Research Article

Spatial Downscaling of TRMM Precipitation Data Using an Optimal Subset Regression Model with NDVI and Terrain Factors in the Yarlung Zangbo River Basin, China

Jinping Liu,1,2 Wanchang Zhang,1 and Ning Nie3

1Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
2University of Chinese Academy of Sciences, Beijing 100039, China
3Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai 200241, China

Correspondence should be addressed to Wanchang Zhang; zhangwc@radi.ac.cn

Received 16 October 2017; Revised 20 January 2018; Accepted 1 February 2018; Published 1 March 2018

High accuracy, high spatial resolution precipitation data is important for understanding basin-scale hydrology and the spatiotemporal distributions of regional precipitation. The objective of this study was to develop a reliable statistical downscaling algorithm to produce high quality, high spatial resolution precipitation products from Tropical Rainfall Monitoring Mission (TRMM) 3B43 data over the Yarlung Zangbo River Basin using an optimal subset regression (OSR) model combined with multiple topographical factors, the Normalized Difference Vegetation Index (NDVI), and observational data from rain gauge stations. After downscaling, the bias between TRMM 3B43 and rain gauge data decreased considerably from 0.397 to 0.109, the root-mean-square error decreased from 235.16 to 124.60 mm, and the $r^2$ increased from 0.54 to 0.61, indicating significant improvement in the spatial resolution and accuracy of the TRMM 3B43 data. Moreover, the spatial patterns of both precipitation rates of change and their corresponding $p$ value statistics were consistent between the downscaled results and the original TRMM 3B43 during the 2001–2014 period, which verifies that the downscaling method performed well in the Yarlung Zangbo River Basin. Its high performance in downsampling precipitation was also proven by comparing with other models. All of these findings indicate that the proposed approach greatly improved the quality and spatial resolution of TRMM 3B43 rainfall products in the Yarlung Zangbo River Basin, for which rain gauge data is limited. The potential of the post-real-time Integrated Multi-satellite Retreivals for Global Precipitation Measurement (IMERG) downscaled precipitation product was also demonstrated in this study.

1. Introduction

Precipitation is an important meteorological variable, and its spatial and temporal distribution and variations are of great significance for the hydrological cycle, land-atmosphere interactions, and water resource utilization [1–3]. The accurate estimation of precipitation, globally and/or within a specific basin, is a crucial component of related research [4, 5]. Accurate precipitation estimation approaches can be roughly divided into two categories: ground observations at surface weather stations and indirect measurements via surface precipitation radar, remote sensing, and similar techniques [1]. Although surface weather stations can be used to obtain fairly accurate precipitation data with limited spatial coverage [6], logistical expenses and maintenance make it difficult to operate stations on cold plateaus such as the Yarlung Zangbo River Basin, where the high altitude and complex terrain effectively prevent the sole use of station observations to obtain accurate, spatially, and temporally resolved precipitation information. The lack of high-resolution data can be attributed primarily
to the generally significant spatial variations in rainfall. The interpolation of observed precipitation data is influenced by the interpolation method and the number and distribution of stations. Therefore, precipitation data obtained by interpolation are usually imprecise when the number of surface observation stations is small. In contrast, surface weather radar can be used to obtain spatial and temporal precipitation information indirectly; however, due to the limited scope of observation, theoretical complexity, and other factors, surface radar cannot be used to derive high-precision precipitation data sufficient for use in hydrological models or for the description of a given basin [7]. Therefore, the development of innovative approaches for accurate precipitation estimation in areas with no or inadequate data is of crucial importance.

During the rapid development of earth observation technologies in recent decades, microwave remote sensing has been shown to possess unique advantages for obtaining accurate spatial and temporal precipitation information over areas lacking meteorological observations [8, 9]. Currently, the global mainstream satellite-based quantitative precipitation estimate (QPE) products available include TRMM (Tropical Rainfall Monitoring Mission) [10], CMORPH (Climate Prediction Centre Morphing Technique) [11], GSMaP (Global Satellite Mapping of Precipitation) [12, 13], and PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) [14, 15]. As studies have shown, TRMM 3B43 (V7) integrates four independent types of precipitation information derived from microwave, near infrared, and other sensor fusion estimates and rain gauge data from the United States National Oceanic and Atmospheric Administration (NOAA) and the Global Precipitation Climatology Centre (GPCC) [1, 10]. Therefore, this dataset is highly consistent with data from rain gauge stations and Doppler radar observations, and, thus, is used in various applications. The global successor to TRMM, the Global Precipitation Measurement (GPM) mission, was launched on February 28, 2014 and provides next-generation global observations of rain and snow [16, 17]. The Integrated Multi-satellite Retrievals for GPM (IMERG) product has much higher spatial resolution (0.1° × 0.1°) and better ability to monitor light rain and snow than does the TRMM product [17, 18], but the IMERG time series is not long enough for use herein.

Coarse spatial resolution severely restricts the use of such precipitation products in regional scale or watershed scale applications [19]. Pioneering studies have indicated that, as a climatic factor, precipitation is closely related to natural environmental factors such as topography and vegetation growth [20–22] and that high-resolution precipitation data can be produced by statistical models using terrain and vegetative growth parameters derived from remote sensing inversions [3]. A number of relevant studies can be found in the literature. One study improved the spatial resolution of TRMM data from 0.25° × 0.25° to 1 km × 1 km by establishing exponential equations between Normalized Difference Vegetation Index (NDVI) and TRMM precipitation data during 2001–2007 [2]. Jia et al. developed a downsampling approach to improve the spatial resolution of TRMM 3B43 data in the Qaidam Basin using DEM and NDVI data during 1999–2009 [3]. Finally, Fang et al. downscaled TRMM 3B42 (V7) data in the Xiao River Basin in southern China based solely on orographic effects and meteorological conditions [23]. However, the regions used in previous studies featured low terrain complexity; also, while NDVI and elevation control factors were generally adopted, the effects of longitude, latitude, slope, and aspect on precipitation were ignored, resulting in increased deviations in the downscaled results.

