THE EFFECT OF ECOLOGICAL LIMITING FACTORS WITH TREE SPECIES ON THE ACCURACY OF PREDICTIVE HABITAT MODELS

Chun-Hao Tu[1] Nan-Jang Lo[2] Wei-I Chang[3] Kai-Yi Huang[*4]
[1]Graduate student, Dept. of Forestry, Chung-Hsing University, Taiwan E-mail: eno88888888@hotmail.com

[2]Specialist, EPMO, Chung-Hsing University, Taiwan E-mail: njl@dragon.nchu.edu.tw

[3]Chief, FP Sec., Forest Bureau, Council of Agriculture, Taipei 100, Taiwan E-mail: weii@forest.gov.tw

[*4]Professor, same as with author-a E-mail: kyhuang@dragon.nchu.edu.tw (corresponding author)
250 Kuo-Kuang Road, Taichung, Taiwan 402, Tel: +886-4-22854663; Fax: +886-4-22854663

KEY WORDS: Limiting Factor, Limit of Tolerance, Long-leaf chinkapin (*Castanopsis carlesii*), Japanese Elaeocarpus (*Elaeocarpus japonicus*), Maximum Entropy (MAXENT), Decision Tree (DT), Discriminant Analysis (DA)

ABSTRACT: Prediction of species' potential habitat distribution has become one of crucial issues in applied ecology. To obtain accurate predictive results, we need to examine any factor that is included a predictive model. In this paper, we aimed to examine whether the limits of tolerance of different tree species would influence the accuracy of models for predicting their suitable habitat. Long-leaf chinkapins (LLC) grow widely over mountain region in central Taiwan, but there is a minimum limit of tolerance in elevation above which this species can only grow in central Taiwan. However, Japanese Elaeocarpus (JE) can grow widely from above the sea level to medium elevation mountains. We used GIS to integrate environmental factors (elevation, slope, aspect, terrain position, vegetation index) and the two species, respectively. Then we developed maximum entropy (MAXENT), decision tree (DT), and discriminant analysis (DA) models to predict the suitable habitats of the two species. The results showed that the *kappa* values of DT and MAXENT models with LLC (0.8 and 0.8) were greater than that of DA (0.7), and the same results were with JE (DT: 0.6, MAXENT: 0.6, and DA: 0.5). More importantly, the accuracies of DT, MAXENT, and DA with LLC were much better than those of the three models with JE. It means that LLC has a narrow ecological amplitude in elevation, which plays a key role on the spatial distribution of LLC, thereby substantially raising the accuracy of predictive models, and the opposite is true with JE. The point may explain why the prediction of a rare species is easier than the prediction of a widespread species. To obtain accurate prediction of a widespread species, we shall attempt to find some significant predictors from high spatial resolution, hyperspectral imagery to discriminate subtle differences between different species in a follow-up study.

1. INTRODUCTION

The prediction of the distribution of species' potential habitat has become a major research focus in applied ecology (Miller *et al.*, 2004). To predict the potential habitat accurately, we must consider many aspects in modeling, like the selection of predictive variables or statistical methods (Guisan and Zimmermann, 2000). Every detail in modeling may influence the predictions.

Expecting the factors in model development, the growing characteristics of different tree species may also affect the prediction. Some limits of tolerance on species would form the ecological limiting factors. The law of limiting factors is that the biomass of an organism is potentially constrained by any of the factors critical to its growth and reproduction (Pianka, 1978). The factor that will be the active constraint on the growth of an organism is called the limiting factor (Kaiser *et al*., 1994). When the boundaries of the survival of the organism caused by these limiting factors are clearer, they might assist models to discriminate potential habitat more explicitly. In this study, we aimed to evaluate the effect of these limiting factors for predicting tree species' potential habitat.

