Estimating Bias of Satellite-Based Precipitation Estimates

THOMAS M. SMITH

NOAA/NESDIS/National Climatic Data Center, and CICS/ESSIC, University of Maryland, College Park, College Park, Maryland

PHILLIP A. ARKIN

ESSIC, University of Maryland, College Park, College Park, Maryland

JOHN J. BATES

NOAA/NESDIS/National Climatic Data Center, Asheville, North Carolina

GEORGE J. HUFFMAN

Science Systems and Applications, Inc., Lanham, and NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 30 June 2005, in final form 1 December 2005)

ABSTRACT

Systematic biases in satellite-based precipitation estimates can be the dominant component of their uncertainty. These biases may not be reduced by averaging, which makes their evaluation particularly important. Described here are several methods of evaluating the biases and their characteristics. Methods are developed and tested using monthly average precipitation estimates from several satellites. Direct estimates of bias are obtained from analysis of satellite–gauge estimates, and they indicate the general bias patterns and magnitudes over land. Direct estimates cannot be computed over the oceans, so indirect-bias estimates based on ensembles of satellite and gauge estimates are also developed. These indirect estimates are consistent with direct estimates in locations where they can be compared, while giving near-global coverage. For both bias estimates computed here, the bias uncertainty is higher than nonsystematic error estimates, caused by random or sampling errors and which have been previously reported by others for satellite estimates.

Because of their greater spatial coverage, indirect-bias estimates are preferable for bias adjustment of satellite-based precipitation. The adjustment methods developed reduce the bias associated with each satellite while estimating the remaining bias uncertainty for the satellite. By adjusting all satellites to a consistent base, the bias adjustments also minimize artificial climate-scale variations in analyses that could be caused by the addition or removal of satellite products as their availability changes.

1. Introduction

The longest available historical records of precipitation are available from gauges. Over land regions gauge observations are available for thousands of locations over the past century, and in some regions sampling is dense. Compared to satellite data, gauge observations have the advantage of being direct in situ measurements of surface precipitation. They are also usually continuous observations throughout the day. But there are large gaps in gauge sampling over oceans and less developed land regions, and satellite estimates are needed for any globally complete analysis of precipitation. Besides incomplete sampling, gauge observations may have biases, such as those due to blowing snow over high latitudes (e.g., Groisman et al. 1991; Huffman et al. 1997; Bogdanova et al. 2002).

Near-global precipitation estimates from satellite-based observations are available beginning in the 1970s. Because satellite observations are dense compared to gauges, satellite-based observations have smaller spatial sampling and random errors (e.g., Xie and Arkin

Corresponding author address: Thomas Smith, NOAA/NESDIS/National Climatic Data Center and Cooperative Institute for Climate Studies, University of Maryland, College Park, ESSIC, 4115 CSS Building, College Park, MD 20742.
E-mail: tom.smith@noaa.gov

1996, 1997). Random or uncorrelated errors are reduced by averaging a number of observations, even if the observations are not well sampled over a spatial region. Observations need to be spread over a spatial region to reduce sampling errors. Both types of errors may be small for satellite data, which tend to have many observations per month over large regions. Satellites do not completely sample the Earth's surface, but compared to in situ observations their sampling is relatively dense. A larger problem with satellites is bias since they remotely measure processes in the atmosphere. Biases may be due to a diurnal sampling bias, tuning of the instrument or the precipitation algorithm, or unusual surface or atmospheric properties that the algorithm does not correctly interpret. These biases have been widely studied over the last several decades (e.g., Scofield 1987; Rosenfeld and Mintz 1988; Morrissey 1991; Xie and Arkin 1997; Gruber et al. 2000; McCollum et al. 2000, 2002; Bowman et al. 2003). The newer satellite algorithms and some of the newer instruments may reduce bias relative to older algorithms and instruments. However, it would be difficult to develop an algorithm that accounts for all possible sources of bias, and the older instrument records are needed for evaluation of variations over the last several decades. Therefore, methods are needed for analyzing satellite biases, relative to each other or to some standard.

Here several methods for evaluating biases in satellite mean monthly precipitation are developed and tested. First, satellite biases are directly evaluated relative to gauges in regions where both estimates are available. The relative biases between different satellite records are evaluated to see how consistent biases are between satellites. Based on those results, an indirect-bias estimate that is not dependent on gauges is developed.

2. Data

All data used in this study are monthly averages of the satellite precipitation and gauge data, averaged to a 2.5° latitude–longitude grid. Most data used are the same data that are used for the Global Precipitation Climatology Project (GPCP; Huffman et al. 1997, Adler et al. 2003). Additional supplementary data are used to extend the analyses and to test the differences caused by using different datasets.

a. Satellite precipitation data

Most of this study is based on eight satellite-based estimates of precipitation: outgoing longwave radiation (OLR) precipitation index (OPI), Geostationary Operational Environmental Sounder (GOES) precipitation index (GPI), adjusted GPI (AGPI), the Special Sensor Microwave Imager (SSM/I) composite (SSM/Ic), SSM/I emission (SSM/Ie), SSM/I scattering (SSM/Is), Televion Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS), and the SSM/I–TOVS estimates. The OPI data are taken from the Climate Prediction Center (CPC) Web site (http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cams_opi.html; Xie and Arkin 1998; Janowiak and Xie 1999). The remaining satellite data are from the GPCP data archive (http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html; GPCP version 2; Huffman et al. 1997; Adler et al. 2003). The OPI data are available beginning in 1979. The other satellite records all begin in 1986 or 1987. Data through 2003 are used for this study.

There are other shorter-period satellite-based estimates, including the Tropical Rainfall Measuring Mission (TRMM) and Advanced Microwave Scanning Radiometer (AMSR-E) estimates. TRMM begins in 1998 and the AMSR in 2002, while the primary satellite products used here all begin in 1987 or earlier. Here the TRMM estimates are used to help validate the indirect method. Because of its brief period of record the AMSR data are not used in this study.

b. In situ precipitation data

The primary gauge data used are the Global Precipitation Climatology Center (GPCC) gauge data (Rudolf et al. 1994). They are the same gauge data used by GPCP. Those gauge data are analyzed and some interpolation is used to fill regions without sampling (Huffman et al. 1997; Adler et al. 2003). Here the number of gauges in each monthly 2.5° square is used to exclude gauge data from squares with no observations (i.e., where GPCC data were interpolated to fill an area). The GPCC gauge data is itself based on several datasets, each using varying degrees of quality control, processing, and adjustments. To evaluate the effect of these processes on bias estimates, the gauge data of Chen et al. (2002) are also used to estimate bias.

3. Direct-bias estimates

Briefly, direct-bias estimates are analyzed differences between satellite and gauge estimates of precipitation. The advantage of direct estimates is that they can quantify for each satellite its bias properties, such as magnitude, variation, and spatial extent. The major disadvantage of direct estimates is that they can only be computed in the neighborhood of gauges, which have a limited spatial coverage.

a. Methods

Direct-bias estimates are computed from analyses of satellite–gauge precipitation differences for each satellite. Data used are the monthly 2.5° GPCP/GPCC data discussed in section 2. The analysis is an optimum interpolation (OI), which produces a smoother and more complete bias estimate compared to simpler methods such as linear interpolation [see appendix A and Reynolds and Smith (1994) for more details on OI]. The OI computes a set of optimal weights. Each observation is assigned a weight that accounts for its noise and its covariance with the interpolation point. Here covariance is modeled as a function of distance from the interpolation point, measured from the centers of the 2.5° regions. Observations with less noise and greater covariance are given greater weight, compared to more noisy or more distant points. For each region the noise of satellite–gauge differences is here defined to be inversely proportional to the number of gauges in the 2.5° region. In regions without gauges there are no differences. Those regions are filled in the OI using surrounding differences. The OI analysis is the weighted sum of the data using these optimal weights. In the version of the OI used here, the analysis is damped toward zero bias when there are few or no data. When data are dense there is practically no damping. Here at least two observations are required before the OI analysis is performed, with no analysis for more sparsely sampled regions. The weights minimize the error of the analysis, assuming that statistics of the biases are known.

These analyses are computed for each month using differences from over 12.5° latitude–longitude squares, to produce an analyzed difference for the central 2.5° region. To test the difference made by using a smaller data-selection region, we also compute the OI using data from 7.5° latitude–longitude regions, using data from three months centered on the analysis month. Comparisons indicate that there is little difference between the 12.5° one-month OI and the 7.5° three-month OI. The zonal and meridional *e*-folding scales of the differences were estimated where sufficient data were available. In general, the meridional scales are most consistent, ranging between 800 and 1200 km at most latitudes. The zonal scales are largest in the Tropics, where they can be 3000 km or larger. Zonal scales decrease to about 1000 km or less between 45° latitude and the poles. Because of the short overlap period when all data are available, and because our analysis assigns a scale lower than most measured correlation scales, as discussed below, we did not evaluate seasonal variations in the bias scales.

