Systematic anomalies over inland water bodies in satellite-based precipitation estimates

Yudong Tian[1] and Christa D. Peters-Lidard[2]

Received 24 May 2007; revised 20 June 2007; accepted 25 June 2007; published 24 July 2007.

[1] We studied two recent high-resolution precipitation datasets derived from satellite remote-sensing, TRMM 3B42 and CMORPH, and compared them with ground-based radar and gauge measurements in the southeastern U. S.. We found there are systematic differences in rainfall estimates between inland water-body pixels and land pixels. On average, there are about twice as many raining days over water bodies than over land pixels in the satellite products, causing much higher false alarm rates over water bodies. The increased false alarms occur mostly in the form of light rain (<2 mm/day), and lead to significantly different rain rate distributions between water-body and land pixels. We speculate that this inconsistency is caused by deficiencies in emissivity characterization for the passive microwave-based rainfall retrievals that serve as input to these merged products. **Citation:** Tian, Y., and C. D. Peters-Lidard (2007), Systematic anomalies over inland water bodies in satellite-based precipitation estimates, *Geophys. Res. Lett.*, *34*, L14403, doi:10.1029/2007GL030787.

1. Introduction

[2] Recently, a few high-resolution precipitation datasets based on satellite remote-sensing estimates have become available [e.g., *Sorooshian et al.*, 2000; *Kidd et al.*, 2003; *Joyce et al.*, 2004; *Turk and Miller*, 2005; *Huffman et al.*, 2007]. These global products' high spatial (0.25-degree or higher) and temporal (3-hour or higher) resolutions make them attractive for many hydrological applications such as watershed modeling and flood prediction, as they can resolve smaller watersheds and water bodies than their precedent products (e.g., Global Precipitation Climatology Project (GPCP) [*Huffman et al.*, 1997]).

[3] These high-resolution products are primarily derived from precipitation estimates from low-Earth-orbit satellites. Today the satellites' extensive spatial and temporal coverage makes production of such measurements possible. In particular, since the launch of the Tropical Rainfall Measuring Mission (TRMM) satellite, the coverage of the Earth's surface by space-borne passive microwave (PMW) sensors has increased significantly. Currently, over 80% of the Earth's surface is covered by at least one of the passive microwave sensors in a 3-hour period [*Huffman et al.*, 2007].

[4] The current practice of producing high resolution multi-sensor precipitation estimates typically combines both PMW-based and infrared (IR)-based retrievals, taking advantage of the more accurate rainfall rate retrievals from PMW sensors and better spatial and temporal coverage from IR-based retrievals [e.g., *Adler et al.*, 1993; *Sorooshian et al.*, 2000; *Kidd et al.*, 2003; *Turk and Miller*, 2005; *Huffman et al.*, 2007]. Alternatively, "morphed" time interpolation of the PMW estimates can be used to provide estimates on a uniform time/space grid [*Joyce et al.*, 2004].

[5] Over land, the individual PMW-based precipitation retrievals have some challenges. The heterogeneity of many land surface parameters and poor characterization of surface temperatures and emissivities make varying contributions to the upwelling microwave radiation, creating background signals similar to rainfall signatures over certain land features [*Grody*, 1991; *Ferraro et al.*, 1998]. Therefore, screening strategies have to be developed to discriminate land pixels with false rainfall signatures, such as snow, ice and semiarid land, in scattering-based retrieval algorithms [*Ferraro et al.*, 1998]. In addition, coastlines usually require special treatment to deal with the emmisivity contrast between land and ocean [*Adler et al.*, 1993]. In land areas where these issues do not exist, such as the subtropical inland southeastern United States (SEUS), it is reasonable to expect the PMW-based rainfall retrievals to perform better.

[6] However, when applying two of the leading TRMM-era combined-sensor precipitation datasets for our hydrological studies in SEUS, we identified systematic anomalies of rainfall retrieval over inland pixels containing small water bodies, such as rivers, lakes and reservoirs. Such anomalies were not discovered in previous studies when the rainfall data were analyzed with long-term averages or over areas too large to reveal spatial details [e.g., *Gottschalck et al.*, 2005; *Ebert et al.*, 2007], or when the datasets' spatial resolution itself is not enough to resolve small water bodies (e.g., GPCP [*Huffman et al.*, 1997]). The anomalies over most sub-footprint-scale inland water bodies are evident when inspected at a daily time scale, and significant on a regional spatial scale such as SEUS. This suggests that at high spatial resolutions, sub-footprint-scale water bodies need separate and careful treatment in PMW-based algorithms, in order to produce highly consistent rainfall estimates even under favorable land surface conditions such as in SEUS.

