Sampling Errors of SSM/I and TRMM Rainfall Averages: Comparison with Error Estimates from Surface Data and a Simple Model

THOMAS L. BELL, PRASUN K. KUNDU,* AND CHRISTIAN D. KUMMEROW[+]

Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland

ABSTRACT

Quantitative use of satellite-derived maps of monthly rainfall requires some measure of the accuracy of the satellite estimates. The rainfall estimate for a given map grid box is subject to both remote sensing error and, especially in the case of low-orbiting satellites, sampling error due to the limited number of observations of the grid box provided by the satellite. A simple model of rain behavior predicts that rms random error in grid-box averages should depend in a simple way on the local average rain rate, and the predicted behavior has been seen in simulations using surface rain gauge and radar data. This relationship is examined using Special Sensor Microwave Imager (SSM/I) satellite data obtained over the western equatorial Pacific during the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Response Experiment. Rms error inferred directly from SSM/I rainfall estimates is found to be larger than was predicted from surface data and to depend less on local rain rate than was predicted. Preliminary examination of Tropical Rainfall Measuring Mission (TRMM) microwave estimates shows better agreement with surface data. A simple method of estimating rms error in satellite rainfall estimates is suggested, based on quantities that can be computed directly from the satellite data.

1. Introduction

Satellite data are now regularly used to produce gridded maps of rainfall averaged over time intervals ranging from hours to many months. It has not been easy, however, to provide accompanying quantitative estimates of the accuracies of the grid-box averages. This is in part because remote sensing techniques do not yet provide sufficient information to allow unambiguous conversion of measurements into rain-rate values for the observed area, and the distribution of errors introduced in the conversion depends on the observed situation in ways that are not always known. The problem is exacerbated by the highly intermittent character of rain, which makes averages of rain data noisy and comparison of remote sensing results with measurements made on the ground difficult.

The Tropical Rainfall Measuring Mission (TRMM) satellite was launched in 1997. Descriptions of TRMM are given by Simpson et al. (1988, 1996) and Kummerow et al. (1998). One of the primary goals of the mission is to provide rain data that are sufficiently accurate that TRMM satellite products can serve as a kind of transfer standard to calibrate rain estimates from other satellite systems and thereby to improve the overall accuracy of global rain maps. To help to reach this goal, the satellite carries several instruments on board, including a precipitation radar and a passive microwave sensor, the latter having higher resolution than most satellite-borne microwave instruments.

An important component of the effort toward reaching this goal is developing quantitative estimates of the accuracy of the gridded products of TRMM. A number of different approaches to this are being tried, including development of models for the error intrinsic to the remote-sensing methods themselves; comparison of satellite products to ground-based measurements from rain gauge arrays, radar sites, and aircraft measurements during field campaigns; and comparison with other satellite observations.

Although much can be learned about sources of error in the TRMM rain estimates from examining individual overlapping coincident snapshots of rain events taken by various TRMM instruments and by other satellites and ground-based observation systems, much can also be learned from comparisons among *averages* of satellite data and ground-based data. As long as the averages of the satellite estimates and ground-based or other-satellite estimates are taken from time intervals

* Additional affiliation: Emergent Information Technologies, Inc., Vienna, Virginia.
[+] Current affiliation: Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado.

Corresponding author address: Dr. Thomas L. Bell, Code 913, Goddard Space Flight Center, Greenbelt, MD 20771.
E-mail: thomas.l.bell@gsfc.nasa.gov

and spatial locations that are believed to have similar statistics, such averages allow enormously more data to be used in the comparisons than can be assembled from coincident observations. Comparisons of averages of data reveal biases in rain estimates. Such biases may be small when compared with discrepancies found in point-by-point comparisons of coincident observations, yet knowledge of these biases is important when averages of TRMM data are used as a transfer standard or for climatological studies, because the contributions of the random discrepancies to averages of the data are reduced by the averaging, but the biases (by definition) are undiminished.

One of the most common methods of comparing satellite estimates of rainfall with ground-based observations and with other satellite estimates is to test the agreement of averages over a spatial domain, such as a grid box on a map, averaged over a sufficiently long time period that the averages are stable enough for the comparison to be informative. Even if the remote sensing techniques are perfectly accurate, such averages will contain sampling errors, because the systems are not measuring rainfall everywhere in the area at every moment. Rain gauges, for example, measure more or less continuously in time but cover very little of the area, whereas radar views irregularly shaped volumes of the atmosphere at frequent but noncontinuous intervals of time, and satellite observations are still more widely spaced in time. Although averages from two different systems may disagree because of inherent errors in the measurement methods, they will almost certainly disagree because they contain different sampling errors.

The bulk of this paper is devoted to developing a quantitative description of sampling-related error for satellite systems like TRMM and the Defense Meteorological Satellite Program's (DMSP) Special Sensor Microwave Imagers (SSM/I). Suppose a satellite system—TRMM, perhaps—gives an estimate $\hat{R}$ for the true average rain rate R_0 in a grid box over some time interval on the order of a month or so. The satellite system makes an (unknown) error

$$\varepsilon = \hat{R} - R_0. \qquad (1)$$

A portion of the difference ε may be due to possible algorithmic and instrumental errors in estimating rain rates when rain is observed, referred to here as retrieval error. The retrieval error may vary with location and time of observation. In addition, depending on the altitude of the satellite orbit and the instrument characteristics, a given area on the earth may be viewed by the satellite only a few times per day and even then sometimes only partially. The quantity ε may therefore include a considerable amount of sampling error due to this intermittent and occasionally incomplete coverage of the grid box by the swath of the remote sensing instrument.

Let us decompose ε in (1) into a portion due solely to the problems of remote sensing and a portion attrib-

utable to the sampling properties of the satellite system. Imagine that the grid-box area is viewed by instruments that are exactly like the ones on the satellite but are somehow stationed above the area so that they are able to provide continuous coverage of the area from the same height as the satellite. Such imaginary instruments would provide an average of the rain occurring in the area during the month as ''seen through the eyes of the satellite,'' that is, with all the errors inherent in remote sensing but without the sampling error of the orbiting satellite instruments. Let us denote by R_* this average based on hypothetical, continuous observation of the entire area with the remote sensing system. Using this hypothetical quantity, we can rewrite ε as

$$\varepsilon = \varepsilon_s + \varepsilon_*, \qquad (2)$$

with

$$\varepsilon_s = \hat{R} - R_*, \quad \text{and} \qquad (3)$$

$$\varepsilon_* = R_* - R_0. \qquad (4)$$

The first term on the right-hand side of (2), ε_s, is the sampling error in the average of the near-instantaneous estimates made when the satellite flies over the grid box, relative to the average that would have been obtained if the entire grid box had been under continuous observation by the remote sensing system. The second term, ε_*, is the area/time average of the retrieval error, under the hypothetical assumption of continuous observation by the satellite. If the remote sensing method were perfect, this term would vanish.

The error term ε_s, though referred to as ''sampling error'' here, differs from the quantity referred to by the same name in most previous publications on this topic. In most other discussions, the term sampling error referred to what the difference $\hat{R} - R_0$ would have been in the absence of retrieval error. As defined here, however, the satellite average $\hat{R}$ is assumed to include the retrieval errors inherent in the system, and, rather than being compared with the true monthly average rain rate R_0, it is compared with the hypothetical continuous time average R_* that includes the retrieval errors of the system superimposed at each instant. The reason for this change is that, as we shall see, estimates of sampling error defined this way can be made directly from the satellite data themselves with the help of some simple modeling motivated by what has been learned about rainfall statistics from ground-based data. These estimates are exactly what are needed to look for retrieval biases by comparing satellite averages with averages of data from other sources.

Before embarking on a discussion of the main topic of this paper, the statistics of ε_s, let us make a few observations about the other component of the error in (2), ε_*, defined in (4). Arguments such as those made by Wilheit (1988) and Bell et al. (1990) suggested that the random component of the retrieval error in $\hat{R}$ might be very small, because $\hat{R}$ is an average over rain-rate

estimates for all the many satellite instrument footprints or fields of view (FOVs) falling within the grid box during the month. Based on such arguments, the random component of ε_* (i.e., departures of ε_* from the mean retrieval bias $\langle \varepsilon_* \rangle$) should be even smaller, because ε_* involves an average over every instant during the entire month instead of an average over just the discrete visits occurring during the month. From that viewpoint, the error ε_* would vary little about the mean bias $\langle \varepsilon_* \rangle$, and $\varepsilon_* \approx \langle \varepsilon_* \rangle$ would change only when the rain statistics change (as might happen in going from one location to another or one season to another). If this were the case, the *random* component of the error ε in (2) would be almost entirely associated with the sampling error component ε_s dealt with in this paper, and all that would remain of the problem of completely characterizing ε would be ferreting out the average retrieval biases $\langle \varepsilon_* \rangle$. Such retrieval biases would be discovered by comparing the satellite averages $\hat{R}$ with averages obtained with other systems with known error characteristics such as rain gauges located in the area under study or a nearby ground-based radar. The accuracy with which such comparisons could be made would be dictated by the uncertainty in $\hat{R}$ due to sampling—the subject of this paper.

It should be noted, however, that the variability in the average retrieval error ε_* may not be so small as is suggested by arguments such as those made by Wilheit (1988) and Bell et al. (1990). Results presented later in this paper suggest, for instance, that retrievals might have biases that change depending on the relative amounts of stratiform and convective rain in the grid box during the month. Because these amounts can shift from month to month because of natural variability in the atmospheric system, the retrieval error ε_* might vary from month to month in ways for which a convenient description will have to be developed.

Let us turn now to the exploration of the statistics of the sampling error ε_s. The statistics of ε_s are likely to depend on many aspects of rain in a given region, such as the amount and types of rainfall, the average synoptic conditions, the season, sea surface temperatures, availability of moisture, aerosol concentrations, and so on, as well as on the sampling and observational characteristics of the satellite and its instruments. Two statistics of central interest are the mean sampling error $\langle \varepsilon_s \rangle$ and the variance of the sampling error

$$\sigma_E^2 \equiv \langle \varepsilon_s'^2 \rangle = \langle \varepsilon_s^2 \rangle - \langle \varepsilon_s \rangle^2, \qquad (5)$$

where the angular brackets indicate an average over a (hypothetical) ensemble of months similar in nature to the month of data used to obtain the estimates $\hat{R}$. (Here the variance of a variable x is defined as $\langle x'^2 \rangle$, where primes indicate deviations from the mean, that is, $x' \equiv x - \langle x \rangle$.) It is, in general, possible for there to be a systematic bias in the sample average; that is, $\langle \varepsilon_s \rangle \neq 0$. For instance, if the satellite is in a sun-synchronous orbit and therefore limited to observing the grid box at certain times of the day, the average $\hat{R}$ may be biased if there

is a diurnal cycle in the statistics of retrieved rain rates. Although it is by no means a trivial matter to correct the estimates $\hat{R}$ for biases due to temporal (or spatial) inhomogeneities in the statistics, it will be assumed that such corrections have been made or that the biases due to the pattern of observations are small when compared with the other kinds of error, so that we may write

$$\langle \varepsilon_s \rangle \approx 0. \qquad (6)$$

From (5), σ_E^2 then simply reduces under this assumption to the mean-squared sampling error $\langle \varepsilon_s^2 \rangle$.

