A New Downscaling-Calibration Procedure for TRMM Precipitation Data Over Yangtze River Economic Belt Region Based on a Multivariate Adaptive Regression Spline Model

Weiwei Tan, Liqiao Tian, Huanfeng Shen, *Senior Member, IEEE*, and Chao Zeng

Abstract—**High-resolution precipitation products are essential for accurate hydrological and meteorological applications. To improve the spatial resolution and accuracy of monthly satellite precipitation products, we developed a new downscaling-calibration framework with three key steps: 1) coarse-resolution satellite precipitation data are downscaled to 1-km resolution precipitation data with multivariate adaptive regression spline (MARS) model; 2) residual correction is applied to bridge the difference between the satellite precipitation data and downscaled precipitation data; and 3) the geographical differential analysis (GDA) calibration method is implemented to improve accuracy by merging the residual-corrected data with rain gauge data. In this study, geolocation variables (longitude and latitude), a digital elevation model (DEM) data, daytime and nighttime land surface temperatures, and four remote sensing indices were used to downscale monthly Tropical Rainfall Measuring Mission (TRMM) 3B43 precipitation datasets over the Yangtze River Economic Belt. The downscaled results showed that the MARS model can avoid "boxy artifact" and pixel-level anomalies, which are often found in geographically weighted regression (GWR) and random forest (RF) results. According to the validation, the step of residual correction is not necessary. With GDA calibration, the MARS-based estimated results were more accurate than the results of the other methods (i.e., GWR and RF) and original TRMM products. Therefore, the developed MARS-based downscaling-calibration procedure can improve not only the spatial resolution but also the quality of the TRMM 3B43 products.**

Index Terms—**"Boxy artifact" anomalies, calibration, pixel-level anomalies, precipitation, spatial downscaling, Tropical Rainfall Measuring Mission (TRMM) 3B43.**

I. INTRODUCTION

PRECIPITATION is an essential component in meteorology, ecology, and hydrology, as it is a major driving force

in the global water cycle [1], [2]. High-spatial-resolution precipitation data are crucial inputs for ecological and hydrological models. Traditional precipitation data are generally derived from point measurements (rain gauges), but point-based measurements cannot provide spatial precipitation information. Gridded precipitation data are typically generated with spatial interpolation methods based on point measurements. To obtain high-spatial-resolution precipitation data, different algorithms for the spatial interpolation of rain gauge data were proposed [3], [4]. However, the sparse density of rain gauges makes interpolation inaccurate and unrepresentative [5], [6], especially in complex mountainous regions [7]. Weather radar systems are also used to provide spatial precipitation information, but the lack of radar-rainfall estimate validation restricts its application for hydrologic modeling [8], [9]. Furthermore, radar-rainfall estimates are affected by various sources of uncertainties [10]. In addition, weather radar systems have a limited range and are generally used to monitor extreme events over a limited span of time, thereby making them less suitable for long-term assessments [11].

Satellite remote sensing has become an important resource for estimating gridded precipitation products at a global scale [12]. Several satellite precipitation products are generated by the Global Precipitation Climatology Project (GPCP) [13]–[15], the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks climate data record (PERSIANN-CDR) [16], the Tropical Rainfall Measuring Mission (TRMM) project [17], the Global Satellite Mapping of Precipitation (GSMap) project [18], the Global Precipitation Measuring (GPM) Core Observatory project [19], and so on. However, such satellite-based products have an approximate spatial resolution of 0.1° or lower [20]. The spatial resolution of such products is too coarse to match high-resolution requirements when applied to hydrological and meteorological studies at the small-basin scale. Therefore, numerous downscaling algorithms were developed to downscale satellite precipitation datasets.

Statistical downscaling methods have been widely used to improve the spatial resolution of satellite-based precipitation data. TRMM 3B43 products were downscaled from 0.25° to 1 km based on the assumption of an exponential relationship between precipitation and the normalized difference vegetation index (NDVI) at the annual scale [11]. As the

spatial patterns of precipitation are influenced by multiple environmental variables, the multiple linear regression (MLR) model was used to downscale TRMM data based on the NDVI and digital elevation model (DEM) data [21]. The MLR downscaling scheme was also used to downscale TRMM 3B43 products based on the relationship between precipitation and related environmental factors, including orographical variables and meteorological conditions [22]. In addition, the geographical differential analysis (GDA) calibration procedure was employed to downscale TRMM 3B43 products with an exponential or quadratic downscaling regression model. Calibrated 1-km monthly TRMM products show satisfactory agreement with rain gauge data [23], [24]. The aforementioned downscaling studies assumed that the relationship between precipitation and explanatory variables is constant in space. However, several studies indicated that the relationship between precipitation and land surface characteristics is spatially varied [25], [26], which contradicts the assumption of regression models. Geographically weighted regression (GWR) is a regional regression method that is widely used to establish the nonhomogeneous relationships between a dependent variable and explanatory variables. Thus, several studies used the GWR model to explore the spatially heterogeneous relationship between precipitation and related environmental factors (e.g., NDVI and DEM). The downscaled results proved that the GWR-based model outperforms the exponential, quadratic, and MLR models [25], [27]–[29].

Apart from the NDVI and DEM as the commonly used factors related to precipitation, other studies have demonstrated that the spatial patterns of precipitation are influenced by other land surface characteristics. Many studies indicated a strong positive correlation between precipitation and the NDVI [30], [31]. As a result, the NDVI was used as one of the key predictors of satellite precipitation data in numerous downscaling studies [11], [21], [24]. As vegetation's sensitivity to precipitation is cumulative rather than instantaneous [32], numerous studies demonstrated that the response of the NDVI to rainfall typically lags by one to three months [33], [34], especially in humid areas. Despite this finding, the enhanced vegetation index (EVI) can improve the relatively poor empirical relationship between precipitation and the saturated NDVI in humid regions [35]. The downscaled results obtained by Shi and Lei [33] showed that the performance of the EVI-based model surpasses that of the NDVI-based model. Because the normalized difference water index (NDWI) is sensitive to vegetation liquid water content, Gao [36] and Ezzine et al. [37] used the NDWI as a satisfactory predictor to downscale TRMM 3B43 data. In addition, the land surface water index (LSWI) was also used to monitor vegetation water content, and Chandrasekar et al. [38] analyzed the relationship between the LSWI and precipitation. Schultz and Halpert [39] found that coupling land surface temperature (LST) with the NDVI improved the precision and accuracy compared with the use of the NDVI alone. Jing et al. [40] demonstrated that introducing LSTs of both daytime and nighttime to the downscaling model to complement the NDVI–precipitation relationship and DEM–precipitation relationship is feasible. Besides, the inclusion of geolocations (longitude and latitude)

can be beneficial to the performance of the downscaling model because precipitation is spatially heterogeneous, and the average variable importance values of geolocations were higher than those of the other variables [40]. Precipitation influences the soil water content; thus, considering the effects of precipitation on soil moisture (SM) is essential [25], [41]. Many studies have found the vegetation temperature condition index (VTCI) can monitor the SM status [42], [43]. As summarized above, geolocation variables (longitude and latitude), the DEM, daytime LST (LSTD) and nighttime LST (LSTN), and four remote sensing indices (i.e., EVI, NDWI, LSWI, and VTCI) were taken into account as land surface characteristics in this study.

As for the spatial downscaling algorithms, machine learning techniques perform better than traditional regression algorithms (e.g., the exponential model, the polynomial model, and the MLR model) due to their ability to deal with the complex nonlinear relationships among variables [33], [40], [44]–[46]. A variety of typical machine learning algorithms were used to downscale satellite precipitation data [20], [33], [40], [41], [44], [45], including the classification and regression tree (CART), k-nearest neighbor (k-NN), support vector machine (SVM), random forest (RF), and Cubist [47] models. The aforementioned machine learning methods have potential application value for downscaling satellite precipitation products. The aforementioned statistical downscaling methods can be classified into traditional regression methods and machine learning methods.

The cumulative precipitation in areas with abundant precipitation is a natural spatially continuous variable at the annual and monthly scales. Therefore, ideally, monthly and annual downscaled rainfall fields should also be relatively smooth, with no sharp changes. However, numerous studies show that downscaled precipitation products encounter two types of problems. The first problem is "boxy artifact" anomalies caused by coarse pixel boundaries from the original TRMM data [see Fig. 1(a) and (b)]. Large fluctuations around edges are unlikely to occur in the real world [48]. The second problem is pixel-level anomalies [41], [48]. For example, in Fig. 1(c), the value of the center downscaled pixel is nearly double the surrounding pixels, which is unrealistic in the downscaled precipitation field [48]. This problem typically derives from two main sources. First, with various high-resolution explanatory factors as inputs, downscaling algorithms may become unstable and produce increased noise. Second, numerous studies found that TRMM products contain systematic anomalies for inland water bodies, which typically leads to overestimation [41], [49], [50]. Such types of coarse-resolution systematic anomalies may also lead to pixel-level anomalies in downscaled products.

Therefore, in this study, we used a new robust downscaling algorithm called the multivariate adaptive regression spline (MARS) model to address the aforementioned problems due to its characteristics of smoothness and regional modeling with smooth nonlinear splines [51]. In addition, we developed an MARS-based downscaling-calibration procedure to improve the spatial resolution and quality of monthly TRMM 3B43 data over the Yangtze River Economic Belt (YREB)
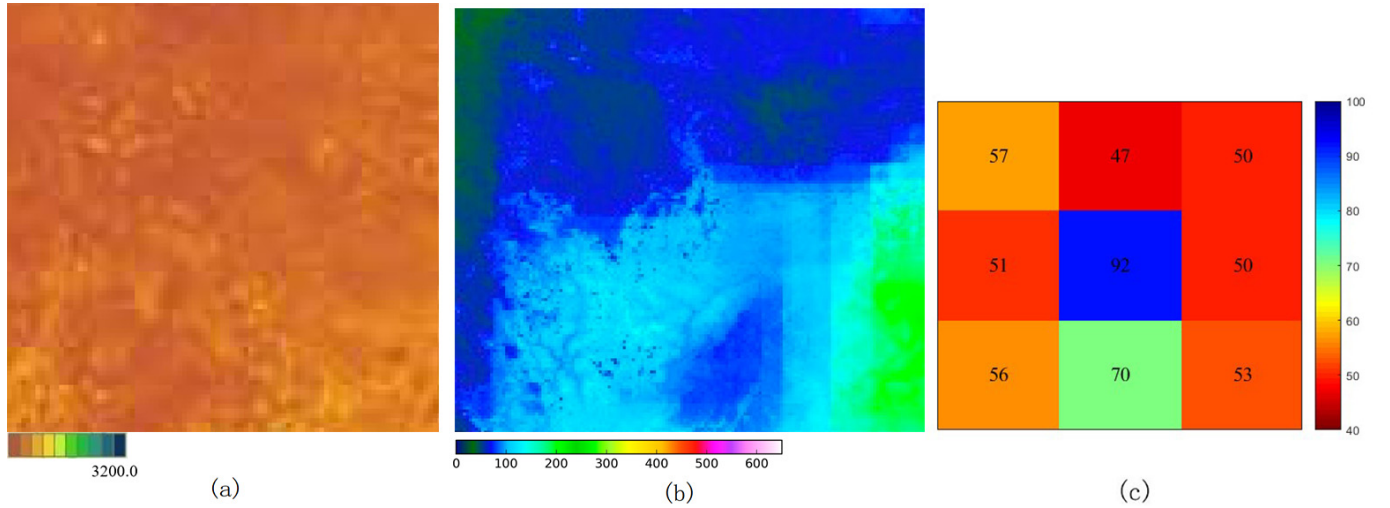
Fig. 1. Example of (a) "boxy artifact" anomaly in downscaled TRMM by using GWR [41], (b) "boxy artifact" anomaly in downscaled TRMM by using RF [45], and (c) pixel-level anomaly in downscaled TRMM [48].

for the period of 2006–2013. As two typical algorithms in downscaled satellite-based precipitation products, GWR and RF were compared with the MARS algorithm. The main objectives of this study were to: 1) introduce a new regression algorithm based on MARS to downscale coarse-resolution satellite precipitation data at the monthly scale; 2) compare the downscaled MARS-based results with the GWR- and RF-based results; and 3) assess whether the downscaled MARS-based results present the most reasonable patterns with respect to spatial variation.

II. STUDY AREA AND DATASETS

A. Study Area

The YREB was selected as the study area (see Fig. 2), which runs across Eastern China, Central China, and Western China, covering coastal areas in the east and vast inland areas [52]. With an area of approximately $2\,050\,000$ km^2 and elevation ranging from -48 to 6738 m, the climate in the YREB is diverse due to its wide area and complex topography. The YREB region experiences abundant rainfall, and the annual precipitation of between 700 and 1800 mm increases from the northwest to the southeast. Fig. 3 shows the average monthly precipitation over the entire region based on TRMM 3B43 data and ground observations for the period of 2006–2013. The TRMM 3B43 data and the ground observations show a similar trend. A significant increase in precipitation amounts can be observed from January to July, but a significant decrease can be seen from August to December.

B. TRMM Satellite Precipitation Dataset

The satellite precipitation dataset used in this study derived from data from the TRMM, which is a joint mission launched by NASA and the National Space Development Agency of Japan in November 1997 [17]. The main instruments in the TRMM are the TRMM microwave imager, precipitation radar sensor, and the Visible and Infrared Radiometer System [17]. TRMM precipitation data were widely used

in recent years due to the relatively large spatial coverage and availability of long-time series observations [53]. Consequently, many studies focused on downscaling TRMM products. The latest version of the monthly TRMM 3B43 data, that is, version 7, was chosen for this study [54]. The TRMM 3B43 data are generated at a spatial resolution of 0.25°, covering between latitude 50° N and latitude 50° S. The dataset for the period of 2006–2013 was downloaded from https://mirador.gsfc.nasa.gov. Moreover, all the monthly TRMM data, which represents precipitation (mm/hour) intensity, were converted to the cumulative amount of precipitation (mm/month).

