Modelling the distribution of *Shorea leprosula* in Kalimantan, Indonesia: a tool for conservation planning

N S Lestari¹*, J Elith²

¹ Dipterocarps Research Center, The Ministry of Environment and Forestry of Indonesia
² School of Biosciences, the University of Melbourne, Australia

*Corresponding author: nurul.silva@gmail.com

**Abstract.** *Shorea leprosula* belongs to the Dipterocarpaceae family, a dominant family in Indonesia’s rainforest. The population of this species has been depleted due to extensive logging, high rates of deforestation and forest degradation in the past several decades. However, the current status of the species’ range and distribution remains unknown. This research project aims to understand and predict the distribution of *S. leprosula* in Kalimantan using species distribution models (SDMs). This study used presence records and presence-absence records from field surveys and the Global Biodiversity Information Facility (GBIF) database. Two modelling methods, MaxEnt and generalized linear models (GLMs), were used to predict species distribution. Prediction maps varied with modelling methods and different datasets, producing different emphases on areas suitable for *S. leprosula*. Even though the descriptive and predictive capabilities of the models are considered modest, the models provide useful insights about environmental factors that affect the distribution pattern of *S. leprosula*. Given the limitations of species records used in this study, the model outputs also need to be cautiously interpreted to avoid prediction biases. This research project also highlights some issues that arise in using small sample sizes in developing the model. Despite its limitations, the prediction maps generated by the models can give some hints to identify the areas with high possibility of the presence of *S. leprosula* in Kalimantan. In addition, this research project also provides some important information on how to improve model predictions for future development to support species conservation in Indonesia’s rainforests.

1. **Introduction**

*Shorea leprosula* is one of several important plant species in the tropical rainforests of Indonesia. However, the population of *S. leprosula* has been depleted due to extensive logging, the high rate of deforestation and forest degradation in the past several decades [1]. The rapid decline of forest areas in tropical regions in general and Indonesia in particular has put significant pressure on many plant species, including *S. leprosula*, which may lead to species extinction [2] . Given all the threats that have been experienced by this species, *S. leprosula* is considered an endangered species according to the IUCN Red List of Threatened Species [3].

*S. leprosula* (common name: light red meranti) has been reported as widely spread across Borneo, Sumatra, Thailand and Peninsular Malaysia [4]. However, the current status of the species’ range and distribution remains unknown. A few studies of this species have been conducted in various locations,
but these mostly focus on the ecological aspects of its growth and abundance. One of the biggest challenges in estimating the correlates of its distribution is that this species is only recorded in relatively few locations. This is a typical characteristic of the availability of species data in tropical regions. Species inventories undertaken across the whole geographical range of the species are not feasible due to financial constraints, lack of adequate infrastructure (making access through dense tropical forests difficult), as well as inadequate numbers of experts working in data collection. In addition, data collection intensities are sometimes unevenly distributed. As a result, some areas may have better records than others [5]. Limited knowledge about species distribution highlights several implications for the conservation of *S. leprosula*. Good conservation planning should be able to select suitable areas for conservation measures based on reliable scientific data to achieve conservation goals. Moreover, knowing the distribution of species may also reduce subjectivity and the uncertainty that may occur in the conservation planning [6].

Species distribution models (SDMs) are a common approach for predicting habitat suitability for species in unsurvey areas. They rely on estimation of the relationship between species and environmental variables. SDMs can provide both ecological inference and species distribution predictive maps using various modelling techniques that include spatial analysis and statistical models. Moreover, SDMs allow us to explore which environmental variables have the strongest influence on the distribution of a given species [7]. SDMs can be used to predict species distribution within the range of environmental variables sampled in species occurrence data in the same temporal frame. SDMs can also be used to predict the potential occurrence of species under environmental changes in different time frames [8].