The topography of the Yarlung Zangbo River Basin features a narrow band of undulating mountains in which the east-west length far exceeds the north-south width; this topography results in local precipitation of moist warm air from the Bay of Bengal in the Indian Ocean, which is blocked by the Himalayas to the south and drifts upward to the artery estuary of the river. To the best of our knowledge, the Yarlung Zangbo River Basin has been the focus of little satellite-based precipitation downsampling research and no upstream deviation estimates, which hampers the application of satellite-based precipitation data in hydrological research.

In this study, an optimal subset regression (OSR) model with integrated NDVI, elevation, longitude, latitude, slope, aspect, and observation data from rain gauge stations was developed over the Yarlung Zangbo River Basin for the spatial statistical downsampling of the TRMM 3B43 precipitation product between 2001 and 2014; observations and deviation distribution from 17 rain gauge stations within this Basin during the same period were used to evaluate performance of the proposed downsampling approach. Additionally, we compared the annual downscaled precipitation results of TRMM 3B43 and IMERG in 2014 to demonstrate the potential of the IMERG product for downsampling precipitation.

The rest of this paper is organized as follows. Section 2 introduces the study area. The data and methodology used for downsampling precipitation and evaluating the accuracy of the downscaled results are presented in Section 3. Section 4 focuses on the results and a discussion thereof. Finally, a brief summary and study conclusions are given in Section 5.

2. Study Area

The Yarlung Zangbo River, which is the fifth longest river in China, originates from the Chamyingdung glacier at an elevation of 5200 m a.s.l. in south-central Tibet and is approximately 2229 km long [24]. The Yarlung Zangbo River Basin, located at 27°49′N–31°7′N, 82°E–97°6′E, has an area of approximately 257.7 × 10^6 km² (Figure 1) and, with its average elevation of more than 4600 m a.s.l., is considered to be the highest river basin in the world [25]. The Yarlung Zangbo River Basin is a part of Indian Ocean hydrological system. Due to its geographic location in the Yarlung Zangbo Suture Zone, the Yarlung Zangbo River Basin contributes to moisture movement from the Indian Ocean to the inner regions of the Qinghai-Tibet Plateau and has a significant effect on the amount of rainfall in the Qinghai-Tibet Plateau [26]. The YZRB is characterized by complex topography, which results in obvious variations in the amount of precipitation between the upstream and downstream regions of the basin [25]; annual precipitation decreases clearly from over 2000 mm in the lower reaches to less than...
Table 1: Geographic information and multiannual mean precipitation for the 17 meteorological stations within the Yarlung Zangbo River Basin. The multiannual mean precipitation values were calculated from daily meteorological data from 2001 to 2014.

<table>
<thead>
<tr>
<th>ID</th>
<th>Station</th>
<th>Longitude (°)</th>
<th>Latitude (°)</th>
<th>Altitude (m)</th>
<th>Precipitation (mm year⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55493</td>
<td>Damxung</td>
<td>91.1</td>
<td>30.48</td>
<td>4200</td>
<td>497.7</td>
</tr>
<tr>
<td>55569</td>
<td>Lazi</td>
<td>87.6</td>
<td>29.08</td>
<td>4000</td>
<td>336.1</td>
</tr>
<tr>
<td>55572</td>
<td>Namling</td>
<td>89.1</td>
<td>29.68</td>
<td>4000</td>
<td>479.9</td>
</tr>
<tr>
<td>55578</td>
<td>Shigatse</td>
<td>88.88</td>
<td>29.25</td>
<td>3836</td>
<td>428.7</td>
</tr>
<tr>
<td>55585</td>
<td>Nimu</td>
<td>90.17</td>
<td>29.43</td>
<td>3809.4</td>
<td>353.1</td>
</tr>
<tr>
<td>55589</td>
<td>Konka</td>
<td>90.98</td>
<td>29.3</td>
<td>3555.3</td>
<td>415.2</td>
</tr>
<tr>
<td>55591</td>
<td>Lhasa</td>
<td>91.13</td>
<td>29.67</td>
<td>3648.9</td>
<td>477.2</td>
</tr>
<tr>
<td>55593</td>
<td>Maizhokunggar</td>
<td>91.73</td>
<td>29.85</td>
<td>3804.3</td>
<td>579.8</td>
</tr>
<tr>
<td>55597</td>
<td>Qonggyai</td>
<td>91.68</td>
<td>29.03</td>
<td>3741</td>
<td>415.5</td>
</tr>
<tr>
<td>55598</td>
<td>Zedang</td>
<td>91.77</td>
<td>29.25</td>
<td>3551.7</td>
<td>465.9</td>
</tr>
<tr>
<td>55680</td>
<td>Jiangzi</td>
<td>89.6</td>
<td>28.92</td>
<td>4040</td>
<td>336.4</td>
</tr>
<tr>
<td>55681</td>
<td>Nagarze</td>
<td>90.4</td>
<td>28.97</td>
<td>4432.4</td>
<td>388.1</td>
</tr>
<tr>
<td>56202</td>
<td>Jiali</td>
<td>93.28</td>
<td>30.67</td>
<td>4488.8</td>
<td>595.5</td>
</tr>
<tr>
<td>56227</td>
<td>Bomi</td>
<td>95.77</td>
<td>29.87</td>
<td>2736</td>
<td>583.7</td>
</tr>
<tr>
<td>56307</td>
<td>Gyaca</td>
<td>92.58</td>
<td>29.15</td>
<td>8260</td>
<td>571.9</td>
</tr>
<tr>
<td>56312</td>
<td>Linzhi</td>
<td>94.33</td>
<td>29.67</td>
<td>2991.8</td>
<td>689.3</td>
</tr>
<tr>
<td>56317</td>
<td>Milin</td>
<td>94.22</td>
<td>29.22</td>
<td>2950</td>
<td>590.6</td>
</tr>
</tbody>
</table>

