ESTIMATING BIAS OF SATELLITE-BASED PRECIPITATION ESTIMATES RELATIVE TO IN SITU MEASUREMENTS

Thomas Smith[1], Phillip A. Arkin[2], George J. Huffman[3], and John J. Bates[1]

[1]NOAA/National Climatic Data Center/NESDIS
[2]ESSIC, Univ. of Maryland, College Park
[3]Science Systems and Applications, Inc. and NASA Goddard Space Flight Center

ABSTRACT

Biases in satellite precipitation estimates and the uncertainty they cause are evaluated, as the first step towards producing an improved merged precipitation analysis with uncertainty estimates. Since biases are not generally random, they may not be reduced by averaging, and it is particularly important to identify and minimize biases in data to be used for climate studies. Here we show how *in situ* data may be used to bias adjust monthly satellite-based precipitation estimates. Uncertainty in the bias adjustment is also estimated. We also discuss possible adjustments and their uncertainty in regions where no *in situ* data are available, by adjusting to the most reliable available satellite estimate. The large-scale biases discussed here can be used to perform a large-scale adjustment to satellite-based precipitation estimates. Because the adjustments have large scales, those satellite estimates will retain their fine-scale features.

1. BIAS EVALUATION

Satellite-precipitation bias is here defined as the bias relative to gauge data, evaluated in the neighborhood of gauge data. Data used are monthly averages of satellite and gauge estimates on a 2.5^o latitude-longitude grid. Eight satellite estimates are evaluated: OLR precipitation index (OPI), GOES precipitation index (GPI), adjusted GPI (AGPI), SSMI composite (SSMIc), SSMI emission (SSMIe), SSMI scattering (SSMIs), TIROS Operational Vertical Sounder (TOVS), and the SSMI/TOVS estimates. The OPI data are taken from the CPC web site[1] (Janowiak and Xie 1999). The remanding data, including the gauge data, are from the GPCP data archive[2] (GPCP version 2, Huffman *et al.* 1997). The GPCP gauge data at this site are analyzed, and some interpolation is used to fill regions without sampling. We use the number of gages in each monthly 2.5^o square to exclude gauge data from squares with no gauges. For comparison we also performed the analysis using the gauge data of Chen *et al.* (2002), similarly masked out where there are no gauges, and found similar overall results.

The satellite-gauge biases are analyzed to each 2.5^o region using optimum interpolation (OI), to produce a smooth large-scale bias estimate. Data from 12.5^o latitude-longitude regions are used for each OI bias analysis. The OI damps the bias to zero in regions where there are few or no

---

[1] http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cams_opi.html
[2] http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html

data, such as over polar latitudes and over most oceanic areas. In regions where raw biases are defined, the OI analyzes them to the center of the region by computing a weight for each value within the region. The OI weights minimize the error of the analysis, assuming that statistics of the raw biases are well known.

The OI may be used for these large-scale estimates because the correlation scales of monthly biases are generally larger than 1000 km. We computed these scales for the bias of each satellite and found similar results: zonal scales range from approximately 1000 km at high latitudes to 2000-3000 km at low latitudes, with low values associated with the ITCZ; meridional scales are approximately 1000 km or slightly larger at nearly all latitudes. Since bias scales could only be computed for land regions, we also computed the scales of satellite-to-satellite differences, which include both land and ocean differences. Those correlation scales are very similar to the bias scales in both zonal and meridional directions. However, to ensure that we do not over-interpolate biases we here set the correlation scale to a constant 750 km in all directions. This reduced scale is large enough to smooth data in a well-sampled region, but small enough to avoid producing strong biases when there is only sampling along one edge of a region.

Analyses are performed month-by-month with no temporal smoothing. As a test analyses were also performed using data from a smaller (7.5º latitude-longitude) spatial region and also from three months centered on the analysis month. There is little difference between these test analyses and the standard analysis, indicating that the analysis is not sensitive to such small changes in data.

Examples of the satellite biases are shown in Fig. 1, for the OPI, SSMIc, 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 mm/day, and their magnitudes are generally about half or less than the typical rainfall in regions where they occur. The OPI estimate is heavily weighted towards climatology, and thus it has the lowest bias. However, the OPI variance is also damped compared to the other products, so it should not be used alone in a precipitation analysis.

