HYDROLOGICAL PROCESSES
*Hydrol. Process.* **28**, 2262–2270 (2014)
Published online 27 March 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9779

# Capabilities of satellite precipitation datasets to estimate heavy precipitation rates at different temporal accumulations

Ali Mehran and Amir AghaKouchak*

*Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA, 92697, USA*

## Abstract:

The importance of satellite datasets as alternative sources of precipitation information has been argued in numerous studies. Future developments in satellite precipitation algorithms as well as utilization of satellite data in operational applications rely on a more in-depth understanding of satellite errors and biases across different spatial and temporal scales. This paper investigates the capability of satellite precipitation data sets with respect to detecting heavy precipitation rates over different temporal accumulations. In this study, the performance of Tropical Rainfall Measuring Mission real time (TRMM-RT), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks and CPC MORPHing (CMORPH) is compared against radar-based gauge-adjusted Stage IV data. The results show that none of the high temporal resolution (3-h) datasets are ideal for detecting heavy precipitation rates. In fact, the detection skill of all products drops as the precipitation thresholds (i.e. 75 and 90 percentiles) increase. At higher temporal accumulations (6, 12 and 24 h), the detection skill improves for all precipitation products, with CMORPH showing a better detection skill compared to all other products. On the other hand, all precipitation products exhibit high false alarm ratios above the heavy precipitation thresholds, although TRMM-RT lead to a relatively smaller level of false alarms. These results indicate that further efforts are necessary to improve the precipitation algorithms so that they can capture heavy precipitation rates more reliably. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS precipitation; satellite; heavy rain rate

*Received 26 August 2012; Accepted 19 February 2013*

## INTRODUCTION

Detecting precipitation reliably, especially heavy precipitation rates, is fundamental to a wide variety of applications including developing early warning system, disaster management strategies and water resources management. Recent developments in remote sensing techniques and precipitation retrieval algorithms have resulted in a number of satellite-based precipitation products (Sorooshian *et al*., 2011). Currently, satellite data sets are being used in numerous applications including flood forecasting (Li *et al*., 2009), drought monitoring (AghaKouchak and Nakhjiri, 2012), terrestrial water cycle analysis (Azarderakhsh *et al*., 2011) and water resources management (Kidd *et al*., 2009).

While satellite data sets provide unique opportunities for uninterrupted and quasi-global precipitation monitoring, their associated uncertainties limit their use in operational applications. Numerous studies stress that quantification of uncertainties and errors associated with satellite data remains a major challenge (Hong *et al*., 2006); (Sorooshian *et al*., 2011); (AghaKouchak *et al*., 2012); (Bellerby, 2007); (Tian *et al*., 2009). Thus far, a great deal of effort has been put into validation and verification of satellite precipitation products (e.g. (Ebert *et al*., 2007); (Turk *et al*., 2008); (Tuttle *et al*., 2008); (Tian *et al*., 2009); (Amitai *et al*., 2009); (Gochis *et al*., 2009); (Shen *et al*., 2010); (Behrangi *et al*., 2011); (Stampoulis and Anagnostou, 2012); (Shen *et al*., 2010); (Zeweldi and Gebremichael, 2009); (Sapiano and Arkin, 2009).

In a recent study, (AghaKouchak *et al*., 2011) evaluated several operational precipitation products over the Southern Great Plains with respect to various extreme precipitation thresholds and concluded that the detection skill of precipitation products reduces as the choice of the extreme threshold increases. Furthermore, the results showed that the commonly used monthly bias adjustment approach did not necessarily improve the detection of high precipitation rates.

Accordingly, this study aims to evaluate the performance of satellite data sets against ground-based measurements across different temporal scales over the conterminous United States (CONUS) in order to understand at what temporal scales satellite precipitation data sets perform better with respect to heavy precipitation rates. Detection of heavy rain rates from three commonly used satellite precipitation products, namely Tropical Rainfall Measuring Mission real time (TRMM-RT) (Huffman *et al*., 2007), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) (Sorooshian *et al*., 2000) and CPC MORPHing (CMORPH) (Joyce *et al*., 2004) were evaluated with respect to Stage IV radar-based gauge adjusted data. Several statistical indices were computed for these three satellite products for different heavy rain rate thresholds (i.e. 75% and 90% quantiles) and time scales (i.e. 3 h, 6 h, 12 h and daily). The results were then compared to evaluate the performance of satellite datasets in detecting extremes over different temporal resolutions.

*Correspondence to: Amir AghaKouchak, Department of Civil and Environmental Engineering, University of California Irvine, Irvine, CA, 92697, USA.
E-mail: amir.a@uci.edu

After this introduction, the study area and data sets are described ('Data sets and study area'). In 'Analysis and results', the methodology and results are presented. Finally, 'Conclusion' summarizes the discussions and conclusions.

## DATA SETS AND STUDY AREA

Satellite data sets used in this study include: (1) The TRMM-RT data (Huffman *et al*., 2007) which is mainly based on microwave data aboard Low Earth Orbit satellites. The TRMM-RT algorithm is primarily based on microwave observations from the low orbiter satellites. In this algorithm, the gaps in microwave observations are filled with infrared (IR) data. (2) The PERSIANN (Sorooshian *et al*., 2000); (Hsu *et al*., 1997), which estimates precipitation based on IR data after real-time adjustment by microwave scans. IR-based precipitation data, including PERSIANN, rely on statistical relationships between IR observations of the cloud-top temperature and precipitation rate. Given the high uncertainties in such statistical relationships, PERSIANN data are calibrated based on microwave observations such as the Special Sensor Microwave Imager, TRMM Microwave Imager and Advanced Microwave Scanning Radiometer-Earth observing system. The calibration process is based on an adaptive training technique (Hsu *et al*., 1997) which updates the PERSIANN neural network parameters when microwave data are available. (3) The CMORPH (Joyce *et al*., 2004) product, which is a microwave-based product advected in time using Geosynchronous IR data. In CMORPH, when microwave observations are not available, the IR observations are used to advect the last microwave scan over time. In addition to advecting precipitation forward in time, the algorithm propagates precipitation backward once the next microwave observation becomes available.

The above mentioned satellite data are available in 3-h intervals and at a 0.25° spatial resolution. In order to evaluate the performance of satellite data at different time intervals, 6-h, 12-h and 24-h precipitation accumulations are computed from 3-h data. Given the coarse resolution of satellite products, stage IV dataset is regridded onto a 0.25° spatial resolution to synchronize it with the satellite data. The spatial regridding is performed by averaging Stage IV precipitation rates within each satellite precipitation pixel. Four years (Jan. 2005 – Dec. 2008) of data from the CONUS were processed for analysis.

