Improving the Global Precipitation Record: GPCP Version 2.1

George J. Huffman¹,², Robert F. Adler¹,³, David T. Bolvin¹,², Guojun Gu¹,³

1: NASA Goddard Space Flight Center, Laboratory for Atmospheres;
   GPCP Global Merge Development Centre
2: Science Systems and Applications, Inc.
3: University of Maryland College Park, ESSIC

Submitted to Geophysical Research Letters

9 July 2009

Corresponding author:
George J. Huffman
NASA/GSFC Code 613.1
Greenbelt, MD 20771
Phone: 301-614-6308
Fax: 301-614-5492
Email: george.j.huffman@nasa.gov
Abstract

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 enforce consistency of the submonthly estimates 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.
1. Introduction

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 (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-orbit–satellite infrared 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. 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 (provided by NESDIS) and the Advanced Infrared Sounder (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.).

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 climatology-anomaly approach. This should provide better results, particularly in undersampled, 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. The same information about the GPCP pentad and daily products ([Xie et al. 2003] and [Huffman et al. 2001], respectively) is not necessary because they have not changed beyond being rescaled to the Version 2.1 monthly product. Throughout, “Version 2” and “Version 2.1” refer to the GPCP products unless otherwise specified.

2. Revisions for Version 2.1

The processing strategy for Version 2.1 is substantially the same as described for Version 2 in 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 satellite-gauge (SG) product is computed by combining the multi-satellite estimate with a precipitation gauge analysis provided by the GPCC.
The primary reason for the upgrade to Version 2.1 is GPCC’s shift to a new climatology-anomaly analysis technique 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 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, 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 analysis is produced by superimposing the anomaly analysis on the month’s climatology.

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 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., following Legates [1987].

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.

Our work in developing Version 2 for the pre-SSM/I era showed calibration-induced biases for the OPI precipitation from TIROS-N (January 1979 – January 1980) and NOAA-6 (February 1980 – August 1981). This is true even though the biases in the OLR itself 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. The precipitation was averaged for each satellite separately over all gridboxes having a valid OPI value, at least one gauge/gridbox, and a gauge estimate of at least 50 mm/month, for all months of TIROS-N, NOAA-6, and NOAA-7 (September 1981 – February 1985). The same averaging is applied to the corresponding gauge estimates for the three periods and compared with the three satellite estimates. The ratios of the satellite averages versus the gauge data were computed. Using the NOAA-7 OPI-gauge ratio as representative, since it appears to be minimally biased, and
assuming that the OPI bias over ocean is similar to that over land, a ratio correction was applied for all grid boxes to the TIROS-N and NOAA-6 data to match the ratio of the NOAA-7 period. Comparison to 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.

The Version 2.1 reprocessing also provided an opportunity to apply several corrections to input datasets. These include:

- substituting a full month of AIRS data for a partial month of TOVS data in February 2004;
- including 5°x5° SSM/I emission-based estimates (i.e., over ocean) for July 1990 through December 1991, completing the 5°x5° time series for use as fill-in when the usual 2.5°x2.5° product failed to converge; and
- corrections in the mid-Pacific overlap region between geo-IR satellites for October and November 1994.

On the other hand, we continue to use the same OPI estimation scheme for Version 2.1 as in Version 2, so the spatial and temporal variances are still underestimated in months when SSM/I data are unavailable.

3. Comparison of Versions 2 and 2.1

The global climatologies for the two versions are quite similar, so only Version 2.1 is displayed in Fig. 1a, while the mean difference map is shown in Fig. 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 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 large gradients in topography. Such gradients 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 steep terrain on coasts that intersect storm tracks – the Pacific coasts of northwestern North America and southern Chile, and New Zealand. 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.

The time series of global and tropical averages for all, land, and ocean regions (Fig. 2) give insight into the aggregate time variation of these differences. 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. A
seven-point boxcar running smoother has been applied to suppress short-interval noise. As in
previous GPCP 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 (Fig. 2a), with
some seasonality apparent in total precipitation for the tropics (Fig. 2b).