These scales are relatively large, considering that precipitation itself can be caused by smaller-scale processes such as topography, frontal boundaries, and thunderstorms. Several factors cause these scales to be larger. One is the monthly averaging of the data, which filters out individual precipitation events. Another is the differencing between satellite and gauge data. Compared to the precipitation, biases responsible for the differences may be caused by larger-scale processes such as systematic satellite instrument or algorithm bias, or algorithm error caused by large-scale environmental variations such as variations in the moisture field.

Although the estimated *e*-folding distances are nearly always larger, the OI analysis uses data from within 12.5° squares and it assumes a constant *e*-folding distance of 750 km. This produces a conservative analysis, to minimize problems that may occur near discontinuities in topography or land–sea boundaries. In addition, the noise-to-signal variance ratio for individual gauges is set to 1 (see Appendix A). Monthly averages are sampled repeatedly over the month from both satellites and gauges, and many 2.5° squares will average several gauges, which would further reduce noise, so it is likely that the noise/signal ratio is smaller than what is assumed here. This setting of the ratio is also conservative, to avoid overinterpolation of the data. As discussed below, these OI statistic settings produce strong analyses when the 12.5° sampling regions are more than about a third full, with exponential damping to zero bias as sampling becomes sparser.

b. Results

Examples of the directly computed satellite biases are shown in Fig. 1, for the SSM/Ic and GPI satellite estimates. The period 1996–2003 is representative, and data are available for all satellites for that period. Biases are on the order of a few millimeters day^{-1}, and their magnitudes are generally about half or less than the typical rainfall in regions where they occur. The SSM/Ic bias is largest in the Northern Hemisphere in winter, when the estimate is too low compared to gauges. This may be due to incorrectly flagged surface snow and ice cover, or from sampling gaps caused by elimination of data due to surface snow and ice. The GPI bias has maximum bias near 40°N in winter, where the GPI estimate is too high. That is near the northernmost limit of GPI data. Near the same time the GPI estimate is also high near the equator. The apparent cycle in Fig. 1 for the GPI and SSM/Ic products suggests that both in the Tropics and in the extratropics these algorithms should be seasonally adjusted, as is done

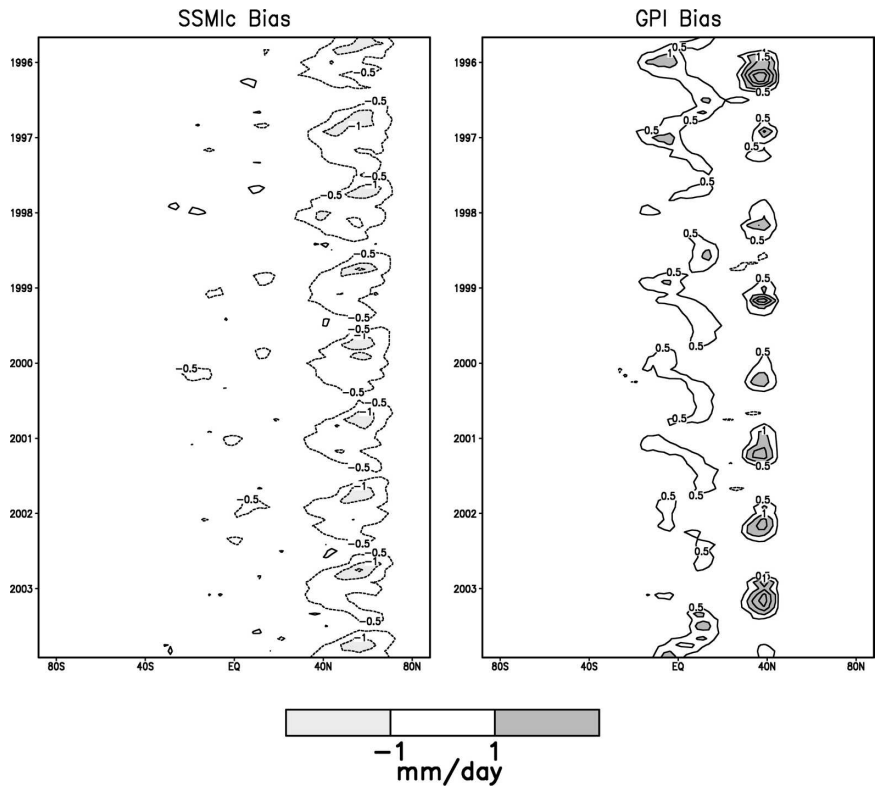

FIG. 1. Directly analyzed zonal-average bias from the SSM/Ic and GPI, in mm day^{-1}. The contour interval is 0.5 mm day^{-1}, with the zero contour omitted and shading as indicated.

with some of the other satellite algorithms (e.g., OPI and TOVS).

The time-average direct bias over 1996–2003 for SSM/Ic and GPI (Fig. 2) indicates that both estimates have a large bias over equatorial Africa. For that region, evidence suggests that the overestimation is caused by virga, which is interpreted as surface rainfall by the satellites (McCollum et al. 2000). The GPI is also high over mountainous regions and in the Tropics around Indonesia. Over mountainous regions the algorithm may be interpreting snow and ice as precipitation. Over the Himalayas the surface temperature is cold enough to make the GPI count it as precipitating clouds, and around Indonesia the abundance of high clouds may be responsible for the bias.

Table 1 gives the direct root-mean-squared bias (RMSB), over 1996–2003. The RMSB is computed from the analyzed satellite–gauge biases. The spatial RMS of the bias is computed globally, but only for regions where the number of defined 2.5° differences, n, is large enough data to minimize damping ($n \geq 20$; see Fig. 3). In addition, the global satellite-to-satellite RMS differences are also given ($\mathrm{RMSD_{sat}}$). For the $\mathrm{RMSD_{sat}}$, values are computed spatially over all regions where satellite pairs are defined, and between the given satellite and all other satellites. Thus, the $\mathrm{RMSD_{sat}}$ is computed over both land and water, while the RMSB is only computed from land data. For both the magnitudes are usually similar. The existence of significant $\mathrm{RMSD_{sat}}$ values indicates that satellite biases are different, and they could be reduced by combining different satellite products. The similarity in magnitude of RMSB and $\mathrm{RMSD_{sat}}$ also suggests that biases over oceans are approximately the same magnitude as those over land.

An estimate of direct-bias error can be estimated from the OI damping of bias as sampling becomes more sparse. With few data, the bias analysis is damped to near zero and the error is approximately the RMSB for the satellite. With many data there is little damping and little error in the bias estimate. Damping with the number of data, n, is shown using average results where bias is assigned a constant value of 1 and the observed sampling is used for all satellite–gauge differences (Fig. 3). This is referred to as the $\mathrm{OI_1}$ analysis. With 20 or more differences there is almost no damping. An exponential fit to $\mathrm{OI_1}$ has the least error using an e-folding scale of $r = 6$, indicated by the dashed line. This simple rela-

