An improved procedure for the validation of satellite-based precipitation estimates

Ling Tang a,b,⁎, Yudong Tian a,b, Fang Yan c, Emad Habib c

a Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, 20740, USA
b Code 617, Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA
c Department of Civil Engineering, University of Louisiana at Lafayette, P.O. Box 42991, Lafayette, LA, 70504, USA

A R T I C L E   I N F O

Article history:
Received 21 March 2014
Received in revised form 19 December 2014
Accepted 23 December 2014
Available online 15 January 2015

Keywords:
Satellite-based precipitation estimates
Validation procedure
Additive and multiplicative error model
Logarithmic transformation
Uncertainty estimation

A B S T R A C T

The objective of this study is to propose and test a new procedure to improve the validation of remote-sensing, high-resolution precipitation estimates. Our recent studies show that many conventional validation methods do not accurately capture the unique error characteristics in precipitation estimates to better inform both data producers and users. The proposed new validation procedure has two steps: 1) an error decomposition approach to separate the total retrieval error into three independent components: hit error, false precipitation and missed precipitation; and 2) the hit error is further analyzed based on a multiplicative error model. In the multiplicative error model, the error features are captured by three model parameters. In this way, the multiplicative error model separates systematic and random errors, leading to more accurate quantification of the uncertainties. The proposed procedure is used to quantitatively evaluate the recent two versions (Version 6 and 7) of TRMM’s Multi-sensor Precipitation Analysis (TMPA) real-time and research product suite (3B42 and 3B42RT) for seven years (2005–2011) over the continental United States (CONUS). The gauge-based National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC) near-real-time daily precipitation analysis is used as the reference. In addition, the radar-based NCEP Stage IV precipitation data are also model-fitted to verify the effectiveness of the multiplicative error model. The results show that winter total bias is dominated by the missed precipitation over the west coastal areas and the Rocky Mountains, and the false precipitation over large areas in Midwest. The summer total bias is largely coming from the hit bias in Central US. Meanwhile, the new version (V7) tends to produce more rainfall in the higher rain rates, which moderates the significant underestimation exhibited in the previous V6 products. Moreover, the error analysis from the multiplicative error model provides a clear and concise picture of the systematic and random errors, with both versions of 3B42RT have higher errors in varying degrees than their research (post-real-time) counterparts. The new V7 algorithm shows obvious improvements in reducing random errors in both winter and summer seasons, compared to its predecessors V6. Stage IV, as expected, surpasses the satellite-based datasets in all the metrics over CONUS. Based on the results, we recommend the new procedure be adopted for routine validation of satellite-based precipitation datasets, and we expect the procedure will work effectively for higher resolution data to be produced in the Global Precipitation Measurement (GPM) era.

1. Introduction

Space-borne precipitation products, with their global coverage, high resolution, frequent sampling and easy access, have been widely used in various applications (e.g., natural hazards, hydrology, agricultural forecasts, and climate studies). However, the errors associated with these satellite precipitation products need quantitative evaluation, because of the highly nonlinear nature of the physical process to measure precipitation from space. The strengths and limitations of those satellite precipitation products need to be understood so they can be interpreted correctly between the data-producing community and data users, especially during the Global Precipitation Measurement (GPM) era with large volume of higher resolution (~10 km, hourly) precipitation data expected to be generated in near future (Huffman et al., 2012).