The Stage IV dataset, (which is a radar-based gauge-adjusted product provided by the National Centre for Environmental Prediction, National Oceanic and Atmospheric Administration) is used as the reference data. These data were collected and mosaicked onto the Hydrological Rainfall Analysis Project national grid system by the National Weather Service River Forecast Centre. The Stage IV is a multi-sensor product that consists of merged and quality controlled radar and gauge measurements. The data set is available in approximately $4 \times 4$ km spatial and hourly temporal resolutions.

The authors acknowledge that Stage IV data, and in principle, radar-based precipitation data, are subject to different types of biases and uncertainties (AghaKouchak *et al*., 2010). Errors and uncertainties in radar-based precipitation data are due to various instrumental and algorithmic issues including non-uniformity in vertical profiles of reflectivity, beam over-shooting, anomalous propagation, inappropriate retrieval relationships, instrumental errors, partial beam filling, hardware calibration, and spatial sampling and random sampling errors (Krajewski and Smith, 2002). Numerous studies have evaluated the State IV data against ground-based observations and have suggested uncertainty models to describe the associated uncertainties (Ciach and Krajewski, 2007) Ciach07; (AghaKouchak *et al*., 2010); (Seo, 1999); (Villarini *et al*., 2008). The current version of Stage IV data is adjusted and calibrated with the automated National Climatic Data Centre gauge data for different biases using an automated algorithm (Lin and Mitchell, 2005).

## ANALYSIS AND RESULTS

The focus of this study is to verify satellite products with respect to detecting heavy precipitation rates (hereafter, extremes) at different temporal accumulations. It should be noted that extremes can be defined based on different viewpoints. In this paper, heavy (extreme) precipitation rate is defined as rain rates above 75% and 90% quantiles. Various statistical and graphical methods are used to assess detection of extremes from satellite datasets. These results are presented for pixels with more than 100 concurrent pairs of reference and satellite precipitation data above 1 mm/h in order to avoid unreliable statistics.

The metrics used for validation and verification include: (1) Probability of Detection (*POD*; (Wilks, 2006)) which represents the ratio of the number of correct identifications of precipitation to the total number of reference precipitation occurrences; (2) Quantile *POD* (*QPOD*; (AghaKouchak *et al*., 2011)), which refers to the ratio of the correct detections above a certain threshold to the total number of occurrences above the same threshold as indicated by the reference - see Equation (2); (3) Volumetric Hit Index (*VHI*), which identifies the volume of precipitation above the threshold ($t$) detected correctly (AghaKouchak *et al*., 2011) - see Equation (3); (4) False Alarm Ratio (*FAR*) which represents the ratio of the number of false identifications of precipitation to the total number of satellite precipitation occurrences (see Wilks, 2006) and (5) Quantile *FAR* (*QFAR*), which is defined as the *FAR* above a certain threshold - see Equation (4). The perfect score (detection skill) for *POD*, *QPOD* and *VHI* is 1 (0 = no detection skill; 1 = perfect detection skill), whereas the perfect score for *FAR* and *QFAR* is 0 (0 = no false alarm; 1 = 100% false precipitation). Finally the equations of the above metrics are presented in the Appendix.

Figures 1, 2 and 3 display *POD* and *QPOD* scores for TRMM-RT, CMORPH and PERSIANN, respectively. The first column of Figures 1–3 show tPOD calculated

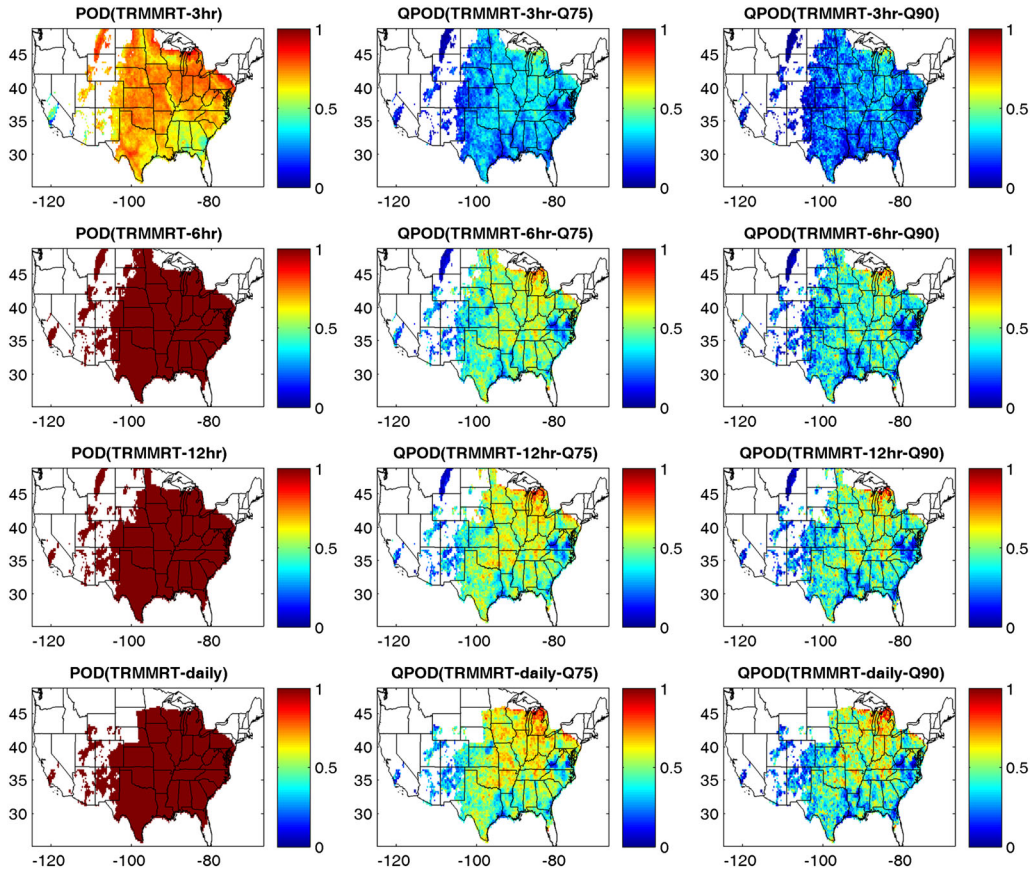


Figure 1. TRMM-RT: Probability of Detection (POD) and Quantile POD (QPOD)

