Full Length Research Paper

Predictive modelling of the distribution of two critically endangered Dipterocarp trees: Implications for conservation of riparian forests in Borneo

Minerva Singh

School of Geography and Environment, University of Oxford, UK and Department of Plant Sciences, University of Cambridge, CB23EA, UK.

Riparian forests of Malaysian Borneo exhibit high tree species diversity. However, many of the tree species found in these riparian forests are conservation dependant, with their current conservation status (*sensu* IUCN Red List) varying from vulnerable to critically endangered. The present study had two objectives. Firstly, to identify the environmental factors associated with the distribution of two critically endangered tree species, *Shorea johorensis* and *Shorea inappendiculata* using a small number of occurrence records. Secondly, the research seeks to predict suitable habitat and distribution of these two species. The occurrence data and environmental variables are incorporated within a maximum entropy (MaxEnt) model to predict the distribution of the species and identify the environmental variables that influence the distribution. The research shows that for a small study area, the bioclimatic variables are relatively insignificant while factors such as, land use, tree cover play a prominent role in determining distribution of tree species.

Key words: MaxEnt, Borneo, riparian forests, land use changes, *Shorea johorensis*, IUCN, critically endangered

INTRODUCTION

Over the past few decades, the forests of South East Asia, especially those in Malaysian Borneo have undergone significant deforestation. The land use type comprises of a mosaic of human modified land uses- logged forests, forest fragments, riparian forests, oil palm plantations and pristine forests. Riparian forest corridors vary from natural riverine forest (such as gallery forests of tropical savannas) to remnant riparian buffer zones that have spared deforestation (Lees and Peres, 2008). The presence of riparian corridors has been associated with increasing regional species richness (Sabo et al., 2005). Riparian corridors of the study area in Malaysian Borneo contain a number of conservation dependant and vulnerable tree species notably the Dipterocarp species,

including IUCN red listed Critically Endangered species like *Shorea johorensis*, *Shorea inappendiculata, Dipterocarpus submellatus* and *Hopea nutans* (IUCN, 2008). Given the continuing levels of deforestation that the region faces, these riparian zones may be the last refuge for critically-endangered Dipterocarp tree species that have lost a significant portion of their original habitat. Hence, it is important to both evaluate the distribution of these species and identify the environmental factors which influence this distribution as a way of establishing conservation priorities.

Species distribution modelling has become a popular technique for the identification of suitable habitat and the evaluation of species' distribution for a wide variety of

E-mail: Ms2127@cam.ac.uk.

taxa. This modelling technique is applied in the present study in order to model the potential distribution and habitats of two critically-endangered Dipterocarp species, *Shorea johorensis* and *Shorea inappendiculata,* which are restricted to lowland forests of Borneo and Sumatra (Indonesia). Specifically, the Maxent modelling is based on a small dataset of presence-only records from riparian margins.

Maxent based SDMs have wide variety of applications. Applications range from estimation of species ranges (Moreno et al., 2011), identification of suitable habitats, establishing conservation priorities (Wilting et al., 2010) and predicting range shifts under future climate change scenarios (Thomas et al., 2004). Owing to its ability to produce useful results with a very small presence data, Maxent has proven itself to be useful for modelling the distribution of rare and endangered species.Tinoco et al. (2009) used Maxent to generate a species distribution model of the Violet throated metal tail hummingbird, a globally endangered bird species which is endemic to south-central Ecuador. The modelling was carried out using a limited species occurrence record. The Maxent model was able to identify that the species was restricted to small pockets in the Andes and has an extent of 2000 sq km. Further, the model helped identify the limiting factors of species distribution which included the presence of deep river canyons. The model also identified three distinct suitable habitats vital for species persistence. Thorn et al. (2009) modelled the distribution of a rare and nocturnal species of primates, the Asian slow loris in Maxent using 20 environmental variables along with information on protected areas to identify both the suitable habitat for the species and to prioritize different areas according to risk. SDMs can aid conservation planning of little known taxa or species with little survey data by highlighting unknown populations, suitable areas for reintroduction, key areas that could be studied in future and provide an assessment of potential risks (Thorn et al., 2009). Kumar and Stohlgren (2009) predicted the distribution and potential habitat of a critically endangered tree species in New Caledonia. Maxent modelling was used to carry out multi-species modelling to model the distribution of 56 endangered pinus tree species in Mexico. The modelling results were further used to evaluate if the pinus tree species are getting sufficient representation in the protected areas (Gutiérrez and Duivenvoorden, 2010).

