Evaluation of satellite-retrieved extreme precipitation rates across the central United States

A. AghaKouchak,[1] A. Behrangi,[2] S. Sorooshian,[1] K. Hsu,[1] and E. Amitai[3,4]

Received 10 July 2010; revised 10 September 2010; accepted 27 October 2010; published 26 January 2011.

[1] Water resources management, forecasting, and decision making require reliable estimates of precipitation. Extreme precipitation events are of particular importance because of their severe impact on the economy, the environment, and the society. In recent years, the emergence of various satellite-retrieved precipitation products with high spatial resolutions and global coverage have resulted in new sources of uninterrupted precipitation estimates. However, satellite-based estimates are not well integrated into operational and decision-making applications because of a lack of information regarding the associated uncertainties and reliability of these products. In this study, four satellite-derived precipitation products (CMORPH, PERSIANN, TMPA-RT, and TMPA-V6) are evaluated with respect to their performance in capturing precipitation extremes. The Stage IV (radar-based, gauge-adjusted) precipitation estimates are used as reference data. The results show that with respect to the probability of detecting extremes and the volume of correctly identified precipitation, CMORPH and PERSIANN data sets lead to better estimates. However, their false alarm ratio and volume are higher than those of TMPA-RT and TMPA-V6. Overall, no single precipitation product can be considered ideal for detecting extreme events. In fact, all precipitation products tend to miss a significant volume of rainfall. With respect to verification metrics used in this study, the performance of all satellite products tended to worsen as the choice of extreme precipitation threshold increased. The analyses suggest that extensive efforts are necessary to develop algorithms that can capture extremes more reliably.

Citation: AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai (2011), Evaluation of satellite-retrieved extreme precipitation rates across the central United States, *J. Geophys. Res.*, *116*, D02115, doi:10.1029/2010JD014741.

1. Introduction

[2] Precipitation plays a significant role in weather research, monitoring, and predictions. Improving our understanding of weather and climate, along with the development of reliable and uninterrupted measurements, are essential for proper assessment of weather conditions. Currently, in situ and radar-based precipitation data are the major input for streamflow forecasts, flash flood warnings, and weather watches across the United States. While some regions have long-term historical in situ precipitation measurements, poor spatial sampling makes the data inadequate to support monitoring, detection, and forecast studies. On the other hand, in most parts of the globe (except in a few developed countries), radar installations for precipitation measurements are not available. In the United States, with one of the most sophisticated radar measurement networks in the world, regions with extensive topographic relief (e.g., the western and southwestern United States) suffer from poor or non-existent radar coverage [*Maddox et al.*, 2002]. In fact, *Maddox et al.* [2002] showed that at lower levels (e.g., 1000 m above ground level), which are closer estimates to ground-level precipitation, the radar coverage area is substantially smaller than at higher levels (e.g., 3000 m above ground level).

[3] Clearly, the lack or absence of ground-based precipitation networks hampers the development and use of flood and drought warning models, hydrological models, and extreme weather monitoring and decision-making systems. Therefore, there exists the need to achieve alternative estimates of precipitation with sufficient sampling density, reliability, and accuracy to enable utilization of data for operational applications. Satellite-derived precipitation estimates have the potential to improve precipitation observation at a global scale. In recent years, the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), and many other international sponsored satellite missions have led to an increase in available precipitation data. These remotely sensed data have several advantages over in situ measurements, including

[1]Department of Civil and Environmental Engineering, University of California, Irvine, California, USA.
[2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA.
[3]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[4]School of Earth and Environmental Sciences, Chapman University, Orange, California, USA.

Copyright 2011 by the American Geophysical Union.
0148-0227/11/2010JD014741

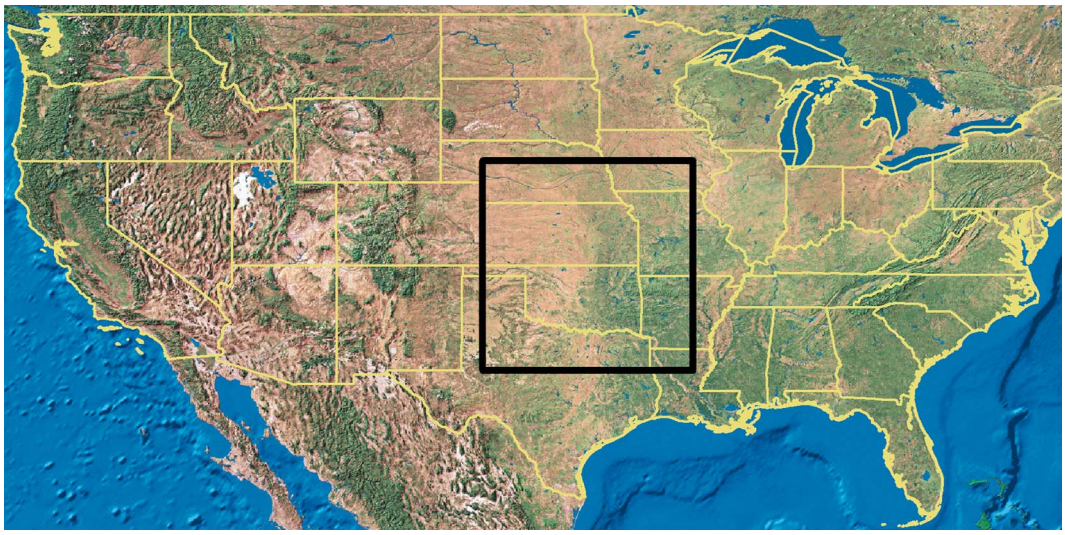

Figure 1. The study area (partly includes states of Texas, Oklahoma, Kansas, Nebraska, Iowa, Missouri, Arkansas, and Louisiana).

higher spatial resolution and uninterrupted coverage. However, these data have not yet been fully integrated into hydrologic and water resources management and decision-making systems, mainly because of undetermined uncertainties associated with satellite data.

[4] Currently, there are several satellite-retrieved precipitation products available in near real time [e.g., CMORPH, *Joyce et al.*, 2004; TRMM 3B42, *Huffman et al.*, 2007; PERSIANN, *Sorooshian et al.*, 2000]. A complete review of satellite rainfall estimation algorithms is beyond the scope of this document. For comprehensive reviews and product comparisons, the interested reader is referred to *Tian et al.* [2009], *Adler et al.* [2001], *Levizzani and Amorati* [2002], *Tian et al.* [2007], and *Kidd* [2001]. Verification of different satellite products is essential for future rainfall retrieval algorithm development, scientific advancements, and integration of data into practical applications. So far, a myriad of studies have been devoted to validation and verification of satellite-based data with respect to ground-based observations (e.g., *Ebert et al.* [2007], *Turk et al.* [2008], *Tuttle et al.* [2008], *Tian et al.* [2009], *Stisen and Sandholt* [2010], *Amitai et al.* [2009], *Durga Rao et al.* [2009], *Zhou et al.* [2008], *Gochis et al.* [2009], *Yilmaz et al.* [2005], *Shen et al.* [2010], *Zeweldi and Gebremichael* [2009], *Feidas et al.* [2009], *Dinku et al.* [2008], *Liu et al.* [2009], *Sapiano and Arkin* [2009]). This study aims to evaluate different satellite products with respect to extremes over a vast area across the southern Great Plains (SGP). Four satellite-retrieved products are investigated with respect to the Stage IV multisensor ground radar-based, gauge-adjusted data. For different extreme value thresholds, quantitative statistics, including probability of detection, false alarm ratio, volumetric probabilities, and areal bias, among others, are computed and used as measures of comparison. Furthermore, the mean error maps with respect to Stage IV data are provided for the satellite precipitation products.

[5] The paper is organized into four sections. After the introduction, the study area and data resources are briefly introduced. The third section is devoted to methodology and evaluation of satellite precipitation products. Section 4 summarizes the results and offers some concluding remarks.

