The Effect of Spatial Resolution on the Correlation Structure of Gate III Rainfall Fields

JOUGH-TAI WANG[1]   AND   SHOICHIRO NAKAMOTO[2]

ABSTRACT

The present study uses a fine scale data set from GATE phase III to investigate spatial effects on rainfall characteristics and its lagged autocorrelation. Due to the finer structure of the present data set, the physical processes invloved for a shorter distance scale can be identified, which has crucial information in the estimation for the sampling errors of a rain field.

It is found that for a rainrate less than 4 mm/hr, a 1 km by 1° and 4 km by 4° data set will not make any noticeable difference to accumulated rainfall statistics. This is directly implied by stratiform rainfall associated with the mesoscale circulation system. The rainfall within that physical region is continuous with a rainrate less than 6 mm/hr. A rainfall rate greater than 40 mm/hr contributes 20 % to the total rainfall in 1 km by 1° data, in contrast to an 8 % contribution from the 4 km by 4° data. The area-averaged processes suppress extreme rainfall considerably (the extreme would have come from the convective scale).

The variance and the autocorrelation calculated in this study reveal that the slope of the variance and autocorrelation for a smaller scale are different from those estimated through a larger area data. This is an indication that the convective scale rainfall field caused this change, and the existence of a horizontal integral length scale for the shorter scale end is postulated. The implications of a horizontal integral length scale and suggestions for further research are discussed.

1. INTRODUCTION

It was understood that outside the tropics, the primary energy source of synoptic disturbances is available potential energy with a strong latitudinal temperature gradient. In the tropics, on the other hand, the storage of available potential energy is very small due to the very weak temperature gradients. The latent heat release associated with convective rainfall appears to be the primary

---

[1] *Institute of Atmospheric Physics, National Central University, Chung-Li, Taiwan 32054, R. O. C.*

[2] *Division of Oceanic Development, Japan Marine Science and Technology Center, Yokosuka, 238 Japan.*

energy source for disturbance there. Most latent heat release in the tropics occurs in convective cloud systems. This large sum of latent heat released through precipitation is vital to a large-scale general circulation pattern, and the water associated with it is essential for human activity. However, the estimation of precipitation over the tropics is a difficult problem. Unlike many meteorological parameters, rainfall is discontinuous in space and time and exhibits a large natural variability. Current observation systems, such as rain gauges and radar, are generally limited to the measurement of precipitation over land. The oceanic rainfall is more frequently extrapolated from other data rather than measured directly.

Satellites have been regarded as a means to circumvent some of the deficiencies of gauge and radar measurement of rainfall. Satellites have the ability to access remote areas of the world on a regular basis and this is a great advantage. However, satellite systems present their own problems. For example, rainfall rates must be indirectly inferred using remote sensing methods. A number of techniques have been developed to infer rainfall from visible or infrared data or to relate it more physically to radiation emitted at microwave frequencies.

Sampling errors due to temporal gaps dominate the error budget of a global rainfall dataset exclusively constructed by a specified satellite. The sampling errors may be defined as the difference between a "true" estimate one would obtain from a temporally continuous data set and an estimate from the discontinuous data. Radar or gauge observed data must be used to estimate the sampling errors and the implication for errors in satellite-measured datasets can then be identified. One of the most frequently used observational data sets for this type of research is from the GATE (Global Atmospheric Research Program, Atlantic Tropical Experiment).

During the GATE in 1974, the approach of combining satellite, radar, and aircraft, with an exceptionally dense array of ships launching upper-air balloons was adopted, to probe the convective cloud field. These data were supplemented by standard synoptic observations, surface-based cloud photography, and a variety of boundary layer measurements. The GATE dataset offered an unprecedented opportunity to investigate the evolution of convective clouds and the rainfall characteristics associated with these events. Many studies had been conducted using both real data taken from the GATE experiment and using stochastic models of rain rates based on the observed data to investigate the lagged space-time correlation of rain fields (McConnell and North, 1987; Laughlin, 1981; Hudlow and Patterson, 1979)

The lagged space-time correlation of the rain rate is needed for the estimation of sampling errors in finite rain-gauge networks, and in the estimation of bias due to beam-filling. The spatial autocorrelation of the rain rate was derived from the GATE phase I and Phase II data. The correlation field is well

described by the formula (Bell, 1987)

$$C_r(s) = \frac{1}{(0.25 \times s + 0.63682)^{2/3}} \quad , \quad 4 \ km \leq s \leq 60 \ km \tag{1}$$

where $C_r(s)$ is the spatial autocorrelation and s is the spatial distance (km). That is the spatial autocorrelation function is described by a power law $s^{-2/3}$. The variance of the area-averaged rain rate decreases much more slowly and is nearly $A^{-1/3}$ where A stands for an area (Laughlin, 1981). This finding about the variance is consistent with the $s^{-2/3}$ spatial correlation power law. This implies that the concept of a conventional horizontal integral length scale is questionable. If there were a finite correlation length, the variance of the area-averaged rain rate would have to decrease as A^{-1}. If there were no integral length scale for the smallest scale, the variance of the small area-averaged rain-field would be large, which then indicates large sampling errors in the estimation of a true correlation. However, the $s^{-2/3}$ power law of spatial autocorrelation is based on data with a minimum resolution of 4 $km \times$ 4 km. The physical events which have a scale smaller than 4 km could not be resolved in the study when deriving Eq. (1). Hence, how the $s^{-2/3}$ power law formulation in Eq. (1) is affected by the smaller-scale events is uncertain.

After the GATE experiment, one of the major findings is the identification of mesoscale disturbances embedded in a convective cloud cluster system (Houze and Betts, 1981). In the GATE region, cloud cluster systems have a life cycle of 12 to 24 hours evolving from individual convective cells to a mature organized mesoscale system and finally dissipating. Rainfall deriving from the mesoscale anvil region is mostly continuous and can last for several hours at a rate of 2 ~ 6 mm/hr, in contrast to the convective core region where rainfall rates are in the range of 10 ~ 100 mm/hr (Leary and Houze, 1979). The characteristic length scale of the convective core region is only on the order of several km, the mesoscale anvil region has a spatial domain of about 100 km. So the convective core regions may not be totally resolved by a 4 km by 4 km data set. This convective disturbance may thus play a role in the spatial lagged correlation statistics. Therefore this study attempts to reexamine the lagged space-time autocorrelation based on a fine scale data set which can resolve the small convective events, in order to investigate the correlation structure for the smaller-scale in order to estimate of the rainfall statistics, and compare them with the conclusions drawn from studies with a coarser grid data. An attempt will be made to physically interpret the statistical results of this study through the observed cloud cluster events.

Section 2 will describe the data set and the techniques used in the analysis. The results and discussion are presented in Sec. 3. Section 4 compares

the present study with other previous studies and a generalized hypothesis is postulated. Conclusions and suggestions for further research follow (Sec. 5).

2. DATA AND ANALYSIS TECHNIQUE

The data set used in this study is based on radar observations from the Canadian ship named "Quadra" during Phase 3 of the GATE experiment. The GATE radar data has been studied extensively. However, most of the radar rainfall analyses were based on the smoothed data set constructed by Hudlow and Patterson (1979). The data density used by Hudlow and Patterson was 4 km by 4 km and the time interval was 15 minutes. This resolution is not capable of deducing behavior smaller than 4 km. The present data set is based on radar observations at 1 km by 1° of sweep angle and at 5 minute intervals. Only 8 Julian days of data are used in this analysis (day 245, 246, 252, 253, 254, 260, 261, 262). The detailed structure of the convective events on these days have been studied by many investigators, e.g, Houze and Betts (1981), Leary and Houze (1979), Esbensen et al. (1988).

