Accuracy evaluation of GPM multi-satellite precipitation products in the hydrological application over alpine and gorge regions with sparse rain gauge network
Jiachao Chen, Zhaoli Wang, Xushu Wu, Xiaohong Chen, Chengguang Lai, Zhaoyang Zeng and Jun Li

ABSTRACT
With the release of Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM products, hydrologists can obtain precipitation data with higher resolution and wider coverage. However, great uncertainties still exist in the accuracy and hydrological utility of these data in alpine and gorge regions with sparse gauge stations. In this study, the Lancang River Basin in China was used as an example, and near real-time products (IMERG-E and IMERG-L) and post-processed products (IMERG-F and TMPA 3B42-V7) were evaluated. Different indexes and methods were applied to evaluate the accuracy of these products. The variable infiltration capacity hydrological model was adopted to evaluate their hydrological utility. The following findings were obtained. (1) Compared with observed precipitation data, the near real-time products tend to underestimate, while the post-processed products tend to overestimate precipitation. The performance of the four products in winter is poor. (2) IMERG products offer improvements in two aspects: first, the near real-time products achieve good accuracy and second, the detectability and the accuracy in gorge areas have been greatly improved. (3) The near real-time products have the potential for hydrological applications. The best simulation result was obtained based on IMERG-F, followed by 3B42-V7, IMERG-E, and IMERG-L. (4) The four products can provide reliable precipitation data for the hydrological application over the Lancang River Basin.

Key words | complex terrain, data-sparse areas, GPM IMERG, TMPA 3B42-V7, VIC model

INTRODUCTION
As an essential input of hydrological research, precipitation datasets play an important role in the research of flood monitoring (Zhang et al. 2018), drought warning, agricultural irrigation, and climate change (Lai et al. 2019; Zhong et al. 2019). The two traditional methods for collecting or estimating precipitation data include rain gauges and ground-based radar. However, both methods show limitations with regard to coverage and consistency (Li & Heap 2008; Teegavarapu et al. 2012).

Over the recent years, multi-satellite observation has become a new method to estimate precipitation data for meteorological research, since it overcomes the shortcoming of incomplete space coverage of traditional collection methods (Gadelha et al. 2019). Measuring precipitation with multi-satellites can provide data for regions where rain gauges are sparse. Over the past decade, many multi-satellite precipitation products, such as Climate Prediction Center Morphing (CMORPH) (Joyce et al. 2003).
Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Hsu et al. 1997), and Tropical rainfall measurement mission Multi-satellite Precipitation Analysis (TMPA) (Huffman et al. 2007; Huffman et al. 2010), have been widely used in both hydrological and meteorological research (Liu et al. 2015; Wang et al. 2017b). Among these, TMPA 3B42 is superior compared with similar multi-satellite precipitation products (Worqlul et al. 2014; Liu et al. 2015; Prakash et al. 2016). However, several studies have indicated that 3B42 is less accurate in the Qinghai–Tibet Plateau, alpine, and gorge regions due to its topography (Liu et al. 2010; Zeng & Li 2011; Yang et al. 2013). Furthermore, the near real-time product, 3B42RT, has limitations in hydrological applications due to a severe overestimation problem (Yong et al. 2010; Romilly & Gebremichael 2011; Hao et al. 2014).

Global Precipitation Measurement (GPM) (Hou et al. 2014) is the successor of Tropical Rainfall Measurement Mission (TRMM) (Rui 2010). This mission is conducted by the National Aeronautics and Space Administration (NASA) with the aim to provide more accurate and timely precipitation data for hydrology, meteorology, and agriculture. Compared with its predecessor (TRMM), the observation methods and inversion algorithms of GPM have been substantially improved. It is more sensitive to capture micro-precipitation and differentiate snow and rainfall events (Draper et al. 2015). The most representative algorithm is the Integrated Multi-satellite Retrievals for GPM (IMERG), which provides precipitation data at a spatial resolution of 0.1° × 0.1° with a coverage range of 60°N–60°S and a time resolution of 0.5 h since 12 March 2014 (Huffman et al. 2012). Similar to TMPA, IMERG provides both near real-time precipitation data and post-processed precipitation data. The near real-time version offers two products, i.e. ‘Early run’ and ‘Late run’, and the post-processed version offers ‘Final run’ (hereinafter referred to as IMERG-E, IMERG-L, and IMERG-F, respectively). GPM will immediately perform the inversion algorithms after obtaining the observation data and provide precipitation estimations. Generally, the first result, IMERG-E, is released 4 h after the observation data is obtained. After receiving more calibration data, GPM will perform the algorithms again and produce the second product, IMERG-L, which is generally released 12 h after the observation. IMERG-F is similar to 3B42-V7, which is calibrated using monthly meteorological data from the Global Precipitation Climatology Centre (GPCC); consequently, a 2.5-month delay is generally unavoidable.

Since the release of the first batch of data from NASA, several scholars have begun to preliminary research GPM IMERG-F. Tang et al. (2016) evaluated the detection capability and accuracy between 3B42-V7 and IMERG-F for mainland China. The results show that IMERG-F product performs better than 3B42-V7 at all time scales. With regard to light precipitation detection, GPM performs better than TRMM. With regard to high latitude and arid regions, IMERG-F still performs better than 3B42-V7 (Gaona et al. 2016). All these results demonstrate that GPM has achieved great progress compared with TRMM. However, the accuracy of IMERG products in regions with special terrain and climate remains unexplored. Ma et al. (2016) compared the performances of the 3B42-V7 and IMERG-F over the Qinghai–Tibet Plateau and found that IMERG-F shows a lower detection probability than 3B42-V7 for altitudes exceeding 4,200 m. Another study was conducted on the Chindwin basin in Myanmar, and the results indicate that IMERG-F presents a non-significant improvement and the detectability of heavy rain is inferior to that of 3B42-V7 in the data-sparse mountainous area (Yuan et al. 2017).