Compared to the application of SDMs in temperate regions, the use of SDMs to support the conservation of single tree species in Indonesia, among other tropical countries, has not been given much emphasis due to lack of studies in this field [5]. Although there has been a growing debate about whether or not environmental variables affect species distribution in tropical rainforests (Bell 2005), this study attempts to use SDM as an approach to predict species distribution based on the relationship between environment and species.

This study aims to understand and predict the distribution of *S. leprosula* in Kalimantan. Kalimantan is an interesting study area because the tropical rainforests in this area are the centre of the distribution of Dipterocarp species in the Southeast Asia region and is a main habitat of *S. leprosula* [4] However, given the limitation of species records in this area and the observation that species in tropical forests tend to be distributed randomly (Hubbell 2005), a clear possibility is that *S. leprosula* cannot be modelled effectively. Therefore an important part of this research was to carefully test whether various typical approaches to modelling could successfully explain and predict the distribution of *S. leprosula*.

2. Methodology

2.1. Species Data

Species occurrence records in this study were obtained from field surveys in 11 locations across Kalimantan conducted by organizations under the Indonesian Ministry of Forestry between 2008 and 2012. Species data were also derived from field surveys in 3 locations in West Kalimantan, conducted by the Fauna and Flora International (FFI) as part of forest carbon stock assessment in the period 2009-2014. The locations were further classified into 5 categories based on forest types as described in table

In addition, species occurrence records were also retrieved from the Global Biodiversity Information Facility (GBIF) [9]. Two datasets were created from the GBIF data: (1) presence records for *S. leprosula* across Kalimantan that consist of 10 records; (2) background records for the model using only Plantae records across Kalimantan that consist of 2,307 records. The reason for using a background sample is to try to account for sampling bias that is likely to exist in the GBIF data. This type of background sample is known as target group background (TGB) [10].
Table 1. Summary of species data from field surveys

| Location             | Number of presence records | Number of absence records | Location category |
|----------------------|----------------------------|---------------------------|-------------------|
| Gunung Lumut         | 14                         | 17                        | 1                 |
| Meratus              | 10                         | 7                         | 1                 |
| Gunung Palung NP     | 6                          | 23                        | 2                 |
| Malinau              | 4                          | 6                         | 1                 |
| Seruyan              | 12                         | 11                        | 1                 |
| Tumbang Nusa         | 13                         | 32                        | 3                 |
| Ketapang             | 3                          | 16                        | 2                 |
| Sangai               | 3                          | 16                        | 1                 |
| Betung Kerihun NP    | 7                          | 58                        | 4                 |
| Bukit Baka NP        | 3                          | 19                        | 4                 |
| Labanan              | 2                          | 30                        | 1                 |
| Danau Siawan         | 0                          | 57                        | 3                 |
| Pematang Gadung      | 0                          | 11                        | 1                 |
| Laman Satong         | 1                          | 6                         | 5                 |
| **Total**            | **78**                     | **309**                   |                   |

Note: Location category: 1 = secondary forest, 2 = primary forest + secondary forest, 3 = secondary forest + wetland, 4 = primary forest, 5 = grassland

2.2. Predictor variables

Eight environmental variables were selected in this study as candidate predictors in modelling the distribution of *S. leprosula*. They are soil depth, soil acidity, soil drainage, soil type, elevation, slope, mean annual rainfall and mean annual temperature described in Table 2. Spatial environmental data were in raster format and retrieved from WRI suitability maps in Kalimantan [11]. For climate variables, an environmental dataset from WorldClim database was also used. All maps were further set at a grid cell size of ~100m.