If we have P data segments in a fixed time interval, and for each segment, there are N numbers of data points, the lagged autocovariance function (ACVF) can be estimated by the formula as follows.

$$C_B^* (m) = \frac{1}{N \times P} \sum_{i=1}^{P} \sum_{n=1}^{N-m} \left[X_{n+m} (i) - \overline{X} \right] \left[X_n (i) - \overline{X} \right] \qquad (2)$$

where the overbar represents an average through all the available data used in the study, namely,

$$\overline{X} = \frac{1}{N \times P} \sum_{i=1}^{P} \sum_{n=1}^{N} X_n (i) \qquad (3)$$

with m as the spatial lag. X can be any meteorological variable of interest. Here the rain rate is used. The subscript of n (i) stands for the data point in the spatial (temporal) domain. The lagged autocorrelation function (ACRF) thus is estimated as

$$\rho_B (m) = \frac{C_B^* (m)}{C_B^* (0)} \qquad (4)$$

$C_B^* (0)$ is the ACVF for zero lag. We name this scheme the "method B" scheme.

There exist different estimators to estimate the ACVF, for example, it can be estimated as

$$C_A^* (i, m) = \frac{1}{N} \sum_{n=1}^{N-m} \left[X_{n+m}(i) - \overline{X}(i) \right] \left[X_n(i) - \overline{X}(i) \right] \qquad (5)$$

where

$$\overline{X}(i) = \frac{1}{N} \sum_{n=1}^{N} X_n(i) \qquad (6)$$

and

$$\rho_A(m) = \frac{1}{P} \sum_{i=1}^{P} \frac{C_A^*(i, m)}{C_A^*(i, 0)} \qquad (7)$$

We call this scheme the "method A" scheme.

The main difference between these two estimates of the ACRF is that method B uses all the possible information contained in the dataset to form an ensemble mean, as indicated in Eq. (3), while method A only used the data in a given time segment to form the mean, as represented by Eq. (6). Further discussion can be found in the Appendix.

For a consistent estimator of the ACVF or ACRF, we want the limit of large N not to lead to a value different from the true ensemble average. In rainfall data, one often has a finite strip of data N, but many different independent realizations of the strip. In such a case, we need to know which is the better estimate of the ACVF. Trenberth (1984) has considered these two estimators with the application to large scale meteorological fields in mind. He found that method B was the better estimator for such an application.

Since his analysis assumed Gaussian statistics which do not hold for rainfall statistics, we have to reexamine this estimator problem. We construct a more generalized model which can treat either the Gaussian or non-Gaussian noise processes. An analytic expression for the bias generated through different methods under a hypothesized AR1 red noise process is derived. By looking at the estimators for an AR1 process within the limit of large N, the asymptotic estimation of the bias generated by different schemes may be written (see Appendix for details) as

$$b(m) = \frac{(1 + \alpha)(\alpha^m - 1)}{(1 - \alpha) \times N^+} \qquad (8)$$

where $b(m)$ is the bias for lag m and α is the AR1 parameter. For method A, N^+ is the number of data points in a given segment. For method B, N^+ is

the number of data points over the **total** segments [or $N \times P$ as in Eq. (3)]. Bias generated by method B is much smaller than that generated by method A. This demonstrates that method B is a better estimator for the estimation of the rainfall statistics. Trenberth (1984) also pointed out method B is a better estimator for his formulation. Our generalized study finds that the Gaussian or non-Gaussian noise only affected the covariance function. It doesn't affect the correlation function. Therefore Eq. (4) is used in this study to estimate the ACRF.

3. RESULTS AND DISCUSSION

Before showing the results of the spatial autocorrelation function computed from the present dataset, some of the mean statistics are presented first to identify the effects of area-averaging on the fundamental rainfield structure.

a. Mean

Figure 1 shows the probability density function of the rainfield, based on the present dataset (which has 1 km by 1° resolution) and the data which are reconstructed from the present set to have 4 km by 4° resolution. The length of a 1° sweep angle depends on it's radius. It varies from about 0.85 km (around a radius of 50 km) to 1.7 km (around a radius of 100 km). The largest probability for the rain rate mode is on the order of 4 mm/hr. 1 km by 1° and 4 km by 4° have a similar probability structure, and the latter has a weaker probability distribution at the 4 mm/hr rate. The corresponding contribution to the total rainfall by these 2 resolutions is shown in Fig. 2. In the 1 km by 1° dataset, the contribution from a rainrate greater than 40 mm/hr contributes about 20 % of the total rainfall, in contrast to the 8 % contribution from the 4 km by 4° resolution. This means that extreme rainfall occurs in a smaller area and the area averaging process suppresses these extreme values. The contribution to the rainfall from the smaller rain rate (say, less than 6 mm/hr) is quite comparable in these two resolutions. This leads to the confirmation that the rainfall from the anvil regions is mostly homogeneous and at rates smaller than 6 mm/hr. This point is clearly identified in Fig. 3. Figure 3 shows the accumulated rainfall contribution from different rain rates for the two resolutions used in Fig. 1. For a rain rate $\leq$ 10 mm/hr, they contribute about 50 % of total rainfall for a 1 km by 1° dataset. Due to smoothing at the 4 km by 4° resolution, it takes a higher rain rate for 1 km by 1° data to achieve the same level of contribution to total rainfall. For example, the accumulated 10 mm/hr rainrate in the 4 km by 4° resolution contributes 55 % of the total rainfall, in contrast to 50 % from the 1 km by 1° resolution.

Below a 6 mm/hr rain rate, the contributions from these two resolutions are almost identical. Physically, this coincidence of two resolutions at a smaller
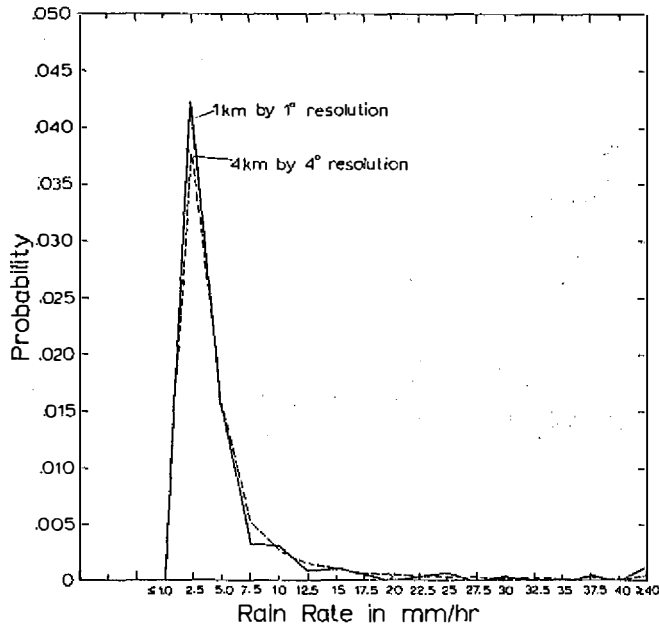
Fig. 1   The probability distribution function of the rainfield based on the GATE phase III data with 1 km by 1° and 4 km by 4° resolutions.

rain rate can be explained by the mechanisms involved in the convective cloud system. It is known that in the GATE region, mature cloud clusters not only has a convective core region, they also have an anvil region associated with them. The rainfall from an anvil region is estimated at a rate of 2 ∼ 6 mm/hr and it usually lasts for several hours. Because of the large horizontal area occupied by the anvil region, these two resolutions actually resolved the physically similar behavior at a smaller rain rate branch.