In a previous paper, Bell and Kundu (1996), hereinafter abbreviated as BK96, derived a simple formula expressing σ_E^2 as a function of the mean rain rate and the "effective" number of samples provided by the satellite. A more general argument for the same formula was subsequently developed by Bell and Kundu (2000), hereinafter abbreviated BK00. They tested this formula using the sampling characteristics of the TRMM satellite and the statistical properties of a number of datasets from ground-based rain gauge and radar measurements. Because the statistics used in the earlier studies were derived from data taken with ground-based instruments, they did not include the effects of satellite remote-sensing errors. In this paper, we continue the investigation by comparing the formula's prediction of the behavior of σ_E^2 for monthly averages obtained from a satellite-derived dataset.

The dataset studied here consists of retrieved rain rates over the western tropical Pacific during the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Research Experiment (TOGA COARE), during November 1992–February 1993. The rain rates are derived from SSM/I data taken from two DMSP satellites that were orbiting at the time, the *F10* and *F11*. The algorithm used in the retrievals is similar to but not so highly developed as the one presently being used for TRMM. Details will be given later. It is found that a fairly simple parameterization of the random error in monthly averages over 2.5° × 2.5° grid boxes seems to describe the data well, but that the dependence on the mean rain rate in the grid box is different from what was predicted by the model and supported by simulations using ground-based data, as summarized in BK00, and the error magnitudes are much higher.

The source of this difference appears to be the very different responses of the satellite microwave instruments and algorithm to the presence of stratiform rain when compared with the ground-based measurements. This explanation will be discussed in a separate paper. Such a rain-type-dependent response has important implications for using one satellite estimate to calibrate another, as is sometimes done in combining datasets to produce global maps of rainfall, or in comparing satellite estimates with ground-validation datasets.

Despite the differences observed here in the sampling error of satellite averages when compared with that predicted from simulations using ground-based averages,

the approach described here can still be used to obtain parameterized estimates of σ_E as a function of the average rain rate in a grid box and thus can be used to supply fairly simple descriptions of the confidence levels to be applied to each grid-box value of rain rate generated from the satellite data. If we assume that the sampling error ε_s is approximately normally distributed, the "two-sigma" ($\approx 95\%$) confidence intervals for the grid-box average rain rate, setting aside the yet-to-be-determined retrieval biases ε_* in the monthly averages, would be

$$R_* = \hat{R} \pm 2\sigma_E. \tag{7}$$

Even when the assumption of normality is not completely justified, though, the above approach to inferring a bias is likely to be a good approximation to the correct one, especially as the area and/or time period averaged over becomes large. This, at least, proved to be the case in a simulation study of sampling errors in monthly averages (Bell et al. 1990).

In the following section we briefly review a model for how the mean-squared sampling error σ_E^2 should depend on rain rate and other factors and how the error estimates obtained with ground-based data compare with those from the simple model. In section 3 we describe the SSM/I-derived rain-rate data and the method of estimating sampling error for SSM/I monthly averages using the differences in the monthly averages from the *F10* and *F11* satellites. The dependence of sampling error on mean rain rate is compared there with the model predictions and with estimates made from ground-based data. It is seen that SSM/I sampling errors vary less with local rain rate than the model predicts and are significantly higher than what are estimated from the ground-based measurements.

In section 4 we pursue further the comparison of SSM/I error estimates with estimates made from surface radar taken in tropical oceanic environments. In particular, we find that SSM/I sampling error displays a simple power-law dependence on local rain rate. In section 5 we show how this power-law dependence of σ_E on the mean rain rate can be understood from the power-law dependence of a number of other statistics derived from the SSM/I data. In section 6 it is shown that sampling error can be predicted well from the temporal variability of area-averaged rain rate in a grid box according to a simple relationship suggested by the theoretical models described in section 2. This suggests an alternative, possibly more robust method of estimating sampling error. In section 7 we report some preliminary results on rainfall statistics observed by TRMM and compare and contrast them with the results from the SSM/I observations. Section 8 summarizes our results and gives some concluding remarks. Some statistical and computational details are provided in two appendices.

2. Review of a simple model for sampling error

A simple theoretical model presented in BK00 suggests how sampling error might depend on the rainfall climate statistics and satellite sampling characteristics for a given grid box. For the reader's convenience and to establish notation, we briefly review the formula and the underlying concepts and definitions. For the detailed derivations see BK00.

As was mentioned in the introduction, most prior discussions of sampling error, including that of BK00, were carried out in a framework in which the contributions of random retrieval errors to the total error were either set aside or assumed to be negligible when averaged over a month's worth of data. In particular, the model described below was developed based on assumptions about the statistical behavior of rain inferred from rain gauge and radar data. Because it is the purpose of this paper to investigate whether the model is consistent with rain statistics as inferred from satellite data, the description of the model is slightly adjusted to take into account the model's intended use here, using some of the notation introduced in the previous section.

a. Definitions and overview

We are interested in an estimate of the space–time-averaged rain rate

$$R_* = (1/T) \int_0^T dt\, R_A(t), \tag{8}$$

where $R_A(t)$ is the instantaneous rain rate averaged over a grid box with area A as it would be estimated by the satellite if it could view the area at time t, and where T is the averaging period, taken here to be one month. In this paper, grid-box sizes are generally on the order of $2.5°$ to $5°$ on a side.

The satellite, in general, views a grid box intermittently and even then sometimes only partially. Thus the instrument provides an estimate $\hat{R}_i$ of the rain rate at times $\{t_i, i = 1, \ldots, n\}$ averaged over an area $A_i \leq A$ corresponding to the region of overlap between the grid box and the instrument swath during the overpass at time t_i. The satellite estimate $\hat{R}$ of the true monthly average is obtained as a weighted average of the individual estimates $\hat{R}_i$:

$$\hat{R} = \frac{1}{n} \sum_{i=1}^{n} w_i \hat{R}_i \tag{9}$$

with suitably chosen weights w_i normalized to

$$(1/n) \sum_{i=1}^{n} w_i = 1. \tag{10}$$

A convenient way to obtain $\hat{R}$ directly from the data is to average the rain-rate estimates from all the instrument footprints that fall within the area A over the period T. (If the footprints are distributed relatively uniformly

over the areas A_i then such an average is equivalent to setting $w_i \propto A_i/A$. If the footprints are nonuniformly distributed but the area-averaged $\hat{R}_i$ has been corrected for this, the same choice for w_i is appropriate. It is shown in BK96 that, to the extent that satellite estimates of rain have statistics similar to that of ground-based estimates, this choice of weights provides a near-optimal estimate of the true monthly average for most grid boxes seen by TRMM except at the highest latitudes.)

The uncertainty in the estimate $\hat{R}$ is measured by the mean-squared error σ_E^2 defined in (5). As mentioned in the introduction, we assume that the satellite-average $\hat{R}$ has been corrected, if necessary, for possible biases due to inhomogeneities in the retrieved rain statistics, such as might occur if there is a diurnal cycle. Equation (6) follows from this assumption, which is equivalent to $\langle \hat{R} \rangle \approx \langle R_* \rangle$ [using the definition of ε_s in (3)]. We can then write

$$\sigma_E^2 = \langle (\hat{R} - R_*)^2 \rangle. \qquad (11)$$

As we have already mentioned, the sampling error can depend on the local rainfall statistics as well as sampling characteristics of the satellite. A simple model for this dependence, described in the following subsection, is based on the straightforward assumption that variations in the total rainfall amount in an area are primarily due to variations in the number of independently evolving precipitating systems present within it rather than variations in the intensity of the individual systems. Such an assumption is present in almost all statistical treatments of rainfall, and some such assumption can be used to justify rain algorithms that estimate areal rainfall from areal coverage (e.g., Short et al. 1993). The assumption is dynamically plausible, at least in the Tropics, because the convective cores of storms are quickly evolving small-scale phenomena, limited in their development by local lapse rates and the availability of moisture. Synoptic-scale lower-level convergence may affect the probability of convective plumes forming, but, once started, they are self-limiting.

Starting from this simple assumption, BK00 obtained the formula

$$\frac{\sigma_E}{R} = C(RAS)^{-1/2}, \qquad (12)$$

where

$$S = \sum_{i=1}^{n} A_i/A \qquad (13)$$

is the "effective number of full-area sweeps" of the grid box A by the satellite instrument swaths, and the prefactor C depends only weakly on a variety of rainfall characteristics consistent with a given value of the mean rain rate $R = \langle R_* \rangle$, as described below. The $S^{-1/2}$ dependence of relative sampling error predicted by (12) was seen by BK96 in simulations with tropical oceanic rain statistics. Behavior very close to this prediction was

also seen in a study of satellite estimates of rainfall by Chang and Chiu (1999). Arguments for an $R^{-1/2}$ dependence of relative sampling error on rain rate like that in (12) were given in BK96, who noted some evidence for it when estimates from simulations with radar data taken over southern coastal Japan (Oki and Sumi 1994) were plotted versus R. A similar analytical dependence of sampling error on R is discussed by Huffman (1997). Quartly et al. (1999) provide a clear review of arguments for (12) and an example of an interesting application of these ideas to a rain climatic description developed with data from the Ocean Topography Experiment/Poseidon satellite dual-frequency altimeter.

Numerous estimates of rms sampling error have been made in the literature using simulated satellite sampling of data taken by ground-based measurement systems in a variety of geographical regions. These estimates therefore omit the contributions of satellite-retrieval errors to the averages given in (8) and (9); instead, they include the errors inherent in rain gauge and radar systems. BK00 examined many of these estimates and found that the dependence of σ_E/R on R was predicted well by (12) in those regions where data were available in sufficient quantities. In particular, as mentioned above, results of simulations with radar data over southern coastal Japan by Oki and Sumi (1994) agree very well (BK96) with (12); and Steiner (1996) obtained error estimates using simulations with rain gauge and radar data from Darwin and Melbourne, Florida, and found that he could fit the dependence of error on R with an expression close to (12).

b. Model explanation

A simple model that yields the relationship (12) can now be described. A more thorough discussion is given by BK00. The model assumes that rainfall consists of individual uncorrelated rain events having, on average, area a and duration $2\tau_a$. From these assumptions they derived an expression for C in (12),

$$C = (ar_c)^{1/2}[1 - 2\tau_a/(T/S)]^{1/2}, \qquad (14)$$

where r_c is the mean nonzero rain rate (subscript c for "conditional"). The ratio T/S can be thought of as the average time interval between two consecutive full-area observations by the satellite. When the sampling is sparse, one has $T/S \gg 2\tau_a$, and in this limit $C \approx \sqrt{ar_c}$. When the effective sampling interval is comparable to τ_a, this simple cell model is no longer applicable, and one must employ a more accurate representation of the statistical properties of the local rain field, an example of which is described next.

