EXTRAPOLATING THE SPATIAL DISTRIBUTION OF TAIWAN RED CYPRESS REVERSELY FROM "TERRAIN-SHELTERBELT" PROTECTING TAIWAN FIR

Bao-Hua Shao (1), Chen-I Hsu (2), Chin-Jou Hsu (3), Ching Chuang (4), Nan-Chang Lo (5), Kai-Yi Huang (6)

[1 4] Graduate Student, Dept. of Forestry, National Chung-Hsing University, No. 250, Kuo-Kuang Road, Taichung 402, Taiwan, R.O.C.
[2 3] Undergraduate Student, Dept. of Forestry, National Chung-Hsing University, No. 250, Kuo-Kuang Road, Taichung 402, Taiwan, R.O.C.
[5] Specialist, Experimental Forest Management Office, National Chung-Hsing University, Kuo-Kuang Road, Taichung 402, Taiwan, R.O.C.
[6] Professor, Dept. of Forestry, National Chung-Hsing University, Kuo-Kuang Road, Taichung 402, Taiwan, R.O.C.
Email: baobao357@gmail.com; bu56fo58@gmail.com ; 10230237bear@gmail.com; ching41041@gmail.com; njl@dragon.nchu.edu.tw; kyhuang@dragon.nchu.edu.tw

KEY WORDS: Species Distribution Model (SDM), topographic sheltering index (TSI), decision tree (DT), logistic multiple regression (LMR), impermeable shelterbelt, terrain-related variable.

ABSTRACT: Estimating the spatial pattern of a species is typically based on ecological parameters (e.g. temperature, rainfall, or wind) that are causal factors for its distribution from a mechanistic viewpoint in forest ecology. However, data for such ecological factors are difficult and expensive to collect and sampling quantities for these point-type ground measurements are usually insufficient. Terrain-related variables are used as surrogates of ecological factors in the studies since these data can be easily acquired by remote sensing. *Chamaecyparis formosensis* (Taiwan red cypress, TRC) forests usually grow in the fog-forest belt (FFB) with elevation above 1,800 m. In contrast to *Abies kawakamii* (Taiwan fir, TF) growing above 3,000 m in depressions (i.e. micro-terrain acting like an impermeable "terrain-shelterbelt") near ridges or peaks with enough soil moisture and nutrients, TRCs can occur at or near dry soils-and-gravels covered ridges or peaks where their leaves have functionality to intercept rich moisture as water. This study used DEMs of three grid sizes (5 m, 20 m, and 40 m) to derive various relating variables such as elevation, slope, aspect, surface curvature (SC), profile curvature (PRC), plan curvature (PLC), and topographic sheltering index (TSI). Subsequently, this study tends to build species distribution models (SDMs) for TRCs and assessed reversely the positive effect of topographic shelters associated with TFs on TRC forests. The decision tree (DT), logistic multiple regression (LMR), and discriminant analysis (DA) models were constructed with the aid of above-mentioned topographic induced variables. For DEMs of all grid sizes, the statistics (mean and median) of TSI differ significantly for both species. Furthermore, all variables concerning the spatial curvature show no statistical differences for both species when the resolution was brought down to 5 m. The aforementioned fact may indicate the existence of specific characteristic scale of curvature, when the resolution is finer than this scale, the idea of curvature may no longer be useful. After simulating various combinations within different variables for SDMs of both TRC and TF, the result indicates that the models should not only include elevation variable but also take other terrain-related variables, such as TSI, profile curvature or aspect, into consideration. Along the way of evaluating different performance of SDMs, this study realized a less intuitive fact that combining all a priori important factors would not always get the best SDM. The follow-up study would try to improve SDMs based on the DEM with grid size of 1m, and then assess the effect of resolution on building various SDMs. Futhermore, the study would try to figure out why TRC forests are absent in the FFB areas of Huisun Experimental Forest Station (HEFS) with the aids of TSI, which is a profound topic in the sense of poor performance in TRC's SDM that the study presented

1. INTRODUCTION

Ecologists have recognized that the influence of macroclimate on the growth and productivity of forests is required to project future growth patterns in the face of global climate change (Zimmermann *et al.*, 2007), and this work could be done by linking species－environment relationships to structural habitat properties with statistical methods in the form of digital data layers in a geospatial information system (GIS). Compared with climatic, edaphic, and biotic data, it is easier to obtain topographic data via remote sensing technology, thereby saving much more resources originally needed for intensive fieldwork. Therefore, this study investigated the relationship between topographic variables and ecological features of species to characterize the ecological traits of species by topographic variables, the surrogates of climatic variables (e.g. wind, humidity, and solar radiation) or possibly even some soil-related variables.