Table 2. Predictor variables for modelling the distribution of *S. leprosula*

| Variable | Description                                                                 | Range               |
|----------|-----------------------------------------------------------------------------|---------------------|
| Type     | Soil type was re-categorized (to make it a more efficient variable for modeling) based on the soil’s capacity to provide water for plants in dry period. Category 1: the combination between Ultisols, Inceptisols and Alfisols. These soil types are able to provide water for plants for more than half of the year or more than 3 consecutive months in dry period. Category 2: other soil types (Entisols, Histosols, Mollisols, Oxisols, Spodosols). | -                   |
| Acid     | Soil acidity (pH)                                                           | < 4.0 - 7.3         |
| Drain    | Soil drainage, classified based on soil capacity to let water pass          | very poor, poor, well, very well-drained |
| Depth    | Soil depth (cm)                                                             | 0-150               |
| Elev     | Elevation (m)                                                               | 0-2294              |
Variable | Description | Range
---|---|---
Slope | Slope (%) | 0-80
Rain | Mean annual rainfall (mm) | 1526.43-3447.28
Temp | Mean annual temperature (°C) | 15.8-27.6

2.3. Modelling methods

2.3.1. Maximum Entropy model (MaxEnt)
In this study, three models were fitted and projected using MaxEnt software (version 3.3.3I). The first MaxEnt model was developed using 10 presence records derived from GBIF and 10,000 background points randomly sampled by MaxEnt by default. The second MaxEnt model was fitted using the Sites with Data (SWD) format [12], with 10 species presence records and TGB from GBIF data. The idea of running the MaxEnt model using the GBIF records is to predict the distribution of *S. leprosula* based on open access data. The third MaxEnt model was developed using 87 presence-only records across Kalimantan that have been collected from both field surveys and the GBIF data. Since MaxEnt cannot use NA values, they should be omitted, resulting in 67 presence records being used to determine the MaxEnt distribution. This model also used a sample of 10,000 points across Kalimantan randomly sampled by MaxEnt software as a background sample.

Seven environmental variables were offered as candidate predictors in the MaxEnt models. Elevation was not included in the model development since it has a strong relationship with mean annual temperature. This study used linear and quadratic features to obtain relatively smooth models, and to limit complexity equally across the different datasets. The default logistic output was selected; this outputs a relative likelihood of species presence in which the values range from 0 to 1. The MaxEnt models were also fitted using 10-fold cross validation to evaluate model performance [8].

2.3.2. Generalized Linear Models (GLMs)
In this study, GLMs were developed using presence-absence records from field surveys in 14 locations across Kalimantan. The predictors used as candidate variables in the GLMs were selected based on the results of a pairwise correlation test. This study only used 5 environmental variables as predictors (mean annual rainfall, mean annual temperature, slope, soil type and soil drainage). In addition, given the structure of the data (locations clustered within sites rather than randomly scattered across the landscape), this study also included locations as a predictor in model development.

The Akaike information criterion (AIC) values between GLMs were compared to find out the relative quality of the GLM models in which the smallest AIC values represent better models [13]. The final model was then used to predict the distribution of *S. leprosula*. GLMs were run in R version 3.2.2 and function GLM, with dismo, car and effect packages were installed. Model performance in GLMs was assessed using % deviance explained, which represents the percentage of null deviance explained by the model [14].
3. Results and discussions

3.1. MaxEnt models

The mapped outputs (see figure 1) show that, in models 1 and 2, \textit{S. leprosula} is predicted more likely to occur in the middle and northern part of Kalimantan. Model 2, which attempts to deal with survey bias by using TGB records from GBIF, predicts more areas with higher likelihood of species presence in the northern Kalimantan compared to model 1. These more northern locations are areas with high mean annual rainfall. Meanwhile, model 3 produced different spatial patterns, in which predicted areas with high likelihood of species presence becomes smaller compared to both model 1 and model 2 with slightly more emphasis on southern parts of Kalimantan. The extended southerly distribution of \textit{S. leprosula} as predicted in model 3 is related to soil type category 1 that occurs in this area. This area also has well-drained soil. These two variables become more important than the other variables in model 3.

The variable importance also varies between models. Models 1 and 2 show a similar result in which mean annual rainfall makes the highest contribution to the models compared to other predictors. However, mean annual rainfall is much more dominant when using background points from the GBIF records across Kalimantan. Both models also show that slope, soil depth and soil drainage have no contribution to the MaxEnt models. Meanwhile, variable importance in model 3 shows a remarkably different result. In model 3, mean annual rainfall only makes a small contribution to the model while both soil depth and soil type and soil type are considered as important predictors (see table 3).