2. Study Area and Data Resources

[6] The study area encompasses most of the SGP. Figure 1 displays the boundaries of the study area, which partly includes the states of Texas, Oklahoma, Kansas, Nebraska, Iowa, Missouri, Arkansas, and Louisiana. The climate condition ranges from humid continental and subhumid subtropical to semiarid steppe climate from east to west. The Great Plains climate is characterized by a strong east-west precipitation gradient [*National Assessment Synthesis Team*, 2001]. The annual precipitation ranges from more than 1300 mm in the southeast to around 450 mm on the western edge of the study area. Extreme precipitation, hail storms, blizzards, floods, droughts, and tornadoes, among other events, are relatively common in the study area.

[7] In this work, the Stage IV radar-based multisensor precipitation estimates (MPE) available from the National Center for Environmental Prediction (NCEP) are used as reference data. The Stage IV (hereafter STIV) data are generated in near real time by aggregating the radar-based precipitation data from the National Weather Service (NWS) River Forecast Centers (RFC) collected over the entire continental United States. The STIV data are available with a temporal resolution of hourly spatial resolution of about 4km on the Hydrologic Rainfall Analysis Project (HRAP) national grid system. The data include rain rates from merged operational radar estimates and rain gauge measurements. It is well known that radar-based precipitation estimates are associated with several types of uncertainties that arise from various factors such as partial beam filling, anomalous propagation, beam overshooting, instrumental errors, non-uniformity in vertical profiles of reflectivity (VPR), inappropriate $Z - R$ relationship, spatial sampling pattern, hardware calibration, and random sampling error [*Seed and Srikanthan*, 1999; *Krajewski and Smith*, 2002]. Additionally, the weather condition may also affect radar rainfall observations [*Steiner and Smith*, 2000]. So far, a great deal of

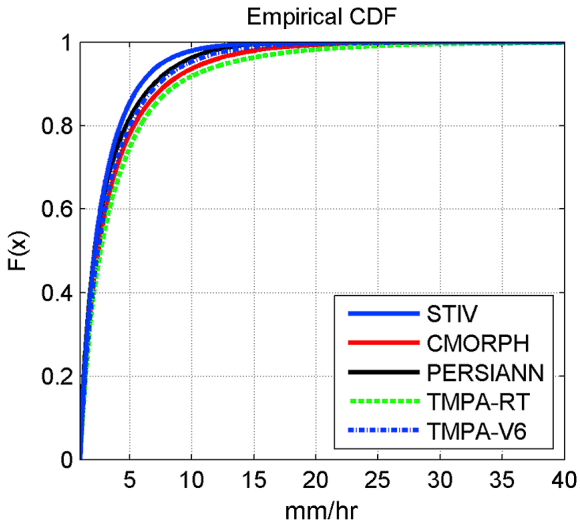

Figure 2. The empirical cumulative distribution functions of the data.

effort has been made to quantify and/or adjust or describe the associated uncertainties [*Seo*, 1999; *AghaKouchak et al.*, 2010a; *Ciach et al.*, 2007; *AghaKouchak et al.*, 2010b]. The STIV data are calibrated and adjusted for different biases using automated rain gauge measurements following various quality control measures to improve the quality [*Lin and Mitchell*, 2005]. The STIV are used as reference data, since the data are routinely quality controlled and are currently being used for forecasts and operational applications.

[8] The satellite-retrieved data used in verification analyses include the following products:

[9] 1. The Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) real time (hereafter TMPA-RT) and TMPA Version 6 (hereafter TMPA-V6) products. The TMPA algorithm collects microwave-based estimates of rainfall from low orbiter satellites and fills the remaining gaps in brackets of 3 hours by infrared (IR)-based rain estimates [*Huffman et al.*, 2007]. Unlike TMPA-RT, TMPA-V6 is an adjusted product which combines precipitation estimates from multiple satellites and gauge analyses [*Huffman et al.*, 2007]. In this algorithm, all precipitation estimates are summed over a calendar month. The bias is then removed using monthly climatological data as described previously [*Huffman et al.*, 2007].

[10] 2. CMORPH (CPC MORPHing technique, *Joyce et al.* [2004]) offers precipitation data derived from available microwave observations (low orbiter satellite) whose features are advected in space and time using IR images from geostationary satellites. In fact, this algorithm utilizes IR data to transport the observed microwave-based precipitation estimates when microwave data are not available. Using a time-weighting interpolation between microwave scans, the intensity and shape of the precipitation estimates are computed in the intervening half-hour periods. The algorithm propagates precipitation features forward and backward in time from the previous and following microwave scans. In this algorithm, any precipitation data from any low orbiter satellite source can be incorporated in retrieving precipitation.

[11] 3. Precipitation Estimation From Remotely Sensed Information Using Artificial Neural Networks (PERSIANN, *Sorooshian et al.* [2000]), which uses grid IR images of the global geosynchronous satellites provided by Climate Prediction Center (CPC), NOAA [*Janowiak et al.*, 2001], as the main source of information. Precipitation estimation from IR data relies on statistical relationships between IR estimates of cloud-top brightness temperature and mean precipitation rate. However, such statistical relationships are associated with high uncertainties because of variability in many factors, including cloud properties (e.g., type, height, thickness) and atmospheric conditions. In this algorithm, using a neural network classification and/or approximation procedures, IR-based estimates are calibrated and adjusted on the basis of microwave data from low-orbital satellites (e.g., TRMM Microwave Imager (TMI) aboard TRMM, Special Sensor Microwave Imager (SSM/I) on Defense Meteorological Satellite Program (DMSP), Advanced Microwave Scanning Radiometer-Earth observing system (AMSR-E) on Aqua spacecraft, and the Advanced Microwave Sounding Unit-B (AMSU-B) aboard the National Oceanic and Atmospheric Administration (NOAA) satellite series). An adaptive training technique [*Hsu et al.*, 1997] updates the neural network parameters whenever microwave data are available (approximately every 3 hours).

[12] The above satellite products are all available with a common spatial resolution of 0.25° and a temporal resolution of 3-hourly. The STIV data are aggregated in space and time into 0.25° spatial grids and 3-hourly accumulations to synchronize STIV data with satellite estimates. The aggregation is performed by averaging precipitation rates from the STIV pixels within each satellite pixel across the area of the satellite pixel and temporally over a window of 3 hours. Four years of 3-hourly data (units in mm/h) from January 2005 to December 2008 are processed and used for the analysis. Figure 2 plots the empirical cumulative distribution functions (CDF) of the data, whereas Table 1 lists summary statistics of precipitation accumulations for the entire study area during the period January 2005–December 2008. The mean and standard deviations are derived on the basis of rainfall rates equal to or greater than 1 mm/h. In this table, the bias is defined as the ratio of total satellite over total reference measurements (STIV). Unlike TMPA-V6, the other products (CMORPH, PERSIANN, and TMPA-RT) are unadjusted to gauge data and exhibit larger bias with respect to STIV (see last column in Table 1). Furthermore, Table 1 indicates that for larger quantiles, the discrepancies between satellite products and STIV data become more significant (compare Q5, Q10, Q25, and Q50 with Q75, Q90, and Q95 in Table 1, Q: quantile).

3. Analysis and Results

[13] As mentioned above, the focus of this study is on verification of satellite products with respect to extremes.

Table 1. Summary Statistics of Rainfall Accumulations (mm/h) for the Entire Study Area[a]

Product	Mean	SD	Q10	Q25	Q50	Q75	Q90	Q95	Bias
Stage IV	3.0	2.4	1.2	1.4	2.2	3.7	5.9	7.7	-
CMORPH	3.8	3.6	1.2	1.5	2.5	4.5	8.1	11.1	1.50
PERSIANN	3.2	2.7	1.1	1.4	2.2	4.0	7.0	9.2	1.43
TMPA-RT	4.2	4.6	1.2	1.6	2.7	5.0	8.9	13.0	1.56
TMPA-V6	3.5	3.1	1.2	1.5	2.4	4.3	7.3	9.7	1.02

[a]Q, quantile; SD, standard deviation.

