Detecting the Diurnal Cycle of Rainfall Using Satellite Observations

THOMAS L. BELL

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

N. REID

Department of Statistics, University of Toronto, Toronto, Canada

(Manuscript received 2 July 1991, in final form 4 March 1992)

ABSTRACT

The diurnal cycle in rainfall varies considerably from region to region in the tropics. Determining this variability is important both for comparing predictions of atmospheric models to real atmospheric behavior and for making sure that estimates of total rainfall from low-altitude satellites are not biased because of their infrequent observations of a given region of the earth. Although there are no data from the proposed Tropical Rainfall Measuring Mission (TRMM) satellite to work with yet, we can ask how well the diurnal cycle in rainfall will be detected when the satellite is eventually collecting data, given the satellite's proposed sampling characteristics. Data analyses for the diurnal cycle are discussed, taking into account the fact that the satellite visits will be irregularly spaced in time. The amplitudes of the first few harmonics will be determined by least-squares fits to the satellite observations, and the tests needed to establish the statistical significance of the fitted amplitudes are discussed. The accuracy with which the first few harmonics of the diurnal cycle can be detected is estimated from several months of satellite data using rainfall statistics observed during the GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment (GATE).

1. Introduction

There is good reason to expect that the probability distribution of rain in a given area will vary with the time of day since atmospheric dynamics are influenced by the daily passage of the sun. Establishing this diurnal cycle quantitatively in rainfall is important for several reasons. Because its amplitude and phase are the result of a subtle interplay between dynamical and radiative processes, the degree of our success in explaining it serves as a useful measure of our understanding of the physics of the atmosphere on this time scale. Knowledge of changes in rainfall statistics with the time of day is also essential in interpreting satellite estimates of rainfall, since satellites (at least nongeosynchronous ones) view a given spot only intermittently, and interpolating between the measurements should be adjusted according to the time of day.

Randall et al. (1991) have recently compared the diurnal cycle in rainfall seen in their general circulation model (GCM) with rainfall climatology where it is known. They remark how variable the diurnal cycle is over the earth, even over the oceans. The physical mechanisms behind these variations are not always obvious. Over the oceans, the diurnal cycle seems controlled by cloud radiative heating and cooling. There is, of course, a dearth of surface rainfall measurements over the oceans, and we must generally rely on satellite estimates of rainfall for establishing the global climatologies needed for testing the GCMs.

There have been some attempts to characterize the diurnal cycle of rainfall based on inferences of convective activity from images taken in the visible and infrared by geosynchronous satellites. These satellites have the advantage that they provide frequent observations during the course of a day over long periods and so can document changes in the statistics with the time of day. Augustine (1984) has, for instance, found afternoon maxima in tropical Pacific rainfall, inferred using the Griffith and Woodley convective infrared satellite technique described in Griffith et al. (1978). Shin et al. (1990), using the Arkin (1979) GOES (Geostationary Operational Environmental Satellite) precipitation index (GPI) to estimate rain rates, find a midday maximum over much of the tropical rainy areas. These findings seem to be in conflict with those of Gray and Jacobson (1977), who find early morning maxima in island raingage reports, and also with Fu et al.'s (1990) finding that deep convective clouds, as inferred from ISCCP (International Satellite Cloud Climatology Project) data, tend to peak in the morning hours. There is some concern that infrared techniques may overestimate rainfall from cold cirrus debris left

Corresponding author address: Thomas L. Bell, NASA/Goddard Space Flight Center, Code 913, Greenbelt, MD 20771.

behind by deep convection and show a peak in rainfall later than it actually occurs.

These and many other questions concerning tropical dynamics require more accurate rainfall measurements with the coverage that only satellites can provide. The National Aeronautics and Space Administration (NASA), in collaboration with the National Space Development Agency of Japan (NASDA), has recently proposed orbiting a satellite specifically designed to observe tropical rainfall from space, the Tropical Rainfall Measuring Mission (TRMM). Simpson et al. (1988) describe its goals and the instruments it will carry (radar, most importantly), although details of the instrument package are still being modified to accommodate various mission constraints.

A number of statistical questions connected with the mission have been examined recently. Laughlin (1981) early on provided a rough estimate of the accuracy with which monthly means could be obtained, given the TRMM revisit times to a given area, which subsequent more elaborate estimates have borne out (Shin and North 1988; Bell et al. 1990). Based on the statistics of rain observed in the GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment (GATE), TRMM should be able to provide monthly means over 5° boxes accurate to about 10%.

We shall investigate here the accuracy with which the diurnal cycle of rainfall can be extracted from TRMM data when it becomes available. The statistical problem is described in section 2. Using the statistical methods developed there, the diurnal cycle in the GATE data is reexamined in section 3. The implications for TRMM are explored in section 4. Problems still unresolved are discussed in section 5. Some technical details are given in the appendixes.

2. Satellite estimation of the diurnal cycle

Although no data from the TRMM satellite are available to work with yet, we can ask whether a diurnal cycle will be detectable when the satellite is eventually collecting data, given the satellite sampling characteristics. That there is something to detect has been amply demonstrated in many studies, especially over the continents (see, for example, Wallace 1975; Hamilton 1981; Desbois et al. 1989). Over the oceans, there is indirect and somewhat conflicting evidence from satellite observations, as has been described in the previous section. Because the diurnal cycle tends to be weaker there, statistical concerns will play a larger role in interpreting the results.

Quite a few simplifying assumptions about the problem will be made in order to obtain some preliminary estimates of the size of the phenomena being dealt with. Some of the assumptions can be relaxed at the cost of greater complexity in the analysis. Others, however, will require careful study in order to deal adequately with the problems raised. These issues will be discussed further in this section and in section 5.

Suppose that estimates for the area-averaged rainfall,

$$R_A(t_m) = \frac{1}{|A|}\int_A R(\mathbf{x}, t_m)d\mathbf{x}, \qquad (2.1)$$

within a given area A are provided at observation times t_m ($m = 0, 1, \cdots$). These observation times t_m will be separated by intervals ranging from roughly 2 to 24 h, depending on the physical location of the area observed. An example is shown in Fig. 1. Near the equator, the observations will be nearly equally spaced at intervals of approximately 12 h. Near 30° latitude, the observations will occur in bursts of 4 or 5, 90 min apart, followed by a gap of some 20 h.

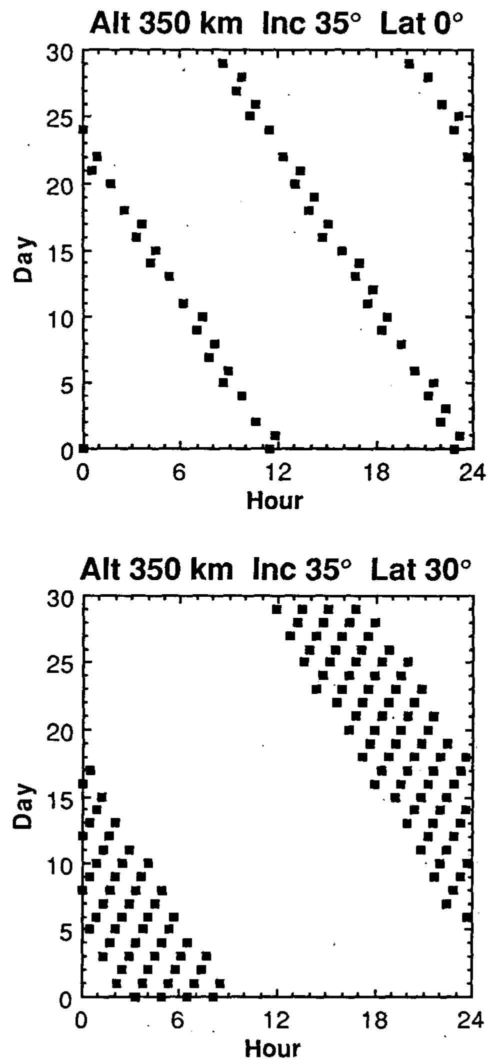

FIG. 1. Local observation times of a 500-km × 500-km area centered on the equator (top) and at 30°N (bottom) for 1 month of observations by a satellite orbiting at 35° inclination, 350-km altitude and viewing a 720-km-wide swath.