With presence-absence data a model is considered to be useful if the AUC value is over 0.75 [15], and evaluations for presence-background data are often based on this same rule of thumb. Based on this value, in model 3 \textit{S. leprosula} has more useful predictive performance (with AUC = 0.77, see table 4) than models 1 and 2 (AUC of 0.69 and 0.74 respectively).

The final selected GLM indicated that rainfall and temperature are the two most important environmental variables predicting the distribution of \textit{S. leprosula} when presence-absence data from field surveys was used. Based on these two variables, the model predicted that the species is more likely to occur in northern and south-eastern Kalimantan, and also in small areas in the western part of Kalimantan. These locations are identified as areas that have moderate mean annual rainfall and temperature (see figure 2).
Table 3. Variable contribution to the MaxEnt models and the evaluation of model performance

| Model | Rain | Temp | Slope | Type | Acid | Depth | Drain | AUC (10 fold CV) |
|-------|------|------|-------|------|------|-------|-------|-----------------|
| 1     | 64.2 | 0.7  | 0     | 15.7 | 19.5 | 0     | 0     | 0.692           |
| 2     | 91.3 | 1    | 0     | 0.5  | 7.2  | 0     | 0     | 0.739           |
| 3     | 1.1  | 3.3  | 17.6  | 31.7 | 0.7  | 43.3  | 2.3   | 0.766           |

Figure 2. The prediction map of the distribution of S. leprosula across Kalimantan using GLM

Furthermore, the model with the smallest AIC value (343.34, see table 4), included rainfall, temperature and locations as covariates, with quadratic terms for the two continuous variables. All these covariates have significant contributions (P < 0.05). After removing outlier observations that might potentially influence the fitted model [16], the deviance explained of this model is about 17.9%. Although the deviance explained is not as high as some GLMs in ecological studies [8], this can be considered a modest level of explanation.

Table 4. Detailed formulas of the GLM models and their outputs

| GLM formulas | AIC value | Deviance explained (%) | AUC value |
|--------------|-----------|------------------------|-----------|
| glm(Shorea_leprosula ~ Location, data=moddat2, family=binomial) | 380.0882 | 4.85 | 0.645 |
| glm(Shorea_leprosula ~ poly(rain,2) + poly(temp,2), data=moddat2, family=binomial) | 344.9645 | 13.89 | 0.749 |
| glm(Shorea_leprosula ~ poly(rain, 2) + poly(temp, 2) + poly (slope, 2), data=moddat2, family=binomial) | 348.9043 | 13.90 | 0.747 |
| glm(Shorea_leprosula ~ poly(rain, 2) + poly(temp, 2) + poly (slope, 2) + type, data=moddat2, family=binomial) | 349.8814 | 14.16 | 0.755 |
| glm(Shorea_leprosula ~ poly(rain, 2) + poly(temp, 2) + poly (slope, 2) + type, data=moddat2, family=binomial) | 350.8276 | 14.43 | 0.754 |
GLM formulas | AIC value | Deviance explained (%) | AUC value
--- | --- | --- | ---
(slope, 2) + type + drain, data=moddat2, family=binomial) | 350.2908 | 14.06 | 0.753
 glm(Shorea_leprosula \sim poly(rain, 2) + poly(temp, 2) + poly(slope, 2) + drain, data=moddat2, family=binomial) | 346.3741 | 14.04 | 0.756
 glm(Shorea_leprosula \sim poly(rain, 2) + poly(temp, 2) + drain, data=moddat2, family=binomial) | 345.9825 | 14.14 | 0.757
 glm(Shorea_leprosula \sim poly(rain, 2) + poly(temp, 2) + type, data=moddat2, family=binomial) | 343.34 | 17.9 | 0.767
 glm(Shorea_leprosula \sim poly(rain, 2) + poly(temp, 2) + Location, data=moddat2, family=binomial) | 343.34 | 17.9 | 0.767

Meanwhile, the estimate of AUC value in the fitted GLM is 0.767, which is similar though not directly comparable to the AUC value of MaxEnt model 3 because the datasets differ in absence vs random background points.