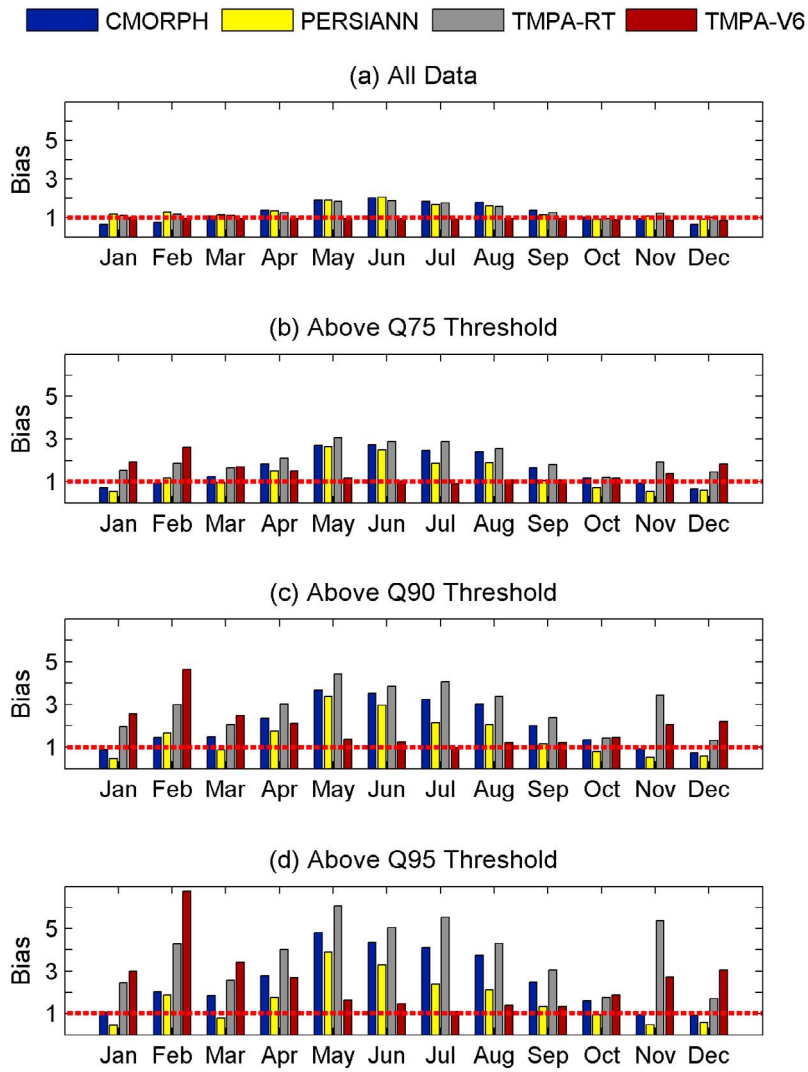

Figure 3. Monthly bias with respect to (a) all data and (b) Q75, (c) Q90, and (d) Q95 thresholds of STIV data (Q, quantile).

Various statistical and graphical methods are used to assess extremes of precipitation observed for different products.

3.1. Monthly Quantile Bias

[14] The monthly quantile bias (*MQB*) is defined as the monthly ratio of satellite-based precipitation accumulations over reference precipitation accumulations above a given threshold:

$$MQB = \frac{\sum_{i=1}^{n}(P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq t)}{\sum_{i=1}^{n}(P_{\mathrm{ref}}|P_{\mathrm{ref}} \geq t)} \quad (1)$$

where P_{sat} are satellite estimates, P_{ref} are reference measurements (e.g., STIV data), t is extreme threshold (i.e., 75, 90, and 95 percentiles of rain rates), and n is number of exceedances.

[15] Notice that throughout this paper, the index i is not shown for P_{sat} and P_{sat} to simplify the equations ($P_{\mathrm{sat}} = P_{\mathrm{sat}_i}$ and $P_{\mathrm{ref}} = P_{\mathrm{ref}_i}$). The *MQB* value of 1 corresponds to no bias in the estimated data. The *MQB* values are computed for the introduced precipitation products with respect to Q75, Q90, and Q95 of STIV data. Figure 3a shows the monthly quantile bias values when the entire data are included in analysis. As shown, CMORPH, PERSIANN, and TMPA-RT exhibit overestimation (*MQB* > 1) over warm months (i.e., May, June, July, and August), while their bias values are around 1 during the rest of the year. The TMPA-V6 data, on the other hand, show no bias throughout the year owing to adjustment applied to the data. As the threshold increases, the *MQB* values consistently increase over warm months, even up to two to three times (compare Figures 3a and 3d). Throughout fall and winter, the bias of CMORPH and PERSIANN either increases or reduces with an increase in the threshold. It is noted that for higher thresholds, the bias values of TMPA-V6 do not change considerably over warm months (i.e., May, June, July, and August), while they significantly increase during fall and winter. In Figure 3, the thresholds are constant values based on the quantiles of the STIV data.

[16] In recent years, many adjustment techniques have been utilized to improve satellite products (e.g., adjusting the probability distribution, removing overall bias; see *Yilmaz*

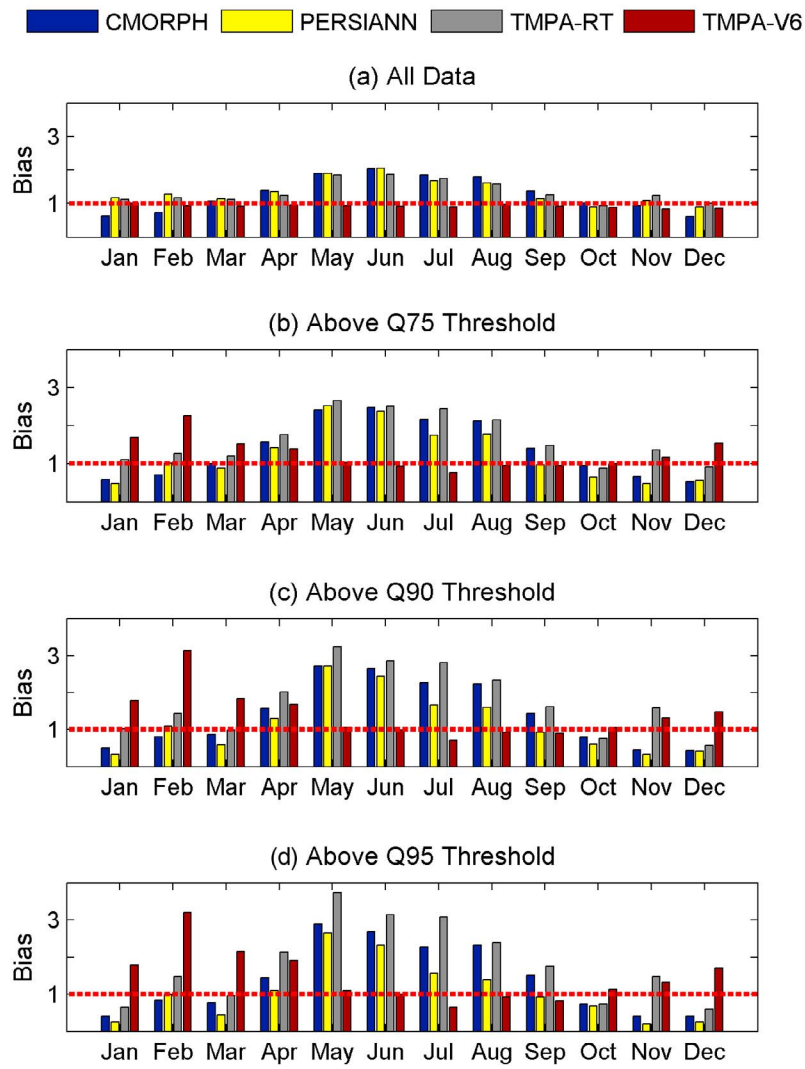

Figure 4. Monthly bias with respect to (a) all data and (b) Q75, (c) Q90, and (d) Q95 thresholds (quantiles are estimated for each product separately).

et al. [2005] and *Huffman et al.* [2007] for examples). Therefore, one may be interested in evaluating the performance of the precipitation products independent of their marginals. In the following, all precipitation data are made uniform and the analyses, mentioned above, are repeated. The *MQB* values are then back-transformed to the original marginals. This means that values above a certain quantile of a satellite product are evaluated with respect to values above a certain threshold of STIV as opposed to comparing values above a fixed threshold of STIV. Figure 4 displays the *MQB* values with respect to relative quantiles of the precipitation products (e.g., Q90 of STIV versus Q90 of CMORPH). One can see that bias values independent of data marginals exhibit similar behavior to that of Figure 3 with less variability in magnitudes of bias (compare Figures 3d and 4d). Both Figures 3 and 4 indicate that satellite estimates significantly overestimate the volumes of extreme precipitation with respect to ground reference data, particularly during warm months. In subsequent analysis, the results are shown with respect to quantiles of STIV data.

