Taiwania 67(1): 9–20, **2022**
DOI: 10.6165/tai.2022.67.9

The effects of true and pseudo-absence data on the performance of species distribution models at landscape scale

Chi-Cheng LIAO[1,*], Yi-Huey CHEN

Department of Life Science, Chinese Culture University, Taiwan, R.O.C. [*]Corresponding author's email: hunter_yyl@yahoo.com.tw; Tel: +886-2-2861-0511 ext. 26233; Fax: +886-2-2861-7507

(Manuscript received 4 July 2021; Accepted 29 November 2021; Online published 2 January 2022)

ABSTRACT: Potential distribution ranges of natural grassland in subtropical humid mountainous areas were predicted by species distribution models (SDMs) to examine the effects of true and pseudo-absence data on model performances that were scarcely assessed by using real data. Climate spaces of potential ranges of natural grassland were then constructed by principal components analysis (PCA). The distribution map projected by six model algorithms based on true absence data had all presented restricted distribution ranges of natural grassland along mountain ridges, whereas that based on pseudo-absence data presented wider distribution ranges. RF model was used to detect the effects of data record number and contribution of climate variables on model performance because of higher True Skill Statistics. Restricted distribution ranges of natural grassland projected by RF based on true absence data were similar to limited climate space quantified by PCA. However, climate variables related to occurrences of natural grassland were not consistent between RF and PCA results. Occurrences of natural grassland associated with treeline at low elevation were presumably determined by multiple climate factors at subtropical mountain ridges, such as relatively lower temperatures, heavy precipitations, and strong winds. Local climate dataset derived from meteorological stations and followed by altitudinal adjustment was available for modeling species distribution range in mountainous areas. Conclusively, true absence data had practically delineated geographical boundaries and characterized the climate environments of natural grassland. True absence data was recommended to collect along a known environmental gradient and used to construct training dataset with pseudo-absence data to improve model performance.

KEY WORDS: High-resolution climate dataset, natural grassland, principal components analysis, species distribution models, Taiwan.

INTRODUCTION

Species distribution models (SDMs) were widely used to predict species distribution range and evaluate potential impacts of climate change on shifting species range (Mohapatra *et al*., 2019; Xu *et al*., 2021; Zhu *et al*., 2018). SDMs provide useful information on understanding how environmental factors control the distribution of species, which is essential for prioritizing places for biodiversity conservation (Brunialti and Frati, 2021; Dubuis *et al*., 2011; Gies *et al*., 2015). Ecologists and conservationists had increasingly relied on model predictions as a means for estimating species distribution patterns and informing conservation and planning management strategies for rare or endangered species (Guillera−Arroita *et al*., 2015; Kier *et al*., 2009; Peng *et al*., 2019; Porfirio *et al*., 2014; Tomlinson *et al*., 2020; Tsoar *et al*., 2007; Xu *et al*., 2021). Since the last century, rare or endangered species have been threatened by continuous expansion of human colonization areas that had seriously caused forest fragmentation and suitable habitat isolation (Anderson-Teixeira *et al*., 2015; Arroyo-Rodriguez *et al*., 2015). Urban expansion has also resulted in complex landscapes with mixed natural and artificial ecosystems, and delineation of conservation areas is a great challenge in such landscapes. Therefore, an accurate map presenting empirical species distribution range is particularly essential for conservation management in landscapes with mixed natural and artificial ecosystems.

SDMs correlate presence-only or presence-absence data of species to relevant environmental variables and project the potential distribution range of species (Elith and Leathwick, 2009; Peterson *et al*., 2011). Previous studies had recommended several global climate datasets that had improved performance of SDMs at continental scales (Booth *et al*., 2014; Fick and Hijmans, 2017; Title and Bemmels, 2018), while others had proposed several high-resolution environmental datasets to powerfully improve the performance of SDMs at landscape scale (Lannuzel *et al*., 2021; Liao *et al*., 2021; Pradervand *et al*., 2014; Tomlinson *et al*., 2020). Local climate dataset, interpolated from meteorological data followed by altitudinal adjustment, was a high-quality climate dataset that had accurately captured climate heterogeneity along a topography and had successfully predicted the potential distribution range of species at landscape scale (Liao and Chen, 2021). In this study, interpolated and altitudinal adjusted climate dataset from local meteorological data was applied to predict species distribution pattern in mountainous areas at landscape scale.

In addition to climate datasets, presence and absence data of species are also critical factors influencing the prediction accuracy of SDMs (Senay *et al*., 2013). Presence data is species georeferenced occurrences directly collected in the field or resulted from efforts to

digitize and reference to geographical coordinates of specimens held in museums and herbaria. Bias collection of presence data is one of the major types of spatial error (Anacker *et al*., 2013). Particularly, mountainous areas support patchy habitats and steep climatic gradients along slopes (Dobrowski, 2011; Lannuzel *et al*., 2021) had caused fragmented and disjunct distributions of plant individuals in mountainous evergreen broadleaved forests that had usually resulted in bias collections of presence data. On the other hand, there are two types of absence data that have effects on performance of SDMs (Dupin *et al*., 2011). Models built with true absence data had the best predictive power (Wisz and Guisan, 2009) when ecologists had ensured unbiased sampling of true absence data across landscapes (Peterson *et al*., 2011). Meanwhile, true absence data is usually not available and may be very difficult to be detected in the field (Hegel *et al*., 2010; Peterson *et al*., 2011; Qiao *et al*., 2019). Thus, true absence data was mostly substituted by pseudo-absence data in model predictions and the most effective method to generate pseudo-absence data is random selection of points from background area (Barbet-Massin *et al*., 2012; Chapman *et al*., 2019; Liang *et al*., 2018; Senay *et al*., 2013). Pseudo-absence data, in comparison with true absence data, is easy to be constructed in model predictions. Pseudo-absence data was geographically and environmentally more distant from presence data that was appropriate to play as a substitute of true absence data when true absence data was not available in model prediction. To our knowledge, true absence data was scarcely examined by using real data in model predictions, particularly in mountainous areas, because unbiased and comprehensive collection of true absence data for SDMs is a difficult task (Senay *et al*., 2013). In this study, model algorithms were calibrated by presence and true/pseudo-absence data to examine the effects of these data on model performances in mountainous area.

In order to examine the effects of true and pseudo-absence data on model performance at landscape scale, natural grassland at subtropical humid mountainous areas was predicted by SDMs. Natural grassland is a prominent and persistent vegetation type distinguished from neighboring evergreen broadleaved forests at subtropical humid mountainous areas with elevations lower than 1000 m above sea level (asl.) (Li *et al*., 2013). True absence data of natural grassland is easily identifiable in field survey, making it possible to examine the effects of true and pseudo-absence data on model performances. Six model algorithms used to project distribution range was applied by the presence and true/pseudo-absence datasets of natural grassland in this study. Model performances were evaluated by True Skill Statistics (TSS) and receiver operating characteristic (ROC) curve. Subsequently, random forest algorithm (RF) was performed to predict potential distribution range of natural grassland and to correlate climate variables and the occurrences of natural grassland. Accuracy of RF prediction power was commonly detected by the area under the receiver operating characteristic curve (AUC) in previous studies (Chapman *et al*., 2019; Lannuzel *et al*., 2021; Lobo *et al*., 2008; Tomlinson *et al*., 2020; Xu *et al*., 2021; Zhu *et al*., 2018) and is used in this study to indicate the accuracy of RF model performance.

Furthermore, principal components analysis (PCA) was performed to correlate occurrences of natural grassland and climate factors in subtropical humid mountainous areas. The occurrences of natural grassland in this area are consistent with the presence of treeline. Treeline is a prominent edge of forest ecosystems that commonly appeared in the alpine zone and was characterized by harsh environments, such as cold soil temperature (Korner, 1998; Liu *et al*., 2011; Smith *et al*., 2009). Limitation of treeline and climate characteristics of natural grassland at low elevation was presumably related to particular climate environments but was seldom studied in subtropical humid mountainous areas.

This study aims to examine the effects of true and pseudo-absence data on the model performance. Model algorithms based on the true absence data was hypothesized to have projected restricted distribution range of natural grassland at mountain ridge, since true absence data was geographically close to the presence data. Model algorithms based on pseudo-absence data was assumed to project a wider ranges of natural grassland along mountain ridges, since pseudo-absence data was random points selected throughout the gridded cells in the study area and was geographically more distant from the presence data. Occurrences of natural grassland at low elevation in subtropical mountainous areas were also correlated to climate factors in this study.

MATERIALS AND METHODS

Study area

The study area is in northern Taiwan (24°57′–25°17′N, 121°24′–122°00′E). In this study, northern Taiwan (NTWN) was divided into five watersheds. The Yangmingshan area (YMSA) was divided into four watersheds, they are northeast (NE), northwest (NW), southwest (SW) and southeast (SE) slopes of YMSA and Pingxi area (PX) is the fifth watershed (Fig. 1). The highest peak of the YMSA is 1,120 m above sea level (asl.) and that of Pingxi area is 757 m asl. The area of study site is about 1,031 square kilometers (103,100 hectares). NTWN is characterized by the subtropical monsoon climate (Chen and Tsai, 1983). The mean monthly temperatures range from 8.6 ℃ in January to 25.5 ℃ in August and the annual total precipitation is more than 3,500 mm. Northeast wind in winter and typhoon in summer constantly transport moisture to the study area which leads to a relatively stable humid conditions and high frequency of cloud cover. There is no significant dry season in the study area.

