
Xungang Yin
Cooperative Institute for Climate Studies, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

Arnold Gruber
NOAA/NESDIS/Office of Research and Applications, Camp Springs, and Cooperative Institute for Climate Studies, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

Phil Arkin
Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

Abstract

The two monthly precipitation products of the Global Precipitation Climatology Project (GPCP) and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) are compared on a 23-yr period, January 1979–December 2001. For the long-term mean, major precipitation patterns are clearly demonstrated by both products, but there are differences in the pattern magnitudes. In the tropical ocean the CMAP is higher than the GPCP, but this is reversed in the high-latitude ocean. The GPCP–CMAP spatial correlation is generally higher over land than over the ocean. The correlation between the global mean oceanic GPCP and CMAP is significantly low. It is very likely because the input data of the two products have much less in common over the ocean; in particular, the use of atoll data by the CMAP is disputable. The decreasing trend in the CMAP oceanic precipitation is found to be an artifact of input data change and atoll sampling error. In general, overocean precipitation represented by the GPCP is more reasonable; over land the two products are close, but different merging algorithms between the GPCP and the CMAP can sometimes produce substantial discrepancy in sensitive areas such as equatorial West Africa. EOF analysis shows that the GPCP and the CMAP are similar in 6 out of the first 10 modes, and the first 2 leading modes (ENSO patterns) of the GPCP are nearly identical to their counterparts of the CMAP. Input data changes [e.g., January 1986 for Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI), July 1987 for Special Sensor Microwave Imager (SSM/I), May 1994 for Microwave Sounding Unit (MSU), and January 1996 for atolls] have implications in the behavior of the two datasets. Several abrupt changes identified in the statistics of the two datasets including the changes in overocean precipitation, spatial correlation time series, and some of the EOF principal components, can be related to one or more input data changes.

1. Introduction

Although more than 200,000 routinely operating precipitation gauges have been established worldwide (Hulme 1995; New et al. 2001), their spatial distribution is inhomogeneous: all the gauges are on land (only very few on atolls and small islands, which may be representative of the ocean), with more in developed areas and many fewer in deserts and other inhospitable places. With 71% of the earth’s surface covered by ocean, gauges represent at most 25%–30% of the whole globe. Most of the gauge measurements commenced in the twentieth century, and at times some of them were abandoned while others have been added. Except in regions that have maintained good socioeconomic conditions, incomplete and short records are common for gauge stations. How to obtain a reliable global precipitation dataset is as difficult a question as it is important. Using satellite observations as proxy data has been one of the most effective solutions for more than a decade. Several merged gauge–satellite precipitation products have already been developed with either global coverage, quasi-global coverage, or regional coverage. Nevertheless, it is still not feasible to conclude which product is the best since there is no such dataset with full global coverage that can be referred to as a standard for verifi-
cation. However, there is value in comparing different global datasets since we may be able to get a better qualitative sense of the distribution of global precipitation by trying to understand differences and similarities.

Since as early as the 1970s with the commencement of using satellite observations to estimate precipitation, there has been an increasing concern on the quality and the credibility of the satellite-based precipitation products. There were two organized intercomparison programs named Precipitation Intercomparison Projects (PIP) (Adler et al. 2001) and Algorithm Intercomparison Projects (AIP) (Ebert and Manton 1998), which compared satellite retrievals to some reference datasets. In addition some researchers have also pursued intercomparison and validation studies (e.g., Spencer 1993; Xie and Arkin 1995; Rudolf et al. 1996; Kondragunta and Gruber 1997; Janowiak et al. 1998; McCollum et al. 2000; Krajewski et al. 2000). These studies have helped the accomplishment of the merged gauge–satellite precipitation products.

Among the gauge–satellite precipitation products, the Global Precipitation Climatology Project (GPCP) (Huffman et al. 1997; Adler et al. 2003; data are available online at http://www.ncdc.noaa.gov/oa/wmo/wdcmh-net.ncdc.html) and the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) (Xie and Arkin 1997; data are available online at ftp://ftp.cpc.ncep.noaa.gov/pub/precip/cmap) are likely the most widely recognized ones that have full global coverage and the longest time span. Gruber et al. (2000) compared the original versions of the two monthly datasets for the period July 1987–December 1998. That study showed that both the GPCP and the CMAP clearly demonstrated the major precipitation patterns. For the average seasonal cycle, the GPCP and the CMAP exhibited very good agreements for area means of both the lands and the oceans in the Southern Hemisphere (SH) and the lands in the Northern Hemisphere (NH). For the oceans in the NH, the seasonal variations of the two datasets still synchronized with each other, but the CMAP was considerably higher in magnitude than the GPCP between May and October.

In 2002 the version 2 GPCP was released (Adler et al. 2003). This latest version supersedes the one used by Gruber et al. (2000). The temporal coverage of its monthly data was extended to January 1979. The CMAP also extended its monthly dataset in the V0207 version (released in July 2002). Now both products cover a period from January 1979 to present with some delay. Compared to its earlier generation version 1c, the GPCP version 2 has two distinct improvements. In addition to extending the time series back to 1979, it developed a method to fill gap in regions where geostationary infrared and microwave retrievals were not reliable, generally poleward of 60°N and 60°S.

In this paper, we compare the latest versions of the monthly GPCP and the CMAP on both temporal and spatial domains. The purpose of this work is to find the similarities as well as the differences that exist between the two datasets and to identify, when possible, the reasons for the differences, in order to provide confidence information to users and quality information to developers. The data and methods are described in section 2. In section 3 we give a general statistical analysis on the two datasets, focusing on large-scale spatiotemporal patterns. Then we will utilize EOF analysis to study the spatial patterns and their temporal evolutions in the two datasets. A general summary and conclusions are given in section 4.