3.2. Mean Quantile Error

[17] The mean quantile error (*MQE*) is defined as the mean of differences in precipitation estimates above a given threshold:

$$MQE = \sum_{i=1}^{n} ((P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq t) - (P_{\mathrm{ref}}|P_{\mathrm{ref}} \geq t))/n, \qquad (2)$$

where P_{sat}, P_{ref}, t, and n are as defined in equation (1). The values of mean error for CMORPH, PERSIANN, TMPA-RT, and TMPA-V6 with respect to Q75, Q90, and Q95 of STIV data are presented in Figures 5a–5c, 5d–5f, 5g–5i, and 5j–5l, respectively. One can see that as the threshold increases, CMORPH and TMPA-RT tend toward more considerable overestimation of extremes, while PERSIANN and TMPA-V6 data show both overestimation and underestimation across the study area (compare Figures 5a–5c with Figures 5d–5f). For the same extreme thresholds, Figure 6 displays monthly *MQE* values. Figure 6 confirms the findings of Figure 3 and provides mean values of error throughout the year. Figure 6 indicates that

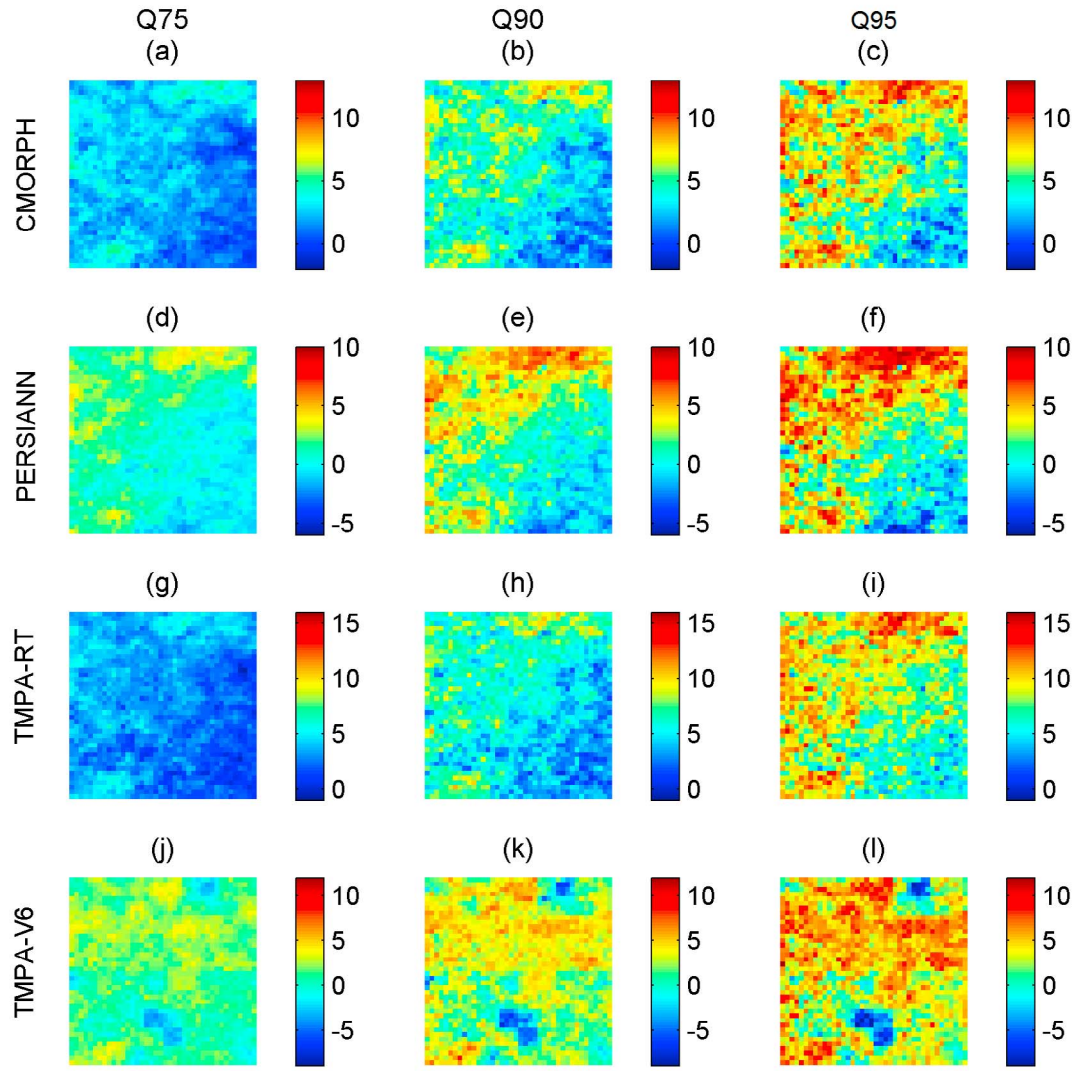

Figure 5. Mean quantile error (*MQE* [mm/h]) with respect to (a, d, g, j) Q75, (b, e, h, k) Q90, and (c, f, i, l) Q95 of STIV data (Q, quantile). Color bars are not identical.

TMPA-RT and PERSIANN exhibit the highest and lowest mean errors, respectively.

3.3. Quantile Probability of Detection

[18] Quantile probability of detection (*QPOD*) is defined as the probability of detection (*POD*; see *Wilks* [2006]) above a certain threshold:

$$QPOD = \frac{\sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{sat}} | P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t)}{\sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{sat}} | P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t) + \sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{ref}} | P_{\mathrm{sat}} < tP_{\mathrm{ref}} \geq t)} \tag{3}$$

where **I** is the indicator function. The *QPOD* represents the ratio of the number of correct identifications of precipitation above a given threshold to the total number of precipitation occurrences above the same threshold as indicated by the reference. The *QPOD* ranges from 0 to 1, with 1 being the perfect *QPOD*. Figure 7 demonstrates *QPOD* values over Q75, Q90, and Q95 of STIV data. For all products, the probability of detection reduces as the threshold increases. Among the precipitation data sets, CMORPH shows the least change in *QPOD* with respect to the choice of threshold, while *QPOD* values of TMPA-RT and TMPA-V6 drop significantly as the threshold increases. Figure 7 indicates that the probability of detection is higher over warm months for all precipitation products, regardless of the choice of threshold. This is consistent with the findings of *Behrangi et al.* [2011].