Quantitative evaluation of satellite precipitation products is critical for both data producers and external users. On one hand, effective error analysis will yield insight into the sources of errors in the precipitation products and possible ways to correct or reduce them. This will lead to the improvement of next generation data algorithms and enhance their data quality. On the other hand, for end users, such evaluation and error characteristics analysis will give better guidance in selecting products for their particular applications, and help them assess the impact of input errors propagated into their applications (Tian et al., 2009).
The Algorithm Inter-comparison Projects (AIP) of the Global Precipitation Climatology Project (GPCP) (e.g., Arkin and Xie, 1994; Ebert et al., 1996), the Precipitation Inter-comparison Projects (PIP) (e.g., Smith et al., 1998; Adler et al., 2001), and a comprehensive validation study at global scale called the Pilot Evaluation of High Resolution Precipitation Products (PEHRPP) (Arkin and Turk, 2006) are examples of major past inter-comparison studies. The International Precipitation Working Group (IPWG, online at www.isac.cnr.it/ipwg/), builds upon the earlier evaluation experiences, has established a validation program to provide both the data producers and external users with up-to-date information on the quality of the precipitation estimates from virtually all the operational satellite algorithms (Ebert et al., 2007). Meanwhile, a number of new multi-sensor precipitation algorithms have been developed to exploit the complementary strengths of three different types of precipitation measuring sensors: Infrared (IR), Passive Microwave (PMW) Radiometers, and Precipitation Radar (PR) (e.g., Sorooshian et al., 2000; Joyce et al., 2004; Huffman et al., 2007). Some multi-sensor precipitation products also incorporate ground-based precipitation measurements, such as rain gauge data. A considerable number of evaluation studies have been devoted to the error analysis and uncertainty quantification for satellite-based precipitation products (e.g., McCollum et al., 2002; Gottschalck et al., 2005; Ebert et al., 2007; Hossain and Huffman, 2008; Lin and Hou, 2008; Tian and Peters-Lidard, 2007; Tian et al., 2007, 2009; Sapiano and Arkin, 2009; Kubota et al., 2009; Habib et al., 2009a; Tian and Peters-Lidard, 2010; Tian et al., 2010; Kirstetter et al., 2012; Tian et al., 2013; Chen et al., 2013a, b; Maggioni et al., 2014; Tang et al., 2014). Most of these researches used conventional error metrics to quantify the uncertainties (e.g., bias, root mean square error). For instance, Ebert et al. (2007) evaluated several operational satellite and numerical weather prediction (NWP) precipitation products, against gauge-based data sets over the continental US, Australia, and Europe, using several conventional error metrics (e.g., correlation, bias ratio, probability of detection and false alarm ratio, etc.). They found that satellite precipitation estimates are more accurate during summer and at lower latitudes. Meanwhile, Hossain and Huffman (2008) proposed a conceptual framework for developing error metrics in three general dimensions: 1) spatial (how does the error vary in space?); 2) retrieval (how “off” is each precipitation estimate from the true value?); and 3) temporal (how does the error vary in time?). They employed formulations for error metrics specific to each dimension, in addition to the conventional error metrics. They applied the error framework on four satellite precipitation products and found that this error framework can identify seasonal and regional differences in uncertainties of data sets more clearly than the conventional error metrics.

Some recent validation studies focused on the performance of the newest version (V7) of Tropical Rainfall Measurement Mission (TRMM) Multi-sensor Precipitation Analysis (TMPA). Yong et al. (2013) compared the performance of TMPA real-time and post-processed products 3B42RT and 3B42 Version 7 (V7) with the previous Version 6 (V6) over two river basins in China, and found that V7 algorithm significantly reduced systematic bias in the low-latitude river basin, while it was ineffective in the high-latitude river basin. Y. Chen et al. (2013) evaluated 3B42 V7 precipitation estimates for tropical cyclone rainfall on two terrain types: low-lying atoll sites (considered as open ocean), and coastal and island sites (land). The results show that 3B42V7 tends to overestimate heavy rain frequency on atoll sites, and underestimate heavy rain frequency on coastal and island sites. Chen et al. (2013a, b) gave comprehensive evaluations of TMPA V7 products over China and continental US against the daily gauge analysis, and found that relative bias and RMSE significantly decreased while correlation increases from V6 to V7. Generally most studies agree that for heavy rainfall, significant underestimation observed in 3B42 V6 is reduced in 3B42 V7 (e.g., Li et al., 2013; Xue et al., 2013; Y. Chen et al., 2013; Chen et al., 2013a, b). The underestimation in heavy rainfall is most severe over higher terrain (e.g., Y. Chen et al., 2013).

The errors in satellite-based precipitation estimates are from two major sources: sampling error and retrieval errors. The former results from estimation the precipitation amount for a continuous spatial and temporal domain with measurements at discrete space and time intervals, such as estimating the daily or monthly total precipitation from instantaneous observations at 3-hour intervals. The sampling error has been studied extensively, and its relationship with rain-rate and spatial/temporal resolution has been well established both empirically and theoretically (e.g., Laughlin, 1981; Huffman, 1997; Bell and Kundu, 1996, 2000, 2003; Bell et al., 2001; Steiner et al., 2003; Nijssen and Lettenmaier, 2004). As well, this part of the errors is beyond the scope of this paper.