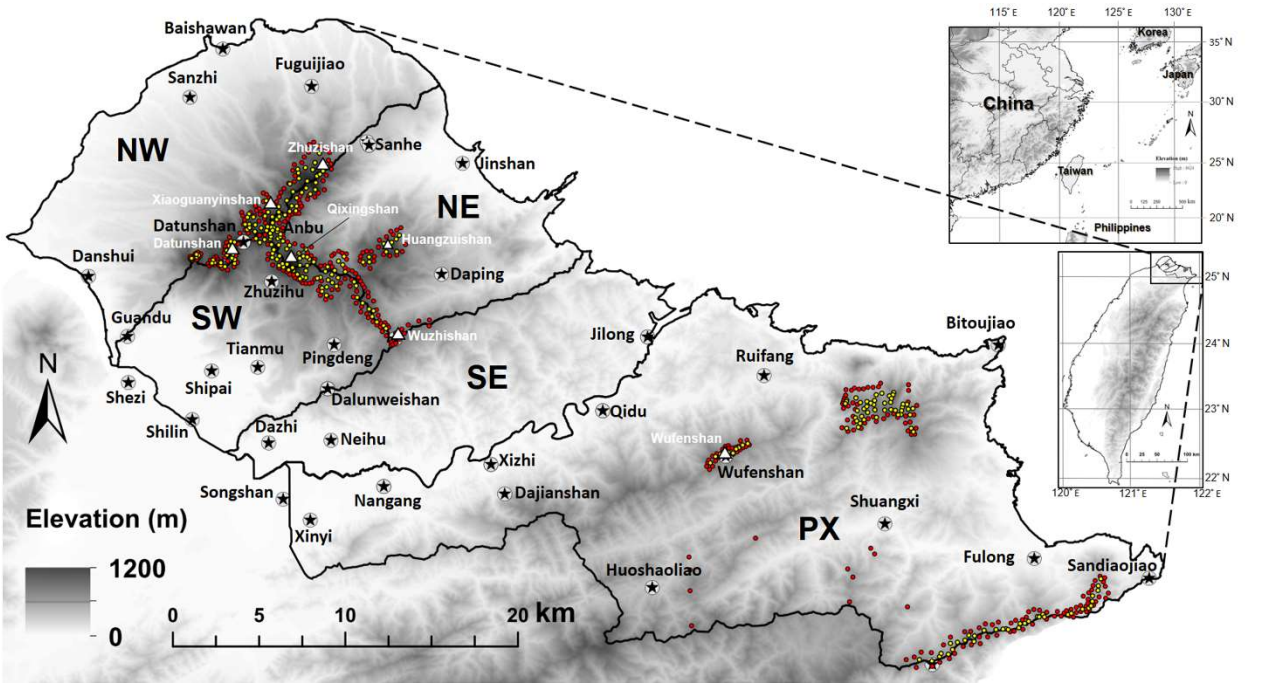

Fig. 1. Maps show geographical location of Taiwan and neighboring countries (upper right map) and georeferenced location of presence (yellow solid circles in the central map) and absence data (red solid circles) of natural grassland in northern Taiwan. Northern Taiwan was divided into five watersheds for calculating empirical lapse rates of climate data. The Yamingshan area (YMSA) was divided into four watersheds, they are northeast (NE), northwest (NW), southwest (SW) and southeast (SE) slopes of YMSA and Pingxi area (PX) is the fifth watershed. The locations of 30 meteorological stations adopted in this study were represented by solid star within a circle. Mountain tops were represented by white triangles.

Evergreen broad-leaved forest is the major vegetation type in NTWN (Hsieh *et al*., 1997; Li *et al*., 2013; Liao *et al*., 2012). The forests are dominated by species of *Castanopsis*, *Cleyera*, *Cyclobalanopsis*, *Dendropanax*, *Elaeocarpus*, *Engelhardia*, *Gordonia*, *Helicia*, *Ilex*, *Keteleeria*, *Limlia*, *Litsea*, *Machilus*, *Meliosma*, *Michelia*, *Pinus*, *Schefflera*, *Symplocos*, and *Trochodendron* with a mean canopy height of 10 m (Li *et al*., 2013). Natural grassland, with the *Miscanthus sinensis* and *Pseudosasa usawai* being the dominant species, free from anthropogenic disturbance was frequently observed at mountain ridge from coast to inland in the study area (Liao *et al*., 2012; Liao *et al*., 2014). Long-term persistence of natural grassland along mountain ridges at low elevation was indirectly indicated by ancient documentations and was empirically related to climate factors (Liao *et al*., 2014), while grasslands around farmland pronouncedly caused by anthropogenic disturbances were not target vegetation in this study.

Vegetation data collection

Presence and absence data of natural grassland were used to construct the data matrix for model evaluations. Presence and absence data for natural grassland were collected along the roads and mountain trails in NTWN. Practically, there is an abrupt transition from evergreen broadleaved forests to natural grassland along mountain slopes in NTWN. Thus, the presence data of grassland was defined as the vegetation without shrub or trees, while absence data of grassland as the closed-canopy forests. Duplicated records of the presence data were spatially verified to ensure only one occurrence within each gridded cell. A total of 252 presence and 372 absence data records were available for modeling the distribution of natural grassland (Fig. 1). As absence data was required for the model evaluations, two types of absence data were used in this study. Absence data collected in the field was re-named as true absence, whereas pseudo-absence data (or background points) were random points selected throughout the gridded cells in the study area. Presence and true/pseudo-absence data were used to construct the training datasets for model predictions.

Climate data

NTWN was divided into gridded cells with spatial resolution of 50 × 50 m^2 and a total of more than 0.4 million gridded cells was generated for construction of the local climate dataset in this study. Spatial size of 50 × 50 m^2 is attempting to capture steep environmental features along mountain slopes and making climate environments over landscapes more prominent and distinguishable (Liao and Chen, 2021). For each gridded cell, longitude, latitude, and elevation were obtained from a digital terrain model (DTM) with a resolution of 20 by 20 meters been developed by the Department of Geography, Chinese Culture University. Climate variables

of gridded cells were extracted from the climate surfaces of watersheds. The climate surfaces were derived from daily data of meteorological stations by means of a downscaling procedure performed by ArcInfo software (ESRI, Redlands, California, USA). Daily data of meteorological stations in NTWN from the year 2000 to 2020 were downloaded from the website of Central Weather Bureau (CWB, https://www.cwb.gov.tw/V7/forecast/). Mean monthly temperature and mean total precipitation were calculated from daily data for each meteorological station. Meteorological stations in NTWN were categorized into five groups based on the five watersheds of the study areas (Fig. 1). The mean monthly data of meteorological stations were imported to ArcInfo software and were used for interpolation by means of Inverse Distance Weighted (IDW) method to generate raster files representing climate surfaces of the watershed. Smooth elevation surface (Elev1) for each watershed was also generated by the IDW method implemented by ArcInfo software based on the elevation of meteorological stations within the watershed. Subsequently, the gridded cells were mapped in the ArcInfo software and overlapped with the raster files of climate surfaces to extract monthly climate data. This was the procedure to create preliminary gridded climate dataset (ClimData1).

The preliminary gridded climate dataset (ClimData1) was then adjusted by empirical elevation lapse rates varied among five watersheds. Empirical lapse rates were calculated on the basis of the meteorological stations within the watersheds. The climate data from Datunshan and Anbu meteorological stations, near the mountain ridge of YMSA represented the climate environments of the mountain ridge, were used to calculate lapse rates of climate data in each of the four watersheds of YMSA. Linear regression model implemented by "stats" package within the R environment (Chambers and Hastie, 1992) was applied to calculate lapse rates (slope and intercept of linear function) for each watershed. For each watershed, mean monthly temperature, wind speed, and monthly total precipitation calculated from recorded data of meteorological stations were the dependent variables, and elevation of the stations was the independent variables of the linear function. Altitudinal adjusted climate data (ClimData2) were calculated based on the preliminary gridded climate dataset (ClimData1), smooth elevation surface (Elev1) interpolated from meteorological stations, and slopes of the regression line fitted elevation and monthly temperature, precipitation, and wind speed of meteorological stations. The preliminary gridded climate dataset (ClimData1) and smooth elevation surface data (Elev1) were then re-projected to actual elevation data from the 20 m DTM (Elev2) to calculate ClimData2. The altitudinal adjusted climate data (ClimData2) were calculated by the function: ClimData2 – ClimData1 = slope × (Elev2 – Elev1). The altitudinal adjusted climate data (ClimData2) were used to construct climate dataset of model algorithms. The climate scenarios were created taking into account the following 13 variables: mean annual temperature (Tann), mean maximum temperature of the warmest month (Twrm), mean minimum temperature of the coldest month (Tcld), mean temperature in summer (Tsmr) and winter (Twnt), temperature differences between warmest and coldest months (Tdif), annual total precipitation (Pann), total precipitation in summer (Psmr) and winter (Pwnt), mean wind speed of the warmest month (WSwrm) and coldest month (WScld), and mean wind speed in summer (WSsmr) and winter (WSwnt).

