# DISSECT THE IMPACT OF DEM WITH DIFFERENT SOURCES ON THE PERFORMANCE OF ESTABLISHING SPECIES DISTRIBUTION MODELS

Chin-Jin Kuo [1], Bao-Hua Shao [2], Nan-Chang Lo[3], Kai-Yi Huang[4]
[1]Dept. of Forestry, Chung-Hsing University, 145 Xingda Rd., Taichung 402, China Taipei,
Email: kuoking37@gmail.com
[2]Pu-Li Workstation, Forest Bureau, 124, Sec. 2, Zhongshan Rd., NanTou 545, China Taipei
Email: baobao357@gmail.com
[3]Experimental Forest Management Office, Chung-Hsing University, 145 Xingda Rd., Taichung 402, China Taipei
Email: njl@nchu.edu.tw
[4]Dept. of Forestry, Chung-Hsing University, 145 Xingda Rd., Taichung 402, China Taipei
Email: kyhuang@dragon.nchu.edu.tw

**KEY WORDS:** Digital Elevation model (DEM), Species Distribution Model (SDM), Aerial stereograms, LiDAR, *Brainea insignis* (cycad-fern)

**ABSTRACT:** This study established cycad-fern's (CF) species distribution model (SDM) by using two types of DEM data measured by LiDAR and Aerial stereograms (AS) to evaluate if overestimated DEM affects model's predictive accuracy. We created three sampling designs (SDs) based on the selected samples to analyze the performance of the model. Four environmental factors elevation, slope, terrain position, and vegetation index were chosen for SDMs. The models were developed with four algorithms maximum entropy, BIOCLIM, DOMAIN, and decision tree. Accuracy assessment contains the *kappa* coefficient and overall accuracy. Results show that the AS DEM overestimates the altitudinal distribution of CFs, which is 18 m higher on average. Besides, the quality of DEM does affect the results of the model. In SD1 mode, the *kappa* value is 4.3–7.7% higher when using LiDAR DEM. Moreover, the model accuracy in SD3 can be improved up to 9.4%, showing that the impact of DEM on SDM cannot be underestimated. However, in SD2, the accuracy dropped significantly about 48.6% when using LiDAR DEM. The reason lies in the precision of DEM. The AS DEM shows that the distribution of CFs in two different habitats differed by about 52 m. However, a portion of the CFs in the two sites grows at similar altitudes. But for the LiDAR data, the average elevation difference between the two sites reached 72.6 m. Furthermore, the two habitats hardly overlap at altitude, so it is not surprising that there are more omission errors in SD2 mode. This study confirmed that AS DEM is not recommended for establishing SDM for mountain species, otherwise there will be deviations. It is believed that the gap will be bridged by incorporating more species-related environmental variables although the current LiDAR DEM model is unfavorable for spatial extrapolation.

## 1. INTRODUCTION

There are complex interactions between environmental factors and plants. Generally, past studies roughly summarize such complex factors into four major items: climatic factor, biotic factor, topographic factor and edaphic factor. From the view point of ecology, four factors obtained from plant's growth site is the core issue. However, there is always a gap between the research and the practical in this field. Climatic and edaphic are direct acting factors, they are usually difficult or expensive to measure. Yet, they often tend to be less precise than pure topographic characteristics (Guisana &.Zimmermann, 2000). Bioclimatic maps are usually developed by spatial interpolations of climate station data and satellite observation data. Satellite data can improved accuracy for bioclimatic maps, particularly for areas with a low station density, although interpolation error remained high in such regions (Fick & Hijmans, 2017). In addition, due to insufficient spatial resolution, it cannot reveal biotic relevant microclimate information. Soil and geology maps are even more difficult to derive, especially in mountain areas. Thus, they are usually generated at very coarse resolution (Guisana &.Zimmermann, 2000). The study of biotic factors is to investigate the social organization formed by the long-term action of the environment and plants in the natural habitat. This requires extensive observation over a long period of time. Moreover, the correlation between most species is not clear. This is why this factor is rarely used as a variable. Among the four, topographic factor is the easiest to obtain large-scale and high-precision data through remote sensing, even in mountainous terrain. Terrain variables such as slope, aspect, and curvature can be easily generated through highly accurate digital elevation model (DEM). Therefore, in the discussion of mountain ecology, the correlation between topographic factors and species should be the primary core research content.

The acquisition modes and processing techniques used in the production of DEMs have been modernized (Okolie & Smit, 2022). Common techniques include: aerial stereograms (AS), radargrammetry, interferometric synthetic aperture radar, and LiDAR. In urban or suburban areas, using these techniques to obtain high-accuracy DEMs is a piece of cake. However, in areas with rugged terrain and high crown-closure forest, the shortcomings of passive optical remote sensing that cannot penetrate the tree canopy is exposed. Since the AS method cannot penetrate the canopy, the measured elevation will be overestimated. Yang & Tseng (2008) found that the error of AS DEM may even reach more than 30 meters, and the height of vegetation cover will be wrongly estimated. Using such data to predict species distribution will inevitably bias the simulation results. Ten years ago, our laboratory published an article about predicting the spatial pattern of rare plant—*Brainea insignis* (cycad-fern, CF). Due to technical limitations, the study used DEM generated by AS. Today, we have obtained a LiDAR DEM in the same study area. Therefore, this study established CF's species distribution model (SDM) by using two types of DEM data measured by LiDAR and AS to evaluate if the quality of the DEM affects model's predictive accuracy.

## 2. STUDY AREA AND TARGET SPECIES

### 2.1 Study area

The study area is located in central Taiwan, and its scope covers the Huisun Experimental Forest Station (HEFS) with an area of about 16,000 ha. Its geographic coordinates (TWD97 TM2) approximately fall within easting 249671.4 m–265421.4 m and northing 2657687.5 m–2667987.5 m. Altitude ranges from 454 m to 2,418 m, including the ecological environment from low altitude to medium-high altitude. There are about 1,100 kinds of plants in HEFS, which is a representative forest in central Taiwan (Lo *et al*., 2011). The location map of the study area is show in figure 1.