A diurnal cycle, if it exists, would be represented by a change in the probability distribution of rain with the time of day. This study will concentrate on determining a single aspect of this change: the variation of the mean rain rate (averaged over A) with the time of day. Other statistics that might be both strongly dependent on time of day and physically interesting are the fraction of area with rain and changes in the parameters of the probability distribution of rain, or of rain type, with time of day, but the statistical problems associated with them will not be investigated here.

To see if the mean rain rate changes with the time of day, averages of the observations are found for the different periods of the day. Break the day up into time intervals of equal duration (each 1 h long, for instance), using t to denote the time interval whose midpoint is at time t. Then the interval averages can be written formally as

$$\bar{R}(t) = \frac{1}{n_t} \sum_{t_m \in t} R_A(t_m), \qquad (2.2)$$

where the notation $t_m \in t$ denotes observations that fall within the interval t and n_t is their number,

$$n_t = \sum_{t_m \in t} 1. \qquad (2.3)$$

If there is a diurnal cycle, the climatic mean rainfall $E\bar{R}(t)$ will vary periodically during the course of the day. Here the expectation is a hypothetical average over a large number of days but taken at the same time of day and during the same season, with all other "climatological influences" held fixed. The diurnal cycle in $r(t) = E\bar{R}(t)$ can be expressed by a Fourier series

$$r(t) = r_0 + r_1 \cos(\omega t - \phi_1) + r_2 \cos(2\omega t - \phi_2) + \cdots, \qquad (2.4)$$

where the first term r_0 represents the daily mean rainfall,

$$\omega = 2\pi(24 \text{ h})^{-1}, \qquad (2.5)$$

is the diurnal frequency, and r_1 and ϕ_1 are the amplitude and phase for simple sinusoidal variation of the mean. The number of harmonics needed to describe the variation of $r(t)$ is limited by the number of intervals into which the day has been broken up. Harmonics higher than the first few will be neglected in what follows.

A natural approach to determining the parameters of the expansion (2.4) is to find ones that minimize the total square difference of the fit

$$D^2 = \sum_t [\bar{R}(t) - r(t)]^2 w_t, \qquad (2.6)$$

where the sum is over all the hourly intervals, and w_t is a weighting factor proportional to the reciprocal of the expected variance of $[\bar{R}(t) - r(t)]$. Because the usual assumptions of independence and normality are not met, the least-squares estimates will not have the usual efficiency properties, although they will be consistent. More efficient estimates could be obtained by generalized least squares if the form of the correlation function was known or by maximum likelihood if a plausible distribution for rain rate was available.

For lack of a better estimate, the weighting

$$w_t \propto n_t \qquad (2.7)$$

will be used, since this would be the case for independently identically distributed differences. It also has the advantage that minimizing (2.6) is essentially equivalent to a least-squares fit of $r(t)$ to $R_A(t_m)$; that is, the parameters in (2.4) can be obtained by minimizing

$$D'^2 = \sum_{m=0}^{N} [R_A(t_m) - r(t_m)]^2, \qquad (2.8)$$

where $N + 1$ is the number of satellite observations. The approximation becomes exact as the number of observations increases. See appendix A for the argument. By recasting the problem in terms of minimizing (2.8), we can draw on standard methods for time series analysis to develop significance tests for the fitted harmonics. The least-squares approach to detecting a periodic signal in unevenly spaced data is discussed, for example, by Press and Teukolsky (1988).

Since we are trying here to obtain an idea of the order of magnitude of the statistical problem posed by the intermittent sampling of the satellite, we shall simplify it by assuming that the satellite samples are at equally spaced intervals

$$t_m = m\Delta t, \quad m = 0, \cdots, N, \qquad (2.9)$$

for one month, so that $N\Delta t$ is approximately one month. The fact that the satellite will not view the entire area on every visit will also be ignored. A more rigorous treatment of the problem would take into account the size of the area averaged over for each observation of area-averaged rain rate.

The harmonic analysis (2.4) of the diurnal cycle is cast in a form amenable to least squares with linear coefficients by writing it as

$$r(t) \approx r_0 + c \cos\omega t + s \sin\omega t, \qquad (2.10)$$

with $c = r_1 \cos\omega\phi_1$ and $s = r_1 \sin\omega\phi_1$. The higher harmonics in (2.10) are neglected, but they can be treated as a straightforward generalization of the approach followed. It can also be shown that if all hours of the day are sampled equally, estimates of the amplitudes of the diurnal harmonics can be made independently of each other because the Fourier modes are orthogonal. In fact, this assumption will be made here so that the mean r_0 can be estimated independently of the first harmonic from a simple average of the data $\{R_A(t_m), m = 0, \cdots, N\}$ as

$$\hat{r}_0 = \frac{1}{N+1} \sum_{m=0}^{N} R_A(t_m). \qquad (2.11)$$

If the satellite observations during the different parts of the day were not approximately equal in number, r_0 could not be estimated in this way independently of r_1. An estimate of the diurnal cycle could be obtained by grouping the observations according to the hour of the day as in (2.2) and minimizing (2.6). Significance tests become more difficult and may require Monte Carlo methods.

The least-squares estimates of s and c are

$$\hat{s} = D^{-1}(A_{cc}S - A_{sc}C), \qquad (2.12a)$$

$$\hat{c} = D^{-1}(-A_{sc}S + A_{ss}C), \qquad (2.12b)$$

where

$$D = A_{ss}A_{cc} - A_{sc}^2,$$

and

$$A_{ss} = \sum_{m=0}^{N} \sin^2 \omega t_m,$$

$$A_{cc} = \sum_{m=0}^{N} \cos^2 \omega t_m,$$

$$A_{sc} = \sum_{m=0}^{N} \sin \omega t_m \cos \omega t_m,$$

and S and C are the sine and cosine components of the rainfall data

$$S = \sum_{m=0}^{N} \sin(\omega t_m)[R_A(t_m) - \hat{r}_0], \qquad (2.13a)$$

$$C = \sum_{m=0}^{N} \cos(\omega t_m)[R_A(t_m) - \hat{r}_0]. \qquad (2.13b)$$

The estimated amplitude is $\hat{r}_1^2 = \hat{s}^2 + \hat{c}^2$ and the estimated phase is $\hat{\phi}_1 = \tan^{-1}(\hat{s}/\hat{c})$. The usual distribution theory for least-squares does not apply here because the observations $R_A(t_m)$ are correlated and the central-limit theorem cannot be directly applied. However, S and C are sums of data extending over times much longer than the correlation time of rain, and it is assumed that they contain enough effectively independent observations to be approximately normally distributed. This assumption will be examined with some simulated data in the next section. It is also assumed that the time series of rain rate is sufficiently long that the variances of $\hat{s}$ and $\hat{c}$ are approximately the same and that the correlation of $\hat{s}$ and $\hat{c}$ may be neglected. It is easy to show that this must be the case in the limit of an infinite time series, and it appears to be a good approximation for the GATE data, as discussed in section 3.

