Taiwania 70(4): 658–676, **2025**
DOI: 10.6165/tai.2025.70.658

Virtual species as baseline for predicting trends of real species distributions in mountainous regions

Chi-Cheng LIAO[1,*], Huan-Yu LIN[2], Su-Wei FAN[3], Yi-Huey CHEN[1]

*1. Department of Life Science, Chinese Culture University, Taiwan, R.O.C. 2. Department of Forestry and Natural Resources, National Ilan University, Yilan County, Taiwan. 3. Taiwan Forestry Research Institute, Taipei, Taiwan. *Corresponding author's email: hunter_yyl@yahoo.com.tw*

(Manuscript received 25 February 2025; Accepted 17 September 2025; Online published 4 October 2025)

ABSTRACT: Species distribution models (SDMs) play a crucial role in predicting species' geographic distributions, especially in mountainous regions where complex terrains pose significant challenges for sampling efforts. To enhance the accuracy of SDM predictions, researchers have employed high-resolution climate datasets derived from meteorological data. These datasets were used to simulate five virtual species across different climatic niches using principal components analysis (PCA). This approach allowed for a thorough examination of various environmental conditions, such as coastal, inland, and leeward slopes, thereby capturing the intricate climatic heterogeneity within mountainous landscapes. In these simulations, eleven algorithms were assessed using the Expected Fraction of Shared Presences (ESP), which measures the overlap between simulated suitable ranges and predicted potential ranges. Ensemble modeling proved superior, particularly when dealing with larger sample sizes, showing minimal impact from geographical distribution variations. However, smaller sample sizes notably reduced ESP values, geographical distance among sample points provide signals of niches that will be identified through accurate predictions of ensemble modeling. Temperature has been determined to be more critical than precipitation for species distribution, heavily influenced by elevation gradients. Climate change projections suggest that by the 2050s and 2090s, species currently found in inland areas or on leeward slopes may experience range expansions. Conversely, species inhabiting coastal areas or windward slopes are projected to face range restrictions or local extinctions. This study supports the development of targeted conservation strategies in topographically complex areas, offering scientific support for conservation planning in mountainous regions and forecasts of species distributions under future climate change.

KEY WORDS: climate change, ensemble modeling, high-resolution climate data, species distribution model, virtual species.

INTRODUCTION

Species distribution models (SDMs) represent a robust analytical framework for elucidating the environmental determinants that underpin species occurrence patterns (Chambers *et al.*, 2013; El−Gabbas and Dormann, 2018; Qazi *et al.*, 2022; Zimmer *et al.*, 2023). By correlating presence-only or presence/absence data with pertinent environmental variables, SDMs facilitate the creation of predictive maps delineating potential species distributions (Peterson *et al.*, 2011). The accuracy of these projected ranges is pivotal for informed decision-making and is especially vital for enhancing conservation strategies for rare or endangered species (Guillera−Arroita *et al.*, 2015; Lannuzel *et al.*, 2021; Hama and Khwarahm, 2023; HamadAmin and Khwarahm, 2023; Zurell *et al.*, 2023). The reliability of SDM predictions significantly depends on adequate sample sizes and unbiased representation of spatial data (Kadmon *et al.*, 2003; Guisan *et al.*, 2007; Elith *et al.*, 2011; Bean *et al.*, 2012; El−Gabbas and Dormann, 2018; Inman *et al.*, 2021). Achieving substantial sample sizes and unbiased spatial sampling presents a formidable challenge, particularly concerning rare or endangered species (Laskey *et al.*, 2020). This issue is compounded in mountainous regions, where complex topography, habitat fragmentation, and steep climatic gradients often result in small, isolated populations and skewed distribution patterns.

SDMs are rarely applied in mountainous areas not only due to the sample size of species but also because of the resolution of climate datasets. Global climate datasets with a 30-arc-second resolution, such as WorldClim (Fick and Hijmans, 2017) or Chelsa (Karger *et al.*, 2017), fail to accurately capture the heterogeneous climatic conditions prevalent in mountainous regions (Fick and Hijmans, 2017). This limitation persists even when these datasets are downscaled to finer resolutions (Dobrowski, 2011; Pradervand *et al.*, 2014; Wang *et al.*, 2016). Recently, a statistical method to generate gridded climate datasets with a spatial resolution of 50 × 50 m^2 for mountainous areas has been developed to improve the performance of SDMs (Liao and Chen, 2021; Liao *et al.*, 2023). The climate dataset introduced by Liao *et al.* (2023) effectively captures the varied climatic environments within mountainous terrains, and when utilized in SDMs, it accurately predicts the potential distribution ranges of grassland ecosystems (Liao *et al.*, 2023). However, this climate dataset has not yet been directly utilized for modeling rare or endangered species, thus its efficacy for predicting their distributions remains uncertain, potentially limiting broader application.

Before being utilized in modeling studies, the high-resolution climate dataset necessitated thorough evaluation. Virtual species are increasingly employed in SDMs primarily because they facilitate the isolation of individual feature effects within complex models (De Marco and Nóbrega, 2018). In this study, virtual species simulations were conducted to determine whether the high-resolution climate dataset accurately captured the heterogeneity of climate environments in mountainous areas. A virtual species was generated by mimicking real species to create a simulated ecological niche within an n-dimensional environmental space (Hirzel *et al*., 2001; Duan *et al*., 2015; Leroy *et al*., 2016; Qiao *et al*., 2016). Specifically, the high-resolution climate dataset developed by Liao *et al*. (2023) was utilized to simulate virtual species through principal component analysis (PCA), involving estimating the probability of each grid cell belonging to the climate niche (Leroy *et al*., 2016). Subsequently, the grid cells classified as presence were mapped to illustrate suitable or geographical ranges within the study area. Different virtual species were simulated across various regions of the PCA diagram, assuming distinct geographical distributions due to the heterogeneous climate environments within the study area.

SDMs were subsequently utilized to project the potential distribution ranges of the virtual species. Virtual species have gained popularity for testing SDMs because they offer known distribution ranges, facilitating a comprehensive understanding of species-environment relationships (De Marco and Nóbrega, 2018; Hirzel *et al*., 2001; Meynard and Kaplan, 2013). These simulated species provide controlled, unbiased presence and absence data, which are typically unavailable to field ecologists (Hirzel *et al*., 2001; Meynard *et al*., 2019). The "true" presence and absence data generated by virtual species are particularly suitable for assessing model performance and mitigating the effects of spatial sampling biases (Hirzel *et al*., 2001; Bombi and D'Amen, 2012; Leroy *et al*., 2016; Qiao *et al*., 2016; Qiao *et al*., 2019).

In addition to predicting current distribution ranges, SDMs were also employed to project future distribution ranges under various climate change scenarios. Climate change represents a powerful force that can induce shifts or expansions in species ranges and significantly heightens the risk of extinction for rare or endangered species (Jiang *et al*., 2022; Qazi *et al*., 2022; Zurell *et al*., 2023). Species characterized by small population sizes, such as rare or endangered species, are particularly vulnerable to the impacts of climate change (IUCN, 2014). Accurately predicting current distribution ranges and evaluating potential future range shifts through SDMs is essential for designing effective conservation strategies for these species (Ning *et al*., 2021; Wan *et al*., 2021; Jiang *et al*., 2022; Ali *et al*., 2023; Zimmer *et al*., 2023; Zurell *et al*., 2023). Therefore, future climate datasets are crucial for assessing species vulnerability to climate change using SDMs (HamadAmin and Khwarahm, 2023; Zurell *et al*., 2023). If the current high-resolution climate dataset generated by Liao *et al*. (2023) accurately reflects climatic heterogeneity in mountainous areas, it can serve as a foundation for generating future climate datasets. The methodology developed by Liao *et al*. (2023) was utilized to create future climate datasets, which were then applied in SDMs to project range shifts for both virtual and real species within the study area.

The objective of this study is to establish current and future climate datasets at a 50 × 50 m^2 resolution for mountainous regions. This high-resolution dataset aims to capture the heterogeneous climate environments of mountainous areas. By utilizing this dataset to simulate virtual species and establish baseline scenarios, we will analyze the current potential distribution ranges of real species and model their future distribution trends. Several key questions were addressed: (1) These simulations of virtual species enabled a comprehensive assessment of the high-resolution climate dataset's ability to reflect climatic heterogeneity and its effectiveness in predicting species distributions under varying environmental conditions. (2) Various virtual species, simulated using the high-resolution climate dataset to determine habitat suitability, were assumed to exhibit distinct geographical distribution ranges within the mountainous area. (3) SDMs were employed to explore the current and future distribution ranges of virtual species in the mountainous area, with the aim of examining the extent of range expansion, restriction, shift, or local extinctions under climate change. (4) Subsequently, real species were used in algorithms to predict both their current potential ranges and future distribution ranges under climate change, with the goal of informing conservation management applications. This approach enhances our understanding of species-environment relationships and aids in developing effective conservation strategies tailored to the impacts of climate change.

MATERIALS AND METHODS

Study area

Taiwan, situated in the subtropical zone at the western fringe of the Pacific Ocean, spans latitudes from 21° 55' to 25° 20' N and longitudes from 119° 30' to 122° 00' E (Fig. 1). Positioned approximately 150 kilometers off the southeastern coast of mainland China, this island experiences a distinct monsoon climate influenced by its geographical location (Chen and Tsai, 1983; Su, 1984). The climatic conditions are significantly affected by two prevailing monsoons: the northeast monsoon during winter and the southwest monsoon during summer. Specifically, the northeast monsoon dominates for half a year, typically from November through April, impacting the northern and eastern flanks of the Central Mountain Range with substantial precipitation and robust winds.

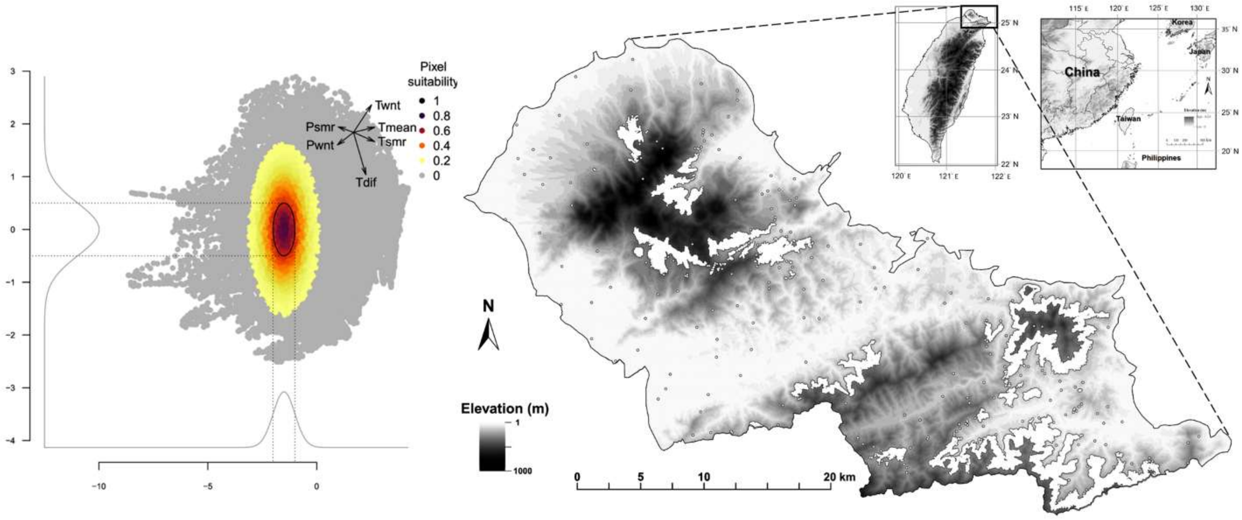

Fig. 1. The fundamental niche of the virtual species (left diagram) generated through the principal components analysis (PCA), and its geographic distribution range in northern Taiwan (right maps). The virtual species was generated using the 'generateSpFromPCA' function from an R package. This function quantified the climatic environments of the study area's gridded cells via PCA. Follow this, suitable environments for the virtual species were identified in the PCA diagram by setting two parameters within the function: means and standard deviation (SDs), showing a color gradient ranging from purple (high suitability) to yellow (low suitability) and grey (unsuitability). The resulting geographic distribution range of the virtual species were shown in right map (white area with dark outline). The right maps also present the study area in northern Taiwan, the geographic locations of Taiwan, and neighboring countries.