This research will use species presence data of afore mentioned tree species within the Maxent modelling framework. The objectives of this research are to: (1) to predict suitable habitat and distribution for the recorded riparian tree species using a small number of occurrence records to inform conservation planning in a mixed landscape in the Malaysian Borneo; (2) to identify the environmental factors associated with species habitat distribution. This research uses species occurrence records, environmental layers (bioclimatic, land use and topogra-

phic data) within the using the Maxent model of maximization of entropy to identify suitable habitats and environmental factors which influence species distribution.

MATERIALS AND METHODS

Study species and occurrence data

Records of species occurrence were collected while carrying out fieldwork in the riparian forests of a mixed landscape comprising of forests that had undergone varying levels of logging, oil palm plantations and intact forests (117.5E, 4.5N) in Sabah, Malaysian Borneo. The entire study area has an approximate size of 75,000 hectares and the riparian zones are a small part of the landscape (Turner et al., 2011). During the course of the fieldwork, it was discovered that many of the tree species in the riparian margins such as *S. johorensis* are on the IUCN Red List of critically endangered species. These tree species (along with other tree species of the riparian forests) are now mostly restricted to patches and strips along the streams of low lying areas. The habitat and survival of these species is being threatened by deforestation and conversion of surrounding areas to oil palm plantations. This study considers two critically endangered tree species belonging to the Dipterocarpacae family- (a) *Shorea johorensis* (b) *Shorea inappendiculata* (IUCN, 2008). Maxent allows multiple species to be modelled simultaneously. Modelling of closely related species or species from the same family in this fashion may help identify the areas where the species occur and provide useful information on biogeographical patterns (Costa et al., 2010).

Environmental variables

Bioclimatic, topographic and land use related variables were used for modelling the distribution of the riparian tree species. A review of literature was carried out to identify which bioclimatic variables may explain the distribution of the tree species. The bioclimatic data included were: (a) Maximum temperature (b) minimum temperature (c) precipitation. These data were obtained from the WorldClim dataset (Hijmans et al., 2005). Digital Elevation Model (DEM) data was obtained from WorldClim dataset. DEM data was further used to calculate slope (in degrees) using the Spatial Analyst functionality of the ArcGIS 10. These data were further resampled to a 1 sq km resolution. In addition to the bioclimatic and topographic variables, the research uses the land use land change (LULC) information of the study area.

LULC map was generated using Landsat TM and ground truth data. An NDVI map of the region was also incorporated as one of the environmental variables. NDVI means Normalized Difference Vegetation Index and is a remote sensing based indicator of live green vegetation in the study area. NDVI can be used as a proxy for the health of the vegetation, plant growth and biomass production. In this research, NDVI was derived from data obtained from the SPOT satellite and was resampled to a 1 sq km resolution. Percent tree cover data was obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) vegetation continuous field (VCF) data. This data has been included for the purpose of giving an indication of tree cover and deforestation in the entire landscape. Deforestation (and ensuing edge effects) in the surrounding landscape can have detrimental effect on the tree species isolated in the riparian margins. Since, the riparian forest zones are located near rivers, hydrological data were also included in the modelling process. A river drainage map and river flow map of the region was obtained from the HydroSHEDS database of the United States Geological Survey (USGS) data archive (USGS, 2012) and incorporated in the analysis. The latter was used to gene-

rate a layer representing Euclidian distance from the river. Distance from the rivers can be seen as a proxy to forest productivity which is higher at a river's edge (Cattau, 2010).

Modelling technique

A body of literature indicates that maximum entropy or Maxent modelling techniques perform better than many different modelling methods (Ortega-Huerta and Peterson, 2008) and may remain effective, despite small sample sizes (Benito et al., 2009). Further, it requires only species presence data and environmental variable (continuous or categorical) layers for the study area. This makes it a suitable choice for the species data and environmental variable data that has been collected in the present study.