300 mm in the upper reaches due to declining monsoon effects [27]. The basin-averaged annual precipitation is ~300–500 mm, and precipitation features large fluctuations in both spatial and temporal distribution [25]. The subpar living environment, difficulty of access, and logistical complexity limit the number of meteorological observation stations in the area and render their distribution uneven, especially in the upstream region. Within this extensive basin, only 17 basic national meteorological stations have been established to date (Figure 1 and Table 1), which has restricted the spatial and temporal description of local climate and hydrology.

3. Data and Methods

3.1. Data Sources and Processing

3.1.1. TRMM 3B43. The TRMM satellite was jointly developed and launched by the National Aeronautics and Space Administration (NASA) of the United States and the National Space Development Agency (NASDA) of Japan in order to understand the global energy and water cycle by studying precipitation and latent heat in the tropics [10]. A number of sophisticated algorithms have been developed since the satellite launch in November, 1997 [28, 29]. Because the TRMM precipitation products are of good quality and also highly consistent with station observations [30], they are often used as reference standards for testing and evaluating other satellite precipitation products. Among TRMM products, TRMM 3B43 algorithm produces the best root-mean-square error (RMSE) for precipitation rate and amount between TRMM and other data sources. The TRMM 3B43 product from January 2001 to December 2014 was selected for the present study; this dataset was archived in Hierarchical Data Format (HDF), had spatial and temporal resolutions of 0.25° and a month, respectively, and covered an extensive area (from 50°N to 50°S). The total precipitation spatial distribution for 2001–2014 was obtained by summing the monthly precipitation derived from TRMM 3B43.

Using only the 17 stations with available precipitation data in the Yarlung Zangbo River Basin for TRMM 3B43 product calibration, Sawunyama and Hughes attempted to calibrate satellite precipitation data via a nonlinear optimization algorithm and achieved good results for use in a hydrological model [31]. The algorithm proposed by Sawunyama and Hughes [31] was used for the TRMM 3B43 product in the
The monthly precipitation values derived from the observation stations were summed up to obtain the total precipitation from 2001 to 2014, with obvious outliers excluded. Equation (1) \((r^2 = 0.54, p < 0.001)\), which was derived between the observation stations and corresponding TRMM 3B43 precipitation grid points (see Figure 2), was established and used to calibrate the TRMM 3B43 data in the Yarlung Zangbo River Basin.

\[
S = 3.437T^{0.756},
\]

where \(T\) (mm) represents TRMM 3B43 annual precipitation and \(S\) (mm) represents annual precipitation observed at meteorological stations between 2001 and 2014.

3.1.2. NDVI Data. The Normalized Difference Vegetation Index (NDVI) is an important indicator used to measure the status of vegetative activity and biomass [32]. NDVI products not only cover a wider range, but also have a higher spatial resolution than do precipitation data. Generally positive correlations have been found between NDVI and precipitation in numerous research projects on the spatial and temporal relationships between NDVI and precipitation both in China and abroad [20, 33]. This study adopted MOD13A2, a Moderate Resolution Image Spectroradiometer (MODIS) NDVI product developed by NASA with a spatial resolution of 1 km × 1 km, a temporal resolution of 1 month, and EOS-HDF format data storage for which the secondary product was developed by calibrating and positioning the primary product. The data selected for the present study, which ranged from January, 2001, to December, 2014, were batch processed into GEO-TIFF format via an IDL code before the commencement of further research.

3.1.3. Topographic Data. Digital elevation model (DEM) data from the Shuttle Radar Topography Mission, Version 4 (SRTM4), were used to characterize the topography of the basin. SRTM is a spacecraft radar topography survey jointly implemented by NASA, the United States National Imagery and Mapping Agency (NIMA), and the German and Italian Space Agencies. SRTM DEM data includes three levels of resolution at 1,000 m, 90 m, and 30 m [34]. To match the spatial resolution of NDVI in the development of a precipitation downscaling approach for the Yarlung Zangbo River Basin, DEM data with a spatial resolution of 1,000 m was used in association with geographic information to calculate topographic features of the basin such as longitude, latitude, slope, aspect, and elevation.

3.1.4. GPM IMERG. IMERG provides three types of products: early run, late run, and final run [16]. Both the early run and late run products are near-real-time, while the final run product is a post-real-time research product and is available beginning from March 12, 2014. Half-hourly and monthly IMERG final run products are available, both of which have a spatial resolution of 0.1° × 0.1°. In this study, the accuracy of the 2014 monthly IMERG final run product was compared with that of TRMM 3B43 in order to demonstrate its potential for use in downscaling precipitation. Please note that the 2014 annual precipitation is in fact the sum of precipitation from March 2014 to February 2015, as the IMERG product is not available before March 2014.

