ANALYZING THE EFFECTS OF TOPOGRAPHIC WETNESS INDEX ON THE PREDICTIVE ABILITY OF SPECIES DISTRIBUTION MODELS

Wei-Kai Lai[1], Nan-Chang Lo[2], and Kai-Yi Huang[3]

[1] *Graduate Student, Dept. of Forestry, Chung-Hsing University, Taiwan, R.O.C., changewiththetime@hotmail.com*

[2] *Specialist, Experimental Forest Management Office, Chung-Hsing University, Taiwan, R.O.C., njl@nchu.edu.tw*

[3] *Professor, Dept. of Forestry, Chung-Hsing University, Taiwan, R.O.C., kyhuang@dragon.nchu.edu.tw*

[123] *250, Kuo-Kuang Rd., Taichung 402, Taiwan, R. O. C.*

KEY WORDS: Species distribution model (SDM), Topographic wetness index (TWI), Solar Irradiance (SIR), Maximum entropy (MAXENT), Decision tree (DT)

ABSTRACT: Species distribution modeling (SDM) is commonly applied to analyze the relationship between species and environment and predicting its distribution. Japanese elaeocarpus (JE), red-stripe rhododendron (RSR) and Chinese guger-tree (CGT) widely distributed in central Taiwan were chosen as target species. JEs are a xeric pioneer and usually grow on uplands with strong sunlight and low soil moisture. RFs have clumpy distribution and form pure stands and can grow on well-drained uplands with sufficient sunlight and acidic soils. CGTs are shade-tolerant and incline to grow in lowlands with wetter soils but need moderate sunlight and have scattered distribution. Hence, the study attempted to develop the SDMs based on terrain-related variables to predict the suitable habitats of these species in the Huisun area in central Taiwan. The base model included elevation, slope, and terrain position and expanded models included the three variables plus topographic wetness index (TWI), and these models were built by maximum entropy (MAXENT), decision tree (DT), and BIOCLIM algorithms. The accuracies of the expanded models with additional TWI were better than that of the base model, regardless of which algorithms being used. The MAXENT and DT models were equally matched in predictive accuracies, and outperformed BIOCLIM. More importantly, the larger the TWI value, the higher likelihood a CGT tree has to grow at a given location, and the opposite is true for either RSR or JE tree. Consequently, this outcome shows that these species have the above-mentioned ecological traits and agrees with our observations from field surveys. The rise in accuracy is relatively limited although this proxy can improve the predictive ability of these SDMs. Therefore, the proxy of solar irradiance (SIR) and high-resolution DEM derived from LiDAR will be incorporated into SDMS so that their predictive ability can be substantially improved.

1. INTRODUCTION

Species distribution models (SDMs) are used to gain ecological and evolutionary insight and to predict distribution across landscapes, sometimes requiring extrapolation in space and times (Elith and Leathwick, 2009). SDMs can provide a measure of a species' occurrence likelihood in areas not covered by biological surveys and consequently becomes an indispensable tool to conservation planning and forest management (Elith *et al*., 2006), and they have become increasingly important in spatial ecology since the latter half of the 20th century (Guisan and Zimmermann, 2000). Broadly speaking, the applications of SDMs usage can be split into two main frames in which one of them is for explanation and the other is for prediction. Each coin has two sides, many studies have strikingly focused on causal drivers of species distribution, and these studies with such purpose are intensely depending on the modeling algorithms and spatially extensive environmental data (Hamazaki, 2002), and that is also critical to the selection of predictors and models in addition, while others are for prediction which was made to new sites within the known range of environments sampled by the training data and to that new and unsampled geographic region within future time or past climate.

Ecological parameters (e.g. rainfall or sunlight) that are causal factors for species distribution are usually used to predict the spatial pattern of species. However, data for such ecological factors are expensive or even difficult to collect and are usually collected from a limited number of stations or field-survey samples. Terrain-related variables are used as proxies of ecological factors in the study since data for them can be easily obtained by remote sensing (Guisan and Zimmermann, 2000). Topographic wetness index (TWI) is a very useful tool in the context of hydrologic simulation (Izham *et al*., 2011). It is one kind of measurements that can represent and control the local situation of topography on hydrological processes. These processes influence the dynamic of surface and stream flows, and are highly related to soil moisture, which is an obvious candidate for being a controller of local plant growth (Moeslund *et al*., 2013). However, data for soil moisture are expensive to collect and are usually collected with insufficient number. Therefore, this study tried to include TWI, the proxy of soil moisture, as a new predictor into the model in order to improve the performance of SDMs.