There are many important rivers in the world that originate in alpine and gorge areas, and these areas often lack sufficient meteorological data. Obtaining reliable precipitation data in such areas is thus difficult for hydrologists. Since the complexity of the terrain impacts the accuracy of multi-satellite observations, and the data deficiency obstructs the calibration of multi-satellite data (Castro et al. 2015; Wang et al. 2018a), an evaluation of the hydrological performance of IMERG products over complex terrain and sparse data conditions is necessary. Moreover, decision-making, such as flood forecasting and drought warning, requires a near real-time precipitation product. Until now, unfortunately, research on GPM near real-time products (IMERG-E and IMERG-L) is inadequate. Moreover, previous studies were mostly conducted at a large spatial scale, with little focus on the basin scale. To overcome these existing limitations, this study was conducted. The Lancang River Basin is an alpine and gorge area with...
sparse data that was selected for this case study. Statistical methods and the variable infiltration capacity (VIC) hydrological model were adopted to evaluate the accuracy and hydrological utility of four products (IMERG-E, IMERG-L, IMERG-F, and 3B42-V7). This study provides a reference for the hydrological application of IMERG products in regions similar to the Lancang River Basin. The remainder of this paper is organized as follows: Section ‘Study area and data’ provides details of the study area and the dataset used in this paper. Section ‘Methods’ introduces the statistical and hydrological simulation methods. Sections ‘Results and analysis’ and ‘Discussion’ present the main results and the discussion, respectively. Finally, the main conclusions are summarized in the last section.

STUDY AREA AND DATA

Study area

The Lancang River (Upstream of the Mekong River) is one of the most important international rivers in China and offers an abundance of both hydropower and biological resources (Wang et al. 2017a; Li et al. 2018). The basin runs from north to south, and the altitude increases from south to north. To eliminate the influencing factors, such as dense meteorological stations and climatic factors on the lower reaches of the river, the river section above the Jiuzhou hydrological station (99.22°E, 25.78°N) was selected as study area (hereinafter referred to as the Lancang River Basin). As shown in Figure 1, the Lancang River Basin (93.84°–99.67°E, 25.5°–33.85°N) is a typical alpine and gorge region with sparse distribution of meteorological stations (also called rain gauge stations). The elevation of the study area ranges from 6,471 m to 1,243 m above the sea level, and the vertical variation of precipitation is obvious. The basin spans eight latitudes from north to south. It is a typical narrow-shaped basin with a large stream gradient and a strong downcutting effect. The basin’s mean annual runoff is 30.6 billion m$^3$, which is supplemented by precipitation, groundwater, and snowmelt water.

The basin area selected in this study is 87,205 km$^2$, and the Jiuzhou hydrological station provides streamflow data. The study area was divided into 188 grids at a spatial resolution of 0.25° × 0.25°, and a total of 19 meteorological stations were adopted for analysis in this study. The stations that were selected in this study exclude the stations used to calibrate IMERG products.

Data

Multi-satellite precipitation products

3B42-V7 is a post-processed product of TMPA that can be downloaded from <https://pmm.nasa.gov/data-access/downloads/trmm>. V7 indicates that 3B42 has been updated...
to version 7. The spatial resolution of this product is 0.25° × 0.25°. The spatial coverage of the data is 50° N–50° S, and the time coverage currently ranges from 1 January 1998 to 30 November 2017. It is recorded using the Coordinated Universal Time (UTC) with the highest time resolution of 3 h.

The GPM IMERG includes near real-time and post-processed products. Similar to the 3B42-V7, the data is recorded in raster form. It can be downloaded from https://pmm.nasa.gov/data-access/downloads/gpm. Both IMERG-E and IMERG-L provide precipitation data from 12 March 2014 to the present, while IMERG-F is currently only released until 30 November 2017 due to the required calibration. The IMERG algorithm has been updated to version 05B. Since the spatial resolution of the IMERG products is 0.1° × 0.1°, to match the spatial resolution, bilinear interpolation was used to process the spatial resolution of the data to 0.25° × 0.25° (Wang et al. 2017c).

Meteorological data

The measured meteorological data of this study are provided by the China Surface Climate Daily Dataset (V3.0) issued by the China Meteorological Data Service Center (http://data.cma.cn/). The meteorological data used in this study includes precipitation, as well as daily maximum, minimum, and average temperatures, daily average wind speed, relative humidity, and sunshine hours. The meteorological forcing data of the VIC hydrological model was interpolated by the inverse distance weight (IDW) into a raster dataset of 0.25° × 0.25°. In addition, the precipitation data recorded by rain gauge stations was also used as a reference for evaluating the accuracy of the multi-satellite products. The meteorological forcing data required for runoff simulations will be obtained by interpolating the station data. Considering the small number of meteorological stations in the basin, to obtain a more reliable interpolation result, this study also considers the data of the stations out of the basin when performing interpolation calculations.

At the same time, to analyze the temporal and spatial error distribution of the multi-satellite precipitation products, China Gauge-based Daily Precipitation Analysis (CGDPA) was adopted as a further reference. The analysis system uses the modified climatology-based optimal interpolation method (Xie et al. 2007) to interpolate precipitation data gathered from 2,419 National Meteorological Stations into 0.25° × 0.25° raster datasets. The obtained results based on this interpolation method can better reflect the topographic influence (Xie et al. 2007). Currently, the data has been successfully applied to the accuracy assessment of multi-satellite precipitation products (Sun et al. 2016; Wang et al. 2017b; Ma et al. 2018).

