



OPEN ACCESS

EDITED BY

Jose Antonio Rodriguez Martin,
Instituto Nacional de Investigación y
Tecnología Agroalimentaria (INIA), Spain

REVIEWED BY

Marco Aurelio González Tagle,
Autonomous University of Nuevo León,
Mexico
Vinicius Manvailer,
Natural Resources Canada, Canada

*CORRESPONDENCE

Tongli Wang
✉ tongli.wang@ubc.ca

RECEIVED 08 January 2026

REVISED 26 April 2026

ACCEPTED 30 April 2026

PUBLISHED 12 June 2026

CITATION

Jiang J, Zhang L, Xu W, Lin H-Y, Xu X,
Zhang S and Wang T (2026) Evaluation
of scale-free climate data in comparison
to widely used gridded data.
Front. Clim. 8:1783859.
doi: 10.3389/fclim.2026.1783859

COPYRIGHT

© 2026 Jiang, Zhang, Xu, Lin, Xu, Zhang
and Wang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Evaluation of scale-free climate data in comparison to widely used gridded data

Jing Jiang¹, Lei Zhang², Wenhuan Xu¹, Huan-Yu Lin³, Xin Xu⁴,
Shiyi Zhang⁵ and Tongli Wang^{1*}

¹Faculty of Forestry, Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC, Canada, ²Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China, ³Department of Forestry and Natural Resources, National Ilan University, Yilan, Taiwan, ⁴Institute of Highland Forest Science, Chinese Academy of Forestry, Kunming, China, ⁵Asia-Pacific Network for Sustainable Forest Management and Rehabilitation, Beijing, China

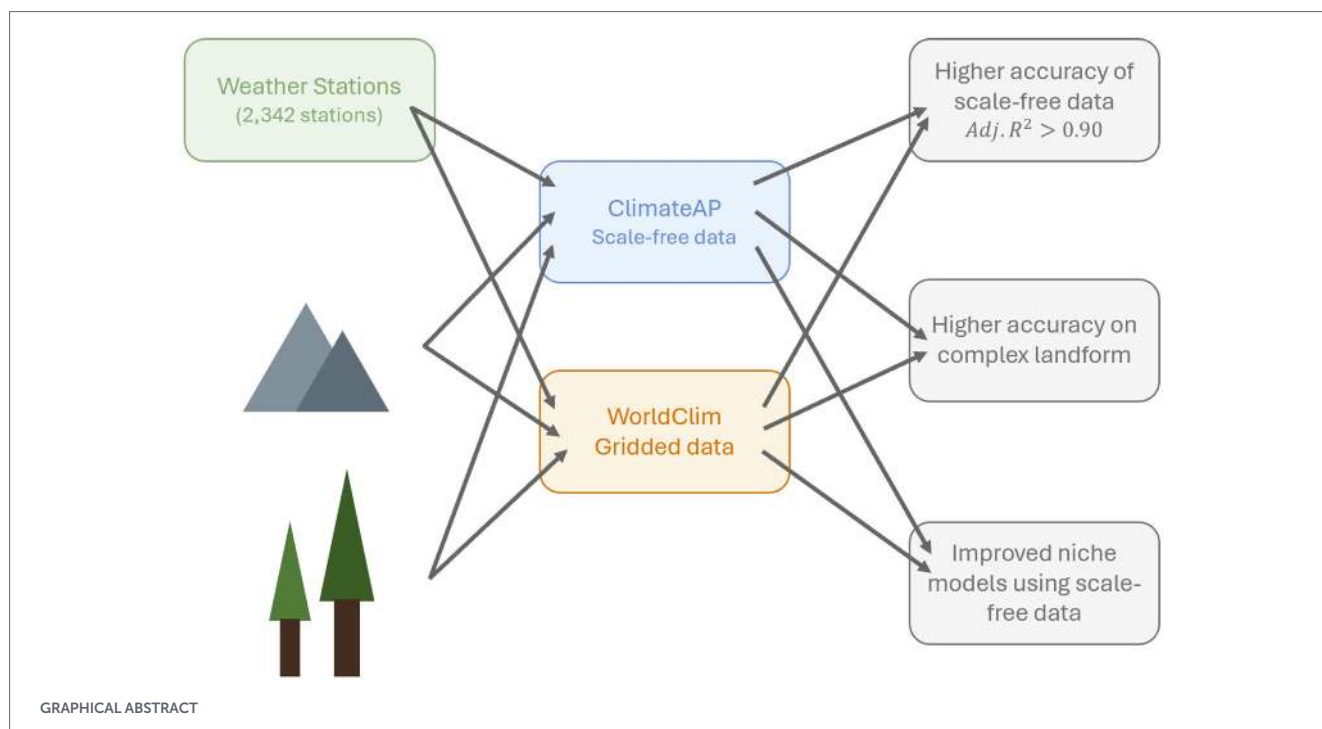
Accurate climate data is fundamental for ecological modeling and forest management to adapt to climate change. Gridded climate datasets often fail to capture local variability, especially in regions with complex terrain. In this study, we compare two open data sources: WorldClim, a widely used gridded dataset, and ClimateAP, a scale-free dataset that generates point-specific climate information for the first time. We evaluated their performance by comparing outputs against observations from 2,340 weather stations (1961 to 1990) across Mainland China. Our analysis showed that ClimateAP yielded higher adjusted R^2 values (>0.90), lower mean absolute error, and higher regression coefficient across temperature and precipitation variables. These improvements were most pronounced in complex terrain and during summer months, where ClimateAP's predictions showed up to 15% higher accuracy than WorldClim. We also applied ClimateAP data to develop climate niche models for *Cunninghamia lanceolata* and *Pinus massoniana* using a Random Forest algorithm. Models based on ClimateAP data show better performance, with lower out of bag error rates (17.70 percent versus 19.23 percent for *Cunninghamia lanceolata* and 15.67 percent versus 17.89 percent for *Pinus massoniana*). Our study provides quantitative evidence to guide the selection of open climate data sources for ecological analyses.

KEYWORDS

climate change adaptation, climate data, ecological niche modeling, forest management, machine learning, species distribution

Highlights

- ClimateAP tested with 2,340 Chinese weather stations vs. WorldClim.
- ClimateAP's scale-free data showed higher accuracy than WorldClim's gridded data format.
- Accuracy gains were most significant in areas with complex terrain.
- Tree species niche models showed improved performance when using ClimateAP data.



1 Introduction

Climate data play a critical role in developing ecological models, assessing the impact of climate change on forest ecosystems and tree species, and developing strategies for forest resource management to mitigate and adapt to climate change. Climatic niche models, in particular, are widely used to evaluate the impact of climate change on tree species and forest ecosystems, as well as to project suitable habitats for them under climate change scenarios (Wang, Wang et al., 2016; Wiens et al., 2009). However, the accuracy and credibility of these ecological models are highly dependent on the accuracy of the climate data used to build the models and make predictions (Abdulwahab et al., 2022; Matsushita et al., 2004). In recent decades, the availability of climate data sources has significantly expanded. Notable sources include WorldClim (Fick and Hijmans, 2017; Hijmans et al., 2005), PRISM (Daly et al., 2002), CRU (Harris et al., 2014, 2020), Daymet (Thornton et al., 2016), ERA5 (Hersbach et al., 2020), and ClimateNA (Wang, Hamann et al., 2016). However, most of these datasets provide gridded data for download, and some are only available in North America. Gridded climate data typically offer grid-averaged values rather than values at specific locations, which can significantly compromise the accuracy of climate data for ecological modeling, particularly in mountainous regions. In these areas, the complex topography—characterized by sharp changes in elevation, slope, and aspect—leads to significant variations in climate within a single grid. As a result, using grid-averaged values for a climate variable across an entire grid may not accurately represent local conditions. To overcome this problem, ClimateNA, a scale-free climate model, has been developed for North America (Wang, Hamann et al., 2016). It applies a dynamic local downscaling algorithm to generate climate variables for specific locations and has been widely used.

In the Asia-Pacific (AP) region, the scarcity and low quality of available climate data have significantly hindered the development of accurate climatic niche models. This limitation restricts the quality of assessments and predictions of climate change impacts on species and ecosystems (Bobrowski and Udo, 2017). To address this gap, ClimateAP¹ was developed in 2017 to generate scale-free climate variables (Wang et al., 2017). ClimateAP applies the same downscaling algorithms and provides scale-free climate data for ecological models, particularly in forestry, to inform adaptive management and conservation strategies. It generates 36 primary, 6 calculated, and 10 derived climate variables. The accuracy of the primary climate variables for the climatic normal period (1961–1990) was evaluated against observations from 1,805 weather stations of the Daily Global Historical Climatology Network (Wang et al., 2017). The variables for the climatic normal period are highly correlated with the observed values, indicating high reliability. Despite its validation, ClimateAP remains underutilized, with only 180 studies citing ClimateAP as their data source. In contrast, the majority of studies in the AP region rely on gridded data from WorldClim (Hijmans et al., 2005).

