Sampling-Induced Conditional Biases in Satellite Climate-Scale Rainfall Estimates

MARK L. MORRISSEY

Oklahoma Climatological Survey, University of Oklahoma, Norman, Oklahoma

JOHN E. JANOWIAK

Climate Analysis Center, National Meteorological Center/NWS/NOAA, Camp Springs, Maryland

(Manuscript received 5 April 1995, in final form 13 September 1995)

ABSTRACT

The effect of temporal sampling error in satellite estimates of climate-scale rainfall is to produce a "conditional" bias where the algorithm overestimates high rainfall and underestimates low rainfall. Thus, the bias is conditional on the value of the estimate. This paper illustrates the problem using satellite infrared rainfall estimates together with a well-known satellite algorithm and shows it to be a function of the averaging scale, the sampling rate, and the temporal autocorrelation structure of the satellite estimates. Using realistic sampling rates, it is shown that significant biases exist in satellite rainfall estimates if polar-orbiting data are used in their construction. A simple correction for this bias based upon the estimated autocorrelation structure is given.

1. Introduction

The hydrologic cycle is the most important process controlling climate and climate change. Temporal and spatial variations in this cycle produce changes in radiation and sensible and latent heat transfer and, thus, are closely associated to the global energy budget. One component of the hydrological cycle, rainfall, plays an especially important role in the Tropics, where it is a direct result of condensation, which releases latent heat to the atmosphere. Latent heating resulting from convection is the major driving force of the tropical atmosphere. A secondary, but still critical role of rainfall in the Tropics is the coupling effect it has between the ocean and the atmosphere. Tropical rainfall and the attendant dynamic and thermodynamic processes associated with convective activity have been shown to produce substantial increases in the upward sensible and latent heat flux from the ocean to the atmosphere (Garstang 1966).

Without knowledge of the temporal variation and spatial distribution of tropical rainfall, the accurate diagnoses and prediction of global climate change is impossible. The critical need for these data has prompted the World Climate Research Program's (WCRP) Global Precipitation Climatology Project (GPCP; Janowiak and Arkin 1988; WMO 1985) to assume the difficult task of assimilating global precipitation measurements into a comprehensive global precipitation dataset. This effort is supported by NOAA's Climate and Global Change Program and is the cornerstone of the Global Energy and Water Cycle Experiment (GEWEX). The GPCP dataset incorporates rain gauge data over land, but must rely on existing satellite data over the open ocean. NASA's Tropical Rainfall Measuring Mission (TRMM) plans to obtain three years of measurements of tropical rainfall with a time and space resolution on the order of 5 days to 1 month and 75 000 km^2 (i.e., a 2.5° × 2.5° box) to 250 000 km^2, respectively (Simpson et al. 1988). This is to be accomplished by sending a satellite dedicated to measuring tropical rainfall into a low inclination orbit. Both of these programs rely upon the development, calibration, and verification of algorithms that transform the radiance properties of clouds and raindrops sensed from satellites to rainfall. Many such algorithms, using visible, infrared, microwave, or combinations of any these three wavelength bands have already been developed (e.g., Kilonsky and Ramage 1976; Richards and Arkin 1981; Arkin and Meisner 1987; Adler and Negri 1988; Spencer et al. 1989; Wilheit et al. 1991; Kummerow and Weinman 1988; Ferraro et al. 1994).

Satellite algorithms designed to estimate climate-scale rainfall can be classified as those that estimate accumulated rainfall over scales greater than 75 000 km^2 and 5 days (WMO 1985). They are either based upon the bulk statistical radiance properties of clouds (e.g., Kilonsky and Ramage 1976; Richards and Arkin 1981; Arkin and Ardanuy 1989) or upon the physical relationships between received radiance and rain rate (e.g., Wilheit et al. 1991; Kummerow and Weinman

Corresponding author address: Mark L. Morrissey, Oklahoma Climatological Survey, University of Oklahoma, Sarkeys Energy Center, Suite 1210, 100 East Boyd, Norman, OK 73019-0628.

1988). For their measurements to be useful, they must be verified over various regions of the globe and different seasons. This is necessary since the assumed relationships upon which the algorithms are based may vary in time and space. The verification process requires accurate samples of surface-measured rainfall over the same time- and space scales as the satellite estimates.

Obtaining useful surface reference data is an especially difficult task over the open ocean where rain gauges are sparsely distributed and radars are practically nonexistent. Existing radars that may be used to estimate open ocean rainfall must be first calibrated before they can be used to validate satellite estimates. Unfortunately, radars that have been proposed for open-ocean satellite verification (Simpson et al. 1988) are generally located along coastlines where rainfall characteristics are unrepresentative of the open ocean. One exception to this is the radar located on Kwajalein atoll in the Marshall islands. Nevertheless, radar is similar to satellite rainfall estimation in that the indirect nature of radar remote sensing requires that a direct measure of rainfall be used in its calibration.