In order to carry out the modelling of the riparian tree species, the research has used the freely available Maxent software, version 3.1 (http://www.cs.princeton.edu/~schapire/maxent/). Maxent is a maximum entropy based machine learning program that estimates the probability distribution for a species' occurrence based on environmental variables (Phillips et al., 2006). The environmental variable values at the presence localities impose constraints on the unknown distribution. The maximum entropy approach then approximates an unknown distribution using the known occurrences and background points (all points/grid cell values in the study region) that maximizes entropy, subject to the constraints imposed by the known occurrences. The result of Maxent shows a map where every grid has a value of 0 to 100 (if the result output format is set as cumulative) or 0-1 (if the result output format is selected as logistic); this represents the estimate of relative probability of species occurrence. Maxent is not strongly influenced by the number of environmental parameters used to build models because it ignores those that are non-informative, and uses regularization techniques to avoid over-parameterization (Phillips et al., 2006).

The study made use of presence records (collected from the field) and the afore-described environmental variables to model the potential habitat of the species under consideration and identification of suitable habitats. Significant sources of uncertainties exist in SDMs making it important to validate the results obtained from these models as a way of verifying the robustness of the model. The model generated by Maxent firstly was evaluated by their area under the curve generated for the model. The test data of models have a value of 0.767 (*S. johorensis*) and 0.864 (*S. inappendiculata*). Models that have an AUC value greater than 0.75 are considered to be useful (Elith, 2006). However, this is not the only method of validating the results of the Maxent model. Araujo et al. (2005) illustrate the use of independent validations and the feasibility of doing so by presenting the case of study of the observed distributional shifts among 116 British breeding bird species over a two decade period. However, it is difficult to obtain an independent datasets in many cases, especially when the species observation data maybe small. Redistribution is widely used in the assessment of accuracy. A part of data is used to calibrate/train the model and the other part is used to validate or test the model. Data partitioning techniques can be used to address the problem associated with redistribution methodologies. Some of these techniques include one-time data splitting of calibration and validation datasets. Although, no exact specifications exists, a review of the literature reveals that the models may be calibrated using 70% of the dataset sample obtained at a given point in time and the predictive accuracy of the model was evaluated using the 30% of the remaining data (Araujo et al., 2005).

However, this approach may not work with a small number of samples because the 'training' and 'test' datasets will be very small (Pearson et al., 2007). Hence it was decided to follow the jack-knife validation methodology developed by Pearson et al. (2007), which is shown to be effective for small sample sizes. Under the principle of these techniques, one locality/occurrence point is removed from the dataset, and the model built using the remaining '*n*–1' locality points. Thus, for a species with '*n*' localities, '*n*' individual models will be built for testing. Model accuracy and significance were evaluated based on the ability of each model to predict the one excluded test locality as present (Pearson et al., 2007).

RESULTS AND DISCUSSION

The Maxent model successfully predicted suitable habitats for both tree species. The predicted probability for the presence of the *S. inappendiculata* and *S. johorensis* is shown in Figure 1.

Probability of presence estimate has been defined on a 0-1 scale (logistic format selected for expressing the output data. This format is preferred over others are it expresses the estimates of probability of occurrence as predicted by included environmental variables, thus providing a comparatively more accurate interpretation of output data; Baldwin, 2009) and areas having a value greater than 0.5 could be considered to have suitable habitat for the persistence of the species (Stabach et al., 2009). However, areas most suitable for supporting the tree species are fragmented, especially in the case of *S. johorensis*. A visual examination of Figure 1 reveals that the distribution of species follows the pattern of river flow, while the presence of steeper slopes restricts the distribution of species; in addition, areas sited at higher altitudes were found to have the least suitable habitat for supporting the tree species under consideration. Further, the Maxent model also allows for performing an internal jack-knife test to quantify the importance of the variables in influencing the distribution of both tree species. The results are shown in Figure 2.