Both the SSMIc and GPI have annual cycles in their bias. The SSMIc bias is largest in the Northern Hemisphere in winter, when the precipitation estimate is too low compared to gauges. This may be due to snow and ice on the surface giving false zero values using this algorithm (Huffman 2004, personal communication). The GPI bias has a Northern Hemisphere maximum in the winter, where the GPI estimate is too high. That is near the northern-most limit of GPI data. Near the same time the GPI estimate is also high near the equator. The GPI uses a simple and globally-constant algorithm to convert cloud-top temperature to the precipitation rate. The apparent cycle in Fig. 1 suggests that both in the tropics and in the extra-tropics this algorithm should be seasonally adjusted, as is done with some of the other satellite algorithms. The OPI, AGPI, SSMI/TOVS, and TOVS satellites all have no apparent cycle of bias.

Biases from different satellites are not perfectly correlated and can even be out of phase, as indicated by Fig. 1. Thus, even in regions where bias can not be removed because of too few in situ data, averaging the precipitation estimates from the different satellites may reduce the bias of their average.

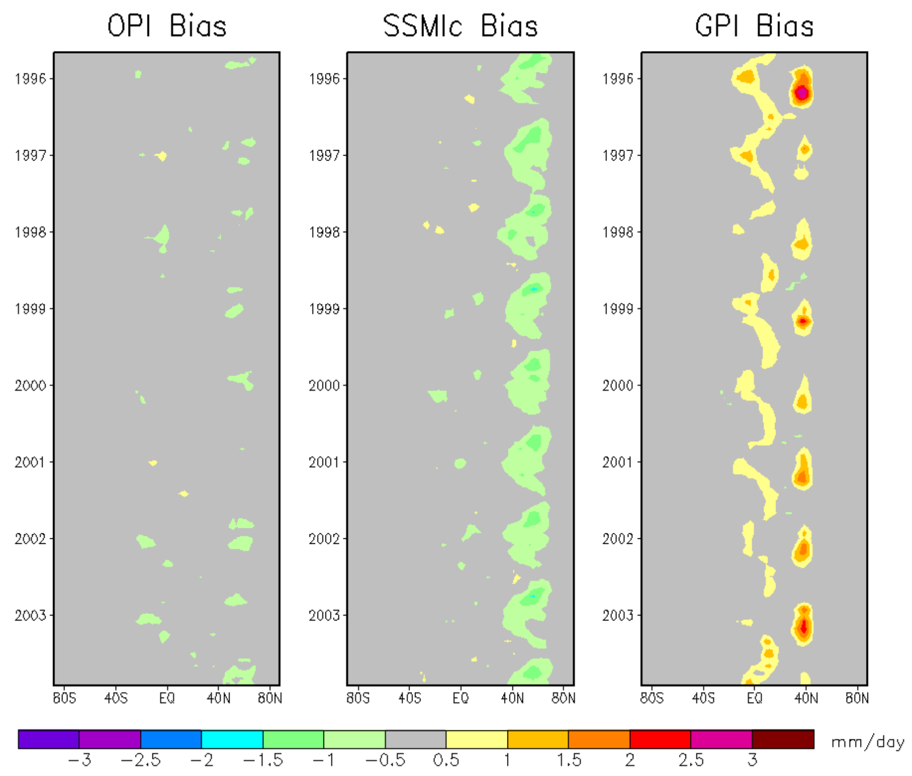

Figure 1. Analyzed zonal-average bias from the OPI, SSMIc, and GPI. Units are mm/day.

2. BIAS UNCERTAINTY

Bias in this analysis can only be evaluated in regions with gauge data. The typical sizes of biases for each satellite are used to estimate bias uncertainty in regions where gauge data are limited or absent. To evaluate bias magnitudes, we use data from 1996-2003, when all satellites were operating. We compute the mean-squared bias (MSB) of the OI analyzed bias, using biases from 2.5^o squares that are analyzed using at least 20 satellite-gauge differences. With at least 20 differences, there is practically no random or sampling error in the analysis. Since we can only compute the MSB over well-sampled regions we use only its global average for our uncertainty estimates. For each satellite the MSB is given in Table 1.

