NSG 5295

Some Analysis on The Diurnal Variation of Rainfall Over The Atlantic Ocean

Tepper Gill, Shien Perng and Arthur Hughes

Department of Mathematics

Howard University

Washington, D.C. 20059

(NASA-CR-164937) SOME ANALYSIS ON THE DIURNAL VARIATION OF RAINFALL OVER THE ATLANTIC OCEAN (Howard Univ.) 55 p HC A04/MF A01 CSCL 04B

N82-11724

Unclas

G3/47 08234

August, 1981

This research was carried out under a grant from the NASA, Goddard Space Flight Center.

Acknowledgement

The authors wish to thank Mr. Charles Laughlin and Dr. Raymond Wexler of National Aeronautics and Space Administration, Goddard Space Flight Center, for their assistance and helpful discussions throughout the period of research, and Mr. Vernon Paterson of National Oceanic and Atomospheric Administration for his assistance in data preparation.

Summary

Diurnal variation of rainfalls over the oceans has long been a great concern of members in the meteorological community. In this paper we examined such variation using part of data collected from the GARP Atlantic Tropical Experiment (GATE), which was conducted in 1974 in the North Atlantic ocean. The data were collected from 10,000 grid points arranged as a 100 x 100 array; each grid covered a 4 square km area. The amount of rainfall was measured every 15 minutes during the experiment periods using c-band radars. We analyzed data collected in Phases I and II of GATE which were conducted between 179th and 197th days and between 209th and 227th days of the year, respectively.

Two types of analyses were performed on the data: Analysis of diurnal variation was done on each of grid points based on the rainfall averages at noon and at midnight, and time series analysis on selected grid points based on the hourly averages of rainfall.

Since there are no known distribution model which best describes the rainfall amount, nonparametric methods were used to examine the diurnal variation. This kind of methods was selected because of its model free nature. Kolmogorov-Smirnov test was used to test if the rainfalls at noon and at midnight have the same statistical distribution. Wilcoxon signed-rank test was used to test if the noon rainfall is heavier than, equal to, or lighter than the midnight rainfall. These tests were done on each of the 10,000 grid points at which the data are available.

Among 10,000 grid points, data are not available at 1,872; these grid points are around the boundary of the square covered by GATE. In addition, there are 1,743 grid points where no conclusion was drawn due to insufficient frequency of rain at noon or at midnight during the experiment periods. Both Kolmogorov-Smirnov and Wilcoxon signed-rank tests were conducted at the rest of 6,385 grid points.

With Kolmogorov-Smirnov test, it was found that at one out of 10 chance of error, the rainfall distributions at noon and at midnight are same at 5,425 grid points and different at 960 grid points. In term of percentage, they are 15.0 and 85.0%, respectively. This is in contrast to 10 and 90%, respectively, if the assumption of no difference between noon and midnight rainfall distributions had been true. A chi-square test showed that this split of percentage was not followed. There are much more grid points where the assertion of difference was made than expected. Thus, overall the temporal rainfall distribution at noon and at midnight are different.

With Wilcoxon signed-rank test it was found that at one out of 10 chance of error, the numbers of grid points at which the noon rainfall is less than, equal to, and more than the midnight rainfall are, respectively, 430, 4,890, and 1,065. In term of percentage, they are 6.7, 76.6, and 16.7%, respectively. This is in contrast to 10, 80, and 10%, respectively, if the assumption of no difference had been true. A chi-square test concluded that the midnight rainfall is not equal to the noon rainfall; the noon rainfall is convincingly higher than the midnight rainfall.

Time series analysis was conducted on some selected grid points in both Phases I and II. This analysis is designated to detect if there is a short term cycle in the rainfall pattern during the experiment periods and to examine the temporal correlational behavior of the rainfall.

For Phase I, 20 grid points are randomly selected from 8,128 grid points at which data are available and for Phase II, 16 grid points are strategically selected. For these selected grid points the hourly averages of rainfall are obtained. For each of these grid points, we thus obtained a time series consisting of the hourly averages of rainfall. There are 36 time series, 20 for Phase I and 16 for Phase II.

Although there were no uniform shape of the autocorrelation functions of these time series, 9 out of 20 in Phase I and 9 out of 16 in Phase II are basically negative exponential curves. The values for the auto-correlation decrease rapidly as the values for lag increase. The medians of the auto-correlations for these series at lags 1, 2, 3 and 4 are 0.46, 0.17, 0.15, and 0.10 for Phase I and 0.41, 0.12, 0.08, and 0.04 for Phase II, respectively. The medians at lags 5 or greater are not significantly different from 0 for series in both Phases.

There are short term cycles detected in some of the time series. These cycles are sparse and irregular. For Phase I, a 10-hour cycle is found in one series, 15-hour cycle in another series and 25-hour cycle in another one. A 12.5-hour cycle is found in two series. For Phase II, the cycles found are usually longer; 13-hour cycle in two series, 50-hour cycle in 3 series and a 100-hour cycle in one series. Thus, short term cycles exist in the oceanic rainfall, but they are not prevail.

It is interesting to note that the rainfall distribution in Phase II is very even among all locations. They vary little from location to location. But in Phase I, the story is different. The value for the mean ranges from 0.0014 cm/hr to 0.1320 cm/hr in Phase I; the latter is 94 times of the former. The variation in the variance is also dramatic in Phase I. The largest variance is 9,349 times of the smallest. For Phase II, it is only 25 times. This phenomenon probably associates with the heavy thunder storm activities usually happen in June or July during which Phase I of GATE was conducted.

Table of Contents

Chapter I: Objectives and Background

1.1 Introduction

Many members of the meteorological community have long been concerned with the possibility of diurnal rainfall variation over the oceans. A comprehensive study of the observational evidence for the acceptance of the hypothesis that such a variation exists was completed by Jacobson (1976). Gray assumes that this is the case and proceeds to attempt an explanation based on a radiational cooling profile theory. Much of the observational data used by Jacobson was based on small island data in the Pacific.

It should be noted that the work of Jacobson points to the existence of a maximum in the morning and a minimum at night; this is in contrast to a recent paper of Weickman, Long and Hoxit (1977), where a maximum appears at night and a minimum appears in the morning. The Weickman (et al) paper is based on GATE ship measurements over the mid-Atlantic ocean. Assuming that Gray's theory explains the variation over the Pacific; the Weickman (et al) study makes it clear that we cannot expect the same theory to apply to the Atlantic.

Questions immediately arise, both of a theoretical and a methodological nature. On the theoretical level, one wants to know if the variation is seasonally dependent and how; is it latitudinally or longitudinally dependent and how; are there any significant fluctuations in the variational distribution from year to year in a given season; how does this affect our current overall estimates of world rainfall rate; how does this affect current sampling plans to estimate the oceanic rainfall budget?

On the methodological level, there have been grumblings concerning the possible belief in any of the reports on rainfall variations over the oceans. This is due to many factors: the lack of quality control over data collected from both small islands and ships; the lack of any reasonable distribution of reliable data collection sources over a given region; the use of unreliable and/or untested data manipulation methods and statistical procedures.

The purpose of this report is to present a comprehensive unbiased analysis of the question of the existence or non-existence of a diurnal rainfall variation over the Eastern Atlantic. Our study is based on use of validated radar data from a dense network of ships in the Eastern Atlantic during phases I and II of GATE. In order to clearly define the scope of our study, the specific research objectives were:

I. To determine if diurnal rainfall variations exist and determine when maxima and minima occur.

II. To identify and analyze periodic behavior in oceanic rainfall.

III. To develope a map showing those areas where rainfall variations exist.

Because of criticisms of previous efforts, we attempted to constrain our approach methodologically in such a manner as to minimize technical objections to our conclusions. This lead us to the following specific methodological objectives.

I. To determine if there was any mathematical difference in the empirical distributions of rainfall at morning verses evening.

II. To determine if there was any statistical difference between the mean hourly rainfall rates in the mornings verses evenings.

III. To determine if there was any temporal correlation between rainfall in morning and evening.

1.2 Background

The GARP Atlantic Tropical Experiment (GATE) was conducted in 1974 using a total of twelve ships arranged in two hexagonal arrays which were exact for the first two phases and distorted during the last phase. It is clear from Figure 1,

Acknowledgement

The author wish to thank Mr. Charles Laughlin and Dr. Raymond Wexler of National Aeronautics and Space Administration, Goddard Space Flight Center for their assistances and helpful discussions throughout the period of research, and Mr. Vernon Paterson of National Oceanic and Atmospheric Administration for his assistance in data preparation.

the extent and nature of the distortion. The ships in the inner hexagon were 165 Km apart while those in the outer hexagon were .445 Km apart. This report is based only on data from phases I and II.

The ships in the array made standard surface synoptic observations on an hourly basis and upper air measurements every six hours, such measurements were increased in frequency during periods of precipitation. Many ships made oceanographic and radiation measurements, collected boundary layer profile data, and recorded surface meterological data; however, the data used in this study come from the use of two c-band radars used to obtain special distributions of rainfall not available from other sources.

The c-band radars were choosen to minimize attenuation problems and maximize spatial resolution. The radars were equipped with automatic digital processing and recording equipment and the antennas were stabilized to compensate for possible roll and pitch.

Data was collected at 15 min. intervals on a 24 hour basis. Antenna tilt sequences of 360° scans at a series of 12 increasing tilt angles were collected out to a range of 250 km. The reader is referred to Hudlow (1975) for details and further discussion on the radar systems along with further references.

1.3 Approach.

The distribution of rainfall has been studied extensively from many different vantage points and a number of probability models have been developed. Thom (1968), Simpson (1972), Johnson and Mielke and Mielke (1973) are among the major workers in this field. There is no single distribution which best describes rainfall and any proposed model may be criticized at a number of levels. In particular, the very nature of the diurnal rainfall variation makes the use of any model problematic.