C. Remote Sensing Index Datasets

In this study, the EVI data were derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD13A3 product at a monthly 1-km resolution to replace NDVI data. The EVI index is given by

$$\text{EVI} = G \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + C_1 \times \rho_{\text{red}} - C_2 \times \rho_{\text{blue}} + L} \tag{1}$$

where ρ is atmospherically corrected or partially atmosphere corrected surface reflectance, L is the canopy background adjustment that addresses nonlinear, differential NIR, and red radiant transfer through a canopy, and c_1 and c_2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band [35].

The two water indices (NDWI and LSWI) were calculated from the surface reflectance Terra MODIS MOD09A1 product at an eight-day/500-m resolution, and the MOD09A1 data were averaged to a monthly 1-km scale. The water indices are given by

$$\text{NDWI} = \frac{\rho_{\text{nir}} - \rho_{\text{swir1}}}{\rho_{\text{nir}} + \rho_{\text{swir1}}} \tag{2}$$

$$\text{LSWI} = \frac{\rho_{\text{nir}} - \rho_{\text{swir2}}}{\rho_{\text{nir}} + \rho_{\text{swir2}}} \tag{3}$$

where ρ_{nir}, ρ_{swir1}, and ρ_{swir2} denote band 2, band 5, and band 7 of the MOD09A1 product, respectively.
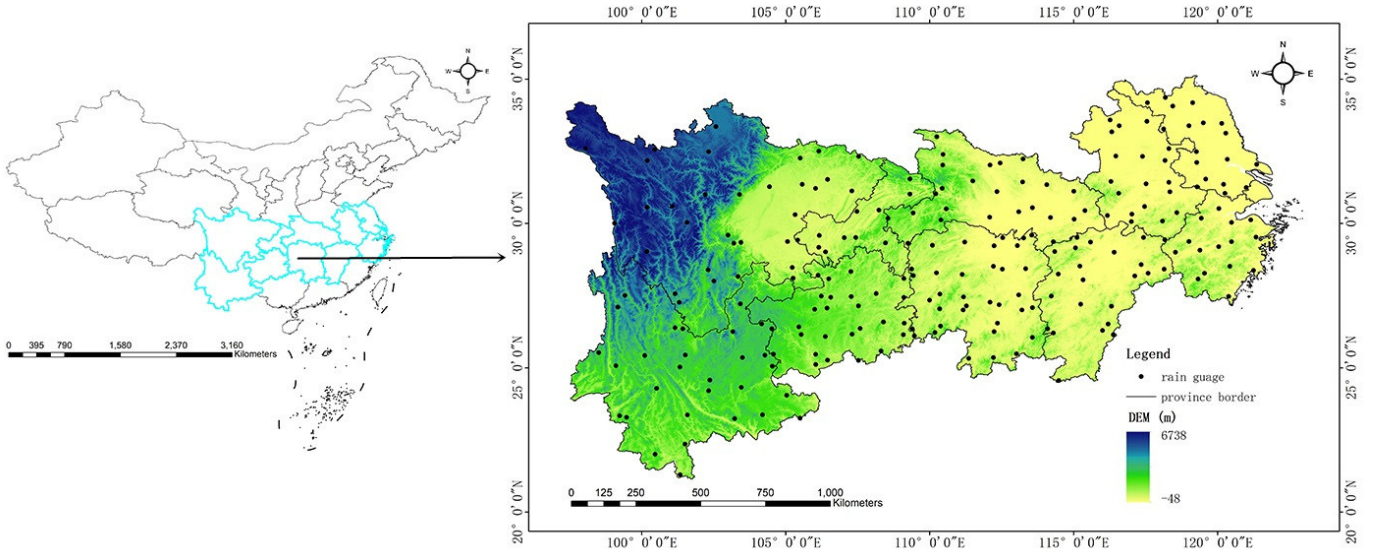
Fig. 2.   Location of the YREB; black dots indicate the location of rain gauges.
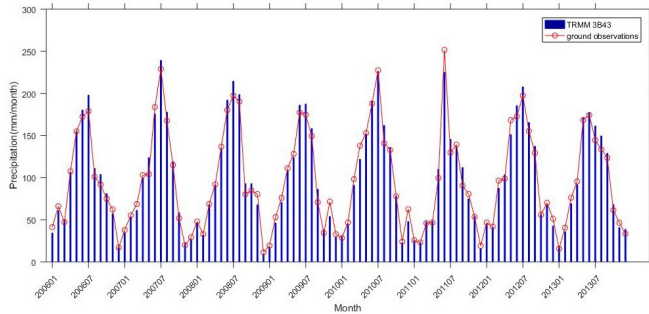

Fig. 3.   Average monthly precipitation in the YREB for 2006–2013 from TRMM 3B43 and rain gauge data.

TABLE I
DATASETS USED IN THIS STUDY

Variable	Dataset	Spatial resolution
Satellite precipitation	TRMM 3B43 V7	0.25°
EVI	MOD13A3	1 km
NDWI/LSWI	MOD09A1	1 km
Day/night LSTs	MOD11A1	1 km
VTCI	MOD13A3, MOD11A1	1 km
DEM	SRTM	90 m
Gauge-based precipitation	——	Point

The VTCI index was calculated based on the LST/EVI feature space [42], [43]. The VTCI index is given by

$$\text{VTCI} = \frac{T_{\max} - T_s}{T_{\max} - T_{\min}} \tag{4}$$

where T_s is the observed LST for a given pixel whose EVI value is EVI_i, and $T_{\max}$ and $T_{\min}$ are the corresponding highest and smallest LST that have the same EVI_i value [43].

D. LST Data

The MOD11A1 product can provide daily LST information, but missing pixels exist in large areas each day; thus, ensuring seamless downscaled results is difficult [55], [56]. The Geospatial Data Cloud platform has composited the MOD11A1 daily LST products to a monthly temporal scale by the maximum value composite method. In this study, LST datasets at a monthly 1-km resolution for LSTD and LSTN were obtained from the Geospatial Data Cloud (http://www.gscloud.cn/). The missing pixels were filled in the monthly composite LST datasets.

E. Shuttle Radar Topography Mission (SRTM) DEM

The DEM data from the SRTM were generated by the National Geospatial-Intelligence Agency and NASA at a

3-arc-second resolution (90 m). The DEM data were obtained from http://www.gscloud.cn and were upscaled to 1-km resolution by pixel averaging, to match other auxiliary data.

F. Rain Gauge Data

For China, rain gauge data from international exchange stations are included in the Global Precipitation Climatology Center monitoring product and used for the bias correction of TRMM products. In this study, 46 international exchange stations in the YREB were excluded for the fair evaluation of TRMM products [38], [56]. Quality-controlled data from 234 automatic rain gauges covering the YREB (see Fig. 2) were used to validate the accuracy of the downscaled TRMM data. The rain gauge dataset was downloaded from http://cdc.cma.gov.cn.

Table I summarizes all datasets that were used in this study, including the TRMM 3b43 dataset, remote sensing datasets, and rain gauge-based measurements.

III. METHODOLOGY

A. Multivariate Adaptive Regression Spline Model

MARS is a nonparametric regression technique [51]. As a flexible regression algorithm, MARS is widely used in science and engineering fields [57], [58]. Specifically, MARS is used to temporally downscale daily rainfall to subdaily

rainfall [59], [60]. However, to the best of our knowledge, experiments have yet to be conducted on the MARS model to spatially downscale satellite-based precipitation products. In this study, the MARS algorithm was used to derive high-spatial-resolution precipitation products.

The MARS algorithm includes the forward selection and backward elimination. In the forward selection process, let be N and P so that $Z = (X, Y)$ is our training dataset with $\dim(Y) = N$ and $\dim(X) = N \times P$. The MARS model has the form of

$$
\begin{aligned}
\overline{Y} &= \sum_{m=0}^{M} a_m B_m(X) \\
&= \sum_{m=0}^{M} a_m B_m(x_1, \ldots, x_p) \\
&= a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [C(x_{v(k,m)} \mid s_{km}, t_{-km}, t_{km}, t_{+km})] \quad (5)
\end{aligned}
$$

where $\overline{Y}$ is the estimated precipitation and x_i denotes the auxiliary predictors, which includes longitude, latitude, EVI, DEM, LSTD, LSTN, VTCI, LSWI, and NDWI. The MARS algorithm has two forms of basis function (BF): the truncated piecewise linear function and the cubic spline function. Friedman [51] pointed that the principal disadvantage of piecewise linear curves is the discontinuity of the first derivative (infinite second derivative) at each knot location. This causes the curve to be cosmetically unappealing to some extent. Therefore, the piecewise linear function was replaced with a corresponding truncated cubic function for ensuring continuous derivatives of the model (see Fig. 4). The truncated cubic spline function was selected, with the form of

$$
\begin{aligned}
&C(x \mid s = +1, t_-, t, t_+) \\
&= \begin{cases} 0, & x \leq t_- \\ p_+(x - t_-)^2 + r_+(x - t_+)^3, & t_- < x < t_+ \\ x - t, & x \geq t_+ \end{cases} \quad (6) \\
&C(x \mid s = -1, t_-, t, t_+) \\
&= \begin{cases} -(x - t), & x \leq t_- \\ p_-(x - t_+)^2 + r_-(x - t_+)^3, & t_- < x < t_+ \\ 0, & x \geq t_+ \end{cases} \quad (7)
\end{aligned}
$$

with $t_- < x < t_+$. In formulas (8)–(11), p_+, r_+, p_-, and r_- were set, with the form of

$$
\begin{aligned}
p_+ &= (2t_+ + t_- - 3t)/(t_+ - t_-)^2 & (8) \\
r_+ &= (2t - t_+ - t_-)/(t_+ - t_-)^3 & (9) \\
p_- &= (3t - 2t_- - t_+)/(t_- - t_+)^2 & (10) \\
r_- &= (t_- + t_+ - 2t)/(t_- - t_+)^3 & (11)
\end{aligned}
$$

to guarantee the continuity and continuous first derivatives of $C(x \mid s = \pm 1, t_-, t, t_+)$. $B_m(X)$ is the mth truncated BF formed with a univariate cubic spline function ($K_m = 1$) or the tensor products of univariate spline functions when the number of splits $K_m > 1$ and $x_{v(k,m)}$ labels the environment predictor, and t_{-km}, t_{km}, t_{+km} represent the values of the knots. In addition, s_{km} takes the value of 1 or -1, which determines the direction (left/right) of the split (i.e., the truncated direction), and a_m is
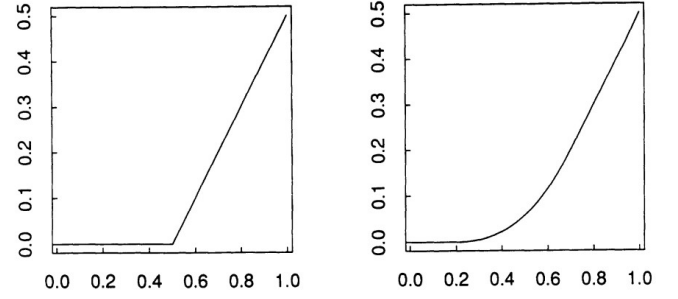

Fig. 4. Comparison of the left truncated linear function with central knot $t = 0.5$ and the corresponding truncated spline function with central knot $t = 0.5$ and side knots at $t_- = 0.2$ and $t_+ = 0.7$ [51].

the coefficient of the mth $B_m(X)$, which can be estimated with ordinary least squares by minimizing the loss function.

The solution of $B_m(X)$ ($0 \sim M$) can be divided into two processes: forward selection and backward elimination. The forward selection iteratively splits the training data and starts with an initial BF set $B_0(x) = 1$. Pairs of BFs are produced at every loop iteration with the form of

$$
\begin{aligned}
B_{2I-1}(x) &= B_l(x) b(x_v \mid p) \\
B_{2I-1}(x) &= B_l(x) b(x_v \mid \overline{p}) \quad (12)
\end{aligned}
$$

where $B_l(X)$ is a BF produced by the last iteration and x_v is an independent variable. $b(x_v \mid .)$ is a pair of truncated functions p and $\overline{p}$. For every loop iteration, the new produced BF can be identified as three parameters, including parent BF $B(x)$, x_v, and p. The object of the forward selection is to solve the three parameters (l^*, v^*, p^*) via ordinary least squares by retaining the largest decrease in training error

$$
\begin{aligned}
(l^*, v^*, p^*) = \arg\min \sum_{n=1}^{N} \Bigg[&\overline{Y}_n - \sum_{i=0}^{2I-2} a_i B_i(x_n) a_{2I-1} B_l(x_n) \\
&\times b(x_{vn} \mid p) a_{2I} B_l(x_n) b(x_{vn} \mid \overline{p}) \Bigg]^2 .
\end{aligned}
$$
$$(13)$$

The solved three parameters (l^*, v^*, p^*) are saved and used for the next iteration until the produced BFs reach the preset maximum BFs number $M_{\max}$ or satisfy the accuracy of the forward selection.

The forward selection can result in an overfitted result with an excessively large model. Thus, backward elimination was incorporated to prune the overfitted model. The backward elimination iteratively deletes the BFs that contribute the smallest increase in the residual squared error at each stage. The backward elimination is based on a lack-of-fit criterion defined by generalized cross validation (GCV) to estimate the best model Y of size λ, which is an approximation of the leave-one-out cross-validation. The GCV criterion is given as follows:

$$
\text{GCV}(\lambda) = \frac{\sum_{i=1}^{N} (y_i - Y(x_i))^2}{\left(1 - \frac{M(\lambda)}{N}\right)} \quad (14)
$$

where N is the number of BFs, $M(\lambda)$ is the effective number of parameters in the model, Y is the best estimated model, and λ is the optimum number of model terms [51], [61].

