Analysis of Darwin Rainfall Data: Implications on Sampling Strategy

QIHANG LI, RAFAEL L. BRAS, AND DANIELE VENEZIANO

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 28 November 1994, in final form 28 June 1995)

ABSTRACT

Rainfall data collected by radar in the vicinity of Darwin, Australia, have been analyzed in terms of their mean, variance, autocorrelation of area-averaged rain rate, and diurnal variation. It is found that, when compared with the well-studied GATE (Global Atmospheric Research Program Atlantic Tropical Experiment) data, Darwin rainfall has larger coefficient of variation (CV), faster reduction of CV with increasing area size, weaker temporal correlation, and a strong diurnal cycle and intermittence. The coefficient of variation for Darwin rainfall has larger magnitude and exhibits larger spatial variability over the sea portion than over the land portion within the area of radar coverage. Stationary and nonstationary models have been used to study the sampling errors associated with space-based rainfall measurement. The nonstationary model shows that the sampling error is sensitive to the starting sampling time for some sampling frequencies, due to the diurnal cycle of rain, but not for others. Sampling experiments using data also show such sensitivity. When the errors are averaged over starting time, the results of the experiments and the stationary and nonstationary models match each other very closely. In the small areas for which data are available for both Darwin and GATE, the sampling error is expected to be larger for Darwin due to its larger CV.

1. Introduction

At present, satellite-based remote sensing is the only viable way to obtain rainfall information on a global scale. One of the objectives of satellite-based rainfall measuring activities is to estimate rainfall aggregates on temporal and spatial scales of climatological interest, say, over areas on the order of 10^5 km^2 and for periods on the order of one month. Errors associated with this type of rainfall estimation include measurement and sampling errors. Measurement error is caused by the brightness temperature to rain-rate conversion algorithm, while the sampling error is caused by the discrete visits to a particular area by the orbiting satellite. In the proposed Tropical Rainfall Measuring Mission (TRMM), for example, the low-altitude, low-inclination satellite flies over a given area at a frequency of about twice daily (Simpson et al. 1988). The sampling error is a function of the satellite geometry (altitude, inclination, swath width, etc.) and the statistical properties of rainfall. Satellite geometry will determine how frequently a given area is visited and what portion of the area is intersected by the swath during each visit. Statistical properties of rainfall that affect the sampling error include the mean, variance, and autocorrelation and the variation of these with time. While the satellite geometry can be controlled through proper design, the rainfall statistics are location dependent and need to be quantified by observation.

To define the sampling error, consider a rain-rate process $r(\mathbf{x}, t)$. The rain rate averaged over an area of interest A is

$$r_A(t) = \frac{1}{A}\int_A r(\mathbf{x}, t)d\mathbf{x}. \tag{1}$$

Further averaging over a period of time T gives the space–time average

$$r_{AT} = \frac{1}{AT}\int_T\int_A r(\mathbf{x}, t)d\mathbf{x}dt. \tag{2}$$

Suppose that during period T the satellite makes N complete snapshots of area A with perfect rainfall measurements. The r_{AT} may be estimated as the sample average

$$\hat{r}_{AT} = \frac{1}{N}\sum_{i=1}^{N} r_A(t_i), \tag{3}$$

where $t_1, t_2, \ldots, t_N$ are times of visit.

The estimation error is

$$\epsilon = \hat{r}_{AT} - r_{AT}. \tag{4}$$

Our interest is in the "sampling error," which is defined here as the standard deviation of ϵ,

$$\sigma_\epsilon = [\langle(\hat{r}_{AT} - r_{AT})^2\rangle]^{1/2}. \tag{5}$$

There have been a number of studies in which sampling errors associated with TRMM-type rainfall mea-

Corresponding author address: Dr. Qihang Li, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Rm. 48-208, Cambridge, MA 02139.
E-mail: hang@athena.mit.edu

surements have been quantified. These studies have used rainfall statistics derived from GATE [Global Atmospheric Research Program (GARP) Atlantic Tropical Experiment], in which rainfall was measured from radars aboard ships in the intertropical convergence zone area of the Atlantic (centered at 8.5°N, 23.5°W) during the summer of 1974. The GATE consists of an 18-day phase 1 experiment and a 15-day phase 2 experiment. The rain data are spatially binned into a 100 × 100 array of 4-km-square boxes. In McConnell and North (1987), GATE rainfall rates are divided into roughly equal volume contribution categories: 0–5, 5–10, 10–20, and greater than 20 mm h^{-1}, and each category is studied separately. An imaginary satellite is flown over the 400 km × 400 km GATE area every 650 min. An ensemble of nine estimates of rain rate is developed for each rainfall category and for each GATE phase. The first member of the ensemble is obtained by starting the first visit at 0800 LST on the first morning of a GATE phase, returning 650 min later, and continuing through the entire period of that phase; similarly, the second member of the ensemble is started at 0900 LST, etc. The nine estimates for each category and each phase are then compared with the true rain rate. The study found that for each category, the bulk of the estimates are within ±10% of the true averages, with the spread being larger for the more intense categories. McConnell and North (1987) concluded that the result is consistent with earlier studies and that their findings can be expected to be stronger, in the sense that some cancellation of random sampling errors may occur when the four intensity categories are put together to obtain the total rain rate. McConnell and North (1987) acknowledged, however, that two factors could invalidate their conclusion that percentage errors of 5%–10% in estimating monthly rainfall over 500-km grid boxes are achievable by twice-daily snapshots: 1) GATE data cover only two relatively short periods and 2) GATE data might not be representative of other tropical areas.

Bell et al. (1990) studied the sampling error problem using a stochastic model of space–time rainfall. The model parameters were determined using the space–time statistics of GATE rainfall. Rainfall was generated over a 512 km × 512 km area, and a satellite was flown to visit the area about once every 12 h surveying the area either completely or partially. The Monte Carlo study showed that for area latitudes of 10°–25°, orbital altitudes of 300–450 km, and an inclination of 30°, the sampling error for monthly rainfall is 6%–10% of the mean. In their conclusions, they pointed out that the model is wholly based on GATE statistics from one 400-km-diameter spot and that even GATE data themselves are subject to considerable uncertainty. Furthermore, the estimates of sampling error do not include the effect of diurnal and spatial variation of the statistics within an averaging box.

North et al. (1993) used a simple rainfall model, in which the stochastic rainfall process $\psi(\mathbf{r}, t)$ is governed by the differential equation

$$\tau_0 \frac{\partial \psi(\mathbf{r}, t)}{\partial t} - \lambda_0^2 \nabla^2 \psi(\mathbf{r}, t) + \psi(\mathbf{r}, t) = F(\mathbf{r}, t), \quad (6)$$

where τ_0 and λ_0 are inherent time and length scales of the rainfall process, and $F(\mathbf{r}, t)$ is a zero-mean noise. This model represents a first-order continuous autoregressive process in time and an isotropic second-order autoregressive process in space. The reason that this model was chosen is that it is easy to analyze spectrally, and it is reasonably accurate in describing the GATE data. Using this model, a 350-km orbit altitude, 35° inclination, and 600-km swath, the sampling error associated with the estimation of monthly rainfall over a 500 km × 500 km area near the equator was estimated to be 11.2% of the mean. When complete coverage is assumed for each visit, the error reduces to 5.3%. In their study, the coefficient of variation of the area-averaged rain rate, defined as the standard deviation divided by the mean, was estimated to be 1 by extrapolating the available data to a 500 km × 500 km area.

Oki and Sumi (1994) used rain gauge and radar data over Japan to estimate the sampling error of rainfall retrieved by a TRMM-type satellite. Their results showed a somewhat larger sampling error than that found using GATE data.

The estimate of sampling error apparently depends on the rainfall statistics or rainfall model used. The rainfall models used in earlier studies are stationary, without diurnal cycles. Indeed, there is little evidence of diurnal cycle in the GATE data. For example, Fig. 5 of Bell et al. (1990) shows no enhanced correlation at a lag time of 24 h. In this study, we analyze radar rainfall data collected in the vicinity of Darwin, Australia, in terms of the mean, variance, temporal correlation, and diurnal variation and study the sampling error associated with such rainfall statistics. At Darwin, the diurnal cycle of rainfall is very pronounced. When appropriate, comparisons are made between Darwin and GATE.