In Figure 2, altitude refers to elevation in metres (obtained from DEM); 'drainage' refers to the river drainage profile and pattern of the region; 'flow_corr' is the distance from the rivers; 'lulc_res' refers to the Landsat based land use land cover (LULC) map of the region which defines the different land use categories of the region; 'max_temp' and 'mean_temp' refer to the maximum and mean temperature of the study area; 'ndvi_spot' is the NDVI map of the study area; 'preceptn' is the annual rainfall; 'slope' is slope of the study area in degrees (obtained from DEM); 'tree_cov' is the percentage tree cover of the area, obtained from MODIS

Figure 2 indicates that in the case of *S. johorensis*, environmental variable with highest gain when used in isolation is 'flow_corr.' (or distance from the rivers), which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is also 'flow_corr.', which therefore appears to have the most information that is not present in the other variables. In the case of *S. inappendiculata*, the environmental variable with highest gain when used in isolation is altitude. This appears to have the most useful information by itself. The environ-

Shorea inappendiculata
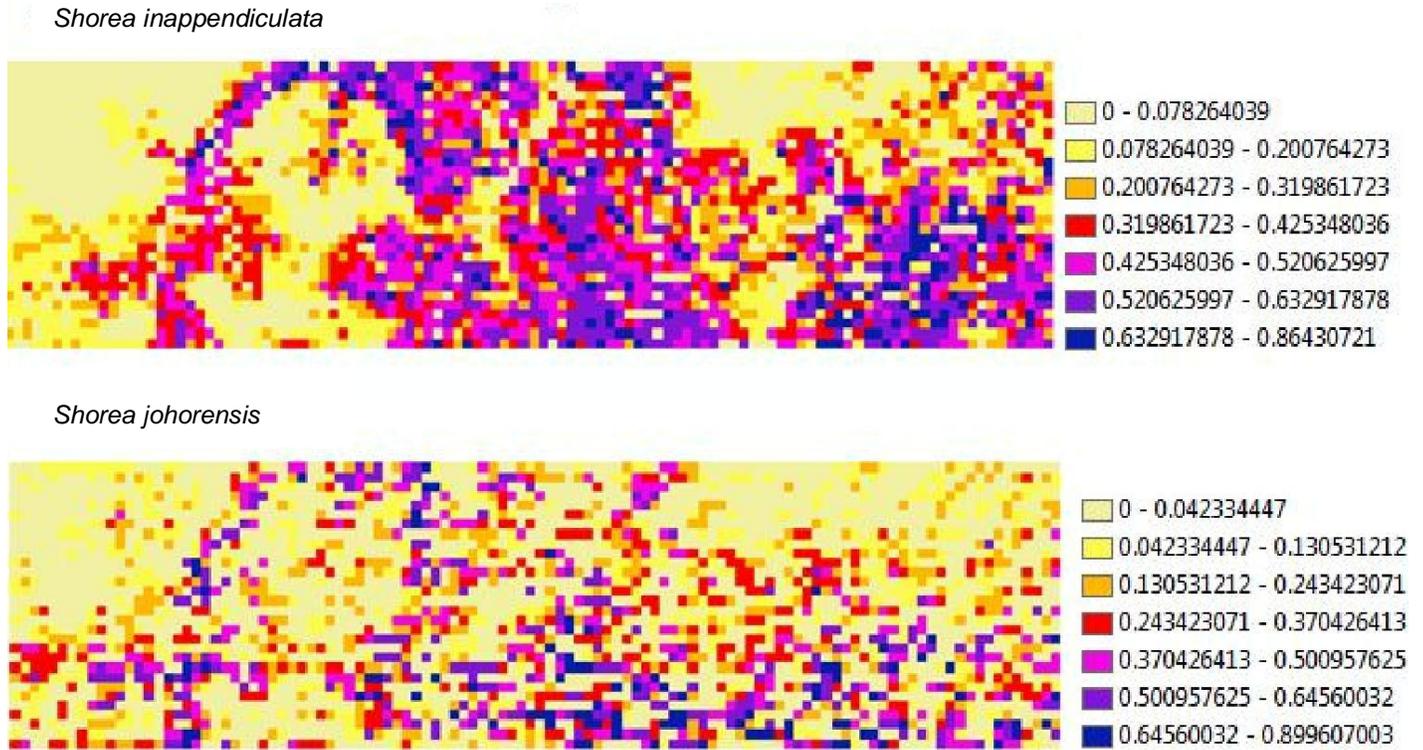
Shorea johorensis