3.2. Methodology

3.2.1. Optimal Subset Regression Model. Optimal Subset Regression (OSR) models can be used to determine an optimal regression equation from all possible subset regressions of the variables according to certain criteria [35]. Regression equations are established through all possible permutations of the independent and dependent variables. In a case with \(n\) variables, \(2n-1\) possible regression equations can be obtained (except in cases with variable-free equations) and an optimal regression subset can thus be determined. This differs from stepwise regression in that the regression equation obtained from optimal subset regression is the best option. In contrast, in stepwise regression, variables are selected or removed based on statistical tests, and the choice of significance level \(\alpha\) is arbitrary. Therefore, it is theoretically difficult to ensure the significance of independent variables selected at any probability. For a more detailed review of OSR models, please refer to Boyce et al. [35].

The Couple Score Criterion (CSC) algorithm [36], which is commonly used in meteorology, was adopted to select a predictor in the present study. The CSC can be used to evaluate the number of independent variables and trend prediction results in terms of climatic prediction characteristics. Therefore, the CSC trend and number ratings are considered in order to improve the model fit and trend prediction. When the CSC score is maximized, the corresponding regression model is optimal. In other words, optimal subset regression is achieved when the CSC is maximized with the fewest variables.

3.2.2. Downscaling Methods. Great effort has been expended to correct the lack of high-resolution precipitation datasets needed to analyze the spatiotemporal variability of precipitation regimes [23] and understand the basin-scale hydrology [2, 37]. Agam et al. proposed a downscaling algorithm...
to sharpen vegetation index-based thermal infrared images [38]. Immerzeel et al. simplified the algorithm by selecting only the most basic functions [2]. The downscaling approach used in this study referenced the simplified algorithm used by Immerzeel et al. [2], which was based on Agam et al. [38], with some revisions. The OSR model was applied to multiple variables to achieve optimal downscaled results consistent with the data derived from observation stations. Specifically, the downscaling consisted of the following five steps. (1) Multiple variables, such as elevation, longitude, latitude, slope, aspect, annual average NDVI, and TRMM 3B43 precipitation data, were retrieved between 2001 and 2014 with a spatial resolution of 0.25° × 0.25° to coordinate with the meteorological observation station data. (2) A regression equation was established for the variables using the OSR algorithm, and the annual total precipitation between 2001 and 2014 was predicted (at 0.25° × 0.25° resolution). (3) Residuals were obtained by subtracting the predicted data from the TRMM 3B43 precipitation data (0.25° × 0.25°), and the spatial resolution was increased to 1 km × 1 km via interpolation. (4) The annual total precipitation (1 km × 1 km) between 2001 and 2014 for the Yarlung Zangbo River Basin was predicted by applying the OSR model to the multiple variables as described in step (1). (5) The downscaled precipitation data (with a spatial resolution of 1 km × 1 km) was then obtained by summing the above-mentioned 1 km × 1 km predicted precipitation data and 1 km × 1 km residuals.

According to previous studies [3], regions with abnormal NDVI values should be examined carefully when establishing regression models. In order to use the NDVI to establish the regression model, a significant correlation must exist between precipitation and NDVI; however, this requirement is not always met, especially in the presence of rivers, lakes, and other water bodies. The Moran Index is an indicator of the NDVI autocorrelation in a given area and can be used to distinguish random pixels. During regression model development, detected random pixels are removed to eliminate their effects, and the annual precipitation at those pixels is obtained through interpolation of the adjacent pixels.

3.2.3. Verification Method. In this study, the downscaled results were verified using the following three methods. (1) The coefficient of determination ($r^2$), bias (B), and RMSE were calculated between the downscaled results and the yearly precipitation data from the 17 observation stations in the Yarlung Zangbo River Basin from 2001 to 2014. The B and RMSE equations can be written as follows:

$$B = \frac{\sum_{i=1}^{n} \hat{y}_i}{\sum_{i=1}^{n} y_i} - 1$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}},$$

where $y_i$ is the annual precipitation observed at the stations, $\hat{y}_i$ is the precipitation extracted from the downscaled results, $i$ is the code number of the station, and $n$ is the number of stations within the basin. (2) Although the precision of the downscaled results at the observation stations (which are mostly located in the middle and lower reaches of the River Basin) were verified by method (1), it is impossible to guarantee their accuracy, especially for the upper reaches, as only data from observation stations in the middle and lower reaches were used to generate the downscaled results. In verification step (2), the restrictions imposed by observation stations were considered, and the deviation distribution was used to analyze and evaluate the significance of the downscaled results throughout the basin. (3) Finally, the effectiveness of the downscaling OSR model herein was compared with other downscaling algorithms, such as Multiple Linear Regression (MLR), Artificial Neural Network (ANN) [39], and Geographically Weighted Regression (GWR) models. Two approaches were used in the ANN comparison, ANN_D and ANN_S; in the ANN_D approach, the downscaled precipitation results were obtained directly through the trained neural network, while the ANN_S approach was similar to the downscaling method discussed in Section 3.2.2, except that steps (1) and (4) were replaced with ANN in the ANN_S approach. Additionally, when using the GWR model, variable collinearity was checked beforehand using Ordinary Least Squares (OLS) regressions to prevent errors.

3.2.4. Statistical Indicators for Spatial Precipitation Trends. In order to examine spatial precipitation trend patterns in the Yarlung Zangbo River Basin, a linear fitting (least squares) method was used to calculate trends in precipitation; $t$-tests were used to assess the statistical significance ($p$) of the trends in this study. $p$ values less than 0.05 indicate that the corresponding pixel is statistically significant.