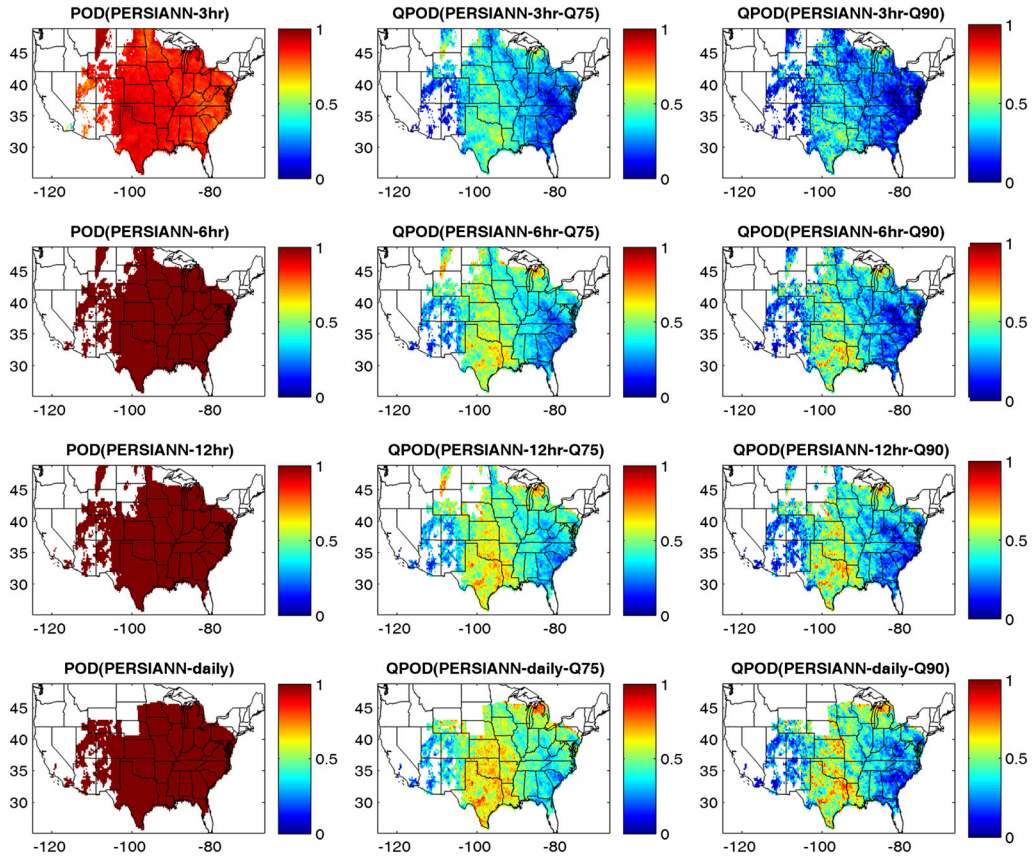


Figure 2. PERSIANN: Probability of Detection (POD) and Quantile POD (QPOD)

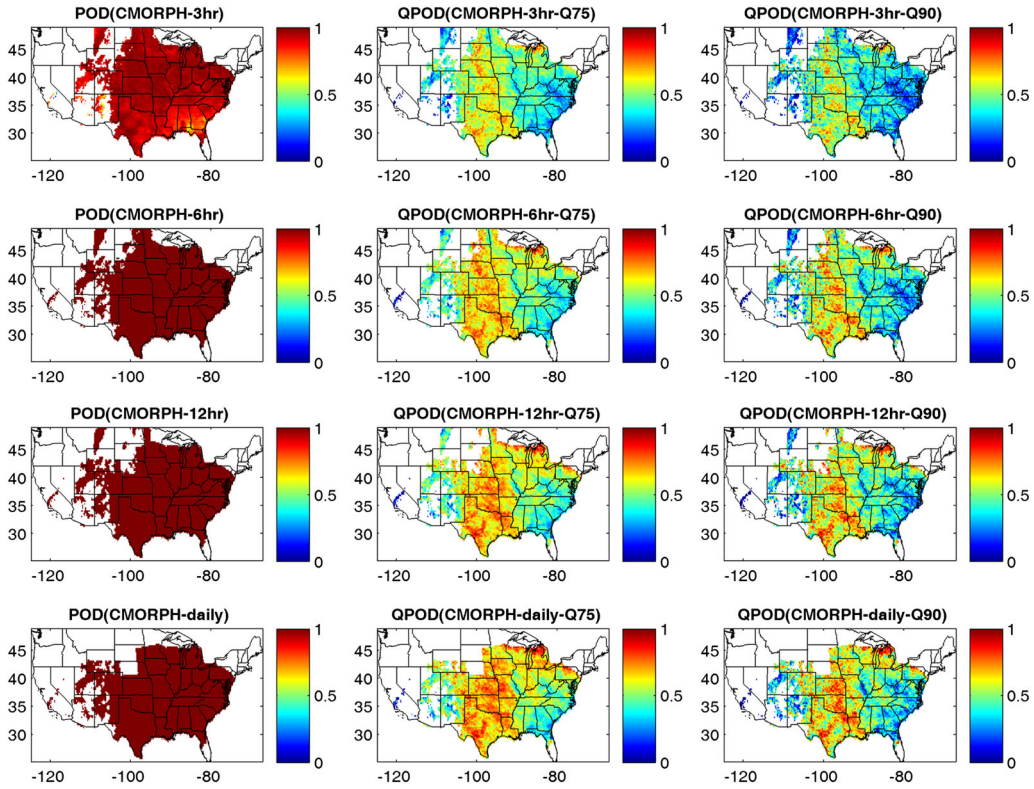


Figure 3. CMORPH: Probability of Detection (POD) and Quantile POD (QPOD)

using the entire distribution of observed precipitation. The second and third columns in Figures 1–3 present *QPOD* for thresholds of 75% quantile (hereafter, Q75) and 90% quantile (hereafter, Q95), respectively. One can see that *POD* and *QPOD* significantly improve as the time interval increases. Especially, this improvement is considerable from 3 h to 6 h (compare the first two row panels in Figures 1–3). The second and third columns in Figures 1–3 highlight that the 3-h *QPOD* values above Q75 and Q90 are very low and often near zero, indicating that there is no skill in detecting heavy precipitation rates.