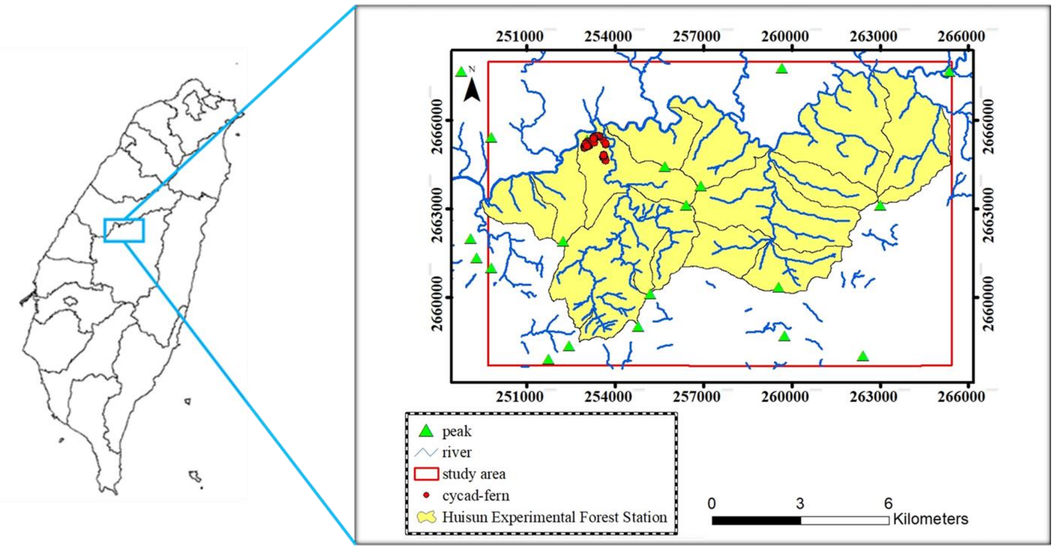


Figure 1 Location map of the study area

### 2.2 Target species—*Brainea insignis* (cycad-fern, CF)

Cycad-fern are perennial ferns of the Blechnaceae family. The plant has a lignification upright trunk which may height up to 70 cm. Its pinnately compound leaves are clustered on the top of the stem. The leaves are oblong, up to 120cm long and 15 to 25cm wide. Figure 2 shows its appearance characteristics. Such plants were abundantly distributed during the Mesozoic Era, but currently *Brainea* is a monotypic genus. It is distributed in tropical Asia, including India, Myanmar, Malay Peninsula, Philippines, Thailand, Taiwan, Vietnam, Indonesia, Sumatra and south China (Kholia & Sharma, 2019). CF is a rare plant in Taiwan, the distribution of this species is relatively few, mainly concentrated in the central part of the island. Among them, HEFS has the largest population. Most people have the impression of ferns as plants that grow in moist areas. However, CF is a shade-intolerant species which grows well in secondary forests with moderate degree of closeness. Zeng *et al* (2016) found out that soil fertility does not limit the development of cycad fern populations, light conditions are the main factor. In addition, this plant is a fire-resistant species. Since its terminal buds have layers of scales that provide good heat insulation, fire won't cause serious damage (Tsia, 1997). Yet, burning will promote CF to spawn spores. In the first month, the percentage of leaves with sorus in the area after the fire was 82% higher than that of the unburned (Lin, 2002). Therefore, periodic forest fires are beneficial to the natural regeneration and survival of this species.

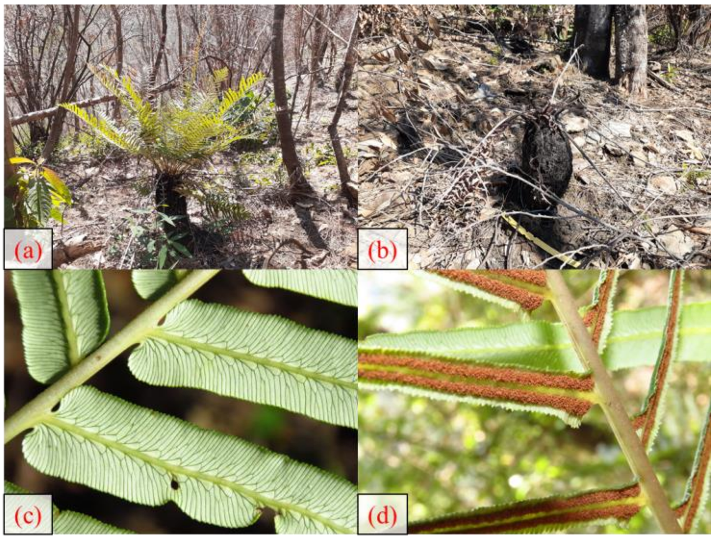

Figure 2 (a) The appearance characteristics of CF; (b) CF that have experienced forest fires; (c) The dorsal surface of the leaf; (d) Sorus of a CF.

## 3. MATERIALS AND METHODS

### 3.1 Digital Elevation Model

This study uses DEM produced by two different techniques. The AS DEM was created using stereo pair images taken by the Forestry Bureau Aerial Survey Office. Its original spatial resolution is 40 m. In order to meet the needs of ecological analysis, this study used the nonlinear rubber sheeting method of ERDAS Imagine software, with fifth-order polynomial interpolation to generate 5 × 5 m pixels. The LiDAR DEM are data from the Satellite Survey Center, Dept of Land Administration, Ministry of the Interior. To match the spatial resolution of AS DEM, we used the resample function of Arc GIS 10.3 software to change the spatial resolution from the original 1m to 5m.