3.2. Species distribution
This study used TGB samples in MaxEnt model 2 to account for bias [17]; [18]. The AUC values indicate that MaxEnt model 2 performs better than MaxEnt model 1, however because the background samples differ across the two datasets the results are not entirely comparable. This finding is consistent with those generated by [10] in which the use of occurrence data as background samples (also known as target group background) can improve presence-only model performance.

The MaxEnt model 3 can be compared, although not directly, with the GLM model. Both models used the similar presence dataset but used a different treatment of absence or background data. MaxEnt model 3 used random background, which may lead to a model and predictions imprinted by bias in collection. In contrast, because GLMs are based on presence and absence data, the data show what was surveyed so bias is rarely problematic [10]. The MaxEnt models assume that species data used in building a model are randomly sampled [19]; [20]. Meanwhile, species data used as samples in the Model 3 data tend to be clustered in one site due to uneven sampling intensity.

In the GLM model, location also appears to become an important predictor in determining the probability of species presence. The model indicates that the location with highest prevalence is primary forests. Since *S. leprosula* has become a major target of illegal logging due to its high economic value, higher species occurrence probability in these locations may also be associated with fewer disturbances. Disturbance level can also be an important predictor that affects species distribution, although it is still rarely incorporated into the models because of lack of spatial data in this field (Elith and Franklin 2013). The GLM model also indicates that *S. leprosula* has high prevalence in locations with the combination of secondary forest (logged-over forest) and wetland. This result is consistent with [4] and [21], which found that *S. leprosula* commonly occurs in swampy areas and concave topography close to streams and rivers. However, because the land cover data used in this study is not resolved to a finer discrimination between wetland types, it is still unclear whether *S. leprosula* has preference for any particular wetland type.

Although locations can explain variable response quite well in the GLM model, using locations as a predictor has limitations. Each location has different environmental characteristics beyond the land use type that might be difficult to further quantify. Moreover, the use of land cover as a baseline to classify locations in this study may lead to biased predictions in the sense that predictions might be more likely to reflect recent land use changes rather than environmental preferences.
3.3. Species vs environmental variables

Even though the descriptive and predictive capabilities of the models are considered modest, the models provide useful insights about environmental factors that affect the distribution pattern of *S. leprosula*. Rainfall appears to be important and makes significant contributions to predict the species distribution in MaxEnt model 1, MaxEnt model 2 and the GLM model. This result may be insufficient to draw a firm conclusion about the relationship between *S. leprosula* and rainfall but may indicate the importance of rainfall variable for species distribution predictions. However, relying only on this variable as a predictor in the models is not really useful to predict species distribution since many plants in tropical rainforests are also strongly related to wet climate [22]. It is therefore important to include other variables in addition to climate variables in model predictions.

In terms of soil variables, soil type appears to be important in MaxEnt model 2 and 3. Both models indicate that *S. leprosula* is more likely to occur in areas where Inceptisols, Ultisols and Alfisols are present. These soils can provide water to plants for a long period of time during the dry season [23]. Soil water availability for growing plants can also be attributable to other measured soil characteristics, such as soil texture and soil moisture level that, may be also influential for *S. leprosula*.

Interpreting ecological inference in SDMs, particularly for tropical rainforests, is quite challenging due to the complexity of the ecosystems. In tropical tree communities, it also has been proposed that the neutral theory widely applies in species and habitat relationships. According to the neutral model under this theory, the distribution of plant communities in tropical regions is more likely to be random rather than influenced by particular environmental variables [24][25]. However, a study conducted by [25] does not appear to validate this view. He further points out that in the neutral models, species distribution pattern is not always random and local dispersal may increase species distribution randomness. By using local selection in heterogeneous environments, the strength of species habitat associations can be estimated based on how well species adapt in each different site. To test this further, it will be necessary to incorporate other environmental variables and increase the number of species records in the models. These additions would help to clarify whether the relationships observed in this study hold true when additional records are available, and whether, with more data, more nuanced relationships can be discovered.

