Mapping 30-m Resolution Bioclimatic Variables During 1991–2020 Climate Normals for Hubei Province, the Yangtze River Middle Reaches

Ruizhen Wang ⓘ, Weitao Chen ⓘ, *Member, IEEE*, Siyang Wan ⓘ, Gaodian Zhou ⓘ, Wenxi He ⓘ, and Lunche Wang ⓘ

Abstract—High-resolution bioclimatic data are crucial to providing fine-scaled insights into biodiversity assessment, forestry, and agricultural management. Existing global bioclimatic datasets often exhibit kilometer-level coarse resolution or have miss the data in recent decades, potentially resulting in the issues of lower spatial accuracy, limited information, and restricted applicability in fine-scaled studies. Hubei Province in Yangtze River Middle Reaches has sparse meteorological stations in high-altitude mountainous areas to map the high-resolution bioclimatic variables directly. This study developed a 30-year averaged bioclimatic dataset for Hubei Province during 1991–2020 at a 30-m spatial resolution, utilizing monthly temperatures and precipitation derived from a downscaling-calibration framework. The downscaling of 1-km resolution climate variables was achieved by using a random forest model with 30-m resolution terrain and spatial covariates. Then, the geographical differential analysis was applied to improve the accuracy of downscaled products by including additional ground data. The mean absolute errors of calibrated monthly maximum, mean, minimum temperatures, and precipitation based on ordinary kriging decreased from 0.74 °C, 0.47 °C, 0.47 °C, and 28.27 mm to 0.43 °C, 0.28 °C, 0.36 °C, and 21.43 mm, respectively. Finally, calibrated climate variables were employed to calculate 19 annual bioclimatic variables, which were subsequently averaged over the 30-year period. The constructed bioclimatic dataset exhibits high overall consistency with the WorldClim dataset according to pixel-based comparison (Spearman correlation coefficients >0.6), with differences mainly attributed to the superior local accuracy of our dataset and climate changes. The dataset will provide fine-scaled, updated, and reliable data supports for local-related studies and decision making.

Index Terms—Bioclimate, geographical differential analysis (GDA), Hubei Province, machine learning (ML), statistical downscaling.

Ruizhen Wang and Siyang Wan are with the Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074, China (e-mail: ruizhen.wang@cug.edu.cn; 1202221952@cug.edu.cn).

Weitao Chen and Gaodian Zhou are with the Faculty of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: wtchen@cug.edu.cn; zhougaodian@cug.edu.cn).

Wenxi He is with the Wuhan Centre of China Geological Survey, Wuhan 430205, China (e-mail: hewenxi.cug@gmail.com).

Lunche Wang is with the Hubei Key Laboratory of Regional Ecology and Environmental Change, School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: wang@cug.edu.cn).

I. INTRODUCTION

CURRENTLY, there is a growing demand for digital spatial climate data on a monthly time step such as monthly mean temperature and monthly total precipitation, as well as their averages over a nominal 30-year period to support agricultural decision making and ecological conservation [1]. Bioclimatic variables, such as mean daily mean air temperatures of the wettest quarter and precipitation amount of the warmest quarter, are specifically designed from these monthly climate variables to capture seasonal trends relevant to the physiological constraints of various species [2], [3], making them more effective in these tasks [4]. Bioclimatic data have been widely applied in species distribution models [5] for invasive species management [6], [7], [8] and assessment of biodiversity [9], [10]. In addition, it provides insights into the impacts of climate on agriculture [11], [12] and forestry [13], [14]. Modern soil digital mapping tasks have also started incorporating bioclimatic data [15], [16], [17], [18] that is more informative than only traditional climate variables in revealing soil formation.

However, the bioclimatic variables used in the aforementioned studies were primarily sourced from the global datasets such as WorldClim [19] and CHELSA [20], which have a resolution of only 30 arc seconds (roughly 1-km spatial resolution). This limitation poses a challenge to local ecological mapping tasks due to the growing importance of fine-gridded climate data ($\leq$ 1 km^2) for indicating detailed environmental variability, especially in regions with complex terrains [21], [22], [23].

Recent studies have attempted to address this limitation by developing the very fine regional bioclimatic dataset. The prevalent approach involves constructing regression models using ground observation climate data along with topographic and spatial covariates, followed by geostatistical methods to correct the differences between predicted and observed values. For instance, the regression kriging was used in constructing a 40-m bioclimatic dataset for Sardinia Island in Mediterranean [23],

and linear regression with thin plate spline (TPS) interpolation was employed to construct a 30-m bioclimatic dataset for Hong Kong [24]. However, applying this method poses challenges in mountainous areas with sparse and low-elevation distributed weather stations, where the lack of high-elevation data requires the model to perform extrapolation for these areas, introducing potential errors and uncertainties [25], [26]. In such cases, downscaling national climate images is a preferred approach, as it aids in filling data gaps by incorporating broader-scale information.

Among the downscaling methods, statistical downscaling has been proved effective to obtain higher resolution climate data [27], [28], [29]. This method involves using regression algorithms to generate high-resolution regional climate data by establishing a statistical correlation between coarser global or national climate products and high-resolution auxiliary data. The 30-m bioclimatic dataset for southern California constructed by California Natural Resources Agency is based on this method (https://data.cnra.ca.gov/dataset/downscaled-climate-grids-at-30m-for-a-variety-of-bioclimatic-variables-over-the-san-j-2001-2099). However, there is often a mismatch between the estimated products and the ground measurements [30], [31]. Errors in estimated climate products, caused by limited availability of ground data or poor satellite performance, can propagate to downscaled data. To address this issue, previous studies included the calibration step by fusing the downscaled products and observation data to minimize the difference [31], [32], [33]. In light of this, the cascade procedure by integrating the statistical downscaling and ground-based calibration may allow the high-resolution products with higher accuracy, but the extent of improvement depends on the availability of observation data.

Hubei Province, situated in the core part of the Yangtze River Middle Reaches, boasts abundant biological resources and contains a global-level biodiversity hotspot and a key area for China's biodiversity [34], [35]. Although the province has established 82 nature reserves for biodiversity preservation, habitat loss due to climate change remains a concern [36], [37]. Due to the absence of high-resolution bioclimatic data in this region, several studies related to species distribution and agricultural planning [36], [37], [38], [39], [40] were conducted by using coarse global bioclimatic datasets like WorldClim, forcing the other high-resolution environmental covariates to be aggregated to this resolution and resulting in the coarse outputs. Moreover, the employed bioclimatic datasets lack data from the latest decades and using the outdated bioclimate data may lead to inaccurate assessments.

Therefore, this study aims to construct a 30-m bioclimatic dataset for Hubei Province, including 19 basic bioclimatic variables from the BIOCLIM package [41] (see Table I) for the latest 1991–2020 climate normals. As Hubei Province faces the challenge of extremely lacking high-elevation weather stations in its mountainous regions, with virtually no available stations in the altitude range of 1000–3000 m, this study utilizes a downscaling-based procedure to provide more reliable climate data for the calculation of bioclimatic variables. The new dataset enhances the spatial resolution of bioclimatic data available in Hubei Province to 30 m, fills the data gaps in the latest climate normal, and improves the data credibility. It will provide fine-scaled, updated, and reliable data supports for the ecological explorations as well as agricultural and forestry managements in the core region of Yangtze River Middle Reaches.

II. Study Site and Available Data

A. Basic Climate Situation and Ground Data

Hubei Province (see Fig. 1) is located in the typical monsoon area of subtropical zone. Except for alpine regions, most regions have a subtropical monsoon humid climate. The average annual temperature here is 15–17 °C. The average annual precipitation in Hubei is between 800 and 1600 mm, showing a decreasing trend from south to north. The precipitation distribution has obvious seasonal changes; it generally achieves the most in summer with rainfall between 300 and 700 mm, and the least in winter with rainfall between 30 and 190 mm.

The ground observation data including monthly maximum temperature, monthly mean temperature, monthly minimum temperature, and monthly total precipitation from 1991 to 2020 were obtained from 82 meteorological stations in Hubei Province. These stations include 5 datum stations, 27 base stations, and 50 ordinary stations. Considering that the construction of the Climatic Research Unit (CRU) and WorldClim datasets included parts of the data from datum and base stations, which are responsible for providing data to international climate organizations [19], [42], the validation stations were all selected from local ordinary stations. In this study, nearly 90% of stations (74 stations in total, five stations lack data on a large number of months) were used for the calibration and nearly 10% of stations (eight stations) were used for the validation.

B. Climate Data

The 1-km monthly minimum temperature, monthly minimum temperature, monthly minimum temperature, and monthly precipitation images from 1991 to 2020 [43], [44], [45], [46] (https://data.tpdc.ac.cn/) were obtained from the downscaled CRU climate data for China mainland by the Delta method [42]. CRU climate data were generated with a spatial resolution of 30 arc seconds based on data from thousands of global weather stations and a function considering altitude, longitude, and latitude [47]. In total, 1440 images (three temperature variables and one precipitation variable in 360 months) were extracted and subset by the extent boundary of Hubei Province.

C. Auxiliary Data

Terrain and spatial variables were usually used in climate downscaling [28], [48], [49]. It has been observed that the inclusion of detailed terrain variables such as slope and aspect makes limited contribution to the accuracy of climate downscaling if they were not considered in original dataset [22]. As the 1-km climate data only considered elevation and spatial position [42], the topographical variable used in this study was only the 30-m resolution elevation obtained from the Shuttle Radar Topography Mission product that can be accessed from the Google Earth Engine platform. The 30-m resolution elevation

TABLE I
DESCRIPTIONS OF 19 BIOCLIMATIC VARIABLES

Symbol	Description	Formula
bio1	Mean annual air temperature (°C)	$bio1 = \dfrac{\sum_{i=1}^{i=12} T_{mean}}{12}$
bio2	Mean diurnal air temperature range (°C)	$bio2 = \dfrac{\sum_{i=1}^{i=12}(T_{max} - T_{min})}{12}$
bio3	Isothermality (%)	$bio3 = \dfrac{bio2}{bio7}$
bio4	Temperature seasonality (°C)	$bio4 = std(\{T_{mean\,1}, T_{mean2}, \dots, T_{mean12}\})$
bio5	Mean daily maximum air temperature of the warmest month (°C)	$bio5 = \max(\{T_{min\,1}, T_{min\,2}, \dots, T_{min\,12}\})$
bio6	Mean daily minimum air temperature of the coldest month (°C)	$bio6 = \min(\{T_{min\,1}, T_{min\,2}, \dots, T_{min\,12}\})$
bio7	Annual range of air temperature (°C)	$bio7 = bio5 - bio6$
bio8	Mean daily mean air temperatures of the wettest quarter (°C)	$bio8 = \dfrac{\sum_{i^*}^{i^*+3} T_{mean\,i^*}}{3},$ $\quad Q_{pre} = \max\left(\left\{\sum_{i=1}^{3} pre_i, \sum_{i=2}^{4} pre_i, \dots, \sum_{i=12}^{2} pre_i\right\}\right)$
bio9	Mean daily mean air temperatures of the driest quarter (°C)	$bio9 = \dfrac{\sum_{i^*}^{i^*+3} T_{mean\,i^*}}{3},$ $\quad Q_{pre} = \min\left\{\sum_{i=1}^{3} pre_i, \sum_{i=2}^{4} pre_i, \dots, \sum_{i=12}^{2} pre_i\right\}$
bio10	Mean daily mean air temperatures of the warmest quarter (°C)	$bio10 = \max\left(\left\{\dfrac{\sum_{i=1}^{3} T_{mean\,i}}{3}, \dfrac{\sum_{i=2}^{4} T_{mean\,i}}{3}, \dots, \dfrac{\sum_{i=12}^{2} T_{mean\,i}}{3}\right\}\right)$
bio11	Mean daily mean air temperatures of the coldest quarter (°C)	$bio11 = \min\left(\left\{\dfrac{\sum_{i=1}^{3} T_{mean\,i}}{3}, \dfrac{\sum_{i=2}^{4} T_{mean\,i}}{3}, \dots, \dfrac{\sum_{i=12}^{2} T_{mean\,i}}{3}\right\}\right)$
bio12	Annual precipitation amount (mm)	$bio12 = \sum_{i=1}^{12} pre_i$
bio13	Precipitation amount of the wettest month (mm)	$bio13 = \max(\{pre_1, pre_2, \dots, pre_{12}\})$
bio14	Precipitation amount of the driest month (mm)	$bio14 = \min(\{pre_1, pre_2, \dots, pre_{12}\})$
bio15	Precipitation seasonality (%)	$bio15 = \dfrac{std(\{pre_1, pre_2, \dots, pre_{12}\})}{mean(\{pre_1, pre_2, \dots, pre_{12}\})} \times 100$
bio16	Precipitation amount of the wettest quarter (mm)	$bio16 = \max\left(\left\{\sum_{i=1}^{3} pre_i, \sum_{i=2}^{4} pre_i, \dots, \sum_{i=12}^{2} pre_i\right\}\right)$
bio17	Precipitation amount of the driest quarter (mm)	$bio17 = \min\left(\left\{\sum_{i=1}^{3} pre_i, \sum_{i=2}^{4} pre_i, \dots, \sum_{i=12}^{2} pre_i\right\}\right)$
bio18	Precipitation amount of the warmest quarter (mm)	$bio18 = \sum_{i^*}^{i^*+3} pre_{i^*},$ $\quad Q_{Tmean} = \max\left(\left\{\dfrac{\sum_{i=1}^{3} T_{mean\,i}}{3}, \dfrac{\sum_{i=2}^{4} T_{mean\,i}}{3}, \dots, \dfrac{\sum_{i=12}^{2} T_{mean\,i}}{3}\right\}\right)$
bio19	Precipitation amount of the coldest quarter (mm)	$bio19 = \sum_{i^*}^{i^*+3} pre_{i^*},$ $\quad Q_{Tmean} = \min\left(\left\{\dfrac{\sum_{i=1}^{3} T_{mean\,i}}{3}, \dfrac{\sum_{i=2}^{4} T_{mean\,i}}{3}, \dots, \dfrac{\sum_{i=12}^{2} T_{mean\,i}}{3}\right\}\right)$

