Advantage and Disadvantage of Global and Local Climate Datasets on Modeling Species Distribution at Continental and Landscape Scales

Chi-Cheng Liao (✉ hunter_yyl@yahoo.com.tw)

Chinese Culture University <https://orcid.org/0000-0002-6409-0506>

Chi-Ru CHANG

Chinese Culture University

Yi-Huey CHEN

Chinese Culture University

Research Article

Keywords: *Bretschneidera sinensis*, Random Forest, Principle components analysis, Taiwan, WorldClim

Posted Date: June 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-595860/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Species distribution model based on global and local climate datasets were hypothesized to have advantages on projecting distribution range at continental and landscape scales, respectively. Random Forest (RF) and principle components analysis (PCA) aimed to project potential distribution range and to construct climate space of *Bretschneidera sinensis* in continental East Asia (CEA) and northern Taiwan (NTWN) based on the WorldClim and local climate datasets. Geographical extent of the endangered species at continental scale was available to be projected by RF based on the WorldClim dataset, whereas RF had projected bias map that presented gridded squares at edges of the potential distribution range. At landscape scale, projection map of RF in NTWN based on the WorldClim dataset presented gridded distribution far from empirical distribution pattern, while that based on local climate dataset presented a distribution pattern relevant to elevation and topography. PCA had revealed climate differentiation between continental and island populations. Evidently, local climate dataset had reflected climate heterogeneity at landscape scale and is essential for identifying local adaptation of island population at geographical margin of the endangered species. However, huge number of gridded cells generated from local climate interpolation method for projecting potential distribution range at landscape scale is not available to expand geographical extent to continental region. Global climate dataset has the advantage on modeling geographical extent of plant species at continental scale, while local climate dataset used for modelling species distribution enables conservationists to delineate reliable conservation areas in fragmented natural habitats at landscape scale.

Introduction

Species distribution models (SDMs) have been widely used in ecological and evolutionary studies and employed as tools for exploring contemporary species distribution ranges (Duque-Lazo et al. 2016; Williams et al. 2009) and conservation applications such as identifying suitable habitats for rare species (Fois et al. 2015; Marmion et al. 2009). Performance of SDMs have heavily relied on the accuracy of georeferenced species occurrences data and high-resolution GIS-based environmental variables that provide discriminatory power regarding presence and absence of species to project maps indicating areas of species distributions (Datta et al. 2020; Maria & Udo 2017; Title & Bemmels 2018). One of the most common sources of climate data is the WorldClim dataset (Fick & Hijmans 2017) that had been widely used as bioclimatic predictors and applied to SDM studies for evaluating geographical extent and ecological characteristics of species (Abdelaal et al. 2019; Datta et al. 2020; Fernández & Hamilton 2015; Lobo et al. 2010; Maria & Udo 2017; Mohapatra et al. 2019). The WorldClim dataset has the advantage on projecting geographical range of plant species at continental region, while the dataset with spatial resolution at 30 seconds (approximately 1 km) is not finer enough to accurately represent local climate characteristics (Bedia et al. 2013).

Some previous studies had downscaled the WorldClim dataset to finer resolutions for local assessment (Lin et al. 2018; Schorr et al. 2012; Wang et al. 2016). Downscaled climate dataset was usually applied to evaluate habitat differentiation and species distribution range at scales of tens to hundreds of kilometers

in continental regions (Godsoe et al. 2015; Maria & Udo 2017; Mohapatra et al. 2019; Zhu et al. 2018). However, downscaled climate dataset from the WorldClim was presumably not finer enough to reflect local climatic heterogeneity in mountainous areas, since local climatic heterogeneity was drastically induced by altitudinal and topographical gradients (Chao et al. 2010; Chen et al. 1997; Ma et al. 2010; McKenzie et al. 2003; Tu et al. 2010). Particularly, distributions of rare species with locally restrict distribution are often critically affected by certain local elements, such as the occurrences of favorable microsites and microclimates (Elith & Leathwick 2009; Heikkinen et al. 2012). Such local special ecological requirements can be difficult to distinguish in the broader-scale climate dataset (Guisan et al. 2007). Alternatively, few SDM studies had used high-resolution environmental data to predict species distribution range in mountainous areas (Lannuzel et al. 2021; Raes et al. 2009; Tomlinson et al. 2020), whereas these models used topographical elements that was not directly related to ecological processes constrained species distribution (Tomlinson et al. 2020). Therefore, uncertainties still exist in the correlation of species distribution with climate factors when the downscaled climate data from the WorldClim dataset or topographical elements had been applied to model species distribution at fine scale (Lannuzel et al. 2021; Tomlinson et al. 2020). In order to model species distribution range and correlate species distribution with climate factors at landscape scale, a previous study had proposed an interpolation method to generate gridded climate dataset with high spatial resolution of $50 \times 50 \text{ m}^2$ from daily data of local meteorological stations (Liao & Chen 2021). The previously proposed method was presumably appropriate for modeling distribution of plant species in relation to climatic heterogeneity induce by elevation and topography and to address climate factors driving species distribution in mountainous area. However, high-resolution local climate dataset generated huge number of gridded cells within a local geographical extent and, thus, it is questionable that the method is available to expand the geographical extent of gridded climate dataset to continental region.