If there is no diurnal cycle, then under these simplifying assumptions, $\hat{s}$ and $\hat{c}$ are approximately independent normal random variables with zero mean and the same variance; $\hat{s}^2/(\text{var}\hat{s}) + \hat{c}^2/(\text{var}\hat{s})$ is approximately distributed as χ_2^2; and

$$\hat{r}_1^2 \sim \chi_2^2(\text{var}\hat{s}) = \frac{1}{2}(E_0\hat{r}_1^2)\chi_2^2, \qquad (2.14)$$

where the latter equality follows from $E_0\hat{r}_1^2 \approx \text{var}\hat{s} + \text{var}\hat{c}$. Here E_0 means expectation as described above but under the assumption that $c = s = 0$ in (2.10). It follows from (2.14), under this "null" hypothesis, that the probability of $\hat{r}_1^2$ being larger than some R^2 is given by

$$\Pr(\hat{r}_1^2 > R^2) = \exp[-R^2/(E_0\hat{r}_1^2)]. \qquad (2.15)$$

In order to compare an estimated value for $\hat{r}_1$ with its expectation in the absence of a diurnal cycle, it is necessary to estimate the variances of $\hat{s}$ and $\hat{c}$, which in turn depend on the variances of S and C in (2.13). From (2.12),

$$E_0(\hat{s}^2 + \hat{c}^2) = D^{-2}[(A_{cc}^2 + A_{sc}^2)E_0S^2 + (A_{sc}^2 + A_{ss}^2)E_0C^2 - 2(A_{cc} + A_{ss})A_{sc}E_0SC]. \qquad (2.16)$$

Equations for E_0S^2, E_0C^2, and E_0SC are derived in appendix B in terms of the lagged correlation of rainfall

$$c_R(\tau) = E[R'_A(t)R'_A(t+\tau)]/\sigma_A^2, \qquad (2.17)$$

where σ_A^2 is the variance of R_A and $R'_A(t)$ is the deviation of $R_A(t)$ from its mean. Zwiers and Hamilton (1986) suggest an alternative estimate based on an average over the spectral power at nearby frequencies.

Under our assumptions, we could, in principle, simplify the estimation of $E_0\hat{r}_1^2$ by only estimating $\text{var}_0\hat{s}$, but the version outlined here may be more reliable under slight deviations from these assumptions.

To proceed further in assessing whether or not TRMM will be able to detect the diurnal cycle, information about the variance and correlations required is needed in the equations above. In the next section, this methodology is applied to some tropical rainfall data, the GATE data. As these data have not to our knowledge been analyzed in this way before, it is hoped that the conclusions will be of interest themselves.

3. Diurnal cycle in the GATE data

The most widely studied tropical rainfall dataset is derived from an experiment carried out in the Atlantic off the west coast of Africa. Ships with meteorological radars took data almost continuously for several periods during the summer of 1974 as part of the GATE. (See Houze and Betts 1981.) The rainfall data were converted into quarter-hourly gridded maps covering an area 400 km in diameter centered on 8°30′N, 23°30′W, with each grid point representing the average

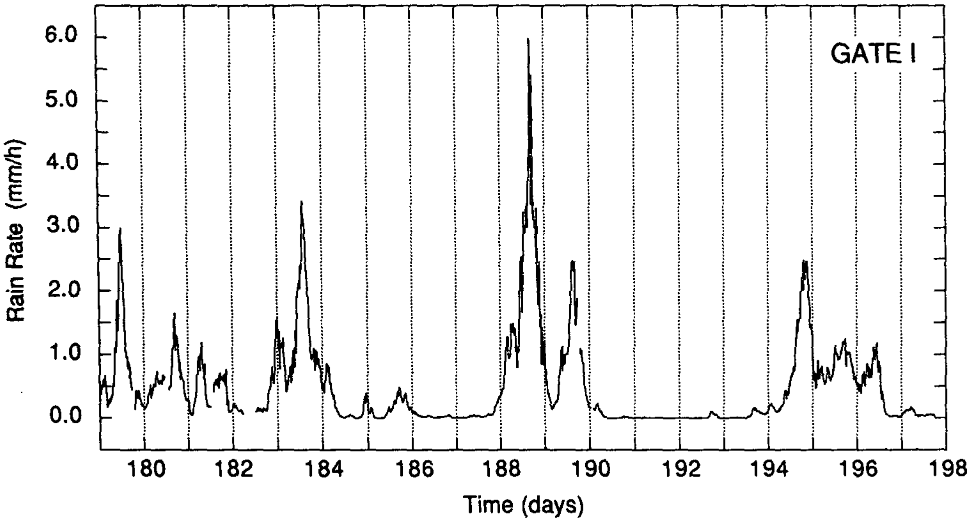

FIG. 2. Rain rate, averaged over a 280-km × 280-km area, for phase I (28 June–16 July 1974) of GATE versus Julian day. Vertical divisions are placed at 0000 UTC.

rain rate within a 4-km × 4-km area. The gridded dataset is described by Hudlow and Patterson (1979).

In order to represent the type of data to be collected by satellite, a time series of area-averaged rainfall was created by averaging the rainfall reported over a 280-km × 280-km area, the size of the largest square fitting within the circular GATE area. The data are available in 15-min intervals, but the data analyzed here use hourly samples. A time series of the data from phase I, 28 June–16 July (Julian days 179–197), is shown in Fig. 2. The gaps represent missing data; there are 14 missing hours in the record and 442 observations. The results did not seem to depend significantly on what interpolation method was used to fill in gaps, since the only large gap (6 h) occurs when there is little rain. The average rainfall $\hat{r}_0$ is 0.49 mm h^{-1}, and the estimated variance $\hat{\sigma}_A^2$ is 0.51 (mm h^{-1})^2. To compute $E_0\hat{r}_1^2$ in (2.14), lagged correlations are needed for lags τ equal to all time intervals between satellite observations. The lagged correlation was fitted to a simple exponential model $c_R(\tau) = \exp(-\tau/\tau_A)$, based on lags 1–20, giving $\hat{\tau}_A = 8.4$ h. A plot of the lagged correlations is shown in Fig. 3, and the exponential fit is shown as the solid line. These values agree with those obtained by Laughlin (1981), with the exception of $\hat{\tau}_A$, which Laughlin estimates as 7.2 h.

In order to separate the influence of σ_A^2 and $c_R(\tau)$ on $E_0\hat{r}_1^2$, write

$$E_0\hat{r}_1^2 = \sigma_A^2 f^2(N, \Delta t, \tau_A), \qquad (3.1)$$

where $f(N, \Delta t, \tau_A) = [E_0(\hat{s}^2 + \hat{c}^2)]^{1/2}/\sigma_A$. The dependence of f on the sampling interval and on the correlation time is expressed in (2.16) and (B.6)–(B.8) of appendix B.