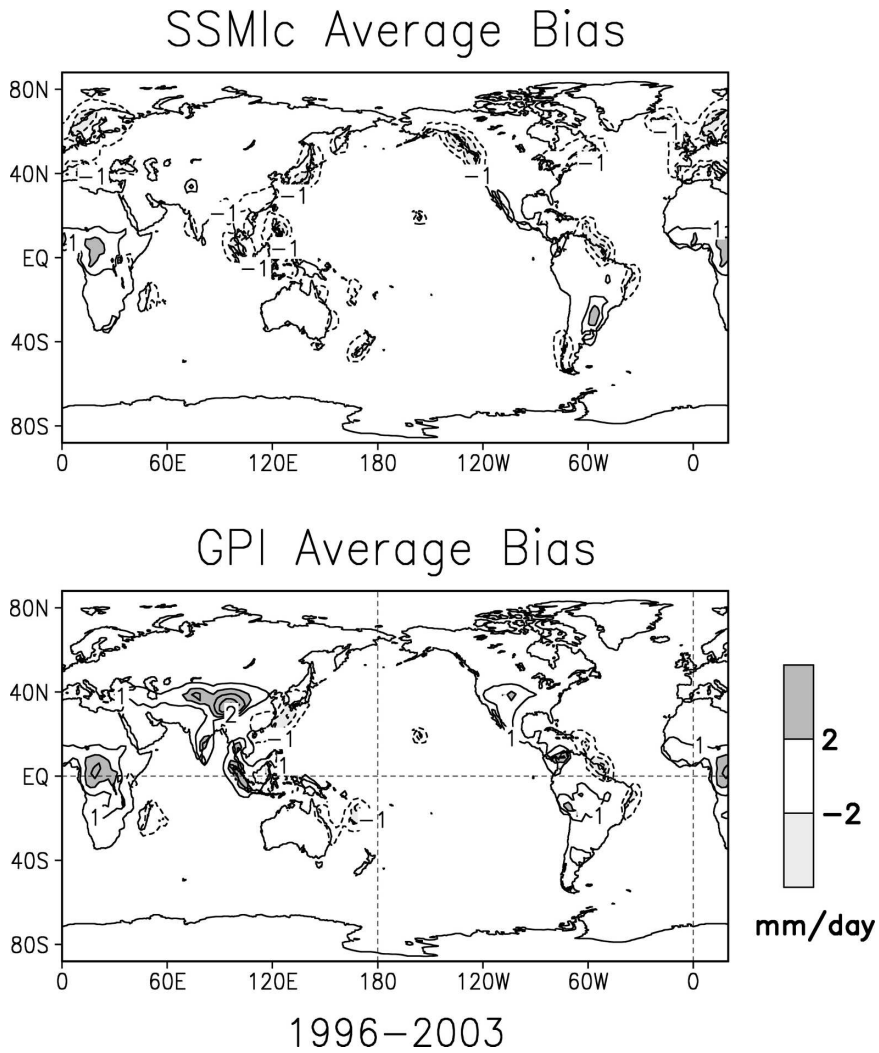

FIG. 2. Time-average precipitation directly analyzed bias (1996–2003), for the SSM/Ic and GPI satellite estimates, in mm day^{-1}. The contour interval is 1 mm day^{-1}, with the zero contour omitted and shading as indicated.

tionship may be used to scale the RMSB to estimate direct-bias uncertainty as a function of n.

To test the sensitivity of these results to details of the gauge dataset, the direct-bias analysis was repeated using the Chen et al. (2002) gauge analysis masked to match the GPCC gauge sampling. Comparison of the results of this analysis to the original analysis shows that they both give nearly identical results. The largest differences occur associated with islands in the western tropical Pacific, where the Chen et al. (2002) data include some gauge data screened out of the GPCC gauges. The global spatial correlations (Murphy and Epstein 1989) between the GPCC-gauge biases and the Chen et al.–based biases are about 0.8 to 0.9 for all products except the OPI, which has correlations of about 0.6 to 0.8. Thus, although they differ in some details, the major bias results presented are not acutely dependent on the choice of the gauge data used.

4. Indirect-bias estimates

Direct-bias estimates are unavailable over oceans and land regions without gauges. To overcome this problem, an indirect-bias estimate was developed. Development of the indirect method is supported by the results of Table 1, which shows that satellites tend to have different biases. Thus, their combination can have a lower bias than any of the individual satellite products. This was found to be the case for biases of satellite-based estimates of sea surface temperature (SST; Reynolds et al. 2004). Their findings suggest that an indirect approach can significantly reduce SST bias in regions without in situ data.

TABLE 1. RMSB, computed using all well-sampled analyzed biases over land from each satellite, from 1996 to 2003 and globally. All SMMI satellites (composite, emission, and scattering) have similar RMSB. Also given is the RMS difference between each satellite and every other satellite, over all regions where the satellites are defined.

Satellite	RMSB (mm day^{-1})	RMSD$_{sat}$ (mm day^{-1})
OPI	0.9	2.0
AGPI	1.5	1.7
GPI	2.0	2.1
SSM/I	1.8	2.0
SSM/I–TOVS	1.5	1.8
TOVS	1.2	2.0

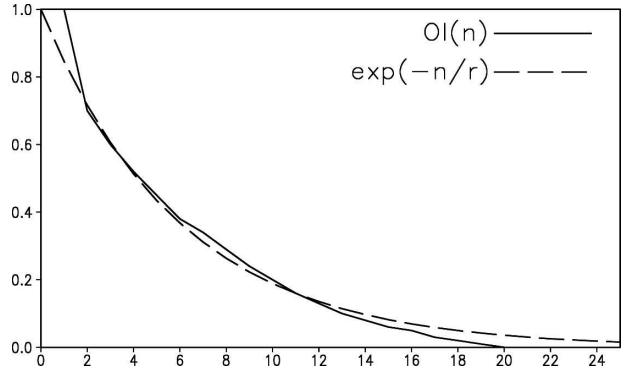

FIG. 3. Error of the OI with all values set to 1, as a function of the number of values, n. Also shown is the exponential function approximation as a function of n, using r = 6.

a. Methods

For the indirect method, the bias is the deviation of the long-term mean (LTM) of a satellite estimate from an estimate of the true LTM precipitation. Long term is defined as the average over a number of years, either for a given month or season, or for all months if the annual mean is considered. Here the method is tested for the annual mean. In general, the long-term average should be at least 4 or 5 yr to prevent one or two unusual years from skewing the estimate (i.e., an ENSO episode). Here an 8-yr average is used to provide sufficient time averaging while allowing several different bias estimates to be computed within the satellite period.

For each precipitation estimate, i, the long-term average is

$$A_i = \frac{1}{N}\sum_{t=1}^{N} P_i(t). \tag{1}$$

Here t is the year, from year 1 to N, and P_i is the precipitation for estimate i, for $i = 1$ to k. If the estimates have different biases that range around zero bias, then an estimate of the true LTM may be computed by combining the k products. There are several ways to combine them, including computing their mean or their median. Here the median is used because it eliminates outliers. Thus, the true LTM is estimated by

$$A_T = \text{median}\{A_i\}, \quad i = 1, \ldots, k. \tag{2}$$

Here the gauge data may be included as an estimate along with the satellite estimates, although the gauge data are only defined over land. Tests of the method are performed with and without gauge data. An estimate's weight in Eq. (2) may be increased by including more than one copy of that estimate. Using the method with gauges, several different gauge weightings are tested. In these tests, the LTMs of the GPCC gauge data are binomially smoothed spatially and gauges associated with isolated islands are removed.

Using Eqs. (1) and (2), the bias of each satellite product is

$$B_i = A_i - A_T. \tag{3}$$

This bias may be removed from each type so that they all have the same LTM value. Such adjustments should minimize the bias. However, while different, the satellite biases are not entirely independent. For example, for the eight satellites that this study uses, the direct-bias estimates have an average spatial correlation of 0.5, with values ranging between about 0.2 and 1.0 for the different estimates. In addition, where gauges are available they may also contain biases. Therefore, these adjustments cannot remove all precipitation biases. They will reduce bias and minimize artificial climate-scale variations that may be caused by changes in data types over the historical record.

The bias correction given by (3) accounts for long-term bias over a number of years. However, it cannot correct for short-period transient bias induced by meteorological variations, such as anomalous snow or ice cover. The shortest-period bias that this can resolve is defined by the long-term averaging period. Direct-bias estimates resolve transient biases. However, those estimates show that most bias is either constant or its variance is dominated by a seasonal cycle. Thus, this limitation should not be a major problem for bias adjustment. However, it would still be desirable at some future time to have bias adjustments for all time scales.

In a bias-adjusted analysis, the bias of each satellite type would be removed, with respect to the estimated LTM. The bias error of the analysis is thus the error of this LTM estimate. The mean-squared error of the LTM may be estimated from the ensemble spread by

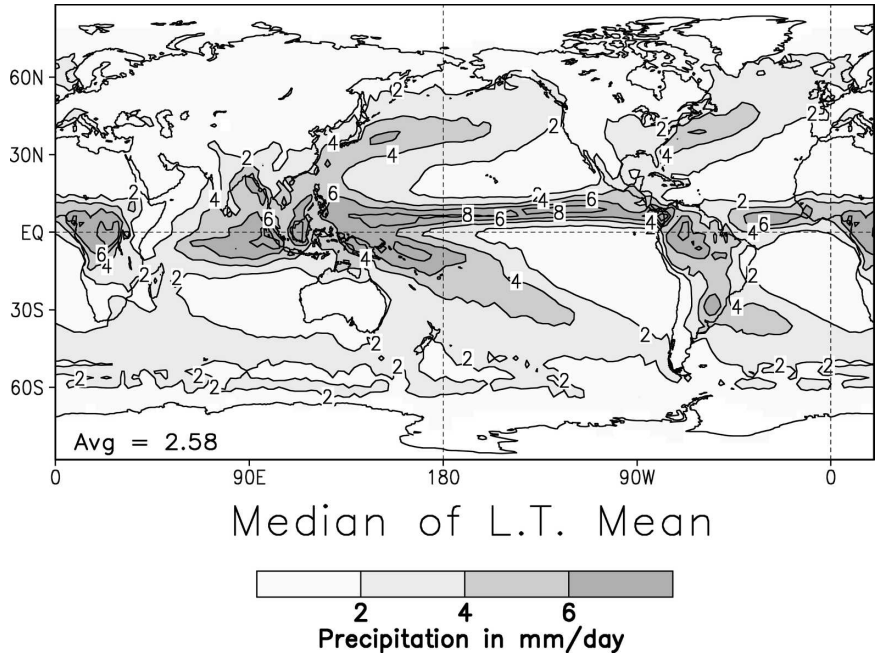

FIG. 4. The ensemble long-term mean precipitation in mm day^{-1}, using only satellite data (gauge weight = 0). The contour interval is 2 mm day^{-1}, with shading as indicated.