2. Data and methods

a. Data

Both the GPCP and the CMAP data used in this study are monthly mean precipitation rates with units of millimeters per day. They are gridded on 2.5° latitude × 2.5° longitude boxes with the first box centered on the point (88.75°S, 1.25°E). We use the observation-only CMAP that has no National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis involved. For the integrity of annual variation statistics, a full 23-yr (276 months) period, January 1979–December 2001, has been selected as the object of this study. The input data details have been described by Huffman et al. (1997) and Adler et al. (2003) for the GPCP and Xie and Arkin (1996, 1997) for the CMAP; thus, only a summarization of the input data is given in Table 1. It should be noted that all the input data (satellite and gauge) to the final GPCP and CMAP precipitation are in the form of gridded analyses. The reader is referred to Table 1 for definitions of the abbreviations of the various input datasets.

The GPCP merged gauge–satellite precipitation is created in two steps (Huffman et al. 1997; Adler et al. 2003). First, in the Special Sensor Microwave Imager (SSM/I) period (July 1987–present, excluding December 1987) the SSM/I-adjusted Geostationary Operational Environmental Satellite (GOES) precipitation index (GPI)/(AGPI) is used between 40°S and 40°N, and merged SSM/I/Television Infrared Observation Sounder (TIROS) Operational Vertical Sounder (TOVS) is used elsewhere; in the pre-SSM/I period calibrated OPI (by GPCP in 1988–97) is used globally for January 1979–December 1985, and for other time AGPI is used between 40°S and 40°N, and the calibrated outgoing longwave radiation (OLR) precipitation index (OPI) is used elsewhere. Then the multisatellite estimate is adjusted toward the large-scale gauge average for each grid box over land. Second, the gauge-adjusted multisatellite estimate and the gauge analysis are combined in a weighted average, where the weights are the inverse error variance of the respective estimates.

The merging method of the CMAP is different from that of the GPCP (Xie and Arkin 1997). First, a base
TABLE 1. Input data information for the GPCP and the CMAP.

<table>
<thead>
<tr>
<th>Input data coverage</th>
<th>Land Coverage</th>
<th>Ocean Coverage</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHCN+CAMS</td>
<td>√</td>
<td></td>
<td>Xie et al. (1996)</td>
</tr>
<tr>
<td>Jan 1979–Dec 1985</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPCC</td>
<td>√</td>
<td></td>
<td>Rudolf (1993), Rudolf et al. (1994)</td>
</tr>
<tr>
<td>Jan 1986–present</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPI</td>
<td>√</td>
<td></td>
<td>Janowiak and Arkin (1991)</td>
</tr>
<tr>
<td>Jan 1986–present</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSM/I emission</td>
<td>√</td>
<td>√</td>
<td>Wilheit et al. (1991)</td>
</tr>
<tr>
<td>Jul 1987–present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSM/I scattering</td>
<td>√</td>
<td>√</td>
<td>Ferraro and Marks (1995)</td>
</tr>
<tr>
<td>Jul 1987–present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOVS</td>
<td>√</td>
<td></td>
<td>Susskind et al. (1997)</td>
</tr>
<tr>
<td>Jul 1987–present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPI</td>
<td>√</td>
<td></td>
<td>Xie and Arkin (1998)</td>
</tr>
<tr>
<td>Jan 1979–Jun 1987</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jan 1979–present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atolls</td>
<td></td>
<td></td>
<td>Morrissey and Greene (1991)</td>
</tr>
<tr>
<td>Jan 1979–present</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Full names of the input datasets are GHCN = Global Historical Climatology Network and CAMS = Climate Anomaly Monitoring System.

The period (July 1987–June 1995) of merged data is created by combining IR, SSM/I, and Microwave Sounding Unit (MSU) satellite data using a maximum likelihood approach in which the weighting coefficients are inversely proportional to the squares of the individual random errors. The errors are determined by comparing with the Global Precipitation Climatology Centre (GPCC) gauge measurements over land and with atoll gauge measurements for an estimate of the errors over the ocean. Second, the possible biases are removed by combining the results of the first step with gauge analyses via a blending technique. Over land it is assumed that the combined satellite estimates can represent the structure of the precipitation distribution and that there is no bias in gauge estimates. Then the merged products can be obtained by solving the structure of precipitation, which is in form of a Poisson equation whose boundary conditions are determined by gauge data. Over the ocean, the bias remaining after the first step is removed by comparison with the atoll gauge data over the Tropics and by subjective assumption regarding the bias structure over the extratropics.

b. Methods

Most of the climate variables can be decomposed into two parts: a long-term average showing the climatological features and an anomaly representing the variation from the average, usually after the annual cycle has been removed. In this study, normalized precipitation anomalies are obtained by subtracting the corresponding calendar-month precipitation climatology from each monthly precipitation and dividing by the standard deviation of that calendar-month precipitation. Since normalization reduces the spatial magnitude difference of the precipitation patterns, in this study the normalized data are utilized when the GPCP and the CMAP are compared in the temporal domain in order to treat all months equally, while the mean and unnormalized anomaly data are employed for the spatial pattern comparisons. Throughout the paper the term “areal mean” is always used with area weighting to account for the grid-area change with latitude. The anomaly and normalized anomaly are all done based on each grid point. Therefore the areal mean normalized precipitation anomaly is calculated by taking the area-weighted average of all the individual normalized precipitation anomalies.