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

Satellite	MSB $(mm/day)^2$	RMSB (mm/day)	$RMSD_{sat}$ (mm/day)
OPI	0.9	0.9	2.0
AGPI	2.2	1.5	1.7
GPI	3.8	2.0	2.1
SSMI	3.1	1.8	2.0
SSMI/TOVS	2.2	1.5	1.8
TOVS	1.3	1.2	2.0

Except for the GPI, all satellites have a root-mean-square bias (RMSB) of less than 2 mm/day. To test how typical these values may be globally, over land and oceans, we also compute the RMS difference between each satellite and every other satellite ($RMSD_{sat}$), over land and oceans. This value is similar to RMSB for most satellites. In addition, the $RMSD_{sat}$ computed using only data from the regions used to compute RMSB is almost identical to that computed using the full area. Thus, we believe that these RMSB values are representative of the global values.

For any sampling of satellite-gauge differences, the typical bias uncertainty can be approximated by how well that sampling reduces the MSB for the satellite. There is no analysis with fewer than two differences and the bias uncertainty is greatest in that case. With 20 or more differences the bias is well resolved and the bias uncertainty is low. In between one and 20 differences, the uncertainty is reduced as a function of *n*, the number of differences.

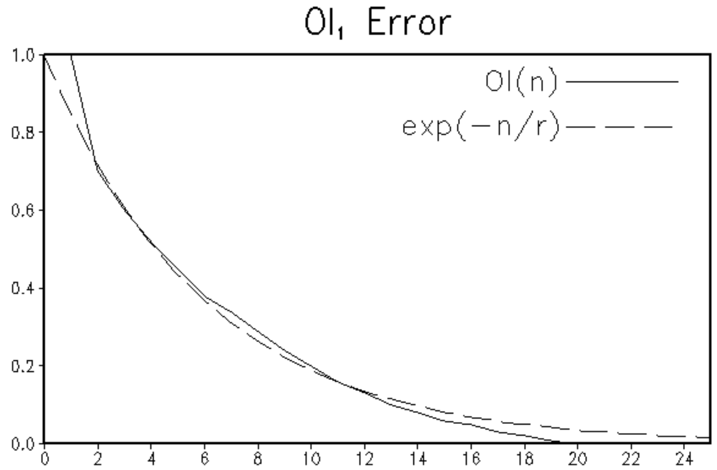

Figure 2. Error of the OI with all values set to 1 as a function of the number of values, and the exponential function approximation.

To evaluate how quickly the OI analysis reduces uncertainty with increasing *n*, we perform a second OI bias analysis as before, except that here all defined raw differences are set to one. We refer to this analysis as OI_1. For each analysis where $n > 1$, we save the analyzed value as a function of *n*. Fig. 2 shows the error averaged globally and over all years, using all satellite-sampling grids. For $n = 2$ only about a quarter of the signal is resolved, but more signal is quickly resolved so that for $n = 5$ most of the variance is resolved, and for $n = 20$ nearly all is resolved.

This relationship can be approximated by $F(n) = \exp(-n/r)$. When we fit this equation to the average data from Fig. 2, we find that the best fit to the OI_1 values is obtained with $r = 6.0$. Thus, to estimate the bias error variance for each satellite type, we use $\exp(-n/r)$ to damp the MSB in regions where $n > 1$,

$$E_{B,k}^2 = MSB_k \exp(-n/r).$$

Here $E_{B,k}^2$ is the error variance for each satellite type, *k*. Bias errors for individual satellites are combined using global estimates of the correlation between the biases.

An example of the bias standard error (the square root of the error variance), computed using all eight satellites equally weighted, is shown for January 2000 (Fig. 3). As expected, it is lowest over land and highest over the oceans. At low latitudes, where there are more satellites, the land errors are generally below 0.4 mm/day. Over the oceans away from gauges, the bias standard error is about 1 mm/day. That is smaller than typical tropical precipitation in many regions. However, it can be a problem over oceanic regions that typically have low precipitation, such as the south-east South Pacific. In those regions, there is no bias correction using this method. The bias error estimate for those regions reflects the fact that we do not know whether or not a correction is needed, and it says how large the unknown correction may potentially be.