WorldClim is a global gridded dataset offering annual and monthly climate variables at four levels of spatial resolutions: 30-arc-second, 2.5-arc-minute, 5-arc-minute, and 10-arc-minute. The highest spatial resolution, which corresponds to 30-arc-second data, is available for 30-year average temperature and precipitation values (Fick and Hijmans, 2017; Hijmans et al., 2005). While this dataset covers extensive areas, its data quality has been questioned. De Clercq et al. (2015) compared species niche models using WorldClim and MODIS²-derived variables, finding that model-based MODIS data outperformed those using WorldClim. Similarly, replacing WorldClim data with MODIS or

¹ ClimateAP is an application that generates scale-free dynamically down-scaled climate data for the Asia-Pacific region (Wang et al., 2017).

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)³ improved the model fit for more species.

ClimateAP data have been utilized, albeit not extensively, in developing species niche models within the AP region (Feng et al., 2020, 2022; Wang, Wang et al., 2016). A knowledge gap exists in comparing biologically relevant variables derived from the scale-free climate data generated by ClimateAP with those from the well-established WorldClim platform. Previous studies have evaluated downscaled datasets in regions such as the northwestern USA (Jiang et al., 2018) and the southeastern USA (Wootten et al., 2014). However, a comparison between ClimateAP and WorldClim in the AP region has not been conducted. Such a comparison is critical for accurately assessing the reliability of the biologically relevant variables produced by ClimateAP and for determining the suitability of this dataset for ecological modeling of climate change impacts on the region's ecosystems and biodiversity. Forests in topographically complex areas, with their rapid climate turnover, provide ideal sites for studying local adaptation and merit more focused research than forests on flat terrains.

In this study, we hypothesize that (1) ClimateAP produces accurate data for areas where testing weather stations are located, and yields more accurate data than conventional gridded sources—WorldClim, (2) the improvement in accuracy when using ClimateAP compared to 30 arc second-gridded data is particularly substantial in landforms of complex topography, and (3) the overall quality of niche models built on ClimateAP data is superior to those using WorldClim data. To test these hypotheses, we collected observed climate data from a large number of weather stations and used them as a reference dataset. We then used these measurements as a

reference to compare the performance of ClimateAP and WorldClim outputs. We conducted univariate regression of observed climate on ClimateAP's scale-free predictions and gridded data from WorldClim. In addition, we built climatic niche models for *Cunninghamia lanceolata* and *Pinus massoniana* using either ClimateAP or WorldClim data. We aim to evaluate the accuracy and reliability of ClimateAP's biologically relevant variables by providing empirical evidence. This should contribute to a better understanding of the strengths and limitations of ClimateAP and its potential for improving ecological modeling in the AP region. Due to the accessibility of weather station data, our study focuses on China.

2 Data and methods

2.1 Climate data from weather stations

Weather station data from China were queried from a national database (Figure 1). It contains the coordinates and elevation, daily average temperature, and accumulated precipitation from 2,365 stations as monthly time-series data for individual years from 1961 to 1990. This dataset was used to derive monthly climate averages for the 1961–1990 normal period (Supplementary Table S1). This period aligns with the reference normal period defined by the World Meteorological Organization's standard for long-term climate change assessments (WMO, 2020), reflecting pre-warming averages. Weather stations that have records shorter than 10 years were excluded from further analysis.

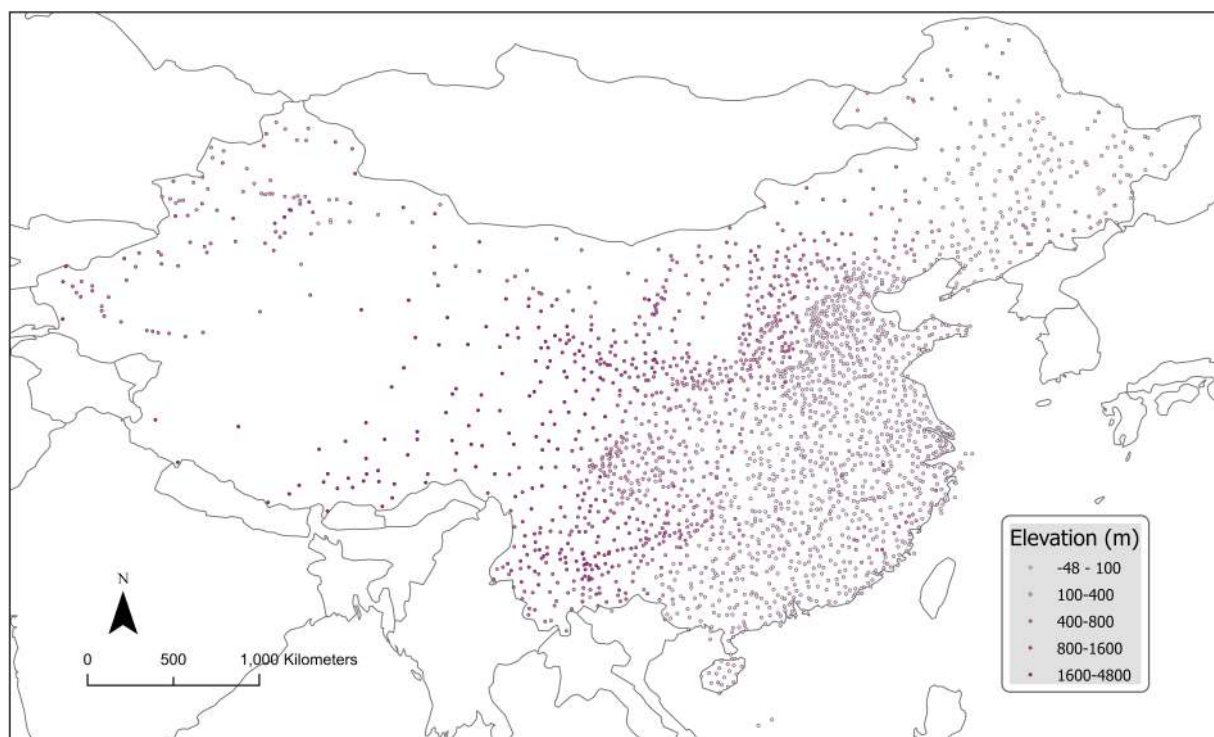


FIGURE 1

Spatial distribution of weather stations. Every dot represents a weather station, and colors represent elevation classes. Coordinates are rounded to the 2nd decimal place. The administrative boundary follows the United Nations Geospatial (2023). The boundaries are derived for cartographic visualization purposes only. The boundaries shown do not imply official endorsement or acceptance. The map extent is optimized for displaying weather station distribution and does not represent any position on geographic coverage.

TABLE 1 Number of weather stations by start and end record years in China.

Record start	Record end year					
Year	1984	1986	1987	1989	1990	Total
1961	1	1	1	2	2,052	2,057
1962	0	0	0	2	34	36
1963	0	0	1	1	19	21
1964	0	0	0	0	15	15
1965	0	0	0	0	26	26
1966	0	0	0	1	22	23
1967	0	0	0	0	19	19
1968	0	0	0	0	4	4
1969	0	0	0	0	4	4
1970	0	0	0	0	15	15
1971	0	0	1	0	28	29
1972	0	0	0	0	33	33
1973	0	0	0	1	12	13
1974	0	1	0	0	19	20
1975	0	0	0	0	8	8
1976	0	0	0	0	4	4
1977	0	0	0	2	9	11
1978	0	0	0	0	12	12
1979	0	0	0	0	9	9
1980	0	0	0	0	5	5
Total	1	2	3	9	2,349	2,364

Climate variables were calculated as both monthly normal values (Table 1) for comparisons at the normal-period level and as monthly values for validating time-series predictions from ClimateAP and WorldClim for the period January 1961–December 1990.

2.2 Scale-free climate data from ClimateAP

ClimateAP v3.1 was employed to generate climate data based on the latitude, longitude, and elevation of weather stations' locations. Two sets of climate data were generated: (1) average monthly climate data for the 30-year normal period of 1961–1990; and (2) monthly time-series climate data spanning from January 1961 to December 1990. Each variable in the ClimateAP dataset corresponds to a climate value calculated from the observed daily or monthly climate.

2.3 Gridded climate data from WorldClim

The coordinates of the weather stations were also used to extract WorldClim climate data using the simple extraction method. We used WorldClim1.4 as gridded data, from which the average climate of the 30-year climate normal period (1961–1990) was extracted, and WorldClim2.1, from which monthly climate data for all months between January 1961 and December 1990 were extracted. We attempted to use the latest version for all purposes in this study, but WorldClim2.1 does not have the data for the corresponding period of 1961–1990 at the spatial resolution of 30 arc seconds. The finest spatial-resolution data for

the normal period from both versions were used: 30 arc seconds for the 30-year average and 2.5 arc minutes for the time-series data.

