Title

Study of the improvement of the multifractal spatial downscaling by the random forest regression model considering spatial heterogeneity

Journal

Journal of Applied Remote Sensing, 17(3)

ISSN

1931-3195

Authors

Zhang, Wei

Ji, Chenjia

Zheng, Shengjie

et al.

Publication Date

2023-07-01

DOI

10.1117/1.jrs.17.034510

Peer reviewed

JARS Journal of Applied Remote Sensing

RESEARCH PAPER

Study of the improvement of the multifractal spatial downscaling by the random forest regression model considering spatial heterogeneity

Wei Zhang,[a,*] Chenjia Ji,[a] Shengjie Zheng,[b] Hugo A. Loáiciga,[c] Wenkai Li,[d] and Xiaona Sun[a]

[a]China University of Geosciences (Wuhan), School of Geography and Information Engineering, Wuhan, China
[b]China Petroleum Pipeline Engineering Co., Ltd., Langfang, China
[c]University of California, Santa Barbara, Department of Geography, Santa Barbara, California, United States
[d]China Railway Siyuan Survey and Design Group Co., Ltd., Wuhan, China

ABSTRACT. Regional hydrological analysis generally requires meteorological inputs with adequate spatial resolution and coverage. Satellite-derived precipitation covers relatively large areas at various temporal scales. The global precipitation measurement (GPM) began releasing a new generation of global precipitation products in April, 2014, i.e., the integrated multi-satellite retrievals for GPM (IMERG), which has a spatial resolution of 0.1 deg (latitude) × 0.1 deg (longitude). However, IMERG does not have the sufficient resolution required by the most advanced fine-scale hydrological models. Meanwhile, due to the randomness of daily precipitation, it is difficult to obtain stable precipitation influence factors. The results of this work show that: (1) The IMERG V06 shows obvious multifractal characteristics, which makes it possible to use the multifractal method to improve its spatial resolution without the help of other elements. However, the accuracy of precipitation products will suffer a certain loss. (2) The downscaling method considering the influence factors of precipitation did well on a monthly dataset, and the maximum CC can reach 0.911. At the same time, the random forest regression model is significantly better than the traditional multiple linear regression model since the former is better matched with the original monthly precipitation data and can produce more local details. (3) The downscaled monthly precipitation data are helpful to the spatial heterogeneity recovery of daily precipitation. The recovery can enrich the spatial details of the daily precipitation and improve the accuracy to a certain extent. Compared with the multifractal (MF) downscaling results, the accuracy of the MF-RFR model was improved by 10.3%, whereas the MF-MLR model improved by only 4.6%. Among them, the accuracy of the MF-RFR model is higher than that of the original IMERG V06 product, with an obvious increase in dry days. The MF-RFR model-derived precipitation would lead to more accurate meteorological disaster assessments and hydrologic analyses than would otherwise be.

 [DOI: 10.1117/1.JRS.17.034510]

Keywords: integrated multi-satellite retrievals for global precipitation measurement; spatial downscaling; multifractal; random forest regression; spatial heterogeneity

Paper 230127G received Mar. 22, 2023; revised Aug. 1, 2023; accepted Aug. 14, 2023; published Sep. 12, 2023.

*Address all correspondence to Wei Zhang, weizhang@cug.edu.cn

1931-3195/2023/$28.00 © 2023 SPIE

1 Introduction

Precipitation is a key climatic factor with high spatiotemporal heterogeneity affecting regional ecology, hydrology, agriculture, and the economy. The resolution of precipitation datasets plays a central role on distributed hydrological modeling.[1,2]

Regional hydrological model analysis generally requires precipitation datasets with a spatial resolution higher than 10 km × 10 km, sometimes of 1 km × 1 km.[3,4] Meteorological stations and ground-based radar provide regional surface precipitation; yet, their distribution is uneven on land.[5] The limited number of observation sensors militate against the accurate estimation of precipitation distribution over large areas.[6] Satellite-derived precipitation products have become the preferred data source for the large-scale synchronous reconstruction of precipitation. This is so because of the spatial continuity, wide-coverage, and fine space–time resolution of satellite measurements. Among the satellite-derived precipitation products, the Tropical Rainfall Measurements Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42 precipitation data have relatively high accuracy due to the TMPA algorithm.[7,8] However, the spatial resolution of TMPA 3B42 is 0.25 deg × 0.25 deg, which at the equator defines spatial cells of ~27.5 km × 27.5 km in size. The global precipitation measurement (GPM) is a new generation of precipitation observation satellites with higher accuracy and increased coverage for more refined hydrological studies.[9] IMERG, the level 3 multi-satellite precipitation algorithm of GPM, can achieve a spatial resolution of 0.1 deg × 0.1 deg, although the resolution is high compared with other precipitation products, this cell size is not fine enough for regional hydrological analyses, especially in applications involving other datasets featuring higher spatial resolution. Therefore, it is necessary to downscale IMERG precipitation products accurately.

There are currently two types of downscaling methods: regional climate models (RCMs) and statistical downscaling models (SDMs). RCMs provide initial and boundary conditions of climate simulations made with higher spatial resolution than that of atmospheric general circulation models (GCMs). RCMs preserve the features of large-scale GCMs and simulate climate scenarios at a higher resolution than GCMs. Wright et al.[10] examined the frequency of tropical cyclones and the characteristics of rainfall in the eastern United States with an 18-km resolution RCM. SDMs rely on empirical mathematical functions between low-resolution dataset and other influencing variables and apply those functions to the high-resolution dataset to make scale-based calculations.[11] Therefore, SDMs lack physical meaning and rely on observational data for fitting purposes. However, SDMs have unique advantages, such as flexible structure, the multiplicity of mathematical formulations, and relatively simple development. Following the survey in Maraun et al.,[12] SDMs are classified into perfect prognosis (PP), model output statistics, and weather generators.

The spatial downscaling of GPM precipitation products has been studied by many scholars at home and abroad. Using the Qilian Mountains as the study area, Wang et al.[13] improved the geographically weighted regression (GWR) model based on the stepwise regression analysis method to achieve GSMaP precipitation products at a 0.1 deg × 0.1 deg scale, increasing the spatial resolution to 1 km while ensuring the spatial distribution of precipitation, and obtaining high-resolution annual and seasonal precipitation data. Zhan et al.[4] constructed GWR models and multiple linear regression (MLR) models by digital elevation model (DEM) and normalized difference vegetation index (NDVI), respectively, to downscale the GPM annual and monthly precipitation data to a spatial resolution of 1 km × 1 km. Shi et al.[14] selected the water vapor factor and vegetation index as influencing factors for the GPM-IMERG monthly precipitation data to obtain high-precision (1 km) precipitation data for the Fujian, Zhejiang, and Jiangxi regions of China.

Machine learning regression models, such as artificial neural networks (ANNs), BP neural network, and support vector machines (SVMs), have produced good results in terms of downscaling.[15] Shen et al.[16] proposed a gradient-propelled decision tree (GBDT) downscaling method to downscale the 2015–2018 GPM-IMERG annual precipitation data for mainland China to 0.01 deg × 0.01 deg, and compared the analysis with RF and SVM. The latter authors concluded that GBDT and RFs produce better downscaling results for IMERG precipitation estimation in mainland China. Ma et al.[3] applied a random forest regression (RFR) model to downscale the IMERG and TRMM annual precipitation data and concluded that the RF produced better downscaling results for IMERG precipitation. Furthermore, it can obtain a more accurate

spatial distribution of precipitation compared to TRMM. However, these SDMs' performance varies across areas according to the choice of parameters. GWR features variable spatial parameters and can capture spatially local characteristics, thus reflecting the spatial heterogeneity of precipitation.[17–20] Numerous studies have shown that DEM and NDVI are explanatory factors that are strongly correlated with rainfall.[21] Xu et al.[22] utilized the GWR model to downscale the TRMM3B43 V7 precipitation data from 0.258 deg to 1 km resolution in the eastern Tibetan Plateau and Tianshan Mountains, demonstrating the feasibility of employing NDVI and DEM for monthly downscaling of TRMM products. Fan et al.[23] conducted an analysis on the impact of seven topographic factors on precipitation and concluded that the topographic relief factor is the most suitable approach for precipitation downscaling in the Tianshan Mountains. Overall, these environmental factors exhibit satisfactory performance in assessing both annual and monthly precipitation patterns. However, it remains challenging to explain the significant dynamic variations in precipitation events due to the temporal invariance of topographic factors.[24] Moreover, daily precipitation exhibits a high level of randomness, and conventional environmental variables are inadequate in simulating its daily fluctuations. Consequently, SDMs that rely on the relationship between these environmental variables and daily precipitation often lack universality and accuracy in predicting precipitation.[25]

Several authors employed fractal theory and discovered multifractal (MF) characteristics in the structure of rain and clouds.[26–29] Lovejoy and Mandelbrot[30] built a space–time rainfall model based on the hyperbolic distribution characteristics of rainfall in time and space. The MF downscaling model is based on a random cascade process that employs the MF spectrum and a statistical distribution model to downscale precipitation from large scale to small scale passing through a sequence of scale levels.[31] MF downscaling, unlike traditional methods, does not rely on environmental factors and has the ability to capture complex rainfall processes using a limited number of parameters. On the other hand, MF downscaling has a homogeneous random structure, which cannot capture the spatial heterogeneity of precipitation accurately. However, Pathirana and Herath[32] downscaled radar-based precipitation with the automated meteorological data acquisition system and proposed a simple method to filter the spatial heterogeneity of precipitation. Posadas et al.[33] successfully downscaled daily TRMM 3B42 data to a spatial resolution of 0.875 km $\times$ 0.875 km. Xu et al.[34] applied ANN regression and the MF model to downscale the monthly TRMM 3B42 to 0.01 deg. The coefficient of determination, bias, and root mean square error (RMSE) of the regressions established precipitation predictions consistent with observational data. The studies cited above fitted precipitation with a lognormal distribution at various time scales. However, the restoration of heterogeneity remains to be studied in depth. Xu et al.[22] showed that many SDMs can be applied, yet, the challenge is to downscale with a suitable time scale while capturing the heterogeneity of rainfall.

This paper proposes a novel approach to downscale daily IMERG precipitation data. The approach is demonstrated with data from Hubei province, China. The model combines the RFR model with the MF model to overcome the limitations of the homogeneous MF model and introduces high-precision monthly precipitation data after RFR downscaling to recover the spatial heterogeneity of the precipitation field.

