only 8 variables that have a higher importance value to the model.
Maxent provides the output in spatial as a map and several graphs that represent the environmental requirement for species distribution [23]. The map was interpreted as a suitable area based on the environmental variable in the background. The spatial output from Maxent, processed using ArcGIS to present the suitable area of *T. sumatrana* distribution and the region that fit with the *T. sumatrana* environmental requirement. Moreover, the Maxent result gives information about the important variables in the model with Jackknife analysis [24]. This analysis defined the variables that increases and decreases the gain of the model.

3. Result and discussion

3.1. Result

The Maxent model successfully predicted the potential area of *T. sumatrana* (Figure 2). Probability of presence estimate has been defined on the 0-1 scale. The suitable area was calculated based on the FAO suitable classification where it started from 0.4-1 [25]. The suitable area is presented in Figure 3.

The total area suitable for *T. sumatrana* occupied only 0.34% of the total land or 1,609 km2 from total 473,481 km2. Three main locations of *T. sumatrana* were Mount Kerinci in Jambi, Karo regency near Lake Toba and Mount Dempo in South Sumatra. However, according to the Maxent result, the potential suitable area for *T. sumatrana* was located along Bukit Barisan mountainous area.

![Figure 2](image-url)  
*Figure 2.* The Maxent result of *T. sumatrana* potential distribution. Red color indicated the highest probability, while blue mean lowest probability.
Figure 3. The suitable area for *T. sumatrana*. The red color shows the suitable area which occupy 0.34% from the total Sumatra island.

The potential distribution of Taxus was noticed in several regions in Sumatra. The provinces which are predicted as the suitable area for *T. sumatrana* are North Sumatra, West Sumatra, Jambi, Bengkulu and Lampung. The potential area was showed in Figure 4.

Figure 4. The potential area of *T. sumatrana*. The colour shows the potential regencies in Sumatra island and the black represents the unsuitable area.

The validation of the model was calculated based on the Receiver Operating Characteristic (ROC) graphic. The graphic indicated the Higher Area Under Curve (AUC) with the score of 0.989 which
signified the good performance for the model. A high AUC indicates that sites with high predicted suitability values tend to be areas of known presence [26]. The highest value in AUC describes the best model by considering all variables used in the model.

**Figure 5.** ROC graph representing the sensitivity and 1-sensitivity graph of *T. sumatrana*

The area suitable for *T. sumatrana* was fragmented in temperature, soil, and altitude due to the Jackknife analysis presented in Figure 6. The minimum temperature of the coldest month represented as bio 6 is the most important variable. The isolation of this variable in Maxent analysis increases the gain for the model. On the other hand, omitting the soil type from the variable will decrease the gain for the model.

**Figure 6.** Jackknife analysis diagram
Two variables important in the model due to Jackknife analysis are shown in Figure 7. Those two variables are bio6 (minimum temperature of the coldest month) and the soil unit.

**Figure 7.** The response of *T. sumatrana* to minimum temperature of coldest month (left) and soil unit (right). Red colour is mean score while the blue is standard deviation of the mean value.

According to the Figure 7, high potential distribution of *T. sumatrana* is in the temperatures below 10°C by 0.5 - 0.8 probability occurrence. For soil unit, the dominant soil types were numbers 5 and 15 which are humic andosols and virtic andosols, respectively. Those soil types have the higher probabilities by more than 0.5 for humic andosols and 0.9 for virtic andosols. Furthermore, the analysis of contribution (in percentage) shows that altitude has the highest contribution followed by the soil and bio14 (precipitation of driest month). The response of the *T. sumatrana* to altitude and bio14 is presented in Figure 8.

**Figure 8.** The response of *T. sumatrana* to altitude (left) and precipitation of driest month (right). Red colour is mean score while the blue is standard deviation of the mean value.

The potential altitude for *T. sumatrana* is higher than 1,500m above sea level. In driest month, *T. sumatrana* survived with precipitation around 50-70 mm/year and declined when the precipitation is higher than 70 mm/year.

### 3.2. Discussion

Maximum entropy model predicts the potential distribution of *T. sumatrana* based on several environmental variables. The validation shows that the model is reliable and can be used to predict the suitable area of *T. sumatrana* in Sumatra Island. The potential area was known in the mountainous area. The *T. sumatrana* mostly occurs in the hill of the mountain and above 1,500m[17]. The altitude shows highest contribution (48.1%) in *T. sumatrana* distribution. It means that altitude will determine the potential area for *T. sumatrana* to distribute. The research in Kebun Percobaan Sipisopiso North Sumatra shows the good performance of *T. sumatrana*. The location has an elevation around 1,400m above sea level[27].
Meanwhile, soil contributes the second highest by 13.6% for species distribution. The suitable soil is andosol type that has dark colour and highly porous. This soil develops from parent material of volcanic origin. Andosol typically occurs in forested highland areas of the continental lands bordering the Pacific Ocean [19]. The altitude and the andosol type explain the potential distribution of the *T. sumatrana* in the mountainous area. However, the detail information about the *T. sumatrana* is still limited and unexplained yet due to its position as the species of less concern in Indonesia [17].