2.4 Presence/absence data for two representative tree species

The coordinates of occurrence data for *C. lanceolata* and *P. massoniana* were obtained from the same source as Wang, Wang et al. (2016). A total of 2,906 records of presence data for *C. lanceolata* (Figure 2) and 16,020 records of presence data for *P. massoniana* (Figure 3) were obtained from the digital version of the Vegetation Map of China (1:1000,000) provided by “Environmental and Ecological Science Data Center for West China, National Natural Science Foundation of China.” It is a regional dataset that provides high-quality tree species occurrence data. Presence and pseudo-absence points were generated following Wang, Wang et al. (2016). Pseudo-absence points were selected only within 2° of both latitude and longitude from the geographic range of the presence points.

Mean values of monthly temperature and precipitation variables for the normal period were generated with the coordinates of the two tree species' presence and pseudo-absence points. When retrieving climate data from ClimateAP, the elevation of the occurrence coordinates was first filled using the Digital Elevation Model (DEM) grid centers at SRTM 90 m spatial resolution. Then, any missing elevation was filled with DEM at a 7.5 arc-second spatial resolution (Danielson and Gesch, 2011).

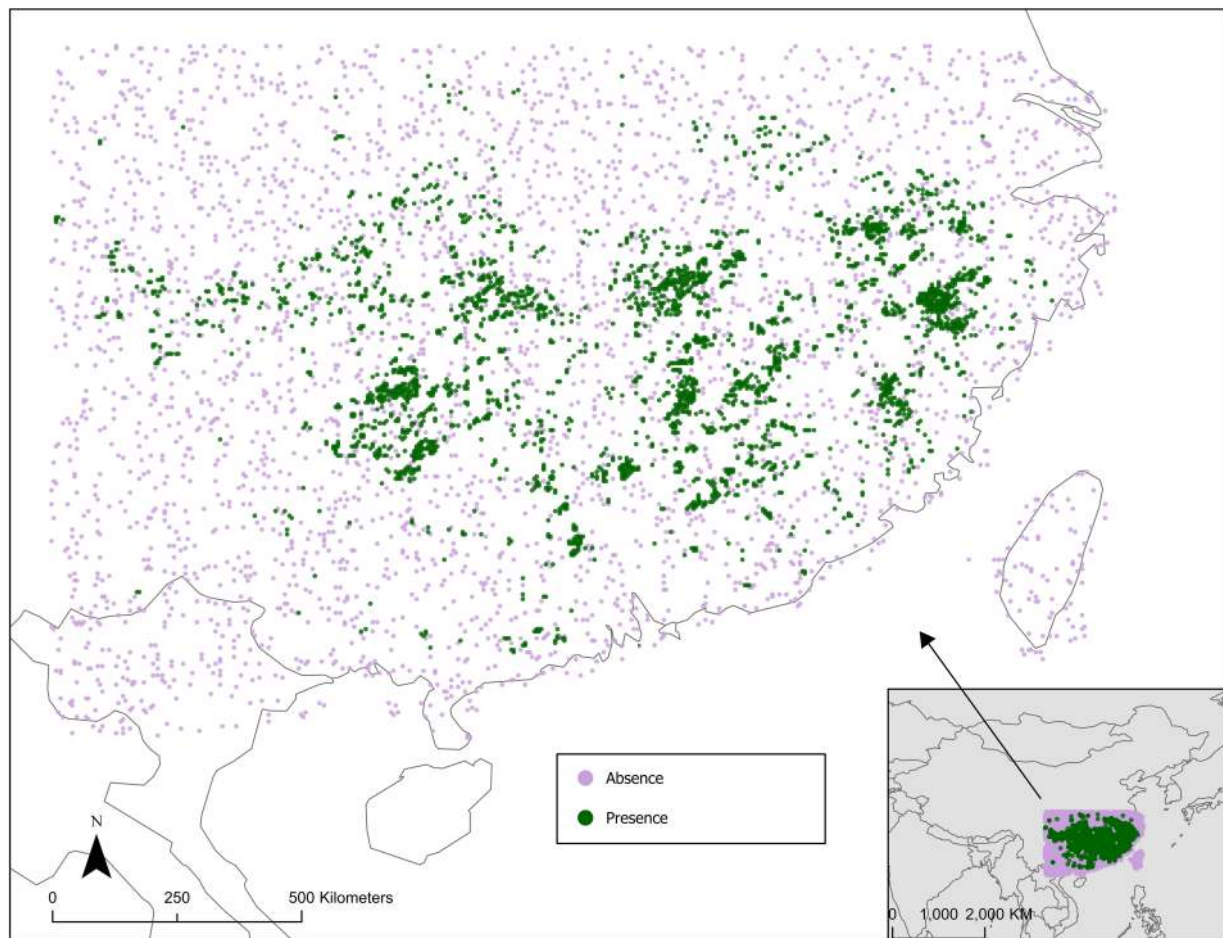


FIGURE 2

Spatial distribution of present-absent data for *Cunninghamia lanceolata* used in training and validating random forest models built with ClimateAP and WorldClim data.

2.5 Statistical analysis

Figure 4 provides an overview of the three analytical approaches in this study. Each part tests one of the three hypotheses. This structured methodology facilitates a comprehensive understanding of the interplay between observed climate and climate modeling predictions.

To test the first hypothesis—the accuracy of ClimateAP's predictions at weather stations, we employed Ordinary Least Squares (OLS) regressions $y = \beta x + \epsilon$ (Wang Hamann et al., 2016; Ye et al., 2023). These observed climate (y) was regressed on predictions generated by ClimateAP v3.1 (x) based on the weather stations' coordinates. For each variable, we reported the adjusted R-squared ($adj.R^2$), MAE, estimated coefficient ($\hat{\beta}$), and p -value. A higher $adj.R^2$ indicate greater variance in observed values explained, lower MAE suggests smaller discrepancy between observed and predicted values, and lower p -values denote statistical significance, and the estimated coefficient ($\hat{\beta}_1$) closer to 1 reveals a 1:1 correlation between per unit of change in the predicted and observed climate.

We conducted the regression analyses for the 30-year normals of 36 monthly climate variables ($Tmax01 - Tmax12$, $Tmin01 - Tmin12$, and $PPT01 - PPT12$). Furthermore, we repeated these regressions for each year from 1961 to 1990, using these 36 monthly variables. We then averaged $adj.R^2$ and MAE values over the 30-year period.

We ran the same set of regressions comparing observed climate data to WorldClim's predictions, evaluating the $adj.R^2$ value, MAE, and estimator. Subsequently, we compared these statistical results with the corresponding statistics from the ClimateAP regressions, enabling a direct comparison of their respective accuracies. WorldClim's 30-year-average climate prediction data are available at four different spatial resolutions. Our regressions include the WorldClim data at the highest available resolution.

To test the second hypothesis, which assesses whether ClimateAP improves accuracy in complex terrain vs. standard 30" grids, we classified the landforms at weather station locations into two groups: plains and complex terrain. We computed the standardized Topographic Position Index (TPI) using a digital elevation model (DEM) at 7.5 arc seconds, ArcGIS Pro 2.9 at dual grid sizes: 5×5 and 11×11 (Weiss, 2001). Considering the varying effectiveness of spatial interpolation in predicting climate in different terrains, we simplified the classification from ten classes in Weiss (2001) to two broader classes.

Plain landscapes, where spatial interpolation performs more precisely due to their relative homogeneity, were grouped. Similarly, complex landscapes, characterized by higher spatial heterogeneity and consequently more significant interpolation errors, formed the other category (Table 2).