2 Study Area and Datasets

2.1 Study Area

Hubei province is located between longitudes 108° and 116°E, and between latitudes 29° and 33°N.[35] Figure 1 shows the terrain elevation, the locations of 75 meteorological stations, and the distribution of the average annual precipitation. It is evident from Fig. 1 that the topography of the study area is complex and variable, with mountains in the northwest, hilly in the southeast and northeast, and plains predominant in its central region. Hubei province has a humid subtropical climate, with four distinct seasons. Equation (1) gives the calculation of average annual precipitation. The average annual precipitation varies in the range of 770 to 1700 mm. Surface water resources are plentiful in the central plains of the study area (within the Jianghan Plain) giving Hubei the name of the "Province of Lakes." Land surface temperature is highest in the southeastern region and lowest in the southwestern and northwestern regions:

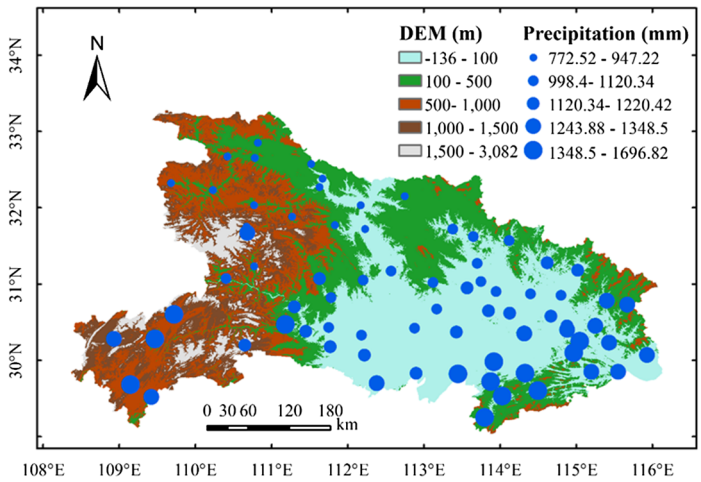

Fig. 1 Distribution of topography, meteorological stations, and precipitation in Hubei province, China.

$$A_n = \frac{\sum_{y=1}^{m} P_n}{m}, \tag{1}$$

where n denotes a site, with a total of 75 sites considered in this article. The variable m represents the year, and P_n represents the annual precipitation of the n'th station.

2.2 Precipitation Products

This paper focuses on the downscaling of IMERG V06 precipitation products and uses precipitation data from 75 meteorological stations for the years 2018 to 2020 as validation data. All the validation data pertain to Hubei province and are aggregated to daily and monthly values.

2.2.1 *IMERG precipitation products*

The GPM, the successor to TRMM, uses satellites to measure Earth's rain and snowfall for the benefit of mankind. Launched by NASA and JAXA on February 27, 2014, GPM is an international mission that sets the standard for spaceborne precipitation measurements. The increased orbital inclination of the GPM allows for wider coverage of precipitation data than the TRMM. The GPM has been adapted for the high-frequency channel, making it more sensitive to observations of trace and solid precipitation and allowing for a more refined rainfall distribution.[36] The GPM-IMERG precipitation data are currently widely used in precipitation studies.[24,37] The IMERG includes three types of precipitation products: "early run," "late run," and "final run."[8,38] Among these products, the first two are near-real time products, which are delayed by 4 and 14 h, respectively, whereas the last one is the final product, which has a delay of 3.5 months. Currently, only the final precipitation products have undergone the final correction process, making them the most accurate. It is widely recognized that the IMERG-final precipitation product is the leading satellite grid-based precipitation product in the current stage.[39] We selected the daily products of IMERG V06 (GPM IMERG final precipitation L3 1 day 0.1 deg × 0.1 deg V06) for the months of July and December from 2018 to 2020 as the target data for spatial downscaling. In addition, we chose the monthly products of IMERG V06 (GPM IMERG final precipitation L3 1 month 0.1 deg × 0.1 deg V06) for the months of July and December from 2016 to 2020 to capture the heterogeneity of daily precipitation. These precipitation products can be accessed on the website of the Goddard Earth Sciences Data and Information Services Center.[40]

2.2.2 *Meteorological stations data*

This study was based on the three-hourly ground-truthed site precipitation data from 75 meteorological stations in Hubei Province provided by China Meteorological Data Service Centre,[41]

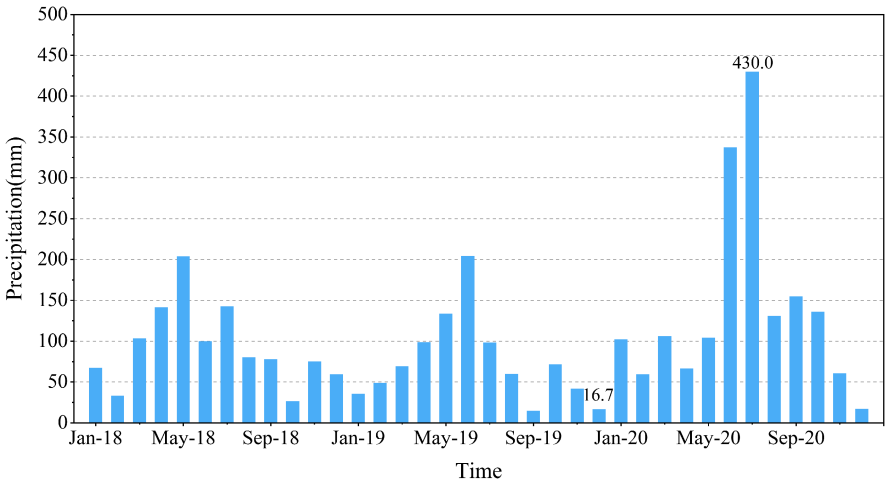

Fig. 2 The cumulative monthly precipitation from 2018 to 2020.

which were processed to obtain daily cumulative precipitation and monthly cumulative precipitation for the years 2018 to 2020. Figure 2 shows the cumulative precipitation for each month, and it is seen that December, 2019, and July, 2020, exhibited the least and the most precipitation, respectively. Therefore, December and July were used in this study as the dry and wet months, respectively.

This work selected for analysis the days with a high number of stations with non-zero values of precipitation. Rainfall was abundant in the wet months, with the average number of stations with precipitation on the days selected being greater than 70% and the number of stations with precipitation on each day being greater than 50%. Based on the above selection criteria, the periods (i) July 3, 2020, to July 8, 2020, (ii) July 14, 2020 to July 15, 2020, and (iii) July 17, 2020, and July 21, 2020, were chosen as being representative of wet days. The dry months have less rainfall, so the dates with a mean number of stations with precipitation greater than 50% were selected for the study, such that December 16, 2019, to December 25, 2019, were determined to be representative of the dry days.

In addition, based on the monthly observation data from 75 meteorological stations in Hubei Province spanning the period from 2016 to 2020, this study evaluates the accuracy of the monthly precipitation downscaling results.

2.3 Environmental Variables

The selection of impact factors plays a crucial role in ensuring the effective performance of statistical downscaling methods. At present, topography, NDVI, and LST are the mainstream influencing factors, which are widely used in the spatial downscaling of satellite precipitation data at the annual or monthly scale.[4,42,43] The influence of topography on precipitation is highly intricate, encompassing dynamic effects, such as topography-induced forced uplift and convergence, as well as the interplay of cloud physics and friction.[44] The topography in Hubei Province exhibits a relatively complex pattern, with a descending trend from west to east, which is closely associated with precipitation patterns. Surface temperature also demonstrates a significant statistical relationship with precipitation.[45] The average temperature in the southeast of Hubei Province is higher than that in the northwest. Correspondingly, precipitation levels tend to be higher in the south and lower in the north, with the northwest exhibiting relatively lower levels of precipitation. Furthermore, vegetation plays a significant role in capturing water vapor, influencing precipitation dynamics. NDVI is often used to study the distribution of vegetation, so NDVI is often used as a vegetation factor to consider the relationship between vegetation and precipitation.[44] In summary, the complex precipitation in Hubei Province is statistically correlated with regional topography, surface temperature, and NDVI. Therefore, this study selects topography, surface temperature difference between day and night, NDVI, latitude, and longitude as input variables for the RFR model and the MLR model.

The DEM data used in this study were all obtained from the SRTMDEM 90M resolution raw elevation data provided by the Geospatial Data Cloud,[46] and six images covering the topography of Hubei Province were selected and resampled by ArcGIS with spatial resolutions of 0.1 deg × 0.1 deg and 0.01 deg × 0.01 deg.

The NDVI is an important parameter reflecting the relevant state of vegetation.[47] The NDVI data selected for this study were all taken from the MOD13A3 product from LAADS DAAC,[48] and the outliers were removed in ArcGIS and then multiplied by 0.0001 to give NDVI values between −1 and 1. Land surface temperature data were obtained from the MOD11A2 product from LAADS DAAC,[49] and these data were an eight-day synthetic product. These data were first processed in ArcGIS as the land surface daytime temperature (LSTD) and the nocturnal land surface temperature (LSTN), then spliced, cropped, and unit converted, and, lastly, the Nibble tool was implemented to fill missing-data cells in land surface temperature data. Finally, the NDVI and LST data need to be resampled in ArcGIS with a spatial resolution of 0.1 deg × 0.1 deg and 0.01 deg × 0.01 deg.

3 Methodology

3.1 Downscaling Method

3.1.1 *Multifractal model for daily precipitation*

MF analysis is a quantitative tool to describe the mass distribution of fractal spaces. Rainfall can be represented as an MF process because of its spatial discontinuity and spatial variability characteristics. The MF model describes the spatial variation characteristics of precipitation as a random cascade process (Fig. 3).

The homogeneous MF precipitation is expressed as

$$M^{l}_{(i,j)} = \begin{cases} M^{l-1}_{(\lfloor i/2 \rfloor, \lfloor j/2 \rfloor)} \times W_{(i,j)}, & l > 0 \\ M^{0}_{(i,j)}, & l = 0 \end{cases}, \tag{2}$$

in which i denotes the row index in the grid, whereas j denotes the column index in grid, and $i, j = 0,1,2,\ldots,2l-1$; $M^{l}_{(i,j)}$ denotes the value of l'th homogeneous MF precipitation in cell (i,j); $W_{(i,j)}$ denotes the random cascade weight applied to the cell (i,j) precipitation to obtain level l homogeneous precipitation; $M^{0}_{(i,j)}$ denotes the homogeneous MF precipitation grid of the first level with cell size equal to 0.1 deg.