4. Results and Discussion

4.1. Excluding Random Pixels. Figure 3 shows the local NDVI Moran Index (exampled by the average NDVI from 2001 to 2014) in the Yarlung Zangbo River Basin, which reflects the spatial autocorrelation of the NDVI distribution in the Basin. According to the definition of the local Moran Index, pixels with negative Moran Index values imply that the corresponding NDVI values were caused by random factors, such as the presence of rivers, lakes, or other water bodies, and therefore should be excluded during the establishment of the optimal regression model. As can be seen in Figure 3,
the random pixels are concentrated in the middle and lower reaches of the Yarlung Zangbo River Basin, where water bodies are densely distributed and water resources utilization is relatively intense, which is consistent with the definition of the Moran Index. Therefore, pixels with negative Moran Index values were removed during the development of the optimal subset regression model, and the final downscaled results for these pixels were obtained via Kriging interpolation by ArcGIS software.

4.2. Spatial Downscaling Process. Table 2 shows all of the possible optimal subsets and the corresponding multiple correlation coefficient (R) and CSC values for the multiple regression analyses of TRMM 3B43 data with elevation, longitude, latitude, slope, aspect, and NDVI in the Yarlung Zangbo River Basin. Table 2 shows that the CSC value of the optimal subset regression model, which consists of five independent variables (namely, elevation, longitude, slope, aspect, and NDVI) as shown in (3), reaches a maximum of 294.62.

\[
y = -0.1x_1 + 17.395x_2 + 160.182x_4 - 8.133x_5 + 545.688x_6 - 725.689
\]

(3)

where \( y \) is the predicted annual precipitation; \( x_1, x_2, x_4, x_5, \) and \( x_6 \) are the elevation, longitude, slope, aspect, and NDVI of the study watershed, respectively, and \( x_3 \) is the excluded variable (latitude).

The factors that affect precipitation in the Yarlung Zangbo River Basin can be ranked according to the OSR model analysis in order of decreasing influence as follows: NDVI, slope, longitude, aspect, and elevation (where the latitude variable was excluded). The NDVI is highly correlated with precipitation [2, 3]. The contributions of slope, longitude, aspect, and elevation to precipitation are markedly different. Latitude has a negligible influence on precipitation.

In the Yarlung Zangbo River Basin, precipitation is a function of longitude but not latitude, for which there are two key physical explanations. First, the basin is very narrow in the north-south direction, with a maximum width of less than 300 km which is far less than the basin length in the west-east direction. Second, approximately 77.9% of the annual precipitation falls during the summer (from June to September); this precipitation arises primarily from moisture from the Bay of Bengal transported by the Indian monsoon and moisture transported by the Western Pacific subtropical high, which converge and are then transported from east to west through the Yarlung Zangbo Suture Zone. Therefore, longitude has a higher correlation with precipitation than does latitude in the Yarlung Zangbo River Basin.

The OSR model was used to determine an optimal regression equation from all possible subsets of multiple independent variables with specific criterion [35]; the CSC adopted in this study was chosen to minimize the number of independent variables and maximize the CSC score to achieve the optimal predicted value [36]. To verify the efficiency of the OSR approach and CSC used in this study, RMSEs were calculated between the predicted precipitation at low spatial resolution (0.25° × 0.25°) and six optimal TRMM 3B43 subsets (Table 2). Clearly, the RMSE is minimized among the 6 optimal regression subsets when the number of independent variables is equal to 5, which demonstrates the advantages of using OSR and the CSC in the TRMM 3B43 downscaling algorithm presented in this study; such advantages will be more pronounced if more independent variables are involved in the downsampling procedure.

Figure 4 shows a scatter diagram of annual mean precipitation derived from calibrated TRMM 3B43 data against annual mean precipitation predicted by OSR during the 2001–2014 period at coarse resolution (0.25° × 0.25°). The \( r^2 \) value is 0.79 and the relationship is significant (\( p < 0.001 \)), indicating that elevation, longitude, slope, aspect, and NDVI can be used in the interpretation of TRMM 3B43 precipitation data.

Figure 5 illustrates the integrated downscaling process used for TRMM 3B43 annual mean precipitation between 2001 and 2014. Figure 5(a) shows the spatial distribution of TRMM 3B43 annual precipitation, and Figure 5(b) shows the predicted annual precipitation with a spatial resolution of 0.25° × 0.25°. The residual data, which feature a resolution of 0.25° × 0.25°, were obtained by subtracting the predicted annual precipitation from the TRMM 3B43 annual precipitation and are shown in Figure 5(c). Negative residual

<table>
<thead>
<tr>
<th>Number of variables</th>
<th>Optimal subsets</th>
<th>Multiple correlation coefficient (R)</th>
<th>RMSE (mm)</th>
<th>CSC results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( x_6 )</td>
<td>0.75</td>
<td>154.81</td>
<td>266.68</td>
</tr>
<tr>
<td>2</td>
<td>( x_1, x_2 )</td>
<td>0.83</td>
<td>144.59</td>
<td>284.81</td>
</tr>
<tr>
<td>3</td>
<td>( x_1, x_4, x_5 )</td>
<td>0.88</td>
<td>138.06</td>
<td>290.07</td>
</tr>
<tr>
<td>4</td>
<td>( x_1, x_2, x_4, x_6 )</td>
<td>0.89</td>
<td>134.83</td>
<td>293.33</td>
</tr>
<tr>
<td>5</td>
<td>( x_1, x_2, x_4, x_5, x_6 )</td>
<td>0.89</td>
<td>133.74</td>
<td>294.62</td>
</tr>
<tr>
<td>6</td>
<td>( x_1, x_2, x_4, x_5, x_6 )</td>
<td>0.89</td>
<td>133.93</td>
<td>293.84</td>
</tr>
</tbody>
</table>

Table 2: All possible optimal regression subsets.
values indicate overestimated annual precipitation, which results from "abnormally" high vegetative NDVI values due to thawing ice or snow rather than precipitation; positive residual values suggest an underestimation of annual precipitation and likely result from "abnormally" low precipitation caused by steep terrain and sparse vegetation. Low-resolution residual data with were transformed into a higher resolution of 1 km × 1 km via spline interpolation using ArcGIS software (Figure 5(d)). The predicted annual precipitation (spatial resolution = 1 km × 1 km) was obtained by applying the OSR
model to elevation, longitude, slope, aspect, and NDVI data at their original spatial resolution (Figure 5(e)). Yarlung Zangbo River Basin annual precipitation with a spatial resolution of 1 km × 1 km is shown in Figure 5(f) and was derived by adding the residual data (shown in Figure 5(d)), which arose in turn from a Kriging interpolation algorithm applied to the random pixels obtained from Moran Index, to the 1 km × 1 km predicted annual precipitation.