Long-leaf chinkapin (LLC, *Castanopsis carlesii*) trees grow widespread in the mountains in central Taiwan and are the important heliophilous species. According to the field survey in the past, this species could only grow above the elevation of **1700** m. Seeds of long-leaf chinkapin have long been identified as an important food source for animals, showing that their value of ecological system has been significant. Japanese Elaeocarpus (JE, *Elaeocarpus japonicus*) trees are the heliophilous and deciduous trees in the mountains in central Taiwan. This species grow throughout from low to mid-high altitude. There is no significant limiting factor for this species.

Maximum entropy (MAXENT) is a novel method, and it has been demonstrated for predictive research in ecology (Elith *et al*., 2006; Hernandez *et al*., 2006; Kumar and Stohlgren, 2009; Peterson *et al*., 2007). Classification techniques like decision tree are common in most investigations (Bourg *et al*., 2005; De'ath and Fabricius, 2000; Felicísimo *et al*., 2004; Landenburger *et al*., 2008; and O'Brien et al., 2005) and regressions like discriminant analysis are also used to analyze the relationship between habitat and environment (Lowell, 1991; and Marnell, 1998), and they can often derive acceptable results.

The objective of this study was to compare the predictions of long-leaf chinkapins and Japanese Elaeocarpus and to evaluate whether the limiting factors of the species affect the predictions. We used a GIS to overlay field tree samples collected with GPS on the layers of altitude, slope, aspect, terrain position (TP), and vegetation indices (VI) derived from SPOT-5 images, to analyze their spatial distribution. Three models, MAXENT, decision tree (DT), and discriminant analysis (DA) models, were developed to predict the potential habitat of the species.

2. STUDY AREA

The study area covers the Huisun Forest Station, which is the property of Chung-Hsing University, in central Taiwan, situated within 24°2´–24°5´N latitude and 121°´–121°7´E longitude (Figure 1). The station has a total area of 7, 477 ha. Its elevation ranges from 454 m to 2, 419 m, and its climate is temperate and humid. Hence, the study area has nourished many different plant species and is a representative forest in central Taiwan. It comprises five watersheds, including two larger, Kuan-Dau watershed at west and Tong-Feng watershed at east. All of the tree samples we collected were from **Tong-Feng watershed** by using a GPS. We developed predictive models by using the samples from Tong-Feng watershed, and then we extrapolated potential habitats throughout the study area.

3. MATERIALS AND METHODS

3.1 Data Collection and Processing

3.1.1 Field Data: Long-leaf chinkapin tree samples were acquired by using a GPS linked with a laser range from the study area. The error of the system would be below one meter after post differential positioning, and the sample data were rectified to the TWD67 (GRS67) Transverse Mercator map projection.

There were 115 long-leaf chinkapins and 104 Japanese Elaeocarpus samples collected in this study, all of them were from the Tong-Feng watershed in the east of the study area. For long-leaf chinkapins, we used 79 out of 115 tree samples for model development (training) and the remaining 36 for model validation in this study. And for Japanese Elaeocarpus, we used 69 out of 104 samples for model development and 35 for validation.

3.1.2 Digital Elevation Model: Digital elevation model (DEM) was acquired from the Aerial Survey Office, Forestry Bureau of the Council of Agriculture. To the requirements of the study, the DEM was interpolated into 5 × 5 m grid size, geo-referenced to TWD67 Transverse Mercator map projection.

The altitude data layer was derived directly from the DEM. Slope and aspect data layers were generated from the DEM by using ERDAS Imagine software.

3.1.3 Orthophoto base maps: We used orthophoto base maps (1:10,000) together with DEM to generate terrain position layer. We calculated the Euclidean distance from each pixel to the nearest ridge and valley, and determined the terrain position by estimating the relative proportions of the distance from each pixel to the ridge and valley. The orthophoto base map was also used to assist in field survey while we took long-leaf chinkapin tree samples.

3.1.3 SPOT-5 satellite images: There were two-date SPOT-5 images we acquired from Center for Space and Remote Sensing Research National Central University. System calibration and geometric correction with level 2B were performed on the images, and then they were rectified to the TWD67 Transverse Mercator map projection and resampled to 5 m resolution by CSRSR, NCU. The information of the SPOT-5 images is shown in Table 1.