1) TEMPORAL AND SPATIAL CORRELATION

Temporal correlation is used to study at each grid the similarity between the GPCP and the CMAP during their time history. Temporal correlation calculation is based on normalized anomaly precipitation data. A temporal correlation map is constructed from all the GPCP–CMAP grid correlations. Based on this map, the GPCP–CMAP correlation distribution is analyzed.
Spatial correlation is used to find out at each time frame the similarity between the GPCP and the CMAP spatial patterns. Calculating spatial correlations for all the time frames, we obtain a time series of GPCP–CMAP spatial correlation for the globe or a selected region. Spatial correlation calculation is based on area-weighted anomaly data.

2) EOF ANALYSIS

The essence of EOF is to decompose a dataset into spatial patterns (EOFs) and time series called principal component (PC) in terms of orthogonal function. The GPCP and the CMAP are able to grasp some of the same precipitation patterns, though they may differ slightly in recording the temporal variations of those patterns. In this case, for convenience, the common EOF (cEOF) technique (Barnett 1999) is employed to study the common spatial patterns in the GPCP and the CMAP. To use this method, we first concatenate the two datasets along the time dimension into one dataset, then apply the normal EOF analysis on this new dataset. Finally, each principal component is divided into two partial PCs, one for the GPCP and the other for the CMAP. However, the spatial pattern is common for the two partial PCs. EOF analysis is based on normalized anomaly data. To have all the precipitation variances in EOF analysis area weighted, every input grid precipitation value is multiplied by the square root of cosine latitude.

3. Results

In this study, land and ocean areas are separately analyzed in the areal mean time series comparison and the spatial pattern comparison. The reason for this separation is that the input data for land and ocean precipitation are quite different in both products, as shown in Table 1. Also, oceanic and land precipitation have different characteristics, and thus it is informative to compare them separately.

a. Long-term mean precipitation field

The GPCP and the CMAP respective long-term-averaged maps for the period January 1979–December 2001 are shown in Fig. 1. In general, the two maps resemble each other closely. Calculated from the two maps, the global mean precipitation rate is 2.64 and 2.65 mm day$^{-1}$ for the GPCP and the CMAP, respectively, equivalent to 964 and 967 mm yr$^{-1}$. The correlation coefficient between the two maps reaches 0.93, indicating the GPCP and the CMAP have a very good agreement in describing the large-scale spatial variations. The patterns that can be clearly discerned in both maps include 1) the intertropical convergence zone (ITCZ), starting from the Indian Ocean almost continuously extending to the West African coast and located mostly north of the equator in the Pacific and the Atlantic Oceans; 2) the South Pacific convergence zone (SPCZ), starting from the Indonesian islands extending all the way to the southeast Pacific Ocean in a NW–SE direction; 3) the South Atlantic convergence zone (SACZ), also in a NW–SE direction starting from the Amazon basin and ending at the mid-Atlantic Ocean around 45°S; 4) midlatitude storm tracks along the east coasts of Asia and North America; 5) dry areas around the horse latitudes including the ocean areas to the west of North America, South America, North Africa, southern Africa, and Australia, as well as the land areas in the Sahara, the Arabian, the Gobi, and the Australian deserts, as well as the deserts in the west of North America.

On the other hand, the difference between the two maps is noticeable. An obvious difference one can see is that in the Tropics and subtropics the CMAP has higher magnitudes. For example, the ITCZ and SPCZ precipitation in the CMAP is roughly one contour interval (2 mm day$^{-1}$) higher than that in the GPCP. Over the coastal oceans to the west of South America and southern Africa, the area of the minimum level precipitation (0.2 mm day$^{-1}$) is smaller in the GPCP than in the CMAP. This situation is reversed in the higher latitudes. In the CMAP, to the west of the Asian and North American continents, the precipitation range associated with storm tracks ends at about 45°N, while in the GPCP it reaches much farther beyond 60°N and also has a wider extent.

To quantitatively demonstrate the difference between
the two climatology means, relative difference (Gruber et al. 2000) is calculated as

\[
\text{Relative Difference} = \frac{\text{GPCP} - \text{CMAP}}{\text{GPCP} + \text{CMAP}} \times 100\%.
\]

(1)

Figure 2 shows the result of the relative difference calculated from the two maps of Fig. 1. Since in a low-precipitation area even very high relative difference does not have practical meaning, the average of the GPCP and the CMAP is also plotted on the figure so that we only focus on areas with at least a moderate precipitation amount. The result is very similar to that of Gruber et al. (2000) on the comparison of the earlier versions of the GPCP and the CMAP. General conclusions that can be drawn from Fig. 2 are as follows:

1) Over land except Antarctica and Greenland, where precipitation is extremely low as well as unreliable, the GPCP is higher than the CMAP, but less than 30% in most areas. The use of wind-corrected gauge data by the GPCP is the main reason for this difference since the corrections always increase the amount.

2) Over the tropical and subtropical ocean, the GPCP is often lower than the CMAP. In these areas, the relative difference is mostly less than 30%, except over the tropical Indian Ocean and the west Pacific Ocean, where precipitation is high, the difference reaches 30%–60%. This difference is related to the use by CMAP of atoll rain gauges in the Pacific tropical ocean to adjust the satellite estimates, but GPCP does not use these gauges. Both the GPCP and the CMAP satellite-only estimates are lower than the atoll gauge precipitation (Xie and Arkin 1995; Adler et al. 2003).

3) Over the high-latitude oceans, the GPCP is higher than the CMAP, with the relative difference ranging from negative or slightly positive to more than 60%. It is believed to be at least attributable to the use of TOVS in the GPCP (Adler et al. 2003).