All three precipitation products are consistent in the fact that their *QPOD* values are very low at the 3-h time scale, although *QPOD* of CMORPH are higher than the other two products. As the temporal accumulation increases to the 6-, 12- and 24-h timescales, *QPOD* improves significantly. However, *QPOD* values are considerably less than *POD* values even at the daily temporal scale (see the last row of Figures 1–3). Overall, the increasing patterns of *QPOD* values for the three products are consistent above Q75. As shown, CMORPH generally exhibits higher *QPOD* values (Q75 and Q95), especially in the central United States. It is noted that all satellite products have limitation in detecting orographic precipitation (see the Appalachian and Rocky Mountains). This issue has been identified as one of the limitations of current satellite precipitation data sets in (Sorooshian *et al.*, 2011). It is expected that the Global Precipitation Measurement (Smith *et al.*, 2007) Mission will improve precipitation estimation using its dual-frequency radar and radiometer equipped with high-frequency channels.

Figures 4, 5 and 6 present the *VHI* (AghaKouchak *et al.*, 2011) that represents the volume of precipitation above a threshold ($t$) detected correctly by satellite data. Similar to *QPOD* plots, *VHI* decreases as the heavy rain rate threshold increases. This indicates that satellite data sets lose their skills at higher thresholds. It should be noted that *VHI* is not the same as bias, and it ranges from 0 to 1, with 1 being the perfect score. While a score of 1 indicates that the satellite data is detecting precipitation above a given threshold (e.g. Q75), it does not show the bias or error of detection (see Equation (3). The first column shows *VHI* for the entire distribution of precipitation, whereas the columns two and three display *VHI* above Q75 and Q90. One can see that the *VHI* values are rather low at the 3-h time scale and that they increase at higher temporal accumulations. Finally, it is worth mentioning that for some areas (e.g. south eastern U.S.), *VHI* values are rather low above Q90 threshold (compare the last rows of Figures 4–6).

It is worth pointing out that *QPOD* (or *POD*) represents the ratio of the number of correct identifications of precipitation to the total number of reference precipitation occurrences. A high *QPOD* does not necessarily indicate capturing a high volume of precipitation, as the detected occurrences by *QPOD* may not constitute a large fraction of the total volume of observed precipitation. Conversely, a high *VHI* does not necessarily lead to a high *QPOD* (e.g. when a small percentage of detected occurrences constitute a large volume of the total rainfall). For this reason, one needs to consider both metrics for evaluation of precipitation products.

Figures 7, 8 and 9 display the *FAR* and the Quantile *FAR* (*QFAR*), respectively for TRMM-RT, CMORPH and PERSIANN. *QFAR* refers to the *FAR* above a threshold ($t$; here, Q75 and Q90). The three columns in Figures 7,

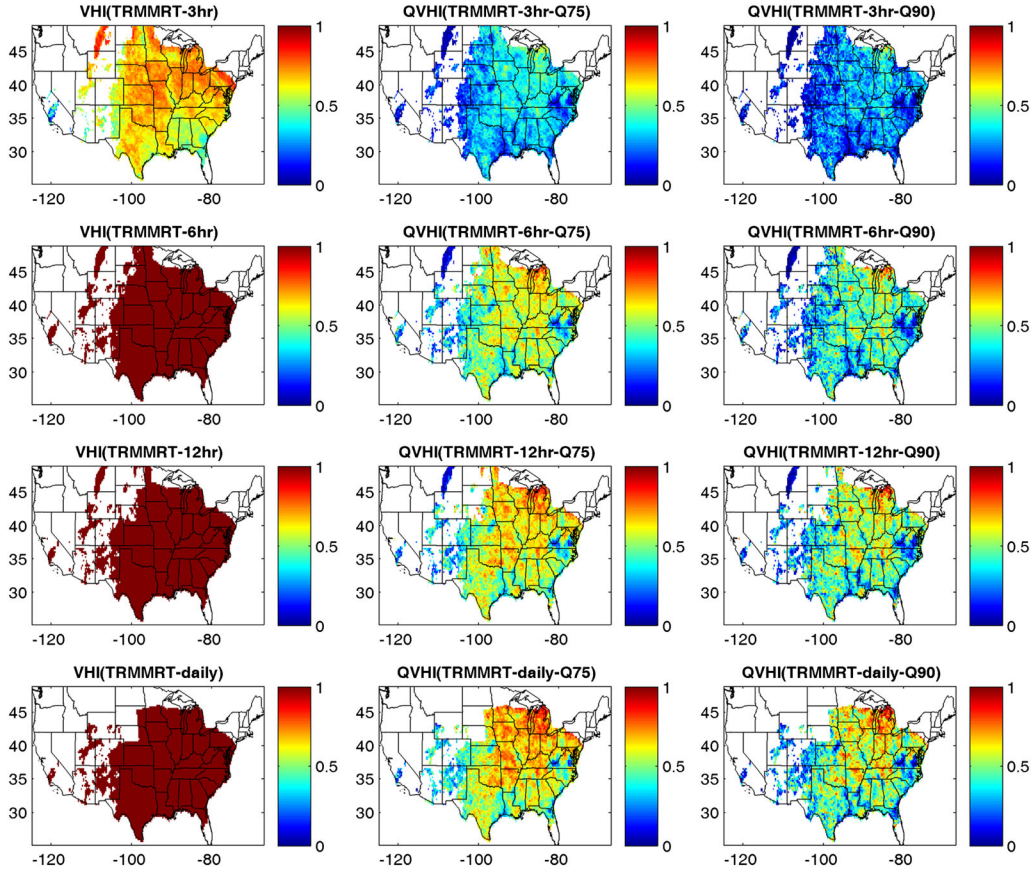


Figure 4. TRMM-RT: Volumetric Hit Index (VHI)