2. Data

[7] The two high resolution satellite-based combination datasets we studied are the TRMM multi-satellite precipitation analysis (TMPA) research dataset produced at Goddard Space Flight Center, NASA (3B42, version 6) [*Huffman et al.*, 2007], and the NOAA Climate Prediction

[1]Goddard Earth Sciences and Technology Center, University of Maryland at Baltimore County, Baltimore, Maryland, USA.
[2]Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.

Copyright 2007 by the American Geophysical Union.
0094-8276/07/2007GL030787

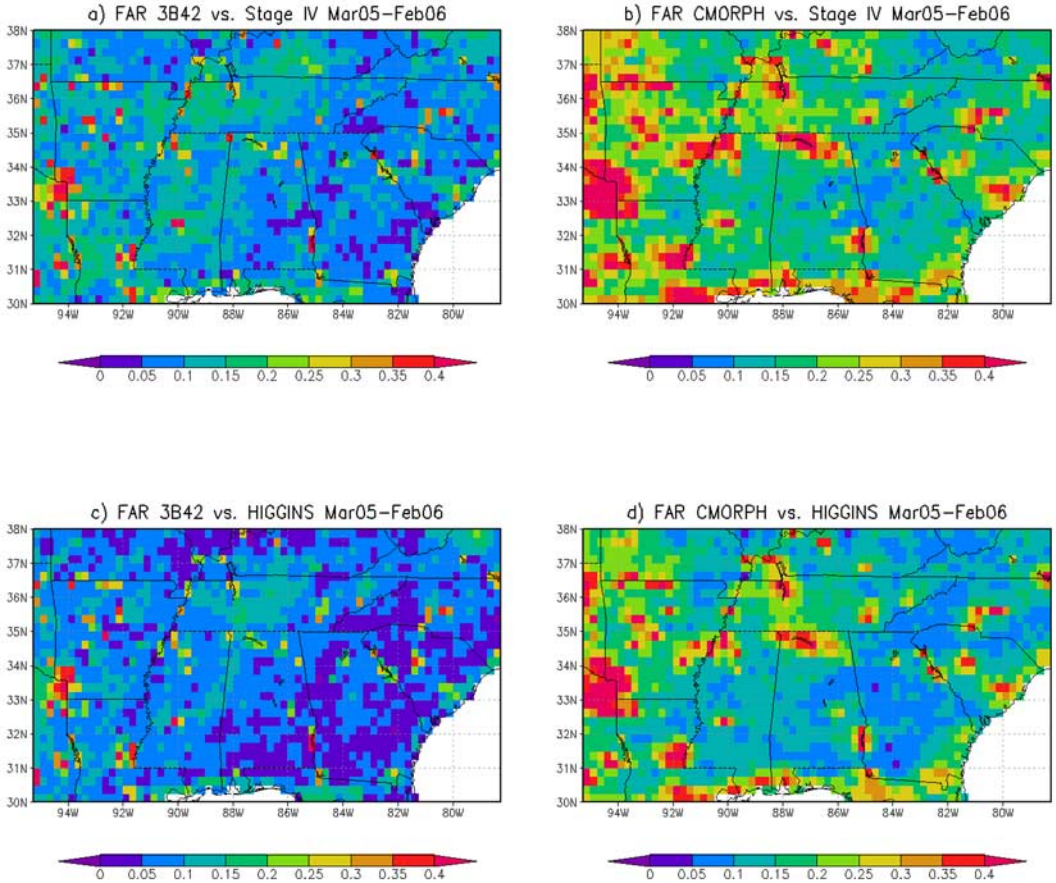

Figure 1. False alarm ratios (FAR) of (left) 3B42 and (right) CMORPH, computed against ground based (top) Stage IV radar and (bottom) Higgins gauge data, for the last year of the 3-year study period, March 2005 through February 2006. A threshold rain rate of 0.1 mm/day is used.

Center (CPC) Morphing technique product (CMORPH) [*Joyce et al.*, 2004].