We used the SPOT-5 images to generate a vegetation index layer by using the difference ratio of NIR and MIR of two SPOT-5 images to discriminate tree species. The formula of the vegetation index (VI) is:

$$\frac{NIR_{autumn} - MIR_{autumn}}{NIR_{\mathrm{summer}} - MIR_{\mathrm{summer}}} \qquad (1)$$

3.2 Database Building and Sampling

We overlaid four topographic variables, including altitude, slope, aspect, terrain position, and a vegetation index from SPOT-5 images to a GIS database, and the tree sample layer was overlaid with five data layers. Then those pixels of the five layers lying at the same position with tree sample pixels were clipped out. Because background sites (non-target) correspond to the vast majority of the study area, larger variation is expected in environmental characteristics for this group, the number of background pixels (sites) should be three times more than that of target pixels to raise the probability of acquiring a more representative sample of the habitat characteristics at background sites (Pereira and Itami, 1991; Sperduto and Congalton, 1996), and the background sample data were taken from data layers by the random sampling technique to minimize spatial autocorrelation in the independent variables (Pereira and Itami, 1991). There were 500 background samples we used for the training dataset and 250 background samples for the test dataset in this study, about seven times more than the number of target samples.

3.3 Model Development

The models for predicting potential habitat of the trees were created using three statistical methods: (1) **maximum entropy** (MAXENT), (2) decision trees (DT), and (3) discriminant analysis (DA). DA and DT models were implemented by using SPSS software package in this study, and **MAXENT** was implemented by using the software freely available on the worldwide website.

3.3.1 Maximum Entropy

Maximum entropy can make predictions or inferences from incomplete information (Phillips *et al*., 2006), and may remain effective from small sample sizes (Kumar and Stohlgren, 2009). The principle of MAXENT is based on the concepts of thermodynamic entropy, referred to asthe measure of disorder, and then is used to described the probability distribution in several domains, and Bayesian statistics is for exploring the probability distribution of each pixel when the entropy reach the maximum that the state would extremely close to uniform distribution. That is, MAXENT would find out the type of probability distribution that is most likely occurring in the general state. The formula for MAXENT is shown in following:

$$P(x) = exp\left[\sum_{n=1}^{p} \lambda_n \frac{f_n(x) - min_n}{max_n - min_n} - linearPredictorNormalizer\right] \Big/ Z \qquad (2)$$

where

$\frac{f_n(x) - min_n}{max_n - min_n}$: hinge feature;

λ_n : weight coefficient;

*linearPredictorNormalize*r: a constant for numerical stability;

Z: a scaling constant that ensures that *P* sums to 1 over all grid cells.

MAXENT software is freely available on the worldwide web (http://www.cs.princeton.edu/~schapire /MAXENT).

3.3.2 Decision Tree

Decision tree (also called Classification and Regression Trees, CART) is a non-parametric classification algorithm for data mining with both classifying and predicting capability. DT could build classified rules from observations or some experiences (Guisan and Zimmermann, 2000). Decision tree algorithm sequentially partitions the dataset with some important predictors in order to maximize differences on a dependent variable. As show in Figure 1, the decision pathways originate from a starting node (root) that contains all observations, then classify step by step into binary subsets based on the important predictors, and so on. Finally, it will end at multiple nodes containing unique subsets of observations. Terminal nodes are assigned a final outcome based on group membership of the majority of observations (De'ath and Fabricius, 2000; Bourg *et al*., 2005; O'Brien *et al*., 2005).