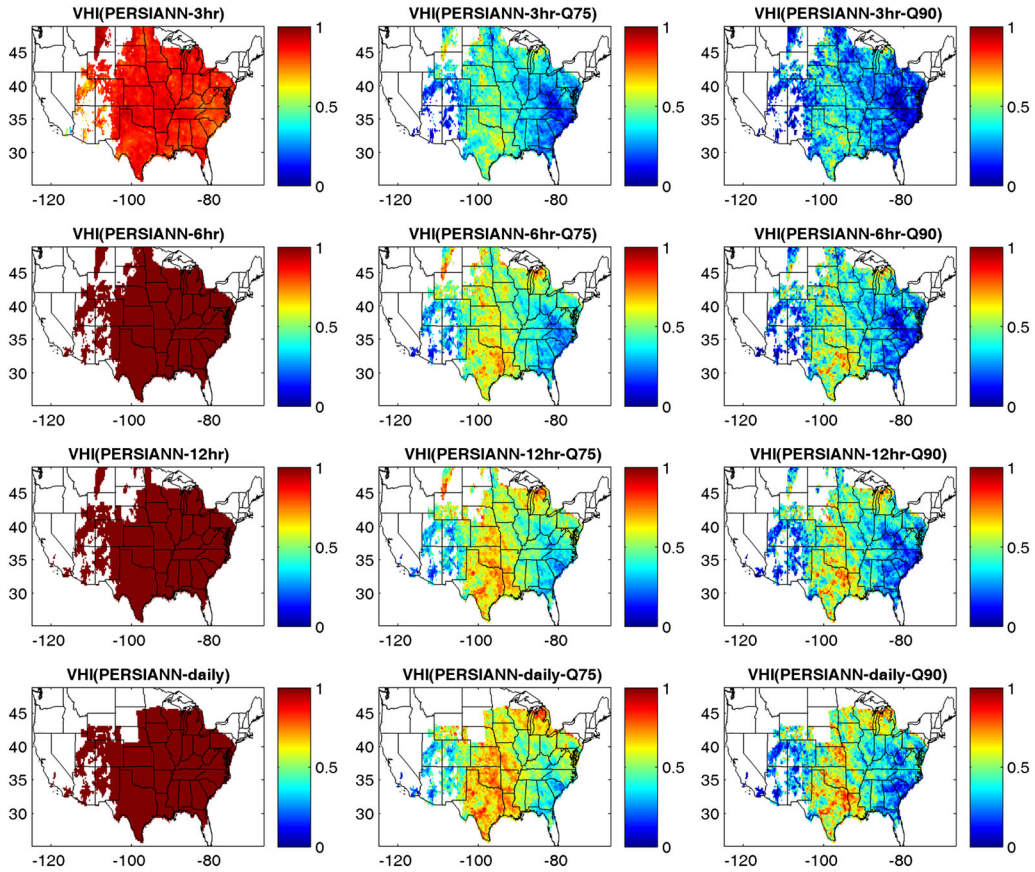


Figure 5. PERSIANN: Volumetric Hit Index (VHI)

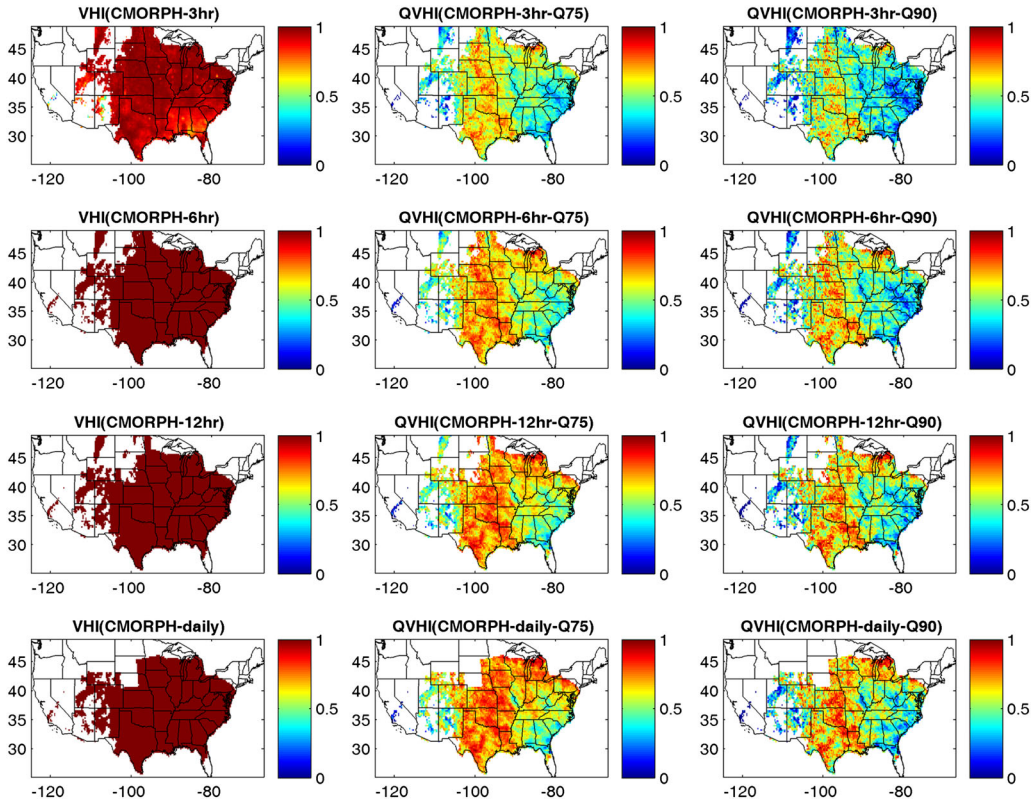


Figure 6. CMORPH: Volumetric Hit Index (VHI)