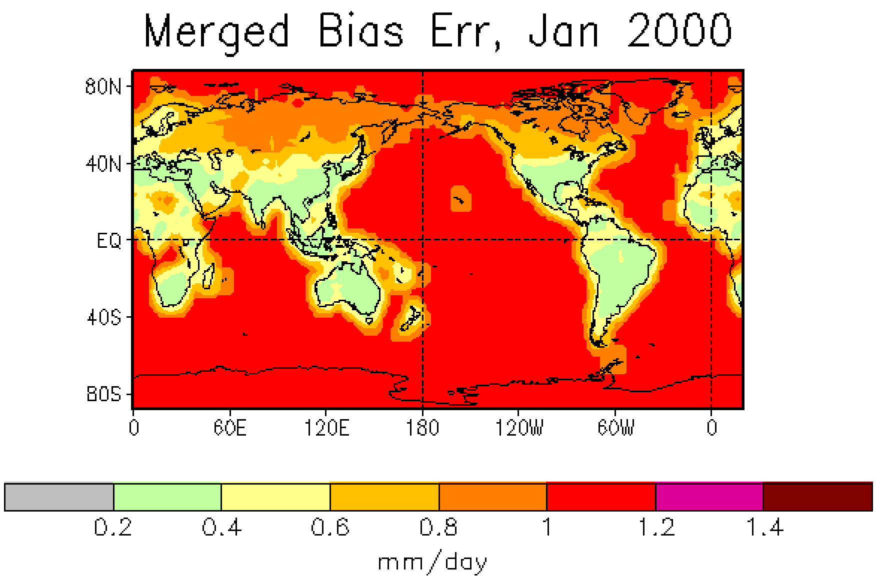

Figure 3. Standard bias error for January 2000, using all eight satellites equally weighted.

3. NEXT STEPS

The methods described above are able to adjust satellite data to minimize bias with respect to the gauge reference, and they are able to show bias uncertainty after adjustment. However, several more topics must be addressed before a global bias adjustment can be produced. One is the problem of gauge corrections at high latitudes, where snow can cause gauge biases. The other is the problem of oceanic bias adjustments.

High-latitude gauge biases are due to blowing snow around gauges and evaporation from gauges. Such biases can be either positive or negative. Although these biases may cancel, they may also be as large as 30% of the total reported precipitation (Groisman and Rankova 2001). Thus, they should be accounted for where they may occur. At present there are several sets of high-latitude gauge adjustments for certain regions. Differences between the adjustments are most critical before 1950, when data are sparser. In the satellite period differences between these adjustments may not greatly affect the overall bias estimates. However, we are currently developing methods to account for this additional uncertainty in our bias-uncertainty estimates over land. As these high-latitude gauge corrections become better understood they may be incorporated into standard precipitation gauge-based data sets. Such improved understanding could reduce bias uncertainty in regions where the corrections are needed.

Oceanic precipitation bias can not generally be adjusted using gauge data from islands and coasts because the correlation scales of the satellite-gauge differences are not large enough. However, an oceanic adjustment can be developed by adjusting to a minimum-bias satellite instead of adjusting to gauges, if a suitable minimum-bias satellite product can be identified. For the region between approximately 40S and 40N, the TRMM TMI precipitation estimate may provide a suitable base for bias adjustments. Bias estimate of TMI in the tropical Pacific indicate that its bias is very low compared to buoy observations (Bowman *et al.* 2003). If we assume that the TMI bias is approximately constant and independent of the biases for other satellites, then we can compute an oceanic bias of each satellite as the TMI bias plus the satellite bias relative to TMI. Thus, we can roughly identify the least-biased oceanic products, and use the best available one for evaluating oceanic bias. Since the best satellite will still contain some bias, this adjustment leaves some residual bias uncertainty after adjustment. However, because of dense satellite coverage there should be little sampling error in this adjustment. Similarly, over land the least-biased satellite relative to gauges can be used to reduce sampling error in the land-bias evaluation.

4. REFERENCES

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. *Geo. Res. Let.*, **30**, NO. 14, 1755, 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. Hydromet.*, **3**, 249-266.

Groisman, P. Ya. and E. Ya. Rankova, 2001: Precipitation trends over the Russian permafrost-free zone: removing the artifacts of pre-processing. *Int. J. Climatology*, **21**, 657-678.

Huffman, G. J. and co-authors, 1997: The Global Precipitation Climatology Project (GPCP) combined data set. *Bull. Amer. Meteor. Soc.*, **78**, 5-20.

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