This seasonal weather pattern underscores the dynamic interplay between Taiwan's topography and atmospheric circulation patterns, which shapes the island's diverse ecological landscapes.

In Northern Taiwan (NTWN), a pronounced precipitation gradient exists, extending from coastal regions to inland areas, which significantly influences the distribution of plant species (Liao and Chen, 2022; Liao *et al*., 2023). This gradient shows a marked decrease in annual precipitation, from over 6,000 mm on the northeastern slopes to approximately 1,900 mm on the southwestern slopes of the mountain ridge in NTWN (Liao *et al*., 2023). The mean monthly temperatures at the mountain ridge fluctuate between 11.3°C during winter and 20.5°C in summer, contrasting with coastal areas where temperatures range from 17.9°C in winter to 26.6°C in summer (Liao and Chen, 2022; Liao *et al*., 2023). These climatic variations highlight the diverse environmental conditions across different elevations within NTWN, thereby shaping the ecological niches and biodiversity patterns observed in this region (Liao and Chen, 2021).

The study area in NTWN extends from 24° 57' to 25° 17' N latitude and from 121° 24' to 122° 00' E longitude (Fig. 1), encompassing a region of approximately 1,031 square kilometers (103,100 hectares). The highest peak within this region is Qixingshan, which reaches an elevation of 1,120 meters above sea level (asl). The predominant vegetation type within NTWN consists of evergreen broad-leaved forests, as documented by previous papers (Hsieh *et al*., 1997; Liao *et al*., 2012; Li *et al*., 2013). These forests are predominantly populated with species such as *Castanopsis*, *Cleyera*, *Cyclobalanopsis*, *Dendropanax*, *Elaeocarpus*, *Engelhardia*, *Gordonia*, *Helicia*, *Ilex*, *Keteleeria*, *Limlia*, *Litsea*, *Machilus*, *Meliosma*, *Michelia*, *Pinus*, *Schefflera*, *Symplocos*, and *Trochodendron*, as reported by Li *et al*. (2013). On average, the canopy height of these forests measures around 15 meters. It is important to note that there are no pure deciduous forests present in NTWN; however, individual native deciduous tree species can be found interspersed throughout the forested areas. Additionally, natural grasslands are frequently observed along the mountain ridges extending from coastal regions into inland territories within NTWN. The elevations at which these natural grasslands occur vary across different ridges within the study area, as indicated by Liao *et al*. (2023). In these mountainous areas, the dominant species of natural grassland are *Miscanthus sinensis* and *Pseudosasa usawai* (Liao *et al*., 2014, 2023). Notably, these two species exhibit similar climatic niches, suggesting potential niche convergence (Liao *et al*., 2023).

Process of Downscaling current climate dataset

In this study, a gridded climate dataset with a spatial resolution of 50 × 50 m² was developed to depict the historical climatic conditions within the study area. The 50 × 50 m² grid was specifically chosen to accurately capture the diverse climatic environments present along the mountain slopes. This dataset was generated utilizing daily meteorological data obtained from the Central Weather Bureau's website (CWB; https://e-service.cwa.gov.tw/HistoryDataQuery/index.jsp). The construction of this dataset involved downloading daily meteorological data from the CWB's online platform, covering the period from 2000 to 2020. According to Liao *et al.* (2023), the methodology for downscaling the historical climate dataset comprised several key steps: (1) Interpolation of the meteorological climate dataset to generate smooth surfaces representing climate variables. (2) Establishment of a grid system consisting of cells with a spatial resolution of 50 × 50 m², from which data were extracted from these smoothed climate variable surfaces. (3) Adjustment of the extracted climate data according to altitude, ensuring that variations in elevation were accurately reflected in the final dataset. By adhering to these procedures, we aimed to produce a comprehensive and detailed representation of the historical climate environments within our study area. This meticulous approach ensures not only the accuracy but also the reliability of the climate data utilized in our study.

Climate data from 30 meteorological stations located within and surrounding the study area were downloaded from the CWB website. Monthly mean temperature and total monthly precipitation data from these stations were imported into ArcInfo software (ESRI, Redlands, California, USA) to produce smooth climate variable surfaces through the application of the Kriging method. This process resulted in the creation of .tif files for the climate variables. Additionally, a grid system consisting of cells with a spatial resolution of 50 × 50 m² was established using ArcInfo software, generating over 0.4 million gridded cells within the study region. For each gridded cell, geographic coordinates (longitude and latitude) and elevation data were extracted from a digital terrain model (DTM) developed by the Department of Geography at the Chinese Culture University. The DTM had a resolution of 20 by 20 meters. The elevation data derived from this DTM was designated as DElev. Subsequently, the 50 × 50 m² gridded cells were overlaid onto the .tif files representing the meteorological climate surfaces to extract relevant climate data. The resultant dataset, referred to as MCD50, represents the meteorological climate data with a spatial resolution of 50 × 50 m². Furthermore, the elevations of the meteorological stations were interpolated using the Kriging method within ArcInfo, yielding a smooth elevation surface named MElev. By overlaying the gridded cells of MCD50 with MElev, elevation data were extracted for each cell. The differences between DElev and MElev were then calculated to perform altitudinal adjustments on MCD50. The adjustment function used was: AdjMCD50 = slope × (DElev – MElev) + MCD50. Here, AdjMCD50 denotes the altitudinally adjusted meteorological climate data at a spatial resolution of 50 × 50 m². The slope parameter in this function, known as the empirical lapse rate, was determined as the slope of the linear relationship between elevation and climate data from the nearest 12 meteorological stations. This linear regression analysis was carried out using the "stats" package within the R environment (Chambers and Hastie, 1992). The methodology employed for generating the current climate dataset closely followed the approach outlined by Liao *et al.* (2023).

The altitudinally adjusted climate data, AdjMCD50, served as the historical climate dataset for modeling species distributions. This comprehensive climate dataset was constructed by incorporating nine key climatic variables: mean annual temperature (Tmean), mean maximum temperature of the warmest month (Twrm), mean minimum temperature of the coldest month (Tcld), mean temperature during summer (Tsmr) and winter (Twnt), the temperature difference between the warmest and coldest months (Tdif), annual total precipitation (Pann), total precipitation during summer (Psmr), and total precipitation during winter (Pwnt). Each of these variables plays a critical role in defining the environmental conditions that influence species distribution patterns. By integrating these variables into the AdjMCD50 dataset, we aimed to create a robust representation of the climatic conditions within the study area. Such detailed climatic information is fundamental for accurately predicting species distributions using species distribution models (SDMs). Precise climate data, like those provided by the AdjMCD50 dataset, enhance the reliability and accuracy of SDMs, thereby contributing to more effective conservation planning and management strategies.

Process of downscaling future climate projections

The history of the Intergovernmental Panel on Climate Change's (IPCC) assessment reports spans several generations of emissions scenarios (IPCC, 2013, 2022; Pedersen *et al.*, 2021). Emission scenarios are typically developed to depict a range of socio-economic and policy choices, enabling the evaluation of diverse potential futures and their implications for long-term climate systems (Kebede *et al.*, 2018). In 2022, the global Shared Socioeconomic Pathways (SSPs) were introduced in the Sixth Assessment Report (AR6), the most recent publication by the IPCC, which integrates socioeconomic developments into future climate scenarios (IPCC, 2022). There are five SSPs (SSP1 to SSP5), each serving as a framework for generating a spectrum of possible future climate and environmental conditions based on different societal trajectories and policy directions (Pedersen *et al.*, 2021; IPCC, 2022). Among these pathways, SSP1 represents the most optimistic scenario, emphasizing sustainable development practices aimed at mitigating climate change and fostering equitable growth. Conversely, SSP2 is characterized as a middle-of-the-road pathway, reflecting current trends and policies with no significant shifts towards either sustainability or unsustainable development (Pu *et al.*, 2020). SSP3 and SSP4 are identified as the least favorable scenarios,

Table 1. Climate variables for model predictions.

Predictors	Climate variable	Unit	Mean ± SD	Min	Max
Psmr	Summer precipitation	dm	7.7 ± 0.8	5.5	13.8
Pwnt	Winter precipitation	dm	10.6 ± 4.4	1.4	23.6
Tmean	Mean annual temperature	℃	21.3 ± 1.3	15.6	23.8
Twnt	Mean winter temperature	℃	16.9 ± 1.0	12.7	19.1
Tsmr	Mean summer temperature	℃	25.8 ± 1.6	19.2	28.7
Tdif	Temperature differences between warmest and coldest monthes	℃	12.9 ± 1.2	9.2	15.4

assuming patterns of unsustainable development that could exacerbate environmental degradation and social inequalities. These pathways highlight the risks associated with continued reliance on non-renewable resources and inadequate adaptation measures. Meanwhile, SSP5 presents an energy-intensive, fossil-fuel-based economy, yet it also envisions relatively optimistic economic development outcomes (Pu *et al*., 2020). By integrating these SSPs with advanced climate models, we can simulate various future climates and assess the effectiveness of different strategies in achieving climate goals.

Concerning future climate scenarios, the working group of the Taiwan Climate Change Information and Adaptation Knowledge Platform (TCCIP) has developed high-resolution 5 × 5 km² gridded climate datasets to project future climatic conditions for Taiwan Island (Wang *et al*., 2021). These projections were refined from 49 General Circulation Models (GCMs) involved in the sixth phase of the Coupled Model Intercomparison Project (CMIP6), ensuring a comprehensive coverage of potential climate futures. For this study, five specific climate system models were selected from the available 49 GCMs: ACCESS-CM2 (Meucci *et al*., 2023), FGOALS-g3 (Pu *et al*., 2020), GFDL-ESM4 (Dunne *et al*., 2020), MIROC6 (Kataoka *et al*., 2020), and TaiESM1 (Wang *et al*., 2021). The selection of these models was based on their performance and relevance to the regional climate characteristics of Taiwan. Utilizing these models, the research team generated downscaled future climate datasets at an enhanced resolution of 50 × 50 m² for the study area under various climate scenarios. This downscaling process is crucial as it allows for more precise climate predictions that are better suited to local environmental conditions.

