Comparative Analysis of Downscaling Methods for TRMM 3B43 Precipitation Data in the Qinghai Lake Basin and Surrounding Areas: Postprint

Authors: Li Yankun

Date: 2023-01-17T00:00:00+00:00

Abstract

Utilizing the Multiple Linear Regression model (MLR), Principal Component Stepwise Regression analysis model (PCSR), and Kriging interpolation method, TRMM 3B43 precipitation data at 0.25° resolution for the Qinghai Lake basin and surrounding regions were downscaled to 0.01° resolution. Observed precipitation data from 20 meteorological stations within the study area were employed to evaluate the downscaling results using correlation coefficient (CC), root mean square error (RMSE), and relative bias (Bias), thereby selecting the most appropriate downscaling method for the study area. The results indicate: (1) The spatial distribution of precipitation in the study area derived from TRMM and the three downscaling methods demonstrates consistency; annual average precipitation and spring, summer, and autumn precipitation all exhibit higher values in the north and lower values in the west and northwest, while winter precipitation shows higher values in the south and northwest and lower values in the central region. (2) Precipitation in the study area, with 3800 m as a threshold, generally exhibits a trend of initially increasing and subsequently decreasing with increasing elevation. (3) Accuracy evaluation results demonstrate that Kriging achieves the best performance at the annual scale; spatially, TRMM and the three downscaled datasets exhibit optimal accuracy in the eastern region. At the seasonal scale, data accuracy follows the order PCSR > Kriging > TRMM > MLR; at the monthly scale, PCSR data demonstrate the highest accuracy. (4) Elevation exerts a minor influence on TRMM and the three downscaled datasets within the study area; however, as elevation increases, remote sensing data gradually exhibit precipitation underestimation, possibly attributable to the underestimation of convective precipitation during microwave precipitation rate retrieval. Based on comprehensive analysis of precipitation spatial distribution consistency and accuracy evaluation, PCSR is identified as the most suitable downscaling method for TRMM 3B43 precipitation data in the Qinghai Lake basin and surrounding regions.

ChinaRxiv

Full Text

Comparative Analysis of Downscaling Methods for TRMM 3B43 Precipitation Data in the Qinghai Lake Basin and Surrounding Areas

LI Yankun[1,2], GAO Liming[1,2,3], ZHANG Lele[1,2,3], WU Xueqing[1,2], LIU Xuanchen[1,2], QI Wen[1,2]

[1]College of Geography Science, Qinghai Normal University, Xining 810008, Qinghai, China
[2]Key Laboratory of Tibetan Plateau Land Surface Processes and Ecological Conservation (Ministry of Education), Qinghai Normal University, Xining 810008, Qinghai, China
[3]Academy of Plateau Science and Sustainability, People' s Government of Qinghai Province and Beijing Normal University, Xining 810008, Qinghai, China

Abstract

This study employs multiple linear regression (MLR), principal component stepwise regression (PCSR), and Kriging interpolation to downscale TRMM 3B43 precipitation data from 0.25° to 0.01° resolution in the Qinghai Lake Basin and surrounding areas. Using measured precipitation data from 20 meteorological stations, the downscaled results are evaluated using correlation coefficient (CC), root mean square error (RMSE), and relative deviation (Bias) to identify the most suitable downscaling method for the study area. The results demonstrate that all four datasets (TRMM and the three downscaled products) show consistent spatial precipitation patterns: annual average precipitation and spring, summer, and autumn precipitation are higher in the north and lower in the west and northwest, while winter precipitation is higher in the south and northwest and lower in the central region. Precipitation increases with elevation up to 3800 m, then decreases at higher altitudes. Accuracy evaluation reveals that Kriging performs best at the annual scale, with all datasets showing optimal accuracy in the eastern region spatially. At the seasonal scale, precision follows the order PCSR > Kriging > TRMM > MLR. At the monthly scale, PCSR demonstrates the highest accuracy. While elevation has minimal impact on the four datasets overall, TRMM and its downscaled products increasingly underestimate precipitation at higher elevations, likely due to underestimation of convective precipitation in microwave retrieval algorithms. Based on comprehensive analysis of spatial distribution consistency and accuracy evaluation, PCSR is identified as the most suitable downscaling method for TRMM 3B43 precipitation data in the Qinghai Lake Basin and surrounding areas.