### 3.2 Sampling Designs (SD)

There were 214 samples of CF collected by our laboratory. These ground surveys were conducted to use Trimble Pro XR GPS with a 5 m telescopic extension antenna and a laser range finder. The background samples are randomly sampled from other areas except the target sample point. The ratio of target sample to background sample is 1:4. This species are mainly distributed in Songfeng Mountain (SFM) and Kuantaoshan forest-road (KFR). We refer to Wang (2012) using three SD methods based on the selected samples. First (SD1), we only selected samples from SM, 2/3 of the dataset for training and the rest for validation. Second (SD2), we used the same training samples as SD1, but the validation samples were all taken from KF. Since the distance between SFM and KFR is about 0.76 km, this design aims to evaluate the ability of spatial extrapolation of each model. The design of the last (SD3) is the same as SD1, but used samples from the two sites. To reduce the impact of sampling deviations, each sampling design is repeated five times, and then the average of the five results will be discussed.

### 3.3 Environmental Factors

Four environmental factors elevation, slope, terrain position (TP), and vegetation index (VI) were chosen for SDMs. The layers for elevation and slope were generated from DEM by ArcGIS software package. The calculation of TP is much more complicated than others.

**3.3.1 TP:** it is the relative position from the ridge line and the valley line. The higher one means that it is closer to the ridge line so there is less accumulation of soils, nutrients, and water. On the contrary, if closer to the valley line, it is easier to accumulate those materials, but at the same time it will receive less sunlight. To calculate TP, first we digitized ridges and valleys from the DEMs, then calculated the Euclidean distance from each cell to the nearest ridge and valley line, and last determined the relative position ratio, the formula is shown as follows:

$$P_{ij} = PV \div (PV + PR) \qquad (1)$$

where $PV$ = Euclidean distance from point $P$ to the nearest valley line
$PR$ = Euclidean distance from point $P$ to the nearest ridgeline
$P$ = a validation point (cell)
$P_{ij}$= the relative position ratio of $j$ row and $i$ column

In this study, the ridgeline is the highest slope position, represented by "8", while the valley line is the lowest slope position, represented by "1". The interval is divided into six levels, so there are eight levels in total.The TP layer derived from the AS DEM is finer, with minor ridge lines and valley lines depicted. Yet, the one produced from the LiDAR DEM recently only depict the main mountain range.

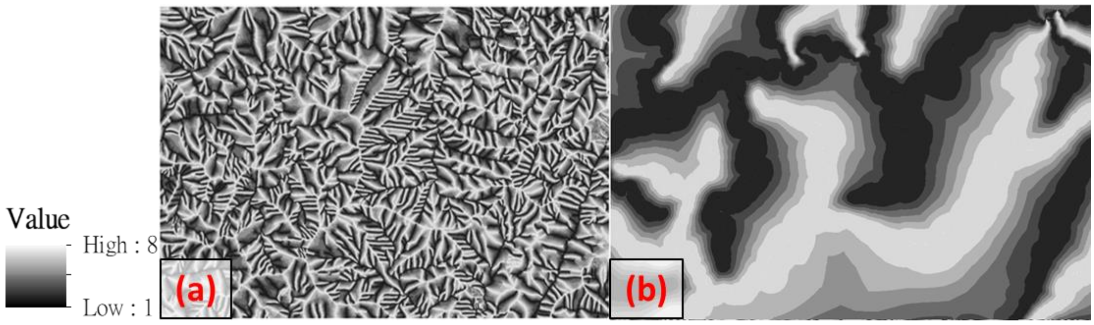


Figure 3 TP layers derived from (a) AS DEM and (b) LiDAR DEM

**3.3.2 VI:** In this study, in terms of two SPOT-5 satellite images, were used to calculate VI. The images were capture on 2004/07/10 and 2005/11/11 respectively. In this sense, this study adopted near infrared (NIR) and middle infrared (MIR) to calculate the VI by the formula as follow (Wang, 2012):

$$[(NIR - MIR)_{fall} / (NIR - MIR)_{summer}] \qquad (2)$$

The strong reflection of NIR by plants is related to the structure of leaves, while the reflection and absorption of MIR is affected by the water content of spongy mesophyll (Jensen, 2015). In general, plants with well-developed leaves and sufficient water results in low MIR reflections, causing (NIR – MIR) > 0. Therefore, the larger the value of the calculation result of the VI, the greater the difference in the leaf water content of the plant in summer and autumn. Conversely, the smaller the value, the less variable the water content of the plant in the two seasons. At present, this indicator is based on the existing data of our laboratory, it will be recalculated in the future with satellite images taken in recent years.

**3.4 Model Development**

This research used the machine learning module from three different software packages to develop SDMs. The software package used for maximum entropy (MAXENT) model construction was downloaded for free on its webpage (https://biodiversityinformatics.amnh.org/open_source/maxent/). BIOCLIM and DOMAIN are modeled using the version of the online free software ModEco 1.0 (http://www.3decology.org/modeco/). As for the decision tree (DT), it is calculated using the module of SPSS 25 version. To maximize the performance of the model, hyperparameter tuning played a key role. The hyperparameters of each model in this study, MAXENT are threshold, DT are child-node, parent-node, and depth, DOMAIN are similarity threshold, and BIOCLIM are percentiles.

**3.4.1 Maximum entropy:** MAXENT is one of the most robust and advanced modeling approaches for presence only data (Qin *et al*., 2020). This method employs the maximum entropy algorithm and species occurrence to predict habitat distribution (Abolmaali *et al*., 2018). Entropy represents the degree of confusion in the system and can also be seen as a state where noise affects useful information in the system. Therefore, when the data distribution reaches the maximum entropy, its theoretical distribution will be closest to the actual distribution state. Research select logic conversion as output format for predicted distribution, it converts the calculation result into an S-shaped curve, which is closer to the biological response to the environment, so it is more suitable for ecological research (Tu, 2013). This study uses the maximum training sensitivity plus specificity given by the software as the segmentation threshold because it has better classification results.

