Annual Report

Project Title: Error Estimates for AMSR-E Rainfall Data

(P.I. Thomas L. Bell, Co-I Prasun K. Kundu)

Project Objectives

The main objectives of the current project that are relevant to the AMSR-E data validation plan are:

(i) Produce estimates of the rms random error in the monthly area-averaged rain rate on a global 5° × 5° grid as measured by AMSR-E. This error arises both from the intermittent and often incomplete observation of a grid box on the earth by the satellite and from random errors in remote sensing. It is determined by the rain statistics and the total area sampled by the satellite in the course of a month.

(ii) Carry out a comparative study of the rain statistics derived from collocated satellite and ground-based radar images at Kwajalein and other ground validation sites.

The error estimates are themselves essential for utilizing the data in scientific investigations such as comparison with other data sets and numerical climate model predictions. Extracting modulating signals like the diurnal cycle and other longer-term trends also requires knowledge of the error distribution. Inter-comparison of ground radar and satellite-derived rain statistics can reveal systematic differences (to the extent that radar biases are understood). These differences are indicative of the presence of retrieval errors and can lead to improvement of the rain retrieval algorithms.

Model of Sampling Error

The validation problem involves comparing two estimates of the monthly grid-box averaged rain rate, one obtained with ground-based techniques ("ground truth") and the other obtained with satellite remote sensing. If the satellite could observe the entire grid box continuously over the month, the difference would be the so-called retrieval error. However, since the satellite estimate is actually obtained from an average over the fields of view (FOVs), or "footprints", that fall within the observed area of the box during the satellite overpasses in the course of a month, the error in the estimate provided by the satellite data is a combination of the retrieval error and a contribution from the error due to intermittent sampling.

The mean squared random error σ_E^2 depends primarily on the variability of precipitation within the box and the total area sampled by the satellite. As described in Bell et al. (2001), this quantity is given by the empirical formula

$$\sigma_E^2 = \gamma\, \sigma_A^2 / S, \qquad (1)$$

where σ_A^2 is the variance of the instantaneous rain rate averaged over the grid box of area A, S is the total area fraction sampled by the satellite and γ is a numerical factor of the order unity whose value depends on the time correlation present in the data. If γ is known, then the other two quantities on the right hand side of (1), and consequently the sampling error, can be directly calculated individually for each grid box for the month. Usefulness of the expression (1) lies in the fact that it can be used to directly generate a gridded map of the error field to accompany the corresponding map of monthly means.

In Bell et al. (2001) we analyzed data from two SSM/I satellites F10 and F11 whose local visit times were roughly 0930/2130 and 0530/1730 respectively. This allowed us to compute the lagged autocorrelation function over a range of time lags starting at 4 hours and obtain an accurate estimate of the correlation time. The value of γ computed from our theoretical model was in good agreement with the value 0.7 calculated directly from data. The AMSR-E observations are by themselves not closely enough spaced in time to yield a reliable estimate of the correlation time. It should eventually be possible to achieve this if data from other satellites such as TRMM and the current SSM/I satellites are utilized to fill in the temporal gaps.

Direct estimation of the variance σ_A^2 for a geographical grid at the $5° \times 5°$ area scale from a time series of gridded rain rates is problematic because often during an overpass a grid box is only partially swept by the swath of the satellite remote-sensing instrument. As described in Bell et al. (2001), the computation of σ_A^2 is carried out by expressing it as an area integral of the spatial covariance function of a pair of spatially separated satellite footprints that fall within the grid box: $\sigma_A^2 = s^2 \Lambda^2/A$, where s^2 and Λ respectively denote the variance and integral correlation length of rain rate averaged at the footprint scale. The quantity Λ^2 in general exhibits some dependence on the size of the grid box A, which limits the integration over the allowed range of spatial separation r of FOV-pairs. If the spatial correlation function of the FOV-averaged rain rate falls off exponentially as $e^{-r/L}$, then as the area A becomes large, Λ approaches the actual correlation length L and we expect the dependence of σ_A^2 on A to be $\sigma_A^2 \sim A^{-1}$.

Estimation of σ_A^2 using the method described above is computationally rather intensive. Moreover it turns out that, for grid boxes with low mean rain rates containing a sparse sample of rainy footprints during a month, the estimate of σ_A^2 can be unreliable because of large sampling errors. In fact, occasionally the area integral that defines Λ^2 can become negative, rendering the correlation length Λ undefined. In order to increase the range of situations for which error estimates can be provided, an alternative method of estimating the variance σ_A^2 was developed. The new method yields estimates of σ_A^2 that are much more stable and do not suffer from the above-mentioned problem.

