Spatial and Temporal Downscaling of TRMM Precipitation with Novel Algorithms^①

HUIHUI ZHANG

School of Resources and Environmental Sciences, Wuhan University, Wuhan, China, and Department of Geography, University of California, Santa Barbara, Santa Barbara, California

HUGO A. LOÁICIGA

Department of Geography, University of California, Santa Barbara, Santa Barbara, California

DA HA

School of Civil Engineering, Tianjin University, Tianjin, China, and Department of Geography, University of California, Santa Barbara, Santa Barbara, California

QINGYUN DU

School of Resources and Environmental Sciences, Key Laboratory of Geographic Information System, Ministry of Education, and Key Laboratory of Digital Mapping and Land Information Application Engineering, Ministry of Natural Resources, Wuhan University, Wuhan, China

(Manuscript received 17 December 2019, in final form 29 April 2020)

ABSTRACT

Tropical Rainfall Measuring Mission (TRMM) satellite products constitute valuable precipitation datasets over regions with sparse rain gauge networks. Downscaling is an effective approach to estimating the precipitation over ungauged areas with high spatial resolution. However, a large bias and low resolution of original TRMM satellite images constitute constraints for practical hydrologic applications of TRMM precipitation products. This study contributes two precipitation downscaling algorithms by exploring the nonstationarity relations between precipitation and various environment factors [daytime surface temperature (LTD), terrain slope, normalized difference vegetation index (NDVI), altitude, longitude, and latitude] to overcome bias and low-resolution constraints of TRMM precipitation. Downscaling of precipitation is achieved with the geographically weighted regression model (GWR) and the backward-propagation artificial neural networks (BP_ANN). The probability density function (PDF) algorithm corrects the bias of satellite precipitation data with respect to spatial and temporal scales prior to downscaling. The principal component analysis algorithm (PCA) provides an alternative method of obtaining accurate monthly rainfall estimates during the wet rainfall season that minimizes the temporal uncertainties and upscaling effects introduced by direct accumulation (DA) of precipitation. The performances of the proposed downscaling algorithms are assessed by downscaling the latest version of TRMM3B42 V7 datasets within Hubei Province from 0.25° (about 25 km) to 1-km spatial resolution at the monthly scale. The downscaled datasets are systematically evaluated with in situ observations at 27 rain gauges from the years 2005 through 2010. This paper's results demonstrate the bias correction is necessary before downscaling. The high-resolution precipitation datasets obtained with the proposed downscaling model with GWR relying on the NDVI and slope are shown to improve the accuracy of precipitation estimates. GWR exhibits more accurate downscaling results than BP_ANN coupled with the genetic algorithm (GA) in most dry and wet seasons.

1. Introduction

Precipitation plays a significant role in hydrological, meteorological, and agricultural research (Li and Shao 2010; Wu and Chen 2012). Conventional observations from rain gauge stations generally provide direct and accurate measurements at local stations (Duan et al. 2016).

^① Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-19-0289.s1>.

Corresponding author: Qingyun Du, qydu@whu.edu.cn

It remains a challenge to acquire accurate precipitation over underdeveloped and mountainous regions due to the sparsity of rain gauge networks (Xu et al. 2015). The development of advanced meteorological satellites and remote sensing technology has made the estimation of precipitation possible over areas lacking rain gauge stations (Mahesh et al. 2011). The Tropical Rainfall Measuring Mission (TRMM) is a widely used precipitation estimation product. The TRMM precipitation is used for evaluating other well-known precipitation estimation products (e.g., PERSIANN-CDR and PERSIANN) over ocean and land in many studies (Sadeghi et al. 2019; Sorooshian et al. 2000). The TRMM precipitation products provide accurate estimation results over ungauged areas compared with various interpolation methods, but their spatial resolution (0.25°) is too coarse for localized studies (Duan and Bastiaanssen 2013).

Much progress has been made to downscale the TRMM products to improve the precipitation estimation's accuracy. Satellite observations of the land surface were found to be useful in downscaling satellite precipitation datasets (Jia et al. 2011; Xu et al. 2015). Jia et al. (2011) downscaled the TRMM datasets for the Qaidam Basin, China, from 0.25° to 1 km at annual temporal scale relying on the normalized difference vegetation index (NDVI) and digital elevation model (DEM) with a multiple regression model. Xu et al. (2015) further demonstrated the possibility of downscaling TRMM product with NDVI and DEM at the monthly temporal scale by downscaling the TRMM3B43 V7 precipitation datasets from 0.25° to 1 km over the eastern Tibetan Plateau and the Tian Shan Mountains with the geographically weighted regression (GWR) model. Temperature also has influence on local precipitation. The temperature–precipitation statistical association has been proposed by several authors (Isaac and Stuart 1992; Madden and Williams 1978; Trenberth and Shea 2005; Zhao and Khalil 1993). Assessing the role played by other environmental factors besides DEM and NDVI [e.g., daytime surface temperature (LTD), slope, latitude, longitude] in the downscaling of precipitation is a timely endeavor. The GWR-model based downscaling algorithm performed better than the traditional downscaling algorithms (univariate regression, multivariate regression and exponential function) when applied to the TRMM3B43 V7 at the monthly scale (Chen et al. 2015; Xu et al. 2015). The artificial neural network method was introduced for the estimation of precipitations in recent years. The performance of the genetic algorithm (GA)-based back-propagation (BP) artificial neural network (GA+BP_ANN) model has been evaluated by forecasting rainfall–runoff in Morocco with the

rainfall and the runoff values observed on the previous time period (Sedki et al. 2009), and by forecasting rainfall in the Upper Parramatta River basin, Australia using rainfall hyetograph of recording rain gauges (Nasseri et al. 2008). The results indicate the well-trained GA-based neural network model gives superior estimations when compared with the back-propagation artificial neural network (BP_ANN) or ANN alone. A substantial effort has been made to advance the TRMM downscaling algorithms. Yet, the comparison of the GWR-based and artificial intelligence-based downscaling algorithm (GA+BP_ANN) for the purpose of downscaling the TRMM precipitation products at the monthly scale has not been previously addressed.

There remain obstacles to the accurate downscaling of TRMM precipitation datasets. The first concerns the errors of the original TRMM precipitation products and related environmental remote sensing datasets. Duan et al. (2016) evaluated the quality of TRMM3B42 V7 precipitation products against interpolated rain gauge data in the Adige Basin in Italy at daily and monthly scales, indicating that further improvement of TRMM3B42 V7 gridded precipitation product is possible. In addition, the bias correction of the original TRMM product considering spatial and temporal scales has been overlooked before downscaling (e.g., Chen et al. 2015; Fang et al. 2013; Xu et al. 2015). Therefore, bias correction based on spatial and temporal scale becomes imperative before downscaling the TRMM3B42 V7 product. The second obstacle concerns the upscaling of daily precipitation to monthly precipitation. According to Mahmud et al.'s (2015) work, upscaling TRMM rainfall datasets by direct accumulation (DA) may introduce large errors in heavy thunderstorms during the wettest season, while the influence of upscaling in the dry season has not been fully addressed.

This study develops and applies a precipitation downscaling algorithm at the monthly temporal scale over regions with complex precipitation–topographic and precipitation–environment patterns of association. This paper develops a methodology to correct the differences between TRMM3B42 precipitation products and gauge station data by means of the probability density function (PDF) considering spatial–temporal scales, and to calculate accurate monthly precipitation from daily precipitation with DA and principal component analysis (PCA) in wet and dry seasons. Furthermore, a new downscaling algorithm is proposed by introducing the GWR model based on LTD, terrain slope, NDVI, altitude, longitude, and latitude as explanatory variables at the monthly temporal scale. This work applies the developed algorithm to downscale the TRMM3B42 V7 precipitation datasets in Hubei Province, China. The performance of the developed algorithm is

FIG. 1. The flowchart of the PDF correction algorithm proposed in this study.

assessed and compared with the GA+BP_ANN-based downscaling algorithm. Two GA+BP_ANN models were trained with datasets of precipitation measured by rain gauges and by TRMM satellite images, respectively.

The outline of this paper is as follows: [section 2](#) presents the error correction and downscaling methodology. [Section 3](#) introduces the study area and provides a brief description of the satellite datasets and used rain gauge station data. [Section 4](#) presents the downscaling results. The conclusions and discussion are summarized in [section 5](#).