**3.4.2 Decision Tree:** The DT method used in the study is the classification and regression tree (CART) proposed by Brieman *et al*. (1984). The principle of decision tree is to establish a dichotomy classification rule of one-to-multilayer tree structure for the input original data, according to this rule to predict the unknown result data. The nodes are the points in a tree where a test is done on the attribute, and branches are test result that leads to another node (Vanfretti and Arava, 2020). There are three kinds of nodes: root node, internal node and leaf node. The root node is on the top, internal nodes are in-between and leaf nodes are assigned a final outcome based on group membership of the majority of observations.

**3.4.3 BIOCLIM:** BIOCLIM was one of the first methods developed for SDM, introduced by Nix (1986). It is not as effective as other modeling methods proposed in recent years. However, it is still used, inter alia, since the algorithm is simple to understand (Sarma *et al*., 2022). BIOCLIM uses a percentile distribution to analyze the environmental variables of each grid in the study area and calculate the probability of species occurrences. That is to say, the SDM can be established based on the environmental variables that match the presence data of the species, there is no need for the absence sample.

**3.4.4 DOMAIN:** The DOMAIN model was proposed by (Carpenter 1993). It is a distance-based model that is estimated found on similarity to the spatial environment, using the Gower metric. It first defines the spatial distance ($d_{AB}$) of two points (A and B) in the *p* dimension by the following formula (Carpenter *et al*., 1993):

$$d_{\mathrm{AB}} = \frac{1}{p}\sum_{k=1}^{p}\left(\frac{|A_{\mathrm{k}} - B_{\mathrm{k}}|}{range\mathrm{k}}\right) \tag{3}$$

where *k* = standardized environmental variables

Next, the complementary similarity measure ($R_{AB}$) will be define:

$$R_{\mathrm{AB}} = 1\text{-}d_{\mathrm{AB}} \tag{4}$$

Finally, the similarity $S_A$ is defined according to the R value calculation formula, which is the maximum similarity between the candidate point A and the known record point $T_j$:

$$S_{\mathrm{A}} = \max_{j=1}^{m} R_{\mathrm{TjA}} \tag{5}$$

DOMAIN produces an index of habitat suitability on a continuous scale (0−100), where higher scores are considered highly suitable (Pandit *et al*., 2017). Studies have used this method for the simulation of woody plants encroachment savanna systems (Arieira *et al*., 2018) or review the effectiveness of existing designated protected areas (Maciel *et al*., 2021). After nearly 30 years, this method is still in use because it only requires presence data to operate and can maintain effective simulation power when the predictors or sample size is small.

### 3.5 Model Validation

The purpose of model validation is to analyze the predictive results and understand the reliability of the model. In this study, split-sample validation approach was taken to split the whole dataset into two subsets. Two-thirds of the samples are used for modeling, and the rest are test samples. Accuracy assessment contains the *kappa* coefficient and overall accuracy (OA). The range of *kappa* value is -1 to 1, but it usually falls between 0 and 1. The higher the *kappa* value is, the more consistent it will be (Wang, 2012). The domain of the OA is between 0 and 1, and the greater the value, the higher the proportion of prediction correct.

## 4. RESULTS AND DISCUSSION

### 4.1 Descriptive Statistics

Table 1 shows the statistics of topographic variables for CF and background samples obtained by AS DEM, while table 2 present the one extracted from LiDAR DEM. After comparison, it can be found that the inability of the AS method to pass through the high crown-closure forest exposes the shortcomings of DEM being overestimated. In terms of elevation, the highest and lowest distributions of CF in the LiDAR data decreased by at least 15 m compared to the AS DEM. As for the average distribution elevation, it dropped from the original 791 m to 778 m, a difference of 13 m. The error concatenation of the DEM makes the slope layer also appear discrepancy. The minimum value of the AS data is 3° higher than that of LIDAR, and the difference in the maximum value is 12°. As mentioned before, the two TP layer differ in the detail with which the terrain is depicted. This is the reason for the large difference in the statistics of the variables between the two. Figure 4 clearly illustrates this phenomenon. It demonstrate that in the SFM and GFR areas where CF appear, this species mostly grow in the area above the mountainside to the ridge line, which does meet the statistics of AS TP layer. However, if the spatial scale is placed on the entire study area, surrounded by mountains of more than 2,000 m, the habitat of the CF, which is only 800 m, becomes an area close to the valley. As for the VI, since the production process does not involve DEM, the statistics of both are the same. Although CF is suitable for growing in areas with abundant sunshine such as adret or open area, it still need sufficient soil moisture. Therefore, it tends to grow in the understory environment with tree layer but have sufficient light-transmitting pores (Wang, 2012). That's why the value range of VI is mostly around 20, which indicates that this species grows under forests without obvious yellowing or deciduous leaves.

Table 1 The statistics of topographic variables for CF and background samples obtained by AS DEM

| statistics | background | | | | CF | | | |
|---|---|---|---|---|---|---|---|---|
| | elevation (m) | slope (°) | TP | VI | elevation (m) | slope (°) | TP | VI |
| average | 1332 | 34 | 5 | 24 | 791 | 32 | 6 | 20 |
| mode | 962 | 40 | 3 | 10 | 819 | 23 | 7 | 10 |
| max | 2326 | 68 | 8 | 70 | 915 | 66 | 8 | 54 |
| min | 521 | 1 | 1 | 0 | 729 | 5 | 3 | 0 |

Table 2 The statistics of topographic variables for CF and background samples obtained by LiDAR DEM

| statistics | background | | | | CF | | | |
|---|---|---|---|---|---|---|---|---|
| | elevation (m) | slope (°) | TP | VI | elevation (m) | slope (°) | TP | VI |
| average | 1320 | 38 | 5 | 24 | 778 | 32 | 2 | 20 |
| mode | 1286 | 40 | 6 | 10 | 763 | 34 | 2 | 10 |
| max | 2324 | 75 | 8 | 70 | 887 | 54 | 2 | 54 |
| min | 520 | 0 | 1 | 0 | 714 | 2 | 2 | 0 |