The estimate of $\hat{\tau}_A$ is sensitive to the number of lags used, but the estimate of $f(N, \Delta t, \tau_A)$ is not. Figure 4 shows the function $f(N, \Delta t, \tau_A)$ plotted for $N\Delta t \approx 18$ days and $\tau_A = 7.2$, 8.4, and 10 h. The value $f = 0.165$, corresponding to $\tau_A = 8.4$ and $\Delta t = 1$ h, was used. The least-squares estimates of c and s in (2.10) are -0.137 and -0.176, respectively, leading to an estimated amplitude $\hat{r}_1 = 0.22$ mm h^{-1} and an estimated phase $\hat{\phi}_1$ of 1600 UTC. Figure 5a shows the hourly rainfall data with the daily averages imposed, and Fig. 5b shows the hourly averages and the estimated sine wave. Using the chi-squared approximation derived in section 2 and the estimated $\hat{\tau}_A = 8.4$, $\hat{\sigma}_A^2 = 0.51$, (2.15) shows that $\hat{r}_1$ is significantly different from zero at about $p = .03$. This is confirmed by the excellent fit to the averages shown in Fig. 5b, and confirms the conclusion of Albright et al. (1981). If Laughlin's estimate of $\hat{\tau}_K$

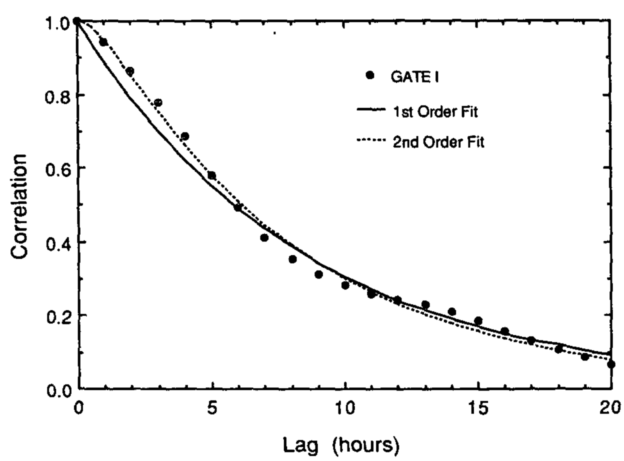

FIG. 3. Sample correlations $\hat{c}_r(\tau)$ plotted as a function of the lag. First-order exponential (solid) and second-order exponential (dashed) fits.

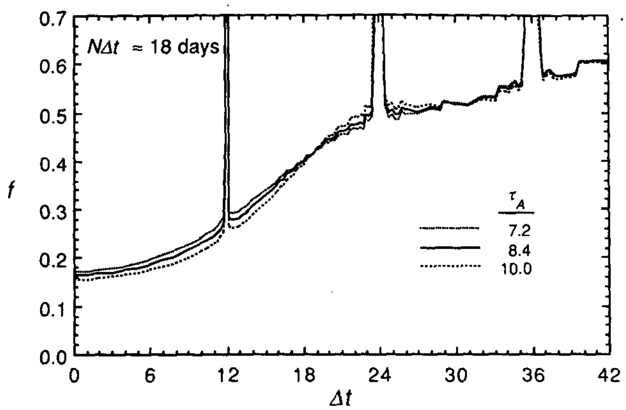

FIG. 4. Plot of $f(N, \Delta t, \tau_A)$ as a function of sampling interval Δt (h) for τ_A = 7.2 (· · ·), 8.4 (——), and 10 (- - -) h.

(7.2 h) is used, f changes to 0.172 but the p value changes only slightly.

To investigate the sensitivity of f to using an exponential fit for the correlation function, the correlation function is also fitted to that of a continuous second-order process,

$$c_R(\tau) = \frac{\tau_1 \exp(-\tau/\tau_1) - \tau_2 \exp(-\tau/\tau_2)}{\tau_1 - \tau_2}. \quad (3.2)$$

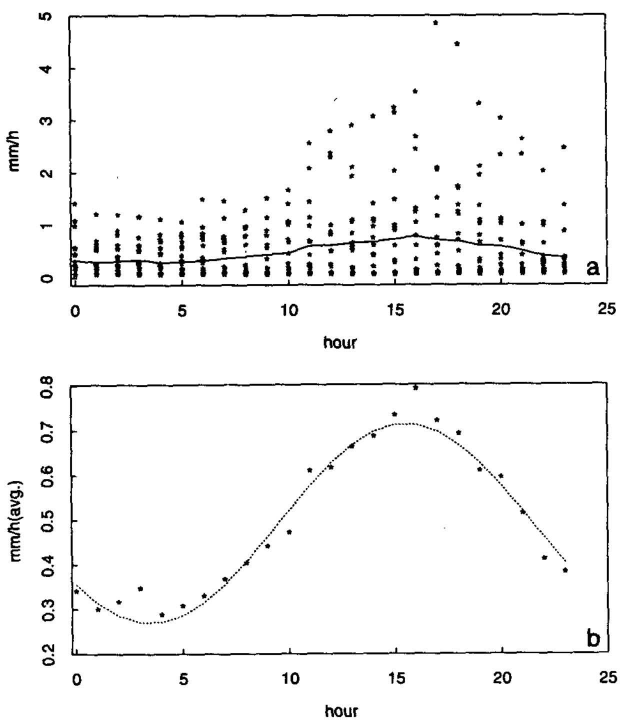

FIG. 5. (a) Hourly rain rates with hourly averages superimposed. (b) Hourly averages with fitted sine wave superimposed. (Times shown are relative to UTC. Local solar "noon" occurred at approximately 0130 UTC.)

Based on lags 1–20, the parameters of the least-squares fit are τ_1 = 7.56 h, τ_2 = 0.85 h. It is shown in Fig. 3 as the dotted line. The corresponding value of f was calculated to be 0.175 and therefore seems relatively insensitive to the choice between these two correlation functions.

As a partial check on some of the assumptions made in section 2, separate sinusoids were fitted by least squares (2.8) for each of the 18 days of GATE phase I. The original data at $\Delta t \approx 0.25$ h spacing were used. The average values and variances of the estimated coefficients are displayed in Table 1. There is a sample correlation of 0.33 between $\hat{c}$ and $\hat{s}$, casting some doubt on the assumption of independence of $\hat{c}$ and $\hat{s}$. The correlation only has a significance level of p = 0.16, however, using a standard test for the correlation of two normally distributed variables (n = 18 samples), described, for example, by Beyer (1968). The variances $E_0(\hat{c}^2) = 0.11$ and $E_0(\hat{s}^2) = 0.19$ can be predicted from equations analogous to (2.16) and agree well with the observed variances. Note that when the fits are obtained from only 1-day-long series, the expected variances of the coefficients $\hat{c}$ and $\hat{s}$ differ somewhat surprisingly by almost a factor of 2. The difference is *not* due to sampling error alone. The chi-squared distribution assumed in (2.14) would not be valid for such short samples.

There was exceptionally high rainfall on day 188 of phase I at 1600 and 1700 UTC (4.8 and 4.4 mm h^{-1}, respectively). As a check on the influence of this on the estimated diurnal cycle, the full least-squares regression is fit to the time series excluding these two observations. The estimated amplitude $\hat{r}_1$ is now 0.19 mm h^{-1}, with an observed significance level of about 0.09. Although in this case the observed outliers are valid observations, the sensitivity of the significance level to these observations is worrisome if there is the possibility that the conversion of radar intensity data to rainfall rates may result in the occasional spurious large observation.

The data from phase II of the GATE study was also analyzed by the same techniques as described above. This data covers 19 days from 26 July to 13 August 1974. A plot of the data is shown in Fig. 6, and the hourly values and superimposed average are shown in Fig. 7. Because there is a large gap between Julian days 211 and 215, only the last 13 days are used in our analysis. Somewhat surprisingly, the evidence for a diurnal cycle is very weak in the dataset. The estimated

TABLE 1. Daily least-squares fits for GATE phase I (18 days).