In the climate variables, the minimum temperature of coldest month and precipitation in the driest month showed an important role for the model. The extreme condition shows the ability of *T. sumatrana* to survive in temperature lower than 10°C and a limited amount of rainfall by 50-70 mm/month. According to the genus of Taxus, most of this species are growing in the sub-tropic area [17], [28], [29]. It is possible that *T. sumatrana* has the same requirement for distribution in the tropic area.

4. **Conclusion**

In general, *T. sumatrana* as the endangered species has been known as the important species for medicine. However, the information about this species is still limited, especially in Indonesia. The information of *T. sumatrana* presence location can be used in SDM using Maxent to predict the higher probability area that satisfies the environmental requirement for *T. sumatrana*.

The result shows that the species has a potential distribution across the mountainous area from North Sumatra, West Sumatra, South Sumatra, Bengkulu, and Jambi region. This species requires a high altitude (1,500 m above sea level and higher) and can survive in extreme conditions with low temperature and a small amount of precipitation. The andosol soil type was suitable for the species to distribute in the region.

**Acknowledgment**

This paper was supported by Environment and Forestry Research and Development Agency of AekNauli. We thank our colleagues from Silviculture laboratories who provided information and also the technical staff for their great help in field data collection.

5. **Appendices**

| Value | Dominant Soil     | FAO SOIL unit | Probability |
|-------|-------------------|---------------|-------------|
| 1     | Orthic Acrisols   | Ao82-2/3b     | < 0.1       |
| 2     | Dystric Fluvisols | Jd9-2/3a      | < 0.1       |
| 3     | Chromic Luvisols  | Le97-3b       | < 0.1       |
| 4     | RENDZINAS         | E18-3c        | < 0.1       |
| 5     | Humic Andosols    | Th17-2c       | > 0.6       |
| 6     | Humic Acrisols    | Ah27-2/3c     | < 0.1       |
| 7     | Orthic Acrisols   | Ao70-2/3b     | < 0.1       |
| 8     | Humic Acrisols    | Ah25-2c       | < 0.1       |
| 9     | Orthic Acrisols   | Ao104-2/3c    | > 0.4       |
| 10    | Humic Ferralsols  | Fh12-2/3c     | < 0.1       |
| 11    | Water             | WAT           |             |
| 12    | Dystric Fluvisols | Jd12-2/3a     | < 0.1       |
| 13    | Albic Arenosols   | Qa17-1a       | < 0.1       |
| 14    | Dystric Fluvisols | Jd10-2/3a     | < 0.1       |
|   | Soil Type                  | Code  | Property  |
|---|---------------------------|-------|-----------|
| 15| Vitric Andosols           | Tv38-1bc | >0.9      |
| 16| Humic Andosols            | Th18-2ab | < 0.1     |
| 17| Orthic Ferralsols         | Fo101-2b | < 0.1     |
| 18| Ferric Acrisols           | Af57-2a | < 0.1     |
| 19| Dystric Cambisols         | Bd63-2/3b | < 0.1     |
| 20| Humic Acrisols            | Ah26-3c | < 0.1     |
| 21| Dystric Histosols         | Od19-a | < 0.1     |
| 22| Orthic Podzols            | Po27-1a | < 0.1     |
| 23| Humic Gleysols            | Gh20-3a | < 0.1     |
| 24| Humic Cambisols           | Bh21-2c | < 0.1     |
| 25| Plinthic Acrisols         | Ap25-2/3a | < 0.1     |
| 26| Xanthic Ferralsols        | Fx32-2ab | < 0.1     |
| 27| Humic Cambisols           | Bh17-2bc | < 0.1     |
| 28| Ferric Acrisols           | Af55-3b | < 0.1     |
| 29| Dystric Cambisols         | Bd58-2c | < 0.1     |
| 30| Orthic Luvisols           | Lo65-2/3b | < 0.1     |
| 31| Humic Nitosols            | Nh9-3a | < 0.1     |
| 32| Ferric Acrisols           | Af11-2/3a | < 0.1     |
| 33| Dystric Histosols         | Od20-a | < 0.1     |
| 34| Orthic Acrisols           | Ao83-2/3c | < 0.1     |
| 35| Eutric Fluvisols          | Je62-2/3a | < 0.1     |
| 36| Cambic Arenosols          | Qc59-1ab | < 0.1     |
| 37| Orthic Acrisols           | Ao105-2/3c | < 0.1     |
| 38| Eutric Cambisols          | Be116-2c | < 0.1     |
| 39| Ferric Acrisols           | Af56-2a | < 0.1     |
| 40| Dystric Nitosols          | Nd54-3b | < 0.1     |
| 41| Eutric Fluvisols          | Je63-2/3a | < 0.1     |
| 42| Humic Gleysols            | Gh21-3a | < 0.1     |
| 43| Dystric Cambisols         | Bd62-2/3a | < 0.1     |

### 6. References

[1] Markman M 1991 Topotecan: an important new drug in the management of ovarian cancer *The Yale Journal of Biology and Medicine* **64**, 583–90

[2] Rajendran L, Rajagopal Kand Subbarayan K 2013 Efficiency of fungal taxol on human liver carcinoma cell lines *American Journal of Research Communication* **1(6)** 112–21.

