GEOPHYSICAL RESEARCH LETTERS, VOL. 36, L17808, doi:10.1029/2009GL040000, 2009

Improving the global precipitation record: GPCP Version 2.1

George J. Huffman,[1,2] Robert F. Adler,[1,3] David T. Bolvin,[1,2] and Guojun Gu[1,3]

Received 9 July 2009; revised 11 August 2009; accepted 12 August 2009; published 10 September 2009.

[1] The GPCP has developed Version 2.1 of its long-term (1979–present) global Satellite-Gauge (SG) data sets to take advantage of the improved GPCC gauge analysis, which is one key input. As well, the OPI estimates used in the pre-SSM/I era have been rescaled to 20 years of the SSM/I-era SG. The monthly, pentad, and daily GPCP products have been entirely reprocessed, continuing to require that the submonthly estimates sum to the monthly. Version 2.1 is close to Version 2, with the global ocean, land, and total values about 0%, 6%, and 2% higher, respectively. The revised long-term global precipitation rate is 2.68 mm/d. The corresponding tropical (25°N-S) increases are 0%, 7%, and 3%. Long-term linear changes in the data tend to be smaller in Version 2.1, but the statistics are sensitive to the threshold for land/ocean separation and use of the pre-SSM/I part of the record.
Citation: Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the global precipitation record: GPCP Version 2.1, *Geophys. Res. Lett.*, *36*, L17808, doi:10.1029/2009GL040000.

1. Introduction

[2] The Global Precipitation Climatology Project (GPCP) is charged with developing global long-term records of precipitation for the international community on behalf of the World Meteorological Organization/World Climate Research Programme/Global Energy and Water Experiment (WMO/WCRP/GEWEX). GPCP is one of several GEWEX global analyses of components of the water and energy cycle organized under the GEWEX Radiation Panel. Such data are essential for quantifying the global water cycle, supporting verification of numerical models, and providing the background climate statistics for many operational water-resource activities. The GPCP datasets are developed and maintained as an international activity. Input datasets are provided by several contributing groups: Retrievals from Special Sensor Microwave/Imager data (SSM/I; Defense Meteorological Satellite Program, U.S.) are provided by L. Chiu (Chinese Univ. of Hong Kong and George Mason Univ., U.S.) and R. Ferraro (National Environmental Satellite Data and Information Service, NESDIS, U.S.). Merged geosynchronous- and low-Earth-orbit infrared (geo- and leo-IR) data are provided by P. Xie (National Oceanic and Atmospheric Administration/Climate Prediction Center, NOAA/CPC, U.S.) using data contributed by NESDIS, the Japanese Meteorological Agency, and the European Organisation for the Exploitation of Meteorological Satellites. As well, P. Xie provides the Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) estimates from NOAA leo-IR data. J. Susskind (National Aeronautics and Space Administration/Goddard Space Flight Center, NASA/GSFC, U.S.) contributes estimates based on the Television Infrared Observation Satellite Operational Vertical Sounder (TOVS; provided by NESDIS) and the Advanced Infrared Sounder (AIRS; provided by NASA). Global precipitation-gauge analyses are provided by B. Rudolf and U. Schneider (Global Precipitation Climatology Centre, GPCC, hosted at the Deutscher Wetterdienst) using data from most countries around the world. The final computations are carried out by the group led by G. Huffman and R. Adler (NASA/GSFC) and the primary archive is hosted by World Data Center A (National Climatic Data Center, U.S.).

[3] It is critical that the dataset be carefully monitored for changes in the input data sources that might cause inhomogeneities in the long-term record. Recently, such an instance arose when the GPCC upgraded their precipitation gauge analysis scheme to a climate anomaly method. This should provide better results, particularly in undersampled regions, but it also would have introduced a jump discontinuity in the GPCP products that was judged unacceptable. To maintain homogeneity, the GPCP determined that the entire record should be reprocessed with the new gauge analysis. After sketching the reprocessing steps, a short analysis of the changes from Version 2 to Version 2.1 of the monthly satellite-gauge (SG) product is presented. A parallel analysis is not needed for the GPCP pentad and daily products (*Xie et al.* [2003] and *Huffman et al.* [2001], respectively) because they have not changed beyond being rescaled to the Version 2.1 monthly product. We denote this first revision as Version 1.1 for both the pentad and daily products. Throughout, "Version 2" and "Version 2.1" refer to the GPCP monthly products unless otherwise specified.