A somewhat different explicit form of the constant C was derived by Bell et al. (1990) using an approach originally due to Laughlin (1981). Assuming that the entire area A is sampled at regular intervals $\Delta t = T/S$, they obtained the formula

$$\sigma_E^2 \approx (\sigma_A^2/S)f(\Delta t/2\tau_A), \qquad (15)$$

where

$$\sigma_A^2 = \langle R_A'^2(t) \rangle \qquad (16)$$

is the variance of $R_A(t)$, the instantaneous rain rate spatially averaged over the area A; τ_A is the corresponding correlation time [$(1/e)$-folding time of the autocorrelation of $R_A(t)$, assumed to be purely exponential]; and

$$f(\nu) = \coth\nu - 1/\nu. \qquad (17)$$

The approximation in (15) assumes $T \gg \tau_A$, which is certainly valid when T is on the order of 1 month, given that τ_A is typically 4–10 h. If rain statistics in the grid box area A can be treated as being homogeneous, Bell et al. (1990) show that the variance σ_A^2 can be expressed as

$$\sigma_A^2 = s^2 \, \Lambda^2/A, \qquad (18)$$

where

$$s^2 = \langle R_{\mathrm{FOV},i}'^2(t) \rangle, \qquad (19)$$

is the variance of $R_{\mathrm{FOV},i}(t)$, the instantaneous rain rate spatially averaged over the ith satellite footprint in the area A at time t, and where Λ^2 can be thought of as the effective area of a fluctuation in the rain-rate field that is statistically independent of other such areas within the grid box A, in analogy with the definition of an "effectively independent sample size" for a time series by Leith (1973), or for data on a sphere by Madden et al. (1993). The value of Λ^2 is given by

$$\Lambda^2 = \frac{A}{N_0^2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_0} \rho(|\mathbf{x}_i - \mathbf{x}_j|), \qquad (20)$$

where $\rho(z)$ denotes the spatial correlation between rain in two footprints separated by a distance z, N_0 is the total number of footprints in A, and the average is performed over all pairs of footprints. Expressions (18) and (20) can be derived from (16) if $R_A(t)$ is written as an average over the satellite FOVs contained within A,

$$R_A(t) = \frac{1}{N_0} \sum_{i=1}^{N_0} R_{\mathrm{FOV},i}(t), \qquad (21)$$

and this is substituted into (16). Defining $\rho(|\mathbf{x}_i - \mathbf{x}_j|) = s^{-2} \langle R_{\mathrm{FOV},i}' R_{\mathrm{FOV},j}' \rangle$ then gives the relationship (18) with the definition of Λ in (20). Note that the value of Λ may in principle vary with both FOV size, which affects $\rho(z)$, and the area A, which affects the range of separations $|\mathbf{x}_i - \mathbf{x}_j|$ encountered in the double sum.

In appendix A it is shown that

$$s^2 = ps_c^2 + p(1 - p)r_c^2, \qquad (22)$$

where r_c and s_c^2 are the mean and variance of nonzero rain rate (i.e., conditional on $R_{\mathrm{FOV}} > 0$). Because p, the rain probability, is generally small,

$$s^2 \approx p(s_c^2 + r_c^2). \qquad (23)$$

Combining (15) and (18) and using $R = pr_c$ and (23), one again obtains formula (12) for the sampling error, with the identification

$$C = \Lambda[r_c(1 + \mu_c^2)]^{1/2}[f(\Delta t/2\tau_A)]^{1/2}, \qquad (24)$$

with $\mu_c \equiv s_c/r_c$. It should be pointed out that although the quantities p, r_c, s_c, and Λ may each depend strongly on the footprint size, our simple theory leads to the expectation that expressions (14) or (24) determining the constant C are insensitive to it. Short et al. (1993) have suggested that the ratio $\mu_c = s_c/r_c$ is relatively constant over a range of footprint sizes, averaging times, types of data (rain gauge or radar) and climates. In the limit of sparse sampling, constancy of μ_c would imply

$$C \approx \mathrm{const} \times r_c^{1/2}\Lambda, \qquad (25)$$

which should be compared with (14). Note that unless A is much larger than a typical rain event, Λ^2 in (20) will depend nontrivially on A and thereby change the A dependence of σ_E in (12). In fact, when A approaches the size of a single footprint, it is easy to see from (20) that $\Lambda^2 \approx A$.

In trying to understand the dependence of sampling error σ_E on the statistical characteristics of rain, a number of parameters describing the rain have been introduced in this section, including R, σ_A, s, τ_A, Λ, p, r_c, s_c, and μ_c. The next sections will attempt to estimate σ_E for monthly averages of SSM/I rain-rate retrievals, and explore how variations in σ_E are related to changes in these parameters.

3. Random error of monthly SSM/I rain rates

Rain estimates made from SSM/I observations provide a way of testing directly the validity of the proposed simple theory of sampling error. Coverage by the SSM/I as measured by S in (13) is very close to that of the TRMM Microwave Imager passive microwave sensor (TMI) for grid boxes at low latitudes, and, if retrieval errors did not differ much between the two systems, the sampling errors should be similar in size as well. In this section we shall investigate the statistical behavior of the retrieved rain rates and the inferred statistics of random errors in gridded monthly averages of retrievals.

a. The SSM/I dataset

The dataset we used consists of rain data from two DMSP satellites, the *F10* and *F11,* in nearly sun-synchronous polar orbits around the earth. The data were taken during the four-month Special Observing Period (SOP) of TOGA COARE from November 1992 to February 1993. Local visit times of the *F10* and *F11* during the SOP were roughly 0930/2130 and 0530/1730, respectively. The SSM/I on each satellite views a given spot on the earth an average of about 30 times per month, so that $S \approx 30$ in (12). (For the TRMM microwave instrument, $S \approx 30$ as well, but local visit times shift over the course of a month.)

Rain rates were derived using the Goddard Profiling Algorithm, which is based on the method described by

Kummerow and Giglio (1994a,b), modified following the description given by Kummerow et al. (1996). The dataset was generated as part of the Third Algorithm Intercomparison Project, as described by Ebert et al. (1996), and in more detail by Ebert and Manton (1998). Rain rates are estimated for footprints, which may be thought of as circles approximately 28 km in diameter, even though in reality they are elliptical in shape, the response of the microwave antenna is nonuniform over the FOV, and there is blurring due to the finite integration time of the SSM/I instruments. Kummerow and Giglio (1994b) provide a more detailed discussion of this topic. The retrieved rain rates are provided along successive arcs, each containing 64 partially overlapping footprints and covering altogether a swath about 1400 km wide.

We study the statistics of rain in the region extending from 10°S to 10°N and from 135° to 175°E in the tropical western Pacific. This region includes the area where the TOGA COARE Intensive Flux Array (IFA) was located. For an optimal choice of the grid-box size for our statistical analysis one needs to strike a compromise among several competing factors. The box needs to be large enough so that rain rates in neighboring boxes can be assumed to be statistically uncorrelated. This is essential for treating collections of grid-box averages as sets of statistically independent samples, so that standard statistical methods of estimating confidence intervals for the averages can be used. On the other hand one would like the boxes to be small so that there are as many boxes as possible, thereby giving us a more detailed, smoother picture of the dependence of the retrieval statistics on local rain rate, as will be clear in the next section. A small box size also increases the likelihood that rain statistics within the box can be regarded as approximately homogenous. With these factors in mind we have chosen a grid-box size $A = 2.5° \times 2.5°$.

b. Estimate of the random error in grid-box averages

The SSM/I dataset itself does not provide access to the monthly average rain-rate R_* appearing in the definition of σ_E in (11). To circumvent this difficulty, we use a procedure, adapted from Chang et al. (1993) and used by Chang and Chiu (1999), to estimate the rms error σ_E for either satellite. Consider the mean-squared difference between the *F10* and *F11* estimates of a grid-box monthly average, $\langle(\hat{R}_{10} - \hat{R}_{11})^2\rangle$. We shall assume that the continuous time average R_* for the rain occurring in the grid box during the month as would be seen by hypothetical, permanently stationed *F10* and *F11* are the same. This assumption means that the same systematic errors would be made in retrieving rain rates using the *F10* or *F11* if they happened to be observing the same spot at the same time. The assumption is reasonable, because the instruments on the two satellites are of the same design, and the satellites orbit at similar altitudes. The fact that the averages of the *F10* estimates

and the *F11* estimates in the dataset, 0.271 and 0.263 mm h^{-1}, respectively, agree to within 3% supports such an assumption.

Based on this assumption, we can write

$$\langle(\hat{R}_{10} - \hat{R}_{11})^2\rangle = \langle(\varepsilon_{10} - \varepsilon_{11})^2\rangle$$
$$= \langle\varepsilon_{10}^2\rangle + \langle\varepsilon_{11}^2\rangle - 2\langle\varepsilon_{10}\varepsilon_{11}\rangle \quad (26)$$

with

$$\varepsilon_{10} \equiv \hat{R}_{10} - R_*, \quad \text{and} \quad (27a)$$

$$\varepsilon_{11} \equiv \hat{R}_{11} - R_*. \quad (27b)$$

We shall argue below that both the squared means $\langle\varepsilon_{10}\rangle^2$ and $\langle\varepsilon_{11}\rangle^2$ and the covariance $\langle\varepsilon_{10}\varepsilon_{11}\rangle$ are small in comparison with the mean-squared errors $\langle\varepsilon_{10}^2\rangle$ and $\langle\varepsilon_{11}^2\rangle$. If the numbers of visits to the grid box by the two satellites are similar, so that $\langle\varepsilon_{10}^2\rangle \approx \langle\varepsilon_{11}^2\rangle$, then, as proposed by Chang and Chiu (1999),

$$\langle(\hat{R}_{10} - \hat{R}_{11})^2\rangle \approx 2\sigma_E^2, \quad (28)$$

where σ_E^2 is the sampling error of either satellite [see definition (5)].