Although rain gauges also suffer from a variety of systematic and random errors (refer to Groisman and Legates 1994), they are the most direct measure of rainfall available and are located on most Pacific islands and atolls (Morrissey et al. 1995). The spatial distribution of rain gauges allow the major rainfall climate regimes in the Pacific to be sampled. The sparsity of their spatial distribution relative to the spatial scale of rainfall systems in the Tropics requires the use of specially developed statistical methods that minimize spatial sampling errors (Morrissey 1991; Morrissey and Greene 1993; Morrissey and Wang 1995; Berg 1993; Wyatt 1994). Efforts must also be made to minimize the effect of orography on the gauge measurements and the known low-bias effects due to wind (Legates and DeLiberty 1993). The problem with direct comparisons of rain gauge and satellite estimates is that rain gauges report accumulated rainfall at a point and satellites provide areal estimates at an instant in time. Thus, averages of rain gauge data are prone to spatial sampling error and averages of satellite data suffer from temporal sampling error.

A new statistical method that minimizes the effect of spatial sampling error from rain gauge averages on the comparisons with satellite data is now available (Morrissey 1991; Morrissey and Greene 1993). Efforts at verifying climate-scale satellite rainfall estimates over the open ocean using this method together with atoll-sited rain gauge data (Morrissey and Greene 1993; Berg 1993; Morrissey and Wang 1995) have shown a tendency for both infrared and microwave satellite rainfall algorithms to overestimate high monthly rainfall and underestimate low monthly rainfall amounts. This can result from a variety of problems including incorrect assumptions regarding the physical or statistical relationships constituting the algorithms or subjective adjustments accounting for the beam-filling problem (Chang et al. 1993). It will be shown that this can also result from temporal sampling error by the satellite. The magnitude of the error is a function of the temporal sampling rate and the autocorrelation structure of the rainfall process.

Climate-scale satellite products generally utilize either 3-h data from the geostationary satellites (Janowiak and Arkin 1991) or once or twice daily data available from the polar orbiters (Spencer et al. 1989; Ferraro et al. 1994). Depending upon the satellite sampling rate and the temporal autocorrelation structure of the rain field, these products can significantly overestimate high values and underestimate low values. Thus, given a theoretically perfect satellite algorithm (i.e., one that perfectly estimates areal rainfall at an instant in time), a bias conditional on the value of the estimate can still be produced simply from temporal sampling error. Evidence of this has been illustrated in the results of Morrissey and Greene (1993) and during the NASA WETNET's PIP-I project where most of the algorithms produced overestimates of high rainfall when compared to surface rain gauge data (Barrett et al. 1995). In this paper, an illustrative example is given of this sampling problem and a quantitative measure of the magnitude of this conditional bias is developed. A simple correction factor is shown that can be applied to these products based upon theoretical considerations.

2. Illustration of the problem

The conditional bias created from satellite temporal subsampling is illustrated using seven years (1987–93) of pentad averages of satellite rainfall estimates for three study regions located in the Tropics. The rainfall estimates were calculated using the GOES precipitation index (GPI; Richards and Arkin 1981; Arkin and Meisner 1987). The GPI uses 3-h geosynchronous infrared satellite data to estimate climate-scale rainfall. The formulation is

$$\mathrm{GPI(mm)} = 3\ \mathrm{mm\ h^{-1}} \times \mathrm{FRAC} \times \mathrm{hours}, \quad (1)$$

where FRAC is the fractional coverage of cloud below 235 K within a 2.5° × 2.5° area and "hours" is the number of hours that FRAC represents. The calibration coefficient of 3 mm h^{-1} was determined using satellite and radar data from the Global Atmosphere Research Project Atlantic Tropical Experiment (GATE) project (Arkin 1979; Richards and Arkin 1981). The three areas selected for study were the GATE region in the equatorial eastern Atlantic centered at 8.75°N, 23.75°W, the Coupled Ocean–Atmospheric Response Experiment (COARE) region centered at 1.25°S, 156.25°E in the equatorial western Pacific, and a region centered over group of atolls just east of the COARE region (1.25°N, 173.75°E; referred to in this paper as the Atoll region). Due to the large volume of satellite

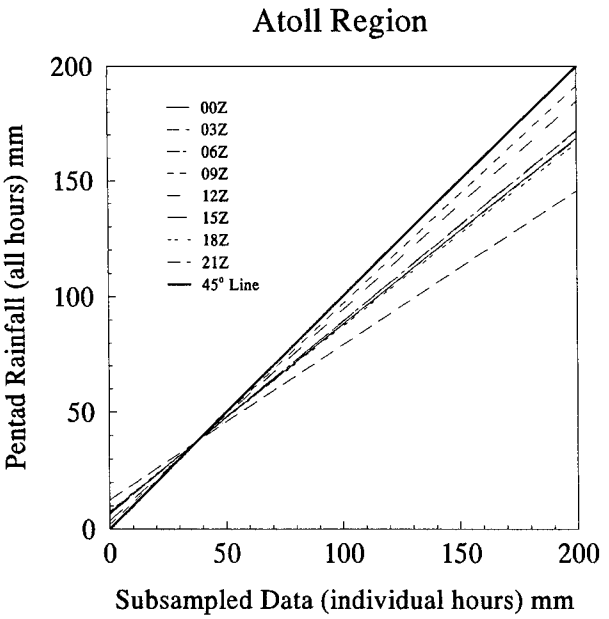

FIG. 1. Regression lines for each pentad average of a particular observation time during the day for the Atoll region defined in the text. The independent variable is the pentad average containing all hours.