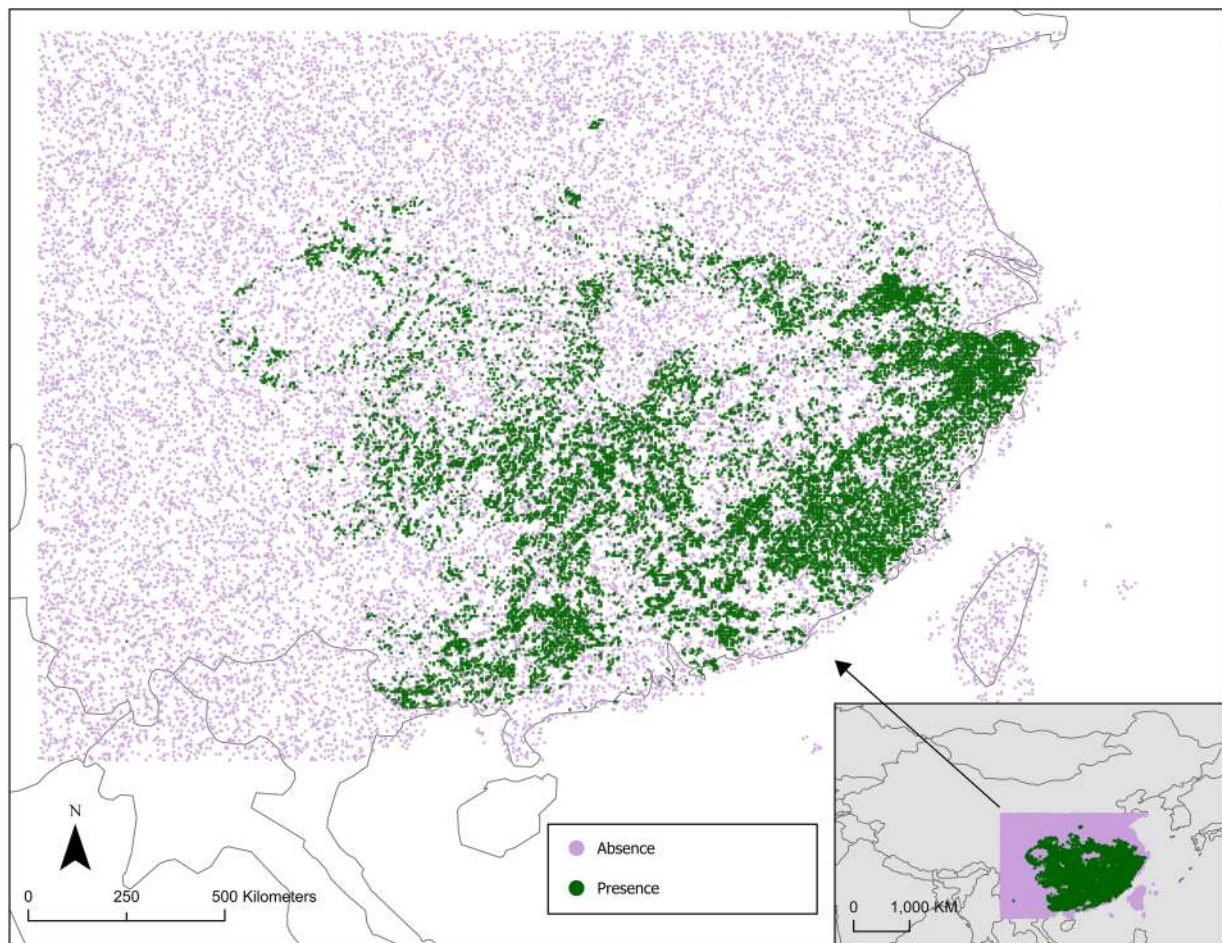


FIGURE 3

Spatial Distribution of present-absent data for *Pinus massoniana* used in training and validating random forest models built with ClimateAP and WorldClim data.

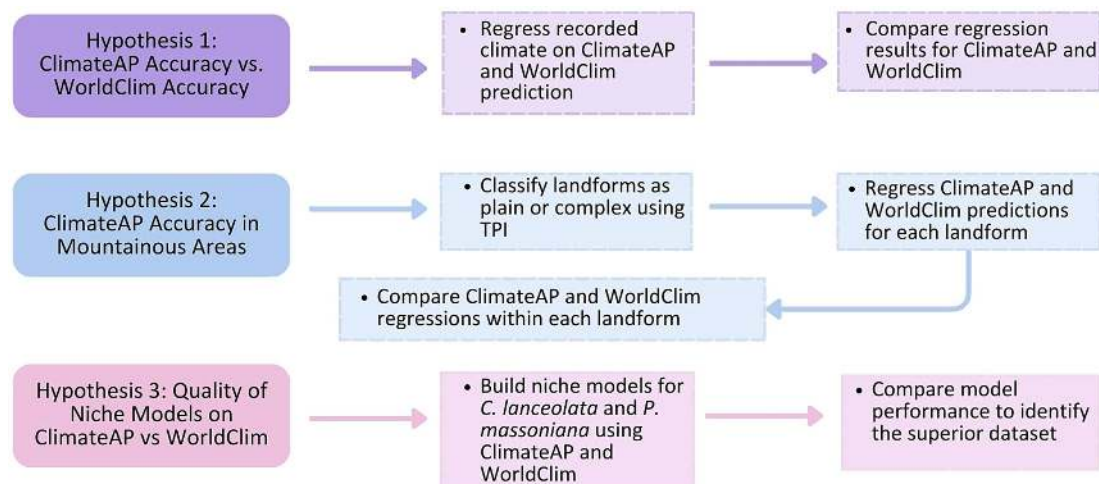


FIGURE 4

Statistical analysis framework for hypothesis testing.

To test the third hypothesis, we used the random forest modeling algorithm to build climatic niche models to predict the occurrence of *C. lanceolata* and *P. massoniana*. From the pool of 2,906 presence and 115,265 absence data points (Figure 4), we first compressed the sample

to a 1.10 absence-to-presence ratio, then allocated 70% as the training set and the remaining 30% as the validation set. This division is not a standard k-fold cross-validation because of the bagging algorithm inherent to random forest models, which obviates the need for

TABLE 2 Landscape classification method using the Topographic Position Index and slope.

Landform classification	TPI 5	TPI 11
Plain landforms	$-\sigma TPI_5 < TPI_5 < \sigma TPI_5$	$-\sigma TPI_{11} < TPI_{11} < \sigma TPI_{11}$
Complex landforms	$TPI_5 = \text{others}$	$TPI_{11} = \text{others}$

The landform classification uses the Topographic Position Index that is standardized using local TPI standard deviations (σTPI_5 and σTPI_{11}). TPI_5 , TPI_{11} and the standard deviations are calculated using a DEM with dual grid sizes: 5×5 and 11×11 , based on Weiss (2001).

cross-validation. Sample points of *P. massoniana* were compressed to 15,000 presence and 20,000 absence, then allocated to the training set and validation set following the same method. Monthly climate data of 30-year averages (1961–1990) from ClimateAPv3.1 and WorldClim1.4 were used to build the two separate models. My evaluation entailed assessing out-of-bag (OOB) errors, predictors' and users' accuracy on the training dataset, alongside model performance on the validation dataset, gauged by the area under the receiver operating characteristic curve (AUC-ROC) and Kappa statistics. AUC-ROC closer to 1 signifies a perfect model, while 0.5 signifies a random guess. Kappa statistics, on the other hand, is a measure of classification accuracy normalized by the imbalance of the classes in the data. It considers the possibility of the agreement occurring by chance; thus, it is a more robust measure than simple accuracy. A high Kappa value indicates better agreement between predictions and observations. The comparison of these metrics allows us to determine if the ClimateAP-based model demonstrates better performance than the WorldClim-based model.

3 Result

3.1 The accuracy of ClimateAP vs. WorldClim

Both ClimateAP and WorldClim achieved high overall accuracy, but ClimateAP consistently outperformed WorldClim across all monthly variables. Averaged over 1961–1990, ClimateAP's $adj.R^2$ exceeded 0.95 for all variables, including precipitation (mean 0.9805 ± 0.0144), T_{max} (0.9670 ± 0.0167), and T_{min} (0.9845 ± 0.0043) (Figure 5A). By contrast, WorldClim's $adj.R^2$ was lower and more variable. Its precipitation predictions $adj.R^2$ reached an average of 0.9078 ± 0.0485 , and the maximum and minimum temperatures were 0.9659 ± 0.0182 and 0.9786 ± 0.0060 . ClimateAP also showed substantially and constantly lower MAE values than WorldClim across all monthly variables, with mean $MAE = 3.13 \pm 2.02$ mm for PPT, 0.486 ± 0.024 °C for T_{max} , and 0.495 ± 0.065 °C for T_{min} , vs. WorldClim's 8.88 ± 5.33 mm for PPT, 0.587 ± 0.048 °C for T_{max} , and 0.672 ± 0.105 °C for T_{min} (Figure 5B). Slope estimates $\hat{\beta}_1$ for ClimateAP were closer to 1.00 with slightly lower monthly variance in precipitation (PPT 1.0158 ± 0.0159) and higher in temperature (T_{max} 0.9762 ± 0.0294 ; T_{min} 0.9891 ± 0.0142) than WorldClim (PPT 1.0076 ± 0.0494 ; T_{max} 0.9800 ± 0.0079 ; T_{min} 0.9906 ± 0.0055), showing similarly minimal proportional bias (Supplementary Table S2).

Seasonal trends were evident in both ClimateAP and WorldClim datasets. Cooler months (November–May for precipitation and November–March for temperatures) yield higher $adj.R^2$ and lower MAE than warmer months. ClimateAP's advantage was most

pronounced in summer. For July precipitation, ClimateAP's $adj.R^2$ was about 0.95 vs. WorldClim's being 0.81 (Figure 6), and MAE was 6.42 mm compared to 18.27 mm (WorldClim). Similarly, ClimateAP's T_{max} MAE fell to 0.37 °C in warm months (peaking at 0.61 °C in cold months), while WorldClim's T_{max} MAE ranged from 0.45 °C (winter) to 0.75 °C (summer). Seasonal trends in T_{min} mirrored those in T_{max} , but with generally lower errors. Examination of the regression slopes ($\hat{\beta}_1$) revealed mixed results (Figure 5c). For PPT, ClimateAP varied closer to 1 than WordClim. For T_{max} , ClimateAP remained closer to 1 from January to April and from November to December, then fell below WorldClim's from May through September (e.g., July: 0.943 vs. 0.978). For T_{min} , ClimateAP remained closer to 1 from January to March and from September to December, then fell lower than WorldClim's from April through August (e.g., July: 0.962 vs. 0.997). All regressions were significant ($p < 0.001$), and slopes remained within ± 0.08 of unity (ClimateAP: 0.93–1.04; WorldClim: 0.94–1.08; Figure 5), showing minimal systematic bias over 30-year normals.