Land cover and soil texture data

The underlying surface data used to drive the VIC hydrological model mainly includes land cover, soil texture parameters, and 3D geographic information in the study area. Among these, the land cover data was obtained from a global 1 km resolution land cover database, which was obtained by the Advanced Very High-Resolution Radiometer (AVHRR) satellite and processed by the University of Maryland (Hansen et al. 1998). The soil texture parameters were obtained from the Harmonized World Soil Database (HWSD) issued by the Food and Agriculture Organization (FAO) of the United Nations (FAO et al. 2012), which provides soil types and their physical and chemical properties worldwide.

METHODS

Statistical methods

The formula for the statistical indexes used in this study is shown in Table 1. The probability of detection (POD), frequency of hit (FOH), and critical success index (CSI) were used to evaluate the detectability of the multi-satellite products. Since the above three indexes do not take the true negative events into account, it is still necessary to use Hansen and Kuipers Score (HKS) and Heidke Skill Score (HSS) to evaluate the detectability of a satellite. HKS considers the impact of accurately identifying non-precipitation events on the scoring results with values ranging between −1 and 1. The HSS was used to evaluate whether the detectability of the product was better than a random forecast. The value range of HSS is [−∞, 1], and values above 0 indicate that the product has a detecting ability. The Pearson correlation coefficient (CC) was used to assess the correlation
between two datasets; the relative bias (BIAS) and the root-mean-square error (RMSE) were used to evaluate the deviation of two datasets; the normalized mean square error (NMSE) was used to evaluate the accuracy of the data considering the inherent deviation of reference data. In general, an NMSE above 1.0 indicates that the application effect of the data is not as good as the mean value of the reference data (Yao & Tan 2000); the Nash–Sutcliffe coefficient of efficiency (NSCE) is usually used to evaluate the quality of a model (Nash & Sutcliffe 1970). In addition, when comparing the accuracy of different multi-satellite precipitation products, a Taylor diagram was adopted as one of the evaluation methods (Taylor 2001). The diagram visually reflects the relationship between different products and reference data, using CC and normalized standard deviation (NSD) displayed as a polar graph. It has been suggested that a closer product representative point to the reference data representative point yields a higher product accuracy. The centered root-mean-square error (CRMSE) was used to describe the distance between two points.

The four multi-satellite precipitation products (IMERG-E, IMERG-L, IMERG-F, and 3B42-V7) and CGDPA, from 12 March 2014 to 30 November 2017, were selected as raw data for statistical analyses. POD, FOH, CSI, HKS, and HSS were used to assess the detecting ability of different products. Statistical analysis was performed at two temporal scales (daily scale and monthly scale) and two spatial scales (grid scale and basin scale). The data processing method at the grid scale compares the precipitation data between the satellite grid and the nearest gauge station. The algorithm of the basin scale is as follows:

1. Calculate the average daily precipitation of all stations in the basin.

### Table 1 | List of formulas for evaluating indexes

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Formula</th>
<th>Optimal value</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD</td>
<td>POD = ( \frac{n_{11}}{n_{11} + n_{10}} )</td>
<td>1</td>
</tr>
<tr>
<td>FOH</td>
<td>FOH = ( \frac{n_{11}}{n_{11} + n_{01}} )</td>
<td>1</td>
</tr>
<tr>
<td>CSI</td>
<td>CSI = ( \frac{n_{11}}{n_{11} + n_{01} + n_{10}} )</td>
<td>1</td>
</tr>
<tr>
<td>HKS</td>
<td>HKS = ( \frac{n_{11}}{n_{11} + n_{10} + n_{01} + n_{00}} )</td>
<td>1</td>
</tr>
<tr>
<td>HSS</td>
<td>HSS = ( \frac{2(n_{11}n_{00} - n_{10}n_{01})}{(n_{11} + n_{10})(n_{10} + n_{00}) + (n_{11} + n_{01})(n_{01} + n_{00})} )</td>
<td>1</td>
</tr>
<tr>
<td>CC</td>
<td>CC = ( \frac{\text{cov}(P, S)}{\sqrt{\text{var}(P)} \cdot \sqrt{\text{var}(S)}} )</td>
<td>1</td>
</tr>
<tr>
<td>BIAS</td>
<td>BIAS = ( \frac{\sum_{i=1}^{n} (S_i - \bar{S}) \cdot (P_i - \bar{P})}{\sum_{i=1}^{n} P_i} \times 100% )</td>
<td>0%</td>
</tr>
<tr>
<td>RMSE</td>
<td>RMSE = ( \sqrt{\frac{1}{n} \sum_{i=1}^{n} (S_i - P_i)^2} )</td>
<td>0</td>
</tr>
<tr>
<td>NMSE</td>
<td>NMSE = ( \frac{1}{n} \sum_{i=1}^{n} \frac{(S_i - P_i)^2}{\text{var}(P)} )</td>
<td>0</td>
</tr>
<tr>
<td>CRMSE</td>
<td>CRMSE = ( \sqrt{\frac{1}{n} \sum_{i=1}^{n} [S_i - \bar{S} - (P_i - \bar{P})]^2} )</td>
<td>0</td>
</tr>
<tr>
<td>NSCE</td>
<td>NSCE = ( 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{(S_i - P_i)^2}{\text{var}(P)} )</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: \( P \), reference sequence; \( S \), sequence to be evaluated; \( \text{var} \), the variance operator; \( \text{cov} \), the covariance operator; \( n_{xx} \), the number of events that satisfy the condition; The first \( x \) indicates whether there is a precipitation event; the second \( x \) indicates whether the satellite detects a precipitation event; 0 indicates no, and 1 indicates yes.
2. Calculate the average daily precipitation of all satellite grids in the basin.
3. Calculate the statistical indexes of the two sets of data.