4) The west coasts of South America and Africa have large negative values, indicating lower GPCP estimates. As discussed by Gruber et al. (2000) this may be related to how coastal grids are treated. For a coastal grid, the GPCP uses gauges based on the ratio of land and ocean in the grid, but CMAP compares the ratio with a threshold to decide either to treat the grid as land or ocean. Over the ocean the CMAP uses SSM/I scattering and emission (SSM/I period), MSU (January 1979–May 1994), GPI (since January 1986), and OPI estimates and atolls, while the GPCP uses SSM/I emission (SSM/I period), GPI (since January 1986), TOVS (SSM/I period), and OPI (pre-SSM/I period). In addition, GPCP modifies its SSM/I emission algorithm over the coast region to avoid land contamination.

5) Subtropical low-precipitation areas including the Sahara Desert and the oceanic areas to the west of South America and southern Africa are distinguished from the surrounding areas in that the GPCP is much higher than the CMAP. It is worth noticing that the ranges of the three areas with high relative difference are coincident with the 0.5 mm day\(^{-1}\) average precipitation contours, so not too much significance should be attributed to those differences.

b. Zonal mean

The January 1979–December 2001 average global zonal mean precipitation profile is presented in Fig. 3 for the GPCP and the CMAP, respectively. The common features in the two products are 1) one primary maximum in the Tropics with its mean position at a few degrees north of the equator; 2) secondary maxima in midlatitudes between 30° and 60°N and 30° and 60°S with mean positions around 40°, except that in the SH the GPCP exhibits another maximum at about 60°S; 3) primary minima in the polar regions; 4) minima in the
Tropics around 20°N and 20°S. One of the differences between the two profiles is that the latitudinal gradient is larger in the CMAP than in the GPCP, a consequence of the larger CMAP tropical values but weaker mid- and high-latitude values. The high tropical values are a direct consequence of the use of the atoll rain gauges by the CMAP to adjust the satellite estimates. The large peak in the GPCP at 57°S is open to question and requires further analysis, as indicated by Adler et al. (2003), since this is the transition zone where TOVS data are input.

The time–latitude profiles for the two products are shown in Fig. 4. We can see that the annual cycles in the two hemispheres are out of phase. From year to year the zonal mean precipitation features are basically repeated with slight modifications from large-scale climate phenomena, particularly the ENSO. In strong El Niño years—for example, 1982/83, 1991/92, and 1997/98—precipitation just north of the equator decreased and the high-precipitation center moved to the SH. The variations associated with El Niño are best seen in the time–longitude (120°E–120°W) section of Fig. 5, which shows the mean precipitation between 10°S and 10°N for each longitude. This figure clearly exhibits that in the El Niño years the rainfall shift starts from the west Pacific Ocean and moves farther to the east Pacific Ocean.

c. Monthly areal mean precipitation time series

The global areal mean precipitation is shown in Fig. 6 separately for land and ocean. For the period January 1979–December 2001 the average overland and over-ocean mean precipitation is, respectively, 2.07 and 2.90
TABLE 2. Jan 1979–Dec 2001 average areal mean precipitation (unit: mm/day$^{-1}$) for the GPCP and the CMAP. The ocean and land are calculated separately.

<table>
<thead>
<tr>
<th></th>
<th>GPCP</th>
<th>CMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global</td>
<td>SH</td>
</tr>
<tr>
<td>Ocean</td>
<td>2.90</td>
<td>2.66</td>
</tr>
<tr>
<td>Land</td>
<td>2.07</td>
<td>2.46</td>
</tr>
</tbody>
</table>

mm day$^{-1}$ for the GPCP and 1.93 and 2.98 mm day$^{-1}$ for the CMAP (Table 2). So for the long-term average of the global mean precipitation, compared to the CMAP, the GPCP is higher over land by 7% but lower over ocean by about 3%. For the four time series in the figure the standard deviation is 0.10 mm day$^{-1}$ for the GPCP over ocean but 0.15 for the rest of the three cases. For the ocean the standard deviation is higher in the CMAP than in the GPCP. This is partly associated with a seemingly decreasing trend as can be seen from the figure. However, this trend, which is only found in the CMAP overocean mean precipitation, is merely an artifact of input data changes as explained below.

In the time series of the CMAP global overocean mean precipitation there are two abrupt change points, January 1986 and January 1996, which are absent in the GPCP time series. The first one at January 1986 corresponds to the starting time of GPI. Before 1986 the temporal variability of the CMAP is much higher than that of the GPCP, while after 1986 the difference has reduced. This change can be associated with the different usages of GPI in the two products. For the GPCP, GPI is adjusted by OPI in the pre-SSM/I period.
TABLE 3. Statistics for the correlation between areal mean normalized GPCP and CMAP. Here, $r$ is calculated correlation; $N_{\text{eff}}$ is effective degree of freedom; and $r_c$ is critical correlation value at the 99% significance level. If $r > r_c$, then $r$ is significant. For land, case A is with Antarctica; case B is without Antarctica.