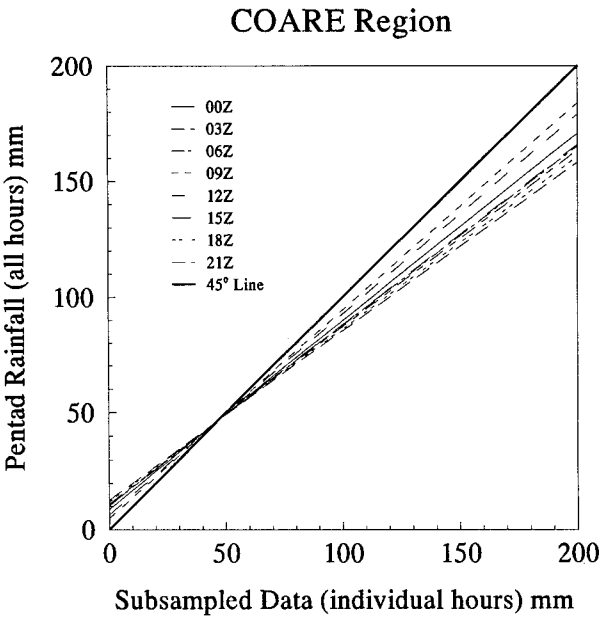

FIG. 2. Regression lines for each pentad average of a particular observation time during the day for the COARE region defined in the text. The independent variable is the pentad average containing all hours.

images required to construct the GPI, the data were collected in a specialized format. For a given pentad period, nine rainfall estimates were computed, one for the pentad average of each 3-h period (i.e., 0000, 0300, . . . , 2100 UTC, eight subsampled series) and one that included the average value from all available hours (i.e., complete series). Thus, each subsampled series contained the average of 5 estimates from a given 3-h period and the complete series was constructed from 40 observations.

A comparison between the complete series and the values from the eight subsampled series was conducted using regression analysis with estimates from the complete series as the dependent variable and estimates from the subsampled series as the independent variable. Results from the three locations (Figs. 1, 2, and 3) show that in every case the slopes are less than one. This indicates that a subsampling rate of one observation per day produces a substantial conditional bias and the magnitude of this bias varies from location to location. The diurnal cycle acts to increase or decrease the overall bias and does not affect the conditional bias. This suggests that subsampling, by itself, creates a conditional bias that should be correctable if the slope and intercepts can be determined theoretically.

3. Theory

Suppose a satellite algorithm has been developed that produces perfect estimates of rainfall over an area at an instant in time and two identical satellite systems are available, one with a very high temporal sampling rate whereby the temporal sampling error is essentially zero and the other with a lower, more realistic sampling

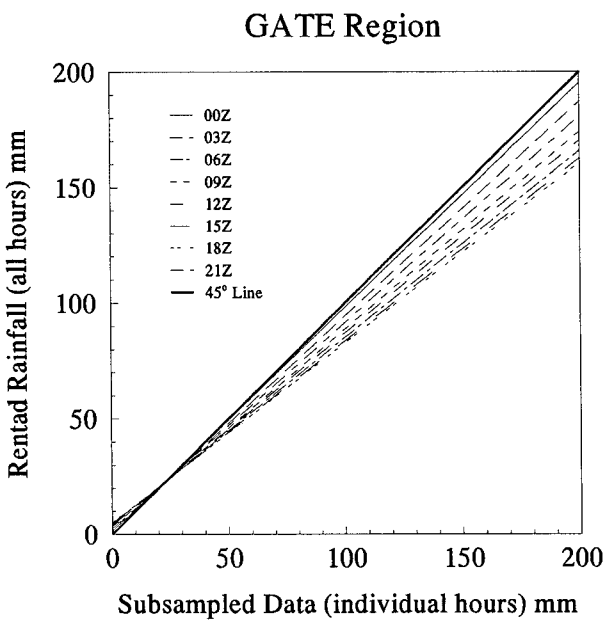

FIG. 3. Regression lines for each pentad average of a particular observation time during the day for the GATE region defined in the text. The independent variable is the pentad average containing all hours.

rate (e.g., 3 or 12 h in between each observation). The quantity to be estimated is time- and space-averaged rainfall over the scales required for climate-scale research (e.g., pentad or monthly values over 77 000 km^2). The conditional bias will be defined as one minus the slope of the regression line between estimates made from the satellite with the high sampling rate (the dependent variable) and the one with the realistic sampling rate (the independent variable). The time–space-averaged rainfall estimated using the satellite with the very high sampling rate is determined using

$$R_m = \frac{\sum_{i=1}^{T} S(t_i)}{T}, \tag{2}$$

where $S(t_i)$ is the satellite rainfall measurement representing time t_i where i is the ith sample during the averaging period (e.g., pentad or month). Since the number of samples T is assumed to be very large, R_m can be considered to be the population value for the given averaging period. It is assumed that the samples are equally spaced in time throughout the averaging period. For the satellite with the lower sampling rate, the corresponding time–space rainfall value is estimated using