Furthermore, the precipitation data at two spatial scales were classified and calculated at the daily scale, and the probability distribution function (PDF) and ratio of rainfall (RR) at different rainfall rates were computed. The RR is the ratio of the rainfall in the same rainfall intensity range to the total rainfall.

The monthly precipitation data from CGDPA and the multi-satellite products were processed into a set of comparison sequences, and statistical indexes (CC, BIAS, and RMSE) were calculated. Time-ordered accuracy indexes were obtained and used to study the variation characteristics of the product error over time. Then, taking CGDPA as a reference, the statistical indexes of the individual grid were calculated to explore the spatial error distribution of the multi-satellite products.

Hydrological simulation

The VIC macroscale hydrological model (http://vic.readthedocs.io/en/master/) is a distributed hydrological model that was developed and improved by Liang et al. (1994) and Liang & Xie (2001). The source code for this model is hosted on GitHub (https://github.com/UW-Hydro/VIC). Compared with the traditional hydrological model, VIC can simultaneously calculate the water balance and energy balance, and simulate the land surface process based on the water-heat transfer process between the atmosphere, land cover, and soil. Spatially, VIC rasterizes the horizontal plane, calculates the streamflow generation separately in each grid, and divides the vertical direction as one layer for land cover and three layers for soil. Although the underlying surface and the runoff mechanism in southwestern China are very complex, the VIC model can fully consider the spatial variation of the underlying surface. Additionally, the model considers the proportion of different vegetation types at the sub-grid scale; in terms of the runoff mechanism, the VIC model also considers the two runoff generation theories of excess infiltration and excess storage. Therefore, the model is applicable to the Lancang River Basin, and several successful cases in southwestern China have already demonstrated its powerful function (Su et al. 2017; Zhong et al. 2018).

The current VIC model has been updated to version 5, which supports parallel operations. At the same time, longwave radiation and shortwave radiation are listed as necessary inputs for this model, which yields more accurate computation results for VIC. The vapor pressure, shortwave radiation, and longwave radiation can be converted from the daily maximum temperature, daily minimum temperature, daily average temperature, relative humidity, and sunshine hours (Martinezlozano et al. 1984; Allen et al. 1994; Konzelmann et al. 1994). The soil hydraulic parameters can be calculated from the soil texture parameters by using the formula proposed by Saxton & Rawls (2006). The model requires a total of seven empirical parameters to be determined. These are the power exponent b that affects the shape of the VIC curve, and the three basic flow parameters of the ARNO model (Famiglietti & Wood 1991): the proportional coefficient Ws, the maximum base flow Dsmax, the base flow speed ratio Ds, and the thickness of the three soil layers (d1, d2, and d3). In this study, large NSCE and small BIAS values were pursued during the calibration period, and then, reasonable calibration results were selected based on the experience. The VIC model requires a start-up period, so that the parameters, such as soil water content, are close to the actual situation. To fully utilize the multi-satellite precipitation data, 1 January 2013–11 March 2014 was set as the start-up period, and the rain gauge data was used to drive the model. However, the simulation results during the start-up period will not be included when statistical indexes are calculated. At present, the VIC hydrological model is widely used due to its good performance (Yuan et al. 2005; Hengade & Eldho 2016). Due to the lack of relevant snowmelt data, this study does not consider the snowmelt process. Fortunately, in the runoff of the Lancang River, the contribution ratio of snowmelt is low; therefore, even if the snowmelt process is not considered, the hydrological simulation results are still reliable.

Since the VIC model only computes the runoff generation in each grid, it is necessary to use a routing model for processing. This study used the routing model developed by Lohmann et al. (1996) as recommended by the VIC official website. The flow direction file can be extracted by Digital Elevation Model (DEM); the flow rate and hydraulic diffusion
parameters are set to 1.5 and 800, respectively. The unit hydrograph of each grid can be estimated according to the above parameters. The model applies the unit hydrographs in the grid for the overland flow concentration and solves the river routing process with the linear Saint-Venant equation.

Hydrological simulations were performed at a daily scale, using two scenarios to fully exploit the potential of multi-satellite precipitation products. Due to a lack of observed runoff data from 2014 to the present, the simulated runoff based on rain gauge data was used as a reference. In hydrological research, the lack of observed data (or satisfying data) is a common occurrence. Yong et al. (2010) suggested that using the measured precipitation data to drive the calibrated hydrological model can yield representative reference data for the hydrological utility evaluation to a certain extent. In a previous study, we have used the observed runoff data from the Jiuzhou hydrological station to calibrate the VIC model and obtained a good result (NSCE: 0.73, BIAS: 8.56%) (Wang et al. 2018b). However, using the runoff simulation result based on rain gauge data as a reference still contains unknown uncertainty. Therefore, the simulation results from the previous study were further analyzed. The impact of using different reference data on the conclusions is compared. The CC, BIAS, RMSE, and NSCE are calculated during both the calibration period and validation period.