<table>
<thead>
<tr>
<th></th>
<th>Globe</th>
<th>Land (NH)</th>
<th>Ocean (NH)</th>
<th>Land (SH)</th>
<th>Ocean (SH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>0.51</td>
<td>0.95</td>
<td>0.64</td>
<td>0.96</td>
<td>0.61</td>
</tr>
<tr>
<td>$N_{\text{eff}}$</td>
<td>131</td>
<td>68</td>
<td>105</td>
<td>118</td>
<td>160</td>
</tr>
<tr>
<td>$r_c$</td>
<td>0.22</td>
<td>0.31</td>
<td>0.26</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

and by SSM/I in the SSM/I period. For the CMAP, there is no adjustment to GPI. The adjustment in the GPCP was done to ensure consistency in the input data, and may have resulted in the reduced variability. Another reason is that prior to 1986 over the ocean the GPCP only used OPI, while the CMAP also used MSU and atolls. The GPI and OPI have smoother variation than does MSU (Xie and Arkin 1997). The second change in the CMAP is a sudden decrease in the magnitude after January 1996. This change can be related to the artificial decrease in the CMAP precipitation as noted in the CMAP V0207 announcement. It is apparently caused by the CMAP adjusting satellite estimates to the atoll data. According to P. Xie (2004, personal communication), after 1996 the number of atoll gauges significantly decreased, causing an artificial drop in precipitation amount compared to previous years, which looks like a trend. In effect this is a sampling problem and represents a time-dependent bias in the CMAP.

The calculated GPCP–CMAP correlation coefficient is 0.95 for global overland mean precipitation time series but only 0.09 for global oceanic mean precipitation time series. Over the ocean the CMAP uses several estimates, while GPCP uses only one estimate globally before 1986 and between 40°S and 40°N since 1986. Therefore the GPCP and the CMAP have much less in common in the input data over the ocean. The overocean correlation is quite likely strongly influenced by various sources of noise in each product. In particular, the atoll information is used throughout the CMAP record, so any variability in the satellite-derived estimate to atoll rainfall relationship will be communicated throughout the global tropical ocean and to a lesser extent into middle latitudes. The sampling problem due to the small number of atoll gauges can be a major error source. On the other hand, from a statistical point of view based on Fig. 6 there are three sources for the low overocean correlation. The first is the out-of-phase character of precipitation between the NH and SH. The bottom two panels of Fig. 6 show the overocean GPCP and CMAP areal mean precipitation time series for the SH and the NH. As the seasons in the two hemispheres are reversed, the timing of the minimum/maximum precipitation is also reversed. As seen from Table 2 the mean overocean precipitation in the NH is higher than in the SH, but the ocean area in the SH is about 30% more than that in the NH. Thus the total precipitation contribution is comparable from two hemispheres, and when they are combined, the annual cycle in the global overocean precipitation almost disappears; thus, the variability largely decreases. The consequence is that the global overocean mean precipitation has very low variability (about 20% of that in each hemisphere), which can be easily contaminated by noise. In comparison, although the overland mean precipitation in the NH is lower than in the SH, the NH land is nearly twice the SH land, and thus the total precipitation contribution from the NH is about 50% more than that from the SH. So the global overland precipitation carries more characteristics of the NH land precipitation, and as a result the annual cycle signal is still strong. The second source comes from the CMAP artificial trend. As shown in Fig. 6, the three stages are more pronounced in the NH, where the CMAP is higher than the GPCP in their high-precipitation phases, while in the SH, the CMAP is generally higher than the GPCP in their low-precipitation phases. During 1979–86 the CMAP–GPCP difference is almost twice that of the GPCP from its long-term mean. When the annual cycle is smoothed out, the noise related to the GPCP–CMAP difference becomes dominant. The third is the spatial distribution of the GPCP–CMAP difference. As shown in Fig. 2, over land the GPCP is almost systematically higher than the CMAP, but over the ocean the distribution of their difference is complicated. This may create another randomlike error when the global overocean mean is calculated. As the areal mean time series comparison gives more weight to regions with high precipitation and involves annual cycles, an effective approach to score the GPCP–CMAP closeness is to normalize the precipitation anomalies in order to remove the annual cycle and to treat all individual grids equally. This is demonstrated by calculating the correlation of areal mean normalized anomaly time series in this section and the temporal correlation of normalized anomalies for each grid in the next section.

The areal mean normalized precipitation anomaly can be considered as an index measuring the relative precipitation variability averaged from all grids regardless of their magnitudes. Simply speaking, the focus of what follows in this section is the analysis of the relative variability rather than the absolute amount. The GPCP–CMAP normalized anomaly correlation is calculated for the globe (land plus ocean), global land, global ocean, and each hemispheric land and ocean separately (Table 3). For the global land and SH land, cases without Antarctica are also studied. In Table 3, the highest corre-
lation is 0.96 for the SH land without Antarctica, and the lowest correlation is 0.36 for the global ocean. To determine the correlation significance, the effective number of degrees of freedom \( N_{\text{eff}} \) is evaluated by taking into account the autocorrelations in the two participating time series (Davis 1976; Chen 1982):

\[
N_{\text{eff}} = \frac{N \Delta t}{\tau},
\]

where \( N \) is the total number of samples, \( \Delta t \) is the sampling interval, and the effective sampling interval \( \tau \) is calculated as

\[
\tau = \left[ 1 + 2 \sum_{i=1}^{N-1} \rho_i(i\Delta t)\rho_j(i\Delta t) \right] \Delta t,
\]

where \( \rho_i(i\Delta t) \) and \( \rho_j(i\Delta t) \) are the respective autocorrelation coefficients of lag \( i\Delta t \) for the two time series.

For this study, \( T = 276 \) and \( \Delta t = 1 \).