Modelling technique

Six model algorithms were implemented through the "biomod2" package in R software (Thuiller *et al*., 2016) to predict the potential distribution range of natural grassland. These models include (1) two machine learning algorithms, random forest (RF) and artificial neural network (ANN); (2) two regression methods, general linear model (GLM) and generalized additive model (GAM); and (3) two classification methods, flexible discriminant analysis (FDA) and classification tree analysis (CTA). The dataset to be evaluated by model algorithms in this study is the gridded cells with altitudinal adjusted climate data (ClimData2). The training datasets were constructed by the presence and true/pseudo-absence data of natural grassland and climate data that was extracted from the closest cells of altitudinal adjusted gridded climate dataset according to the coordinates of presence and true/pseudo-absence data. To assess model accuracy, a random set of 80% of the presence and absence data was used to train the model, and the remaining 20% was used for evaluation. The training dataset was modeled 100 times with the resampled training dataset by model algorithms. Prediction results (Fig. S1) and model accuracy represented by True Skill Statistics (TSS) and receiver operating characteristic (ROC) curve of the six model algorithms (Fig. S2) were presented in the supplement. Among the six model algorithms, RF algorithm had the highest TSS value and was further utilized for evaluating impacts of the number of data records on model performance and for correlating climate variables and potential distribution range of natural grassland. RF algorithm is a machine learning method that handles numerous variables and is well suited to the complex data set (Breiman, 2001). RF is capable of detecting complex relationships among model variables without making a prior assumption about the type of relationship (Breiman, 2001).

Presence and true/pseudo-absence data of natural grassland were randomly re-sampled 50, 100 and 200 data records to create different sizes of training datasets and subsequently modeled 100 times by RF algorithm with the resampled training datasets to quantify uncertainties in predictions. These procedures allowed us

Table 1. Importance of predictor variables generated by Random Forest based on the true and pseudo-absence data. Predictor variables contributed most to the models based on the true and pseudo-absence data were different.

	True absence data			Pseudo-absence data		
	50	100	200	50	100	200
Tann	6.9 ± 1.1	7.3 ± 0.9	7.3 ± 0.7	14.5 ± 3.6	14.8 ± 2.5	13.2 ± 2.7
Twrm	7.9 ± 1.4	8.3 ± 1.1	8.1 ± 0.9	16.9 ± 3.7	17.8 ± 3.7	18.7 ± 3.1
Tcld	6.5 ± 1.4	7.4 ± 1.1	7.8 ± 0.8	9.0 ± 3.0	8.1 ± 2.2	6.6 ± 1.8
Tsmr	7.4 ± 1.4	8.1 ± 1.2	7.9 ± 0.7	19.1 ± 4.4	19.4 ± 3.6	19.8 ± 2.8
Twnt	6.8 ± 1.3	7.2 ± 0.9	7.1 ± 0.7	9.5 ± 2.8	9.2 ± 2.1	7.9 ± 1.8
Tdif	8.4 ± 1.9	9.4 ± 1.4	10.8 ± 1.1	4.0 ± 1.5	5.0 ± 1.4	6.6 ± 1.4
Pann	11.7 ± 2.9	8.1 ± 1.3	7.8 ± 0.9	3.0 ± 1.6	3.2 ± 1.3	2.6 ± 0.7
Psmr	8.6 ± 2.0	8.0 ± 1.2	6.9 ± 0.8	4.0 ± 2.8	2.6 ± 0.9	2.1 ± 0.6
Pwnt	9.3 ± 1.7	7.8 ± 1.2	7.7 ± 0.7	2.0 ± 1.1	2.5 ± 0.8	2.4 ± 0.5
WSwrm	6.4 ± 1.2	7.2 ± 1.0	7.5 ± 0.7	3.6 ± 1.6	3.2 ± 1.0	3.6 ± 0.9
WScld	7.0 ± 1.6	7.3 ± 1.2	7.0 ± 0.6	5.4 ± 2.4	6.3 ± 2.4	7.5 ± 2.1
WSsmr	6.6 ± 1.1	7.3 ± 1.0	7.4 ± 0.6	3.7 ± 1.6	3.5 ± 1.3	3.7 ± 1.0
WSwnt	6.2 ± 1.1	6.4 ± 0.8	6.6 ± 0.5	5.2 ± 2.4	5.2 ± 1.7	5.4 ± 1.5

Note: The number presented in the column names are sample sizes of training datasets. The value in each cell is mean ± standard deviation. Tann: mean annual temperature; Twrm: mean maximum temperature of the warmest month; Tcld: mean minimum temperature of coldest month; Tdif: temperature differences between warmest and coldest months; Tsmr: mean temperature in summer; Twnt: mean temperature in winter; Pann: annual total precipitation; Psmr: total precipitation in summer; Pwnt: total precipitation in winter; WSwrm: mean wind speed of the warmest month; WScld: mean wind speed of the coldest month; WSsmr: mean wind speed in summer; WSwnt: mean wind speed in winter.

to generate a range of training datasets with contrasting sizes corresponding to the bias field collections of georeferenced data in mountainous areas. The RF was implemented by the "randomForest" library within the R software (Breiman, 2001; Liaw and Wiener, 2002). The area under the receiver operating characteristic curve (AUC) was used to assess the RF model performance (Fois *et al*., 2015; Lannuzel *et al*., 2021; Qiao *et al*., 2019; Xu *et al*., 2021).

Quantification of climate spaces by principal components analysis (PCA)

Principal components analysis (PCA) implemented by "prcomp" package in R software was performed to correlate climate variables and occurrences of natural grassland. The climate space of the natural grassland was quantified by climate variations along significant axes, defining ecological preferences and climate environments. Six datasets were used for PCA quantification of climate spaces, and they were the background points, potential ranges of natural grassland projected by RF based on true and pseudo-absence data, presence and true absence data of natural grassland, and meteorological stations. PCA was applied to scaled data for 13 climate variables corresponding to the formation of climate spaces of natural grassland. Among the 13 climate variables implemented to PCA, Pann, Psmr, and Pwnt were rescaled from mm to dm. The 13 climate variables were thought to provide climate preferences for the distributions of natural grassland and the ecological demands were distilled into three principal components, the first, second, and third axes from a PCA. Analysis of variance (ANOVA) and a Tukey's HSD post-hoc test were performed to identify differences of climate data among meteorological stations to evaluate ecological preference and climate environments of natural grassland.

RESULTS

Distribution map projected by the six model algorithms based on true absence data presented a restricted distribution range of natural grassland along mountain ridges in the study area (Fig. S1), whereas that based on pseudo-absence data presented a wider distribution range (Fig. S1). The effect of true absence data played a role in restricting the potential distribution range of natural grassland when modeling by SDMs. TSS and ROC scores based on the presence and true/pseudo-absence data showed no conspicuous trend among the six model algorithms (Fig. S2).

Among the six models, RF algorithm had the highest TSS values based on the presence and pseudo-absence data and was further used to evaluate the effect of data record numbers on the model performance and to correlate climate factors and the presence of natural grassland. Interestingly, the number of presence/absence data records has negligible effect on the RF model performances. Projection map of RF evaluated by 50 data records (Fig. 2A and 2B) of training dataset had a similar potential range contrasted to the maps evaluated by 100 (Fig. 2B and 2E) or 200 data records (Fig. 2C and 2F). Conclusively, true and pseudo-absence data had evident effects on evaluating species distribution range, while different numbers of presence/absence data records had weak effects on projecting species distribution range, regardless of true or pseudo-absence data (Fig. 2).