$$R_m^* = \frac{\sum_{i=1}^{T} S(t_i)\delta(t_i)}{T'}, \tag{3}$$

where

$$\delta(t_i) = \begin{cases} 1, & \text{if a sample is present at time } t_i \\ 0, & \text{otherwise} \end{cases}$$

and T' is the actual number of samples taken during the month; that is,

$$T' = \sum_{i=1}^{T} \delta(t_i). \tag{4}$$

The delta function is used simply as a place holder in time so that different satellite sampling schemes may be tested (i.e., samples irregularly spaced in time may be used). To simplify the derivation, the long-term areal mean rainfall μ has been removed from the instantaneous satellite estimates to produce anomalies. The slope b_1 and intercept b_0 of the regression relationship between the dependent and independent variables is defined as shown by Draper and Smith (1981),

$$b_1 = \frac{\sigma_{R_m}}{\sigma_{R_m}^*}\rho(R_m, R_m^*)$$

$$b_0 = \mu - b_1\mu, \tag{5}$$

where σ_{R_m} is the temporal standard deviation of the dependent variable, $\sigma_{R_m}^*$ is the temporal standard deviation of the independent variable, and $\rho(R_m, R_m^*)$ is the correlation between R_m and R_m^*. Expanding Eq. (5) using expectations yields

$$b_1 = \frac{E[R_m^2]^{1/2}}{E[R_m^{*2}]^{1/2}} \frac{E[R_m R_m^*]}{E[R_m^2]^{1/2} E[R_m^{*2}]^{1/2}} = \frac{E[R_m R_m^*]}{E[R_m^{*2}]}, \tag{6}$$

where the expectation is taken over time. Substituting the expressions for R_m and R_m^* into the numerator of Eq. (6) yields

$$\begin{aligned} E[R_m R_m^*] &= \frac{1}{TT'} E\left[\sum_{i=1}^{T}\sum_{i'=1}^{T} S(t_i)S(t_{i'})\delta(t_{i'})\right] \\ &= \frac{1}{TT'}\left\{E\left[\sum_{i=1}^{T} S(t_i)^2\delta(t_i)\right]\right. \\ &\quad + E\left[\sum_{i=1}^{T-1}\sum_{i'=i+1}^{T} S(t_i)S(t_{i'})\delta(t_{i'})\right] \\ &\quad \left. + E\left[\sum_{i=1}^{T-1}\sum_{i'=i+1}^{T} S(t_i)S(t_{i'})\delta(t_i)\right]\right\}. \end{aligned} \tag{7}$$

Rearranging terms in Eq. (7) and taking expectations produces

$$\begin{aligned} E[R_m R_m^*] &= \frac{\sigma_s^2}{T} + 2\frac{\sigma_s^2}{T^2}\sum_{L=1}^{T-1}\rho[S(t_i), S(t_{i+L})](T-L) \\ &= \frac{\sigma_s^2}{T}\left[1 + \frac{2}{T}\sum_{L=1}^{T-L}(T-L)\rho_s(L)\right], \end{aligned} \tag{8}$$

where the time-lag index between samples is L, the variance of the individual sample measurements about the long-term mean is σ_s^2, and the sum over T of $\delta(t_i)$ is T'. It is also assumed that

$$\rho[S(t_i), S(t_{i+L})] = \frac{E[S(t_i), S(t_{i+L})]}{\sigma_s^2} = \rho_s(L) \tag{9}$$

is the lagged autocorrelation between measurements. In this development

$$\sum_{i=1}^{T-L}\delta(t_i) \approx \sum_{i=L+1}^{T}\delta(t_i) \tag{10}$$

has been assumed, which is approximately valid for satellite data where the sampling scheme is not extremely skewed to either end of the sampling period. It is common for polar-orbiting satellites not to sample at regular intervals depending on orbital parameters and the latitude sampled, but generally a sample is taken at least once every day or every other day over a given location. The delta functions can easily take any sampling irregularities into account by simply modifying their values to adjust for the actual sampling scheme.

The expected value of the autocorrelation over the averaging period, m, is equal to

$$E[\rho|m] = \frac{2}{T(T-1)} \sum_{L=1}^{T-1} (T-L)\rho(L)$$

$$= \int_0^T f(L)\rho(L)dL, \qquad (11)$$

where $f(L)$ is the frequency function of the number of lags of a given time difference (i.e., L) within T. This equals $2(T-L)T^{-2}$. By substituting the expression given in Eq. (11) into Eq. (8), the numerator of Eq. (6) equals

$$E[R_m R_m^*] = \frac{\sigma_s^2}{T}\{1 + (T-1)E[\rho|m]\}$$

$$= E[R_m^2] = \sigma_{R_m}^2. \qquad (12)$$

This well-known equation (Rodriguez-Iturbe and Mejia 1974; Yevjevich 1982; Morrissey 1991) expresses the temporal variance of an average of T autocorrelated samples as a function of the expected value of the autocorrelation over the sampling period m. Thus, the expected value of the product of the averages of the complete and subsampled values over m is simply the variance of the average of the series with the higher sampling rate. As the number of samples T taken during the sample period, m, approaches infinity, Eq. (12) becomes