	Mean (mm h^{-1})	Variance (mm h^{-1})^2
c	−0.12	0.10
s	−0.19	0.17
$r_1 = (c^2 + s^2)^{1/2}$	0.23	—
$\phi_1 = \tan^{-1}(s/c)$	15.9 h	—

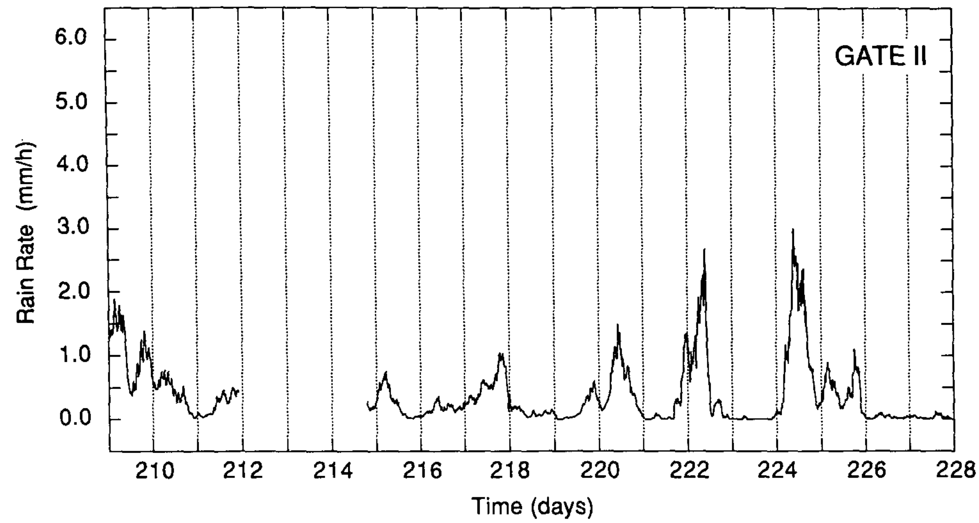

FIG. 6. Area-averaged rain rate for phase II of GATE: 26 July–13 August 1974.

amplitude is just 0.11 mm h^{-1}, with rainfall peaking at 0900 UTC. The phase estimate is not really relevant, given the weakness of the signal. The 13 phase estimates can be thought of as observations on a circle, and a simple cosine test for uniformity (Cox and Hinkley 1974, p. 67) has a p value of about 0.27.

The value of $E_0\hat{r}_1^2$ is more sensitive to the assumed value for σ_A^2 than τ_A. The estimate 0.51 (mm h^{-1})^2 obtained from the phase I data has been used, although it would be preferable to have an independent estimate of σ_A^2 for assessing the effect of the diurnal cycle. The estimated variance goes down to 0.42 (mm h^{-1})^2 if the two largest observations are deleted, so the significance level reported above is probably conservative. McGarry and Reed (1978) combined the phase II series with data from phase III (30 August–19 September 1974) and found some evidence for a diurnal cycle peaking at about 1530 UTC. This suggests the possibility that there is some nonstationarity in the diurnal cycle through this summer.

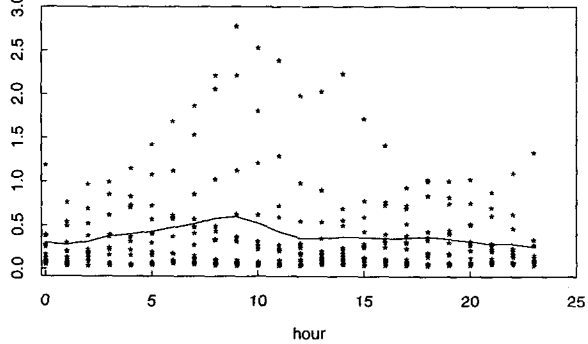

FIG. 7. Hourly rain rates for phase II of GATE, with hourly average superimposed.

It is straightforward to include a semidiurnal cycle in the least-squares fitting procedure, replacing (2.10) with

$$r(t) \approx r_0 + c_1 \cos\omega t + s_1 \sin\omega t + c_2 \cos 2\omega t + s_2 \sin 2\omega t. \quad (3.3)$$

The estimated amplitude of the semidiurnal component was negligible in both phases of the GATE data. It is possible that a signal might be detectable from a longer series or a larger-area average and, in this case, it is of interest to note that $r_2^2 = s_2^2 + c_2^2$ is estimated independently of r_1^2.

The plot of $f(N, \Delta t, \tau_A) = [E_0(\hat{s}^2 + \hat{c}^2)]^{1/2}/\sigma_A$ given in Fig. 4 is equivalent to a theoretical version of the periodogram (sample spectrum), but with the sampling frequency Δt varying instead of the cyclical frequency of the time series. The spectrum of the GATE data does not, however, indicate a large-frequency component at 24 and 12 h, as might be hoped. This is probably a reflection of the fact that the average diurnal fluctuation is relatively small (0.22 mm h^{-1}) and the variance of the residuals after this fluctuation has been accounted for is still large. Of course the main objective in this section is not to model the rainfall series with a sum of Fourier components but rather to investigate the effect of area averaging involved in satellite observation on our ability to detect a signal at a fixed frequency, expected a priori from physical arguments.

As discussed earlier, the chi-squared approximation (2.14) for the distribution of the diurnal amplitude under the null hypothesis is assumed to be adequate for testing the significance of a diurnal fit. This is not entirely obvious, since individual rain rates (even averaged over the GATE area) are far from being normally distributed, and the sums in (2.13) may not contain a sufficient number of samples for the central-limit theorem to be applicable.

The distribution of area-averaged rain rate $R_A(t)$ itself is not easy to predict. Large values of R_A occur with a frequency far greater than a Gaussian or exponential distribution would predict. A lognormal distribution is sometimes postulated for large rain rates, and in fact for many other properties of rainfall as well (Crane 1985, 1986; Biondini 1976; Crow et al. 1979; Houze and Cheng 1977; Lopez 1977; Kedem et al. 1990). That the distribution of $R_A(t)$ for the GATE data has a very strong tail is evident from the frequency plot in Fig. 8, where the GATE rainfall data for the 18-day phase I period and the 15-day phase II period are histogrammed with logarithmic bins of size $\Delta \ln r_A = 0.3$. Because the data are strongly correlated in time, however, it is difficult to gauge the size of sampling errors in the histogram values.

To get some impression of the sort of sampling error present, a two-dimensional stochastic model of rainfall (Bell 1987; Bell et al. 1990) with point statistics and spatial and temporal correlations adjusted to be as close as possible to those of GATE phase I was used to simulate 16 successive 18-day periods, and the histograms for each 18-day period were computed. The mean and standard deviation of frequencies for each bin is superimposed on the GATE data in Fig. 8 and gives some impression of the possible amount of variability due to sampling. The distribution of rain rates greater than 0.12 mm h^{-1} in the model, which contribute more than 95% of the accumulated rainfall in the model, is fit quite well by a lognormal distribution. It is, however, clear that the lognormal distribution would not fit the

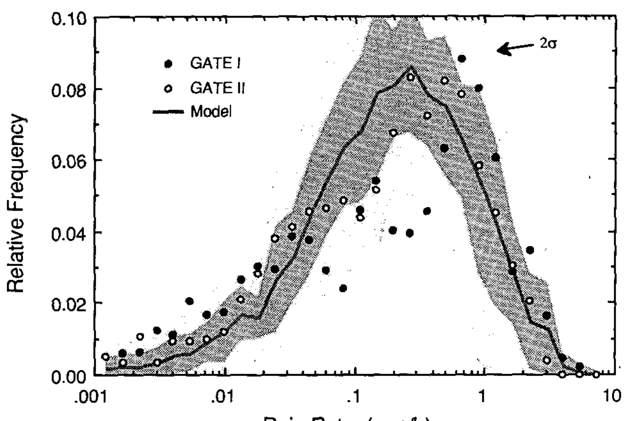

FIG. 8. Histogram of rain rate $R_A(t)$ averaged over 280-km × 280-km area plotted as fraction of time ln R_A falls within bins of width 0.3. The solid and open circles show GATE phases I and II rain-rate frequencies, respectively (recorded quarter-hourly). The smooth curve at the center of the stippled region is the average of 16 model simulations, each 18 days long, and the stippled regions denote the one and two standard-deviation limits estimated from the 16 simulated periods for each bin.