Only in the convective core region, which is associated with higher rainfall, does the contribution to the total accumulated rainfall through these two resolutions begin to differ. The two different characteristic rainfall mechanisms in the GATE cloud clusters do have an impact on the accumulated rainfall contribution with respect to different spatial resolutions. This suggests that it would be interesting to investigate rainfall characteristics in other climatologically different regimes to see if this phenomenon persists. Such information will be useful for generalizing the conclusions drawn based on the GATE data.

b. Variances

The variances of these two resolutions with different radii of radar measurement are shown in Table 1. The general pattern is that the 1 km by 1° spatial resolution has large variances as expected. However, the point which really concerns us is the slope of this increase from a larger spatial domain to
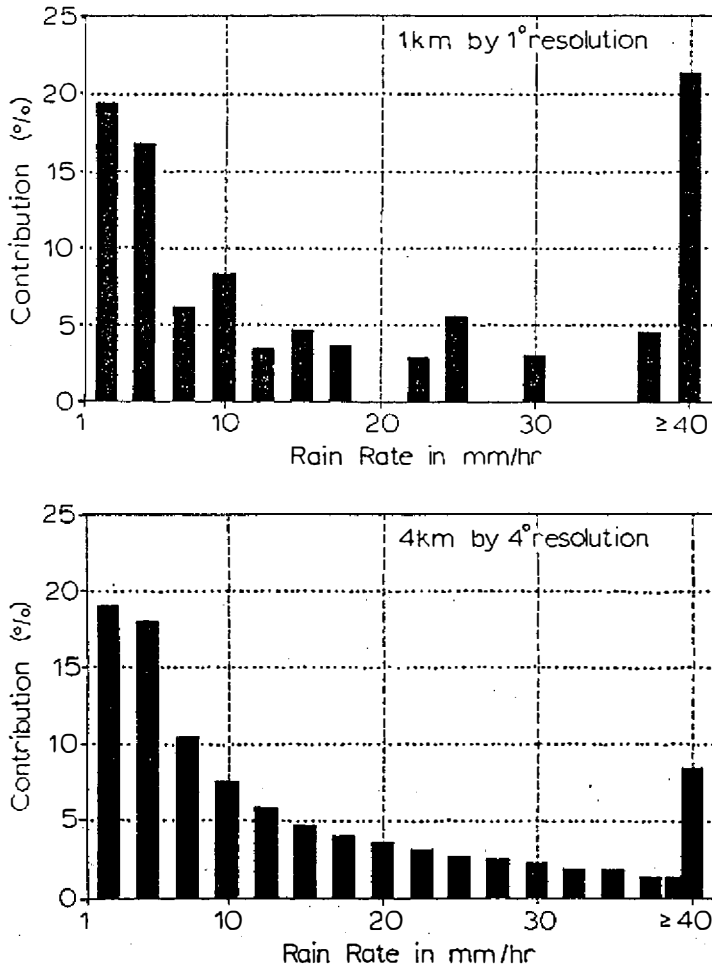
Fig. 2   The contribution of various rainrates to the total rainfall for 1 *km* by 1° and 4 *km* by
          4° resolutions.

a smaller region. Table 2 shows the estimated slope between these two res-
olutions. Some of the data in Table 2 is plotted in Fig. 4a. The slopes as
shown from Table 2 are all concentrated in the neighborhood of 0.166, with
0.024 as the standard deviation. In general, the variance is in logarithmics
$V = C - 0.166 \times A$, where C is an empirical constant, A is the area under
investigation and V is the logarithm of the variance asociated with the area A.

The slope is apparently larger than those deduced from GATE phases I
and II which used the 4 *km* by 4 *km* grid as the smallest spatial resolution.
Figure 4b from North (1987) shows that the decrease of variance for an area-
averaged rain rate decreases at about power law $-1/3$. The finding is consistent
with the $-2/3$ spatial correlation power law found by Bell (1987). They used
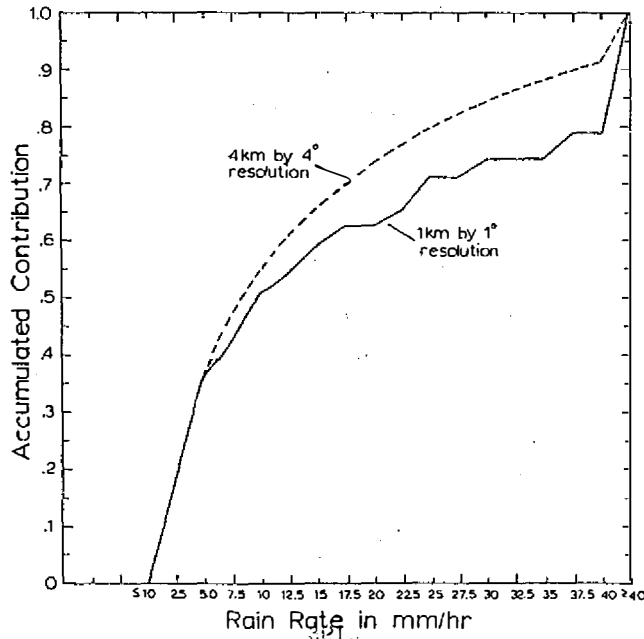these results to argue the non-existence of the horizontal integral length scale

Fig. 3   The accumulated rainfall contribution from different rainrate for the two resolutions used in Fig. 1.

Table 1.   The variance of 1 *km* by 1° (D1) and 4 *km* by 4° (D4) resolutions with various radar sweep radii.

VARIANCE

	D1 date	in log	in log difference	D4 date	in log
50 km	9.357	0.971	0.1945	5.979	0.777
54 km	9.162	0.962	0.2233	5.479	0.739
58 km	7.337	0.866	0.175	4.898	0.690
62 km	7.783	0.891	0.140	5.638	0.751
66 km	8.384	0.923	0.151	5.345	0.730
70 km	6.944	0.841	0.192	4.459	0.649
74 km	7.263	0.861	0.187	4.718	0.674
78 km	6.859	0.836	0.209	4.235	0.627
82 km	6.200	0.792	0.223	3.707	0.569
86 km	5.396	0.732	0.238	3.114	0.493
90 km	4.529	0.656	0.217	2.750	0.439
94 km	3.927	0.594	0.209	2.426	0.385
98 km	3.904	0.592	0.234	2.279	0.358

The expected difference in the variance from the −.3392 power law is 0.41.

Table 2.   The estimated slope for the variances at 1 *km* by 1° (D1) resolution to reach the area associated with 4 *km* by 4° (D4) resolution.

	Estimated Slope
50 km	-0.162
54 km	-0.186
58 km	-0.146
62 km	-0.117
66 km	-0.126
70 km	-0.16
74 km	-0.156
78 km	-0.174
82 km	-0.185
86 km	-0.198
90 km	-0.181
94 km	-0.174
98 km	-0.195

Mean=−0.166

Standard Dev.=0.024

in the rain field.

As illustrated in Fig. 4a, the solid line should be the estimated variance increase as the power law of −0.33, and the dashed line is the actual increase given by the present study. Here the variance which approached from the even larger area is assumed to be at a rate of −0.33 power law. Our present study indicates that the change of variance for the area-averaged rain rate between 1 *km* by 1° and 4 *km* by 4° decreases at a power law of −0.166, or we may write it as V is proportional to $A^{-0.166}$. This small power suggests that the variance associated with the smaller spatial resolution falls off more slowly. That could be an indication of the existence of an integral length scale. This is in contrast to the spatial region which does not seem to have a preferrable horizontal length scale. We are hypothesizing that there exists a horizontal length scale somewhat smaller than 4 *km*. Physically, this may be related to the rainfall characteristic associated with the convective scale. The next section contains more discussion on this.