3.4. Quantile False Alarm Ratio

[19] Quintile false extreme alarm ratio (*QFAR*) is defined as the false alarm ratio (*FAR*; see *Wilks* [2006]) above a certain threshold:

$$QFAR = \frac{\sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{sat}} | P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} < t)}{\sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{sat}} | P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t) + \sum_{i=1}^{n} \mathbf{I}(P_{\mathrm{ref}} | P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} < t)} \tag{4}$$

[20] The *QFAR* represents the ratio of the number of false identifications of precipitation above a given threshold to

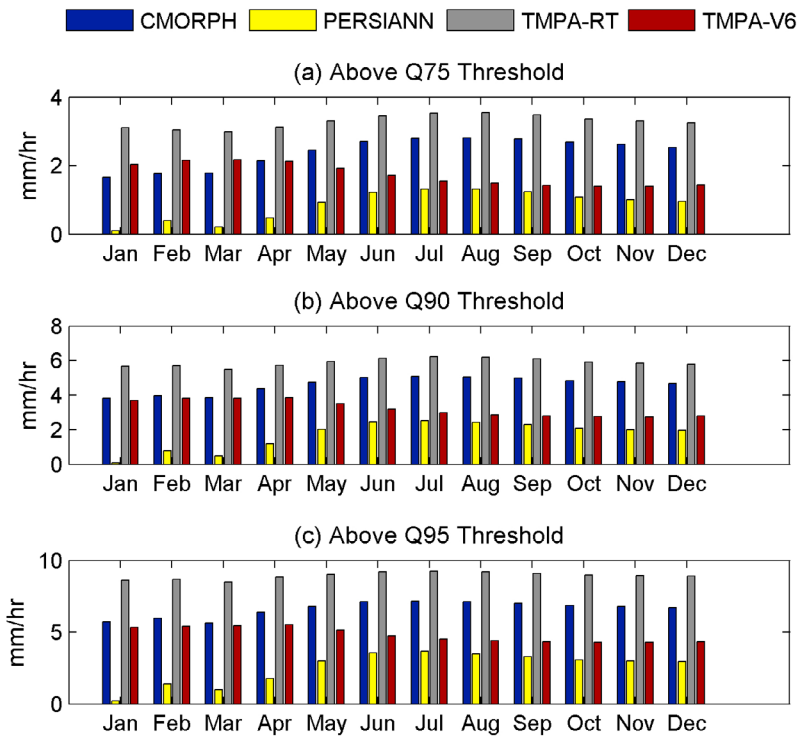

Figure 6. Monthly mean quantile error (*MQE* (mm/h)) with respect to (a) Q75, (b) Q90, and (c) Q95 of STIV data.

the total number of correct and false occurrences over the same threshold as indicated by the reference. The *QFAR* ranges from 0 to 1, with 0 being the perfect *QFAR*. As shown in Figure 8a, all products exhibit higher false alarm during cold months (November, December, January, and February). Overall, PERSIANN and CMORPH estimates are subject to higher false alarm than TMPA-RT and TMPA-V6. As the threshold increases, the *QFAR* values of TMPA-RT and TMPA-V diminish, while *QFAR* values of PERSIANN and CMORPH slightly increase (e.g., see Figure 8b). Furthermore, for higher thresholds (Q75, Q90, and Q95), the false alarm ratios are higher over warm months than the findings given in Figure 8a (all data). This implies that with respect to higher quantiles (extremes), the false alarm ratios are higher over warm months.

3.5. Volumetric Hit Index

[21] In addition to probability of detection, one may be interested in the volume of precipitation detected correctly. The Volumetric Hit Index (*VHI*), defined below, indicates the percentage of the volume of precipitation above a certain threshold being detected correctly:

$$VHI = \frac{\sum_{i=1}^{n}(P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t)}{\sum_{i=1}^{n}(P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t) + \sum_{i=1}^{n}(P_{\mathrm{ref}}|P_{\mathrm{sat}} < tP_{\mathrm{ref}} \geq t)} \times 100. \tag{5}$$

[22] The *VHI* ranges from 0 to 100, with 100 being the perfect *VHI*. Figure 9a presents the *VHI* when all precipitation data are included. Figures 9b, 9c, and 9d present the *VHI* values for Q75, Q90, and Q95 of STIV data, respectively. Figures 9b–9d indicate that considering the entire distribution (Figure 9a), the *VHI* values are higher over warm months for all precipitation products and choices of threshold. It is also noted that the *VHI* values decrease as the threshold increases. Comparing the *VHI* values for all precipitation products reveals that CMORPH and PERSIANN exhibit the highest *VHI*.

3.6. Volumetric False Alarm Ratio

[23] Volumetric false alarm ratio (*VFAR*) represents the percentage of the volume of precipitation above a certain threshold being identified falsely:

$$VFAR = \frac{\sum_{i=1}^{n}(P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} < t)}{\sum_{i=1}^{n}(P_{\mathrm{sat}}|P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} \geq t) + \sum_{i=1}^{n}(P_{\mathrm{ref}}|P_{\mathrm{sat}} \geq tP_{\mathrm{ref}} < t)} \tag{6}$$

[24] The *VFAR* ranges from 0 to 100, with 0 being the perfect *VFAR*. Figure 10 displays *VFAR* for the selected products. One can see that the *VFAR* values increase significantly when considering high thresholds (compare Figures 10a and 10b). Figure 10 shows that PERSIANN and CMORPH exhibit higher *VFAR* compared to TMPA-RT and TMPA-V6. The same behavior was observed with respect to *VHI*.

4. Discussion and Concluding Remarks

[25] Reliable management and decision making in the field of water resources requires accurate measurements of weather variables. Reliable estimation of precipitation is essential for practical applications, since it is believed that precipitation uncertainties will propagate into land surface and

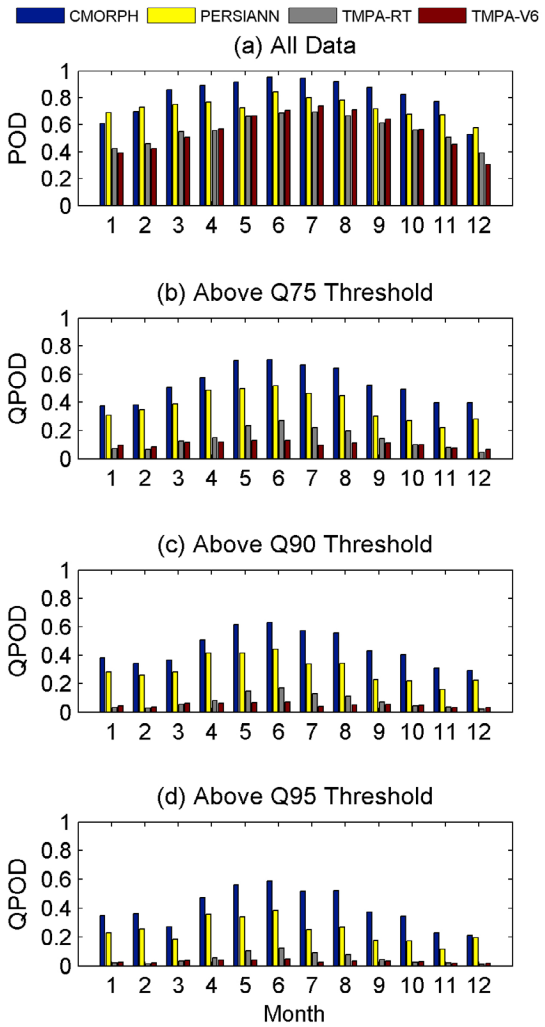

Figure 7. Probability of detection (*POD*) for (a) all data and (b) above Q75, (c) above Q90, and (d) above Q95 thresholds (*QPOD*, quantile probability of detection).

hydrologic modeling predictions [*Nijssen and Lettenmaier*, 2004]. Hydrologists have long relied on in situ data as the main source of precipitation measurements. However, in situ data yield poor spatial coverage and lack of areal representation over land, particularly over the oceans. On the other hand, the need for high-resolution data in space and time is well recognized for regional and global scale land-surface, hydrology, and climate prediction and forecasting studies. The emergence of various satellite-based precipitation products with high spatial resolution and global coverage could be considered good alternatives to in situ measurements. However, because of lack of information about the associated uncertainties and reliability of these products, they are not well integrated into operational and decision-making applications.

[26] In this study, several satellite-derived precipitation products (CMORPH, PERSIANN, TMPA-RT, and TMPA-V6) are evaluated with respect to STIV (radar-based and gauge-adjusted) precipitation data. On the basis of the results of this study, the following conclusions and remarks can be made.

[27] 1. The unadjusted precipitation products (CMORPH, PERSIANN, and TMPA-RT) tend to overestimate precipitation estimates, particularly over warm months. For higher thresholds of reference data (intense precipitation rates), the overestimation of all satellite products increase significantly.