$$E[R_m R_m^*] = \sigma_s^2 E[\rho|m] = \sigma_{R_m}^2. \qquad (13)$$

Expanding the denominator in Eq. (6) by substituting in Eq. (3) for R_m^* and taking expectations over time yields

$$E[R_m^{*2}] = E\left[\left(\sum_{i=1}^{T} \frac{S(t_i)\delta(t_i)}{T'}\right)^2\right]$$

$$= E\left[\frac{1}{T'^2}\sum_{i=1}^{T}\sum_{i'=1}^{T} S(t_i)S(t_{i'})\delta(t_i)\delta(t_{i'})\right] = \frac{\sigma_s^2}{T'^2}$$

$$\times\left\{T' + 2\sum_{i=1}^{T-1}\sum_{i'=i+1}^{T}\rho[S(t_i), S(t_{i'})]\delta(t_i)\delta(t_{i'})\right\}$$

$$= \frac{\sigma_s^2}{T'^2}\left[T' + 2\sum_{L=1}^{T-1}\rho(L)\sum_{i=1}^{T-L}\delta(t_i)\delta(t_{i+L})\right]. \qquad (14)$$

The variable T, which expresses the number of samples within the averaging period, does not have to be taken as infinity, but should be set as the total possible number of samples in the averaging period (e.g., one month) for a given satellite system. For example, a polar-orbiting satellite may take two samples a day over a given area and, thus, 60 samples are possible for a 30-day month (i.e., $T = 60$). Frequently, samples from a particular satellite are missing for some reason. This is accounted for in Eq. (14) by defining T' to be the actual number of samples during the month ($T' \leq T$). The delta function $\delta(t_i)$ is one for a sample that was taken at time t_i and zero if a sample is missing at that time. Equation (14) is simplified by defining a weighting factor,

$$W(L) = \sum_{i=1}^{T-L} \delta(t_i)\delta(t_{i+L}), \qquad (15)$$

which accounts for temporal arrangement between missing samples. Thus, Eq. (14) then becomes

$$E[R_m^{*2}] = \frac{\sigma_s^2}{T'}\left[1 + \frac{2}{T'}\sum_{L=1}^{T-1}\rho(L)W(L)\right]. \qquad (16)$$

By recombining the numerator and denominator of Eq. (6) using Eqs. (13) and (16), the expression for the slope becomes

$$b_1 = \frac{E[R_m R_m^*]}{E[R_m^{*2}]} = \frac{E[\rho|m]}{\frac{1}{T'} + \frac{2}{T'^2}\sum_{L=1}^{T-1}\rho(L)W(L)}, \qquad (17)$$

and, therefore, the conditional bias is one minus this quantity. It can be readily seen that as T' approaches infinity the slope approaches one since $W(L)$ becomes $T - L$ and the denominator of Eq. (17) approaches $E[\rho|m]$ [refer to Eq. (11)].

4. Examining the conditional bias

Using Eq. (17) it is possible, given an estimated autocorrelation function and a known satellite sampling rate, to predict the conditional bias using available satellite data. Once the bias is determined, it can be removed from satellite estimates. By assuming an exponential autocorrelation structure of the form

$$\rho(L) = e^{-L/\tau_A}, \qquad (18)$$

Eq. (17) can be solved to determine the magnitude of the slope with different e-folding autocorrelation times τ_A and different sampling rates. To determine the magnitude of the conditional bias in the GPI estimates, the temporal autocorrelation function representing the GPI is required. Ideally, GPI estimates spanning the entire 8-yr period in which the GPI has been produced should be used in the estimation of the autocorrelation function over a specified region. Unfortunately, due to data volume considerations the GPI values are not in a format where the autocorrelation structure is obtainable. However, a dataset of hourly infrared values taken from the Japanese Geostationary Meteorological Satellite (GMS) for the TOGA COARE experiment were available for the 4-month period of that experiment (November 1992 through February 1993). Thus, as a first approximation, GPI rainfall estimates were constructed for the COARE region using these data from which autocorrelation coefficients at different lag times were then estimated. This was done for GPI estimates over

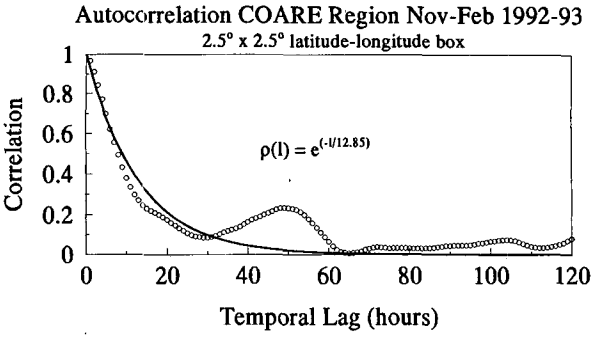

FIG. 4. The time-lagged autocorrelation determined from four months of hourly GMS infrared data over a 2.5° × 2.5° area located in the COARE region. The least squares fit exponential function is also shown.