Species distribution models (SDMs) can provide useful predictions for insufficiently surveyed areas and provide guidelines for conservation planning tasks, such as identifying suitable habitats and seeking new populations for rare species (Fois et al. 2015; Mi et al. 2017). In the last century, economic development, urbanization, and population growth led to a highly modification of natural ecosystems, while habitat of the plant species were fragmented and currently formed by complex mosaics of rural or urban areas, abandoned grasslands, agricultural landscapes, and secondary forests (Dong et al. 2019; Kier et al. 2009). Thus, accurate prediction of current distribution range for the designation of effective protected areas is necessary for the conservation of rare or endangered plant species. Random Forest (RF), one of the SDMs, is a powerful machine learning classifier to provide accurate predictions of species distributions (Boulesteix et al. 2012; Breiman 2001; Liaw & Wiener 2002). RF has become a popular method to identify geographical extent and climate factors of rare or endangered species (Evans & Cushman 2009; Hu et al. 2017; Iturbide et al. 2018; Mi et al. 2017; Mohapatra et al. 2019; Williams et al. 2009). On the other hand, principle components analysis (PCA) was usually used for quantifying climate space and to assess ecological preferences of plant species (Early & Sax 2014; Edwards & Still 2008).

In this study, the WorldClim and local climate datasets were applied for RF and PCA to project potential distribution range and to construct climate space of the endangered plant species, *Bretschneidera sinensis* Hemsl. The endangered plant species is majorly distributed in Southern China and marginal populations of the species extend to Indo China, India and Taiwan (Dong et al. 2019; Hu et al. 2014; Kumar et al. 2017; Li et al. 2016; Wang et al. 2018). Populations of the species are quite small and have suffered from isolation and fragmentation of habitats (Dong et al. 2019). Distribution patterns of the endangered plant species in continental region and local area were projected by RF based on the WorldClim and local climate datasets to evaluate geographical extent and locally restricted distribution range and were expected to provide valuable information for effective conservation.

The objectives of this study were to identify the advantage and disadvantage of the global and local climate datasets at continental and landscape scales. Two hypotheses were assessed in this study. (1) The WorldClim dataset was hypothesized to have advantage on projecting geographical distribution range and quantifying climate space of plant species at continental region, whereas it was presumably not appropriate for projecting species distribution at landscape scale. (2) Local climate dataset was presumably available to capture climatic heterogeneity induced by elevation and topography in mountainous areas and has the advantage on projecting species distribution range and quantifying climate space at landscape scale. Model prediction based on local climate dataset was further expected to capture fine-scale ecological characteristics and to provide valuable geographical information for developing effective conservation management of the endangered species.

Material And Methods

Plant species and collections

Bretschneidera sinensis Hemsl., a deciduous tree species of monotypic genera belongs to the family Akaniaceae, is a relic species of the Tertiary tropical flora mainly occurring in evergreen broad-leaved or mixed evergreen and deciduous forests in mountainous areas at altitudes of 300-1,700 m above sea level in Southern China (Dong et al. 2019; Li et al. 2016; Wang et al. 2018). The species is listed as endangered on the IUCN Red List of Threatened Species (IUCN 2020, <http://www.iucnredlist.org>). According to recent studies, most of the populations are seriously threatened (Hu et al. 2014) and the species also suffers from reproductive failure due to low seed productivity and seed germination rates (Qiao et al. 2012). Geographical range of the endangered species extends from Southern China to India, Indo China, Vietnam, and Taiwan (Dong et al. 2019; Hu et al. 2014; Kumar et al. 2017; Li et al. 2016; Wang et al. 2018). Taiwan is a continental island possessed isolated population of *B. sinensis* at geographical margin of the Eurasia continent. In Taiwan, *B. sinensis* is a narrow ranges species usually scattered in remnant or secondary broadleaved forests only in northern part of the island (Li et al. 2013). Genetic evidence indicated that population of *B. sinensis* from mountainous areas in Taiwan was probably derived from the southeastern Yunnan through postglacial range expansion to its present range (Wang et al. 2018).

Georeferenced occurrences of the *B. sinensis* in continental East Asia (CEA) was downloaded from the Global Biodiversity Information Facility (GBIF). A total of 158 georeferenced data records in CEA were obtained from the GBIF (Fig. 1). Presence data of *B. sinensis* in northern Taiwan (NTWN) was downloaded from the herbarium of the Taiwan Forestry Research Institute, herbarium of National Taiwan University, and herbarium of the Academia Sinica, Taipei. A total of 72 data records were collected for NTWN (Fig. 1).

Climate datasets and variables selection

The WorldClim dataset (available at <https://www.worldclim.org/>) with 30-sec spatial resolution (approximately 1 km in equator) is one of the most commonly used climate dataset in modeling species distribution (Fick & Hijmans 2017) and the dataset was applied to predict current potential distribution range of *B. sinensis* in CEA. The CEA had been divided into gridded cells with spatial resolution at 30 minutes (approximately 50 km) to capture climate variables from the WorldClim dataset. A total of 1.5 million gridded cells was generated in CEA including Mongolia, China, India and Indo China and was applied for model prediction. Projection map in CEA based on the WorldClim dataset were presented to show the distribution pattern of the *B. sinensis* at continental scale. In this study, 13 climate variables, 9 temperature and 4 precipitation variables, were adopted for constructing the RF model (Table 1).

Table 1
Bioclimatic predictors used for modeling potential distribution range of *Bretschneidera sinensis*Hemsl.