c.   Autocorrelation

The lagged autocorrelation of these two resolutions are computed from Eq. (3). Figure 5 shows the lagged autocorrelation of the rainfield with respect to the distance for two resolutions at a radar sweep radius of 102 *km*. The 4 *km* by 4° data has a higher correlation distance for a given correlation value than the 1 *km* by 1° data set. This fact can be illustrated by the e-folding distance. The e-folding distance of 1 *km* by 1° data is about 5 *km*, while the 4 *km* by
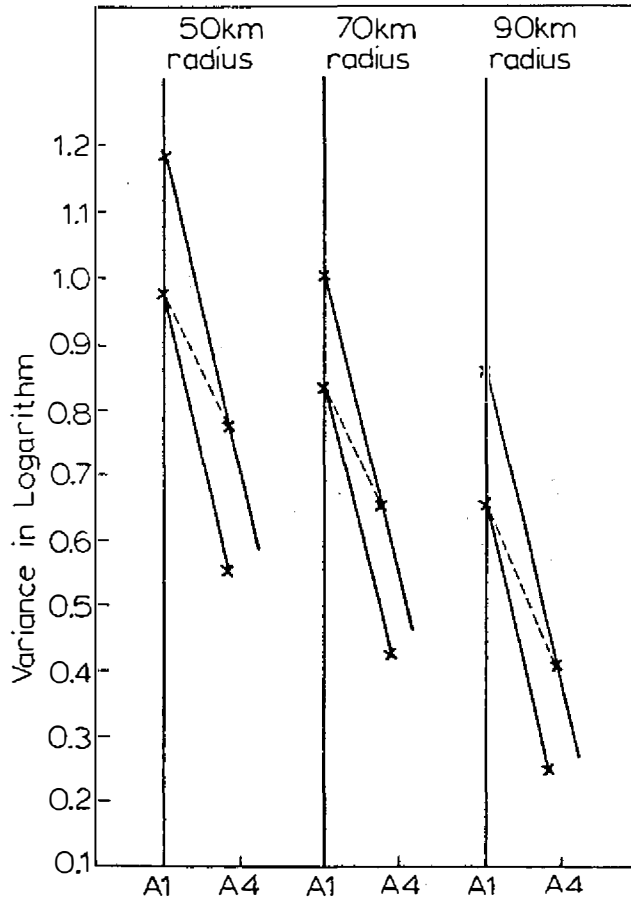
Fig. 4a   The plot of some of the data points from Table 2. The solid line indicates that variance
increased at -0.33 power law, while the dashed line is the actual computation to reach
the 1 km by 1° area.

4° data is about 25 km. The decay of the autocorrelation in the logarithm
for the 4 km by 4° data on the large correlation end is close to a straight line
with a slope of around -0.6. Over the small correlation end, the shape of the
autocorrelation is not a straight line. For a straight line in a log-log plot, the
$ACRF(s)$ is proportional to s^{-k}, and k is the estimated slope in the log-log plot.
For a large correlation (or smaller distance), the 4 km by 4° autocorrelation
behaves as a -0.6 power law.

For the 1 km and 1° data, the logarithmic ACRF decays close to a straight
line with a slope between -0.58 and -0.667 in the large correlation end. How-
ever, at distances of $2 \sim 4$ km, the decaying slope is slower than the estimated
straight line slope. Figure 6a shows the lagged ACRF for 1 km by 1° data at
various radar sweep radii from the center of radar measurement. Although the
detailed structure of the ACRF with respect to various radii differ, they do have
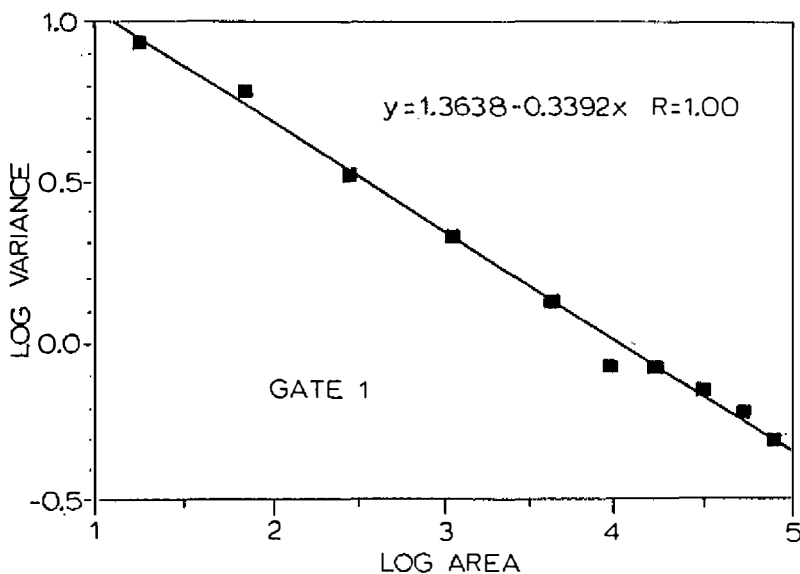
Fig. 4b   The logarithmic plot of the area-averaged variance as a function of the area, which had a 4 *km* by 4 *km* resolution, from North (1987).
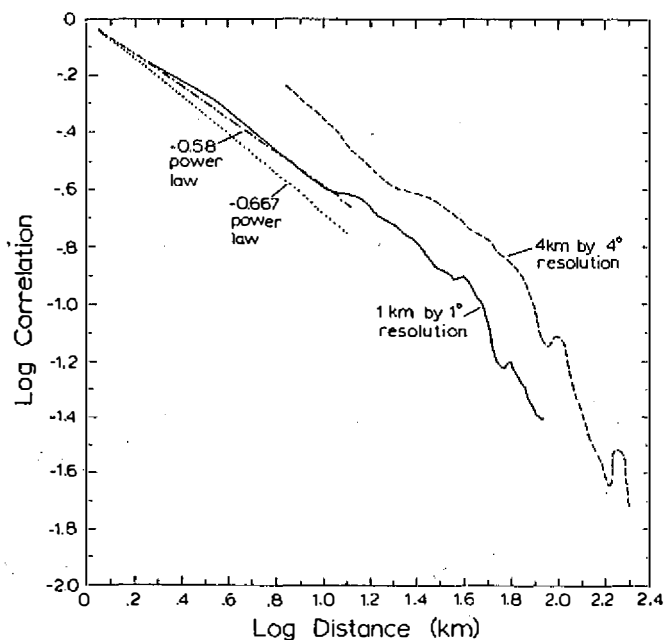

Fig. 5   The logarithmic plot of lagged ACRF of two resolutions used in Fig. 1, with the method described in Eq. (3) for a radius of 102 *km*.

