EXTRAPOLATING THE SPATIAL PATTERN OF CHINESE GUGER-TREE AND FORMOSAN RHODODENDRON VIA 3S AND AI

Chen-I Hsu (1), Bao-Hua Shao (2), Nan-Chang Lo (3), Kai-Yi Huang (4)

[1] Dept. of Forestry, Chung Hsing University, 145 Xingda Road, Taichung 402, Taiwan (R.O.C.)
[2] Nan-Tou Division Office, Forest Bureau, Council of Agriculture, 456 Shihguan Rd., NanTou 542, Taiwan (R.O.C.)
[3] Experimental Forest Management Office, Chung Hsing University, 145 Xingda Rd., Taichung 402, Taiwan (R.O.C.)
[4] Dept. of Forestry, Chung Hsing University, 145 Xingda Road, Taichung 402, Taiwan (R.O.C.)
Email: bu56fo58@gmail.com; baobao357@gmail.com; njl@nchu.edu.tw; kyhuang@dragon.nchu.edu.tw

Keywords: Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Topographic Sheltering Index (TSI), Species Distribution Modeling (SDM)

ABSTRACT:

Various machine-learning techniques have been used to model species potential habitat. However, compared to species with a clumped distribution, building SDM for widely-dispersed species is more challenging because of their broader ecological amplitude, which makes the relationships between tree species and environmental factors more complex. Thus, we chose two representative species: widespread species, *Schima superba* (CGT) and semi-cluster species, *Rhododendron formosanum* (FR) in the Huisun study area in Taiwan as modeling target. This study performed a comprehensive assessment of different combinations of algorithms and environmental variables, examining whether model predictions are associated with ecological characteristics of species. We took samples randomly from in-situ datasets of two species with the 3S technique, applied 8 algorithms to perform species distribution modeling (SDM): support vector machine (SVM), decision tree (DT) and random forest (RF) with Gini impurity and information gain, k-nearest neighbors (KNN), and discriminant analysis (DA). Considered environmental factors includes elevation, slope, aspect, three types of curvatures (surface, plan, and profile), and topographic sheltering index (TSI), all resampled into 40/20/5 m, then build and evaluate SDM separately in three resolutions. Model performance was evaluated with Kappa index and overall accuracy. The overall modeling accuracies of FR species were higher than those of CGT species for each algorithms. RF_gini, and RF_entropy are the most capable of predicting the tree's potential habitat. Elevation, slope, and TSI are the most important environmental variables, models built with combinations of them are more accurate than all the others. Model built with grid size of 5 m has the highest Kappa value among all resolutions. However, the outcome clearly indicates that the models merely based on topographic variables performed poorly on spatial extrapolation over the entire study area. To improve model performance, follow-up study will use DEM with finer spatial resolution, add more environmental variables such as humidity and solar radiation or their surrogates developed from topographic variables and deep learning algorithms, such as convolution neural network (CNN).

1. INTRODUCTION

Traditional vegetation ecology had heavily relied on field surveys to collect data for species (plants or animals) and environmental variables (climate and soil), and then estimated species distribution based on environmental factors derived from a limited number of sampling plots and stations, which may lead to severe bias, or possibly even erroneous results. Recently, with the fast technological advance in sensors, communication, and computing power, geospatial information systems (GIS), global navigation satellite system (GNSS), and remote sensing (RS) have been tightly integrated as a 3S system. Hence, this system has made it possible to do more detailed survey on vegetations with higher efficiency and positioning accuracy, to provide high-quality data over large scales to build species distribution models (SDM), and eventually to perform spatial extrapolation with higher accuracy and credibility. Nowadays, the evolution of 3S system conduces towards the popularization of precision forestry or even intelligent forestry, also the global climate change has made it more urgent to extrapolate species distribution with higher accuracy for ecological conservation, and thus building a high precision SDM is very important. SDM builds models with multivariate statistics or machine learning algorithms, utilizing the 3S system coupled with field survey data to perform big data analysis, is able to inferring interactions between species and the environment to forming the spatial pattern. SDM can provide information for decision-making, indicating specie's potential distribution and the impact of global climate change on species' distribution, etc. These are all important in intelligent forestry.