The retrieval error arises from the remote-sensing procedures involved to convert satellite observations (brightness temperature) to rain rate. This error type is more complex, because of its dependencies on many factors, including sensor type (conical vs. cross-track, active vs. passive microwave), sensor resolution and viewing geometry, precipitation type, surface type, atmospheric condition, cloud microphysics, and retrieval algorithm itself (e.g., Arkin and Xie, 1994; Sorooshian et al., 2000; Adler et al., 2001; McCollum et al., 2002; McCollum and Ferraro, 2003; Gottschalck et al., 2005; Hossain and Anagnostou, 2006; Ebert et al., 2007; Tian et al., 2007; Tian and Peters-Lidard, 2010; Tian et al., 2010; Kirstetter et al., 2012; Tian et al., 2013; Chen et al., 2013a, b; Maggioni et al., 2014; Tang et al., 2014). For the retrieval error, it can be further decomposed into systematic and random errors. The systematic errors reflect predictable, consistent error behaviors often associated with instrument or algorithm characteristics (e.g., mis-calibration). This can be further decomposed into missed, false, and hit errors (Hossain and Anagnostou, 2006; Tian et al., 2008; Habib et al., 2008b; Maggioni et al., 2014). Assessment for sources of the systematic error helps to gain additional insight into performance accuracy of those satellite-based precipitation estimates (Gebregiorgis et al., 2012; Habib et al., 2012). Meanwhile, it was also revealed that the amplitudes of the three components were often larger than the systematic error, because different error components can cancel each other (Tian et al., 2009). Therefore, it is not enough to evaluate the performance of one satellite precipitation product solely based on its total retrieval error as most existing studies do. The random error is the stochastic component whose magnitude directly determines the uncertainty. Therefore correct error analysis and modeling to separate the different error components are essential to the quantification of uncertainty, which is also the major purpose of this paper.

Meanwhile, most conventional error metrics (e.g., correlation coefficients) are based on the assumption of additive, Gaussian errors. However, this proves to be unrealistic for precipitation estimates which have a non-Gaussian distribution. Habib et al. (2001) demonstrated that the estimation of linear correlation coefficients in precipitation is not accurate due to departures from the normal distribution assumption, and is skewed by extreme precipitation rates. In addition, it has been suggested that some transformation be applied to precipitation prior to computing errors, e.g., Stephenson et al. (1999) used the square root transformation on the extreme daily precipitation events to improve the predictability of ensemble forecasts of the mean Indian rainfall. Tian et al. (2013) demonstrated that the conventional additive error model is not applicable to high-resolution precipitation data, and error evaluation should be based on a multiplicative error model. In this multiplicative error model, a logarithmic transformation was applied to the precipitation estimates. The error model was then evaluated at precipitation logarithmic space.

Considering these new developments and the deficiencies in the conventional approaches, this study systematically proposes and tests a new procedure for the validation of the satellite-based high-resolution precipitation estimates. The new validation procedure has two steps: 1) we employ the error decomposition scheme of Tian et al. (2009) to separate the total error into three independent components: hit error, missed precipitation, and false precipitation; and 2) the hit error is further analyzed based on a multiplicative error model provided by Tian et al. (2013). In the multiplicative error
model, the error features are captured by three model parameters, with each describing a unique aspect of the systematic and random errors. These three parameters can largely replace the variety of conventional error metrics, and produce more accurate quantification of both systematic and random errors.

To demonstrate the effectiveness of the proposed new procedure, we applied it to quantitatively evaluate the recent two versions (Version 6 and 7) of TMPA real-time and post-real-time product suite (3B42RT and 3B42) for six years (2005–2010) over the continental U.S. (CONUS). The gauge-based National Centers for Environmental Prediction (NCEP) Climate Prediction Center (CPC) near-real-time daily precipitation analysis at 0.25° was used as the reference (Chen et al., 2008). This data set uses solely rain gauge measurements from the dense gauge network over CONUS. In addition, the radar-based NCEP Stage IV precipitation data (Lin and Mitchell, 2005; Kitzmiller et al., 2013) were also evaluated with the same procedure as a reference, as we already understand its error characteristics fairly well (e.g., Tian et al., 2007).

The approach used for error decomposition is introduced in Section 2. The multiplicative error model and its three parameters are discussed in Section 3. Section 4 describes the four TMPA data sets, as well as the reference data. Section 5 presents the results on error components, and evaluations of the three parameters in the multiplicative error model. Finally, the results are summarized in Section 6, with recommendations for future improvement of next version TMPA algorithm based on the findings in this paper.

2. Error decomposition

In the first step, we decomposed the total error into three different components using the error decomposition approach derived in Tian et al. (2009) and Habib et al. (2009b). During the retrieval process, the total error (E) in the estimate can be from three possible causes: 1) the sensor detects the precipitation event, but the estimated precipitation amount is different from the truth. This difference is defined as hit bias (H) – the event is detected but the amount is biased, and it can be either negative or positive; 2) a precipitation event is not detected by the sensor at all. This leads to missed precipitation (–M), which is obviously always negative; and 3) the sensor detects some false signature of rain (e.g., Tian and Peters-Lidard, 2007) while there is no precipitation in reality, and the mistakenly measured amount is called false precipitation (F), and is always positive. It is easy to prove (Tian et al., 2009) that the total error can be decomposed as:

\[ E = H - M + F \] (1)

The three components have substantial magnitudes with large spatial and temporal variations. Sometimes, the amplitude of the individual components is larger than that of the total error, simply because the three components may cancel one another. Thus this approach reveals much more of the error characteristics.