Figure 1. Predicted suitable habitat for *Shorea inappendiculata* and *Shorea johorensis.*
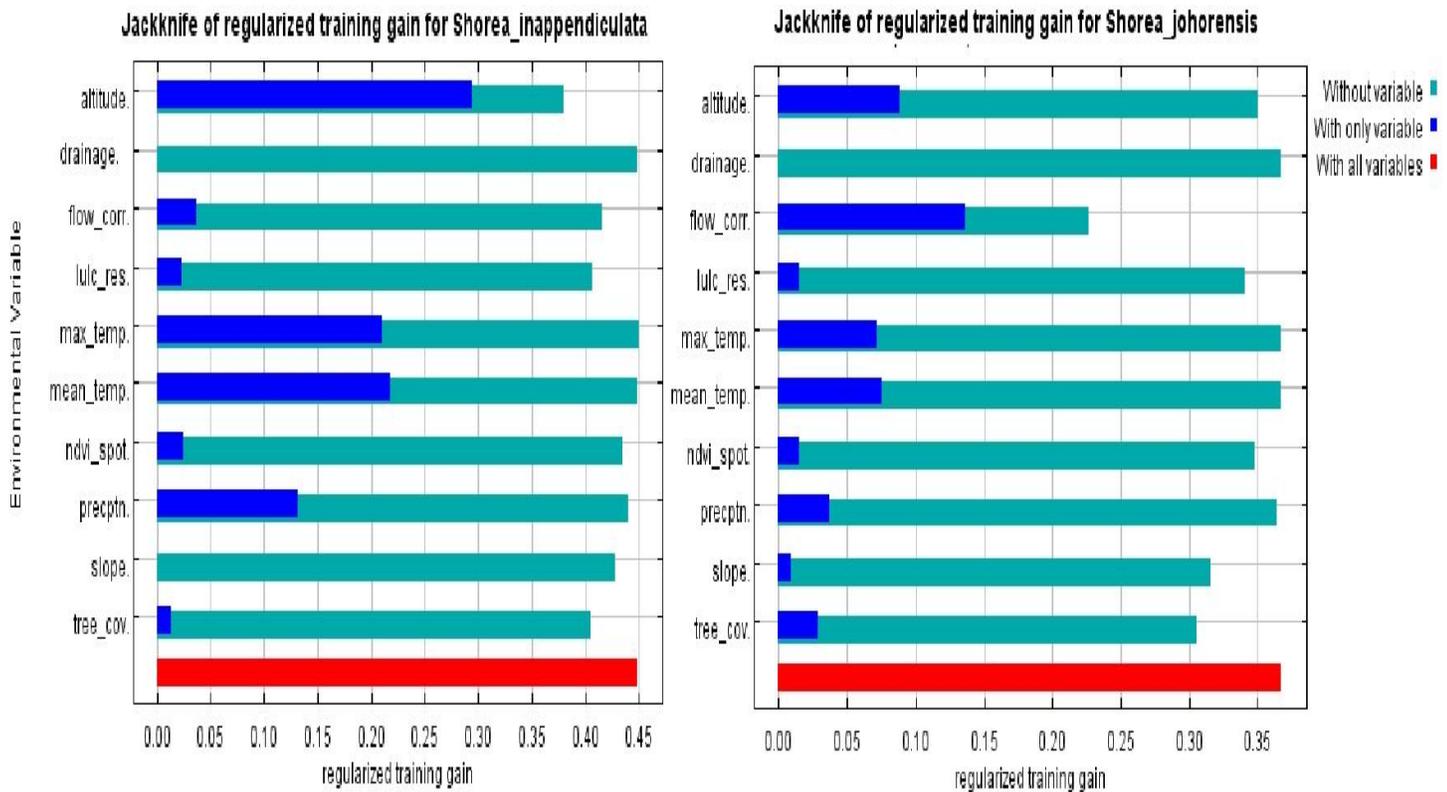
Figure 2. Results of jack-knife evaluations indicating the relative importance of predictor variables for *Shorea inappendiculata* and *Shorea johorensis* in the Maxent model.