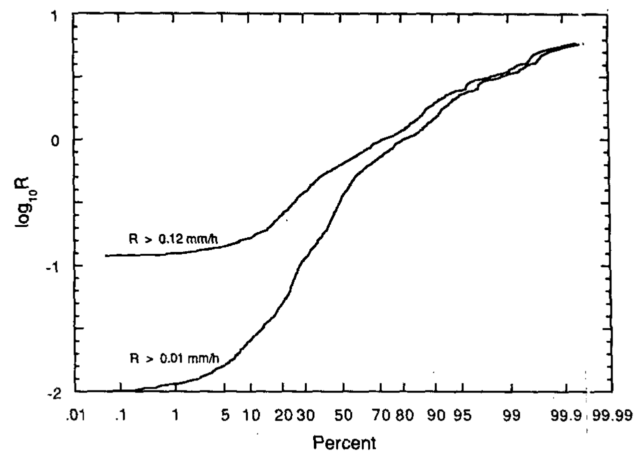

FIG. 9. Lognormal probability plots for phase I data, recorded quarter-hourly: one for rain rates $R_A > 0.01$ mm h^{-1}, the other for $R_A > 0.12$ mm h^{-1}. A straight line would indicate a lognormal distribution.

distribution of rain rates over the entire range of values displayed in Fig. 8.

Lognormal probability plots for the phase I data are shown in Fig. 9. Two curves are shown, one for the cumulative distribution of rain rates $R_A > 0.01$ mm h^{-1} and the other for $R_A > 0.12$ mm h^{-1}. (Over 99.9% of the total rain volume that fell in GATE phase I is due to rain rates $R_A > 0.01$ mm h^{-1}.) The plots use the quarter-hourly data; the hourly plots are very similar. The curve for $R_A > 0.01$ mm h^{-1} indicates that the area-averaged rates have a shorter right-hand tail and longer left-hand tail than the lognormal distribution would predict. This is consistent with Jensen's inequality if the rain rate $R(\mathbf{x}, t)$ at a fixed point in time and space follows a lognormal distribution. By increasing the cutoff to 0.12 mm h^{-1}, the fit to a lognormal distribution improves, as indicated by the other curve in Figs. 8 and 9.

Given the highly skewed distribution of R_A, then, there is reason for concern that the length of the time series used is sufficiently long enough to assure convergence of the statistics to the approximate chi-squared one proposed in (2.14). To explore this assumption for the GATE data, a simulated time series of hourly GATE rain rates has been generated using the stochastic model mentioned above, which reproduces the temporal correlations and highly skewed distribution of area-averaged rain rate but has no intrinsic diurnal cycle. Diurnal fits to 200 segments each 6 days long were obtained, and the probability distribution of the quantity

$$\rho^2 = \hat{r}_1^2(\overline{\hat{r}_1^2}/2)^{-1} \qquad (3.4)$$

is compared with the χ_2^2 distribution in Fig. 10. It can be seen that the chi-squared distribution is in rough agreement with the actual distribution over the usual

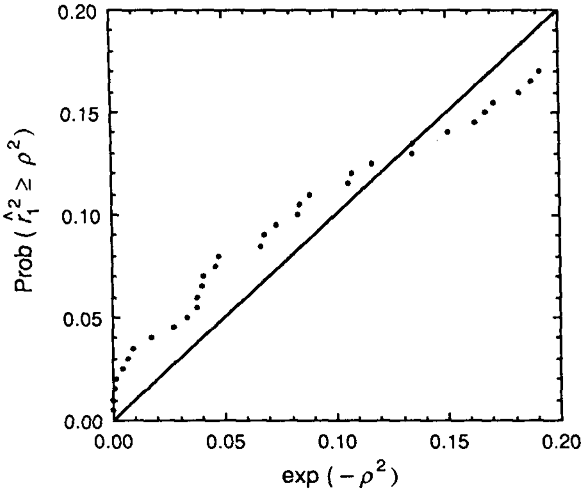

FIG. 10. Comparison of the χ_2^2 distribution with the distribution of the diurnal amplitude fits $\hat{r}_1$ to 200 6-day samples of simulated data with GATE-like distribution and correlation time but no diurnal cycle.

significance levels but provides a very poor representation of the extreme tail of the distribution. The chi-squared approximation is of course better for GATE-length time series (18 days) and is still better when applied to 30-day time series of rain rates averaged over the 500-km × 500-km areas of interest in the TRMM experiment.

4. Implications for TRMM

The estimated amplitude $\hat{r}_1$ obtained from the phase I data is significantly different from 0 using the χ^2 test derived in section 2, but the evidence is not overwhelming, and the apparent lack of a significant signal in phase II is somewhat worrying. The analysis of the diurnal cycle of convective cloudiness over tropical Africa and surrounding oceans by Desbois et al. (1989) shows a diurnal cycle that is strong over the continent and persists weakly out to the GATE area; the cycle may therefore be stronger there than over more remote oceanic regions. The easterly waves appear to extend to the GATE area (e.g., Albright et al. 1981). Nevertheless, assume that a diurnal cycle as large as 0.22 mm h^{-1} may occur, and ask whether a satellite sampling roughly every 12 to 24 h would be able to detect it.

Figure 11 shows the critical factor f for $N\Delta t \approx 30$ days and correlation time $\tau_A = 8.4$ h. When sampling takes place every 12 h, but differently enough from 12 h that the entire diurnal cycle is observed during 1 month, we find $f \approx 0.2$. Under the assumptions that rainfall will be averaged over an area of size 600 × 600 km^2, and that the variance σ_A^2 decreases inversely as the area so that it is about one-fourth the value found in GATE, we could conclude that a 5% critical value for $\hat{r}_1$ is 0.12 mm h^{-1}; that is, that a diurnal cycle whose amplitude is about 25% of the mean would be detectable. Averaging over longer periods or greater areas would further reduce the critical value for detecting the diurnal cycle. Because the time period considered is longer than the 18 days of GATE phase I, the sampling interval Δt is longer; this result for the satellite's performance is probably less sensitive to our assumptions than was the result for the GATE diurnal cycle. The diurnal cycle seen in phase I of GATE should be easily seen in a month of data taken by the satellite.

5. Discussion

Within the context of a statistical model, we have calculated the probability that the changes in the mean rainfall in GATE with the time of day could have been due to chance, given the small number of days for which data are available. The amplitude of the first harmonic was found to be significant at the 0.03 level during phase I but not to be significant at the 0.05 level in phase II. The methods used could be applied to the kind of data TRMM would produce, and a diurnal cycle like the one seen in GATE could be detected with a month of such data averaged over a 5° latitude–longitude box.

There are, however, a number of issues not addressed by the statistical method used that deserve more attention and about which we would like to offer our observations here.