$$E_m^2 = \frac{1}{k}\sum_{i=1}^{k}(A_T - A_i)^2. \quad (4)$$

Outliers that are filtered out of the median estimate [Eq. (2)] may influence this mean-squared estimate. To limit their influence, the high and low $A_i(m)$ values can be discarded, and the error computed using the remaining k-2 values. In discussions below, this is referred to as the truncated error estimate. This bias error is independent of other types of analysis errors such as random and sampling errors.

b. Results

Ensemble LTM precipitation is computed using Eq. (2), using the eight GPCP satellite estimates alone and also using the satellites plus gauge data. Here data averaged over the 8-yr period 1996–2003 are used for all months to show the annual LTM estimate. At least 10 individual months are required for the LTM of each type, and at least four LTMs are required to compute the ensemble median. These restrictions eliminate estimates over polar latitudes. The ensemble shows the familiar mean patterns of precipitation (Fig. 4; for comparisons see, e.g., Xie and Arkin 1997).

Compared to gauge LTMs, the ensemble LTM from satellites is similar over most regions. However, over central Africa the satellite-based ensemble LTM is up to 6 mm day^{-1} larger than the gauge estimate. There tends to be an overestimate of all the satellite estimates in that region (McCollum et al. 2000). Because the satellite biases are locally correlated, they are not reduced in the ensemble. Including gauges in the ensemble greatly reduces the differences. Before including gauge data in the ensemble, the gauge LTM is smoothed and filled slightly using spatial binomial filters. Smoothed gauge data are not allowed to extend beyond the coasts, and data from isolated islands are excluded from the ensemble. Weighting the gauge data the same as every other satellite has little effect over central Africa since there are generally seven satellite estimates for that region. However, weighting the gauge LTM to the equivalent of four satellites greatly reduces the difference over central Africa (Fig. 5). The factor of 4 was found by testing several factors and evaluating the results. A factor of 9 would guarantee that the satellite estimates were filtered out of the median where gauge estimates are available. Computing the median using a factor of 4 will filter out the four most extreme satellite estimates.

Averaged globally, the LTM ensembles from both satellite-only and satellite plus gauges are almost the same, as indicated on the lower left of the panels (2.58 and 2.60 mm day^{-1}, respectively). However, over land the global mean-absolute difference from gauges for the satellite-only ensemble is 0.79 mm day^{-1}, while including gauges reduces it to 0.47 mm day^{-1}. Differences

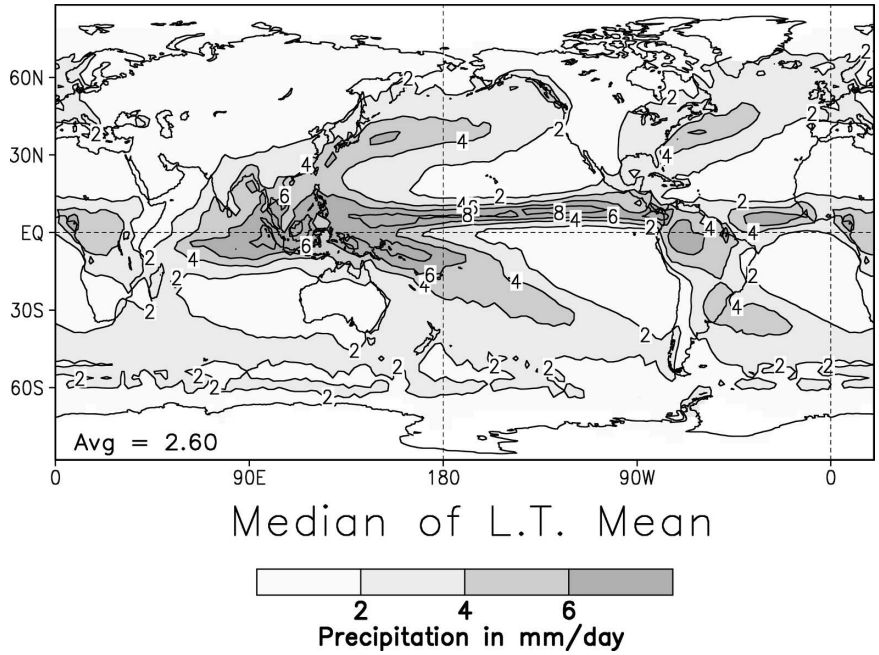

FIG. 5. The ensemble long-term mean using both satellite and gauge data, in mm day^{-1}. The gauge data are weighted times 4 where available (gauge weight = 4). The contour interval is 2 mm day^{-1}, with shading as indicated.

remain because of the use of ensembles and because the gauge data used for the ensemble are smoothed GPCC gauges, while the validation gauge data are the unsmoothed GPCC gauges.

The root-mean-square error estimates [RMSE; the square root of the error variance computed using Eq. (4)] are computed for the LTMs shown above. Comparisons are shown of the RMSE divided by the ensemble mean, with and without gauges in the ensemble and also with and without truncation in the estimate of the RMSE.

Using satellite estimates only, without truncation, the ratio is small over most regions (Fig. 6, upper panel). However, it is large over regions that receive low precipitation, where both the RMSE and the mean are small. It is also large over high latitudes, including over the oceans south of 45°S. The global average of the ratio (on the lower left of the panel) is 0.33. This average ratio is higher than the uncertainty ratio estimate based on the global energy balance study of Kiehl and Trenberth (1997), which gives an estimated ratio of 0.2. This energy-balance-based uncertainty ratio is itself only a crude overall estimate of uncertainty, based on uncertainty estimates in components of the energy balance. It is used here only for rough comparisons to our results. Our overall results are supported by the rough consistency with estimates based on the independent energy-balance method.

To test the stability of this estimate we compute the truncated version of the ratio (Fig. 6, lower panel). The patterns in the truncated version are similar, but the high values are reduced because of removal of the extremes. The global average value for the truncated version is also reduced to 0.22, close to the value estimated from the global energy balance. In both of the satellite-only ensemble estimates, the RMSE over central Africa is relatively low, since all satellites tend to be biased high in that region.

Including gauges (weighted four times) in the ensemble yields similar overall results to the ratios without gauges in the ensemble, and the global averages are similar for the full and truncated estimates (Fig. 7). Including the gauges also increases the uncertainty in central Africa, because the satellites all tend to give higher values than the gauges.

The global indirect RMSB (Table 2) is computed using values wherever they are defined for the given satellite. These global values are smaller than the direct method RMSB values, discussed earlier. Both the direct-method RMSB and the global $RMSD_{sat}$ are between about 1 and 2 mm day^{-1}, while the ensemble-method RMSB values are all smaller. For both methods the RMSB of the GPI is largest, indicating that it tends to be the most biased. The direct method RMSB and $RMSD_{sat}$ are larger because they account for month-to-month variations, while the indirect RMSB only ac-