The study aimed to evaluate the effects of TWI on the predictive performance of species distribution models. It developed the base model including elevation, slope and terrain position and the expanded model including the same three variables plus TWI for predicting the suitable habitat of three species in the Huisun area in central Taiwan. These models were built by three algorithms of maximum entropy (MAXENT), decision tree (DT), and BIOCLIM.. The study compared the predictive performances between the base model and the expanded model to determine the effectiveness of TWI for improving the performance of SDMs.

2. MATERIALS AND METHODS

2.1. Study Area

The study area is a rectangular area, which encompasses Huisun Experimental Forest Station (HEFS) with irregular shape (7, 477 ha), and it has a total area of 17,136 ha. The HEFS is located in the 24˚2´–24˚5´ N latitude and 121˚3´–121˚7´ E longitude (figure 1). The entire study area ranges in elevation from 454 m to 2,418 m, and its climate is temperate and humid. The environment has nourished a wide variety of plant species more than 1,100 and is a representative forest in central Taiwan. The study area comprises five watersheds, including two larger watersheds, Kuan-Dau at west and Tong-Feng at east. All of the tree samples (three species) were collected from the Tong-Feng and Kuan-Dau watersheds in HEFS by using a GPS.

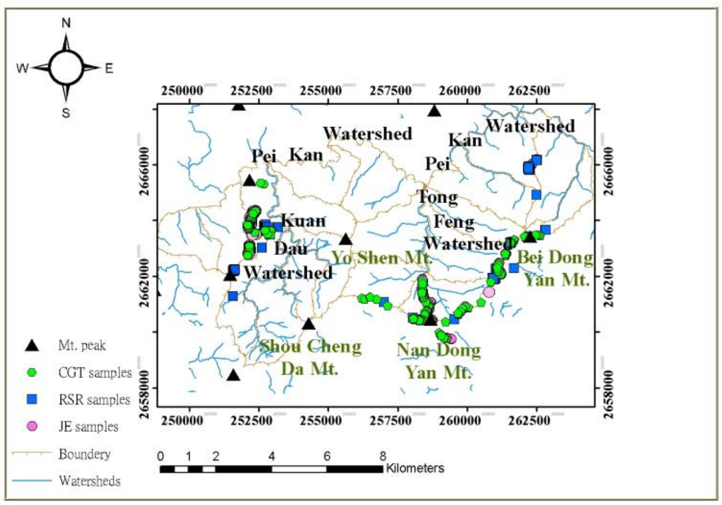

Figure 1 Location map of the study area.

2.2. Target Species

The study chose Japanese elaeocarpus (*Elaeocarpus japonicas*, JE), red-stripe rhododendron (*Rhododendron formosanum*, RSR) and Chinese guger-tree (*Schima superba*, CGT) as target species in order to examine the relationships between their ecological characteristics and environment. JEs are a xeric pioneer, usually grow on uplands with strong sunlight and low soil moisture, and are often accompanied by pine and RSR species. RSRs have clumpy distribution and usually form pure stands, can grow on well-drained uplands with sufficient sunlight and acidic soils, and are often accompanied by pine and JF species, but stay at shrub pattern due to their low growth rate that cannot compete with other trees. CGTs are shade-tolerant evergreen trees and incline to grow in lowlands with wetter soils but still need moderate sunlight and have scattered distribution.

2.3. Data Processing

In situ RSR, CGT and JE samples were collected by using a GPS linked with a 5-m expandable rod and a laser range finder, and then performed a post-processed differential correction that makes them have an accuracy of sub-meters.

The dataset was eventually converted into ArcGIS shapefile format for later use. The sample sizes of the three tree species are shown in table 1. Both digital elevation model (DEM) of 5 m resolution and orthophoto base maps (1:10,000) were also collected. Elevation and slope were derived from DEM by ERDAS Imagine software module. Digitized main ridges and valleys in the study area were used together with DEM to generate terrain position layer.

Table 1 The sample sizes of the three tree species

Watershed/Species	No. of JEs	No. of RSRs	No. of CGTs
Tong-Feng	126	118	130
Kuan-Dau	103	61	64
Total	229	179	194

The calculation of TWI usually uses the gridded DEM and formula as $TWI=\ln(\alpha/\tan\beta)$, where α is the upslope contributing area per unit length of contour, and β is the topographic slope of the cell. The value of TWI depends on the algorithm to calculating α and estimation of $\tan\beta$ (Qin *et al.*, 2007). Here D8 algorithm was adopted and implemented in ArcGIS 10.0 software to compute α. D8 algorithm is one of Single Flow Direction (SFD) that assumes all water from a pixel should flow into one and only one neighboring pixel, which has the lowest elevation (O'Callaghan and Mark, 1984). TWI was computed based on the formula by using Map Algebra tool/Raster Calculator in ArcGIS.