Abbreviation	Climatic variable
Temperature	
BI01	annual mean temperature
BI02	mean diurnal temperature range [mean of monthly (maximum temperature - minimum temperature)]
BI03	isothermality ($BI02 / BI07 \cdot 100$)
BI04	temperature seasonality (standard deviation of monthly temperature)
BI05	maximum temperature of the warmest month
BI06	minimum temperature of the coldest month
BI07	temperature range ($BI06 - BI05$)
Bio10	mean temperature of summer
BI011	mean temperature of winter
Precipitation	
BI012	annual precipitation
BI015	precipitation seasonality (standard deviation of monthly precipitation)
BI018	precipitation of the summer
BI019	precipitation of the winter

In order to predict the current distribution range of *B. sinensis* at landscape scale, NTWN was divided into gridded cells with spatial resolution of $50 \times 50 \text{ m}^2$ and more than 0.4 million gridded cells were generated in NTWN. 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 built by Department of Geology, Chinese Culture University. Climate variables of the gridded cells in NTWN were independently extracted from the WorldClim and local climate datasets performed by ArcInfo software (ESRI, Redlands, California, USA). The WorldClim dataset was downloaded from the website, while local climate dataset was constructed in this study based on the data of local meteorological stations in NTWN. Local climate dataset was generated from daily data of 30 meteorological stations downloaded from the website of the Central Weather Bureau, Ministry of Transportation and Communications, Taiwan (<https://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp>). The daily data downloaded from local meteorological stations were calculated and interpolated by Inverse Distance Weighted (IDW) method performed by ArcInfo software (ESRI, Redlands, California, USA) and were subsequently corrected by empirical elevation laps rates varied among slopes of the five areas in NTWN (Fig. 1). In order to make the models based on the two climate datasets comparable, the same climate variables independently selected from the two sources of climate datasets were applied to the RF model.

Random Forest algorithm

RF algorithm implemented by “randomForest” package within the R environment (Breiman 2001; Liaw & Wiener 2002) was applied to project geographical extent of the *B. sinensis* at continental scale and to predict local distribution pattern relevant to topography and elevation. A total of 158 and 72 georeferenced data records were available for modeling plant distribution in CEA and NTWN, respectively. The presence dataset of *B. sinensis* were spatially verified to eliminate duplicate records. The coordination of presence data records was adopted to extract climate variables from the WorldClim and local climate datasets. The presence data with identical climate variables derived from the two climate datasets were imported to the RF as two training datasets. False absence points (or background points) were randomly selected within the study range as the same number of presence data. False absence points were combined with the presence data to construct the training dataset of the model and subsequently modelled 100 times with the resampled background points to quantify uncertainties in prediction. The two testing datasets were the gridded cells with climate variables derived from the WorldClim and local climate datasets. RF algorithm categorized gridded cells into presence and absence and gridded cells with 50 times of presences categorized by RF were utilized to project potential distribution range of the endangered plant species. The distribution range projected by RF based on the WorldClim and local climate datasets in NTWN were verified in field by expert knowledge to determine the accuracy of the projection maps.

Quantification of climate spaces by principle components analysis (PCA)

Further analysis aims to make a comparison between climate spaces of the *B. sinensis* populations in CEA and NTWN constructed by the WorldClim and local climate datasets. Principle components analysis (PCA) was performed to quantify climate spaces of the endangered plant species and its geographical range to explicitly understand the ecological preferences and climate environments represented by the WorldClim and local climate datasets. The RF had classified the gridded cells with climate variables into presence and absence cells in CEA and NTWN. For the gridded cells in CEA, 0.15 million cells with presence and absence data of *B. sinensis* projected by RF were randomly selected to construct two of the datasets for quantification of climate spaces. The other four datasets for quantification of climate spaces were the presence and absence data of *B. sinensis* projected by RF in NTWN based on the WorldClim and local climate datasets. PCA was applied to scaled data for 13 studied climate variables corresponding to the formation of climate spaces of *B. sinensis*. The 13 climate variables were thought to provide climate preferences for the plants’ distributions and the ecological demands were distilled into three principal components, the first, second, and third axis from a PCA.

Results

Potential distribution range of B. sinensis

Distribution maps of *B. sinensis* in CEA were projected by RF based on the construction of presence and absence data extracted climate variables from the WorldClim dataset (Fig. 2a). The AUC (area under operating characteristic curves) value of the model was 0.995. Distribution map in CEA projected by RF presented suitable habitats of the plant species in Southern China. The model had effectively presented potential distribution range of the plant species at continental scale. However, projection map based on the WorldClim dataset presented no isolation and fragmentation of potential habitats at Southern China that is contradict to the empirical distribution pattern of the endangered species. Furthermore, projection map presented a pattern of gridded squares at the edge of the potential distribution range in a larger scale map (Fig. 2b and 2c). Thus, projection map based on the WorldClim dataset may not reflect the empirical distribution pattern of the plant species and there were uncertainties of the distribution pattern based on the WorldClim dataset at continental scale.

Potential distribution range of the plant species were further examined in NTWN and the map projected by RF based on the WorldClim and local climate datasets had presented completely different distribution patterns at landscape scale (Fig. 3a and 3b). Distribution map of *B. sinensis* based on the WorldClim dataset was irrelevant to the topography and elevation, since that had presented gridded distribution pattern around the presence data records (Fig. 3a). Gridded distribution pattern in mountainous areas of NTWN projected by RF model was similar to the edge of the potential distribution range of the species in CEA and it was an unrealistic distribution pattern. In contrast, distribution map of *B. sinensis* based on local climate dataset was pertinent to topography and elevation (Fig. 3b), that is more likely to present empirical distribution pattern at landscape scale.

The error rates (ERR_{00B}) of the two projection results in NTWN presented no significant differences by statistical test (0.9177 for the WorldClim and 0.9103 for local climate datasets). Moreover, the climate variables contributed most to the model performances were different between the two climate datasets (Fig. 4). Temperature (Bio2) derived from the WorldClim dataset was the most important factor affecting model performance, while water availability (Bio19) contributed most to the model performance based on the local climate dataset.