4.3. Verification of the Downscaled Results. In this study, the downscaled annual precipitation results were verified using three methods: First, the downscaled results at the location of the 17 observation stations from 2001 to 2014 were extracted (yielding 238 samples in total) and a scatter diagram of the observed annual precipitation (\( P_{\text{obs}} \)) and the downscaled annual precipitation (\( P_{\text{ds}} \)) was produced (Figure 6). The calculated \( r^2 \) value between \( P_{\text{obs}} \) and \( P_{\text{ds}} \) from 2001 to 2014 is 0.61 (\( p < 0.001 \)); the bias \( B \) is 0.109; and the RMSE is 124.60 mm. The \( r^2 \) (0.54), \( B \) (0.397), and RMSE (235.16 mm) values for the actual annual precipitation against the TRMM 3B43 data before downscaling were also calculated for 2001–2014. The results above show that the annual precipitation \( r^2 \) value is greatly improved after downscaling, while \( B \) and RMSE are reduced somewhat; in other words, the downscaled annual precipitation is more accurate than the original TRMM 3B43 annual precipitation. Therefore, the use of the OSR model and elevation, longitude, slope, aspect, and MODIS NDVI data in spatial TRMM 3B43 annual precipitation downscaling is highly efficient.

Although the observation stations were used to validate the accuracy of the 2001–2014 downscaled results in the first step, this does not address the reliability of the downscaled results for the entire basin, especially for upstream areas where no observation stations exist. In this study, deviation distributions were used to evaluate the downscaled results for the entire basin under the basic assumptions that the normalized standard deviation, \( (P_{\text{ds}} - P_{\text{obs}})/P_{\text{obs}} \), of all pixels within the basin are normally distributed and that the errors of the observed data accurately represent the entire basin. The Kolmogorov-Smirnov Method was adopted to test the normality of the normalized standard deviation distribution (238 samples) for the 17 stations between 2001 and 2014; the normalized standard deviations of the 238 samples are normally distributed, and the mean and standard deviations, \( \mu \) and \( \sigma \), are 0.1593 and 0.2553, respectively. According to the characteristics of the normal distribution curve, the probability of samples falling within the \( (\mu-\sigma) \) to \( (\mu+\sigma) \) range is 68.26%. In the present study, there is a 68.26% probability that the relative deviation for the downscaled result at a given pixel falls between −0.096 and 0.4146. For example, the average downscaled annual precipitation between 2001 and 2014 is 551.2 mm, and the probability of the deviation falling within the −52.92 mm to 228.53 mm range is 68.26%. The deviations and significance of the downscaled results for the entire Yarlung Zangbo River Basin can be evaluated based on the above analysis.

Table 3 compares results from different downscaling models, including OSR, MLR, ANN_D, ANN_S, and GWR. These results indicate that, among these five models, the OSR model performs best at downsampling precipitation data over the Yarlung Zangbo River Basin. In contrast, the ANN_D model downscaled results are the least accurate, which may be explained by terrain complexity and climate diversity, as well as the limited number of observations. The GWR model introduced numerous errors into the final downscaled results when the regression coefficients were interpolated to high spatial resolution data to obtain high spatial resolution downscaled precipitation, leading to its relative low performance in comparison to OSR.

4.4. Precipitation Trends over the Yarlung Zangbo River Basin. The downscaled results were verified using observed precipitation data, which has shown that the downscaled results were suitable for practical applications, such as spatiotemporal precipitation variability analysis and basin-scale hydrology. Figure 7 illustrates the spatial precipitation rates of change and their corresponding \( p \) value statistics derived from the downscaled results and TRMM 3B43 products during the 2001–2014 period. Figures 7(a) and 7(c) show spatial distributions very similar to those in Figures 7(b) and 7(d), respectively. There are large errors in the TRMM precipitation
products at the site scale; however, preexisting studies have shown that TRMM products provide accurate precipitation spatial patterns [1, 22, 23], which implies that the downscaled results produced in this study are accurate, even in regions without observed precipitation data.

According to the precipitation trends and corresponding $p$ value statistics derived from the downscaled results (Figures 7(a) and 7(c), resp.), precipitation increases in approximately half (53%) of the pixels in the Yarlung Zangbo River Basin, and 13.4% of the pixels show significant increases, mostly in the western and southern regions of the basin. The average rate of increase in precipitation over the entire Yarlung Zangbo River Basin was 7.7 mm year$^{-1}$. Precipitation decreased in 47% of the pixels; however, only 5.2% of the pixels show significant decreases, most of which are concentrated in small regions in the westernmost and eastern parts of the basin. The average rate of decrease in precipitation over the entire basin was relatively low at $-5.5$ mm year$^{-1}$. In total, the average precipitation rate of change over the Yarlung Zangbo River Basin was $1.6$ mm year$^{-1}$, which indicates that precipitation in the basin increased overall during the 2001–2014 period.