Keywords: TRMM 3B43; downscaling; multiple linear regression; principal component stepwise regression; Kriging; Qinghai Lake Basin

Introduction

Precipitation is a critical parameter in climate systems and a vital component of the water cycle. Accurate precipitation data are essential for understanding regional and global climate change and hydrological processes. Currently, precipitation data are obtained through three primary methods: ground meteorological station observations, weather radar, and satellite remote sensing. Among these, ground observations provide the most accurate measurements, and a dense network of rain gauges can theoretically capture accurate spatial precipitation distribution characteristics. However, in high-altitude mountainous watersheds, observation stations are sparse and precipitation exhibits high spatial heterogeneity, making accurate characterization of precipitation patterns challenging. Weather radar estimates precipitation through echo intensity, reducing errors associated with point-based station measurements, but its application is constrained by numerous limiting conditions and shows poor applicability. Satellite remote sensing offers broad coverage, high spatiotemporal resolution, and fewer constraints, effectively compensating for deficiencies in ground observations. With advancing remote sensing technology, satellite precipitation products have been increasingly applied in meteorological and hydrological research.

Nevertheless, existing remote sensing precipitation products generally suffer from coarse resolution, making them unsuitable for small watershed studies in mountainous areas. Consequently, downscaling research for remote sensing precipitation products has become a research hotspot. Previous studies have employed various downscaling approaches. For instance, some researchers have used Google Earth Engine and Google Cloud Computing to analyze monthly precipitation downscaling, while others have constructed principal component stepwise regression models using TRMM data in the Tianshan Mountains, demonstrating significantly improved accuracy compared to original TRMM data. Additional studies have applied downscaling to GPM IMERG data for spatiotemporal precipitation analysis and drought monitoring in central China, and others have combined multivariate geographically weighted regression with Kriging (MGWRK) for spatial statistical downscaling of TRMM precipitation over the Tibetan Plateau. Random forest algorithms have also been applied for downscaling TRMM data in the Tibetan Plateau region. Comparative assessments of different downscaling methods have been conducted in various watersheds, including the Lake Urmia Basin. Despite these efforts, no universal downscaling standard has emerged.

The Qinghai Lake Basin, located in the northeastern Tibetan Plateau, represents an important ecological barrier. The basin contains only a few meteorological stations with publicly available national meteorological data, significantly limiting hydrological research in the area. Previous studies have evaluated TRMM 3B42V7, PERSIANN, and CMORPH remote sensing precipitation datasets in the Qinghai Lake Basin and surrounding areas, finding PERSIANN to be the most accurate. However, the coarse spatial resolution of TRMM 3B43

(0.25°$×$0.25°) limits its application in small watershed hydrological process studies. Therefore, obtaining higher-resolution precipitation data through downscaling is necessary. Due to the scarcity of meteorological stations within the Qinghai Lake Basin, this study expands the research area to include surrounding regions with 20 meteorological stations. Using MLR, PCSR, and Kriging methods, TRMM 3B43 data are downscaled and evaluated against measured data at annual, seasonal, and monthly scales. The primary objective is to identify the most suitable downscaling method for the Qinghai Lake Basin, providing data support for high-resolution ecological and hydrological process modeling.

1.1 Study Area Overview

The Qinghai Lake Basin is located between 97°50 -101°20 E and 36°15 -38°20 N, covering a watershed area of approximately 29,661 km^2 with a lake area of about 4,400 km^2. The basin is bounded by the Qinghai Nanshan Mountains to the south, Amunikushan to the west, Riyue Mountain to the east, and Datong Mountain to the north, forming a complete closed inland plateau basin. The watershed extends in a northwest-southeast direction, with terrain higher in the northwest and lower in the southeast. Due to the scarcity of meteorological stations within the Qinghai Lake Basin, the study area was expanded to include surrounding regions with 20 meteorological stations, with geographic coordinates ranging from 35°–39°N and 97°–103°E. The study area is characterized by plateau topography, with elevations ranging from 1,330 to 5,663 m, higher in the west and lower in the east. Precipitation exhibits uneven seasonal distribution, with summer precipitation dominant and concentrated, followed by spring and autumn, and minimal winter precipitation.