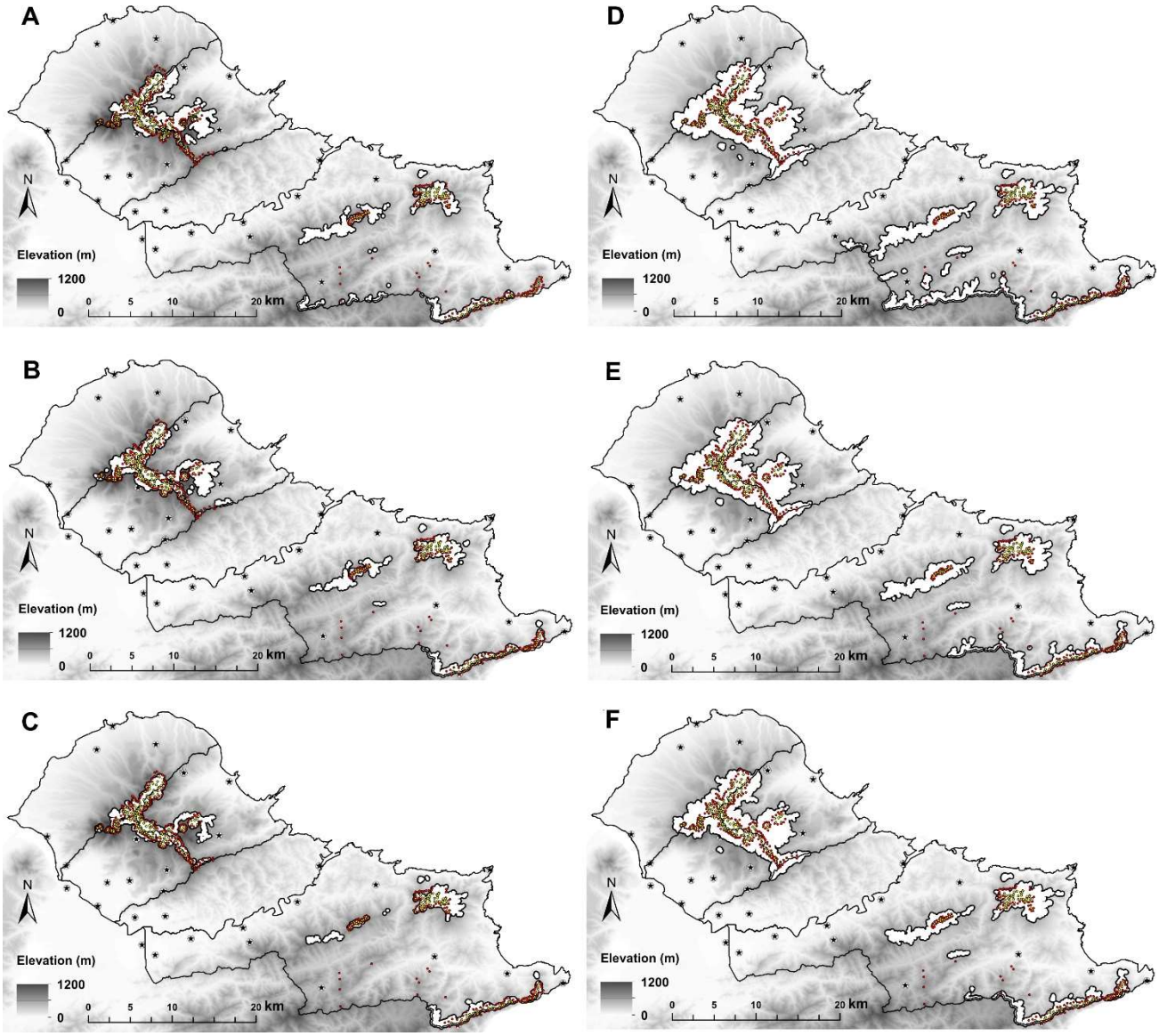

Fig. 2. Contemporary potential distribution range of natural grassland in northern Taiwan (white mask with dark outline in the maps) projected by Random Forest. Randomized re-samples of presence and true absence data (A, B, and C) had precisely projected distribution range of grassland at mountain ridge, while that of presence and pseudo-absence data (D, E, and F) had projected wider range of grassland along mountain ridge. The sample sizes of training datasets were 50 presence and absence data records (A and D), 100 records (B and E), and 200 records (C and F). The locations of 30 meteorological stations adopted in this study were represented by solid star within a circle.

Pann and Tdif contributed most to the RF model predictions based on the presence and true absence data, whereas Tsmr, Twrm, and Tann were the most important predictors that strongly influenced model performance based on the presence and pseudo-absence data (Table 1). Important predictors that contributed to the model predictions were not consistent between the two types of absence data. The AUC scores based on the presence and true absence data were lower than those based on the presence and pseudo-absence data (Fig. 3). Typically, higher AUC scores indicated better performance of SDMs. Accordingly, higher AUC score of RF predictions based on presence and pseudo-absence data was supposed to indicate better model performance. However, the potential range projected by RF based on pseudo-absence data was wider than the geographical range of true absence data and was certainly wider than the realistic range of natural grassland. Wider potential range leads to an inaccurate model performance since true absence data were locations of evergreen broadleaved forests and was geographically close to the boundaries of natural grassland. True absence data delineate the natural contemporary species distribution range guaranteed more accurate model prediction result.

Table 2. The first three axes of the principal components analysis (PCA) on the correlation matrix of climate variables from the local climate dataset. Pann and WSwnt had significantly correlated with the axes I and II of PCA, respectively.

	PC1	PC2	PC3
Tann	-0.1155	-0.2389	0.2733
Twrm	-0.1067	-0.3290	0.2157
Tcld	-0.1319	-0.2335	0.4586
Tsmr	-0.1191	-0.2375	0.2560
Twnt	-0.1122	-0.2432	0.2953
Tdif	-0.0212	-0.0189	0.0078
Pann	0.8562	-0.2245	-0.0772
Psmr	0.0615	-0.0219	-0.1609
Pwnt	0.4091	-0.1841	0.4021
WSwrm	0.0280	0.1772	0.1051
WScld	0.0302	0.2146	0.1867
WSsmr	0.1334	0.4935	0.4019
WSwnt	0.0830	0.5084	0.3388

Tann: mean annual temperature; Twrm: mean maximum temperature of the warmest month; Tcld: mean minimum temperature of coldest month; Tdif: temperature differences between warmest and coldest months; Tsmr: mean temperature in summer; Twnt: mean temperature in winter; Pann: annual total precipitation; Psmr: total precipitation in summer; Pwnt: total precipitation in winter; WSwrm: mean wind speed of the warmest month; WScld: mean wind speed of the coldest month; WSsmr: mean wind speed in summer; WSwnt: mean wind speed in winter.

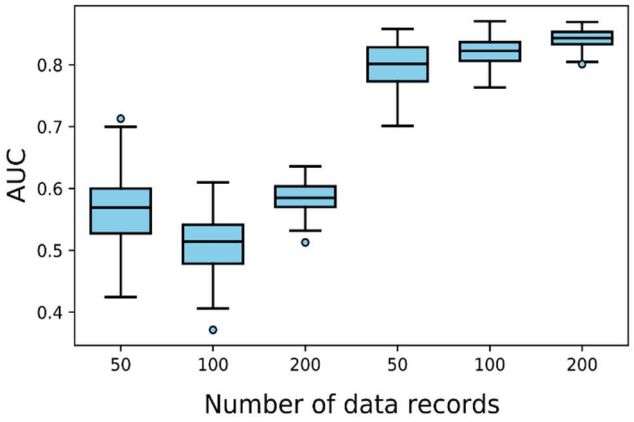

Fig. 3. AUC (area under the receiver operating characteristic curve) scores of RF model based on true (left three boxes) and pseudo-absence data (right three boxes). The number at x-axis are sample sizes of training dataset that were randomly re-sampled from presence/true-absence and presence/pseudo-absence datasets.

Climate spaces quantified by PCA

PCA had quantified climate spaces of presence data, true-absence data, the potential range of natural grassland based on the presence and true/pseudo-absence data, background points and meteorological stations (Fig. 4). Principal component 1 (PC1) accounted for 81.32% of the variation, while principal component 2 (PC2) accounted for 11.67%. Water availability and wind speed had evidently played as the major role for the quantification of climate spaces, since they were significantly correlated with the PC1 and PC2, respectively (Table 2).

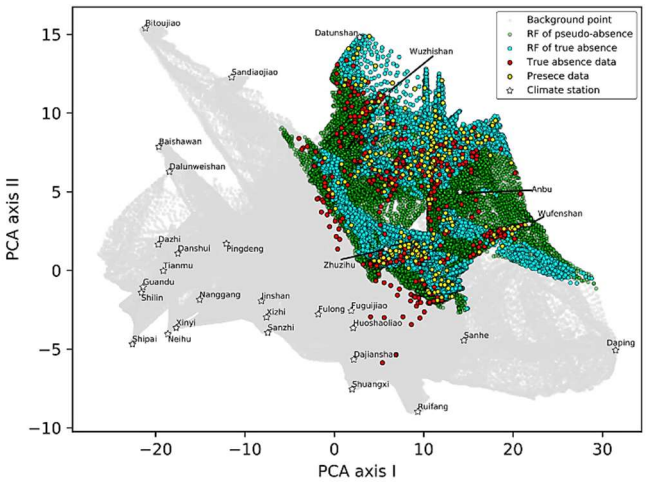

Fig. 4. Climate spaces quantified by principal components analysis (PCA). PCA coordination constructed by PCA I and II. The grey circles are background points. Green and sky-blue points are climate spaces of grassland's potential distribution range projected by RF based on the true absence and pseudo-absence data, respectively. Yellow and red points are presence and true absence data of natural grassland, respectively. Empty stars represent meteorological stations. Locations of four meteorological stations, Anbu, Datunshan, Wuzhishan, and Wufenshan, were overlapped with the presence data, absence data, and potential range of natural grassland. Zhuzihu station is distant from the potential range of natural grassland along PCA axis III in the diagram.

Climate spaces of presence data, true absence data, and potential range of natural grassland projected by RF based on the two types of absence data were overlapped in the PCA diagram (Fig. 4). The potential range of natural grassland based on pseudo-absence data had wider climate space than that based on true absence data. Wider climate space quantified by PCA was consistent with wider potential range of natural grassland projected by RF algorithm. In PCA diagram, locations of true absence data were close to the locations of presence data that was similar to the small geographical distances between presence and true-absence data. Four meteorological stations, Datunshan, Anbu, Wuzhishan, and Wufenshan, were geographically close to the natural grassland that had represented climate characteristics of the natural grassland. Climate environments of the four meteorological stations were characterized by significantly lower temperatures, higher annual and winter precipitations, and strong winds that were significantly different from most of the other meteorological stations detected by ANOVA and Tukey HSD post-hoc statistical test (Table 3).