2. Methodology

a. Bias correction with the PDF

The basic assumption of the PDF algorithm is that the gauge-based analysis of daily precipitation is unbiased. This renders the probability density function of TRMM3B42 products consistent with that of the rain gauge precipitation by adjusting the satellite precipitation and correcting the systematic error of the TRMM3B42 precipitation data. The principle of this algorithm is based on the work of [Xie and Xiong \(2011\)](#), who corrected the bias of CMORPH product by matching the PDF of the satellite data with that of the gauge data in China.

[Figure 1](#) presents a flowchart of the PDF correction procedure. The steps of the data-correction algorithm are as follows:

- 1) First, the daily precipitation is estimated by direct accumulation of a series of 3-hourly TRMM3B42 rainfall ([Duan et al. 2016](#)). The PDF algorithm does not operate at the hourly scale because there are many zero-valued distributions. Second, selecting the size of the spatiotemporal moving window: the central point of the spatial moving window is at each of the 27 rain gauge stations. Substitute the value of each pixel in TRMM daily precipitation with the weighted summation value of the surrounding pixels in the spatial moving window. Last, substitute the value of each pixel obtained from the spatial moving window with the weighted summation value of the surrounding pixels in the temporal moving window. The precipitation obtained from the rain gauge stations is processed with the temporal moving window.
- 2) The number of collected sample data must over 350 to ensure a stable probability density distribution. The threshold of 350 used in this study may be not suitable for other regions. The selection of the sample data must be based on the data's statistical stability characteristics ([Xie and Xiong 2011](#)).
- 3) Calculate the cumulative probability densities of the TRMM precipitation data and of rain gauge data. The precipitation difference Δr is calculated as follows:

$$\Delta r = R_T - R_G, \quad (1)$$

where R_T denotes the value of satellite derived precipitation; R_G denotes the value of gauge station rainfall. The same cumulative probability density value is consistent with R_T and R_G .

- 4) The revised TRMM precipitation R_S is calculated as follows:

$$R_S = R_T - \overline{\Delta r}. \quad (2)$$

where $\overline{\Delta r}$ denotes the correction value, which is estimated by the average Δr over the entire study area.

b. Monthly precipitation derived from daily precipitation with the PCA

1) BRIEF DESCRIPTION OF THE PCA

The monthly TRMM3B42 precipitation data are generated by applying the PCA to the corrected daily precipitation. PCA is widely applied in dimensional reduction and noise elimination (Lichtert and Verbeeck 2013; Urška et al. 2012). The monthly precipitation generation algorithm implemented in this study is based on the work of Mahmud et al. (2015). The accuracy of PCA monthly rainfall estimates is evaluated by comparing them with those obtained by DA. The PCA was introduced by Pearson (1901). Jensen (1996) applied PCA to compress information from multispectral satellite remote sensing images. Details of the PCA can be found in Urška et al. (2012). The steps of the PCA are as follows:

- 1) Each daily precipitation data matrix ($m \times n$) is reduced to one column.
- 2) PCA computes the eigenvalues and eigenvectors of all daily TRMM rain images in one month and creates an equal number of principal components corresponding to the corresponding input daily precipitation images. Thus, the new matrix is $\mathbf{X} [(m \times n) \times q]$, where q denotes the number of days in one month.
- 3) The singular value decomposition (SVD) algorithm sorts the eigenvectors by decreasing eigenvalues and chooses k eigenvectors. The selected number k corresponds to the threshold of cumulative variance proportion that are over 95% to form a $[(m \times n) \times k]$ dimensional matrix $\mathbf{W}$.
- 4) The total rainfall for each principal component (PC) is recalculated in this step. This is done by multiplying the original rainfall value by the corresponding eigenvector.

- 5) Sum the products from every PC to form one column of total rainfall value, and next form a one-dimensional $[(m \times n) \times 1]$ matrix that represents the monthly precipitation.
- 6) Convert the one-dimensional $[(m \times n) \times 1]$ matrix to an $m \times n$ dimensional matrix.

2) VALIDATION

The monthly rainfall from the PCA and that calculated by DA are compared with the reference precipitation for evaluation of predictive accuracy. The kriging-interpolated monthly precipitation obtained from rain gauge stations constitute the reference precipitation (Khandu et al. 2016). The comparison is based on three goodness-of-fit indexes. These are the Kullback–Leibler divergence (KLD) that has been used to estimate the accuracy of satellite precipitation product (Prakash et al. 2018), the modified law of cosines that has been used in image matching (Kaur and Aggarwal 2013), and the root-mean-square error (RMSE). The first two algorithms are aimed to mine the similarity patterns from satellite images; the classic RMSE is used to further evaluate the PCA. The rain gauge data obtained from 27 stations are applied as real precipitation data, that is, the reference data. The rain gauge data are spatially interpolated by kriging in the Hubei study area and have a resolution consistent with the precipitation calculated with PCA and by DA.

The KLD measures the similarity between two distributions (Kullback and Leibler 1951). The KLD is also known as information divergency, information gain, or relative entropy between two variables random X and Y (Wu 2018). It is calculated as follows:

$$\text{KL}(f, g) = \int_{-\infty}^{+\infty} f(x) \log \frac{f(x)}{g(x)} dx, \quad (3)$$

where X and Y are two continuous nonnegative random variables, with probability density functions f and g , respectively, and $\log(\cdot)$ denotes the natural logarithm operator. The smaller the KL value, the more similar the variables. The detail steps of KLD evaluation are as follows.

First, calculate the probability distribution of kriged rainfall from ground stations. This represents the reference precipitation distribution f . Second, the probability distribution of precipitation estimated by PCA and by DA are calculated, and are respectively called approximate precipitation distributions g_1 and g_2 . Calculate the KL between the reference precipitation distribution and each of the two approximate precipitation distributions, and designate them by KL_1 and KL_2 , which quantify the similarity between

f and g_1 , and f and g_2 , respectively. Finally, compare KL_1 and KL_2 to evaluate whether PCA-based precipitation or DA-based precipitation provides the better fit to the reference precipitation.

The modified cosine similarity (MLC) measure the matching similarity between two nonzero vectors. The cosine similarity $\cos(\theta)$ is expressed as follows:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n (A_i B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \sqrt{\sum_{i=1}^n (B_i)^2}}, \quad (4)$$

where A_i and B_i are the components of vectors $\mathbf{A}$ and $\mathbf{B}$, respectively. The range of $\cos(\theta)$ is between -1 and 1 . The value $\cos(\theta) = 1$ means exact equality, and the value $\cos(\theta) = -1$ means exactly the opposite. Cosine similarity captures directional differences, while it is not sensitive to the absolute value of a number due to the scaled nature of the cosine function. Because of these attributes, the modified cosine similarity is applied in this study to assess the accuracy of similarity matching between images. The comparison of precipitation estimates from PCA and by DA proceeds as follows.

First, subtract the average value of each of the three sets of precipitation estimates from the values of each set's estimates. The three sets of precipitation estimates are (i) kriging-interpolated monthly precipitation obtained from rain gauge stations (this is the reference data), from which its average value is subtracted according to the first step; (ii) TRMM monthly precipitation calculated by the PCA, from which its average value is subtracted according to the first step; and (iii) TRMM monthly precipitation estimated by the DA, from which its average value is subtracted according to the first step. The second step calculates the cosine similarities between precipitation (ii) and precipitation (i), and between precipitation (iii) and precipitation (i). These comparisons resolve the predictive accuracy of the calculated precipitation with cosine similarity.