Because the equatorial crossing times of the *F10* or *F11* are confined to two times of day differing by about 12 h, and each satellite averages about 15 morning and 15 evening observations, the means $\langle\varepsilon_{10}\rangle$ and $\langle\varepsilon_{11}\rangle$ might differ significantly from zero if there is a diurnal variation in the mean rain rate. (The diurnal variation would have to be more complex than a simple first-harmonic sinusoid to contribute in this way, because of the 12-h difference in the two viewing times for either satellite.) There is considerable evidence for a diurnal cycle in rain statistics over the western tropical Pacific. Hendon and Woodberry (1983), for example, map the amplitudes of the diurnal cycle based on an index for deep convection obtained from satellite-measured infrared brightness temperatures. The amplitudes tend, however, to be relatively weak, except over land. Short et al. (1997) find a diurnal variation in the rainfall observed with radar in the TOGA COARE IFA, with an amplitude that is about 25% of the mean rain rate. Values of $\langle\varepsilon_{10}\rangle$ and $\langle\varepsilon_{11}\rangle$ generated by diurnal variability at these levels are unlikely to contribute significantly to the mean-squared errors $\langle\varepsilon_{10}^2\rangle$ and $\langle\varepsilon_{11}^2\rangle$, except possibly over some land areas. Because the portion of the area with land included in the dataset is less than 25% of the total area, it is unlikely that diurnal cycles in the statistics will affect estimates of σ_E^2 made using (28). Moreover, the close agreement of the means $\langle\hat{R}_{10}\rangle$ and $\langle\hat{R}_{11}\rangle$ already discussed and the lack of significant lagged correlations in rain rate at lags near 24 h, discussed later, also argue for neglecting contributions to (28) from diurnal-cycle effects.

The neglect, in (28), of the contribution of the covariance term $\langle\varepsilon_{10}\varepsilon_{11}\rangle$ in (26) would be justified if the observations by the two satellites were far enough apart in time to be nearly uncorrelated. Although the legitimacy of this assumption may be suspect, given that the

satellites can in principle view the same scene only 4–5 h apart, several factors may justify the approximation. Each satellite visits a grid box only once per day on average, and the visits of one satellite are generally well separated from the other's. Moreover, some simple calculations based on Laughlin's (1981) approach show that, for two satellites with idealized sampling like that of the *F10* and *F11,* expression (28) is very accurate, even though the two averages $\hat{R}_{10}$ and $\hat{R}_{11}$ are not, in fact, statistically independent. It should be noted, however, that the same calculation indicates that the approximation (28) is not so good if the satellites were to have closer sampling times or, more surprisingly, if one satellite's visit times were exactly midway between the other's. Last, this approximation was corroborated by performing sampling error calculations using the method developed in BK96 and the exact sampling patterns of the *F10* and *F11* satellites, and the approximation (28) is borne out at the level of 5% accuracy.

c. Statistical analysis of the data

Monthly averages of retrieved rain were obtained for each $2.5° \times 2.5°$ grid box in the TOGA COARE SOP dataset described above, yielding a total of 512 samples (128 grid boxes, 4 months of data). Grid-box results were also segregated according to whether the grid boxes contain mostly land, mostly ocean, or a mixture, but the differences in the statistics for these subsets were, for the most part, difficult to discern. They will be discussed later.

The coverage provided by the two satellites can vary from grid box to grid box and month to month. To gauge this, let us define S_{10} and S_{11} as the effective numbers of full viewings of a grid box by the *F10* and *F11,* respectively, as measured by (13). To compute S_{10} and S_{11}, a method is needed for estimating the areal fraction A_i/A for each satellite visit i.

1) ESTIMATION OF S_{10} AND S_{11}

If the number of footprints required to cover the entire area A is known, the ratio of the actual number of footprints in A to the full-coverage number provides an estimate of the fraction A_i/A for that particular visit. A possible method of determining the full-coverage footprint number is to examine the distribution of the number of footprints observed in many overflights of a grid box. Because the SSM/I swath is wide in comparison with the grid-box size, we would expect a histogram of the number of footprints observed in a box to peak at the maximum possible number. In reality, such histograms are not so simply behaved. This is, in part, because the density of footprints varies with location in the instrument swath, being largest near the swath's edges. Sporadic data loss due to instrumental and algorithmic problems can also occur. As a result, the histogram of footprint counts displays a somewhat broadened peak

at the largest footprint counts. Although a more exact method of determining the fractions A_i/A could certainly be devised, it is sufficient for our purposes to define the full-coverage footprint count as the number of counts N_{max} where the histogram peaks. We estimate the fraction A_i/A for a given visit to a grid box to be the ratio of the actual footprint count to N_{max}. This estimate can sometimes be greater or less than 1 even though the swath completely covers the grid box, but the monthly sums S_{10} and S_{11} that result from this choice are reasonably good approximations to the values that would be obtained from more geometric estimates, and in addition take account of occasional data dropouts. For the SSM/I dataset, we found $N_{max} \approx 120$. Values of S_{10} and S_{11} computed this way for the 512 cases ranged between 15 and 34, with a mean value of about 28, indicating considerable variations in the satellite sampling. (Note that variation in the number of days available in each month is also a contributing factor.)

2) REMOVING EFFECTS OF VARIABLE COVERAGE

Because our chief concern here is with how well (12) predicts the dependence of σ_E on local rain rate, it would be preferable if we could minimize the effects on our analysis of the varying coverage by the satellites. Arguments very similar to those used in deriving (12) predict

$$\langle (\hat{R}_{10} - \hat{R}_{11})^2 \rangle = C^2 \frac{R}{A} \left(\frac{1}{S_{10}} + \frac{1}{S_{11}} \right), \qquad (29)$$

where S_{10} and S_{11} are the effective numbers of full viewings of a grid box by the *F10* and *F11,* respectively, as measured by (13). By defining a "mean" coverage S for the two satellites by

$$2/S = 1/S_{10} + 1/S_{11}, \qquad (30)$$

we can recast (29) in a form identical to (12) even if the relative coverage by the two satellites varies. As in (12), the coefficient C in (29) may depend on local rain statistics in ways suggested by (14) or (24), but it should be relatively insensitive to changes in coverages S_{10} or S_{11}. (It bears repeating, however, that the rain statistics determining C are those of the "measured" rain, including the effects of randomly varying retrieval error.)

Consider the result of multiplying (29) by $S/2$:

$$S \langle (\hat{R}_{10} - \hat{R}_{11})^2 \rangle / 2 = C^2 R / A. \qquad (31)$$

The ensemble average $\langle \ \rangle$ here indicates an average over many different sequences of rain events all having the same monthly mean R and observed by the two satellites. Because changes in S have relatively little effect on the right-hand side of (31), the left-hand side will be insensitive to changes in S as well. This fact allows us to obtain estimates of the right-hand side of (31) from averages of data with differing values of S, so that we can write

$$C^2 R/A = \langle S(\hat{R}_{10} - \hat{R}_{11})^2 \rangle / 2, \qquad (32)$$

where now the angular brackets are meant to indicate an average over an ensemble of months with varying rain sequences with monthly average R *and* varying satellite sampling as measured by S.

3) DEPENDENCE OF RMS ERROR AND OTHER STATISTICS ON R

Guided by (32), then, we investigate the dependence of σ_E^2 on rain rate by first computing the mean rain rate

$$\hat{R} = (\hat{R}_{10} + \hat{R}_{11})/2, \qquad (33)$$

for each of the 512 grid boxes and months. The 512 pairs of estimates from the *F10* and *F11* are sorted into eight bins in order of increasing values of $\hat{R}$, with 64 samples to a bin. For each bin, an average over the 64 values of $S(\hat{R}_{10} - \hat{R}_{11})^2/2$ gives us an estimate of $C^2 R/A$, using (32), at the mean R for that bin. The various FOV-scale rain statistics introduced in section 2, namely s^2, p, s_c^2, r_c, μ_c, and Λ, are also computed for each rain-rate bin R and are discussed in a later section. The binning process destroys information regarding the geographical location of a particular box and the observation month, because samples containing similar monthly averaged rain rates are lumped together regardless of their location or time of observation. Although rain statistics no doubt change as various factors affecting the formation and development of precipitating systems within each grid box change, the operating assumption is the same as that of the simple model: if the frequency with which rain events occur in a grid box is known, all other rain statistics at that location can be predicted reasonably well.

It was mentioned in section 2 that sampling errors for monthly averaged TRMM TMI data have been estimated with simulations using ground-based measurements in a variety of rain environments. Near the equator, the TMI and a DMSP satellite carrying SSM/I provide almost identical coverage, as measured by S, if both instruments are providing rain estimates from the entire instrument swaths during the month. With perfect coverage, S is approximately 30 for both satellites. To compare our SSM/I results to these earlier TRMM studies, (12) and our estimates of $C^2 R/A$ from (32) can be used to compute what the random error σ_E in monthly averages of SSM/I data would be for the same coverage $S_0 = 30$ assumed in the TRMM studies, via

$$\sigma_E = \left[\frac{\langle S(\hat{R}_{10} - \hat{R}_{11})^2 \rangle / 2}{S_0} \right]^{1/2}. \qquad (34)$$

Figure 1 shows a plot of σ_E/R estimated for a single SSM/I providing maximum possible coverage during a month (i.e., assuming an average of 30 visits per month). Results are plotted versus the average R for each bin. Error bars are 95% confidence limits obtained under the assumption that differences in monthly means behave
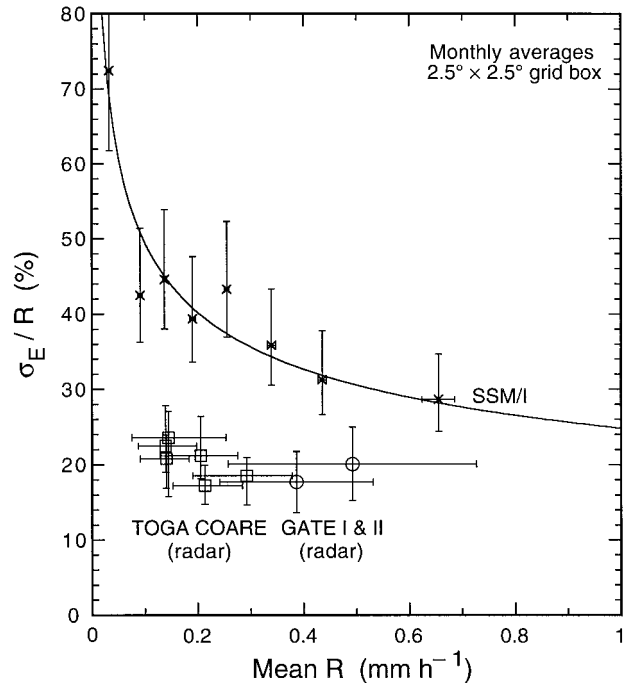

FIG. 1. Relative sampling error of monthly grid-box averages over the equatorial western Pacific as a function of mean local rain rate R. SSM/I estimates have been corrected for missing data using (34). A power-law fit, $\sigma_E/R = 0.26 R^{-0.27}$, is shown. Estimates using surface radar data assume coverage identical to what is provided by the TRMM microwave instrument, averaging 30 visits per month, very close to the SSM/I sampling. GATE radar data were taken during 1974. TOGA COARE radar data were taken contemporaneously with the SSM/I data.

statistically like independent, normally distributed variables.