[8] Both 3B42 and CMORPH derive their precipitation estimates primarily by merging most PMW scans available from the array of PMW sensors including TMI, SSM/I, AMSU-B and the more recent AMSR-E. Both datasets use infrared (IR) data from geostationary satellites as well, to handle PMW coverage gaps in different ways. The TRMM 3B42 product uses PMW-calibrated IR-based rainfall estimates directly, to fill PMW coverage gaps [*Huffman et al.*, 2007], whereas CMORPH uses the high resolution IR imagery to infer the motion of rainfall patterns between PMW scans, and use this advection information to obtain a smooth "morphing" of PMW rain patterns between PMW snapshots [*Joyce et al.*, 2004]. In addition, 3B42 incorporates monthly surface gauge measurement information for bias correction, while CMORPH is purely satellite-derived, with no ground-based measurement incorporated. TRMM 3B42 collection starts from 1998, while CMORPH starts from December 2002, both with a spatial resolution of 0.25 by 0.25 degree and a time resolution of 3 hours (although a higher-resolution 8-km, 30-minute version is also available).

[9] We used two ground-based precipitation estimates as reference data to evaluate the two satellite-based datasets. One is the NCEP Stage IV data [*Lin and Mitchell*, 2005], which are primarily based on the Next-Generation Weather Radar (NEXRAD) measurements, optimally merged with hourly gauge reports based on the Multi-sensor Precipitation Estimator (MPE) [*Seo*, 1998] algorithm, and with manual quality control. This hourly dataset has a spatial resolution of approximately 4 km. The other is the NCEP CPC near-real-time daily precipitation analysis [*Higgins et al.*, 2000], denoted here as "Higgins", which is a daily, 0.25 by 0.25-degree product. This dataset is mainly derived from the daily reports of 6000–7000 rain gauges from NWS River Forecast Centers over US, with quality control measures including duplicate station checks, buddy checks and standard deviation checks against climatology.

[10] We selected an overlapping time-span of three complete years for CMORPH and 3B42, from March 2003 through February 2006, as our study period. All the datasets are reprojected if necessary onto a common 0.25 by 0.25-degree spatial grid and aggregated to daily accumulation.

3. Results

[11] The anomalous signature of rainfall estimates is readily observable from the False Alarm Ratio (FAR) analysis. A false alarm is a day on which a satellite-based estimate reports rain while the ground reference does not, and the false alarm ratio is the ratio between the total false alarms and the total raining days detected by the satellite data. The FAR for the last year (Mar. 2005–Feb. 2006) of the study period is shown as an example in Figure 1.

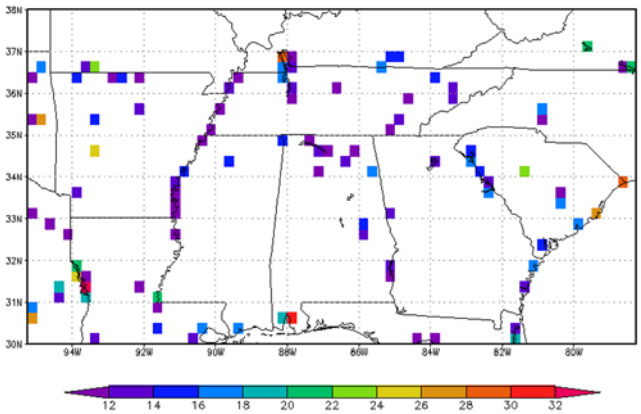

Figure 2. Pixels with 10% to 35% of total area covered by water bodies. The color bar shows the fraction of water-body area (percent). Pixels with over 35% of water-body coverage are mostly over the ocean and a few along the coast, and are masked out from the domain of our study as ocean. There are a total of 118 such pixels in this region.