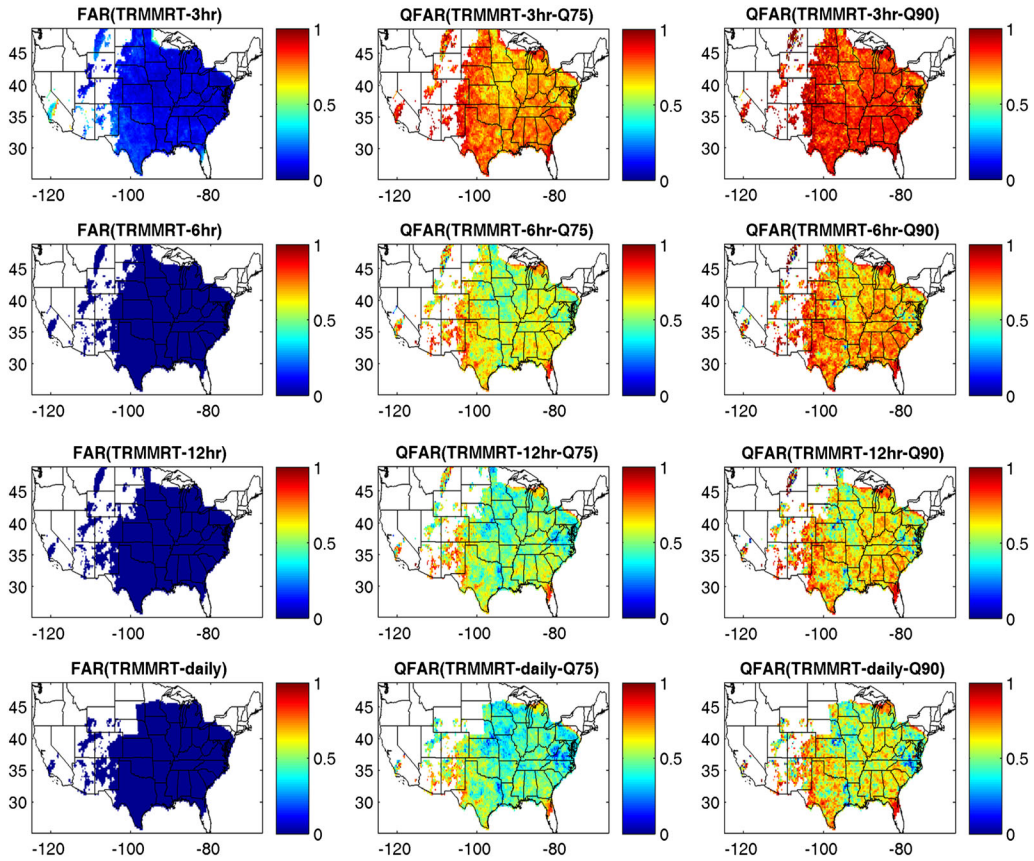


Figure 7. TRMM-RT: False Alarm Ratio (FAR) and Quantile FAR (QFAR)

8 and 9 show *FAR* and *QFAR* above Q75, and *QFAR* above Q90, respectively. While the *FAR* of all precipitation products are rather low, *QFAR* values significantly increase with the heavy precipitation thresholds, especially at the 3-h temporal scale (compare the first row panels in Figures 7–9).

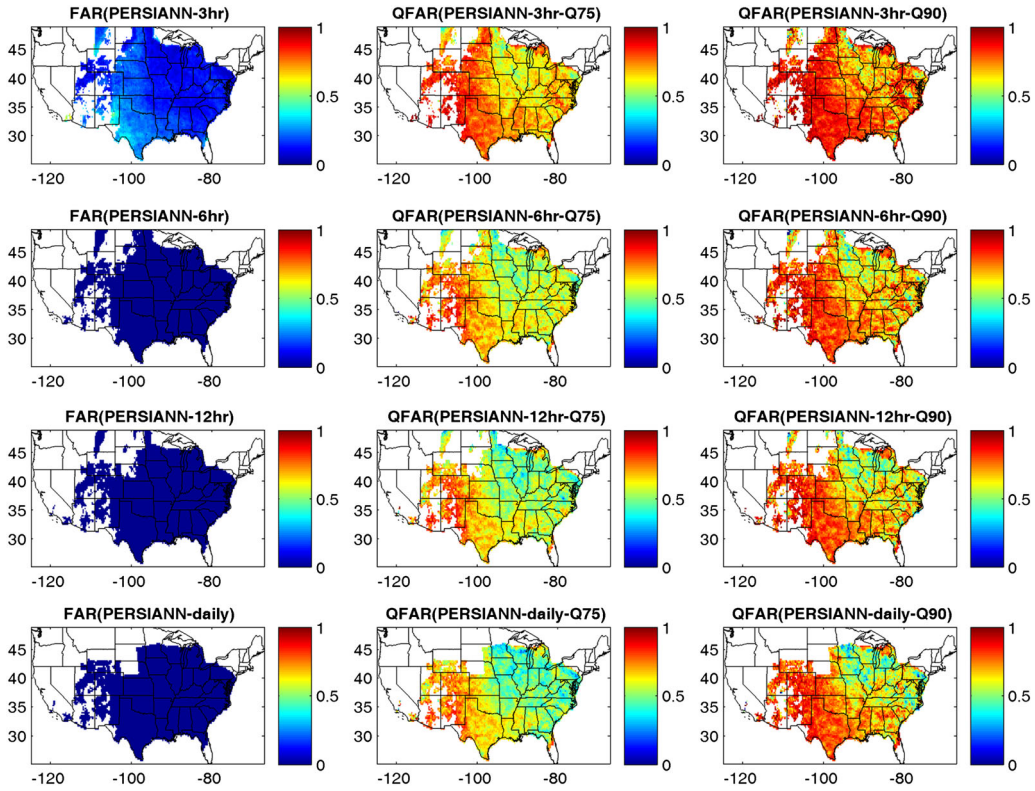


Figure 8. PERSIANN: False Alarm Ratio (FAR) and Quantile FAR (QFAR)