Also shown in Fig. 1 are sampling-error estimates based on two radar datasets collected from ships stationed over open ocean. The two estimates labeled "GATE" use the statistics of data taken over the eastern tropical Atlantic during phases I and II of the Global Atmospheric Research Program Atlantic Tropical Experiment (GATE) in 1974. The six estimates labeled "TOGA COARE" use the statistics of radar data from two ships during the three cruises of the SOP. The methods used in obtaining these estimates are described fully in BK00. Comparison of the SSM/I estimates with the TOGA COARE estimates is particularly appropriate because the data were taken during the same four months, although the radar data cover only a limited region around 2°S, 156°E.

Figure 1 brings out two salient characteristics of the SSM/I error estimates: 1) estimated errors in SSM/I averages, which may include random retrieval errors, are 30% or more of monthly mean rain rates and are considerably larger than previous error estimates based on surface radar data (which do not, of course, include satellite remote sensing errors but do include the errors in the radar-derived rain rates) and 2) even though both

the simple model and experience (though admittedly limited) with ground-based data suggest that σ_E/R might be described by a power law with exponent $-\frac{1}{2}$, the SSM/I errors are better described by a power law with an exponent of about -0.27. As we shall see later in section 5, this departure from the simple $R^{-1/2}$ dependence in (12) can be accounted for at least in part by the R dependence of some of the other rainfall statistics that determine the prefactor C via equations such as (24).

Note that a number of sampling error estimates have been made with ground-based data other than those shown in Fig. 1. They are reviewed by BK00. Two extensive studies, by Oki and Sumi (1994) and Steiner (1996), yielded sampling-error estimates that are comparable in magnitude to the SSM/I values in Fig. 1, except at the highest rain rates, where the SSM/I estimates are larger. Because these studies used data from southern coastal Japan and from Darwin, on the northern coast of Australia, however, it is not clear that comparison with the SSM/I results is appropriate here. Rain in tropical coastal areas is very different in character from rain over the open ocean. The SSM/I statistics we used are largely determined by rain over oceanic areas. The TOGA COARE radar statistics shown in Fig. 1 are from an area and time period included in the SSM/I dataset and so would be most nearly comparable.

It is interesting to note that Chang et al. (1993) and Chang and Chiu (1999) also obtained rms error as a function of the mean rain rate on a $5° \times 5°$ grid, using global oceanic monthly estimates of rainfall obtained with a microwave emission–based algorithm. If Chang et al.'s (1993) results are converted to the format used here, they can be fitted to $\sigma_E/R \approx 0.26R^{-0.26}$ (R: mm h^{-1}). Similar results are reported in Chang and Chiu's (1999) study. The relative errors they found are roughly 50% higher than the corresponding errors for $5° \times 5°$ boxes we found (not shown) using the SSM/I dataset studied here. We conjecture that, because the grid boxes in Chang et al.'s (1993) study were all $5° \times 5°$ regardless of location, boxes at higher latitudes that contributed to their statistics had smaller physical areas, and (12) predicts that they would have higher rms errors than would boxes near the equator. Thus, the higher errors of extratropical grid boxes may have been averaged with the errors for tropical grid boxes and resulted in an overall increase in average error, whereas our analysis covers only equatorial areas.

Figure 1 has shown that, where they can be compared, the statistics of the microwave-retrieved rain rates clearly differ in important ways from the statistics of surface radar data. In the sections that follow we shall try to identify where the differences occur, propose some useful diagnostics for these differences, and suggest how (12) might be modified to take them into account.

4. Exploration of ground radar–SSM/I differences

The assumptions of the simple model in section 2 lead to predictions for sampling error such as (15),

where mean-squared error is the product of the variance of area-averaged rain rate, σ_A^2, and a factor $f(\Delta t/2\tau_A)/S$ determined by the temporal sampling pattern of the satellite and by the correlation time τ_A of area-averaged rain rate. We can rewrite it somewhat schematically as

$$\sigma_E^2 \approx \sigma_A^2[f(T/2\tau_A S)/S]. \quad (35)$$

In reality, when satellite visits are not evenly spaced and the area A is not viewed in its entirety on each visit, the dependence of f/S on a satellite's sampling pattern is more complicated than the simple dependence on S in (35) suggests. Based on an earlier study (BK96) with TRMM sampling, however, (35) apparently captures much of the change in sampling error with satellite sampling.

As we shall see later, the correlation times of SSM/I-retrieved rain rates tend to be similar in size to the correlation times seen in radar data and small when compared with the typical time interval between SSM/I visits. We therefore conclude that the factor f cannot explain the differences in sampling errors in Fig. 1. Most of the difference may be due to differences in variability of area-averaged rain rate as reported by satellite and ground-based systems, and we turn now to investigating the differences in σ_A^2 for the two.

By combining (12) and (35) it is easy to show that the simple model predicts that σ_A^2 should increase linearly with R, so that the ratio σ_A^2/R should remain constant with changing local rain rates. Figure 2 shows this quantity plotted as a function of R using the same binning procedure as in Fig. 1. To improve the legibility of the figure, only error bars (95% confidence intervals) for the ratio computed from GATE radar data are shown. They are representative of the estimated errors in the other plotted quantities. (Also shown are corresponding values obtained from TRMM TMI retrievals. These will be discussed later.) Given the level of uncertainty, it could be argued that the surface radar statistics are consistent with the constancy with R predicted by the simple model, though synoptic conditions at the two radar sites are sufficiently different that some underlying changes in the statistics may also be occurring. Whether or not this is so, it is evident from Fig. 2 that variances in SSM/I area averages are significantly larger than for the same averages obtained with surface radar, and they also probably increase faster with R than the surface data.

Equation (18) indicates that σ_A^2 is determined by the variance of the individual SSM/I "point" estimates of rain rate (i.e., s^2 for FOV estimates) and by Λ^2, the area of statistically independent rain events. Figure 3 shows the dependence of Λ on R, calculated using (20). The calculation of Λ had to be adapted to handle the actual spatial distribution of SSM/I footprints and is described in the appendix. In this figure and the plots that follow, the statistics for each value of R are averages over 64 grid-box/months with monthly means in the neighborhood of R, just as in Figs. 1 and 2. SSM/I estimates for regions with monthly rain rates similar to those observed
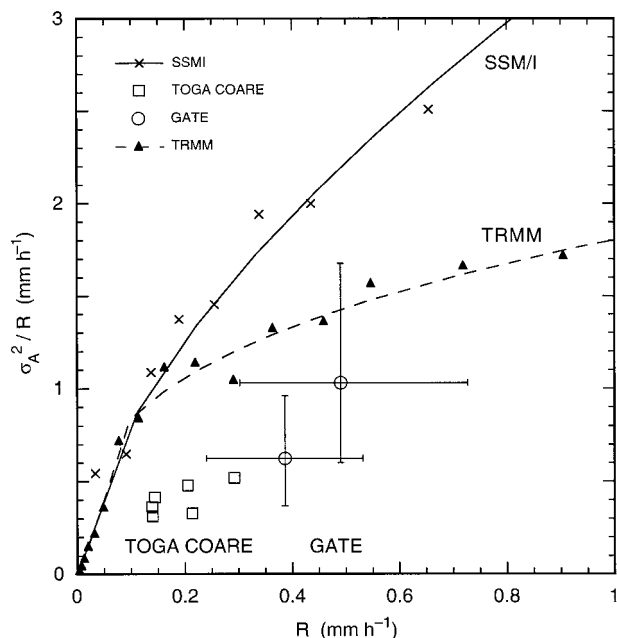
FIG. 2. Ratio of the variance of instantaneous area-averaged rain rate $R_A(t)$ to R, $A = 2.5° \times 2.5°$, computed following the procedures used for Fig. 1. The simple model predicts that this quantity should be insensitive to local rain rate. Error bars (95% confidence limits) are shown only for GATE, but others would have similar errors. A power-law fit to the SSM/I points, $\sigma_A^2/R = 3.4R^{0.62}$, is shown. Corresponding statistics derived from TRMM TMI data are also plotted and are discussed in section 7.

by the surface radar in TOGA COARE, $R \approx 0.2$ mm h^{-1}, yield values of $\Lambda \approx 100$ km (corresponding to a "correlation distance" of about 40 km—see appendix). If the TOGA COARE radar data are smoothed to a spatial resolution corresponding to the scale of the SSM/I footprint area, about $\sqrt{\pi}(28/2) \approx 25$ km, and are used to calculate Λ, a value of Λ very close to the SSM/I value is obtained. It is therefore the larger values of s^2 for the SSM/I rather than differences in Λ that are mostly responsible for the larger values of σ_A^2 seen in Fig. 2.

Equation (23) relates values of s^2 to the average areal coverage by rain p and the mean and variance of nonzero rain rates, r_c and s_c^2. Figure 4 shows the conditional mean $r_c = R/p$ and standard deviation s_c of nonzero rain seen by SSM/I, and also the ratio $\mu_c = s_c/r_c$, as a function of R. The statistics are comparable in size to those reported for GATE data by Short et al. (1993), especially μ_c. The ratio μ_c is nearly constant, a phenomenon also noted by Short et al. (1993) in other rain data. As Conner and Petty (1998) have remarked, however, there are subtle threshold-dependent effects in the conditional statistics that make intercomparison of the radar and SSM/I statistics problematic. The radar is able to detect much smaller rain rates than the SSM/I can. When values of r_c, s_c, and μ_c are calculated from surface TOGA COARE radar data smoothed to a spatial resolution corresponding to the scale of an SSM/I FOV [$\approx (25 \text{ km})^2$],
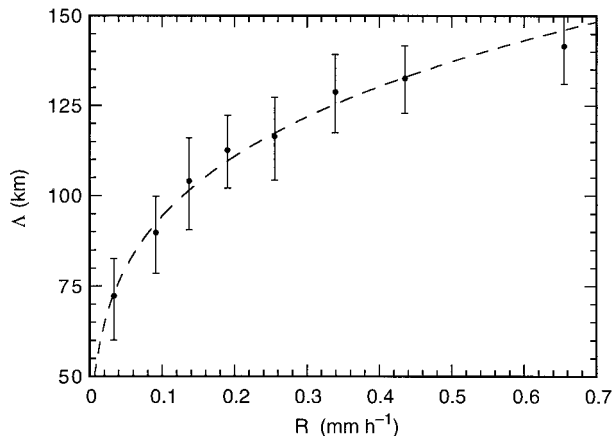

FIG. 3. The scale of "statistically independent rain events," for SSM/I data in $2.5° \times 2.5°$ grid boxes, from (20). If spatial correlations decreased exponentially as $\exp(-z/\lambda)$ with separation z and the dimensions of A are large when compared with λ, then $\lambda = \Lambda/\sqrt{2\pi}$. See appendix for details.

we find values $r_c = 0.5$ mm h^{-1}, $s_c = 1.4$ mm h^{-1}, and $\mu_c = 2.7 \pm 0.3$. They are very different from the satellite values. For example, we see in Fig. 4 that, for the SSM/I data, μ_c ranges between 1.21 and 1.44. The difference in the values of μ_c obtained by us from TOGA COARE radar data and the values obtained from SSM/I data and in the analyses of surface data by others suggests that μ_c may depend on the threshold of detectability of rain in a way that was fortuitously absent in other studies.