The TCCIP has published 5 × 5 km² gridded climate datasets that cover the period from 1960 to 2100. Given that the spatial resolution of the AdjMCD50 dataset is 50 × 50 m², the TCCIP's future climate projections at this finer resolution were recalculated based on the relative changes observed in historical and future climate data provided by TCCIP. For this study, we selected three key time periods from the TCCIP's climate datasets: early (2000-2020), mid (2045-2055), and end (2091-2100) of the 21st century. These datasets were utilized to calculate the mean monthly temperature and total monthly precipitation for each period. Subsequently, differences in monthly temperature and total monthly precipitation between the early and mid-century, as well as the early and end-century, were computed to generate relative changes in climate data. These relative changes in temperature and precipitation were represented as 5 × 5 km² gridded data. To produce smooth climate surfaces, these gridded data were used to generate .tif files using the Kriging method within ArcInfo software. The high-resolution grids with a spatial resolution of 50 × 50 m² were then overlaid with the .tif files representing smoothed climate surfaces. This overlay process allowed for the extraction of relative changes in climate data at the higher resolution. Finally, the historical climate datasets, specifically the AdjMCD50, were overlaid with the relative changes derived from the 50 × 50 m² gridded climate datasets. This step was crucial for projecting future climates under different scenarios. By employing this methodology, researchers can provide detailed projections that are closely aligned with local climate conditions, thereby enhancing the accuracy and relevance of climate impact assessments.

Simulations of virtual species

The gridded cells containing historical climatic data (spanning from 2000 to 2020) underwent analysis through Principal Components Analysis (PCA). This statistical approach was chosen for its efficacy in reducing dimensionality while simultaneously addressing issues of collinearity among environmental variables, thus facilitating a more precise quantification of environmental overlap (Qiao *et al*., 2016; Meynard *et al*., 2019; Journé *et al*., 2020). Initially, nine climate variables were incorporated into the PCA. During the analysis, significant correlations were identified between Pwnt and Pann, Tsmr and Twrm, as well as Twnt and Tcld. Consequently, to minimize redundancy and improve model clarity, Pann, Twrm, and Tcld were excluded from further PCA computations. The remaining variables subjected to PCA included Tmean, Tsmr, Twnt, Tdif, Psmr, and Pwnt (Table 1). Tmean provides an overall measure of thermal conditions throughout the year, while Twrm and Tcld highlight the extremes that organisms must endure. The variables Tsmr and Twnt offer insights into seasonal temperature fluctuations, which can be particularly significant for species with specific temperature requirements during different life stages. Meanwhile, Pann, Psmr, and Pwnt are essential for understanding moisture availability across seasons, a crucial factor for many species' survival and reproduction. These variables were effectively summarized by the first two principal components (PC1 and PC2), which collectively accounted for 84.1% of the total variance observed within the dataset.

In this study, virtual species serve as controlled experimental units to isolate topographic effects—the primary driver of heterogeneous climate environments in the study area—from biological noise arising from species' responses to these environments. Virtualspecies (Leroy *et al*., 2016) constitutes a comprehensive R package offering a complete workflow encompassing environmental response simulation, presence-absence conversion, dispersal limitations, and sampling bias integration. It supports diverse environmental response functions (linear, Gaussian, logistic) and combination methods, including multiplicative approaches. Virtualspecies differs from other methods, such as NicheLim (Huang *et al*., 2016), which represents a specific algorithmic approach designed to address the "substitution effect"—where high-suitability environmental variables compensate for unsuitable conditions in others—by emphasizing ecological limiting factor theory such as Liebig's Law of the Minimum during variable combination. Given that real species often exhibit continuous, probabilistic distributions rather than threshold-driven presences, virtualspecies' probabilistic conversion via the convertToPA function's logistic transformation aligns with this ecological reality. Furthermore, by generating environmental suitability through a PCA-based approach (generateSpFromPCA), the package inherently accommodates complex environmental interactions and variable correlations. This avoids ecologically contradictory combinations—such as simultaneous requirements for high maximum and low minimum temperatures—that might arise when applying independent thresholds, an advantage not inherently addressed in NicheLim. The need for ecological realism in virtual species generation is critical, as these serve as benchmarks for real species applications. In mountainous environments characterized by intricate climatic gradients, species occurrence typically follows gradual probabilistic patterns rather than abrupt thresholds. Our study encompasses both rare and common species within such topographically complex systems, where NicheLim's strict emphasis on limiting factors may not be universally optimal for capturing distributional nuances across diverse mountain-dwelling taxa.

Consequently, virtual species with rare-type niche spaces deliberately selected provide a rigorous stress test of our 50 × 50m² dataset's capacity to resolve fine-scale microclimate gradients. To generate a variety of spatial patterns in habitat suitability, the "generateSpFromPCA" function, available within the R package "virtualspecies," was utilized to simulate virtual species (Leroy *et al*., 2016). This approach facilitated the creation of distinct environmental niches and suitability landscapes by leveraging principal component analysis (PCA). In employing this function, the parameter `nb.points` was set to 5000, aiming to provide sufficient resolution for the simulated distributions. Furthermore, the parameters `means` and `sds` were meticulously adjusted to position five virtual species across various sectors of the PCA diagram, thereby representing diverse ecological niches and suitable environments (Leroy *et al*., 2016). The implementation of these settings led to the emergence of distinctive geographical distribution patterns across the study area, as depicted in Figs. 1 and S1. Specifically, the first virtual species was strategically positioned at the center of the PCA diagram, resulting in a geographic distribution that encompasses the windward slopes adjacent to mountain ridges (Fig. 1). Meanwhile, the remaining four virtual species were allocated to the lower, left, right, and upper extremities of the PCA diagram (Fig. S1), respectively. By positioning these virtual species in different regions of the PCA space, the simulation ensured that each occupied a unique environmental niche, thereby achieving wide-ranging geographical distributions throughout the study area. This methodology not only underscores the flexibility and applicability of the "virtualspecies" package in simulating complex ecological scenarios but also highlights its potential in exploring the impacts of varying environmental conditions on species distribution patterns.

Inventory of real species

In this study, we selected eleven plant species and grassland areas for modeling analysis. The eleven species and grassland areas were selected for modeling analysis based on two core criteria: verifiable field presence within the study area, and ecological representativeness through distinct distribution patterns (including both geographically restricted and widespread species) across diverse taxonomic groups (encompassing woody angiosperms and herbaceous angiosperms, and ferns). Among these, five species are rarely encountered within the study area: *Maackia taiwanensis*, *Benthamidia japonica*, *Lilium speciosum*, *Rhododendron pseudochrysanthum*, and *Bretschneidera sinensis*. Their scarcity can be attributed to their restricted geographic distribution on this island, predominantly limited to the mountainous regions of NTWN, with only a few occurrences during fieldwork. The remaining six species consist of four woody angiosperms—*Rhododendron nakaharai*, *Euscaphis japonica*, *Ficus fistulosa*, and *Saurauia tristyla* var. *oldhamii*—and two fern species, *Dipteris conjugata* and *Sphaeropteris lepifera*. The presence of these eleven plant species and grasslands was recorded along roadsides and mountain trails within the study area. Field data were meticulously geographically verified to eliminate any duplicate occurrence records, ensuring the accuracy and reliability of our dataset.

Modelling technique

In this investigation, we utilized a suite of eleven predictive algorithms to delineate the potential distribution ranges for both virtual and real species inhabiting mountainous terrains (Table S1). These methodologies encompassed an array of machine learning and statistical techniques: Artificial Neural Networks (ANN), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), Generalized Additive Models (GAM), Generalized Boosting Models (GBM; often referred to as boosted regression trees), General Linear Models (GLM), Multivariate Adaptive Regression Splines (MARS), Maximum Entropy (MAXENT), Random Forest (RF), Surface Range Envelopes (SRE; also known as BIOCLIM), and Extreme Gradient Boosting (XGBOOST). Furthermore, an ensemble modeling framework was adopted, integrating all aforementioned algorithms to enhance prediction accuracy for species distributions within these environments. The execution of these models, including the ensemble approach, was facilitated through the utilization of the "biomod2" package within the R programming environment (Thuiller *et al*., 2016).

For each virtual species, occurrence and background points were aggregated to create a comprehensive modeling dataset. Specifically, the geographic coordinates of these occurrences and background points were utilized to extract corresponding climate data from predefined climate surfaces. Occurrence points were randomly selected from the known geographic distribution range of the virtual species, as illustrated in Fig. 1 and detailed further in Fig. S1. To evaluate the influence of sample size on model performance, varying numbers of sample points were incorporated into the predictive models: 5, 20, 50, 100, and 200. For each scenario, the number of background points was set to be 100 times the number of occurrence points, randomly distributed across the study area. In implementing the models, a stratified random sampling approach was employed, whereby 80% of both occurrence and background data were designated for training purposes, while the remaining 20% served for model evaluation. Such an approach is crucial for gauging the effectiveness of different sample sizes and understanding their implications for ecological predictions.

Model performance

To evaluate the accuracy of individual algorithms and ensemble modeling approaches, the training dataset underwent resampling and modeling processes ten separate times. This iterative approach was designed to quantify the uncertainties inherent in the predictive outputs. The true skill statistic (TSS) and the area under the receiver operating characteristic curve (AUC) served as the primary metrics for assessing the performance of species distribution models (Fois *et al*., 2015; Qiao *et al*., 2019; Lannuzel *et al*., 2021; Xu *et al*., 2021). These measures are particularly useful because they provide insights into how well a model can distinguish between positive and negative cases, with AUC being especially robust in scenarios where class imbalance is present. For the construction of the final ensemble models, stringent criteria were applied to ensure only the most reliable predictions were incorporated. Specifically, models were included in the ensemble if they achieved a TSS score exceeding 0.8, indicating a high level of accuracy and reliability according to previous studies that have established this threshold as indicative of superior model performance (Khan and Verma, 2022).

This rigorous selection process helps to mitigate the risk of incorporating less accurate models, thereby enhancing the overall predictive power and robustness of the ensemble model. The iterative resampling and modeling not only allowed for the assessment of model stability across different subsamples but also provided a means to estimate the variability within the predictions. By averaging the outcomes from multiple runs, the ensemble approach effectively reduces the impact of random fluctuations and enhances the generalizability of the model predictions. Furthermore, by retaining models above the specified TSS threshold, the ensemble benefits from the collective strength of individually strong performers, potentially leading to improved ecological niche estimations and more reliable projections of species distributions under various environmental conditions.