2. Revisions for Version 2.1

[4] The processing strategy for Version 2.1 is substantially the same as described for Version 2 of *Adler et al.* [2003], so the summary here focuses on the differences. In general terms, multiple sources of satellite data are combined, according to their periods of availability and estimated accuracies, then the monthly SG product is computed by combining the multi-satellite estimate with a precipitation gauge analysis provided by the GPCC.

[5] The primary reason for the upgrade to Version 2.1 is GPCC's shift to a new climate anomaly analysis method for precipitation gauge data [*Schneider et al.*, 2008]. Precipitation gauge reports are archived from a time-varying collection of over 70,000 stations around the globe, both from

[1]Global Precipitation Climatology Project, Laboratory for Atmospheres, Global Merge Development Centre, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[2]Science Systems and Applications, Inc., Lanham, Maryland, USA.
[3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA.

0094-8276/09/2009GL040000$05.00

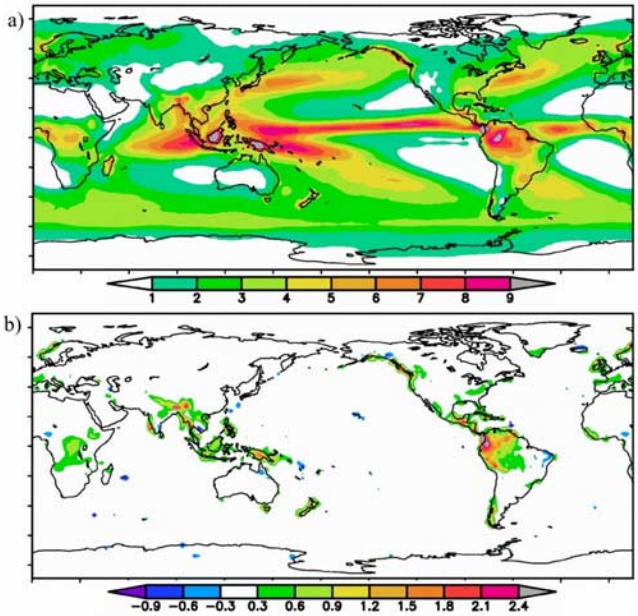

Figure 1. (a) Thirty-year (1979–2008) climatology for GPCP Version 2.1 in mm/d, and (b) (Version 2.1–Version 2) difference averaged over 1979–2007 in mm/d.

Global Telecommunications System (GTS) reports, and from other world-wide or national data collections. The extensive quality-control system, featuring both an automated screening and then a manual check, feeds into a two-step analysis. First, a long-term climatology is assembled from all available gauge data, focusing on the period 1951–2000. The lack of complete consistency in period of record for individual stations has been shown to be less important than the gain in detail that adding more stations provides, particularly in complex terrain. Then for each month, the individual gauge reports are converted to deviations from climatology, and are analyzed into gridded values using a variant of the SPHEREMAP spatial interpolation routine [*Willmott et al.*, 1985]. Finally, the month's precipitation values (in mm/month) are produced by superimposing the anomaly analysis on the month's climatology.

[6] The GPCC creates multiple products, and two are used in the GPCP Version 2.1. The Full Data Reanalysis (Version 4) is a retrospective analysis that covers 1901–2007, and it is used in GPCP for 1979–2007. Thereafter we use the GPCC Monitoring Product (Version 2), whose data source is limited to GTS reports. The advantages of these changes are that: 1) we no longer need to use the separate and differently prepared gauge analysis based on the Global Historical Climate Network and Climate Analysis and Monitoring System (GHCN + CAMS) for 1979–1985, and 2) the numbers of gauges used are much higher for much of the record. When the Full Data Reanalysis is updated to a longer record in the future we expect to reprocess the GPCP datasets to take advantage of the improved data. We continue the GPCP's long-standing practice of correcting all gauge analysis values for climatological estimates of systematic error due to wind effects, side-wetting, evaporation, etc., once again using the *Legates* [1987] coefficients.