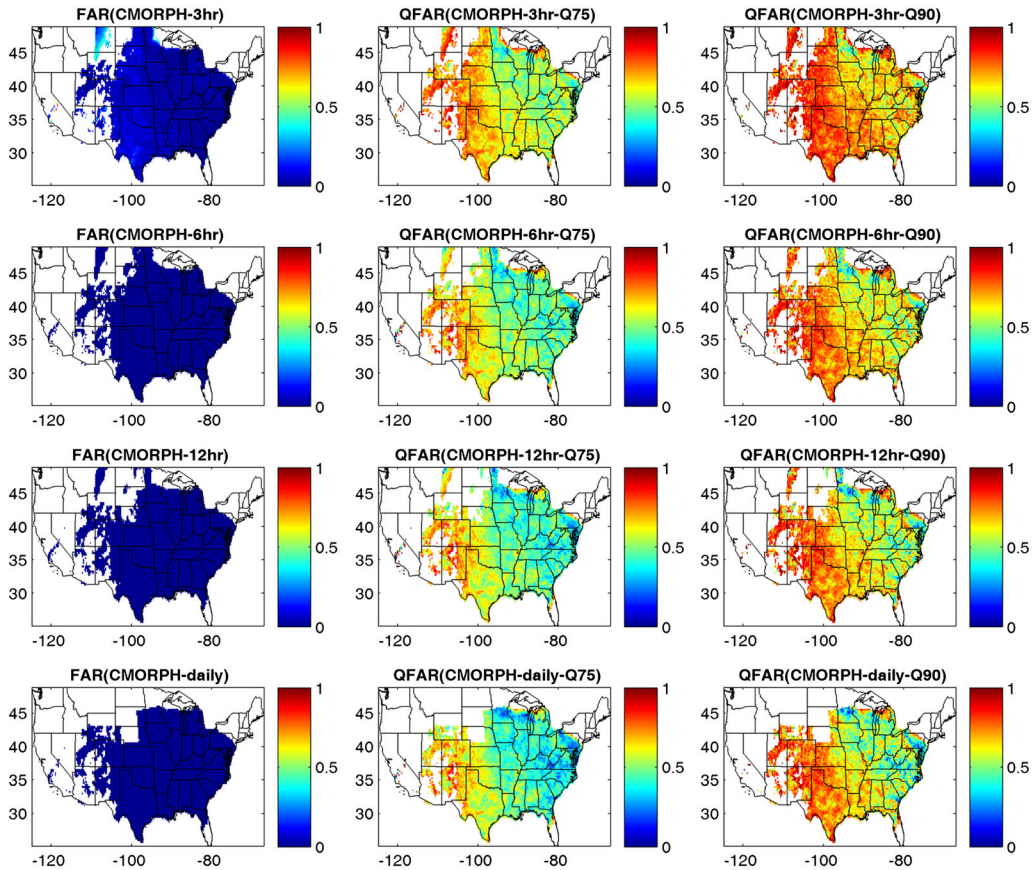


Figure 9. CMORPH: False Alarm Ratio (FAR) and Quantile FAR (QFAR)

In all products, as the temporal accumulation increases, *QFAR* values decrease. However, the increase is more significant for TRMM-RT compared to the other precipitation products. Overall, TRMM-RT exhibits the lowest *QFAR* values at high thresholds and temporal accumulations (compare the last row panels in Figures 7–9). The main reason is that TRMM-RT is primarily a microwave-based product. Previous studies show that microwave-based precipitation data sets lead to smaller false alarm compared to IR-based precipitation products (Tian *et al*., 2009); (AghaKouchak *et al*., 2011). While CMORPH is also a microwave-based product, we speculate that the CMORPH's forward interpolation using IR data results in higher *QFAR* relative to TRMM-RT. For this reason, one can see similarities in the *QFAR* patterns of PERSIANN and CMORPH (see Figures 8 and 9).

Note that *QFAR* (e.g. Q75) values only indicate false precipitation above a certain threshold. In practice, both satellite and reference data may have detected precipitation at a given pixel; however if satellite products are above a threshold (e.g. Q75) and reference data is below the same threshold, then *QFAR* would be above zero.

## CONCLUSIONS

The importance of satellite data sets as alternative sources of precipitation information has been argued in numerous studies (Sorooshian *et al*., 2000) and (Sorooshian *et al*., 2011). Future developments in satellite precipitation algorithms as well as utilization of satellite data in operational applications rely on a more in-depth understanding of satellite errors and biases across different spatial and temporal scales. In hydrologic and land-surface modeling, errors and biases in precipitation propagate into model predictions (Nijssen and Lettenmaier, 2004). A modeler might want to know at what spatial and temporal scales the input data are less subject to uncertainty. In a flood warning system, the detection of precipitation is a key component and one needs to know the capability of the target data set to detect precipitation, especially heavy precipitation rates. This paper contributes to ongoing validation and verification efforts with a focus on several validation metrics above high precipitation thresholds and over different temporal accumulations.

This study surveys precipitation detection and *FAR*s of three satellite-retrieved precipitation data sets: TRMM-RT, PERSIANN and CMORPH. These datasets are compared with the radar-based gauge-adjusted Stage IV data above Q75 and Q95 thresholds. It can be concluded that none of the 3-h products can be considered ideal for detecting heavy precipitation rates. Actually, the detection skill of all products decreases as the extreme thresholds increase. Moreover, at higher temporal accumulations (6, 12 and 24 h), the detection skill increases for all precipitation products. Overall, CMORPH seems to offer a better detection skill at higher precipitation rates.

With respect to false precipitation signals, the 3 h data sets show high *QFAR* values above the extreme precipitation thresholds. Also, the false signals reduce as the temporal accumulation increases. The results show that TRMM-RT lead to less *FAR* and *QFAR* compared to the other two products (compare Figures 7, 8 and 9).

This study indicated that none of these precipitation products can be considered ideal with respect to detecting extremes at high temporal resolutions (e.g. 3 h). One can argue that integrating high temporal resolution data into hydrological models or land-surface models for flood forecasting (or applications sensitive to heavy precipitation rates) may lead to biased results. However, accumulating data into longer temporal accumulations significantly improves the data. The authors argue that further efforts are necessary to improve precipitation retrieval algorithms so that they can capture heavy precipitation rates more reliably. Furthermore, improvements can also be achieved by developing bias adjustment techniques designed to improve satellite data with respect to extremes (e.g. quantile-based or PDF-based adjustment techniques). Given the significance of applications such as flood forecasting, improvements in detecting extremes could lead to a major advance in integration of satellite data into operational applications.

ACKNOWLEDGEMENTS

We thank the Editor and reviewers for their thoughtful comments and suggestions which led to substantial improvements in the revised version of the manuscript. This study is funded by the United States Bureau of Reclamation (USBR) Award No. R11AP81451 and the Army Research Office (Award No.W911NF-11-1-0422).

## REFERENCES

AghaKouchak, A., Bárdossy, A., Habib, E., 2010. Copula-based uncertainty modeling: Application to multi-sensor precipitation estimates. *Hydrological Processes* **24**(15): 2111–2124.

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., Amitai, E., 2011. Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research-Atmospheres* **116**.

AghaKouchak, A., Mehran, A., Norouzi, H., Behrangi, A., 2012. Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters* **39**(9): L09406.

AghaKouchak, A., Nakhjiri, N., 2012. A near real-time satellite-based global drought climate data record. *Environmental Research Letters* **7**(4): 044037, doi:10.1088/1748-9326/7/4/044037.

Amitai, E., Llort, X., Sempere-Torres, D., 2009. Comparison of trmm radar rainfall estimates with noaa next-generation qpe. *Journal of the Meteorological Society of Japan* **87**: 109–118.

Azarderakhsh, M., Rossow, W. B., Papa, F., Norouzi, H., Khanbilvardi, R., 2011. Diagnosing water variations within the Amazon basin using satellite data. *Journal of Geophysical Research-Atmospheres* **116**.

Behrangi, A., Khakbaz, B., Jaw, T., AghaKouchak, A., Hsu, K., Sorooshian, S., 2011. Hydrologic evaluation of satellite precipitation products at basin scale. *Journal of Hydrology* **397**: 225–237.

Bellerby, T., 2007. Satellite rainfall uncertainty estimation using an artificial neural network. *Journal of Hydrometeorology* **8**(6): 1397–1412.

Ciach, G., Krajewski, W., Villarini, 2007. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with nexrad data. *Journal of Hydromet.* **8**: 1325–1347.