1.1.Target Tree Species

Schima superba (Chinese guger-tree, CGT) is a fine board-leaf arbor in Theaceae family, widely distributed in Taiwan, most dispersed in elevation of 300–2,300 m. CGT's wood is dense and resistant to pests; its leaves with high water-content, CGT has an excellent fire resistance characteristics, and thus it can be intertwined with pine, fir, or camphor tree plantations to prevent the forest fire from spreading. *Rhododendron formosanum* (Formosan rhododendron, FR) is an endemic evergreen board-leaf arbor of Taiwan, distributed in the cloud forest belt at medium elevation, usually clustered on upper slopes to wide ridges or flat mountain tips, often growing under cypress forest. The 3S system coupled with machine learning techniques has been widely applied to building SDM for rare plant species, but seldom used on widely-dispersed species such as CGT. Also, compared to species with a clumped distribution such as FR, it is more challenging to build SDM for widely-dispersed species because they have a broader ecological amplitude, which makes the relationships between tree species and environmental factors more complex (Hernandez *et al.*, 2006). Accordingly, this study selected these two target species and took samples from them to build models and to evaluate the performance of predictive modelling for extrapolating their spatial patterns.

1.2.Study area

The study area is situated in central Taiwan, encompassing the Huisun Experimental Forest Station (HEFS), and it has a total area of 17,300 ha. HEFS is the property of National Chung-Hsing University, located in Nantou, Taiwan, with the geospatial extent 24◦2′–24◦5′ N and 121◦3′–121◦7′ E, with a total area of 7,477 ha. The entire study area ranges in elevation from 450 m to 2419 m, and its climate is temperate and humid. In addition, the study area has nourished many plant species more than 1,100 and is a representative forest in central Taiwan. It comprises five watersheds, including two larger watersheds, Kuan-Dau at west and Tong-Feng at east. All the tree samples were collected from the Tong-Feng and Kuan-Dau sites in Huisun.

2. MATERIAL AND METHOD

2.1.Species occurrence data

There are 929 CGT and 931 FR 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 to makes them have an accuracy of sub-meters. Then the dataset was converted to ESRI Shapfile format for later use.

2.2.Environmental variables

This study used a digital elevation model (DEM) produced with lidar and provided by ministry of the interior, ROC (Taiwan). The DEM was interpolated and resampled into 5m, 20m, and 40m, then build and evaluate SDMs seperately for each resolution (5/20/40 m). Slope, aspect, three types of curvatures (surface, plan, and profile), are derived from DEM with ArcGIS. To build TSI layers, first we extract and manually select ridge lines from DEMs with ArcGIS, digitize and convert to raster map, assigned 255 as the ridge otherwise assigned 0. Then compute the topographic sheltering index (TSI) with :

$$\mathrm{S}_{ij} = \sum_{k=1}^{8} \frac{\frac{w_k}{d_k}}{\mathrm{STD}(d'_k)} \tag{1}$$

where S_{ij} = the topographic index of the test cell at i row and j column of ridge layer, w = the weight of the test cell in the k cardinal compass direction, d_k =the distance from the test cell to the ridge cell in the k cardinal compass direction, $STD(d'_k)$ =the standard deviation of scaled distances (d'_k) from a test cell to ridge cells in eight directions, and $d'_k = \frac{d_k - \mathrm{m}(d_k)}{\mathrm{m}(d_k) - min(d_k)}$ (Huang, 2002).

We built TSI layers with ridges of 4 different level, and with two different weight $w_k = 2.0, 2.0, 1.0, 1.0, 1.0, 1.0, 1.0, 2.0$ and $w = 10.0, 10.0, 1.0, 1.0, 1.0, 1.0, 1.0, 10.0$). Totally we have 14 different environmental variable layers.

2.3.Model devleopment

The study built SDM with 8 different algorithm for CGT and RFH, including :

1. SVMlinear : support vector machine with linear kernel
2. DA : discriminant analysis
3. DT_gini : Decision tree with Gini impurity splitting criteria
4. DT_entropy : Decision tree with information gain splitting criteria
5. KNN : K-nearest neighbors
6. MLP : Multilayer perceptron
7. RF_gini : Random forest with Gini impurity splitting criteria
8. RF_entropy : Random forest with information gain splitting criteria

All the algorithms are implemented with python 3.8.0 and scikit-learn 0.22.1. Python is a dynamic, interpreted programming language, commonly used on data analysis, and scikit-learn is a free software machine learning library for the python programming language to build models.