The random cascade weights $W_{(i,j)}$ in Eq. (1) influence the downscaling results directly, and they are calculated by Eqs. (3)–(5):[30,32]

$$W_{(i,j)} = BY_{(i,j)}, \tag{3}$$

$$P[B = 0] = 1 - b^{-\beta}, \quad P[B = b^{\beta}] = b^{-\beta}, \tag{4}$$

$$Y_{(i,j)} = b^{-\delta^2 \frac{\ln b}{2} + \delta X_{(i,j)}}, \tag{5}$$

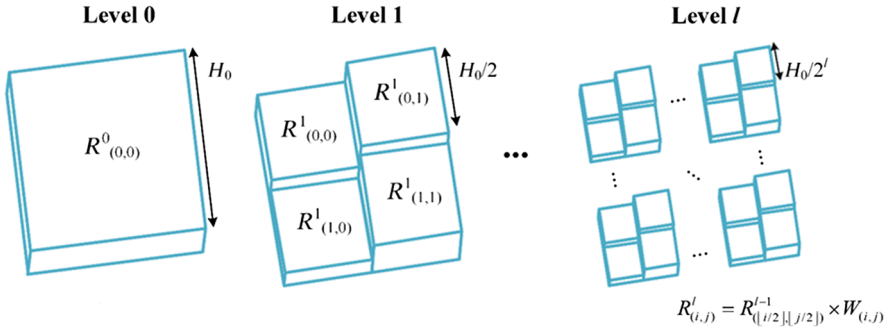

Fig. 3 The random cascade processes. See Eq. (12) for $R^{l}_{(i,j)}$, the l'th level estimated precipitation in cell (i,j).

where β and δ are the parameters of the β-lognormal model; b denotes the branching number with value of 4 in this paper. $X_{(i,j)}$ denotes the random variable following a standard normal distribution according to classical MF theory.

Combining Eqs. (3)–(5), and the Mandelbrot–Kahane–Peyriere equations,[28,29,32,50,51] $W_{(i,j)}$ yields the set of Eqs. (6)–(10):

$$P[W_{(i,j)} = 0] = 1 - b^{-\beta}, \tag{6}$$

$$P\left[W_{(i,j)} = b^{\beta-\delta^2\frac{\ln b}{2}+\delta X_{(i,j)}}\right] = b^{-\beta}, \tag{7}$$

$$\beta = 1 + \frac{\tau'(q)}{2} - \delta^2\frac{\ln\, b}{2}(2q-1), \tag{8}$$

$$\tau(q) = \lim_{l\to\infty}\frac{\log_2 M^l(q)}{-\log_2 \lambda_l}, \tag{9}$$

$$\lambda_l = b^{-l/2}, \quad l = 0,1,2,\ldots., \tag{10}$$

where λ_l represents the level l scale ratio; $M^l(q)$ denotes the q'th order statistical moment of $M^l_{(i,j)}$, and q is in the range $[-10,10]$ in this work; $\tau(q)$ denotes the quality index describing the fractal characteristic of precipitation; and $\tau'(q)$ and $\tau''(q)$ denote the first and second derivatives of $\tau(q)$, respectively. The parameter δ is defined as

$$\delta^2 = \frac{\tau''(q)}{2\,\ln\, b}. \tag{11}$$

The homogeneous MF precipitation fields $M^l_{(i,j)}$ are calculated with Eq. (12):

$$R^l_{(i,j)} = M^l_{(i,j)} \cdot G^l_{(i,j)}, \tag{12}$$

in which $R^l_{(i,j)}/G^l_{(i,j)}$ denotes the value of the l'th level estimated heterogeneous precipitation in cell (i,j); Eq. (13) calculates the level 3 heterogeneous precipitation corresponding to the IMERG precipitation:

$$G^3_{(i,j)} = N^3 \times \frac{A_{(i,j)}}{\sum\limits_{i,j} A_{(i,j)}}, \tag{13}$$

in which $A_{(i,j)}$ denotes the monthly precipitation value of R^{IMERG} in cell (i,j) in a specific month during the period of analysis; N^3 denotes the total number of level 3 grid cells.

From Eqs. (12) and (13), it is possible to calculate $M^3_{(i,j)}$ from the IMERG data. The expected value of $W_{(i,j)}$ equals 1; therefore, $M^0_{(i,j)}$ equals the average of $M^3_{(i,j)}$ from Eqs. (8) and (11) with which to calculate the higher resolution homogeneous precipitation ($M^l_{(i,j)}$).

The recovering of level l precipitation's heterogeneity is achieved with Eq. (14):[28,29,32,50,51]

$$R^l_{(i,j)} = T^l_{(i,j)} \times \frac{M^l_{(i,j)}G^l_{(i,j)}}{\sum_{i,j} M^l_{(i,j)}G^l_{(i,j)}}, \tag{14}$$

in which T^l denotes the large-scale forcing factor, which is calculated with Eq. (15):

$$T_{(i,j)}{}^l = N^l \times R^3_{(\bar{i},\bar{j})} \quad \bar{i} = [i/2^{l-3}], \quad \bar{j} = [j/2^{l-3}]. \tag{15}$$

3.1.2 *Random forest regression model for monthly precipitation*

The random forest (RF) algorithm is a machine learning algorithm that integrates multiple trees based on the idea of integrated learning, with the basic building blocks being decision trees that can be used for classification and regression.[43,52] The RFR model is a multiple regression tree constructed for continuous random variables, where the final prediction is the average of the outputs of all the decision trees. The selection of features for each tree at each step is random,

so each tree corresponds to multiple random vectors. Assume that the original training set is K, the model is divided into the following main steps.[52]

Boostrap sampling with put-back was applied to the original training set K to introduce randomness in the model and improve model diversity. Ultimately, multiple sampling is performed to generate different training sets depending on the number of regression trees to be constructed. The resulting training set is K_f, $f = 1,2,3,\ldots,S$.

Regression trees are constructed separately for each sub-sample set. In the split construction process of each regression tree, M number of features are randomly selected from all the attribute features, and the optimal feature is chosen among them based on the minimum mean squared error until the tree grows to the maximum.

The final regression result is calculated with Eq. (16):

$$g(x_t) = \frac{1}{S}\sum_{f=1}^{S} h_f(x_t), \tag{16}$$

where x_t, $h_f(x_t)$, and S denote the sample to be tested, the predicted results for each regression tree, and the number of trees, respectively. Figure 4 shows the prediction process of the RFR model.

The random forest model typically involves adjusting two parameters: the number of decision trees in the forest and the number of predictor variables randomly selected on each node of the trees ($mtry$).[53,54] In this study, following the parameter settings proposed by Guan et al.,[55] an initial RFR model is constructed with 100 trees, and the value of M ($mtry$) is set as the square root of the number of features. The accuracy of the model is then evaluated based on site data to identify the months with the highest accuracy for wet months (July) and dry months (December) during 2016 to 2020. Subsequently, the parameters of the RFR model for these two months are further adjusted using cross-validation. The optimal number of trees in the forest is determined as 433 for the wet month and 42 for the dry month. The model implementation is carried out using the scikit-learn package in Python.

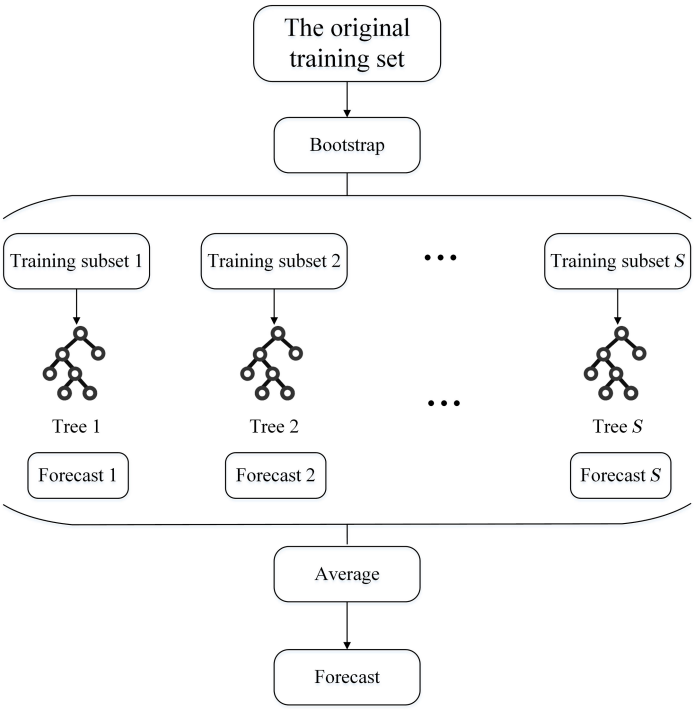

Fig. 4 Schematic of the RFR model.

3.1.3 *Multiple linear regression model for monthly precipitation*

The MLR model is a widely used regression model that involves analyzing the correlation between two or more independent variables and a dependent variable to establish a prediction model. Jia et al.[56] were among the first to apply the MLR model in the context of precipitation downscaling. The MLR model can be described by Eq. (17):

$$P = a + b1 * x1 + b2 * x2 + b3 * x3, \tag{17}$$

where P represents the precipitation, x represents the environmental variable that has a linear relationship with precipitation, b represents the coefficient of the variable, and a represents the intercept of the linear regression model.

3.1.4 *Recovering the heterogeneity of the MF model with monthly precipitation*

It is seen in Eq. (14) the MF model requires the l'th level heterogeneous precipitation to recover heterogeneity. The high spatial resolution monthly precipitation is essentially representative of the spatial distribution characteristics of precipitation for that month. Therefore, the MF-RFR model introduces high resolution (0.01 deg × 0.01 deg) monthly precipitation after RFR downscaling to improve the heterogeneity of high resolution daily precipitation. Recall in Eq. (14) $G^{l}_{(i,j)}$ denotes the value of l'th level heterogeneous precipitation in cell (i, j), and it equals the monthly precipitation downscaled with the RFR at a 0.01 deg × 0.01 deg spatial resolution. A flowchart of the MF-RFR downscaling process is displayed in Fig. 5.