The MARS algorithm iteratively splits coarse TRMM precipitation data using truncated piecewise BF pairs produced by additively considering one environmental variable at each loop iteration. Consequently, the MARS model was used to divide the YREB TRMM images into subregions, which were determined by the nonzero range of $B_m(X)$ [see formulas (5)–(7)].

B. Brief Description of Geographically Weighted Regression and Random Forest

The GWR model can be expressed as

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \qquad (15)$$

where u_i and v_i are the longitude and latitude of location i; Y is the TRMM precipitation; x_{ik} are the kth explanatory variable of location i; $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ are the regression coefficients of location i; and ε_i is the residual. The solution of the coefficients is given as follows:

$$\beta(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \qquad (16)$$

where $W(u_i, v_i)$ is the weighting matrix obtained with kernel functions, such as the Gaussian function and bisquare function, and observations near the specific point have larger weighted values. Selection criteria, such as the Akaike information criterion (AIC), small-sample bias-corrected AIC, and cross validation (CV), are used to estimate the kernel bandwidth. Raster data have a uniform geographical distribution, and the Gaussian function is universally used in GWR [27], [28]. In this study, the fixed Gaussian function was used as the kernel function, and the CV was employed as the selection criterion. Additional details on the GWR model can be found in Brunsdon *et al.* [62].

RF is a popular ensemble learning method for classification and regression, which constructs a multitude of CARTs by randomly selecting subsets from an entire dataset. The decision-making ability of the model depends on each CART decision tree. The RF process can be divided into the following steps.

1) M ntree (number of trees) sample sets are randomly drawn from the original training dataset with replacements, and each sample set is a bootstrap sample.
2) A regression tree is constructed for each bootstrap sample. The RF model randomly trains each individual tree on the M bootstrap samples of the training dataset. For each split of the node based on the minimum Gini value, N_{try} variables are randomly selected.
3) The predicted data are estimated by averaging the predictions of the M regression trees, as follows:

$$f = \sum_{m=1}^{M} \frac{f_m(x)}{M} \qquad (17)$$

where $f_m(x)$ is the prediction of each regression tree.
4) The unsampled elements, which are called out-of-bag (OOB) data, are used to estimate the prediction error, which is called the OOB error (E_{OOB})

$$E_{OOB} = \frac{\sum_{i=1}^{n} [\overline{Y}(X_i) - Y_i]^2}{n} \qquad (18)$$

where n is the number of OOB samples, $\overline{Y}(X_i)$ is the prediction of the RF model at sample X_i, and Y_i is the target data. Two important parameters, that is, M and N_{try}, must be determined [40], [45]. The mean squared error (MSE) was chosen as the loss function. In this study, $N_{try} = \sqrt{N} = 3$ was set according to Breiman [63]. To determine the optimal value of M, the dataset was first randomly split into training and test datasets, containing 90% and 10%, respectively. Then, the training dataset was used to train the RF model, and the trained RF model was applied to the test datasets. Finally, the RF algorithm was executed with the ntree ranging from 600 to 1200; 1000 ntrees were determined as the best parameter because of the minimum MSE value for the test datasets [64]–[66]. Additional details on the RF model can be found in Breiman [63].

C. MARS-Based Downscaling-Calibration Framework

The detailed process of the MARS-based downscaling-calibration framework is given as follows.

1) *Data Preprocessing:* The significant relationship between precipitation and auxiliary factors is interrupted by regions covered with rivers, lakes, and snow. Pixels covering water bodies were detected by referencing the threshold of EVI < 0 or a constant DEM value. The detected pixels were substituted by a low-pass filter using an 11×11 moving window. The window size selection was based on Lu *et al.* [67] and Xu *et al.* [25].
2) *Data Aggregation:* To acquire coarse-resolution auxiliary data, all the auxiliary data for each month were resampled to a 0.25° resolution by averaging.
3) *MARS Model Building at Coarse Spatial Resolution:* In the forward selection process, the three key parameters to be determined are the form of the BFs and two parameters related to the stopping criterion, that is, the maximum BF number M_{max} and stopping threshold. The cubic spline function was selected as the form of the BFs. The established model in this forward selection process will gradually stabilize as M_{max} increases. M_{max} was set to 120, and the threshold was set to the default value of 1e−4 [68]. The forward selection process stops when the number of added BFs reaches the maximum number M_{max} or when adding a new BF changes R^2 of the training model to less than the threshold. In the MARS backward elimination step, the GCV criterion was used for pruning the model. The best BF number and prediction model were determined with the minimum GCV [68].
4) *TRMM Downscaling*: To generate the preliminary downscaled precipitation products at a 1-km resolution, the MARS model established in step 3) at a 0.25° resolution was applied to all the environmental factors at a 1-km resolution.
5) *Residual Correction*: In any model, differences will inevitably exist between the original TRMM data and downscaled precipitation products. To eliminate differences, residual correction was commonly adopted in
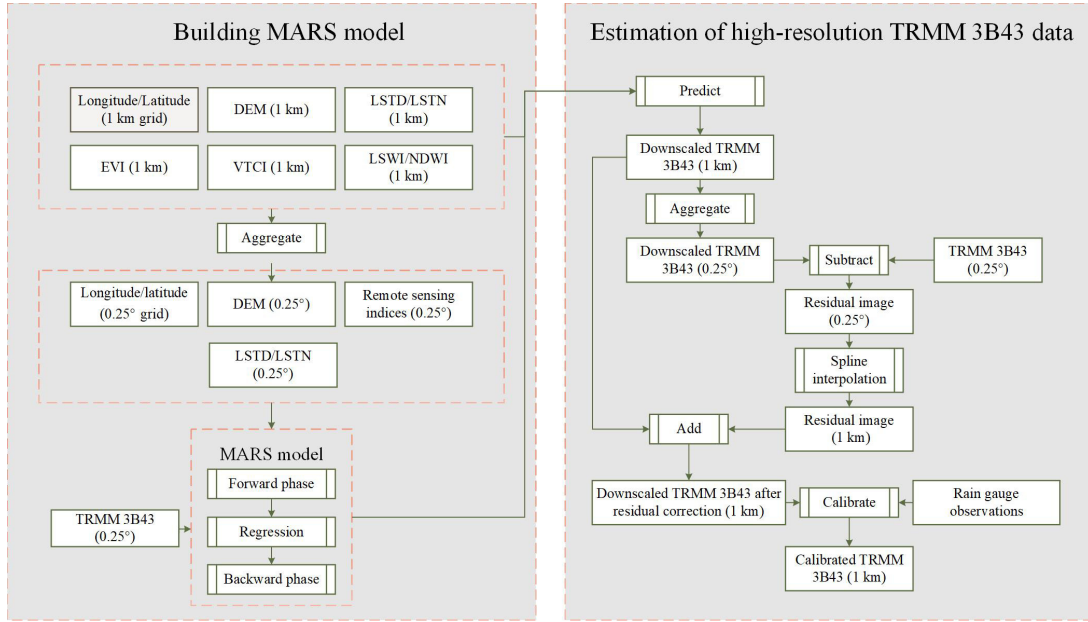
Fig. 5. Flowchart of the MARS-based downscaling-calibration framework.

previous studies [11], [21], [24], [25], [40], 41]. First, the predicted downscaled precipitation results obtained in step 4) were aggregated at a 0.25° resolution. Second, the residual images were obtained by subtracting the aggregated results from the original TRMM data and then interpolated them to a 1-km resolution with a minimum curvature spline interpolator. Previous studies demonstrated that the spline interpolator outperforms other interpolators, such as inverse distance weighting (IDW) and kriging [11], [24]. Finally, the corrected downscaled results were obtained by adding the interpolated residual images to the estimated downscaled images in step 4).

6) *GDA Calibration*: The GDA calibration method was incorporated into the calibration of the downscaled TRMM data [24], [33]. The main GDA steps are given as follows.

   a) The differences between the downscaled TRMM data and rain gauge data were computed.

   b) The point-based difference data were interpolated to a 1-km resolution with the IDW method.

   c) The 1-km difference map was added to the downscaled TRMM data to acquire the final calibrated result.

The flowchart of the MARS-based downscaling-calibration process is provided in Fig. 5, and the main steps for the GWR- and RF-based downscaling-calibration processes are the same as those for the MARS.

D. Validation

The meteorological station rainfall dataset was used to evaluate the accuracy of the original TRMM products and the downscaled precipitation products based on the different models. In this study, to assess the calibrated TRMM products,

a tenfold CV method was employed to use all the rain gauge data for the GDA calibration or verification. The rain gauge data were randomly divided into ten equal sample sets. The tenfold CV method was employed to randomly take nine sample sets from the ten sample sets for the GDA calibration. The remaining sample set was used to validate the calibrated TRMM data. The validation step was repeated ten times until all the rain gauge data were included in the validation. In addition, a hand-out CV method was also used to evaluate the estimated precipitation products. Three commonly used criteria were chosen for the validation, that is, the coefficient of determination (R^2), the mean absolute error (MAE), and the bias. These criteria are calculated as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}(Y_i - \overline{Y})(O_i - \overline{O})\right]^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{19}$$

$$\text{MAE} = \sum_{i=1}^{n} \frac{|Y_i - O_i|}{n} \tag{20}$$

$$\text{Bias} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} O_i} - 1 \tag{21}$$

where Y_i and O_i denote the TRMM precipitation data and rain gauge precipitation data at location i, respectively.

IV. RESULTS

A. Downscaling and Calibration Results

In this study, we downscaled the monthly TRMM 3B43 data for 2006–2013 over the YREB. Separate datasets for each month comprised the TRMM precipitation data, and aggregated auxiliary data were independently imported into the MARS model. The wettest month in 2012, namely, July, was selected as an example. The original TRMM data, the preliminary downscaled TRMM data, the downscaled TRMM data after residual correction, the downscaled TRMM data after
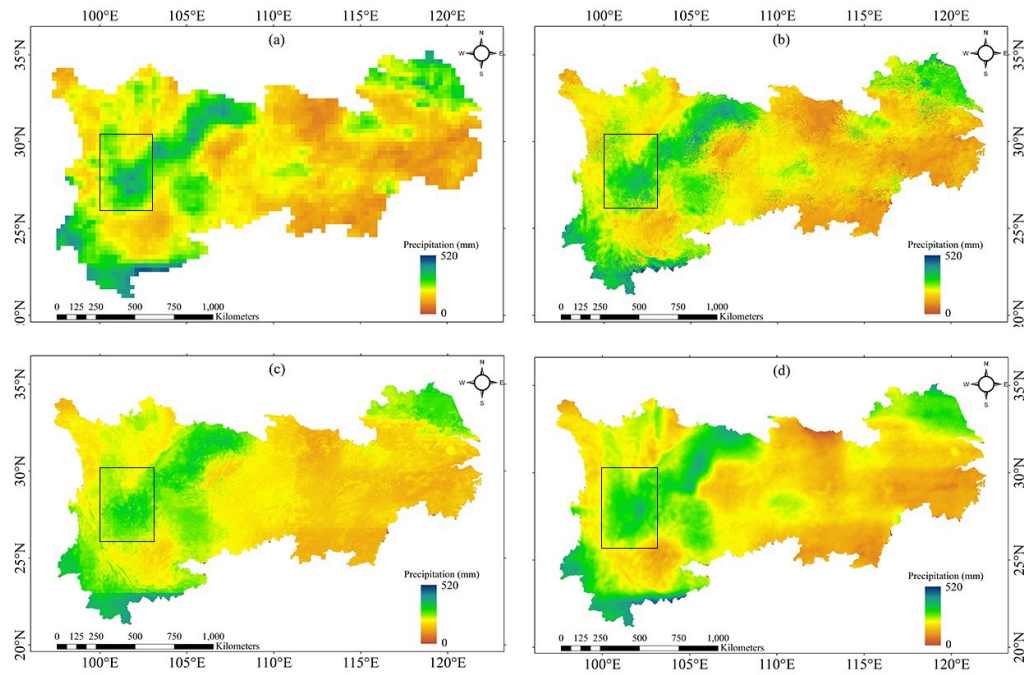
Fig. 6. Comparison between (a) TRMM_Raw, (b) GWR_Raw, (c) RF_Raw, and (d) MARS_Raw over the YREB in July 2012.
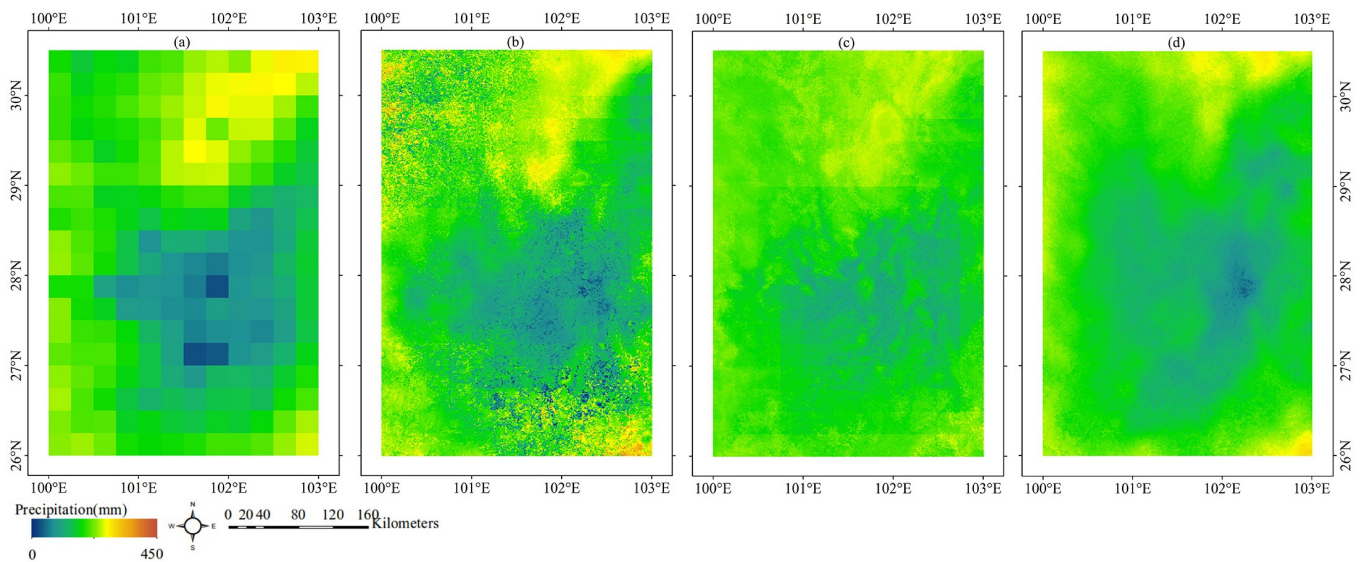

Fig. 7. Comparison between (a) TRMM_Raw, (b) GWR_Raw, (c) RF_Raw, and (d) MARS_Raw for a subregion in the YREB in July 2012.