[7] During the period before SSM/I data became available the GPCP uses the Outgoing Longwave Radiation (OLR) Precipitation Index (OPI) product [*Xie and Arkin*, 1998], calibrated to an overlap period with the GPCP monthly product in the SSM/I era. The OPI depends on the correlation of colder OLR radiances to higher cloud tops, and thus increased precipitation rates. It is necessary to define "cold" locally, so a regression relationship is developed for anomalies in both OLR and precipitation. In use, the total precipitation inferred is the estimated anomaly plus the local climatological value. A backup direct OLR-precipitation regression is used when the anomaly approach yields unphysical values. In Version 2 the calibration was computed for 1988–1995, but for Version 2.1 it was extended to 1988–2007. This spatially and seasonally varying climatological calibration is then applied to the independent OPI data covering the span 1979–1987 to fill all months lacking SSM/I data.

[8] Our work in developing Version 2 for the pre-SSM/I era showed calibration-induced biases for the OPI precipitation from the TIROS-N (January 1979–January 1980) and NOAA-6 (February 1980–August 1981) satellites, while NOAA-7 (September 1981–February 1985) results appeared relatively unbiased. This is true even though the biases in all of the OLR data are small (less than 1%), and this continued to be true for Version 2.1. Accordingly, we re-applied the scheme used in Version 2 to adjust the bias of the first two satellites. Matched satellite and gauge precipitation estimates are separately accumulated for each individual satellite over all gridboxes having 1) a valid OPI value, 2) at least one gauge in the gridbox, and 3) a gauge estimate of at least 50 mm/month, for all months of TIROS-N, NOAA-6, and NOAA-7. Each satellite's ratio of the OPI accumulation to the gauge accumulation is computed. Taking the NOAA-7 OPI-gauge ratio as representative, and assuming that the OPI bias over ocean is similar to that over land, a ratio correction is applied for all grid boxes to the TIROS-N and NOAA-6 data to match the ratio of the NOAA-7 period. The same computation using an alternative OLR data set [*Lee et al.*, 2007] shows very similar results, and confirms that biases are consistent between land and ocean. Nonetheless, the first two satellites still appear to be biased low and will be re-examined in the next upgrade. The new Version 2.1 adjustments for the TIROS-N and NOAA-6 periods are +8% and +0.4%, compared to +12% and +3% in Version 2.

[9] The Version 2.1 reprocessing also provided an opportunity to apply several corrections to input datasets. These include: (1) Substituting a full month of AIRS data for a partial month of TOVS data in February 2004, (2) Including $5^\circ \times 5^\circ$ SSM/I emission-based estimates (i.e., over ocean) for July 1990 through December 1991, completing the $5^\circ \times 5^\circ$ time series for use as fill-in when the usual $2.5^\circ \times 2.5^\circ$ product failed to converge, (3) Corrections in the mid-Pacific overlap region between geo-IR satellites for October and November 1994.

[10] On the other hand, we continue to use the same OPI estimation scheme for Version 2.1 as in Version 2, so the