2. Sampling error formulation

Let the process $r_A(t)$, which has mean $m(t)$, variance $\sigma^2(t)$, and autocorrelation $\rho(t_1, t_2)$, be sampled N times during period T at times $t_1, t_2, \ldots, t_N$. Then (5) can be expanded as

$$\sigma_\epsilon = (X + Y + Z)^{1/2}, \quad (7)$$

where

$$X = \left\langle \left[\frac{1}{N} \sum_1^N r_A(t_i) \right]^2 \right\rangle = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N [m(t_i)m(t_j) + \sigma(t_i)\sigma(t_j)\rho(t_i, t_j)],$$

$$Y = \left\langle \frac{1}{T^2}\left[\int_0^T r_A(t)dt\right]^2 \right\rangle = \frac{1}{T^2}\int_0^T\int_0^T [m(t_1)m(t_2) + \sigma(t_1)\sigma(t_2)\rho(t_1, t_2)]dt_1 dt_2,$$

and

$$Z = \left\langle -\frac{2}{NT}\sum_1^N r_A(t_i)\int_0^T r_A(t)dt \right\rangle = -\frac{2}{NT}\sum_{i=1}^N \int_0^T [m(t_i)m(t) + \sigma(t_i)\sigma(t)\rho(t_i, t)]dt.$$

When the rainfall statistics are time dependent, (7) is the proper expression to evaluate sampling error. In the case where the rainfall process is stationary and visits are at constant interval $\Delta t = T/N$, (7) reduces to

$$\sigma_\epsilon = \sigma\left[\frac{1}{N} + \frac{2}{N^2}\sum_{k=1}^{N-1}(N-k)\rho(k\Delta t) + \frac{2}{T^2}\int_0^T\int_0^{t_1}\rho(|t_1 - t_2|)dt_2 dt_1 - \frac{2}{NT}\sum_{i=1}^N\int_0^T \rho(|t - t_i|)dt\right]^{1/2}$$

or

$$\sigma_\epsilon = \sigma\phi[\rho(\tau), N, T], \qquad (8)$$

where σ is constant, $\rho(\tau)$ depends only on lag time τ, and $\phi(\rho, N, T)$ represents the impact on sampling error by the autocorrelation and the sampling frequency.

In the stationary case, the sampling error as a fraction of the process mean is simply

$$\frac{\sigma_\epsilon}{m} = \frac{\sigma}{m}\phi(\rho, N, T) = \mathrm{CV}\phi(\rho, N, T), \qquad (9)$$

where CV is the coefficient of variation of the area-averaged rainfall process $r_A(t)$.

The sampling error can be evaluated explicitly if the time-dependent or time-independent mean, variance, and autocorrelation of the rainfall process are known and the sampling period T and sampling interval Δt are given. In the above formulation, complete coverage of the area A by the sensor during each visit is assumed, which leads to a lower bound on the sampling error for the case with partial coverage.

3. Data

The Down Under Doppler and Electricity Experiment (DUNDEE) was carried out during two wet seasons (November 1988–February 1989 and November 1989–February 1990) (see Rutledge et al. 1992). The C-band Doppler radar was installed at Darwin to collect detailed rainfall data (Fig. 1). The radar covered three types of surfaces: continental land, sea, and an island. During rainfall events, PPI scans were made at time intervals ranging from 5 to 20 min. Ranges of 56, 113, and 226 km were used, depending on the storm location. (The range is the radius of the circular area covered by the radar scan.) The intermediate range, 113 km, was used most often. During periods of no significant rain (mostly during the early morning periods), measurements could be terminated based on the operator's judgment. In this study, zero rain is assumed for any period of no measurement, except for periods of known instrument failure. There are several short periods (a few hours) of no measurement between periods of heavy rain. Such no-measurement periods have been filled in with rainfall estimated through linear interpolation from the rainfall measurements on both sides. Our analysis has been limited to the second season, because its record is much more continuous and has finer temporal resolution than that of the first season. Except for two relatively long breaks, the data used cover the period from 10 November 1989 to 18 February 1990 nearly completely. In the original polar coordinates, the beam angle is 1.4°, and the widths are 0.25, 0.5, and 1 km for ranges of 56, 113, and 226 km, respectively. The polar coordinates have been converted into Cartesian coordinates with a spatial resolution of 1 km × 1 km. For individual pixels, the reflectivity has been converted to rain rate using the relationship $Z = 400R^{1.3}$, where Z is in its standard units ($\mathrm{mm^6\ m^{-3}}$) and R is in millimeters per hour. For widespread monsoon-type rainfall events, however, Williams et al. (1992) have found that using this relationship results in total rainfall estimates that are smaller

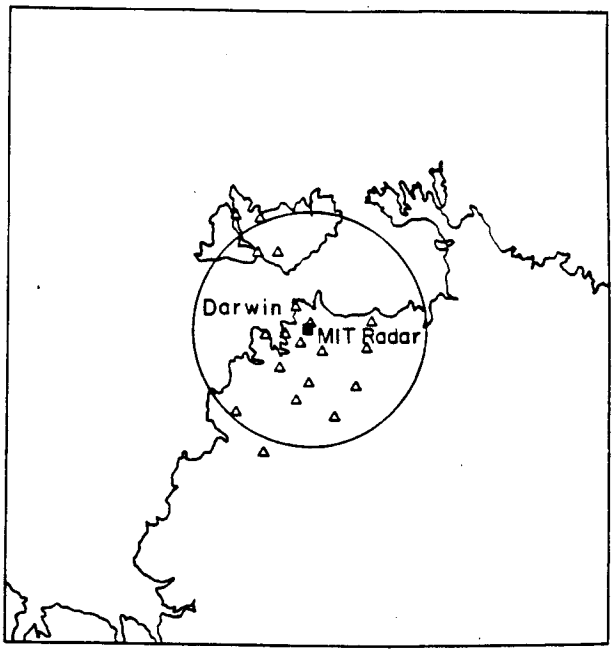

FIG. 1. Vicinity of Darwin and location of the MIT radar. Triangles indicate rain gauges.

than rain-gauge-based measurements by a factor of 5. Thus, in the present application, the rain rate derived from the relationship above has been multiplied by 5 to obtain areal rain.

Figure 2 is an example of the rainfall data, which shows the rain rate averaged over a 113-km-radius circular area (113-km range) scanned by the radar. The horizontal axis represents time from 0000 LST 1 November 1989. Records of rain rate averaged over areas of various sizes have also been obtained from the raw data. Notice that no measurements were made from day 50 to day 60 and from day 88 to day 91; these two periods have been excluded from our analysis. Also notice that relatively small amounts of rain occur during the initial 20 days, particularly during the second 10 days. These 20 days might represent a transition from the dry season to the wet season and, hence, correspond to a different rainfall regime than the rest of the dataset. In our analysis, these 20 days have also been excluded.

The C-band radars are known for their strong attenuation by heavy rain. To investigate this issue for the Darwin data, the average rainfall intensity is computed as a function of the rain depth traveled by the radar beam. Intensity–rain depth relationships are shown for the land and sea surfaces in Fig. 3. When the rain depth is not very large (localized convective storms), the attenuation is small, and for large rain depth (monsoon-type rainfall), the attenuation can be as large as a factor of 3–4.

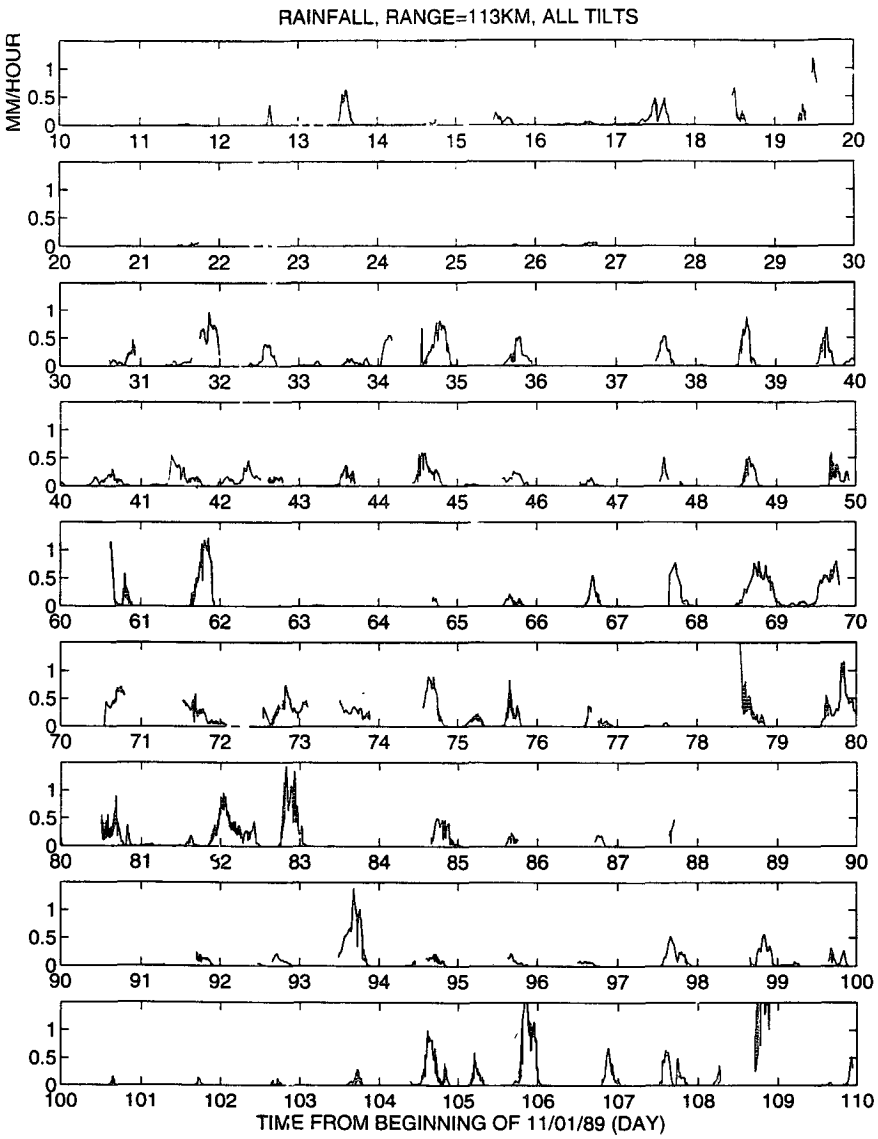

FIG. 2. Rain rate averaged over the 113-km range, 10 November 1989–18 February 1990.