Table 1. Selected environmental variables and their percentage contribution.

Variable	*Shorea johorensis*	*Shorea inappendiculata*
Altitude	24.4	53.3
Flow_corr	39.3	23.2
Tree_cov	15.4	6.5
Slope	9.2	3.3
NDVI_spot	3.5	2
LULC_res	5.8	6.7
Max_temp	1.5	0
Preception	0.8	0
Drainage	0	0

mental variable that decreases the gain the most when it is omitted is altitude. This therefore appears to have the most information that is not present in the other variables.

The Maxent model also quantifies the percentage contribution of the predictor variables in influencing the distribution of the tree species as shown in Table 1.

On the basis of this analysis, it may be argued that distance from the rivers and altitude are the most environmental variable that influence the presence and distribution of the riparian tree species under consideration. Both *S. johorensis* and *S. inappendiculata* are species of lowland forests and altitude can be seen as a limiting factor to their distribution. However, *S. inappendiculata* is less restrictive in its choice of habitat as compared to *S. johorensis* and is mainly dependant on topographic features. On the other hand, *S. johorensis* in addition to requiring suitable topographical conditions is also significantly dependant on eco-geographical characteristics such as distance from rivers and land use and quality characteristics such as percentage tree cover and NDVI.

This study has achieved its main objective of predicting the distribution and suitable habitats of the recorded riparian tree species on the basis of a relatively small number of presence records and environmental predictor variables. Further, the study has been able to identify the environmental factors that influence and limit the distribution of the species under consideration. The latter establishes the role played by topographic and land use patterns in influencing the tree species distribution in the study area. The research shows that for the small study area discussed in this paper, bioclimatic variables play a relatively insignificant role in determining the presence and distribution of tree species. However, other factors such as land use and vegetation quality factors such as tree cover, land use land change dynamics play an important role in influencing the presence and distribution of species. Additionally topographic factors such as altitude, distance from the rivers and slope also play an important role in influencing the presence and distribution of species.

These findings in turns have deep ramifications for conservation planning. Firstly, even though both the critically endangered species belong to the same family, their distributions are significantly different and environmental variables influenced their distribution differently. Hence, different conservation strategies may be required for conservation of species which belong to the same family and have the same conservation status. The results indicate that distance from rivers is an important determining factor for distribution of both species. For instance, it may be argued that in the immediate future it is important to focus on land use and vegetation quality factors as a way of ensuring the persistence of the tree species. One of the instances that can be noted is that of the Brazilian forestry legislation which requires the maintenance of riparian corridors on all private land holdings and these are required to have a pre-determined width (Lees and Peres, 2008). Given the role that distance from rivers plays in influencing the presence and distribution of tree species, laws similar to the aforementioned law may be considered as a way of protecting tree species in the study area.

In addition to land use changes, climate change is regarded as an important driver of biodiversity change. However, in this study, bioclimatic variables play a non-significant role in influencing the presence and distribution of the tree species. This may be attributed to two reasons. Firstly, the study area is small (covering only a small part of Malaysian Borneo).

Hence, it may be argued that while land use, topography and vegetation vary significantly over the landscape, climatic factors remain fairly constant throughout the study area. Secondly, bioclimatic variables are based on the interpolations of global climate data and thus have a coarse resolution. Hence, it is important to evaluate if the inclusion of finer scale climatic variables could improve the predictions of the SDM, especially when considering species response under future climate change scenarios.

Conclusions

The results of this study and the literature discussed previously indicates that Maxent can be useful in predicting species distribution and subsequently establishing conservation priorities both at local and regional scales. However, SDMs are fraught with significant uncertainties. Mainly uncertainties in the predictions from SDMs stem from the basic assumptions of the models, algorithms used, parameterization, the variables included for analysis or even the spatial scale of the variables (as demonstrated by Randin et al., 2009). Most SDMs are based on species data collected as a result of sampling carried out at a given point in time/space and the working postulate justifying the use of these data is that the species in question are in pseudo-equilibrium with their environment (Guisan and Thuiller, 2005). Further, sampling design too

can introduce biases such as those stemming from incomplete sampling to focus on a particular geographic space as opposed to random sampling (Zimmerman et al., 2010). While Maxent offers the advantage of being able to use small samples, the accuracy of the models maybe compromised by sampling bias. Further, the transferability of Maxent results between sampled and unsampled areas needs to be interpreted with caution (Baldwin, 2009). It is important to minimize the inherent uncertainties in SDMs. This may be accomplished by the use of link different models, fine scale data (as opposed to coarse scale) and collection of detailed field records.