Topographic shelters formed by mountains protect *Abies kawakamii* (Taiwan fir, TF) in the Hehuanshan from a high wind so that TFs can grow in depressions or gullies (i.e. impermeable shelterbelt) with elevation above 3,000m near ridges or peaks with enough soil moisture and nutrients rather than on them (Huang, 2002). The study further assessed reversely the positive effect of topographic shelters associated with TF trees, the surrogates of wind (and moisture) variable, which was applied to predicting the spatial distribution of *Chamaecyparis formosensis* (Taiwan red cypress, TRC).

2. MATERIALS AND METHODS

2.1 Study Area

The study area is situated in central Taiwan, and it contains Huisun Experimental Forest Station (HEFS), Pai-Ku Mountains, Hehuanshan and Qilaishan as shown in Fig. 1. The elevation of the study area falls within 392–3,606 m, partially contained in the FFB.

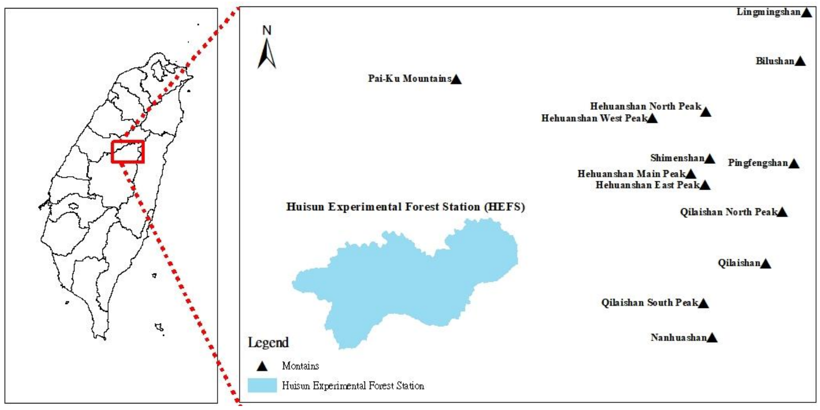

Fig. 1. Location map of the study area.

2.2 Data

Elevation, slope, aspect, and surface curvature (SC) layers from DEM by ArcGIS software packages. Furthermore, the study developed TSI which was derived from the concepts of permeable shelterbelt (windbreak) using a point-in-polygon operation in GIS (Huang, 2002). The index considers the downwind (or horizontal) distance protected by a “shelterbelt”, i.e. impermeable topographic obstacle. The downwind distance protected by a topographic obstacle is roughly 15 – 20 times the relative height difference between the topographic obstacle and a certain point in the downwind direction (Lapen and Martz 1993; Huang 2002). Moreover, the index considers topographic sheltering effects toward multiple terrain obstacles at a given position in the entire study area.

Samples were selected from the digital maps of land use and land cover extracted from the fourth inventory of forest resources which is conducted by Forestry Bureau. The number of background samples randomly selected was three times more than that of target samples (TRCs and TFs) in order to avoid spatial autocorrelation. The study simulating various combinations between different variables for SDMs of both TRC and TF using decision tree (DT), logistic multiple regression (LMR), and discriminant analysis (DA) to predict the suitable habitat TRCs and evaluated them by the kappa statistic (the coefficient of agreement) using an independent sample set.

2.3 Model Development

2.3.1 Decision Tree (DT)

DT is a sequential partitioning of the dataset in order to maximize differences on a dependent variable. Decision pathways originate from a starting node (root) that contains all observations and end at multiple nodes containing unique subsets of observations. Terminal nodes were assigned a final outcome based on group membership of the majority of observations (Bourg *et al.*, 2005; O’Brien *et al.*, 2005). The two most widely used algorithms of DT are that of CHAID (Chi-square Automated Interaction Detection) and CART (Classification and Regression Trees). Since the study focused only on the binary outcomes, we implemented the machine learning package scikit-learn in python so that DT algorithm corresponds to CART.

2

2.3.2 Discriminant Analysis (DA)

DA 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 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). DA was achieved by implementing the machine learning package scikit-learn in python.

2.3.3 Logistic Multiple Regression (LMR)

Logistic regression is a classic approach for identifying datasets with dichotomous target variables. It resembles linear regression method, but with addition of sigmoid transformation. Also, the cost function between these two are quite different, in the case of linear regression, the corresponding cost function is just ordinary least square error, whereas logistic regression takes cross entropy as cost function (Hosmer, 2013). This model outputs continuous probability, and these are transformed into binary cases with different setting of thresholds. A commonly used default threshold is 0.5. Obviously, with different threshold settings, one gets different model agreements, which, in our case, the Cohen's kappa. We then naturally consider the best threshold as the one that optimizes the Cohen's kappa. LMR was achieved by implementing the machine learning package scikit-learn in python.