Quantification of climate spaces by PCA

PCA were conducted to identify climate characteristics of the CEA and NTWN and to depict climate space of the plant species. Principle component 1 (PC1) accounted for 97.28% of the variation, while principle component 2 (PC2) 2.34% of the variation (Fig. 5). PC1, PC2, and principle component 3 (PC3) were strongly related to Bio12, Bio19 and Bio18 variables, respectively (Table 2). Apparently, these three variable were all water availability that had evidently played as the most important role for the quantification of climate spaces of *B. sinensis*.

Table 2

The first three axes of the principle components analysis (PCA) on the correlation matrix of climate variables from the WorldClim and local climate datasets.

Abbreviation	Climatic variable
Temperature	
BIO1	annual mean temperature
BIO2	mean diurnal temperature range [mean of monthly (maximum temperature - minimum temperature)]
BIO3	isothermality ($BIO2 / BIO7 \cdot 100$)
BIO4	temperature seasonality (standard deviation of monthly temperature)
BIO5	maximum temperature of the warmest month
BIO6	minimum temperature of the coldest month
BIO7	temperature range ($BIO6 - BIO5$)
Bio10	mean temperature of summer
BIO11	mean temperature of winter
Precipitation	
BIO12	annual precipitation
BIO15	precipitation seasonality (standard deviation of monthly precipitation)
BIO18	precipitation of the summer
BIO19	precipitation of the winter

Climate environment in NTWN based on local climate dataset had wider range along PC1 than that based on the WorldClim dataset presented by PCA (Fig. 5). *B. sinensis* occupies three different climate spaces based on the WorldClim and local climate datasets represented by dark circles in the PCA diagram (Fig. 5). Particularly, the plant species occupies distinct climate spaces in NTWN based on the WorldClim and local climate datasets (Fig. 5). It is interesting that the same geographical coordination locations of the plant species presented distinct climate spaces based on the different climate datasets. It is necessary to verify the empirical distribution pattern of the plant species in the field based on the expert knowledge to identify which distribution pattern projected by RF and which climate space constructed by PCA were more close to empirical distribution pattern at landscape scale. Furthermore, statistical tests were performed to detect which variable was the most responsible for the different climate spaces in the PCA diagram. Surprisingly, results of analysis of variance (ANOVA) and Tukey's test presented that all the variables have significant differences among the three climate spaces of the plant species (Table 3).

Table 3

Statistical test by analysis of variance (ANOVA) and Tukey's test for comparisons of the climatic variables from three datasets, WorldClim in continental East Asia (CEA) and WorldClim in northern Taiwan (NTWN), and local climate dataset in NTWN.

	PC1	PC2	PC3
Bio1	0.0030	0.0089	-0.0214
Bio2	0.0000	-0.0119	0.0115
Bio3	0.0038	-0.0203	0.0144
Bio4	-0.0013	-0.0038	0.0073
Bio5	0.0013	-0.0011	-0.0125
Bio6	0.0050	0.0202	-0.0372
Bio7	-0.0037	-0.0212	0.0247
Bio10	0.0013	0.0052	-0.0150
Bio11	0.0050	0.0125	-0.0293
Bio12	0.9349	0.1645	-0.3075
Bio15	0.0023	-0.1203	0.1233
Bio18	0.1493	0.6133	0.7748
Bio19	0.3219	-0.7620	0.5346
	WorldClim in CEA	WorldClim in NTWN	Local climate dataset in NTWN
Bio1	5.34 ± 7.66 ^a	20.53 ± 0.96 ^b	21.41 ± 1.44 ^c
Bio2	12.36 ± 2.56 ^a	5.41 ± 0.32 ^b	15.01 ± 0.97 ^c
Bio3	30.55 ± 6.16 ^a	28.94 ± 0.88 ^b	55.21 ± 1.55 ^c
Bio4	10.88 ± 3.19 ^a	4.82 ± 0.16 ^b	4.81 ± 0.18 ^c
Bio5	24.92 ± 6.94 ^a	30.08 ± 1.19 ^b	34.34 ± 1.56 ^c
Bio6	-16.55 ± 10.97 ^a	11.40 ± 0.99 ^b	7.18 ± 1.81 ^c
Bio7	41.47 ± 9.76 ^a	18.68 ± 0.66 ^b	27.16 ± 1.14 ^c
Bio10	18.09 ± 7.06 ^a	26.22 ± 1.06 ^b	25.87 ± 1.47 ^c
Bio11	-8.66 ± 9.87 ^a	14.53 ± 0.92 ^b	17.03 ± 1.44 ^c

	PC1	PC2	PC3
Bio12	500.23 ± 456.55 ^a	3319.73 ± 175.59 ^b	3567.89 ± 88.263 ^c
Bio15	92.97 ± 22.70 ^a	24.56 ± 3.47 ^b	111.74 ± 33.05 ^c
Bio18	262.13 ± 194.90 ^a	813.02 ± 80.60 ^b	767.37 ± 87.38 ^c
Bio19	29.98 ± 48.93 ^a	678.55 ± 97.83 ^b	1041.72 ± 449.09 ^c

Climate characteristics were drastically different between CEA and NTWN based on the WorldClim and local climate datasets. Histogram had showed wide range of mean diurnal temperature range (Bio2) in CEA based on WorldClim (Fig. 6), since the CEA had broad latitude range. However, annual precipitation in NTWN based on local climate dataset had extraordinary broad range in contrast to that in CEA and NTWN based on the WorldClim dataset (Fig. 6). Wide range of NTWN's climate environment along the PC1 was caused by extraordinary high annual precipitation that was partly a consequence of high winter precipitation at coastal area and lower at inland area. Local climate dataset generated from daily data of meteorological stations was more likely to capture the variation of local climate characteristics and was available to reflect climate heterogeneity in NTWN at landscape scale that was not available to be exhibited by the WorldClim dataset.