$T_{mean}, T_{max}, T_{min}$ and pre means the monthly mean, maximum, minimum temperature and monthly precipitation, respectively. i means the month and i^* means the first month of an annual quarter (Q) with 3-month interval.

map was resampled to 1 km for model training by using bilinear interpolation, which can preserve the gradients and continuity of pixel values in the image and avoid introducing artifacts. The coordinates (longitude and latitude) at both 30-m and 1-km resolutions were derived from elevation images at respective resolutions.

III. METHODOLOGY

The cascade procedure comprises three main steps: Initially, the 1-km climate variables encompassing monthly mean, maximum, minimum temperatures, as well as precipitation, underwent statistical downscaling using a machine learning (ML) model that utilized altitude, longitude, and latitude derived from 30-m resolution digital elevation model (DEM) data. Subsequently, the geographical differential analysis (GDA) method was employed to calibrate the downscaled climate data with ground measurements. Finally, the calibrated 30-m climate data were aggregated to derive 19 annual bioclimatic variables and the average values of the variables from 1991 to 2020 were calculated. A visual representation of the workflow can be found in Fig. 2.

A. ML-Based Downscaling

ML has been widely used in downscaling tasks of air temperature [28], [50] and precipitation [51]. Given the extensive modeling and predicting demands for this long time-series study,
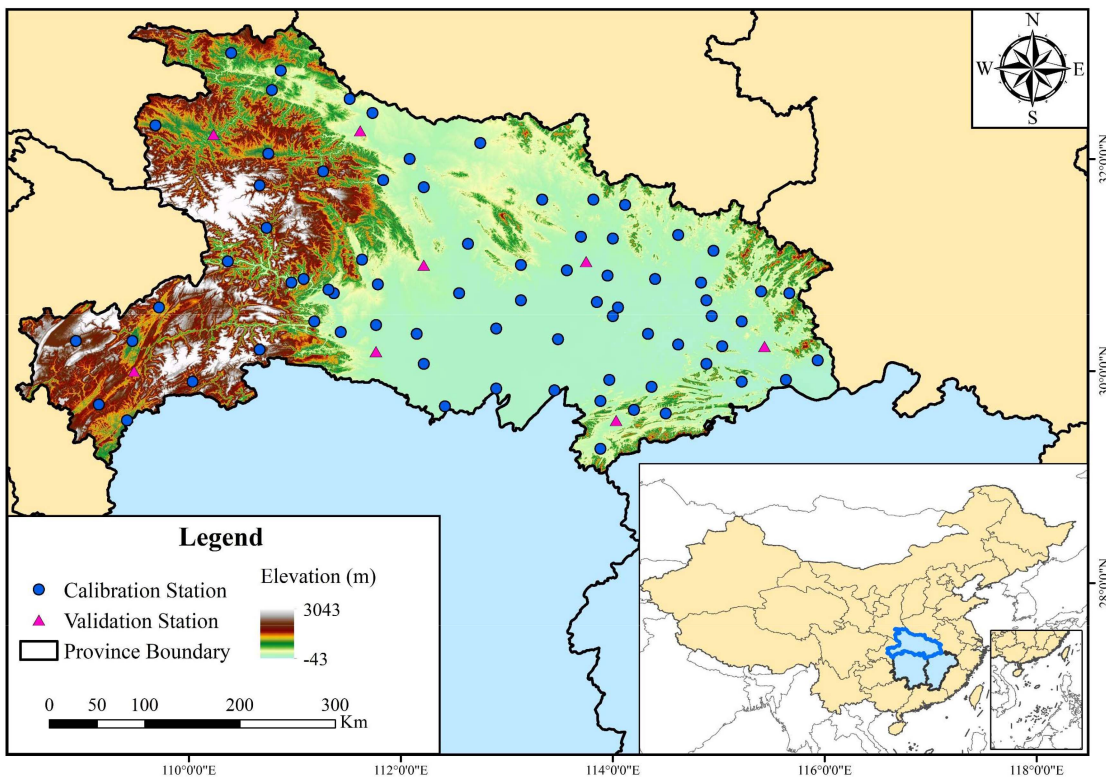
Fig. 1. Location of the meteorological stations used in this study in Hubei Province (with blue edge), the Yangtze River Middle Reaches (with blue areas). The validation stations were randomly selected to validate the performance of interpolation-based calibration with observation data.
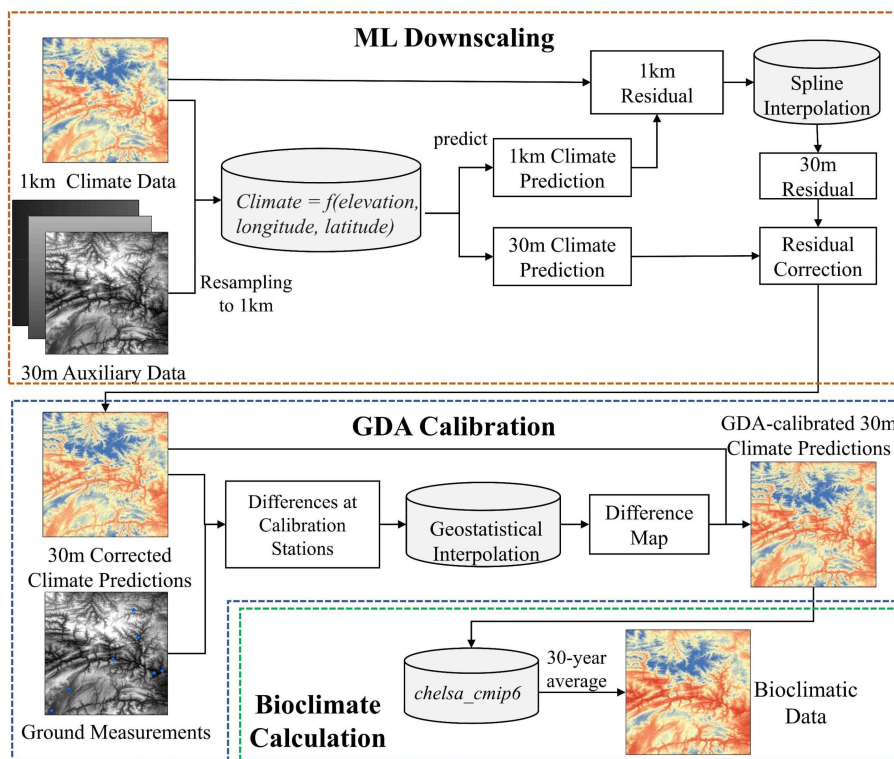

Fig. 2. Flowchart of the bioclimate dataset construction procedure.

the choice of models should consider the balance between accuracy and efficiency, with the benefits of shorter runtimes, ease of operation, and fewer artifacts besides high accuracy. Consequently, the random forest (RF) method was ultimately chosen for this task, which can leverage parallel computing and bootstrap sampling to attain high efficiency and satisfactory accuracy, while also offering favorable transparency and interpretability [52]. The 1-km elevation, longitude, and latitude of extent Hubei Province were used as the predictors. In total, 1440 1-km climate images (four target variables in 360 months) as the target variables with the predictors were combined to construct 1440 RF regression models. Each image contains 930×508 pixels. Then, the well-trained models were used to make prediction of the climate variables at 30-m resolution

$$\widehat{\mathrm{Clim}}_{30\mathrm{m}} = \frac{1}{n} \sum_{i=1}^{n} (\beta_{0i} + \beta_{1i} \times \mathrm{Ele}_{30\mathrm{m}} + \beta_{2i} \times \mathrm{Lon}_{30\mathrm{m}}$$
$$+ \beta_{3i} \times \mathrm{Lat}_{30\mathrm{m}}) \qquad (1)$$

where $\widehat{\mathrm{Clim}}_{30\mathrm{m}}$ is the 30-m climate variables, n is the total number of decision trees in the RF, and β_is are the parameters learned by the ith decision tree.

A residual correction was then performed to refine the accuracy of downscaled climate data by addressing biases to some extent. First, the prediction of the climate variables at 1-km resolution was performed by using the corresponding trained RF models with 1-km elevation, longitude and latitude maps

$$\widehat{\mathrm{Clim}}_{1\mathrm{km}}^{\mathrm{predict}} = f(\mathrm{Ele}_{1\mathrm{km}}, \mathrm{Lon}_{1\mathrm{km}}, \mathrm{Lat}_{1\mathrm{km}}). \qquad (2)$$

Then, 1440 residual maps at 1-km resolution were obtained by the computation of difference between the 1-km original variables and corresponding predicted variables, which indicates the unpredictable part of the variables. The bivariate spline interpolation (in *scipy.interpolate* library, Python 3.9) was utilized to interpolate the 1-km residuals to obtain the 30-m residuals. The spline method is suitable for interpolation of regularly spaced data [32], [53], as the regularity in the spacing of data points can provide implicit constraints that contribute to the smoothness of the interpolated surface. The process can be represented by following equations:

$$\widehat{\mathrm{Res}}_{1\mathrm{km}} = \widehat{\mathrm{Clim}}_{1\mathrm{km}}^{\mathrm{origin}} - \widehat{\mathrm{Clim}}_{1\mathrm{km}}^{\mathrm{predict}} \qquad (3)$$

$$\widehat{\mathrm{Res}}_{30\mathrm{m}} = \mathrm{Spline}\left(\widehat{\mathrm{Res}}_{1\mathrm{km}}\right). \qquad (4)$$

Afterwards, the predicted 30-m target variables were corrected by the addition of the corresponding 30-m residual values as

$$\widehat{\mathrm{Clim}}_{30\mathrm{m}}^{\mathrm{correct}} = \widehat{\mathrm{Clim}}_{30\mathrm{m}} + \widehat{\mathrm{Res}}_{30\mathrm{m}}. \qquad (5)$$

The downscaled 30-m climate variables were evaluated by using the observation data from 82 ground meteorological stations. Coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and bias were also included for a more comprehensive evaluation.

In addition, the semivariograms of temperatures and precipitation were generated in order to check if the downscaled maps preserved the similar spatial variability of the original maps. The equation of semi-variogram [54] is

$$\gamma(h) = \frac{1}{2N(h)} \times \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_{i+h}))^2 \qquad (6)$$

where $\gamma(h)$ is the semivariogram of lag distance h, $N(h)$ is the number of pairwise pixels, and $Z(x_i)$ is the value of the target variable at location x_i. The calculations were performed with *gstat* package [55] in *R*.