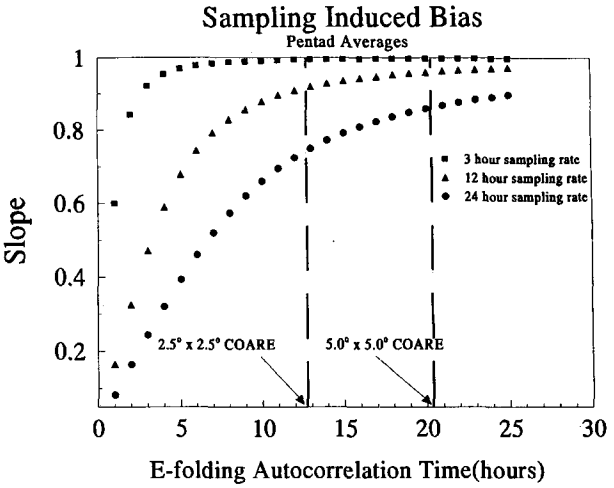

FIG. 6. The results for pentad averages, using Eq. (17), of slope values given an exponential autocorrelation function with different e-folding times for three different sampling rates. The dashed lines indicate the e-folding times determined using satellite data for the COARE region over the spatial scales indicated.

a 2.5° × 2.5° area and a 5.0° × 5.0° area. The exponential autocorrelation function shown in Eq. (18) was then fit to both sets of coefficients using a least squares routine. The results (Figs. 4 and 5) indicate that an exponential function provides a reasonable fit to the estimates with the exception of the secondary peak in the estimates at approximately 50-h lag time. If used to investigate the bias in a particular satellite, it would be advisable to attempt to fit a function that better matches the behavior of the autocorrelation coefficients. However, for testing purposes we preferred to utilize the simple exponential function, which does not provide an optimal fit but does allow for easy manipulation of the single e-folding time parameter. This provides us with the opportunity to observe the variation in the slope with changes in this parameter.

The e-folding time for the 2.5° × 2.5° area is 12.85 h and for the 5.0° × 5.0° area the value was 20.44 h. Using these values, Figs. 6 and 7 show the slope values for various sampling rates for the COARE region for pentad and monthly averages, respectively. For a 3-h geostationary sampling rate the slope is close to one for both sized areas and time averages. However, if the GPI is computed using polar-orbiter data having once or twice a day sampling rates the conditional bias (i.e., 1 minus slope) can be significant (i.e., >0.15), especially for the smaller sized areas.

The e-folding autocorrelation times vary from season to season and from place to place. Laughlin (1981) determined the e-folding distance τ_A using GATE radar rainfall data for a 280 km × 280 km area to be 7.16 h (for Phase I). Bell (1987) and Bell et al. (1990) de-

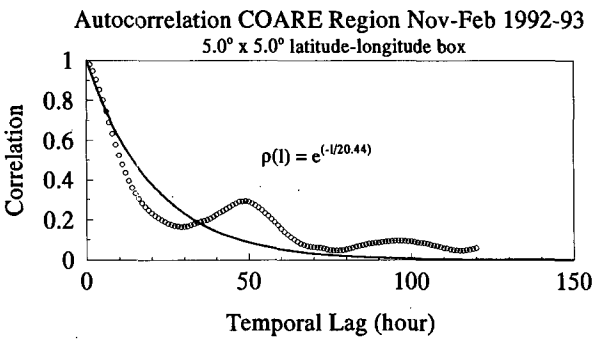

FIG. 5. The time-lagged autocorrelation determined from four months of hourly GMS infrared data over a 5.0° × 5.0° area located in the COARE region. The least squares fit exponential function is also shown.

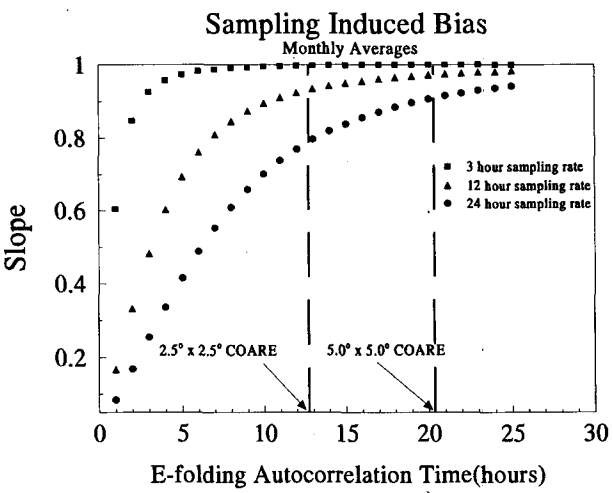

FIG. 7. The results for monthly averages, using Eq. (17), of slope values given an exponential autocorrelation function with different e-folding times for three different sampling rates. The dashed lines indicate the e-folding times determined using satellite data for the COARE region over the spatial scales indicated.