GDA calibration with residual correction, and the downscaled TRMM data after GDA calibration without residual correction based on MARS model were labeled as TRMM_Raw, MARS_Raw, MARS_Res, MARS_ResGda, and MARS_Gda, respectively, and so were the other two models.

The coarse TRMM precipitation data over the YREB in July 2012 was shown in Fig. 6(a). The preliminary downscaled results based on GWR, RF, and MARS were shown in Fig. 6(b)–(d), respectively. The downscaled results [see Fig. 6(b)–(d)] show more spatial information than Fig. 6(a). A subregion was shown in Fig. 7 for an intuitive comparison. "Boxy artifacts" phenomenon can still be found in

some regions, such as the box region in Fig. 6 (refer to Fig. 7). Apparently, Fig. 7(b) and (c) retains some borderlines derived from coarse TRMM images, while the "boxy artifacts" structures disappear in Fig. 7(d). It is essential to find out the effects of residual correction and the GDA calibration on downscaled results. Downscaled results after residual correction based on GWR, RF, and MARS were shown in Fig. 8(b)–(d), respectively. All of the corrected results were highly consistent with the original TRMM; thus, the residual correction was an effective technique for rectifying information distortion. The calibrated results with residual correction were shown in Fig. 9. As shown in Fig. 9(b)–(d), the spatial
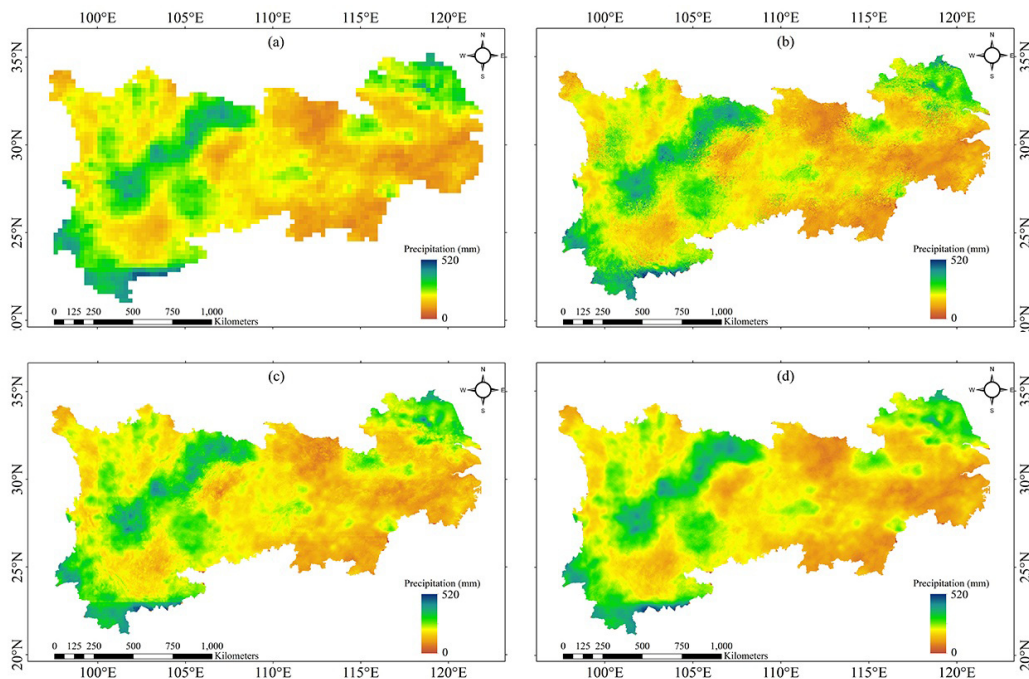
Fig. 8. Comparison between (a) TRMM_Raw, (b) GWR_Res, (c) RF_Res, and (d) MARS_Res over the YREB in July 2012.

Fig. 9. Comparison between (a) TRMM_Raw, (b) GWR_ResGda, (c) RF_ResGda, and (d) MARS_ResGda over the YREB in July 2012.

pattern of the GDA-calibrated results was altered because the gauge-based precipitation information was merged with the residual-corrected data. In addition, the calibrated results without residual correction were also provided in Fig. 10(b)–(d). We found that the calibrated results with residual correction were closer to the TRMM_Raw, and the differences between the calibrated results with and without GDA calibration were not so large.

B. Validation Results

The accuracy of the original TRMM data for each month in 2006–2013 was first evaluated as a standard reference for a comparison with the downscaled results. The assessment results were displayed in Table II, which shows that the original TRMM 3B43 products over the YREB were overestimated (average bias = 8.38%). The values of the three comparison metrics (i.e., R^2, MAE, and bias) for the downscaled TRMM
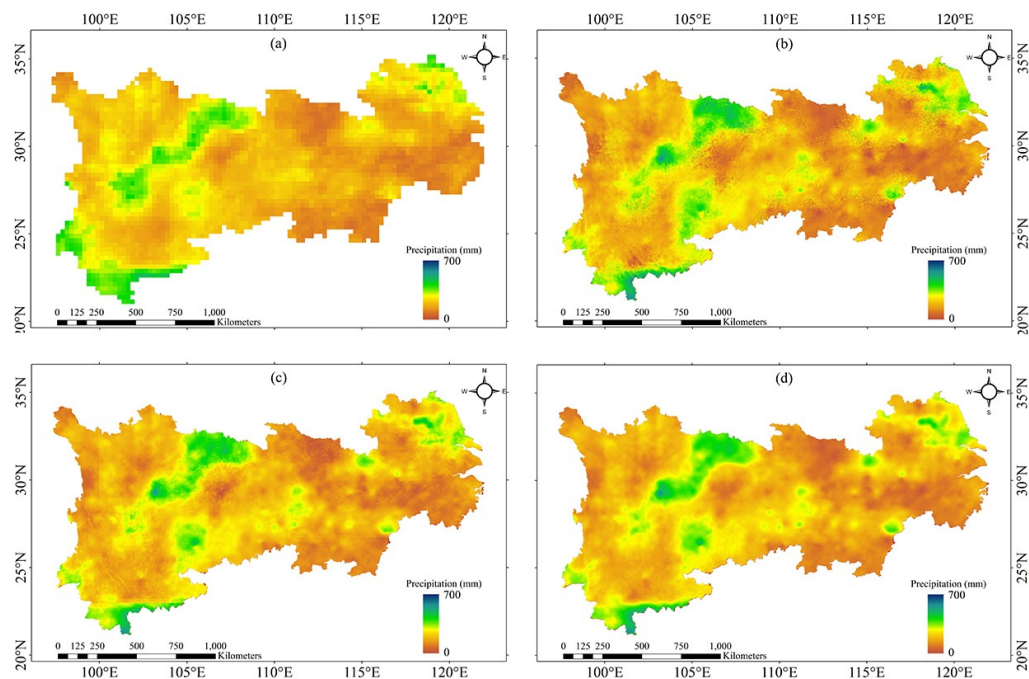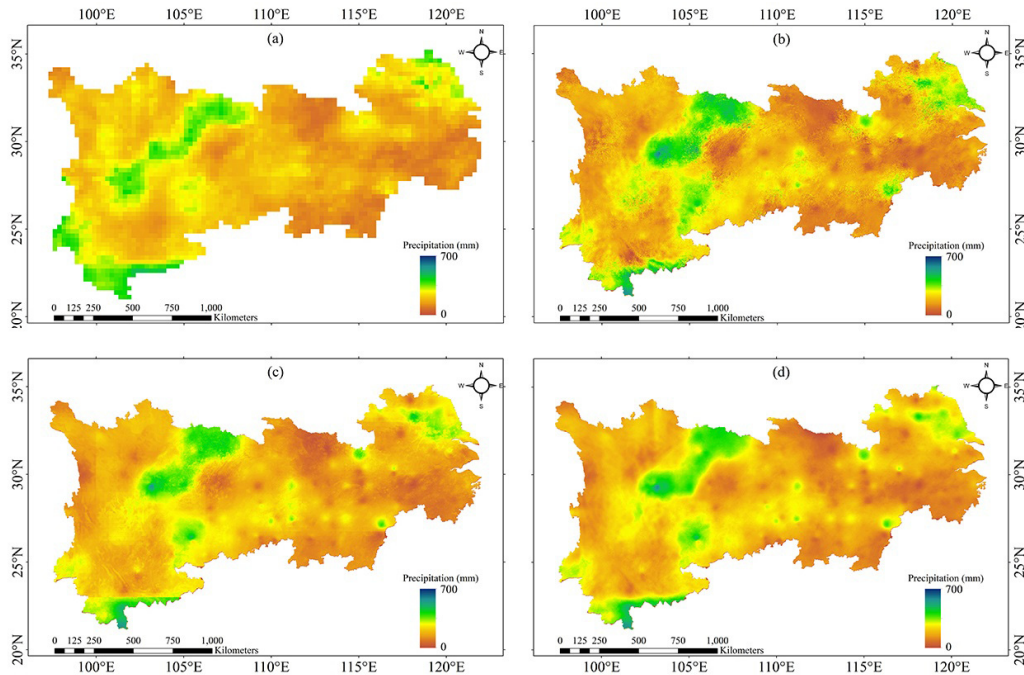
Fig. 10. Comparison between (a) TRMM_Raw, (b) GWR_Gda, (c) RF_Gda, and (d) MARS_Gda over the YREB in July 2012.

TABLE II

AVERAGED METRICS FOR THE ORIGINAL TRMM DATA, DOWNSCALED TRMM DATA, AND DOWNSCALED TRMM DATA WITH RESIDUAL CORRECTION BASED ON THE THREE DIFFERENT ALGORITHMS, VALIDATED WITH THE RAIN GAUGE DATA EACH MONTH DURING 2006–2013

Month	TRMM_Raw			GWR_Raw			RF_Raw			MARS_Raw			GWR_Res			RF_Res			MARS_Res		
	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)		MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)
Jan.	0.77	9.58	14.45	0.70	11.22	15.34	0.77	9.75	14.10	0.80	9.35	14.95	0.69	11.37	14.77	0.75	9.94	14.25	0.77	9.51	14.66
Feb.	0.83	12.26	10.46	0.76	14.41	10.19	0.83	12.40	9.20	0.85	12.08	11.39	0.76	14.77	10.04	0.81	12.72	9.72	0.83	12.27	10.57
Mar.	0.85	17.16	9.35	0.80	20.11	10.39	0.84	16.96	7.74	0.86	16.52	9.92	0.79	20.71	10.38	0.82	17.71	7.84	0.85	17.00	9.57
Apr.	0.76	25.31	9.43	0.66	29.13	8.84	0.74	25.13	8.05	0.77	25.40	9.46	0.66	29.85	9.09	0.74	25.75	8.84	0.76	25.32	9.30
May.	0.65	34.14	7.50	0.55	39.04	6.56	0.62	35.18	6.74	0.65	34.43	7.65	0.56	39.30	6.86	0.64	34.41	6.87	0.65	34.42	7.65
Jun.	0.63	48.47	7.17	0.52	55.27	6.61	0.58	53.20	6.75	0.62	49.71	7.55	0.54	54.28	6.83	0.61	50.61	7.45	0.63	48.44	7.50
July.	0.60	52.50	5.81	0.47	61.58	5.94	0.50	60.84	8.47	0.56	55.09	6.12	0.53	58.24	5.88	0.58	55.02	7.98	0.61	52.14	5.91
Aug.	0.54	45.35	6.44	0.45	50.03	6.85	0.46	49.82	8.31	0.50	47.53	6.71	0.48	49.27	7.40	0.53	46.50	8.03	0.55	45.07	7.08
Sept.	0.55	31.41	7.89	0.46	34.89	7.45	0.50	34.55	10.45	0.54	31.85	7.98	0.49	34.61	8.25	0.52	33.13	10.58	0.56	31.30	8.33
Oct.	0.70	18.41	4.89	0.63	20.49	4.74	0.67	20.14	6.35	0.72	17.87	4.77	0.63	20.78	4.20	0.68	19.20	6.18	0.71	17.88	4.31
Nov.	0.68	15.89	7.36	0.60	17.93	7.32	0.65	16.78	8.14	0.68	15.72	7.28	0.60	18.24	7.08	0.65	16.82	8.38	0.68	15.97	7.28
Dec.	0.64	10.51	9.77	0.59	12.09	8.81	0.66	10.39	9.16	0.67	10.20	10.14	0.58	12.34	8.60	0.61	10.65	8.94	0.64	10.43	9.73
Average	0.68	26.75	8.38	0.60	30.52	8.25	0.65	28.76	8.62	0.68	27.14	8.66	0.61	30.31	8.28	0.66	27.71	8.75	**0.69**	**26.65**	8.49

data with and without residual correction based on the three models were also provided in Table II. The downscaled results using MARS ($R^2 = 0.68$ and MAE = 27.14) on average, in general, performed better than GWR ($R^2 = 0.60$ and MAE = 30.52) and RF ($R^2 = 0.65$ and RMSE = 28.76) before residual correction, but the Bias (Bias = 8.66%) based on MARS is greater than GWR (Bias = 8.25%) and RF (Bias = 8.62%). In the further, we can find that the downscaled TRMM based on those models is not accurate than the original TRMM data ($R^2 = 0.68$). After residual correction, the accuracy of downscaling products based on

MARS ($R^2 = 0.69$ and MAE = 26.65) was slightly improved better than the original TRMM data ($R^2 = 0.68$ and MAE = 26.75), but GWR ($R^2 = 0.61$ and MAE = 27.14) and RF ($R^2 = 0.66$ and MAE = 27.71) still produced lower accuracy than the original TRMM data in terms of R^2 and MAE. Besides, all the residual-corrected downscaling results still have been largely overestimated. The values of the three metrics for the residual-corrected TRMM data show only a slight or no improvement with the residual correction step.