The calibrated VIC model based on gauge data was defined as the observed VIC model. The IMERG products and 3B42-V7 provide input into the observed VIC model to investigate whether the multi-satellite precipitation product is suitable for hydrological simulation, only using existing observation data (Scenario I). The calibrated VIC model based on 3B42-V7 data is defined as the multi-satellite VIC model. The IMERG products and 3B42-V7 provide input into the multi-satellite VIC model to investigate whether better simulation results can be obtained by recalibrating the VIC model using multi-satellite precipitation products (Scenario II). The calibration and validation of the observed VIC model and the multi-satellite VIC model have been conducted previously and will not be repeated here (Wang et al. 2018b).

RESULT AND ANALYSIS

Statistical evaluation

The average annual precipitation of IMERG products (IMERG-E, IMERG-L, and IMERG-F) and 3B42-V7 are shown in Figure 2. The spatial distribution patterns of precipitation from multi-satellite products are relatively consistent. The precipitation increases from north to south, and large rainfall was detected near 31°N–32°N. Among them, the average annual precipitation of the near real-time products varies from 200 to 1,100 mm/year, while the range of the post-processed products varies from 400 to 1,500 mm/year, which exceeds the former by 200 mm/year.

![Figure 2](https://i.imgur.com/3Q5Q5Q5.png)

**Figure 2** | Spatial distribution of IMERG products and 3B42-V7 average annual precipitation in the Lancang River Basin.
The detection ability of the four multi-satellite products was evaluated using the CGDPA precipitation product. The daily precipitation of all grids was treated as a set of comparison sequences, and the overall indexes are shown in Table 2. The POD of IMERG products reached 0.82, while the FOH was slightly less than that of 3B42-V7. This phenomenon indicates that the false negative event of IMERG products is less than 3B42-V7, and the false positive event is more. The CSI results show that the detectability of all IMERG products was generally higher than that of 3B42-V7. The detectability of the products was further quantified using two detect skill scores: HKS and HSS. The results of the skill scores show that IMERG products have good detecting ability with 3B42-V7, and the former is better. Overall, all IMERG products showed better detectability than 3B42-V7, and each detecting index between the near real-time products and the post-processed products of IMERG was quite close, which may be due to the sparse GPCC network. This situation weakens the calibration effect on the post-processed products, and therefore, the detecting ability of the post-processed products improved less.

The detecting indexes for each grid were calculated separately. The distribution of different indexes is shown in Figure 3, and the spatial distribution is shown in Figure 4. Box plots are used to display the upper and lower limits, upper and lower quartiles, median, and outliers (indicated by dots) of a dataset. The distributions of the same index of IMERG products were similar. In addition to FOH, other indexes of IMERG were significantly better than those of 3B42-V7, with IMERG-L and IMERG-F performing slightly better than IMERG-E. Except for the normal values of indexes, the outliers in Figure 3 are also cause for concern. As shown in Figure 4, the upper outliers mainly appear in the southern part of the basin, which may be due to the sudden decrease of the altitude in the south and the increased rainfall. Low altitude and abundant rainfall are favorable for satellite observation; therefore, the indexes in this part perform significantly better than others. However, the upper outliers are closer to the upper limit and do not greatly impact the overall quality. In contrast, the lower outliers are farther from the lower limit, which may imply instability of the product quality. The lower outliers are mainly found in 3B42-V7 and mainly occur in the gorge area south of the basin. This indicates that the detecting capacity of IMERG products in the gorge area has been significantly improved compared with that of 3B42-V7.

The accuracy evaluation results are shown in Table 3 and Figure 5. The results of 3B42-V7 are similar to those of our previous study (Wang et al. 2018b). At the daily grid scale, the performance is poor, but with the expansion of temporal and spatial scales, the accuracy could be greatly improved, and the overall performance of the product was satisfying. At the daily grid scale, IMERG products showed less error and higher correlation than 3B42-V7. However, in the estimation of the total precipitation (see BIAS), IMERG products performed poorly. The near real-time products underestimated total precipitation, while the post-processed products overestimated it. At the daily basin scale, IMERG products still showed a higher correlation than 3B42-V7, and the error was further reduced. The BIAS of IMERG products decreased significantly, but the underestimation from IMERG-E and IMERG-L was more severe. At the monthly grid scale, the CCs of the four products were greatly improved. IMERG-F achieved the highest CC among the four products (CC = 0.91). At the monthly basin scale, all indexes were close to the optimal value.

Table 2 | Comparison of overall forecasting capacity of four satellite precipitation products

<table>
<thead>
<tr>
<th>Products</th>
<th>POD</th>
<th>FOH</th>
<th>CSI</th>
<th>HKS</th>
<th>HSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMERG-E</td>
<td>0.82</td>
<td>0.68</td>
<td>0.59</td>
<td>0.48</td>
<td>0.47</td>
</tr>
<tr>
<td>IMERG-L</td>
<td>0.82</td>
<td>0.68</td>
<td>0.60</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>IMERG-F</td>
<td>0.82</td>
<td>0.68</td>
<td>0.60</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td>3B42-V7</td>
<td>0.62</td>
<td>0.73</td>
<td>0.51</td>
<td>0.43</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Figure 3 | Box plot of indicators for each of the four satellite precipitation products.