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 scaling and the replacement of the
GHNC+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 following OPI satellites.
Over land, Version 2 contains a data boundary at the start of 1986, when the GHNC+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
GHNC+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 (Fig. 2a) than any other years,
confirming earlier impressions that the GHNC+CAMS was making the Version 2 land estimates
in those years somewhat inconsistent with the rest of the record.

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%).

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. Fig. 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
Electronic Data Supplement. 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 Table 2. 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 Table 2, 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.
4. Concluding Remarks

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 were: 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 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. The pentad and daily GPCP data sets are now considered to be at Version 1.1, since this is the first reprocessing since they were released. Version 2.1 provides an estimate of the global total of 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, particularly for the more certain SSM/I era.

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 Version 2.1 products will continue to be produced for some time after the release of Version 3 to provide continuity for the user community.


Table Captions

Table 1 Global- and tropical-average land, ocean, and total precipitation for Versions 2.1 and 2 in mm/d. 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.

Table 2 Global- and tropical-average linear changes in mm/d/decade a) for the entire study period (1979-2007), and b) for the SSM/I era (1988-2007). “Ocean” and “land” regions are defined by 100% and <100% coverage by water.

Figure Captions

Fig. 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.

Fig. 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.

Fig. 3 Map of linear change for Version 2.1 for the SSM/I era (1988-2007) in mm/d/decade.
Table 1 Global- and tropical-average land, ocean, and total precipitation for Versions 2.1 and 2 in mm/d. 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.

<table>
<thead>
<tr>
<th>1979-2007</th>
<th>Global 90°N-90°S</th>
<th>Tropical 25°N-25°S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>2</td>
<td>2.1 (+2%)</td>
</tr>
<tr>
<td>Land and Ocean</td>
<td>2.62</td>
<td>2.68 (+2%)</td>
</tr>
<tr>
<td>Land</td>
<td>2.39</td>
<td>2.53 (+6%)</td>
</tr>
<tr>
<td>Ocean</td>
<td>2.78</td>
<td>2.78 (+0%)</td>
</tr>
</tbody>
</table>
Table 2 Global- and tropical-average linear changes in mm/d/decade a) for the entire study period (1979-2007), and b) for the SSM/I era (1988-2007). “Ocean” and “land” regions are defined by 100% and <100% coverage by water.

### a) 1979-2007

<table>
<thead>
<tr>
<th>Version</th>
<th>Global 90°N-90°S</th>
<th>Tropical 25°N-25°S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2.1</td>
</tr>
<tr>
<td>Land and Ocean</td>
<td>+0.0115</td>
<td>+0.0069</td>
</tr>
<tr>
<td>Land</td>
<td>+0.0018</td>
<td>−0.0118</td>
</tr>
<tr>
<td>Ocean</td>
<td>+0.0184</td>
<td>+0.0199</td>
</tr>
</tbody>
</table>

### b) 1988-2007

<table>
<thead>
<tr>
<th>Version</th>
<th>Global 90°N-90°S</th>
<th>Tropical 25°N-25°S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>2.1</td>
</tr>
<tr>
<td>Land and Ocean</td>
<td>+0.0343</td>
<td>+0.0169</td>
</tr>
<tr>
<td>Land</td>
<td>+0.0630</td>
<td>+0.0252</td>
</tr>
<tr>
<td>Ocean</td>
<td>+0.0144</td>
<td>+0.0111</td>
</tr>
</tbody>
</table>
Fig. 1  a) Thirty-year climatology for GPCP Version 2.1 in mm/d, and b) (Version 2.1 – Version 2) difference averaged over 1979-2007 in mm/d.
Fig. 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.
Fig. 3  Map of linear change for Version 2.1 for the SSM/I era (1988-2007) in mm/d/decade.
Electronic Data Supplement to:

Improving the Global Precipitation Record: GPCP Version 2.1

George J. Huffman\textsuperscript{1,2}, Robert F. Adler\textsuperscript{1,3}, David T. Bolvin\textsuperscript{1,2}, Guojun Gu\textsuperscript{1,3}