B. GDA Calibration of Downscaled Data

The GDA calibration procedure was developed by Cheema and Bastiaanssen [56] for fusing ground measurements and estimated products by interpolating the difference between them to unmeasured areas, which reduces the impact of geostatistics and has been proven to indicate good performance in many studies [31], [32], [57]. The difference between the 30-m downscaled climate variables and corresponding measurements for the 74 calibration meteorological stations was first calculated by using

$$\Delta\mathrm{Clim}_{(x,y)} = \mathrm{Clim}_{(x,y)}^{30\mathrm{m}} - \mathrm{Clim}_{(x,y)}^{\mathrm{obs}} \qquad (7)$$

where $\Delta\mathrm{Clim}_{(x,y)}$ is the climate variable difference between the 30-m downscaled data and meteorological station data at the location with coordinate of (x, y), and $\mathrm{Clim}_{(x,y)}^{30\mathrm{m}}$ and $\mathrm{Clim}_{(x,y)}^{\mathrm{obs}}$ are the 30-m downscaled climate value and meteorological station measured value at the location. If a station measurement is unavailable for a specific month, it will not be included in the procedure for the month.

Then, the differences at calibration stations for a specific month were used to generate the difference map. Following interpolation methods used in prior researches [31], [56], [58] that have reported desirable performance, we employed the IDW and OK interpolation techniques separately to generate the spatially interpolated difference map for all target variables and subsequently conducted a comparison. The IDW interpolation is the spatially weighted average sample values in the searching neighborhood [59] that is robust in estimation and has no string and screening effect [60]. IDW was applied with the exponent power of 2. OK is based on Gaussian process governed by prior covariances that can give the best linear unbiased prediction at target locations [61]. Gaussian model was used as the variogram model in this study. The $\Delta\mathrm{Clim}^{\mathrm{ip}}$ in the following equation is the spatially interpolated difference map generated by interpolation:

$$\Delta\mathrm{Clim}^{\mathrm{ip}} = \begin{cases} \mathrm{IDW}\left(\Delta\mathrm{Clim}_{(x,y)}\right) \\ \mathrm{OK}\left(\Delta\mathrm{Clim}_{(x,y)}\right) \end{cases}. \qquad (8)$$

After the difference maps were prepared, the values of the difference maps were subtracted from the 30-m downscaled climate variable maps to get the calibrated maps by using

$$\widehat{\mathrm{Clim}}_{30\mathrm{m}}^{\mathrm{cal}} = \widehat{\mathrm{Clim}}_{30\mathrm{m}}^{\mathrm{correct}} - \Delta\mathrm{Clim}^{\mathrm{ip}} \qquad (9)$$

where $\widehat{\mathrm{Clim}}_{30\mathrm{m}}^{\mathrm{cal}}$ is the calibrated climate variable map and $\widehat{\mathrm{Clim}}_{30\mathrm{m}}^{\mathrm{correct}}$ is the 30-m downscaled climate variable map. The benefit of GDA calibration is that it works on the differences

of ground truth and image rather than the measured values, which can reduce the impact of geostatistic methods on the final calibrated products [56].

C. Calculation of Bioclimatic Variables

The better calibrated datasets in previous step will be used for the calculation of bioclimatic variables. The 19 bioclimatic variables were aggregated according to the definition of a series studies [3], [62] and by using the provided formulas (see Table I). Detailed description of the bioclimatic variables can be found in [3]. The Python library *chelsa_cmip6* developed by Karger et al. [62] was used as the basis for constructing the bioclimatic variable calculation program, but due to the huge image dataset with 30-m spatial resolution, the original *xarray.Dataset* based calculation in *Bioclim.py* file was modified by using array calculation on the dimension of time to achieve a higher computational speed.

To obtain the annual quarter with a 3-month interval, in total 12 consecutive sets were constructed by using the previous and following month with the focal month. The first and last sets were by using December, January, February, and November, December, January, respectively. Annual quarter calculation was achieved by *quarter_class* class in *chelsa_cmip6.Bioclim* Python library.

Finally, the grid bioclimatic variables were postprocessed by the smoothing filter designed by Daly et al. [1] to reduce noise and artifacts. The filter performs a distance-weighted average of all surrounding grid cells within the filtering window and ensures a smoother bioclimatic field in low-gradient areas without affecting the high-gradient areas, and the filter can be expressed as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \frac{1}{d_i^a}}{\sum_{i=1}^{n} \frac{1}{d_i^a}}, a = \begin{cases} a_{\max}; \overline{\Delta x} \geq \Delta x_{\max} \\ a_{\max}\left(\frac{\overline{\Delta x}}{\Delta x_{\max}}\right); \overline{\Delta x} < \Delta x_{\max} \end{cases} \quad (10)$$

where $\bar{x}$ is the averaged value of center pixel, x_i is the value at pixel i, d_i is the distance between the centre pixel and grid cell i, a is the distance weighting exponent, $a_{\max}$ is the maximum exponent, $\overline{\Delta x}$ is the mean absolute difference between the centre pixel and all surrounding pixels within the filtering window, and $\Delta x_{\max}$ is the specified maximum average difference.

To cross validate the constructed dataset and study the variation in different climate normals, the comparison with World-Clim global bioclimate products [19] was also implemented. The final calibrated bioclimatic dataset was further resampled back to 1-km spatial resolution by using bilinear interpolation to calculate the difference maps of two datasets [24]. Spearman correlation coefficients, RMSE, and normalized RMSE (nRMSE) were used for comparison evaluation.

IV. Results

A. Evaluation of the Downscaling Procedure

In total, 1440 RF models were constructed for downscaling the four monthly climate variables in the 30-year period. The downscaled products were validated by ground observation data from 82 meteorological stations, and the 1-km climate variables were also included as the reference. A better performance can be
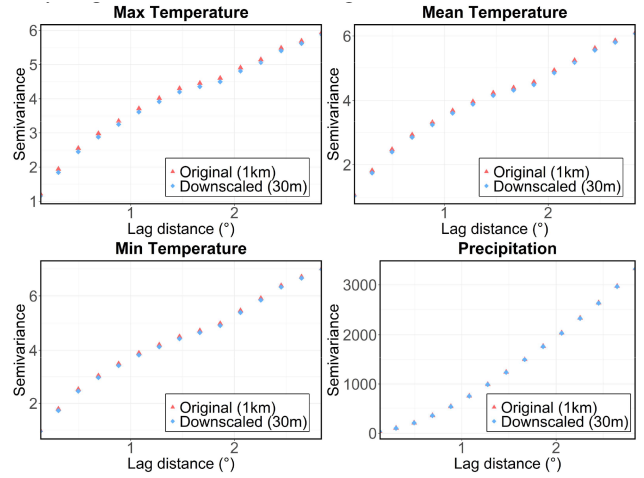

Fig. 3. Semi-variograms for original and downscaled monthly maximum, mean, minimum temperatures, and precipitation (June 2020 as an example).

observed in Table II that three temperature variables downscaled by our procedure show slightly higher R^2s, as well as lower RMSEs and MAEs. The downscaled monthly mean temperatures exhibit a higher bias compared to the original values, leaning toward underestimation. The downscaled precipitation shows a slightly higher RMSE, while a lower R^2, MAE, and bias. But overall, according to the station-based validation, variations of the accuracy caused by the downscaling procedure are negligible. Semivariograms and spatial patterns were also analyzed to ensure the downscaled images maintained the spatial structure of original images. Figs. 3 and 4 were conducted by using original 1-km variables and downscaled 30-m variables in a specific month as an example. The results indicate that the spatial structure of all downscaled variables closely aligns with those of the original variables.

Fig. 5 illustrates the MAEs of downscaled temperatures and precipitation at each meteorological station, providing insights into the local performance of the products. In the current stage, the temperature variables exhibit relatively larger errors in western regions of Hubei Province, which is corroborated by corresponding maps for R^2 and RMSE (see Figs. S1 and S2 in the Supplementary Material), possibly due to the sparser availability of observation data in high-elevation regions during the development of original dataset. In terms of precipitation, the southeastern region exhibits very high MAE values, primarily attributed to the local high precipitation intensity [63]. For the similar reason, the low precipitation intensity in the northwestern arid areas results in a favorable MAE value but a low R^2 (see Fig. S1 in the Supplementary Material). Furthermore, southwestern region also exhibits relatively larger errors. This mirrors the case of temperatures where insufficient ground precipitation data during the construction of the original dataset may be the primary factor responsible for this outcome.

B. Evaluation of GDA-Calibrated Variables

As the original gridded climate data were constructed with limited observation data and the downscaling procedure has a very limited impact on the data accuracy, the downscaled
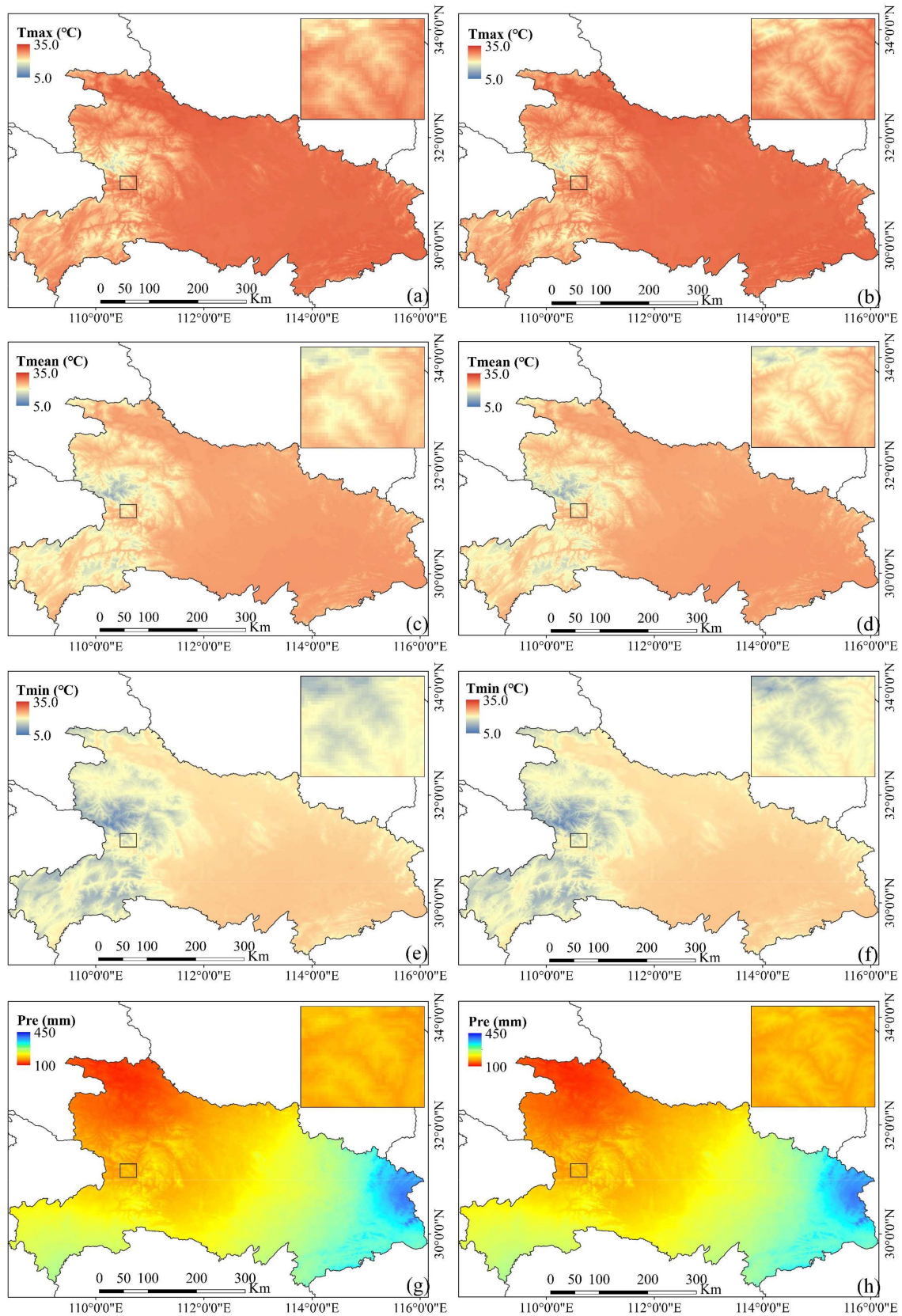
Fig. 4.   Comparison of 1-km original (left images) and downscaled 30-m (right images) (a) and (b) monthly maximum, (c) and (d) mean, (e) and (f) minimum temperatures, and (g) and (h) monthly precipitation with data in June 2020 as an example.