DISCUSSION

In this study, six model algorithms had all projected similar patterns of potential distribution range of natural grassland and RF algorithm had correlated climate data

Table 3. Climate characteristics of meteorological stations. Climate environments of the four meteorological stations located near natural grassland (the first four rows) were significantly different from the other 25 stations. Temperature of the four meteorological stations were significantly lower than the other 25 stations, while precipitation and wind speed were significantly higher than most of the stations.

Station	Tann	Tmax	Tmin	Tsmr	Twnt	Tdif	Pann	Psmr	Pwnt	WSwrm	WScld	WSsmr	WSwnt
Datunshan	10.2 ± 1.2^{n}	27.9 ± 1.1^{o}	0.7 ± 1.0^{f}	20.5 ± 0.6^{n}	11.3 ± 0.6^{l}	12.9 ± 1.2^{abc}	3586.4 ± 565.6^{fghi}	838.0 ± 301.2^{ab}	825.8 ± 226.1^{defg}	4.1 ± 0.6^{c}	3.9 ± 0.4^{b}	4.9 ± 0.4^{a}	4.1 ± 0.2^{c}
Anbu	11.7 ± 1.2^{m}	29.2 ± 0.8^{o}	1.9 ± 2.0^{f}	21.8 ± 0.5^{m}	12.6 ± 0.7^{k}	13.0 ± 1.1^{abc}	4844.5 ± 970.1^{bcd}	959.1 ± 351.8^{a}	1267.6 ± 437.6^{bc}	2.9 ± 0.6^{de}	3.2 ± 0.3^{d}	2.4 ± 0.3^{c}	3.2 ± 0.3^{d}
Wuzhishan	12.7 ± 1.1^{l}	30.5 ± 0.7^{mn}	3.4 ± 1.7^{ef}	22.6 ± 0.6^{l}	13.6 ± 0.6^{j}	13.0 ± 1.0^{abc}	3892.0 ± 735.6^{efg}	766.7 ± 263.5^{ab}	1147.7 ± 440.7^{cde}	4.8 ± 0.7^{c}	6.2 ± 0.6^{ab}	4.7 ± 0.3^{a}	6.6 ± 0.5^{b}
Wufenshan	13.4 ± 1.1^{lm}	30.6 ± 1.2^{no}	4.0 ± 1.7^{def}	22.2 ± 0.4^{lm}	14.1 ± 0.3^{ij}	11.5 ± 0.9^{c}	5490.5 ± 397.6^{ab}	918.2 ± 196.8^{ab}	1881.2 ± 335.2^{b}	3.9 ± 0.4^{cd}	4.4 ± 0.2^{c}	3.3 ± 0.3^{b}	4.7 ± 0.1^{c}
Baishawan	17.9 ± 0.4^{ghi}	33.4 ± 0.9^{hijkl}	12.9 ± 0.1^{a}	26.8 ± 0.5^{defg}	18.6 ± 0.1^{abcd}	12.2 ± 0.6^{bc}	1812.5 ± 76.0^{kl}	588.0 ± 127.5^{b}	387.0 ± 86.0^{fghi}	3.6 ± 0.4^{bc}	5.8 ± 0.0^{cd}	3.1 ± 0.2^{bc}	6.4 ± 0.2^{b}
Bitoujiao	18.2 ± 0.8^{cdef}	33.0 ± 1.1^{kl}	10.6 ± 1.9^{a}	26.5 ± 0.4^{fg}	18.8 ± 0.4^{abcd}	11.9 ± 0.8^{c}	1518.5 ± 454.3^{l}	412.1 ± 171.9^{b}	403.9 ± 220.9^{fghi}	4.2 ± 0.8^{a}	7.1 ± 0.6^{c}	3.6 ± 0.4^{b}	7.2 ± 0.5^{a}
Dajianshan	15.0 ± 1.1^{ijk}	33.8 ± 0.9^{ghijk}	5.1 ± 1.8^{de}	24.6 ± 0.4^{ij}	15.7 ± 0.6^{h}	13.1 ± 0.9^{abc}	4066.7 ± 857.0^{efgh}	833.5 ± 235.9^{ab}	937.0 ± 222.8^{cde}	0.4 ± 0.1^{kl}	0.3 ± 0.1^{jk}	0.3 ± 0.1^{lm}	0.3 ± 0.1^{m}
Dalunweishan	14.3 ± 1.1^{ijk}	33.4 ± 1.1^{ijkl}	4.8 ± 1.9^{de}	24.2 ± 0.3^{j}	15.1 ± 0.5^{hi}	13.3 ± 1.0^{ab}	1943.9 ± 405.7^{kl}	613.4 ± 204.1^{b}	248.4 ± 104.6^{hi}	1.3 ± 0.5^{ef}	3.2 ± 0.9^{ghij}	1.1 ± 0.4^{hij}	3.1 ± 0.9^{de}
Danshui	17.3 ± 1.2^{def}	35.9 ± 0.9^{cdef}	7.2 ± 1.9^{bcde}	27.1 ± 0.5^{def}	18.0 ± 0.8^{d}	13.5 ± 1.0^{ab}	2126.4 ± 579.5^{jkl}	644.8 ± 298.2^{ab}	342.2 ± 157.6^{ghi}	1.7 ± 0.3^{gh}	2.1 ± 0.3^{def}	1.6 ± 0.3^{efg}	2.0 ± 0.3^{fgh}
Daping	14.7 ± 1.1^{ij}	34.0 ± 1.2^{ghij}	5.3 ± 1.6^{de}	24.8 ± 0.6^{hij}	15.5 ± 0.6^{h}	13.5 ± 1.0^{ab}	6423.6 ± 1022.3^{a}	903.9 ± 317.4^{ab}	2632.6 ± 817.9^{a}	2.1 ± 0.5^{de}	3.2 ± 0.6^{d}	2.2 ± 0.2^{cd}	3.3 ± 0.4^{d}
Dazhi	18.0 ± 1.1^{abcd}	35.5 ± 0.8^{cdef}	8.8 ± 2.0^{abc}	27.8 ± 0.6^{abcd}	18.8 ± 0.6^{abcd}	13.4 ± 1.0^{ab}	2000.3 ± 503.2^{jkl}	669.5 ± 222.9^{ab}	218.0 ± 97.6^{hi}	2.0 ± 0.3^{gh}	2.2 ± 0.3^{de}	1.8 ± 0.3^{def}	2.3 ± 0.3^{f}
Fuguijiao	16.0 ± 1.1^{hi}	33.5 ± 0.8^{hijk}	7.4 ± 1.8^{abcd}	26.0 ± 0.4^{g}	16.8 ± 0.6^{fg}	13.6 ± 1.1^{a}	3896.3 ± 787.5^{efg}	779.3 ± 339.7^{ab}	1285.7 ± 468.2^{bc}	2.0 ± 0.3^{gh}	2.0 ± 0.3^{de}	2.0 ± 0.2^{de}	2.1 ± 0.4^{fg}
Fulong	17.4 ± 1.0^{fg}	35.0 ± 0.9^{fgh}	8.0 ± 2.0^{abc}	26.3 ± 0.5^{g}	18.1 ± 0.5^{cd}	12.3 ± 0.8^{bc}	3500.5 ± 695.8^{fgh}	667.5 ± 245.0^{ab}	1171.9 ± 418.4^{cd}	1.4 ± 0.3^{gh}	2.0 ± 0.4^{efg}	1.2 ± 0.2^{ghij}	1.9 ± 0.3^{fghi}
Huoshaoliao	14.4 ± 1.3^{ijk}	34.1 ± 1.1^{ghijk}	4.0 ± 2.3^{ef}	24.4 ± 0.6^{ij}	15.5 ± 0.9^{h}	12.5 ± 1.5^{abc}	4066.3 ± 689.8^{cdef}	953.2 ± 283.6^{a}	916.2 ± 312.3^{cde}	0.6 ± 0.2^{ijk}	0.7 ± 0.2^{ijk}	0.4 ± 0.1^{lm}	0.7 ± 0.1^{lm}
Jinshan	17.2 ± 1.1^{fg}	35.8 ± 1.2^{bcd}	8.8 ± 1.9^{abc}	26.6 ± 0.5^{efg}	17.9 ± 0.6^{de}	13.3 ± 0.9^{ab}	2925.4 ± 591.1^{hij}	721.1 ± 283.9^{ab}	816.5 ± 295.1^{def}	1.1 ± 0.3^{hi}	1.8 ± 0.4^{ghi}	1.0 ± 0.2^{ijk}	1.8 ± 0.3^{ghi}
Nanggang	17.7 ± 1.1^{bcdef}	36.1 ± 1.0^{abcd}	8.6 ± 2.0^{abc}	27.7 ± 0.5^{bcd}	18.5 ± 0.6^{abcd}	13.8 ± 1.0^{a}	2581.8 ± 601.6^{ijk}	739.1 ± 257.4^{ab}	357.2 ± 148.0^{ghi}	1.6 ± 0.3^{hij}	1.4 ± 0.5^{efg}	1.4 ± 0.3^{fgh}	1.5 ± 0.5^{ijk}
Neihu	18.1 ± 1.1^{abc}	36.7 ± 0.9^{ab}	8.7 ± 2.0^{abc}	28.1 ± 0.5^{abc}	18.8 ± 0.5^{abcd}	13.7 ± 1.1^{a}	2240.7 ± 507.0^{jkl}	730.2 ± 233.3^{ab}	266.9 ± 106.4^{hi}	0.9 ± 0.4^{jkl}	1.0 ± 0.5^{hijk}	0.8 ± 0.3^{jkl}	1.1 ± 0.5^{kl}
Pingdeng	15.7 ± 1.1^{hi}	32.9 ± 0.5^{ijkl}	7.3 ± 1.4^{abcde}	25.5 ± 0.8^{gh}	16.5 ± 0.2^{efg}	12.4 ± 0.5^{abc}	2697.0 ± 430.7^{ijkl}	767.2 ± 303.2^{ab}	601.8 ± 85.7^{efghi}	1.4 ± 0.2^{fgh}	2.4 ± 0.4^{efghi}	1.4 ± 0.1^{fghij}	2.6 ± 0.1^{ef}
Ruifang	16.6 ± 1.0^{gh}	35.1 ± 1.3^{defg}	7.7 ± 2.1^{abcd}	26.2 ± 0.3^{g}	17.5 ± 0.5^{def}	12.8 ± 1.0^{abc}	4660.6 ± 818.2^{bcde}	778.0 ± 199.4^{ab}	1675.1 ± 555.5^{b}	0.6 ± 0.2^{hijk}	1.0 ± 0.3^{ijk}	0.5 ± 0.1^{klm}	1.0 ± 0.1^{kl}
Sandiaojiao	17.7 ± 0.9^{efg}	34.3 ± 1.1^{ghi}	9.7 ± 1.8^{a}	26.2 ± 0.4^{g}	18.4 ± 0.5^{bcd}	11.9 ± 0.9^{c}	2376.5 ± 436.9^{jkl}	507.7 ± 187.5^{b}	739.4 ± 242.8^{efgh}	4.8 ± 0.8^{ab}	6.5 ± 0.7^{a}	4.6 ± 1.0^{a}	6.2 ± 0.5^{b}
Sanhe	15.6 ± 1.1^{i}	32.8 ± 0.9^{jkl}	6.8 ± 1.9^{bcde}	25.2 ± 0.6^{hi}	16.4 ± 0.7^{g}	13.0 ± 1.0^{abc}	5090.7 ± 778.2^{bc}	901.8 ± 372.5^{ab}	1781.7 ± 579.2^{b}	1.9 ± 0.7^{gh}	1.7 ± 0.5^{def}	1.5 ± 0.5^{fgh}	1.7 ± 0.4^{ghi}
Sanzhi	17.0 ± 0.9^{defg}	34.6 ± 1.3^{efgh}	9.4 ± 2.3^{ab}	27.4 ± 0.6^{cde}	18.1 ± 0.5^{cd}	13.8 ± 1.0^{a}	3068.8 ± 549.2^{hijk}	695.7 ± 305.9^{ab}	1045.2 ± 391.1^{cde}	1.6 ± 0.2^{hij}	1.6 ± 0.1^{defg}	1.6 ± 0.1^{efgh}	1.7 ± 0.1^{ghij}
Shilin	18.3 ± 1.1^{ab}	36.1 ± 0.6^{abcd}	9.1 ± 2.1^{ab}	28.2 ± 0.5^{ab}	19.1 ± 0.7^{ab}	13.5 ± 1.0^{a}	1895.0 ± 506.5^{kl}	648.0 ± 237.9^{ab}	210.5 ± 102.1^{hi}	1.3 ± 0.2^{ijk}	1.1 ± 0.1^{fghi}	1.2 ± 0.1^{hij}	1.1 ± 0.1^{kl}
Shipai	18.5 ± 1.3^{a}	37.0 ± 0.8^{a}	8.7 ± 2.2^{abc}	28.3 ± 0.5^{a}	19.2 ± 0.7^{a}	13.4 ± 1.1^{ab}	1908.2 ± 553.1^{kl}	655.1 ± 260.8^{ab}	197.3 ± 107.1^{i}	0.4 ± 0.1^{l}	0.4 ± 0.1^{jk}	0.4 ± 0.1^{lm}	0.4 ± 0.1^{m}
Shuangxi	16.3 ± 1.1^{gh}	36.3 ± 1.2^{bcd}	6.1 ± 1.9^{cde}	26.1 ± 0.5^{g}	17.1 ± 0.7^{efg}	13.1 ± 1.0^{abc}	3997.9 ± 762.0^{defg}	768.1 ± 233.4^{ab}	1286.2 ± 503.7^{bc}	0.4 ± 0.2^{jkl}	0.2 ± 0.1^{k}	0.2 ± 0.1^{m}	0.2 ± 0.1^{m}
Tianmu	18.2 ± 1.3^{abc}	35.6 ± 0.8^{bcde}	8.4 ± 2.2^{abc}	27.8 ± 0.7^{abc}	19.0 ± 0.8^{abc}	13.1 ± 1.1^{abc}	2126.8 ± 595.8^{jkl}	677.9 ± 260.1^{ab}	208.4 ± 101.7^{hi}	1.3 ± 0.3^{fg}	1.6 ± 0.3^{fghi}	1.1 ± 0.2^{hij}	1.6 ± 0.3^{hij}
Xinyi	17.5 ± 1.0^{abcde}	36.6 ± 0.8^{abc}	9.2 ± 2.6^{ab}	27.8 ± 0.5^{abcd}	18.5 ± 0.6^{abcd}	13.5 ± 1.1^{ab}	2396.0 ± 407.4^{jkl}	895.7 ± 279.6^{ab}	341.4 ± 136.2^{ghi}	0.7 ± 0.3^{jkl}	1.0 ± 0.4^{hijk}	0.6 ± 0.2^{jklm}	1.1 ± 0.3^{jkl}
Xizhi	17.4 ± 1.1^{bcdef}	36.0 ± 1.0^{bcdef}	8.8 ± 2.4^{abc}	27.6 ± 0.5^{bcd}	18.4 ± 0.5^{abcd}	13.2 ± 0.8^{abc}	3141.4 ± 400.4^{ghij}	855.3 ± 242.3^{ab}	702.6 ± 202.8^{defgh}	1.5 ± 0.3^{hi}	1.8 ± 0.2^{efgh}	1.3 ± 0.1^{fghi}	1.9 ± 0.1^{fghi}
Zhuzihu	13.4 ± 1.1^{k}	31.4 ± 0.6^{lm}	3.6 ± 1.9^{ef}	23.4 ± 0.4^{k}	14.3 ± 0.6^{ij}	12.9 ± 1.1^{abc}	4222.2 ± 800.9^{bcdef}	834.1 ± 314.7^{ab}	1046.9 ± 401.0^{cde}	1.2 ± 0.2^{ef}	2.9 ± 0.4^{ghi}	1.3 ± 0.2^{ghij}	2.8 ± 0.2^{e}