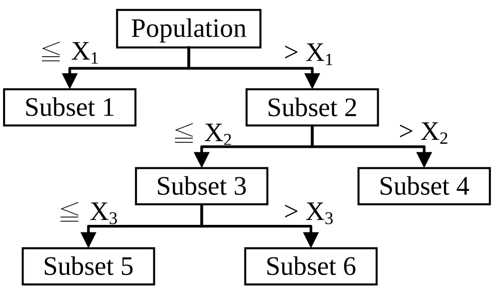

Figure 1. The diagram of the classified process of DT

3.3.3 Discriminant Analysis

Discriminant analysis is a technique, which discriminates among k classes (objects) based on a set of independent or predictor variables. The objectives of DA are to (1) find linear composites of n independent variables which maximize the ratio of among-groups to within-groups variability; (2) test if the group centroids of the *k* dependent classes are different; (3) determine which of the n independent variables contribute significantly to class discrimination; and (4) assign unclassified or “new” observations to one of k classes (Lowell, 1991). The variates for a discriminant analysis, also known as the discriminant function, takes the following form:

$$Y_{j\,k} = \alpha + \beta_1 X_{1\,k} + \beta_2 X_{2\,k} + \ldots + \beta_n X_{n\,k} \qquad (2)$$

where

$Y_{j\,k}$ = discriminant Y score of discriminant function j for object (class) k

α = intercept

β_i = discriminant weight for independent variable i

$X_{i\,k}$ = independent variable i object (class) k

3.5 Model Validation

Accuracy assessment contains the overall accuracy and *kappa* coefficient of agreement of the predictions for two species. The *kappa* coefficient is a measure of agreement between predictive values and observations. The *kappa* value of 1 indicates perfect agreement and the value of 0 indicates agreement equivalent to chance (Viera and Garrett, 2005), and the value higher than 0.8 indicate stronger agreement and the value lower than 0.4 indicate poorer agreement (Jensen, 2005).

4. RESULTS AND DISCUTION

Table 1 is the statistics of field sample data of two species, it shows that the altitude range of JE was much more than that of LLC, that were about 1720–2100 m and 1050–2030 m, respectively. The average slope of JE was also steeper than that of LLC and the range of slope of JE also wider than that of LLC. Especially for altitude, the experience of the field survey in the past also shows that LLC trees could only grow in elevation above about **1700** m. It means that there are significant limit in altitude and slope for LLC, and the limiting factors of JE were fewer than those of LLC.

Table 1. The statistics of long-leaf chinkapins and Japanese Elaeocarpus in study area.

Statistics	Study area					Long-leaf chinkapins					Japanese Elaeocarpus				
	Altitude (m)	Slope (°)	Aspect (°)	TP	VI	Altitude (m)	Slope (°)	Aspect (°)	TP	VI	Altitude (m)	Slope (°)	Aspect (°)	TP	VI
Mean	1314	34	—	5	24	1910	13	—	7	24	1517	24	—	7	27
Mode	1239	37	127	6	22	2095	4	262	7	22	1869	22	339	7	22
Max	2418	89	361	8	119	2097	33	359	8	73	2027	46	358	8	60
Min	445	0	0	1	0	1718	1	6	2	20	1075	2	2	2	20

TP: Terrain Position; VI: Vegetation Index

Because the importance of altitude, slope, and TP was much greater than that of other two factors in five predictors, as show in Table 2, the results as follows were predicted using these three factors. Table 3 shows the accuracies of three models for the two species. The *kappa* values of MAXENT were similar to DT, both in two species, and they were significantly greater than that of DA. It means that the predictions of MANENT and DT were better than that of DA. On the other hand, the accuracies of LLC were much higher than that of JE, all in three models. It means that the prediction of LLC was easier than JE, because there are more growing limits for LLC that can assist to discriminate its habitat from background easily.

Figure 2a-c show the predictive potential habitat maps of LLC and Figure 3a-c show that of JE. These maps indicate that the potential habitat of LLC was more concentrative on the sites at which LLC has higher altitude in the study area and narrower distribution. In contrast, the potential habitat of JE was more dispersed that extended to the area with lower altitude and had wider distribution all in three models. It also indicates that the predictions of LLC had more explicit boundaries for classification.