All of these data layers were geo-referenced to the coordinate system, TWD67 (Taiwan Datum, spheroid: GRS1967) and transverse Mercator map projection over two-degree zone with the central meridian 121°E. There were a total four variables, elevation, slope, terrain position, and TWI used in the predictive models.

2.4. Sampling Design and Model Development

The study took a split-sample approach to develop and evaluate predictive models. One sampling design (SD) was created for model development and evaluation through different combinations of tree samples taken from two watersheds in the HEFS. The study merged the tree samples collected from the two watersheds into a dataset and separate the dataset into two subsets, the first subset containing two-thirds of the dataset (2/3) for model calibration and the second subset containing the remaining (1/3) for model evaluation. The base model was built with three predictor variables (elevation, slope, terrain position) and the expanded model was built with the three variables mentioned earlier plus TWI variable. These two types of SDMs were developed by maximum entropy (MAXENT), BIOCLIM, and decision tree (DT) for predicting the distribution of the three species.

2.5. Model Evaluation

Model validation (evaluation) can be done by split-sample validation approach, as mentioned previously. Predictions of each model were compared to the validation data set to form a confusion matrix, from which Cohen's *Kappa* was calculated. The *Kappa* statistic ranges from -1 to +1, where +1 indicates perfect agreement and values of zero or less indicate a performance no better than random (Cohen, 1960; Lillesand *et al*., 2008).

3. RESULTS AND DISCUSSION

3.1. Comparison between Base Model and Expanded Model

Table 2 shows the statistics (mean, mode, maximum, and minimum) of elevation, slope, terrain, and TWI calculated from the entire study area (HEFS) and the target species samples. TWI was the focus of this study. A large TWI value represents a given location with high soil moisture; a small TWI value represents a given location with low soil moisture. By comparison, the mean and mode of TWI values for CGTs (4.8 and 5.9) were not only greater than those of the entire study area (4.3 and 3.5) but also greater than those of JEs (4.2 and 3.5) and those of RSRs (4.2 and 2.7). This outcome indicates that CGTs have a preference for lowlands with wetter soils, and RFs can be better suited to grow on well-drained uplands than JEs because the mode of TWI values for RSRs (2.7) was smaller than that of JEs (3.5). Consequently, the greater the TWI value at a given point, the higher likelihood an GCT tree has to grow at that point, and vice versa, the smaller the TWI value at a given point, the higher an RF tree has the likelihood to grow at that point. JEs frequently accompany with RSRs as TWI is relatively small.

Table 2 The statistics of topographic factors for the entire study area and target species samples

Site	Statistic	Elevation (m)	Slope (°)	TP	**TWI**	Site	Elevation (m)	Slope (°)	TP	**TWI**
Entire study area	Mean	1314	34.9	5	**4.3**	JE sample site	1568	21.9	6	**4.2**
	Mode	1247	37.7	6	**3.5**		1225	19.9	7	**3.5**
	Maximum	2419	79	8	**19.8**		2028	42.7	7	**8.9**
	Minimum	445	0	1	**0.2**		1104	2.9	4	**2.4**
CGT sample site	Mean	1678	19.7	6	**4.8**	RSR sample site	1704	17.7	7	**4.2**
	Mode	1881	1.8	7	**5.9**		1697	8.0	7	**2.7**
	Maximum	2054	36.2	8	**7.8**		2024	47.5	8	**5.7**
	Minimum	1149	1.8	1	**2.1**		1062	8.0	4	**2.7**

TP: Terrain position; TWI: Topographic wetness index

Table 3 shows the accuracies of the base model and expanded model built by MAXENT, DT, and BIOCLIM for predicting the suitable habitat of the three species. For BIOCLIM, the accuracy of the expanded model was better than that of the base model for each of the three species, with a rise in accuracy of 0.05–0.12. For MAXENT, the accuracy of the expanded model was better than that of the base model for each of JE and RSR species, with a rise in accuracy of 0.03–0.04, while the accuracy of the expanded model was less than that of the base model for CGT species, with a drop in accuracy of 0.05. On the contrary, the opposite was true for DT. These results indicate that TWI was useful to improve the predictive ability of SDMs in a greater or lesser degree, depending on modeling species and algorithms. However, the rise in model's accuracy is relatively limited although TWI can improve the predictive ability of SDMs.