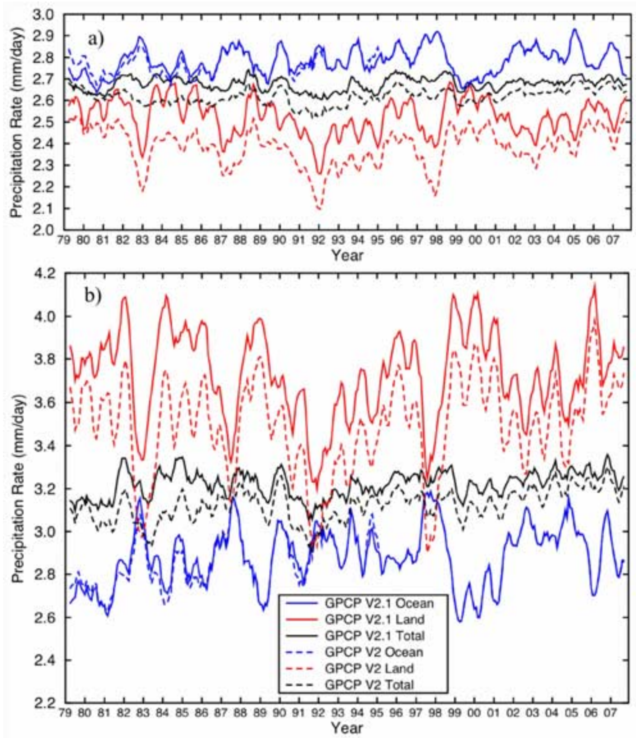

Figure 2. Time series of (a) global-average and (b) tropical-average land, ocean, and total precipitation (red, blue, black) for Versions 2.1 (solid) and 2 (dashed) for the entire study period (1979–2007) in mm/d. ''Ocean'' and ''land'' regions are defined by 100% and <100% coverage by water, and the monthly data have been smoothed with a seven-point running boxcar filter.

spatial and temporal variances are still underestimated in months when SSM/I data are unavailable.

3. Comparison of Versions 2 and 2.1

[11] The global climatologies for the two versions are quite similar, so only Version 2.1 is displayed in Figure 1a, while the mean difference map is shown in Figure 1b. Differences over oceanic regions are generally small, representing a compromise between essentially-zero differences in the SSM/I era and the mean differences during the pre-SSM/I era. The freckles of difference in the oceans mostly correspond to island locations used in the previous GPCC Monitoring Product that have been eliminated from the new Full Data Reanalysis. Land regions have more substantial differences, mostly due to the mean differences

Table 1. Global- and Tropical-Average Land, Ocean, and Total Precipitation for Versions 2.1 and 2 in mm/d[a]

	Global 90°N–90°S		Tropical 25°N–25°S	
1979–2007 Version	2	2.1	2	2.1
Land and Ocean	2.62	2.68 (+2%)	3.12	3.22 (+3%)
Land	2.39	2.53 (+6%)	3.49	3.73 (+7%)
Ocean	2.78	2.78 (+0%)	2.88	2.88 (+0%)

[a]The percentage increase of Version 2.1 over Version 2 is given in parentheses. ''Ocean'' and ''land'' regions are defined by 100% and <100% coverage by water.

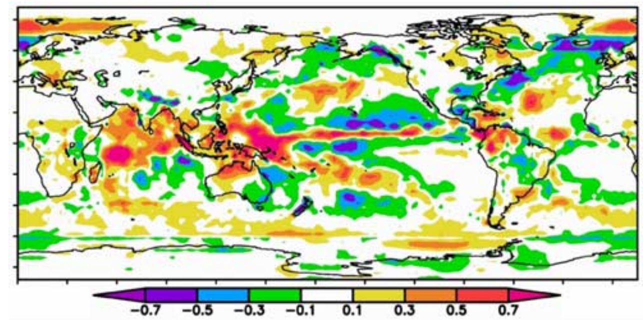

Figure 3. Map of linear change for Version 2.1 for the SSM/I era (1988–2007) in mm/d/decade.

between the versions of the gauge analyses throughout the period of record. The largest differences occur in northwestern South American and Mesoamerica. The new gauge analysis is attributing as much as 50% more precipitation to parts of this region, which is characterized by mountainous terrain. Such areas typically feature higher precipitation at higher elevations (within the limits of the 2.5° resolution), which the previous GPCC Monitoring Product tended to miss. Similar tropical topographic regimes are highlighted in Papua New Guinea, the Himalayas, and along the east coast of the Bay of Bengal. The change in central Africa is an improvement over the Version 2 data set, in which persistent gaps in gauge coverage over central Africa coincided with the climatological maximum. Under such conditions, the previous GPCC analysis scheme, and therefore the GPCP satellite-gauge product, tended to underestimate the climatological maximum month after month. At higher latitudes the major increases occur in mountainous terrain on coasts that intersect storm tracks – the Pacific coasts of northwestern North America and southern Chile, New Zealand, and the west coasts of Norway and the British Isles. Finally, the new GPCC analyses do not cover Antarctica, so the satellite adjustments in the high southern latitudes continue to be based on the (very approximate) mean gauge precipitation climatology computed in Version 2 from these gauges as contained in the previous GPCC Monitoring Product.