3.4. Model limitations

The model outputs presented above only provide modest knowledge about the association between species and environmental variables. To some extent, this limitation is due to the species data. The occurrence data used in this study may be uncertain due to inaccuracy in geo-referencing processes, especially species records derived from the GBIF database that were mostly obtained from herbarium specimens. Moreover, only relatively few records were available. This might lead to inaccurate predictions of species distribution [10]. The survey data used in this study are also clustered spatially within parks and other locations. This type of data may fail to represent the heterogeneity of environments across the study area. Cautious interpretation of the model outputs is therefore required to avoid prediction biases.

Another emerging issue that needs to be considered in this study is the use of various datasets from different sources in building the models. This can lead to errors since each dataset uses different sampling methods and different sampling efforts. SDMs assume that each sample used in the models is independent and that sampling effort is consistent across the sample [25]. A problem with combining datasets is that replicate records are frequently found at one site or close by [10] point out that sampling bias in this type of data can be minimized by only using background points in grid cells where at least one species was identified and that inaccessible areas with no data can be excluded from model estimation.

Environmental datasets also contribute to uncertainty in predictions for species distributions. Many of these are only available at coarse resolution. The complexity of species habitat associations may be not well represented due to the limitations of the geospatial database in providing fine scale information that can describe all parameters that are relevant to habitat suitability. Moreover, errors may occur in
the available environmental datasets due to errors in interpolation processes [25]; [19]. Given that this study only used seven environmental covariates, further research in this area could include a broader suite of environmental variables that have been reported in the literature review as influential for the distribution of similar tropical tree species.

4. Conclusions
SDM is a useful tool to predict species distributions based on species habitat associations, but several constraints may hinder robust predictions. Using limited species records and environmental data as predictors, the models developed in this study produced varying predicted distribution maps. The number and type of species presence and background points used in the models also influence model outputs. According to the MaxEnt models 1 and 2 that use the GBIF presence records, S. leprosula has a higher likelihood of occurring in the middle and northern part of Kalimantan. These areas are identified as areas with high mean annual rainfall. Meanwhile, MaxEnt model 3 and the GLM model, which use all presence records, predict less emphasis on middle and northern Kalimantan and some emphasis on south eastern Kalimantan.

In terms of variable importance, mean annual rainfall was the most important predictor in all models except MaxEnt model 3. Location also appears to become an important predictor in the GLM model in which S. leprosula has higher probability of occurrence in locations with primary forests, secondary forests or wetlands. This suggests the use of forest types will provide a more direct approach for predicting species distribution. Soil type is also potentially an important predictor since all MaxEnt models indicate that S. leprosula has preference for soil types that are able to provide water for plants for a long period of time, but finer classification of soil type is necessary to obtain more refined predictions. However, all of these results need to be carefully interpreted due to the characteristics of the data.

Although the models cannot account for large amounts of variation in species distribution, the outputs provide useful information that can be used to improve model performance. Better-planned sampling methods and data sharing mechanisms as well as incorporating more environmental predictors with finer resolution are required to generate more meaningful models. Model development is also important to test hypotheses about whether the distribution pattern of S. leprosula is affected by environmental variables or is merely randomly distributed, as suggested by the neutral theory of species distributions in tropical areas. This research project can also be used as a preliminary study for further development of species distribution models, particularly for plant species in tropical rainforests across Indonesia to support species conservation.