Notably, a prior study posited that error metrics such as the TSS and the AUC do not necessarily reflect the accuracy of suitability predictions. These indices provide a single-number discrimination measure across all possible threshold ranges, which may not always correspond to model accuracy (Lobo *et al*., 2008). High TSS and AUC values are therefore insufficient on their own to guarantee accurate model performance (Liao and Chen, 2022). Consequently, this study introduces an alternative similarity index known as the Expected Fraction of Shared Presences (ESP) to evaluate model performance. The ESP was adapted from the Sorenson similarity index to quantify the degree of overlap between potential ranges of two species (Godsoe, 2014; Inman *et al*., 2021). In this context, the ESP has been tailored to compare the suitable range of a virtual species simulated through Principal Component Analysis (PCA) against its predicted potential ranges using eleven different algorithms and ensemble modeling techniques. The mathematical formulation of the ESP is defined as follows:

$$\mathrm{ESP} = \frac{2\Sigma_1^j P_{s(j)} P_{p(j)}}{\Sigma_1^j (P_{s(j)} + P_{p(j)})}$$

Herein, $P_{s(j)}$ represents the presence of a suitable range at a specific cell j, whereas $P_{p(j)}$ denotes the presence of a potential range at the same cell j. The term $P_{s(j)}P_{p(j)}$ indicates that cell j is simultaneously within both the

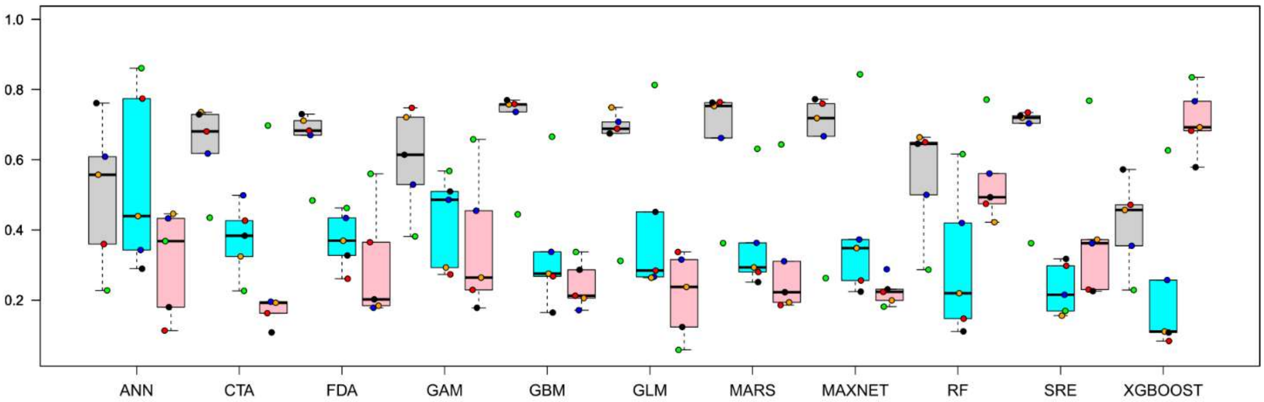

Fig. 2. The boxplot shows the index of the expected fraction of shared presence (ESP, grey box), false presence (cyan box), and false absence (pink box). The ESP was calculated by the degree of overlap between the suitable ranges of the virtual species at center of principal components analysis (PCA) diagram and the potential ranges predicted by 11 algorithms. A relatively high ESP value indicates better performance of the algorithms. False presence (commission error) and false absence (omission error) represent prediction errors of the 11 algorithms. The colored points in the boxplots represent the sample sizes used for predictions by the algorithms. The sample sizes are 5, 20, 50, 100, and 200 random points of occurrences, represented by green, blue, orange, red, and black colors, respectively. Their interquartile range are shown by boxes and with whiskers for their full range. Their interquartile ranges are shown by boxes, with whiskers indicating their full range.

suitable ($P_{s(j)}$) and potential ranges ($P_{p(j)}$). An ESP value of 1 signifies perfect concordance between the suitable and potential ranges for the virtual species, while a value of 0 indicates complete geographic disjunction (Godsoe, 2014; Inman *et al*., 2021).

This adaptation of the ESP provides a more nuanced approach to evaluating the performance of species distribution models by focusing on the spatial agreement between modeled suitable and potential ranges. Unlike traditional metrics like TSS and AUC, which focus primarily on the discriminatory power of the model, the ESP emphasizes the spatial congruence of the predicted distributions. This makes it particularly useful in ecological applications where understanding the extent of overlap between different species' ranges is crucial for conservation planning and biodiversity assessments.

RESULTS

Suitable ranges of virtual species

Five virtual species were simulated by selecting gridded cells from various regions within the PCA diagram, each representing distinct suitable ranges for the five virtual species within the study area. The central gridded cells in the PCA (PCAC) diagram corresponded to a distribution range characterized by windward slopes near mountain ridges within the study area (Fig. 1). Conversely, the gridded cells located at the lower (PCAO), left (PCAL), right (PCAR), and upper (PCAU) extremities of the PCA diagram represented distribution ranges associated with coastal areas, mountain ridges, inland regions, and leeward slopes near mountain ridges, respectively (Fig. S1). The creation of these five virtual species, each with unique climatic niches identified within the PCA diagram, demonstrated clear distinctions in their geographical distribution ranges across the study area. This evidence strongly suggests that heterogeneous climate environments can offer a variety of suitable habitats conducive to species growth. Through the simulation of these virtual species, we identified the high-resolution gridded climate dataset derived from meteorological data as an effective tool for reflecting the climatic heterogeneity inherent to mountainous terrains.

Performances of 11 algorithms

The potential distribution ranges of five virtual species were assessed using eleven distinct algorithms. This evaluation was designed with the objective of identifying the most suitable algorithms for predicting the distribution ranges of real species in mountainous regions. Subsequently, the calculation of ESPs was employed to quantify the degree of overlap between the suitable ranges of virtual species, as depicted in the PCA diagram, and the predicted potential ranges by these algorithms (Fig. 2). Among the evaluated algorithms, FDA, GBM, GLM, MARS, MAXENT, and SRE exhibited relatively higher mean ESP values. These elevated mean ESP scores indicated a greater extent of overlap between the simulated suitable ranges and the projected potential ranges. Consequently, this suggested superior performance of these six algorithms in accurately predicting the potential distributions of the virtual species within the studied area. The analysis not only underscores the effectiveness of specific algorithms but also highlights their applicability in ecological modeling for real-world species in complex terrains. The findings provide valuable insights into algorithm selection for future studies aiming at ecological predictions in mountainous landscapes.

In the context of the virtual species approach, ESP

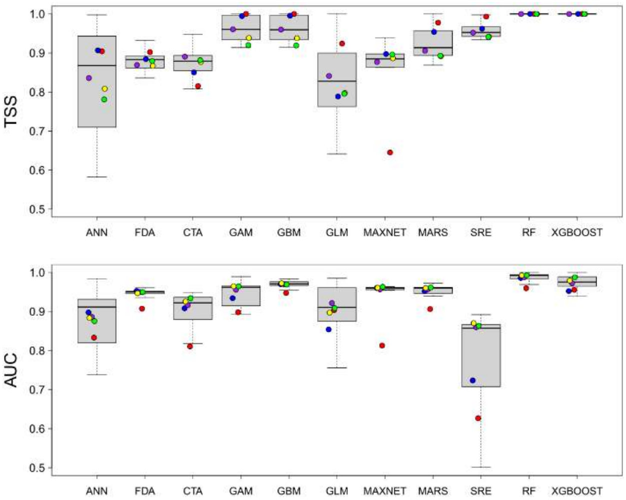

Fig. 3. Boxplots show the values of two error indices, TSS and AUC, used to assess predictive performances of 11 algorithms. The 11 algorithms predicted potential ranges of the virtual species at the center of the principal components analysis (PCA) diagram. The colored points in the two boxplots represent sample sizes used for model predictions. The sample sizes are 5, 20, 50, 100, and 200 random points of occurrences, represented by red, blue, purple, yellow, and green colors, respectively.

associated with false presences and false absences serve as indicators for discerning algorithmic characteristics. Lower ESP values typically signify a higher likelihood of commission errors (false presences) and omission errors (false absences) (Fig. 2). High rates of false presences and false absences suggest overestimation and underestimation of the potential ranges of virtual species, respectively. A modest sample size can reveal the distinct attributes of various algorithms. Specifically, when the number of sample points was limited to five, the predicted ranges generated by CTA, SRE, and XGBOOST exhibited relatively lower rates of false presence but significantly elevated rates of false absences. This pattern indicates that CTA, SRE, and XGBOOST are prone to higher omission errors or underestimations of the species' potential ranges. Conversely, GLM and MAXENT produced predictions characterized by relatively high rates of false presences and slightly reduced rates of false absences. These results imply that GLM and MAXENT tend to incur higher commission errors or overestimate the species' potential ranges. The insights gained from algorithms that either overestimate or underestimate potential ranges offer valuable technological options. Such knowledge can provide scientific support for designing conservation strategies tailored to specific ecological needs and environmental conditions.

The potential ranges predicted by algorithms utilizing a large sample size demonstrated a high degree of overlap with the simulated suitable ranges of virtual species. These findings underscore that sample size significantly influences model performance, especially when the sample size is limited to five points. At such a small sample size, it becomes evident that the predictive capacity of models is constrained; this is reflected in the ESP values, which were consistently below 0.45 across all evaluated algorithms. Our results indicate that with only five sample points, the effectiveness of algorithms in accurately predicting potential species distributions is notably diminished. This limitation highlights the necessity for a more robust sampling strategy to improve model accuracy and reliability. Adequate sample sizes are crucial for enhancing the precision of ecological predictions, thereby providing a stronger foundation for conservation planning and management decisions.

In addition, the distinct geographical ranges of the five virtual species have influenced the ESP values to some extent. Specifically, for the virtual species distributed in coastal areas (PCAO), the range of ESP values is observed to be marginally narrower, as depicted in the first diagram of Fig. S2. Conversely, for the virtual species located on the leeward slopes near mountain ridges (PCAU), the range of ESP values is somewhat broader, as illustrated in the lowest diagram of Fig. S2. These variations in ESP value ranges can be attributed to the unique environmental conditions associated with each geographical setting.

In many previous investigations, the TSS and the AUC have been utilized as key metrics for evaluating the predictive performance of various algorithms. These indices are generally understood to reflect the efficacy of a model, with higher scores suggesting accurate performance of a model. Despite this common understanding, our study has identified that high TSS and AUC values do not invariably correspond to enhanced model outcomes. Specifically, when the ESP values were elevated, there was no consistent correlation with increased TSS or AUC values, which sometimes appeared at odds with the ESP index. Moreover, the analysis conducted in this research did not reveal any discernible patterns within the TSS and AUC values (Figs. 3 and S3). This suggests that these metrics may not be reliable indicators under all circumstances, particularly in contexts where other factors such as ESP play a significant role. Consequently, it can be concluded that while TSS and AUC remain valuable tools for assessing algorithmic performance, they may not be suitable as standalone measures for predicting the effectiveness of models in every scenario. It is imperative to consider additional metrics and contextual variables to ensure a comprehensive evaluation of model performance.