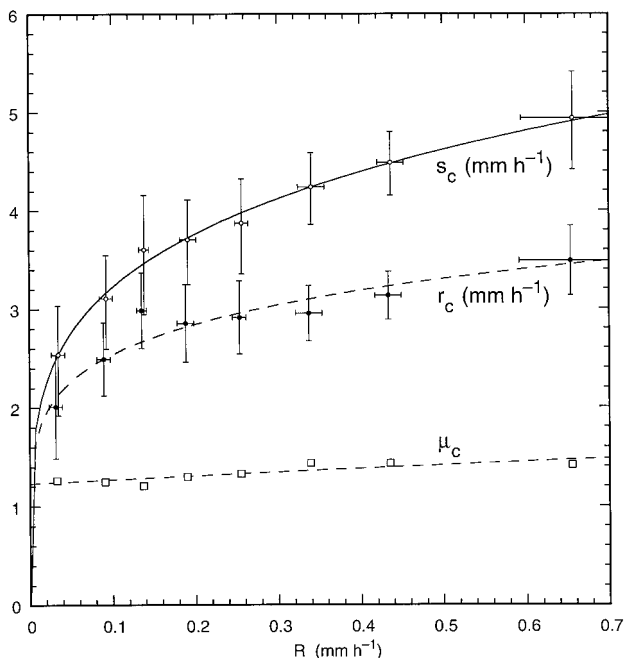To study temporal correlations of area-averaged SSM/I rain estimates, a time series of the average rain

FIG. 4. Mean r_c and standard deviation s_c of SSM/I rain rates for FOVs with nonzero rain, and the ratio $\mu_c = s_c/r_c$.
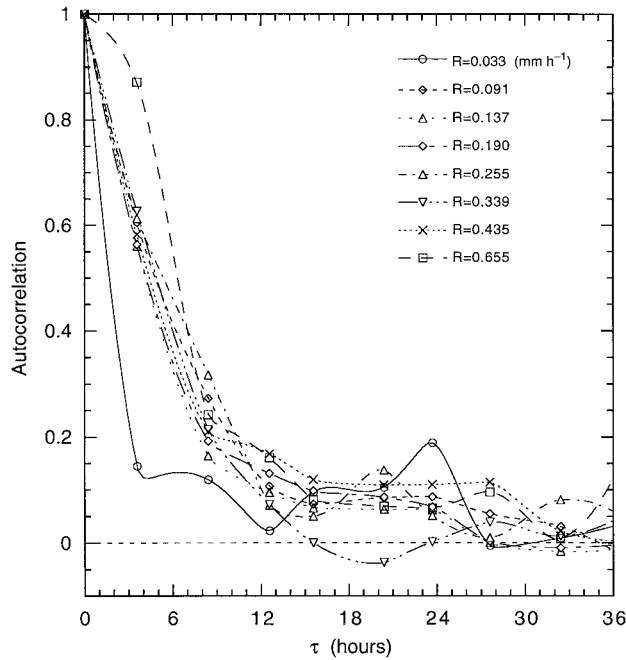
FIG. 5. Autocorrelation of SSM/I rain rate averaged over 2.5° × 2.5° grid boxes for various categories of monthly R. Correlations are shown only when more than about 400 pairs of observations are available at a given separation τ. Curves through data points are smoothed interpolations.

rate for full-area observations at each grid-box location was obtained. All visits with greater than about 85% coverage, determined from the footprint counts as explained in section 3c(1), were included to get a time series that is sufficiently dense. Because the visit times of the *F10* and *F11* sometimes differed by as little as 3h, these series had sufficient time resolution for useful time correlations to be obtained. Figure 5 shows the lagged autocorrelations of $R_A(t)$ sorted into the same eight climatological rain-rate bins used in the previous figures; that is, autocorrelations for a given R represent the statistics of 64 time series with monthly means in the neighborhood of R. For each of the eight rain-rate categories, we fitted the lagged autocorrelation function of the area-averaged rain rate to a simple exponential form $\exp(-|t - t'|/\tau_A)$. The correlation times τ_A were found to be about 6 h and nearly independent of R, except at the lowest and highest rain rates. Spectral analysis of the time series indicated enhanced spectral power at frequencies corresponding to periods of 2–5 days and 40–50 days. The former may possibly be related to the convective disturbances with that timescale discussed by Takayabu and Nitta (1993), and the latter may be related to the Madden-Julian oscillation (Madden and Julian 1972; Chen and Yanai 2000).

It is well known that the statistical behavior of rainfall differs over land and ocean. To investigate this fact quantitatively, we employed a land/ocean mask at 2.5° spatial resolution. Of the 128 grid boxes in the chosen

area, 97 are categorized as covered by ocean, 23 as mostly covered by land—largely concentrated around New Guinea in the southwest quadrant of the area we studied—and 8 as containing substantial amounts of both. The statistics of land-containing grid boxes were sorted into only four bins with increasing rain-rates R in order to have a reasonable number of samples in each bin. Monthly rain rates in the land-containing boxes tended to range over values less than half as large as for the ocean-covered boxes. Most land–ocean differences in the statistics were indistinguishable from variability caused by small-sample effects. The conditional means r_c, however, were 50% to 75% larger over land, unlike the values of s_c, which were, perhaps surprisingly, a little smaller. The ratio μ_c ranged from 1.43 to 1.56 over ocean and from 0.85 to 1.0 over land. A pronounced peak in spectral power was found in the time spectrum of rain over land-covered boxes at a frequency of 1 day^{-1}, indicating the presence of a strong diurnal cycle. No spectral peak was evident in oceanic rain rates at that frequency. There is also little sign of any enhanced autocorrelation at $\tau = 24$ h in Fig. 5, except perhaps for grid boxes with the smallest rain rates, indicating that statistics tended to be dominated by the statistics of the oceanic grid boxes.

5. Power-law descriptions of SSM/I statistics

The statistics of the SSM/I retrieved rain rates are described very well by simple power-law dependences on R, as can be seen from the power-law fits shown in Figs. 1–4. Because this approach provides a much more concise description of the statistics, we present these results here.

It is convenient to express the various statistical quantities as powers of the dimensionless quantity p rather than R. We introduce the three basic exponents α, β, and γ through the relations

$$r_c = r_0 p^\alpha, \qquad s_c^2 = s_0^2 p^\beta, \quad \text{and} \quad \Lambda^2 = \Lambda_0^2 p^\gamma. \quad (36)$$

Note that in the simple model all the exponents would vanish. From the definition $R = pr_c$ it follows that

$$R = r_0 p^{1+\alpha}, \quad (37)$$

and if we treat the ratio μ_c as approximately constant, (23) gives

$$s^2 \approx (s_0^2 + r_0^2)p^{1+\beta}. \quad (38)$$

(Strict constancy of μ_c would imply $\beta = 2\alpha$.)

The expression (18) for σ_A^2 implies the power-law relation

$$\sigma_A^2 = (s_0^2 + r_0^2)(\Lambda_0^2/A)p^{1+\beta+\gamma}. \quad (39)$$

Because p and R are related by (37), the exponents α, β, and γ can be derived from the exponents obtained with error-weighted least squares power-law fits to the statistics in Figs. 3 and 4. We find that the SSM/I statistics can be reasonably well explained by the values

$$\alpha = 0.17, \quad \beta = 0.53, \quad \text{and} \quad \gamma = 0.53. \quad (40)$$

The coefficients of the power-law fits in (36) were found to be

$$r_0 = 4.5 \text{ mm h}^{-1}, \quad s_0 = 7.6 \text{ mm h}^{-1}, \quad \text{and}$$

$$\Lambda_0 = 220 \text{ km}. \quad (41)$$

If the relatively small effects on sampling error σ_E due to changes in τ_A with R are neglected, (35) and (39) imply

$$\sigma_E^2 \propto p^{1+\beta+\gamma}. \quad (42)$$

The relative sampling error for a single SSM/I satellite shown in Fig. 1, when fitted to a power law in R,

$$\sigma_E/R \propto R^\delta, \quad (43)$$

gives an exponent $\delta = -0.27$ (instead of the -0.5 predicted by the simple model). The power laws (36) would predict

$$\delta = -(1/2)[1 - (\beta + \gamma - \alpha)/(1 + \alpha)], \quad (44)$$

or $\delta = -0.12$ when the exponents in (40) are substituted. The discrepancy in the exponent obtained by directly fitting σ_E/R to a power law and the exponent predicted using the other empirical exponents may be due in part to the changes in the correlation time of $R_A(t)$ at the smallest and largest rain rates R seen in Fig. 5 and the resulting changes in the factor f in (35). The value of δ that best fits the data is made still more negative by the deviation from the power-law fit—possibly fortuitous—of σ_A^2 for the smallest rain-rate bin (Fig. 2).