The calculated effective number of degrees of freedom and the critical correlation coefficient at the 99% significance level of Student’s \( t \) test distribution are given in Table 3. The result shows that for all the cases the areal mean normalized GPCP is significantly correlated with the normalized CMAP. However, the normalized anomaly correlation based on the areal mean is much lower than that based on each individual grid in most of the areas, particularly the ocean. The three error sources for the low correlation of global ocean mean time series are also applied to this overocean correlation, which is significant but can only explain 13% of the variance. From the result it is apparent that in analyzing the normalized precipitation anomaly Antarctica plays an important role. If the continent is included in the calculation (case A), the correlation for the global land and SH land is, respectively, 0.81 and 0.64; if it is excluded (case B), the two correlations increase to 0.94 and 0.96, respectively. Since the precipitation condition in Antarctica is poorly understood from either gauge measurements or satellite observations, caution should be taken when using high-latitude land and ocean data.

d. Temporal correlation of individual grid

The correlation between the normalized anomalies of the GPCP and the CMAP is calculated for each grid based on the period January 1979–December 2001 (Fig. 7). It is seen from the figure that in general the GPCP–CMAP correlations are higher over land than over the ocean, as might be expected because of more common input data described earlier. Areas over land where the correlations are low tend to occur when the precipitation is low. There the noise probably dominates so the correlations will be low. This is also true over the oceanic areas with extremely low precipitation, for example, the west coast of South America and southern Africa. Over most of the tropical and midlatitude ocean the correlations are moderately high (>0.7), and the highest ones are around the mean position of the ITCZ in the Pacific Ocean. The ITCZ is dominated by convective rainfall type, which is well retrieved from satellite observations, both infrared and microwave. Low correlations poleward of 60°, particularly in the SH, are the results of uncertainties in both datasets as described earlier.

e. Spatial correlation

The GPCP–CMAP spatial pattern correlations are calculated for each individual months using anomaly data, separately for the land and the ocean, exhibited in Fig. 8. The two panels show that from pre-SSM/I to the SSM/I period the overall correlation is getting better for both land and ocean. This is particularly true for the ocean, for which the number of months that have a correlation coefficient less than 0.7 in the two periods are 10 and 4, respectively, that is, 10% and 2% of the total months in the corresponding periods. This is again a result of input data change in both products. In the figure, it is also noticeable that over ocean the frequency of low correlation is higher during June–October, which is the southern winter. This phenomenon is mainly caused by the SH overocean precipitation. In each hemisphere from summer to winter the precipitation is getting lower, and thus the noise-to-signal ratio is getting higher. As a result the precipitation variability becomes more difficult to estimate in the winter than in the summer by remote sensing technique. So for each hemisphere the spatial correlations are lower in the hemispheric winter as shown in Fig. 9, although it is less evident for the NH, whose ocean areas are mostly in the low latitudes. Since the ocean area in the SH is much larger than in the NH, the global overocean precipitation carries more characteristics of the SH. Accordingly, the global overocean GPCP–CMAP correlation is lower in the southern winter, when the uncertainty in the SH oceanic precipitation becomes higher.

Since over land the GPCP and the CMAP have more
common input data, particularly the gauge data for large-scale adjustment, the overland spatial correlation is not expected to have obvious seasonal variation. However, during the April–November season the correlations are still slightly lower, especially in the pre-SSM/I period. This can be attributed to the data quality and the precipitation seasonal distribution in Antarctica. We have already discussed that satellite techniques have difficulty in estimating precipitation accurately in Antarctica. This continent covers more than 10% of the global land but has only about 20–30 gauge stations that are mainly on its coastal area. The GPCC wind correction scheme may increase the precipitation in Antarctica by 100%–200%. The Antarctica precipitation is poorly represented by both the GPCP and the CMAP but it has considerable influence on the global overland GPCP–CMAP correlation. In the meantime, for Antarctica both the GPCP and the CMAP show maximum precipitation in the SH wintertime, as evidenced in Fig. 10. This is opposite to any other large-scale regions on the earth. The incompatibility between the two areal means is much higher during months April–November as the CMAP has higher variability. For the SH alone, the overland correlation is much lower in the SH winter than in the SH summer. When Antarctica is excluded from the calculation, the seasonal difference in the GPCP–CMAP correlation map almost disappears. This is clearly demonstrated by two examples, July 1987 and April 1998, that are analyzed in this section.

Over land the lowest spatial correlation is 0.66, found in July 1997. In order to discover the source of this low correlation, a scatterplot of GPCP and CMAP anomalies with latitude correction is shown in the top panel of Fig. 10.
The slope \( k \) of the best-fit line in the plot is calculated using one of the following equations:

\[
k = \tan \theta \quad \text{or} \quad k = \frac{-1}{\tan \theta},
\]

whichever has the same sign as the correlation coefficient, and where

\[
\tan 2\theta = \frac{2(\bar{xy} - \bar{x} \bar{y})}{(\bar{x}^2 - \bar{x}^2) - (\bar{y}^2 - \bar{y}^2)},
\]

where \( x \) and \( y \) represent the GPCP and the CMAP precipitation, respectively, and the intersection on the \( y \)-axis is evaluated as

\[
y_0 = \bar{y} - k \bar{x}.
\]

Then for each GPCP precipitation \( x \), the expected CMAP precipitation on the best-fit line can be calculated by the following equation:

\[
\bar{y} = y_0 + kx.
\]

Define the standard deviation of \( y \) based on the best-fit line as

\[
\sigma = \sqrt{\frac{1}{N-2} \sum_{n=1}^{N} (y_n - \bar{y})^2},
\]

where \( N \) is the total number of grids. Select a number, say \( q \), as a criterion so that for a given grid if

\[
\frac{|y - \bar{y}|}{\sigma} \geq q,
\]

then for this grid the difference between the GPCP and the CMAP anomalies is significant. In practice, \( q \) is no less than 3.