This study also utilizes virtual species distributions to assess the predictive accuracy of ensemble modeling. The ensemble approach effectively delineated the potential geographic ranges of the five virtual species (Figs. 4 and S4). Notably, ensemble modeling outperformed the individual 11 algorithms, as evidenced by ESP values that were predominantly close to or exceeding 0.8 when sample sizes surpassed five points. The influence of sample sizes

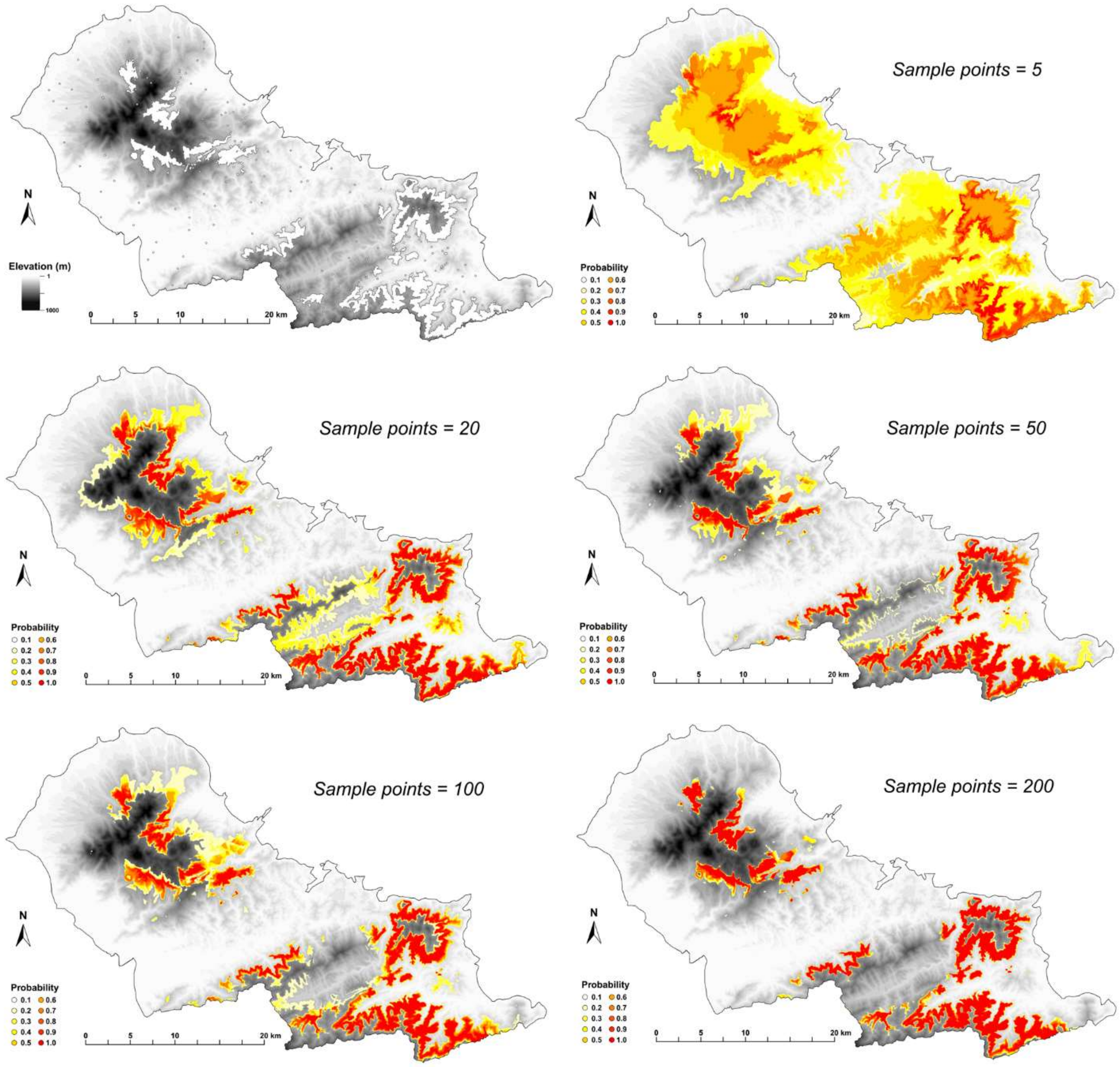

Fig. 4. Maps show the geographical distribution ranges of virtual species (upper left map) generated from the principal components analysis (PCA) and the projected results by the ensemble modeling with different sample sizes of occurrences. The virtual species is located at the center of the PCA diagram (Fig. 1). The gridded cells that were categorized by ensemble modeling as presence points are shown by colored points. The points with a color gradient from yellow to red represent the probability of presence. Sample size had significant effects on the performance of ensemble modeling, while five sample points are evidently the limitation for the predictions of ensemble modeling.

and geographical distribution range on ensemble modeling was found to be relatively minimal. Nevertheless, extremely small sample sizes exerted a significant impact on the projection outcomes of ensemble modeling. For instance, when only five sample points were used, ensemble modeling yielded lower ESP values, indicative of reduced model performance (Fig. 5). Conversely, ensemble modeling exhibited superior performance when the number of sample points exceeded twenty, leading to a precise alignment between the simulated suitable habitat range and the predicted potential range for the virtual species. Given that ensemble modeling accurately projected the potential distribution ranges of the virtual species, it emerges as a robust tool for predicting species' potential distribution ranges. Importantly, all five virtual species could be reliably predicted using this method, irrespective of their ecological characteristics. This suggests that ensemble modeling is not only effective but also versatile across different ecological contexts. Our findings underscore the reliability of ensemble modeling in generating accurate information regarding the potential ranges of species.

Table 2. The importance values of predictors for 11 algorithms and ensemble modeling in predicting virtual species. The virtual species are located at the center of PCA diagram.

Predictors	ANN	FDA	CTA	GAM	GBM	GLM	MAXNET	MARS	SRE	RF	XGBOOST	Ensemble
Psmr	0.11	0.00	0.00	0.05	0.03	0.00	0.01	0.04	0.09	0.06	0.07	0.04
Pwnt	0.08	0.01	0.01	0.07	0.03	0.02	0.01	0.05	0.09	0.06	0.16	0.05
Tmean	0.30	0.31	0.00	0.18	0.02	0.15	0.00	0.10	0.20	0.21	0.04	0.14
Twnt	0.16	0.27	0.33	0.25	0.23	0.27	0.25	0.28	0.18	0.18	0.26	0.24
Tsmr	0.20	0.32	0.51	0.24	0.28	0.50	0.38	0.35	0.22	0.26	0.18	0.31
Tdif	0.15	0.09	0.15	0.21	0.42	0.05	0.35	0.18	0.22	0.23	0.30	0.21

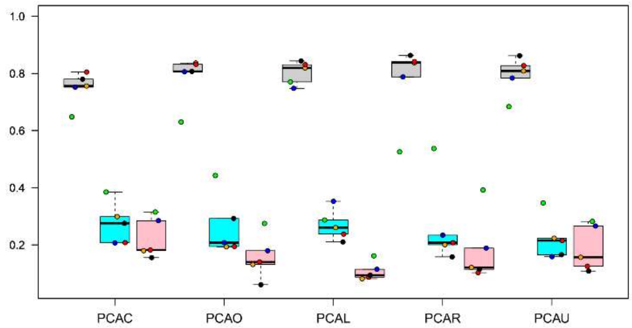

Fig. 5. Boxplots show index of expected fraction of shared presence (ESP, grey box), false presence (cyan box), and false absence (pink box). The ESP value was calculated by degree of overlap between the suitable ranges of five virtual species generated from principal components analysis (PCA) and the potential ranges predicted by ensemble modeling. Relatively high ESP values indicate better performance of ensemble modeling. The performance of ensemble modeling is better than that of the 11 algorithms (Fig. 2). The colored points in the boxplots represent the number of sample points used for predictions in ensemble modeling. The numbers of sample points are 5, 20, 50, 100, and 200 and the corresponding colors are green, blue, orange, red, and black, respectively. Their interquartile ranges are shown by boxes and their full ranges are depicted with whiskers.

Importance of predictors

For the five virtual species examined in this study, the relative importance of predictor variables varied across different algorithms, with temperature consistently emerging as a more significant factor compared to precipitation (Tables 2 and S2). The study region is characterized by complex topography and elevation gradients that undoubtedly influence species distributions. Given that temperature is highly correlated with elevation gradients, it plays a crucial role in predicting the distribution of virtual species. Conversely, precipitation patterns differ markedly between windward and leeward slopes within the study area. Two species, *Bretschneidera sinensis* and *Benthamidia japonica*, are predominantly found on windward slopes where higher precipitation levels are recorded. Consequently, the distribution of these two species may be closely associated with the elevated precipitation observed on the windward slopes. However, the predictive models did not significantly capture the relationship between plant distribution and the precipitation gradient within the study area. It is important to note that while temperature appears to be a critical determinant in species distribution modeling, the lack of a clear representation of precipitation effects highlights potential limitations or complexities within the model. This could imply that additional factors or interactions not fully accounted for in the current analysis might play a role in shaping species distributions. Further research may be necessary to better understand how precipitation interacts with other environmental variables to influence species presence and abundance.

Model evaluation of real species

Given the successful application of ensemble modeling in predicting the distributions of virtual species, this methodology was then applied to estimate the potential distribution ranges of 11 plant species and grassland ecosystems (Fig. 6). Sample size was identified as the most influential factor affecting model performance. Acquiring a large sample size for species, particularly in mountainous terrains where rare or endangered species may reside, poses significant challenges. When the number of occurrence records was below 50, all occurrences were included in the ensemble modeling predictions. For species with over 50 occurrence records, a random subset of 50 samples was selected for the modeling process. Ensemble modeling successfully and reasonably projected the potential distribution ranges of the 11 plant species and grassland areas (Fig. 6), highlighting its effectiveness even under conditions of variable sample sizes and complex geographies.

The impact of a small sample size on the potential distribution ranges of actual species is profound. *Lilium speciosum* and *Maackia taiwanensis*, for example, have notably limited sample sizes, with only 9 and 14 occurrence records respectively. However, these two species exhibit dramatically different patterns in their potential distribution ranges. For *L. speciosum*, despite its small sample size, the species demonstrates a broad potential geographical range, likely due to its widespread initial distribution. In contrast, *M. taiwanensis* shows both its occurrences and potential distribution ranges constrained to a smaller geographical area within the study zone. These observations indicate that although sample size is a crucial factor, it is not the only one affecting potential distribution ranges. The spatial arrangement and distances between occurrence points also contribute significantly to the model's predictions of species distribution.

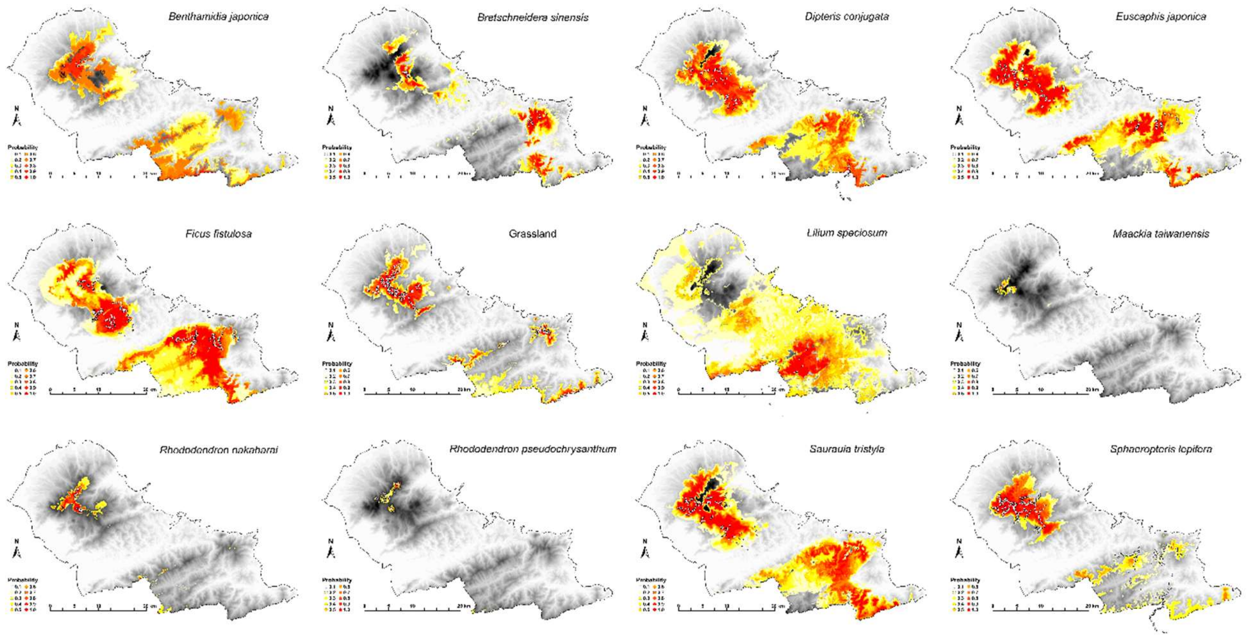

Fig. 6. The maps show the occurrences of the 11 species and grassland in northern Taiwan (depicted as open dots) and their potential distribution ranges predicted by ensemble modeling. The gridded cells categorized as presence are represented by colored points. The points exhibit a color gradient from yellow to red, indicating the probability of presence.