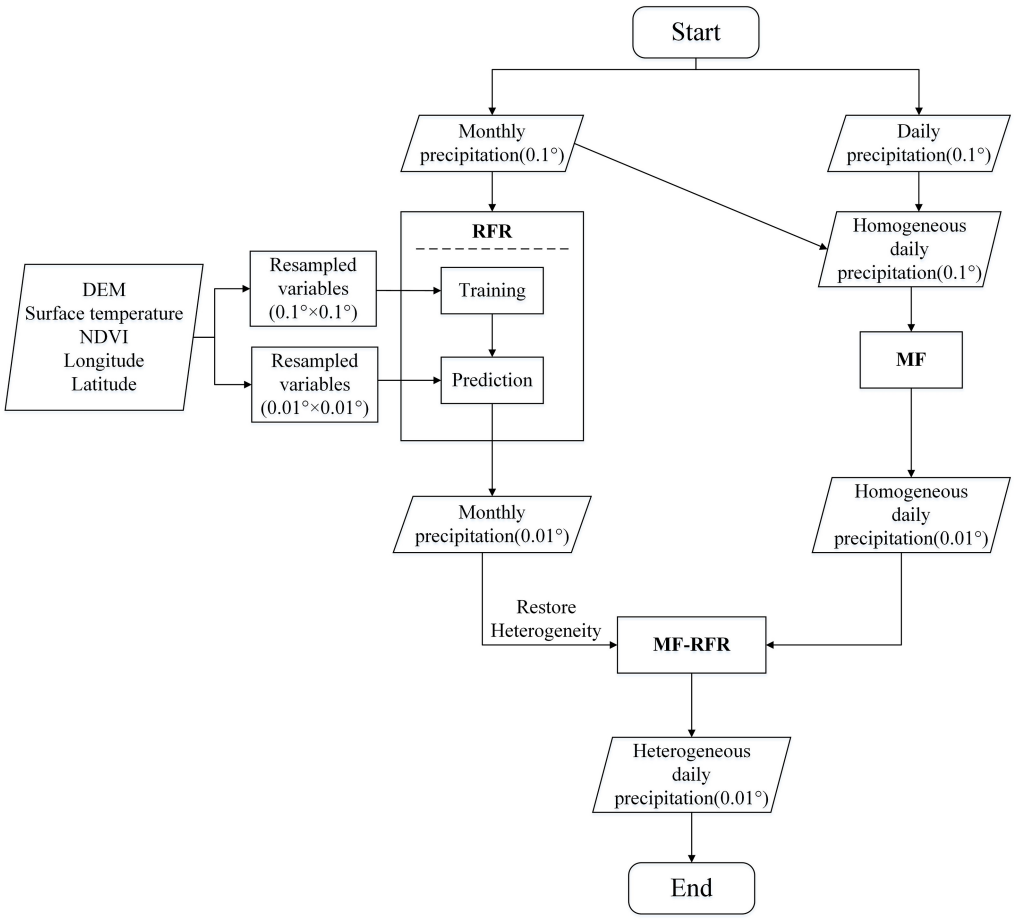

Fig. 5 Flowchart of the MF-RFR downscaling process.

The model construction approach for the MF-MLR model is identical to that of the MF-RFR model. In other words, the high-resolution monthly precipitation data obtained after the MLR model are incorporated into the MF model to restore the heterogeneity of daily precipitation.

3.2 Validation

Most of the existing methods for evaluating the accuracy of precipitation data rely on using observed precipitation data from meteorological stations as reference data. The evaluation of accuracy is then performed by considering multiple evaluation indices.[13,57]

This work was conducted to evaluate the experimental results based on precipitation data from 75 meteorological stations in Hubei Province. This paper adopts the RMSE, bias, mean absolute error (MAE), and the correlation coefficient (CC) to measure the accuracy of precipitation predictions. CC close to 1 (or 0) indicates linear statistical relation (or no relation). The bias, MAE, and RMSE close to zero are desirable from the perspective of calculating predictions nearly identical to observations. These performance criteria are given by Eqs. (18)–(21):

$$\mathrm{CC} = \frac{\sum_{z=1}^{Z}(D_z - \overline{D})(E_z - \overline{E})}{\sqrt{\sum_{z=1}^{Z}(D_z - \overline{D})^2 \sum_{z=1}^{Z}(E_z - \overline{E})^2}}, \tag{18}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{z=1}^{Z}(D_z - E_z)^2}{Z}}, \tag{19}$$

$$\mathrm{Bias} = \frac{\sum_{z=1}^{Z}(E_z - D_z)}{\sum_{z=1}^{Z} D_z} \times 100\%, \tag{20}$$

$$\mathrm{MAE} = \frac{\sum_{z=1}^{Z}|D_z - E_z|}{Z}, \tag{21}$$

where z and Z denote the index of the validation data and the total gauge number; D_i and E_i denote the predicted and gage-measured precipitation, respectively; $\overline{D}$ and $\overline{E}$ represent the average predicted precipitation and average measured precipitation, respectively.

4 Results and Discussion

4.1 Analysis of Multifractal

This work selected the IMERG data for Hubei province with 90×55 cells for downscaling analysis of daily precipitation. The spatial distribution of precipitation is commonly modeled with a logarithmic normal distribution. The same phenomenon has been reported by Kedem and Chiu,[58] who reported the lognormal distribution fits well the daily precipitation.

First, calculate $\log_2 \lambda$, $\log_2 M(\lambda, q)$, $\tau(q)$, $f(\alpha)$, α, and $D(q)$ from the daily precipitation after homogenization. According to MF theory, a regional process that obeys the power law is called non-scaling, which is a requirement of fractal phenomena, in which case the $\log_2 \lambda$ and $\log_2 M(\lambda, q)$ plot linearly in a log–log diagram. For the quality index $\tau(q)$ versus q, the relationship is not a simple linear. $f(\alpha)$ is the continuous spectrum and reflects the dimension of MF subsets. When the object is MF, $f(\alpha)$ is a single peak and a convex function of α. The generalized dimension $D(q)$ decreases as q increases.

This work subjected the daily precipitation data for a total of 20 days in the wet and dry months to MF analysis, and the variation pattern of each parameter indicates that the daily precipitation conforms to the MF characteristics. Figure 6 shows the MF feature analysis for July 5, 2020. It shown in Fig. 6 that there is essentially a good linear statistical association in the log–log diagram, and variation of the quality index $\tau(q)$ with respect to q reveals the basic characteristics of the MFs. The rest of the parameters are also consistent with multiple fractal phenomena. Figure 7 displays the MF analysis for December 25, 2019. It is shown in Fig. 7 that there is a single peak in the MF spectra and the generalized dimension $D(q)$ is monotonically decreasing, which means the daily precipitation is non-scaling within the scale range. This means the daily precipitation conforms to the fractal structure. These findings suggest plausible downscaling of daily precipitation characterized by the random cascade process.

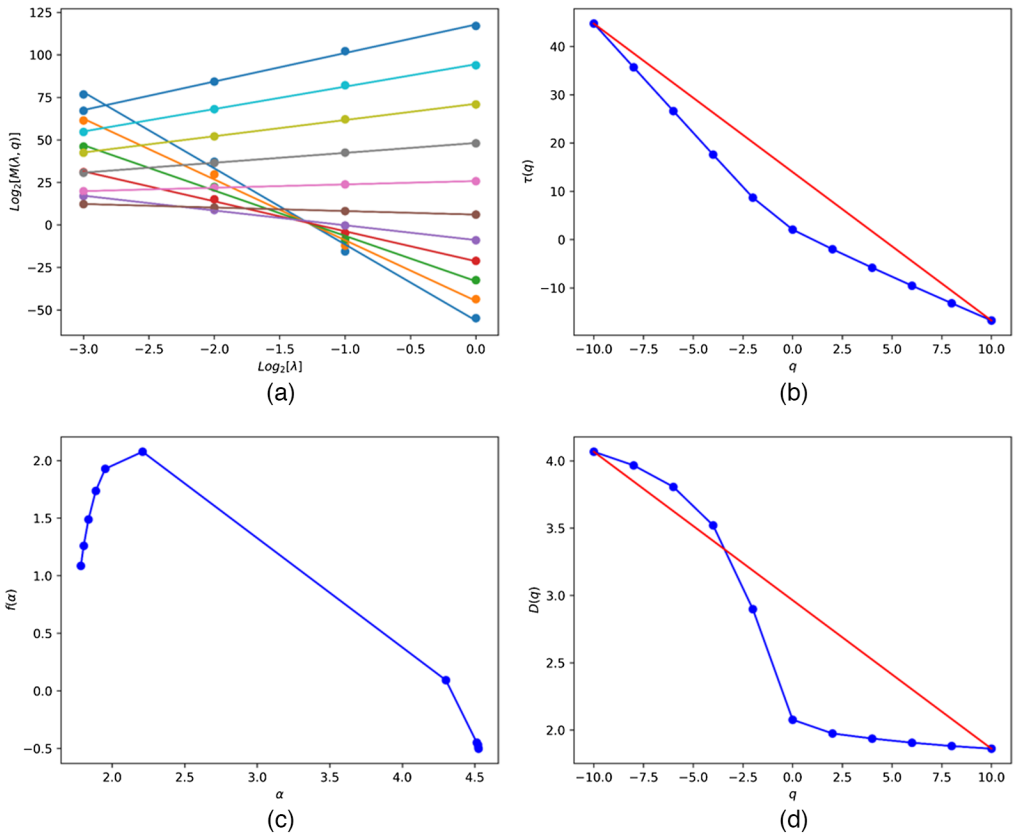

Fig. 6 MF characteristics of wet daily IMERG precipitation (July 5, 2020). (a) Log–log diagram, (b) quality index, (c) convex function, and (d) generalized dimension.

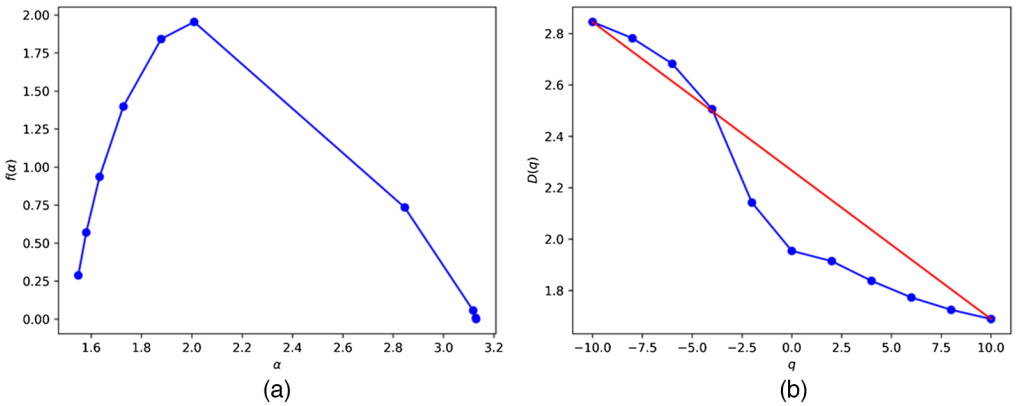

Fig. 7 MF characteristics of dry daily IMERG precipitation (December 25, 2019). (a) Convex function and (b) generalized dimension.