From the precipitation data one can construct a time series $\{R_i, t_i, A_i\}$ ($i = 1, 2, \ldots n$) corresponding to the satellite overpasses of the grid box A consisting of sub-areas A_i swept out by the swath at time t_i and the instantaneous area-averaged rain rates R_i within it. An estimate of σ_A^2 is then sought in the form

$$s_A^2 = \frac{\sum_{i=1}^{n} w_i [R_i - \overline{R}]^2}{\sum_{i=1}^{n} w_i} \qquad (2)$$

where

$$\overline{R} = \frac{\sum_{i=1}^{n} A_i R_i}{\sum_{i=1}^{n} A_i} \qquad (3)$$

is the mean monthly rain rate in the box A and w_i are a set of suitably chosen weight factors for each overpass. A statistically optimal choice for the weight factor is given by

$$w_i = \sigma_A^2 / \sigma_{A_i}^2 \qquad (4)$$

where the quantity σ_{Ai}^2 in the denominator represents the true variance of the instantaneous rain rate averaged over the sub-area A_i covered by the swath. [It is to be noted that if $A_i = A$, *i.e*. the grid box is completely seen during each overpass, then the quantity s_A^2 defined in (2) becomes the usual estimator for the true variance σ_A^2.] The estimate (2) requires knowing the area dependence of the variance but is not very sensitive to exact dependence, since (*i*) for box sizes under consideration the whole grid box is generally seen (i.e. $A_i = A$ for most *i*), and (*ii*) statistical optimality implies that the estimate is insensitive to the first order difference between the supposed and the correct dependences on area. We have therefore used a power law representation of the area dependence,

$$\sigma_{A_i}^2 \propto A_i^{-\xi} \qquad (5)$$

in the formula (4) for the weights w_i to compute the estimates of variance through the expression (2). The representation is based on our experience with TRMM, SSM/I and surface radar data statistics, which suggest an exponent $1/3 < \xi < 1$. We use $\xi = 2/3$, with the intention of refining our prescription for the variance as more data are analyzed. We also exclude partial visits with $A_i < 0.1A$.

Accomplishments during the Reporting Period

While awaiting the accumulation of a substantial amount of level 2 rainfall data from AMSR-E, we have now undertaken an extensive project of studying the full multi-year global TRMM Microwave Imager (TMI) data with the operational support of the TRMM data-mining program at the Goddard DAAC. This allows us to investigate geographical and seasonal trends in the rain statistics in greater depth than has been possible until now since we are able to avoid storing and processing large volumes of archived data. The algorithm we submitted to the Goddard DAAC contained several improvements over the algorithms that we had previously used for sampling error computation. The most important refinement consisted of a more accurate estimation of the fractional area $f_i = A_i/A$ covered during an overpass. In our previous work, including Bell et al. (2001), we took f_i to be the ratio n/n_0, with n denoting the number of footprints in the grid box during the overpass, and n_0 denoting the "typical" number of footprints in a grid box completely covered by the swath. This latter number was crudely estimated from the peak value of a histogram of FOV counts in a box. This approach implicitly assumed the FOV spacing to be uniform across the swath, which is in reality not the case. The improved approach takes into account the non-uniform spacing of the footprints along a scan line by assigning an "effective area" to each footprint along the line and furthermore corrects for any potential change of the FOV-area from one scan line to another due to a change in the altitude of the satellite. The corresponding difference between the average rain rate resulting from weighted FOV estimates and the standard average with equal weighting of each FOV is of the order 5-10% where mean rain rates are large.

We have obtained estimates of σ_E from TMI version 5 data for the 6-year period 1998-2003 using both our original proposed method and the new method based on Eqs. (2-3). Figure 1 is an example of a map of the seasonal mean rain rate field (upper panel) for the northern hemisphere summer of 1999 on a 2.5° grid with the accompanying error estimates (lower panel) expressed as relative error σ_E/R, obtained with the original proposed method. Mean rain rates use weighting based on the effective areas of FOVs, as described above. Errors for rain rates $R < 0.05$ mm h^{-1} have been masked (brown) because they tend to be noisy. Boxes colored pink are ones where the integral correlation length based on the data was undefined, which tends to occur where average rain rate is small. Error estimates for the most pole-ward grid boxes are also masked because the TMI is unable to view the entire boxes at any time and calculation of Λ is difficult.