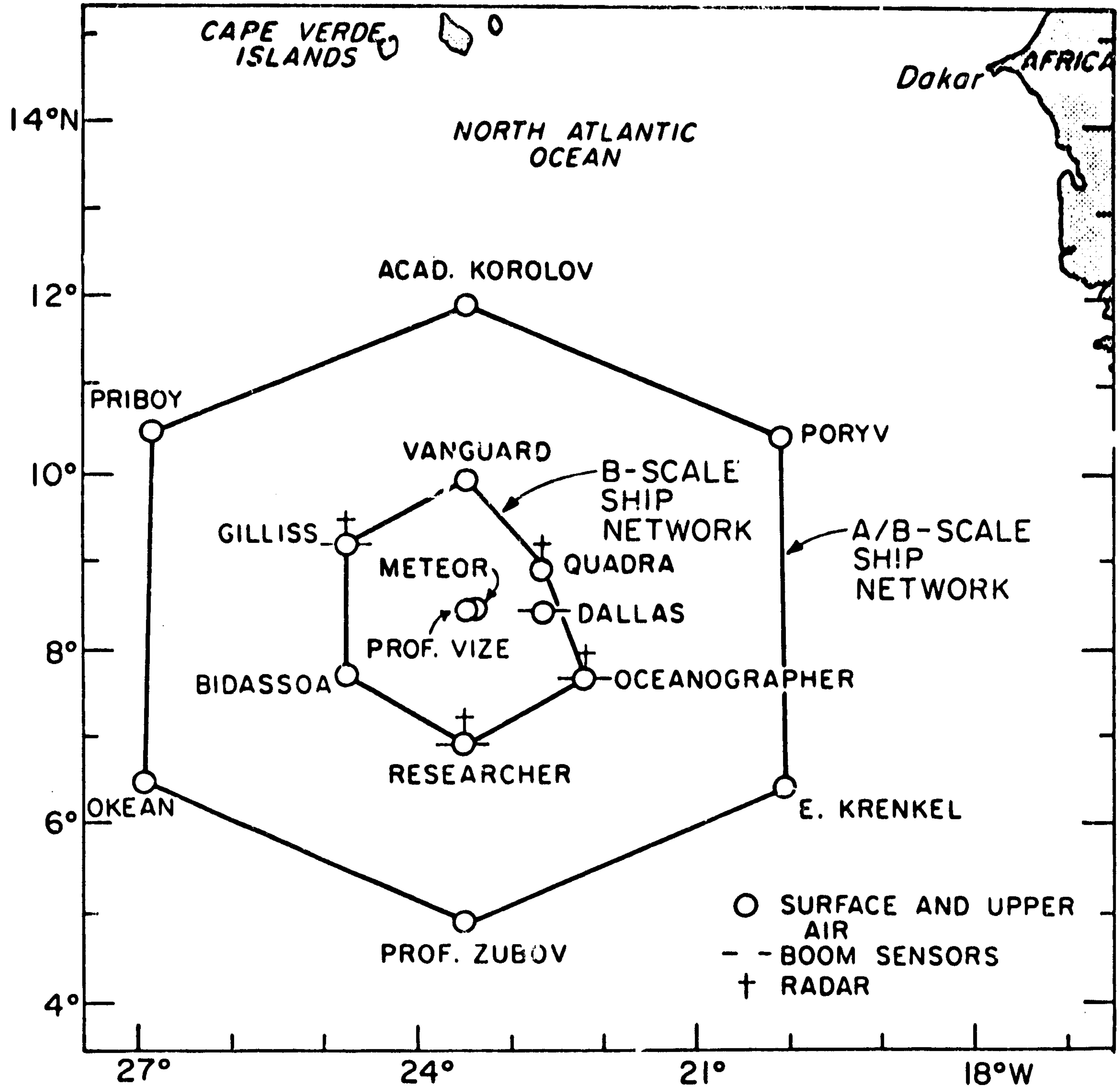

CAPE VERDE
ISLANDS
Dakar
AFRICA
NORTH ATLANTIC
OCEAN
14°N
12°
10°
8°
6°
4°
27°
24°
21°
18°W
ACAD. KOROLOV
PRIBOY
PORYV
VANGUARD
B-SCALE
SHIP
NETWORK
GILLISS
QUADRA
A/B-SCALE
SHIP
NETWORK
METEOR
DALLAS
PROF. VIZE
BIDASSOA
OCEANOGRAPHER
RESEARCHER
OKEAN
E. KRENKEL
SURFACE AND UPPER
AIR
BOOM SENSORS
RADAR
PROF. ZUBOV

Fig. 1

The approach taken in this study is based on the use of three powerful methods in mathematical statistics. The first two methods are used to accomplish the first research objective and part of the others. We use two non-parametric statistics, the Kolmogorov-Smirnov and the Wilcoxon signed rank. Non-parametric statistics are by definition, those statistics that are independent of any particular model for the distribution which generates the data, and thus, all conclusions drawn are model independent.

This approach only depends on the assumption that the underlying distribution is of continuous type; that is that measurements take values in a continuous interval or union of intervals and does not take any specific value with positive probability. It is clear that rainfall measurement is not of continuous type, and in fact takes the value zero with large probability. For our study, this is not a major problem since we are only interested in comparisons conditioned on the event of rainfall, thus our approach uses the method of conditional probability. The measurement of comparisons conditioned on the event of nonzero rainfall is a continuous distribution. For related applications, see McAllister (1969) and Gringarten (1970).

The third method used in this report is based on the assumption that the hourly rainfall averages may be viewed as a time series. This approach allows us to examine the serial correlation of the hourly rainfall, to detect short term cycles, to measure and compare rainfall at different locations and to establish confidence intervals for the mean value of hourly rainfall at different locations. We are thus able to accomplish our last two research objectives within imposed methodological constraints.

In the second chapter we discuss the Kolmogorov-Smirnov and Wilcoxon signed Rank statistics and analyze their application to the problem of rainfall variation between noon and midnight. In the third chapter, we apply time series methods to the hourly rainfall data and provide a detailed analysis of results. In Chapter 4, we report results of the diurnal analyses.

Chapter 2: Statistical Methodologies

In this section we discuss the statistical foundations which underpin our approach. Since the Wilcoxon test is well documented in the literature and well known to applied researchers, we denote a major portion of this section to the explication of the Kolomogorov-Smirnov statistic which is less used and less known in applied research circles.

2.1 Kolmogorov-Smirnov.

We would like to compare the distribution of rainfall and rainfall rate at noon with that occuring at midnight. We expect that if there is a diurnal variation then these two distributions may be different. It is possible that there could be a diurnal variation and yet the distributions are the same, so that this assumption is biased in favor of no variation.

Let $(x_1, y_1), \ldots (x_n, y_n)$ represent n-observations of rainfall over some region G. The x's represent the measurement at midnight and the y's at noon. In order to satisfy our assumption of continuity, we only consider those pairs (x, y) for which either $x > 0$ or $y > 0$. Let $F^1(x)$ be the cumulative distribution function for midnight and $F^2(y)$ be the cumulative distribution for noon. These are the unknown true conditional distributions. We can now state that if there is no variation, then we expect that $F^1(x) = F^2(y)$; that is, it is desired to test the statistical hypothesis:

$$H: \quad F^1(y) = F^2(y) \tag{2.1}$$

verses the alternate hypothesis:

$$A: \quad F^1(x) \neq F^2(y) \tag{2.2}$$

Under this framework, the basic hypothesis (H) is that the distribution of midnight rainfall measure is the same as the distribution of noon rainfall measure. Based on the data, we would like to confirm or reject this hypothesis. In rejecting H we would be making the conclusion that the alternate hypothesis (A) is true.

Since we don't have direct access to $F^1(x)$ and $F^2(y)$ we construct and analyze the empirical distributions $F_n^1(x)$ and $F_n^2(y)$, where n denotes the number of sample points. Define $\varepsilon(\lambda)$ by:

$$\varepsilon(\lambda) = \begin{cases} 1 & \text{if } \lambda \geq 0 \\ 0 & \text{if } \lambda < 0 \end{cases} \tag{2.3}$$

and set:

$$F_n^1(x) = \frac{1}{n} \sum_{j=1}^{n} \varepsilon(x - x_j)$$
$$F_n^2(y) = \frac{1}{n} \sum_{j=1}^{n} \varepsilon(y - y_j) \tag{2.4}$$

Since the superscript identifies the distribution, we shall let λ denote the independent variable so that equation (2.4) now becomes:

$$F_n^i(\lambda) = \frac{1}{n} \sum_{j=1}^{n} \varepsilon(\lambda - \lambda_j) \qquad i = 1,2 \tag{2.5}$$

where $\lambda_j = x_j$ if $i = 1$ and $\lambda_j = y_j$ if $i = 2$. The following theorems show how the empirical distributions are related to the true distributions. We interpret $F_n^i(\lambda)$ to be the proportion of λ_j which are less than λ.

Theorem 2.1. The expectation and covariance of $F_n^i(\lambda)$ satisfy:

1) $E\{F_n^i(\lambda)\} = F^i(\lambda)$ (2.6)

2) $\text{Cov}\{F_n^i(\lambda), F_n^i(\bar{\lambda})\} = \frac{1}{n}\{\min[F^i(\lambda), F^i(\bar{\lambda})] - F^i(\lambda)\,F^i(\bar{\lambda})\}$ (2.7)

Theorem 2.2

1) (Strong law of large numbers)

$$F_n^i(\lambda) \to F^i(\lambda) \qquad \text{(with probability 1)}$$

2) $\lim_{n\to\infty} \frac{\sqrt{n}\,|F_n^i(\lambda) - F^i(\lambda)|}{\log \log n} = F^i(\lambda)\,[1 - F^i(\lambda)]$

3) $\sup_{-\infty<\lambda<\infty} |F_n^i(\lambda) - F^i(\lambda)| \to 0$ (with probability 1) (Cantilli-Glivenko lemma).

Proofs of the above results require very delicate analysis and are presented in order to provide us with the relationship between true and empirical distributions, the relevent literature includes: Gihman 1953, 1954, Glevenko 1933, Gnedenko and Kolmogorv 1933, 1941, and Smirnov 1936, 1937, 1939.

The next two theorems are fundamental to our testing procedure, define E_n^i and D_n^{12} by: $(i = 1, 2)$

$$E_r^i = \sqrt{n} \sup_{-\infty<\lambda<\infty} [F_n^i(\lambda) - F^i(\lambda)] \tag{2.8}$$

$$D_n^{12} = \sqrt{\frac{n}{2}} \sup_{-\infty<\lambda<\infty} [F_n^1(\lambda) - F_n^2(\lambda)]$$

$$D_n^{21} = - D_n^{12} \tag{2.9}$$

Theorem 2.3

1) D_n^{12} is independent of $F^i(\lambda)$ $\quad i = 1, 2$

2) $\lim_{n\to\infty} P_n[\sqrt{\frac{n}{2}}\, D_n^{12} < \gamma] = 1 - e^{-2r^2} = \phi(\gamma)$ for $0 \le \gamma < \infty$

Proof: We shall obtain the proof of theorem 2.3 as a special case of theorem 2.4.

Theorem 2.4. Set $c = [\![2n\gamma]\!]$ (greatest integer) then:

$$\Pr[\sqrt{\frac{n}{2}}\, D_n^{12} < \gamma] = \begin{cases} 1 - \dfrac{\binom{2n}{n-c}}{\binom{2n}{n}} & \dfrac{1}{\sqrt{2n}} \le \gamma \le \sqrt{\dfrac{n}{2}} \\ 0 & \gamma < \sqrt{\dfrac{1}{2n}} \\ 1 & \gamma > \sqrt{\dfrac{n}{2}} \end{cases}$$

Proof: Let us arrange our 2n-observations $(x_1, y_1) \ldots (x_n, y_n)$ in increasing order of magnitude $X_1, \ldots X_{2n}$. Consider the new set of random variables $Y_1, \ldots Y_{2n}$, defined by:

$$Y_k = \begin{cases} 1 & \text{if } X_k \text{ is a midnight observation} \\ -1 & \text{if } X_k \text{ is a noon observation} \end{cases} \tag{2.11}$$

Set $S_0 = 0$ and $S_k = \sum_{j=1}^{k} X_k$. It is easy to see that $nD_n^{12} = \sup_{0 \leq k \leq 2n} S_k$. Let us note that $S_{2n} = 0$; if we plot the points (k, S_k) for $k = 0,1,.-.2n$ in the (u,v) plane and connect these points by straight-line segments, the v component will increase by one unit at n points (corresponding to rainfall observation at noon) and will decrease by one unit at the remaining n points.

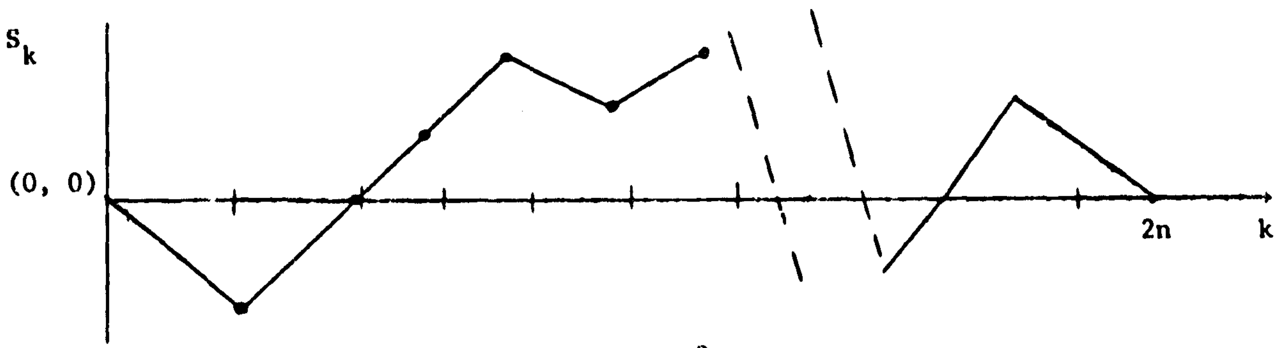


Fig. 2

Figure 2 represents a typical trajectory of the process, the dotted lines indicate that the curve has been broken. Since there are n increases and n decreases, the total number of possible trajectories is $\binom{2n}{n}$; furthermore as our null hypothesis is that both distributions are the same, this means that all trajectories are equally likely. (Another bias in favor of the null hypothesis) Hence, the probability of each trajectory is $\frac{1}{\binom{2n}{n}}$. In our geometric interpretation, the required probability satisfying $\phi_n(\gamma) = \Pr[\sqrt{\frac{n}{2}}\, D_n^{12} < c]$ may be stated as the probability that the entire trajectory will be below the line $u = c$. Now the total number of trajectories which do not cross the line $u = c$ is the same as $\binom{2n}{n}$ - (the number of trajectories which reach the line $u = c$), as there are $n - c$ possible ways to reach the line $u = c$, there are $\binom{2n}{n-c}$ possible distinct trajectories which reach the line $u = c$; thus the number of trajectories which do not reach the line $u = c$ is $\binom{2n}{n} - \binom{2n}{n-c}$ so that:

$$\phi_n(\gamma) = \frac{\binom{2n}{n} - \binom{2n}{n-c}}{\binom{2n}{n}} = 1 - \binom{2n}{n-c}/\binom{2n}{n}$$