The value in each cell is mean ± standard deviation. Different letters within each cell indicate significant differences at 0.05 level among meteorological stations by Tukey-HSD test. Gandu recorded only precipitation and was not included in this table. For climate variables, Tann = mean annual temperature, Twrm = mean maximum temperature of the warmest month, Tcld = mean minimum temperature of coldest month, Tsmr = mean temperature in summer, Twnt = mean temperature in winter, Pann = annual total precipitation, Psmr = total precipitation in summer, Pwnt = total precipitation in winter.

with georeferenced occurrences data to predict potential distribution range and to determine climate environments of natural grassland. Occurrences of natural grassland along mountain ridges at low elevation is consistent with the presence of low elevation treeline in Taiwan. Many studies had previously proposed climate characteristics of treeline appeared at alpine zone and alpine treeline is commonly characterized by harsh environments (Germino *et al*., 2002; Hoch and Korner, 2003; Korner, 1998; Liu *et al*., 2011; Mohapatra *et al*., 2019; Stevens and Fox, 1991). A global comparison of alpine treeline positions in humid regions revealed very similar mean growing season temperatures at the treeline between 5 and 7 °C, irrespective of the geographical latitude (Hoch and Korner, 2003; Korner, 1998). From the results of model evaluation in this study, temperature variation is the most important determinant of low elevation natural grassland in Taiwan, particularly based on pseudo-absence data (Table 1). Based on the climate characteristics of the four meteorological stations near natural grassland (Table 3), Tcld reaches 4 °C is lower than the mean growing season temperatures at the alpine treeline. That is, lowest temperature in coldest month is the probable factor limiting the growth of evergreen broadleaved trees near mountain ridge at subtropical humid mountainous areas. On the other hand, the results of PCA had indicated that heavy precipitation and strong wind speed were the probable factors related to the occurrences of natural grassland at mountain ridge of low elevations in NTWN. The major climate factor related to the occurrence of natural grassland is not consistent between RF model and PCA. Thus, occurrences of natural grassland as well as limits of treeline at low elevation was presumably determined by multiple factors, such as relatively low temperature, heavy precipitations, and strong winds. Further studies will be necessary for identifying limiting factors of evergreen tree species at mountain ridge of low elevations in humid subtropical mountainous areas.