Table 1: Modeling algorithm implemented in this paper:

Model	Core concept	Source
SVM	Hyper-plane	(Cortes and Vapnik, 1995)
DA	Discriminant function	(Hastie *et al.*, 2008)
DT_gini	Classification tree with Gini impurity	(Breiman *et al.*, 1984a)
DT_entropy	Classification tree with information gain	(Breiman *et al.*, 1984a)
KNN	nearest neighbors	(Goldberger *et al.*, 2005)
MLP	network of perceptrons and backward propagation of errors	(Rumelhart *et al.*, 1986)
RF_gini	Random forest with Gini impurity	(Breiman, 2001)
RF_entropy	Random forest with information gain	(Breiman, 2001)

2.3.1. SVM: An SVM classifier maps input data into a high-dimensional space, construct a hyper-plane or a set of hyperplane in a high dimension space, to seperate all the data point into 2 subsets. A good seperation is archieved when the hyper-plane has the maximum distance to the nearest training data point of any class. The SVM algorithm applied in this study is C-support vector classification (C-SVC)(Chang and Lin, 2011).

Given training vectors $x_i \in R^n, i=1,\cdots,l,$ in two classes, and an indicator vector $y \in R^l$ such that $y_i \in \{1,-1\}$, C-SVC solves the following primal optimization problem.

$$\begin{array}{c} \min_{w,b,\xi} \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i,\ \xi \geq 0, i=1,\ldots,l \end{array} \tag{2}$$

where ϕ maps x_i into a higher-dimensionial space and C>0 is the regularization parameter. Its dual is:

$$\begin{array}{c} \min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \text{subject to} \quad y^T\alpha = 0, 0 \leq a_i \leq C, \quad i=1,\ldots,l, \end{array} \tag{3}$$

where $e=[1,\ldots,1]^T$ is the vector of all ones, Q is an l by l positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i,x_j)$, and $K(x_i,x_j) \equiv \phi(x_i)^T \varphi(x_j)$ is the kernel function.

After problem (3) is solved, using the primal-dual relationship, the optimal w satisfies

$$\mathrm{w} = \sum_{i=1}^{l} y_i \ \alpha_i \phi(x_i) \tag{4}$$

and the decision function is

$$\text{sgn} \ \ (\mathrm{w}^T\phi(x)+b) = \text{sgn}(\sum_{i=1}^{l} y_i\alpha_i K(x_i,x)+b) \tag{5}$$

2.3.2. DA: The DA algorithm implemented is linear discriminant analysis (LDA). LDA is a way of finding a linear transformation of data that reduces the number of dimensions required to represent it. It is often used for dimensionality reduction prior to classification, but can also be used as a classification technique itself. LDA uses labeled data. The data is modeled by a multivariate Gaussian distribution for each class c, with mean μ_c and a common covariance matrix Σ . Because the covariance matrix for each class is assumed to be the same, the posterior distribution over the classes has a linear form, and for each class a linear discriminant function (6) is computed, where n_c is the number of examples of class c and n the total number of examples. The data is classified by choosing the largest y_c (Witten *et al.*, 2016a).

$$\mathrm{y}_c = x^T \Sigma^{-1} \mu_c - \tfrac{1}{2} \mu_c^T \Sigma^{-1} \mu_c + \log(n_c/n) \tag{6}$$

2.3.3 DT: DT build a tree with the training data, then predict with the tree. Constructing a decision tree is a recursive process. First, select an attribute to place at the root node, and make one branch for each possible value, splits up the data set into subsets, one for every value of the attribute. Repeat the process recursively for each branch, using only those instances that actually reach the branch. If all instances at a node have the same class, stop developing that part of the tree. In seek of small trees, we would like this to happen as soon as possible. To choose the attribute that produces the purest daughter nodes, most DT algorithms applied a measure of the purity of each node (Witten *et al.*, 2016b). We applied two different measure of the purity for the decision tree. One is information gain (entropy), another is Gini index.

The information measure (7) relates to the amount of information obtained by making a decision, and is based on the physical entropy concept. Decisions can be made in a single or several stages, and the amount of information involved is the same in both cases.

$$\text{Entropy } (\,\mathrm{p}_1, p_2,\ \ldots\ , p_n) = -p_1 \log p_1 - p_2 \log p_2 ... - p_n log p_n \tag{7}$$

where $p_i, i=1,\ldots,n$ is the proportion of the number of instance which belongs to class i.