Range shifts of virtual and real species under future climate change

Virtual species characterized by distinct contemporary distributions displayed marked range shifts, contractions, or expansions under the impacts of climate change (Fig. S5). Species positioned on windward slopes, mountain ridges, or coastal zones will have significant range shifts by mid-century and faced range contractions by the end of the century (Figs. S5.1, S5.2, and S5.3), particularly under the SSP585 climate model scenario. In contrast, virtual species inhabiting inland areas and leeward slopes near inland regions exhibited trends toward range expansion both by mid-century and towards the century's close (Figs. S5.4 and S5.5).

Ensemble modeling has revealed that the consequences of climate change could yield diverse outcomes for the 11 plant species and grassland ecosystems (Fig. S6). Predictive results from ensemble modeling suggest that climate change might cause local extinctions among certain rare species, such as *M. taiwanensis* (Fig. S6.8), *R. nakaharai* (Fig. S6.9), and *R. pseudochrysanthum* (Fig. S6.10). However, not all rare species will necessarily suffer from range contractions or local extinctions; *L. speciosum*, for instance, is projected to expand its range by mid-century and into the end of this century (Fig. S6.7). A number of species are expected to experience range shifts during the mid and latter parts of this century, including *Br. sinensis* (Fig. S6.2), *D. conjugata* (Fig. S6.3), *S. tristyla* var. *oldhamii* (Fig. S6.11), and *Sp. lepifera* (Fig. S6.12). In contrast, *Be. japonica* (Fig. S6.1) and *F. fistulosa* (Fig. S6.5) are anticipated to show range expansions over the same timeframe. For the remaining species, *E. japonica* (Fig. S6.4) and grassland ecosystems (Fig. S6.6), projections based on different climate models indicate either range restrictions or potential local extinctions.

DISCUSSION

Simulation of virtual species

The utilization of virtual species simulations has been advocated as an essential preliminary step for testing novel methodologies in SDM before their application to empirical datasets (Austin *et al*., 2006; Austin, 2007). Prior research has demonstrated the effectiveness of employing such simulations to evaluate how environmental configurations impact SDM performance, the implications of various data aggregation strategies, and considerations regarding spatial resolution and scale (Hirzel *et al*., 2001; Meynard *et al*., 2019; Qiao *et al*., 2019). In this study, we extended these methodologies by simulating virtual species within a mountainous landscape. Our objective was to investigate the influence of climate dataset heterogeneity on model outcomes, assess the characteristics of different algorithms employed, examine the effects of varying sample sizes, and explore potential range shifts or local extinctions under projected climate change scenarios.

Heterogeneity of climate dataset

The WorldClim dataset (Fick and Hijmans, 2017) is extensively utilized in SDM studies. Despite its global coverage, the WorldClim dataset contains uncertainties in environmental data, particularly in mountainous regions and areas with sparse observation stations (Fick and Hijmans, 2017). Although variables from WorldClim have been downscaled to finer resolutions for SDM applications, this process may only preserve the information from the original data source without enhancing the spatial resolution of the variables themselves (Peterson *et al*., 2011; Sillero and Barbosa, 2021). In this study, we generated high-resolution historical and future climate datasets using a proposed statistical method (Liao *et al*., 2023). These high-resolution climate datasets incorporated horizontal variations in climate features through the interpolation of meteorological data. Additionally, calculations of lapse rates were employed to adjust the climate data of gridded cells, thereby accounting for climate variations along elevation gradients. Consequently, the high-resolution climate dataset generated in this study effectively captured the topographical and altitudinal variations of climate environments within mountainous areas. The simulation of virtual species successfully certified the heterogeneous climate environments present in the study area. The use of these high-resolution climate datasets is deemed appropriate for addressing issues pertinent to SDM studies, as they provide a more nuanced representation of climatic conditions across complex terrains.

The spatial resolution of future climate datasets derived from GCMs typically stands at 5×5 km², which may not be sufficiently detailed to accurately capture the heterogeneous climate environments characteristic of mountainous regions. To address this limitation, this study employed a statistical method developed by Liao *et al*. (2023) to downscale future climate datasets from five GCMs to a finer resolution. The resulting downscaled future climate datasets, with a spatial resolution of 50×50 m², demonstrate clear reliability for predicting the future distribution ranges of virtual species within these complex terrains. This statistical approach confidently enables the generation of high-resolution historical and future climate datasets, thereby offering significant potential for application in other mountainous areas where capturing fine-scale climatic variations is crucial.

Performances and characteristics of algorithms

The virtual species approach serves as an effective tool for exploring the predictive accuracy of various algorithms. The TSS and the AUC, evaluated through split-sample validation, are frequently utilized metrics for assessing model accuracy. However, it is important to note that while TSS and AUC provide insights into a model's discrimination capacity, they may not fully reflect the actual predictive ability of SDMs. Specifically, these metrics tend to present overly optimistic evaluations of model performance, particularly when predictions are made under current conditions and even more so when projected into future scenarios (Santini *et al*., 2021). In this study, we observed that the TSS and AUC values for eleven different algorithms were predominantly close to or exceeded 0.8, except in cases where the sample size was limited to five (Fig. 3 and S3). Despite these high values, it is crucial to recognize that elevated TSS and AUC scores do not necessarily ensure accurate predictions of species' potential distributions. Indeed, our analyses revealed that the TSS and AUC values exhibited only minor variations across the different algorithms tested. Consequently, these metrics alone may not be sufficient for adequately presenting the performance of SDMs.

The results demonstrate that despite high TSS and AUC values, significant overestimation or underestimation persists in the predicted distribution ranges of virtual species. To address this limitation, our study proposes the ESP metric for assessing model performance. Within mountainous regions—where complex topography drives exceptionally heterogeneous climates—the combination of virtual species and the ESP index establishes a robust predictive benchmark for evaluating real species' potential distributions. This approach not only clearly delineates algorithmic predictive capabilities with superior effectiveness compared to TSS and AUC but also achieves critical advantages unattainable through traditional metrics: it effectively detects species distribution ranges while precisely quantifying how survey point quantity, unbiased point distribution, and algorithm selection impact model predictions. Furthermore, it reveals explicit alignment between model outputs and species distributions shaped by climatic gradients. Notably, when applied to ensemble models, this method yields exceptionally low omission and commission errors (Fig. 5).

Despite its advantages, the ESP's effectiveness is notably compromised by small sample sizes. Our analyses reveal that five sample points represent a critical limitation for model predictions, as the ESP values consistently fall below 0.45 across all evaluated algorithms. Furthermore, the ratios of false presence and false absence are particularly pronounced when the sample size is limited to five. Consequently, it is evident that a small sample size significantly diminishes the accuracy of algorithm performance.

By incorporating the ESP into the evaluation framework, researchers can obtain a more comprehensive understanding of how well their models predict not just the presence or absence of a species, but also the spatial extent of its suitable habitat. This additional layer of analysis complements existing metrics and offers valuable insights into the reliability of species distribution models, especially when dealing with complex ecological datasets involving multiple species or environmental variables. Thus, the use of the ESP alongside other standard measures enhances the robustness of ecological modeling efforts and contributes to more informed decision-making in environmental management contexts.

Future distribution range of species

Our analysis has revealed that climate change exerts distinct influences on climatic environments across different regions, thereby eliciting varied responses among the five virtual species under consideration. Climatic projections for coastal areas suggest a shift towards higher temperatures and reduced precipitation levels by the middle to late 21st century. Consequently, the future climatic conditions in coastal zones are anticipated to resemble those currently observed in inland areas. Species that are presently distributed within inland regions or on the leeward slopes of our study area are expected to expand or relocate their distribution ranges towards coastal areas in response to these climatic shifts. Conversely, species currently inhabiting coastal regions or mountain ridges are projected to experience range contractions or even local extinctions. This is primarily due to the limited potential for these species to track suitable climatic conditions within the study area under ongoing climate change. The divergent responses exhibited by various virtual species in our study clearly correlate with their specific climatic requirements. Species adapted to cooler and wetter conditions, typical of current coastal climates, may struggle to find suitable habitats as these environments become warmer and drier. In contrast, species accustomed to the more temperate conditions of inland areas might benefit from the shifting climatic patterns, potentially expanding their ranges into previously unsuitable territories. Thus, our findings underscore the critical role that climatic preferences play in determining species' adaptive capacity and distributional shifts under climate change. These insights highlight the necessity for tailored conservation strategies that account for the unique vulnerabilities and adaptive potentials of different species in the face of global environmental changes.

The differing responses of virtual species to climate change provide a foundational understanding for studying the potential fates of real plant species. Several of our study species, including *Be. japonica*, *Br. sinensis*, *E. japonica*, *M. taiwanensis*, *R. nakaharai*, and *R. pseudochrysanthum*, are at risk of local extinction. These six species have suitable ranges primarily located in coastal areas or on windward slopes near mountain ridges. Other species, such as *D. conjugata*, *Sa. tristyla* var. *oldhamii*, *Sp. lepifera*, and grassland are projected to experience range restrictions by the middle or end of this century. The suitable range for grassland is widely distributed along mountain ridges, whereas the three species (*Sa. tristyla* var. *oldhamii*, *Sp. lepifera*, and *D. conjugata*) are found across both leeward and windward slopes. Interestingly, only two species, *F. fistulosa* and *L. speciosum*, are expected to expand their distribution ranges under climate change. It is reasonable to predict that *F. fistulosa* will expand its suitable range due to its current wide distribution within the study area. However, the anticipated expansion of *L. speciosum* is somewhat surprising. Historically, *L. speciosum* was widespread in the lowlands of northern Taiwan but has suffered from extensive collection for trade or cultivation since the last century. Consequently, its population size has significantly diminished in northern Taiwan, with only a few scattered individuals remaining within the study area today. Due to human influences, the current distribution of *L. speciosum* largely avoids environmentally suitable areas, occupying only a fraction of its fundamental niche within the study area. This suggests that the absence of the species from suitable habitats is driven by artificial factors rather than climatic suitability. As noted by Elith *et al*. (2011), the presence records used in modeling may reflect these anthropogenic absences, which can be projected by algorithms. Furthermore, the predicted range expansion of *L. speciosum* by the middle or end of this century indicates that if a rare or endangered species is not threatened by climate change but rather by human activities, it may not face extinction under future climatic conditions.