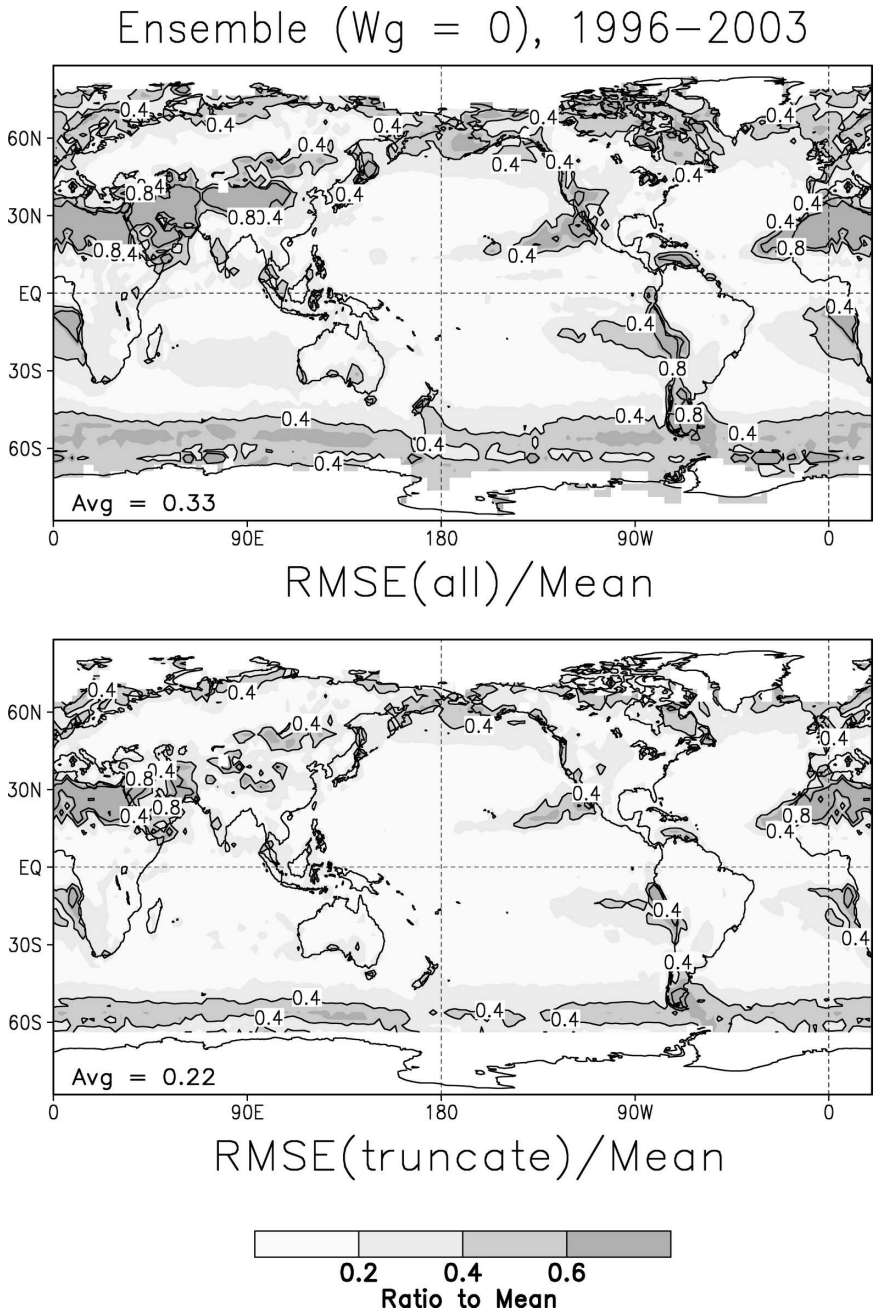

FIG. 6. Ensemble estimates of RMSE divided by ensemble mean using only satellite data, using (top) all estimates and (bottom) truncating the high and low estimates of the ensemble. The 0.4, 0.6, and 0.8 contours are drawn, with shading is as indicated.

counts for the LTM bias, averaging out shorter-period variations.

Why the ensemble method should reduce the LTM bias and validation against several sets of measurements is given in appendix B. Here validation against TRMM precipitation estimates is summarized. The TRMM estimates are only available beginning in 1998 for the region 40°S–40°N. But for their limited period, they may give less biased precipitation estimates. At a few oceanic locations where in situ data are available, Bowman et al. (2003) found that the TRMM estimates tend to have low biases. Their validation is limited to a few sites, but they are encouraging and they suggest that TRMM should be useful for comparisons over the

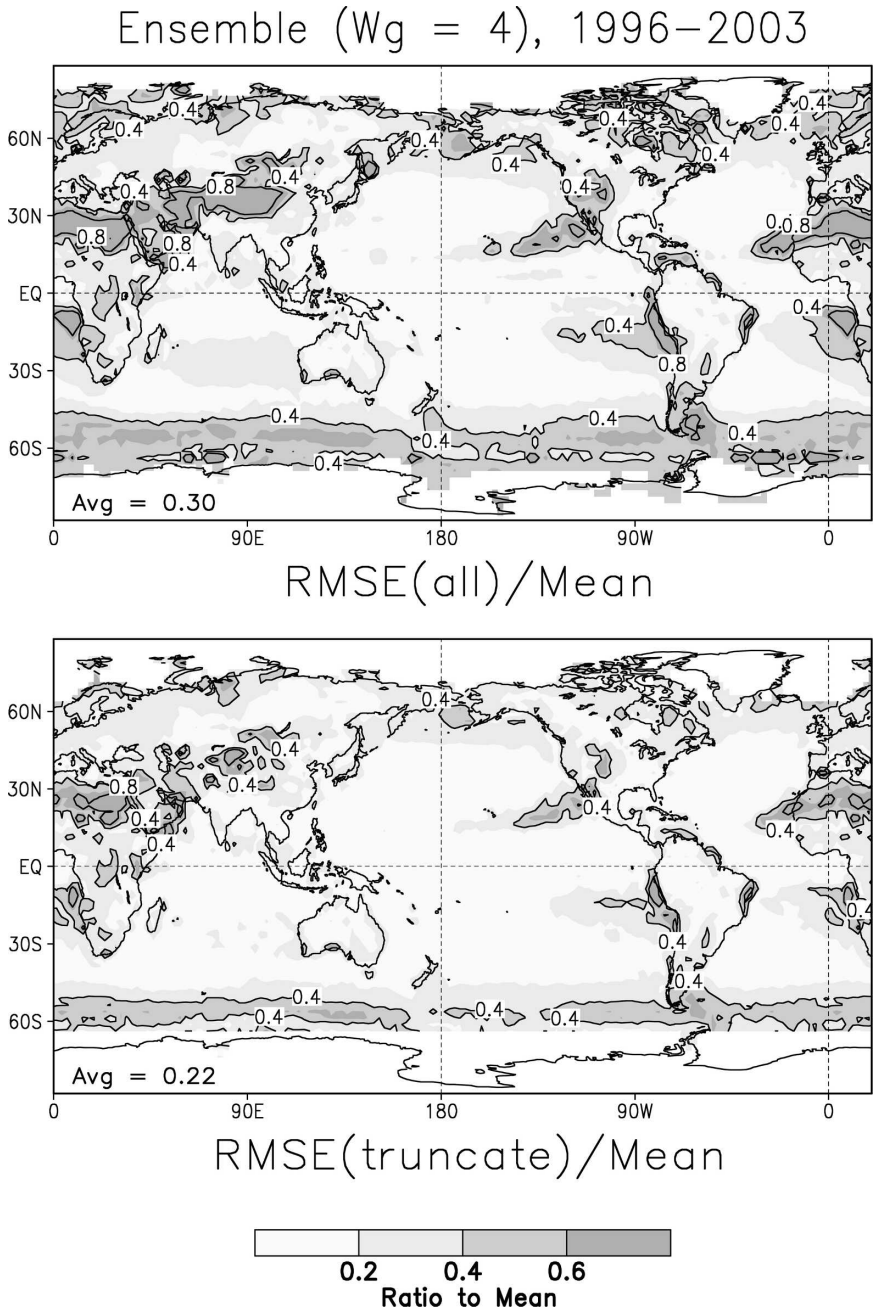

FIG. 7. Ensemble estimates of RMSE divided by ensemble mean using both satellite and gauge data, using (top) all estimates and (bottom) truncating the high and low estimates. The 0.4, 0.6, and 0.8 contours are drawn, with shading as indicated.

larger region. As a test, the satellite-only ensemble for 1998–2003 was compared to both TRMM precipitation products: the microwave (TMI) and the radar (PR) estimates. For each, comparisons of the average absolute bias over the entire region were computed, for each individual satellite and for the ensemble. For the individual satellites, the bias relative to TMI is on average 0.68 mm day^{-1} and the bias relative to PR is on average 0.93 mm day^{-1}. Compared to both TMI and PR, the SSM/Ic and AGPI have low biases while the OPI and GPI biases are high. For the ensemble without gauge data, the bias relative to TMI is reduced to 0.41 mm day^{-1} (a 40% reduction from the average bias) and relative to PR it is reduced to 0.73 mm day^{-1} (a 22%

TABLE 2. RMSB relative to the ensemble estimate of the long-term mean. The time-averaging period is 1996–2003 with weighting of gauges by 4 and averaged globally wherever the satellite product is defined.

Satellite	RMSB (mm day^{-1})
OPI	0.6
AGPI	0.5
GPI	1.1
SSM/I	0.7
SSM/I–TOVS	0.7
TOVS	0.8

reduction). This comparison confirms a reduction in bias.