the same trend to slower decay for much smaller distances. The corresponding field for a 4 *km* by 4° data set at a different radius is shown in Fig. 6b. In the
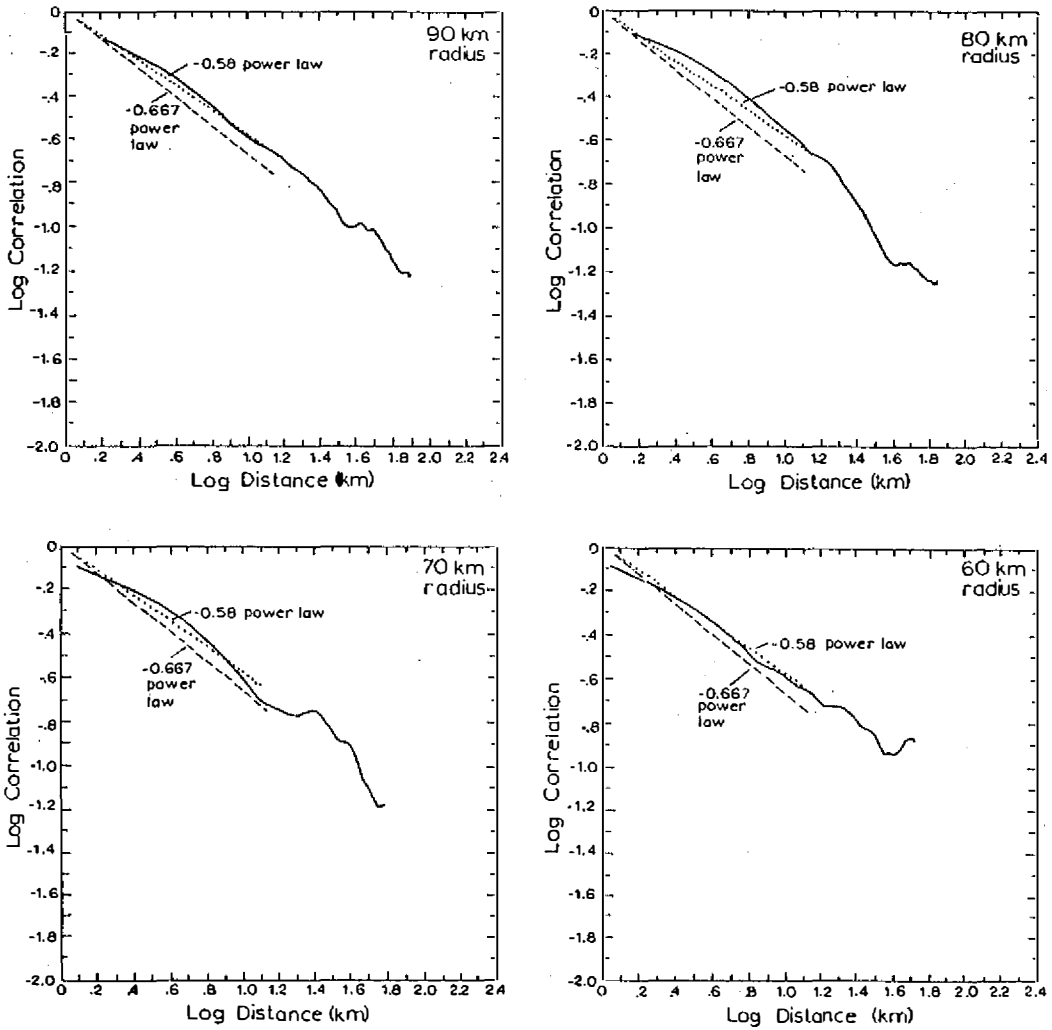
Fig. 6a   The logarithmic plot of lagged ACRF for a 1 *km* by 1° resolution, for various radii from the radar measurement center.

shorter distance end, the lagged ACRF are all close to a straight line, and are quite different from those of the 1 *km* by 1° resolution.

This slower decay of the ACRF at the shorter distance end has significant implications. This aspect is consistent with the variance results discussed earlier. This is also an indication that an integral length scale exists somewhere for a shorter distance to make the characteristics of the ACRF differ when compared to the outer distance end, due to the physical characteristic of the convective scale associated with the rainfield. The typical convective rain scale is only on the order of several *km*. On the other hand, the rainfall from the mesoscale stratiform cloud region will extend for a much larger outer distance.
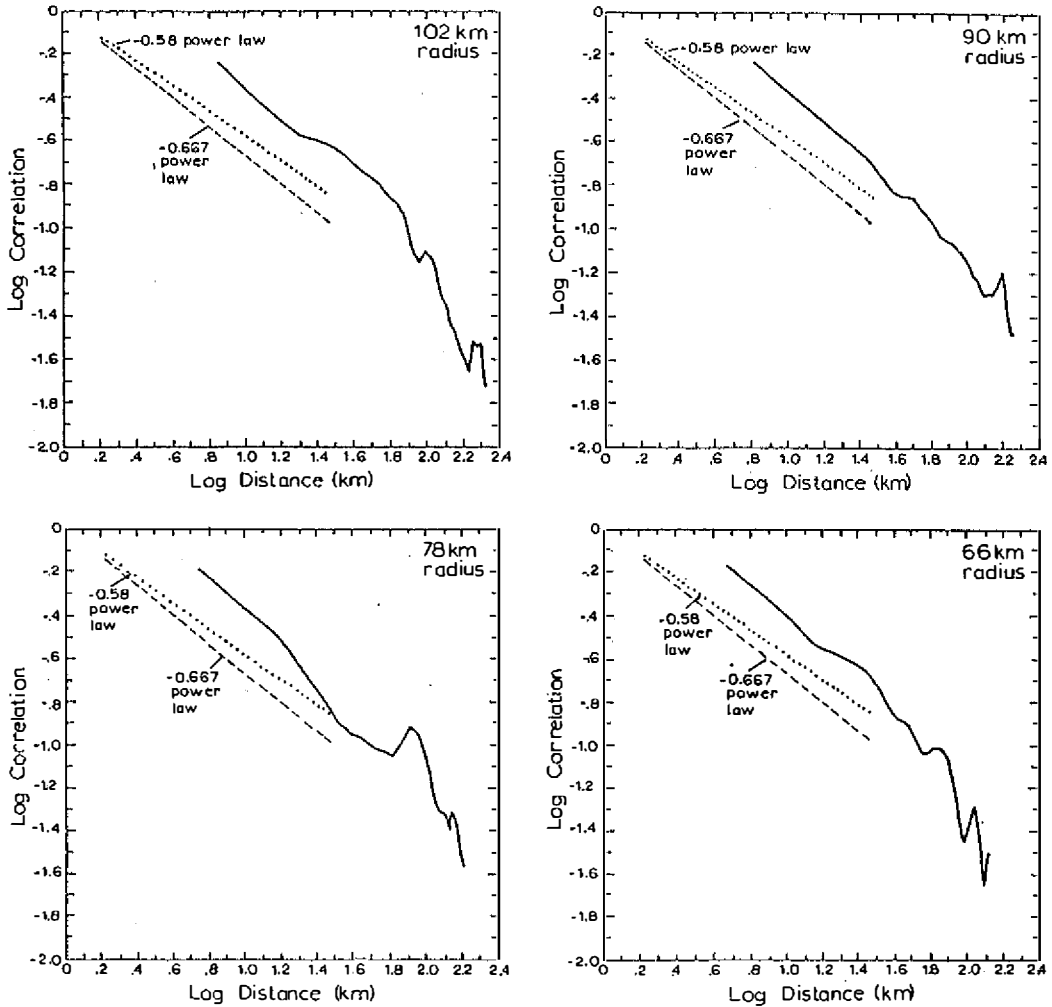
Fig. 6b   Same as Fig. 6a, except for the 4 *km* by 4° resolution. However, they are at different radii.

For the small correlation end, the slope of the logarithmic ACRF is also far from a straight line, both in 4 *km* by 4° and 1 *km* by 1° data. We attributed this departure from a straight line is due to the smallness of the sample size used in the present study, which only included 8 partial days of rainfall data, in contrast to the 18 days of data used by Laughlin (1981) and Bell (1987). Due to the small size of the data, the results for large spatial distances have not been able to respond to the ample information contained in the horizontal stratiform rainfall region. This will be improved a great deal when more data (in a temporal sense) is included.

4.   COMPARISON AND IMPLICATIONS OF THE RESULT

The present study used the GATE phase III data from "Quadra" to estimate the lagged autocorrelation. The original data is in a 1 km by 1° grid. The 4 km by 4° data set used here is constructed from the original data by simple area weighted averaging. Some previous studies of this problem used data from Phases I and II of the GATE (Laughlin, 1981; McConnel and North, 1987; Bell, 1987). A 4 km by 4 km grid was used as the basic data set. Physical characteristics having a scale smaller than 4 km can't be resolved entirely. Figure 7 from Bell (1987) shows the spatial correlation of the GATE I rainfall. The corresponding lagged spatial ACRF in logarithmics is shown in Fig. 8 (North, 1987). Their results show a $s^{-0.58}$ power law decay of the ACRF, and the long-tail of a spatial correlation with a value of about 0.2 extends to a farther distance (at least up to 70 km). Also the power law decay of ACRF can extend to around 70 km before the fitting begins to deteriorate. Laughlin (1981) studied the effect of spatial averaging on the temporal autocorrelation in rainfall fields. He found the e-folding time scale changed from around 1 hr (in 4 by 4 km^2) to more than 10 hr (in 280 by 280 km^2). The temporal correlation has an implicit implication on the spatial correlation. This indicates that the larger area-averaged data should contain a slower correlation decay in the outer distance.