The RMSE is also employed as a goodness-of-fit index of precipitation predictions. The RMSE is given by Eq. (5):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P_{ei} - P_{ri})^2}{N}}, \quad (5)$$

where P_{ei} and P_{ri} denote the PCA calculated satellite precipitation and the rain gauge based DA precipitation, respectively; $i = 1, 2, \dots, N$; N denotes the total number of pixels.

c. Geographically weighted regression model

1) BRIEF DESCRIPTION OF THE GWR MODEL

The GWR model extends traditional regression by allowing parameters to be estimated locally (Huang et al. 2010). Recall that the multiple linear regression model reflects stationary and homogeneous statistical associations whose coefficients are fixed. In contrast, the GWR model captures local effects and reflects the nonstationary and scale-dependent characteristics of the statistical associations between the dependent and explanatory variables by adding the geographically location information to the regression model. The GWR is expressed as follows:

$$Y_i = \beta_0(u_i, v_i) + \sum_{k=1}^n \beta_k(u_i, v_i) X_{ik} + \varepsilon_i, \quad (6)$$

where (u_i, v_i) denotes the coordinates of the sampling or measurement point i ; $\beta_0(u_i, v_i)$ and $\beta_k(u_i, v_i)$ denote respectively the intercept and slope estimated at the i th point. Parameters Y_i and X_{ik} represent the dependent variable and the k th independent variable of the local regression model, respectively. Parameter ε_i denotes the regression residual at the i th point. The essential idea underlying the GWR model is that data points closer to the sample point have higher impact on the local coefficients than those that are farther away. Thus, a distance decay function is introduced to determine the weights of the data points. The bisquare weighting function to specify the weights is expressed as follows:

$$W_{ik} = \left[1 - \left(\frac{d_{ik}}{b} \right)^2 \right]^2, \quad \text{when } d_{ik} \leq b, \quad (7)$$

$$W_{ik} = 0, \quad \text{when } d_{ik} > b, \quad (8)$$

where W_{ik} denotes the geographical weight, d_{ik} denotes the distance between k th point and the neighboring observation i , and b denotes the kernel bandwidth. The optimal bandwidth is selected according to the Akaike information criterion (AIC) (see, e.g., Fotheringham et al. 2002).

The GWR model involves coefficients σ (which represents the standard deviation of the residuals), the goodness-of-fit criteria (the AIC), the adjusted R^2 and the local R^2 . The adjusted R^2 is an index to measure of the overall goodness-of-fit of the model. A large adjusted R^2 indicates a better model fit. In addition, small σ and AIC imply a better model fit than otherwise.

Ordinary least squares (OLS) is applied to select the explanatory factors before entering them to the GWR model in case of multicollinearity.

2) PRECIPITATION DOWNSCALING ALGORITHM BASED ON THE GWR MODEL

The GWR-based downscaling algorithm proposed in this study is based on the work of Xu et al. (2015). The new, hybrid, downscaling algorithm rests on two assumptions: 1) the precipitation is statistically associated with the environment factors and 2) the GWR model defined at a spatial resolution can be used to predict the precipitation at finer resolution with the high-resolution environmental factors. The TRMM monthly precipitation data are input to the GWR model. The detailed steps of the GWR downscaling algorithm are as follows:

- 1) Remove pixels with 1-km resolution with $\text{NDVI}_{1\text{km}} < 0$, because they include rivers, lakes, and snow-covered pixels that are not statistically associated with precipitation.
- 2) Yin et al. (2008) proposed that the association between precipitation and NDVI may be affected by soil moisture, crop growing, and irrigation, leading to higher or lower values of NDVI compared to the surrounding pixels. Thus, an averaging filter is used to identify the outliers introduced by human and natural factors. This algorithm eliminates pixel values that are unrepresentative of their surroundings. The details of the NDVI smoothing process are as follows:
 - (i) The $\text{NDVI}_{1\text{km}}$ eliminated in step 1 is smoothed by the averaging filter with a $5\text{ km} \times 5\text{ km}$ moving window. The selection of the moving window must consider the NDVI resolution and the influences of the disturbance factors. The new pixels of $\text{NDVI}_{1\text{km}}$ are substituted by the average values of the surrounding pixels in the moving window.
 - (ii) Subtract the smoothed NDVI from the original NDVI, that is,

$$\text{A_NDVI}(j) = \text{NDVI}_{1\text{km}}(j) - \frac{\sum_{i=1}^{25} \text{NDVI}_{1\text{km}}(n)}{25}, \quad (9)$$

where $\text{NDVI}_{1\text{km}}(j)$ denotes the original value of the $\text{NDVI}_{1\text{km}}$ in pixel j ; $\text{NDVI}_{1\text{km}}(n)$ denotes the value of the $\text{NDVI}_{1\text{km}}$ around pixel j in the $5\text{ km} \times 5\text{ km}$ moving window. Set threshold of the $\text{A_NDVI}(j)$ to identify the outliers of $\text{NDVI}_{1\text{km}}$ to account for the fact that if the vegetation in a given location is significantly different from the surrounding pixels, it leads to a much higher or lower $\text{A_NDVI}(j)$ value. It must be pointed out that the threshold determination of

the $\text{A_NDVI}(j)$ must consider the topographic factors. The revised NDVI (NDVI_R) is produced by eliminating the outliers from the original $\text{NDVI}_{1\text{km}}$.

- 3) The environmental factors ($\text{LTD}_{1\text{km}}$) and topographic factors ($\text{DEM}_{1\text{km}}$, $\text{slope}_{1\text{km}}$) including the revised $\text{NDVI}_{1\text{km}}$ are resampled to 25-km resolution (termed as $\text{LTD}_{25\text{km}}$, $\text{DEM}_{25\text{km}}$, $\text{slope}_{25\text{km}}$, $\text{NDVI}_{25\text{km}}$), and collocated with the 0.25° resolution of TRMM monthly precipitation data. This work applies the nearest neighbor method to resample the data with the resampling tool of the ArcGIS 10.6 software. This resampling method assigns the digital designation of the closest input pixel (in terms of coordinates) to the corresponding output pixel. The detailed description of the resampling method of ArcGIS 10.6 software can be found at <https://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/performing-analysis/cell-size-and-resampling-in-analysis.htm>. The value of $\text{LTD}_{25\text{km}}$, $\text{DEM}_{25\text{km}}$, $\text{slope}_{25\text{km}}$, $\text{NDVI}_{25\text{km}}$ at the same location are stored in the attribute information of the fishnets to generate the training sample of GWR model in ArcGIS 10.6 or other suitable software.
- 4) The multicollinearity of environmental factors is examined with the OLS method. Further investigate whether a single factor or the combination of different factors show good performance in downscaling the TRMM monthly precipitation datasets. This work employs ArcGIS 10.6 to implement the GWR model. The kernel and bandwidth parameters are important in defining the GWR model. The selection criterion for these two parameters can be found in <http://desktop.arcgis.com/en/arcmap/10.6/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm>. The kernel types are fixed Gaussian and adaptive Gaussian. There are three bandwidth selection methods associated with the GWR model: the corrected Akaike information criterion (AICc), cross validation (CV), and bandwidth parameter. The detailed explanations of the three bandwidths can be found in Fotheringham et al. (2002, p. 59). This work implements the adaptive Gaussian as the kernel function to calculate the weighting matrix, and the AICc is selected to determine the kernel bandwidth.
- 5) The prediction of finer-scale TRMM 1-km monthly data is based on the established GWR model with 25-km scale. ArcGIS provides a tool to predict the dependent variable over the unsampled locations with the established GWR and the 1 km resolution explanatory variables, that is, the environmental factors that are selected to build the GWR model.

- 6) The residuals of the GWR model are calculated by subtracting the simulated TRMM values from the original TRMM monthly data, and such residuals are interpolated to 1 km and added back to the TRMM 1-km data calculated in step 5. The spline interpolation method is operated in ArcGIS 10.6 software or other suitable software, and the tensor is selected as the spline type.

d. Artificial neural networks

Artificial neural networks (ANN) have been widely used in data mining and data downscaling. The multi-layer feed forward neural network is applied in this study. The BP algorithm is one of the most effective methods of ANN. BP_ANN has strong capabilities for nonlinear modeling and exploring the associations between impact factors (Zhang and Wang 2008). It is very efficient in forecasting. The genetic algorithm (GAs) is a metaheuristic technique inspired by natural evolution. The GA was introduced by Holland (1975). It has been widely used to optimize neural networks (Mohsen et al. 2007). BP applies the Levenberg–Marquardt optimization algorithm (LM). The GA improves the performance of the LM (Zheng et al. 2019) by finding a suboptimal solution from a global search space. Subsequently, the solution obtained with the GA serves as an initial search point to launch the LM to achieve a near-global solution that avoids local, suboptimal, solutions. The hybrid algorithm is herein named the GA+BP_ANN downscaling algorithm.