[28] 2. While the adjusted TMPA-V6 exhibits a low bias during warm months (even over high thresholds), it significantly overestimates high precipitation rates (above high thresholds of Q75, Q90, and Q95) during cold months. Analyzing bias as well as mean error of precipitation products independently of their marginals (uniform precipitation data) shows that both bias and mean error can be reduced if the distribution functions of products are adjusted with respect to the reference data. It is emphasized that we do not intend to introduce an adjustment technique in this study, and neither do we claim that applying the distribution function of observations is sufficient to improve the products with respect to their bias. Further in-depth research on adjustment methods and bias removal techniques is required to improve the precipitation products.

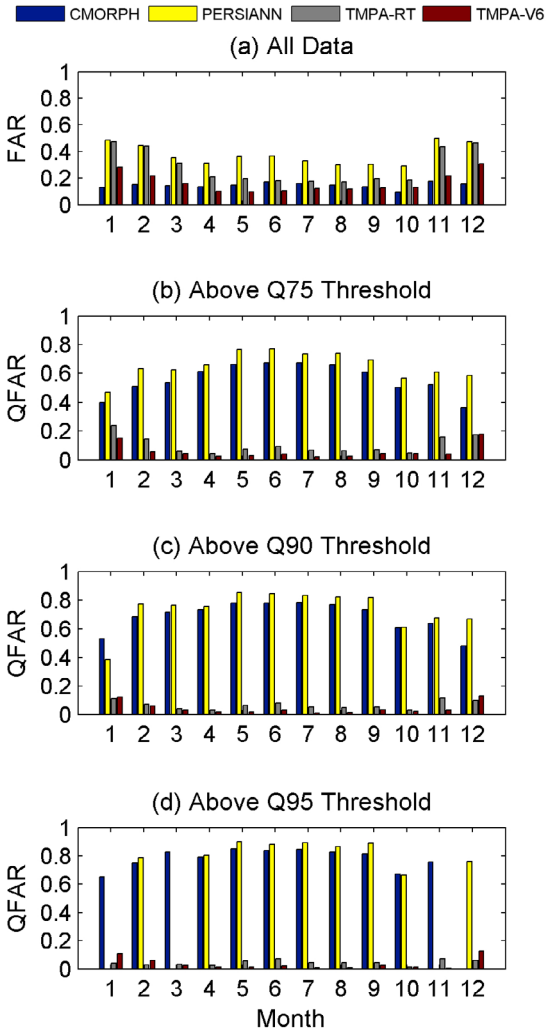

Figure 8. False alarm ratio (*FAR*) for (a) all data and (b) above Q75, (c) above Q90, and (d) above Q95 thresholds (*QFAR*, quantile false alarm ratio).

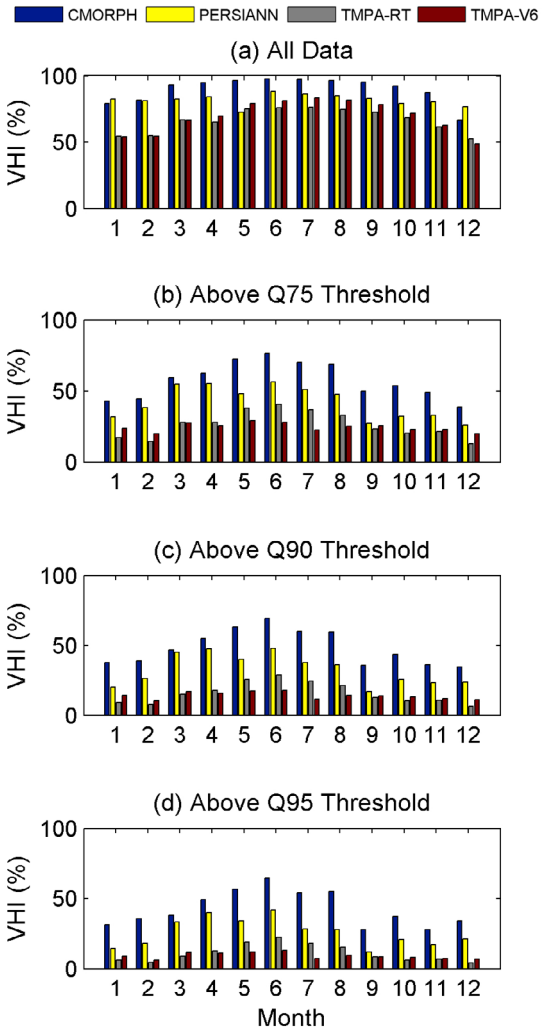

Figure 9. Volume of hit index (*VHI* (%)) with respect to the observed precipitation: (a) all data and (b) above Q75, (c) above Q90, and (d) above Q95 thresholds of STIV data.

[29] 3. The presented results underline the seasonal variability of precipitation algorithms' skills in detecting precipitation rates. Figure 11 summarizes products' skills with respect to different metrics during warm (left) and cold (right) months. Notice that $1 - FAR$, $1 - VFAR$, and $1 - VMI$ are plotted instead of *FAR*, *VFAR*, and *VMI*, so that the best score of all indicators is 1. Figure 11 is created by computing the above-mentioned statistics on the basis of all warm (cold) month data in the period of analysis (2005–2008). Figure 11 clearly highlights the performance of algorithms in different climate conditions. Arguably, further analysis of error sources in conjunction with more sophisticated, self-adaptive, climate-based algorithms should be developed to improve precipitation retrieval techniques.

[30] 4. Among the unadjusted products, PERSIANN precipitation estimates consistently exhibit less bias, regardless of the choice of threshold. On the other hand, TMPA-RT estimates show higher bias over all thresholds.

[31] 5. The reported probability of detection values indicate that all satellite products are more likely to detect precipitation occurrence over warm months. Considering *QPOD* values above high thresholds, CMORPH and PERSIANN seem to perform better in detecting precipitation areas correctly throughout the year. The results show that all products lose their skill to detect volume of precipitation correctly (*VHI*) as the threshold increases. With respect to *VHI*, CMORPH and PERSIANN seem to be less variable with the choice of threshold. It is noted that a precipitation product may offer a high probability of detection, but at the cost of a high false alarm ration. Therefore, high values of *POD*, *QPOD*, and *VHI* should be judged along with the corresponding false alarm ratios.

[32] 6. The estimated *QPOD* and *VHI* values imply that for the entire distribution of precipitation, all satellite products miss more precipitation over cold months (note that the total volume of precipitation consists of the volume of hit and volume of missed precipitation). Overall, CMORPH and PERSIANN miss the least amount of precipitation. For intense precipitation (above higher thresholds), the

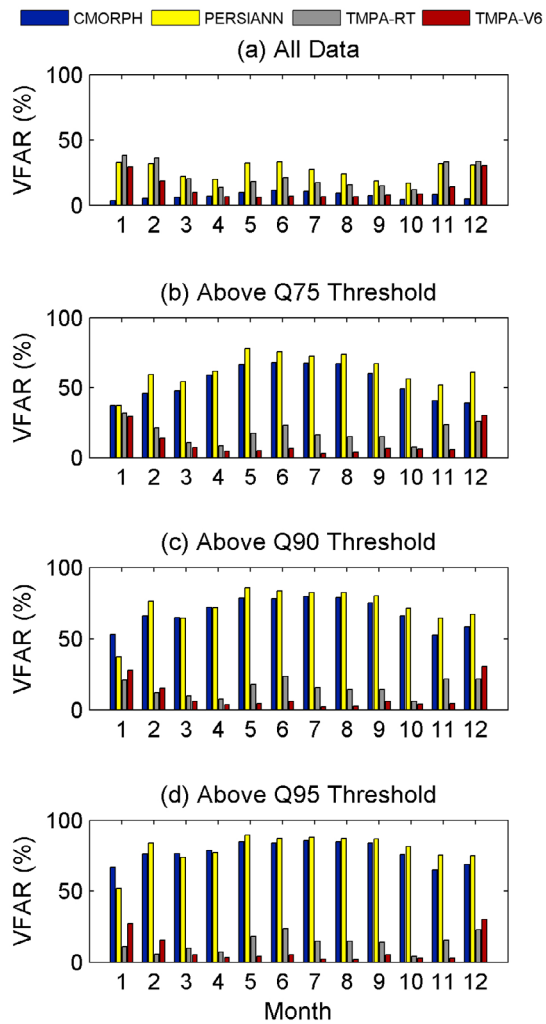

Figure 10. Volumetric false alarm ratio (*VFAR* [%]) with respect to the observed precipitation: (a) all data and (b) above Q75, (c) above Q90, and (d) above Q95 thresholds of STIV data.