Figure 5 | The hydrological utility of GPM products in alpine and gorge region.
Figure 6 visually shows the relative proximity of the four multi-satellite products to the rain gauge data. The Taylor diagram was drawn based on the polar axis, the black arc represents the NSD of the rain gauge data, and the green arc shows the ratio of CRMSE to NSD. As shown in Figure 6, the near real-time products have a significantly different error pattern compared with the post-processed products. The near real-time products tend to underestimate, while the post-processed products tend to overestimate. The performance of the four products at the monthly grid scale is similar, and the IMERG-F performs slightly better. The near real-time products outperform the post-processed
Table 3 | Accuracy evaluation result of four satellite precipitation products on different temporal and spatial scales

<table>
<thead>
<tr>
<th>Temporal scale</th>
<th>Products</th>
<th>Grid scale</th>
<th>Basin scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CC</td>
<td>BIAS%</td>
</tr>
<tr>
<td>Daily scale</td>
<td>IMERG-E</td>
<td>0.41</td>
<td>-14.62%</td>
</tr>
<tr>
<td></td>
<td>IMERG-L</td>
<td>0.42</td>
<td>-18.77%</td>
</tr>
<tr>
<td></td>
<td>IMERG-F</td>
<td>0.43</td>
<td>19.01%</td>
</tr>
<tr>
<td></td>
<td>3B42-V7</td>
<td>0.35</td>
<td>8.18%</td>
</tr>
<tr>
<td>Monthly scale</td>
<td>IMERG-E</td>
<td>0.87</td>
<td>-14.62%</td>
</tr>
<tr>
<td></td>
<td>IMERG-L</td>
<td>0.87</td>
<td>-18.77%</td>
</tr>
<tr>
<td></td>
<td>IMERG-F</td>
<td>0.91</td>
<td>19.01%</td>
</tr>
<tr>
<td></td>
<td>3B42-V7</td>
<td>0.89</td>
<td>8.18%</td>
</tr>
</tbody>
</table>

Figure 5 | Scatterplots for IMERG products and 3B42-V7, vs. rain gauge measurements on different temporal and spatial scales.
products at three other scales. In general, the near real-time products are significantly less accurate than the post-processed products because they have not been calibrated with GPCC monthly meteorological data. With regard to the good performance in both IMERG-E and IMERG-L in this study, it can be concluded that the high-quality raw data and algorithms of GPM enable the product to perform well before calibration. However, the advantages of the post-processed products are not obvious under the condition of sparse gauge station distribution, and moreover, the accuracy improvement caused by calibration is limited. Figure 7 shows the PDF and RR for each precipitation data at different rainfall rates. At the grid scale, IMERG products underestimate the PDF of no-precipitation event and overestimate the PDF at a rainfall rate of 0–1 mm/day, which accounts for less than 10% of the total rainfall. The PDF of the precipitation with a rainfall rate of >1 mm/day is similar to the gauge data. At the basin scale, all four products tend to underestimate the PDF of no-precipitation and overestimate at a rain rate of 0–1 mm/day, which indicates that IMERG may still suffer from excessive sensitivity to light precipitation. The PDF at a rainfall rate >1 mm/day is still similar to the gauge data, and the distribution of RR for the post-processed products is consistent with the gauge data, while the near real-time products show a different distribution.

Figure 8 shows the time series of indexes for monthly precipitation from both IMERG products and 3B42-V7. CC achieved a relatively high level overall. However, from November to December 2014, December 2016 to January 2017, and November 2017 abnormally small values appeared, indicating that the products may have poor performance in winter. A CC close to zero means that the precipitation in winter is spatially different from CGDPA, and the BIAS close to −100% indicates that the products severely underestimate daily rainfall. However, considering
that the winter rainfall is relatively small, this will not affect the subsequent hydrological simulation. Overall, although the near real-time products have not been calibrated by the GPCC monthly data, their correlation is still close to that of the post-processed products. The satellites have accurate estimates of precipitation for most of the time (see BIAS). The near real-time products tend to underestimate the actual precipitation. RMSE is affected by rainfall and increased during periods of heavy rainfall. Overall, the BIAS and RMSE of the near real-time products deviate farther from the optimal values than that of the post-processed products.

The spatial distribution of the error is shown in Figure 9. 3B42-V7 achieved a poor correlation on the northern part of the basin and the gorge area. This has been significantly improved in IMERG products, of which IMERG-F performed best. Spatially, the near real-time products still tend to underestimate precipitation (negative BIAS), while the post-processed products tend to overestimate it (positive BIAS). The precipitation in the basin showed less error in the time dimension; therefore, the RMSE calculation result was small. Due to the large rainfall in the south, the RMSE has also increased accordingly. IMERG-L performs slightly better on RMSE.

Hydrological evaluation

The uncertainty analysis result is listed in Table 4. Using measured data as a reference or using the gauge-based simulation result as a reference can lead to the same conclusion (CC above 0.85 and NSCE above 0.70). The difference of the RMSE obtained based on these two references remains within 100 m³/s, indicating that the error of the result can be correctly recognized even if the gauge-based simulation result is used as a reference. However, it should be noted that using the gauge-based result as a reference may lead to a slightly overly optimistic conclusion. The conclusion consistency and the optimism occur during both the calibration period and the validation period and are not an accidental phenomenon. In addition, for the BIAS, it is
best to refer only to its absolute value. Its positive or negative sign (simulated runoff is higher/lower than the reference data) has little reference value since its sign is mainly affected by the deviation between the measured runoff and the gauge-based simulation result. In summary, there is indeed uncertainty in the hydrological evaluation method, but the further evaluation result shows that the error caused by this uncertainty is actually very small; therefore, the following evaluation is still reliable.