The results of the year-specific climate regressions showed a similar trend, with ClimateAP constantly outperforming WorldClim in terms of both $adj.R^2$ values and MAE for all monthly climate variables (Figure 7). For temperatures, ClimateAP had $adj.R^2$ values varying between 93.15% (T_{max05}) and 98.39% (T_{min01}) and MAE varying between 0.57 (T_{min07}) to 0.88 (T_{min12}). While WorldClim had $adj.R^2$ values varying between 90.87% (T_{max05}) and 98.11% (T_{min12}), and MAE varying between 0.79 (T_{max06}) and 1.00 (T_{min02}). The average difference in temperature between the two data sources ranged from 0.07 (T_{max05}) to 0.28 (T_{max12}) for maximum temperature, and 0.02 (T_{min09}) to 0.36 (T_{min12}) for minimum temperature. The estimated slope ($\hat{\beta}_1$) varied seasonally. For T_{max} , ClimateAP's $\hat{\beta}_1$ was closer to 1.00 than WorldClim's in March, April, and November, but dropped below both 1.00 and WorldClim's from May through September. For T_{min} , ClimateAP's $\hat{\beta}_1$ remained nearer unity in March, April, and August–November, then fell below WorldClim's and below 1.00 in the other months.

For precipitation regressions, ClimateAP had $adj.R^2$ varying between 0.68 (PPT07) and 0.92 (PPT02) and MAE varying between 4.13 (PPT01) and 41.68 (PPT07). In contrast, WorldClim had $adj.R^2$ between 0.60 (PPT07) and 0.87 (PPT02), and MAE between, 4.95 (PPT01) and 46.83 (PPT07). There was an average difference of 5.15 mm in predicting July's precipitation every year from 1961 to 1990 between WorldClim and ClimateAP. From January through April, ClimateAP's $\hat{\beta}_1$ exceeded WorldClim's and was closer to 1.00. In the remaining months, WorldClim's $\hat{\beta}_1$ was nearer unity.

3.2 Prediction accuracy by landform

On complex landforms, ClimateAP consistently outperformed WorldClim in terms of $adj.R^2$ and MAE values across all temperature and precipitation variables (Figure 8). Mean monthly $adj.R^2$ rose by

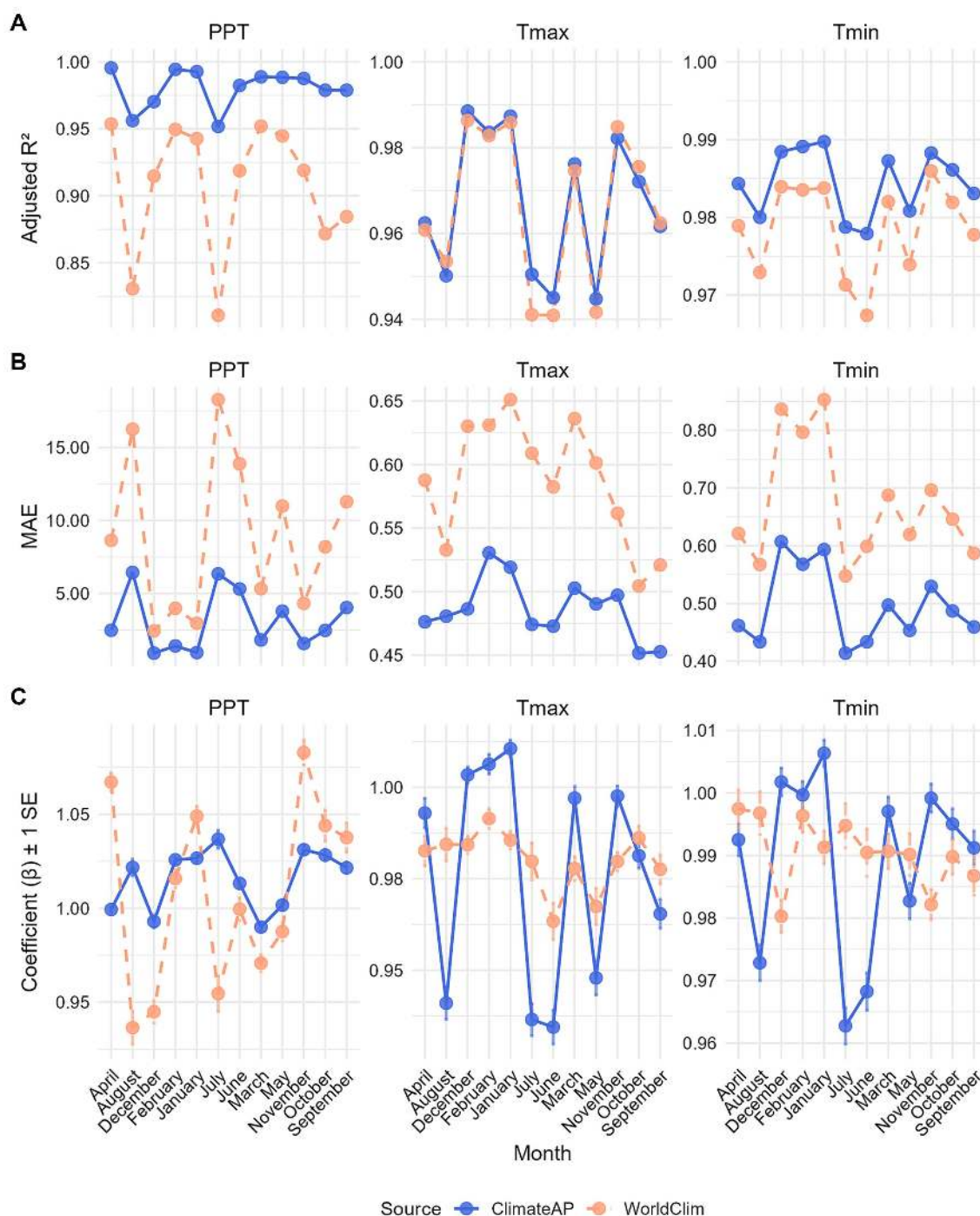


FIGURE 5

Comparisons in prediction accuracies between ClimateAP and WorldClim for monthly maximum temperature (Tmax), monthly minimum temperature (Tmin), and monthly precipitation (PPT) in terms of adjusted R-squared (A), MAE (B), and estimated coefficient β_1 (C) for the 1961–1990 normal period.

approximately 3% for maximum temperature, 2.5% for minimum temperature, and 9% for precipitation under ClimateAP relative to WorldClim, while MAE declined by 40–60% for both temperature metrics and by 58–72% for precipitation. Regression slopes (β_1) remained within $\pm 2\%$ of unity year-round for ClimateAP vs. deviations up to 8% for WorldClim. However, mixed results were shown on plain landforms. ClimateAP outperformed WorldClim for monthly minimum temperatures but did not for maximum temperatures (Figure 8). Although on plain landforms, ClimateAP's precipitation MAE was still lower than those of WorldClim by 48–68%, temperature MAE's advantages were limited (5–18%), and WorldClim marginally outperformed ClimateAP

for Tmax between April and August. For regression slopes, two data sources showed similar results for precipitation, while ClimateAP showed higher values in winter months and lower values in summer months than WorldClim. Thus, while ClimateAP consistently demonstrated strength in complex landscapes, it showed seasonal mixed results on plain landforms.