In the residual correction stage, the rain gauge information was not merged with the residual-corrected products; thus,

TABLE III

AVERAGED METRICS FOR CALIBRATED TRMM DATA WITH AND WITHOUT RESIDUAL CORRECTION BASED ON THE THREE DIFFERENT ALGO-
RITHMS VALIDATED BY TENFOLD CV METHOD EACH MONTH DURING 2006–2013

Month	GWR_ResGda			RF_ResGda			MARS_ResGda			GWR_Gda			RF_Gda			MARS_Gda		
	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)
Jan.	0.79	8.16	1.23	0.82	7.48	1.01	0.82	7.27	1.10	0.72	9.50	1.62	0.85	7.28	1.38	0.88	6.27	1.38
Feb.	0.86	10.34	1.56	0.87	9.81	1.60	0.88	9.52	1.31	0.79	12.07	1.57	0.89	9.21	1.38	0.91	8.31	1.36
Mar.	0.83	15.81	1.12	0.86	14.57	1.22	0.86	13.93	1.18	0.78	17.42	1.13	0.86	13.82	0.88	0.88	12.48	1.11
Apr.	0.78	21.98	0.87	0.80	21.29	0.99	0.80	20.52	0.91	0.71	25.65	1.07	0.80	20.92	1.08	0.83	19.26	1.11
May.	0.66	31.53	0.79	0.66	32.6	1.93	0.69	29.68	0.58	0.61	34.85	0.54	0.66	31.53	0.62	0.69	29.50	0.38
Jun.	0.63	46.45	1.13	0.64	45.02	1.17	0.66	44.11	0.77	0.58	52.58	1.14	0.65	46.52	0.98	0.68	44.44	0.84
July.	0.63	48.57	0.79	0.63	48.03	0.72	0.65	47.04	0.41	0.53	56.86	0.62	0.59	51.70	0.68	0.62	49.61	0.49
Aug.	0.57	41.97	0.92	0.58	41.76	0.79	0.57	41.62	0.67	0.49	45.86	0.89	0.51	44.67	0.89	0.54	43.60	0.75
Sept.	0.58	29.41	1.04	0.59	28.61	0.62	0.60	28.47	0.52	0.51	32.00	0.61	0.57	28.85	1.36	0.60	28.26	0.80
Oct.	0.74	16.70	0.96	0.76	15.84	0.72	0.77	15.55	1.01	0.69	18.62	0.66	0.76	16.22	0.87	0.79	14.89	0.76
Nov.	0.73	15.50	1.17	0.74	14.49	0.96	0.75	13.98	0.83	0.63	17.45	0.87	0.73	14.75	0.92	0.76	13.73	0.76
Dec.	0.66	10.62	1.66	0.69	10.07	1.67	0.69	10.09	1.20	0.62	11.75	1.12	0.73	9.76	0.99	0.74	9.02	0.88
Average	0.70	24.75	1.10	0.72	24.13	1.12	0.73	23.48	0.87	0.64	27.88	1.53	0.72	24.60	1.30	0.74	23.28	0.89

the GDA calibration step is essential to enhance the accuracy of the downscaled TRMM products. In this study, the tenfold CV method was repeated 100 times, and the averaged value was taken as the final estimate. The monthly averaged validation results for the calibrated TRMM data with and without residual correction based on the three models were provided in Table III. Compared with the TRMM_Raw data, all the calibrated results with residual correction (GWR_ResGda, RF_ResGda, and MARS_ResGda) showed improved accuracy with increased R^2 and decreased MAE and Bias on average, and the MARS model performed best among the three models and, then, the RF model. When the residual correction step was ignored, the accuracy of GWR_Gda data rapidly dropped ($R^2 = 0.64$ and MAE $= 27.88$), and the result of RF_Gda data changed little, while the accuracy of MARS_Gda data was further slightly improved with higher R^2 (0.73–0.74) and decreased MAE (23.48–23.28). Therefore, the GWR algorithm is relatively not a stable downscaling approach. As for the GWR and MARS, the residual correction step was invalid for the improvement of accuracy. The proposed MARS-based downscaling-calibration procedure can improve the spatial resolution and quality of the TRMM 3B43 products with the best performance results.

Apart from the tenfold CV method, the hand-out CV method was also employed to test the effectiveness of GDA calibration; 90% of the total rain gauge stations (210 stations) were exposed to the GDA calibration procedure, and the remaining 10% (24 stations) were used for validation purposes only. Because downscaling and residual correction are difficult to improve the TRMM accuracy, Table IV only displays the accuracy of TRMM_raw and the validation results with GDA calibration. From Table IV, the performances of the three models were similar with the results validated by the tenfold CV method. The MARS model outperformed the GWR and RF, and the residual correction step was still not necessary for the MARS-based downscaling approach.

V. DISCUSSION

A. Spatial Variations of Different Estimated TRMM Results

In Section IV, the study results showed that the GWR- and RF-based downscaled TRMM data contained some "boxy artifact" structures. In this study, a diagnostic method was used to distinguish these structures. The coefficient of variation (C_V) is a statistical measure of the dispersion of data points in a data series. This index is defined as the ratio of the standard deviation σ to the mean μ, $C_V = (\sigma/\mu)$, which is a useful statistic for comparing the degree of variation from one data series to another even if the means (μ) differ from one another. Therefore, the C_V index was chosen as a quantitative index to evaluate the degree of spatial variation for the different downscaled TRMM products. In this study, C_V was calculated with a 5×5 moving window. A low C_V value represented a relatively smooth region, whereas a high C_V value represented a region with sharp changes. Therefore, the C_V image can reveal the "boxy artifact" structures in the estimated TRMM image.

The C_V results of the calibrated TRMM products with residual correction [a subregion in Fig. 9(b)–(d)] were displayed in Fig. 11(b)–(d). In addition, the original TRMM data [see Fig. 8(a)] were resampled to a 1-km resolution with the unchanged pixel values by nearest-neighbor interpolation, and the C_V image for this resampled TRMM data was also calculated [see Fig. 11(a)]. The C_V results shown in Fig. 11(a) contained regular stripes around the borders of the original TRMM pixels, whereas the results shown in Figs. 11(b)–(c) contained some visible stripes. Fig. 11(d) indicates that the boxy artifacts disappeared completely. Thus, the GWR- and RF-based downscaling methods can not completely avoid the visible differences, and the steps of residual correction and GDA calibration were invalid for the removal of "boxy artifact" structures as well, but the MARS-based method can totally remove the "boxy artifact" structures.

TABLE IV

AVERAGED METRICS FOR THE ORIGINAL TRMM DATA, CALIBRATED TRMM DATA WITH AND WITHOUT RESIDUAL CORRECTION BASED ON THE DIFFERENT ALGORITHMS VALIDATED BY THE HAND-OUT CV METHOD EACH MONTH DURING 2006–2013

| Month	TRMM_Raw			GWR_ResGda			RF_ResGda			MARS_ResGda			GWR_Gda			RF_Gda			MARS_Gda		
	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)	R^2	MAE (mm)	Bias (%)
Jan.	0.79	8.57	17.44	0.79	8.86	7.31	0.81	7.82	4.86	0.84	6.86	3.60	0.76	8.64	6.09	0.84	7.10	5.87	0.88	5.98	3.60
Feb.	0.85	10.12	9.15	0.75	12.15	7.63	0.85	9.90	6.30	0.89	8.56	5.01	0.77	12.18	9.26	0.87	8.82	5.89	0.89	8.08	4.22
Mar.	0.84	18.26	14.84	0.76	20.36	6.42	0.82	15.60	3.78	0.85	14.38	3.70	0.80	16.94	4.19	0.87	12.83	4.01	0.88	12.52	4.04
Apr.	0.74	25.59	14.35	0.70	26.80	6.58	0.77	22.78	6.06	0.79	20.76	6.05	0.66	28.07	5.16	0.76	21.45	3.06	0.79	20.60	3.70
May.	0.72	35.90	12.12	0.67	37.70	8.01	0.74	33.14	5.46	0.75	31.66	5.93	0.68	36.77	7.23	0.70	33.04	5.88	0.72	32.31	5.68
Jun.	0.64	48.13	15.59	0.54	58.14	11.37	0.64	45.05	5.82	0.67	42.30	7.21	0.57	51.81	10.37	0.65	42.51	5.75	0.65	43.17	8.29
July.	0.62	47.48	15.34	0.51	54.36	8.31	0.60	45.59	7.49	0.64	42.41	8.36	0.51	53.33	9.65	0.61	44.74	6.96	0.63	41.33	7.99
Aug.	0.44	47.81	18.47	0.51	43.50	11.56	0.48	40.58	8.58	0.48	40.12	9.98	0.45	45.16	14.82	0.46	41.34	9.72	0.48	41.47	9.23
Sept.	0.56	33.40	21.17	0.50	35.13	11.17	0.57	31.65	11.17	0.56	31.56	12.3	0.47	34.12	11.46	0.55	31.00	8.79	0.56	31.03	9.38
Oct.	0.69	20.08	11.40	0.71	19.28	5.72	0.73	18.57	8.46	0.75	17.10	7.43	0.64	21.45	8.33	0.72	18.80	8.99	0.76	17.08	7.49
Nov.	0.73	13.66	18.29	0.72	14.24	17.44	0.78	12.53	13.35	0.80	11.39	12.31	0.74	13.56	11.06	0.78	11.55	10.68	0.80	10.85	8.30
Dec.	0.77	8.03	12.19	0.72	9.56	7.59	0.79	7.33	5.07	0.82	6.30	4.53	0.77	8.48	10.83	0.84	6.26	4.11	0.84	5.39	3.69
Average	0.70	26.42	15.03	0.66	28.34	9.09	0.71	24.21	7.20	0.74	22.78	7.20	0.65	27.54	9.04	0.72	23.29	6.64	0.74	22.49	6.30
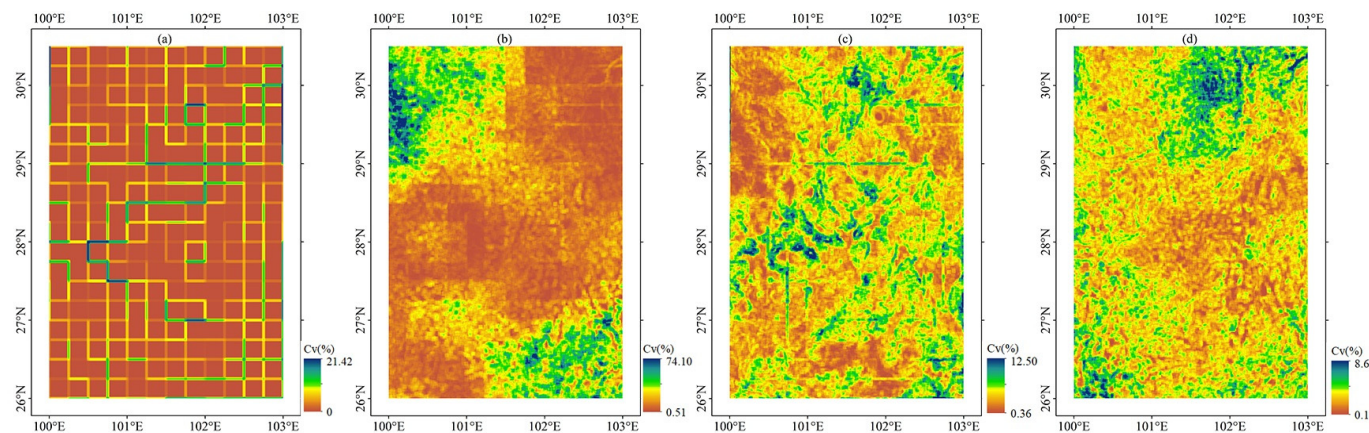
Fig. 11. Comparison between C_V images of (a) resampled TRMM 3B43 precipitation data for a local area in the YREB in July 2012 and corresponding (b) GWR_ResGda, (c) RF_ResGda, and (d) MARS_ResGda.
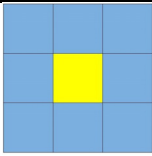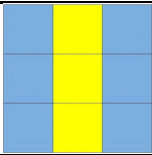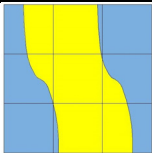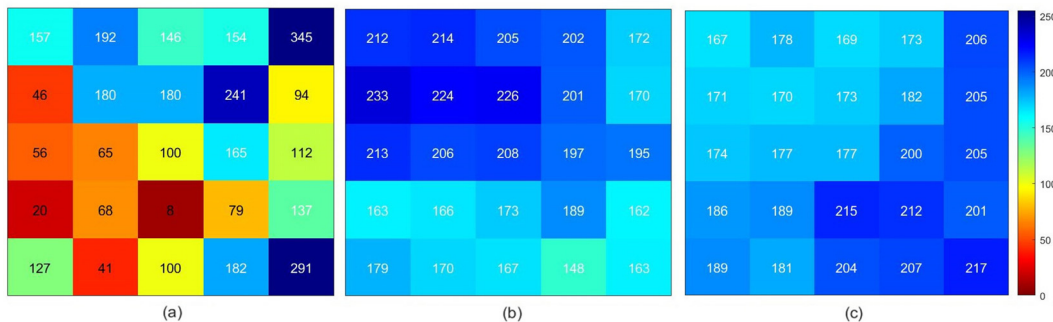

Fig. 12. Corresponding 5 × 5 matrices for (a) GWR_ResGda data, (b) RF_ResGda data, and (c) MARS_ResGda data with the maximum C_V value for a local area in the YREB in July 2012.