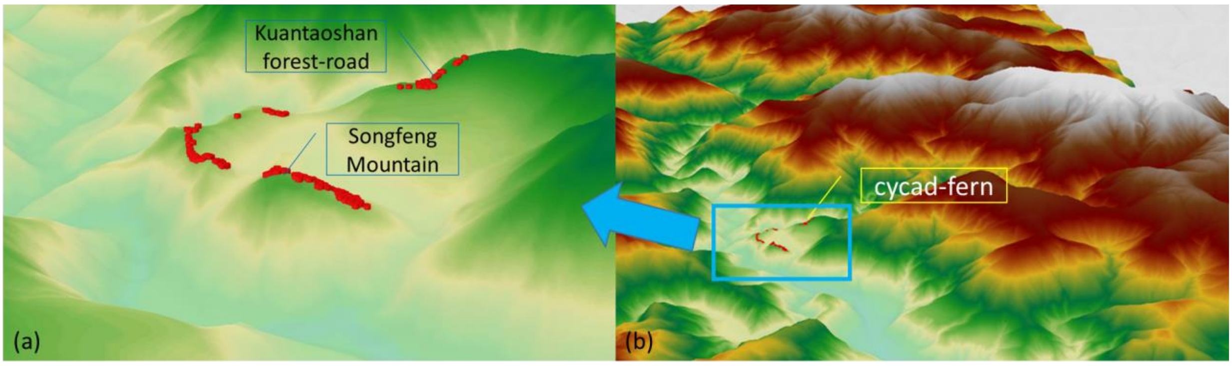


Figure 4 CF habitat at different spatial scales (a) small spatial scale (b) large spatial scale

### 4.2 Assessment of the importance of environmental factors

All four models used in this study can assess the importance of selected variables. MAXENT provide the output value includes percent contribution and permutation importance, while DT output predictor importance and normalized importance. BIOCLIM and DOMAIN understand the importance of each variable to the model by presenting the *kappa* value as a histogram. If there is a factor that has an important impact on the distribution of species, whether it is included in the model will greatly affect the *kappa* value. Due to page limitations, Table 3 only lists the percent contribution of MAXENT and the predictor importance of DT. It can be seen that elevation is basically the most important variable, while TP is the secondary, and the remaining two are less important. However, TP is no less important than elevation when modeling with LIDAR data. Figures 5 and 6 show the importance of the four environmental factors to the BIOCLIM and DOMAIN model using AS DEM and LIDAR DEM, respectively. Similar with the previous two model, elevation and TP are also important factors affecting BIOCLIM, the importance of TP is also higher when using LIDAR data rather than adopting AS data. As for DOMAIN, elevation was the only variable that had a significant impact on the distribution simulation of CF. In conclusion, for all four algorithms, elevation plays a pivotal factor.

Table 3 The importance of four environmental factors in MAXENT model and DT model

| | AS DEM | | | | LiDAR DEM | | | |
|---|---|---|---|---|---|---|---|---|
| | MAXENT percent contribution | | DT predictor importance | | MAXENT percent contribution | | DT predictor importance | |
| | SD1 | SD3 | SD1 | SD3 | SD1 | SD3 | SD1 | SD3 |
| elevation | 88.12 | 84.36 | 0.242 | 0.234 | 59.72 | 50.18 | 0.250 | 0.233 |
| slope | 0.08 | 0.36 | 0.008 | 0.010 | 2.74 | 4.8 | 0.011 | 0.051 |
| TP | 11.12 | 15.02 | 0.035 | 0.063 | 36.84 | 44.88 | 0.207 | 0.164 |
| VI | 0.7 | 0.26 | 0.008 | 0.004 | 0.68 | 0.14 | 0.008 | 0.005 |

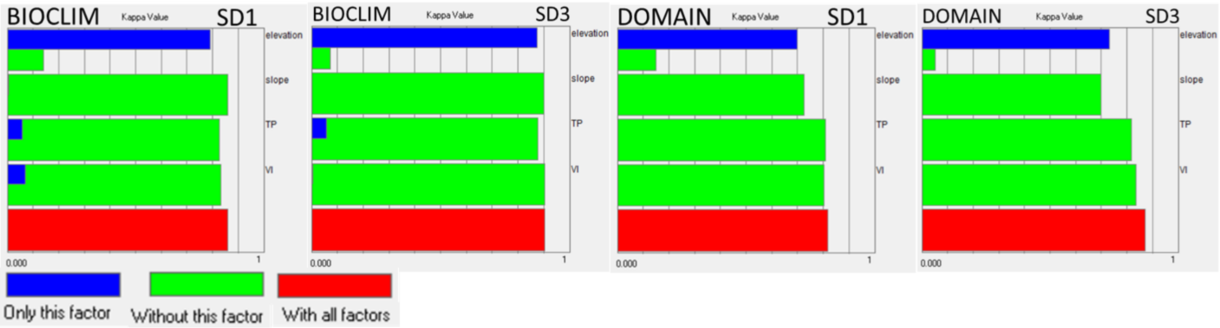

Figure 5 The importance of the environmental factors to the BIOCLIM and DOMAIN model using AS DEM