The random variables of the MF model obey the standard normal distribution. This work simulated the MF downscaling model 20 times and employs the average of the results as the downscaling result. Figure 8 shows the original IMERG product for a wet day (July 17, 2020) and its high-resolution precipitation obtained with the MF model. Figure 9 shows the original IMERG product for a dry day (December 18, 2019) and its high-resolution precipitation obtained with the MF model. The downscaled image has a distinct mosaic-like appearance despite the presence of detailed features. This is mainly because of the size of each image element after the MF downscaling model is independent of the size of the neighboring image elements and only changes within its image element. Precipitation after downscaling is closely related with the initial precipitation according to Eqs. (3)–(5). The average precipitation remains the same no matter how many downscaling levels are employed.

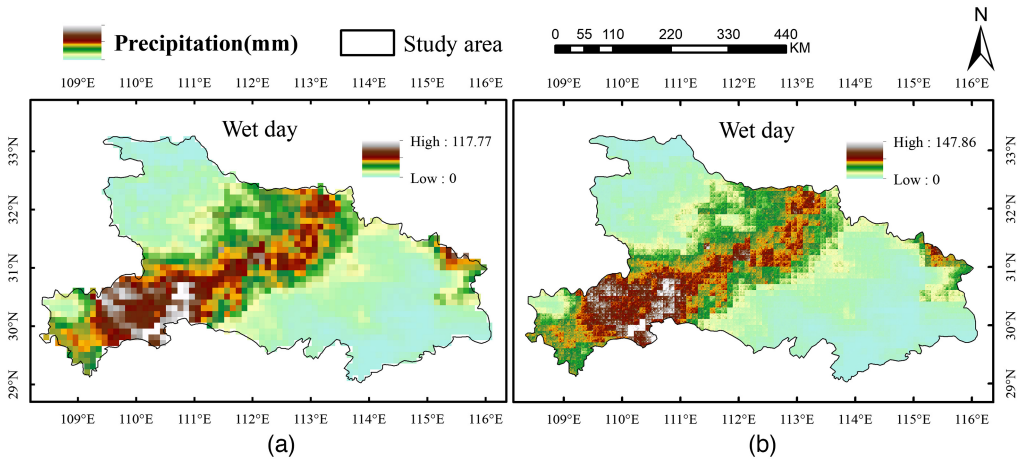

Fig. 8 Precipitation maps corresponding to two stages on the wet day (July 17, 2020). (a) IMERG precipitation and (b) downscaled results with the MF model.

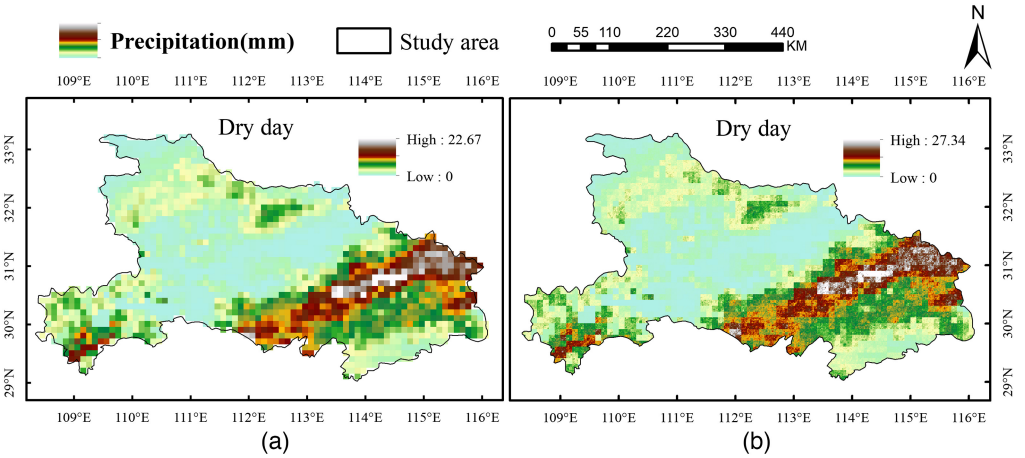

Fig. 9 Precipitation maps corresponding to two stages on the dry day (December 18, 2019). (a) IMERG precipitation and (b) downscaled results with the MF model.

This work evaluates the accuracy of daily precipitation and the average accuracy is obtained for 10 wet days and 10 dry days. The accuracy evaluation of the MF downscaling model is presented in Table 1. The precipitation with 0.01 deg × 0.01 deg resolution on wet days is somewhat more accurate than on dry days. This means that the MF downscaling model has better applicability on wet days, but the overall accuracy is poor for both wet and dry days. This work employed heterogeneous precipitation with low-resolution to recover the spatial heterogeneity of high-resolution precipitation according to Eq. (14). This affects the overall accuracy of the calculated results.

Table 1 Average of accuracy assessments of wet and dry daily precipitation downscaled with the MF model.

Model	Time	CC	Bias	MAE	RMSE
IMERG	Wet day	0.694	0.282	13.078	21.184
	Dry day	0.467	5.524	2.131	3.097
MF	Wet day	0.681	0.296	13.426	21.594
	Dry day	0.426	5.744	2.140	3.149

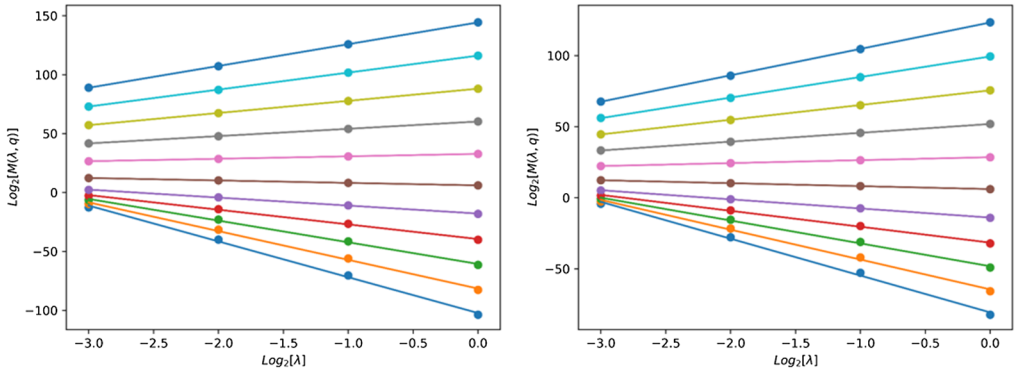

Fig. 10 MF characteristics of July, 2016, (wet month) and December, 2016, (dry month). (a) July 2016 and (b) December 2016.

Table 2 Average of accuracy assessments of wet and dry monthly precipitation downscaled with the MF model.

Model	Time	CC	Bias	MAE	RMSE
IMERG	Wet month	0.697	0.154	70.349	87.631
	Dry month	0.757	0.418	10.285	13.001
MF	Wet month	0.602	0.151	77.537	96.792
	Dry month	0.727	0.416	11.096	14.760

To further confirm the MF characteristics of precipitation data, this study conducted the MF downscaling experiments on IMERG monthly precipitation data from July and December spanning the years 2016 to 2020. The analysis revealed that the variation of parameters across all months aligns with MF characteristics. Figure 10 takes July 2016 (wet month) and December 2016 (dry month) as examples to show the changes of some parameter characteristics. The diagram demonstrates a strong linear relationship between $\log_2 \lambda$ and $\log_2 M(\lambda, q)$, indicating that the monthly precipitation exhibits scale-free properties and adheres to MF characteristics within the specified range of scale variations. Table 2 presents the average accuracy for wet and dry months from 2016 to 2020. Compared to the original IMERG monthly precipitation data, the MF downscaling results show a slight improvement in data deviation, with an average enhancement of 0.25%. However, they exhibit poorer performance in terms of CC, MAE, and RMSE. Among them, the average loss of CC for wet and dry months amounts to 6.25%. In summary, IMERG precipitation data demonstrate favorable MF characteristics at both daily and monthly scales, and the MF model can be used to improve the spatial resolution of precipitation data. The overall accuracy of the MF downscaling approach is comparable to the original data, albeit with varying degrees of loss. Therefore, it is necessary to introduce high-precision monthly precipitation data to restore the heterogeneity within the MF model.

4.2 Analysis of the RFR and MLR Models

The statistical data from meteorological stations revealed that July, 2020, are the month with the heaviest precipitation and December 2019 is the month with the lowest precipitation. Therefore, in this work, July and December represent the months with most and least precipitation, respectively. Monthly precipitation has a more stable spatial distribution, and it is suitable to represent the spatial distribution of daily precipitation within the same month. Therefore, this paper uses downscaled high-precision monthly precipitation to restore the spatial heterogeneity of daily precipitation. Monthly data from 2016 to 2020 were used in the RFR model, increasing the spatial resolution of monthly precipitation from 0.1 deg × 0.1 deg to 0.01 deg × 0.01 deg. The monthly downscaling results with the highest accuracy were then used to restore the