[12] In the SEUS region shown, FAR shows exceptionally high values over scattered land patches, especially in the CMORPH data (Figures 1b and 1d). Such features are also present in 3B42 (Figures 1a and 1c), but are less pronounced and more confined in space. Such features are present when evaluated against either Stage IV or Higgins as the reference, with slight variations. Further inspection reveals that most of these FAR anomalies are located over pixels containing small water bodies. We illustrate in Figure 2 the 0.25 by 0.25-degree pixels that contain more than 10% of water surface area, determined from a much higher resolution (1 km) land cover dataset [*Hansen et al.*, 2000]. If one compares Figures 1 and 2, the collocation of high FAR pixels and water-body-containing ones is evident. For example, the pixels covering Walter F. George Reservoir (on lower Alabama-Georgia border), Pickwick Lake (northwest corner of Alabama), or Thurmond Lake (near the middle of Georgia-South Carolina border), are consistently associated with much higher FAR values (as much as 40%) for both 3B42 and CMORPH. These FARs are many times higher than the surrounding land-only areas, mostly with FAR values of 10–20%. A pixel does not need to contain a large percentage of water-body area: 10–15% of water surface in a pixel is enough to make a big difference (e.g., Walter F. George Reservoir).

[13] TRMM 3B42 has slightly lower FAR values overall in these water-body areas, with smaller spatial extent, yet the FAR values are still significantly higher than their surrounding areas. The FAR patterns for the other two years of our study period show qualitatively similar features, with smaller magnitudes. Further inspection of Figure 1 reveals similar anomalies for pixels along coastlines, but this issue is recognized and known to be related to joining the retrievals from the different land-based and ocean-based PMW algorithms [*Adler et al.*, 1993].

[14] The anomalously high FAR values over water-body pixels can be further diagnosed by computing the four conventional event-based measures: hits, misses, zeros (also called nulls) and false alarms, as shown in Table 1. For comparison, we also randomly picked the same number (118) of "land-only" pixels in the same SEUS region, defined as pixels containing less than 10% of water-body area. Table 1 shows there are no significant differences in total number of hits and misses between land and water-body pixels. However, over water body pixels, the number of false alarms is consistently higher than its counterpart over land pixels. Over the 3 years, 3B42 has more than twice the false alarms over water body pixels as over land pixels (95 vs. 38), and CMORPH has 50% more false alarms. The difference in year 3 is most outstanding, with 31 vs. 10, and 60 vs. 26 false alarms for 3B42 and CMORPH, respectively. The increased number of false alarms over water bodies is compensated mostly by the reduced number of zeros.

[15] The anomalous rainfall signature over water bodies is inherent in the satellite estimates, independent of ground-based reference data. Figure 3 shows the number of raining days averaged over the total number of pixels (118), as a function of rain rate. Both 3B42 and CMORPH exhibit

Table 1. Accuracy Measures of Event-Based Statistics: Total Annual Hits, Misses, Zeros and False Alarms, With 3B42 and CMORPH Evaluated Against Stage IV[a]

Surface Type	Study Period	Hits, Days	Misses, Days	Zeros, Days	False Alarms, Days	Total False Alarms, Days
			3B42			
Land	Year 1	102	45	205	13	38
	Year 2	104	51	195	15	
	Year 3	90	55	211	10	
Water-body	Year 1	94	49	198	24	95
	Year 2	106	46	173	40	
	Year 3	84	56	194	31	
			CMORPH			
Land	Year 1	122	25	194	24	78
	Year 2	124	31	181	28	
	Year 3	107	37	195	26	
Water-body	Year 1	115	28	194	28	127
	Year 2	121	31	174	39	
	Year 3	107	33	165	60	

[a]A hit (zero) is a day when both a satellite estimate and its ground reference report a rain (no-rain) day. A miss (false alarm) is a day when a satellite estimate reports a no-rain (rain) day while its ground reference reports otherwise. A threshold rain rate of 0.1 mm/day is used.