6. Alternative approach to estimating σ_E

Equation (35) suggests that mean-squared error in monthly averages can be estimated from the variance of instantaneous grid-box-averaged rain rate σ_A^2, the sample volume S, and a factor f that depends on the sampling pattern of the satellite and the time correlations of the grid-box-averaged rain rate $R_A(t)$. Because the correlation time of rain is somewhat smaller than the typical interval between satellite visits, the dependence of f on τ_A and S is relatively weak, which suggests that $\sigma_E^2 S$, estimated by the right-hand side of (32), is proportional to σ_A^2. This result apparently is borne out by Fig. 6, where $\langle S(\hat{R}_{10} - \hat{R}_{11})^2 \rangle/2$ is plotted against σ_A^2, each point representing an average over 64 grid-box/months with similar mean R as in Figs. 1–5. As (35) predicts, the dependence of σ_E on R is mostly determined by the R dependence of σ_A. Their relationship is described empirically by

$$\sigma_E^2 = (0.72 \pm 0.08)\sigma_A^2/S. \quad (45)$$

It is interesting to compare the empirical coefficient in (45) with the theoretical model estimate based on (15) and (17). If we use the correlation time $\tau_A = 6$ h found in section 4 for most R and the mean monthly
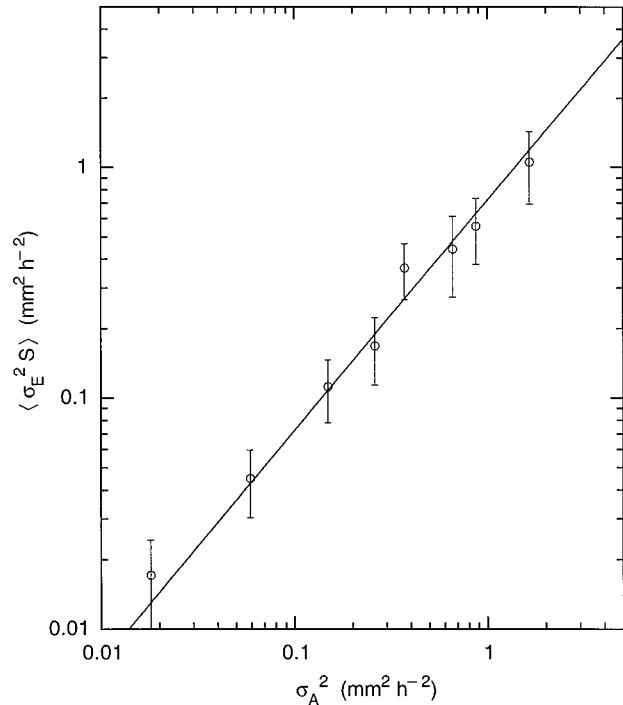

FIG. 6. Estimates of the average product of SSM/I sampling error σ_E^2 and coverage S [defined in (13)], obtained from the statistic $\langle S(\hat{R}_{10} - \hat{R}_{11})^2 \rangle/2$ as discussed in section 3, plotted against variance of instantaneous grid-box-averaged rain rate σ_A^2. Each point represents the average of 64 grid-box/months binned according to monthly rain rate, as in the previous figures. The 95% confidence limits for the slope of the linear, zero-intercept, weighted least squares fit to the points are 0.72 ± 0.08.

areal coverage $S = 28$, we calculate $f(T/2\tau_A S) = 0.56$. Given the crude nature of the estimate, which assumes exponential autocorrelation of $R_A(t)$ and equally spaced observations in time by the satellite, the extent of agreement with the observed value 0.72 is perhaps remarkable. The fact that the empirical coefficient is larger than the prediction may be an indication that the geometric estimate of S used here overestimates the actual *effective* amount of coverage by the satellite instruments.

7. Some preliminary TRMM results

The analysis so far described was motivated in part by the need to supply a measure of the random error for gridded monthly rain-rate products produced by TRMM. From a rainfall-retrieval-algorithm point of view, the TRMM's TMI has an advantage over the SSM/I because the TRMM satellite orbits closer to the earth, giving the instruments improved spatial resolution and the TMI includes a lower-frequency dual-polarization 10.7-GHz channel in addition to SSM/I's four higher-frequency channels. Although the random error in TRMM monthly rain climate data will be more thoroughly explored in a subsequent paper, it is interesting

TABLE 1. Power-law dependence of r_c, s_c, and Λ on p, defined in (36), and power-law dependence of σ_A/R on R defined in (46), for TRMM TMI statistics over the western tropical Pacific. As can be seen in Fig. 2, fits to the data must be obtained separately for small and large R.

	α	β	γ	r_0 (mm h^{-1})	s_0 (mm h^{-1})	Λ_0 (km)	w_0	δ
$R < 0.1$ mm h^{-1}	1.02	3.40	0.46	24.6	379	175	3.27	0.05
$R > 0.1$ mm h^{-1}	1.44	0.90	0.44	104	14	165	1.34	-0.34

to compare the performance of TRMM with what has been learned about SSM/I here.

a. TRMM data

We used TMI surface rainfall retrievals made available by the Goddard Space Flight Center Distributed Active Archive Center as official TRMM product 2A12, version 4, for the 4-month period of January–April 1998 over the same geographical area as the one used in the SSM/I study here. The TMI rain product has benefited not only from the instrumental advantages mentioned above, but also from the use of a version of the algorithm more advanced than the one used with the SSM/I data. The most important change in the algorithm is probably the addition of a step that adjusts for the relative amounts of convective and stratiform rain present in each FOV, as described by Hong et al. (1999).

b. Data analysis results

The dependence of the statistics of TMI rain-rate data on local rain rate R was determined in the same manner as before, by binning the statistics for each $2.5° \times 2.5°$ grid box and month according to the monthly mean R. A plot of σ_A^2/R for TRMM is shown in Fig. 2. The number of bins was increased to 16 when it became apparent that the statistics change in character above and below $R \sim 0.1$ mm h^{-1}, so that each point represents an average of 32 rather than 64 grid-box results. It is encouraging to see that the TRMM statistic has moved closer to the radar values. The improvement is especially marked at the higher rain rates, where the ratio is both more nearly constant with R and considerably lower than the SSM/I results.

Good fits of the TRMM results to power laws in R can be obtained if a fairly sharp crossover of the exponent values for rain rates above and below $R = 0.1$ mm h^{-1} is allowed. The parameters of the fits in both regimes are given in Table 1. The parameters for the conditional rain statistics for TRMM are very different from those of the SSM/I statistics given in (40) and (41). Table 1 also gives the parameters w_0 and δ of a power-law fit to the TMI data,

$$\sigma_A/R = w_0 R^\delta, \qquad (46)$$

from which estimates of TRMM rms error can be obtained using (45). Note of course that these parameters were obtained from fits to data from a very limited portion of the globe and, in particular, are more representative of oceanic data than of retrievals over land.

Further research is clearly needed to determine whether similar parameterizations of error are possible globally and seasonally.

8. Summary and conclusions

SSM/I rain-rate data taken during the TOGA COARE experiment were used to estimate the rms sampling error as defined in (5) in monthly averages over $2.5°$ grid boxes in the western tropical Pacific. The satellite algorithm that was used is a predecessor of the one currently used to process TRMM microwave data. The error estimates were made two different ways: one estimate was obtained from the rms differences of the monthly averaged rain rates given by the *F10* and *F11* satellites; a second estimate was obtained from the variance σ_A^2 of instantaneous area-averaged rain rates $R_A(t)$ and a rough estimate of the temporal correlations of $R_A(t)$. The two estimates agreed well. This result suggests that reasonable estimates of random error in gridded monthly averages might be made from σ_A^2 and an approximate characterization of the time correlations of $R_A(t)$—quantities that can be obtained from the satellite data themselves. Such estimates will include the contributions of random retrieval errors to the total error.

Over the ocean, both the magnitude of σ_A^2/R and its dependence on local rain rate R are clearly different for the SSM/I rain estimates and surface radar estimates. The higher variance of SSM/I estimates of $R_A(t)$ in comparison with radar appears to be due mostly to the larger variance of individual footprint estimates, measured by s^2, rather than greater spatial correlations of the rain data—to the extent they are measured by Λ. It will be shown in a separate paper that the SSM/I estimates are highly correlated with stratiform rain as identified in the TOGA COARE surface radar data and are not so well correlated with rain identified as convective; the SSM/I rain estimates where there is stratiform rain are much larger than the corresponding radar estimates, whereas rain estimates where the radar reports convective rain tend to be estimated as smaller by SSM/I. The net effect is to make s^2 large for SSM/I FOV estimates. These conclusions apply, of course, only to the rain data generated by the particular algorithm used to produce the dataset investigated here.

Little has been said here about how sampling error depends on the grid-box area A. As was seen in (12), the simple model would predict $\sigma_E \propto A^{-1/2}$. Equations (15) and (18), however, indicate that this result is only true if the area A is much larger than Λ^2. The $2.5° \times 2.5°$ boxes studied here are not quite large enough in

this respect. Although increasing the box size to 5° × 5° reduces the number of samples per bin when the statistics are binned by R, as was done in section 3, such an experiment shows that the power-law dependence of σ_A^2 on R is almost the same for the two box sizes but that the dependence of σ_A^2 on A is consistent with $\sigma_A \propto A^{-0.33}$ rather than with $A^{-1/2}$. Thus, increasing the box size from 2.5° to 5° does not decrease sampling error as much as the simple model would have predicted if A were larger.

Based on our results, it is recommended that future algorithm intercomparison projects include comparisons of σ_A^2/R for grid-box sizes on the order of 2.5° or larger, in addition to comparing the mean rain rates themselves. The ratio is easy to calculate and, as has been shown here, can serve to bring out some aspects of the algorithms that can be missed in point-by-point comparisons but are important for climatological use of the data. This quantity has the advantage that, other things being equal, it is not so sensitive to instrument resolution and so makes intercomparison of different measurement systems conceptually easier. The quantity σ_A^2 can reveal the presence of correlated retrieval errors in the satellite product, a possible byproduct of the reason for its being larger in the SSM/I data than in the radar data, as will be discussed in a subsequent paper.

An especially important result is that the quantity σ_A^2 can be used to estimate the accuracy of monthly averages of rain data via a relation like (45). Such an estimate avoids some of the assumptions used in parameterizing error in terms of average R, though it requires that the satellite dataset supply values of σ_A in addition to R for each grid box.

Whether because of better resolution and additional channels in the TMI or because of improvements in algorithms, the statistics of TRMM TMI (version 4) rain estimates from the western tropical Pacific may be significantly closer to oceanic surface radar statistics than are the SSM/I statistics. An improved TMI algorithm is now being used to process TMI data, and we expect even better agreement with ground-based data. This possibility will be examined in a future paper.

APPENDIX A

Relation for s, p, r_c, and s_c

A derivation of (22) is given here for completeness. Consider a set of N footprint-averaged rain-rate values

$\{r_i \mid i = 1, \ldots, N\}$ of which the subset of N_c values $\{r_{i_\alpha} \mid \alpha = 1, \ldots, N_c\}$ is nonzero. (We assume N and $N_c \gg 1$ and that the rain statistics are homogeneous.) The fraction of nonzero values is $p = N_c/N$. The average rain rate for the entire set is $R = \sum_{i=1}^{N} r_i/N$. Of course only the nonzero terms contribute to the sum. The average rain rate conditional on nonzero rain is then

$$r_c = \sum_{\alpha=1}^{N_c} r_{i_\alpha}/N_c = R/p.$$

The variance of the entire set $\{r_i\}$ is given by

$$s^2 = \sum_{i=1}^{N} (r_i - R)^2/N = \sum_{i=1}^{N} r_i^2/N - R^2.$$

The variance conditional on nonzero rain can be computed similarly for the subset:

$$s_c^2 = \sum_{\alpha=1}^{N_c} (r_{i_\alpha} - r_c)^2/N_c = \sum_{\alpha=1}^{N_c} r_{i_\alpha}^2/N_c - r_c^2 = \sum_{i=1}^{N} r_i^2/N_c - r_c^2.$$

The above two relations can be immediately rearranged to yield

$$\sum_{i=1}^{N} r_i^2 = N(s^2 + R^2) = N_c(s_c^2 + r_c^2).$$

This implies

$$s^2 + R^2 = p(s_c^2 + r_c^2),$$

from which (22) follows.