1.2 Data Sources and Preprocessing

TRMM 3B43 precipitation data were obtained from NASA Earthdata (https://search.earthdata.nasa.gov) with a spatial resolution of 0.25°$×$0.25° and monthly temporal resolution, covering the period from 1998 to 2019. The original data are in NetCDF format and were processed using ArcGIS for format conversion and projection transformation. The original precipitation units were mm/hr, which were converted to monthly precipitation by multiplying by the number of hours in each month.

SRTM DEM data at 90 m resolution were obtained from the Geospatial Data Cloud (http://www.gscloud.cn/search). Slope and aspect data were derived from this DEM using ArcGIS hydrological analysis tools. The DEM was resampled to 0.25° and 0.01° resolutions for the study area. MOD13A3 NDVI data were obtained from NASA Earthdata (https://search.earthdata.nasa.gov) with a spatial resolution of 1 km and monthly temporal resolution for the period 2000–2019. These data were mosaicked, processed for outliers, converted in format, and projected and clipped for the study area using ArcGIS.

Meteorological station precipitation data were obtained from the China Meteorological Data Network (http://data.cma.cn). Twenty meteorological stations are located within the study area: Tuole, Yeniugou, Qilian, Delingha, Gangcha, Menyuan, Wulan, Dulan, Gonghe, Xining, Guide, Minhe, Xinghai, Guinan, Tongren, Chaka, Wushaoling, Wuwei, Yongchang, and Shandan [Figure 1: see original paper]. The observation period matches that of TRMM data (2000–2019), with daily resolution data aggregated to monthly, seasonal, and annual scales.

1.3 Research Methods

Three downscaling methods were employed: multiple linear regression (MLR), principal component stepwise regression (PCSR), and ordinary Kriging interpolation.

The MLR method uses multiple influencing factors as independent variables to explain variations in the dependent variable (precipitation). Precipitation formation is influenced by various factors, including longitude, latitude, elevation, slope, aspect, and normalized difference vegetation index (NDVI). The general regression equation is:

$$y = a_0 + a_{1x}1 + a_{2x}2 + \cdots + a_{nx}n + k$$

where y represents TRMM precipitation data, a_i are regression coefficients, x_i are regression factors, and k is the precipitation residual.

The PCSR method transforms multiple influencing factors into several independent components for analysis. This approach maximizes information retention while eliminating non-significant variables, thereby improving efficiency and clarity. The model is constructed as follows:

$$Y = b_0 + b_{1X}1 + b_{2X}2 + \cdots + b_{nX}n + (b_{n+1} + \cdots) + K$$

where Y represents downscaled precipitation data, b_i are principal component regression coefficients, X_i are principal component regression factors, and K is the precipitation residual.

The downscaling procedure involves: (1) extracting grid center pixel values from TRMM data, longitude/latitude, elevation, slope, aspect, and NDVI at 0.25° resolution using Matlab; (2) calculating regression coefficients and precipitation residuals; (3) interpolating 0.25° residuals to 0.01° resolution using ordinary Kriging; and (4) applying the regression coefficients to high-resolution factors and adding the interpolated residuals to obtain final 0.01° downscaled precipitation data.

1.4 Accuracy Evaluation

Using ArcGIS extraction tools, downscaled annual, seasonal, and monthly precipitation values were extracted for the 20 meteorological stations and compared with observed data using three metrics: correlation coefficient (CC), root mean square error (RMSE), and relative deviation (Bias). CC ranges from -1 to 1, with values closer to 1 indicating stronger consistency between gridded and observed precipitation. RMSE ranges from 0 to $+\infty$, with values closer to 0 indicating higher accuracy. Bias ranges from $-\infty$ to $+\infty$, with values closer to 0 indicating smaller deviation from observed precipitation.

Results

2.1 Spatial Distribution Characteristics of Precipitation in the Qinghai Lake Basin

Following downscaling and temporal processing, the spatial distribution of annual average precipitation in the Qinghai Lake Basin was obtained [Figure 2: see original paper]. Results show consistent spatial patterns between downscaled annual precipitation and original TRMM data. High precipitation areas are primarily located north of Qinghai Lake, while low precipitation areas concentrate in the west and northwest. In terms of magnitude, the maximum precipitation values from MLR and PCSR match TRMM data, while Kriging shows slightly higher maximum values. All three downscaled datasets have lower minimum values than TRMM data.