To assess the pixel-level anomalies, the maximum value in a 5 × 5 window was calculated with Fig. 9(b)–(d). The three corresponding 5 × 5 matrices in Fig. 9(b)–(d) were displayed in Fig. 12(a)–(c). Fig. 12(a) shows that the pixel values change dramatically within a 1-km distance, which contradicts the basic assumption in this study. The element values of Fig. 12(b) and (c) demonstrated the rationality of the downscaled precipitation fields obtained with RF or

TABLE V
DIFFERENCES BETWEEN THE THREE DOWNSCALING ALGORITHMS

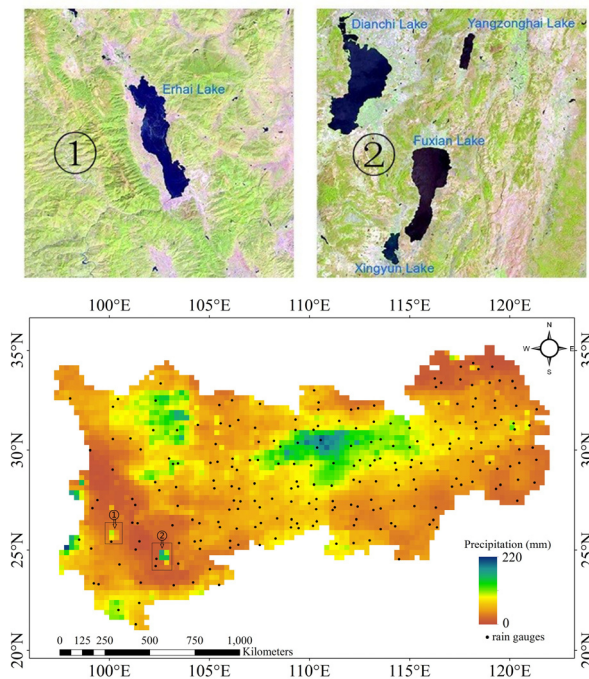| Algorithm	Downscaling formula	Effective domain of each downscaling formula
GWR	"unsmooth" multiple regression models	
RF	"unsmooth" regression trees	
MARS	truncated cubic spline functions	

Fig. 13. Systematic anomalies in the original TRMM data covering Erhai Lake, Dianchi Lake, Yangzonghai Lake, Fuxian Lake, and Xingyun Lake in October 2012.

Fig. 14. Ratio of average values of anomalies to surrounding TRMM pixels each month in 2006–2013.

MARS due to the acceptable differences in rainfall amounts. In summary, the downscaled results obtained with GWR and RF contained some unrealistic "boxy artifact" structures. Moreover, the large difference in the GWR-based downscaled rainfall field was unacceptable, but the MARS method can completely overcome the two deficiencies.

Table V lists the differences between the three downscaling algorithms. As for the GWR downscaling algorithm, the estimation of the regression coefficients is obtained by the coarse TRMM data pixel by pixel at a 0.25° resolution (as shown in Table V), which caused the unsmooth characteristic in
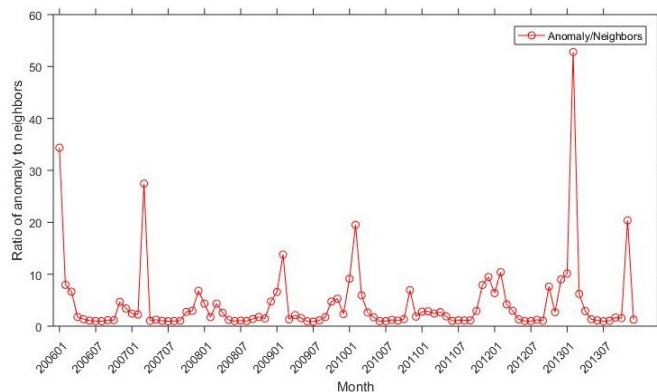
formula (15). Some boxy artifacts in GWR-based downscaled TRMM have resulted from the "unsmooth" multiple regression models. The splitting mechanism of the RF model divides the coarse TRMM pixels into multiple clusters and constructs each regression tree for each corresponding TRMM cluster (as shown in Table V). The recursive partitioning regression models (e.g., CART and RF) are piecewise constant and sharply discontinuous at split boundaries [40]. Thus, the estimated 1-km resolution TRMM data based on RF were also lacking in smoothness.

As a regression tree generalization, the difference between the MARS and other regression trees (e.g., CART and RF) is that the "hard" binary splits are replaced by "smooth" BFs [69]. The MARS model divided the entire study area into several subregions with the selected BFs. The scope of each subregion was determined by the effective domain of the associated BF (as shown in Table V) when the BF is not equal to 0 but not constricted by the boundaries of the coarse TRMM pixels. The high-resolution precipitation in each subregion was then estimated by each independent cubic spline function. It is clear that the characteristic of each independent cubic
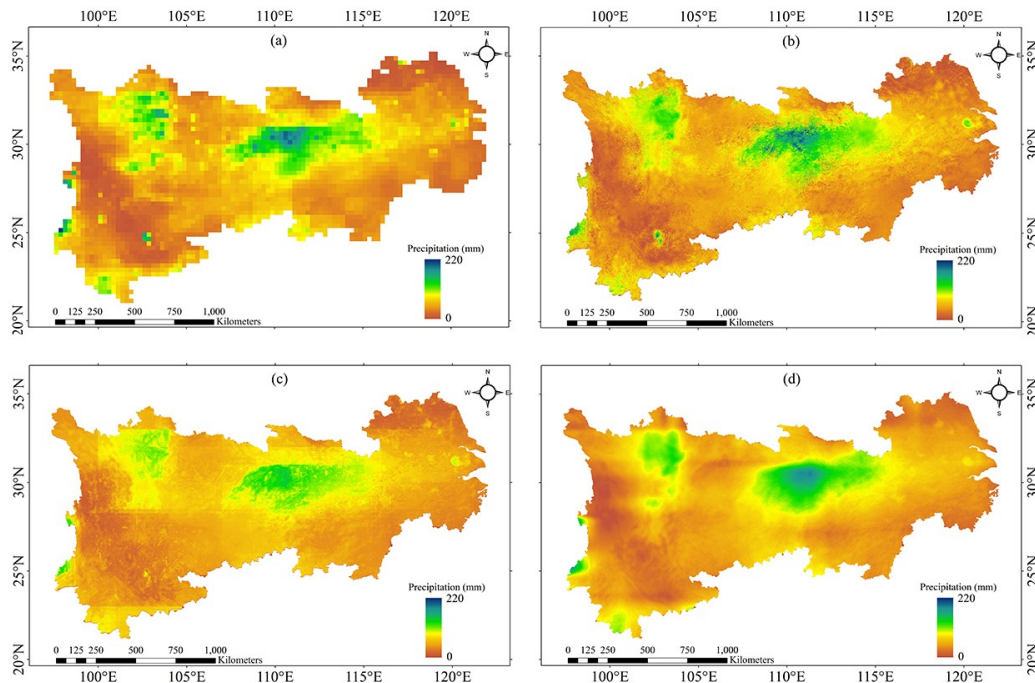
Fig. 15. Comparison between (a) TRMM_Raw, (b) GWR_Raw, (c) RF_Raw, and (d) MARS_Raw over the YREB in October 2012.

BF was smooth in each independent subregion. In addition, the cubic spline BFs had continuous derivatives at the splitting knots to ensure the smoothness of the entire study area [see formulas (8)–(11)] [48]. Therefore, MARS is a highly adaptive and flexible algorithm for simulating precipitation distribution at a fine scale.

B. Denoising Effect of MARS-Based Downscaling Model

In this study, we found that systematic anomalies in the original TRMM products over the YREB existed in numerous months in 2006–2013. These overestimated outliers in the TRMM data were in the location of various lakes, including Erhai Lake, Dianchi Lake, Yangzonghai Lake, Fuxian Lake, and Xingyun Lake. Fig. 13 displays the outliers in the original TRMM data in October 2012 as an example. Assessing the outliers by comparing them with neighboring TRMM pixels is feasible despite the absence of rain gauges in the location of the outliers. The ratios of the average values of the anomalies (for Dianchi Lake, Yangzonghai Lake, Fuxian Lake, and Xingyun Lake) to the average values of the eight neighbors adjacent to the outliers in 2006–2013 are shown in Fig. 14. The ratio of 7.63 in October 2012 confirmed the obvious overestimation of the anomalous pixels.

To assess the ability of the different models to cope with the systematic anomalies in the TRMM data, Fig. 15 exemplifies the different downscaled results to evaluate the robustness of each model. Fig. 15(b) and (c) indicates that the downscaling models using GWR and RF were prone to causing pixel-level anomalies at a 1-km resolution. GWR is a local regression algorithm that generates regression coefficients by searching adjacent data to estimate a multiple regression model [see formula (15)] for each point. For the anomalous TRMM pixels, the closer adjacent anomalous pixels are weighted greater than the adjacent normal TRMM pixels; thus, GWR is invalid for the noise reduction of TRMM anomalies. In areas with water bodies, the pixel values of the anomalous TRMM data and auxiliary environmental data at a 1-km resolution changed sharply. RF leads to pixel-level anomalies in areas with systematic anomalies in the original TRMM products, as it produces discontinuous output with an uneven response surface [70]. Inversely, the high-resolution precipitation estimation based on MARS was computed using smooth spline functions. Thus, the downscaled result was less sensitive to the original pixels with sharp variance.

To compare the downscaled results in Fig. 15 more clearly, Fig. 16 further displays the downscaling results of the TRMM anomalies in Fig. 13 over the region marked with ②. Fig. 16 demonstrates the proposed method outperformed the other two methods in noise reduction. The values of the original TRMM_Raw data ranged from 4.31 to 169.93 [see Fig. 16(a)], whereas the MARS_Raw data ranged from 1.89 to 54.55 [see Fig. 16(d)]. By contrast, the GWR_Raw and RF_Raw data retained overestimated values, which ranged from 0 to 245.12 [see Fig. 16(b)] and 11.23 to 107.91 [see Fig. 16(c)], respectively. The GWR- and RF-based downscaled results were obviously influenced by the TRMM anomalies, but the downscaled TRMM was improved by the use of MARS, with its effective noise reduction capability. After applying the GDA calibration, the calibration results with and without residual correction were displayed in Fig. 17(a)–(c) and (d)–(f), respectively. We can find the step of residual correction added the noise from the TRMM_Raw data into the GWR_ResGda, GWR_ResGda, and GWR_ResGda data [41]. When ignoring the residual correction step and taking advantage of the noise reduction ability of the MARS model, the MARS-based
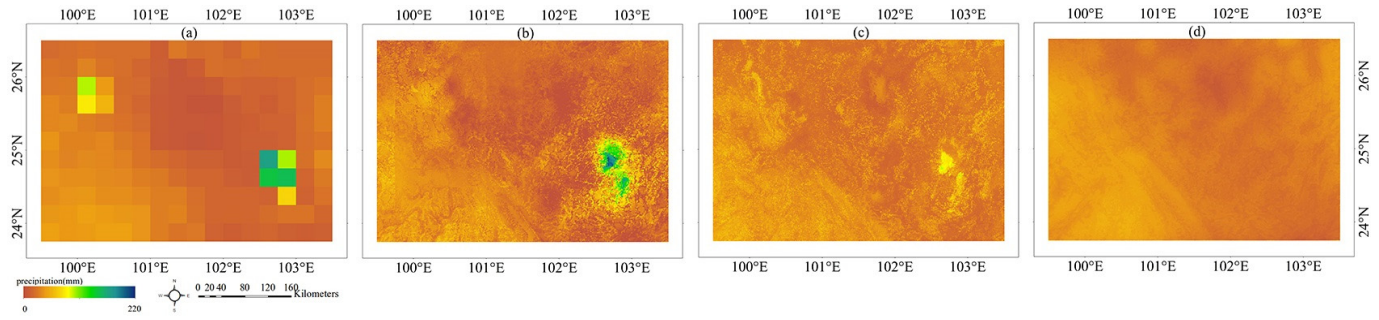
Fig. 16. Comparison between (a) TRMM_Raw data with systematic anomalies in October 2012 and corresponding (b) GWR_Raw, (c) RF_Raw, and (d) MARS_Raw.
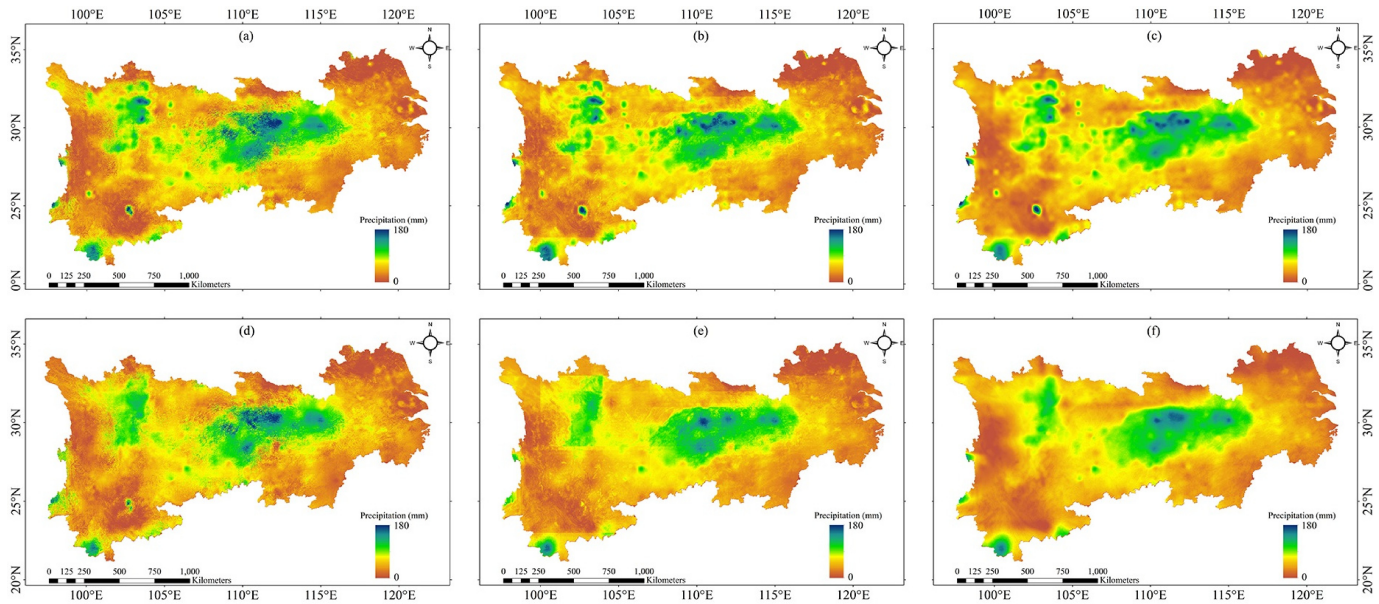

Fig. 17. Comparison between (a) GWR_ResGda, (b) RF_ResGda, (c) MARS_ResGda, (d) GWR_Gda, (e) RF_Gda, and (f) MARS_Gda in October 2012.

downscaling-calibration procedure can achieve the noise-free result [see Fig. 17(f)].