TABLE II
EVALUATION OF 1-KM CLIMATE VARIABLES AND DOWNSCALED 30-M CLIMATE VARIABLES BASED ON OBSERVATION DATA FROM 82 METEOROLOGICAL STATIONS
IN HUBEI PROVINCE

Variables	1-km climate variables				30-m downscaled corrected climate variables			
	R^2	RMSE	MAE	Bias	R^2	RMSE	MAE	Bias
T_{max} (°C)	0.9844	1.06	0.78	0.35	0.9859	1.01	0.75	0.32
T_{mean} (°C)	0.9920	0.75	0.58	-0.27	0.9923	0.73	0.57	-0.30
T_{min} (°C)	0.9867	0.95	0.74	0.41	0.9873	0.93	0.72	0.39
Pre (mm)	0.7197	49.01	29.26	1.50	0.7195	49.02	29.25	1.49
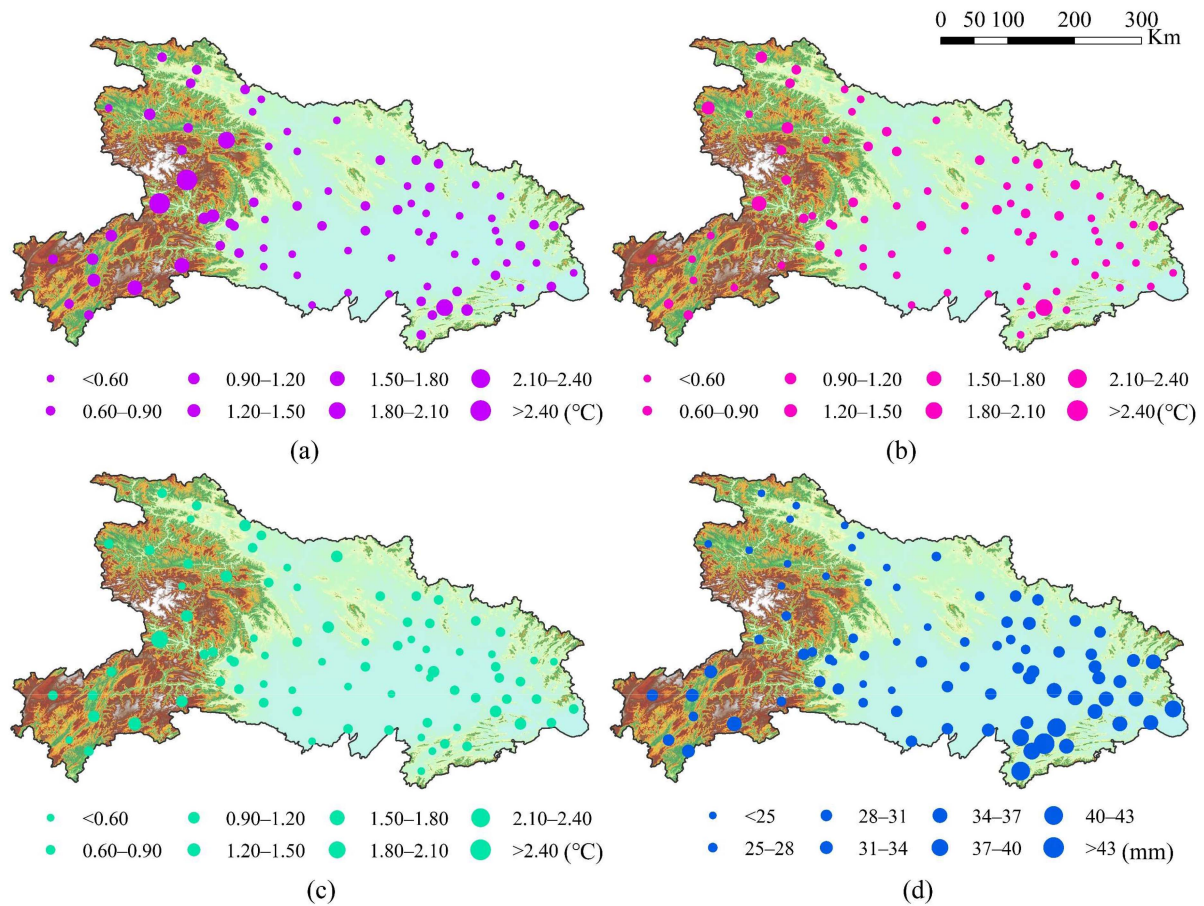
Fig. 5. MAE maps for the downscaled (a)–(c) monthly maximum, mean, and minimum temperatures, and (d) monthly precipitation at the 82 meteorological stations in Hubei Province from 1991 to 2020.

products still have significant errors. Two geostatistical methods, IDW and OK, were separately used as the interpolators for the GDA calibration to correct the errors with additional local observation data. Table III illustrates the evaluation metrics of GDA-calibrated downscaled 30-m climate variables using the observation data from eight validation stations. Significant improvements have been achieved for all variables by using both GDA calibration approaches. The IDW calibration allows the R^2s of temperature variables to increase and surpass 0.995, and the calibrated precipitation also exhibits a notable increase of R^2 to 0.8819 (by 21.3%). While the OK-calibrated temperatures demonstrate slightly higher performance than IDW, a contrasting observation is noted in precipitation calibration, where the performance is less favorable. Both calibration methods result

in notable reductions in RMSEs and MAEs across monthly maximum, mean, and minimum temperatures. Similarly, substantial improvements can be also observed in the calibration of monthly precipitation, with consistent decreases in the two metrics. Biases of monthly maximum, mean, and minimum temperature are all improved significantly in both methods, but get a bit worse for monthly precipitation. This arises from the high positive bias of a single station (refer to Tables S2–S5 in the Supplementary Material), which, in the original dataset, neutralizes the negative biases of some other stations. Calibrations result in a significant reduction of this bias, leading to an overall increase in negative bias. And OK-calibration causes a lower bias for temperature variables and a higher bias for precipitation compared to IDW-calibration. Overall, according to the metrics,

TABLE III
EVALUATION OF UNCALIBRATED DOWNSCALED CLIMATE VARIABLES AND GDA-CALIBRATED CLIMATE VARIABLES BASED ON 2877 OBSERVATION DATA FROM 8
VALIDATION METEOROLOGICAL STATIONS IN HUBEI PROVINCE

Variables	Uncalibrated 30-m downscaled climate variables				IDW-calibrated 30-m downscaled climate variables				OK-calibrated 30-m downscaled climate variables			
	R^2	RMSE	MAE	Bias	R^2	RMSE	MAE	Bias	R^2	RMSE	MAE	Bias
T_{max} (°C)	0.9871	0.97	0.74	0.37	0.9956	0.57	0.43	0.12	0.9958	0.55	0.43	0.07
T_{mean} (°C)	0.9947	0.61	0.47	-0.25	0.9980	0.38	0.29	0.10	0.9981	0.36	0.28	0.05
T_{min} (°C)	0.9887	0.87	0.47	0.69	0.9967	0.47	0.36	0.12	0.9967	0.47	0.36	0.10
Pre (mm)	0.7270	45.99	28.27	-0.40	0.8819	30.25	18.58	-1.75	0.8376	35.48	21.43	-1.82
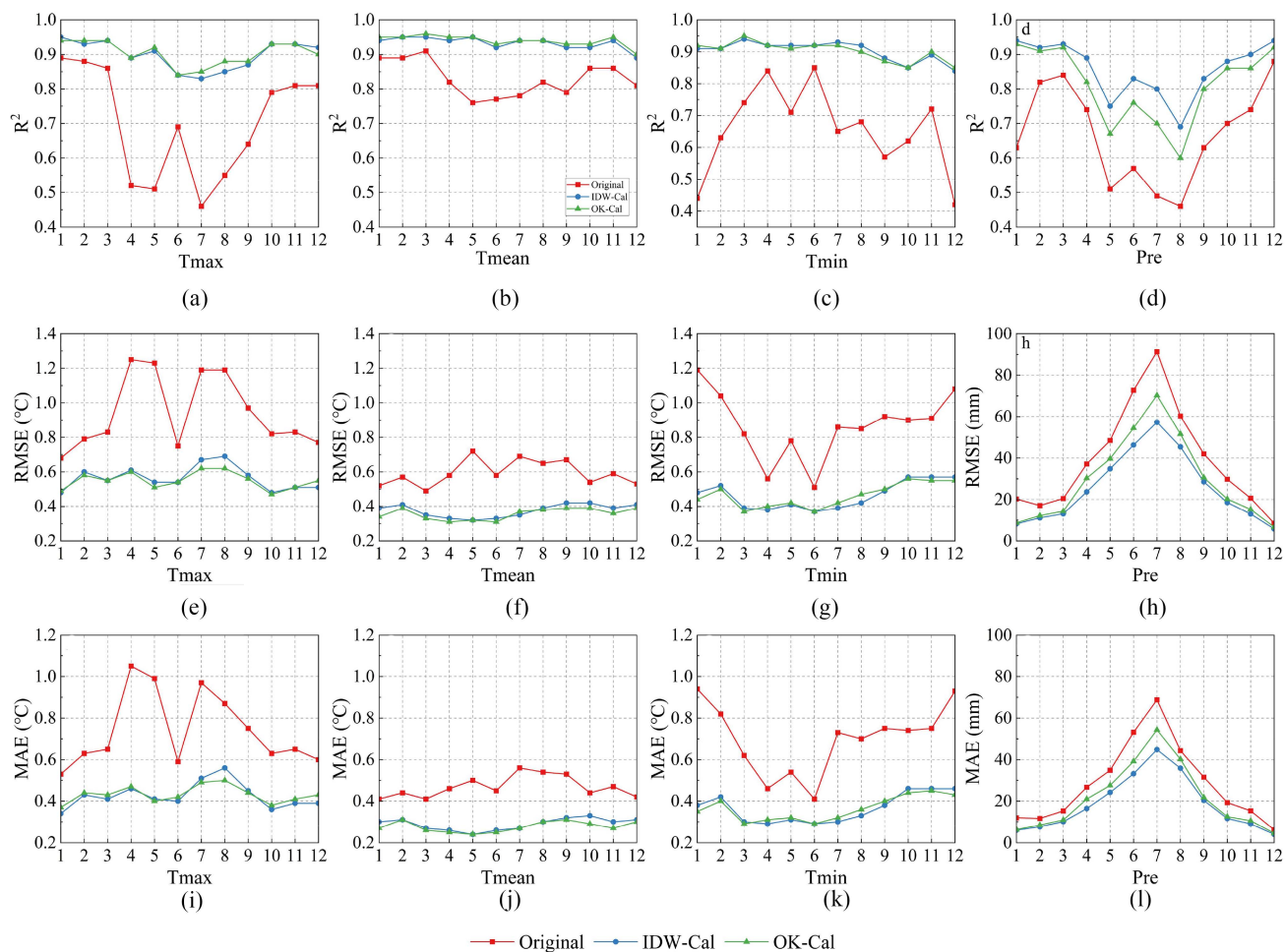
Fig. 6. (a)–(d) R^2s, (e)–(h) RMSEs, and (i)–(l) MAEs of original 1-km climate variables and GDA-calibrated downscaled climate variables for each month based on 2877 observation data from eight validation meteorological stations in Hubei Province for 1991–2020 climate normals.

OK performs better on the calibration of temperatures, while IDW performs better on the calibration of precipitation.

Fig. 6 is provided to ensure the applicability of calibration methods across various months and to facilitate a performance comparison. Both the IDW and OK calibration methods lead to improvements in monthly temperatures and precipitation for each month. R^2s exhibit consistent increases across the variables, while RMSEs and MAEs for these variables show notable decreases. While the performance of OK calibration does not consistently surpass that of IDW calibration

for temperature, the IDW calibration always outperforms for precipitation.