Where the mean can be validated against gauge or TRMM estimates, the indirect method reduces the bias. The indirect method also has a major advantage over the direct method: it can analyze the bias with or without gauges. This advantage makes the indirect method preferable to the direct method.

c. *Reconstruction of indirect estimates*

A problem with the indirect method is that it can only be used when there are a number of products for computing an ensemble, and for 1979–86 there are only gauges and the OPI satellite. Since the ensemble LTM cannot be directly estimated for periods before 1986, a method was developed to reconstruct most of the LTM variance using the available data.

The reconstruction is based on the OPI data and empirical orthogonal function (EOF) modes of the LTM, computed from the more recent data (Fig. 8). The LTM was computed for 10 overlapping 8-yr periods: 1987–94, 1988–95, . . . , 1996–2003. This is minimal data for computing EOFs, so only the first two modes are considered, accounting for 86% of the LTM variance. For these EOFs, undersampled regions are filled using spatial binomial filtering. Most EOF variations are in the Tropics. The first mode indicates a trend in the LTM, which is largest over the Indonesian region with teleconnections across the Pacific and into the extratropics. The second appears to be linked to low-frequency variations associated with ENSO precipitation.

To reconstruct the time series for the first two modes, the OPI data are used to compute OPI LTMs. Those OPI LTM estimates are projected onto each of the EOF modes to minimize the mean-squared error of each fit. This gives the relative variance of the OPI associated with the mode. However, the OPI data are themselves biased, so the time series must be further bias corrected. This time series bias correction is done by linear regression of the OPI-projected series against the EOF series. The resulting coefficients are used to correct the OPI-projected series.

The OPI data are able to almost completely reconstruct the first-mode variance (Fig. 9, upper panel), and they are also able to reconstruct most of the second-mode variance (lower panel). Over the base period the first mode appears to be a trend, but over the extended period there is a suggestion of a decadal oscillation. The OPI monthly values have damped variance compared to some other satellite products, but these reconstructions indicate that their LTM values can be used to reconstruct the ensemble LTM variance.

Much of the variance in the first two modes is over land, especially for the first mode. Therefore, reconstructions of the LTM using gauge data are also tested. For the first mode the correlation is high for the base period (0.99), but there is less variance in the independent period compared to the OPI projections. For the second mode, the gauge-projection correlation is lower (0.54) compared to the OPI-projection correlation (0.83). These comparisons indicate that the greater sampling of the OPI should make its reconstruction significantly more accurate than the gauge-based reconstruction.

The reconstruction LTM error occurs because of damping of the LTM variance. Total LTM variance, σ^2_{LTM}, is computed from the same 10-yr base period used for the EOFs and is the sum of the signal variance, σ^2_s, and the error variance, E^2_m,

$$\sigma^2_{\mathrm{LTM}} = \sigma^2_s + E^2_m. \tag{5}$$

The error variance is computed from the means of individual products, and it cannot be assumed to be random over the EOF analysis period. Here the error variance is assumed to be part of the EOF analysis, and it is damped as much as the signal variance. Thus, if f_s is the total fraction of the variance accounted for by the set of EOFs, then the EOF-damping error is $(1 - f_s)(\sigma^2_s + E^2_m)$. In addition to the EOF-damping error, there is also error from the reconstructed error variance, $f_s E^2_m$. The sum of these gives the total error variance of the reconstruction LTM,

$$E^2_R = E^2_m + (1 - f_s)\sigma^2_s. \tag{6}$$

This error estimate has one term reflecting uncertainty in the estimate of the LTM and another term reflecting uncertainty in estimates of temporal changes of the mean.

The values of σ^2_{LTM} and E^2_m can be computed using the base-period data, allowing an estimate of σ^2_s to be computed using Eq. (5). The temporal variation in the LTM is relatively small, and in most regions $\sigma^2_s < 0.1\ E^2_m$.

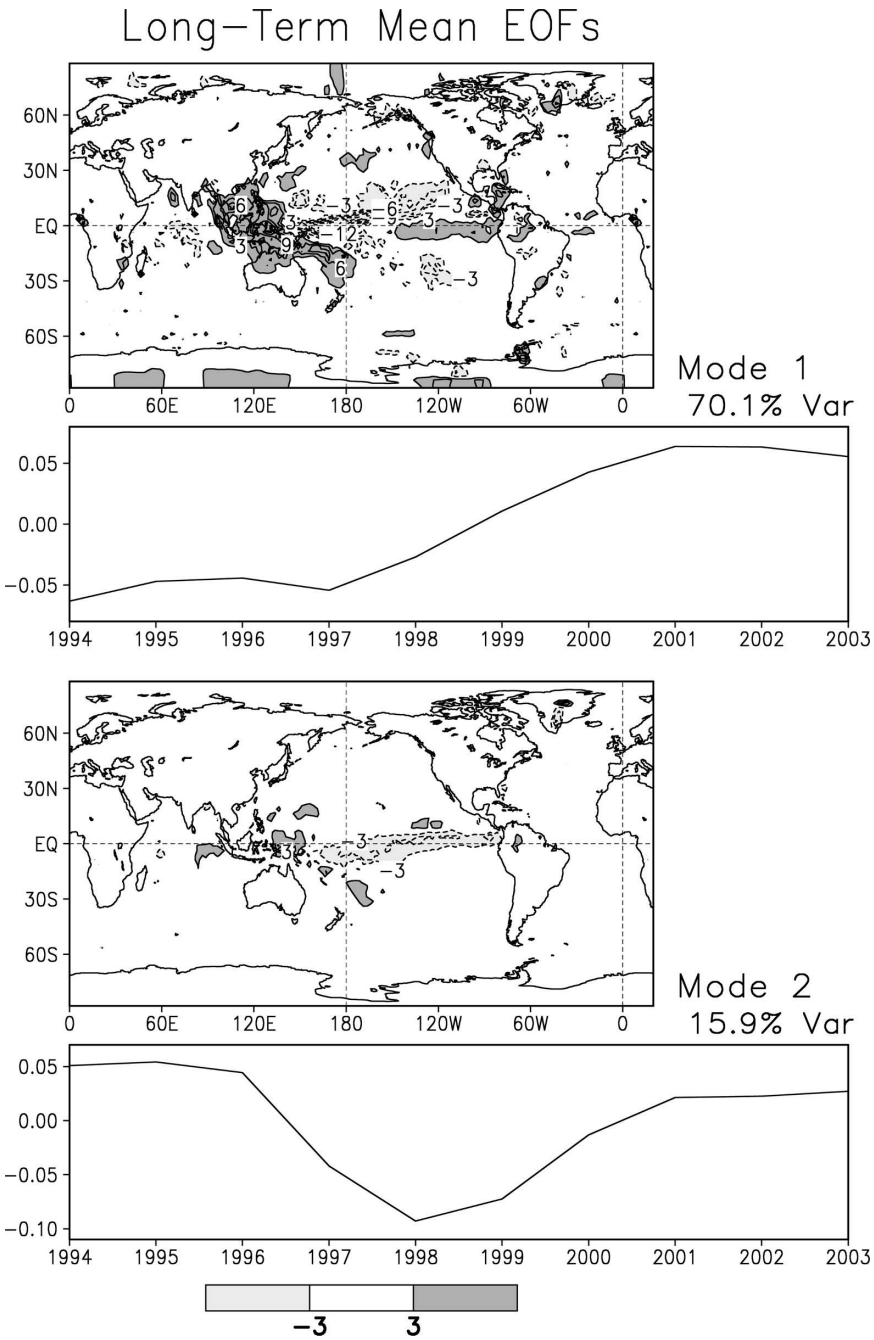

FIG. 8. EOF modes 1 and 2 of the ensemble long-term mean. There are ten 8-yr overlapping periods. The time series is labeled with the last of the 8 yr in the period. The contour interval is 3, with the zero contour omitted and shading as indicated.

Using the leading two EOF modes, the smallest possible signal-damping error is $0.14\sigma_s^2$, when both are perfectly reconstructed. In practice, the error is dependent on how well each mode may be reconstructed, which is reflected by the explained variance of the mode reconstruction, r^2. Reconstruction statistics for the first-two-mode reconstructions, based on OPI and gauge data, are listed in Table 3. For the OPI reconstruction, $f_s = 0.81$ and damping error $= 0.19\sigma_s^2$, while for gauges-based reconstructions, $f_s = 0.74$ and damping error $= 0.26\sigma_s^2$.