Usually the logarithms are expressed in base 2, and then the entropy is in units called bits—just the usual kind of bits used with computers. When all the instance in the current node has the same class, the entropy has the minimum value zero, whereas if the number of instance of each class is the same, the entropy has a maximum value of one.

The Gini impurity can be computed with :

$$\textstyle\sum_{j\neq 1} p(i|t)p(j|t) = \sum_j p(j|t)(1 - p(j|t)) \tag{8}$$

For objects in a node t , use the rule that assigns an object selected at random from the node to class i with probability $p(i|t)$. The estimated probability that the item is actually in class j is $p(j|t)$. Therefore, the estimated probability of misclassification under this rule is the Gini index. The Gini index is simple and can be quickly computed (Breiman *et al.*, 1984b).

2.3.4. KNN: For each item of the query dataset, the classification based on KNN locates the k closest members, generally using the Euclidean distance, of the training dataset. The category mostly represented by the k closest members is assigned to the considered item in the query dataset because it is statistically the most probable category for this item. With k goes up, the computing time goes up, but the advantage is that higher values of k provide smoothing that reduces vulnerability to noise in the training data (Garcia *et al.*, 2008). We use the Minkowski distance with p=2 (equivalent to Euclidean distance) in the KNN implementation.

2.3.5. MLP: Percptron was proposed by (Rosenblatt, 1958) as the first supervised learning model, it is the simplest form of a neural network used for the classification of linearly separable patterns. The goal of the perceptron is to correctly classify the input set of externally applied stimuli $x_1, x_2, \dots, x_m$ into one of two classes, c_1 or c_2 . For each input vector $x_1, x_2, \cdots, x_m$, the percptron find an equation:

$$v=\sum_{i=1}^{m} w_i x_i + b \tag{9}$$

where

- w_i is the synaptic weights,
- and b is externally applied bias.

The decision rule for the classification is to assign the point to class c_1 if the perceptron output y is +1 and to class c_2 if it is -1. Perceptron built with a single neuron is limited to performing binary classification, but with more than one neuron, the MLP is capable of performing classification with more classes (Haykin, 2009) . MLP is a network of percptrons contains one or more layers that are hidden from both the input and output nodes. MLP has a high degree of connectivity, the extent of which is determined by synaptic weights of the network. Each neuron in an MLP network includes a nonlinear differentiable activation function. Training an MLP had been a difficult task before the development of the back-propagation algorithm in the mid-1980s ,which provided a computationally efficient method for the training of multilayer perceptrons (Haykin, 2009).

2.3.6. RF: Random forest is an ensemble learning approach which construct classifier with combination of several decision tree predictors. Each trees in the forest's training data is randomly sampled independently with the same distribution from the same dataset. Compared to traditional decision tree method, RF is less sensitive to noise and generally has higher accuracy (Breiman, 2001) .

2.4.Model evaluation

The dataset is split into training (2/3) and testing(1/3) subset to perform a split-sample validation. Prediction accuracy of each model were measured with the Cohen's *kappa* agreement coefficient and the overall accuracy score computed from a confusion matrix. The *kappa* value ranges from -1 to +1, where 0.81-1.0 indicate almost perfect agreement, 0.41 - 0.60 is moderate agreement, and if the *kappa* is less than 0, the strength of agreement is poor (Landis and Koch, 1977) . The target of this study were to compare SDMs built with several machine learning algorithm and different combinations of environmental variables, to find the best approach to obtain the SDM with highest accuracy.

3. RESULT AND DISCUSSION

3.1.Descriptive statistics

According to the descriptive statistics data of CGT (Table 2) and FR (Table 3), CGT samples mainly distributed in elevation of 1316-1838 m, it is a wide-spread species, but have a tendency to grow on relatively higher elevation in our research area. FR samples mainly distributed in elevation of 1683-1999 m, which is much higher than background examples and CGT. As for slope, CGT samples is within 1.2-50.6 degree, with Q1=11.9 and Q3=30.8 degree, which is smaller than background examples. This phenomena may be caused by CGT's light-demanding characteristic. FR samples' slope has a range of 1.1-58.0 degree, with Q1=16.5 degree and Q3=32.6 degree, compared to background examples. The lower value may indicate that their habitat preference of wide flat ridges as CGT. Both CGT and FR mostly grow in areas with higher TSI value (CGT: 0-3.1484 with Q1=0.0041, and Q3=0.6182;FR: 0-17.9580 with Q1=0.0042, and Q3=0.8590), which mean the shading effect of ridges may have a positive effect on their occurence.