Fig. 6 also illustrates that the large errors in maximum temperature are concentrated in late spring and summer (April to September), while errors in minimum temperature are concentrated in winter (December to February). This phenomenon may be attributed to the inherent biases or limitations of interpolation algorithms used during the development of the original dataset that might not accurately represent the extreme temperatures in these seasons. Notably, the data for these months have been
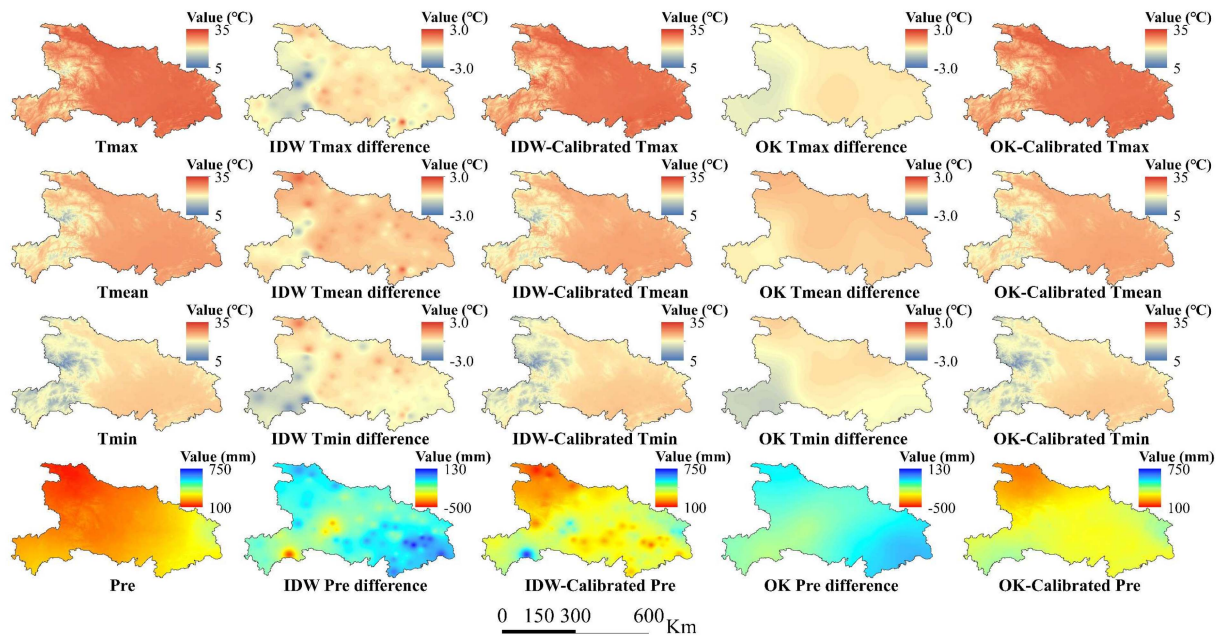
Fig. 7.   IDW-based and OK-based calibrations of the downscaled monthly maximum, mean, minimum temperature, and monthly precipitation by subtracting the difference with station observation data in June 2020.

significantly improved with both calibration methods. OK generally performed better on these months when calibrating the temperature variables. In addition, large errors of precipitation mainly occur in summer (May to August). The performances of IDW and OK are very similar in the dry season, but IDW tended to perform much better in wet season. Evaluation metrics for data of eight validation stations can be found in Table S1 in the Supplementary Material. And a randomly selected partial period (2006–2010) is visualized in Figs. S3–S6 in the Supplementary Material. In most cases, the calibrated values indicated higher consistency with the observed values compared to the original data.

To assess the spatial distribution of calibrated differences and compare the final spatial patterns of climate variables calibrated by two methods, we visualized the data of June 2020 as an example and generated Fig. 7. The impacts of calibrations on temperatures are generally insignificant, with most pixels showing low absolute difference values, which can mainly be attributed to the high consistency between the original image pixel values and ground measurements at calibration stations. In contrast, precipitation calibration exhibits more pronounced effects, particularly in the southwest region. Given that the original precipitation data failed to capture the small-scale precipitation extreme in this region due to the incorporation of limited weather stations during its production, the differences are substantial.

According to the figure, the primary difference between the two calibration methods is that the IDW interpolated difference maps reveal a few higher absolute difference values, creating distinct "high-value circles," while the OK interpolated difference maps present a smoother variation without these "circles." Despite not being immune to the extreme values in precipitation calibration, the OK interpolated difference map appears considerably smoother comparing to the IDW interpolated differences.

C. Bioclimatic Dataset Construction

Both of the IDW-calibrated and OK-calibrated monthly climate variables were then aggregated into annual bioclimate variables by using 48 climate variables (four variables in 12 months) for each year. The average values of 19 bioclimatic variables in 1991–2020 Climate Normals generated on the basis of IDW-calibrated and OK-calibrated data are shown in Figs. 8 and 9, respectively. From the maps, the overall spatial distribution of the bioclimatic variables does not indicate distinct difference between two calibration approaches. The main difference between the two dataset is that the effect of extreme values caused by IDW interpolation leads to obvious circles with deeper color even though these are 30-year average value maps, especially in bio9, bio13, and bio15, making the maps lose the spatial continuity. In light of this, on the basis of the comprehensive consideration of station-based validation accuracy and spatial pattern in this study, the bioclimatic data calculated by using OK-calibrated climate variables are more recommended to be further processed, evaluated, and used, which can simultaneously indicate the enough accuracy and reasonable spatial distribution.

D. Comparison With WorldClim Bioclimatic Dataset

To cross validate the generated bioclimatic dataset and indicate the general variation happened between the two climate normals, the bioclimatic variables for 1971–2000 climate normals were downloaded from WorldClim [19] dataset. The OK-calibrated bioclimatic variables were finally processed through Daly's smoothing filter [1] to further reduce the influence of noise and artifacts. They were then used in this comparison by being resampled to 1-km resolution that is same to WorldClim data, calculating the Spearman correlation coefficients, RMSE, and nRMSE that are show in Table IV. According to Spearman
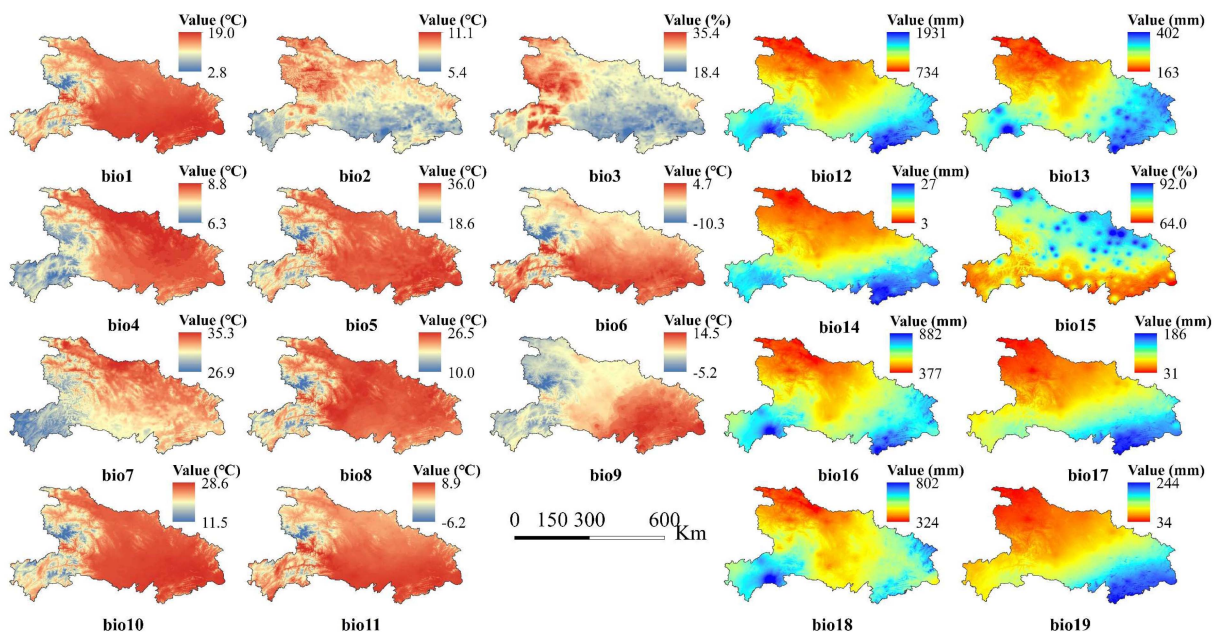
Fig. 8. IDW-calibrated 30-m resolution bioclimatic variables bio1–bio19 in Hubei Province for 1991–2020 climate normals.
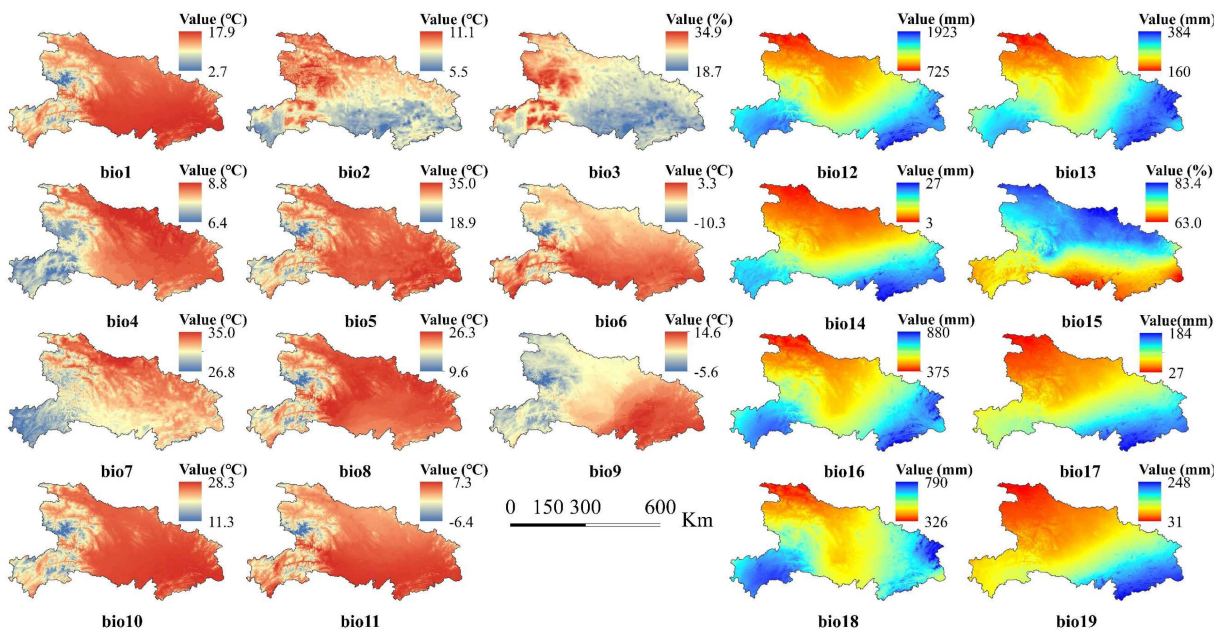

Fig. 9. OK-calibrated 30-m resolution bioclimatic variables bio1–bio19 in Hubei Province for 1991–2020 climate normals.

correlation coefficient and P-value, there are significant linear correlations for all the variables between two datasets. Most variables indicate very high correlation coefficient (> 0.80), while the seasonality trend of precipitation (bio15) was most poorly correlated, which is followed by diurnal air temperature range (bio2, 0.72). The nRMSE of precipitation (bio12–19) variables shows an overall higher discrepancy than that of temperature variables (bio1–11).