This work implements the GA to optimize the weights between nodes in BP_ANN and to improve its prediction accuracy. The computational steps of the GA are as follows:

- 1) The initial population of H chromosomes (i.e., potential solutions) is randomly generated and becomes the current population.
- 2) The fitness of each chromosome in the current population is calculated.
- 3) Create the empty successor population by repeating the following steps until H successor chromosomes have been created in the successor population:
 - (i) Select two chromosomes $c1$ and $c2$ from the current population based on proportional fitness selection.
 - (ii) Obtain a child chromosome c by one-point crossover of $c1$ and $c2$.
 - (iii) Uniform mutation is applied to the child chromosome c to produce C , and the produced C is added to the successor population.
- 4) The successor population replaces the current population.
- 5) Return to step 2 until the stop criteria has been met.

A description of BP_ANN can be found in the work of Fernandes et al. (1998).

TABLE 1. The statistical indexes used in the evaluation of downscaling precipitation products. Notation: n refers to the number of samples; P_i means the precipitation estimates by a model at the location of rain gauge i ; G_i means observed precipitation from rain gauge station i ; $\bar{P}$ denotes the mean value of the estimated precipitation at all the locations with rain gauges; and $\bar{G}$ denotes the mean value of all rain gauge observations.

Statistical index	Equation	Optimal value
Correlation coefficient (CC)	$CC = \frac{\sum_{i=1}^n (P_i - \bar{P})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^n (G_i - \bar{G})^2}}$	1
Mean error (ME)	$ME = \frac{\sum_{i=1}^n (P_i - G_i)}{n}$	0
Mean absolute error (MAE)	$MAE = \frac{\sum_{i=1}^n P_i - G_i }{n}$	0
Root-mean-square error (RMSE)	See Eq. (5)	0
Relative bias (RB)	$RB = \frac{\sum_{i=1}^n P_i}{\sum_{i=1}^n G_i} - 1$	0

e. Validation of the downscaling products

Five commonly used statistical indexes are computed for evaluation purposes (Tan et al. 2015; Yong et al. 2010; Duan et al. 2016). They are the correlation coefficient (CC), mean error (ME), mean absolute error (MAE), RMSE, and relative bias (RB) (Table 1). The CC reflects the degree of linear correlation between downscaling satellite precipitation products and gauge observations. The other four indexes describe the error and bias. Ground observations from independent gauge stations are applied for ground truthing of the downscaled products estimated by the GWR and the GA+BP_ANN.

3. Study area and data

a. Study area

Hubei Province was selected as the study area. Hubei Province is located in the middle part of China (108°21'42"–116°07'50"E, 29°01'53"–33°06'47"N) near the Yangtze River, leading to abundant precipitation over this area. The annual precipitation in the study area ranges between 800 and 1609 mm, and most precipitation occurs between April and September (Wang and Li 2016). The elevation of this area ranges from −44 to 3002 m with three diverse topographic zones, including mountainous (about 56%), plain (about 20%), and hill zones

(about 24%) (Yu 2004). The topography of the study area has pronounced spatial heterogeneity. In the northern, eastern, and western regions the terrain is high, while in the middle of Hubei Province, the terrain features minimal relief. Due to the subtropical monsoonal climate, its annual temperature ranges from 13° to 18°C with the warm temperature and rain occurring in the same period. The wet season of Hubei Province starts in May and lasts through September. The other months (October–April) comprise the dry season.

b. Data

1) RAIN GAUGE DATA

The ground-based observations of daily precipitation (2005–10) used in this study were provided by the National Meteorological Information Center, China. The 27 gauge stations feature unique morphological characteristics that cover all landform types in Hubei Province. The daily precipitation data of rain gauge observations were statistically accumulated to monthly precipitation. The spatial distribution of gauge stations is displayed in Fig. 2a. The slope map of Hubei Province including main river networks is shown in Fig. 2b. The DEMs of Hubei Province and neighboring provinces is shown in Fig. 2c.

2) ENVIRONMENT FACTORS

The environment factors considered in this work are the NDVI, LTD, DEM, and slope data. The NDVI data reflect the vegetative cover. The MOD13A3 monthly NDVI product with 1-km spatial resolution from 2005 to 2010 used in this study was obtained from International Scientific and Technical Data Mirror Site, Computer Network Information Center of the Chinese Academy of Sciences (<http://www.gscloud.cn>). The MODLT1M monthly LTD dataset with 1-km spatial resolution between 2005 and 2010 was obtained from International Scientific and Technical Data Mirror Site, Computer Network Information Center, the Chinese Academy of Sciences (<http://www.gscloud.cn>). The DEM data employed in this study were obtained from <http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp> with 90-m resolution. The DEM was resampled to 1 km to be consistent with the spatial scale of this study. The resampling method is described in section 2c(2)(iii). The slope data used in this study were derived from the resampled DEM data with the slope tool of the ArcGIS10.6 software (see Fig. 2b).

3) THE TRMM SATELLITE PRECIPITATION DATASET

The Tropical Rainfall Mission is a joint mission of NASA and the Japan Aerospace Exploration Agency. It was launched in 1997 to study rainfall for weather and

FIG. 2. (a) The gauge stations in Hubei Province, which are labeled 1–27; (b) the slope map of Hubei Province with main river networks; and (c) the DEMs of Hubei Province and neighboring provinces.

climate research with three rainfall measuring sensors. The three sensors—Precipitation Radar (PR), TRMM Microwave Imager (TMI), and Visible and Infrared Scanner (VIRS)—are loaded on the mission's satellites (Xu et al. 2015). TRMM3B42 is the version 7 TRMM of the real-time multisatellite 3-hourly precipitation products

FIG. 3. Absolute daily bias of precipitation from May through September in 2006 at nine gauge stations.

with spatial resolution of 0.25° (25 km at the equator). The precipitation products were released in May 2012 (Huffman et al. 2007). Precipitation files are produced every 3 h on synoptic observation hours (0000, 0300, . . . , 2100 UTC). The TRMM3B42 products used in the study were obtained from <https://disc.gsfc.nasa.gov/datasets?keywords=rainfall%20TRMM%203B42&page=1>.

4. Results

Figure 3 presents the daily absolute bias of precipitation from May through September 2006 (wet season of Hubei) that is calculated by Eq. (10):

$$D_{\text{bias}} = P_{\text{TRMM}} - P_{\text{observed}}, \quad (10)$$

where P_{TRMM} and P_{observed} denote the daily precipitation derived from the original TRMM product and the precipitation obtained from rain gauge stations, respectively. Negative values mean underestimation, while positive values mean overestimation in Eq. (10).

Figure 3 displays the absolute daily bias of precipitation at nine gauge stations. These stations have different elevations and cover the study area. Compared with the red line in Fig. 3 whose precipitation error is 0 the peak values of the absolute error appear in the first 30 days (May), and within the 80th through 90th days (July). The low values below the red line are focused in the first 30 days, and within the 70th through 90th days (July). Figure 3 demonstrates that the overestimation of low values and underestimation of high values is notable, especially in May and July. Thus, the bias correction is essential before downscaling the TRMM precipitation product.

a. Effectiveness of PDF algorithm in bias correction of daily precipitation

1) BIAS CORRECTION BY THE PDF ALGORITHM

The size of the spatial moving window (shown in step 1 in section 2a) is 3×3 (with a dimensional unit equal to 0.25°) and the center of the window is the target grid,

namely, the location of the rain gauge station. The spatial kernel (termed as the weighting matrix) is determined according to the Euclidean distance and Tobler's first law. The Euclidean distance is calculated as follows:

$$\beta = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}, \quad (11)$$

where (x_1, y_1) and (x_2, y_2) denote the target grid and surrounding grid coordinates, respectively; β denotes the Euclidean distance between points (x_1, y_1) and (x_2, y_2) . The kernel is set according to β , therefore, the nearer to the target grid, the larger weight a value is given, and vice versa. The spatial kernel equals $[1/16, 1/8, 1/16; 1/8, 1/4, 1/8; 1/16, 1/8, 1/16]$.

The next step is determining the temporal moving window. Selecting each 20 days as the size of the temporal window and setting the 20th day as the baseline; $a = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]$. Parameter $i = 1, 2, \dots, 20$; i is the distance in days from the baseline. The weight b is given by $b = a(i)/\text{sum}(a)$. The data of each target grid are estimated from the outputs of the spatial and temporal windowing. Collecting sample data of each target grid. Finally, 27 sample datasets are collected, and they are arranged according to label of the gauge stations (labels 1, 2, $\dots$, 27) to form the datasets of the study area.