We may now provide a proof of theorem 2.3:

Proof: Let us prove part 2) first; set

$$J = \binom{2n}{n-c}/\binom{2n}{n} \tag{2.12}$$

If we use Stirlings formula:

$$k! \cong k^{k+1/2}\ e^{-k}(1 + 0(k))(\sqrt{2\pi}) \tag{2.13}$$

where $\underline{0}(k) \to 0$, $k \to \infty$, and note that:

$$J = \frac{(n!)^2}{(n + c)!\ (n - c)!}$$

we have:

$$J \cong \frac{[n^{n+1/2}\ e^{-n}(1 + 0(n))]^2}{(n + c)^{n+c+1/2}\ (n - c)^{m-c+1/2}\ e^{-(n-c+1/2)}}$$

$$= \frac{n^{2n+1}\ e^{-2n}(1 + \theta(n))}{n^{n+c+1/2}(1 + c/n)^{n+c+1/2}\ e^{-(n+c)}\ n^{n-c+1/2}(1 - c/n)^{n-c+1/2}\ e^{-(n-c)}}$$

$$= (1 + c/n)^{-(n+c+1/2)}\ (1 - c/n)^{-(n-c+1/2)}(1 + \theta(n)) \tag{2.14}$$

Using the Maclaurin expansion for log J, we get to the first order:

$$\log J = -\frac{c^2}{n} + 0(n) \qquad (2.15)$$

Recall that $c = [\![2n\gamma]\!]$ so that $\frac{c^2}{n} \cong 2\gamma^2$, equation (2.15) becomes:
$\log J \cong -2\gamma^2 + 0(n)$ so that:

$$J = \exp\{-2\gamma^2\} \cdot (1 + 0(n)) \qquad (2.16)$$

If we now use equation (2.10) we get

$$\phi_n(\gamma) = 1 - \exp\{-2\gamma^2\}\ (1 + 0(n)) \Rightarrow \phi(\gamma) = \lim_{n\to\infty} \phi_n(\gamma) = 1 - \exp\{-2\gamma^2\}$$

Let us now prove part 1, we will show that D_n^{12} is independent of $F^1(\lambda)$ by showing that E_n^1 is independent of $F^1(\lambda)$; it suffices to show this for E_n^1 since the same method applies to E_n^2 .

Proof:

Define $\nu(t)$ as the inverse function of the event $F^1(s) \leq t$ this means that $\{s \leq \nu(t)\}$ is the inverse event, and this event has probability $F^1(\ (t)) = t$, hence if we set $t = F^1(s)$ then E_n^1 remains unchanged if $F^1(s)$ is replaced by the uniform distribution. Our proof is complete if we recall that given any probability distribution, there exists a transformation which transforms it to the uniform distribution. Hence any statistic which is invariant under the uniform distribution is invariant under all distributions.

Returning to our test procedures, two types of possible error may be committed in confirming or rejecting the basic hypothesis: type I and type II. Type I error occurs if the basis hypothesis H is rejected while it is true and type II error occurs if H is confirmed while it is not true. Type II error is usually used to determine the testing approach. Since our approach has been determined by other constraints, we shall only concern ourselves with type I error.

Let α be the probability of type I error, that is:

$$\alpha = \Pr\ [\text{H is true and H is rejected}]$$

This number α is referred to as the significance level of the test. In this study, we choose α be .10. This implies that there is a 1/10 chance of rejecting H while H is actually true. This level was determined by the number of data points available.

In order to implement our procedure, we must determine a constant γ such that, assuming H is true:

$$\Pr\ [D_n^{12} \geq \gamma] = .10$$

$$\Pr\ [D_n^{12} < \gamma] = .90$$

This gives us a 90% chance of accepting H when H is true. The results of this test are reported in Chapter 4.

2.2 Wilcoxon Signed Rank Test.

Instead of the alternative hypothesis A in the Kolomogorov-Smirnov test, we consider two alternative hypotheses A_1 and A_2 defined as

A_1: The midnight rainfall is (stochastically) greater than the noon rainfall,

A_2: The midnight rainfall is (stochastically) less than the noon rainfall.

Using the notations in sub-section 2.1, A_1 and A_2 can be denoted symbolically as

$$A_1:\ F^1 > F^2 \quad \text{and} \quad A_2:\ F^1 < F^2.$$

Thus, the two testing problems considered in the Wilcoxon signed rank test one.

$$H:\ F^1 = F^2 \quad \text{vs} \quad A_1:\ F^1 > F^2$$

and

$$H:\ F^1 = F^2 \quad \text{vs} \quad A_2:\ F^1 < F^2$$

However, these two testing problems are statistically equivalent, respectively, to H_1: $F^2 \leq F^2$ vs A_1: $F^1 > F^2$ and H_2: $F^1 > F^2$ vs A_2: $F^1 < F^2$. If A_1 is not rejected in the first problem and A_2 is not rejected in the second problem then we conclude that H: $F^1 = F^2$ holds.

As in the K - S test, there are two types of possible error in each of these two problems. Again, we consider only type I errors. To test H vs A_1, type I error is committed if H (or H_1) is rejected while H: $F^1 = F^2$ is actually true, i.e. confirming A_1: $F^1 < F^2$ while H: $F^1 = F^2$ is actually true. The significance level α_1 is given by

$$\alpha_1 = P_H[\text{rejecting } H] = P_H[\text{accepting } A_1].$$

To test H vs A_2, type I error is committed if A_2: $F^1 > F^2$ is accepted while H: $F^1 = F^2$ is actually true. The significance level is given by

$$\alpha_2 = P_H[\text{rejecting } H] = P_H[\text{accepting } A_2].$$

Note that rejecting H implies accepting A_1 in the first testing problem H vs A_1 and accepting A_2 in the second testing problem H vs A_2. In this study, we adopted $\alpha_1 = \alpha_2 = .10$.

Let $(x_1,y_1),\ldots, (x_n,y_n)$ be as defined in subsection 2.1. The testing procedure of Wilcoxon signed test consists of deriving a statistic V_n based on $(x_1,y_1),\ldots, (x_n,y_n)$ and choosing constants C_1 and C_2 $(C_2 < C_1)$ corresponding to α_1 and α_2, respectively, such that for testing H vs A_1

rejecting H (or accepting A_1) if and only if $V_n \geq C_1$

and for testing H vs A_2

rejecting H (or accepting A_2) if and only if $V_n \leq C_2$.

For more detail, the reader is referred to Lehmann (1975).

Note from the above discussion that when H is true, $P_H[V_n \leq C_2] = .10$,

$P_H[C_2 < V_N < C_1] = .£C$ ind $P_H[V_{\text{a}} \geq C_1] = .10$. Thus, when H is true, the chance of concluding that midnight rainfall is (stochastically) greater than, equal to and less than the noon rainfall are, respectively 10%, 80% and 10%.

The result of this test is reported in Chapter 4.

2.3. Chi-square Test.

The tests describe in subsection 2.1 and 2.2 are applied to the midnight and noon rainfall measures at each grid of the GATE data. It was seen that when H is true, the chances of drawing the conclusion that $F^1 = F^2$ and $F^1 \neq F^2$, are, respectively, 90% and 10% and that $F^1 < F^2$, $F^1 = F^2$ and $F^1 > F^2$ are, respectively 10%, 80% and 10%. We use χ^2-test to test whether these percentages are actually obtained in the result.

Let p_j be the probability of making the jth conclusion, $j = 1,2,\ldots, k$, where k is the number of different conclusions (k = 2 in K-S test and k = 3 in Wilcoxon test). Let n_j be the number grids at which jth conclusion is made, $j = 1,2,\ldots, k$ and $n = n_1 + n_2 +\ldots+ n_k$, the number of grids in the study. Define

$$\chi_k^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j}$$

If the given percentages of different conclusions are followed, χ_k^2 should be relatively small. Thus, if χ_k^2 is large, we may conclude that the percentages do not hold true and should have been otherwise. For instance, in the Wilcoxon test, $P_1 = .10$, $P_2 = .80$ and $P_3 = .10$. If χ_3^2 is too large, then the percentages of 10, 80 and 10% do not hold and thus the hypothesis H: $F^1 = F^2$ is not true. To determine "how large" is "too large", a cutoff constant can be ound corresponding to k and the desired significance level of the χ^2-test. With $\alpha = .05$ (i.e. 5% chance of making wrong conclusion), C = 3.84 for k = 2 and C = 5.99 for k = 3.

Results of the χ^2-test are reported in Chapter 4.

Chapter 3

Temporal Analysis on Hourly Rainfall Averages

The hourly rainfall averages were considered as observations in a time series. Analyses on the time series were undertaken for both Phase I and II data in experiment. Properties of such time series at selected locations in the GATE experiment were investigated. In the following sections, the objective of the study, the data and sampling design as well as the analyses and their results are reported.

3.1. Objective of The Study

The objectives of this study are:

- To examine the serial correlation of the hourly rainfall.
- To detect the short term cycles in the rainfall activities.
- To measure and compare the power of rainfall activities at different localities.
- To establish a confidence interval for the mean of hourly rainfall averages.

The serial correlation of hourly rainfall averages reveal the linear dependence of the rainfall on the rainfall of the preceeding hours. These statistics are closely related to the duration of rainfall activities.

High autocorrelation with large lags implies long duration and vice versa. The serial correlation is also related to an assumption in the studies in Chapter 2: The diurnal analysis of rainfall. It was implicitly assumed in the study that the noontime and midnight rainfall rates are independent. Here we consider the corelation between rainfalls several hours apart. While no correlation between rainfalls in different hours do not imply independence between them, they may be considered close enough for practical purpose.

It has been suspected that a two-day cycle exists in oceanic rainfall. One objective of the study is either to confirm or disconfirm this conjecture.

The third objective is to establish a quantity to measure the "ample" of rainfall activities. Power is a term in time series analysis designated to measure the amount of variation in the series. Since for most of the time there is no rain. A large value of this quantity may be a good indication of frequent rainfall or even thunderstorm activity.

3.2. Sampling Design and Data

3.2.1. Sampling Design

In order to have unbiased results as well as to account for the variation of rainfall in different locations, both remdomly selected sample and strategically selected sample were used in the study.

The region of the experiment was put into a coordinate system as shown in Figure 3.2.1. Each grid is assigned a pair of integers between 0 and 100 as longtitude and latitude coordinates. The grid represents a square of 4 km x 4km, which is the spatial resolution for the experiment.

There are 10,000 (100 x 100) grids in the coordinate system. Rainfall measures were consistently observed in 8,128 of these grids, scattering around the center of the big square. For this study, a sample of grids is taken for each phase from those grids in which the rainfall measure is available.

For the Phase I data a randomly selected sample of 20 grids were used. For each grid in the sample, we have a time series consisting of hourly rainfall averages. Thus, from Phase I data of the experiment, we obtain 20 time series, each represents the hourly rainfall averages in one of the selected grids. These 20 selected grids are shown in Figure 3.2.1.