The spatial distribution of differences between the two datasets was calculated by using our dataset values to subtract WorldClim values (see Fig. 10). First, the differences for

the annual mean temperature (bio01) of two datasets mainly distribute in the high-altitude mountainous areas, with difference values higher than 1 °C. The high differences of bio2–7, bio10-11 can be observed in western mountainous areas, with the peak value of 4.2 °C for the maximum temperature of the warmest month (bio5). Mean temperature of wettest quarter (bio8) indicates high difference in western high-altitude areas. Abrupt changes can be found in southeast regions in bio8 and bio9, leading to the peak value of 8.3 °C for the difference of mean air temperatures of the driest quarter (bio9). Regarding precipitation bioclimatic variables, the difference of annual
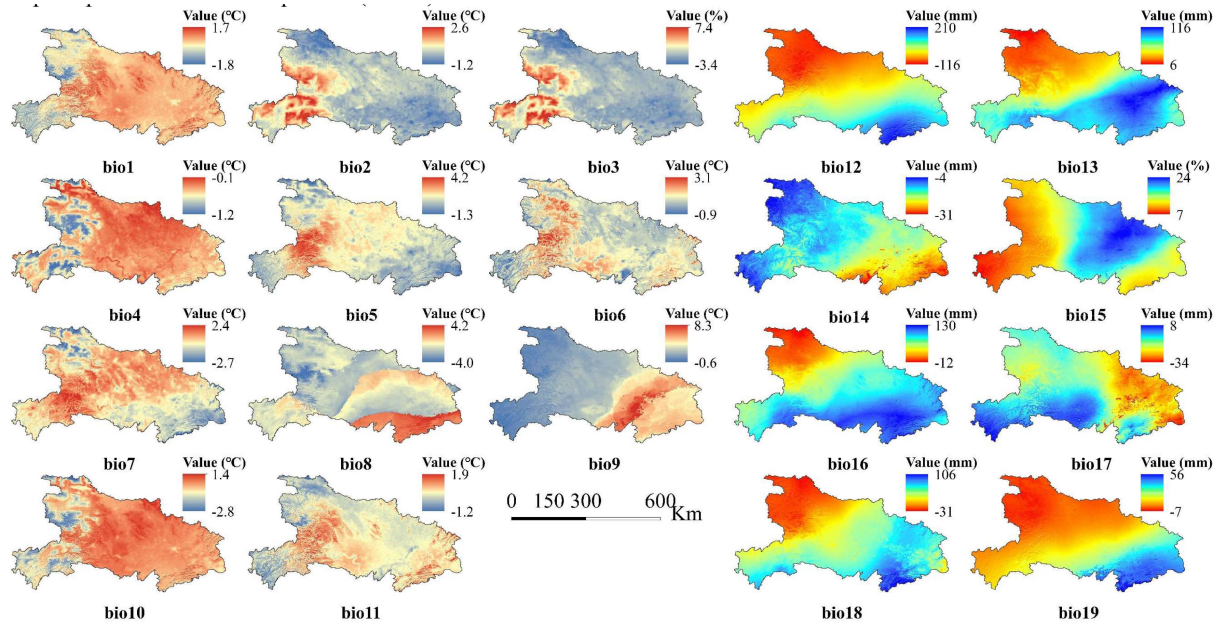
Fig. 10. Difference maps of bioclimatic variables in constructed dataset and WorldClim dataset for Hubei Province.

TABLE IV
COMPARISON OF THE GENERATED AND WORLDCLIM BIOCLIMATIC VARIABLES
IN TERMS OF SPEARMAN'S CORRELATION COEFFICIENT, RMSE, AND nRMSE

Symbol	Spearman correlation	RMSE	nRMSE	P-value
bio1	0.99	0.49	0.03	<0.001
bio2	0.72	0.64	0.12	<0.001
bio3	0.80	1.77	0.12	<0.001
bio4	0.96	0.53	0.22	<0.001
bio5	0.94	1.57	0.10	<0.001
bio6	0.97	1.15	0.09	<0.001
bio7	0.91	0.65	0.08	<0.001
bio8	0.84	1.23	0.08	<0.001
bio9	0.91	3.17	0.16	<0.001
bio10	0.99	0.52	0.03	<0.001
bio11	0.98	0.52	0.04	<0.001
bio12	0.98	61.69	0.05	<0.001
bio13	0.83	62.66	0.28	<0.001
bio14	0.95	12.36	0.51	<0.001
bio15	0.62	15.71	0.79	<0.001
bio16	0.96	64.58	0.13	<0.001
bio17	0.99	8.17	0.05	<0.001
bio18	0.95	32.41	0.07	<0.001
bio19	1.00	17.60	0.08	<0.001

precipitation (bio12) shows an obvious gradient: WorldClim data are higher in the northwest areas and lower in the southeast areas, with the highest differences of 210 mm. Difference of precipitation of wettest month (bio13) is totally positive, while that of precipitation of driest month (bio14) is opposite, but they

nearly follow the gradient of bio12. The difference of precipitation seasonality (bio15) is much higher in central and eastern flatten low-altitude areas. Differences of precipitation of wettest quarter (bio16) and wettest and coldest quarter (bio18–19) share a similar spatial pattern, with higher values in southeast areas, while the precipitation of the driest quarter (bio17) also shows a higher difference in southeast, but it is negative.

V. DISCUSSION

A. Performance of the Bioclimate Dataset Construction Framework

The framework used for bioclimatic dataset construction in this study is based on the downscaling-calibration procedure [32], of which the customized steps were achieved by using the RF model with a residual correction and GDA, respectively. The outstanding comprehensive performance makes RF suitable for this task. It is noticed that the downscaling procedure hardly affected the accuracy of its products, with only a very slight accuracy improvement for temperature variables and decrease for precipitation observed, similar to [64] but being different from others [32], [33], [53]. This may first attribute to the quality of original data [65], which has already considered the auxiliary data used by us, allowing the high consistency of the spatial variability between original and downscaled data. In addition, for the pixels of original data at or around the weather stations, when corresponding fine pixels of auxiliary data exhibit low spatial heterogeneity, downscaling may not cause substantial variations. Thus, the downscaled products still contain certain errors against ground observations due to the error propagation from coarse resolution products [31]. As the coarse data were constructed with limited ground measurements in the study sites [42], the calibration by including data from additional local stations can provide more information related to local climate pattern. Data
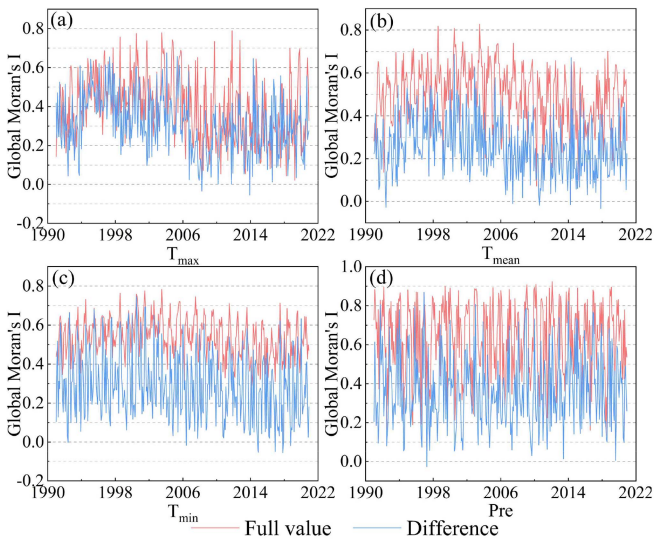
Fig. 11. Monthly variation of moran index for spatial autocorrelation evaluation of ground measurements and their differences with downscaled data.

from tens of local new stations were incorporated for the GDA calibration, significantly improving accuracy according to the enhanced evaluation metrics.

B. Comparison of IDW and OK Methods for the Calibration

This study considered IDW and OK calibration methods to fuse the downscaled data and ground measurements. While overall evaluation metrics favor OK calibration for three temperature variables (see Table III), a closer look at monthly assessments (see Fig. 6) reveals nuances, aligning with previous findings [66]. The seasonal difference of the comparison performance can be linked to principles of the two methods. OK assumes spatial continuity and correlation between neighboring locations, making it suitable for direct temperature spatial interpolation [67], [68]. When OK is used for calibration, the differences between temperature image and ground measurements also show spatial autocorrelation, although usually much weaker (see Fig. 11). In this case, the superiority of the OK method may vary across months as its performance highly relies on the significance of variable's spatial variability and the ability of the variogram to capture it for a specific month. In contrast, the IDW method disregards the spatial variability [69], [70] and only focuses on distances between the interpolation points and the surrounding data points, performing stable calibration across months. Hence, this indicates that the higher accuracy method may vary in different seasons. Differently, IDW outperforms OK in precipitation calibration throughout all months, which might be mainly attributed to the quality of source data. Given the limited quality of sourced precipitation data, the differences with ground measurements will exhibit extreme values in areas where the small-scale precipitations are not well captured. IDW assumes that the near values are more related than distant values [71] and the precipitation usually has obvious proximity in its spatial distribution, thus the interpolation results can be influenced strongly by these extreme values [72], allowing the

station-surrounded regions to be well calibrated. While the OK interpolation focuses more on large-extent trend, leading to a narrower range of values in the interpolated map compared to the IDW-interpolated map. Thus, the surrounding pixels of the stations with extreme values cannot be well calibrated.

Regarding the spatial pattern of calibrated maps, OK demonstrates more favorable outcomes than IDW. As IDW handles extreme values by giving higher weights to nearby data points, it tends to pull the interpolated surface toward those extreme values of the difference, resulting in enhanced local signals and less smooth maps (see Fig. 7) [73]. Although the good quality of sourced temperature images (see Table II) cause fewer extreme values of differences with the ground measurements, allowing IDW-calibrated temperature maps to be smoother than precipitation map, these "circle" signals can still be observed and are undesired in generation of bioclimatic maps (see Fig. 8). Furthermore, IDW calibration fails to integrate the neighboring stations well. This is evident in subfigure "IDW precipitation difference" in Fig. 7 where the same deeper color circles around meteorological stations become shallower in their intermediate areas. These intermediate areas should ideally have deeper color than the pixels around surrounding stations, reflecting the central precipitation. This phenomenon can be attributed to the sparse distribution of adjacent stations and the power of the distance weight. Although reducing the power value can increase the influence of more distant data points on the interpolation results that may lead to a better spatial connection [60], as this study worked with the power of 2, there is not much room for adjustment anymore to improve the calibration pattern.

C. Potential Reasons of the Differences Between WorldClim and Our Bioclimatic Datasets

Differences of bioclimatic variables can be observed between the new dataset and WorldClim dataset (see Table IV and Fig. 10). As there is a temporal discrepancy between WorldClim dataset (1970–2000) and our study (1991–2020), inherent errors and climate change can both contribute to these differences. Previous studies [22], [23], [24] have reported that the regional bioclimatic data are critical for local applications, which are restricted by the coarse resolution and lower accuracy of global dataset. The WorldClim dataset was developed by regression modeling and TPS interpolation, utilizing data from a part of global weather stations [22]. However, the limited incorporation of data from local weather stations within Hubei Province during its development [74] may result in lower accuracy in some specific areas, which underscores the greater importance of improving the dataset accuracy through the addition of ground data than focusing on refining the methodology.

The differences caused by climate change are noteworthy. For example, the large differences in bio2-6 concentrate in the western mountainous areas. The positive differences in bio2, bio3, bio5, and bio6 typically suggests an increasing trend of overall temperatures, with greater variations in temperature between day and night. This can result in more uniform climate throughout the year. This can also be proved by negative large differences in bio4 and bio10 and positive difference in bio11
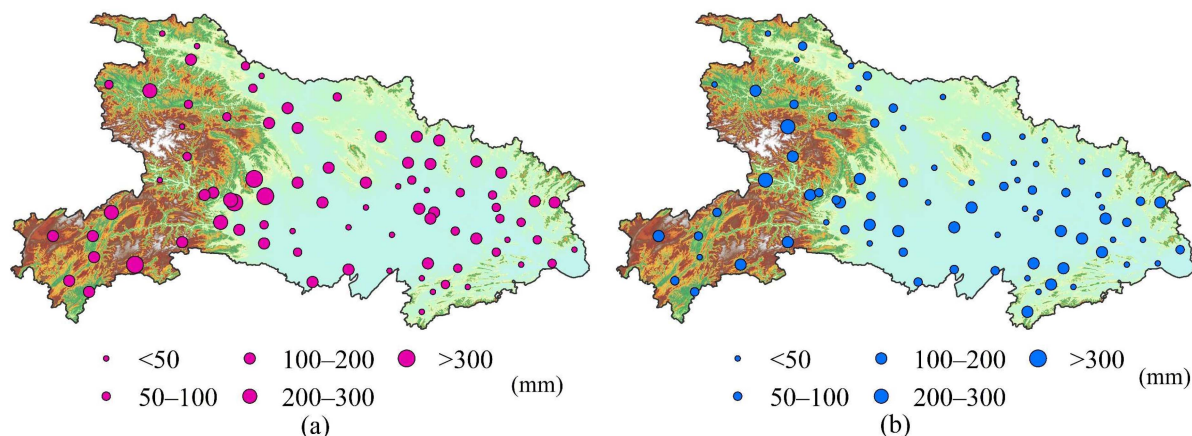
Fig. 12. Absolute error of monthly total precipitation (a) between sourced data in this study [42] and ground observation data, and (b) between ERA5-Land [77] and ground observation data, in Hubei Province in June 2020.

in western Hubei, which overall follows the statement that mountainous environments experience more rapid temperature changes than plain areas [75]. The differences in bio8 and bio9 between the two datasets are significant, as the eastern Hubei is divided into three distinct different parts (see Fig. 10). This may mainly attribute to the changes in precipitation patterns, leading to different months for the wettest or driest seasons. Difference of precipitation mainly concentrates in eastern areas where more precipitation in the latest climate normal leads to the large positive difference in bio12 and bio16, as well as small negative difference in bio17. There is a positive difference in the bio13 and a negative difference in bio14 that might be caused by the changes in rainfall seasonal distribution. These changes can also be indicated in the bio15, bio18, and bio19.