Results from the present study illustrate the correlation characteristics of the rainfall data for the shorter distance end. The spatial ACRF estimated by 4 km by 4° data in the present study shows a power law decay of s^{-k}, k is between (−0.667 and −0.58), which occurs in a comparable structure with the results shown by North (1987). However, the results from a 1 km by 1° dataset revealed a different characteristic of the ACRF for the shorter end. There exist ranges where the ACRF decay is slower than the power law established for the larger area (namely, 4 km by 4° and beyond). The actual slope of this decay in the shorter end varies from radius to radius, but in general, the trend to have a slower decay rate for ACRF exists, and this slower decay of the lagged ACRF is consistent with the $A^{-0.166}$ power law estimated through variance analysis.

As discussed in previous sections, this point is adopted as the basis of our postulation. We thus postulated that the slower decay of the spatial ACRF for the shorter end is due to the existence of a horizontal integral length scale. This length scale is around 2 ∼ 3 km. Physically, we interpreted this length scale to be due to the rainfall characteristic on the convective scale. This observation couldn't have been determined with a 4 km by 4° or 4 km by 4 km data. The characteristic length scale of the convective scale is on the order of several km. The larger data set has a great impact in suppressing the convective scale rainfall characteristics.
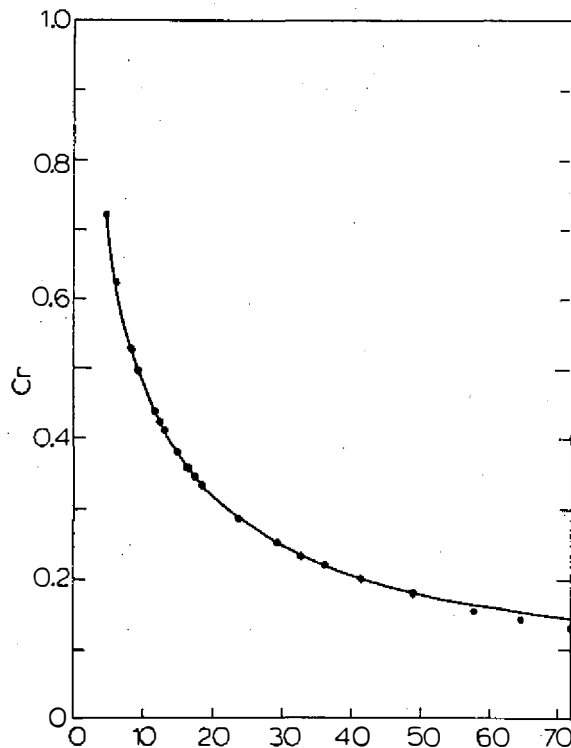
Fig. 7   The spatial correlation of GATE I rainfall from Bell (1987).

So as determined from the GATE study (Houze and Betts, 1981), the convective scale, as well as the mesoscale structure, is an integral part of tropical cloud clusters. The rainfall characteristic associated with these two distinctively different branches have a profound impact on the spatial ACRF structure. The long-tailed structure found by Bell (1987) is associated with continous stratiform rainfall produced from the mesoscale circulation.

The existence of a horizontal integral length scale has a great significance when relating the problems that the TRMM (Tropical Rainfall Measurement Mission, see Simpson *et al.*, 1988 for more details) project encounters. In short, the TRMM is a proposed project to measure tropical rainfall from space through microwave channels and radar. Sampling errors for this project, as well as raingauge measurement, depends on the characteristics of tropical cloud clusters. Present results indicates the existence of an integral length scale. This also implies more numbers of independent samples, which will reduce the rainfall sampling errors estimated through the proposed TRMM satellite.
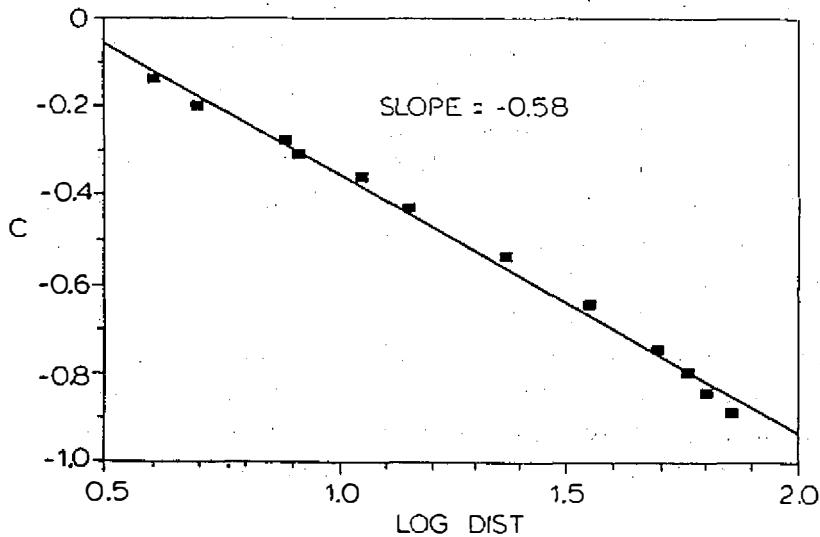
Fig. 8   The logarithmic plot of the lagged ACRF based on GATE phase I data, which had a 4 *km* by 4 *km* resolution as a function of the distance, from North (1987).

5.   CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The present study used a finer data set from the GATE phase III to investigate the spatial effects on rainfall characteristics and its lagged autocorrelation. Because of the finer structure of this present data set, we are able to identify the physical processes invloved in a shorter distance scale, which has crucial information for estimating sampling errors in the TRMM project.

It is found that for a rainrate less than 4 *mm/hr*, the 1 *km* by 1° and 4 *km* by 4° data set will not make any noticeable difference to accumulated rainfall statistics. This is a direct implication from the stratiform rainfall associated with the mesoscale circulation (or rainfall from the mesoscale anvil). The rainfall within that physical region is continuous with a rainrate less than 6 *mm/hr*. A rainfall rate larger than 40 *mm/hr* contributes 20 % of the total rainfall in a 1 *km* by 1° dataset, in contrast to the 8 % contribution from the 4 *km* by 4° data set. The area-averaged process suppresses the extreme rainfall considerably (the extreme would have come from the convective scale). This is another way to indicate that convective scale rainfall always has a smaller spatial scale and an extremely large rainrate.

The variance and the autocorrelation calculated in this study revealed that the slope of the variance and autocorrelation for the smaller scale are different from those estimated through the large area data. This is an indication that the convective scale rainfall field caused this change. The existence of an integral

length scale is a reasonable consequence of this scale of disturbance.

The slow decay of the autocorrelation found by Bell (1987) for large distances (say, up to 70 *km*) is due to the characteristics of homogeneous stratiform rainfall. This fact leads to an interesting problem, since all the studies so far used data which contains these two different scales of motion within tropical cloud clusters (Laughlin, 1981; Bell, 1987; McConnel and North, 1987; and the present study). It will be helpful then to investigate how spatial effects on the autocorrelation field within an area which contains only stratiform rainfall. With this observational study and some rainfall modeling studies that simulate rainfall from the convective scale, we may learn more about the interaction of rainfall statistics (convective scale and mesoscale) as well as how these two physically different processes evolve, reaching the final rainfall field as envisioned from the tropical region. These areas will be the topics for further research.