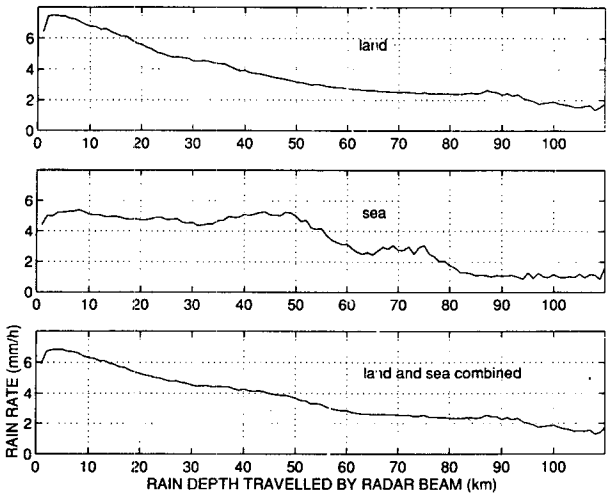

FIG. 3. Average rainfall intensity versus rain depth traveled by the radar beam.

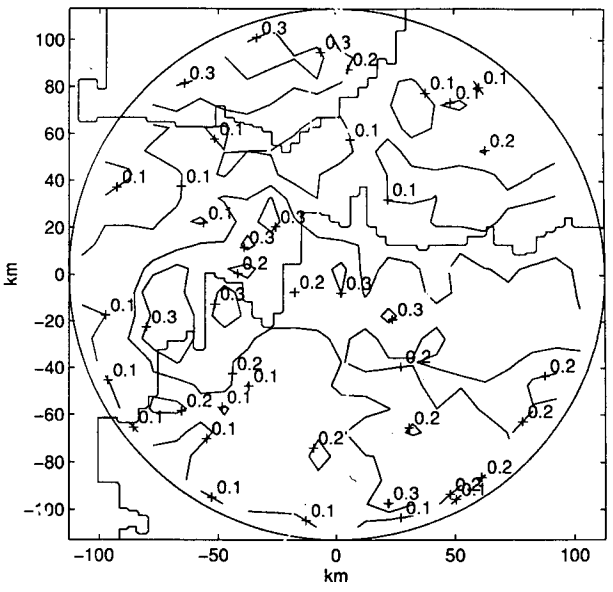

FIG. 5. Spatial distribution of the standard deviation of rain rate averaged over a 10 km × 10 km area.

4. Statistics of rain

a. Mean, variance, diurnal cycle, and intermittence

Rain statistics have been computed for the land and sea portions over various averaging areas. The spatial characteristics can be demonstrated by using relatively small areas such as boxes of 10 km × 10 km. The local mean rainfall within such boxes is shown in Fig. 4. It can be observed that, on average, mean rainfall intensity over the land is larger than that over the sea by about a factor of 2, and variability exists within each type of surface. The standard deviation, which is shown in Fig. 5, shows larger values over the land than over the sea. However, owing to difference in the mean, the coefficient of variation, is larger and more variable over the sea (see Fig. 6). Specifically, the CV varies between 9 and 13 over the sea but only between 8 and 9 over the land. It is appropriate to point out that, because of the large uncertainties associated with radar-derived rain rates, absolute values of rainfall cannot be determined with high confidence. But if we assume that the radar-derived rain rate differs from the true rain by a

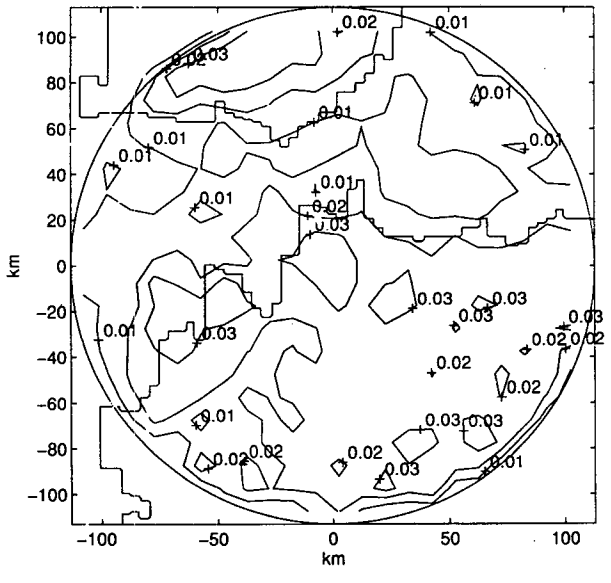

FIG. 4. Spatial distribution of mean rain for a 10 km × 10 km area.

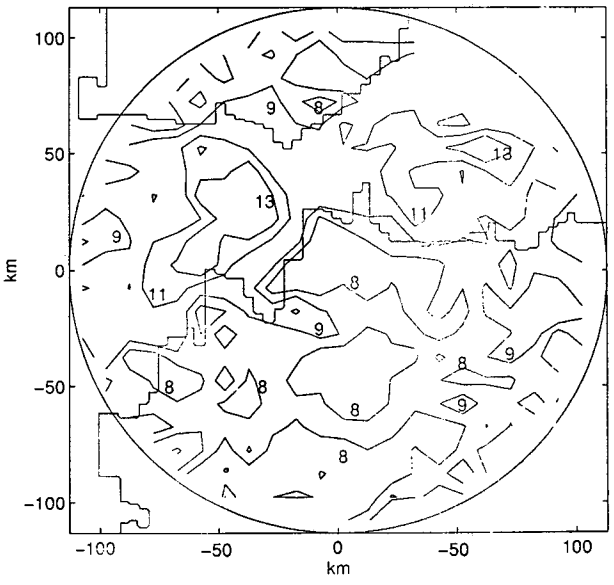

FIG. 6. Spatial distribution of CV for rain averaged over a 10 km × 10 km area.

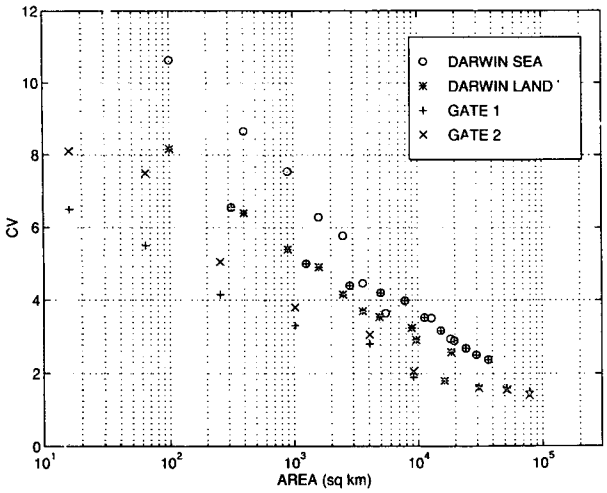

FIG. 7. Relationship between CV and area size. The ⊕ signs are for discs centered at the radar site.