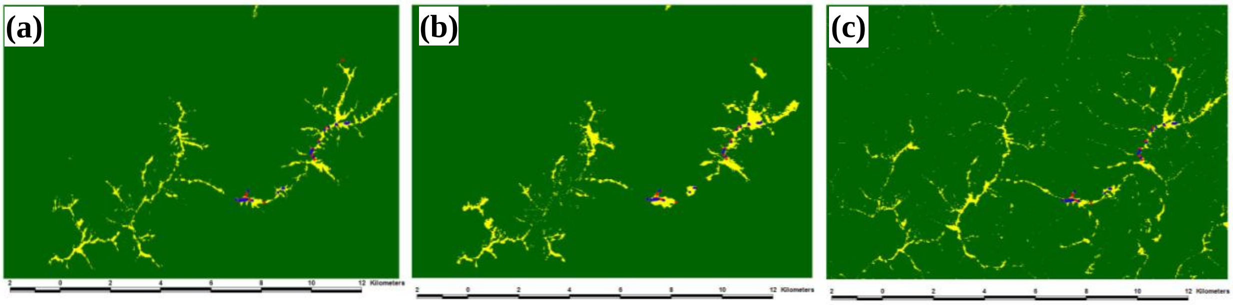

Figure 2. Maps of potential habitat generated from three models of long-leaf chinkapins: (a) MAXENT (The proportion of habitat is 4.2%); (b) DT (2.6%); (c) DA (2.8%).

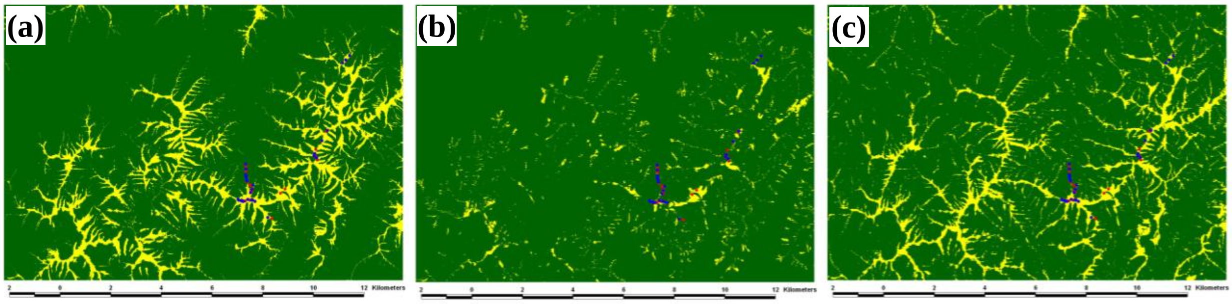

Figure 3. Maps of potential habitat generated from three models of Japanese Elaeocarpus: ; (a) MAXENT (11.7%); (b) DT (3.2%); (c) DA (10.7%).

Table 2. The importance of five predictor variables with three models of two species.

Variable	Long-leaf chinkapins			Japanese Elaeocarpus		
	Percent contribution of MAXENT	Importance of DT	Standardized function coefficients of DA	MAXENT	DT	DA
Altitude	57.2%	89.5%	0.530	38.8%	100.0	0.459
Slope	32.8%	100.0%	-0.661	13.4%	62.0	-0.519
Aspect	0.5%	11.8%	0.072	4.5%	27.2	0.022
TP	9.3%	34.1%	0.337	42.4%	48.4	0.561
VI	0.2%	2.5%	-0.051	0.9%	22.8	-0.065

TP: Terrain Position; VI: Vegetation Index

Table 3. Accuracies of three models for predicting the potential habitats of two species.

Model	Sample set	Long-leaf chinkapins		Japanese Elaeocarpus	
		OA(%)	*kappa*	OA(%)	*kappa*
MAXENT	Training	95.8	0.82	86.6	0.52
	Test	95.8	0.80	88.8	0.57
DT	Training	97.0	0.87	94.4	0.70
	Test	95.4	0.80	92.3	0.61
DA	Training	89.7	0.65	80.3	0.41
	Test	92.6	0.72	83.9	0.50

OA: Overall Accuracy; PA: Producer's Accuracy; UA: User's Accuracy.

5. CONCLUTIONS

The accuracies of MAXENT and DT models were much higher than that of DA model. It indicates that MAXENT and DT models were better suited for predicting these two species.