REFERENCES

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

Esbensen, S. K., J.-T. Wang and E. I. Tollerud, 1988: A composite life cycle of non-squall mesoscale convective systems over the tropical ocean. Part II: Heat and moisture budget, *J. Atmos. Sci.*, **45**, 537–548.

Houze, R. A. Jr., 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 Report, available from US Government Printing Office, Washington, DC 20402, **158**, pp.

Jenkins, G. M., and D. G. Watts, 1968: *Spectoral Analysis and Its Applications*. Holden Day, **523**, pp.

Laughlin, C., 1981: On the effect of temporal sampling on the observation of mean rainfall, *Precipitation Measurement from Space*, edited by D. Atlas and O. Thiele, pp. D59–D66, NASA Goddard Space Flight Center, Greenbelt, MD.

Leary, C. A., and R. A. Houze, Jr., 1979: The structure and evolution of convection in a tropical cloud cluster, *J. Atmos. Sci.*, **36**, 437–457.

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

North, G. R., 1987: Sampling studies for satellite estimation of rain, *Preprint volume on the 10th Conference on Probability and Statistics in Atmospheric Sciemces*. Amer. Met. Soc., Edmonton, Alberta, Canda.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed tropical rainfall measurement mission (TRMM) satellite, *Bull. Amer. Met. Soc.*, **69**, 278–295.

Trenberth, K. E., 1984: Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations, *Mon. Wea. Rev.*, **112**, 2359–2368.

APPENDIX

Bias Analyses for Different Autocorrelation Estimators

In this appendix we will consider the problem that rain fields are so distinctly nongaussian that this might bear strongly on the choice of the statistical estimator. We are interested to know whether the statistial estimators are input dependent or not. Our approach is to build a "time" series using an AR1 process and we will have an exact theoretical answer to compare with our estimator results. We wil use the same estimators as Trenberth (1984) and demonstrate anlytically how these affect the bias of the estimator and compare it with the autocorrelation coefficient obtained from a numerically generated first-order autoregressive process.

a. A first-order autoregressive process

A first-order autoregressive process is assumed to represent a meteorological variable X_n at a typical location:

$$X_n - \mu = \alpha \left(X_{n-1} - \mu \right) + \epsilon_n \tag{A1}$$

where X_n is the value of the variate at time step n, μ a process mean (assumed constant here), α the autoregressive coefficient such that $|\alpha| < 1$ for stationarity, and ϵ_n can be Gaussian or Poisson distributed white noise with variance σ_ϵ^2. Trenberth (1984) employed Gaussian distributed white noise input to derive an approximation of the expected value for the autocorrelation estimator. Note that the Poisson distributed input noise is appropriate for application to the random rain fields process.

For the stationary process, the theoretical values of mean and autocovariance are

$$\mu = \langle X_n \rangle \tag{A2}$$

$$C(m) = \langle (X_n - \overline{X})(X_{n+m} - \overline{X}) \rangle$$
$$= \sigma^2 \alpha^m \tag{A3}$$

where m is the time-lag, and the bracket represents the ensemble averaging procedure and σ^2 is related to the variance of the noise σ_ϵ^2 by:

$$\sigma^2 = \frac{\sigma_\epsilon^2}{1 - \alpha^2} \tag{A4}$$

Notice that the variance of the AR1 process is input noise dependent. The mean estimator for a finite sample size is expressed as:

$$\begin{aligned}
\overline{X} &= \frac{1}{N} \sum_{n=1}^{N} X_n \\
&= \mu + \frac{1}{N} \sum_{n=1}^{N} \phi_n
\end{aligned} \tag{A5}$$

where N is the sample size and the over-bar indicates a time average across the summation of the stationary time series. Even if the data are stationary, the above time mean estimator depends on N since the second term in (A4) does not vanish for finite sample sizes. The random time series $X_n = \mu + \phi_n$ is made up of random fluctuation ϕ_n around the constant mean μ so that the time averaged mean estimators are also random variables. That is, the ensemble average of the above time mean is constant and the above estimator is unbiased but has random error due to finite sample size N.

We are interested in how the uncertainty of the mean estimator affects the estimation of the second-order moments. We will find an analytical form for the autocovariance estimator when the time averaged mean value has an error which is not negligible for finite sample sizes.

b.   The ACRF estimator and its input noise dependency

Taking the ensemble average of the ACVF estimator, we obtain the following expression for method A (from Eq. (5) in the text):

$$\begin{aligned}
\langle C_A^* (i, m) \rangle = \frac{1}{N} \sum_{n=1}^{N-m} \langle X_n (i)\, X_{n+m} (i) \rangle - \frac{1}{N} \sum_{n=1}^{N-m} \langle X_n (i)\, \overline{X}(i) \rangle \\
- \frac{1}{N} \sum_{n=m+1}^{N} \langle X_n (i)\, \overline{X}(i) \rangle + \frac{(N-m)}{N} \langle \overline{X}(i)^2 \rangle
\end{aligned} \tag{A6}$$

Neglecting the differences between the means computed over $n = 1$, $N - m$ and $n = m + 1$, N, Trenberth (1984) obtained the following expression,

$$\langle C_A^* (i, m) \rangle = \left(1 - \frac{m}{N} \right) \left[\sigma^2 - \langle \overline{X}(i)^2 \rangle \right] \tag{A7}$$

Notice that the ACVF estimator is biased but $\langle C_A^* (i,m) \rangle \to 0$ due to the devision of N instead of $N - m$.

Since the presence of pronounced cycles or nonstationarity in the data time series, the differences between the means mentioned above are significant in general so that we use (A6). Inserting Eq. (6) (from the text) into (A6), the expected value of the autocorrelation estimator for method A becomes as follows:

$$\langle C_A^* (i,m) \rangle = \left(1 - \frac{m}{N}\right) \sigma^2 \left[\alpha^m - B(N, \alpha, m)\right] \tag{A8a}$$

where

$$B(N, \alpha, m) = \frac{1}{N^2(1-\alpha)}\left\{(1+\alpha)(N-2m) - \frac{2\alpha}{1-\alpha}(1 - \alpha^{N-2m})\right.$$

$$+ \frac{2\alpha}{1-\alpha}(1 - \alpha^{N-m})(1 - \alpha^m) - \alpha^{N-m}\left[m(\alpha^{1-m} - \alpha^m) - \frac{D}{1-\alpha}\right]$$

$$\left. + \frac{2\alpha}{1-\alpha}(\alpha^{N-m} - \alpha^N) + \frac{2\alpha}{1-\alpha}\frac{m}{N^3}(\alpha^N - 1)\right\} \tag{A8b}$$

$$D = 4\alpha - \alpha^2 - 1 - \alpha^{3-m} - \alpha^{m-1} - (m-1)(\alpha^m + \alpha^{2-m})$$

$$+ (m-1)(\alpha^{m+1} - \alpha^{1-m}) \tag{A8c}$$

Notice that the above ACVF estimator depends on input noise through relation (A4). If both the coefficient of the first-order autoregressive process, α, and the time-lag, m, are zero, the above expression agrees with the variance for independent sampling σ^2/N. The second term of (A8a) is a random error which comes from the random uncertainty of the mean estimator defined by (A5) and vanishes for infinite sample size.