Reference
[1] Syfert M M, Smith M J and Coomes D A 2013 The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models PloS. One. 8(2) 10-20
[2] Margono B A, Potapov P V, Turubanova S, Stolle F and Hansen M C 2014 Primary forest loss in Indonesia over 2000-2012 Nat. Clim. Change. 4 730–735
[3] Ashton P 1998 Shorea leprosula The IUCN Red List of Threatened Species. Version 2014.3. [online], available: http://www.iucnredlist.org
[4] Ashton P S 1982 Dipterocarpaceae / P.S. Ashton, Flora Malesiana: Series I Spermatophyta vol. 9, pt. 2 (The Hague: Boston: M. Nijhoff)
[5] Cayuela L, Golicher D, Newton A, Alkemade J and Pérez A 2009 Species distribution modelling in the tropics: problems, potentialities, and the role of biological data for effective species conservation Trop. Conserv. Sci. 2(3) 319-352
[6] Rondinini C, Wilson K A, Boitani L, Grantham H and Possingham H P 2006 Tradeoffs of different types of species occurrence data for use in systematic conservation planning Ecology. Letters. 9(10) 1136-1145
[7] Franklin J 2010 Mapping Species Distributions: Spatial Inference and Prediction (Cambridge: Cambridge University Press)
[8] Elith J and Leathwick J 2009 The contribution of species distribution modelling to conservation prioritization Spatial Conservation Prioritization: Quantitative Methods (Oxford: Oxford University Press) pp 70-93

[9] Bakre E and Rycroft S 2015 BioAcoustica: Wildlife Sounds Database. (Copenhagern: GBIF)

[10] Phillips S J, Dudík M, Elith J, Graham C H, Lehmann A, Leathwick J and Ferrier S 2009 Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data Ecological Applications Ecological. Society. of. America. 19(1) 181-197

[11] McGray H, Hammill A, Bradley R., Schipper L and Parry J E 2007 Weathering the storm: options for framing adaptation and development (Washington, DC: World Resources Institute) p 57

[12] Phillips S 2005 A Brief Tutorial on Maxent (New York: AT&T Research)

[13] Burnham K and Anderson D 2002 Model Selection And Inference? A Practical Information-Theoretic Approach (Berlin: Springer)

[14] Ferrier S and Watson G 1997 An Evaluation of the Effectiveness of Environmental Surrogates and Modelling Techniques in Predicting the Distribution of Biological Diversity (Camberra: Departement Environment Australia)

[15] Pearce J and Ferrier S 2000 Evaluating the predictive performance of habitat models developed using logistic regression Ecol. Model. 133(3) 225-245

[16] Fox J and Weisberg S 2011 Diagnosing Problems in Linear and Generalized Linear models in An R Companion to Applied Regression (Thousand Oaks, CA: SAGE Publications Inc.) pp 285-328.

[17] Syfert M M, Smith M J and Coomes D A 2013 The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models PloS. One. 8(2) 10-20

[18] Austin M 2007 Species distribution models and ecological theory: a critical assessment and some possible new approaches Ecol. Model. 200(1) 1-19

[19] Phillips S J, Anderson R P and Schapire R E 2006 Maximum entropy modeling of species geographic distributions Ecol. Model. 190(3) 231-259

[20] Margrove J A, Burslem D F R P, Ghazoul J, Khoo E, Kettle C J. and Maycock C R 2015 Impacts of an Extreme Precipitation Event on Dipterocarp Mortality and Habitat Filtering in a Bornean Tropical Rain Forest Biotropica. 47(1) 66-76

[21] Wright J S 2002 Plant diversity in tropical forests: a review of mechanisms of species coexistence Oecologia. 130(1) 1-14.

[22] USDA N 1998 Keys to Soil Taxonomy (Washington DC: USDA)

[23] Hubbell SP 2005 Neutral theory in community ecology and the hypothesis of functional equivalence Functional. Ecology. 19(1) 166-172

[24] Bell G 2005 The co-distribution of species in relation to the neutral theory of community ecology Ecology. 86(7) 1757-1770.

[25] Elith J and Franklin J 2013 Species Distribution Modeling (Amsterdam: Elsevier Inc)