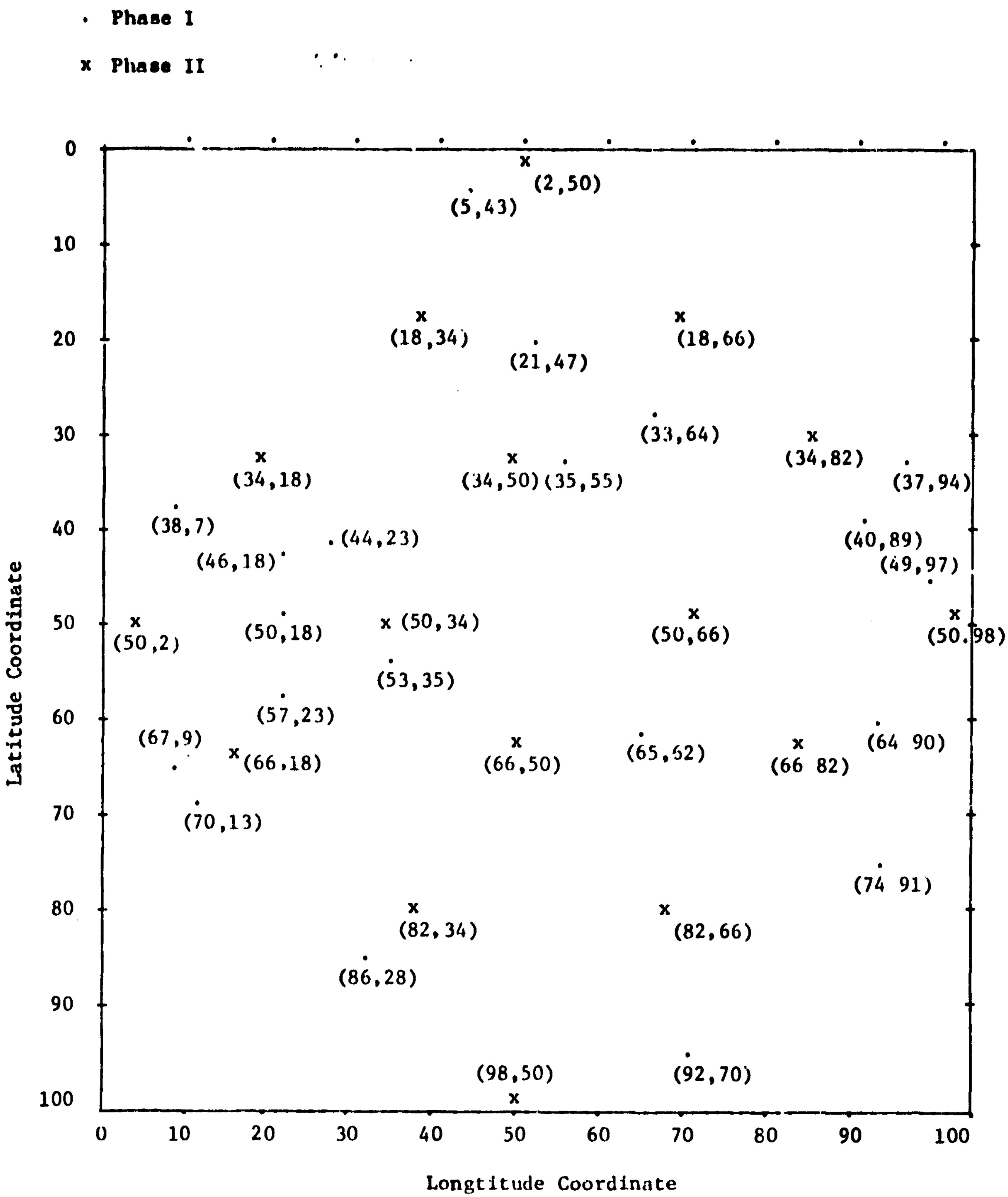


Figure 3.2.1. Grids Selected for Study in Temporal Analysis.

For Phase II data, a strategically selected sample of 15 grids were used. The sampled grids are also shown in Figure 3.2.1.

3.2.2. Data and Missing Values.

Hourly rainfall averages are computed by adding the rainfall measured in the hour and then dividing by the number of rainfall measurements within the hour. This is done for each hour in the duration of Phases I and II in the GATE experiment.

The nature of the study calls for an uninterrupted sequence of hourly rainfall averages. It was found that there are hours in both Phases I and II which do not have rainfall measurements. This situation usually occured in the earlier stage of both phases. We decided to cut short the time series to accomodate our needs. For Phase I, the actual duration used for this study is from the 12th hour of 182nd day to the 17th hour of the 197th day. (The Phase I experiment lasts from the 0th hour of the 179th day to the 17th hour of the 197th day.) (All Julian hours and days). For Phase II the actual duration used is from the 20th hour of the 214th day to the 21st hour of the 227th day. (The Phase II experiment lasts from the 0th hour of the 209th day to the 21st hour of the 227 day.)

In both Phases I and II there still exists an hour without measurement. This is filled with method of interpolation. It is expected that the effect of this is negligible.

We could have used the original rainfall measure of every quarter hour as the observation (term) of the time series, instead of the hourly rainfall average. However, there are so many missing values in the quarter-hourly measures throughout the experiment in both Phase I and II that it is not advisable to use them directly.

3.3. Auto Correlation

It is well known that the hourly rainfall averages of adjacent hours are positively correlated. That is given that there was an above average rainfall in a specific hour, it is very likely that there would be an above average rainfall in the next hour, and possibly in the hour thereafter, etc. Auto-correlation measures the correlation of the hourly rainfall of hours apart. The computation formula are given in subsection 3.3.1 and results reported in the subsequent sections.

3.3.1. Computation Formulas.

Let $x_1, x_2, \ldots, x_n$ denote the hourly rainfall averages at a given grid of the study. Note that n is the number of hourly rainfall averages in the grid. Define the sample mean and sample variance as

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t \qquad S^2 = \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2 .$$

Then the rth auto-correlation of the series is defined to be

$$R_r = R(r) = \frac{1}{n} \sum_{t=1}^{n-r} (x_t - \bar{x})(x_{t+r} - \bar{x})/S^2, \quad r = 1,2,\ldots,k.$$

The number r in the formula is called <u>lag</u> of the auto-correlation, and k is the maximum number of lags for the auto-correlation to be computed. The auto-correlation $R_r = R(r)$ considered as a function of r is called an <u>auto-correlation function.</u>

It should be pointed out that, unlike in a random sample (from a population of rainfalls), $x_1, x_2, \ldots, x_n$ are not independent and identically distributed. The x_t denote the t-th hourly rainfall average starting at some specific hour. The length n of time series is 366 for series in Phase I data and 328 for series in Phase II data.

3.3.2. Correlograms.

The correlogram of a time series is the plot of its auto-correlation function R(r) against the lag r. Just as the histogram is studying sampling problems, the correlogram is descriptive and informative in studying the time series. It is a visual device which is useful to perceive the linear relationship of rainfalls several hours apart and to identify the mechanism generating the rainfall series. However, it is not an objective of this study for the latter, i.e. to identify the model of the series.

In the following paragraphs, we make some observations of the correlograms for time series in both phases I and II.

As said earlier, the length of the time series is 366 for phase I and 328 for phase II. Corresponding to these sample sizes, the approximate 95% confidence interval for the correlation coefficient of 0 is (- .12, .12). That is, if it is hypothesized that the correlation coefficient is zero, the chance of having the computed coefficient to be out of interval (- .12, .12) is 5% or 1 out of 20. The 5% chance of error is a generally accepted level in practice. Thus, if the auto-correlation of a time series at any given lag is in the interval (- .12, .12), we may safely assume the value 0 (with 5% chance of being wrong.)

Examining the correlogram, it is found that there are no uniform and specific shape for the correlograms for all time series. The shape varies greatly from series to series. The length of these time series may be blamed for the non-uniformity in the correlogram. In general, a time series of length 366 or 328 should be long enough to obtain a meaningful result. But for the time series on hand, this may not be true, because there are too many zeros (more than 4 out of 5) for most series) for the hourly rainfall average. It is very likely that just a few hours of heavy or even moderate amount of rainfall at a different time would change the shape of the correlogram dramatically.

First consider the correlograms for rainfall series in phase I.

Even though there are no uniform shape for the correlogram, nine (9) of them share the same basic (negative) exponential shape. These are for time series at grids (5,43), (33,64), (38,7), (40,89), (44,23), (46,18), (49,97), (50,18), and (92,70). For most of these time series, the auto-correlations of lag 5 or higher are small and within the interval (-.12,.12), hence may be considered zero with 95% confidence.

Time series at grids (64,90) and (74,91) have very similar lobed exponential shape of correlogram. This may not come as a surprise because these two grids are only about 40 km apart. The correlogram for the time series at (21,47) have moderately high auto-correlation at lags 10,30, and 42. This may be an indication 10 hours cycle and will be examined in next section (see 4.2).

Time series at grid (40,89) have consistently large auto-correlation at small to moderate lag numbers. This is largely due to a long string of hours with persistent rainfall on 183rd and 184th Julian days at the location.

The magnitudes of the auto-correlation for time series at grids (35,55), (37,94), (38,7), (57,23), (65,62) and (67,9) do not seem to depend on the lag number; the magnitudes do not decrease as the lag number increase. In particular, those at (37,94), (38,7) and (65,62) are almost all negligible at 95% confidence level.

Next, consider the correlogram in Phase II.

The correlogram forms essentially a negative exponential curve for the rainfall averages at grids (18,34), (50,2), (50,34), (50,66), (50,98), (66,18), (66,50), (66,82) and (82,34). Thus 9 out of 16 correlograms assume this shape. The auto-correlations of first order are positive number of moderate magnitude and decrease exponentially as the order increases.

For some of the time series, auto-correlations are significantly non-zero for large lags. For the series at grid (2,50), R(38) is .25, and for the series at (18,66), R(45) is .18 while R(1) is only .24. At (34,18), the auto-correlations raised for lags between 38 and 46 to as high as .24.

At (98,50), the correlogram behaves like a sine wave with peak at lags 15, 28 and 40. The power spectral density will be examined closely if cycles exist.

3.3.3. Median and confidence interval.

The median and its confidence interval of the auto-correlations are obtained and reported in this section. Since the correlation does not seem to be symmetrically distributed and the number of correlations available for the study is limited (20 for phase I and 16 for phase II) we use non-parametric approach instead of the more conventional one where normal (ball shape) distribution are assumed.

Let $y_1 \leq y_2 \leq \cdots \leq y_n$ be the auto-correlation (of any given lag) ordered in increasing order. In Phase I, n = 20 where the 25th and 75th percentiles are y_5 and y_{16} respectively and the medium $m = \frac{1}{2}(y_{10} + y_{11})$. In Phase II, n = 16, the 25th and 75th percentiles are y_4 and y_{13}, respectively and the median $m = \frac{1}{2}(y_8 + y_9)$.

It is well known (see e.g. [33], p. 181) that for $0 \leq i \leq j \leq n$

$$P[y_i < m < y_j] = \sum_{x=i}^{j-1} \binom{n}{x} (\tfrac{1}{2})^x (\tfrac{1}{2})^{n-x} = r.$$

Thus (y_i, y_j) is a 100r % confidence for m, the median. By using a binomial table (e.g. [33]) it is found that for n = 20 (phase I)

$$P[y_5 < m < y_{16}] = .9941 - .0059 = .9882 \cong .99$$

and for n = 16 (Phase II)

$$P[y_4 < m < y_{13}] = .9788 \cong .98.$$

Hence (y_5, y_{16}) is a 99% confidence interval for the medium of the auto-correlation with any given lag in Phase I and (y_4, y_{13}) a 98% confidence interval in Phase II. The chances of error are 1% and 2% in Phases I and II, respectively, in saying that the medium is in the interval.

Table 3.3.1 and Table 3.3.2 summarize the means, medians, 25th and 75th percentiles, ranges, lengths of confidence intervals of the auto-correlations for lags 1 through 12 in Phases I and II data, respectively. The range is defined to be the maximum of correlations minus minimum of the same. Figure 3.3.1 and Figure 3.3.2 show the medium, 25th and 75th percentile along with confidence interval for the medium in Phases I and II data respectively.

It is seen that in both Figure 3.3.1 and Figure 3.3.2, the median with lag 1 is moderate (in 40's) and there is a considerable drop between lags 1 and 2. The mediums with lags 2 and 3 are about the same magnitude. Those with lag 5 or higher are so small that they are negligible. In fact, their confidence intervals contain 0 for most of them. Thus, the hypothesis that median is 0 would not have been rejected.