Figure 4 displays the probability density function of daily precipitation in 2006. Only data with nonzero TRMM estimates and gauge observations were included. After estimating by the spatial and temporal moving windows, 486 grids of daily precipitation data at rain gauge stations and satellite derived data were collected, respectively. The cumulative probability distribution was obtained. The principle of the error correction rests on adjusting the average satellite derived precipitation based on the mean deviation between the rain gauge precipitation and the satellite derived data. It is seen in Fig. 4 the cumulative probability density value is consistent with the value of satellite derived precipitation (Fig. 4a) and rainfall observed at gauge stations (Fig. 4b). An example of the application of Eq. (1) shows that when the cumulative probability density is 25% the value of the satellite derived data equals 3 mm and that of the rain gauge data equals 3.5 mm, in which case the error of TRMM daily precipitation data equals -0.5 mm.

The average error over the entire research area equals 0.551 mm. The correction formula is therefore

$$R_C = R_s + 0.551, \quad (12)$$

where R_C and R_s denote the corrected TRMM precipitation data and the original TRMM precipitation data, respectively.

a. PDF of TRMM precipitation (Year 2006)

b. PDF of gauge stations precipitation (Year 2006)

FIG. 4. Probability density function of daily precipitation.

2) THE VALIDATION OF TRMM PRECIPITATION CORRECTED BY THE PDF ALGORITHM

The spatial distributions of monthly precipitation obtained by accumulation of the TRMM3B42 daily precipitation adjusted by the PDF algorithm in June and December within Hubei Province are displayed in Fig. 5. Figure 6 depicts the monthly bias between gauge stations and TRMM monthly precipitation. The monthly bias corrected by the PDF algorithm are minimized compared with the original TRMM (Figs. 6a,b) in June and December. In June the maximum positive bias decreased from 98.82 to 90.28, while the negative bias declined from 60.75 to 58.46 after PDF adjusting. The bias for December follows the same trend as that of June, which demonstrates the PDF algorithm corrected the overestimation of low values and the underestimation of high values.

b. Comparison of PCA-based and DA-based precipitation benchmark against the in situ monthly precipitation

The TRMM daily data were corrected by the PDF algorithm, and the monthly precipitation was estimated based on the daily rainfall calculated with PCA and with DA in June (Fig. 7a) and December (Fig. 7b), respectively. The in situ monthly precipitation at gauge stations is shown in Fig. 8.

FIG. 5. The spatial distributions of (a) monthly precipitation before PDF matching in June 2006, (b) the differences between before and after PDF matching in June 2006, (c) monthly precipitation before PDF matching in December 2006, and (d) the differences between before and after PDF matching in December 2006.

Table 2 lists the estimation accuracy of DA-gauge station and PCA-gauge station areal interpolations in several months. The KLD and the modified cosine similarity (MLC) indexes nearly follow a consistent trend except for August and September. Table 2 indicates that in the wet season (except for August and September), the monthly precipitation estimated with the PCA is more similar to the precipitation measured at gauge stations than the precipitation estimated with the DA algorithm, having a small KLD score and a score near 1 calculated by the MLC. The monthly precipitation estimated with the DA algorithm is more similar to the precipitation measured at gauge stations in the dry season, having a small KLD score and a large MLC score. With regard to August and September, the results estimated by KLD and RMSE (the matching similarity of PCA is high) are opposite to that by the MLC (the matching similarity of DA is high). The KLD index and RMSE show consistent estimation results compared with the MLC.

Table 3 lists the RMSE of precipitation estimated by the PCA, which is lower than that obtained with DA

from May through September 2006. In other months the RMSE of PCA is higher than the DA's, which demonstrates that compared with DA, the PCA transformation increases the estimation accuracy of monthly rainfall during the wet season (from May through September). However, in the dry season the accuracy of monthly precipitation estimated by DA is higher than that of the PCA.

c. The GWR downscaling algorithm

Figure 9 presents the spatial distribution of A_NDVI_{1km} in June and December 2006 to exemplify identifying the effects caused by human activity and the surrounding environment on the NDVI.

The algorithm [described in section 2c(2)(ii)] demonstrates that a much higher or lower A_NDVI_{1km} would be obtained if vegetation at a given location changes significantly compared to the surrounding areas. Comparing A_NDVI_{1km} (see Fig. 9) with slope (including river networks) of Hubei Province (see Fig. 2b) shows most high and low values are found along rivers. The $A_NDVI_{1km} > 0.15$ and $A_NDVI_{1km} < -0.15$ were

FIG. 6. Differences between monthly precipitation of gauge stations and TRMM product before and after PDF correction of 2006 over Hubei Province (only values of grids with stations are shown).

identified as outliers based on a subjective adjustment. In this manner the snow or water body outliers were eliminated from the NDVI. The thresholds of A_NDVI_{1km} applied in this study might not be suitable for other regions. The TRMM monthly precipitation is estimated based on topographic and environmental factors: LTD,

NDVI, altitude, and slope, which influence the monthly precipitation. The corrected NDVI, LTD and DEM were resampled to 25 km. The effective predictors are validated respectively, including LTD and altitude (termed as GWR_{LA}), LTD and NDVI (termed as GWR_{LN}), LTD and slope (termed as GWR_{LS}), NDVI

FIG. 7. (a) The monthly precipitation of Hubei Province estimated with the PCA (shown at left) and with DA (shown at right) in June 2006. (b) As in (a), but for December 2006.

FIG. 8. The in situ monthly precipitation of Hubei Province in June and December 2006.

and altitude (termed as GWR_{NA}), NDVI and slope (termed as GWR_{NS}), LTD, altitude and slope (termed as GWR_{LAS}), LTD, NDVI and altitude (termed as GWR_{LNA}), LTD, NDVI and slope (called GWR_{LNS}), NDVI, altitude and slope (termed as GWR_{NAS}), and the combination of four factors (GWR_{LNAS}). The coefficient of determination (R^2) was applied to evaluate the effective predictors of the GWR_{LNAS} model. The evaluation results of topographic (altitude, slope) and environmental factors (NDVI, LTD) are listed in Table 4 (choosing June and December in 2006 as an example).

The results in Table 4 indicate that in June and December the prediction accuracy of GWR_{NS} is the

highest among all the models with $R^2 = 0.63$. NDVI and slope are selected as the environmental and topographic factors for downscaling purposes. The GWR prediction model was trained with 25 km resolution satellite images and it was trained to estimate the monthly precipitation with resolution of 1 km. The residuals were interpolated by spline tensor interpolator. The interpolated residuals (Figs. 10a,b) are added back to the prediction results. The final downscaling monthly precipitation are displayed in Figs. 10c and 10d.

Figures 11a and 11b shows the accuracy of the original TRMM monthly precipitation after PDF and PCA or DA corrections with a spatial resolution of 0.25° , and Figs. 11c and 11d displays the accuracy of the downscaled results

TABLE 2. 2006 monthly precipitation estimated by PCA and DA and validated with the KLD and MLC indexes.

Month date	KLD		MLC		Month date	KLD		MLC	
	PCA	DA	PCA	DA		PCA	DA	PCA	DA
1	6.556	6.362	0.442	0.515	7	0.027	0.039	0.665	-0.246
2	0.114	0.097	0.631	0.694	8	0.057	0.065	0.675	0.697
3	0.070	0.048	0.725	0.856	9	0.074	0.077	0.536	0.702
4	0.055	0.048	-0.235	0.062	10	0.070	0.062	0.561	0.825
5	0.052	0.062	0.525	0.411	11	0.138	0.098	0.131	0.530
6	0.096	0.118	0.506	-0.192	12	3.074	3.073	0.454	0.465

TABLE 3. RMSE of monthly precipitation estimated by PCA and DA.