As an illustration of the integral correlation lengths Λ that enter into the estimates of σ_E, Figure 2 shows maps of mean rain rate (top panel), relative error (middle panel), and Λ (bottom panel) for Jan-Mar 1998, with rain rates $R < 0.02$ mm h^{-1} masked. Grid boxes colored pink indicate where integral correlation lengths were undefined.

In Bell et al. (2001) it was found that in the tropical Western Pacific region Λ obeys a power law relation with R, increasing slowly with increase of R. It is now seen from the global maps that the value of Λ apparently also depends on the geographical region. In the equatorial region near the ITCZ Λ is relatively small in the range of about 30-50 km even though the mean rain rates there are high. On the other hand, over the Pacific Ocean at the northernmost TRMM latitudes Λ is relatively large, of the order 60-80 km, although the rain rates are only moderate. This may be attributable to the different rain types at the two locations. While rain near the ITCZ is dominated by strong localized convective activity, the rain at northern latitudes is more stratiform in nature and is spread over larger areas.

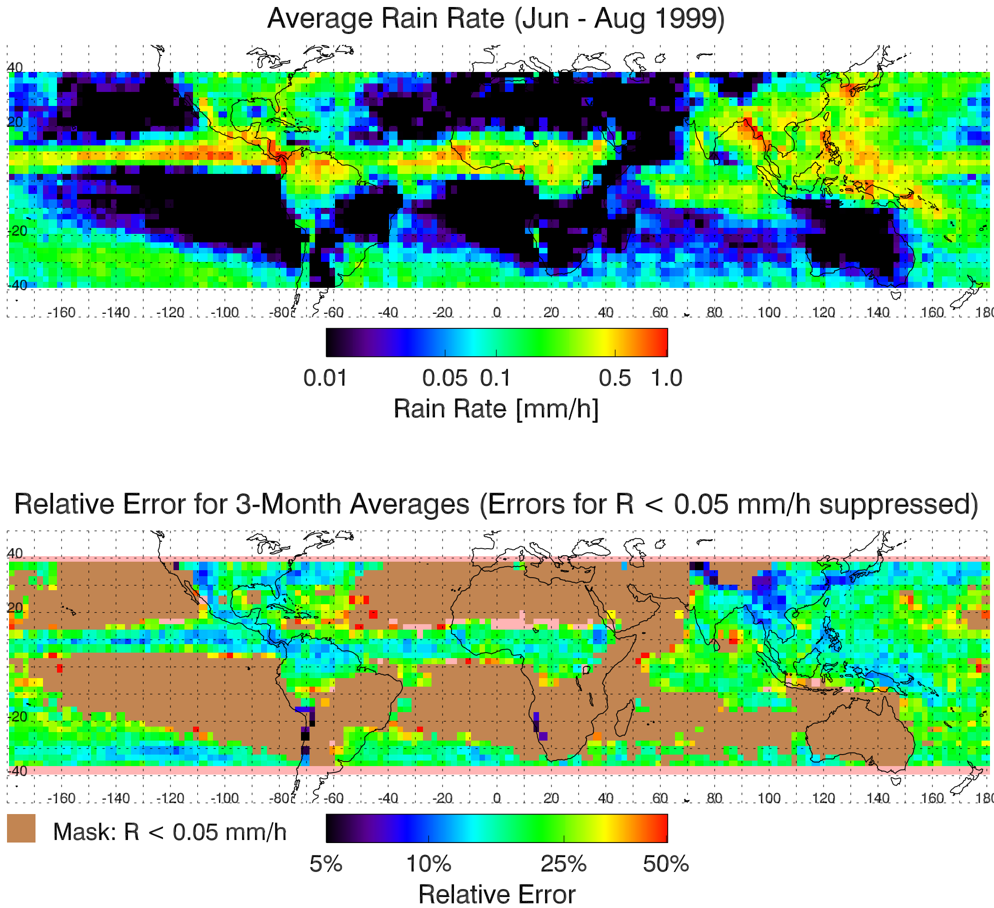


Figure 1.