3. Analysis of hit errors

In the second step, a multiplicative error model (Tian et al., 2013) is applied to evaluate the hit errors derived from the error decomposition in the first step. Compared with the conventional additive error model, this multiplicative error model can better separate the systematic and random errors (Tian et al., 2013).

Normally, the additive error model of is defined as:

\[ Y_i = a + bX_i + e_i \] (1)

where \( i \) is the index of the datum, in this paper, it refers to a single precipitation measurement at a certain grid box; \( X_i \) is the “truth” reference data, indicates CPC data; \( Y_i \) represents a certain satellite estimate, refers to TMPA products. \( a \) is the offset; \( b \) is a scale parameter to represent the different dynamic range of measurements with the reference data. \( e_i \) represents random error which has zero mean and variance \( \sigma^2 \). Therefore, the error model is defined by three independent parameters, namely, \( a \), \( b \) and \( \sigma \), where \( a \) and \( b \) jointly specify the systematic error and the random error in the measurements \( Y_i \) is quantified by \( \sigma \) (Tian et al., 2013).

The multiplicative error model is usually defined as

\[ Y_i = \alpha X_i e^{\beta} \] (2)

In the multiplicative error model, the systematic error (defined by \( a \) and \( b \)) is now a non-linear function of the reference data. The random error \( e^{\beta} \) is a multiplicative factor, with zero mean and variance \( \sigma^2 \). Note here the values of \( \sigma \) in Eqs. (1) and (2) will be different, indicating that the uncertainty definition depends on the error model formulation. After a logarithmic transformation in Eq. (2), the multiplicative model becomes:

\[ \ln(Y_i) = \alpha + \beta \ln(X_i) + e_i \] (3)

where \( \alpha = \ln a \), \( \beta = b \), this equation is a simple linear regression in the transformed domain, and the three parameters \( (\alpha, \beta \text{ and } \sigma) \) can be easily estimated with the ordinary least squares (OLS) method.

The effect of the three error model parameters is illustrated in Fig. 1. The PDFs of the true rainfall \( (X_i) \), the true rainfall corrupted with systematic error only \( (Y_{true} = aX_i) \), and the one corrupted with both systematic and random errors \( (Y_i = aX_ie^{\beta}) \), are shown in log-scale for, for two different sets of parameter values extracted from Table 1 (yearly 3B42V7 and Stage IV). The offset parameter \( \alpha \) simply shifts the PDF of \( X_i \) to the left (\( \alpha < 0 \)) or right (\( \alpha > 0 \)) in the logarithmic scale of \( X_i \) (for both 3B42V7 and Stage IV, shifting the PDF to the right); while the scale parameter \( \beta \) expands (\( \beta > 1 \)) or shrinks (\( \beta < 1 \)) the PDF of \( X_i \) (for both 3B42V7 and Stage IV, shrinking the PDFs). The addition of the random error given by \( \epsilon \), uniformly broadens the resultant PDF. Therefore \( \alpha \) and \( \beta \) capture the systematic differences in the position (offset) and width (scale) of the probability distribution within the respective range of \( X_i \) and \( Y_i \). This is more representative than simply using “bias” to quantify the systematic error in the additive model.

4. Data

We focused specifically on two TMPA three-hourly microwave-IR merged estimates (Huffman et al., 2007, 2010) – the real-time product known as 3B42RT and the post-real-time with gauge adjustment known as 3B42. Our new validation procedure was applied to on both the previous version (V6) and the newest version (V7) of the two products. Hereafter, we name them 3B42RT-V6 and 3B42RT-V7 (for 3B42RT Version 6 and 7), 3B42V6 and 3B42V7 (for 3B42 Version 6 and 7). The near-real-time production of 3B42RT requires several simplifications compared to 3B42. The main focus of 3B42RT data is timeliness, while 3B42 is strongly recommended for any research work not specifically focused on real-time applications (ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf). Since TMPA V6 was ended on 30 June 2011 (superseded by version 7 after then), we selected a common period of seven years (2005–2011) for this study. Each of the original TMPA data sets is available at spatial and temporal resolution of 3 hourly and 0.25°. We aggregated them into daily resolution, to match the reference data CPC gauge analysis.