Seasonal precipitation patterns align with annual distributions for spring, summer, and autumn. Winter precipitation shows a distinct pattern with high values in the south and northwest and low values in the central basin. To analyze elevation effects, the study area was divided into 500 m elevation intervals. Results indicate that precipitation increases with elevation from 1,330–2,800 m, peaks at 2,800–3,800 m, then decreases above 3,800 m [Figure 3: see original paper].

2.2 Evaluation and Comparison of Downscaling Models

2.2.1 Annual Precipitation Evaluation Scatter plots comparing measured and downscaled annual precipitation reveal that all three downscaled datasets show improved correlation with observations compared to original TRMM data, with Kriging performing best [Figure 4: see original paper]. Spatial analysis dividing the study area at 100.7°E into western and eastern regions shows that all datasets exhibit higher CC values and lower RMSE values in the eastern region, indicating better performance [Figure 5: see original paper]. Bias analysis reveals underestimation in the west and overestimation in the east, with Kriging showing the smallest deviation. Boxplot comparisons confirm that all three downscaled datasets achieve higher accuracy in the eastern region, with Kriging demonstrating the best overall performance at the annual scale [Figure 6: see original paper].

2.2.2 Seasonal Precipitation Evaluation Seasonal accuracy varies among methods [Figure 7: see original paper]. In spring, Kriging shows the best performance with the highest CC and lowest RMSE. Summer performance is similar across methods, though PCSR exhibits slightly better CC values. Autumn results favor PCSR for correlation, while Kriging shows the smallest Bias magnitude. Winter presents challenges for all methods, with large positive Bias values indicating systematic overestimation. This may be attributed to measurement errors in solid precipitation, as rain gauges poorly capture snowfall, leading to observed precipitation values that are lower than actual precipitation in this cold region. Overall, seasonal precision ranks as PCSR > Kriging > TRMM > MLR.

2.2.3 Monthly Precipitation Evaluation Monthly scatter plots comparing TRMM, downscaled data, and observations show that PCSR correlates best with measured data, followed by MLR, TRMM, and Kriging [Figure 8: see original paper]. RMSE analysis indicates that PCSR and MLR are closest to observations, while Kriging shows the highest dispersion and poorest accuracy. Bias results rank as Kriging < PCSR = TRMM < MLR. Overall, PCSR demonstrates the best monthly-scale performance.

2.3 Analysis of Elevation Effects on TRMM and Downscaled Data

Elevation is a critical factor influencing mountain precipitation. Statistical analysis of station elevations and corresponding accuracy metrics shows no clear trend between elevation and CC values, indicating minimal elevation impact on correlation performance [Figure 9: see original paper]. However, Bias values transition from positive to negative with increasing elevation, revealing a gradual underestimation of precipitation at higher altitudes. This pattern is consistent with findings from previous studies. The underestimation may be attributed to TRMM' s microwave precipitation retrieval algorithm, which tends to underestimate convective precipitation. Since convective precipitation constitutes a large proportion of total precipitation in mountainous areas above 2,800 m, this retrieval bias leads to systematic underestimation at higher elevations.

Conclusions

This study compares downscaling methods for TRMM 3B43 precipitation data in the Qinghai Lake Basin and surrounding areas using MLR, PCSR, and Kriging. Key findings include:

1. All datasets produce consistent spatial precipitation distributions. Annual and seasonal (spring, summer, autumn) precipitation is higher in the north and lower in the west and northwest, while winter precipitation is higher in the south and northwest and lower centrally. Precipitation increases with elevation up to 3,800 m, then decreases.
2. Accuracy evaluation reveals PCSR as the most reliable method overall. At the annual scale, Kriging performs best, with all datasets showing

optimal accuracy in the eastern region. Seasonal precision ranks as PCSR > Kriging > TRMM > MLR. Monthly-scale accuracy is highest for PCSR.

3. While elevation minimally impacts dataset performance, TRMM and downscaled products increasingly underestimate precipitation at higher elevations, likely due to underestimation of convective precipitation in microwave retrieval algorithms.

Based on comprehensive analysis of spatial distribution consistency and accuracy evaluation, PCSR is recommended as the most suitable downscaling method for TRMM 3B43 precipitation data in the Qinghai Lake Basin and surrounding areas.

References

[?] [?]

Note: Figure translations are in progress. See original paper for figures.