APPENDIX B

Computation of the Length Scale Λ

In this appendix we discuss in more detail the computation and interpretation of Λ^2 defined in (20). It is helpful in developing an interpretation of Λ to assume that the footprints are sufficiently densely and evenly distributed that they can be treated as if arranged in a regular rectangular array completely filling the area $A = L^2$. Each footprint occupies a box of side $d = L/N$. The number of footprints is then $N_0 = N^2$. The quantity Λ so defined in general depends both on the area size L and the footprint size d. For instance, if the area is small enough to be covered by a single footprint, then obviously $\Lambda = L$. More generally, however, Λ is closely related to the scale over which the data are spatially correlated, as we now show.

Using the identity

$$\sum_{i=1}^{N} \sum_{j=1}^{N} f(i - j) = \sum_{m=-N}^{N} (N - |m|)f(m) \quad \text{(B1)}$$

for a function $f(i)$ defined at each integer i, $|i| \leq N - 1$, we can write (20) as

$$\Lambda^2 = \frac{A}{N^4} \sum_{i,j=1}^{N} \sum_{k,l=1}^{N} \rho(|\mathbf{x}_{ij} - \mathbf{x}_{kl}|)$$

$$= \frac{A}{N^4} \sum_{m_1=-N}^{N} \sum_{m_2=-N}^{N} (N - |m_1|)(N - |m_2|)\rho(|\mathbf{m}|d).$$

(B2)

This equation formally transforms the sum over the correlation between all pairs of footprints in the $N \times N$ array into a weighted sum of the correlation between each footprint in an equally spaced $(2N + 1) \times (2N + 1)$ array and a footprint located at the center of the array. If $\rho(|\mathbf{m}|d)$ is sufficiently smooth, (B2) can be treated as a discrete numerical approximation to a continuous double integral. The approximation becomes exact in the limit $d \to 0$ ("point footprint"). Introducing the separation vector $\mathbf{s} = \mathbf{m}d$, and using the relations $A = L^2$ and $L = Nd$ we can express Λ^2 in this limit as an area integral over a $2L \times 2L$ square:

$$\Lambda^2 \approx \frac{1}{L^2} \int_{-L}^{L} ds_1 \int_{-L}^{L} ds_2 \ (L - |s_1|)(L - |s_2|)\rho(|\mathbf{s}|).$$

By going to polar coordinates this can be reduced further to the one-dimensional integral

$$\Lambda^2 = 4 \int_0^{\sqrt{2}L} sg(s)\rho(s) \ ds, \qquad (B3)$$

with the angular integral replaced by the areal weighting factor

$$g(s) = \int_{\varphi_0(s)}^{\varphi_1(s)} d\varphi \left(1 - \frac{s}{L} \cos\varphi \right)\left(1 - \frac{s}{L} \sin\varphi \right), \quad (B4)$$

where

$$\varphi_0(s) = \begin{cases} 0, & s \le L; \\ \cos^{-1}(L/s), & s > L; \end{cases}$$

and

$$\varphi_1(s) = \pi/2 - \varphi_0(s).$$

Carrying out the integrations in (B4) we get

$$g(s) = \begin{cases} \pi/2 - 2s/L + s^2/(2L^2), & 0 \le s \le L; \\ \pi/2 - 1 - 2\cos^{-1}(L/s) + 2(s^2/L^2 - 1)^{1/2} - s^2/(2L^2), & L < s \le \sqrt{2}L. \end{cases}$$

When the footprints are small compared with spatial correlation lengths and the grid-box size A is large, one can easily show that

$$\Lambda^2 \to 2\pi L_{\text{int}}^2, \qquad (B5)$$

where

$$L_{\text{int}}^2 = \int_0^{\infty} s\rho(s) \ ds. \qquad (B6)$$

The quantity L_{int} is an "integral correlation length," which is just the usual correlation length, the $(1/e)$-folding distance, if the correlation $\rho(s)$ decreases exponentially.

Although the continuous integral representation of Λ^2 given by (B3) in the limit of infinite resolution is conceptually illuminating, estimation of the integral from the finite-resolution data in practice takes one back to a discrete sum. We estimated Λ^2 for each 2.5° grid-box area as follows: The footprint pairs are binned according to their mutual distance of separation in units of $d/2$, where d is the nominal diameter of an SSM/I footprint (about 28 km). For all the pairs belonging to the kth separation bin ($k = 0, 1, 2, \ldots, k_{\max} = [2\sqrt{2}L/d]$, where $[x]$ denotes the integer part of x) we compute the correlation coefficient ρ_k, the mean separation $\bar{s}_k$ and the angular factor $g_k = g(kd/2)$. In terms of these quantities,

a reasonably accurate estimate of Λ^2 is given by the Riemann-sum approximation

$$4 \sum_{k=0}^{k_{\max}} \frac{1}{2}(\rho_{k+1}\bar{s}_{k+1}g_{k+1} + \rho_k\bar{s}_kg_k)(\bar{s}_{k+1} - \bar{s}_k).$$

This method of proceeding does not require the assumption that the footprints be uniformly distributed in the area A that was used to develop the interpretation (B3) for Λ. We have tested the accuracy of the approximation by plotting $s^2\Lambda^2/A$ against σ_A^2. Our results are closely fitted by a straight line with unit slope.

REFERENCES

Bell, T. L., and P. K. Kundu, 1996: A study of the sampling error in satellite rainfall estimates using optimal averaging of data and a stochastic model. *J. Climate,* **9,** 1251–1268.
——, and ——, 2000: Dependence of satellite sampling error on monthly averaged rain rates: Comparison of simple models and recent studies. *J. Climate,* **13,** 449–462.
——, A. Abdullah, R. L. Martin, and G. R. North, 1990: Sampling errors for satellite-derived tropical rainfall: Monte Carlo study using a space–time stochastic model. *J. Geophys. Res.,* **95,** 2195–2205.
Chang, A. T. C., and L. S. Chiu, 1999: Nonsystematic errors of monthly oceanic rainfall derived from SSM/I. *Mon. Wea. Rev.,* **127,** 1630–1638.
——, ——, and T. T. Wilheit, 1993: Random errors of oceanic monthly rainfall derived from SSM/I using probability distribution functions. *Mon. Wea. Rev.,* **121,** 2351–2354.

Chen, B., and M. Yanai, 2000: Comparison of the Madden–Julian oscillation (MJO) during the TOGA COARE IOP with a 15-year climatology. *J. Geophys. Res.,* **105,** 2139–2149.

Conner, M. D., and G. W. Petty, 1998: Validation and intercomparison of SSM/I rain-rate retrieval methods over the continental United States. *J. Appl. Meteor.,* **37,** 679–700.

Ebert, E. E., and M. J. Manton, 1998: Performance of satellite rainfall estimation algorithms during TOGA COARE. *J. Atmos. Sci.,* **55,** 1537–1557.

——, P. A. Arkin, R. J. Allam, G. E. Holpin, and A. Gruber, 1996: Results from the GPCP Algorithm Intercomparison Projects. *Bull. Amer. Meteor. Soc.,* **77,** 2875–2887.

Hendon, H. H., and K. Woodberry, 1993: The diurnal cycle of tropical convection. *J. Geophys. Res.,* **98,** 16 623–16 637.

Hong, Y., C. D. Kummerow, and W. S. Olson, 1999: Separation of convective and stratiform precipitation using microwave brightness temperature. *J. Appl. Meteor.,* **38,** 1195–1213.

Huffman, G. J., 1997: Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteor.,* **36,** 1191–1201.

Kummerow, C., and L. Giglio, 1994a: A passive microwave technique for estimating rainfall and vertical structure information from space. Part I: Algorithm description. *J. Appl. Meteor.,* **33,** 3–18.

——, and ——, 1994b: A passive microwave technique for estimating rainfall and vertical structure information from space. Part II: Applications to SSM/I data. *J. Appl. Meteor.,* **33,** 19–34.

——, W. S. Olson, and L. Giglio, 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sens.,* **34,** 1213–1232.

——, W. Barnes, T. Kozu, J. Shiue, and J. Simpson, 1998: The Tropical Rainfall Measuring Mission (TRMM) sensor package. *J. Atmos. Oceanic Technol.,* **15,** 808–816.

Laughlin, C. R., 1981: On the effect of temporal sampling on the observation of mean rainfall. *Precipitation Measurements from Space, Workshop Report,* D. Atlas and O. W. Thiele, Eds., NASA Publication, D59–D66. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.]

Leith, C. E., 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.,* **12,** 1066–1069.

Madden, R. A., and P. R. Julian, 1972: Description of global-scale circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.,* **29,** 1109–1123.

——, D. J. Shea, G. W. Branstator, J. J. Tribbia, and R. D. Webber, 1993: The effects of imperfect spatial and temporal sampling on estimates of the global mean temperature: Experiments with model data. *J. Climate,* **6,** 1057–1066.

Oki, R., and A. Sumi, 1994: Sampling simulation of TRMM rainfall estimation using radar–AMeDAS composites. *J. Appl. Meteor.,* **33,** 1597–1608.

Quartly, G. D., M. A. Srokosz, and T. H. Guymer, 1999: Global precipitation statistics from dual-frequency TOPEX altimetry. *J. Geophys. Res.,* **104,** 31 489–31 516.

Short, D. A., D. B. Wolff, D. Rosenfeld, and D. Atlas, 1993: A study of the threshold method utilizing rain gauge data. *J. Appl. Meteor.,* **32,** 1379–1387.

——, P. A. Kucera, B. S. Ferrier, J. C. Gerlach, S. A. Rutledge, and O. W. Thiele, 1997: Shipboard radar rainfall patterns within the TOGA COARE IFA. *Bull. Amer. Meteor. Soc.,* **78,** 2817–2836.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission satellite. *Bull. Amer. Meteor. Soc.,* **69,** 278–295.

——, C. Kummerow, W.-K. Tao, and R. F. Adler, 1996: On the Tropical Rainfall Measuring Mission (TRMM). *Meteor. Atmos. Phys.,* **60,** 19–36.

Steiner, M., 1996: Uncertainty of estimates of monthly areal rainfall for temporally sparse remote observations. *Water Resour. Res.,* **32,** 373–388.

Takayabu, Y. N., and T. Nitta, 1993: 3–5 day-period disturbances coupled with convection over the tropical Pacific Ocean. *J. Meteor. Soc. Japan,* **71,** 221–246.

Wilheit, T. T., 1988: Error analysis for the Tropical Rainfall Measuring Mission (TRMM). *Tropical Rainfall Measurements,* J. S. Theon and N. Fugono, Eds., A. Deepak, 377–385.