The CPC unified gauge analysis was used as the reference data. To verify the effectiveness of the multiplicative error model mentioned in Section 3, we also used the NOAA Next Generation Weather Radar (NEXRAD) Stage IV data (Lin and Mitchell, 2005; Kitzmiller et al., 2013) as another reference (see Section 5.2 on how it works). Note here, the Stage IV data was model-fitted using the CPC data as the reference, to verify the effectiveness of the multiplicative error model we
used. No satellite estimates were validated against Stage IV data in this paper. Based on previous research (Tian et al., 2010), over the eastern CONUS, the differences between the two reference datasets are small, because of the high density of gauge stations and relatively flat terrain. The uncertainties of the CPC and Stage IV data are one order of magnitude lower than the satellite-based precipitation datasets, which will not affect the uncertainty quantification in TMPA datasets. However, over the west CONUS, because of the complex terrain and relatively sparse radar coverage, the uncertainties in Stage IV data are expected to be much larger than the CPC unified gauge analysis, especially in winter (December - February) (Tian et al., 2010). Stage IV data severely underestimate precipitation than the CPC unified gauge analysis, which will result in the discrepancy when model-fitted with our multiplicative error model (Section 3).

5. Results

Fig. 2 shows an example of daily mean precipitations in winter 2006 and summer 2005 for the four TMPA data sets, compared with the ground reference Stage IV and CPC data during the same periods. In winter, the mean precipitation of the real-time products 3B42RT-V6 or 3B42RT-V7 shows obvious overestimation over the Midwest, especially for 3B42RT-V7. Error decomposition analysis (see Section 5 and Fig. 3) indicates this overestimation is dominated by false precipitation (F). After gauge correction, the overestimation is well removed from 3B42V6 and 3B42V7. Meanwhile, all the four TMPA data sets show different levels of underestimation along the coastal area in the Northwest, largely due to the challenging in detecting warm, coastal precipitation.

Table 1
Area-averaged values of the three error model parameters (α, β, and σ) over CONUS for the four TMPA products, for year-around, winter and summer, respectively. The parameter values for Stage IV data are also shown for comparison (last row).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Years</th>
<th>Winters</th>
<th>Summers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>α</td>
<td>β</td>
<td>σ</td>
</tr>
<tr>
<td>3B42RT-V6</td>
<td>1.12</td>
<td>0.40</td>
<td>0.95</td>
</tr>
<tr>
<td>3B42RT-V7</td>
<td>1.08</td>
<td>0.42</td>
<td>0.9</td>
</tr>
<tr>
<td>3B42V6</td>
<td>0.93</td>
<td>0.42</td>
<td>0.86</td>
</tr>
<tr>
<td>3B42V7</td>
<td>1.02</td>
<td>0.47</td>
<td>0.88</td>
</tr>
<tr>
<td>Stage IV</td>
<td>0.37</td>
<td>0.74</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Fig. 1. Illustration of the effects of the three multiplicative error model parameters (α, β, and σ) on the PDF of the 3B42V7 (top panel) and Stage IV (bottom panel). Alpha and beta define the systematic errors (in y0), with the offset parameter (α) shifting the measurements PDF (dashed line) to the left (α < 0) or right (α > 0) relative to the reference PDF (x, solid line), while the scale parameter (β) shrinking (β < 1) or expanding (β > 1) the measurements PDF. When random error defined by σ is included (y, dotted line), it only broadens the PDF (dotted line).
Among the four TMPA data sets, only 3B42RT-V6 could not capture the precipitation pattern in the Southeast. It is consistent with earlier findings (e.g., Yong et al., 2013; Chen et al., 2013a) that V6 products (especially for the real-time products) exhibit significant underestimation of moderate to high rain rates for winter.

In summer, both 3B42RT-V6 and 3B42RT-V7 show obvious overestimation over the Midwest. This is well corrected in the research version data sets 3B42V6 and 3B42V7, which have very similar spatial patterns of precipitation as the Stage IV and CPC data. Not much difference is observed in 3B42V6 and 3B42V7, except that 3B42V7 has slightly higher precipitation in the Southeast, which is in accordance with the ground reference.

These initial comparisons reveal that the newest TMPA V7 research version shows substantial improvements, while the real-time version sees mixed performance gain and degradation. We present results of error analysis of the four TMPA products, using the new validation procedure described above. In Section 5.1, we discuss the spatial characteristics of the error components for each TMPA product for winter and summer seasons, respectively, as well as the distributions of error components as a function of precipitation intensity for both seasons.

In Section 5.2, we present results using the multiplicative error model, and studied the spatial, seasonal and probability distribution of the model parameters. The physical meanings of the model parameters are also illustrated.

5.1. Error decomposition

We decomposed the total error of daily data over the CONUS into the three error components described above. Figs. 3 and 4 show the spatial distributions of total error and its three components for winter and summer seasons, respectively. The seasonal sum of the total bias ($E$) and its three error components: hit bias ($H$), missed precipitation ($−M$), false precipitation ($F$) have the relationship as: $E = H - M + F$. This relationship still holds after averaging over the seasons.