3.3.4. Concluding Remarks.

The auto-correlation function for each time series in both Phases I and II were studied. The length of series is 366 in Phase I and 328 in Phase II. Since most (more than 80%) of the rainfall are 0, the auto-correlation function seem unstable in sense that the shape of the function are quite different from series to series. However, 9 out of 20 in Phase I and 9 out of 16 in Phase II are essentially negative exponential curves. For these series, the auto-correlation with lag 1 ranges from .46 to .82 in Phase I and from .34 to .68 in Phase II. The value decrease exponentially as the lag increase. Three of the correlation functions show some repeating peaks and valleys. These series may have short term cycles and will be examined for such in the next section. At lag 1, the median of the auto-correlation is .46 for phase I series and .41 for phase II series. At lags 2 and 3, the values for the auto-correlation drop considerably for both phases I and II. The values are not significant for lag larger than or equal to 5. Thus we may say with caution that rainfalls of 5 or more hours apart are uncorrelated. This statement is not always true. There are occasions that the auto-correlation with larger lag is significantly non-zero.

3.4. Power Spectrum

The power spectrum is useful tool to detect a cycle in the time series. If the power at a frequency is large comparing to the powers of neighboring frequencies, then we may conclude that there is a cycle in the time series corresponding to the frequency.

In this section, we use power spectrum estimate of hourly rainfall averages to detect short term cycles. We start with a brief review of the methology in subsection 3.4.1. The analysis and results are reported in subsection 3.4.2.

3.4.1. Power spectrum of the time series.

To facilitate the interpretation of the results of spectral analysis an over-simplified version of time series representation and its analysis is presented here. This version is intended for readers who have no background in time series analysis and want to get hold of some conceptual meaning of the result to be reported in the following subsection. Readers with background in spectral analysis of time series may skip this subsection and proceed to subsection 3.4.2.

We assume that a time series x(t) may be written as

$$x(t) = \sum_{j=-\infty}^{\infty} z_j \, e^{i\lambda_j t}$$

where $i = \sqrt{-1}$, $\lambda_0 = 0$, $\lambda_{-j} = -\lambda_j$ and $z_{-j} = \bar{z}_j$, the z's are random variables (with $E[z_j] = 0$ for $j \neq 0$). The restrictions on the λ_0 and z's imply that x(t) is a real-valued random variable for each t, the time. The representation of x(t) says that

Table 3.3.1 Mean, median and other statistics of Auto-correlations in Phase I

Lags	Mean	min.	25th percentile (y_5)	Median	75th Percentile (y_{16})	Max.	Range	75% - 25%
1		.00	.14	.46	.60	.82	.82	.46
2		-.01	.03	.17	.36	.67	.68	.33
3		-.01	-.01	.15	.22	.59	.60	.23
4		-.04	.01	.10	.18	.51	.55	.17
5		-.03	-.01	.02	.10	.53	.56	.11
6		-.01	-.01	.02	.08	.49	.50	.09
7		-.02	-.01	.02	.04	.40	.42	.05
8		-.02	-.01	.02	.09	.33	.35	.11
9		-.03	-.01	.02	.08	.25	.28	.09
10		-.02	-.01	.02	.15	.29	.31	.16
11		-.03	-.01	.01	.12	.25	.28	.13
12		-.03	-.01	.01	.11	.42	.45	.12

Table 3.3.2 Mean, Median and Other Statistics of Auto-Correlations in Phase II

Lags	Mean	Min.	25th Percentile (1/4)	Median	75th Percentile (1/13)	Max.	Range (max-min)	(75%-25%)
1		.25	.31	[illegible]	.56	.68	.43	.25
2		.01	.08	.12	.30	.44	.43	.22
3		-.02	.04	.08	.29	.35	.37	.25
4		-.02	.00	.04	.26	.30	.32	.26
5		-.02	-.01	.01	.07	.25	.27	.08
6		-.04	-.03	-.01	.03	.23	.27	.06
7		-.05	-.03	-.01	.00	.27	.32	.03
8		-.04	-.02	-.02	-.03	.21	.25	.05
9		-.05	-.02	-.01	.03	.16	.21	.05
10		-.04	-.02	-.01	.02	.11	.15	.04
11		-.04	-.03	-.02	.00	.12	.16	.03
12		-.04	-.03	-.02	-.01	.07	.11	.02

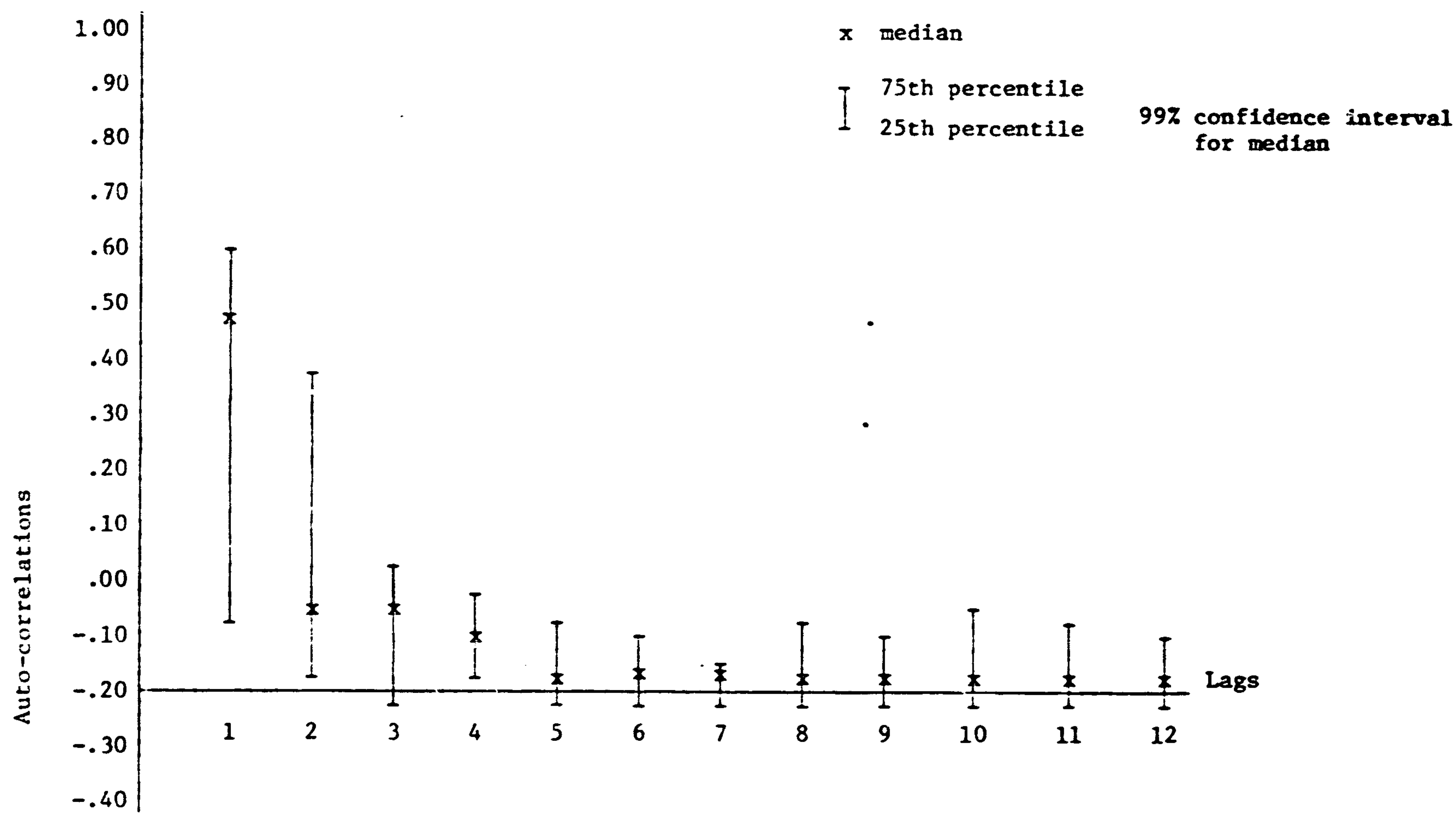


Fig. 3.3.1 Median, its confidence Interval, and 25th and 75th percentiles, of Auto-Correlations in Phase I.

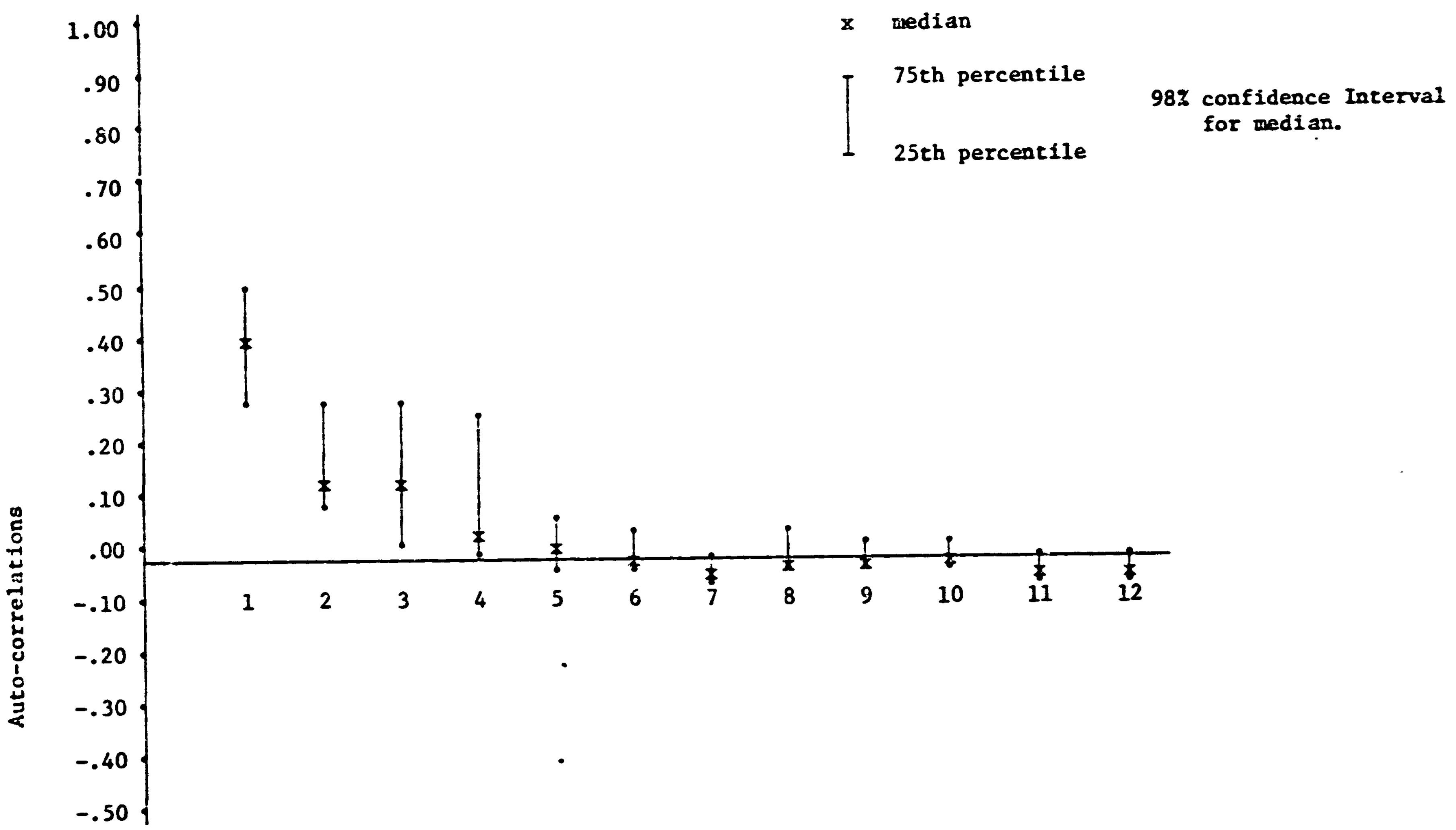


Figure 3.3.2. Median, its confidence Interval, and 25th and 75th percentiles of Auto-correlation in Phase II.

the series can be decomposed into contribution of harmonic at frequency λ_j or that the series is the superimposed sum of harmonics at frequency λ_j.