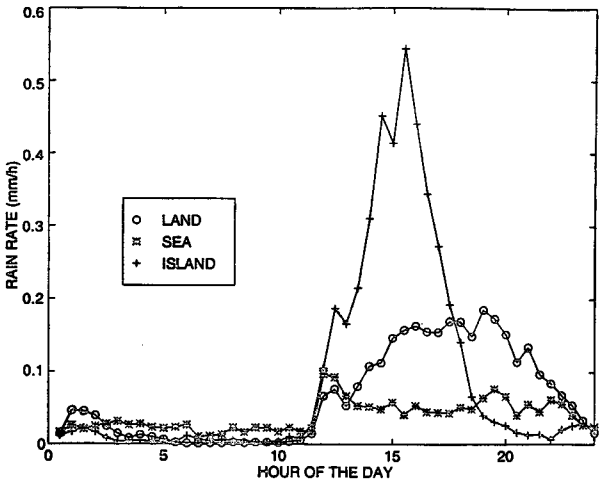

FIG. 8. Diurnal cycle of mean rain over the land, sea, and island portions within the 113-km range.

fixed factor, then CV can be determined correctly. Notice that the three figures cited above are for 10 km × 10 km boxes. As the size of the box increases, the mean remains stable, while the standard deviation decreases, leading to smaller CVs. The dependence of CV on the area size is presented in Fig. 7, in which the average CV for area size A is plotted against $\log A$ for the land and the sea separately and combined. Results for GATE, adopted from North et al. (1993), are also plotted for comparison. Figure 7 shows that for Darwin the sea portion generally has larger CV than the land but that the difference becomes smaller as the area increases. In addition, Darwin as a whole shows larger CV than GATE. However, CV for Darwin decreases with increasing area at a faster rate, which may be due to the fact that Darwin rainfall is mainly contributed by isolated local convective storms that have weak spatial correlation. When rainfall is averaged spatially, the variance for Darwin rainfall is reduced more effectively than for the GATE rainfall, which is spatially more widespread. The ⊕ signs in Fig. 7 are for areas in which both land and sea are included. These areas have been taken as discs centered at the radar site. Below 30-km radius (about 3000 km^2), the discs remain entirely on the land, so the ⊕ signs are consistent with the star signs (which represent land). Above 30-km radius, however, the discs begin to include sea surface, which has larger CV, resulting in the deviation of the ⊕ signs away from the stars. It should be pointed out here that the portion of the sea covered by the radar is a coastal water surface that is strongly influenced by the nearby continent, and its rain statistics should not be interpreted as representative of open oceans.

The Darwin rainfall exhibits a strong diurnal cycle, especially for the island and the continental land portions. Figure 8 shows the diurnal cycle of mean rain over the land, sea, and the island within the 113-km-radius radar coverage. Figure 9 presents the diurnal cycles of the mean, standard deviation, and CV in 1-h bins for the areal rain within the 113-km-radius coverage area (i.e., in this case, the land, sea, and the island are combined). As shown in Fig. 8, rainfall over the sea also shows a diurnal pattern, although it is less pronounced. It is interesting to note that there is generally more rain over the sea than over the land during the night, although both are small in magnitude; during the day rainfall over the land is much larger than that over the sea. The diurnal cycle over the island is extraordinarily strong, with a well-defined peak in the afternoon. This is likely due to the abundant vapor supply from the surrounding water surface coupled with the heated

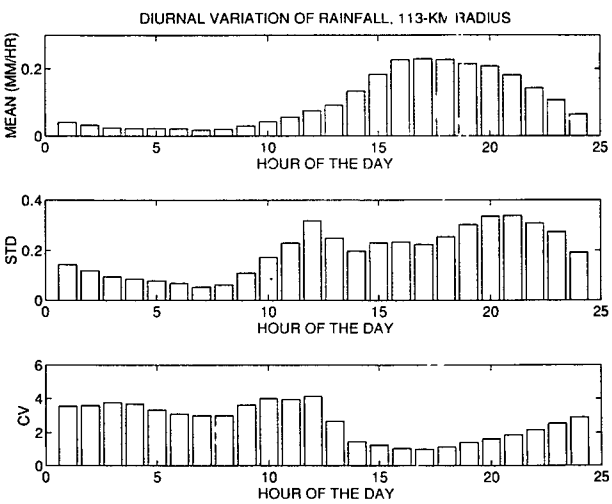

FIG. 9. Diurnal variation of mean, standard deviation, and CV for the areal rainfall within 113-km range.

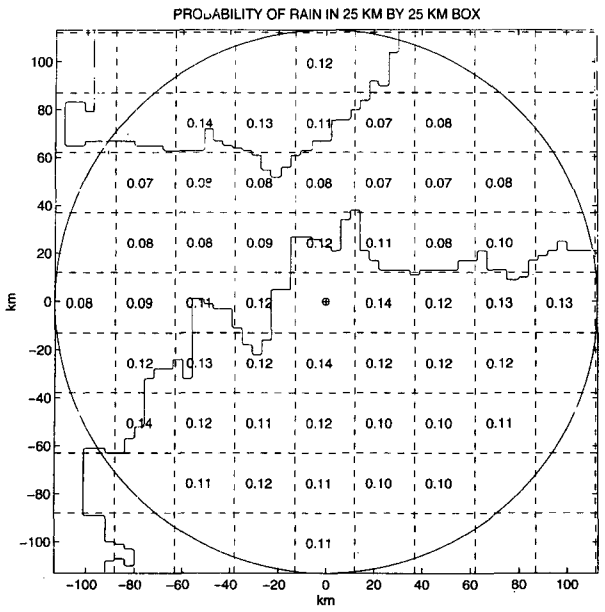

FIG. 10. Probability of rain (in time) within a 25 km × 25 km box.

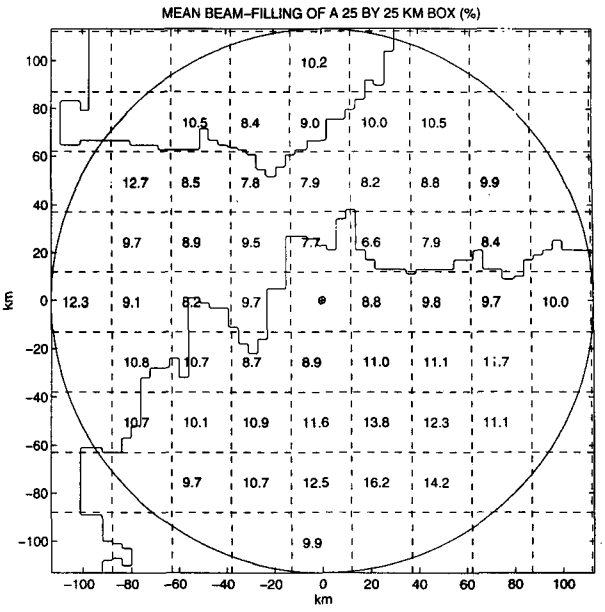

FIG. 11. Percentage of area covered by rain within a 25 km × 25 km box.

ground surface over the island in the afternoon. Intermittence is another distinctive feature of Darwin rainfall. Most rainfall occurs in the afternoon hours, and the duration of each storm is short. There is not much rainfall during the early morning hours. Analysis shows that at a random point of time the probability of rain anywhere within a 25 km × 25 km box is about 0.1, with smaller values over sea than over land (see Fig. 10). Darwin rainfall is very localized, which is associated with the intermitance. In most rainfall events, there are only a few isolated rain cells within the radar coverage. Figure 11 shows the average fractional raining rate within a 25 km × 25 km box, which has been based on the entire dataset. During rainfall, only about 10% of a 25 km × 25 km box is covered by the rain.

b. *Autocorrelation*

Since rainfall has been measured at irregular intervals, a new dataset with a constant time interval of 10 min has been created by linearly interpolating the original data. The autocorrelation is computed using the new dataset. Notice that we are dealing with the area-averaged rain rate. The correlation can be computed in two different ways. One is to treat the whole rainfall process as stationary with constant mean and variance and with correlation that depends only on lag time. The other is to account for the diurnal cycle and treat the process as nonstationary, with a diurnally varying mean and variance and with correlation that depends not only on lag time but also on absolute time.

Examples of autocorrelation resulting from the stationary treatment are shown in Fig. 12. The correlation has peaks around multiples of 24 h, which is an indication of diurnal cycle. However, the peaks are not always located at exact multiples of 24 h, due to the fact that throughout the whole period, heavy rain does not always occur at a fixed time of day. For example, in the early part of the DUNDEE project, most rainfall occurs from noon to 1600 LST, while in the later period heavy rainfall occurs mainly from 1600 to 2000 LST. At small lag time τ, $\rho(\tau)$ decays rapidly, indicating weak temporal correlation within each storm (Fig. 13). If a correlation timescale τ_0 is defined as the lag time at which the correlation drops

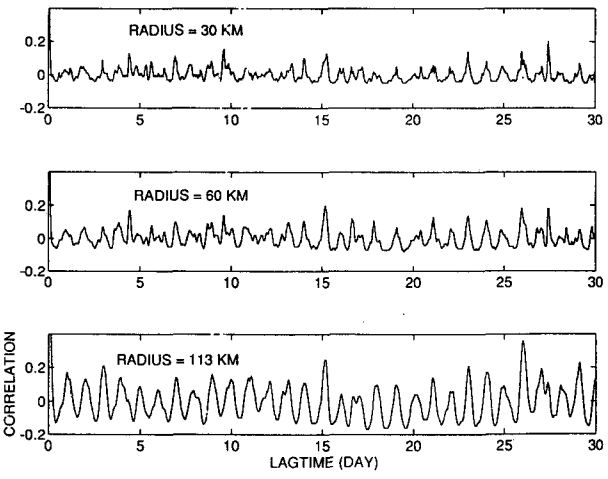

FIG. 12. Autocorrelation of areal rain rate over 30-, 60-, and 113-km radii for lag times up to 30 days; stationary model.