In winter, four TMPA data sets show different spatial patterns of total error and three error components (Fig. 3). For real-time product 3B42RT, both V6 and V7 products show obvious overestimation over the Midwest. The overestimation is largely dominated by the false
precipitation, probably due to the more complex land surface conditions during winter (inaccurate snow-cover mask, frozen soil, etc.). After gauge-correction, false precipitation is well reduced from 3B42V6 and 3B42V7 over the Midwest.

Another significant difference is the underestimation along the west coastal areas and the Rockies. The underestimation is primarily from the missed precipitation, and second from the negative hit bias. Missed precipitation is either the consequence of snow cover on the ground at high latitudes or over the Rockies, or the lack of capability to capture the warm rain process and convective rainfall at low latitudes with passive microwave radiometers (PMW) (Tian et al., 2009). Obviously, the V7 products do not overcome this issue. The negative hit bias also contributes to the underestimation, but is only confined to the west coast.

The large variations of hit bias and missed precipitation over the Northeast and the Southeast, are probably caused by the shortcomings of the inter-calibrated TMPA algorithm. The positive hit bias over the southeast US is an example of the overcompensation in newest TMPA algorithm observed on the V7 products. Meanwhile, the 3B42 product was corrected using the TMPA monthly satellite-gauge product 3B43, this particular gauge-correctly strategy could not make up the missed precipitation over the Northeast.

In summer, the spatial patterns of total errors and error components for all data sets are quite different comparing to its winter counterpart (Fig. 4). Real-time products show quite significant overestimation over the central US. Different from its winter counterpart (overestimation is from false precipitation), the overestimation observed in summer is primarily from the hit bias. The gauge-corrected scheme improves the quality of 3B42 products over this region, especially for 3B42V7. For 3B42V6, however, excessive reduction of precipitation causes negative hit bias around the Great Lakes and the East US.

Both missed precipitation and false precipitation are not very pronounced in summer for all four data sets. Most missed precipitation takes place along the Atlantic, Gulf coast and Florida, which is probably caused by the under-sampling of convective thunderstorms. V7 performs slightly better than V6 for both real-time and post-real-time products.

Fig. 5 shows the distributions of total precipitation and error components as the function of precipitation intensities. In Fig. 5a, distributions of total precipitation show difference among four TMPA data sets for both winter and summer seasons. In winter, when precipitation intensity is less than 10 mm/day, the total precipitation of 3B42RT-V6 and 3B42RT-V7 are close to the reference data. Over 10 mm/day, they go towards the opposite directions, with 3B42RT-V6 underestimates the total precipitation and 3B42RT-V7 overestimates. On the contrary, Gauge-corrected product 3B42V6 constantly underestimates precipitation till around 50 mm/day, while matches well with the reference data for extreme precipitation (>50 mm/day). 3B42V7 underestimates when rain intensity is less than around 25 mm/day but turns out to overestimate afterwards. In summer, the distribution of total precipitation in 3B42V7 aligns the closest to the reference data. Three data sets (3B42RT-V6, 3B42RT-V7, and 3B42V7) are well overlapped with the distribution of the reference data.
when the precipitation intensity is less than 16 mm/day. When over 16 mm/day, the overestimates behave as 3B42RT-V6 > 3B42RT-V7 > 3B42V7. 3B42V6 consistently underestimates moderate precipitation (around 8 ~ 40 mm/day) while well following the reference data on two sides (<8 mm/day and >40 mm/day). Overall, 3B42V7 performs better than the other three products in total precipitation in winter. Due to the spatial average process, this trend is not as obvious in summer. 3B42V6 constantly underestimates and 3B4RT-V7 constantly overestimates total precipitation in both summer and winter. Hit precipitation (Fig. 5b) shows similar distributions as the total precipitation (Fig. 5a), except that all the TMPA data sets underestimate hit precipitation from the low-end of the precipitation intensity. This is reasonable because the satellite-based passive microwave radiometers are less sensitive to the light rain than the ground measurements. There is not much difference in the missed precipitation for the four data sets in both winter and summer (Fig. 5c). In winter, the distributions of false precipitation are slightly separated to two groups with more false precipitation in the real-time products 3B42RT-V6 and 3B42RT-V7 for both winter and summer (Fig. 5d).

The study on spatial distributions of the total bias and three error components shows distinguished error patterns in winter and summer. The winter total bias is dominated by the missed precipitation over the west coastal area and the Rockies, and the false precipitation over large areas in Midwest. The summer total bias is largely from the hit bias in Central US. The study on precipitation intensity distribution demonstrates the improvement of TMPA V7 algorithm. The new version tends to produce more rainfall in the higher rain rates, which moderates the significant underestimation exhibited in the V6 products.