[12] The time series of global and tropical averages for all, land, and ocean regions (Figure 2) give insight into the aggregate time variation of the two versions. Experience has shown that such regionalization is somewhat sensitive to the choice of regions. Although the most realistic land/water distribution is provided by a threshold for coverage by water of 75%, we wish to provide a clean ''open ocean'' comparison of the data sets. Thus, throughout this paper ''ocean'' and ''land'' regions are defined as having 100% and <100% coverage by water, respectively. In both versions we see that the seasonal cycle over land, primarily driven by the boreal seasons, is almost exactly balanced by changes over the ocean on the global scale (Figure 2a), with some seasonality apparent in total precipitation for the tropics (Figure 2b).

[13] In the pre-SSM/I era (i.e., before mid-1987) the revised scaling for the OPI raises the oceanic mean for most of the period. The exception is for the first two satellites (January 1979–August 1981). Even though the same bias adjustment procedure is used in both Versions 2 and 2.1, as described above, we find that the revised OPI

Table 2a. Global- and Tropical-Average Linear Changes in mm/d/decade for the Entire Study Period (1979–2007)[a]

	Global 90°N–90°S		Tropical 25°N–25°S	
1979–2007 Version	2	2.1	2	2.1
Land and Ocean	+0.0115	+0.0069	+0.0480	+0.0252
Land	+0.0018	−0.0118	+0.0432	−0.0030
Ocean	+0.0184	+0.0199	+0.0511	+0.0438

[a]"Ocean" and "land" regions are defined by 100% and <100% coverage by water.

Table 2b. Global- and Tropical-Average Linear Changes in mm/d/decade for the SSM/I Era (1988–2007)[a]

	Global 90°N–90°S		Tropical 25°N–25°S	
1988–2007 Version	2	2.1	2	2.1
Land and Ocean	+0.0343	+0.0169	+0.0714	+0.0497
Land	+0.0630	+0.0252	+0.1327	+0.0889
Ocean	+0.0144	+0.0111	+0.0308	+0.0238

[a]"Ocean" and "land" regions are defined by 100% and <100% coverage by water.

scaling and the replacement of the GHCN + CAMS gauge analysis with the new GPCC Full Data Reanalysis work together to produce almost no change from Version 2 to Version 2.1, unlike the OPI computed from the following satellites. Over land, Version 2 contains a temporal data boundary at the start of 1986, when the GHCN + CAMS was replaced by the then-current GPCC Monitoring Product. Reasonable continuity in the Version 2 time series itself, as well as comparison with the Version 2.1 time series, reveals that the change in gauge analysis in January 1986 is relatively unimportant at the tropical or global scale, although locally there can be noticeable differences (not shown). However, within the GHCN + CAMS record there is an issue. The first few years of GPCP Version 2 are closer to the corresponding Version 2.1 data for the global-average land (Figure 2a) than any other years, confirming earlier impressions that the GHCN + CAMS was making the Version 2 land estimates in those years somewhat inconsistent with the rest of the record.

[14] The total, land, and ocean averages for each of the Versions are given in Table 1. The global and tropical regions are consistent in showing modestly higher values in Version 2.1, with essentially all of the change occurring over land (and coast, since the threshold is 100%).