ACKNOWLEDGEMENTS

REFRENCES

Araujo MB, Pearson RG, Thuiller Erhard WM (2005). Validation of species–climate impact models under climate change. Glob. Chang. Biol. 11: 1504-1513.

Baldwin RA (2009). Use of maximum entropy modelling in wildlife research. Entropy11:854–866.

Benito BM, Martınez-Ortega MM, Munoz LM, Lorite J, Penas J (2009). Assessing extinction-risk of endangered plants using species distribution models: a case study of habitat depletion caused by the spread of greenhouses. Biodiv. Conserv. 18:(9)2509-2520

Cattau M (2010). Using the Ecosystem Service Value of Habitat Areas For Wildlife Conservation: Implications of Carbon Rich Peat swamp Forests For the Bornean Orang-utan, http://www.outrop.com/uploads/7/2/4/9/7249041/cattau_megane_mastersproject.pdf

Costa G, Nogueira C, Machado R, Colli G (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. Biodivers. Conserv. 19:883–899

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Guisan A, Thuiller W (2005). Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8: 993-1009.

Gutiérrez JA, Duivenvoorden JF (2010). Can we expect to protect threatened species in protected areas? A case study of the genus *Pinus* in Mexico Revista Mexicana De Biodiversidad 81: 875-882.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25: 1965-1978

Kumar S, Stohlgren TJ (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree Canacomyrica monticola in New Caledonia. J. Ecol. Nat. Environ. 1(4): 94-98.

IUCN (2008). 2008 IUCN Red List of threatened species. <http://www.iucnredlist.org>

Lees A, Peres CA (2008). Conservation value of remnant riparian forest corridors of varying quality for Amazonian birds and mammals. Conserv. Biol. 22: 439-449.

Moreno R, Zamora R, Molina JR, Vasquez A, Herrera MA (2011). Predictive Modelling of Microhabitats For Endemic Birds in South Chilean Temperate Forests Using Maximum Entropy (MaxEnt). Ecol. Inform. 6: 364-370

Ortega-Huerta MA, Peterson AT (2008). Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. Revista Mexicana De Biodiversidad 79: 205-216.

Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeo. 34: 102- 117.

Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modelling of species geographic distributions. Ecol. Model. 190: 231-259.

Randin CF, Engler R, Normand S, Zappa M, Zimmermann, NE, Pearman, PB, Vittoz P Thuiller W. and Guisan A. 2009: Climate change and plant distribution: local models predict high-elevation persistence. Chang. Biol. 15:1557-1569.

Sabo JL, Sponseller R, Dixon M, Gade K, Harms T, Heffernan J, Jani A, Katz G, Soykan C, Watts J, Welter J (2005). Riparian Zones Increase Regional Species Richness by Harbouring Different, Not More Species. Ecology 86(1)56-62

Stabach JA, Laporte N, Olupot W (2009). International Journal of Biodivers. Conserv. 1(5):177-186

Tinoco BA, Pedro X, Latta SC, Graham CG (2009). Distribution, ecology and conservation of an endangered Andean humming bird Violet throated Metaltail. Bird Conserv. Int. 19:63-76

Thorn JS, Nijman V, Smith D, Nekaris KAI (2009). Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). Divers. Distrib. 15: 289-298.

Thomas CD, Cameron A, Green RE et al. 2004. Extinction risk from climate change. Nature 427:145–147.

Turner EC, Snaddon JL, Ewers RM, Fayle TM, Foster WA (2011). The impact of oil palm expansion on environmental change: putting conservation research into context.in I. M. A. dos Santos Bernardes, editor. Environmental Impact of Biofuels. InTech

United States Geological Survey (USGS) 2012. <http://gisdata.usgs.gov/website/HydroSHEDS/viewer.php>