For the case of July 1997, set \( q = 4 \) and only remove the outliers above the best-fit line. The removal result is given in the bottom panel of Fig. 11. A total of 15 points have been removed and the GPCP–CMAP correlation dramatically increases to 0.85. Also the slope, which is too large (=2) in the original scatterplot, now becomes reasonable (=1). Almost all of the removed grids are in equatorial West Africa. Examination of the GPCC dataset indicates that in this area most of the grids have no gauge station at all in July 1997, while in the Nicholson’s (1993) African rainfall archive there are 49 gauge stations available at this time. We then compare the GPCP and the CMAP precipitation map with Nicholson’s (1993) rainfall (top two panels of Fig. 12), and the result shows that in July 1997 the CMAP has given too high a rainfall amount in equatorial West Africa. This GPCP–CMAP discrepancy can be explained by the differences in merging procedures between the two products. As we know both the GPCP and CMAP use gauge analyses for large-scale adjustment. If there are too few or no gauges, there is no adjustment in the GPCP, but the CMAP still uses the gauge analyses regardless the gauge number. Figure 12 also shows the GPCC gauge number together with GPCC analyses precipitation and the GPCP multisatellite estimates for equatorial West Africa in July 1997. Obviously for this case, the CMAP largely followed the precipitation pattern of the GPCC analyses, while the GPCP is mainly decided by the multisatellite result, which is in much better agreement with the Nicholson data.

Similarly, for September 1985, the second lowest overland correlation month with \( r = 0.67 \), we set \( q = 4 \) and remove both positive and negative outliers. After the removal of 31 grids, which are mainly in Southeast Asia (high CMAP) and Central America (high GPCP), the correlation has increased to 0.84. June 1987 and April 1988 are the other 2 months that have low overland correlations, which are 0.70 and 0.69, respectively. The scatterplots in Fig. 13 show for both months there is a group of points congregating along the \( y \) axis, indicating those grids have low GPCP anomalies but high CMAP anomalies. Although according to Eq. (9) most of these points are not outliers, their collective contribution to the low correlation is significant. Analysis concludes that for both months those grids are from
almost the same area in Antarctica, where there are high uncertainties in precipitation records. When Antarctica is excluded, the GPCP–CMAP correlations for the two months rise to 0.84 and 0.81, respectively.

f. EOF analysis

The normalized GPCP and CMAP anomalies are used for EOF analysis. The study area is between 50°S and 50°N since the CMAP has a large number of missing data in the high latitudes. First we have applied the EOF to the normalized GPCP and CMAP anomalies separately in the period January 1979–December 2001. We now compare each pair of GPCP and CMAP EOF modes by calculating their correlation coefficient for the EOFs and PCs. The statistical results for the first 10 modes are given in Table 4.

Unlike those of temperature, the spatial pattern of precipitation is highly variable in space and time. Therefore each EOF mode of precipitation only explains a small percent of the total variability as shown in Table 4. Such modes sometimes may not have clear physical meanings since they can be easily contaminated by noise. Nevertheless, the EOF result shows that the GPCP
and the CMAP are in good agreement in describing the large-scale precipitation patterns since 6 out of first 10 of the EOF modes have higher positive GPCP–CMAP correlations in both spatial patterns and time series, particularly the first 2 leading ones. In what follows we use the common EOF method to demonstrate the ENSO precipitation revealed by the first two EOFs and to identify the sources for the difference in the third EOF.

Figure 14 shows the cEOF-1 and the two partial PCs, one for the GPCP and the other for the CMAP. This mode explains 4.3% of the total variance. It is an ENSO pattern that features two distinct regions with opposite variations. In the 23-yr period, moderate to strong El Niño events were observed in years 1982/83, 1986/87, 1991/92, 1994/95, and 1997/98, and La Niña events in years 1988/89, 1995/96, and 1998/99. It can be seen from the PC time series that almost all of the negative and positive peaks are associated with ENSO events. In all the five El Niño years the amplitudes of both PCs are significantly positive, indicative of decreased rainfall in the equatorial west Pacific but enhanced rainfall in the central to the eastern Pacific Ocean along the equator. During the four La Niña events, the negative peaks in the two PCs denote that the El Niño rainfall pattern is reversed with above-normal rainfall in the west equatorial Pacific Ocean but below normal rainfall in the central to east equatorial Pacific Ocean.

Figure 15 is the second cEOF result. The explained variance is 2.6%. The major rainfall pattern in cEOF-2 is very similar to that of cEOF-1, only it has moved to the east by about 30°. This is another ENSO pattern, but it only appears in strong El Niño years including 1982/83 and 1997/98, characterized by enhanced rain-
fall in the central to east Pacific Ocean over the equator and a belt region between Southeast Asia and Australia, and weakened rainfall in a belt region north of the equator roughly between 150°E–120°W. In the two strongest El Niño years, the PC-2 peaks lag the corresponding PC-1 peaks by about 4–6 months. The cEOF-1 is a universal pattern that exists in all ENSO events, representing their mature stages. For an El Niño, the PC-1 pattern moves to the east and starts to decay. For a weak event, the pattern disappears on its way to the east, but for a strong event the decaying process takes such a long time that the El Niño pattern can reach the east Pacific Ocean after several months.