Our projections highlight critical priorities for biodiversity management under climate change in northern Taiwan. Furthermore, within this study region, species primarily suited to windward slopes are projected to experience continued range contraction due to climate change, while those thriving on leeward slopes are highly likely to expand their distribution ranges. For species at high risk of local extinction (e.g., *Be. japonica*, *Br. sinensis*, *R. nakaharai*), which have suitable ranges primarily located in coastal areas or on windward slopes near mountain ridges, conservation efforts must prioritize in situ protection of these habitats, coupled with urgent ex situ measures like seed banking and assisted migration to climatically stable refugia. Species facing range restrictions (e.g., *D. conjugata*, *Sp. lepifera*, *Sa. tristyla* var. *oldhamii*), found across both leeward and windward slopes, require habitat corridors to facilitate natural dispersal and adaptive management of these slope ecosystems. Conversely, projected range expansions (e.g., *F. fistulosa*) offer opportunities for proactive habitat restoration in newly suitable areas. Notably, *L. speciosum*'s paradox—predicted expansion versus anthropogenic decline—emphasizes that human-driven threats (e.g., illegal collection) may override climatic suitability. Thus, while SDMs identify climatically vulnerable species and patterns like slope aspect vulnerability, they also reveal cases where targeted policy interventions (e.g., anti-poaching enforcement, reintroduction programs) could enable populations to reclaim suitable habitats. Integrating these projections with threat assessments will prioritize actions: protect refugia for climate-sensitive endemics (especially on vulnerable windward slopes), mitigate non-climatic pressures for "climate-resilient" species like *L. speciosum*, and monitor areas of potential range shift (particularly leeward slopes for expanding species) for early intervention.

Niche study

In complex mountainous terrain, climatic heterogeneity creates intricate species niche-geographic distribution relationships, posing dual challenges for traditional SDMs: global climate datasets (e.g., WorldClim) inadequately resolve micro-topographic gradients, while methodological limitations hinder differentiation of true niche requirements from sampling bias. To address this, we implement a virtual species framework—leveraging the R virtualspecies package (Leroy *et al*., 2016; Qiao *et al*., 2016)—to establish a verifiable benchmark that precisely delineates niche-distribution dynamics and supports endangered species conservation. Grounded in niche theory (Hutchinson, 1957), we employ PCA to reduce dimensionality, identifying six key climatic variables (Tmean, Tsmr, Twnt, Tdif, Psmr, Pwnt) that define multidimensional niche space and quantify gradients (e.g., coastal-inland, windward-leeward). Within this space, five virtual species representing distinct niches (e.g., PCA center = ridge windward, PCA edge = inland leeward) were generated using response functions ensuring geographic distribution diversity (Figs. 1 and S1). These virtual niches provide critical benchmarks for real species: anthropogenic disturbance has skewed *L. speciosum*'s current distribution, yet models reveal its climate-driven expansion potential, guiding conservation targeting; for endangered *M. taiwanensis*, predictions signal range contraction to coastal/windward refugia, serving as an extinction early-warning; while for expansion-prone species like *F. fistulosa*, strategies prioritize ecological corridor establishment. This PCA-based approach, integrating high-resolution gridded climate datasets with virtual species construction, confidently defines fundamental niches at landscape scales, explicitly linking climatic demands to geographic distributions. The transferable framework enhances conservation strategy efficiency in global mountain ecosystems by resolving microclimatic complexity and predicting climate-change impacts on biodiversity.

Evaluating Modeling Practices against the GMP framework

GMP (Good Modeling Practices) refers to a set of assessment principles for evaluating the modeling process, primarily focusing on aspects such as model purpose, problem scoping, model formulation and evaluation, and model application (Jakeman *et al*., 2024). This study developed high-resolution contemporary meteorological data layers to analyze the potential distribution ranges of mountain species. By creating virtual species and mapping their geographical distributions in mountainous areas, the feasibility and reliability of the high-resolution meteorological data were validated with considerable confidence. Based on the contemporary high-resolution meteorological data, future climate data layers were constructed to assess the future distribution trends of the virtual species. After model evaluation, the future distribution trends of the virtual species aligned with findings from other studies (Ackerly *et al*., 2020; HamadAmin and Khwarahm, 2023; Liao *et al*., 2023; Xu *et al*., 2021; Yin *et al*., 2022; Zhu *et al*., 2018). The results from the virtual species demonstrated that the high-resolution contemporary and future climate data layers could adequately capture the steep climatic variation trends in complex mountainous terrains, with moderate validation achieved.

Building on this, the current potential distribution ranges and future distribution shifts of real species were evaluated. The results indicated that common species are highly likely to expand their distribution ranges, while most of rare species face extinction risks. However, one exception was noted: species whose population declines are due not to climatic conditions but to human disturbances may have a high probability of population recovery if appropriate conservation measures are implemented in the future. The study area covers the entire Yangmingshan National Park, and the findings can provide effective and rational conservation information to support administrative decision-making. The modeling process in this study incorporated factors related to human disturbance, avoiding preconceived notions about the extinction risks of rare species, and striving to achieve the goals of GMP to deliver valuable conservation insights.

CONCLUSION

This study demonstrates the utility of virtual species as an effective methodological tool for three critical applications: 1) verifying the accuracy of high-resolution climate data reflecting climatic heterogeneity in mountainous regions, 2) evaluating the characteristics of algorithms employed in ecological modeling, and 3) establishing benchmark for assessing both contemporary potential distribution ranges and future range shifts of real species. The use of real landscapes in virtual species studies offers distinct advantages by enabling direct correlation between explanatory environmental variables and observed geographical distribution patterns. Our analyses revealed that ensemble modeling outperformed eleven alternative algorithms in capturing the complex processes linking climatic requirements with species' geographical distributions. This approach proved particularly effective for predicting potential ranges of rare or endangered species, demonstrating its practical conservation value. Notably, ensemble modeling achieved optimal predictive accuracy when applied with adequate sample sizes, highlighting the critical relationship between data quality and model performance. The evaluation metric ESP emerged as more reliable than traditional measures (TSS and AUC) for assessing model accuracy, particularly when modeling rare or endangered species exhibiting scattered or elevation-biased distributions in mountainous terrain. Our findings empirically demonstrate that such distribution patterns in mountainous regions can generate either expansive or constrained potential geographical ranges, depending on species-specific ecological requirements. These results underscore the necessity of implementing robust sampling strategies to ensure ecologically meaningful predictions. Virtual species studies provide essential baselines for evaluating predicted contemporary and future distribution ranges of real species under climate change scenarios. Our research emphasizes the critical importance of long-term monitoring programs for both validating model predictions and tracking species-specific responses to evolving climatic conditions. For species projected to undergo range expansions, these findings present novel opportunities for ecosystem restoration initiatives, necessitating the development of adaptive management strategies that explicitly incorporate future climate projections. This research makes significant contributions to ecological modeling by demonstrating the superior performance of ensemble methods and emphasizing the importance of incorporating multiple environmental predictors in distribution modeling. Future studies should build upon these findings to develop comprehensive strategies that address the complexities of species-environment interactions while promoting sustainable ecosystem management. Through continued methodological advancements and enhanced collaboration among researchers, practitioners, and policymakers, conservation planners can design more effective protected area networks and management strategies for mountain ecosystems. This approach provides crucial scientific support for biodiversity conservation efforts in an era of rapid climate change.

ACKNOWLEDGMENTS

This study was supported by research grant "Study on nature-based solutions and adaptation strategies for responding to climate change impacts in the Lanyang River Basin" (NSTC 114-2321-B-004-001) funded by National Science and Technology Council. The authors would like to thank Mr. Shi-Chieh Kuo and several students from the Department of Life Science at Chinese Culture University, Taipei, Taiwan for their assistance in data collection. Particularly, Ms. Weng, Yu-Ting and Zhang, Qiao-Zhu gave the greatest help in the field works. The authors are grateful to Mr. Kai-Jie Yang from the Institute of Geography at Chinese Culture University, Taipei, Taiwan for providing technical supports for ArcGIS software. This work was partially supported by Yangmingshan National Park, Construction and Planning Agency, Ministry of the Interior, Executive Yuan, Taipei, Taiwan. Professor Hung-Yang Tseng from the Department of Atmospheric Sciences at Chinese Culture University assisted in coordinating with the Taiwan Climate Change Projection Information and Adaptation Knowledge Platform (TCCIP) to obtain climate model data. We would like to express our gratitude for his assistance. The authors also express their gratitude to TCCIP for providing gridded datasets on the magnitude of temperature and precipitation changes predicted by climate models for the 21st century.

LITERATURE CITED

(Included in the supplementary file)

Ackerly, D.D., Kling, M.M., Clark, M.L., Papper, P., Oldfather, M.F., Flint, A.L., Flint, L.E. 2020 Topoclimates, refugia, and biotic responses to climate change. Front. Ecol. Environ. **18(5):** 288–297.

Ali, F., Khan, N., Khan, A.M., Ali, K., Abbas, F. 2023 Species distribution modelling of *Monotheca buxifolia* (Falc.) A. DC.: Present distribution and impacts of potential climate change. Heliyon. **9(2):** e13417.

Araújo, M.B., New, M. 2007 Ensemble forecasting of species distributions. Trends Ecol. Evol. **22(1):** 42–47.

Austin, M. 2007 Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. **200(1-2):** 1–19.

Austin, M., Belbin, L., Meyers, J.a.A., Doherty, M., Luoto, M. 2006 Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecol. Model. **199(2):** 197–216.

Bean, W.T., Stafford, R., Brashares, J.S. 2012 The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography **35(3):** 250–258.

Bombi, P., D'Amen, M. 2012 Scaling down distribution maps from atlas data: a test of different approaches with virtual species. J. Biogeogr. **39(4):** 640–651.

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

Breiman, L., Friedman, J., Olshen, R.A., Stone, C. J. 2017 Classification and Regression Trees. Chapman and Hall/CRC.

Busby, J.R. 1991 BIOCLIM-a bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P., Eds., Nature Conservation: Cost Effective Biological Surveys and Data Analysis, CSIRO, Canberra, 64–68.

Chambers, D., Périé, C., Casajus, N., de Blois, S. 2013 Challenges in modelling the abundance of 105 tree species in eastern North America using climate, edaphic, and topographic variables. For. Ecol. Manage. **291:** 20–29.

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

Chen, T., Guestrin, C. 2016 Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794.

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

Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K. T., Gibson, J., Lawler, J.J. 2007 Random forests for classification in ecology. Ecology **88(11):** 2783–2792.

De Marco, P., Nóbrega, C.C. 2018 Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. PloS one. **13(9):** e0202403.

De'ath, G., Fabricius, K.E. 2000 Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology **81(11):** 3178–3192.

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

Duan, R.Y., Kong, X.Q., Huang, M.Y., Wu, G.L., Wang, Z.G. 2015 SDMvspecies: a software for creating virtual species for species distribution modelling. Ecography **38(1):** 108–110.