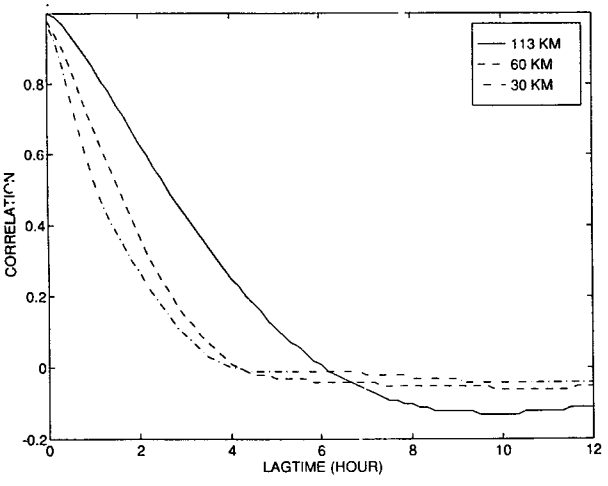

FIG. 13. Autocorrelation of areal rain rate over 30-, 60-, and 113-km radii for short lag times.

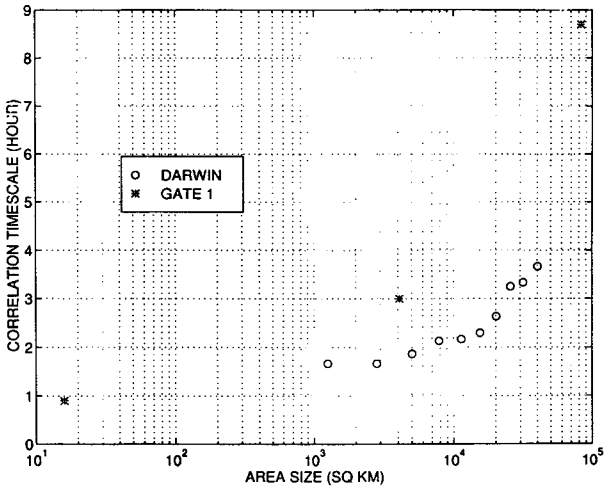

FIG. 14. Relationship between correlation timescale and area size.

to e^{-1}, then a τ_0 can be computed for each area size. Figure 14 shows the relationship between the correlation timescale and the area size. Results for GATE 1, adopted from Bell et al. (1990), are also plotted. For a given area size, Darwin rainfall has a much shorter correlation timescale.

Autocorrelations for the rain rate averaged over the land and sea portions are shown in Fig. 15. As the two regions have roughly the same area, Fig. 15 indicates that the land has a slightly stronger correlation than the sea.

In the nonstationary treatment, the autocorrelation is computed as

$$\rho(t_1, t_2) = \frac{\mathrm{cov}(t_1, t_2)}{\sigma(t_1)\sigma(t_2)},$$

where $\mathrm{cov}(t_1, t_2) = E\{[r(t_1) - m(t_1)][r(t_2) - m(t_2)]\}$, and $m(t)$ and $\sigma(t)$ are the time-dependent mean and standard deviation. Here, we consider only diurnal variation, that is, $m[t + 24n(h)] = m(t)$ and $\sigma[t + 24n(h)] = \sigma(t)$. To reflect the diurnal cycle of rain, a whole day is divided into 24 1-h bins, with bin 1 covering 0000 to 0100 LST, bin 2 covering 0100 LST to 0200 LST, etc. Each bin has a mean and a variance. The correlation $\rho(t_1, t_2)$ is computed and made into a lookup table as a function of the bins and the number of days apart that correspond to t_1 and t_2. A portion of this table is shown in Table 1 for the region with 113-km radius centered at the radar site. For example, supposing t_1 and t_2 are within the same day, t_1 is 0130 LST and t_2 is 0220 LST (meaning that t_1 belongs to bin 2 and t_2 belongs to bin 3), then from the upper portion of the table, one finds that the correlation is $\rho(t_1, t_2) = \rho(t_2, t_1) = 0.88$. The computation of correlation involves averaging within a bin. Therefore, even if both t_1 and t_2 belong to the same bin (but they are usually not exactly the same time), the correlation is less than 1. Suppose now that t_1 and t_2 are not located within the same day but that t_1 is one day ahead of t_2. Say, t_1 is 2230 LST 1 December 1989 and t_2 is 0120 LST 2 December 1989. Then we look at the lower portion of Table 1 with $t_1 = 23$ and $t_2 = 2$ and find that the correlation is $\rho(t_1, t_2) = 0.35$. The various columns in the upper portion of the table correspond to the correlation between a particular bin (t_2) and all the other bins (t_1). It can be observed that the correlation decays with increasing lag time. It is interesting, however, to note that the correlation between morning and night hours of the same day is quite high (e.g., for $t_1 = 23$ and $t_2 = 7$, the correlation is 0.77). This is because there is usually no rain during these two periods of the day, and occasional rainfall that occurs both during the morning and the night

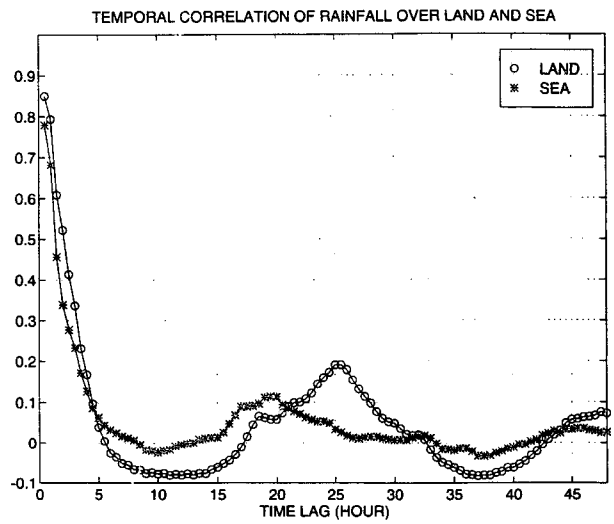

FIG. 15. Autocorrelation of areal rain rate over land and sea.

TABLE 1. A portion of the correlation lookup table—t_1 and t_2 are located within the same calendar day.