C. Importance of Variables

To explore the contribution of the variables to precipitation, the relative variable importance can be estimated with analysis of variance (ANOVA) decomposition criteria provided by the MARS method. To assess the variable importance, we selected the "nSubsets" criteria, which count the number of subsets including the variable. The larger the number of subsets containing the variable, the more important the variable.

Fig. 18(a) shows the results of the variable importance estimation for each environmental factor. Longitude and latitude provided significant contributions to the MARS model. Conversely, the EVI provided no contribution to the MARS model for July 2012. To prove the rationality of the built MARS model for July 2012, the correlation coefficients (CCs) between precipitation and each environmental variable were shown in Fig. 18(b). The MARS model deleting the EVI factor, as its relationship with precipitation (CC = 0.03) was the weakest in July 2012 over the YREB, is reasonable because the time-lagged vegetation index demonstrated poor explanatory
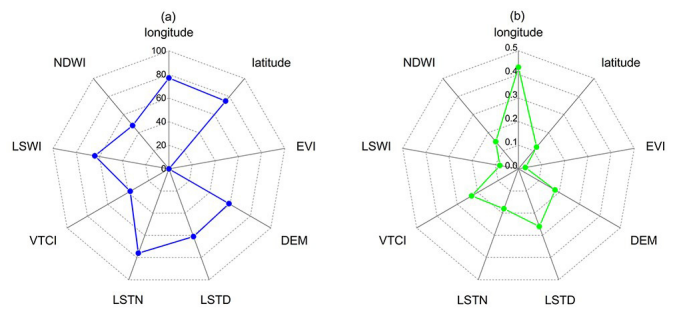

Fig. 18. Radar charts of (a) selected nSubsets of each environmental variable and (b) CCs between precipitation and environmental variables in July 2012.

power for precipitation in several months. It should be noted that the orders of the importance and correlation of variables are not necessarily identical (e.g., LSWI) because of the different calculation approaches.

Fig. 19(a) displays the selected numbers for each factor in the MARS models and average variable importance for 96 months in 2006–2013. Fig. 19(b) indicated that each MARS model selected longitude, latitude, and the DEM.
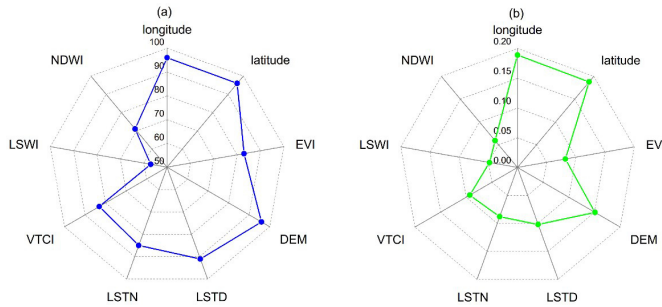
Fig. 19. Radar charts of (a) selected numbers of each environmental variable in the MARS model for all months and (b) average importance of each environmental variable in the MARS model for all months.

Fig. 20. Change curve of GCV in the forward selection and backward elimination processes.

Meanwhile, the average variable importance values of the two geolocation variables were higher than those of the other factors, followed by the DEM, LSTD, and VTCI, which can be attributed to the spatial heterogeneity of precipitation [27] and is consistent with the conclusion of Jing *et al.* [40]. The EVI and LSTN indices showed similar average importance values (approximately 0.08). However, the two water indices (i.e., LSWI and NDWI), with the lowest average importance values, were excluded in many months, which implied that water indices were not indispensable factors when explaining the precipitation over the YREB. Therefore, the MARS model has the advantage of being able to detect effective variables related to monthly precipitation, which can avoid certain relatively ineffective variables being added to the model, thereby allowing it to achieve the best precipitation estimation results in this study.

D. Advantages and Limitations of Proposed MARS-Based Downscaling Scheme

Numerous studies highlighted the accuracy of monthly downscaled precipitation products. However, few studies focused on the details of spatial variation in the downscaled rainfall field. Although Verlinde [48] pointed out the irrationality of "boxy artifact" anomalies in monthly downscaled precipitation products, these phenomena were neglected in many of the aforementioned studies. In this study, we paid attention to the spatial variation characteristics of the downscaled products and developed an MARS-based downscaling framework. The downscaled results proved that the MARS model can effectively avoid the "boxy artifact" phenomenon through its characteristics of local modeling and smoothness. In addition, the MARS model can adaptively select the effective variables in each specific month, whereas the performance of the other models, such as the GWR model, will degrade rapidly if unnecessary variables are considered.

The estimation of TRMM results based on the MARS model achieved an improved spatial resolution and higher accuracy compared with the original TRMM data. However, this downscaling-calibration scheme has several limitations, which may cause uncertainties in the downscaled results. The TRMM datasets over areas where precipitation was interrupted by human factors (e.g., agriculture and buildings) or natural factors (e.g., water bodies and snow) could not be downscaled directly using the MARS model. Aux-
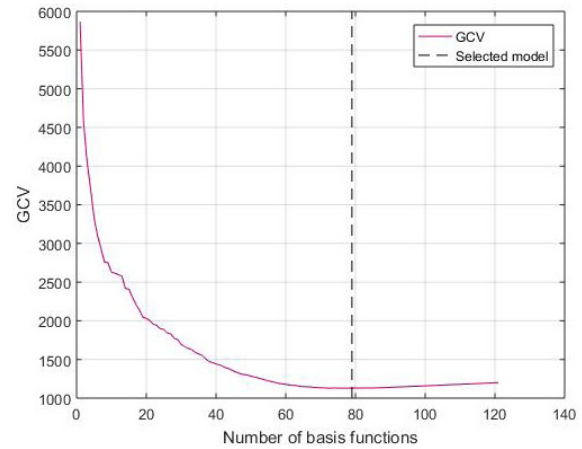
iliary data preprocessing was performed to estimate such areas by averaging the surrounding pixels, which may cause some uncertainties in the areas. Furthermore, the proposed MARS-based downscaling-calibration procedure can only be applied to basins with gauge measurements. The inconsistency between the original TRMM data and rain gauge data may enlarge the difference between the downscaled TRMM data and real precipitation data when the GDA calibration is applied over several months.

VI. CONCLUSION

In this study, a new downscaling-calibration procedure based on the MARS model is developed to estimate satellite-based TRMM 3B43 monthly precipitation data over the YREB from $0.25°$ to 1 km. The MARS model flexibly separates the study area into subregions and constructs the relationship between precipitation and explanatory variables over each subregion by using the smooth spline function. As a result, the MARS model can successfully avoid "boxy artifact" anomalies in the downscaled TRMM data, which are common problems in several widely used regression models (e.g., polynomial regression and GWR) and other regression tree models (e.g., CART and RF). With the satisfactory generalization ability of the MARS model, the downscaled TRMM data obtained with the MARS model show the most reasonable spatial distribution patterns and the best denoising results. However, pixel-level anomalies occur in the GWR- and RF-based downscaled results. The performance of the downscaled products based on the different downscaling algorithms is validated using ground observations from rain gauges. The MARS model demonstrates the best estimation performance, but the validation accuracy shows that the residual correction step is invalid for improving the quality of the downscaled TRMM data and even produces noise data from system anomalies in the original TRMM data. Thus, the residual correction step should not be considered in the downscaling-calibration framework. After the GDA calibration step, the calibrated downscaled TRMM data based on MARS are more accurate than the original TRMM data with the effect of noise reduction, whereas the results based on GWR and RF models show lower accuracy. Therefore, the MARS-based

downscaling-calibration procedure is a promising approach for satellite precipitation downscaling research. Through the backward elimination pruning step, the MARS model can select the most effective variables in each month. In this study, the three variables of longitude, latitude, and the DEM are selected by all the MARS models and, thus, can be considered as essential predictors of precipitation over the YREB. The variable importance of the geographic variables is higher than that of the other variables in the MARS model, and the two water indices (i.e., LSWI and NDWI) provide the least contribution to precipitation.

In future works, we will attempt to further enhance the quality of satellite-based downscaled precipitation products. First, the combination of satellite precipitation datasets and hydrological models for downscaling is worth exploring, which can be extended to small-basin-scale precipitation downscaling. Second, a highly robust calibration method should be developed to improve the calibration accuracy of low-quality TRMM data. Third, the MARS model shows considerable potential in generating high-temporal-resolution precipitation data. Therefore, exploring the MARS-based downscaling framework or other techniques to obtain high-temporal-resolution (daily or subdaily) satellite-based precipitation datasets would be worthwhile.

APPENDIX

The process of establishing the MARS model for July 2012 was shown in the following. Fig. 20 illustrates the curve of the GCV value of the established MARS model as the selected BF numbers changed. The GCV value decreased rapidly as the BF numbers increased from 1 to 78 and then decreased slowly as the BF numbers changed from 78 to 120. The forward selection process stopped when 120 preset BFs were added to the model. After the backward elimination process, in this example, 78 cubic BFs were selected based on the minimum GCV value. To test the stability of this established MARS model, a tenfold CV strategy was used to perform an in-depth assessment of the robustness of this model. All the TRMM precipitation samples and aggregated auxiliary data for July 2012 were randomly divided into ten equal sample sets. The idea was to take nine sample sets from the ten sample sets to establish the MARS model based on the GCV criterion. This step was repeated ten times. The selected BF numbers for the ten MARS models were 71, 77, 77, 72, 74, 81, 76, 70, 79, and 77. The selected BF numbers are all close to 78, which confirmed that the proposed MARS model was adequately robust to choose the best BF numbers based on the GCV criterion.

REFERENCES

[1] P. Goovaerts, "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall," *J. Hydrol.*, vol. 228, nos. 1–2, pp. 113–129, Feb. 2000.

[2] M. Li and Q. Shao, "An improved statistical approach to merge satellite rainfall estimates and raingauge data," *J. Hydrol.*, vol. 385, nos. 1–4, pp. 51–64, May 2010.

[3] A. Di Piazza, F. L. Conti, L. V. Noto, F. Viola, and G. La Loggia, "Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 3, pp. 396–408, Jun. 2011.

[4] S. Ly, C. Charles, and A. Degré, "Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale: A review," *Biotechnologie, Agronomie, Société Environ.*, vol. 17, no. 2, pp. 392–406, 2013.

[5] P. C. Kyriakidis, J. Kim, and N. L. Miller, "Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics," *J. Appl. Meteorol.*, vol. 40, no. 11, pp. 1855–1877, Nov. 2001.

[6] M. Strauch *et al.*, "Adjustment of global precipitation data for enhanced hydrologic modeling of tropical andean watersheds," *Climatic Change*, vol. 141, no. 3, pp. 547–560, Apr. 2017.

[7] B. Henn, A. J. Newman, B. Livneh, C. Daly, and J. D. Lundquist, "An assessment of differences in gridded precipitation datasets in complex terrain," *J. Hydrol.*, vol. 556, pp. 1205–1219, Jan. 2018.

[8] W. F. Krajewski and J. A. Smith, "Radar hydrology: Rainfall estimation," *Adv. Water Resour.*, vol. 25, nos. 8–12, pp. 1387–1394, Aug. 2002.

[9] P. V. Mandapaka, W. F. Krajewski, G. J. Ciach, G. Villarini, and J. A. Smith, "Estimation of radar-rainfall error spatial correlation," *Adv. Water Resour.*, vol. 32, no. 7, pp. 1020–1030, Jul. 2009.

[10] Q. Dai, D. Han, M. A. Rico-Ramirez, L. Zhuo, N. Nanding, and T. Islam, "Radar rainfall uncertainty modelling influenced by wind," *Hydrol. Processes*, vol. 29, no. 7, pp. 1704–1716, Mar. 2015.

[11] 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.

[12] J. S. Famiglietti, A. Cazenave, A. Eicker, J. T. Reager, M. Rodell, and I. Velicogna, "Satellites provide the big picture," *Science*, vol. 349, no. 6249, pp. 684–685, Aug. 2015.

[13] G. J. Huffman *et al.*, "The global precipitation climatology project (GPCP) combined precipitation dataset," *Bull. Amer. Meteorol. Soc.*, vol. 78, no. 1, pp. 5–20, Jan. 1997.