Differences in the LTM reconstruction, relative to the full LTM over 1996–2003, are small and mostly in the Tropics (Fig. 10). Note that this difference is much

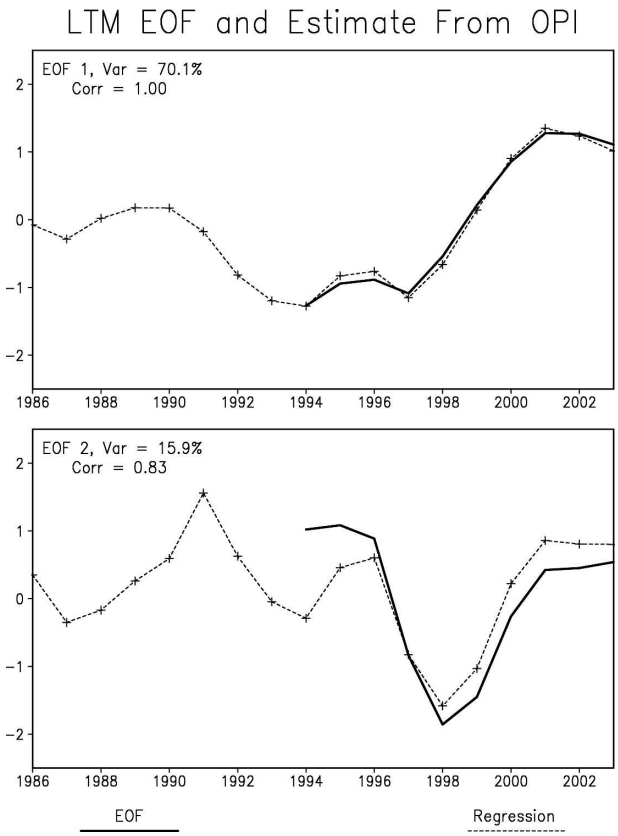

FIG. 9. Time series for the LTM EOFs 1 and 2 and similar series computed by projecting OPI data onto the modes. The regression time series is the projected time series corrected for bias, based on the common period. The % variance of the mode and the correlation with the EOF time series is indicated in each panel.

less than the LTM values (shown in Fig. 5). The reconstruction RMSE for the same period is slightly larger in the tropical Pacific. However, the temporal variance being reconstructed by the modes is still a small fraction of the total LTM error.

5. Summary and conclusions

Two methods for evaluating the bias of satellite-based precipitation estimates are developed and tested. A direct method gives higher temporal resolution, but it is not able to evaluate bias over oceans. The direct method is based on local analyses of satellite–gauge differences. An indirect method gives near-global spatial resolution, but it is only able to resolve the LTM bias. The indirect method is based on ensembles of the various precipitation products. In the neighborhood of gauges, those gauges can be incorporated into the indirect method. In addition, most of the variance of this indirect method can be reconstructed using its leading modes, allowing the indirect bias to be reconstructed for some years when only the OPI satellite product is available. Because it can be computed near globally, the indirect method based on ensembles is preferred for bias adjustments.

TABLE 3. Statistics needed for error estimation of reconstructed LTM, using both OPI- and gauge-based reconstructions.

	OPI				Gauge		
Mode	F	r	r^2	r^2F	r	r^2	r^2F
1	0.70	1.00	1.00	0.70	0.99	0.98	0.69
2	0.16	0.83	0.69	0.11	0.54	0.29	0.05
Sum	0.86			0.81			0.74

Indirect biases are typically about 0.5–1 mm day^{-1}, with standard errors typically about 0.2 times the mean precipitation. Bias errors are computed from the spread of the ensemble members, and they tend to be largest in regions where the mean precipitation is large. In tests of the indirect-bias estimate computed without gauge data, the indirect-bias estimates are consistent with satellite–gauge direct-bias estimates over most regions where the comparison can be made. Although it is difficult to remove all satellite bias using these indirect-bias estimates, bias-adjusted satellite precipitation has several advantages over unadjusted satellite precipitation. The global adjustment reduces the bias of each satellite estimate, over oceans and land. The uncertainty estimate for the adjustment allows users to make an estimate of how the remaining bias may be affecting low-frequency changes in the precipitation. In addition, by adjusting all satellites to a common base, artificial changes in precipitation analyses are minimized. Artificial changes are possible whenever the mix of satellite data in an analysis is changed if the satellites have different biases. Adjusted data will all contain the same, reduced bias.

Acknowledgments. We thank R. Ferraro, A. Gruber, J. Janowiak, C. Kummerow, R. Reynolds, R. Vose, P. Xie, and X. Yin, and an anonymous reviewer for reviews and helpful discussions and suggestions. The views and opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision.

APPENDIX A

A Brief Discussion of the Optimum Interpolation Method

Optimum interpolation is a method of statistical interpolation, developed for meteorological data by Gan-

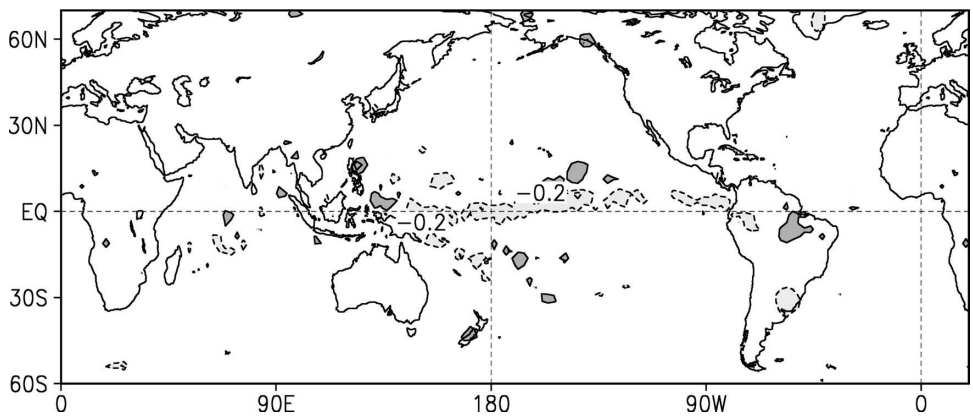

FIG. 10. The difference between the OPI-based reconstruction and the ensemble estimate (contour interval = 0.2 mm day^{-1}, with the zero contour omitted). Gauge data are included in the ensemble estimate, weighted × 4.

din (1963). The OI method used here is similar to the OI described in detail by Reynolds and Smith (1994). The OI interpolates data to a point, k, using a weighted sum of the n surrounding observations,

$$P_k = \sum_{i=1}^{n} w_i p_i. \tag{A1}$$

Here P_k is the interpolated value, p_i is the ith observation, assigned a weight of w_i. The set of weights assigned to the observations minimizes the mean-squared error of the interpolation value, assuming that certain data statistics are known. Because the method in theory minimizes the error it is called optimum. However, because the data statistics are only approximately known, the method is only approximately optimum. Here only nearby data are used for interpolation, to minimize the influence of uncertainties in the statistics.

Statistics needed to compute the weights include the correlations between spatial points on the analysis region, the variance across the region, and the random error or noise associated with each observation. Here the observations are satellite–gauge precipitation differences. The statistics are used to define the set of weights. In practice, each weight is roughly proportional to the correlation with the analysis point, and also roughly inversely proportional to the random error. Here the correlation is estimated as a function of distance between points in the region, and for simplicity the variance is assumed to be constant across the analysis region. The random error of each satellite–gauge difference interpolated is estimated using the number of stations in each monthly gauge 2.5° area, n. Here the noise/signal variance ratio is assumed to be proportional to $1/n$ for each 2.5° difference, with a minimum ratio of 0.1 in regions where $n \geq 10$.

Most observations have correlations less than one and all have some random error associated with them, so all sets of weights are at least slightly damped. If there are only a few observations with greatly reduced weights, then the sum of the weights may be much less than one and the interpolation greatly damped. For example, there may be several observations with large values, but if they are assigned weak weights because of low correlations or high errors the interpolation value will have a lower, damped value. Damping is greatest in situations when data are sparse and not reliable enough to produce a strong analysis. In our analysis damping tends to be by a factor of 0.5 or less when there are fewer than five observations (as indicated by Fig. 3). There is little damping when the number of observations is 10 or more. In cases with little or no damping the analysis may be referred to as a strong analysis.