4.5. Uncertainties in the Downscaled Results. The spatial resolution and fitting $r^2$ were greatly improved by downscaling; however, certain errors remain, as indicated by $B$ and RMSE ($0.109$ and $124.60$ mm, resp.) in the downscaled precipitation result from the 2001–2014 period in the Yarlung Zangbo River Basin. The bias, $B$, indicates an overestimation of rainfall over the basin; this overestimation may be caused partially by TRMM product errors [40, 41], which arise from mixed water-body and land pixels within the passive microwave radiometric field of view and poor performance regarding the differences in emissivity and brightness temperature of water surfaces in the passive microwave frequencies used by the retrieval algorithms [1, 9]. The error propagation process during downscaling can be detailed as follows. (1) Overall, the annual precipitation derived from original TRMM products is overestimated as compared to observed precipitation during the 2001–2014 period. (2) Overestimation and underestimation are more likely to occur in the relatively wet and relative dry regions, respectively, after the data are calibrated with observed precipitation data. (3) An overall overestimation exists in the predicted high spatial resolution data due to meteorological station elevations below the average elevation, which results in (4) the underestimation of residual data in most regions (see Figure 5(d)) and residual data absolute errors less than those for predicted data. (5) Thus, the final downscaled precipitation results feature overall overestimation.

Rain gauges, which have a sparse and uneven spatial distribution in the basin, may also be partially responsible for this overestimation [42]. Because they are sensitive to sensor and retrieval algorithm accuracy, precipitation products derived from TRMM 3B43 are relatively inconsistent with observations. Therefore, original products from TRMM 3B43 products at the site scale; however, preexisting studies have shown that TRMM products provide accurate precipitation spatial patterns [1, 22, 23], which implies that the downscaled results produced in this study are accurate, even in regions without observed precipitation data.

According to the precipitation trends and corresponding $p$ value statistics derived from the downscaled results (Figures 7(a) and 7(c), resp.), precipitation increases in approximately half (53%) of the pixels in the Yarlung Zangbo River Basin, and 13.4% of the pixels show significant increases, mostly in the western and southern regions of the basin. The average rate of increase in precipitation over the entire Yarlung Zangbo River Basin was 7.7 mm year$^{-1}$. Precipitation decreased in 47% of the pixels; however, only 5.2% of the pixels show significant decreases, most of which are concentrated in small regions in the westernmost and eastern parts of the basin. The average rate of decrease in precipitation over the entire basin was relatively low at $-5.5$ mm year$^{-1}$. In total, the average precipitation rate of change over the Yarlung Zangbo River Basin was $1.6$ mm year$^{-1}$, which indicates that precipitation in the basin increased overall during the 2001–2014 period.

4.5. Uncertainties in the Downscaled Results. The spatial resolution and fitting $r^2$ were greatly improved by downscaling; however, certain errors remain, as indicated by $B$ and RMSE ($0.109$ and $124.60$ mm, resp.) in the downscaled precipitation result from the 2001–2014 period in the Yarlung Zangbo River Basin. The bias, $B$, indicates an overestimation of rainfall over the basin; this overestimation may be caused partially by TRMM product errors [40, 41], which arise from mixed water-body and land pixels within the passive microwave radiometric field of view and poor performance regarding the differences in emissivity and brightness temperature of water surfaces in the passive microwave frequencies used by the retrieval algorithms [1, 9]. The error propagation process during downscaling can be detailed as follows. (1) Overall, the annual precipitation derived from original TRMM products is overestimated as compared to observed precipitation during the 2001–2014 period. (2) Overestimation and underestimation are more likely to occur in the relatively wet and relative dry regions, respectively, after the data are calibrated with observed precipitation data. (3) An overall overestimation exists in the predicted high spatial resolution data due to meteorological station elevations below the average elevation, which results in (4) the underestimation of residual data in most regions (see Figure 5(d)) and residual data absolute errors less than those for predicted data. (5) Thus, the final downscaled precipitation results feature overall overestimation.

Rain gauges, which have a sparse and uneven spatial distribution in the basin, may also be partially responsible for this overestimation [42]. Because they are sensitive to sensor and retrieval algorithm accuracy, precipitation products derived from TRMM 3B43 are relatively inconsistent with observations. Therefore, original products from TRMM 3B43

![Figure 7: Changes in annual precipitation derived from (a) downscaled results (DS-P) and (b) TRMM 3B43 and the corresponding $p$ value statistics for the changes in (c) DS-P and (d) TRMM-P in the Yarlung Zangbo River Basin during the 2001–2014 period.]
must be calibrated with local rain gauge station observations before they are applied to a specific study area. However, with only 17 rain gauge stations in the Yarlung Zangbo River Basin (all of which are located in the middle and lower reaches of the basin), gauge measurements are scarce, which greatly impact the effective calibration of TRMM 3B43 data. This lack of observations thus affects the effectiveness of the downscaling algorithm.

In this study, the downscaling model was established upon the basic assumption that precipitation amount has a strong relationship with NDVI, topographical factors, and geographical location in the Yarlung Zangbo River Basin. Numerous studies have found positive correlations between precipitation and NDVI [2, 43, 44] and between precipitation and topographical factors (i.e., elevation) [3]. In addition, geographical location (i.e., longitude) must be considered, as it is an important factor due to the long and narrow terrain in the Yarlung Zangbo River Basin. Although rainfall has a strong relationship with NDVI, there is a time lag of 1–3 months between rainfall and the response NDVI in monthly data [45–47]; it is well-known that NDVI may become saturated when the precipitation exceeds a certain value in humid regions, which may impact the downscaled results. In the Yarlung Zangbo River Basin specifically, hydrological conditions have particularly obvious impacts on NDVI in the midstream and downstream regions, where water bodies such as rivers, lakes, and groundwater runoff disrupt the relationship between precipitation and NDVI [3].