2.4 Model Validation

Model validation could be done by split-sample validation or cross validation. Split-sample validation can implement by dividing a dataset into two subsets, the first one typically comprising 30% to 70% of all data and the other comprising one-third to one-half of all data. The first one is used to build a model; the other one is used to calibrate the model. For each model, the study predicted the response of the remaining data, and calculated the error between the predictions and the observed ground truth with the aid of Cohen's kappa.

3. RESULTS AND DISCUSSION

Table 1 and Table 2 below shows the result of independent sample t-test by comparing background and target sites for each species. In the case of TF, the statistics of PLC is not significantly different between those two sites for all resolutions. Also, under the resolution of 5 m, all variables that are related to curvatures is not significantly different.

Table1. Independent sample t-test of TF

Grid	Elevation	Slope	Aspect	SC	PLC	PRC	TSI
5	-239.24***	11.17***	6.88***	-0.91	-0.34	1.27	-103.56***
20	-230.98***	21.93***	7.14***	-3.19***	0.76	6.81***	-75.21***
40	-209.93***	21.33***	8.06***	-3.27***	1.74	8.71***	-34.20***

Table2. Independent sample t-test of TRC

Grid	Elevation	Slope	Aspect	SC	PLC	PRC	TSI
5	-239.24***	11.17***	6.88***	-0.91	-0.34	1.27	-103.56***
20	-67.23***	-6.70***	20.07***	-4.76***	-3.42***	4.86***	7.39***
40	-61.05***	-4.07***	21.25***	-6.04***	-3.46***	7.58***	3.19***

The table 3 below shows the best Cohen's kappa in the study for each corresponding situations. For TRF and TF, both elevation and TSI are included in both cases brings more confidence for studying and realizing the effect of terrain sheltering. For TRC's SDM, the best combination of variables contains elevation, profile curvature and TSI, and for TF's SDM behaves the best with combination of elevation, slope, aspect, profile curvature and TSI. Furthermore, comparing models, DT and LMR are better than DA in the case of TF, whereas the kappa value for TRF is inarguably low for almost all algorithms. Furthermore, the result shows that the statistics of elevation, slope, aspect and TSI at both TRC and TF forests are strikingly different from those at the background sites as the grid size of DEM decreases from 40 m to 5 m (i.e. spatial resolution becomes much higher and can show the details of micro-terrain).

3

Table3. the Kappa of TRF’s and TF’s SDM

	TRF			TF		
Grid	DA	DT	LMR	DA	DT	LMR
5m	0.292048	0.435425	0.256971	0.666850	0.790523	0.790729
20m	0.289881	0.358009	0.264348	0.672847	0.777515	0.783841
40m	0.276639	0.348052	0.272037	0.666850	0.790523	0.773017

[1] Variables for TRF’s SDM: elevation, TSI and profile curvature.
[2] Variables for TF’s SDM: elevation, TSI, slope and aspect.

4. CONCLUSIONS

The study first analyzed relatively importance for building SDMs with variables containing elevation, slope, aspect, curvature and TSI. Intuitively, for variables with higher statistical differences will be effective factors to discriminate the existence and non-existence for a specific species. The statistical results reveal that almost all factors considered here show explicit differences, except for curvature terms in resolution of 5 m. While incorporating all variables suggested by the logic above for each resolution, our study does not get the model with best performances. This result reveals the fact that statistical differences does not always imply the ability for the model to tell discrepancies effectively. For TRC’s SDM, the best combination of variables contains elevation, profile curvature and TSI, as for TF’s SDM, it behaves best with combination of elevation, slope, aspect, profile curvature and TSI. Indeed, incorporating TSI into modelling construction for both species does improve its performances. This result, indicates that topographic sheltering has its own value. The usefulness of TSI would be re-examined in the future study. In a nutshell, the defining mathematical structure of TSI need to be revised in order to incorporate the finer idealization of sheltering effects, such as multiple sheltering effect, decreasing effect in wind speed and moisture.

5. 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), 2793-2804.

Hosmer, D. W. 2013. Applied logistic regression (3rd ed. / David W. Hosmer, Stanley Lemeshow, Rodney X. Sturdivant. ed.). Hoboken, N. J: Wiley.

Huang, K. Y. 2002. Evaluation of the topographic sheltering effects on the spatial pattern of Taiwan fir using aerial photography and GIS. International Journal of Remote Sensing, 23 (10), 2051-2069.

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), 175–191.

O’Brien, C. S., Steven, S. R., John, J. H., Jill, L. B. and Susan, R. B. 2005. Landscape - level models of potential habitat for Sonoran pronghorn. Wildlife Society Bulletin, 33 (1), 24–34.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

Skidmore, A. K., 1990. Terrain position as mapped from a grided digital elevation model. International Journal of Geographical Information System , 4 (1), 33–49.

4