Table 3 The predictive accuracies of the base model and expanded model validated by independent sample dataset

	Base model			Expanded model		
	Kappa value			*Kappa* value		
	BIOCLIM	MAXENT	DT	BIOCLIM	MAXENT	DT
JE	0.35	0.59	0.55	0.43	0.63	0.50
RSR	0.30	0.72	0.77	0.35	0.75	0.70
CGT	0.20	0.56	0.50	0.32	0.51	0.55

Also shown in table 3, the MAXENT and DT models equally matched in predictive accuracy, and outperformed BIOCLIM. With the exception of BIOCLIM, the accuracies of MAXENT and DT models for estimating RSR species were higher than those of the two models for estimating CGT and JE species, and the ecological traits of the three species substantially affected the predictive performance of these models. JE and CGT species were found to have a broad and scattered distribution, while RSR species has a specialized, narrow, and clustery distribution, usually forming a pure forest. Specifically, RSR species was hard to compete with many other species in a good environment due to its slow growth rate, but it can grow in a poor environment with thin, acidic, and infertile soils where most species almost cannot grow. Consequently, the ecological traits of species can affect modeling accuracy, and species with a widespread distribution (or broad ecological amplitude) like CGT and JE are generally more difficult for modeling than species with a clustery distribution like RSR.

BIOCLIM is very simple and nearly identical to parallelepiped classifier, and it defines the ecological niche of a species as the bounding hyper-box that encloses all the records of the species in the core-climate based on rectangle (or parallelepiped). However, BIOCLIM often leads to higher commission errors (erroneously assigns many background pixels to target species) and in turn decreases omission errors, thereby raising both overall accuracy and *Kappa* value. In contrast, MAXENT and DT, the two machine learning methods, can make more refined distinction between target species pixels and background pixels, thereby resulting in lower commission errors but higher omission errors and leading to lower overall accuracy and *Kappa* value. The results in table 4 demonstrated the arguments as mentioned above as well. Figure 2 shows the spatial patterns of the three species estimated by the base model and expanded model built by MAXENT (maps produced from DT and BIOCLIM are not shown here). Regardless of predicting which species, the area of suitable habitat of any species estimated by BIOCLIM was much greater than that of any species estimated by either MAXENT or DT.

Table 4 Distribution of the suitable habitat of the three species estimated by the base model and expanded model

		Base model			Expanded model		
	Algorithms	BIOCLIM	MAXENT	DT	BIOCLIM	MAXENT	DT
JE	Area (ha)	5,908.9	1,622.4	1,542.0	4,844.0	1,504.4	1,395.5
	Total area (%)	34.48%	9.46%	8.99%	28.27%	8.77%	8.14%
RSR	Area (ha)	5,908.4	556.0	615.1	4,892.8	534.8	808.8
	Total area (%)	34.47%	3.24%	3.59%	28.55%	3.12%	4.71%
CGT	Area (ha)	8,500.0	1,390.0	1,302.5	6,379.1	1,304.3	1,229.0
	Total area (%)	49.60%	8.11%	7.60%	37.22%	7.61%	7.17%