APPENDIX B

Reduction of Bias in Indirect Estimates: Theory and Validation

Bias is a type of partly correlated error. The assumption behind indirect-bias estimates is that biases from different sources have different causes, and therefore they are not perfectly correlated. This was shown to be so using the satellite direct-estimate biases with respect to the gauge data. Spatial correlations were computed between direct-estimate biases of each satellite and every other satellite. Those direct-estimate bias spatial correlations are generally between 0.4 and 0.8, with an average of 0.6. The direct-estimate biases also tell us the approximate magnitude of the biases from the different sources. Table 1 shows that, compared to gauges, most have a monthly RMS bias between 1.2 and 2.0 mm

day^{-1}, with an average of 1.5 mm day^{-1}. The lower value for OPI is to be expected since OPI is tuned to the gauges, and this is likely an underestimate of global OPI RMS bias, including regions without gauges. In the same table, the monthly RMS difference between each satellite and every other satellite yields similar values, with an average of 1.9 mm day^{-1}. Combined with the RMS bias estimate, this intersatellite bias suggests that the satellites used here have similar monthly RMS bias values of around 2 mm day^{-1}.

Because the correlation between biases is less than one, combining the biases can reduce the RMS bias of that ensemble. If the correlation were one then there would be only one degree of freedom between the estimates and the ensemble bias would simply be the average RMS bias without any reduction. With a correlation less than one the ensemble RMS bias is reduced:

$$E^2 = \frac{\langle E_k^2 \rangle}{n'}. \tag{B1}$$

Here $\langle E_k^2 \rangle$ is the average of the individual bias error variance estimates and E^2 is the reduced error variance, and n' is the number of independent degrees of freedom. For completely uncorrelated errors $n' = n$, the number of individual observations. Because of this reduction, random errors tend to be small whenever a large number of observations are combined, which is typical for monthly satellite observations. Bias errors are unlikely to be completely uncorrelated, and in general we can expect to find $n' < n$. The larger n', the more the reduction of the ensemble bias uncertainty. As long as $n' > 1$, the ensemble error variance will be less than the average of the individual estimates.

The equation for computing the variance of the sum of correlated random variables can be found in many books on statistics (e.g., Larson 1982). That equation can be simplified by setting all weights to $1/n$, assuming that all individual variances are constant, and that all correlations between random variables are constant. These simplifications make it simple to estimate the number of degrees of freedom, and to show its dependence on correlation, r. Making these simplifications allows us to write

$$\frac{1}{n'} = \left[\frac{1}{n} + \frac{n-1}{n} r\right]. \tag{B2}$$

As noted above, the bias spatial correlation is approximately $r = 0.6$ for the monthly values. Thus, there are approximately 1.5 degrees of freedom. For the temporal LTM estimates, biases from many months are averaged, so there may be many more degrees of freedom further reducing the bias in the ensemble estimate LTM.

TABLE B1. Spatial (40°S–40°N) averages of RMS bias relative to three standards: gauge data, TMI, and PR data. Bias is computed over 1998–2003, in mm day^{-1}. The average bias of the satellites used in this study (Avg sat) and the bias of the satellite ensemble (Ens sat) is given. The left columns give average of monthly values, and the right columns give bias for time-averaged (LTM) values.

	Monthly validation		LTM (1998–2003) validation	
	Avg sat	Ens sat	Avg sat	Ens sat
Gauge	2.54	2.23	1.41	1.27
TMI	1.46	1.08	0.68	0.41
PR	1.86	1.58	0.93	0.73

To show how this works in practice, validation is shown against three standards: the gauge data (over land only), the TRMM Microwave Imager (TMI; over oceans only) precipitation estimate, and the TRMM Precipitation Radar (PR; over land and oceans). Since the TRMM estimates begin in 1998, validation is given for the 1998–2003 period. All validations are over the area 40°S–40°N. Gauges give a direct precipitation measurement, and therefore can be expected to have lower biases than remote estimates from satellites, as discussed in the text. In addition, measurements have shown that the TRMM estimates have relatively low bias compared to the other satellite estimates used in this study (Bowman et al. 2003). Although these standards cover different regions, and none of them are completely free of bias, the combined results help to validate the indirect method. Encouraging is the consistency with the expected error reduction based on Eqs. (B1) and (B2). Spatial averaged values are given in Table B1.

The table shows that both forming ensembles and time averaging reduces the bias, to the point where the LTM ensemble RMS bias is about a third to a half of the average monthly bias. Comparison of the monthly average and ensemble RMS bias values indicates between 1.3 and 1.8 degrees of freedom, consistent with the estimate of 1.5 based on spatial correlations and discussed above.

Comparison of the ensemble estimates between the monthly and LTM estimates shows that there are between 3 and 7 temporal degrees of freedom in this averaging period, which reduce the LTM RMS bias compared to the monthly values. As an additional check, the direct-estimate RMS bias in Table 1 was recomputed. For Table 1 the mean-squared bias was computed for each month and those values were then av-

eraged in time. For this check the data were first averaged in time, and the mean-squared bias of those averages was computed. Comparison of the direct-estimate errors of the mean to the mean of the errors also indicates approximately 3 to 7 degrees of freedom in the time period, consistent with the temporal degrees of freedom in the indirect estimates.

These comparisons show that forming ensembles or weighted means of the different satellite estimates reduces the bias on all time scales. Thus, multisatellite products will tend to have lower biases than the average bias of the individual products. However, there is greater bias reduction in LTM estimates. Both the temporal degrees of freedom and the satellite-to-satellite degrees of freedom contribute to reduction of the LTM ensemble bias.

REFERENCES

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.,* **4,** 1147–1167.

Bogdanova, E. G., B. M. Ilyin, and I. V. Dragomilova, 2002: Application of a comprehensive bias-correction model to precipitation measured at Russian North Pole drifting stations. *J. Hydrometeor.,* **3,** 700–713.

Bowman, K. P., A. B. Phillips, and G. R. North, 2003: Comparison of TRMM rainfall retrievals with rain gauge data from the TAO/TRITON buoy array. *Geophys. Res. Lett.,* **30,** 1757, doi:10.1029/2003GL017552.

Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin, 2002: Global land precipitation: A 50-yr monthly analysis based on gauge observations. *J. Hydrometeor.,* **3,** 249–266.

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields* (in Russian). Gidrometeoizdat, 238 pp.

Groisman, P. Ya., V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl, 1991: Overcoming biases of precipitation measurement: A history of the USSR experience. *Bull. Amer. Meteor. Soc.,* **72,** 1725–1733.

Gruber, A., X. Su, M. Kanamitsu, and J. Schemm, 2000: The comparison of two rain gauge-satellite precipitation datasets. *Bull. Amer. Meteor. Soc.,* **81,** 2631–2644.

Huffman, G. J., and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined dataset. *Bull. Amer. Meteor. Soc.,* **78,** 5–20.

Janowiak, J. E., and P. Xie, 1999: CAMS_OPI: A global satellite–rain gauge merged product for real-time precipitation monitoring applications. *J. Climate,* **12,** 3335–3342.

Kiehl, J. T., and K. E. Trenberth, 1997: Earth's annual global mean energy budget. *Bull. Amer. Meteor. Soc.,* **78,** 197–208.

Larson, H. J., 1982: *Introduction to Probability Theory and Statistical Inference*. John Wiley & Sons, 637 pp.

McCollum, J. R., A. Gruber, and M. B. Ba, 2000: Discrepancy between gauges and satellite estimates of rainfall in equatorial Africa. *J. Appl. Meteor.,* **39,** 666–679.

——, W. F. Krajewski, R. R. Ferraro, and M. B. Ba, 2002: Evaluation of biases of satellite estimation algorithms over the continental United States. *J. Appl. Meteor.,* **41,** 1065–1080.

Morrissey, M. L., 1991: Using sparse raingages to test satellite-based rainfall algorithms. *J. Geophys. Res.,* **96,** 18 561–18 571.

Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.,* **117,** 572–581.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea-surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

——, C. L. Gentemann, and F. Wentz, 2004: Impact of TRMM SSTs on a climate-scale SST analysis. *J. Climate,* **17,** 2938–2952.

Rosenfeld, D., and Y. Mintz, 1988: Evaporation of rain falling from convective clouds as derived from radar measurements. *J. Appl. Meteor.,* **27,** 209–215.

Rudolf, B., H. Hauschild, W. Rueth, and U. Schneider, 1994: Terrestrial precipitation analysis: Operational method and required density of point measurements. *Global Publications and Climate Change,* M. Desbois and F. Desalmand, Eds., NATO ASI Series, Vol. I26, Springer-Verlag, 173–186.

Scofield, R. A., 1987: The NESDIS operational convective precipitation estimation technique. *Mon. Wea. Rev.,* **115,** 1773–1792.

Xie, P., and P. A. Arkin, 1996: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate,* **9,** 840–858.

——, and ——, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.,* **78,** 2539–2558.

——, and ——, 1998: Global monthly precipitation estimates from satellite-observed outgoing longwave radiation. *J. Climate,* **11,** 137–164.