Ebert, E., Janowiak, J., Kidd, C., 2007. Comparison of near real time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society* **88**: 47–64.

Gochis, D., Nesbitt, S., Yu, W., Williams, S., 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.

 

Hong, Y., Hsu, K., Moradkhani, H., Sorooshian, S., 2006. Uncertainty quantification of satellite precipitation estimation and monte carlo assessment of the error propagation into hydrologic response. *Water Resources Research* **42**(8): w08421.

Hsu, K., Gao, X., Sorooshian, S., Gupta, H., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology* **36**: 1176–1190.

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Stocker, E., Wolff, D., 2007. The trmm multi-satellite precipitation analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale. *Journal of Hydrometeorology* **8**: 38–55.

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

Kidd, C., Levizzani, V., Turk, J., Ferraro, R., JUN 2009. Satellite Precipitation Measurements for Water Resource Monitoring. *Journal of the American Water Resources Association* **45**(3): 567–579.

Krajewski, W., Smith, J., 2002. Radar hydrology: rainfall estimation. *Journal of Hydrology* **25**: 1387–1394.

Li, L., Hong, Y., Wang, J., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Korme, T., Okello, L., JUL 2009. Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational flood prediction system in Nzoia Basin, Lake Victoria, Africa. *Natural Hazards* **50**(1): 109–123.

Lin, Y., Mitchell, K., 2005. The ncep stage ii/iv hourly precipitation analyses: development and applications. In: Preprints, 19th Conf. on Hydrology, American Meteorological Society, San Diego, CA, 9–13 January 2005, Paper 1.2.

Nijssen, B., Lettenmaier, D., 2004. Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the global precipitation measurement satellites. *Journal of Geophysical Research* **109** (D02103).

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

Seo, D., 1999. Real-time adjustments of mean field and range-dependent biases in WSR-88d rainfall estimation. 79th AMS Annual Meeting, 14th Conference on Hydrology.

Shen, Y., Xiong, A., Wang, Y., Xie, P., JAN 30 2010. Performance of high-resolution satellite precipitation products over china. *Journal of Geophysical Research-Atmospheres* **115**.

Smith, E., Asrar, G., Furuhama, Y., Ginati, A., Mugnai, A., Nakamura, K., Adler, R., Chou, M., Desbois, M., Durning, J., *et al*., 2007. International global precipitation measurement (gpm) program and mission: An overview. *Measuring Precipitation From Space*, **611–653**.

Sorooshian, S., AghaKouchak, A., Arkin, P., Eylander, J., Foufoula-Georgiou, E., Harmon, R., Hendrickx, J. M. H., Imam, B., Kuligowski, R., Skahill, B., Skofronick-Jackson, G., 2011. Advanced concepts on remote sensing of precipitation at multiple scales. *Bulletin of the American Meteorological Society* **92**(10): 1353–1357.

Sorooshian, S., Hsu, K., Gao, X., Gupta, H., Imam, B., Braithwaite, D., 2000. Evolution of the persiann system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society* **81**(9): 2035– 2046.

Stampoulis, D., Anagnostou, E., 2012. Evaluation of global satellite rainfall products over continental europe. *Journal of Hydrometeorology* **13**(2): 588–603.

Tian, Y., Peters-Lidard, C., Eylander, J., Joyce, R., Huffman, G., Adler, R., Hsu, K., Turk, F., Garcia, M., Zeng, J., 2009. Component analysis of errors in satellite-based precipitation estimates. *Journal of Geophysical Research* **114**(D24101).

Turk, F. J., Arkin, P., Ebert, E. E., Sapiano, M. R. P., 2008. Evaluating high-resolution precipitation products. *Bulletin of the American Meteorological Society* **89**(12): 1911–1916.

Tuttle, J., Carbone, R., Arkin, P., DEC 2008. Comparison of ground-based radar and geosynchronous satellite climatologies of warm-season precipitation over the united states. *Journal of Applied Meteorology and Climatology* **47**(12): 3264–3270.

Villarini, G., Serinaldi, F., Krajewski, W., 2008. Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Advances in Water Resources* **31**(12): 1674–1686.

Wilks, D., 2006. Statistical Methods in the Atmospheric Sciences. Academic Press, Burlington, MA (2nd edition).

Zeweldi, D., Gebremichael, M., 2009. Sub-daily scale validation of satellite-based high-resolution rainfall products. *Atmospheric Research* **92**(4): 427–33.

## APPENDIX: EQUATIONS

Probability of Detection (POD; (Wilks, 2006)):

$$POD = \frac{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} > 0 \& P_{ref_i} > 0\right)}{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} > 0 \& P_{ref_i} > 0\right) + \sum_{i=1}^{n} \mathrm{I}\left(P_{ref_i} | P_{sat_i} > 0 \& P_{ref_i} > 0\right)} \quad (1)$$

where $P_{sat}$ = satellite data; $P_{ref}$ = reference measurements (here Stage IV data); I = the indicator function; and $n$ = number exceedances.

Quantile Probability of Detection (QPOD; (AghaKouchak *et al*., 2011)):

$$QPOD = \frac{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} \geq t \& P_{ref_i} \geq t\right)}{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} > t \& P_{ref_i} \geq t\right) + \sum_{i=1}^{n} \mathrm{I}\left(P_{ref_i} | P_{sat_i} < t \& P_{ref_i} \geq t\right)} \quad (2)$$

where $t$ = threshold (i.e. 75, 90 Quantiles).

Volumetric Hit Index (VHI; (AghaKouchak *et al*., 2011)):

$$VHI = \frac{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} \geq t \& P_{ref_i} \geq t\right)}{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} \geq t \& P_{ref_i} \geq t\right) + \sum_{i=1}^{n} \mathrm{I}\left(P_{ref_i} | P_{sat_i} < t \& P_{ref_i} \geq t\right)} \quad (3)$$

False Alarm Ratio (FAR; (Wilks, 2006)): represents the ratio of the number of false identifications of precipitation to the total number of satellite precipitation occurrences.

Quantile False Alarm Ratio:

$$QFAR = \frac{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} \geq t \& P_{ref_i} < t\right)}{\sum_{i=1}^{n} \mathrm{I}\left(P_{sat_i} | P_{sat_i} \geq t \& P_{ref_i} \geq t\right) + \sum_{i=1}^{n} \mathrm{I}\left(P_{ref_i} | P_{sat_i} \geq t \& P_{ref_i} < t\right)} \quad (4)$$