Precipitation, which is an extremely complex meteorological phenomenon, is influenced by many factors other than NDVI and topography (i.e., monsoon and temperature), as evidenced by the residual errors in Figure 5(c). Theoretically, the dominant factors controlling precipitation vary in different locations within the Yarlung Zangbo River Basin; thus, deviations remain in the downscaled results despite the inclusion of multiple factors and the removal of nonsignificant factor(s) in this study.

The errors in the downscaled results arise partially because of the sparse and uneven rain gauge spatial distribution in the Yarlung Zangbo River Basin; however, the correct spatial pattern provided by TRMM 3B43 guarantees the rationality of the results. Moreover, the accuracy is improved greatly by downscaling. Thus, this study shows that the downscaling model presented herein can be used to produce high spatial resolution precipitation data in the Yarlung Zangbo River Basin.

4.6. The Potential of IMERG for Downscaling Precipitation. Figure 8(a) illustrates the accuracy of TRMM 3B43 and IMERG products before downscaling and indicates that IMERG is closer to observations than is TRMM 3B43 in the Yarlung Zangbo River Basin; TRMM 3B43 overestimates precipitation amount to a greater degree than does IMERG. Both TRMM-based and IMERG-based downscaled precipitation are closer to observations after downscaling than before downscaling (Figure 8(b)).

As shown in Table 4, IMERG has lower B and RMSE values than does TRMM 3B43 before downscaling, which indicates that the IMERG product is more accurate than the TRMM 3B43 product. However, the IMERG $r^2$ is lower than that for TRMM 3B43, which may lead to a less accurate calibration using observed precipitation. After downscaling, the lower B and RMSE and higher $r^2$ values indicate great improvements in precipitation accuracy in both the TRMM 3B43 and IMERG data along with the higher spatial resolution. It is worth noting that the TRMM-based downscaled precipitation is more accurate than the IMERG-based results; one possible explanation for this result is that the poor linear, or nonlinear, correlation between IMERG data and observed precipitation leads to a poor IMERG calibration, which suggests that the downscaling algorithm presented herein is more suitable for TRMM 3B43. Therefore, a new downscaling algorithm must be developed for IMERG.

### Table 4: Statistical results for comparisons of TRMM 3B43 and IMERG with observed precipitation before and after downscaling for 2014.

<table>
<thead>
<tr>
<th></th>
<th>Before downscaling</th>
<th>After downscaling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRMM 3B43</td>
<td>IMERG</td>
</tr>
<tr>
<td>B</td>
<td>0.508</td>
<td>0.223</td>
</tr>
<tr>
<td>RMSE (mm)</td>
<td>283.94</td>
<td>171.58</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.54</td>
<td>0.31</td>
</tr>
</tbody>
</table>

5. Conclusions

The precipitation downscaling process performed in this study yielded a number of primary findings. Using a nonlinear optimization method, annual precipitation from TRMM 3B43 was effectively calibrated with observations from 17 meteorological stations within the Yarlung Zangbo River Basin taken from 2001 to 2014. After downscaling, both the spatial resolution and the accuracy of the precipitation data were greatly improved; the bias between TRMM 3B43 and rain gauge data decreased considerably from 0.397 to 0.109, the RMSE decreased from 235.16 to 124.60 mm, and $r^2$ increased from 0.54 to 0.61. In addition, a deviation distribution method was used to derive deviation estimates for the downscaled precipitation in regions without observation stations; the results show a 68.26% probability that the normalized standard deviation for the downscaled result at a given pixel falls between $-0.096$ and 0.4146. Therefore, the final downscaled results are statistically significant for the entire Yarlung Zangbo River Basin.

The precipitation rates of change and their corresponding $p$ value statistics based on the downscaled results were consistent with those based on TRMM 3B43 data, which suggests that the proposed downscaling approach performed quite well in the Yarlung Zangbo River Basin. Its high performance in downscaling precipitation was proven again by comparing with MLR, ANN, and GWR models over the Yarlung Zangbo River Basin. Precipitation over the study area featured a dominant increasing trend at an average rate of 1.6 mm year$^{-1}$ during the period from 2001 to 2014; in individual pixels, increases in precipitation were more common, and greater in magnitude, than were decreases in precipitation.
IMERG was more accurate than TRMM 3B43 when compared with observed precipitation, but had a low correlation with observed precipitation, resulting in the introduction of numerous errors during calibration. Therefore, the accuracy of the IMERG-based downscaled precipitation was lower than that of the TRMM-based downscaled result, which suggests that the downsampling method presented herein is more suitable for TRMM 3B43 than for IMERG.

Overall, errors do likely exist in the downscaled precipitation data produced in this study. However, the spatial downsampling algorithm presented herein can provide high spatial resolution, high accuracy precipitation datasets in the Yarlung Zangbo River Basin using a number of relevant factors, even in regions with extremely sparse meteorological stations. These downscaled datasets offer an opportunity to better understand spatial variations in precipitation in the Yarlung Zangbo River Basin and provide foundational data for the development of monthly and daily precipitation datasets that can be used to drive hydrological models.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors’ Contributions

The first two authors (Jinping Liu and Wanchang Zhang) contributed equally to this paper.

Acknowledgments

This study was supported by the National Key R&D Program of China under Grants 2016YFA0602302 and 2016YFB0502502. The authors would like to thank Professor Shaofeng Jia from Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, for his helpful suggestions.

References