	t_2									
t_1	1	2	3	4	5	6	7	8	9	10
1	0.97	0.86	0.57	0.44	0.35	0.35	0.25	0.34	0.20	0.14
2	0.86	0.97	0.88	0.77	0.56	0.45	0.29	0.41	0.28	0.19
3	0.57	0.88	0.99	0.95	0.69	0.46	0.26	0.37	0.26	0.17
4	0.44	0.77	0.95	0.99	0.79	0.51	0.27	0.33	0.20	0.13
5	0.35	0.56	0.69	0.79	0.93	0.81	0.49	0.35	0.14	0.08
6	0.35	0.45	0.46	0.51	0.81	0.95	0.78	0.52	0.16	0.08
7	0.25	0.29	0.26	0.27	0.49	0.78	0.97	0.69	0.25	0.07
8	0.34	0.41	0.37	0.33	0.35	0.52	0.69	0.90	0.78	0.57
9	0.20	0.28	0.26	0.20	0.14	0.16	0.25	0.78	0.96	0.92
10	0.14	0.19	0.17	0.13	0.08	0.08	0.07	0.57	0.92	0.98
11	0.08	0.13	0.11	0.08	0.03	0.01	0.01	0.48	0.88	0.96
12	0.00	0.02	0.03	0.00	−0.04	−0.06	−0.05	0.42	0.84	0.92
13	−0.03	0.00	0.01	−0.02	−0.07	−0.10	−0.08	0.27	0.69	0.79
14	−0.10	−0.07	−0.06	−0.08	−0.13	−0.17	−0.12	0.10	0.35	0.42
15	−0.14	−0.11	−0.08	−0.09	−0.14	−0.18	−0.15	−0.01	0.15	0.19
16	−0.17	−0.13	−0.08	−0.06	−0.11	−0.16	−0.16	−0.06	0.07	0.11
17	−0.17	−0.10	−0.03	−0.01	−0.05	−0.11	−0.13	−0.09	−0.02	0.01
18	−0.15	−0.05	0.05	0.09	0.09	0.11	0.30	0.14	0.00	−0.05
19	−0.01	0.08	0.15	0.18	0.21	0.27	0.50	0.29	0.04	−0.04
20	0.17	0.25	0.29	0.33	0.50	0.56	0.61	0.37	0.09	0.02
21	0.12	0.19	0.23	0.27	0.51	0.63	0.62	0.32	0.02	−0.04
22	0.22	0.24	0.21	0.23	0.45	0.63	0.72	0.41	0.07	−0.02
23	0.32	0.30	0.21	0.20	0.46	0.65	0.77	0.45	0.10	0.01
24	0.14	0.14	0.10	0.09	0.32	0.56	0.75	0.40	0.07	−0.01
t_1 is one calendar day ahead of t_2										
1	0.08	−0.02	−0.03	−0.05	−0.07	−0.07	−0.06	−0.07	−0.05	−0.04
2	0.13	−0.01	−0.02	−0.03	−0.06	−0.07	−0.08	−0.07	−0.05	−0.04
3	0.12	−0.02	−0.03	−0.03	−0.05	−0.07	0.07	−0.07	−0.05	−0.05
4	0.10	−0.02	−0.03	−0.04	−0.06	−0.07	−0.07	−0.06	−0.04	−0.05
5	0.14	−0.01	−0.02	−0.02	−0.06	−0.08	−0.08	−0.06	−0.05	−0.06
6	0.28	0.08	0.03	0.04	−0.01	−0.06	−0.09	−0.08	−0.06	−0.06
7	0.51	0.22	−0.02	−0.04	−0.06	−0.06	−0.08	−0.08	−0.06	−0.06
8	0.29	0.09	−0.05	−0.07	−0.08	−0.08	−0.07	0.06	−0.04	−0.05
9	0.07	0.01	−0.03	−0.03	−0.06	−0.06	−0.01	0.05	0.04	0.00
10	0.00	0.00	0.01	0.00	−0.02	−0.02	0.04	0.15	0.11	0.03
11	−0.02	−0.03	−0.02	−0.02	−0.05	−0.05	−0.01	0.06	0.04	0.01
12	−0.04	−0.03	−0.02	−0.03	−0.04	−0.05	−0.02	0.02	0.01	−0.02
13	−0.07	−0.05	−0.04	−0.04	−0.03	−0.02	0.00	0.00	−0.03	−0.05
14	−0.11	−0.07	−0.04	−0.02	0.09	0.14	0.19	0.05	−0.04	−0.07
15	−0.14	−0.11	−0.08	−0.02	0.20	0.29	0.31	0.08	−0.06	−0.07
16	−0.15	−0.13	−0.10	−0.04	0.18	0.26	0.24	0.02	−0.11	−0.12
17	−0.15	−0.15	−0.12	−0.08	0.06	0.08	0.04	−0.08	−0.15	−0.16
18	0.27	0.08	−0.09	−0.10	−0.05	0.01	0.06	−0.04	−0.11	−0.13
19	0.43	0.19	−0.05	−0.10	−0.11	−0.09	0.00	−0.06	−0.10	−0.11
20	0.44	0.19	−0.02	−0.07	−0.07	−0.08	−0.04	−0.08	−0.09	−0.09
21	0.42	0.18	0.01	−0.02	−0.05	−0.04	−0.02	−0.06	−0.08	−0.07
22	0.59	0.28	0.03	−0.01	−0.04	−0.03	−0.01	−0.02	−0.04	−0.05
23	0.65	0.35	0.09	0.03	0.00	0.01	0.04	0.03	−0.01	−0.02
24	0.83	0.56	0.26	0.19	0.13	0.15	0.13	0.17	0.08	0.05

on the same day will contribute to such high correlation. However, because the mean rain rate during the morning and the night hours is very low, this high correlation has only a small impact on sampling error. Another feature to note in Table 1 is that there are no peaks around multiples of 24 h. The diagonal elements in the lower portion of the table correspond to correlation at 24-h lag time and they are very small. This means that the diurnal cycle has been accounted for in the diurnally varying mean.

5. Quantification of sampling error

a. Stationary process

Assuming a stationary rainfall process and using information about the mean, CV, and correlation ob-

tained in the previous section, the sampling error associated with the estimate of space–time rainfall for various sampling frequencies can be evaluated using (9). Equation (9) includes two factors, CV and ϕ, where CV is the coefficient of variation of the rainfall process and ϕ is a function of the rainfall autocorrelation and the sampling frequency. When the measuring period T is fixed at 30 days, ϕ can be computed for various sampling frequencies, as shown in Fig. 16. One can see from Fig. 16 that, although large areas tend to have smaller values of ϕ, the differences are generally small except at some particular hours. At $\Delta t = 12$ h, for example, the four areas shown in the figure all have ϕ values around 0.1. Generally, ϕ increases with increasing Δt, but sometimes a larger Δt produces a smaller ϕ. This is due to the particular shape of the empirical autocorrelation function. At $\Delta t = 24$ h, ϕ has a jump, and larger areas tend to have larger ϕ values, which is the result of the diurnal cycle. (Rainfall averaged over larger areas has stronger diurnal signal, see Fig. 12.)

To get an estimate of the sampling error percentage, one just needs to multiply the ϕ function by the coefficient of variation CV. The coefficient of variation can be obtained from Fig. 7. For example, given radar coverage radii of 30, 60, 90, and 113 km (i.e., over land–sea combined discs), CV is found to be 4.4, 3.5, 2.6, and 2.3, respectively. Multiplication of ϕ by CV produces the plots in Fig. 17, which give the sampling error as a fraction of the mean for various areas and sampling frequencies. For example, at $\Delta t = 12$ h, sampling errors are about 42%, 38%, and 26% for radii of 30, 60, and 113 km, respectively. For an estimate of the sampling error for larger areas, say 500 km × 500 km, visual extrapolation of Fig. 7 yields a CV value of around 1.0, if such extrapolation can be validly made. Assuming that the ϕ function for the 500 km × 500 km area does not change from that for the 113-km radius, the sampling errors at $\Delta t = 12$ h and $\Delta t = 24$ h are found to be about 11% and 37% for complete coverage of the area at each visit.

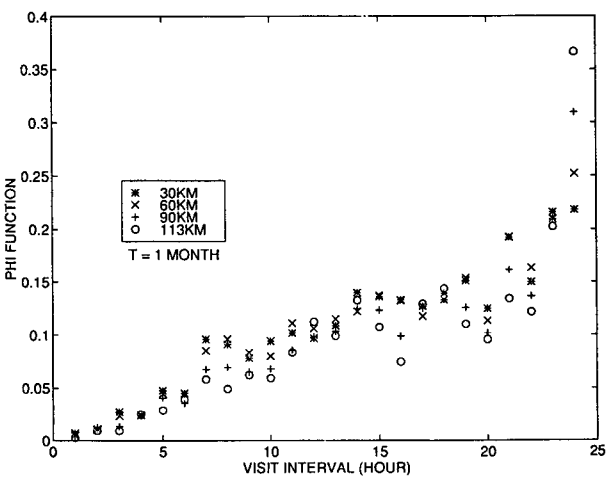

FIG. 16. Parameter ϕ as a function of sampling interval for 30-, 60-, 90-, and 113-km radii.

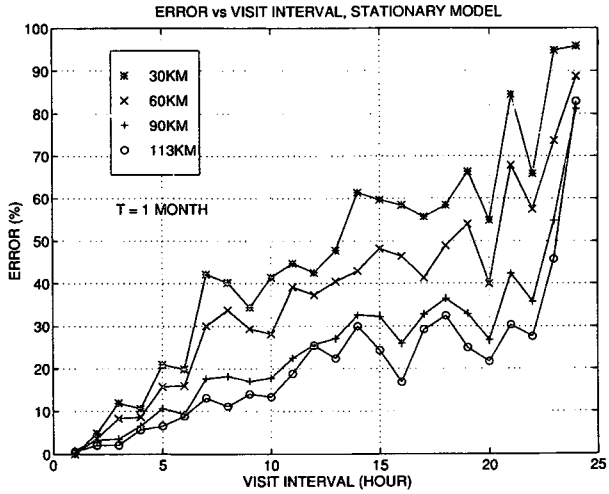

FIG. 17. Sampling error as a function of sampling interval for 30-, 60-, 90-, and 113-km radii; stationary model.