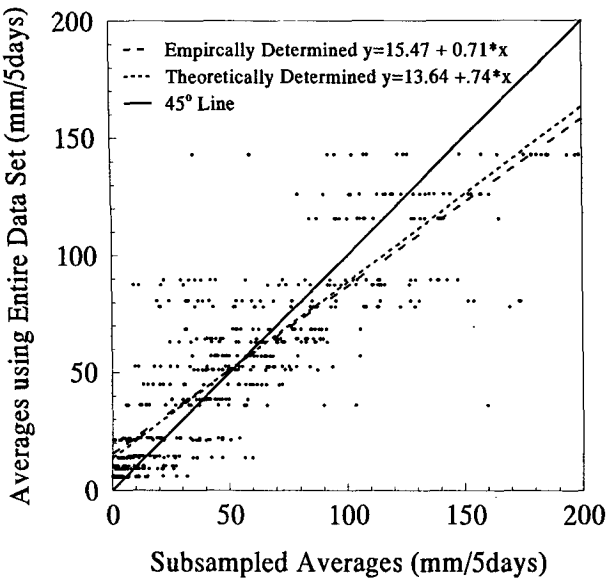

FIG. 8. A scatter diagram comparing pentad averages of GPI-estimated pentad rainfall averages for the TOGA COARE region with the same constructed using only 1-h estimates per day. Both empirically determined and theoretically determined regression lines are shown along with the 45° line.

veloped a stochastic rainfall model representing the GATE region and determined the e-folding distance representing the extrapolated area of 580 km × 580 km to be approximately 10.6 h. Thus, the conditional bias will change according to many factors.

5. Adjustment for the conditional bias

The conditional bias is a direct result of the central limit theorem, which states that a sampling distribution of averages will converge to the population distribution with an increased number of independent samples within the averaging period. Thus, the variance of the sample distribution is always greater than the population distribution. This results in overestimates of high rainfall and underestimates of low rainfall. This can be adjusted if the slope and intercepts are known using equations developed in this paper. Once the slope is computed, the intercept can be derived using Eq. (5) and the following adjustment can be made to the satellite estimates:

$$R'_m = b_1 R^*_m + b_0, \qquad (19)$$

where R^*_m are original values and R'_m are the transformed values in which the sampling-induced conditional bias was removed. It should be noted that other biases due to algorithm deficiencies may remain in the estimates.

As a test of this adjustment, the TOGA COARE hourly GMS infrared satellite data were used to construct rainfall estimates using the GPI formula. Pentad averages were constructed using data for a given hour producing 24 (hours) × 24 (pentad) values (i.e., 576 values). Each pentad for a given hour was compared to the pentad value constructed using the entire data. A scatter diagram was produced from these comparisons (Fig. 8) and a simple linear regression function was to fit the points. Using an e-folding time of 12.85 h (Fig. 4) Eqs. (17) and (5) were used to obtain the theoretical slope and intercept. Figure 8 illustrates the very small differences observed between the empirical and theoretical regression coefficients. The adjusted values using the theoretical regression parameters in Eq. (19) are shown in Fig. (9). The regression line fitted to these values produces a slope very near to 1.0 and an intercept close to 0.0. While the scatter about the fitted regression lines remains unchanged (correlation coefficient is 0.84 in both cases) the conditional bias is essentially removed.

6. Conclusions

This paper demonstrates the effect of temporal sampling error on the estimation of time and space averaged rainfall from satellite algorithms. The stochastic relationship between sampling and the conditional bias was developed, and a correction adjustment was given. It was also demonstrated that the conditional bias given realistic sampling rates, averaging scales, and temporal correlation structures is significant, especially if polar-orbiting data are used in the construction of rainfall averages. This is quite relevant for microwave rainfall algorithms using the Special Sensor Microwave/Imager (SSM/I) onboard the Defense Meteorological Satellite Program (DMSP) satellite.

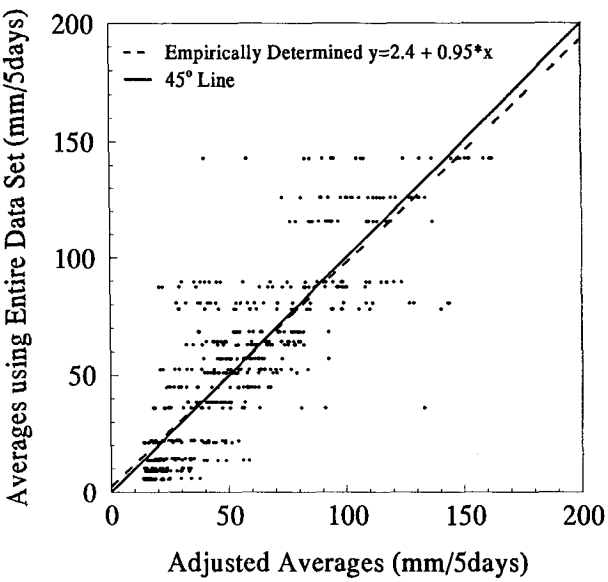

FIG. 9. Same as in Fig. 8 except using values produced using Eq. (19) in the text.

It was also noted that the conditional bias changes in time and space depending upon the variation in the temporal correlation structure of the satellite estimates. This could explain the variations in the accuracy of satellite estimates from one test site to another. This problem, while significant, is easily corrected using the correction adjustment described in this paper. It is recommended that a conditional bias correction be the first step in algorithm validation so that the accuracy of algorithm estimates may be more easily determined. This requires the estimation of the autocorrelation structure. This can be accomplished using the available satellite data over different regions of the globe. Unfortunately, the low sampling rates of the polar orbiters make this estimation difficult. One solution may be that a polar-orbiting algorithm be applied to geostationary data representing a given region and these data used to estimate this structure.