The daily runoff simulation results are shown in Figure 10 (Scenario I) and Figure 11 (Scenario II). Figure 10 shows that the runoff simulated by the observed VIC model shows a high correlation with the reference runoff, but the error in the products is further amplified and transmitted to the simulation results. Compared with the statistical results in Table 3 (daily basin scale), the products error propagation through the VIC model is obtained. The CCs of the four products have improved significantly. Table 3 shows CC values of only 0.60 for IMERG-F and 0.95 for the simulated streamflow. The error propagation of BIAS shows an increasing trend, and the increasing trend of the post-processed products is more apparent; the 3B42-V7

Figure 9 | Spatial distribution of error from IMERG products and 3B42-V7.

Table 4 | Uncertainty analysis result of using the gauge-based runoff simulation result as a reference

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
<td>BIAS</td>
</tr>
<tr>
<td>Measured vs. gauge-based</td>
<td>0.86</td>
<td>8.56%</td>
</tr>
<tr>
<td>Gauge-based vs. 3B42-V7-based</td>
<td>0.89</td>
<td>2.78%</td>
</tr>
<tr>
<td>Measured vs. 3B42-V7-based</td>
<td>0.86</td>
<td>11.58%</td>
</tr>
</tbody>
</table>
increases from 8.65 to 30.89%. All four products can reproduce the medium- and low-flow well, while the outputs based on the near real-time products underestimate several of the peaks, and the post-processed product-based outputs overestimate all peaks. It is worth noting that although the BIASs are high, IMERG-E and IMERG-L still exhibit specific hydrological application potentials with NSCEs of 0.67 and 0.61, respectively (Moriasi et al. 2007). Using IMERG-F and 3B42-V7 as meteorological forcing data leads to severe overestimations of the peak flow. Low NSCE and high BIAS indicate that the observed VIC model is not suitable for the post-processed products.

Figure 11 shows that the multi-satellite VIC model is suitable for IMERG-F and 3B42-V7. This model effectively
reduces the error of the simulation results based on the post-processed products, and the BIAS decreases sharply (2.69% for IMERG-F and 1.55% for 3B42-V7). The NSCEs also increased to above 0.8. However, the multi-satellite VIC model was not applicable to IMERG-E and IMERG-L. The NSCEs of IMERG-E and IMERG-L were only 0.15 and 0.07, respectively. Compared with Table 3, it can be concluded that the multi-satellite VIC model can effectively improve the CC of the satellite precipitation products. Furthermore, for the post-processed products, the model can produce reduced BIAS simulation results. Such a model can effectively control and reduce the transmission of errors. However, for the near real-time products, this model even enlarges its BIAS. This may be due to the
different error distribution patterns between the near real-time products and the post-processed products. Runoff results based on the near real-time products perform poorly in high, medium, and low flow. The post-processed product-based outputs can reproduce the medium- and high-flow well; only a few peaks are overestimated, and the low flow is generally overestimated. Therefore, the multi-satellite VIC model, which was calibrated based on the post-processed product (3B42-V7) was suitable for the post-processed products, while it was not suitable for the near real-time products.

Combined with the accuracy evaluation results, the near real-time products achieved a similar accuracy than the post-processed products and should exhibit similar hydrological utility. Considering that the VIC model has a specific tolerance level for satellite precipitation products (Yong et al. 2010), it is possible to obtain a model that is more suitable for the near real-time products. Since the multi-satellite VIC model is calibrated based on the post-processed 3B42-V7, while the near real-time products have different error distribution patterns, the model may need to be recalibrated to better demonstrate the hydrological utility of the near real-time products. Here, the VIC model was recalibrated using the reference runoff and IMERG-E to reveal the potential of hydrological utility, and the model was defined as the near real-time VIC model. Then, IMERG-L was applied to the near real-time VIC model, setting 12 March 2014–31 December 2015 as the calibration period and 1 January 2016–30 November 2017 as the validation period. As shown in Figure 12, the BIAS decreased significantly after recalibration and the NSCEs increased above 0.8, showing similar runoff simulation capability as the post-processed products. Due to limitations in data and other factors, this model was not calibrated by the observed runoff data but by the simulation result based on the rain gauge data. The parameters of the three models used in this study are also listed in Table 5 for reference.

Overall, both IMERG-E and IMERG-L exhibit a certain degree of hydrological utility when the observed VIC model is used. The use of the multi-satellite VIC models allows the

| Table 5 | Calibrated parameters of the VIC model |
|-----------------|-------------------|-------------------|-------------------|-------------------|
| VIC model | b | Ds | Dsmax | Ws | d₁ | d₂ | d₃ |
| Observed | 0.55 | 0.13 | 29.24 | 0.93 | 0.10 | 0.34 | 0.44 |
| Remote sensing | 0.30 | 0.05 | 12.15 | 0.66 | 0.10 | 0.56 | 0.48 |
| Near real-time | 0.55 | 1.00 | 8.00 | 0.98 | 0.10 | 0.10 | 0.80 |
post-processed products to better perform their hydrological utility, of which IMERG-F is slightly better than 3B42-V7.

**DISCUSSION**

During the evaluation, several problems emerged that are worth exploring in depth. While conducting an accuracy evaluation of the multi-satellite precipitation products, the performance of the products at the daily scale was found to be unsatisfactory. NMSE above 1 indicates that the precipitation estimate is not as accurate as simply using the average of rain gauge data. However, this statistical index cannot reflect the accuracy of the spatial distribution pattern, while the hydrological model can. Therefore, to understand the hydrological utility of IMERG products in a data-sparse area with alpine and gorge terrain, it is necessary to use a hydrological model for evaluation. The satisfying simulation results indicate that the multi-satellite products have an accurate spatial distribution pattern, which cannot be achieved simply by replacing the mean value. For the post-processed products, the statistical results show that they are not better than the near real-time products. However, the precipitation spatial distribution may be more accurate than the latter, and the overestimation was corrected by the parameters of the hydrological model. Therefore, the hydrological simulation results of the post-processed products are better.