Dunne, J.P., Horowitz, L., Adcroft, A., Ginoux, P., Held, I., John, J., Krasting, J.P., Malyshev, S., Naik, V., Paulot, F. Shevliakova ,E., Stock, C.A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K.A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P.P.G., Griffies, S.M., Guo, H., Hallberg, R.W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P.C.D., Nikonov, S., Paynter, D.J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B.G., Robinson, T., Schwarzkopf, D.M., Sentman, L.T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A.T., Wyman, B., Zeng, Y., Zhao, M. 2020 The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. J. Adv. Model. Earth Sy. **12(11):** e2019MS002015.

El-Gabbas, A., Dormann, C.F. 2018 Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. Ecol. Evol. **8(4):** 2196–2206.

Elith, J., Leathwick, J.R., Hastie, T. 2008 A working guide to boosted regression trees. J. Anim. Ecol. **77(4):** 802–813.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J. 2011 A statistical explanation of MaxEnt for ecologists. Divers. Distrib. **17(1):** 43–57.

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

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

Friedman, J.H. 1991 Multivariate adaptive regression splines. Ann. Stat. **19(1):** 1–67.

Friedman, J.H. 2001 Greedy function approximation: a gradient boosting machine. Ann. Stat. **29(5)**: 1189–1232.

Godsoe, W. 2014 Inferring the similarity of species distributions using Species' Distribution Models. Ecography **37(2):** 130–136.

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

Guisan, A., Edwards Jr, T.C., Hastie, T. 2002 Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. **157(2-3):** 89–100.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S., Peterson, A.T. 2007 What matters for predicting the occurrences of trees: techniques, data, or species characteristics? Ecol. Monogr. **77(4):** 615–630.

Hama, A.A., Khwarahm, N.R. 2023 Predictive mapping of two endemic oak tree species under climate change scenarios in a semiarid region: Range overlap and implications for conservation. Ecol. Inform. **73:** 101930.

HamadAmin, B.A., Khwarahm, N.R. 2023 Mapping impacts of climate change on the distributions of two endemic tree species under Socioeconomic Pathway Scenarios (SSP). Sustainability **15(6):** 5469.

Hastie, T., Tibshirani, R. 1990 Exploring the nature of covariate effects in the proportional hazards model. Biometrics **46(4)**:1005–1016.

Hastie, T., Tibshirani, R., Buja, A. 1994 Flexible discriminant analysis by optimal scoring. J. Am. Stat. Assoc. **89(428):** 1255–1270.

Hirzel, A.H., Helfer, V., Metral, F. 2001 Assessing habitat-suitability models with a virtual species. Ecol. Model. **145(2-3):** 111–121.

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

Huang, M., Kong, X., Varela, S., Duan, R. 2016 The Niche Limitation Method (NicheLim), a new algorithm for generating virtual species to study biogeography. Ecol. Model. **320:** 197–202.

Hutchinson, G.E. 1957 Concluding remarks. - Cold Spring. Harb. Symp. **22:** 415–427.

Inman, R., Franklin, J., Esque, T., Nussear, K. 2021 Comparing sample bias correction methods for species distribution modeling using virtual species. Ecosphere **12(3): e**03422.

IPCC 2013 The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC 2022 Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In: Shukla, P.R., Skea, J., Slade, R., Khourdajie, A.A., Diemen, R. v., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., Malley, J. (eds.) Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA.

IUCN 2014 Red List of Ecosystems Workshop Report. Cambridge, United Kingdom, 22–23 January 2014. Commission on Ecosystem Management and Species Survival Commission, International Union for Conservation of Nature, Gland.

Jakeman, A.J., Elsawah, S., Wang, H.-H., Hamilton, S.H., Melsen, L., Grimm, V. 2024 Towards normalizing good practice across the whole modeling cycle: its instrumentation and future research topics. SESMO. **6:** 18755–18755.

Jiang, R., Zou, M., Qin, Y., Tan, G., Huang, S., Quan, H., Zhou, J., Liao, H. 2022 Modeling of the potential geographical distribution of three *Fritillaria* species under climate change. Front. Plant Sci. **12:** 749838.

Journé, V., Barnagaud, J.Y., Bernard, C., Crochet, P.A., Morin, X. 2020 Correlative climatic niche models predict real and virtual species distributions equally well. Ecology. **101(1):** e02912.

Kadmon, R., Farber, O., Danin, A. 2003 A systematic analysis of factors affecting the performance of climatic envelope models. Ecol. Appl. **13(3):** 853–867.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., Kessler, M. 2017 Climatologies at high resolution for the earth's land surface areas. Sci. Data. **4(1):** 170122.

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., Imada, Y., Shiogama, H., Kimoto, M., Watanabe, M. 2020 Seasonal to decadal predictions with MIROC6: Description and basic evaluation. J. Adv. Model. Earth Sy. **12(12):** e2019MS002035.

Kebede, A.S., Nicholls, R.J., Allan, A., Arto, I., Cazcarro, I., Fernandes, J.A., Hill, C.T., Hutton, C.W., Kay, S., Lázár, A.N. 2018 Applying the global RCP–SSP–SPA scenario framework at sub-national scale: A multi-scale and participatory scenario approach. Sci. Total Environ. **635:** 659–672.

Khan, S., Verma, S. 2022 Ensemble modeling to predict the impact of future climate change on the global distribution of *Olea europaea* subsp. *cuspidata*. Front. For. Glob. Change. **5:** 977691.

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. Biodivers. Conserv. **30(4):** 963–990.

Laskey, H., Crook, E.D., Kimball, S. 2020 Analysis of rare plant occurrence data for monitoring prioritization. Diversity **12(11):** 427.

Leathwick, J. R., Elith, J., Hastie, T. 2006 Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol. Model. **199(2):** 188–196.

Lek, S., Guégan, J.-F. 1999 Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. **120(2-3):** 65–73.

Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F. 2016 virtualspecies, an R package to generate virtual species distributions. Ecography **39(6):** 599–607.

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

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

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

Liao, C.C., Chen, Y.H. 2022 The effects of true and pseudo-absence data on the performance of species distribution models at landscape scale. Taiwania **67(1):** 9–20.

Liao, C.C., Kuo, S.C., Chang, C.R. 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., Lin, H.Y., Fan, S.W. 2023 A statistical method to generate high-resolution climate datasets for modeling plant distribution range and range shifts under climate change in mountainous areas. Taiwania **68(1):** 8–22.

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

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W. 2009 Evaluation of consensus methods in predictive species distribution modelling. Divers. Distrib. **15(1):** 59–69.

McCullagh, P. 2019 Generalized Linear Models. Routledge.

Meucci, A., Young, I.R., Hemer, M., Trenham, C., Watterson, I. G. 2023 140 years of global ocean wind-wave climate derived from CMIP6 ACCESS-CM2 and EC-Earth3 GCMs: Global trends, regional changes, and future projections. J. Clim. **36(6):** 1605–1631.

Meynard, C.N., Kaplan, D.M. 2013 Using virtual species to study species distributions and model performance. J. Biogeogr. **40(1):** 1–8.

Meynard, C.N., Leroy, B., Kaplan, D.M. 2019 Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? Ecography **42(12):** 2021–2036.

Ning, H., Ling, L., Sun, X., Kang, X., Chen, H. 2021 Predicting the future redistribution of Chinese white pine *Pinus armandii* Franch. under climate change scenarios in China using species distribution models. Glob. Ecol. Conserv. **25:** e01420.

Nix, H.A. 1986 A biogeographic analysis of Australian elapid snakes. Atlas of elapid snakes of Australia. **7:** 4–15.

Olden, J.D., Jackson, D. A. 2002 Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. **154(1-2):** 135–150.

Pedersen, J.S.T., Santos, F.D., van Vuuren, D., Gupta, J., Coelho, R.E., Aparício, B.A., Swart, R. 2021 An assessment of the performance of scenarios against historical global emissions for IPCC reports. Global Environ. Change. **66:** 102199.

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

Phillips, S.J., Anderson, R. P., Schapire, R. E. 2006 Maximum entropy modeling of species geographic distributions. Ecol. Model. **190(3-4):** 231–259.

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

Pu, Y., Liu, H., Yan, R., Yang, H., Xia, K., Li, Y., Dong, L., Li, L., Wang, H., Nie, Y. 2020 CAS FGOALS-g3 model datasets for the CMIP6 scenario model intercomparison project (ScenarioMIP). Adv. Atmos. Sci. **37(10):** 1081–1092.

Qazi, A.W., Saqib, Z., Zaman-ul-Haq, M. 2022 Trends in species distribution modelling in context of rare and endemic plants: a systematic review. Ecol. Process. **11(1):** 1–11.

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

Qiao, H., Peterson, A.T., Campbell, L.P., Soberón, J., Ji, L., Escobar, L.E. 2016 NicheA: creating virtual species and ecological niches in multivariate environmental scenarios. Ecography **39(8):** 805–813.

Santini, L., Benítez-López, A., Maiorano, L., Čengić, M. and Huijbregts, M.A. 2021 Assessing the reliability of species distribution projections in climate change research. Divers. Distrib. **27(6):** 1035–1050.

Sillero, N., Barbosa, A.M. 2021 Common mistakes in ecological niche models. Int. J. Geogr. Inf. Sci. **35(2):** 213–226.

Su, H.J. 1984 Studies on the climate and vegetation types of the natural forests in Taiwan (II) Altitudinal vegetation zones in relation to temperature gradient. Q. J. Chin. Forestry **17:** 57–73.

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

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J. 2022 Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecol. Monogr. **92(1):** e01486.

Wan, J.-N., Mbari, N. J., Wang, S.-W., Liu, B., Mwangi, B. N., Rasoarahona, J.R., Xin, H.-P., Zhou, Y.-D., Wang, Q.-F. 2021. Modeling impacts of climate change on the potential distribution of six endemic baobab species in Madagascar. Plant Divers. **43(2):** 117–124.

Wang, T., Wang, G., Innes, J., Nitschke, C., Kang, H. 2016 Climatic niche models and their consensus projections for future climates for four major forest tree species in the Asia–Pacific region. For. Ecol. Manage. **360:** 357–366.

Wang, Y.C., Hsu, H.H., Chen, C.A., Tseng, W.L., Hsu, P.C., Lin, C.W., Chen, Y.L., Jiang, L.C., Lee, Y.C., Liang, H. C. 2021 Performance of the Taiwan earth system model in simulating climate variability compared with observations and CMIP6 model simulations. J. Adv. Model. Earth Sy. **13(7):** e2020MS002353.

Wood, S.N. 2017 Generalized Additive Models: An introduction with R. chapman and hall/CRC.

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(4):** 429.

Yin, Y., He, Q., Pan, X., Liu, Q., Wu, Y. and Li, X. 2022. Predicting current potential distribution and the range dynamics of *Pomacea canaliculata* in China under global climate change. Biology **11(1):** 110.

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. Glob. Ecol. Conserv. **16:** e00501.

Zimmer, S.N., Holsinger, K.W., Dawson, C.A. 2023 A field-validated ensemble species distribution model of *Eriogonum pelinophilum*, an endangered subshrub in Colorado, USA. Ecol. Evol. **13(12):** e10816.

Zurell, D., Fritz, S. A., Rönnfeldt, A., Steinbauer, M.J. 2023 Predicting extinctions with species distribution models. Cambridge Prisms: Extinction. **1:** e8.

Supplementary materials are available from Journal Website