Discussion

The WorldClim dataset provided global climate surface that had been widely used in many previous SDMs studies as bioclimatic predictors and had accurately projected species distribution range at continental regions (Abdelaal et al. 2019; Datta et al. 2020; Heikkinen et al. 2012; Mi et al. 2017; Mohapatra et al. 2019). In this study, projection of potential distribution range and quantification of climate space was available at continental scale and was evidently the advantage of the WorldClim dataset. Model projections aims to project geographical extent, identify suitable key sites, and find suitable environmental factors based on the presence/absence data and climate variables that can provide scientific information for developing conservation and management strategies at continental scale (Abdelaal et al. 2019; Hu et al. 2017; Lobo et al. 2010; Maria & Udo 2017; Mi et al. 2017; Mohapatra et al. 2019; Williams et al. 2009). Modeling efforts to asses potential distribution range had been made for suggesting conservation areas of plant species in China (Wan et al. 2017; Xu et al. 2021; Yu et al. 2017; Zhang et al. 2017). However, projection map based on the WorldClim would be the disadvantage at landscape scale. Restricted elevation ranges and isolation and fragmentation of suitable habitat were not presented by the model projection in CEA. In addition, projection map based on the WorldClim dataset presented gridded squares at the edge of the potential distribution range in CEA and that also presented gridded squares in NTWN. Gridded squares of the WroldClim dataset was evidently unavailable to reflect climate heterogeneity induced by elevation and topography and gridded distribution pattern is an unrealistic distribution pattern far from empirical distribution of the plant species. Model predictions

based on the global climate dataset had evidently generated bias projection map and provided misleading results of species geographical distribution at landscape scale.

To avoid misleading result caused by the WorldClim dataset, local climate dataset was suggested to apply as bioclimatic predictors of model prediction at landscape scale. The projection maps of the plant species in NTWN based on the WorldClim and local climate datasets were verified in the field based on expert knowledge and field examination had identified that projection map based on local climate dataset was more close to empirical distribution pattern of the plant species. Accordingly, model performance was much better when model was calibrated by local climate dataset at landscape scale. Bioclimatic predictors from local climate dataset had precisely reflected climate characteristics induced by elevation and topography (Fig. 7) and were suggested to use for model projection to guarantee the accuracy of SDMs performance in mountainous area. Model projections at landscape scale is the advantage of local climate dataset. However, local climate dataset generated huge number of gridded cells within a local geographical area and there were more than 0.4 million gridded cells in NTWN that is only 1,041 km². Because of huge number of gridded cells, it is very difficult to expand the geographical range of SDMs study based on local climate dataset. Thus, model prediction based on local climate dataset would be the disadvantage at continental scale.

On the other hand, previous studies had proposed that RF model with low error rate (ERR_{OOB}) were considered to have more accurate performance. Our study had confirmed that low error rate (ERR_{OOB}) can be achieved even when model did not project accurate distribution range and projection map was evidently far from empirical distribution pattern of the plant species. Low error rate (ERR_{OOB}) did not guarantee an accurate projection map of species distribution.

RF model performance might be affected by bias collections and might have over-fitted to local conditions. Bias collection of presence data commonly proposed by many previous studies (Datta et al. 2020; El-Gabbas & Dormann 2018; Ferro & Flick 2015; Lannuzel et al. 2021; Tomlinson et al. 2020), since comprehensive collection is a great challenge to be conducted in field survey. Bias collection may lead to bias projection of distribution. Bias projection maps in continental scale is unavailable to be corrected in the field, whereas that at landscape scale can be corrected in the field examination by expert knowledge. Field examination based on the projection map is helpful to compensate bias collection of presence data by correcting presence and absence areas of the projection results at landscape scale. On the other hand, over-fit may lead to loss of generality of the model and may have poor performance when this method is expanded to other regions or other species. However, over-fit may not be a deficit of this method, since a previous study had accurately predicted species distribution range at landscape scale based on the same method (Liao & Chen 2021). More studies were expected to identify the generality of model prediction of this method. Despite of, this method generated precise distribution map of the plant species in mountainous area at landscape scale that is very close to its empirical distribution range and is very useful for further applications.

Marginal population of continental species had showed striking genetic differences from central populations in neighboring continent (Sexton et al. 2009; Thompson et al. 2005; Wake et al. 2009; Wu et al. 2001), while this study had further identified ecological differentiation between marginal and central populations. *B. sinensis* had locally adapted to climate environments in NTWN. Water availability had evidently played as a major climate factor related to the potential distribution range of the *B. sinensis* in NTWN and the plant species was evidently, locally adapted to extraordinary high precipitation in NTWN. Local adaptation of marginal population to novel habitats at geographical margin is akin to niche evolution (Sexton et al. 2009). Climatic characteristics of marginal population distinguished from central populations may consequently contribute to the genetic variation. Thus, marginal population on continental islands warrant a high priority in biodiversity conservation (Kier et al. 2009; Weigelt & Kreft 2013), since genetic variation (Li et al. 2016; Wang et al. 2018; Wu et al. 2001) and ecological differentiation both existed between central and marginal populations.