We have also investigated the effect of introducing the improved method of estimation of the variance. The two panels in Figure 3 show scatter-plots for January 1999 and July 2000 TMI data of the variance estimates computed by the two methods and linear regression lines. Each point on the plot represents one grid box. As is seen from the plots, the slopes of the regression lines are close to unity and the intercepts are very small, showing that the simple power law weighting method reproduces the "true" sample variance σ_A^2 with little bias. The "noisiness" of the estimates is probably representative of the accuracy with which σ_A^2 can be estimated from a single month of data. The method needs to be explored further to settle on an accurate value of the exponent. As more data is analyzed, it is likely that the best value of the exponent may depend on rain climatology and other factors.

Future Work

As more L2B rainfall data from AMSR-E become available, we expect to be able to carry out investigations of the rain statistics and generate error estimates for the gridded global AMSR-E precipitation data set along the lines outlined above. The future goals of the project are to generate monthly error maps for each grid box on the globe seen by AMSR-E and carry out an inter-comparison study of the satellite data with collocated ground radar and/or rain gauge data to explore any systematic differences in their statistical properties, as well as to investigate how best to combine different satellite rain estimates when error estimates for each are available.

References

- Bell, T.L., P.K. Kundu and C.D. Kummerow, Sampling errors of SSM/I and TRMM rainfall averages: Comparison with error estimates from surface data and a simple model, *J. Appl. Meteor.*, **40**, 938-954, 2001.

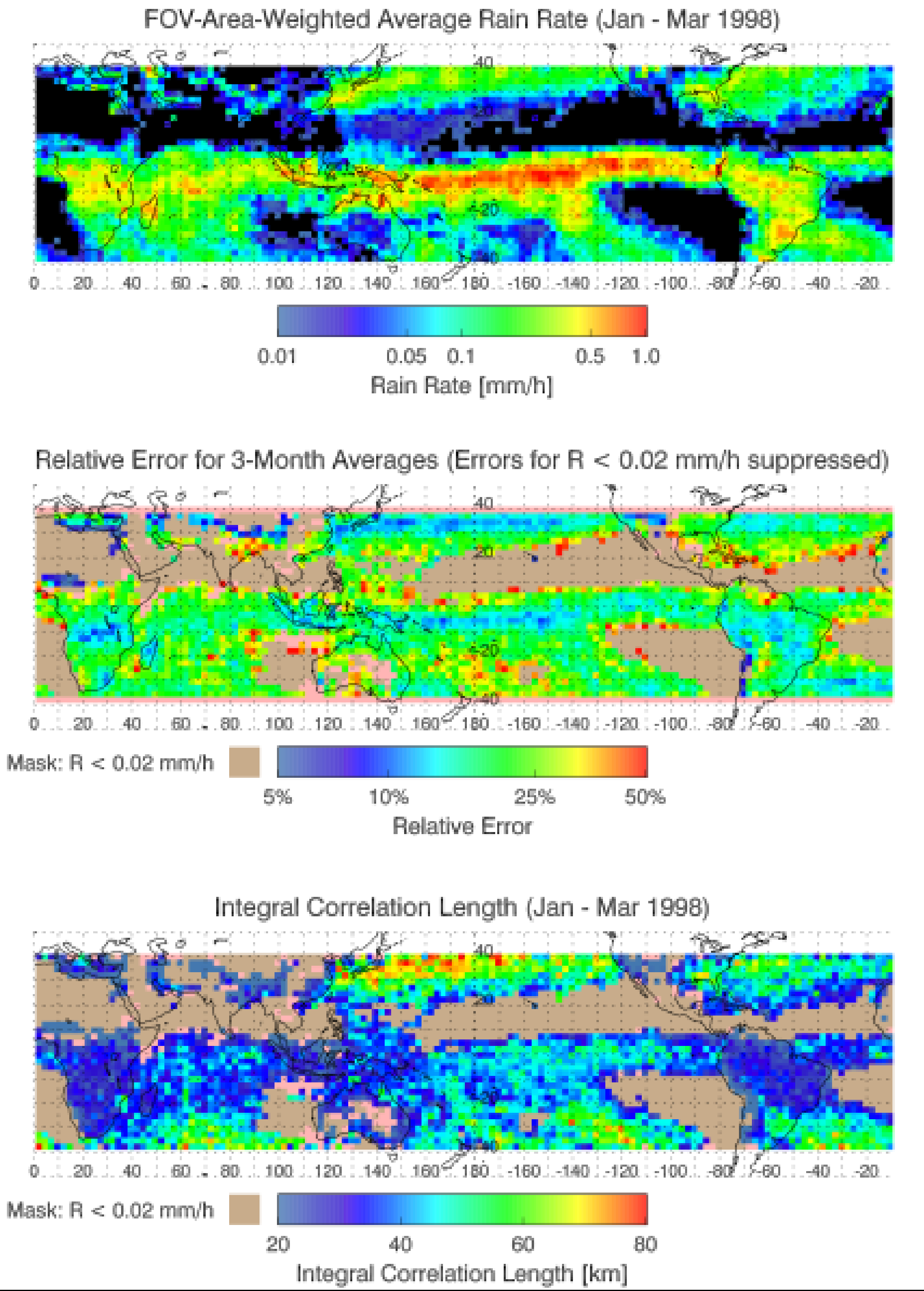
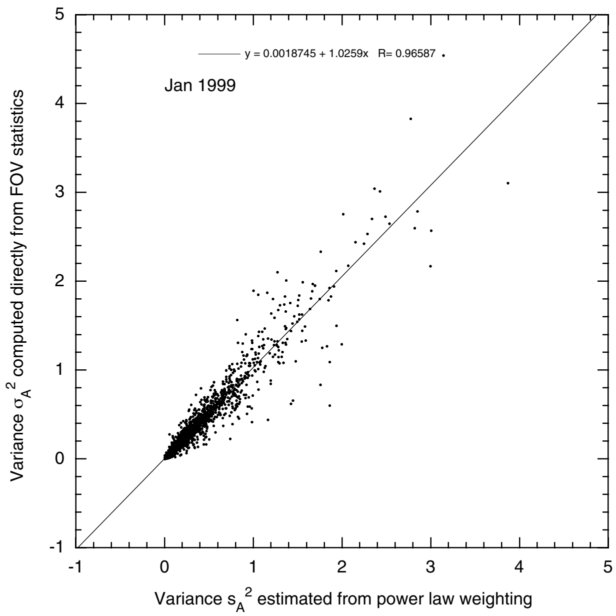