5.2. The multiplicative error model

In this section, the multiplicative error model is used to further quantify the hit bias. The three parameters of the multiplicative error model (Eq. (3) in Section 3) were evaluated for the four TMPA data sets in both winter and summer. The parameters were estimated for only the hit rain events with a threshold of 0.25 mm/day, as the lighter rain events are statistically unreliable for either gauge data or satellite measurements.

Spatial distribution of $\alpha$ is shown for four TMPA data sets in winter and summer, respectively (Fig. 6). The closer $\alpha$ is to zero, the better. From Fig. 6, $\alpha$ is positive across the CONUS for both winter and summer. This is consistent with the results shown in Fig. 5b, from which all the precipitation intensity distributions shift towards the right (larger precipitation intensity) in both winter and summer. For winter, $\alpha$ is increased in V7 products in the Midwest, which is probably caused by the overcompensation of hit precipitation in V7 algorithm (for significant underestimation in V6 algorithm). For summer, large hit biases over the Central US in 3B42RT-V6 and 3B42RT-V7 (shown in Fig. 3) are also observed here (with $\alpha > 1.6$). The V7 algorithm shows effect on decreasing $\alpha$ in 3B42RTV7. After gauge-correction, the values of $\alpha$ are significantly reduced for both 3B42V6 and 3B42V7. We also model-fitted the Stage IV data using the CPC data as the reference, to verify the effectiveness of the multiplicative error model we used. Stage IV data have much smaller $\alpha$ (average 0.30 in winters and 0.45 in summers), compared to larger than 1 in winters and summers for TMPA data sets (Table 1). This is consistent with reality. Note here the gridded Stage IV data behave a bit smoother than the gauge analysis, which also results in the values of model parameters $\alpha$, $\beta$ and $\sigma$. 
Similar to Fig. 6, Fig. 7 presents results on the spatial distribution of shape parameter $\beta$ for the four TMPA data sets for winter and summer, respectively. The perfect situation, beta equals to 1, means an identical dynamic range for the TMPA data sets and the reference data. Generally, four TMPA data sets show very similar spatial patterns of $\beta$ in both winter and summer. For winter, $\beta$ is relatively larger cross the East and in West coastal area, with the largest over the Southeast. This is reasonable as southeast regions share the most various climate patterns with extreme rain events throughout the year. On the contrary, Midwest and the Rocky mountain areas share the lowest $\beta$ values, where liquid precipitation is seldom formed in winter. Dynamic ranges of satellite estimates are small in both regions. For summer, spatial distributions of $\beta$ are quite similar among all the TMPA data sets, with relatively large values in the East and Central US, and small values in the West. For Stage IV data, most areas show a much larger $\beta$ (closer to 1), with average $\beta$ equals to ...
0.75 for winter and 0.71 for summer, compared to around 0.4 for winter and summer for the TMPA data sets (Table 1). This is also consistent with reality.

The standard deviation of the random error (Fig. 8) $\sigma$ is shown for winter and summer, respectively. For winter, real-time products 3B42RT-V6 and 3B42RT-V7 exhibit fairly random errors over the Midwest. Although the random error is obviously reduced in 3B42V6, the precipitation compensation process in V7 algorithm seems to add more random error into 3B42V7 over this region. The Gauge-correction in 3B42V6 and 3B42V7 adds more random error over the West coastal area. Another interesting finding is the random error is smaller along the Rocky Mountain region. This region is covered with snow during the winter, and less variance is observed for liquid precipitation. For summer, the difference in random errors is

---

Fig. 6. Spatial distributions of the scale parameter $\alpha$ in the error model. Results are compared for winters (DJF 2006–2011) and summers (JJA 2005–2010) for TMPA data sets and Stage IV data. The values of $\alpha$ closer to zero means smaller offset errors.
very marked for real-time products and gauge-corrected products. 3B42RT-V6 exhibits significant random error over the Central US and East coastal area. Random error is effectively reduced in 3B42RT-V7. After gauge-correction, random error is notably reduced in 3B42V6 and 3B42V7 over the central US. Stage IV data shows much smaller random error over the CONUS, compared to the TMPA data sets. The average $\sigma$ for Stage IV is 0.48 for winters and 0.60 for summers, while TMPA data sets have an average $\sigma$ around 0.87 for both winter and summer (Table 1).