More importantly, the results in this study indicate that the ecological limiting factors affected the prediction of the two species' potential habitat because the limiting factors may assist model to distinguish the potential habitat of LLC. However, the growing characteristic of many species are similar to that of JE that has no significant ecological limiting factors, so we tried to find some information from remote sensing images. The vegetation indices from SPOT-5 images could not improve the accuracy of these models since the spatial and spectral resolutions were not enough to discriminate the subtle difference between species, so we shall attempt to extract more information from high-spatial resolution and hyperspectral imagery to improve model ability for predicting potential habitat of species.

6. REFERENCES

Bourg, N. A., Mcshea, W. J., and Gill, D. E., 2005. Putting a CART before search: successful habitat prediction for a rare forest herb. Ecology, 86 (10), pp. 2793-2804

De'ath, G. and Fabricius, K. E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81 (11), pp. 3178-3192

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L, G., Loiselle, B, A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., and Zimmermann, N. E., 2006. Novel methods improve prediction of species' distributions from occurrence data. ECOGRAPHY, 29, pp. 129-151

Felicísimo, A. M. and Gómez-Muñoz, A., 2004. GIS and predictive modeling: a comparison of methods applied to forestal management and decision-making. Proceedings of GIS Research UK, pp. 143-144

Guisan, A. and Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. Ecological Modeling, 135, pp. 147-186

Hernandez, P. A., Graham, C. H., Master, L. L., and Albert, D. L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. ECOGRAPHY, 29, pp. 773-785

Jensen, J. R., 2005. Introductory Digital Image Processing: A Remote Sensing Perspective, 3rd Edition, Pearson Education, Inc., New Jersey.

Kaiser, M. S., Speckman, P. L., and Jones, J. R., 1994. Statistical models for limiting nutrient relations in inland waters. Journal of the American Statistical Association 89, pp. 410-423.

Kumar, S. and Stohlgren, T. J., 2009. Maxent modeling for predicting suitable habitat for threatened and endangered tree Canacomyrica monticola in New Caledonia. Journal of Ecology and Natural Environment, 1(4), pp. 94-98.

Landenburger, L., Lawrence, R. L., Podruzny, S., and Schwartz, C. C., 2008. Mapping regional distribution of a single tree species: white bark pine in the Greater Yellowstone Ecosystem. Sensors, 8, 4983-4994.

Lowell, K., 1991. Utilizing discriminant function analysis with a geographical information system to model ecological succession spatially. International Journal of Geographical Information System, 5 (2), pp. 175-191

Marnell, F., 1998. Discriminant analysis of the terrestrial and aquatic habitat determinants of the smooth newt (Triturus vulgaris) and the common frog (Rana temporaria) in Ireland. Journal of Zoology, 244, pp. 1-6

Miller, J. R., Turner, M. G., Smithwick, E. A. H., Dent, C. L., and Stanley, E. H., 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. BioScience, 54 (4), pp. 310-320.

O'Brien, C. S., Rosenstock, S. S., Hervert, J. J., Bright, J. L., and Boe, S. R., 2005. Landscape – level models of potential habitat for Sonoran pronghorn. Wildlife Society Bulletin, 33 (1), pp. 24-34

Peterson, A. T., M. Papeş¸ and M.Eaton, 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography, 30, pp. 550-560

Pereira, J. M. C. and Itami, R. M., 1991. GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering & Remote Sensing, 57 (11), pp. 1475-1486

Phillips, S. J., Anderson, R. P., and Schapire, R. E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model, 190, pp. 231-259

Pianka, E. R., 1978. Evolutionary Ecology, 2nd Edition. New York: Harper and Row.

Sperduto, M. B. and Congalton, R. G., 1996. Predicting rare orchid (small Whorled Pogonia) habitat using GIS. Photogrammetric Engineering & Remote Sensing, 62 (11), pp. 1269-1279

Viera, A.J. and Garrett, J. M., 2005. Understanding interobserver agreement: the kappa statistic. Family Medicine, 37(5), pp. 360-363.