The growing impacts of climate change on plant species calls a request to evaluate geographical extent where species with narrow distribution ranges exist or likely exist in order to enhance their conservation and restoration (Abdelaal et al. 2019). However, designation and effective management of conservation areas is a great challenge on islands because anthropogenic disturbances caused fragmentation and isolation of natural habitats and, most importantly, the complex mosaics of natural and artificial ecosystems (Kier et al. 2009). Conservation and management planning, such as the selection of representative conservation sites, may critically depend on the detailed knowledge of empirical species distribution patterns (Allouche et al. 2006; Kier et al. 2009). Model predictions based on local climate dataset had illustrated projection map accurately showing distribution patterns of plant species along elevation and topography that is useful to delineate conservation area in fragmented habitat at landscape scale. In fact, projection map of *B. sinensis* in NTWN had provided useful information for delineating effective conservation area to protect isolated populations in fragmented habitats.

This study suggested a two-step procedure to identify climate characteristics of central and marginal populations of plant species at continental scale and project local distribution of marginal population at landscape scale. Climate characteristics of plant species at continental scale was available to be projected based on the WorldClim dataset. Meanwhile, local climate dataset apparently provides more reliable basis for constructing climate space and delineate effective conservation area at landscape scale. The method proposed in this study is also a powerful tool to correlate climate factors with species distribution at landscape scale that offer detailed geographical and ecological information for assessing the impact of future climate change on shifting distribution range of plant species.

Conclusion

Our results demonstrated that global climate dataset has the advantage on projecting geographical extent of plant species and is evidently appropriate to evaluate effects of climate change on species distribution range at continental scale. High-resolution local climate dataset had accurately captured climate heterogeneity induced by elevation and topography that is available for SDMs to project accurate

distribution map represented species distribution pattern at landscape scale. Accurate distribution map based on local climate dataset offers an advantage in designing conservation areas for plant species in mountainous areas with isolation and fragmentation of natural habitat, since it delineated accurate geographical boundaries of the species along elevation and topography. Model predictions based on local climate dataset was also a powerful tool to correlate species distribution with climate factors that can be confidently applied for assessing the impacts of future climate change on shifting species distribution range at landscape scale.

Declarations

Acknowledgement

The authors deeply appreciate assistant researcher Lin, Huan-Yu in Taiwan Forestry Research Institute for his helping and sharing technique of climate data interpolation. The authors want to thank for Mr. Shih-Chieh Kuo and many students in the Department of Life Science, Chinese Culture University, Taipei, Taiwan for their helps in data collections. The authors appreciate Mr. Kai-Jie Yang in the Institution of Geology, Chinese Culture University, Taipei, Taiwan for his technical support of ArcGIS software. This work was in part supported by Yangmingshan National Park, Construction and Planning Agency Ministry of the Interior, Executive Yuan, Taipei, Taiwan.

Data Availability

The data analyzed in this study was downloaded from Global Biodiversity Information Facility (GBIF) and herbarium of the Taiwan Forestry Research Institute, herbarium of National Taiwan University, and herbarium of the Academia Sinica, Taipei.

Consent to Publish (Ethics)

All the authors agree to publish the manuscript.

Author Contribution

CC Liao conceived and designed the study and analyzed data. CR Chang and YH Cheng discussed with CC Liao and contributed novel ideas of the study. All authors had read and approved the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This work was in part funded by Yangmingshan National Park, Construction and Planning Agency Ministry of the Interior, Executive Yuan, Taipei, Taiwan.

References

1. Abdelaal M, Fois M, Fenu G, Bacchetta G (2019) Using MaxEnt modeling to predict the potential distribution of the endemic plant *Rosa arabica* Crép. in Egypt. *Ecological Informatics* 50: 68–75
2. Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232
3. Bedia J, Herrera S, Gutiérrez JM (2013) Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. *Global Planet Change* 107:1–12
4. Boulesteix AL, Janitza S, Kruppa J, König IR (2012) Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining Knowledge Discovery* 2:493–507
5. Breiman L (2001) Random forests. *Machine learning* 45:5–32
6. Chao WC, Song GZM, Chao KJ, Liao CC, Fan SW, Wu SH, Hsieh TH, Sun IF, Kuo YL, Hsieh CF (2010) Lowland rainforests in southern Taiwan and Lanyu, at the northern border of Paleotropics and under the influence of monsoon wind. *Plant Ecol* 210:1–17
7. Chen ZS, Hsieh CF, Jiang FY, Hsieh TH, Sun IF (1997) Relationships of soil properties to topography and vegetation in a subtropical rain forest in southern Taiwan. *Plant Ecol* 132:229–241
8. Datta A, Schweiger O, Kühn I (2020) Origin of climatic data can determine the transferability of species distribution models. *NeoBiota* 59:61
9. Dong F, Liu X, Zhang Y, Lu R, Zheng Z, Shen X, Li Y (2019) Characterization of the complete plastid genome sequence of the endangered species *Bretschneidera sinensis*. *Conservation Genetics Resources* 11:287–290
10. Duque-Lazo J, Van Gils H, Groen T, Navarro-Cerrillo R (2016) Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecol Model* 320:62–70
11. Early R, Sax DF (2014) Climatic niche shifts between species' native and naturalized ranges raise concern for ecological forecasts during invasions and climate change. *Glob Ecol Biogeogr* 23:1356–1365
12. Edwards EJ, Still CJ (2008) Climate, phylogeny and the ecological distribution of C4 grasses. *Ecol Lett* 11:266–276
13. El-Gabbas A, Dormann CF (2018) Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecology Evolution* 8:2196–2206

14. Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst* 40:677
15. Evans JS, Cushman SA (2009) Gradient modeling of conifer species using random forests. *Landscape Ecol* 24:673–683
16. Fernández M, Hamilton H (2015) Ecological niche transferability using invasive species as a case study. *Plos One* 10:e0119891
17. Ferro ML, Flick AJ (2015) “Collection Bias” and the Importance of Natural History Collections in Species Habitat Modeling: A Case Study Using *Thoracophorus costalis* Erichson (Coleoptera: Staphylinidae: Osoriinae), with a Critique of GBIF.org. *The Coleopterists Bulletin* 69:415–425
18. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology* 37:4302–4315
19. Fois M, Fenu G, Lombrana AC, Cogoni D, Bacchetta G (2015) A practical method to speed up the discovery of unknown populations using Species Distribution Models. *Journal for Nature Conservation* 24:42–48
20. Godsoe W, Murray R, Plank MJ (2015) Information on biotic interactions improves transferability of distribution models. *Am Nat* 185:281–290
21. Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007) What matters for predicting the occurrences of trees: techniques, data, or species characteristics? *Ecol Monogr* 77:615–630
22. Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276–288
23. Hu X-G, Wang T, Liu S-S, Jiao S-Q, Jia K-H, Zhou S-S, Jin Y, Li Y, El-Kassaby YA, Mao J-F (2017) Predicting future seed sourcing of *Platycladus orientalis* (L.) for future climates using climate niche models. *Forests* 8:471
24. Hu Z-y, Lin L, Deng J-f, Wang S-h (2014) Genetic diversity and differentiation among populations of *Bretschneidera sinensis* (Bretschneideraceae), a narrowly distributed and endemic species in China, detected by inter-simple sequence repeat (ISSR). *Biochem Syst Ecol* 56:104–110
25. Iturbide M, Bedia J, Gutiérrez JM (2018) Background sampling and transferability of species distribution model ensembles under climate change. *Global Planet Change* 166:19–29
26. Kier G, Kreft H, Lee TM, Jetz W, Ibisch PL, Nowicki C, Mutke J, Barthlott W (2009) A global assessment of endemism and species richness across island and mainland regions. *PNAS* 106:9322–9327
27. Kumar V, Dash SS, Panday S, Lahiri S, Sinha BK, Singh P (2017) Akaniaceae: A New Family Record for Flora of India and Lectotypification of the Name *Bretschneidera sinensis*. *Bulletin of the Botanical Survey of India* 59:8–10
28. Lannuzel G, Balmot J, Dubos N, Thibault M, Fogliani B (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. *Biodiversity and Conservation*

29. Li CF, Chytrý M, Zelený D, Chen MY, Chen TY, Chiou CR, Hsia YJ, Liu HY, Yang SZ, Yeh CL (2013) Classification of Taiwan forest vegetation. *Appl Veg Sci* 16:698–719
30. Li M, Chen H, Wang Z, Zhang S (2016) Isolation and characterization of polymorphic microsatellite markers in the endangered species *Bretschneidera sinensis* Hemsl. *Genetics and Molecular Research* 15
31. Liao CC, Chen YH (2021) Improving performance of species distribution model in mountainous areas with complex topography. *Ecological Research*: 1–15
32. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2:18–22
33. Lin HY, Hu JM, Chen TY, Hsieh CF, Wang G, Wang T (2018) A dynamic downscaling approach to generate scale-free regional climate data in Taiwan. *Taiwania* 63:251–266
34. Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103–114
35. Ma S, Concilio A, Oakley B, North M, Chen J (2010) Spatial variability in microclimate in a mixed-conifer forest before and after thinning and burning treatments. *For Ecol Manage* 259:904–915
36. Maria B, Udo S (2017) Why input matters: Selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. *Ecol Model* 359:92–102
37. Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W (2009) Evaluation of consensus methods in predictive species distribution modelling. *Divers Distrib* 15:59–69
38. McKenzie D, Peterson DW, Peterson DL, Thornton PE (2003) Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. *J Biogeogr* 30:1093–1108
39. Mi C, Huettmann F, Guo Y, Han X, Wen L (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5:e2849
40. Mohapatra J, Singh CP, Hamid M, Verma A, Semwal SC, Gajmer B, Khuroo AA, Kumar A, Nautiyal MC, Sharma N (2019) Modelling *Betula utilis* distribution in response to climate-warming scenarios in Hindu-Kush Himalaya using random forest. *Biodiversity Conservation* 28:2295–2317
41. Qiao Q, Chen H, Xing F, Wang F, Zhong W, Wen X, Hou X (2012) Pollination ecology of *Bretschneidera sinensis* (Hemsley), a rare and endangered tree in China. *Pak J Bot* 44:1897–1903
42. Raes N, Roos MC, Slik J, Van Loon EE, Steege Ht (2009) Botanical richness and endemism patterns of Borneo derived from species distribution models. *Ecography* 32:180–192
43. Schorr G, Holstein N, Pearman P, Guisan A, Kadereit J (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine Primula. *Ecology Evolution* 2:1260–1277
44. Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annu Rev Ecol Evol Syst* 40:415–436
45. Thompson JD, Lavergne S, Affre L, Gaudeul M, Debussche M (2005) Ecological differentiation of Mediterranean endemic plants. *Taxon* 54:967–976