3.3 Climate niche model for tree species

Two sets of RF models for both *C. lanceolata* and *P. massoniana* showed that models built with ClimateAP data demonstrated higher

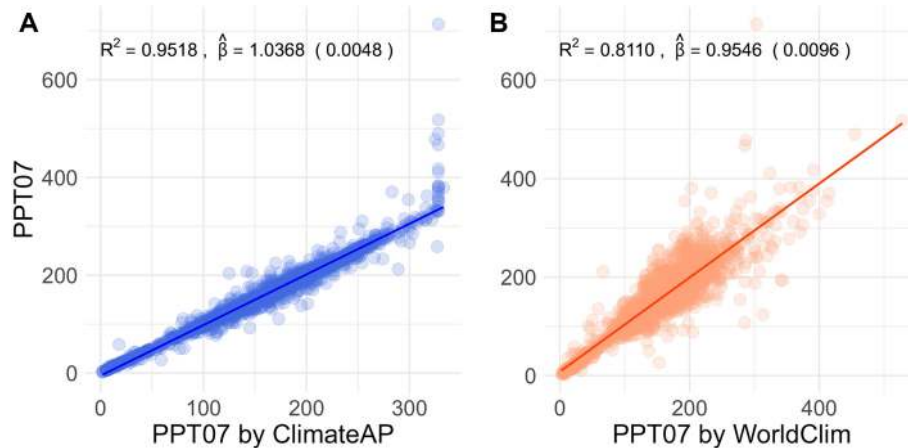


FIGURE 6

Scatter plots of observations of monthly accumulated precipitation in July against predictions (the trend lines) from ClimateAP (A) and WorldClim (B).

prediction accuracy. For *C. lanceolata*, the ClimateAP-based model achieved a lower OOB error rate (17.70%) compared to the WorldClim model (19.23%). Classification errors for absence and presence data were also lower for the ClimateAP model (21.16 and 13.89%) than the WorldClim model (22.76 and 15.55%). In addition, the ClimateAP model demonstrated higher AUC (0.90 vs. 0.88), higher precision (0.84 vs. 0.83), and identical recall (0.78 vs. 0.78) (Table 3). Similarly, the RF models for *P. massoniana* built with ClimateAP data also showed slightly lower OOB error (15.67% vs. 17.89%). However, differences in other terms were less pronounced than for *C. lanceolata*. Due to missing data points, the final presence/absence ratios were 1:1.10 and 1:1.04 for *C. lanceolata*, and 1:1.33 and 1:1.06 for *P. massoniana* in ClimateAP and WorldClim models, respectively.

4 Discussion

This study evaluated the accuracy of scale-free climate data generated by ClimateAP compared with widely used gridded climate data from WorldClim for locations in China for the first time. The results supported our three hypotheses and provided insights into the accuracy of these datasets for ecological modeling, particularly in the Asia-Pacific (AP) region. Our findings contribute to the growing body of research on climate data accuracy and its implications for ecological modeling (Fick and Hijmans, 2017; Wang, Hamann et al., 2016), particularly to enhance ecological modeling in mountainous regions where forests mostly occupy.

4.1 Prediction accuracy of ClimateAP vs. WorldClim

Overall, data from both sources showed high prediction accuracy against station observations for 1961–1990 monthly normals, a critical period for establishing climate normals and building ecological models using climate normals (Harris et al., 2020). On average across all 12 months, ClimateAP outperforms WorldClim, supporting Hypothesis 1. ClimateAP's mean *adj. R*²

for precipitation and temperature values are higher than those of WorldClim. Likewise, ClimateAP's MAE for these variables remains well below that of WorldClim. Although for four late-summer/early-autumn months (Tmax in August–November), WorldClim's adjusted *R*² slightly exceeds ClimateAP's, ClimateAP maintains lower MAEs in those months. The superior performance of ClimateAP, which produces climate data for specific locations, aligns with previous studies that have highlighted the importance of high-resolution climate data in complex terrain (Daly et al., 2008; Wang et al., 2012).

ClimateAP's regression slopes ($\beta^{\wedge}$) drop below 1 and WorldClim's from May through September for Tmax and April through August for Tmin, reflecting the dynamic local adjustment of PRISM-based smoothing of 30-year normals, which dampens monthly extremes (Daly et al., 2008; Wang et al., 2017). In contrast, WorldClim's thin-plate spline interpolation retains greater curvature, yielding slopes closer to 1.00 in summer. In other months, however, ClimateAP's $\beta^{\wedge}$ exceeds WorldClim's and remains nearer to 1, indicating a more accurate one-to-one correspondence.

The scale-free climate data from ClimateNA or ClimateAP have been tested against observations from weather stations (Wang et al., 2017; Wang Hamann et al., 2016), but rarely tested at the monthly level for individual years. The higher *adj. R*² values, and lower MAE for ClimateAP across all monthly climate variables except for four monthly maximum temperatures show better performance over WorldClim. This is particularly evident in precipitation predictions, where ClimateAP's *adj. R*² values exceed 0.95, compared to WorldClim's more variable range of 0.81 to 0.95. These findings are consistent with other regional studies that have emphasized the challenges in accurately modeling precipitation patterns (Sun et al., 2018; Zhang et al., 2011). ClimateAP's superior precipitation predictions can be largely attributable to its use of the PRISM data (Daly et al., 2008), which was considered to have exceptional quality (Wang et al., 2017). The superior performance of ClimateAP during summer months is especially noteworthy, as this period is critical for many ecological processes and biodiversity patterns (Piao et al., 2019). The scatter plots for July's precipitation further illustrate ClimateAP's closer alignment with observed data, highlighting its robustness in capturing climate variability.

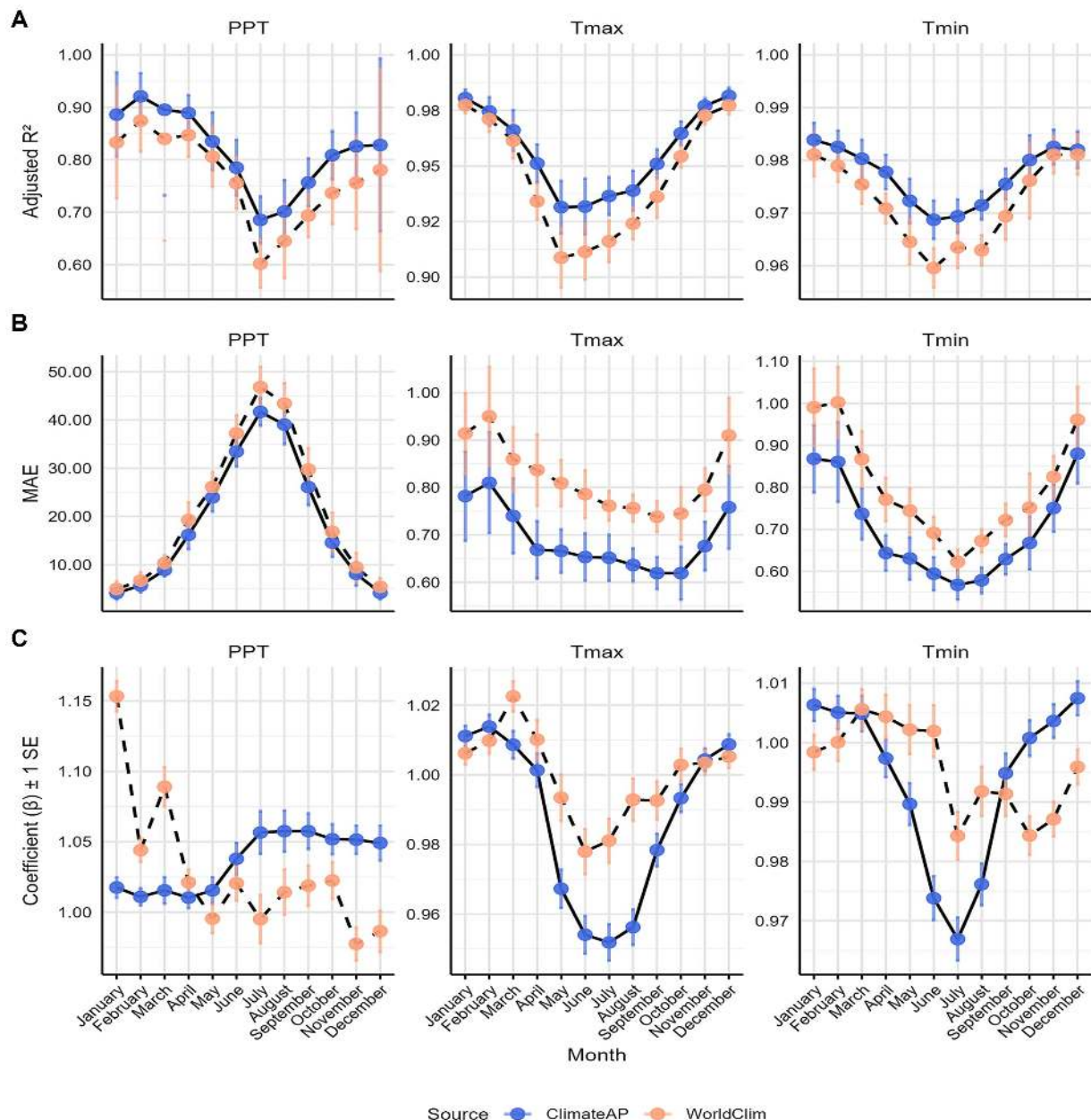


FIGURE 7

Comparisons in prediction accuracies between ClimateAP and WorldClim for monthly maximum temperature (Tmax), monthly minimum temperature (Tmin), and monthly precipitation (PPT) in terms of Adjusted R-squared (A), MAE (B), and estimated coefficient β_1 (C) for individual years between 1961 and 1990. Error bars show the standard deviation of R-squared and MAE values across all years, and the averaged standard error for the estimated coefficient.