1: NASA Goddard Space Flight Center, Laboratory for Atmospheres; GPCP Global Merge Development Centre
2: Science Systems and Applications, Inc.
3: University of Maryland College Park, ESSIC

Submitted to \textit{Geophysical Research Letters}
9 July 2009

Corresponding author:
George J. Huffman
NASA/GSFC Code 613.1
Greenbelt, MD 20771
Phone: 301-614-6308
Fax: 301-614-5492
Email: george.j.huffman@nasa.gov
Fig. EDS-1 Maps of linear change for Versions 2.1 and 2 for the entire study period (1979-2007; a and b, respectively), and for the SSM/I era (1988-2007, c and d, respectively), all in mm/d/decade.
Improving the Global Precipitation Record: GPCP Version 2.1
George J. Huffman¹,², Robert F. Adler¹,³, David T. Bolvin¹,², Guojun Gu¹,³

1: NASA Goddard Space Flight Center, Laboratory for Atmospheres;
GPCP Global Merge Development Centre
2: Science Systems and Applications, Inc.
3: University of Maryland College Park, ESSIC

Submitted to Geophysical Research Letters
9 July 2009

Popular Summary

How much precipitation fell last month around the entire globe? What’s the long-term average? It might surprise most Americans that these are hard questions because they are accustomed to flicking on the television to watch weather reports that detail daily accumulations from numerous weather stations and feature composite radar movies of the current precipitation patterns. Globally, the situation is much different, with a nearly total lack of surface data over the oceans, which cover about 2/3 of the globe, and large gaps in many land areas. It has been clear from the dawn of the Space Age that satellites provide an excellent platform to address this problem, but it was not until the late 1970’s that the first generation of sensors was continuously available in space to start the precipitation record. It took another 15 years and a second generation of satellite-based sensors, starting in mid-1987, for scientists to develop the first long-term global precipitation data set that took advantage of these then-new satellite data, while still incorporating raingauge data where available. The Global Precipitation Climatology Project (GPCP) was created by the World Climate Research Program (part of the World Meteorological Organization) to carry out the necessary innovative international cooperation among the operational weather satellite agencies in the U.S., Europe, and Japan, as well as the international precipitation science community.

Now, almost two decades later, the GPCP continues to extend and improve the global record of precipitation. Recently, the Global Precipitation Climatology Centre, a companion organization based in the Deutscher Wetterdienst (German Weather Service), released a new long-term archive of raingauge data. The improvements include a much higher number of stations and a better analysis scheme, and resolve some deficiencies that had been affecting the GPCP analysis. The GPCP has now adopted this new gauge analysis and recomputed the entire record of global precipitation, from 1979 to the present. During the reprocessing, special attention was paid to ensuring consistency among the various sensor types so that the entire record, now spanning 30 years, can be used for climate studies.

So, what is the long-term global average? This new “Version 2.1” dataset averages 2.68 mm per day (a tad over an inch a day), which is about 2% higher than for the old “Version 2”. The change essentially all comes from changes over land due to the new gauge analysis. The important point, of course, is that the estimated values over particular locations at particular times can vary by a great deal more. Notably higher values are now analyzed in tropical
mountain regions (northwestern South America, Papua New Guinea, the Himalayas, and along the east coast of the Bay of Bengal), along the mountainous Pacific coasts of northwestern North America and southern Chile, in New Zealand, and in central Africa. These are all high-precipitation areas that lacked sufficient gauge coverage in the previous analysis to capture the highest values.

Note that the various data sets produced in the highly successful Tropical Rainfall Measuring Mission (TRMM, a joint NASA – Japanese Aerospace Exploration Agency satellite mission) are created at much finer time and space scales than the GPCP data, but for a shorter period and only the tropical and subtropical regions. Both TRMM and GPCP have benefited greatly from intercomparing datasets and exchanging analysis concepts since the start of TRMM, and this is expected to continue with the development of the Global Precipitation Measurement (GPM) satellite mission. One important new area of comparison will arise from GPM’s coverage over polar latitudes.