Month date	RMSE (mm)		Month date	RMSE (mm)	
	PCA	DA		PCA	DA
1	25.60	24.90	7	44.49	128.36
2	31.09	24.12	8	34.50	49.06
3	17.14	10.04	9	41.23	56.60
4	51.85	40.98	10	37.59	20.81
5	49.35	66.95	11	34.81	20.24
6	55.13	121.63	12	14.76	11.43

after downscaling by GWR_{NS} with a spatial resolution of 1 km. It is shown in Figs. 11c and 11d that the corrected results of GWR_{NS} exhibits better accuracy than those before corrections. It is obvious that the GWR_{NS} algorithm improves the accuracy of precipitation. The accuracy of estimations is improved in June (from 0.45 to 0.59) and December (from 0.12 to 0.27) after downscaling by GWR_{NS} algorithm, but the data correlation between precipitation observed from rain gauge stations and derived from downscaled TRMM in December remains low. The downscaling results of satellite prediction are more accurate in the wet season because of the accuracy of the original TRMM product.

d. The GA+BP_ANN downscaling algorithm

This type of downscaling involves model training and testing. Two networks are trained based on different datasets, respectively. The first training network applied the monthly precipitation data at gauge stations measured in 2005–10 (see the output of matrix **B**). The gauge stations have been interpolated by kriging to cover the study area with a resolution of 0.25° . The second training network relied on the

monthly precipitation data derived from TRMM monthly satellite images that have been adjusted by PDF and PCA from 2005 to 2010 as the datasets of precipitation (see the output of matrix **D**). Data from the first 66 months constitute the training set and the 6 months remaining data are used as the testing set in both networks. All the satellite images used in this study had resolution of 0.25° . The input–output data matrices are designed respectively as follows:

$$\mathbf{B} = [\text{input}(\text{lat, lon, NDVI, altitude, slope, LTD}), \text{output}(\text{gauge stations})],$$

$$\mathbf{D} = [\text{input}(\text{lat, lon, NDVI, altitude, slope, LTD}), \text{output}(\text{TRMM}_{25\text{km}})],$$

where the columns of matrix **B** represent latitude, longitude, NDVI, altitude, slope, LTD, and precipitation of 27 rain gauge stations. The rows are the corresponding factors arranged according to months and years. The only difference between the two matrices is in the last column. The input–output dataset of **B** and **D** are used to train the network with a resolution of 0.25° , respectively. The well-trained networks of GA+BP_ANN_B and GA+BP_ANN_D are applied to forecast precipitation at 1 km, respectively.

The optimal weights and bias were determined with the genetic algorithm. The population size was set to 50. The MAXGEN was set to 100. Once the population size is determined the initial population is randomly generated. Potential solutions are ranked by their fitness, and the GA algorithm improves the current population of solutions by mutation and cross-over iteratively until meeting a stopping criterion. The BP_ANN network was trained and tested based on the observed data of rain gauge stations with 6 layers

FIG. 9. The spatial pattern of the $A_NDVI_{1\text{km}}$ over Hubei Province in June and December 2006.

FIG. 10. The spatial pattern of the interpolated residuals generated by GWR_{NS} in (a) June and (b) December 2006, and the downscaled results by GWR_{NS} in (c) June and (d) December 2006.

including one input layer, four hidden layers and one output layer. The numbers of neurons in the hidden layer are 7, 7, 5, and 3, respectively. Adjusting the learning rate from 0.001 to 0.1 and then the learning rate and epochs were determined as 0.01 and 500, separately. Typically, 90% of the data was used for training purposes. The other 10% of data was applied for validation. The average coefficient of determination versus GA-BP_ANN predictions equals 0.80 with $RMSE = 24.37$ mm. The trained network was applied to downscale the TRMM of June 2006 from 25- to 1-km resolution. The result is displayed in Fig. 12a.

The structure of the training data of another network is displayed as matrix **D**. The number of records equals 37 010 after eliminating the zeroes and null values. The structure of the network is the same as that used in the first training network. The numbers of neurons in the hidden layer are 9, 9, 7, and 6. The average coefficient of determination versus GA-BP_ANN predictions to verification data points is 0.79 with $RMSE = 27.82$ mm. The downscaled result is shown in Fig. 12b.

Compare the downscaling results of Fig. 12a with those of Figs. 8c, 10c, and 12b. The spatial distribution pattern of heaviest precipitation in the southeast is consistent with observations (Fig. 8c), while in the southwest of Hubei Province there are differences. There are two possible reasons for causing the difference between the monthly precipitation estimated by the two models: first, the number of actual gauge stations might limit the accuracy

TABLE 4. The GWR model prediction results corresponding to June and December 2006.

Model	R^2 (June)	R^2 (December)
GWR_{LNAS}	0.22	0.35
GWR_{LAS}	0.16	0.38
GWR_{LNA}	0.19	0.37
GWR_{LNS}	0.34	0.38
GWR_{NAS}	0.46	0.60
GWR_{LA}	0.20	0.42
GWR_{LN}	0.40	0.41
GWR_{LS}	0.24	0.45
GWR_{NA}	0.59	0.61
GWR_{NS}	0.63	0.63

FIG. 11. The scatterplot between the rain gauge stations and (a) the original TRMM 3B42 V7 after PCA correction in June, (b) the original TRMM3B42 V7 after PCA correction in December, (c) the downscaled results of the GWR_{NS} algorithm in June, and (d) the downscaled results of the GWR_{NS} algorithm in December.

of BP_ANN. Second, the disparity of measurement methods of precipitation between ground observations and satellite based (Smith et al. 2006).

e. Comparison of downscaled results with the GA+BP_ANN and with the GWR algorithm

The evaluation results for June 2006 are listed in Table 5. The downscaling product estimated by GA+BP_ANN_B had the worst performance in terms of all five statistical indexes. The downscaling product estimated by GA+BP_ANN_D was the second worst performing product. The lowest ME ($-15.22 \text{ mm month}^{-1}$), MAE ($20.24 \text{ mm month}^{-1}$) and RMSE values ($39.07 \text{ mm month}^{-1}$) was achieved by downscaling product estimated by GWR_{NS} .

The downscaling results estimated by GWR_{NS} corresponding to June 2006 represent the most accurate products compared to the others. It is also seen in Fig. 12b that the heaviest precipitation occurs in the northern part of Hubei Province. However, in the southwest of Hubei Province the rainfall is relatively low. Our results show the spatial distribution of precipitation estimated by GA+BP_ANN_B (Fig. 12a) or by GA+BP_ANN_D (Fig. 12b) showed a significant difference, which demonstrates the accuracy of precipitation is poor with this algorithm. In addition, Fig. 13 presents the MAE and RMSE evaluated by the downscaling algorithms from January through December 2006 in Hubei Province. The GWR_{NS} generated lower MAE and RMSE of predictions

FIG. 12. (a) The downscaled results by GA+BP_ANN based on the precipitation measured at rain gauge stations in June 2006 and (b) the downscaled results by GA+BP_ANN based on the precipitation of TRMM satellite images in June 2006.

for the dry season (except for April and October) than GA+BP_ANN_D and GA+BP_ANN_B. The GWR_{NS} generated lower MAE and RMSE in May, June, July, and September (the wet season). The GWR_{NS} provided better predictions of the TRMM3B42 V7 in most months.

5. Conclusions and discussion

This paper introduced two algorithms to downscale TRMM3B42 precipitation from 25 to 1 km. The performances of the algorithms were assessed with ground observations measured by rain gauge networks within Hubei Province. Based on the study we conclude the following.

The bias correction is crucial for TRMM3B42 precipitation datasets. The performances of the downscaling algorithms are largely dependent on the accuracy of the original satellite precipitation dataset. It is effective to apply the PDF to improve the spatial and temporal accuracy of TRMM products. The application of PCA procedures introduced in this work provided an

TABLE 5. Statistical indexes evaluating the downscaling algorithms in June 2006.

Statistical index	GWR	GA+BP_ANN _D	GA+BP_ANN _B
CC	0.77	0.71	0.67
ME (mm month ⁻¹)	-15.22	26.27	35.35
MAE (mm month ⁻¹)	20.24	35.22	39.73
RMSE (mm month ⁻¹)	39.07	57.86	63.76
RB	0.18	-0.36	-0.47

alternative method of obtaining accurate monthly rainfall estimates from corrected TRMM3B42 datasets during the wet rainfall season in a region with sparse rain gauges. However, the monthly precipitation estimated by directly accumulation is more accurate in the dry season than in the wet season. This paper applied the Kullback–Leibler divergence and the modified cosine similarity combined with a conventional index (RMSE) by mining of the contents of satellite images to evaluate the accuracy of monthly precipitation estimated by PCA. The Kullback–Leibler divergence algorithm produced the most stable evaluation results consistent with the RMSE.