4.2 Prediction accuracy by landform

The analysis of prediction accuracy across different landforms reveals that ClimateAP consistently outperforms WorldClim in complex terrains. This supports our hypothesis that ClimateAP's scale-free approach is better suited for regions with significant topographic variation, where microclimates can differ substantially within short distances (Daly et al., 2008; Wang Hamann et al., 2016). This finding aligns with previous studies demonstrating that topographic complexity significantly influences local climate patterns and the accuracy of climate predictions (Dobrowski, 2011; Meineri and Hylander, 2017).

However, in plain landforms, the superiority of ClimateAP was less pronounced. This pattern is consistent with other studies

showing that interpolation methods perform well in areas with relatively uniform terrain (Fick and Hijmans, 2017). This demonstrates the overall superiority of ClimateAP, which is mainly attributable to its scale-free downscaling approach in mountainous regions where climate varies significantly over small distances (Wang et al., 2012). Traditional gridded averages, such as those used in WorldClim, often struggle to capture this fine-scale variation due to the averaging effect across grid cells (Maclean et al., 2019).

The superior performance of ClimateAP in complex landforms can be primarily attributed to its dynamic local downscaling algorithm. This sophisticated approach enables the estimation of location-specific lapse rates through local ordinary least squares (OLS)

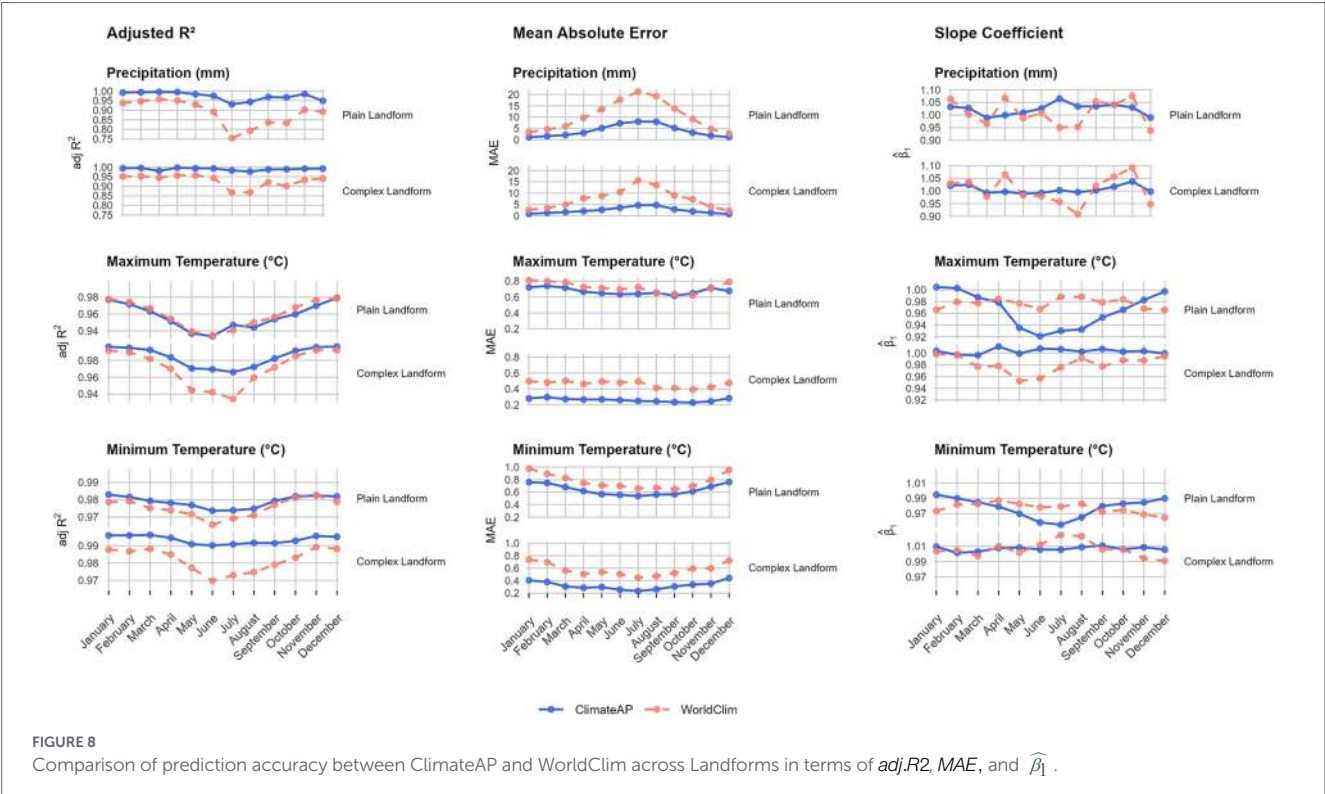


TABLE 3 Out-of-bag error, area under curve, precision, and recall in each random forest model.

Metrics	<i>Cunninghamia lanceolata</i>		<i>Pinus massoniana</i>	
	ClimateAP	WorldClim	ClimateAP	WorldClim
OOB	17.70%	19.23%	15.67%	17.89%
AUC-ROC	0.8967	0.8836	0.9045	0.9033
Kappa	0.6220	0.6105	0.6558	0.6487
Precision	0.8394	0.8289	0.8694	0.8607
Recall	0.7800	0.7778	0.7831	0.7860

regression based on adjacent grid cells (Wang, Hamann et al., 2016). The method's effectiveness stems from its ability to account for varying lapse rates across different terrains while generating point-specific (scale-free) estimates rather than grid cell averages (Minder et al., 2010).

Given that elevation is a primary driver of microclimate variation in mountainous regions (Dobrowski, 2011; Körner, 2007), ClimateAP's point-specific climate estimates more accurately represent the actual environmental conditions experienced by organisms. Notably, ClimateAP achieves this performance despite using a lower resolution baseline data (2.5 arc minutes) compared to WorldClim's 30 arc seconds (Hijmans et al., 2005). This suggests potential for further improvement in prediction accuracy if higher-resolution baseline data were incorporated in future versions.

However, previous studies have identified limitations in the dynamic local downscaling algorithm's effectiveness for precipitation variables (Wang et al., 2012). The marked superiority of ClimateAP in precipitation prediction likely stems from the high-quality precipitation data provided by PRISM (Parameter-elevation Relationships on Independent Slopes Model) (Daly et al., 2008). This highlights the

critical importance of baseline data quality in climate modeling accuracy.

4.3 Impact of climate data source on climate niche models for tree species

The random forest models developed for *C. lanceolata* and *P. massoniana* demonstrated comparable overall performances using either ClimateAP or WorldClim data, but with notable differences in specific metrics. Models built with ClimateAP data consistently showed superior performance across multiple evaluation metrics, including lower out-of-bag (OOB) error rates, higher Area Under the Curve (AUC), Kappa, and precision values (Elith and Leathwick, 2009). The superior performance of ClimateAP-based models aligns with theoretical expectations, as more accurate climate data should lead to more reliable species distribution predictions (Peterson, 2011). The improvement in model performance can be attributed to ClimateAP's scale-free climate estimates, which better capture the microclimate variations that influence species distributions, particularly in topographically complex regions (Maclean et al., 2019; Suggitt et al., 2018).

4.4 Limitations and future research

The weather stations used in this study may have contributed to the development of either the WorldClim or the PRISM datasets, or both. Thus, the prediction accuracies may be somewhat overestimated. However, our objective was to assess the relative performance of the two data sources. Supporting this approach, the superiority of the dynamic local downscaling algorithm has been observed with independent weather stations at a local scale (Huan-Yu et al., 2018; Ye et al., 2023). An additional methodological limitation lies in comparing ClimateAP and WorldClim outputs. As ClimateAP generates point-based data while WorldClim provides gridded data, it is difficult to determine how much of the improved prediction accuracy stems from the different interpolation methods vs. the inherent higher precision of point-based climate representations. Future studies could address this by developing methods to decompose these two sources of improvement.

C. lanceolata and *P. massoniana* served as the model species in this study, selected for their economic significance, susceptibility to climate change, and the availability of substantial observed presence data. While these factors affirm its suitability for this analysis, the results' applicability may vary with different species due to unique traits and adaptations. Therefore, additional sensitivity analyses with diverse species could clarify how climate data sources impact niche modeling. Future research should validate climate data sources across varied species or ecosystems and regions to strengthen their credibility. Identifying the most suitable data sources for different models could provide valuable insights for ecological modelers and enhance understanding of climate change impacts.