D. Limitations and Further Perspectives

This study aligns with previous research by using climate variables from the same source [28], [76], but the quality of precipitation is comparatively lower. The comparison of precipitation from our sourced data and ERA5-Land dataset (see Fig. 12) indicates that the ERA5-Land precipitation product has a generally better performance. Especially, it captures more of the small-scale rainfall extremes. Thus, further improvements may involve the selection of better data sources for a specific climate variable, and integration of multisource data to comprehensively improve the data quality.

In term of methodology, RF was chosen for the downscaling process to strike a balance between accuracy and efficiency. Using deep learning models such as convolutional neural network could be a better way to preserve more information of original images as they can capture the latent associations between climate variables and predictors [78], despite the black-box models will make it difficult to explain their inner workings. In addition, OK-calibrated outputs have finally remained as they achieve acceptable evaluation metrics and smoother maps, but the lower accuracy of precipitation calibration should be improved in further studies. The possible

solutions include using alternative variogram models or interpolation methods. Designing specific calibration methods for the different seasons could also be useful in achieving a more accurate representation of the climate conditions during that time.

Finally, this study only used DEM and coordinates as the auxiliary data. And the absence of additional ground details poses a limitation. For the fusion of estimated and observed data, integrating more ground surface information into the ML models that account for spatial autocorrelation and nonstationary [78] could be a more accurate solution.

V. CONCLUSION

This study aims to construct a 30-m bioclimatic dataset for Hubei Province during the latest 1991–2020 climate normals. To achieve this, it demonstrates a bioclimatic mapping procedure on the basis of statistical machine learning downscaling of coarse national-scale climate maps and interpolation-based calibration with observation data from local meteorological stations. It was found the downscaling procedure can generate reasonable downscaling products that can almost capture the spatial pattern of original maps. From the evaluation metrics, GDA calibration with the OK method achieved an overall slightly higher accuracy on three temperature variables while that with the IDW method indicated higher accuracy on precipitation. But from the spatial evaluation, OK-based calibration allowed the generated maps to be smoother and have more reasonable spatial distribution. The constructed bioclimatic dataset shows high consistency with global bioclimatic dataset according to the pixel-based comparison, but is of higher local accuracy. The latest high-resolution bioclimatic dataset is available to support the new generation of studies on ecology, as well as agricultural and forestry ordinations in Hubei Province. Stakeholders, researchers, and policymakers can utilize the dataset to gain a deeper understanding of Hubei's climate patterns and their implications on their concerned realms. The constructed bioclimatic dataset for Hubei Province during 1991–2020 can be accessed on https://doi.org/10.5281/zenodo.10057926.

REFERENCES

[1] C. Daly et al., "Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States," *Int. J. Climatol.*, vol. 28, no. 15, pp. 2031–2064, Dec. 2008, doi: 10.1002/joc.1688.

[2] H. A. Nix, "A biogeographic analysis of Australian elapid snakes," *Atlas Elapid Snakes Australia*, vol. 7, pp. 4–15, 1986.

[3] M. S. O'Donnell and D. A. Ignizio, "Bioclimatic predictors for supporting ecological applications in the conterminous United States," US Geological Survey Data Series, U.S. Geological Survey, Reston, VA, USA, 2012.

[4] M. D. Jennings and G. M. Harris, "Climate change and ecosystem composition across large landscapes," *Landscape Ecol.*, vol. 32, no. 1, pp. 195–207, Jan. 2017, doi: 10.1007/s10980-016-0435-1.

[5] J. Elith and J. R. Leathwick, "Species distribution models: Ecological explanation and prediction across space and time," *Annu. Rev. Ecol. Evol. Systematics*, vol. 40, no. 1, pp. 677–697, Dec. 2009, doi: 10.1146/annurev.ecolsys.110308.120159.

[6] H. Yan et al., "Predicting the potential distribution of an invasive species, Erigeron canadensis L., in China with a maximum entropy model," *Glob. Ecol. Conservation*, vol. 21, Mar. 2020, Art. no. e00822, doi: 10.1016/j.gecco.2019.e00822.

[7] B. W. Low, Y. Zeng, H. H. Tan, and D. C. J. Yeo, "Predictor complexity and feature selection affect Maxent model transferability: Evidence from global freshwater invasive species," *Diversity Distributions*, vol. 27, no. 3, pp. 497–511, Mar. 2021, doi: 10.1111/ddi.13211.

[8] F. Zhang, C. Wang, C. Zhang, and J. Wan, "Comparing the performance of CMCC-BioClimInd and WorldClim datasets in predicting global invasive plant distributions," *Biology*, vol. 12, no. 5, Apr. 2023, Art. no. 652, doi: 10.3390/biology12050652.

[9] J. Morales-Barbero and J. Vega-Álvarez, "Input matters matter: Bioclimatic consistency to map more reliable species distribution models," *Methods Ecol. Evol.*, vol. 10, no. 2, pp. 212–224, Feb. 2019, doi: 10.1111/2041-210X.13124.

[10] L. Li et al., "Elevational patterns of phylogenetic structure of angiosperms in a biodiversity hotspot in eastern Himalaya," *Diversity Distributions*, vol. 28, no. 12, pp. 2534–2548, Dec. 2022, doi: 10.1111/ddi.13513.

[11] E. Cano et al., "Mitigating climate change through bioclimatic applications and cultivation techniques in agriculture (Andalusia, Spain)," in *Sustainable Agriculture, Forest and Environmental Management*, M. K. Jhariya, A. Banerjee, R. S. Meena, and D. K. Yadav, Eds. Berlin, Germany: Springer, 2019, pp. 31–69, doi: 10.1007/978-981-13-6830-1_2.

[12] I. Charalampopoulos, F. Droulia, and J. Evans, "The bioclimatic change of the agricultural and natural areas of the Adriatic Coastal countries," *Sustainability*, vol. 15, no. 6, Mar. 2023, Art. no. 4867, doi: 10.3390/su15064867.

[13] B. Evans, C. Stone, and P. Barber, "Linking a decade of forest decline in the south-west of Western Australia to bioclimatic change," *Australian Forestry*, vol. 76, no. 3–4, pp. 164–172, Dec. 2013, doi: 10.1080/00049158.2013.844055.

[14] R. Costa, H. Fraga, P. M. Fernandes, and J. A. Santos, "Implications of future bioclimatic shifts on Portuguese forests," *Regional Environ. Change*, vol. 17, no. 1, pp. 117–127, Jan. 2017, doi: 10.1007/s10113-016-0980-9.

[15] A. Ramcharan et al., "Soil property and class maps of the conterminous United States at 100-meter spatial resolution," *Soil Sci. Soc. Amer. J.*, vol. 82, no. 1, pp. 186–201, Jan. 2018, doi: 10.2136/sssaj2017.04.0122.

[16] L. Poggio et al., "SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty," *Soil*, vol. 7, no. 1, pp. 217–240, Jun. 2021, doi: 10.5194/soil-7-217-2021.

[17] T. Hengl et al., "African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning," *Sci. Rep.*, vol. 11, no. 1, Mar. 2021, Art. no. 6130, doi: 10.1038/s41598-021-85639-y.

[18] F. Liu et al., "Mapping high resolution national soil information grids of China," *Sci. Bull.*, vol. 67, no. 3, pp. 328–340, Feb. 2022, doi: 10.1016/j.scib.2021.10.013.

[19] S. E. Fick and R. J. Hijmans, "WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas," *Int. J. Climatol.*, vol. 37, no. 12, pp. 4302–4315, Oct. 2017, doi: 10.1002/joc.5086.

[20] D. N. Karger et al., "Climatologies at high resolution for the Earth's land surface areas," *Sci. Data*, vol. 4, no. 1, Sep. 2017, Art. no. 170122, doi: 10.1038/sdata.2017.122.

[21] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *Int. J. Climatol.*, vol. 25, no. 15, pp. 1965–1978, Dec. 2005, doi: 10.1002/joc.1276.

[22] J. J. Lembrechts et al., "Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing," *Glob. Ecol. Biogeogr.*, vol. 28, no. 11, pp. 1578–1596, Nov. 2019, doi: 10.1111/geb.12974.

[23] E. Bazzato, L. Rosati, S. Canu, M. Fiori, E. Farris, and M. Marignani, "High spatial resolution bioclimatic variables to support ecological modelling in a Mediterranean biodiversity hotspot," *Ecol. Model.*, vol. 441, Feb. 2021, Art. no. 109354, doi: 10.1016/j.ecolmodel.2020.109354.

[24] B. Morgan and B. Guénard, "New 30 m resolution Hong Kong climate, vegetation, and topography rasters indicate greater spatial variation than global grids within an urban mosaic," *Earth Syst. Sci. Data*, vol. 11, no. 3, pp. 1083–1098, Jul. 2019, doi: 10.5194/essd-11-1083-2019.

[25] P. V. Bolstad, L. Swift, F. Collins, and J. Régnière, "Measured and predicted air temperatures at basin to regional scales in the southern Appalachian mountains," *Agricultural Forest Meteorol.*, vol. 91, no. 3–4, pp. 161–176, Jun. 1998, doi: 10.1016/S0168-1923(98)00076-8.

[26] S. Terzago, E. Palazzi, and J. Von Hardenberg, "Stochastic downscaling of precipitation in complex orography: A simple method to reproduce a realistic fine-scale climatology," *Natural Hazards Earth Syst. Sci.*, vol. 18, no. 11, pp. 2825–2840, Nov. 2018, doi: 10.5194/nhess-18-2825-2018.

[27] Z. A. Holden, J. T. Abatzoglou, C. H. Luce, and L. S. Baggett, "Empirical downscaling of daily minimum air temperature at very fine resolutions in complex terrain," *Agricultural Forest Meteorol.*, vol. 151, no. 8, pp. 1066–1073, Aug. 2011, doi: 10.1016/j.agrformet.2011.03.011.

[28] U. Mital, D. Dwivedi, J. B. Brown, and C. I. Steefel, "Downscaled hyper-resolution (400 m) > gridded datasets of daily precipitation and temperature (2008–2019) for the East–Taylor subbasin (western United States)," *Earth Syst. Sci. Data*, vol. 14, no. 11, pp. 4949–4966, Nov. 2022, doi: 10.5194/essd-14-4949-2022.

[29] X. Zhang et al., "Deep learning-based 500 m spatio-temporally continuous air temperature generation by fusing multi-source data," *Remote Sens.*, vol. 14, no. 15, Jul. 2022, Art. no. 3536, doi: 10.3390/rs14153536.

[30] P. Xie and P. A. Arkin, "Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions," *J. Climate*, vol. 9, no. 4, pp. 840–858, 1996.

[31] R. Li, T. Huang, Y. Song, S. Huang, and X. Zhang, "Generating 1 km spatially seamless and temporally continuous air temperature based on deep learning over Yangtze River Basin, China," *Remote Sens.*, vol. 13, no. 19, Sep. 2021, Art. no. 3904, doi: 10.3390/rs13193904.

[32] Z. Duan and W. G. M. Bastiaanssen, "First results from version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure," *Remote Sens. Environ.*, vol. 131, pp. 1–13, Apr. 2013, doi: 10.1016/j.rse.2012.12.002.

[33] N. Lei et al., "Performance evaluation and improvement of CMFD's precipitation products over Shanghai City, China," *Earth Space Sci.*, vol. 10, no. 3, Mar. 2023, Art. no. e2022EA002690, doi: 10.1029/2022EA002690.

[34] N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca, and J. Kent, "Biodiversity hotspots for conservation priorities," *Nature*, vol. 403, no. 6772, pp. 853–858, Feb. 2000, doi: 10.1038/35002501.

[35] H. Le, C. Zhao, W. Xu, Y. Deng, and Z. Xie, "Anthropogenic activities explained the difference in exotic plants invasion between protected and non-protected areas at a northern subtropics biodiversity hotspot," *J. Environ. Manage.*, vol. 345, Nov. 2023, Art. no. 118939, doi: 10.1016/j.jenvman.2023.118939.