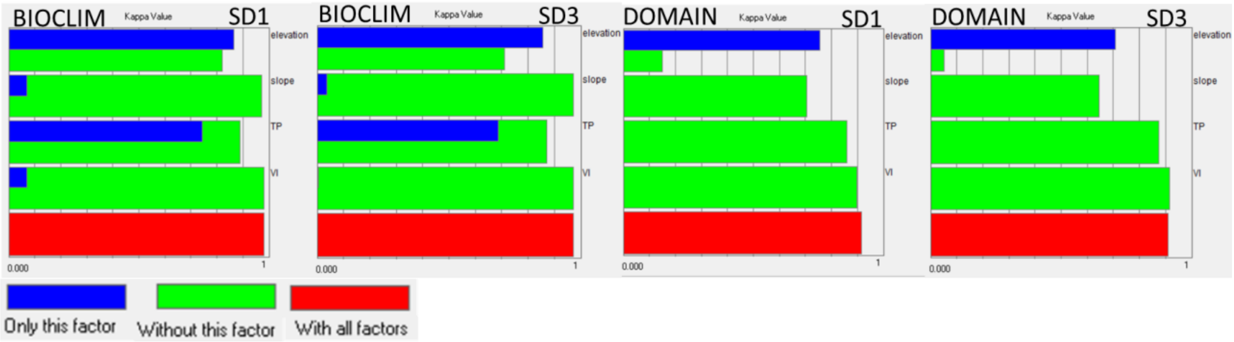

Figure 6 The importance of the environmental factors to the BIOCLIM and DOMAIN model using LiDAR DEM

**4.3 Model performance**

By comparing the accuracy assessments of Tables 4, it is clear that the accuracy of the LiDAR-based model will be better than that of the AS-based. In SD1 mode, the *kappa* value is 4.3–7.7% higher when using LiDAR DEM. Moreover, the model accuracy in SD3 can be improved up to 9.4%, showing that the impact of DEM on SDM cannot be underestimated. Among the four algorithms, BIOCLIM is the most affected, while DT is the lightest, but the model accuracy is also different by 4.4% on average. However, in SD2, except for DOMAIN, the accuracy dropped significantly about 48.6% when using LiDAR DEM. The LIDAR-based model is almost completely inferior to the AS model in terms of spatial extrapolation. Figure 7 presents the CF SDMs using BIOCLIM in SD2 mode. It is clear that the two models have good prediction accuracy in SFM, but the LiDAR-based model almost completely misses the cycad fern population in KFR. This seems to contradict the previous conclusion that the LiDAR data perform better. Nevertheless, after careful analysis, it is found that the reason for this result is still related to the precision of DEM. From the variable importance analysis in the previous section, the key factor affecting the distribution of CF is elevation. Therefore, Table 5 summarizes the elevation distribution statistics of this species in SFM and KFR. The AS DEM shows that the distribution of CFs in SFM is on average 52 m lower than that of the KFR population. However, the highest distribution of the former is 915.8 m, which is higher than the latter's 886.2 m. It demonstrates that a portion of the CFs in the two sites grows at similar altitudes. Thus, in the SD2 mode, the algorithm classifies part of the KFR area as suitable habitat for CF to growth. But for the LiDAR data, the average elevation difference between the two sites reached 72.6 m. Furthermore, the highest distribution in SFM (817.8 m) is only slightly higher than the lowest distribution in KF (815.1 m). The two habitats hardly overlap at altitude, so it is not surprising that there are more omission errors in SD2 mode.

Table 6 demonstrated the potential habitat area of CF estimated by various models. The comparison shows that

LIDAR-based models have smaller estimated habitats. This means that using LIDAR DEM can precisely target CF's habitat, resulting in a reliable SDM which will be more helpful for subsequent applications such as the designation of conservation area or the exploration of new populations. Among the four algorithms, the estimated area of DOMAIN is the largest, while the looser forecast gives it the best spatial extrapolation effect on LIDAR-based models. Through Figure 8, it can be found that this model has a good ability to grasp the known distribution points of CF, but the accuracy of the mode is still not ideal due to the commission error of background samples.

Table 4 Comparison of the testing accuracy of the four models under three sampling designs (%)

| | aerial stereograms | | | | | | LiDAR | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | SD 1 | | SD 2 | | SD 3 | | SD 1 | | SD 2 | | SD 3 | |
| | *kappa* | OA | *kappa* | OA | *kappa* | OA | *kappa* | OA | *kappa* | OA | *kappa* | OA |
| DT | 89.3 | 96.4 | 51.3 | 91.6 | 87.7 | 95.9 | 93.7 | 98.0 | 9.3 | 88.5 | 92.1 | 97.4 |
| MAXENT | 85.6 | 94.7 | 72.9 | 93.1 | 85.1 | 94.8 | 93.3 | 97.8 | 23.6 | 89.7 | 91.8 | 97.2 |
| BIOCLIM | 86.0 | 95.1 | 58.1 | 90.8 | 83.7 | 94.3 | 93.4 | 97.9 | 3.7 | 88.1 | 93.1 | 97.8 |
| DOMAIN | 86.0 | 95.1 | 58.4 | 90.9 | 84.9 | 94.7 | 90.3 | 96.7 | 66.7 | 93.3 | 90.5 | 96.8 |

Table 5 Elevation distribution statistics of CF in two different habitats (meter)

| statistics | aerial stereograms | | LiDAR | |
|---|---|---|---|---|
| | Songfeng Mountain | Kuantaoshan forest-road | Songfeng Mountain | Kuantaoshan forest-road |
| average | 782.2 | 834.3 | 766.6 | 839.2 |
| max | 915.8 | 886.2 | 817.8 | 886.5 |
| min | 728.3 | 808.4 | 714.9 | 815.1 |

Table 6 Potential habitat area of CF estimated by various models (hectare)

| | aerial stereograms | | | LiDAR | | |
|---|---|---|---|---|---|---|
| | SD 1 | SD2 | SD 3 | SD 1 | SD2 | SD 3 |
| DT | 447.3 | 447.3 | 516.4 | 214.7 | 214.7 | 302.8 |
| MAXENT | 814.1 | 814.1 | 838.4 | 382.8 | 382.8 | 403.7 |
| BIOCLIM | 867.3 | 867.3 | 1208.5 | 151.25 | 151.25 | 257.6 |
| DOMAIN | 1039 | 1039 | 941.9 | 565.3 | 565.3 | 633.9 |