5 Conclusion

This study provides strong evidence for the accuracy and utility of scale-free climate data generated by ClimateAP for ecological modeling in China compared with WorldClim. Through rigorous comparisons with weather station observations and the widely used WorldClim gridded dataset, we demonstrated that ClimateAP consistently produced highly accurate temperature and precipitation estimates across all months and years evaluated. This performance was especially pronounced in topographically complex regions, highlighting the value of ClimateAP's downscaling approach in capturing fine-scale climate variation that is often missed by gridded averages. Moreover, we showed that the improved accuracy of ClimateAP's climate data translated into better predictions of species distributions. Ecological niche models for two representative tree species built with ClimateAP predictors outperformed equivalent models using WorldClim data, with lower error rates and higher accuracy metrics. This demonstrates the potential for ClimateAP to enhance the quality and reliability of ecological assessments and climate change impact projections in the Asia-Pacific region. Our study provides a robust validation of ClimateAP's accuracy and utility in China and highlights its potential to enhance ecological modeling and climate change impact assessments across the Asia-Pacific region.

Data availability statement

Data from ClimateAP (<https://climateap.net/>) and WorldClim (<https://www.worldclim.org/>) is publicly available. Chinese weather station data is restricted to users with existing permissions; station identifiers and species distribution data is available upon request from the corresponding author.

Author contributions

JJ: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. LZ: Resources, Writing – review & editing. WX: Resources, Writing – review & editing. H-YL: Data curation, Resources, Writing – review & editing. XX: Resources, Writing – review & editing. SZ: Project administration, Writing – review & editing. TW: Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by the Asia-Pacific Network for Sustainable Forest Management and Rehabilitation to Tongli Wang, UBC (Project ID: 2020P4-UBC).

Acknowledgments

We also would like to thank users of ClimateAP from the Asia-Pacific region for their help with the data acquisition.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy,

including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product

that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fclim.2026.1783859/full#supplementary-material>

References

- Abdulwahab, U. A., Hammill, E., and Hawkins, C. P. (2022). Choice of climate data affects the performance and interpretation of species distribution models. *Ecol. Model.* 471:110042. doi: 10.1016/j.ecolmodel.2022.110042
- Bobrowski, M., and Udo, S. (2017). Why input matters: selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. *Ecol. Model.* 359, 92–102. doi: 10.1016/j.ecolmodel.2017.05.021
- Daly, C., Gibson, W., Taylor, G., Johnson, G., and Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Clim. Res.* 22, 99–113. doi: 10.3354/cr022099
- Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., et al. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.* 28, 2031–2064. doi: 10.1002/joc.1688
- Danielson, J. J., and Gesch, D. B. (2011). Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey open-file report 2011–1073 (pp. 1–26).
- De Clercq, E. M., Leta, S., Estrada-Peña, A., Madder, M., Adehan, S., and Vanwambeke, S. O. (2015). Species distribution modelling for *Rhipicephalus microplus* (Acari: Ixodidae) in Benin, West Africa: comparing datasets and modelling algorithms. *Prev. Vet. Med.* 118, 8–21. doi: 10.1016/j.prevetmed.2014.10.015
- Dobrowski, S. Z. (2011). A climatic basis for microrefugia: the influence of terrain on climate. *Glob. Change Biol.* 17, 1022–1035. doi: 10.1111/j.1365-2486.2010.02263.x
- Elith, J., and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* 40, 677–697. doi: 10.1146/annurev.ecolsys.110308.120159
- Feng, L., Sun, J., Shi, Y., Wang, G., and Wang, T. (2020). Predicting suitable habitats of *Camptotheca acuminata* considering both climatic and soil variables. *Forests* 11:891. doi: 10.3390/F11080891
- Feng, L., Tian, X., El-Kassaby, Y. A., Qiu, J., Feng, Z., Sun, J., et al. (2022). Predicting suitable habitats of *Melia azedarach* L. in China using data mining. *Sci. Rep.* 12:12617. doi: 10.1038/s41598-022-16571-y
- Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315. doi: 10.1002/JOC.5086
- Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations – the $\times$ CRU TS3.10 dataset. *Int. J. Climatol.* 34, 623–642. doi: 10.1002/joc.3711
- Harris, I., Osborn, T. J., Jones, P., and Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* 7:109. doi: 10.1038/s41597-020-0453-3
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146, 1999–2049. doi: 10.1002/qj.3803
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276
- Huan-Yu, L. I., Jer-Ming, H. U., Tze-Ying, C. H. E. N., Guangyu, W. A. N. G., and Tongli, W. A. N. G. (2018). A dynamic downscaling approach to generate scale-free regional climate data in Taiwan. *Taiwania*. 63.
- Jiang, Y., Kim, J. B., Still, C. J., Kerns, B. K., Kline, J. D., and Cunningham, P. G. (2018). Inter-comparison of multiple statistically downscaled climate datasets for the Pacific northwest, USA. *Sci. Data* 5:180016. doi: 10.1038/sdata.2018.16
- Körner, C. (2007). The use of 'altitude' in ecological research. *Trends Ecol. Evol.* 22, 569–574. doi: 10.1016/j.tree.2007.09.006
- Macleán, I. M. D., Mosedale, J. R., and Bennie, J. J. (2019). Microclima: an R package for modelling meso- and microclimate. *Methods Ecol. Evol.* 10, 280–290. doi: 10.1111/2041-210X.13093
- Matsushita, B., Xu, M., Chen, J., Kameyama, S., and Tamura, M. (2004). Estimation of regional net primary productivity (NPP) using a process-based ecosystem model: how important is the accuracy of climate data? *Ecol. Model.* 178, 371–388. doi: 10.1016/j.ecolmodel.2004.03.012
- Meineri, E., and Hylander, K. (2017). Fine-grain, large-domain climate models based on climate station and comprehensive topographic information improve microrefugia detection. *Ecography* 40, 1003–1013. doi: 10.1111/ecog.02494
- Minder, J. R., Mote, P. W., and Lundquist, J. D. (2010). Surface temperature lapse rates over complex terrain: lessons from the Cascade Mountains. *J. Geophys. Res. Atmos.* 115:2009JD013493. doi: 10.1029/2009JD013493
- Peterson, A. T. (2011). *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton University Press.
- Piao, S., Liu, Q., Chen, A., Janssens, I. A., Fu, Y., Dai, J., et al. (2019). Plant phenology and global climate change: current progresses and challenges. *Glob. Change Biol.* 25, 1922–1940. doi: 10.1111/gcb.14619
- Suggitt, A. J., Wilson, R. J., Isaac, N. J. B., Beale, C. M., Auffret, A. G., August, T., et al. (2018). Extinction risk from climate change is reduced by microclimatic buffering. *Nat. Clim. Change* 8, 713–717. doi: 10.1038/s41558-018-0231-9
- Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. (2018). A review of global precipitation data sets: data sources, estimation, and intercomparisons. *Rev. Geophys.* 56, 79–107. doi: 10.1002/2017RG000574
- Thornton, P. E., Thornton, M. M., Mayer, B. W., Wei, Y., Devarakonda, R., Vose, R. S., et al. (2016). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3 (Version 3, p. 0 MB) [NetCDF]*. Oak Ridge, TN: ORNL Distributed Active Archive Center.
- United Nations Geospatial. (2023). United Nations geospatial data, simplified features [shapefile]. Available online at: <https://geoportal.un.org/arcgis/home/item.html?id=f86966528d5943efbdb83fd521dc0943>
- Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS One* 11:e0156720. doi: 10.1371/journal.pone.0156720
- Wang, T., Hamann, A., Spittlehouse, D. L., and Murock, T. Q. (2012). Climate-wna-high-resolution spatial climate data for western North America. *J. Appl. Meteorol. Climatol.* 51, 16–29. doi: 10.1175/JAMC-D-11-043.1
- Wang, T., Wang, G., Innes, J., Nitschke, C., and 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. Manag.* 360, 357–366. doi: 10.1016/j.foreco.2015.08.004
- Wang, T., Wang, G., Innes, J. L., Seely, B., and Chen, B. (2017). ClimateAP: an application for dynamic local downscaling of historical and future climate data in Asia Pacific. *Front. Agric. Sci. Eng.* 4, 448–458. doi: 10.15302/J-FASE-2017172
- Weiss, A. D. (2001). "Topographic position and landforms analysis," in Poster Presentation, ESRI User Conference, (Seattle WA: The Nature Conservancy, Northwest Division).
- Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., Snyder, M. A., Grinnell, J., et al. (2009). Niches, models, and climate change: assessing the assumptions and uncertainties a brief perspective on niches the foundations for thinking about ecological niches were laid. *Proc. Natl. Acad. Sci. USA* 106, 19729–19736. doi: 10.1073/pnas.0901639106
- WMO (2020). WMO Climatological Normals. Available online at: <https://community.wmo.int/en/wmo-climatological-normals> (Accessed November 26, 2023).
- Wootten, A., Smith, K., Boyles, R., Terando, A., Stefanova, L., Misra, V., et al. (2014). Downscaled climate projections for the Southeast United States: evaluation and use for ecological applications. doi: 10.3133/ofr20141190
- Ye, Z., O'Neill, G. A., and Wang, T. (2023). Climate data for field trials: onsite micro stations versus ClimateNA. 550:121509. doi: 10.1139/cjfr-2022-0051
- Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim. Change* 2, 851–870. doi: 10.1002/wcc.147