[3] Nisar M, Khan I, Simjee US, Gilani AH, Obaidullah and Perveen H 2008 Anticonvulsant, analgesic and antipretic activities of TaxuswallichianaZonec. *J. of Ethno-Pharmacology* **116** 490-4. DOI: 10.1016j-jep.2007.12.021.

[4] Tachibana S, Ishikawa K and Itoh K 2005 Antifungal activities of compounds isolated from the leaves of *Taxus cuspidata* var. nana against plant pathogenic fungi *Journal of Wood Science* **51** 181–4.

[5] Guang-li S, Bing B and Chang-hu L 2010 Seed rain and seed bank of Chinese yew
(Taxuschinensis var. mairei) population in Tianmu mountain ActaEcologicaSinca30(5) 276-79

[6] Deforce Kand Bastiens J 2013 The Holocene History Of The Holocene History Of Taxusbaccata (Yew) In Belgium And Neighbouring Regions Belgian Journal of Botany 140(2) 222–37.

[7] Farr K 2008 Genus–level approach to Taxus species(Mexico City).

[8] Harahap RMS and Izudin I 2002 Konifer di Sumatera Bagian UtaraKonifera1/XVII 61–71.

[9] Hussain A, Qarshi IA, Nazir H, Ullah I, Rashid M and Shinwari ZK 2013 In vitro callogenesis and organogenesis in TaxuswallichianaZuccThe Himalayan yew J. Bot 45(5) 1755–59.

[10] Garcia D, Zamora R, Hodar JA, Gomez JM and Castro J 2000 Yew (Taxusbaccata L.) regeneration is facilitated by fleshly-fruited shrubs in Mediterranean environmentsBiological Conservation95 31-8.

[11] Miller J2010 Species distribution modeling Geography Compass 4 490–509.

[12] Phillips SJ, Anderson RP and Schapire RE2006 Maximum entropy modeling of species geographic distributions Ecological Modelling190 231–59.

[13] S Wirjohamidjojo, Y Swarinoto 2010klm kawasan Indonesia (dari Aspek Dinamik - Sinoptik)(Jakarta, Indonesia: Badan Meteorologi Klimatologi Dan Geofisika Jakarta)

[14] Badan Pusat Statistik 2016Laporan Uji coba lahan pulau sumatera(Badan Pusat Statistik)

[15] Suharyani, Qisthi N, Siregar L, et al 2016Deskripsi Peta Ekoregion Sumatera Skala 1:250.000pp 165.

[16] Kusmana C and Hikmat A2015 The Biodiversity of Flora in Indonesia. Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan5 187–98.

[17] Hidayat A, Rachmat HH and Subiaxto A2014Taxus sumatrana: Mutiara Terpendam dari Zamrud Sumatra (FORDA PRESS).

[18] Hijmans RJ, Cameron S, Parra J2005worldclim:very high resolution interpolated climate surfaces for global land areas(http://worldclim.com/version1)

[19] FAO 1974Soil map of the world(Epub ahead of print) DOI: 10.1038/1791168c0.

[20] Phillips SJ, Dudík M and Schapire RE2004 A maximum entropy approach to species distribution modeling Twentyfirst international conference on Machine learning ICML 469-83.

[21] Phillips SJ and Dudík M 2008 Modeling of species distribution with Maxent: new extensions and a comprehensive evaluationEcography31 161–75.

[22] O’Donnell MS and Ignizio DA2012 Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States US Geological Survey Data Series 691 10.

[23] Baldwin RA2009 Use of maximum entropy modeling in wildlife research Entropy11 854–66.

[24] Elith J, Phillips SJ, Hastie T, et al 2011 A statistical explanation of MaxEnt for ecologists Diversity and Distributions1743–57.

[25] Loi N Van 2008 Use of GIS Modelling in Assessment of Forestry Land’s Potential in Thua Thien Hue Province of Central Vietnam.

[26] Hijmans RJ and Elith J2013 Species distribution modeling with R Introduction 71.

[27] Sunandar AD and Saputra MH2015 pertumbuhan awal dua provenans taxus sumatrana di kebun percobaan sipiso-piso SEMINAR NASIONAL MAPEKI XVII PRO. medan: Masyarakat
Peneliti Kayu Indonesia Pusat Penelitian Biomaterial LIPI, 2015, pp. 230–36.

[28] Nimasow G, Nimasow OD, Rawat JS, et al 2016 Remote sensing and GIS-based suitability modeling of medicinal plant (Taxus baccata Linn.) in Tawang district, Arunachal Pradesh, India. *Current Science* **110** 219–27.

[29] Sunandar AD and Saputra MH2016 Pertumbuhan Stek Taxus sumatrana Miq. de Laub Umur Satu tahun di persemaian *Prosiding Ekspose Hasil-hasil Penelitian* (Bogor ,Indonesia: Pusat Penelitian dan Pengembangan Hasil Hutan BP2I Kementerian Lingkungan Hidup dan Kehutanan)