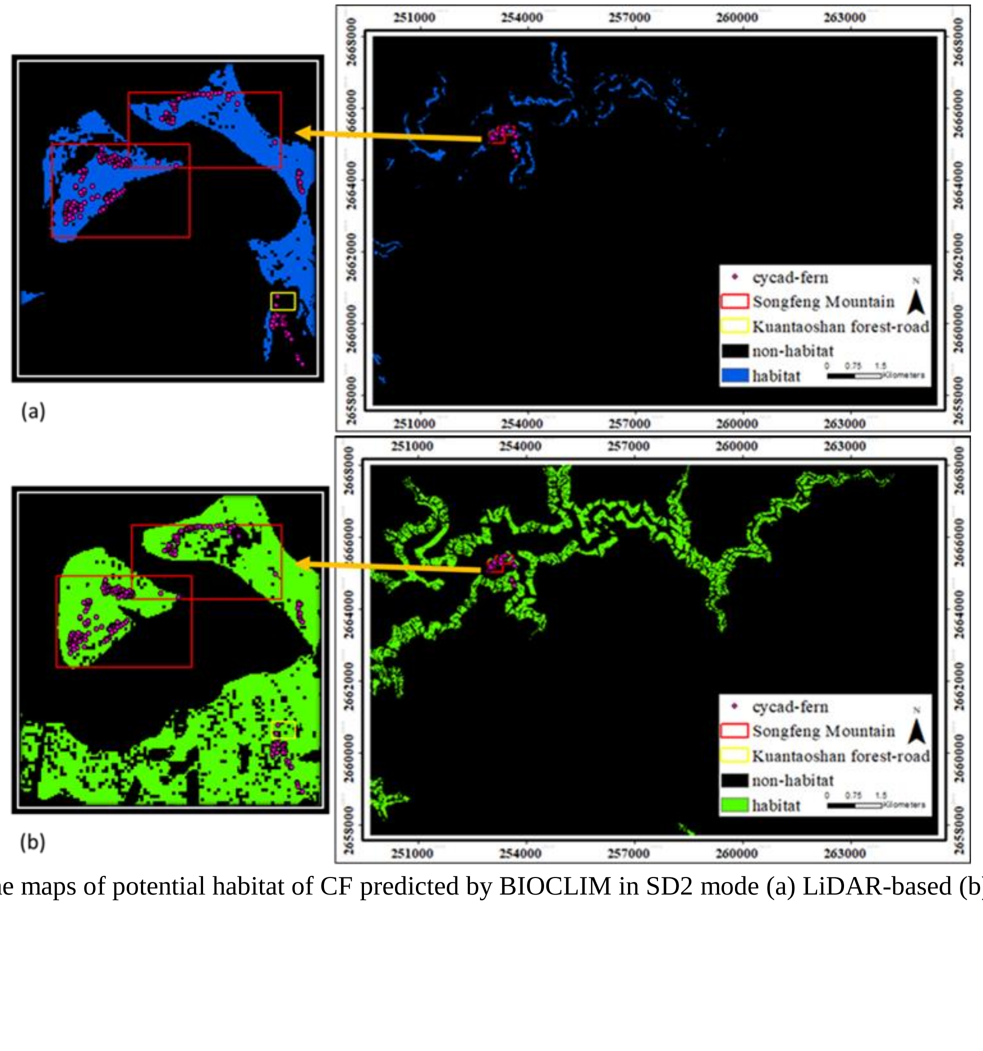

Figure 7 The maps of potential habitat of CF predicted by BIOCLIM in SD2 mode (a) LiDAR-based (b) AS-based

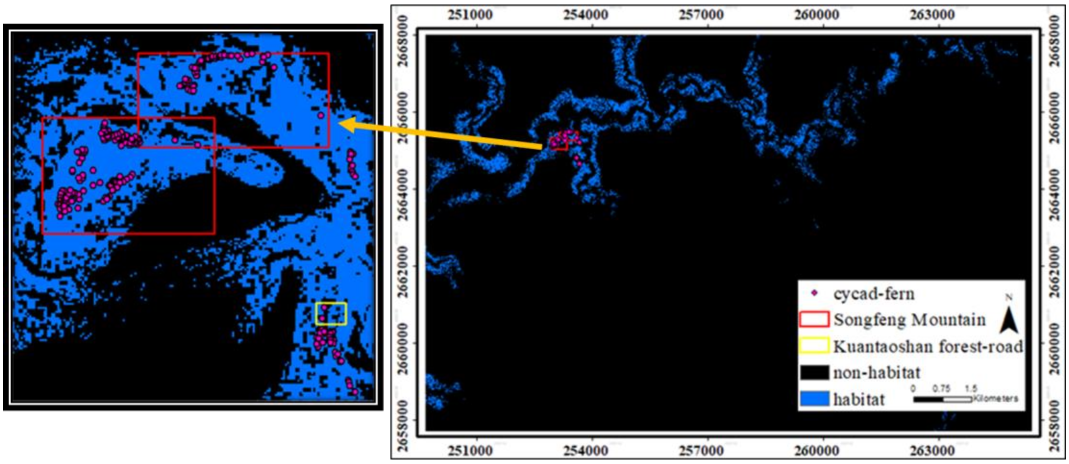

Figure 8 The map of potential habitat of CF predicted by DOMAIN in SD2 mode (LiDAR-based)

## 5. CONCLUSIONS

Only when research based on accurate and reliable data can be trusted. This study confirmed that AS DEM is not recommended for establishing SDM for mountain species, otherwise there will be deviations. In particular, the most important variable of CF's SDM is elevation, the inaccurate DEM makes its predicted habitat area overestimated. It is believed that this gap will be bridged by incorporating more species-related environmental variables (such as area solar radiation, topographic wetness index) although the current LiDAR DEM model is unfavorable for spatial extrapolation. In the past, due to technical limitations, AS DEM was used as a last resort to establish SDM, but now that LIDAR and radar technologies are familiar, DEM produced by these technologies should be used to correctly analyze mountain ecology.