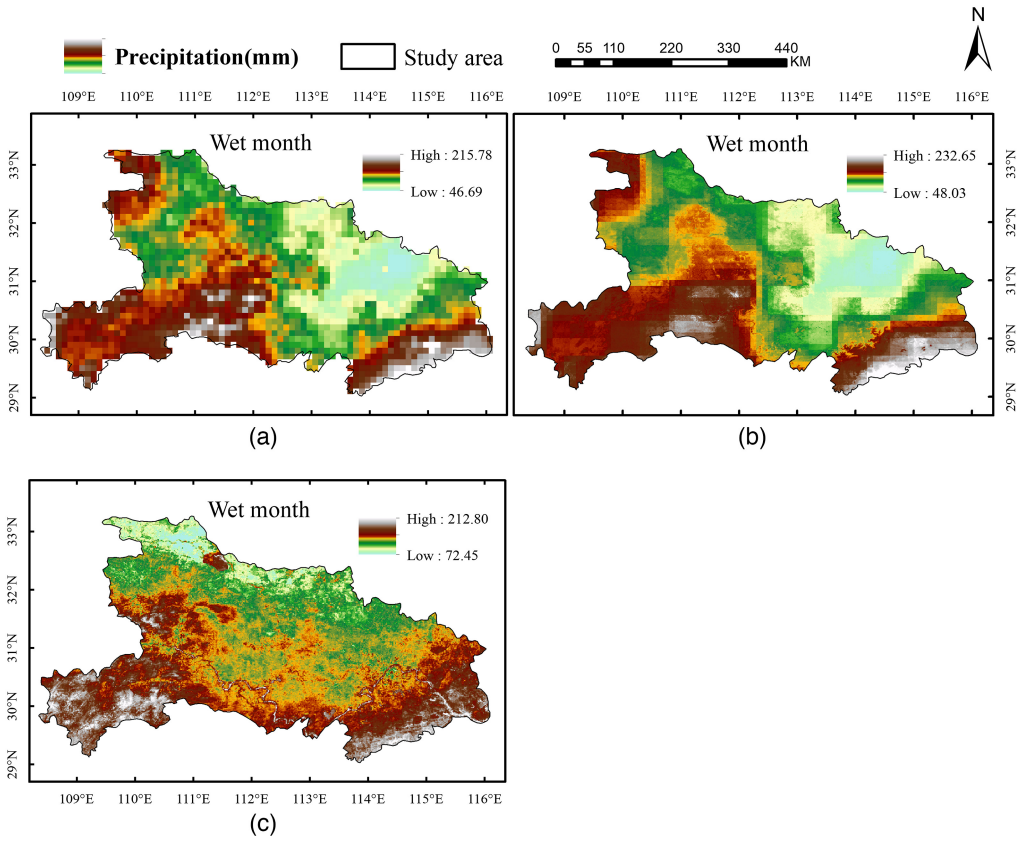

Fig. 11 (a) Original precipitation, (b) downscaled precipitation by the RFR model, and (c) downscaled precipitation by the MLR model in the wet month (July, 2019), Hubei province.

MF heterogeneity. July, 2019, (wet month) and December, 2018, (dry month) were selected as the research period. The MLR method was also applied for monthly downscaling during these two months, and its results were compared with the RFR model method.

Figure 11 shows the comparison of downscaling results from various models in July 2019. As observed in the figure, the spatial distribution characteristics of the monthly precipitation data obtained after downscaling with the RFR model are more similar to the original data when compared with the MLR model. The correspondence of high-value precipitation regions is more accurate in the RFR downscaling results, and local details are more prominent. Figure 12 showcases the comparison of downscaling results from different models in December, 2018. The map reveals that both the RFR model and the MLR model are capable of improving the spatial resolution of the IMERG data. However, the RFR model preserves the original spatial distribution characteristics of precipitation to a greater extent and provides richer local detail information, particularly in the southwestern mountainous areas. These findings indicate that, in terms of spatial distribution, the downscaled precipitation data obtained through the RFR model outperform that of the MLR model.

Table 3 presents the accuracy evaluation of the RFR model and the MLR model for July 2019 (wet month) and December 2018 (dry month). The evaluation results for July 2019 demonstrate that the RFR downscaling results exhibit higher fitting accuracy compared to the non-downscaled data. The deviation between the monthly precipitation and ground station observations decreases, indicating improved accuracy of the downscaled data. The MLR model does not perform as well as the RFR model, with lower accuracy indicators compared to the original IMERG data. Regarding the evaluation results for December, 2018, the RFR model shows varying degrees of improvement in the accuracy indicators after downscaling. The MLR model also demonstrates some improvement in bias, RMSE, and MAE; however, its CC is lower than that of the original data. In summary, the RFR model outperforms the MLR model in both wet and dry months, demonstrating higher accuracy in generating high spatial resolution monthly precipitation data.

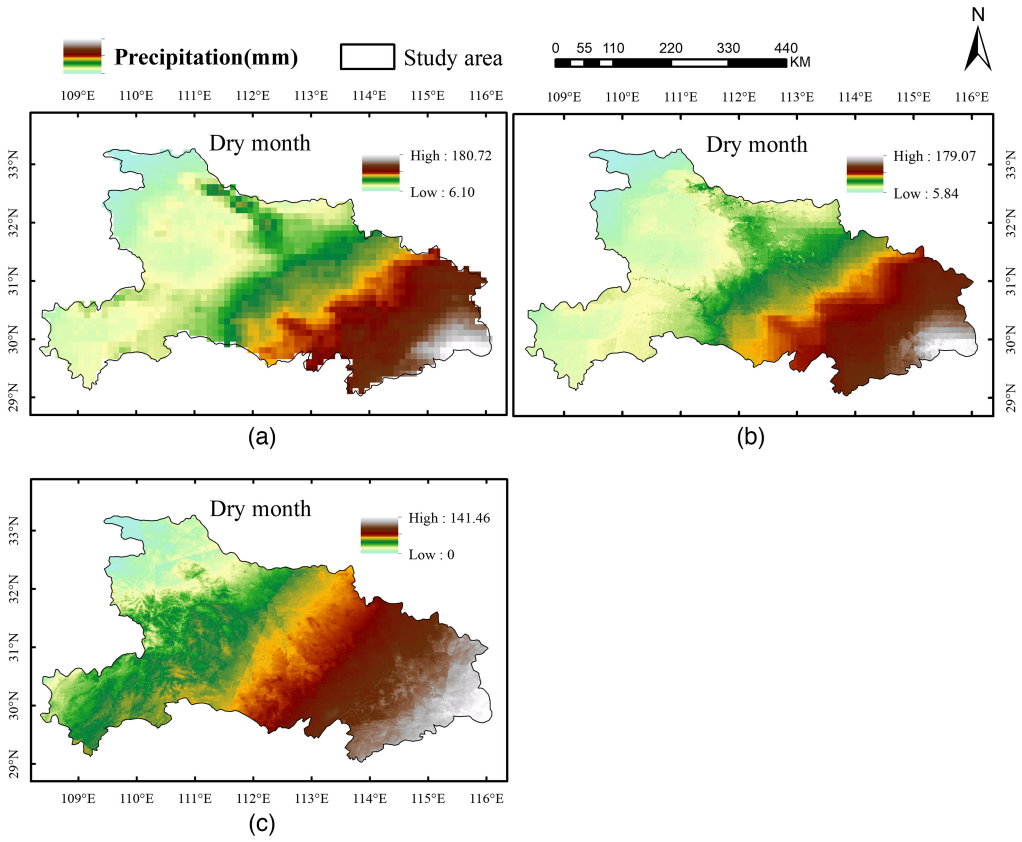

Fig. 12 (a) Original precipitation, (b) downscaled precipitation by the RFR model, and (c) downscaled precipitation by the MLR model in the dry month (December, 2018), Hubei province.

Table 3 Accuracy assessment of the RFR and MLR models.

Time	CC	Bias	MAE	RMSE
Wet month precipitation downscaled by RFR	0.520	0.054	39.610	51.365
Wet month precipitation downscaled by MLR	0.200	0.376	61.757	69.644
Original wet month precipitation	0.484	0.060	41.342	52.757
Dry month precipitation downscaled by RFR	0.911	0.155	15.270	18.675
Dry month precipitation downscaled by MLR	0.899	0.110	14.489	18.105
Original dry month precipitation	0.907	0.155	15.379	19.194

4.3 Analysis of the MF-RFR Model

4.3.1 *Comparative analysis of downscaling results from different models*

Based on the selection of wet days (10 days) and dry days (10 days) outlined in Sec. 2.2.2, this study further chose representative days from each category. Ultimately, July 5, 2020 (the day with the highest rainfall) and December 22, 2019 (the day with the lowest rainfall) were selected as the research period. The spatial resolution of IMERG daily precipitation data was enhanced to 0.01 deg by the MF-RFR and MF-MLR models.

Figure 13 shows the comparison between the original precipitation field and the downscaled precipitation fields using different models. In terms of the spatial distribution trend, all three downscaled precipitation fields generally preserve the overall distribution characteristics of the original IMERG precipitation field and accurately depict the spatial distribution of strong and weak precipitation. Regarding local detail features, both the MF-RFR and MF-MLR models

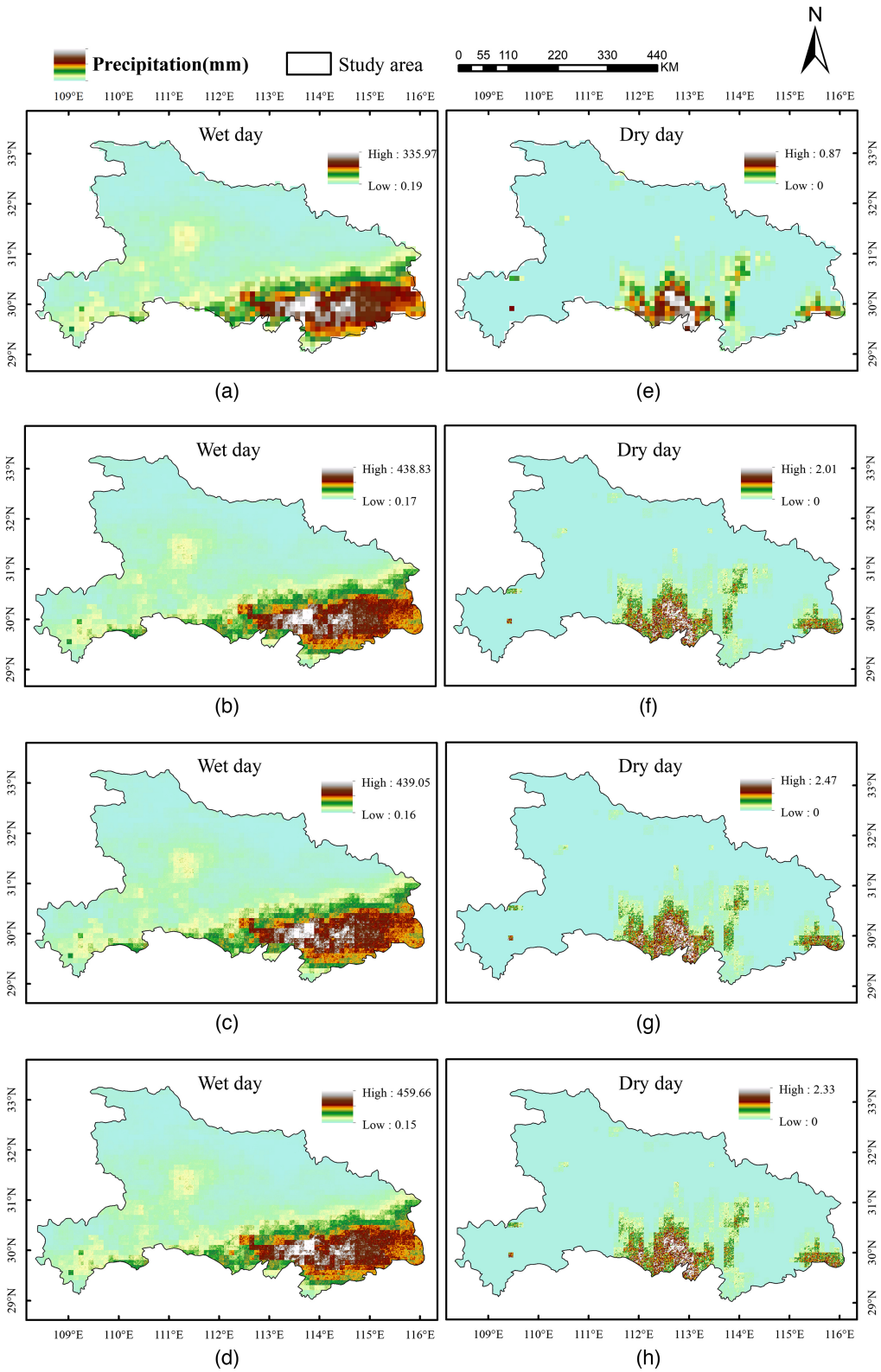

Fig. 13 Precipitation maps corresponding to several stages on the wet day (July 5, 2020) and the dry day (December 22, 2019). (a), (e) IMERG precipitation, (b), (f) downscaled precipitation with the MF model, (c), (g) downscaled precipitation with the MF-RFR model, and (d), (h) downscaled precipitation with the MF-MLR model.