Acknowledgments. The authors would like to thank Mr. Ralph Ferraro and Dr. Phillip Arkin for their helpful comments concerning the formulation given in this paper. The first author would like to note the support of this work by the National Oceanic and Atmospheric Administration through Grants NA37RJ0202 III-17 and NA90RAH0078.

REFERENCES

Adler, R. F., and A. J. Negri, 1988: A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteor.*, **27,** 30–51.

Arkin, P. A., 1979: The relationship between fractional coverage of high cloud and rainfall accumulation during GATE over the B-scale array. *Mon. Wea. Rev.*, **107,** 1382–1387.

——, and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud in the Western Hemisphere during 1982–84. *Mon. Wea. Rev.*, **115,** 51–74.

——, and P. E. Ardanuy, 1989: Estimating climate-scale precipitation from space: A review. *J. Climate,* **2,** 1229–1238.

Barrett, E. C., and Coauthors, 1994: The first WETNET precipitation intercomparison project (PIP-1): Interpretation of results. *Remote Sens. Rev.*, **11,** 1–12.

Bell, T. L., 1987: A space–time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res.*, **92,** 9631–9643.

——, 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,** 2195–2206.

Berg, W. K., 1993: Estimation and analysis of climate-scale rainfall over the tropical Pacific. Colorado Center for Astrodynamics Research, 202 pp.

Chang, A. T., L. S. Chiu, 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.

Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis.* John Wiley & Sons, 709 pp.

Ferraro, R., N. Grody, D. Forsyth, R. Carey, A. Basist, J. Janowiak, F. Weng, G. F. Marks, and R. Yanamandra, 1994: Microwave measurements produce global climate hydrologic data. *Eos, Trans. Amer. Geophys. Union,* **75,** 337–338.

Garstang, M., 1966: Sensible and latent heat exchange in low latitude synoptic scale systems. *Tellus,* **19,** 492–508.

Groisman, P. Y., and D. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75,** 215–227.

Janowiak, J. E., and P. A. Arkin, 1988: The Global Precipitation Climatology Project: description, goals nd progress. Proc. *13th Annual Climate Diagnostics Workshop,* Cambridge, MA, 343–346.

——, and ——, 1991: Rainfall variations in the Tropics during 1986–1989, as estimated from observations of cloud-top temperature. *J. Geophys. Res.*, **96,** 3359–3373.

Kilonsky, B. J., and C. S. Ramage, 1976: A technique for estimating tropical open-ocean rainfall from satellite observations. *J. Appl. Meteor.*, **15,** 972–975.

Kummerow, C., and J. A. Weinman, 1988: Determining microwave brightness temperature from horizontally finite and vertically structured clouds. *J. Geophys. Res.*, **93,** 3720–3728.

Laughlin, C. R., 1981: On the effect of temporal sampling on the observation of mean rainfall. Precipitation measurements from space. Workshop Rep. NASA, Goddard Space Flight Center, Greenbelt, MD, 56–66.

Legates, D. R., and T. L. DeLiberty, 1993: Precipitation measurement biases in the United States. *Water Resour. Bull.*, **29,** 855–861.

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

——, and J. S. Greene, 1993: Comparison of two satellite-based rainfall algorithms using Pacific atoll raingage data. *J. Appl. Meteor.*, **32,** 411–425.

——, and Y. Wang, 1995: Verifying satellite microwave rainfall estimates over the open ocean. *J. Appl. Meteor.*, **34,** 794–804.

——, M. A. Shafer, S. E. Postawko, and B. Gibson, 1995: The Pacific rain gauge rainfall database. *Water Resour. Res.*, **31,** 2111–2113.

Richards, F., and P. A. Arkin, 1981: On the relationship between satellite-observed cloud cover and precipitation. *Mon. Wea. Rev.*, **109,** 1081–1093.

Rodriguez-Iturbe, I., and J. M. Mejia, 1974: The design of rainfall networks in time and space. *Water Resour. Res.*, **10,** 713–728.

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

Spencer, R. W., H. M. Goodman, and R. E. Hood, 1989: Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. *J. Atmos. Oceanic Technol.*, **2,** 254–263.

Wilheit, T. T., A. T. C. Chang, and L. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometric measurements using the probability distribution function. *J. Atmos. Oceanic Technol.*, **8,** 118–136.

World Meteorological Organization, 1985: Review of the requirements of area-averaged precipitation data, surface-based and space-based estimation techniques, space and time sampling, accuracy and error; Data exchange. WMO No. 115, 57 pp.

Wyatt, A., 1994: Estimation of area-average rain rates using WSR-88D (NEXRAD) reflectivity measurements and a dense raingage network. M.S. thesis, Dept. of Meteorology, University of Oklahoma, 112 pp.

Yevjevich, V., 1982: *Probability and Statistics in Hydrology.* Water Resources Publications, 302 pp.