From our results, it is apparent that the multiplicative error model effectively captured the error characteristics in both the satellite-based and ground radar-based datasets, and their error characteristics are succinctly summarized in the three model parameters.
6. Conclusions and discussions

In this paper, we proposed a new procedure for the quantitative validation of high-resolution satellite-based precipitation estimates. To test the procedure, a comprehensive error analysis was performed for the newest version 7 TMPA products and its predecessor version 6 products using daily 0.25° CPC data over the CONUS as the reference for a period of six years (2005–2011). The new procedure was separated into two steps: first, an error decomposition process (Tian et al., 2009; Habib et al., 2009b) to decompose the total error into three difference error components, hit bias, missed precipitation and false precipitation; then a multiplicative error model (Tian et al., 2013) is employed to further analyze the hit error. The effectiveness of the new procedure is verified by applying it to the

Fig. 8. Same as Fig. 6, except for the random error parameter sigma, which represents the standard deviation of the random error (unit: mm/day).
well-understood radar-based Stage IV data. The major findings are summarized as follows:

(1) The new procedure proposed in this paper is effective in quantiﬁxing both the systematic errors and random errors. It overcomes many of the deﬁciencies in the conventional approaches (e.g., cancellation among error components, skewed error metrics from Gaussian error assumption, etc.), and results in more accurate description, separation and quantiﬁcation of the error components, and of the systematic and random errors. This leads to more accurate assessment of the uncertainty in the data. Therefore, we recommend the new procedure be adopted for routine validation of satellite-based precipitation data sets.

(2) With the new procedure, we performed analysis of spatial distributions of the total bias and three error components (Figs. 3 and 4) and examined their error patterns in winter and summer for four TMPA data sets. The winter total bias is dominated by the missed precipitation over the west coastal area and the Rocky Mountains, and the false precipitation over the large areas in Midwest. The summer total bias is largely from the hit bias in central United States.

(3) Analysis of precipitation intensity distribution (Fig. 5) demonstrates the improvement of TMPA V7 algorithm. The new version tends to produce more rainfall in the higher rain rates, which moderates the signiﬁcant underestimation exhibited in the previous V6 products.

(4) All the three parameters in the multiplicative error model, work together, are effective in evaluating both the systematic and random error in the hit error in the four studied TMPA products.

As mentioned earlier, TMPA V7 products are still quite recent, and not many investigations have been conducted to evaluate their systematic and random error characteristics compared to their V6 predecessors, especially for potential use in various hydrological applications. In this paper, we presented a comprehensive error analysis on Version 7 TMPA real-time and post-real-time products 3B42 and 3B42RT, in an effort to draw a consistent conclusion on the advancement of TMPA V7 algorithm. Meanwhile, this analysis provides input to the generation of the next version TMPA algorithm by constructing the characteristics of both the systematic and random errors.

The same validation procedure can be applied to other areas, with other satellite-based datasets, such as CMORPH [Joyce et al., 2004], or the upcoming IMERG [http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.1.pdf]. Moreover, the proposed validation procedure can also be used on quantitative precipitation forecasted from models.

Acknowledgements

This research was supported by the NASA Earth System Data Records Uncertainty Analysis Program (Martha E. Maiden) under solicitation NNH10ZDA001N-ESDERR. Computing resources were provided by the NASA Center for Climates Simulation. We appreciate the very helpful suggestions from two anonymous reviewers.

References


Huffman, G.J., Stocker, E.F., Huffman, G.J., Tan, Y., Qian, Y., Hong, Y., Hong, G., Mitchell, D., 2012. Comparison of TRMM 3B42 and 3B42RT, in an effort to draw a consistent conclusion on the advancement of TMPA V7 algorithm. The new version tends to produce more rainfall in the higher rain rates, which moderates the significant underestimation exhibited in the previous V6 products.

(4) All the three parameters in the multiplicative error model, work together, are effective in evaluating both the systematic and random error in the hit error in the four studied TMPA products.

As mentioned earlier, TMPA V7 products are still quite recent, and not many investigations have been conducted to evaluate their systematic and random error characteristics compared to their V6 predecessors, especially for potential use in various hydrological applications. In this paper, we presented a comprehensive error analysis on Version 7 TMPA real-time and post-real-time products 3B42 and 3B42RT, in an effort to draw a consistent conclusion on the advancement of TMPA V7 algorithm. Meanwhile, this analysis provides input to the generation of the next version TMPA algorithm by constructing the characteristics of both the systematic and random errors.

The same validation procedure can be applied to other areas, with other satellite-based datasets, such as CMORPH [Joyce et al., 2004], or the upcoming IMERG [http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.1.pdf]. Moreover, the proposed validation procedure can also be used on quantitative precipitation forecasted from models.

Acknowledgements

This research was supported by the NASA Earth System Data Records Uncertainty Analysis Program (Martha E. Maiden) under solicitation NNH10ZDA001N-ESDERR. Computing resources were provided by the NASA Center for Climates Simulation. We appreciate the very helpful suggestions from two anonymous reviewers.

References


Chen, Y., Xie, P., et al., 2008. CPC uniﬁed gauge-based analysis of global daily precipi-