Table 4 Accuracy assessment of wet and dry daily precipitation downscaled with different models.

Model	Time	CC	Bias	MAE	RMSE
IMERG	Wet day	0.585	0.109	37.061	58.277
	Dry day	0.051	9.674	0.060	0.144
MF	Wet day	0.569	0.140	38.206	61.442
	Dry day	−0.001	9.722	0.060	0.154
MF-RFR	Wet day	0.585	0.073	36.460	57.214
	Dry day	0.104	8.339	0.052	0.124
MF-MLR	Wet day	0.579	0.139	38.350	60.902
	Dry day	0.047	10.202	0.062	0.170

exhibit more spatial details compared to the MF model, particularly in the southeastern hilly areas. This further emphasizes the importance of using high-precision monthly precipitation data to restore the heterogeneity of daily precipitation, which plays a crucial role in the downscaling process. Furthermore, the comparison of precipitation fields indicates that the overestimation in the MF-MLR model is significantly greater than that of the MF model and the MF-RFR model on the wet day. Overall, the MF-RFR model developed in this study yields more accurate precipitation distribution.

Table 4 provides the accuracy for the original IMERG, the MF, the MF-MLR, and the MF-RFR models. On the wet day (July 5, 2020), both the MF-RFR and MF-MLR models outperform the MF downscaling model. The MF-RFR model enhances the accuracy of the original IMERG precipitation data. Specifically, the CC remains relatively unchanged, whereas the bias is reduced by 3.6%, the MAE is reduced by 0.601 mm/d, and the RMSE is reduced by 1.063 mm/d. The downscaling effect of the MF-MLR model is notably inferior to that of the MF-RFR model, with lower accuracy indicators compared to the original precipitation data. However, the MF-MLR model does improve the accuracy of the MF model to some extent, emphasizing the importance of restoring heterogeneity. On the dry day (December 22, 2019), the downscaled results from the MF-RFR model exhibit the best performance across the four accuracy indicators compared to the original IMERG data. The CC increases by 5.3% and the bias, MAE, and RMSE decrease to varying degrees. The accuracy of the MF-MLR downscaling model is very similar to that of the MF model, which significantly improves the CC of the MF model. In addition, the MF-RFR model demonstrates better applicability on the dry day, and the overall accuracy is improved more. However, it is important to note that the CC of dry days remains low, indicating that further verification of the applicability of this method in dry weather is necessary in future research. In addition, it is worth exploring the potential impact of the accuracy of the original IMERG precipitation data on the results. Future studies could consider integrating correction techniques with the downscaling model to improve the overall accuracy and applicability of the approach.

Overall, the introduction of high-resolution monthly precipitation data can enhance the numerical accuracy of the MF model to some extent. Among the three models, the MF-RFR model exhibits the highest numerical accuracy. This signifies that the MF-RFR model provides improved predictions of high-resolution precipitation, which are valuable for meteorological disaster assessment, hydrological analysis, and other applications.

4.3.2 *Stability analysis of the MF-RFR model*

To further validate the applicability of the MF-RFR downscaling model, this study performed daily downscaling using the MF-RFR model for the months of July and December in each year from 2018 to 2020.

Table 5 presents the percentage of days in both wet and dry months where the MF-RFR model outperforms the MF model and the MF-RFR model surpasses the original IMERG daily

Table 5 Percentage of days with improved performance by the MF-RFR model.

		Compared with the MF model (%)	Compared with IMERG (%)
CC	Wet day	85	90
	Dry day	86	86
Bias	Wet day	83	73
	Dry day	57	60
MAE	Wet day	68	76
	Dry day	76	73
RMSE	Wet day	54	43
	Dry day	69	50

precipitation data. In wet months, compared to the MF model, the MF-RFR model performs better on the four indicators, with over 50% of the days showing improved accuracy or a gap within 0.05. Notably, the most significant improvements are observed in CC and bias, where over 80% of the days exhibit enhanced accuracy or a gap within 0.05. These results demonstrate the ability of the MF-RFR model to enhance the accuracy of the MF model. When compared with the original IMERG precipitation data, the CC, bias, and MAE of the precipitation field after MF-RFR downscaling have more than 70% of the days with better effect or the gap is only within 0.05. Among them, 90% of the days exhibit enhanced data consistency. However, the improvement in RMSE is slightly lower compared to the other three indicators. This discrepancy may be attributed to the influence of heavy precipitation and frequent rainstorms, which aligns with the findings of Gu et al.[21] Nevertheless, nearly half of the days still demonstrate an improvement in the downscaling effect. Overall, the proposed MF-RFR model performs well during wet months, as it either improves or basically maintains the numerical accuracy of the original precipitation data. Furthermore, it enhances the performance of the MF model to varying degrees.

The accuracy statistics for dry months are presented in Table 5. The MF-RFR model proposed in this paper significantly enhances the data accuracy of the MF model. The four indicators have more than 50% of the days better than the MF model or the gap is within 0.05. Notably, CC exhibits the most significant improvement, with over 86% of the days demonstrating higher data consistency compared to the MF model or a gap within 0.05. When compared to the original IMERG precipitation data, the MF-RFR model outperforms in CC, BIAS, MAE, and RMSE. Over half of the days exhibit improved accuracy or maintain the accuracy of the original data. These findings highlight the overall superiority of the MF-RFR model in dry months, with stronger stability and higher accuracy indices. Overall, the MF-RFR model can obtain daily precipitation with higher spatial resolution and accuracy in both wet and dry months.

5 Conclusions

This paper introduced a downscaling model called MF-RFR to create high-resolution daily precipitation from the IMERG dataset. The models were applied to Hubei province precipitation (China). The RFR model was employed to downscale the IMERG monthly precipitation data from 0.1 deg × 0.1 deg to 0.01 deg × 0.01 deg. The selected influencing factors include DEM, LSTD, LSTN, NDVI, latitude, and longitude. The downscaling effect was compared and analyzed with the MLR model. In addition, the IMERG daily precipitation data were downscaled from 0.1 deg × 0.1 deg to 0.01 deg × 0.01 deg using the random cascade process of the MF model. The monthly precipitation data after downscaling with the RFR and MLR models were introduced to restore the spatial heterogeneity of daily precipitation. Finally, based on site observation data, a comparison of the downscaling effects of the MF, MF-MLR, and MF-RFR models was conducted, along with an analysis of the stability of the MF-RFR model. The main conclusions include the following:

(1) IMERG precipitation products exhibit evident MF characteristics at both daily and monthly scales, suggesting the feasibility of downscaling precipitation using the random cascade process. In comparison to the original IMERG precipitation field, the MF downscaling model increases the spatial resolution by 10 times. However, some degree of numerical accuracy loss was observed. On average, the daily-scale correlation experienced a loss of 2.7%, whereas the monthly scale correlation experienced a loss of 6.25%.

(2) Regarding monthly precipitation data, the downscaling results of the RFR model showed higher data quality. Compared to the traditional MLR model, the RFR model preserved the spatial distribution characteristics of precipitation before and after downscaling, enhancing the accuracy of the original precipitation products in both dry and wet months, with an overall accuracy improvement. The CC can reach up to 0.911. Consequently, the RFR model can further enhance the performance of the MF model.

(3) This paper proposes the use of high-precision monthly precipitation data to restore the spatial heterogeneity of daily precipitation, which is a crucial step in the downscaling process. Both the MF-RFR model and the MF-MLR model improve upon the traditional MF model, enhancing its accuracy and the ability to capture local details. Among them, the MF-RFR model demonstrates significantly better downscaling performance than the MF-MLR model, with accuracy improvements of up to 10.3%. Moreover, the MF-RFR model achieves higher accuracy compared to the original precipitation data, providing high-resolution precipitation data of superior quality. This is particularly beneficial during the dry day, where substantial overall accuracy improvements are observed.

(4) Considering the statistical analysis results across multiple research periods, the MF-RFR model exhibits good stability in both wet and dry months. The accuracy evaluation after downscaling shows that over 50% of the days demonstrate improved accuracy or maintain the accuracy of the original IMERG precipitation data. Notably, over 86% of the days show improvements in CC. Furthermore, over 54% of the days outperform the MF model in terms of accuracy indicators. These findings indicate that the MF-RFR downscaling method is well-suited for the downscaling of satellite daily precipitation data and holds promising application prospects.

Data Availability

All the IMERG precipitation data created or used during this study are openly available from the Goddard Earth Sciences Data and Information Services Center website at https://disc.gsfc.nasa.gov. The DEM data used in this study were all obtained from the SRTMDEM 90M resolution raw elevation data provided by the Geospatial Data Cloud (http://www.gscloud.cn). The NDVI data selected for this study were all taken from LAADS DAAC (https://ladsweb.modaps.eosdis.nasa.gov/search/). Land Surface Temperature data were obtained from the MOD11A2 product from LAADS DAAC (https://ladsweb.modaps.eosdis.nasa.gov/search/).

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant Nos. 41501584 and 41871304) and supported by the Key Research and Development Program of Hubei (Grant No. 2022BCA080). The authors declare no conflicts of interest.