Climate environment in NTWN was highly affected by monsoon winds (Chen and Tsai, 1983). The meteorological stations at the coastal range of the study area (Table 1), including Daping, Fuguijiao, Fulong, Ruifang, Sanhe, Sanzhi, Shuangxi, had received high winter precipitation that was accompanied by winter monsoon wind. Winter precipitation from winter monsoon wind have resulted in high annual total precipitation of the coastal range in NTWN. On the contrary, the meteorological stations at inland area of NTWN, including Dazhi, Nangang, Neihu, Shipai, Shilin, Tianmu, Xinyi, possessed higher summer precipitation (Table 1) that is accompanied by summer monsoon wind. The differentiation of temperature and precipitation between coastal and inland areas as well as between windward and leeward slopes derived from monsoon wind had affected on the model performances in the study area. It warrants more attentions in future studies of modeling plant distributions.

Gridded climate dataset with 50 × 50 spatial resolution developed in this study was available to precisely project potential distribution range and to quantify climate space of natural grassland at landscape scale in subtropical humid mountainous areas. The local climate dataset interpolated from daily data of meteorological stations and followed by altitudinal adjustment to generate gridded climate dataset was available to model potential distribution range of natural grassland at landscape scale. The local climate dataset had evidently and effectively reflected habitat heterogeneity between coastal and inland areas of NTWN. Gridded climate dataset used in this study was suggested to apply for modeling potential distribution range of rare or endangered plant species in NTWN.

SDMs correlated high-resolution climate data and georeferenced occurrence data to predict potential distribution range of species and characterized climatic dimensions of a species' niche (Evans *et al*., 2009; Peterson *et al*., 2011). In this study, natural grassland was restricted along mountain ridges in NTWN and predominated by two species, *Miscanthus sinensis* and *Pseudosasa usawai* (Liao *et al*., 2014). Distribution range and climatic environments were indiscriminate between the two species. Niche convergence among phylogenetic distantly related species has played a primary role in driving community assembly in local vegetation along altitudinal gradient (Pearse and Hipp, 2012; Qian, 2017). These two species were assumed to have convergence of climatic niche along environmental gradient in subtropical mountainous areas, and further studies will be necessary to conduct on this topic.

The effect of true absence data on modeling species distribution range was distinct from that of pseudo-absence data. The potential range projected by RF based on true absence data had accurately reflected real geographical range of natural grassland. Distribution range of natural grassland is easily to verify in the field because of distinct boundaries between natural grassland and evergreen broadleaved forests. True absence data was geographically close to the boundaries of natural grassland, and the potential range projected by RF based on pseudo-absence was geographically wider than the true absence data (Fig. 2D, 2E and 2F). Thus, RF prediction based on pseudo-absence was an inaccurate model performance in this study, even though the AUC scores were higher. AUC was frequently used to detect model performance (Chapman *et al*., 2019; Lannuzel *et al*., 2021; Lobo *et al*., 2008; Tomlinson *et al*., 2020; Xu *et al*., 2021; Zhu *et al*., 2018). However, AUC does not provide sufficient information about the model errors (Lobo *et al*., 2008). Pseudo-absence data more environmentally distant from the presence data lead to higher AUC scores but not guarantee an accurate

distribution map projected by SDMs.

The magnitude of over- and under-prediction of species distribution range would greatly affect management strategies for conservation of species (Early and Sax, 2014), particularly in mountainous areas with isolated and fragmented suitable habitats. In this study, model prediction based on pseudo-absence data had evidently over-predicted species potential distribution range, since RF had projected a distribution map with wider potential range along mountain ridge based on pseudo-absence data (Fig. 2). The risk of true absence data may have resulted in under-prediction of potential range when true absence data were geographically close to the boundaries of the natural distribution range. Accurate contemporary distribution range of rare or endangered species is particularly important for their conservation in mountainous areas. However, it is challenging to project an accurate contemporary distribution range of plant species based on true absence data. Thus, an accurate contemporary distribution map of rare or endangered species was suggested to complement potential distribution ranges projected by SDMs based on the two types of absence data. Although comprehensive collection of true absence data is a difficult task in mountainous areas, collections of true absence data along a known environmental gradient, for example altitudinal gradient, is a costly but easier task. Model algorithms will perform better when the models were calibrated by some true absence data and some pseudo-absence data. Over-prediction of pseudo-absence data and under-prediction of true absence data could have a complemented result in the model predictions. A more accurate map of species distribution range will be generated, and that will be better for planning conservation management in mountainous areas with complex topography and mixed natural and artificial ecosystems.

ACKNOWLEDGMENTS

The author deeply appreciates assistant researcher Lin, Huan-Yu in Taiwan Forestry Research Institute for his helping and sharing technique of climate data interpolation. The authors appreciate Mr. Kai-Jie Yang in the Institution of Geography, Chinese Culture University, Taipei, Taiwan for technical support of ArcGIS software.

LITERATURE CITED

Anacker, B. L., M. Gogol-Prokurat, K. Leidholm and S. Schoenig. 2013. Climate change vulnerability assessment of rare plants in California. Madroño **60(3):** 193–210.

Anderson-Teixeira, K.J., S.J. Davies, A.C. Bennett, E. B. Gonzalez-Akre, H.C. Muller-Landau, S.J. Wright, K.A. Salim, A.M.A. Zambrano, A. Alonso, J.L. Baltzer, Y. Basset, N.A. Bourg, E.N. Broadbent, W.Y. Brockelman, S. Bunyavejchewin, D.F.R.P. Burslem, N. Butt, M. Cao, D. Cardenas, G.B. Chuyong, K. Clay, S. Cordell, H.S. Dattaraja, X. Deng, M. Detto, X. Du, A. Duque, D.L. Erikson, C.E.N. Ewango, G.A. Fischer, C. Fletcher, R.B. Foster, C.P. Giardina, G.S. Gilbert, N. Gunatilleke, S. Gunatilleke, Z. Hao, W.W. Hargrove, T.B. Hart, B.C.H. Hau, F. He, F.M. Hoffman, R.W. Howe, S.P. Hubbell, F.M. Inman-Narahari, P.A. Jansen, M. Jiang, D.J. Johnson, M. Kanzaki, A.R. Kassim, D. Kenfack, S. Kibet, M.F. Kinnaird, L. Korte, K. Kral, J. Kumar, A.J. Larson, Y. Li, X. Li, S. Liu, S.K.Y. Lum, J.A. Lutz, K. Ma, D.M. Maddalena, J.-R. Makana, Y. Malhi, T. Marthews, R.M.Serudin, S.M. McMahon, W.J. McShea, H.R. Memiaghe, X. Mi, T. Mizuno, M. Morecroft, J.A. Myers, V. Novotny, A.A. de Oliveira, P.S. Ong, D.A. Orwig, R. Ostertag, J. den Ouden, G.G. Parker, R.P. Phillips, L. Sack, M.N. Sainge, W. Sang, K. Sri-Ngernyuang, R. Sukumar, I-F. Sun, W. Sungpalee, H.S. Suresh, S. Tan, S.C. Thomas, D.W. Thomas, J. Thompson, B.L. Turner, M. Uriarte, R. Valencia, M.I. Vallejo, A. Vicentini, T. Vrška, X. Wang, X. Wang, G. Weiblen, A. Wolf, H. Xu, S. Yap, J. Zimmerman 2015. CTFS-forestGEO: A worldwide network monitoring forests in an era of global change. Glob. Chang. Biol. **21(2):** 528–549.

Arroyo-Rodríguez, V., F.P. Melo, M. Martínez-Ramos, F. Bongers, R.L. Chazdon, J.A. Meave, S.J. Wright, N. Norden, B.A. Santos, I.R. Leal, M. Tabarelli. 2015. Multiple successional pathways in human-modified tropical landscapes: New insights from forest succession, forest fragmentation and landscape ecology research. Biol. Rev. **92(1):** 326–340.

Barbet-Massin, M., F. Jiguet, C. H. Albert and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: How, where and how many? Methods Ecol. Evol. **3(2):** 327–338.

Booth, T. H., H. A. Nix, J. R. Busby and M. F. Hutchinson. 2014. BIOCLIM: The first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions **20(1):** 1–9.

Breiman, L. 2001. Random forests. Mach. Learn. **45(1):** 5–32.

Brunialti, G. and L. Frati. 2021. Modeling of species distribution and biodiversity in forests. Forests **12(3):** 319.

Chambers, J. and T. Hastie. 1992. Linear Models. Chapter 4 of statistical models in S. Wadsworth & Brooks/Cole

Chapman, D., O. L. Pescott, H. E. Roy and R. Tanner. 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. J. Biogeogr. **46(5):** 1029–1040.