Table 2: Descriptive statistics of topographic variables for CGT

		elevation	**aspect**	**slope**	**curvature**	**curvature (plan)**	**curvature (profile)**	**tsi2**
Taget Count: 168	mean	1,610.58	176.19	22.23	1.42	0.65	-0.77	0.41
	std	310.90	134.37	12.22	7.60	4.08	4.56	0.70
	min	674.75	2.32	1.18	-33.76	-14.31	-17.56	0.00
	Q1	1,316.18	41.08	11.88	-2.13	-1.58	-2.86	0.00
	Q3	1,838.42	323.61	30.78	4.20	2.26	1.42	0.62
	max	2,082.72	358.23	50.57	33.72	16.16	19.44	3.15
Background Count: 750	mean	1,302.09	189.68	39.08	-0.24	-0.09	0.15	0.02
	std	377.91	104.12	13.12	14.83	7.85	8.72	0.14
	min	472.77	2.00	0.78	-117.76	-62.21	-39.82	0.00
	Q1	1,025.37	102.80	33.58	-5.21	-3.14	-3.39	0.00
	Q3	1,563.27	278.21	47.23	4.80	2.83	3.30	0.01
	max	2,374.16	358.96	72.03	79.36	39.54	82.29	2.49

		tsi3_	**tsi4_**	**tsi5_**	**tsi2_10_**	**tsi3_10_**	**tsi4_10_**	**tsi5_10_**	**TP**
Target	mean	0.56	0.56	0.80	1.27	1.71	1.70	2.43	6.31
	mean	0.73	0.71	0.77	2.64	2.66	2.51	2.70	1.36
	std	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	min	0.02	0.02	0.16	0.01	0.03	0.03	0.45	6.00
	Q1	0.88	0.88	1.26	1.09	2.21	2.31	3.63	7.00
Background	Q3	2.87	2.65	3.08	15.31	13.97	11.39	12.04	8.00
	max	0.05	0.05	0.14	0.08	0.15	0.16	0.42	4.53
	mean	0.16	0.16	0.29	0.56	0.59	0.57	1.01	1.80
	std	0.00	0.00	0.00	0.01	0.00	0.00	0.00	1.00
	min	0.01	0.01	0.04	0.01	0.03	0.04	0.11	3.00
	Q1	0.03	0.03	0.12	0.03	0.09	0.10	0.36	6.00
	Q3	2.49	2.49	3.60	10.04	10.00	10.00	14.86	8.00

Table 3: Descriptive statistics of topographic variables for FR (5m)

		elevation	**aspect**	**slope**	**curvature**	**curvature (plan)**	**curvatur (profile)**	**Tsi2**
target count: 181	mean	1,744.21	150.3305	25.54	4.3383	1.99	-2.3528	1.04
	std	300.48	116.2382	12.37	18.7346	11.86	10.7171	2.44
	min	1,036.43	2.4514	1.07	-54.9805	-30.49	-57.3595	0.00
	0.25	1,683.31	57.1631	16.49	-4.9927	-3.55	-6.1449	0.00
	0.75	1,998.75	247.9756	32.57	9.021	7.42	3.3129	0.86
	max	2,077.00	358.0634	57.99	94.0063	49.08	26.8416	17.96
background count: 750	mean	1,302.19	189.9374	40.02	-2.6434	-0.89	1.751	0.02
	std	377.91	103.5837	14.25	42.2937	23.94	28.6892	0.15
	min	474.95	0.0557	1.32	-265.9912	-186.29	-245.7614	0.00
	0.25	1,023.17	101.5846	32.50	-10.997	-7.57	-6.6522	0.00
	0.75	1,562.98	277.8747	48.89	8.017	5.92	7.0496	0.01
	max	2,372.86	359.4726	77.34	438.9893	193.23	293.9571	3.00

		tsi3_	**tsi4_**	**tsi5_**	**tsi2_10_**	**tsi3_10_**	**tsi4_10_**	**tsi5_10_**	**TP**