[15] One convenient way to summarize time changes in the data sets is to compute the long-term linear rate of change for each grid box. Figure 3 displays the most certain results, namely Version 2.1 during the SSM/I era, while all the fields summarized below are shown in the auxiliary material.[1] Note that we compute the linear-change statistic with no assumption or implication of a particular dynamic or secular trend. Furthermore, the shift in input satellite data from OPI to SSM/I led us to compute the linear changes both for the entire data set, and for the SSM/I era (1988–2007). In general, there is consistency both between the longer and shorter period results and between the Version 2 and 2.1 results. The more precise, somewhat shorter, and more recent SSM/I-era data mostly show larger linear change values than the entire data record, while Version 2.1 shows smaller linear changes than in Version 2. See *Gu et al.* [2007] and *Adler et al.* [2008] for analyses of the Version 2 results. The global- and tropical-average linear changes for both versions are listed in Tables 2a and 2b. The increase in linear change from Version 2 to Version 2.1 for global ocean across the entire data set, which is the only such increase in Tables 2a and 2b, is driven by the somewhat questionable behavior of the OPI in the first 2.5 years of the data record. The revisions to the linear change over land from Version 2 to 2.1 are driven by the gauge data and tend to be focused in the regions previously noted for having large mean differences between the two Versions. As stated previously, the statistics under discussion are somewhat sensitive to the definition of land and ocean. These subtle, but important changes in the data set will be the subject of further analysis.

[1]Auxiliary materials are available in the HTML. doi:10.1029/2009GL040000.

4. Concluding Remarks

[16] It was necessary to develop Version 2.1 of the GPCP data sets to prevent a discontinuity due to changes to the new GPCC gauge analysis scheme that forms a key part of the GPCP input data. The only substantive changes are: use of the new GPCC Full Data Reanalysis (Version 4) for 1979–2007 and the new GPCC Monitoring Product (Version 2) thereafter, and recalibration of the OPI data to a longer (20-year) record of the new SSM/I-era GPCP data. All GPCP products have been reprocessed for their entire periods of record, including the monthly, pentad, and daily products, and are labeled Version 2.1, 1.1, and 1.1, respectively. The estimated global average for Version 2.1 monthly data is 2.68 mm/d, about 2% higher than in Version 2, with changes almost entirely over land. Linear changes over the data set tend to be smaller in Version 2.1, particularly for the more certain SSM/I era.

[17] Development work in GPCP will now focus on the upcoming Version 3, which is planned to feature finer time and space resolution. At the same time, it is likely that the current GPCP products will continue to be produced for some time after the release of Version 3 to provide continuity for the user community.

References

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, *4*, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Adler, R. F., G. Gu, J.-J. Wang, G. J. Huffman, S. Curtis, and D. T. Bolvin (2008), Relationships between global precipitation and surface temperature on inter-annual and longer time scales (1979–2006), *J. Geophys. Res.*, *113*, D22104, doi:10.1029/2008JD010536.

Gu, G., R. F. Adler, G. J. Huffman, and S. Curtis (2007), Tropical rainfall variability on interannual-to-interdecadal/longer-time scales derived from the GPCP monthly product, *J. Clim.*, *20*, 4033–4046, doi:10.1175/JCLI4227.1.

Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. Susskind (2001), Global precipitation at one-degree daily resolution from multisatellite observations, *J. Hydrometeorol.*, *2*, 36–50, doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2.

Lee, H.-T., A. Gruber, R. G. Ellingson, and I. Laszlo (2007), Development of the HIRS outgoing longwave radiation climate dataset, *J. Atmos. Oceanic Technol.*, *24*, 2029–2047, doi:10.1175/2007JTECHA989.1.

Legates, D. R. (1987), A climatology of global precipitation, *Climatology*, *40*, 85.

Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf (2008), Global precipitation analysis products of the GPCC, 12 pp., Deutscher Wetterdienst, Offenbach am Main, Germany. (Available at ftp://ftp-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf)

Willmott, C. J., C. M. Rowe, and W. D. Philpot (1985), Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring, *Am. Cartogr.*, *12*, 5–16, doi:10.1559/152304085783914686.

Xie, P., and P. A. Arkin (1998), Global monthly precipitation estimates from satellite-observed outgoing longwave radiation, *J. Clim.*, *11*, 137–164, doi:10.1175/1520-0442(1998)011<0137:GMPEFS>2.0.CO;2.

Xie, P., J. E. Janowiak, P. A. Arkin, R. F. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis (2003), GPCP Pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates, *J. Clim.*, *16*, 2197–2214, doi:10.1175/2769.1.

R. F. Adler, D. T. Bolvin, G. Gu, and G. J. Huffman, NASA Goddard Space Flight Center, Code 613.1, Greenbelt, MD 20771, USA. (george.j.huffman@nasa.gov)