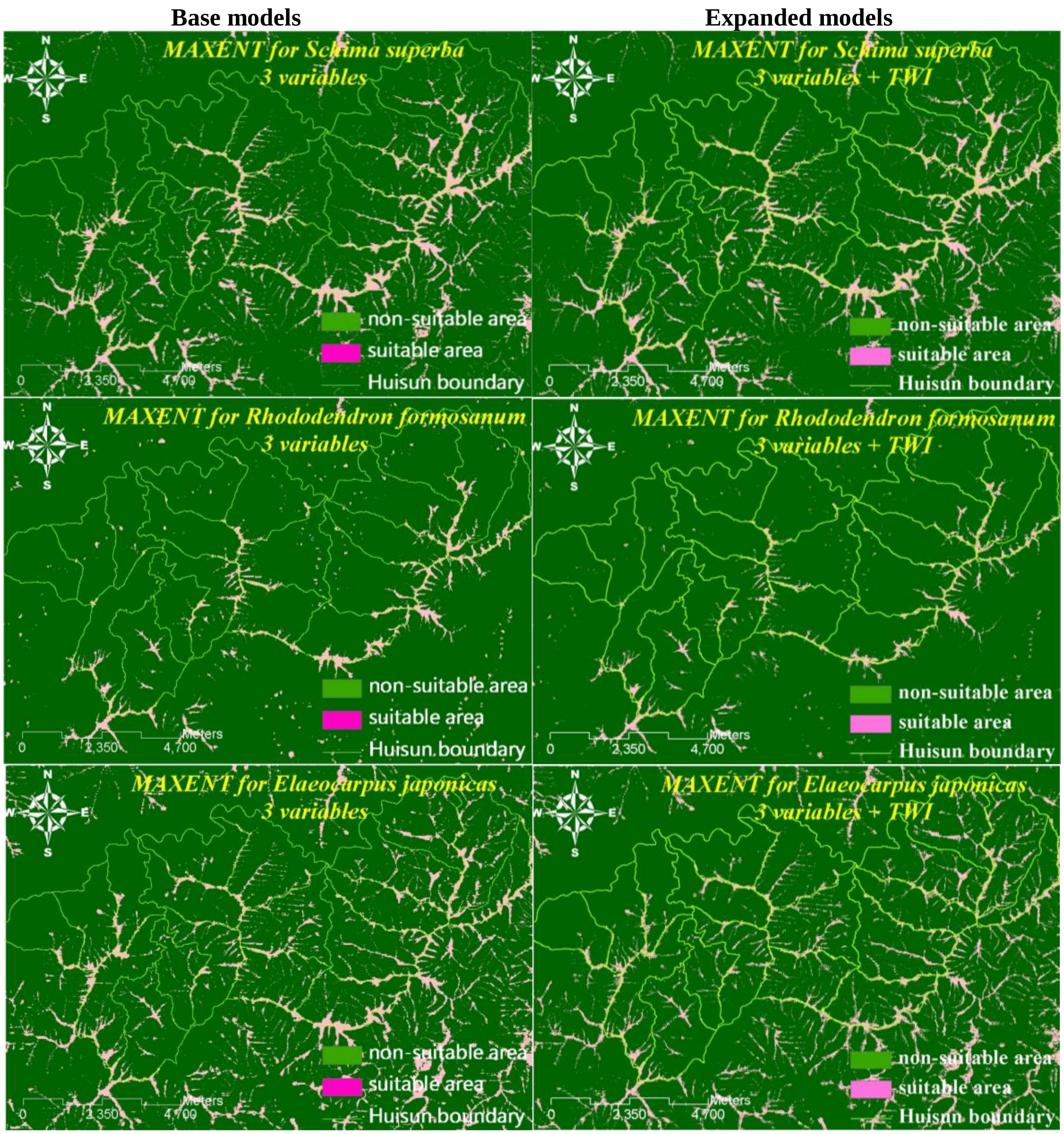

Figure 2 Map of the suitable habitat of three species generated from the base model and expanded model built by MAXENT (left column is base model; right column is expanded model; CGT, RSR and JE from top to bottom, respectively; maps produced by BIOCLIM and DT are not shown here)

4. CONCLUSIONS

The accuracies of the expanded models with additional TWI were better than that of the base model, regardless of which algorithms being used. The MAXENT and DT models were equally matched in predictive accuracies, and outperformed BIOCLIM. More importantly, the larger the TWI value (i.e. represent more soil moisture), the higher likelihood an CGT tree has to grow at a given location, and the opposite is true for either an RSR or a JE tree, i.e. the lower the TWI value (less soil moisture), the higher likelihood either an RSR or a JE tree has to grow at that location, but an RSR tree has higher likelihood than a JE tree does. Consequently, this outcome shows that these species have the above-mentioned ecological traits and agrees with the team's observations from field surveys. The rise in accuracy is relatively limited although this proxy can improve the predictive ability of these SDMs. Therefore, the proxy of solar irradiance (SIR) and high-resolution DEM derived from LiDAR will be incorporated into SDMS so that their predictive ability can be substantially improved.

5. REFERENCES

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20 (1): 37–46.

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

Elith, J., and J. R. Leathwick, 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677–697.

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

Hamazaki T., 2002. Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic ocean (From Cape Hatteras, North Carolina, USA to Nova Scotia, Canada). Mar. Mamm. Sci. 18:920–39

Izham, M. Y., Muhamad Uznir U., Alias, A. R., Ayob, K., Wan Ruslan I., 2011. Influence of georeference for saturated excess overland flow modeling using 3D volumetric soft geo-objects. Comput Geosci 37(4): 598－609.

Lillesand, T. M., Kiefer, R. W. and Chipman, J. W., 2008. Remote sensing and image interpretation. 5th Edition. John Wiley & Sons. Inc., New York.

Moeslund J. E., Arge L, Bøche,r P. K., Dalgaardm, T., Ejrnæs, R., Odgaard, Mv., Svenning, J. C. (2013) Topographically controlled soil moisture drives plant diversity patterns within grasslands. Biodivers Conserv 22: 2151－2166.

O'Callaghan, J. F. and Mark, D. M., 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics, and Image Processing 28: 323–344.

Qin, C., Zhu, A., Yang, L., Li, B., Pei, T., 2007. Topographic Wetness Index Computed using Multiple Flow Direction Algorithm and Local Maximum Downslope Gradient.