target count: 181	mean	1.71	1.6963	1.91	3.3605	5.42	5.339	5.92	6.52
	std	2.75	2.6501	2.87	10.1701	11.60	10.9983	11.79	1.07
	min	0.00	0	0.00	0	0.00	0	0.00	3.00
	0.25	0.26	0.2553	0.50	0.0118	0.36	0.3578	0.71	6.00
	0.75	1.81	1.8041	2.09	2.4004	4.18	4.2179	4.58	7.00
	max	16.18	14.5034	16.65	77.4572	69.78	62.3743	71.58	8.00
background count: 750	mean	0.07	0.0833	0.20	0.0805	0.19	0.2542	0.61	4.53
	std	0.50	0.5898	0.76	0.6379	1.28	1.848	2.71	1.80
	min	0.01	0.006	0.02	0.006	0.01	0.0116	0.04	1.00
	0.25	0.01	0.0131	0.04	0.0107	0.03	0.0365	0.11	3.00
	0.75	0.03	0.0343	0.12	0.0333	0.09	0.1014	0.37	6.00
	max	12.76	12.7644	9.84	12.5318	30.32	36.8513	46.55	8.00

3.2. Model performance

The best environmental combination for CGT is C1:elevation, slope, tsi3_, tsi4_, tsi5_, and for FR the best is C2:elevation, tsi3_, tsi4_10_, tsi5_10,the accuracy assessment result are presented in table 4 and table 5. The results show that all the models built with 5m resolution data (*Kappa* value: 0.67) are superior to others built with 20m and 40m resolution data (0.60 and 0.59). RF (gini and entropy) have similar accuracy (*Kappa* values:0.79 and 0.78 for CGT, 0.84 and 0.86 for FR) , they are the most capable of predicting both the tree species' potential habitat. The order of average predictive accuracy for models built with other algorithms from highest to lowest is MLP, DT_entropy, DA, DT_gini, SVM, KNN (*Kappa* values: 0.79, 0.76, 0.53, 0.52, 0.52, 0.52) for CGT; DT_entropy,DT_gini, KNN, MLP, DA, SVM for RF (*Kappa* values: 0.77, 0.75, 0.75, 0.73, 0.63, 0.60). Although the MLP model is not the best among them, it still has a noticeable improvement in the accuracy on high-resolution data (5m) (Validation *Kappa* value: CGT=0.73 and FR=0.76) as compared to those shallow machine learning algorithms such as KNN.

Table 4: Model's performance of CGT for enviromental variable combination C1: elevation, slope, tsi3_, tsi4_, tsi5_, for each algorithm.

		DA			**SVM**			**DT_gini**			**DT_entropy**		
	resolution(m)	40	20	5	40	20	5	40	20	5	40	20	5
OA	training	0.81	78%	0.81	83%	0.81	81%	0.83	81%	0.81	96%	0.93	0.92
	validation	0.74	77%	0.82	71%	0.78	84%	0.71	78%	0.84	85%	0.83	0.89
	average	0.78	77%	0.81	77%	0.80	82%	0.77	80%	0.82	90%	0.88	0.90
Kappa	training	0.58	0.51	0.56	0.59	0.56	0.55	0.59	0.56	0.55	0.90	0.84	0.82
	validation	0.43	0.49	0.60	0.33	0.48	0.62	0.33	0.48	0.62	0.66	0.62	0.75
	average	0.51	0.50	0.58	0.46	0.52	0.58	0.46	0.52	0.58	0.78	0.73	0.78
		RF_gini			**RF_entropy**			**MLP**			**KNN**		
	resolution(m)	40	20	5	40	20	5	40	20	5	40	20	5
OA	training	0.96	92%	0.95	96%	0.93	92%	0.98	94%	0.96	88%	0.85	0.86
	validation	0.84	88%	0.87	84%	0.90	88%	0.82	88%	0.88	69%	0.73	0.79
	average	0.90	90%	0.91	90%	0.91	90%	0.90	91%	0.92	78%	0.79	0.82
Kappa	training	0.92	0.82	0.89	0.92	0.84	0.82	0.95	0.85	0.90	0.69	0.65	0.65
	validation	0.65	0.74	0.70	0.64	0.76	0.73	0.61	0.72	0.73	0.27	0.36	0.50
	average	0.78	0.78	0.80	0.78	0.80	0.77	0.78	0.79	0.81	0.48	0.50	0.58