It can be shown then that the power of the time series can be written as

$$\text{power of } x(t) = \sum_j [\text{power of } x(t) \text{ at frequency } \lambda_j].$$

The power of x(t) at frequency λ_j measure the amount of variation, or intensity, of the series at frequency λ_j. In particular if $z_j = c_j$, a fixed constant (i.e. not a chance variable), then the power at λ_j is c^2_j. Note that $|c_j|$ is the magnitude of the term $c_j e^{i\lambda jt}$ in the representation. (i.e. $|c_j| = |c_j|e^{i/jt|}$). When z_i is a chance variable, the power of x(t) at frequency λ_j is the variance of z_j, $E[z_j^2]$, and the power of x(t) is the total variance at all frequencies, which is also the variance of the ensemble x(t), t = 1,2.....N.

The power spectrum of the time series describes the distribution of the total power of x(t) at different frequencies. Define the power of x(t) at frequency λ_j as a function of λ_j. The function so defined is the power spectrum of the series x(t). At frequency λ_j, the value of this function measures the contribution of the harmonic $e^{i\lambda_j t}$ to the total power of x(t). If the contribution at the frequency λ_j is large, comparing to the neighboring frequency, then there is a cycle imbedded in the time series x(t) at frequency λ_j. It is understood that the power spectrum, or spectral analysis in general, is useful in other respects such as model building, prediction, filtering and control simulation and optimization, etc. We shall restrict our study to explore the short term cycles of the series. For detail discussion, the interested readers are referred to Jenkins and Watts [37] and Koopmans [42].

3.4.2. Power spectrum estimate.

The power spectrum estimate is computed using FT-FREQ subroutine of the International Mathematical and Statistical libraries (IMSL). Due to the limited accessibility of the IMSL to the authors, only preliminary exploratory study of the analysis is undertaken. The study is restricted to explore the short term cycles in the data. It is noted that the length of the time series is short for detecting cycles: 366 for series in Phase I and 328 in Phase II.

In Phase I, the series at grid (21, 47) show strongly a 10 hour-cycle, at grid (53,35) a 25 hour cycle, at grid (64,90) and (65,62) a 12.5 hour cycle, and at (74,91) a 15 hour cycle.

In Phase II, the series at (98,50) shows a 13 hour cycle, and a 5 hour cycle, at (82,34) a slight 50 hour cycle, at (66,82) a 100 hour cycle, at (66,50) a slight 33 hour cycle, at (50,98) a 33 hour cycle, at (50,2), 14, 7 and 5 hour cycles, at (34,18) a 50 hour cycle and at (18,34) a 50 hour cycle.

From the observation in the last two paragraphs, it does not seem to have dominant cycle prevail to all time series. A cycle of 12-13 hours is observed in two of Phase I series and one of Phase II series. A cycle of 50 hours (or more likely 2 days) is observed in three series in Phase II.

3.5. Distribution of Total Power

It was observed in the last section that the power varies wildly among series. This is true for both total power and power at all frequencies. To some extent, the power measure the "amount" of rainfall activities at a specific frequency. The total power is in fact the variance of the hourly rainfall averages (the ensemble). Since most of the hourly rainfall averages are zero, large variance would indicate a large of rainfall from time to time or maybe frequent thunderstorm activities.

Figures 3.5.1 and 3.5.2 show the plot of the total powers of times series at selected grids in Phase I and Phase II, respectively. Tables 3.5.1 and 3.5.2 list the values.

It is observed that the values of the total power of the series in Phase I vary from .002 to 1.589. This variation is dramatical considering that there are

366 observations involved. The value of 1.589 occured at grid (65.62). It is found that at this grid there was exceptionally large amount (23.71 cm/hour) of rainfall at one time. At grids (37,97) and (53,35) the total powers are .187 and .131, respectively. These values are also considerably large comparing to the rest.

The total power of series in Phase II ranges from .002 to .050. The largest value is 25 times of the smallest values. The ratio is moderate if one notes that the corresponding ratio in Phase I is 9,349. Thus the rainfall activities were somewhat similar among all localities during Phase II of the experiment while they were dramatically different during Phase I.

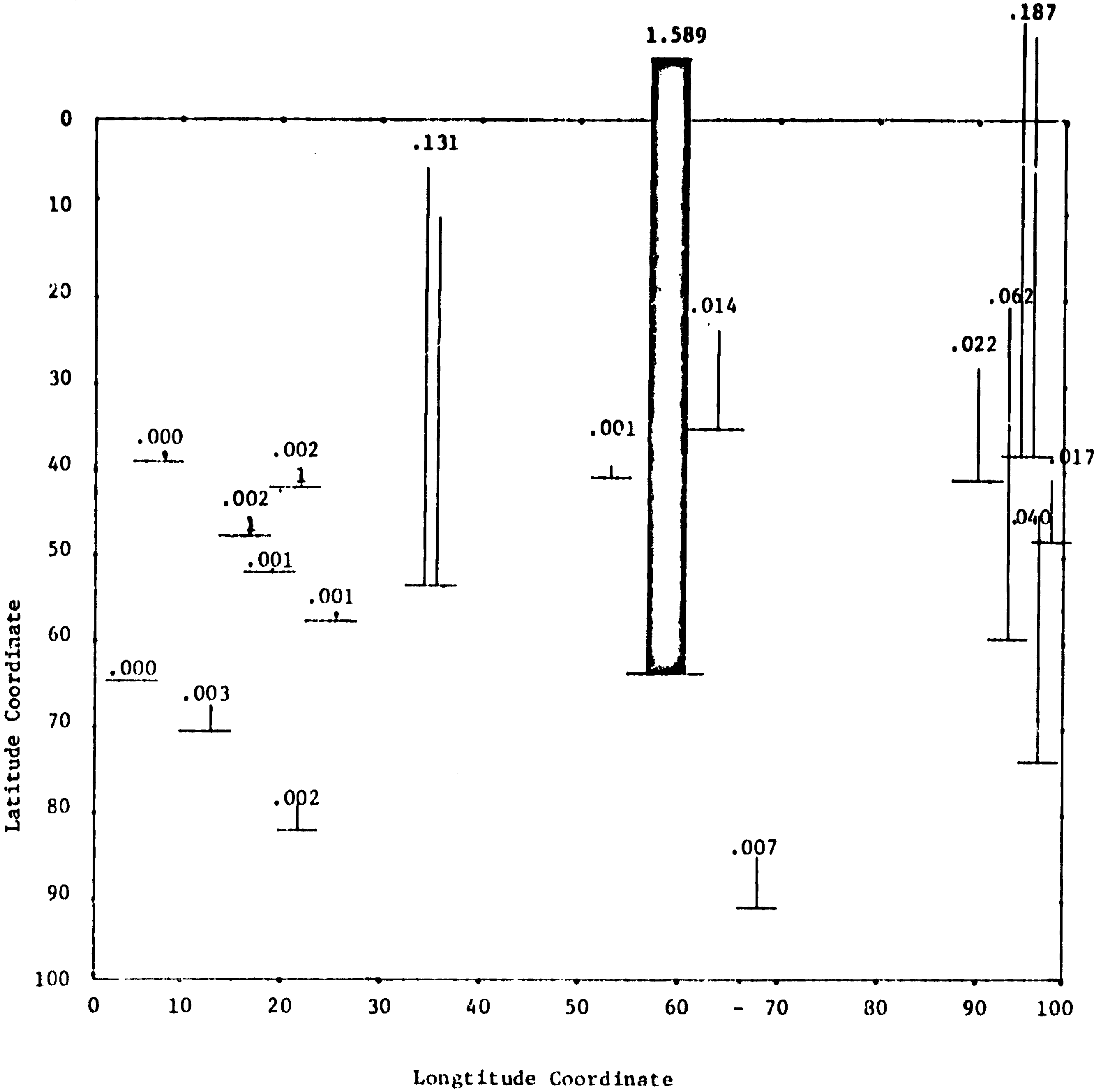


Figure 3.5.1. Total Power of Hourly Rainfall Averages in Phase I at Selected Grids.

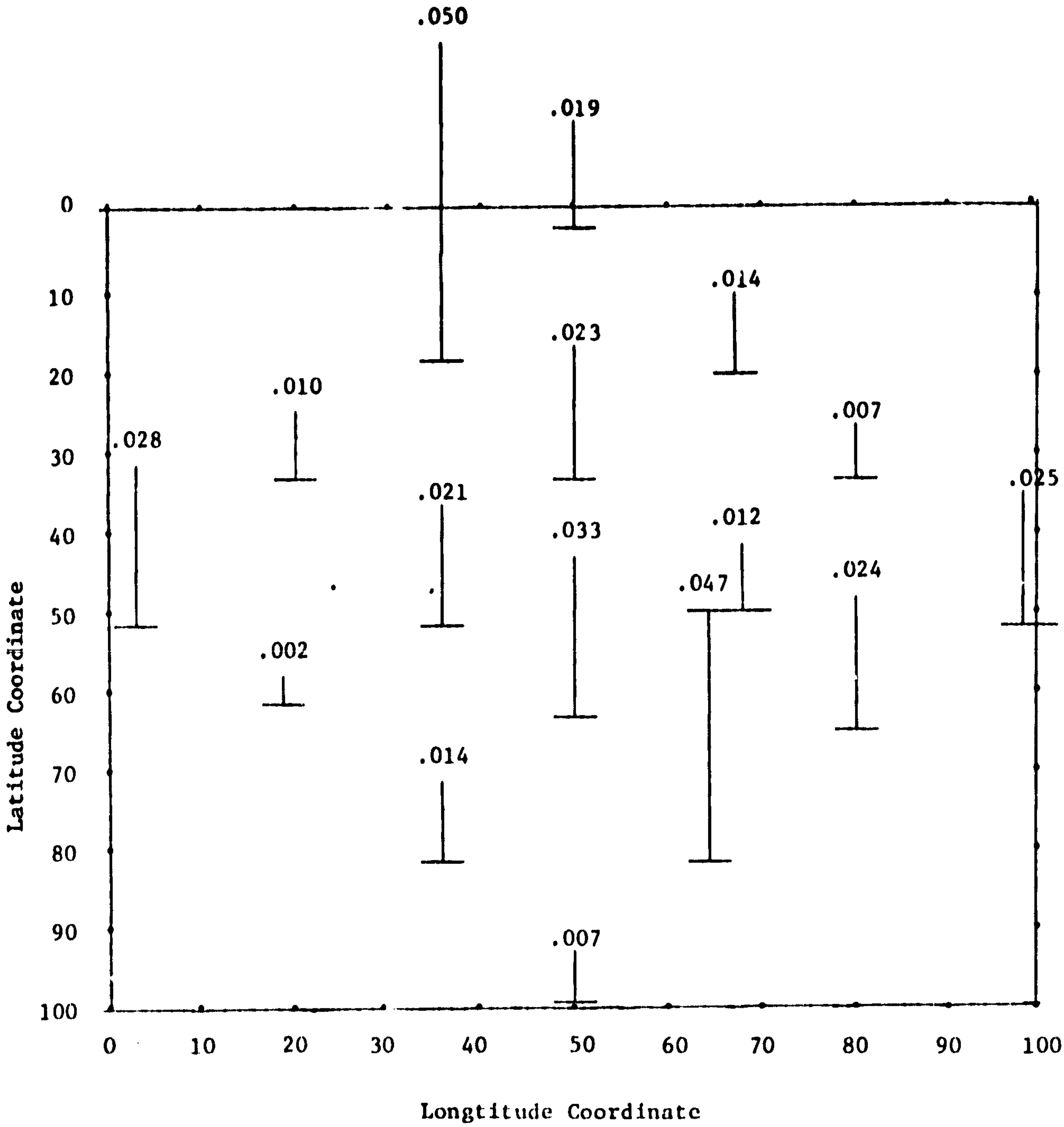


Figure 3.5.2. Total Power of Hourly Rainfall Averages in Phase II at Selected Grids.