(a) The lagged correlations used do not capture longer time-scale phenomena that are known to occur over GATE and over the tropical oceans in general. The easterly waves that pass over the GATE region at intervals of about 4 or 5 days create large-scale conditions favorable to convective activity. These pulses of activity are clearly visible in the time series (Fig. 2). Phase I shows a rise in autocorrelation at lag 5 days that appears significant in spite of the shortness of the

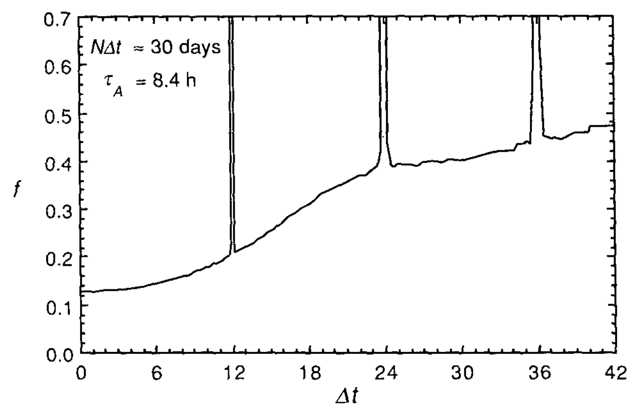

FIG. 11. Plot of $f(N, \Delta t, \tau_A)$ as a function of sampling interval Δt (h) for $\tau_A = 8.4$ h.

time series. A crude attempt to introduce this additional correlation at lag 5 days in the calculation of the critical level for detecting a diurnal cycle with TRMM sampling at the equator produced very little change in the critical level. More data with which to generate better estimates of the autocorrelation out to many days' lag are needed to pursue this further.

At an even longer time scale, the 30–60-day oscillation (Madden and Julian 1971; Lau and Shen 1988) could cause significant aliasing of power at this frequency into the diurnal cycle for TRMM observations at higher latitudes, where the time of day of the overflights shifts slowly during a month (see Fig. 1). Solutions to this may require either combining months of data from several years or interpolating between TRMM measurements using observations from other satellites.

(b) The effects of measurement error on the weighting of the observations used to obtain the least-squares fit have not been taken into account. They may be less important than the sampling errors mainly dealt with here because we have used averages over large spatial and temporal domains, and sampling error seems likely to dominate measurement error in such averages, as discussed, for example, by Bell et al. (1990).

(c) Our estimates used equally spaced observations and assumed the entire area A was viewed by TRMM at each pass. The calculations of $E(S^2)$, $E(C^2)$, and $E(SC)$ in appendix B could be carried out for irregularly sampled data but would require a model for the lagged covariance of partial area averages $R_{A_1}(t_1)$ and $R_{A_2}(t_2)$ (the area-averaged rain rates over area A_1 at time t_1 and area A_2 at time t_2, where the A_i are the portions of A observed at the times t_i). This would be possible in principle if the statistics are sufficiently homogeneous over the area A.

(d) Once a detectable diurnal cycle is found, one would like to know the confidence limits for $\hat{r}_i$ and $\hat{\phi}_i$; that is, how well can one determine the amplitude and phase of the cycle for a given amount of data? A rough estimate of these limits may be obtained by treating the diurnal cycle amplitude as a vector $(\hat{c}, \hat{s})$, with the 95% confidence "interval" being the circle centered on the end of the vector $(\hat{c}, \hat{s})$ in its two-dimensional vector space, whose radius is estimated from the null hypothesis estimate (2.15) for the desired confidence level. We have not been able to discover a discussion in the statistical literature for the confidence limits of $\hat{\phi}$ alone relevant to our problem. Resampling strategies, taking into account the problems of temporal correlation (see, for example, Zwiers 1990), may be helpful here.

(e) Data from other satellites may be necessary for resolving some of the aliasing problems at higher latitudes. These will be less accurate, though, and appropriate weighting schemes for a least-squares treatment will be needed.

(f) A basic problem that needs more attention is that the diurnal cycle seems to be the result of a combination of influences, one of them certainly the diurnal heating by the sun, but also the current synoptic situation in the region determining large-scale convergence in the lower troposphere. This is certainly the case in GATE. There is a nonlinear interaction between these two time scales, as discussed in Albright et al. (1981). As McGarry and Reed (1978) point out, the diurnal cycle is much clearer during the passage of easterly wave troughs. This interaction is not described well by the linear model used here. Although our linear approach can be justified for long time series, more sensitive methods of establishing the diurnal cycle might be developed if based on models that are physically motivated. This problem of describing the waxing and waning of rain over many time scales is an aspect of rainfall that bedevils all efforts to deal with it statistically: it is a combination of dry periods and areas, with a finite probability of occurring, and wet periods with highly skewed statistics.

Acknowledgments. This paper had its origins in part from material presented at the American Statistical Association meeting in San Francisco, August 1987, in a session entitled "Statistical Problems in Climatology." We are grateful to F. Zwiers for organizing the session. We would like to thank L. S. Chiu, D. R. Cox, A. Feuerverger, D. A. S. Fraser, B. Kedem, G. R. North, S. Ravipati, and D. A. Short for helpful comments. Research support from the Natural Sciences and Engineering Research Council Canada is also acknowledged.

APPENDIX A

Fitting Hourly Averages Versus Fitting Time Series to Daily Harmonics

Rewrite $r(t)$ in (2.4) so that the parameters to be fit appear linearly:

$$r(t) = r_0 + c_1 \cos\omega t + s_1 \sin\omega t + \cdots$$
$$= \sum_{\alpha} u_\alpha f_\alpha(t), \qquad (A.1)$$

so that the unknown coefficients u_α are r_0, c_1, s_1, etc., and the functions $f_\alpha(t)$ are 1, $\cos\omega t$, $\sin\omega t$, etc. Then the total square differences (2.6), with choice (2.7), can be written

$$D^2 = \sum_t [\bar{R}(t) - \sum_\alpha u_\alpha f_\alpha(t)]^2 n_t, \qquad (A.2)$$

and the linear equations for u_α that minimize (A.2) are

$$\sum_{\alpha'} [\sum_t f_\alpha(t) f_{\alpha'}(t) n_t] u_{\alpha'} = \sum_t \bar{R}(t) f_\alpha(t) n_t, \qquad (A.3)$$

which are obtained by requiring the derivatives of (A.2) with respect to u_α to vanish. Substitute (2.2) and (2.3) into (A.3) to obtain

$$\sum_{\alpha'} [\sum_{t} \sum_{t_m \in t} f_\alpha(t) f_{\alpha'}(t)] u_{\alpha'} = \sum_{t} \sum_{t_m \in t} R_A(t_m) f_\alpha(t). \quad (A.4)$$

If the intervals t are small enough that the variations of $f_\alpha(t')$ over the interval can be neglected, then the sums $\sum_t \sum_{t_m \in t}$ can be replaced by a sum over all sample times, and (A.4) can be rewritten as

$$\sum_{\alpha'} [\sum_{m=1}^{N} f_\alpha(t_m) f_{\alpha'}(t_m)] u_{\alpha'} = \sum_{m=1}^{N} R_A(t_m) f_\alpha(t_m), \quad (A.5)$$

which is exactly the equation obtained by minimizing (2.8). This equivalence becomes more nearly exact as the number of samples becomes large and the size of the intervals t in (2.2) can be made small.