References

1. C. Ma et al., "Changes in precipitation and temperature in Xiangjiang River Basin, China," *Theor. Appl. Climatol.* **123**, 859–871 (2016).
2. Z. Wang et al., "Combination GRACE and TRMM estimate of water storage capacity and flood potential in Afghanistan," *Geomat. Inf. Sci. Wuhan Univ.* **41**(1), 58–65 (2016).
3. Z. Ma et al., "Comparisons of spatially downscaling TMPA and IMERG over the Tibetan Plateau," *Remote Sens.* **10**(12), 1883 (2018).
4. C. Zhan et al., "Spatial downscaling of GPM annual and monthly precipitation using regression-based algorithms in a mountainous area," *Adv. Meteorol.* **2018**, 1506017 (2018).
5. C. Kidd et al., "So, how much of the Earth's surface is covered by rain gauges?" *Bull. Am. Meteorol. Soc.* **98**(1), 69–78 (2017).

6. Z. Sun et al., "How China's Fengyun satellite precipitation product compares with other mainstream satellite precipitation products," *J. Hydrometeorol.* **23**(5), 785–806 (2022).
7. Z. Duan et al., "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 (2016).
8. X. Wang et al., "Similarities and improvements of GPM IMERG upon TRMM 3B42 precipitation product under complex topographic and climatic conditions over Hexi region, Northeastern Tibetan Plateau," *Atmos. Res.* **218**, 347–363 (2019).
9. Q. Yan et al., "Study of sea-surface slope distribution and its effect on radar backscatter based on global precipitation measurement Ku-band precipitation radar measurements," *J. Appl. Remote Sens.* **12**(1), 016006 (2018).
10. D. B. Wright, T. R. Knutson, and J. A. Smith, "Regional climate model projections of rainfall from US landfalling tropical cyclones," *Clim. Dyn.* **45**, 3365–3379 (2015).
11. X. Su et al., "Multi-site statistical downscaling method using GCM-based monthly data for daily precipitation generation," *Water* **12**(3), 904 (2020).
12. D. Maraun et al., "Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user," *Rev. Geophys.* **48**(3), RG3003 (2010).
13. H. Wang et al., "A GWR downscaling method to reconstruct high-resolution precipitation dataset based on GSMaP-Gauge data: a case study in the Qilian Mountains, Northwest China," *Sci. Total Environ.* **810**, 152066 (2022).
14. L. Shi, Q. Q. He, and J. Yang, "Downscaling modeling of the GPM IMERG precipitation product and comparative analysis in the Fujian-Zhejiang-Jiangxi region," *J. Geo-Inf. Sci.* **21**(10), 1642–1652 (2019).
15. E. Sharifi, B. Saghafian, and R. Steinacker, "Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques," *J. Geophys. Res.: Atmos.* **124**(2), 789–805 (2019).
16. Z. Shen and B. Yong, "Downscaling the GPM-based satellite precipitation retrievals using gradient boosting decision tree approach over Mainland China," *J. Hydrol.* **602**, 126803 (2021).
17. C. Chen et al., "An improved spatial downscaling procedure for TRMM 3B43 precipitation product using geographically weighted regression," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **8**(9), 4592–4604 (2015).
18. F. Chen et al., "Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity," *Int. J. Remote Sens.* **35**(9), 3074–3093 (2014).
19. Y. Peng et al., "A geographically and temporally weighted regression model for spatial downscaling of MODIS land surface temperatures over urban heterogeneous regions," *IEEE Trans. Geosci. Remote Sens.* **57**(7), 5012–5027 (2019).
20. M. Wang et al., "Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China," *Remote Sens.* **9**(12), 1278 (2017).
21. J. J. Gu et al., "A high-precision spatial downscaling method for remotely sensed precipitation data in the Luanhe River basin," *South-to-North Water Transf. Water Sci. Technol.* **19**(5), 862–873 (2021).
22. S. Xu et al., "A new satellite-based monthly precipitation downscaling algorithm with non-stationary relationship between precipitation and land surface characteristics," *Remote Sens. Environ.* **162**, 119–140 (2015).
23. X. Fan and H. Liu. "Downscaling method of TRMM satellite precipitation data over the Tianshan Mountains," *J. Nat. Resour.* **33**(3), 478–488 (2018).
24. F. Chen et al., "Downscaling satellite-derived daily precipitation products with an integrated framework," *Int. J. Climatol.* **39**(3), 1287–1304 (2019).
25. A. Jafarzadeh et al., "Examination of various feature selection approaches for daily precipitation downscaling in different climates," *Water Resour. Manage.* **35**, 407–427 (2021).
26. R. Deidda, R. Benzi, and F. Siccardi, "Multifractal modeling of anomalous scaling laws in rainfall," *Water Resour. Res.* **35**(6), 1853–1867 (1999).
27. S. Lovejoy, "Area-perimeter relation for rain and cloud areas," *Science* **216**, 185–187 (1982).
28. B. B. Mandelbrot and B. B. Mandelbrot, "Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier," *J. Fluid Mech.* **62**(2), 331–358 (1974).
29. T. M. Over and V. K. Gupta. "Statistical analysis of mesoscale rainfall: dependence of a random cascade generator on large-scale forcing," *J. Appl. Meteorol. Climatol.* **33**(12), 1526–1542 (1994).
30. S. Lovejoy and B. B. Mandelbrot, "Fractal properties of rain, and a fractal model," *Tellus A* **37**(3), 209–232 (1985).
31. D. Schertzer and S. Lovejoy, "Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes," *J. Geophys. Res.: Atmos.* **92**(D8), 9693–9714 (1987).
32. A. Pathirana and S. Herath, "Multifractal modelling and simulation of rain fields exhibiting spatial heterogeneity," *Hydrol. Earth Syst. Sci.* **6**(4), 695–708 (2002).
33. A. Posadas et al., "Spatial random downscaling of rainfall signals in Andean heterogeneous terrain," *Nonlinear Processes Geophys.* **22**(4), 383–402 (2015).

34. G. Xu et al., "Spatial downscaling of TRMM precipitation product using a combined multifractal and regression approach: demonstration for South China," *Water* **7**(6), 3083–3102 (2015).
35. J. Xia et al., "Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004–2011," *Malar J.* **14**(1), 1–10 (2015).
36. Q. Zhuang et al., "The evaluation and downscaling-calibration of IMERG precipitation products at sub-daily scales over a metropolitan region," *J. Flood Risk Manage.* **16**(3), e12902 (2023).
37. A. Gemitzi, N. Koutsias, and V. Lakshmi, "A spatial downscaling methodology for GRACE total water storage anomalies using GPM IMERG precipitation estimates," *Remote Sens.* **13**(24), 5149 (2021).
38. J. Xu et al., "Calibrating GPM IMERG late-run product using ground-based CPC daily precipitation data: a case study in the Beijing-Tianjin-Hebei urban agglomeration," *Remote Sens. Lett.* **12**(9), 848–858 (2021).
39. R. J. Joyce and P. Xie, "Kalman filter-based CMORPH," *J. Hydrometeorol.* **12**(6), 1547–1563 (2011).
40. Global precipitation measurement precipitation processing system, ftp://gpmweb2.pps.eosdis.nasa.gov/pub/GPMfilespec/filespec.GPM.pdf (2020).
41. China Meteorological Data Service Centre, Ground-based basic meteorological observations in China, Real Time, http://data.cma.cn.
42. W. Jing et al., "A spatial downscaling algorithm for satellite-based precipitation over the Tibetan plateau based on NDVI, DEM, and land surface temperature," *Remote Sens.* **8**(8), 655 (2016).
43. X. Yan et al., "A downscaling-merging scheme for improving daily spatial precipitation estimates based on random forest and cokriging," *Remote Sens.* **13**(11), 2040 (2021).
44. Y. Shao, The Study of Downscaling Method for TRMM Products in China Region, Nanjing University of Information Science and Technology (2015).
45. K. E. Trenberth and D. J. Shea, "Relationships between precipitation and surface temperature," *Geophys. Res. Lett.* **32**(14), L14703 (2005).
46. The Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences, SRTM (Shuttle Radar Topography Mission) DEM, http://www.gscloud.cn (2023).
47. C. Potter, "Differing ecosystem responses of vegetation cover to extreme drought on the Big Sur coast of California," *J. Appl. Remote Sens.* **12**(2), 026031 (2018).
48. K. Didan, "MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006," distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13A3 (2015).
49. Z. Wan, S. Hook, and G. Hulley, "MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006," distributed by NASA EOSDIS Land Processes Distributed Active Archive Center, https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD11A2 (2015).
50. J. P. Kahane and J. Peyriere, "Sur certaines martingales de Benoit Mandelbrot," *Adv. Math.* **22**(2), 131–145 (1976).
51. T. M. Over and V. K. Gupta, "A space-time theory of mesoscale rainfall using random cascades," *J. Geophys. Res.: Atmos.* **101**(D21), 26319–26331 (1996).
52. X. Cui et al., "A study of predicted porosity based on random forest regression algorithm," *West-China Explor. Eng.* **11**, 99–102, 105 (2019).
53. H. Ebrahimy et al., "Downscaling MODIS land surface temperature product using an adaptive random forest regression method and Google Earth Engine for a 19-years spatiotemporal trend analysis over Iran," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 2103–2112 (2021).
54. M. Kühnlein et al., "Precipitation estimates from MSG SEVIRI daytime, nighttime, and twilight data with random forests," *J. Appl. Meteorol. Climatol.* **53**(11), 2457–2480 (2014).
55. H. Guan et al., "Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests," *Int. J. Remote Sens.* **34**(14), 5166–5186 (2013).
56. S. Jia et al., "A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China," *Remote Sens. Environ.* **115**(12), 3069–3079 (2011).
57. Q. Ma et al., "Performance evaluation and correction of precipitation data using the 20-year IMERG and TMPA precipitation products in diverse subregions of China," *Atmos. Res.* **249**, 105304 (2021).
58. B. Kedem and L. S. Chiu. "On the lognormality of rain rate," *Proc. Natl. Acad. Sci. U. S. A.* **84**(4), 901–905 (1987).

Wei Zhang is an assistant professor at the China University of Geosciences (Wuhan). She received her BS degree in geographic information system in 2003 from the China University of Geosciences (Wuhan), and then received her MS and PhD degrees in cartography and geographic information engineering from the same university in 2006 and 2012. She is the author of more than 10 journal papers. Her current research interests include spatial downscaling, spatiotemporal analysis, and prediction of meteorological data.

Biographies of the other authors are not available.