Table 5: Model's performance of FR for enviromental variable combination C2

		DA			**SVM**			**DT_gini**			**DT_entropy**		
	resolution(m)	40	20	5	40	20	5	40	20	5	40	20	5
OA	training	0.67	68%	0.86	84%	0.80	86%	0.98	94%	0.93	92%	0.97	0.98
	validation	0.72	75%	0.88	81%	0.81	86%	0.75	86%	0.89	79%	0.81	0.92
	average	0.70	71%	0.87	82%	0.80	86%	0.87	90%	0.91	85%	0.89	0.95
Kappa	training	0.53	0.49	0.78	0.61	0.49	0.66	0.95	0.85	0.83	0.80	0.93	0.95

	validation	0.59	0.61	0.81	0.56	0.57	0.68	0.45	0.67	0.76	0.52	0.60	0.82
	average	0.56	0.55	0.79	0.59	0.53	0.67	0.70	0.76	0.79	0.66	0.77	0.89
		RF_gini			**RF_entropy**			**MLP**			**KNN**		
	resolution(m)	40	20	5	40	20	5	40	20	5	40	20	5
OA	training	0.96	96%	0.97	96%	0.97	97%	0.91	90%	0.95	90%	0.89	0.94
	validation	0.85	90%	0.95	87%	0.89	96%	0.85	78%	0.89	85%	0.86	0.90
	average	0.90	93%	0.96	91%	0.93	97%	0.88	84%	0.92	87%	0.87	0.92
Kappa	training	0.91	0.91	0.92	0.91	0.92	0.94	0.79	0.75	0.89	0.78	0.73	0.86
	validation	0.65	0.78	0.89	0.70	0.76	0.92	0.65	0.52	0.76	0.68	0.68	0.79
	average	0.78	0.84	0.91	0.80	0.84	0.93	0.72	0.63	0.82	0.73	0.71	0.83

4. CONCLUSIONS

This paper built SVM with various combinations of environmental factors and algorithms, assess their performance to determine the best model. We also compare how their performance varies with spatial resolution. Models built with random forset (there is nearly no difference between RF_gini and RF_entropy) is the most accurate and robust algorithm for both species, the MLP model's performance is slightly lower than FR, and sometimes outperformed DT_gini and DT_entropy. The difference among algorithms and combinations of environmental variables for FR and CGT is mostly similar, but most of the models perform better on FR, also the KNN model performs far better on FR than CGT, since it usually clusters to form pure forests, while CGT's has a wide-dispersed distribution pattern. However, the outcome clearly indicates that the models merely based on topographic variables did not perform well on spatial extrapolation over the entire study area. Thus, the models developed from topographic variables can only be applied within a limited geographical extent without a significant error. Follow-up studies will continue to overcome difficulties encountered although spatial extrapolation of species pattern over a large area is extremely difficult. These studies will generate more simulated data for environmental variables (causal factors) that include rainfall, humidity, soil moisture and thickness, sunlight, and others through geostatistical modelling. Also, these studies will develop topographic variables as the surrogates of these causal factors, and the raster data of these topographic variables will be calibrated by combining on-site measurements of causal factors with fine-resolution, high-precision remote sensing imagery and spatially extrapolated over larger scales using machine learning algorithms, convolution neural network (CNN) and deep learning. Since the MLP model's performance has been demonstrated, other neural networks may has a high potential on building high accuracy and robust SDM. The above-mentioned algorithms will also be applied to performing species distribution modelling so that the accuracy and credibility of predictive modelling can be greatly improved.

5. REFERENCES

Breiman, L., 2001. Random Forests. Machine Learning, 45 (1), pp. 5-32.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984a. Classification and regression trees, The Wadsworth statistics / probability series. Chapman & Hall, New York.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984b. 4.3.1 The Gini Criterion. In: Classification and Regression Trees, The Wadsworth Statistics / Probability Series. Chapman & Hall, New York, pp. 109-110.

Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2 (3), pp. 1-27.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning, 20 (3), pp. 273-297.

Garcia, V., Debreuve, E., Barlaud, M., 2008. Fast k nearest neighbor search using GPU. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-6.

Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R., 2005. Neighbourhood Components Analysis. Advances in Neural Information Processing Systems, 17, pp. 513-520.

Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer-Verlag, New York.

Haykin, S., 2009. Neural networks and learning machines, 3rd ed. Prentice Hall, New York.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29 (5), pp. 773-785.

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), pp. 2051-2069.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics, 33 (1), pp. 159.

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65 (6), pp. 386-408.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature, 323 (6088), pp. 533-536.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016a. 8.3 Projections: Linear Discriminant Analysis. In: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, pp. 310.

Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016b. 4.3 Divide-and-Conquer: Constructing Decision Trees. In: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, pp. 203-204.