[14] G. J. Huffman *et al.*, "Global precipitation at one-degree daily resolution from multisatellite observations," *J. Hydrometeorol.*, vol. 2, no. 1, pp. 36–50, Feb. 2001.

[15] G. J. Huffman, R. F. Adler, D. T. Bolvin, and G. Gu, "Improving the global precipitation record: GPCP version 2.1," *Geophys. Res. Lett.*, vol. 36, no. 17, 2009, Art. no. L17808.

[16] K.-L. Hsu, H. V. Gupta, X. Gao, and S. Sorooshian, "Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation," *Water Resour. Res.*, vol. 35, no. 5, pp. 1605–1618, May 1999.

[17] C. Kummerow, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, "The tropical rainfall measuring mission (TRMM) sensor package," *J. Atmos. Ocean. Technol.*, vol. 15, no. 3, pp. 809–817, Jun. 1998.

[18] T. Kubota *et al.*, "Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2259–2275, Jul. 2007.

[19] G. J. Huffman *et al.*, "NASA global precipitation measurement (GPM) integrated multi-satellite retrievals for GPM (IMERG)," *Algorithm Theor. Basis Document*, vol. 4, p. 30, Nov. 2015.

[20] Z. Ma, Y. Zhou, B. Hu, Z. Liang, and Z. Shi, "Downscaling annual precipitation with TMPA and land surface characteristics in China," *Int. J. Climatol.*, vol. 37, no. 15, pp. 5107–5119, Dec. 2017.

[21] S. Jia, W. Zhu, A. Lű, and T. Yan, "A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3069–3079, Dec. 2011.

[22] J. Fang, J. Du, W. Xu, P. Shi, M. Li, and X. Ming, "Spatial downscaling of TRMM precipitation data based on the orographical effect and meteorological conditions in a mountainous area," *Adv. Water Resour.*, vol. 61, pp. 42–50, Nov. 2013.

[23] 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.

[24] 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.

[25] S. Xu, C. Wu, L. Wang, A. Gonsamo, Y. Shen, and Z. Niu, "A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics," *Remote Sens. Environ.*, vol. 162, pp. 119–140, Jun. 2015.

[26] G. M. Foody, "Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship," *Remote Sens. Environ.*, vol. 88, no. 3, pp. 283–293, Dec. 2003.

[27] F. Chen, Y. Liu, Q. Liu, and X. Li, "Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity," *Int. J. Remote Sens.*, vol. 35, no. 9, pp. 3074–3093, May 2014.

[28] C. Chen, S. Zhao, Z. Duan, and Z. Qin, "An improved spatial downscaling procedure for TRMM 3B43 precipitation product using geographically weighted regression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 9, pp. 4592–4604, Sep. 2015.

[29] Y. Zhang, Y. Li, X. Ji, X. Luo, and X. Li, "Fine-resolution precipitation mapping in a mountainous watershed: Geostatistical downscaling of TRMM products based on environmental variables," *Remote Sens.*, vol. 10, no. 1, p. 119, Jan. 2018.

[30] H. Iwasaki, "NDVI prediction over mongolian grassland using GSMaP precipitation data and JRA-25/JCDAS temperature data," *J. Arid Environ.*, vol. 73, nos. 4–5, pp. 557–562, Apr. 2009.

[31] J.-M. K. Onema and A. Taigbenu, "NDVI–rainfall relationship in the Semliki watershed of the equatorial Nile," *Phys. Chem. Earth, A/B/C*, vol. 34, nos. 13–16, pp. 711–721, Jan. 2009.

[32] U. Gessner, V. Naeimi, I. Klein, C. Kuenzer, D. Klein, and S. Dech, "The relationship between precipitation anomalies and satellite-derived vegetation activity in Central Asia," *Global Planet. Change*, vol. 110, pp. 74–87, Nov. 2013.

[33] Y. Shi and S. Lei, "Spatial downscaling of monthly TRMM precipitation based on EVI and other geospatial variables over the Tibetan Plateau from 2001 to 2012," *Mountain Res. Develop.*, vol. 35, no. 2, pp. 180–194, 2015.

[34] W. W. Immerzeel, R. A. Quiroz, and S. M. de Jong, "Understanding precipitation patterns and land use interaction in Tibet using harmonic analysis of SPOT VGT-S10 NDVI time series," *Int. J. Remote Sens.*, vol. 26, no. 11, pp. 2281–2296, Jun. 2005.

[35] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, nos. 1–2, pp. 195–213, Nov. 2002.

[36] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, Dec. 1996.

[37] H. Ezzine, A. Bouziane, D. Ouazar, and M. D. Hasnaoui, "Downscaling of open coarse precipitation data through spatial and statistical analysis, integrating NDVI, NDWI, elevation, and distance from sea," *Adv. Meteorol.*, vol. 2017, pp. 1–20, Jan. 2017.

[38] K. Chandrasekar, M. V. R. S. Sai, P. S. Roy, and R. S. Dwevedi, "Land surface water index (LSWI) response to rainfall and NDVI using the MODIS vegetation index product," *Int. J. Remote Sens.*, vol. 31, no. 15, pp. 3987–4005, Aug. 2010.

[39] P. A. Schultz and M. S. Halpert, "Global analysis of the relationships among a vegetation index, precipitation and land surface temperature," *Int. J. Remote Sens.*, vol. 16, no. 15, pp. 2755–2777, Oct. 1995.

[40] W. Jing, Y. Yang, X. Yue, and X. Zhao, "A comparison of different regression algorithms for downscaling monthly satellite-based precipitation over North China," *Remote Sens.*, vol. 8, no. 10, pp. 138–144, 2016.

[41] Z. Ma, Z. Shi, Y. Zhou, J. Xu, W. Yu, and Y. Yang, "A spatial data mining algorithm for downscaling TMPA 3B43 V7 data over the Qinghai–Tibet Plateau with the effects of systematic anomalies removed," *Remote Sens. Environ.*, vol. 200, pp. 378–395, Oct. 2017.

[42] I. Sandholt, K. Rasmussen, and J. Andersen, "A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status," *Remote Sens. Environ.*, vol. 79, nos. 2–3, pp. 213–224, Feb. 2002.

[43] J. Peng, A. Loew, S. Zhang, J. Wang, and J. Niesel, "Spatial downscaling of satellite soil moisture data using a vegetation temperature condition index," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 558–566, Jan. 2016.

[44] Y. Shi *et al.*, "Mapping annual precipitation across mainland China in the period 2001–2010 from TRMM3B43 product using spatial downscaling approach," *Remote Sens.*, vol. 7, no. 5, pp. 5849–5878, May 2015.

[45] X. Zhao, W. Jing, and P. Zhang, "Mapping fine spatial resolution precipitation from TRMM precipitation datasets using an ensemble learning method and MODIS optical products in China," *Sustainability*, vol. 9, no. 10, p. 1912, Oct. 2017.

[46] X. He, N. W. Chaney, M. Schleiss, and J. Sheffield, "Spatial downscaling of precipitation using adaptable random forests," *Water Resour. Res.*, vol. 52, no. 10, pp. 8217–8237, Oct. 2016.

[47] J. R. Quinlan, "Combining instance-based and model-based learning," in *Proc. 10th Int. Conf. Mach. Learn.*, 1993, pp. 236–243.

[48] J. Verlinde, "TRMM rainfall data downscaling in the Pangani Basin in Tanzania," M.S. thesis, Dept. Civil Eng. Geosci., Delft Univ. Technol., Delft, Netherlands, 2011, pp. 1–72.

[49] G. Tang, Y. Ma, D. Long, L. Zhong, and Y. Hong, "Evaluation of GPM day-1 IMERG and TMPA version-7 legacy products over mainland China at multiple spatiotemporal scales," *J. Hydrol.*, vol. 533, pp. 152–167, Feb. 2016.

[50] Y. Tian and C. D. Peters-Lidard, "Systematic anomalies over inland water bodies in satellite-based precipitation estimates," *Geophys. Res. Lett.*, vol. 34, no. 14, 2007, Art. no. L14403.

[51] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Statist.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.

[52] Y. Xu, "Logistic development along the Yangtze River economic belt," *Contemporary Logistics in China*. Singapore: Springer, 2016, pp. 121–152.

[53] D. Fan *et al.*, "Temporal downscaling of TRMM precipitation products using AMSR2 soil moisture data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 7733–7736.

[54] G. J. Huffman *et al.*, "The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *J. Hydrometeorol.*, vol. 8, no. 1, pp. 38–55, Feb. 2007.

[55] C. Zeng, D. Long, H. Shen, P. Wu, Y. Cui, and Y. Hong, "A two-step framework for reconstructing remotely sensed land surface temperatures contaminated by cloud," *ISPRS J. Photogramm. Remote Sens.*, vol. 141, pp. 30–45, Jul. 2018.

[56] G. Yang, W. Sun, H. Shen, X. Meng, and J. Li, "An integrated method for reconstructing daily MODIS land surface temperature data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 1026–1040, Mar. 2019.

[57] S. Kuter, Z. Akyurek, and G.-W. Weber, "Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines," *Remote Sens. Environ.*, vol. 205, pp. 236–252, Feb. 2018.

[58] X. Zhang, Z. Nan, J. Wu, E. Du, T. Wang, and Y. You, "Mountain permafrost distribution modeling using multivariate adaptive regression spline (MARS) in the Wenquan area over the Qinghai-Tibet Plateau," *Sci. Cold Arid Regions*, vol. 4, no. 5, pp. 361–370, 2012.

[59] X. Beuchat, B. Schaefli, M. Soutter, and A. Mermoud, "Toward a robust method for subdaily rainfall downscaling from daily data," *Water Resour. Res.*, vol. 47, no. 9, pp. 1995–2021, Sep. 2011.

[60] J. Corte-Real, X. Zhang, and X. Wang, "Downscaling GCM information to regional scales: A non-parametric multivariate regression approach," *Climate Dyn.*, vol. 11, no. 7, pp. 413–424, Sep. 1995.

[61] H. Trevor, T. Robert, and F. Jh, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.

[62] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: A method for exploring spatial nonstationarity," *Geograph. Anal.*, vol. 28, no. 4, pp. 281–298, Sep. 2010.

[63] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[64] L. Edel, J.-F. Rysman, C. Claud, C. Palerme, and C. Genthon, "Potential of passive microwave around 183 GHz for snowfall detection in the arctic," *Remote Sens.*, vol. 11, no. 19, p. 2200, Sep. 2019.

[65] A. Adhikari, M. R. Ehsani, Y. Song, and A. Behrangi, "Comparative assessment of snowfall retrieval from microwave humidity sounders using machine learning methods," *Earth Space Sci.*, vol. 7, no. 11, Nov. 2020, Art. no. e2020EA001357.

[66] M. R. Ehsani *et al.*, "Assessment of the advanced very high-resolution radiometer (AVHRR) for snowfall retrieval in high latitudes using CloudSat and machine learning," *J. Hydrometeorol.*, vol. 22, pp. 1591–1608, Apr. 2021.

[67] X. Lu *et al.*, "Correcting GPM IMERG precipitation data over the tianshan mountains in China," *J. Hydrol.*, vol. 575, pp. 1239–1252, Aug. 2019.

[68] G. Jekabsons. (2011). *ARESLab: Adaptive Regression Splines Toolbox for Matlab/Octave*. [Online]. Available: https://www.cs.rtu.lv/jekabsons

[69] R. Nisbet, J. Elder, and G. Miner, *Handbook of Statistical Analysis and Data Mining Applications*. New York, NY, USA: Academic, 2009.

[70] J. Elith and C. H. Graham, "Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models," *Ecography*, vol. 32, no. 1, pp. 66–77, Apr. 2009.

Weiwei Tan received the M.S. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2020.

He is a Research Assistant with the Southern Marine Science and Engineering Guangdong Laboratory, Zhuhai, China. His research interests focus on remote sensing image processing, remote sensing applications, and physical geography.

Huanfeng Shen (Senior Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

He is a Distinguished Professor with Wuhan University, where he serves as the Associate Dean of the School of Resource and Environmental Sciences. His research interests include remote sensing image processing, multisource data fusion, and intelligent environmental sensing. He has been a PI of two projects supported by the National Key Research and Development Program of China and six projects supported by the National Natural Science Foundation of China. He has authored or coauthored more than 150 journal citation report (JCR) articles (over 60 in IEEE journals) and published 4 books as the chief editor. His articles received 6675 citations in Web of Science (WoS) (as of May 2020).

Dr. Shen is also a fellow of the Institution of Engineering and Technology (IET) and a member of the Education Committee of the Chinese Society for Geodesy Photogrammetry and Cartography and the Theory Committee of the China Association for Geospatial Information Society. He has been named in the lists of Highly Cited Chinese Researchers by Elsevier and The World's Top 2% Scientists by Stanford University. He was a recipient of the First Prize in the Natural Science Award of Hubei Province, China, in 2011, the First Prize in Nature Scientific Award of China's Ministry of Education in 2015, and the First Prize in Scientific and Technological Progress Award of the Chinese Society for Geodesy Photogrammetry and Cartography in 2017. He is also a Senior Regional Editor of Journal of Applied Remote Sensing and an Associate Editor of Geography and Geo-Information Science and Journal of Remote Sensing.

Liqiao Tian was born in Hubei, China, in March 1980. He received the Ph.D. degree in cartography and geographic information system from Wuhan University, Wuhan, China, in June 2008.

He has been with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, since July 2008. His research interests are applications of remote sensing.

Chao Zeng received the B.S. degree in resources environment and urban–rural planning management, the M.S. degree in surveying and mapping engineering, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009, 2011, and 2014, respectively.

He was a Post-Doctoral Researcher with the Department of Hydraulic Engineering, Tsinghua University, Beijing, China. He is with the School of Resources and Environmental Science, Wuhan University. His research interests focus on remote sensing image processing and hydrological remote sensing applications.