In fact, to understand the impact of data-sparsity and alpine gorge terrain on IMERG products, it is also necessary to compare this result with the relevant research in other areas. Using the Beijiang River basin as an example, both dimensionless indexes, CC and BIAS, show that the accuracy of the satellite products is better than the results in this study at both the grid scale and the basin scale (Wang et al. 2017c). The topography and the data of the Lancang River Basin are not conducive to the satellite estimation and the algorithm calibration.

The choice of the study area is very important. He et al. (2017) also conducted research on IMERG-F in the upper reaches of the Mekong River. However, their study did not investigate the influence of alpine gorge and data-sparsity on the quality of multi-satellite precipitation products since the lower reaches of the Lancang River Basin have dense rain gauge stations. It is obvious that if too many rain gauge stations are concentrated in the lower reaches, the precipitation data used for calculation will have a higher proportion of data from the lower reaches. The topography and precipitation of the upper and lower reaches of the Lancang River vary greatly. Since the investigated problem is the hydrological utility of IMERG products under the conditions of the alpine, gorge, and sparse station network, it is necessary to select only the basin controlled by the Jiuzhou hydrological station as the study area.

The Lancang River Basin is a complex basin, and the spatial continuity of its precipitation is poor, and precipitation is affected by terrain. Therefore, in theory, there must be sufficient rain gauge stations to reflect the spatial distribution of precipitation in the basin. However, the actual situation is that the rain gauge stations in the Lancang River Basin are sparsely distributed. To make the most of the data from rain gauge stations, the rainfall station around the basin was also considered when interpolating. However, this action raises another question: do the precipitation records outside the basin strongly affect the precipitation interpolation results in the basin? In fact, with the current technique, it is difficult to know the real situation of the spatial distribution of precipitation in the basin or to quantify the uncertainty; therefore, this study only applies the interpolation results to hydrological simulation. Regardless, from the previous research results with NSCE of 0.73 and BIAS of 8.56% (Wang et al. 2018b), the simulation results of this study proved to be reliable.

The results of the runoff simulation indicate that it is feasible to use the different inputs to calibrate the model to compensate for errors in the input data. As long as the input data contains signals (rather than absolute values) that are in-sync with the true observed value, the model can perform fairly well. However, is this the commonality of conceptual hydrological models? What percentage of the sync signal in the input must be available to obtain reliable simulation results? These questions need to be answered by further research.

It should be noted that the real-time application of the near real-time products is still limited. At present, the release of the near real-time products still suffers from a 4–12 h delay; therefore, it is almost impossible to apply real-time hydrological forecasting. However, the simulation results
of this study indicate that compared with the near real-time algorithm of TMPA, the IMERG algorithm can provide near real-time products that have the potential of runoff simulation, and the research results of Wang et al. (2017c) also support this conclusion. This suggests the following revelation: the staff that develop the next-generation multisatellite precipitation product inversion algorithm should focus on optimizing the algorithm, reducing the delay of product release, and realizing real-time application in the true sense.

CONCLUSIONS

In this study, IMERG products (and their predecessors, 3B42-V7) were selected to explore their accuracies and hydrological utility under data-sparse, alpine, and gorge conditions. The Lancang River Basin was used as an example. The conclusions are as follows:

1. IMERG products have a similar spatial distribution pattern for precipitation to 3B42-V7. The post-processed products (IMERG-F and 3B42-V7) tend to overestimate precipitation levels, while the near real-time products (IMERG-E and IMERG-L) tend to underestimate these.

2. The three products of IMERG achieve similar detection accuracy and perform better than 3B42-V7. The near real-time products are of good quality, while due to data-sparcity, the post-processed products are not better than the near real-time products in the gorge area.

3. The four products have high accuracy, and the error of IMERG products is smaller than that of 3B42-V7; however, 3B42-V7 is superior for estimating the total precipitation. The near real-time products have similar accuracy than the post-processed products. The RR distribution pattern of the post-processed products differs from the near real-time products and achieves a better match with the rain gauge data.

4. The performance of IMERG products in the gorge region is improved compared with 3B42-V7, and IMERG-F performed best; however, the performance of the four products is still poor in winter.

5. IMERG-F can provide better precipitation data than the 3B42-V7 for the Lancang River Basin (a data-sparse region with alpine and gorge terrains) and achieves better hydrological simulation results. Both IMERG-E and IMERG-L can also provide reliable precipitation data.

ACKNOWLEDGEMENTS

The research is financially supported by the National Natural Science Foundation of China (Grant Nos. 51579105, 51709117, 91547202, 51879107); China Postdoctoral Science Foundation (2017M612662); the Science and Technology Program of Guangzhou City (201707010072); the Science and Technology Planning Project of Guangdong Province, China (2017A040405020); and the Special Fund of Water Resources Conservation and Protection of Guangdong Province (2017).

REFERENCES


FAO, ISRIC, ISSCAS 2012 Harmonized World Soil Database (Version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Gadelha, A. N., Coelho, V. H. R., Xavier, A. C., Barbosa, L. R., Melo, D. C. D., Xuan, Y., Huffman, G. J., Petersen, W. A. & Almeida, C. d. N. 2019 Grid box-level evaluation of IMERG over Brazil at various space and time scales. Atmos. Res. 218, 231–244.


First received 8 March 2019; accepted in revised form 2 September 2019. Available online 18 October 2019.