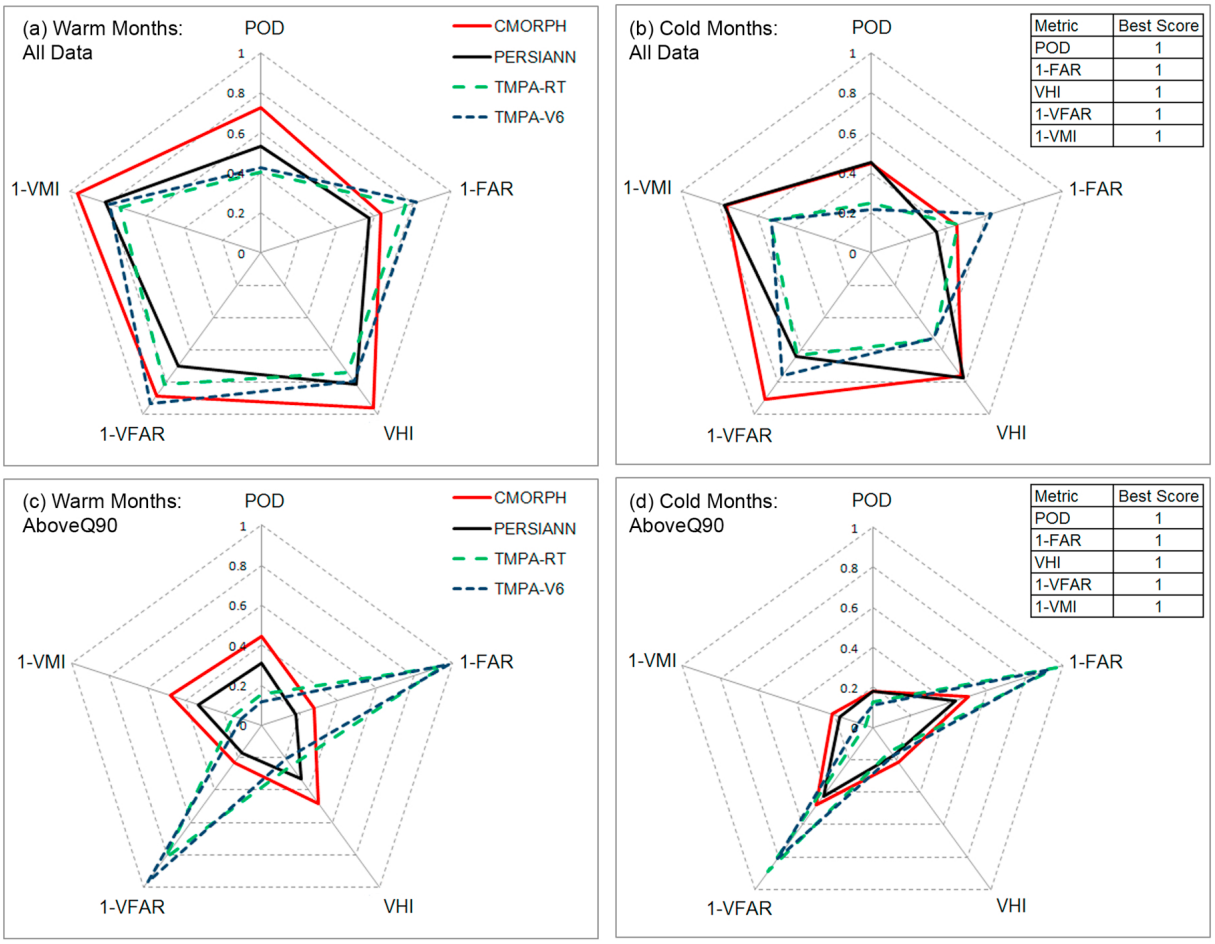

Figure 11. Variability of precipitation algorithm skills in detecting precipitation during (a) warm months (all data), (b) cold months (all data), (c) warm months (above Q90), and (d) cold months (above Q90).

volume of missed precipitation significantly increases for all precipitation products. It is worth reminding the reader that in this study, values above a certain threshold are considered. Therefore, low *POD* of a product should not be interpreted as no rain detection. In fact, the product may have detected precipitation, but below the selected threshold.

[33] 7. The computed false alarm values indicate that all satellite products tend to provide more false precipitation over cold months. As the threshold increases, *QFAR* values of TMPA-RT and TMPA-V6 diminish, while CMORPH and PERSIANN consistently exhibit high false alarm ratios. Similar behavior is observed for volumetric false alarm magnitudes.

[34] The results suggest that none of the precipitation products can be considered ideal for detecting extreme events. In fact, satellite products lose their accuracy as the choice of extreme precipitation threshold increases. This indicates that extensive efforts are necessary to further develop algorithms that can capture extremes more reliably. In addition to improving current algorithms, developing adjustment techniques to advance the application of satellite-derived precipitation in detecting extremes merits a great deal of research. The current limitations on accurate estimation of extremes prevent us from designing short-term warning systems on the basis of satellite data. Future advancement in the detection of extremes may lead to a quantum advancement in early warning systems and hazard mitigation.

[35] It is emphasized that the above conclusions are based on exploratory data analysis using available satellite-based and radar-based data sets. The authors acknowledge that spatial and temporal sampling uncertainties may exist when comparing different satellite products with STIV data [*Sieck et al.*, 2007]. However, given the available data, STIV data are the best possible approximation of the true areal average rainfall values [*AghaKouchak et al.*, 2010c]. This work was intended to contribute to the ongoing research on evaluation of satellite-based precipitation data with respect to extremes. Improving the current algorithms to reduce false alarm and missed precipitation may result in a major advancement in the utilization of satellite data in practical applications. Efforts are underway by the authors to validate satellite-derived data with regard to extremes over different elevations and climate conditions. Such studies may reveal additional information on the reliability of satellite-based extreme precipitation rates at regional scales. Because of the potential significance of extremes and their spatial

extent to current and future research, we predict that in the near future more research efforts will focus on particular treatments required to capture extreme precipitation events more reliably.

[36] **Acknowledgments.** The financial support for this study was made available by NOAA/NESDIS/NCDC (prime award NA09NES4400006, NCSU CICS sub-award 2009-1380-01).

References

Adler, R., C. Kidd, G. Petty, M. Morissey, and H. Goodman (2001), Intercomparison of global precipitation products: The Third Precipitation Intercomparison Project (PIP-3), *Bull. Am. Meteorol. Soc.*, *82*, 1377–1396.

AghaKouchak, A., E. Habib, and A. Bárdossy (2010a), Modeling radar rainfall estimation uncertainties: Random error model, *J. Hydrol. Eng.*, *15*(4), 265–274, doi:10.1061/(ASCE)HE.1943-5584.0000185.

AghaKouchak, A., A. Bárdossy, and H. Habib (2010b), Copula-based uncertainty modeling: Application to multi-sensor precipitation estimates, *Hydrol. Processes*, *24*(15), 2111–2124, doi:10.1002/hyp.7632.

AghaKouchak, A., A. Bárdossy, and E. Habib (2010c), Conditional simulation of remotely sensed rainfall data using a non-gaussian v-transformed copula, *Adv. Water Resour.*, *33*(6), 624–634, doi:10.1016/j.advwatres.2010.02.010.