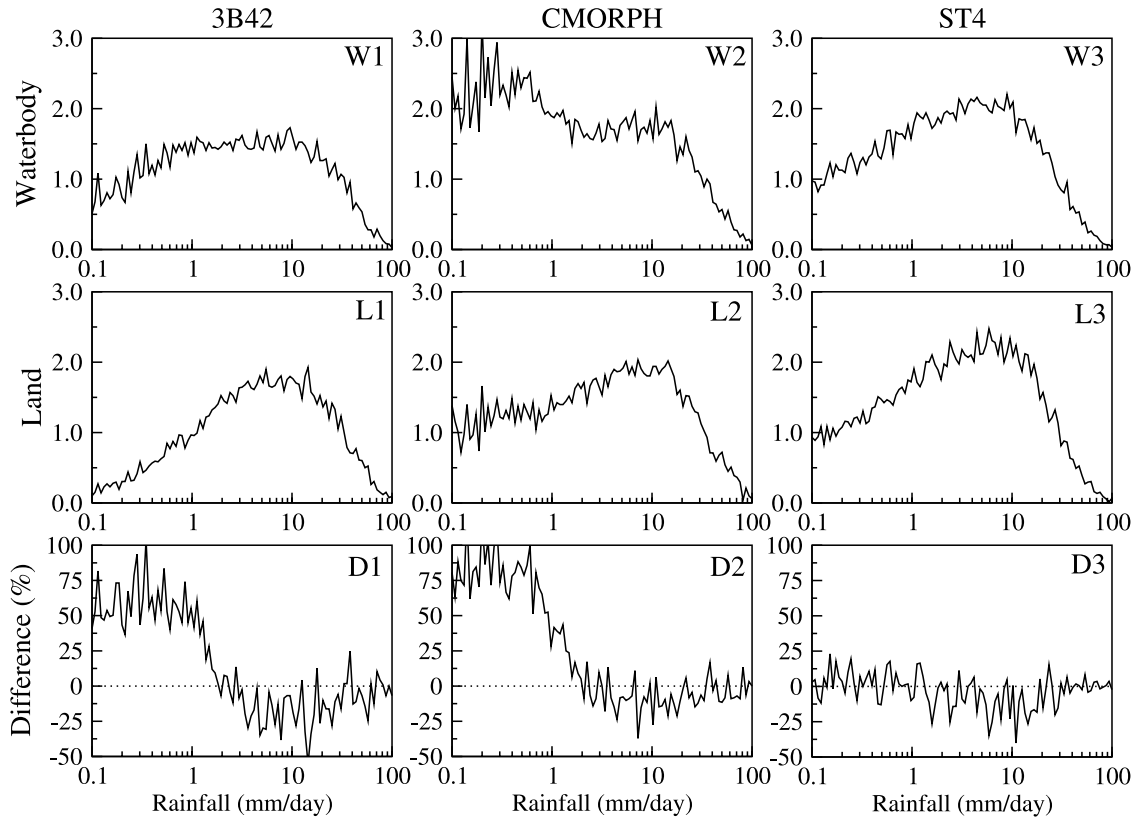

Figure 3. Rain rate distribution of the annual number of rain events, averaged over the total number (118) of (top) water-body and (middle) land pixels for 3B42 and CMORPH, for the period of Mar. 2005 through Feb. 2006. Stage IV data are also shown for comparison. (bottom) The differences in the distributions between water bodies and land are also shown as a percentage of total averaged raining days over the land pixels. The 0.1–100 mm/day range on the abscissa is discretized into 100 bins in a logarithmic scale for the accounting of the raining days.

remarkable differences in their raining-day distribution between water-body (top) and land pixels (middle). In particular, over water bodies, both 3B42 and CMORPH show much broader distributions, with a much higher number of raining days at lower daily rainfall intensities (<2 mm/day) than the land pixels in this region. For example, in the light rain range (0.1–1 mm/day), 3B42 has about 50% more raining days over water-bodies, and CMORPH has about 75% more (Figure 3 (bottom)). At higher rainfall intensities (>2 mm/day), the difference becomes insignificant. For comparison, Stage IV distributions are also shown over the same sets of water-body and land pixels, and they do not possess such dramatic differences (Figure 3 (right)).

4. Discussion

[16] We speculate that the systematic rainfall anomalies over water bodies in the SEUS region are attributed to the individual PMW-based input components in both datasets. PMW-based retrievals are known to be sensitive to land surface heterogeneity, including contrasts in temperature and emissivity. Such anomalies are unlikely to result from climatological differences over different land surface types, because the water bodies in our study are relatively small-scale (~10-km) features. In addition, the Stage IV radar data do not show systematic differences over the same set of water-body and land pixels in this region (Figure 3).

[17] It is more likely these anomalies are caused by the poor characterization of the differences in emissivity and temperature of water surfaces in the PMW frequencies used by the retrievals. The water bodies produce a rainfall-like signature when higher-frequency PMW channels are used for scattering-based algorithms tuned to land surfaces [e.g., *Grody*, 1991]. We also examined rainfall distributions in the merged products over ocean pixels around the SEUS region (not shown here), and they do not have the anomalous light-rain boost seen in Figure 3. That provides additional support for our hypothesis that the PMW algorithms over land are affected by the small inland water bodies shown in Figure 2. Recently, we were informed of improvements to Version 6 of the GPROF algorithm [*Kummerow et al.*, 2001] since November 2006 (R. Joyce, personal communication, 2007), and we therefore repeated the same analysis for the period of November 2006 through March 2007. The anomalies still exist with a similar magnitude. Further studies are needed to trace the contribution of individual PMW sensors to these errors contained in these multi-sensor products.