Environmental factors related to precipitation are altitude, slope, LTD, NDVI, latitude, and longitude, which were vetted in the process of downscaling satellite precipitation datasets. Furthermore, the GWR_{NS} provided better estimates of the TRMM values with NDVI and slope in most dry and wet seasons than GA+BP_ANN based on the monthly ground-measured precipitation or that from TRMM satellite images. The downscaling results for the dry season exhibited large bias due to the low accuracy of the TRMM3B42 dataset. The bias correction of the TRMM3B42 precipitation dataset is based on the accuracy of precipitation measured at gauge stations in this study. However, rain gauge measurements may have errors from evaporation loss particularly in the dry season (Tang et al. 2018; Ye et al. 2004; Zhang et al. 2004). The GWR_{NS} is reliable in regions with low and midlatitudes, and in wet seasons (from May through September). The highly variable precipitation in mountainous regions hinders the accuracy of downscaling with GWR_{NS}, and this deserves further testing in future studies. The GWR_{NS} may have poor predictive skill of winter precipitation (January, November, December) because of two reasons. First, snowfall, which is the main form of precipitation, causes low accuracy of the original TRMM3B42 dataset, especially in winter (Shen et al. 2010). Moreover, the statistical association between vegetation and precipitation is relatively weak in winter and strong in the growing season (Zhang et al. 2018). The new downscaling algorithm may not be applicable in areas where the precipitation of the

FIG. 13. MAE and RMSE of the simulated values using GWR with NDVI and slope as explanatory variables, the GA+BP_ANN with matrix **D** as training datasets (GA+BP_ANN_D) and the GA+BP_ANN with matrix **B** as training datasets (GA+BP_ANN_B) compared to the observations of gauge stations in Hubei Province between January and December in 2006 corresponding to the (a) dry season and (b) wet season.

TRMM satellite precipitation datasets have no significant statistical association with the NDVI and Slope (such as desert, snow covered areas, temperature-limited ecosystems, radiation-limited ecosystems, and water bodies). The performance of GA+BP_ANN_B and GA+BP_ANN_D may have been limited in this study by the accuracy of the training precipitation datasets. The results indicate the high-resolution precipitation datasets have higher accuracy than the original TRMM3B43 V7 datasets when downscaling the satellite precipitation by GWR_{NS}. Finding a suitable regression method is crucial to downscaling the monthly precipitation within regions with complex precipitation–topographic and precipitation–environment relationships.

Experimental results using data from Hubei Province, China, provide us with insights as to how the new

downscaling algorithm may be applied and expected to perform in other regions with complex topographic and environment characteristics. There are also limitations with the proposed algorithm that may lead to systemic or random errors in downscaled precipitation datasets. First, the spatial variation of precipitation in 0.25° grids is influenced primarily by monsoons. The statistical association between precipitation and slope cannot always be represented accurately at the hillslope scale in mountainous regions with downscaling approaches (Zhang et al. 2018). Yong et al. (2010) evaluated the performance of TRMM3B42 V6 datasets in the Laohahe basin, China, and reported good performance at lower elevation and poor performance at higher elevations. However, the study by Xu et al. (2015) provides a different insight in downscaling TRMM3B43 V7 with the GWR method. Xu et al. (2015) proposed that the residuals featured irregular characteristics in space over the Tibetan Plateau, with an elevation range from 1671 to 6043 m. The average daily precipitation bias of gauge stations and slopes from May through September in 2006 determined in this study are listed in Table S1 in the online supplemental material [derived from Eq. (10)]. The precipitation structure is highly variable over the mountainous regions at those scales due to environmental heterogeneity, such as topography, and the output is with a 1 km × 1 km grid. Therefore, many uncertainties may arise in the precipitation downscaling process. The second limitation is introduced by smoothing effects that stem mainly from the interpolation methods. Smoothing arises when the NDVI and DEM are resampled from 1 to 25 km, introducing loss of detail about the NDVI and DEM. The third limitation is the number of gauge stations. This paper's downscaling results of GWR_{NS} were validated with the ground observations from gauge stations, and also by the China Gauge-Based Daily Precipitation Analysis (CGDPA) product (see Fig. S1, take year 2006 as an example), which has high accuracy (Shen and Xiong 2016; Tang et al. 2020). Nevertheless, spatial coverage is sparse for precipitation downscaling in mountainous regions.

Future work by these authors will explore the influence of global environmental forcing factors (e.g., monsoons, ocean currents) on regional precipitation. The proposed downscaling algorithm will be verified to investigate whether these factors are feasible for downscaling satellite precipitation datasets in a wide range of regions. Further efforts will be devoted to downscaling daily precipitation (see, e.g., Smith et al. 2006).

Acknowledgments. This study was supported by the National Key R&D program of China (2016YFC0803106). We are grateful to the editorial team and three anonymous reviewers for their efforts to improve this manuscript. We

also thank Dr. Yan Shen from China's Meteorological Administration for providing the CGDPA (2006) precipitation product for China. The authors report no potential conflicts of interest.

REFERENCES

- Chen, C., S. Zhao, Z. Duan, and Z. Qin, 2015: An improved spatial downscaling procedure for TRMM3B43 precipitation product using geographically weighted regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **8**, 4592–4604, <https://doi.org/10.1109/JSTARS.2015.2441734>.
- Duan, Z., and W. G. M. Bastiaanssen, 2013: First results from version 7 TRMM3B43 precipitation product in combination with a new downscaling-calibration procedure. *Remote Sens. Environ.*, **131**, 1–13, <https://doi.org/10.1016/j.rse.2012.12.002>.
- , J. Liu, Y. Tuo, G. Chiogna, and M. Disse, 2016: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Sci. Total Environ.*, **573**, 1536–1553, <https://doi.org/10.1016/j.scitotenv.2016.08.213>.
- Fang, J., J. Du, W. Xu, P. Shi, M. Li, and X. Ming, 2013: Spatial downscaling of TRMM precipitation data based on the orographical effect and meteorological conditions in a mountainous area. *Adv. Water Resour.*, **61**, 42–50, <https://doi.org/10.1016/j.advwatres.2013.08.011>.
- Fernandes, M. A. C., A. D. D. Neto, and J. B. Bezerra, 1998: A neural network model applied to the detection of digital signals. *SBT/IEEE Int. Telecommunications Symp.*, Sao Paulo, Brazil, IEEE, 279–283, <https://doi.org/10.1109/ITS.1998.713132>.
- Fotheringham, A. S., C. Brunsdon, and M. E. Charlton, 2002: *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, 284 pp.
- Holland, J. H., 1975: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 183 pp.
- Huang, B., B. Wu, and M. Barry, 2010: Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.*, **24**, 383–401, <https://doi.org/10.1080/13658810802672469>.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- Isaac, G. A., and R. A. Stuart, 1992: Temperature–precipitation relationships for Canadian stations. *J. Climate*, **5**, 822–830, [https://doi.org/10.1175/1520-0442\(1992\)005<0822:TRFCS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1992)005<0822:TRFCS>2.0.CO;2).
- Jensen, J. R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*. 2nd ed. Prentice Hall, 316 pp.
- Jia, S., W. Zhu, A. Lü, and T. Yan, 2011: A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sens. Environ.*, **115**, 3069–3079, <https://doi.org/10.1016/J.RSE.2011.06.009>.
- Kaur, S., and D. Aggarwal, 2013: Image content based retrieval system using cosine similarity for skin disease images. *Adv. Comput. Sci. Int. J.*, **2**, 89–95.
- Khandu, J. L. Awange, and E. Forootan, 2016: An evaluation of high-resolution gridded precipitation products over Bhutan (1998–2012). *Int. J. Climatol.*, **36**, 1067–1087, <https://doi.org/10.1002/joc.4402>.
- Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86, <https://doi.org/10.1214/aoms/1177729694>.
- Li, M., and Q. X. Shao, 2010: An improved statistical approach to merge satellite rainfall estimates and rain gauge data. *J. Hydrol.*, **385**, 51–64, <https://doi.org/10.1016/j.jhydrol.2010.01.023>.
- Lichtert, S., and J. Verbeeck, 2013: Statistical consequences of applying a PCA noise filter on EELS spectrum images. *Ultramicroscopy*, **125**, 35–42, <https://doi.org/10.1016/j.ultramicro.2012.10.001>.
- Madden, R. A., and J. Williams, 1978: The correlation between temperature and precipitation in the United States and Europe. *Mon. Wea. Rev.*, **106**, 142–147, [https://doi.org/10.1175/1520-0493\(1978\)106<0142:TCBTAP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1978)106<0142:TCBTAP>2.0.CO;2).
- Mahesh, C., S. Prakash, V. Sathiyamoorthy, and R.M. Gairola, 2011: Artificial neural network based microwave precipitation estimation using scattering index and polarization corrected temperature. *Atmos. Res.*, **102**, 358–364, <https://doi.org/10.1016/j.atmosres.2011.09.003>.
- Mahmud, M. R., H. Matsuyama, T. Hosaka, S. Numata, and M. Hashim, 2015: Temporal downscaling of TRMM rain-rate images using principal component analysis during heavy tropical thunderstorm seasons. *J. Hydrometeorol.*, **16**, 2264–2275, <https://doi.org/10.1175/JHM-D-14-0233.1>.
- Mohsen, S., A. Morteza, and A. Y. Varjani, 2007: Design of neural networks using genetic algorithm for the permeability estimation of the reservoir. *J. Petrol. Sci. Eng.*, **59**, 97–105, <https://doi.org/10.1016/j.petrol.2007.03.007>.
- Nasseri, M. K., K. Asghari, and M. J. Abedini, 2008: Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Syst. Appl.*, **35**, 1415–1421, <https://doi.org/10.1016/j.eswa.2007.08.033>.
- Pearson, K., 1901: On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, **2**, 559–572, <https://doi.org/10.1080/14786440109462720>.
- Prakash, S., A. K. Mitra, A. AghaKouchak, Z. Liu, H. Norouzi, and D. S. Pai, 2018: A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.*, **556**, 865–876, <https://doi.org/10.1016/j.jhydrol.2016.01.029>.
- Sadeghi, M., A. A. Asanjan, M. Faridzad, V. A. Gorooh, P. Nguyen, K. Hsu, S. Sorooshian, and D. Braithwaite, 2019: Evaluation of PERSIANN-CDR constructed using GPCP V2.2 and V2.3 and a comparison with TRMM3B42 V7 and CPC unified gauge-based analysis in global scale. *Remote Sens.*, **11**, 2755, <https://doi.org/10.3390/rs11232755>.
- Sedki, A., D. Ouazar, and E. E. I. Mazoudi, 2009: Evolving neural network using real coded genetic algorithm for daily rainfall-runoff forecasting. *Expert Syst. Appl.*, **36**, 4523–4527, <https://doi.org/10.1016/j.eswa.2008.05.024>.
- Shen, Y., and A. Xiong, 2016: Validation and comparison of a new gauge-based precipitation analysis over mainland China. *Int. J. Climatol.*, **36**, 252–265, <https://doi.org/10.1002/joc.4341>.
- , A. Xiong, Y. Wang, and P. Xie, 2010: Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res.*, **115**, D02114, <https://doi.org/10.1029/2009JD012097>.
- Smith, T. M., P. A. Arkin, J. J. Bates, and G. J. Huffman, 2006: Estimating bias of satellite-based precipitation estimates. *J. Hydrometeorol.*, **7**, 841–856, <https://doi.org/10.1175/JHM524.1>.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSS>2.3.CO;2).
- Tan, M. L., A. Ibrahim, Z. Duan, A. Cracknell, and V. Chaplot, 2015: Evaluation of six high-resolution satellite and ground-based