The sampling error is proportional to the value of CV. Since CV for the Darwin rainfall decreases with increasing area at a much faster rate than the GATE CV, its sampling error also decreases at a faster rate. For smaller areas, the sampling error for Darwin is expected to be larger than that of GATE because of the larger CV. For larger areas, the difference in the sampling errors for the two sites may be reduced, with Darwin having a somewhat larger error than those reported by Bell et al. (1990) and North et al. (1993).

b. Nonstationary process

The stationary treatment described in section 5a ignores the fact that rainfall has a diurnal cycle in the mean and variance. It treats this diurnal cycle only implicitly in the autocorrelation function. The error estimates obtained from such stationary treatment do not depend on the specific times at which the samples are taken. To allow the error estimates to vary with the absolute sampling time, the nonstationary model of (7) must be used. This model requires the quantification of the time-dependent mean, variance, and autocorrelation of the rainfall process. These quantities have been computed and discussed in section 4. Using the model, sampling errors for monthly rainfall over areas of various sizes can be evaluated as a function of sampling frequency and starting sampling time. Figure 18 shows the sampling errors of monthly rainfall for the 113-km radius. Each small plot corresponds to a sampling frequency, and the bars in the plot correspond to various starting sampling times. For example, the first plot (up-

per left) has a visit interval of $dt = 1$ h, and the only bar represents the sampling error of monthly rainfall when the sampling starts from 0100 LST of the first day. Since the visit interval is 1 h, starting from 0100 or any other hour will result in the same hours being visited and, hence, in the same sampling error. Thus, only one bar is shown in this first plot. Generally, the sampling errors increase with increasing visit interval. For some visit intervals such as 12-h intervals, the errors change with the starting time and the maximum error can be twice as large as the minimum. For some other visit intervals such as 5-h intervals, however, the errors stay nearly the same for all starting times. This is because for certain visit intervals, not all hours of the day are sampled during the month, and different starting times result in different sets of hours being visited. For example, for a 12-h interval, if the sampling starts at 0100 of the first morning, the only two times that are sampled during the month are 0100 and 1300. Similarly, if the sampling starts at 0200 of the first morning, only 0200 and 1400 are sampled. Such different sampling times result in different sampling errors, due to the diurnal cycle of rain. For other visit intervals, almost every hour of the day is sampled during the month regardless of the starting time, resulting in a nearly constant error estimate. At an interval of 5 h, for example, each hour of the day is sampled six times during the month no matter when the sampling starts (although, of course, the starting time must be located within the first day).

Cyclicity can be observed in some of the plots in Fig. 18. For example, with visits at an 18-h interval

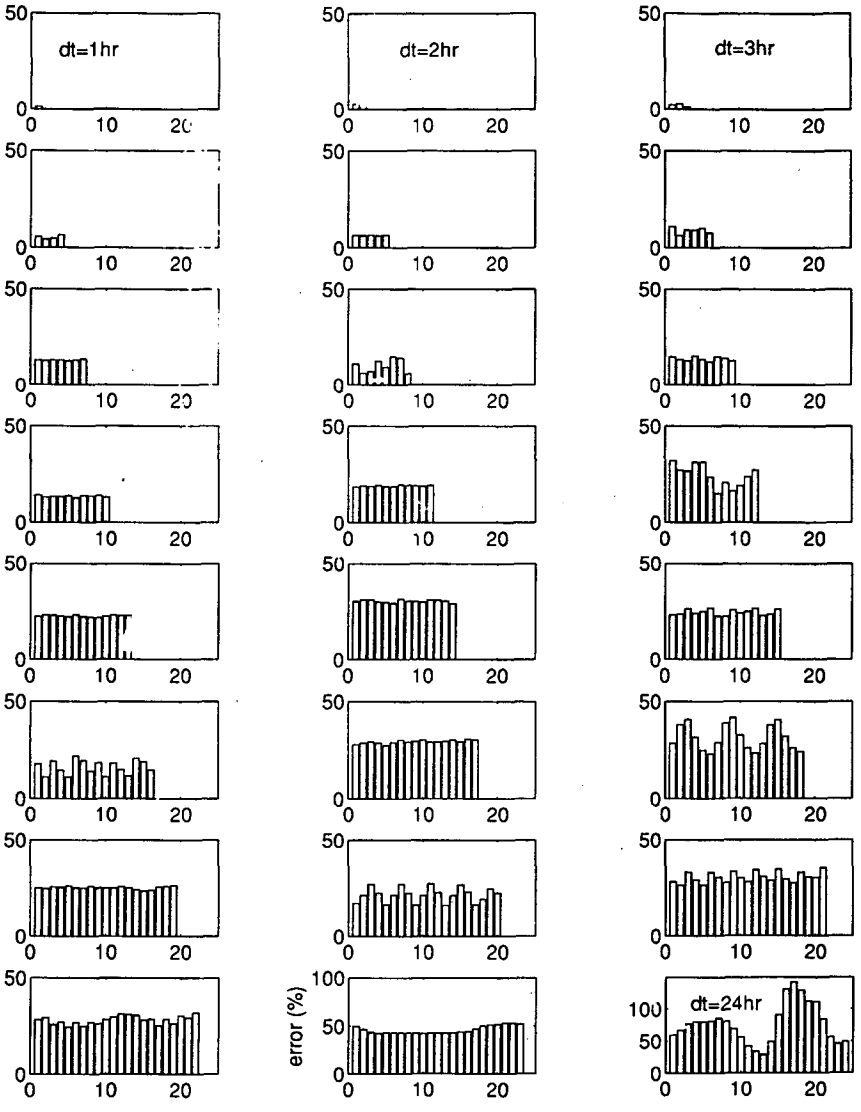

FIG. 18. Sampling error vs sampling interval and starting time; nonstationary model.

(last column, third plot from bottom), there are three cycles in the bar chart. The reason is that starting at 0100, 0700, and 1300 results in the same set of hours of the day being visited. Likewise, starting at 0200, 0800, and 1400 results in another identical set of hours being visited, thus producing cyclicity. The same phenomenon happens for the 20-h interval (middle column, second plot from bottom), where there are five cycles of period 4 in the bar chart.

Figure 18 is intended to show the effect of starting time on the sampling error when the diurnal cycle is explicitly considered in the nonstationary model. Compared with the stationary model, the nonstationary model provides more detailed information. When the results of the nonstationary model are averaged over starting times, they reduce to the stationary model results. This is shown in Fig. 19, where the circles represent stationary model results, and the cross signs are averages of the nonstationary model errors over all starting times for various visit intervals. The match is quite good.

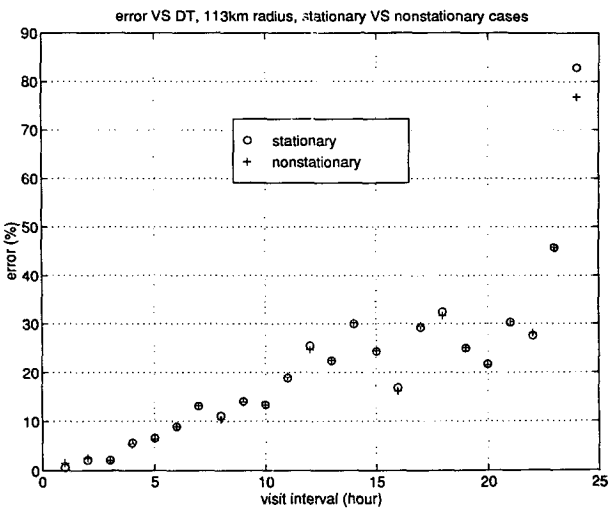

FIG. 19. Comparison between stationary and nonstationary models.

c. *Sampling experiments using real data*

Another way of estimating the sampling error is to use the original rainfall data to perform sampling experiments. The sampling errors obtained this way are realizations of a sampling process and may deviate considerably from the pattern of standard deviations produced by the models. McConnell and North (1987) have used this method to study GATE rainfall and found that the errors are within 10% of the true rain for a 400 km × 400 km box and for periods of 15–18 days. Their results do not depend on the exact sampling time, apparently due to the lack of diurnal cycle in the GATE rainfall.

We consider a period of 30 days, consisting of days 61 to 88 and days 92 to 93 (since there were no measurements from days 89 to 91; see Fig. 2).

The rainfall process is sampled at intervals ranging from 1 to 24 h and with various starting times. These experiments have been done with the 113-km-radius data. The sampling errors are presented in Fig. 20 as percentages of true rain. It can be observed that the sampling error varies with starting times. For example, for a 12-h sampling interval, starting at 0100, 0200, 0900, 1000, 1100, and 1200 causes underestimation of the monthly rainfall, while starting at 0400, 0500, 0600, 0700, and 0800 causes overestimation of rainfall.

As this is only a single experiment, the errors for each starting time and each visit interval cannot be easily compared with the statistical results of the nonstationary model. However, when an average is taken over various starting times, comparison with the stationary model should be possible. This is shown in Fig. 21, where the cross signs represent the averaged results (standard deviation) of the experiments, and the circles are the stationary model results. Considering that the experiments consist of only one month of data, the correspondence is quite good.