FOV-Area-Weighted Average Rain Rate (Jan - Mar 1998)
0.01
0.05
0.1
0.5
1.0
Rain Rate [mm/h]
Relative Error for 3-Month Averages (Errors for R < 0.02 mm/h suppressed)
Mask: R < 0.02 mm/h
5%
10%
25%
50%
Relative Error
Integral Correlation Length (Jan - Mar 1998)
Mask: R < 0.02 mm/h
20
40
60
80
Integral Correlation Length [km]

Figure 2.


y = 0.0018745 + 1.0259x R= 0.96587
Jan 1999
Variance σ_A^2 computed directly from FOV statistics
Variance s_A^2 estimated from power law weighting

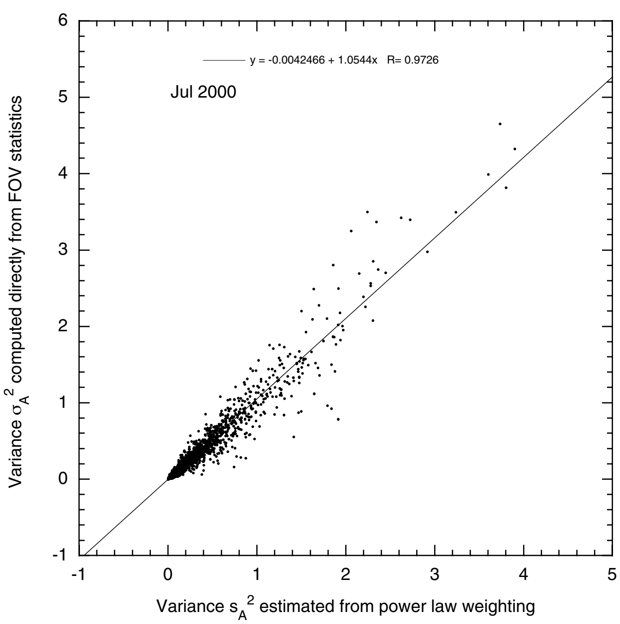

y = -0.0042466 + 1.0544x R= 0.9726
Jul 2000
Variance σ_A^2 computed directly from FOV statistics
Variance s_A^2 estimated from power law weighting

Figure 3.