Chen, W. K. and C. Y. Tsai. 1983. The climate of Yangmingshan National Park. Yangmingshan National Park, Construction and Planning Agency Ministry of the Interior, Executive Yuan, Taipei, Taiwan

Dobrowski, S. Z. 2011. A climatic basis for microrefugia: The influence of terrain on climate. Glob. Chang. Biol. **17(2):** 1022–1035.

Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J. P. Theurillat and A. Guisan. 2011. Predicting spatial patterns of plant species richness: A comparison of direct macroecological and species stacking modelling approaches. Divers. Distrib. **17(6):** 1122–1131.

Dupin, M., P. Reynaud, V. Jarošík, R. Baker, S. Brunel, D. Eyre, J. Pergl, D. Makowski, S. Thrush. 2011. Effects of the training dataset characteristics on the performance of nine species distribution models: Application to *Diabrotica virgifera virgifera*. Plos One **6(6):** e20957

Early, R. and D. F. Sax. 2014. Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. Glob. Ecol. Biogeogr. **23(12):** 1356–1365.

Elith, J. and J. R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. Ann. Rev. Ecol. Evol. Syst. **40(1):** 677–697.

Evans, M. E., S. A. Smith, R. S. Flynn and M. J. Donoghue. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, Sections Anogra and Kleinia). Am. Nat. **173(2):** 225–240.

Fick, S. E. and R. J. Hijmans. 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. **37(12):** 4302–4315.

Fois, M., G. Fenu, A. C. Lombrana, D. Cogoni and G. Bacchetta. 2015. A practical method to speed up the discovery of unknown populations using species distribution models. J. Nat. Conserv. **24:** 42–48.

Germino, M. J., W. K. Smith and A. C. Resor. 2002. Conifer seedling distribution and survival in an alpine-treeline ecotone. Plant Ecol. **162(2):** 157–168.

Gies, M., M. Sondermann, D. Hering and C. K. Feld. 2015. Are species distribution models based on broad-scale environmental variables transferable across adjacent watersheds? A case study with eleven macroinvertebrate species. Fundamental and Applied Limnology/Archiv für Hydrobiologie **186(1-2):** 63–97.

Guillera-Arroita, G., J.J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P.E. Lentini, M.A. McCarthy, R. Tingley, B.A. Wintle. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. **24(3):** 276–292.

Hegel, T.M., S.A. Cushman, J. Evans and F. Huettmann. 2010. Current State of the Art for Statistical Modelling of Species Distributions. In: **Cushman, S.A. and F. Huettmann** (eds). Spatial complexity, informatics, and wildlife conservation, 273–311 pp. Springer, Tokyo.

Hoch, G. and C. Korner. 2003. The carbon charging of pines at the climatic treeline: A global comparison. Oecologia **135(1):** 10–21.

Hsieh, C. F., W. C. Chao, C. C. Liao, K. C. Yang and T. H. Hsieh. 1997. Floristic composition of the evergreen broad-leaved forests of Taiwan. Nat. Hist. Res. **4:** 1–16.

Kier, G., H. Kreft, T. M. Lee, W. Jetz, P. L. Ibisch, C. Nowicki, J. Mutke, W. Barthlott. 2009. A global assessment of endemism and species richness across island and mainland regions. PNAS **106(23):** 9322–9327.

Korner, C. 1998. A re-assessment of high elevation treeline positions and their explanation. Oecologia **115(4):** 445–459.

Lannuzel, G., J. Balmot, N. Dubos, M. Thibault and B. Fogliani. 2021. High-resolution topographic variables accurately predict the distribution of rare plant species for conservation area selection in a narrow-endemism hotspot in New Caledonia. Biodivers. Conserv. **30(4)**: 963–990

Li, C.F., M. Chytrý, D. Zelený, M.Y. Chen, T.Y. Chen, C.R. Chiou, Y.-J. Hsia, H.-Y. Liu, S.-Z. Yang, C.-L. Yeh, J.-C. Wang, C.-F. Yu, Y.-J. Lai, W.-C. Chao, C.-F. Hsieh, H. Bruelheide 2013. Classification of Taiwan forest vegetation. Appl. Veg. Sci. **16(4):** 698–719.

Liang, W., M. Papeş, L. Tran, J. Grant, R. Washington-Allen, S. Stewart and G. Wiggins. 2018. The effect of pseudo-absence selection method on transferability of species distribution models in the context of non-adaptive niche shift. Ecol Model **388:** 1–9.

Liao, C.C., S.C. Kuo and C.R. Chang. 2012. Forest distribution on small isolated hills and implications on woody plant distribution under threats of global warming. Taiwania **57(3):** 242–250.

Liao, C.C., C.R. Chang, M.T. Hsu and W.K. Poo. 2014. Experimental evaluation of the sustainability of dwarf bamboo (*Pseudosasa usawai*) sprout-harvesting practices in Yangminshan National Park, Taiwan. Environ. Manage. **54(2):** 320–330.

Liao, C. C. and Y. H. Chen. 2021. Improving performance of species distribution model in mountainous areas with complex topography. Ecol. Res. **36(4):** 648–662.

Liaw, A. and M. Wiener. 2002. Classification and regression by random forest. R news **2:** 18–22.

Liu, B., E. Liang and L. Zhu. 2011. Microclimatic conditions for *Juniperus saltuaria* treeline in the Sygera Mountain, Southeastern Tibetan plateau. Mt. Res. Dev. **31(1):** 45–53.

Lobo, J. M., A. Jiménez-Valverde and R. Real. 2008. AUC: A misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. **17(2):** 145–151.

Mohapatra, J., C. P. Singh, M. Hamid, A. Verma, S. C. Semwal, B. Gajmer, A.A. Khuroo, A. Kumar, M.C. Nautiyal, N. Sharma, H.A. Pandya. 2019. Modelling *Betula utilis* distribution in response to climate-warming scenarios in Hindu-Kush Himalaya using random forest. Biodivers. Conserv. **28(8-9):** 2295–2317.

Pearse, I. S. and A. L. Hipp. 2012. Global patterns of leaf defenses in oak species. Evolution **66(7):** 2272–2286.

Peng, D., L. Sun, H. W. Pritchard, J. Yang, H. Sun and Z. Li. 2019. Species distribution modelling and seed germination of four threatened snow lotus (*Saussurea*), and their implication for conservation. Glob. Ecol. Biogeogr. **17:** e00565.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura and M. B. Araújo. 2011. Ecological Niches and Geographic Distributions (MPB-49). Princeton University Press.

Porfirio, L.L., R.M. Harris, E.C. Lefroy, S. Hugh, S.F. Gould, G. Lee, N.L. Bindoff, B. Mackey, L. Kumar. 2014. Improving the use of species distribution models in conservation planning and management under climate change. Plos One **9(11):** e113749.

Pradervand, J.-N., A. Dubuis, L. Pellissier, A. Guisan and C. Randin. 2014. Very high resolution environmental predictors in species distribution models: Moving beyond topography? Prog. Phys. Geog. **38(1):** 79–96.

Qian, H. 2017. Climatic correlates of phylogenetic relatedness of woody angiosperms in forest communities along a tropical elevational gradient in South America. J. Plant Ecol. **11(3):** 394–400.

Qiao, H., X. Feng, L. E. Escobar, A. T. Peterson, J. Soberón, G. Zhu and M. Papeş. 2019. An evaluation of transferability of ecological niche models. Ecography **42(3):** 521–534.

Senay, S. D., S. P. Worner and T. Ikeda. 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. Plos One **8(8):** e71218.

Smith, W.K., M.J. Germino, D.M. Johnson and K. Reinhardt. 2009. The altitude of alpine treeline: A bellwether of climate change effects. Bot Rev **75(2):** 163–190.

Stevens, G. C. and J. F. Fox. 1991. The causes of treeline. Annu Rev Ecol Syst **22(1):** 177–191.

Thuiller, W., D. Georges, R. Engler, F. Breiner, M. D. Georges and C. W. Thuiller. 2016. Package 'biomod2'. Species distribution modeling within an ensemble forecasting framework.

Title, P. O. and J. B. Bemmels. 2018. ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography **41(2):** 291–307.

Tomlinson, S., W. Lewandrowski, C. P. Elliott, B. P. Miller and S. R. Turner. 2020. High-resolution distribution modeling of a threatened short-range endemic plant informed by edaphic factors. Ecol. Evol. **10(2):** 763–777.

Tsoar, A., O. Allouche, O. Steinitz, D. Rotem and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. Divers. Distrib. **13(4):** 397–405.

Wisz, M. S. and A. Guisan. 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecology **9(1):** 1–13.

Xu, Y., Y. Huang, H. Zhao, M. Yang, Y. Zhuang and X. Ye. 2021. Modelling the effects of climate change on the distribution of endangered *Cypripedium japonicum* in china. Forests **12(4):** 429.

Zhu, Y., W. Wei, H. Li, B. Wang, X. Yang and Y. Liu. 2018. Modelling the potential distribution and shifts of three varieties of *Stipa tianschanica* in the eastern Eurasian steppe under multiple climate change scenarios. Glob. Ecol. Biogeogr. **16:** e00501.

Supplementary materials are available from Journal Website.