46. Title PO, Bemmels JB (2018) ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* 41:291–307
47. Tomlinson S, Lewandowski W, Elliott CP, Miller BP, Turner SR (2020) High-resolution distribution modeling of a threatened short-range endemic plant informed by edaphic factors. *Ecology Evolution* 10:763–777
48. Tu WG, Gao XF, Wu N, Liu SH (2010) Altitudinal patterns of vascular plant richness in high mountains: Applying of generative additive model. *Polish Journal of Ecology* 58:231–240
49. Wake DB, Hadly EA, Ackerly DD (2009) Biogeography, changing climates, and niche evolution. *PNAS* 106:19631–19636
50. Wan J-Z, Wang C-J, Yu F-H (2017) Spatial conservation prioritization for dominant tree species of Chinese forest communities under climate change. *Clim Change* 144:303–316
51. Wang M-N, Duan L, Qiao Q, Wang Z-F, Zimmer EA, Li Z-C, Chen H-F (2018) Phylogeography and conservation genetics of the rare and relict *Bretschneidera sinensis* (Akaniaceae). *Plos One* 13:e0189034
52. Wang T, Hamann A, Spittlehouse D, Carroll C (2016) Locally downscaled and spatially customizable climate data for historical and future periods for North America. *Plos One* 11:e0156720
53. Weigelt P, Kreft H (2013) Quantifying island isolation—insights from global patterns of insular plant species richness. *Ecography* 36:417–429
54. Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. *Divers Distrib* 15:565–576
55. Wu J-E, Huang S, Wang J-C, Tong W-F (2001) Allozyme variation and the genetic structure of populations of *Trochodendron aralioides*, a monotypic and narrow geographic genus. *J Plant Res* 114:45–57
56. Xu Y, Huang Y, Zhao H, Yang M, Zhuang Y, Ye X (2021) Modelling the effects of climate change on the distribution of endangered *Cypripedium japonicum* in China. *Forests* 12:429
57. Yu F, Skidmore AK, Wang T, Huang J, Ma K, Groen TA (2017) Rhododendron diversity patterns and priority conservation areas in China. *Divers Distrib* 23:1143–1156
58. Zhang M-G, Slik JF, Ma K-P (2017) Priority areas for the conservation of perennial plants in China. *Biol Cons* 210:56–63
59. Zhu Y, Wei W, Li H, Wang B, Yang X, Liu Y (2018) Modelling the potential distribution and shifts of three varieties of *Stipa tianschanica* in the eastern Eurasian Steppe under multiple climate change scenarios. *Global Ecology Conservation* 16:e00501

Figures

Figure 1

Georeferenced occurrences of *Bretschneidera sinensis* Hemsl. in continental East Asia (dark circles in upper right map) and northern Taiwan (grey circle with dark outline in the central map). Five areas were delineated in northern Taiwan for calculating empirical lapse rates of climate data. They are northeast (NE), northwest (NW), southwest (SW) and southeast (SE) slopes of Yamingshan area and Pingxi area (PX). The locations of 30 climate meteorological stations adopted in this study were represented by dark star within grey circle. Mountain tops were represented by white triangles. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

Georeferenced occurrences (dark circles) and contemporary potential distribution range of *Bretschneidera sinensis* in continental East Asia (dark grey areas in the maps) projected by Random Forest based on the WorldClim dataset (a). Distribution pattern of the plant species presents gridded squares when the larger scale maps present distribution range at two local areas, Zhangjiajie, Hunan (b) and Wenshan Prefecture, Yunnan (c). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 3

Contemporary potential distribution range of *Bretschneidera sinensis* in northern Taiwan projected by Random Forest algorithm with the WorldClim (a) and local climate datasets (b). The symbols of meteorological stations are the same as in the Fig. 1. The georeferenced occurrences of *B. sinensis* are represented by gray solid circles and white areas bordered by dark lines are the potential distribution range of the endangered plant species. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 4

Relative importance of bioclimatic predictors in explaining species distribution of *Bretschneidera sinensis* in northern Taiwan. Relative importance of bioclimatic predictors independently derived from the WorldClim (a) and local climate datasets (b).

Figure 5

Climate spaces quantified by principle components analysis (PCA). The grey circles enclosed by dashed line, thick solid line, and thin solid line were climate spaces of continental East Asia (CEA) and northern Taiwan (NTWN) based on the WroldClim dataset and NTWN based on local climate dataset, respectively. The three groups of dark circles were climate spaces of *B. sinensis* based on the WorldClim and local climate datasets in CEA and NTWN.

Figure 6

Histogram of three bioclimatic predictors in continental East Asia (CEA) based on the WorldClim dataset (a, b, and c) and northern Taiwan (NTWN) based on the WorldClim dataset (d, e, and f) and NTWN based on local climate datasets (g, h, and i). The grey histograms represent the relative frequencies of background cells. The blank histograms within the grey histograms are the cells with presences of *Bretschneidera sinensis* evaluated by Random Forest. Winter precipitation based on the WorldClim dataset presented wide range in CEA and gridded cells with winter precipitation higher than 200 mm were showed in the subplot at the upper central figure. The x and y scales and intervals were different among histograms.

Figure 7

Characteristics of bioclimatic predictors from the WorldClim (left column) and local climate datasets (right column). Bioclimatic predictors from the WorldClim presented significant gridded squares that were not available to reveal climate heterogeneities induced by elevation and topography, while that from local climate dataset better reflect elevation and topographical features. The first row is annual temperature (Bio1) from the WorldClim (a) and local climate dataset (b). The second row (c and d) is annual precipitation (Bio12). The third row (e and f) is winter precipitation (Bio19). The dark stars are meteorological stations used in this study. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research

Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.