6. Discussion and conclusions

Rainfall data from Darwin have been analyzed for the variability in mean, variance, temporal correlation, and the errors associated with discrete sampling schemes. Within the radar coverage, the rainfall over the central portion of the sea surface generally has a smaller mean than on land, while near the coast, the rainfall over the sea is strongly influenced by local land. The standard deviation of area-averaged rain shows a similar pattern, being larger over the land and smaller over the sea. The coefficient of variation is smaller and is spatially more stable over the land, varying from 8 to 9 for a 10 km × 10 km area, while over the sea, CV varies from 9 to over 13 for the same area size. Up to the maximum area available within the radar coverage ($\sim 2 \times 10^4$ km^2), CV remains larger over the sea than over the land. However, CV decreases with area faster over the sea than over the land, and over large areas (beyond radar coverage), they may converge. Compared with GATE, the Darwin rainfall has larger CV over small areas for which data are available, which means that the sampling error is expected to be larger for Darwin.

The temporal correlation of area-averaged rain is weak at short lag times compared with GATE, a manifestation of short-lived storms. On the other hand, a strong diurnal cycle exists, which peaks in the afternoon. The diurnal cycle is strongest for the island, followed by the continental land and the sea. Although rainfall over both land and sea is rare and has moderate intensity during the early morning hours, the sea receives more rain than the land during that period. Dur-

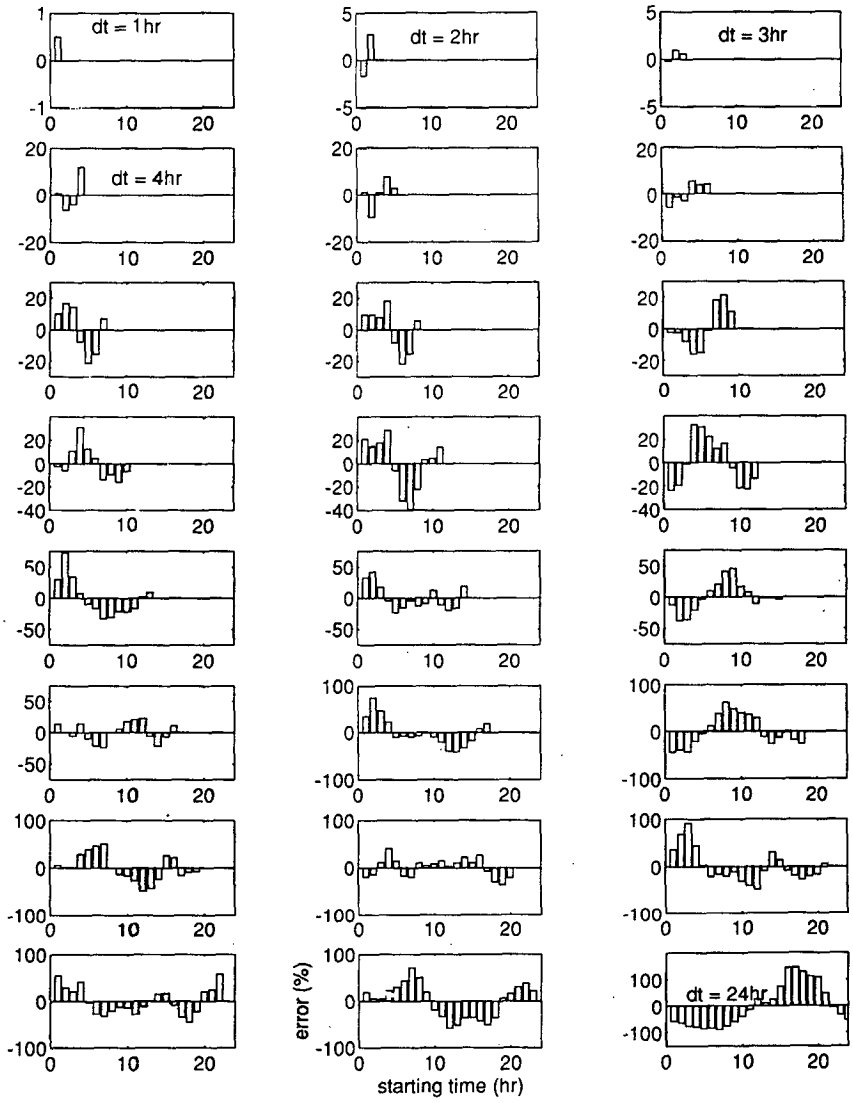

FIG. 20. Sampling error versus sampling interval and starting time, from data.

ing the day, this trend is reversed, and the land receives much more rain than the sea.

The intermittent and localized nature of the Darwin rainfall has also been studied. For a 25 km × 25 km box, the probability of rain at a generic point in time is about 0.1, with lower values over the sea (0.07–0.11) than over the land (0.10–0.14). On average, only about 10% of the area within such a box receives rain when the box as a whole experiences rain. This percentage value is smaller over the sea (6.6–12.7) than over the land (8.8–16.2). These results have implications for space-based rainfall measurement, in that low probability of rain makes it difficult for the satellite to catch rainy scenes and thus increases sampling error. Also, a highly localized rain cell is smaller than the sensor's field of view and thus decreases the sensitivity of the sensor to the rain.

Both stationary and nonstationary models have been used to quantify the sampling error associated with space-based measurements of space–time rainfall. The nonstationary model provides important information about the effect of the diurnal cycle on the sampling error. This effect is absent in stationary rainfall representations. When diurnal cycle is present, for certain sampling frequencies, the sampling error is sensitive to the starting sampling time. For other frequencies, the starting time is not important, due to the even sampling of almost every hour of the day during the period of estimation (i.e., one month). When the results of the nonstationary model are averaged over various starting times, they reduce to the stationary model results.

Also, empirical sampling experiments, in which the sampling errors are calculated directly from rainfall data, show sensitivity of the sampling error to the start-

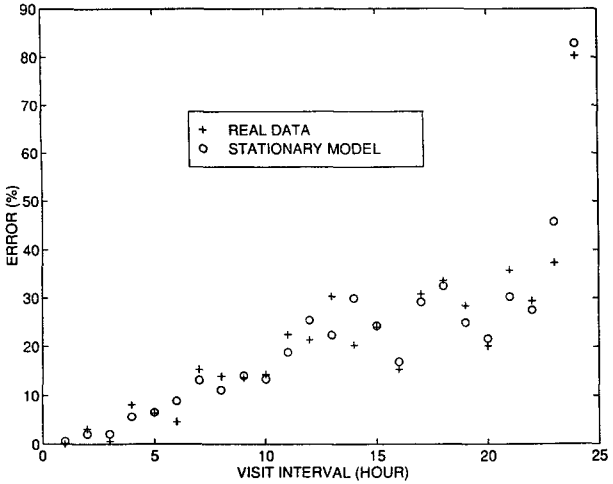

FIG. 21. Comparison between model and empirical results.

ing time. This is not the case for the GATE results (McConnell and North 1987). When the results of the experiments are averaged over various starting times, they match the stationary model results quite well.

Acknowledgments. We would like to thank Dr. E. Williams and D. Boccippio of MIT Radar Laboratory for their generous help with the radar data, and Drs. D. Entekhabi and E. A. B. Eltahir of MIT Parsons Laboratory for their useful comments and suggestions. This research was supported in part by the National Science Foundation under Grant ATM-9020832 and in part by the National Aeronautics and Space Administration under Grant NAG5-1615.

REFERENCES

Bell, T. L., 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**(D3), 2195–2205.

Bras, R. L., and I. Rodrigues-Iturbe, 1985: *Random Functions and Hydrology.* Addison-Wesley, 559 pp.

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

McConnell, A., and G. R. North, 1987: Sampling errors in satellite estimates of tropical rain. *J. Geophys. Res.*, **92,** 9567–9570.

North, G. R., and S. Nakamoto, 1989: Formalism for comparing rain estimation designs. *J. Atmos. Oceanic Technol.*, **6,** 985–992.

——, S. S. P. Shen, and R. Upson, 1993: Sampling errors in rainfall estimates by multiple satellites. *J. Appl. Meteor.*, **32,** 399–410.

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

Rutledge, S. A., E. R. Williams, and T. D. Keenan, 1992: The Down Under Doppler and Electricity Experiment (DUNDEE): Overview and preliminary results. *Bull. Amer. Meteor. Soc.*, **73,** 3–16.

Seed, A., and G. L. Austin, 1990: Variability of summer Florida rainfall and its significance for the estimation of rainfall by gages, radar and satellite. *J. Geophys. Res.*, **95,** 2207–2216.

Shin, K.-S., and R. R. North, 1988: Sampling error study for rainfall estimate by satellite using a stochastic model. *J. Appl. Meteor.*, **27,** 1218–1231.

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

Williams, E. R., S. A. Rutledge, S. G. Geotis, N. Renno, E. Rasmussen, and T. Rickenbach, 1992: A radar and electrical study of tropical "hot towers." *J. Atmos. Sci.*, **49,** 1386–1395.