APPENDIX B

Details of Calculations of $E(S^2)$, $E(C^2)$, and $E(SC)$ in (2.16)

The variance of S defined in (2.13a) is given by

$$E(S^2) = \sum_{m_1=0}^{N} \sum_{m_2=0}^{N} \{\sin(\omega t_{m_1}) \sin(\omega t_{m_2}) \times E[R'_A(t_{m_1}) R'_A(t_{m_2})]\}. \quad (B.1)$$

The lagged covariance of rainfall $R'_A(t)$ (primes indicating deviation from the mean) is

$$E[R'_A(t) R'_A(u)] = \sigma_A^2 c_R(|t-u|),$$

where σ_A^2 is the variance of R_A and $c_R(\tau)$ its lagged correlation. Use a trigonometric identity for sin$A$ sin$B$ to write (B.1) as

$$E(S^2) = \frac{\sigma_A^2}{2} \sum_{m_1=0}^{N} \sum_{m_2=0}^{N} \{\cos[\omega \Delta t(m_1 - m_2)] - \cos[\omega \Delta t(m_1 + m_2)]\} c_R(|m_1 - m_2| \Delta t). \quad (B.2)$$

Change summation variables from (m_1, m_2) to (w, u) using the identity

$$\sum_{m_1=0}^{N} \sum_{m_2=0}^{N} f(m_1 + m_2) g(m_1 - m_2) = \sum_{u=-N}^{N} \sum_{w=0}^{N-|u|} f(|u| + 2w) g(u).$$

Equation (B.1) becomes

$$E(S^2) = \frac{\sigma_A^2}{2} \sum_{u=-N}^{N} \sum_{w=0}^{N-|u|} \{\cos(u\omega\Delta t) - \cos[(|u| + 2w)\omega\Delta t]\} c_R(u\Delta t). \quad (B.3)$$

Define

$$V(u) = \sum_{w=0}^{N-|u|} \cos[(|u| + 2w)\omega\Delta t], \quad (B.4)$$

which can be explicitly summed to

$$V(u) = \frac{\sin[(N+1-|u|)\omega\Delta t]}{\sin(\omega\Delta t)} \cos(N\omega\Delta t). \quad (B.5)$$

Substituting this into (B.3) gives

$$E(S^2) = \frac{\sigma_A^2}{2} \sum_{u=-N}^{N} [(N+1-|u|) \times \cos(u\omega\Delta t) - V(u)] c_R(u\Delta t). \quad (B.6)$$

Similar techniques can be used to obtain

$$E(C^2) = \frac{\sigma_A^2}{2} \sum_{u=-N}^{N} [(N+1-|u|) \times \cos(u\omega\Delta t) + V(u)] c_R(u\Delta t) \quad (B.7)$$

and

$$E(SC) = \frac{\sigma_A^2}{2} \sum_{u=-N}^{N} W(u) c_R(u\Delta t), \quad (B.8)$$

with

$$W(u) = \sum_{w=0}^{N-|u|} \sin[(|u| + 2w)\omega\Delta t], \quad (B.9)$$

which can be explicitly summed to

$$W(u) = \frac{\sin[(N+1-|u|)\omega\Delta t]}{\sin(\omega\Delta t)} \sin(N\omega\Delta t). \quad (B.10)$$

Note that the coefficients in (2.12) can be expressed in terms of the sums (B.4) and (B.9) as

$$A_{ss} = \frac{1}{2}[N + 1 - V(0)],$$

$$A_{cc} = \frac{1}{2}[N + 1 + V(0)],$$

and

$$A_{sc} = \frac{1}{2} W(0).$$

REFERENCES

Albright, M. D., D. R. Mock, E. E. Recker, and R. J. Reed, 1981: A diagnostic study of the diurnal rainfall variation in the GATE B-scale area. *J. Atmos. Sci.*, **38**, 1429–1445.

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

Augustine, J. A., 1984: The diurnal variation of large-scale inferred rainfall over the tropical Pacific Ocean during August 1979. *Mon. Wea. Rev.*, **112**, 1745–1751.

Bell, T. L., 1987: A space–time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res.*, **92D**, 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.*, **95D**, 2195–2205.

Beyer, W. H., 1968: *Handbook of Tables for Probability and Statistics*, 2d ed. CRC Press, 642 pp.

Biondini, R., 1976: Cloud motion and rainfall statistics. *J. Appl. Meteor.*, **15**, 205–224.

Cox, D. R., and D. V. Hinkley, 1974: *Theoretical Statistics*. Chapman and Hall, 511.

Crane, R. K., 1985: Evaluation of global and CCIR models for estimation of rain rate statistics. *Radio Sci.*, **20**, 865–879.

——, 1986: Horizontal small scale structure of precipitation. *Proc. 23rd Conf. on Radar Meteorology*, Snowmass, Colorado, Amer. Meteor. Soc., 181–184.

Crow, E. L., A. B. Long, J. E. Dye, and A. J. Heymsfield, 1979: Results of a randomized hail suppression experiment in Northeast Colorado. Part II: Surface data base and primary statistical analysis. *J. Appl. Meteor.*, **18**, 1538–1558.

Desbois, M., T. Kayiranga, and B. Gnamien, 1989: Diurnal cycle of convective cloudiness over tropical Africa observed from Meteosat: Geographic characterization and interannual variations. *Ann. Geophys.*, **7**, 395–404.

Fu, R., A. D. Del Genio, and W. B. Rossow, 1990: Behavior of deep convective clouds in the tropical Pacific deduced from ISCCP radiances. *J. Climate*, **3**, 1129–1152.

Gray, W. M., and R. W. Jacobson, Jr., 1977: Diurnal variation of deep cumulus convection. *Mon. Wea. Rev.*, **105**, 1171–1188.

Griffith, C. G., W. L. Woodley, P. G. Grube, D. W. Martin, J. Stout, and D. N. Sikdar, 1978: Rain estimation from geosynchronous satellite imagery—Visible and infrared studies. *Mon. Wea. Rev.*, **106**, 1153–1171.

Hamilton, K., 1981: A note on the observed diurnal and semi-diurnal rainfall variation. *J. Geophys. Res.*, **86C**, 12122–12126.

Houze, R. A., and C.-P. Cheng, 1977: Radar characteristics of tropical convection observed during GATE: Mean properties and trends over the summer season. *Mon. Wea. Rev.*, **105**, 964–980.

——, and A. K. Betts, 1981: Convection in GATE. *Rev. Geophys. Space Phys.*, **19**, 541–576.

Hudlow, M. D., and V. L. Patterson, 1979: GATE Radar Rainfall Atlas. NOAA Special Rep., 158 pp. [Available from U.S. Government Printing Office, Washington, DC 20402].

Kedem, B., L. S. Chiu, and G. R. North, 1990: Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res.*, **95D**, 1965–1972.

Lau, K.-M., and S. Shen, 1988: On the dynamics of intraseasonal oscillations and ENSO. *J. Atmos. Sci.*, **45**, 1781–1797.

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

Lopez, R. E., 1977: The lognormal distribution and cumulus cloud populations. *Mon. Wea. Rev.*, **105**, 865–872.

Madden, R., and P. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708.

McGarry, M. M., and R. J. Reed, 1978: Diurnal variation in convective activity and precipitation during Phases II and III of GATE. *Mon. Wea. Rev.*, **106**, 101–113.

Press, W. H., and S. A. Teukolsky, 1988: Search algorithm for weak periodic signals in unevenly spaced data. *Comp. Phys.*, **2**, 77–82.

Randall, D. A., Harshvardhan, and D. A. Dazlich, 1991: Diurnal variability of the hydrologic cycle in a general circulation model. *J. Atmos. Sci.*, **48**, 40–62.

Shin, K.-S., and G. R. North, 1988: Sampling error study for rainfall estimate by satellite using a stochastic model. *J. Appl. Meteor.*, **28**, 1218–1231.

——, ——, Y.-S. Ahn, and P. A. Arkin, 1990: Time scales and variability of area-averaged tropical oceanic rainfall. *Mon. Wea. Rev.*, **118**, 1507–1516.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed tropical rainfall measuring mission satellite. *Bull. Amer. Meteor. Soc.*, **69**, 278–295.

Wallace, J. M., 1975: Diurnal variations in precipitation and thunderstorm frequency over the conterminous United States. *Mon. Wea. Rev.*, **103**, 406–419.

Zwiers, F. W., 1990: The effect of serial correlation on statistical inferences made with resampling procedures. *J. Climate*, **3**, 1452–1461.

——, and K. Hamilton, 1986: Simulation of solar tides in the Canadian Climate Centre general circulation model. *J. Geophys. Res.*, **91D**, 11 877–11 896.