The expected value of the lagged ACRF estimator

$$\langle \rho_A (m) \rangle = \frac{1}{P} \sum_{i=1}^{P} \frac{\langle C_A^* (i,m) \rangle}{\langle C_A^* (i,0) \rangle} \tag{A9}$$

is a consistent estimator of the autocorrelation coefficient in the sense that the limit of large N does not lead to a value different from the true ensemble average. A more important finding is that the ACRF estimator is independent of the input noise, either Gaussian or Poisson. This is because the variance for the AR1 process simply depends linearly on the variance of the input noise as

shown in (A4). In the next section we will compare the exact form of the ACRF estimator for method A and B with the numerically obtained ACRF.

c.   <u>Biases of the different ACRF estimators</u>

For finite sample size N, the random error of ACVF estimate (A6) depends on the size of the data sample such that the ACRF estimator (A8a) also depends on the sample size. The expected value of the ACRF estimator for method A is thus written as:

$$\langle \rho_A(m) \rangle = \frac{1}{P} \sum_{i=1}^{P} \frac{\langle C_A^*(i,m) \rangle}{\langle C_A^*(i,0) \rangle}$$
$$= \frac{\alpha^{|m|} - B(N, \alpha, m)}{1 - B(N, \alpha, 0)} \qquad (A10)$$

which is independent of the input noise, either Gaussian or Poisson.

We obtain a similar form for the expected value of the autocorrelation estimator with method B as:

$$\langle \rho_B(m) \rangle = \frac{\alpha^{|m|} - B(N^+, \alpha, m)}{1 - B(N^+, \alpha, 0)} \qquad (A11)$$

Notice that the above expression has the same form as expression (A16) except for the sample size N^+. Both estimators (A10) and (A11) are plotted in Fig. A1. The upper curve is method B for observations in all the strips $(N^+ = 10000)$. The lower is method A for observations in a single strip $(N = 100)$. Method B is so close to the exact answer, α^m, that the two are nearly coincident. Notice that the smaller the coefficient of the first-order autoregressive process, α, the larger the discrepancy. The numerical estimates are plotted in Fig. A2, for method A $(N = 100)$ and for method B $(N^+ = 10000)$. The upper curve is for method B and the lower is for method A. Since $|\alpha| < 1$ and the number of data points N is usually large compared with time-lag m, the asymptotic estimations of the bias generated by different methods may be written to order $O(\alpha)$ as:

$$b(m) = \frac{(1+\alpha)(\alpha^m - 1)}{(1-\alpha)N^+} \qquad (A12)$$

for method A, N^+ is the number of members of the strip $(N^+ = N)$. Whereas, for method B, N^+ is the number of members of the entire set including all strips $(N^+ = N \times P)$. The above expression for the bias of both methods is used in Equation (8) in the text. Since the autocorrelation time for the series
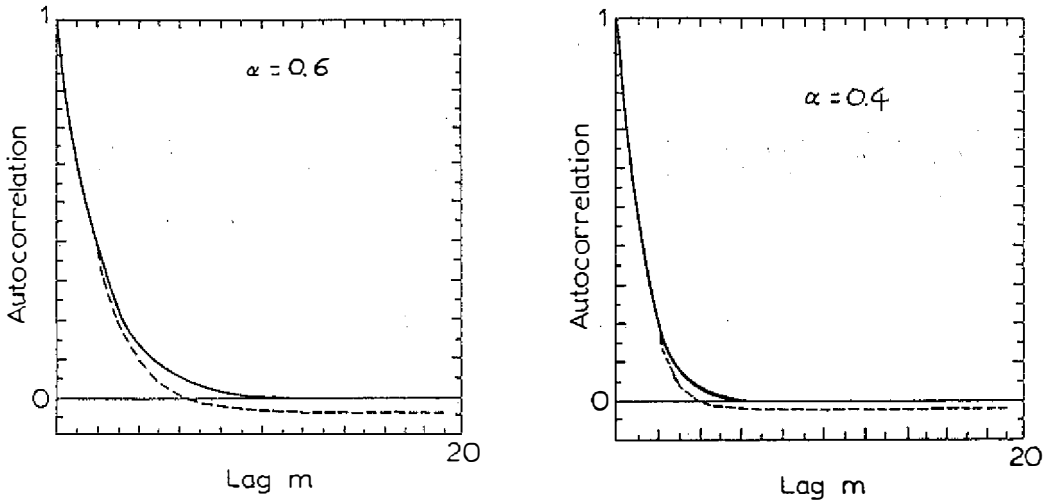
Fig. A1   Theretical autocorrelations for the AR1 process. The upper curve shows method B with 10000 data points, which coincides with the exact value α^m. The lower curve shows method A for strips of length 100. The autocorrelation is shown out to 20 lags for $\alpha = 0.4$ and 0.6.

is $1/(1 - \alpha)$, the product $N^+ (1 - \alpha)$ is just twice the number of independent samplings for the aggregate N^+. Thus the bias is inversely proportional to the independent samples so that method B is a better estimator even if it provides only one estimate of the ACRF form from the data set of sample size $N^+ \times P$.
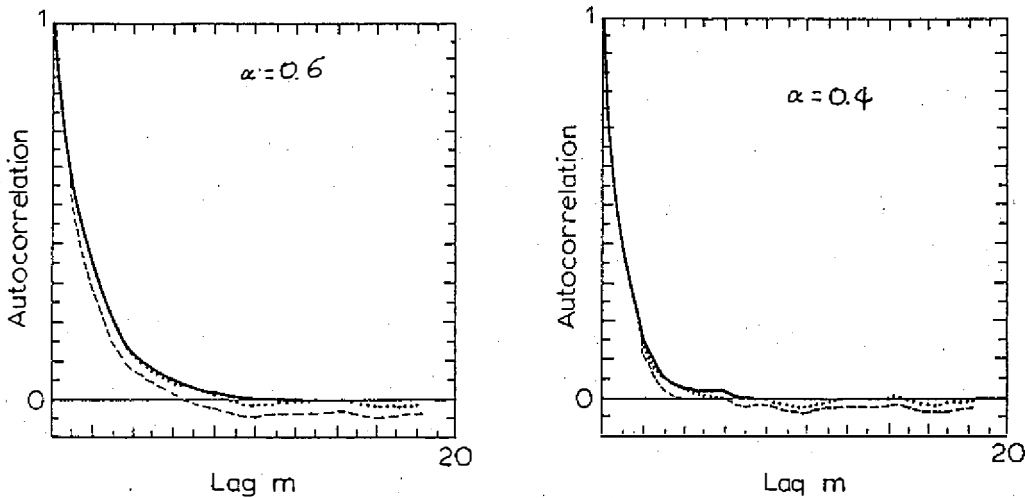

Fig. A2   As in Fig. A1 but for a typical time series of simulated AR1 process with white noise forcing. Notice that the above two autocorrelations identically agree with the theoretical autocorrelations shown in Figure A1.

空間解析度對熱帶降水特性之影響

王作臺

國立中央大學大氣物理研究所

中本正一郎

日本海洋科學技術中心

摘要

本文利用太平洋熱帶實驗中一解析度較好的雷達降水資料，來研究不同空間解析度的取樣對降水總量及 Lag 相關係數之影響，此資料能解析和一些較小尺度相關的物理過程，而這些物理過程是了解及估算降水場抽樣誤差所須要知道的。

此研究發現在降水率小於 $4\,mm/hr$ 時，由 $1\,km \times 1°$ 的資料和由 $4\,km \times 4°$ 得到的累積降水統計沒有明顯的差別。這結果顯然是由中尺度系統中的層雲內降水特性造成的，層雲內的降水是連續而且小於 $6\,mm/hr$，在 $1\,km \times 1°$ 資料中， 20% 的降水量是由大於 $40\,mm/hr$ 的降水率造成的，然而在 $4\,km \times 4°$ 的資料中，此降水率區只有 8% 的貢獻，區域平均的結果，使得極大之降水率被平滑了許多。

變異數及自相關係數之計算顯示了此二變數對空間分佈的斜率而言，由 $1\,km \times 1°$ 及 $4\,km \times 4°$ 的資料中得到的是有顯著的不同，這說明了小對流尺度的降水特性造成的差異，由此吾人推論在較小尺度的干擾中存在一水平積分尺度，本文同時討論了些水平積分尺度存在的影響及意義。