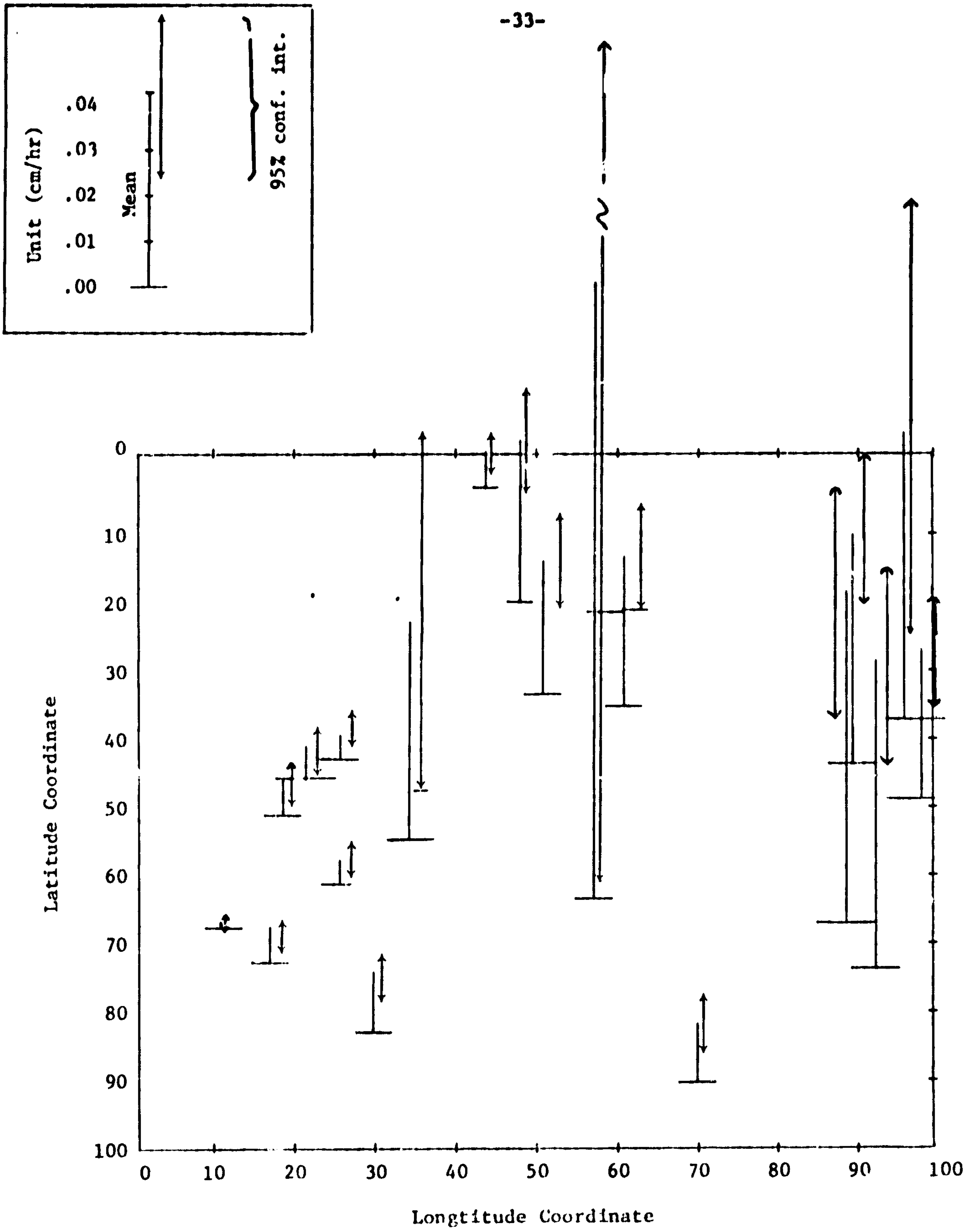


Figure 3.6.1. The Mean of Hourly Rainfall and its 95% Confidence Interval in Phase I at Selected Grids.

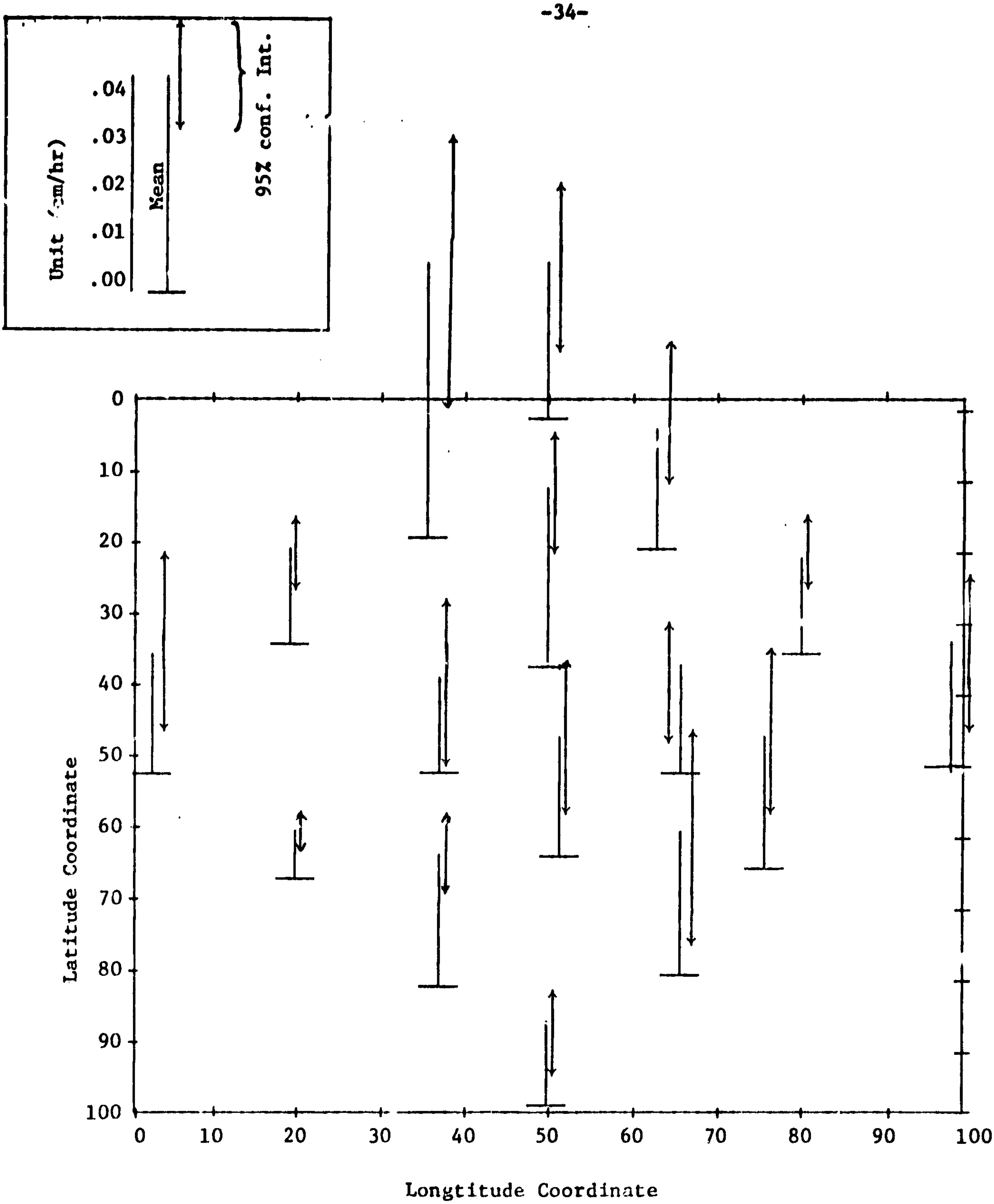


Figure 3.6.2. The Mean of Hourly Rainfall and its 95% Confidence Interval in Phase II at Selected Grids.

3.6. Rainfall Average

The mean of the hourly rainfall averages and its 95% confidence interval were listed in Tables 3.5.1 and 3.5.2. Although the mean of averages should be closer to normal distribution than the mean of raw rainfall, the distribution of the mean of averages is far from being normal. In addition, the hourly rainfall averages are not independent either, as was observed in section 3 of this chapter. Therefore Tables 3.5.1 and 3.5.2 should be viewed with cautions.

Figures 3.6.1 and 3.6.2 display the mean, along with its 95% confidence interval, of the hourly averages, according to the location of observations. It is interesting to note that the rainfall distribution in Phase II is very even among all locations. The magnitude of the mean and the length of its confidence interval are comparable. They vary very little from location to location. But in Phase I, the story is completely different. The value of the mean ranges from .0014 cm/hr. to .1320 cm/hr; the latter is 94 times of the former. The length of the confidence interval varies dramatically from location to location also.

Table 3.5.1 Mean and Variance of Hourly Rainfall Averages in Pnase I

(cm/hr)

Grid	Mean	Variance	95% Confidence Interval for Mean
(5, 43)	.0071	.0017	(.0029, .0113)
(21, 47)	.0389	.0439	(.0175, .06028)
(33, 64)	.0268	.0143	(.0146, .0390)
(35, 55)	.C260	.0085	(.0166, .0354)
(37, 94)	.0585	.1869	(.0144, .1026)
(38, 7)	.0014	.0004	(.0012, .0016)
(40, 89)	.0464	.0221	(.0312, .0616)
(42, 23)	.0066	.0021	(.0019, .0113)
(46, 18)	.0066	.0020	(.0020, .0112)
(49, 97)	.0337	.0171	(.0204, .0470)
(50, 18)	.0069	.0013	(.0032, .0106)
(53, 35)	.0486	.1314	(.0114, .0855)
(57, 23)	.0070	.0011	(.0035, .0104)
(64, 90)	.0663	.0622	(.0409, .0917)
(65, 62)	.1320	1,5892	(.0035, .2605)
(67, 9)	.0015	.0002	(.0012, .0027)
(70, 13)	.0045	.0027	(-.0009, .0098)
(74, 91)	.0669	.0404	(.0464, .0874)
(86, 28)	.0143	.0029	(.0088, .0198)
(92, 70)	.0181	.0068	(.0097, .0265)

Table 3.5.2 Mean and Variance of Hourly Rainfall Averages in Phase II

Grid	Mean (cm/hr)	Variance	95% Confidence Interval for Mean
(2, 58)	.0267	.0135	(.0120, .0414)
(18, 34)	.0497	.0503	(.0255, .0739)
(18, 66)	.0229	.0141	(.0101, .0357)
(34, 18)	.0231	.0096	(0125, .0337)
(34, 50)	.0305	.0234	(.0140, .0470)
(34, 82)	.0193	.0075	(.0100, .0286)
(50, 2)	.0242	.0277	(.0063, .0421)
(50, 34)	.0157	.0206	(.0002, .0312)
(50, 66)	.0199	.0119	(.0082, .0316)
(50, 98)	.0269	.0254	(.0097, .0441)
(66, 18)	.0099	.0022	(.0049, .0149)
(66, 50)	.0276	.0326	(.0082, .0470)
(66, 82)	.0315	.0240	(.0148, .0482)
(82, 34)	.0253	.0142	(.0115, .0481)
(82, 66)	.0337	.0457	(.0107, .0567)
(98, 50)	.0178	.0068	(.0089, .0267)

Chapter 4: Results on Diurnal Analysis

4.1 Introduction.

In this part of our report, the results of the Kolmogorov-Smirnov, Wilcoxon Signed Rank and the Chi-square Goodness of fit tests are reported as they were applied to the noon and midnight average rainfall rates at each grid (4 square Km) of the 160,000 square Km array of GATE.

Special programs were written to perform the analysis utilizing standard subprograms from S.S.P. One program prints the results of both tests in a tabular form while two other programs were designed to present the output of the results pictorially as a 100 x 100 cartesian array of alpha-numeric symbols, each representing the outcome of the statistical procedures indicated above. The graphical approach makes it easy to detect clustering and/or periodic behavior in any region of the array.

Data condi oned on the event of rain for noon or midnight were combined for the first two phases of GATE to produce a temporal resolution of 35 days at each point of the 100 x 100 array. Noon and midnight rate vectors of length 35 were generated.

Fifteen minute instantaneous radar precipitation data for phases one and two were used in the study. To minimize the effects of missing data, there was 107 records used from phase one and 188 used from phase two. The noon (12th hour) and midnight (o th hour) rainfall rates were obtained by taking averages of all data both one hour before and one hour after noon and midnight. Thus the neighborhood about both noon and midnight was a one hour radius.

4.2 Correlation

In Chapter 3, a detailed analysis of the time-series approach to the question of temporal correlations was presented. Recall that a major (implicit) assumption of the Kolmogorov-Smirnov and the Wilcoxon Signed Rank tests is that the noon and midnight data are independent. A fundamental conclusion from the temporal analysis, Fig. 3.3.1 and 3.3.2, (pages 26 and 27) is that with a 98% confidence level, rainfall 5 or more hours apart are uncorrelated.