[18] Though our study is limited to the SEUS region, we suspect other regions suffer this issue as well. However, in most areas, other factors, such as topography, snow cover, precipitation regimes (convective vs. stratiform), as well as ground reference data quality, may mask this effect [*Tian et*

al., 2007]. In addition, since the anomalies are more apparent over small water bodies and in light rain intensities, they will usually have modest impact on large scale, low resolution applications. But for small scale, high resolution applications, such as satellite algorithm validation, watershed simulation, and river flow management, the issue documented in this study may not be ignored.

[19] **Acknowledgments.** This research was supported in part by the NASA Precipitation Measurement Missions Program and the Terrestrial Hydrology Program under solicitation NRA-02-OES-05 (PI: Peters-Lidard) and the Air Force Weather Agency MIPR F2BBAJ6033GB01 (PI: Peters-Lidard). The authors wish to thank George Huffman, Robert Adler, Robert Joyce, and John Janowiak for helpful discussions and Ying Lin and Yelena Yarosh for assistance with data access and questions.

References

Adler, R. F., P. R. Keehn, and I. M. Hakkainen (1993), Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data, *J. Appl. Meteorol.*, *32*, 335–356.

Ebert, E. E., J. E. 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.

Ferraro, R. R., E. A. Smith, W. Berg, and G. J. Huffman (1998), A screening methodology for passive microwave precipitation retrieval algorithms, *J. Atmos. Sci.*, *55*, 1583–1600.

Gottschalck, J., J. Meng, M. Rodell, and P. Houser (2005), Analysis of multiple precipitation products and preliminary assessment of their impact on Global Land Data Assimilation System land surface states, *J. Hydrometeorol.*, *6*, 573–598.

Grody, N. C. (1991), Classification of snow cover and precipitation using the Special Sensor Microwave Imager, *J. Geophys. Res.*, *96*, 7423–7435.

Hansen, M. C., R. S. DeFries, J. R. Townsend, and R. Sohlberg (2000), Global land cover classification at 1 km spatial resolution using a classification tree approach, *Int. J. Remote Sens.*, *21*, 1331–1364.

Higgins, R. W., W. Shi, and E. Yarosh (2000), Improved United States precipitation quality control system and analysis, in *NCEP/Climate Prediction Center Atlas*, vol. 7, 40 pp., Natl. Cent. for Environ. Predict., Camp Springs, Md.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolph, and U. Schneider (1997), The global precipitation climatology project (GPCP) combined precipitation dataset, *Bull. Am. Meteorol. Soc.*, *78*, 5–20.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM Multi-satellite Precipitation Analysis (TMPA): Quasi-global, multi-year, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, *8*, 38–55.

Joyce, R. J., J. E. Janowiak, P. A. 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., D. R. Kniveton, M. C. Todd, and T. J. Bellerby (2003), Satellite rainfall estimation using combined passive microwave and infrared algorithms, *J. Hydrometeorol.*, *4*, 1088–1104.

Kummerow, C. Y., et al. (2001), The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors, *J. Appl. Meteorol.*, *40*, 1801–1840.

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

Seo, D. J. (1998), Real-time estimation of rainfall fields using radar rainfall and rain gauge data, *J. Hydrol.*, *208*, 37–52.

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

Tian, Y., C. D. Peters-Lidard, B. J. Choudhury, and M. Garcia (2007), Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications, *J. Hydrometeorol*, in press.

Turk, F. J., and S. D. Miller (2005), Toward improving estimates of remotely-sensed precipitation with MODIS/AMSR-E blended data techniques, *IEEE Trans. Geosci. Remote Sens.*, *43*, 1059–1069.

C. D. Peters-Lidard, Hydrological Sciences Branch, NASA Goddard Space Flight Center, Mail Code 614.3, Greenbelt, MD 21042, USA.

Y. Tian, GEST Center, Mail Code 614.3, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (yudong.tian@nasa.gov)