Higher-order EOF modes are composed of high-frequency oscillations in both the spatial patterns and time series. Their physical meaning is much less clear than those of the first two leading modes. As an example, the third EOF modes of the GPCP and the CMAP are shown separately in Fig. 16 for comparison. First, we notice that in the 1980s there is a short decreasing and then an increasing trend in the GPCP time series. This trend at least reflects the rainfall variation in many regions in Africa as can be derived from both the GPCP and the CMAP. It has been confirmed by other studies (e.g., Nicholson 2001; Hulme 1992) that particularly the Sahelo–Sahara and Sahel have been experiencing a long-term drought starting since the late 1960s, though there was a short-term wet period in late 1970s to early 1980s. Without the consideration of this trend, the GPCP PC-3 variability is lower in the pre-SSM/I period than
in the SSM/I period. As stated in section 3c, in the pre-SSM/I period the GPCP used only OPI and GPI, which are known to be smoother. The decreasing trend in the CMAP PC-3 is nearly the same as the one found in the global overocean mean precipitation time series (section 3c). The first change point is also January 1986 when GPI becomes available, but the second one in the CMAP is mid-1994, the termination time of MSU, which is known to have larger variability (Xie and Arkin 1997). The after-1996 artificial decrease due to the atoll sampling problem in the CMAP may also contribute to the CMAP PC-3 trend. Similar changes that can be attributed to the input data changes can also be found in other EOF modes.

4. Summary and conclusions

The GPCP and the CMAP monthly precipitation datasets are compared in this study on a full 23-yr period from January 1979 to December 2001. The results indicate that the two products are in very good agreement in describing the large-scale precipitation spatial patterns and temporal variations. For the long-term-averaged precipitation field, both the GPCP and the CMAP have successfully represented those large-scale patterns including the ITCZ, the SPCZ, storm tracks, etc. Global mean normalized GPCP and CMAP are significantly correlated. In most of the months the precipitation spatial patterns represented by the GPCP and the CMAP are significantly close. With the large-scale gauge adjustment in both products, the GPCP–CMAP correlation is generally higher over land than over the ocean. EOF analysis further demonstrates the large-scale agreements between the two products, since the first two leading modes of the GPCP and the CMAP are almost identical in exhibiting the ENSO events, which are among the strongest precipitation signals in the Tropics. In spite of the encouraging large-scale agreement between the GPCP and the CMAP, there are still some significant differences between them that appear to be related to the input data and/or the method of merging the satellite and gauge data. The most obvious is the differences in the tropical oceans and the high-latitude land areas—a consequence of the input data. Also, the CMAP is very clearly sensitive to the atoll gauge data, which by virtue of how they incorporate it, has a significant impact on their tropical oceanic precipitation estimates indicating an apparent artificial trend. GPCP is sensitive to the TOVS data, which may be the cause of a questionable maximum at 60°S. On a regional scale (equatorial West Africa) there was some evidence that the CMAP analysis procedure yields erroneously high values. This, however, can be traced to a poor gauge analysis, suggesting that CMAP has to pay closer attention to how the gauge analysis is incorporated into the final product. The GPCP analysis procedure avoided this pitfall by judiciously using the gauge density information in their merging procedures.

In general, over land the GPCP and the CMAP are close but over the ocean precipitation represented by the GPCP is more reasonable mainly in that the CMAP has an artificial decreasing trend in the global mean oceanic precipitation. This trend results from the input data change and atoll gauge sampling problem. Meanwhile the correlation between the GPCP and the CMAP time series of global mean oceanic precipitation is significantly low. This can be explained by the fact that the two products have much less in common in the input data over the ocean than over land. Specifically, the atoll data have been improperly used by the CMAP for constructing the oceanic precipitation. However, this problem is expected to be solved in the new version of the CMAP that is in development (P. Xie 2004, personal communication). The coexistence of the GPCP and the CMAP as well as other satellite-based precipitation products may bring confusion to the general users. However, although our comparison shows that for oceanic precipitation the GPCP is better than the CMAP, it is still not feasible to conclude that one product supersedes the other because of the lack of a reliable oceanic reference dataset. For example, despite the artificial trend it may be possible that the magnitude of the CMAP could be more realistic than that of the GPCP. Since the GPCP and CMAP are created independently using different analysis methods, they together can be used for cross validation in order to assess and ultimately improve each dataset.

The application of the satellite-based precipitation products should be done with caution. Over land, with more common input data, particularly the use of gauge analyses, the GPCP and the CMAP are better correlated over land than over the ocean. It appears that the gauge analyses and the satellite data complement each other. Over land, gauge measurements are used for large-scale magnitude adjustment and satellite observations are used for large-scale distribution. But in areas with low gauge density or no gauge at all at the different merging procedures used by the two products can make considerable difference (e.g., equatorial West Africa). Over the ocean, the difference between the GPCP and the CMAP is higher than over land. In the tropical and subtropical regions, the two normalized products have high correlations, indicating that even though the two products may have difference in their means, their variabilities are closely represented by the GPCP and the CMAP. Both products have problems in the high latitudes, not only because there are very few gauge stations available, but also because satellite techniques have difficulties retrieving precipitation in areas over snow- and ice-covered surfaces and retrieving cold-season precipitation in general.

Finally, trends in some of the precipitation statistics (e.g., monthly mean, EOF time series) derived from either of the products may not always be true in nature. Some trends are only artifacts of the input data changes, such as the commencement of GPI in 1986 and SSM/I
in 1987, the termination of MSU in 1994, and the sampling problem of atolls after 1996. In view of the diversity of the input data sources over time, considerable caution must be exercised in using either dataset for trend analysis.

Acknowledgments. This study was supported and monitored by the Office of Research and Applications of the National Oceanic and Atmospheric Administration (NOAA) under Grant NA17EC1483. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration or U.S. Government position, policy, or decision.

REFERENCES


---, 2001: Climatic and environmental change in Africa during the last two centuries. Climate Res., 17, 123–144.


---, and ---, 1996: Analysis of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. J. Climate, 9, 840–858.