Amitai, E., X. Llort, and D. Sempere-Torres (2009), Comparison of TRMM radar rainfall estimates with NOAA next-generation QPE, *J. Meteorol. Soc. Jpn.*, *87*, 109–118.

Behrangi, A., B. Khakbaz, T. C. Jaw, A. AghaKouchak, K. Hsu, and S. Sorooshian (2011), Hydrologic evaluation of satellite precipitation products over a mid-size basin, *J. Hydrol. Amsterdam*, doi:10.1016/j.jhydrol.2010.11.043, in press.

Ciach, G., W. Krajewski, and G. Villarini (2007), Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data, *J. Hydrometeorol.*, *8*, 1325–1347.

Dinku, T., S. Chidzambwa, P. Ceccato, S. Connor, and C. Ropelewski (2008), Validation of high-resolution satellite rainfall products over complex terrain, *Int. J. Remote Sens.*, *29*(14), 4097–4110.

Durga Rao, K., V. Bhanumurthy, and P. Roy (2009), Application of satellite-based rainfall products and SRTM DEM in hydrological modelling of brahmaputra basin, *J. Indian Soc. Remote Sens.*, *37*, 587–600.

Ebert, E., J. Janowiak, and C. Kidd (2007), Comparison of near real time precipitation estimates from satellite observations and numerical models, *Bull. Am. Meteorol. Soc.*, *88*, 47–64.

Feidas, H., G. Kokolatos, A. Negri, M. Manyin, N. Chrysoulakis, and Y. Kamarianakis (2009), Validation of an infrared-based satellite algorithm to estimate accumulated rainfall over the mediterranean basin, *Theor. Appl. Climatol.*, *95*(1–2), 91–109.

Gochis, D., S. Nesbitt, W. Yu, and S. Williams (2009), Comparison of gauge-corrected versus non-gauge corrected satellite-based quantitative precipitation estimates during the 2004 name enhanced observing period, *Atmosfera*, *22*(1), 69–98.

Hsu, K., X. Gao, S. Sorooshian, and H. Gupta (1997), Precipitation estimation from remotely sensed information using artificial neural networks, *J. Appl. Meteorol.*, *36*, 1176–1190.

Huffman, G., R. Adler, D. Bolvin, G. Gu, E. Nelkin, K. Bowman, E. Stocker, and D. Wolff (2007), The TRMM multi-satellite precipitation analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale, *J. Hydrometeorol.*, *8*, 38–55.

Janowiak, J., R. Joyce, and Y. Yarosh (2001), A real-time global half-hourly pixel-resolution infrared dataset and its applications, *Bull. Am. Meteorol. Soc.*, *82*, 205–217.

Joyce, R., J. Janowiak, P. Arkin, and P. Xie (2004), CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, *J. Hydrometeorol.*, *5*, 487–503.

Kidd, C. (2001), Satellite rainfall climatology: A review, *Int. J. Climatol.*, *21*, 1041–1066, doi:10.1002/joc.635.

Krajewski, W., and J. Smith (2002), Radar hydrology: Rainfall estimation, *J. Hydrol.*, *25*, 1387–1394.

Levizzani, V., and R. Amorati (2002), A review of satellite-based rainfall estimation methods: A look back and a perspective, paper presented at the 2000 Meteorological Satellite Data User's Conference, Eur. Organ. for the Exploit. of Meteorol. Satell., Bologna, Italy.

Lin, Y., and K. Mitchell (2005), The NCEP stage II/IV hourly precipitation analyses: Development and applications, preprint paper 1.2 presented at 19th Conference on Hydrology, Am. Meteorol. Soc., San Diego, Calif., 9–13 January 2005.

Liu, Z., H. Rui, W. Teng, L. Chiu, G. Leptoukh, and S. Kempler (2009), Developing an online information system prototype for global satellite precipitation algorithm validation and intercomparison, *J. Appl. Meteorol. Climatol.*, *48*(12), 2581–2589.

Maddox, R., J. Zhang, J. Gourley, and K. Howard (2002), Weather radar coverage over the contiguous United States, *Weather Forecast.*, *17*, 927–934.

National Assessment Synthesis Team (2001), *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change, National Assessment Synthesis Team, US Global Change Research Program*, Cambridge Univ. Press, Cambridge, U. K.

Nijssen, B., and D. Lettenmaier (2004), Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the global precipitation measurement satellites, *J. Geophys. Res.*, *109*, D02103, doi:10.1029/2003JD003497.

Sapiano, M., and P. Arkin (2009), An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data, *J. Hydrometeorol.*, *10*(1), 149–166.

Seed, A., and R. Srikanthan (1999), A space and time model for design storm rainfall, *J. Geophys. Res.*, *104*, 31,623–31,630.

Seo, D. (1999), Real-time adjustments of mean field and range-dependent biases in WSR-88d rainfall estimation, paper presented at 79th AMS Annual Meeting, 14th Conference on Hydrology, Am. Meteorol. Soc., Dallas, Tex.

Shen, Y., A. Xiong, Y. Wang, and P. Xie (2010), Performance of high-resolution satellite precipitation products over China, *J. Geophys. Res.*, *115*, D02114, doi:10.1029/2009JD012097.

Sieck, L., S. J. Burges, and M. Steiner (2007), Challenges in obtaining reliable measurements of point rainfall, *Water Resour. Res.*, *43*, W01420, doi:10.1029/2005WR004519.

Sorooshian, S., K. Hsu, X. Gao, H. Gupta, B. Imam, and D. Braithwaite (2000), Evolution of the PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Meteorol. Soc.*, *81*(9), 2035–2046.

Steiner, M., and J. Smith (2000), Reflectivity, rain and kinetic energy flux relationships based on raindrop spectra, *J. Appl. Meteorol.*, *39*, 1923–1940.

Stisen, S., and I. Sandholt (2010), Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling, *Hydrol. Processes*, *24*, 879–891.

Tian, Y., C. Peters-Lidard, B. Choudhury, and M. Garcia (2007), Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications, *J. Hydrometeorol.*, *8*(6), 1165–1183, doi:10.1175/2007JHM859.1.

Tian, Y., C. Peters-Lidard, J. Eylander, R. Joyce, G. Huffman, R. Adler, K. Hsu, F. Turk, M. Garcia, and J. Zeng (2009), Component analysis of errors in satellite-based precipitation estimates, *J. Geophys. Res.*, *114*, D24101, doi:10.1029/2009JD011949.

Turk, F. J., P. Arkin, E. E. Ebert, and M. R. P. Sapiano (2008), Evaluating high-resolution precipitation products, *Bull. Am. Meteorol. Soc.*, *89*(12), 1911–1916.

Tuttle, J., R. Carbone, and P. Arkin (2008), Comparison of ground-based radar and geosynchronous satellite climatologies of warm-season precipitation over the United States, *J. Appl. Meteorol. Climatol.*, *47*, 3264–3270, doi:10.1175/2008JAMC2000.1.

Wilks, D. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., Academic Press, Burlington, Mass.

Yilmaz, K., T. Hogue, K. Hsu, S. Sorooshian, H. Gupta, and T. Wagener (2005), Intercomparison of rain gauge, radar and satellite-based precipitation estimates with emphasis on hydrologic forecasting, *J. Hydrometeorol.*, *6*(4), 497–517.

Zeweldi, D., and M. Gebremichael (2009), Sub-daily scale validation of satellite-based high-resolution rainfall products, *Atmos. Res.*, *92*(4), 427–433.

Zhou, T., R. Yu, H. Chen, A. Dai, and Y. Pan (2008), Summer precipitation frequency, intensity, and diurnal cycle over China: A comparison of satellite data with rain gauge observations, *J. Clim.*, *21*(16), 3997–4010.

A. AghaKouchak, K. Hsu, and S. Sorooshian, Department of Civil and Environmental Engineering, University of California, E4130 Engineering Gateway, Irvine, CA 92697, USA. (amir.a@uci.edu)

E. Amitai, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

A. Behrangi, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., MS 183-301, Pasadena, CA 91109, USA.