[36] Z. Luo et al., "Impacts of climate change on the distribution of Sichuan snub-nosed monkeys (*Rhinopithecus roxellana*) in Shennongjia area, China," *Amer. J Primatol.*, vol. 77, no. 2, pp. 135–151, Feb. 2015, doi: 10.1002/ajp.22317.

[37] Y. Zhang et al., "Identifying refugia and corridors under climate change conditions for the Sichuan snub-nosed monkey (*Rhinopithecus roxellana*) in Hubei Province, China," *Ecol. Evol.*, vol. 9, no. 4, pp. 1680–1690, Feb. 2019, doi: 10.1002/ece3.4815.

[38] D. T. Liu, J. Y. Chen, and W. B. Sun, "Distributional responses to climate change of two maple species in southern China," *Ecol. Evol.*, vol. 13, no. 9, Sep. 2023, Art. no. e10490, doi: 10.1002/ece3.10490.

[39] A. An, Y. Zhang, L. Cao, Q. Jia, and X. Wang, "A potential distribution map of wintering Swan Goose (Anser cygnoides) in the middle and lower Yangtze River floodplain, China," *Avian Res.*, vol. 9, no. 1, Dec. 2018, Art. no. 43, doi: 10.1186/s40657-018-0134-5.

[40] H. Q. Li, X. H. Liu, J. H. Wang, L. G. Xing, and Y. Y. Fu, "Maxent modelling for predicting climate change effects on the potential planting area of tuber mustard in China," *J. Agricultural Sci.*, vol. 157, no. 5, pp. 375–381, Jul. 2019, doi: 10.1017/S0021859619000686.

[41] T. H. Booth, H. A. Nix, J. R. Busby, and M. F. Hutchinson, "bioclim+: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies," *Diversity Distrib.*, vol. 20, no. 1, pp. 1–9, Jan. 2014, doi: 10.1111/ddi.12144.

[42] S. Peng, Y. Ding, W. Liu, and Z. Li, "1 km monthly temperature and precipitation dataset for China from 1901 to 2017," *Earth Syst. Sci. Data*, vol. 11, no. 4, pp. 1931–1946, Dec. 2019, doi: 10.5194/essd-11-1931-2019.

[43] S. Peng, "1-km monthly maximum temperature dataset for China (1901-2022)," National Tibetan Plateau Data Center, Beijing, China, Apr. 2020, doi: 10.5281/zenodo.3114194.

[44] S. Peng, "1-km monthly mean temperature dataset for China (1901-2022)," National Tibetan Plateau Data Center, Beijing, China, Apr. 2020, doi: 10.11888/Meteoro.tpdc.270961.

[45] S. Peng, "1-km monthly minimum temperature dataset for China (1901-2022)," National Tibetan Plateau Data Center, Beijing, China, Apr. 2020, doi: 10.5281/zenodo.3114194.

[46] S. Peng, "1-km monthly precipitation dataset for China (1901-2022)," National Tibetan Plateau Data Center, Beijing, China, Mar. 2020, doi: 10.5281/zenodo.3185722.

[47] M. New, M. Hulme, and P. Jones, "Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology," *J. Climate*, vol. 12, no. 3, pp. 829–856, Mar. 1999, doi: 10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2.

[48] S. Mouatadid, S. Easterbrook, and A. R. Erler, "A machine learning approach to non-uniform spatial downscaling of climate variables," in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2017, pp. 332–341, doi: 10.1109/ICDMW.2017.49.

[49] S. Praskievicz, "Downscaling climate-model output in mountainous terrain using local topographic lapse rates for hydrologic modeling of climate-change impacts," *Phys. Geogr.*, vol. 39, no. 2, pp. 99–117, Mar. 2018, doi: 10.1080/02723646.2017.1378555.

[50] B. Pang, J. Yue, G. Zhao, and Z. Xu, "Statistical downscaling of temperature with the random forest model," *Adv. Meteorol.*, vol. 2017, pp. 1–11, 2017, doi: 10.1155/2017/7265178.

[51] A. Elnashar et al., "Downscaling TRMM monthly precipitation using Google Earth Engine and Google Cloud Computing," *Remote Sens.*, vol. 12, no. 23, Nov. 2020, Art. no. 3860, doi: 10.3390/rs12233860.

[52] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[53] W. W. Immerzeel, M. M. Rutten, and P. Droogers, "Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula," *Remote Sens. Environ.*, vol. 113, no. 2, pp. 362–370, Feb. 2009, doi: 10.1016/j.rse.2008.10.004.

[54] G. Matheron, "Principles of geostatistics," *Econ. Geol.*, vol. 58, no. 8, pp. 1246–1266, Dec. 1963, doi: 10.2113/gsecongeo.58.8.1246.

[55] E. J. Pebesma, "Multivariable geostatistics in S: The GSTAT package," *Comput. Geosci.*, vol. 30, no. 7, pp. 683–691, Aug. 2004, doi: 10.1016/j.cageo.2004.03.012.

[56] M. J. M. Cheema and W. G. M. Bastiaanssen, "Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus Basin," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2603–2627, Apr. 2012, doi: 10.1080/01431161.2011.617397.

[57] W. Tan, L. Tian, H. Shen, and C. Zeng, "A new downscaling-calibration procedure for TRMM precipitation data over Yangtze River economic belt region based on a multivariate adaptive regression spline model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4702819, doi: 10.1109/TGRS.2021.3087896.

[58] Y. Chen et al., "A new downscaling-integration framework for high-resolution monthly precipitation estimates: Combining rain gauge ob-

servations, satellite-derived precipitation data and geographical ancillary data," *Remote Sens. Environ.*, vol. 214, pp. 154–172, Sep. 2018, doi: 10.1016/j.rse.2018.05.021.

[59] R. Franke, "Scattered data interpolation: Tests of some methods," *Math. Comp.*, vol. 38, no. 157, pp. 181–200, 1982, doi: 10.1090/S0025-5718-1982-0637296-4.

[60] O. Babak and C. V. Deutsch, "Statistical approach to inverse distance interpolation," *Stochastic Environ. Res. Risk Assessment*, vol. 23, no. 5, pp. 543–553, Jul. 2009, doi: 10.1007/s00477-008-0226-6.

[61] B. S. Daya Sagar, Q. Cheng, and F. Agterberg, Eds., *Handbook of Mathematical Geosciences: Fifty Years of IAMG*. Berlin, Germany: Springer, 2018, doi: 10.1007/978-3-319-78999-6.

[62] D. N. Karger, Y. Chauvier, and N. E. Zimmermann, "CHELSA - CMIP6 1.0: A python package to create high resolution bioclimatic variables based on CHELSA ver. 2.1 and CMIP6 data," *Ecography*, vol. 2023, no. 6, Apr. 2023, Art. no. e06535, doi: 10.1111/ecog.06535.

[63] H. Chen, B. Yong, Y. Shen, J. Liu, Y. Hong, and J. Zhang, "Comparison analysis of six purely satellite-derived global precipitation estimates," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124376, doi: 10.1016/j.jhydrol.2019.124376.

[64] L. E. Flint and A. L. Flint, "Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis," *Ecol. Processes*, vol. 1, no. 1, Dec. 2012, Art. no. 2, doi: 10.1186/2192-1709-1-2.

[65] Z. Q. Ma et al., "Long-term precipitation estimates generated by a downscaling-calibration procedure over the Tibetan Plateau from 1983 to 2015," *Earth Space Sci.*, vol. 6, no. 11, pp. 2180–2199, Nov. 2019, doi: 10.1029/2019EA000657.

[66] H. Chai, W. Cheng, C. Zhou, X. Chen, X. Ma, and S. Zhao, "Analysis and comparison of spatial interpolation methods for temperature data in Xinjiang Uygur Autonomous Region, China," *Natural Sci.*, vol. 03, no. 12, pp. 999–1010, 2011, doi: 10.4236/ns.2011.312125.

[67] S. Kim, W. Lee, K. Shin, M. Kafatos, D. J. Seo, and H. Kwak, "Comparison of spatial interpolation techniques for predicting climate factors in Korea," *Forest Sci. Technol.*, vol. 6, no. 2, pp. 97–109, Dec. 2010, doi: 10.1080/21580103.2010.9671977.

[68] M. Wang et al., "Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China," *Remote Sens.*, vol. 9, no. 12, Dec. 2017, Art. no. 1278, doi: 10.3390/rs9121278.

[69] L. Mabit, "Erosion/disposition data derived from fallout radionuclides (FRNs) using geostatistics," in *Impact of Soil Conservation Measures on Erosion Control and Soil Quality*. Vienna, Austria: International Atomic Energy Agency, 2011, pp. 185–194.

[70] Y. Liao, D. Li, and N. Zhang, "Comparison of interpolation models for estimating heavy metals in soils under various spatial characteristics and sampling methods," *Trans. GIS*, vol. 22, no. 2, pp. 409–434, Apr. 2018, doi: 10.1111/tgis.12319.

[71] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan," *Paddy Water Environ.*, vol. 10, no. 3, pp. 209–222, Sep. 2012, doi: 10.1007/s10333-012-0319-1.

[72] Y. Shi, L. Li, and L. Zhang, "Application and comparing of IDW and Kriging interpolation in spatial rainfall information," in *Proc. Geoinformatics*, Nanjing, China, Jun. 2007, Art. no. 67531I, doi: 10.1117/12.761859.

[73] D. Brown and A. Comrie, "Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA," *Clim. Res.*, vol. 22, pp. 115–128, 2002, doi: 10.3354/cr022115.

[74] R. J. Hijmans, S. Cameron, and J. Parra, "WorldClim: Global weather stations," 2010. [Online]. Available: https://databasin.org/datasets/15a31dec689b4c958ee491ff30fcce75/

[75] Mountain Research Initiative EDW Working Group, "Elevation-dependent warming in mountain regions of the world," *Nature Clim. Change*, vol. 5, no. 5, pp. 424–430, May 2015, doi: 10.1038/nclimate2563.

[76] T. Wang, A. Hamann, D. Spittlehouse, and C. Carroll, "Locally downscaled and spatially customizable climate data for historical and future periods for North America," *PLoS One*, vol. 11, no. 6, Jun. 2016, Art. no. e0156720, doi: 10.1371/journal.pone.0156720.

[77] J. M. Sabater, "ERA5-Land monthly averaged data from 1950 to present," Copernicus Climate Change Service (C3S), Climate Data Store (CDS), 2019, doi: 10.24381/cds.68d2bb30.

[78] Y. Jing, L. Lin, X. Li, T. Li, and H. Shen, "Cascaded downscaling–calibration networks for satellite precipitation estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 1506105, doi: 10.1109/LGRS.2022.3214083.

Ruizhen Wang received the B.E. degree in vehicle engineering from Hunan University, Changsha, China, in 2015, and the M.E. degree in environmental engineering from the University of New South Wales, Sydney, Australia, in 2019. He is currently working toward the Ph.D. degree in geology with Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan, China.

His research interests include machine learning and remote sensing of environment.

Gaodian Zhou received the B.S. degree in resource prospecting engineering and the M.S. degree in mineral resource prospecting and exploration from the China University of Geosciences, Wuhan, China, in 2014 and 2017, respectively.

His research interests include semantic segmentation, remote sensing image processing, and Big Data.

Weitao Chen (Member, IEEE) received the B.E. degree in land resource management from Jiaozuo Institute of Technology, Jiaozuo, China, in 2003, the M.E. degree in quaternary geology from China University of Geoscience, Wuhan, China, in 2006, and the Ph.D. degree in environmental science and engineering from China University of Geosciences, Wuhan, China, in 2012.

He is currently a Professor with the School of Computer Science, CUG. He has authored and coauthored more than 70 papers. His main research interests include machine learning and remote sensing of environment.

Wenxi He received the B.E. degree in cartography and geographic information engineering from the China University of Geoscience, Wuhan, China, in 2012.

He is a Senior Engineer with the Wuhan Geological Survey Center, Wuhan. His main research interests include mineral resources development and mine ecological restoration remote sensing monitoring, natural resources supervision, and law enforcement support and other related work.

Siyang Wan received the B.E. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2021. He is currently working toward the M.Sc. degree in resource and environment with the Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan.

His main research interests include machine learning and remote sensing of environment.

Lunche Wang received the B.S. degree in geographic information system (GIS) from Chuzhou University, Chuzhou, China, in 2010, and the M.S. and Ph.D. degrees in geography and photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2012 and 2015, respectively.

He is currently a Professor with the School of Earth Sciences, China University of Geosciences, Wuhan. His research interests include climate change and remote sensing of the environment.