CF is a rare species. If a nature reserve or an ex situ conservation is to be set up for it in the future, correct distribution information must be required. Otherwise, there is no protected object in the reserve and the ex situ conservation cannot allow this species to survive, which will only make this species endangered. In addition, the rugged topography of the mountainous area makes each survey quite arduous and costly. When reliable SDMs are used, it is possible to directly visit the areas where this species is most likely to occur, with minimal investigation cost. However, inferior SDM could make the entire investigation a void. Researchers must pay close attention to the quality of the data when building simulations of species distribution, whether it is the coordinates of the species or the environmental variables used.

Follow-up research will continue to explore the impact of data quality on SDM performance, such as samples with different positioning accuracy measured by global navigation satellite system (GNSS), using coarse-resolution direct acting factors instead of high-resolution terrain derived surrogate index, and using DEM-derived environmental variables at different resolutions (1 m, 5 m, 20 m, 40 m). Expect better development in the field of species distribution modeling.

## 6 REFERENCES

Abolmaali, S., Tarkesh, M., and Bashari, H., 2018. MaxEnt modeling for predicting suitable habitats and identifying the effects of climate change on a threatened species, Daphne mucronata, in central Iran. Ecological Informatics, 43, pp. 116-123.

Arieira, J., Padovani, C. R., Schuchmann, K.-L., Landeiro, V. L., & Santos, S. A. 2018. Modeling climatic and hydrological suitability for an encroaching tree species in a Neotropical flooded savanna. Forest Ecology and Management, 429, pp. 244-255.

Breiman, L., Friedman, J., Olshen, R., and Stone, C., 1984. Classification and Regression Trees. Belmont: Wadsworth International Group.

Carpenter, G., Gillison, A. N., & Winter, J. 1993. DOMAIN: A flexible modelling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation, 2(6), pp. 667-680.

Fick, S., & Hijmans, R., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37.

Guisan, A., and Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135 (2-3), pp. 147-186.

Jensen, J. R., 2015 Thematic Information Extraction: Pattern Recognition. Introductory Digital Image Processing: A Remote Sensing Perspective, Prentice-Hall Inc, USA.

Kholia, B. s, Sharma, S., & Sinha, B., 2019. Brainea insignis (Hook.) J.Sm. – A conservation priority fern of North East India. Current science, 116, pp. 32-34.

Lin, C. H., 2002. The establishment of pteridophytic vegetation in the disturbed sites of Guandaushi forest ecosystem. Dept. of Botany, Chung-Hsing University, master thesis.

Lo, N. C., Chang, W. I., Huang, K. Y., 2011. Application of 3S and Multivariate Statistics to Predict the Potential Habitat of *Elaeocarpus Japonicus* and *Castanopsis carlesii.* Quarterly Journal of Forest Research, 33 (3), pp. 55-70.

Maciel, E. A., Oliveira-Filho, A. T., Sobral-Souza, T. S., Marimon, B. S., Cupertino-Eisenlohr, M. A., José-Silva, L., & Eisenlohr, P. V. 2021. Climate change forecasts suggest that the conservation area network in the Cerrado-Amazon transition zone needs to be expanded. Acta Oecologica, 112, pp. 103764.

Nix, H.A. 1986. A biogeographicanalysis of Australian elapid snakes. In' Atlas ofElapid Snakes ofAustralia', ed R. Longmore, Australian Flora and Fauna Series, 7, pp. 4-15.

Okolie, C. J., & Smit, J. L., 2022. A systematic review and meta-analysis of Digital elevation model (DEM) fusion: Pre-processing, methods and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 188, pp. 1-29.

Pandit, S. N., Maitland, B. M., Pandit, L. K., Poesch, M. S., & Enders, E. C. 2017. Climate change risks, extinction debt, and conservation implications for a threatened freshwater fish: Carmine shiner (Notropis percobromus). Science of the Total Environment, 598, pp. 1-11.

Qin, A., K. Jin, M. E. Batsaikhan, J. Nyamjav, G. Li, J. Li, Y. Xue, G. Sun, L. Wu, T. Indree, Z. Shi, and W. Xiao., 2020. Predicting the current and future suitable habitats of the main dietary plants of the Gobi Bear using MaxEnt modeling. Global Ecology and Conservation, 22, e01032.

Sarma, K., Roy, S. J., Kalita, B., Baruah, P. S., Bawri, A., Nath, M. J., Baruah, U. D., Sahariah, D., Saikia, A., & Tanti, B. 2022. Habitat suitability of Gymnocladus assamicus—A critically endangered plant of Arunachal Pradesh, India using machine learning and statistical modeling. Acta Ecologica Sinica, S1872203222000361.

Tsai, C. H., 1997. Studies on the regeneration of distributed sites at Guandaushi forest ecosystem. Dept. of Botany, Chung-Hsing University, master thesis.

Tu, C. H., 2013. The effects of different statistical methods on the ecological niche model of *Castanopsis carlisii*. Dept. of Forestry, Chung-Hsing University, master thesis.

Vanfretti, L., and V. S. N. Arava., 2020. Decision tree-based classification of multiple operating conditions for power system voltage stability assessment. International Journal of Electrical Power & Energy Systems, 123, pp 106251.

Wang, W. C., 2012. Using species distribution models in GIS to predict the spatial pattern of rare and endangered plant—*Brainea insigni*. Dept. of Forestry, Chung-Hsing University, master thesis.

Yang, S. Z., & Tseng, Y. H., 2008. Quality Assessment for Digital Elevation Models. Journal of Photogrammetry and Remote Sensing, 13 (3), pp. 195-206.

Zeng, S. J., Ma, D. W., Liao, Y. F., Lin, W. T., Wang, P., Liu, Q., Hong, W. J., Zhu, C. Y., Zhuang, X. Y., 2016. Communities Characteristics with the Populations of Brainea insignis in Lianhuashan-Baipenzhu Nature Reserve, Guangdong. Jour of Fujian Forestry Sci and Tech. 43 (2), pp. 8-13.