- precipitation products over Malaysia. *Remote Sens.*, **7**, 1504–1528, <https://doi.org/10.3390/rs70201504>.
- Tang, G., A. Behrangi, D. Long, C. Li, and Y. Hong, 2018: Accounting for spatiotemporal errors of gauges: A critical step to evaluate gridded precipitation products. *J. Hydrol.*, **559**, 294–306, <https://doi.org/10.1016/j.jhydrol.2018.02.057>.
- , M. P. Clark, S. M. Papalexiou, Z. Ma, and Y. Hong, 2020: Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sens. Environ.*, **240**, 111697, <https://doi.org/10.1016/J.RSE.2020.111697>.
- Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Lett.*, **32**, L14703, <https://doi.org/10.1029/2005GL022760>.
- Urška, D., P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone 2012: Principal component analysis on spatial data: An overview. *Ann. Assoc. Amer. Geogr.*, **103**, 107–128, <https://doi.org/10.1080/00045608.2012.689236>.
- Wang, R. H., and C. Li, 2016: Spatiotemporal analysis of precipitation trends during 1961–2010 in Hubei province, central China. *Theor. Appl. Climatol.*, **124**, 385–399, <https://doi.org/10.1007/s00704-015-1426-x>.
- Wu, C., and J. M. Chen, 2012: The use of precipitation intensity in estimating gross primary production in four northern grasslands. *J. Arid Environ.*, **82**, 11–18, <https://doi.org/10.1016/j.jaridenv.2012.02.014>.
- Wu, P., 2018: Performance monitoring of MIMO control system using Kullback-Leibler divergence. *Can. J. Chem. Eng.*, **96**, 1559–1565, <https://doi.org/10.1002/cjce.23090>.
- Xie, P. P., and A. Y. Xiong, 2011: A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res.*, **116**, D21106, <https://doi.org/10.1029/2011JD016118>.
- Xu, S. G., C. Wu, L. Wang, A. Gonsamo, Y. Shen, and Z. Niu, 2015: A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics. *Remote Sens. Environ.*, **162**, 119–140, <https://doi.org/10.1016/j.rse.2015.02.024>.
- Ye, B., D. Yang, Y. Ding, T. Han, and T. Koike, 2004: A bias-corrected precipitation climatology for China. *J. Hydrometeorol.*, **5**, 1147–1160, <https://doi.org/10.1175/JHM-366.1>.
- Yin, Z. Y., X. Zhang, X. Liu, M. Colella, and X. Chen, 2008: An assessment of the biases of satellite rainfall estimates over the Tibetan Plateau and correction methods based on topographic analysis. *J. Hydrometeorol.*, **9**, 301–326, <https://doi.org/10.1175/2007JHM903.1>.
- Yong, B., L.-L. Ren, Y. Hong, J.-H. Wang, J. J. Gourley, S.-H. Jiang, X. Chen, and W. Wang, 2010: Hydrologic evaluation of multisatellite precipitation analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. *Water Resour. Res.*, **46**, W07542, <https://doi.org/10.1029/2009WR008965>.
- Yu, H., 2004: Distribution of plant species richness along elevation gradient in Hubei Province, China. *Proc. Bridging Scales and Epistemologies: Linking Local Knowledge and Global Science in Multi-Scale Assessments*, Alexandria, Egypt, Millennium Ecosystem Assessment, 378–386, <https://www.millenniumassessment.org/documents/bridging/papers/hua.yu.pdf>.
- Zhang, Q. W., and C. Wang, 2008: Using genetic algorithm to optimize artificial neural network: A case study on earthquake prediction. *Second Int. Conf. on Genetic and Evolutionary Computing*, Hubei, China, IEEE, 128–131, <https://doi.org/10.1109/WGEC.2008.96>.
- Zhang, T., B. Li, Y. Yuan, X. Gao, Q. Sun, L. Xu, and Y. Jiang, 2018: Spatial downscaling of TRMM precipitation data considering the impacts of macro-geographical factors and local elevation in the three-river headwaters region. *Remote Sens. Environ.*, **215**, 109–127, <https://doi.org/10.1016/j.rse.2018.06.004>.
- Zhang, Y., T. Ohata, D. Yang, and G. Davaa, 2004: Bias correction of daily precipitation measurements for Mongolia. *Hydrol. Processes*, **18**, 2991–3005, <https://doi.org/10.1002/hyp.5745>.
- Zhao, W. N., and M. A. K. Khalil, 1993: The relationship between precipitation and temperature over the contiguous United States. *J. Climate*, **6**, 1232–1236, [https://doi.org/10.1175/1520-0442\(1993\)006<1232:TRBPAT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1232:TRBPAT>2.0.CO;2).
- Zheng, G., D. Ha, H. Loaiciga, H. Zhou, C. Zeng, and H. Zhang, 2019: Estimation of the hydraulic parameters of leaky aquifers based on pumping tests and coupled simulation/optimization: Verification using a layered aquifer in Tianjin, China. *Hydrogeol. J.*, **27**, 3081–3095, <https://doi.org/10.1007/s10040-019-02021-z>.