It is well known that uncorrelated data need not be independent, however from a practical point of view, this is all that can be expected. In a more technical vain, there can be no difference between the two concepts unless we are a priori given the joint distribution function, which is a part of the unknown information in this study.

4.3 Kolmogorov-Smirnov Test

A major objective was to determine if there is any mathematical difference in the empirical distribution of rainfall at noon verses midnight. This is equivalent to our test of the null hypothesis that noon and midnight distributions are the same verses the alternate hyopthesis that they are different.

In this study we surveyed 10,000 grid points; 1,872 were excluded because they had too few rainfall events for analysis; 1,742 had noon and midnight data distributed in such a way that we could not reach any conclusion. In 5,425 grid areas, we found that the null hypothesis could be accepted and in 960 grid areas we found that the null hypothesis must be rejected and the alternate hypothesis accepted.

Figure 4.2.1 displays the results pictorially in a 100 x 100 array. The letter D indicates that there is a significant difference between noon and midnight data. (reject null hypothesis) A blank indicates that there is no difference. It is easy to see that zero rainfall rates in both phases dominate the western border

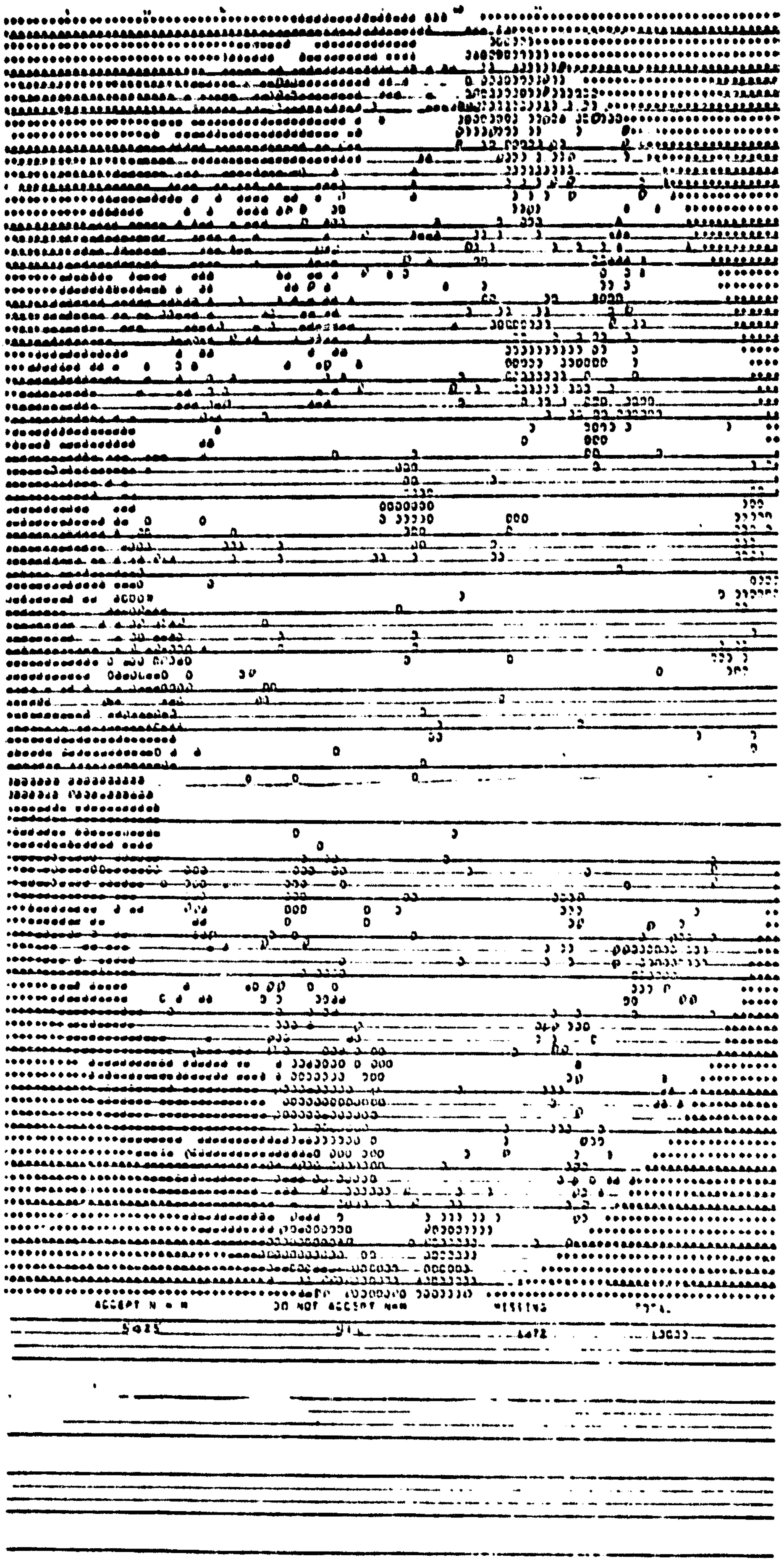


Fig. 4.3.1

of the array. Heavy clustering of D's can be detected in the north eastern and south western region of the array. We expect that this is due to the occurance of heavy rain in these regions throughout the thirty five day period of phase 1 and 2. This also implies that there may be a periodic spatial distribution of areas where there is a definite diurnal rainfall variation surrounded by regions where there is no variation.

4.4 The Wilcoxon Signed Rank Test

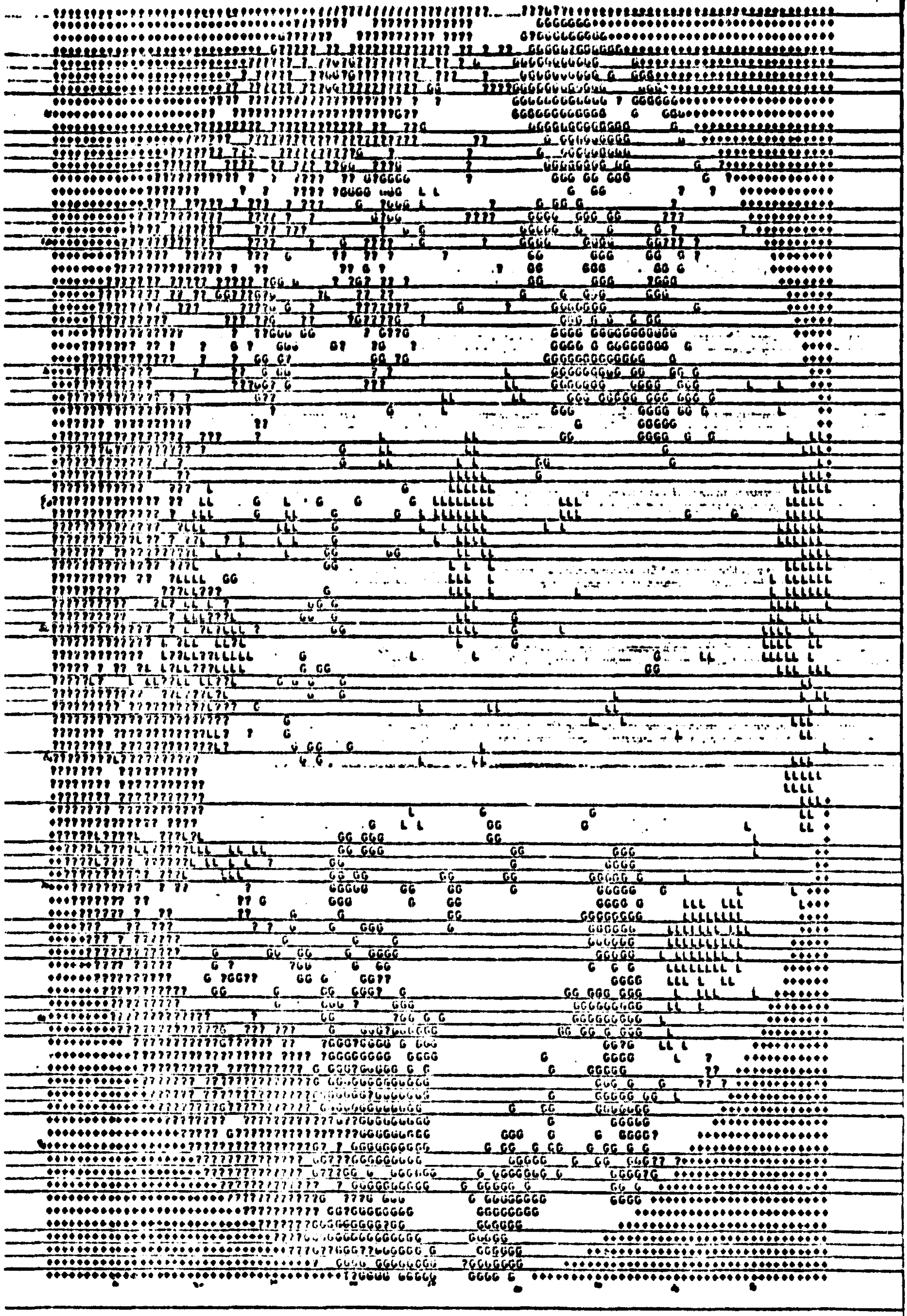
The Wilcoxon test, allows us to test the null hypothesis that the mean distribution of rainfall rate for noon and midnight are the same verses the alternative hypothesis that one is greater than the other.

In this study, 1,872 grid areas were excluded because they had too few events for analysis; for 1,743 we could reach no conclusion. In 4,890 grid areas we found that the null hypothesis could be accepted while in 1,495 of the grid areas we found that the null hypothesis must be rejected and the alternate hypothesis accepted.

This result is expected since the Wilcoxon test is less conservative than the Kolmogorov-Smirnov test. It is possible for the distributions of rainfall at noon and midnight to be the same and yet the means are different.

Figure 4.3.1 displays the results pictorially in a 100 x 100 array. The letter L indicates that noon rainfall is significantly less than midnight rainfall, the letter G indicates that noon rainfall is significantly greater than midnight and a blank indicates no difference.

The results of Wilcoxon signed-rank test match those of Kolmogorov-Smirnov test, as far as whether there is difference in the rainfall distribution at noon and in the midnight is concerned. From Figure 4.4.1, it is observed that the rainfall activities at noon and in the midnight during the experiment period are rare in the western and north-western parts of the area, as indicated by "?". The noon rainfall is greater than the midnight rainfall in the southern and northeastern



NOON < MIDNIGHT(L)	MIDNIGHT < NOON(G)	NOON = MIDNIGHT( )	NO CONCLUSION (?)	MISSING (*)	TOTAL
430	1065	4890	1743	1872	10000

Fig. 4.4.1

parts, as indicated by "G". The places where midnight rainfall is greater are scattering.

4.5 Chi-Square Test

The chi-square test is used to provide a further check on the results of the Kolmogorov-Smirnov and Wilcoxon signed rank tests. This test allows us to determine whether or not the percentages used for analysis using the other two statistics were actually preserved in the results of the analysis. In this case the χ^2-test is defined by: (see Chapter 2)

$$\chi_k^2 = \sum_{j=1}^{k} \frac{(n_j - nP_j)^2}{nP_j}$$

where $n = \sum_{n=1}^{k} n_j$; n_j is the number of grids for which the j-th conclusion was made; P_j is the probability of making the n-th conclusion.

We use an $\alpha = .05$ for the level of significance for each experiment.

The χ^2 test applied to the K-S experiment.

In the K-S test we have the following information:

$k = 2 \qquad j = 1,2.$

$n_1 = 5425 \qquad P_1 = .90 \qquad P_2 = .10$

$n_3 = 960$

$n = n_1 + n_2 = 6385$

$c_{.05,2} = 3.84$

So

$$\chi_2^2 = \frac{(5425 - 6385\ (.90))^2}{6385\ (.90)} + \frac{(960 - 6385\ (.10))^2}{6385\ (.10)}$$

$$= \frac{(321.5)^2}{5746.5} + \frac{(321.5)^2}{638.5} = 17.99 + 161.88 = 179.87 > 3.84$$

hence, the percentages of 80 and 20 do not hold at the 5% level of significance and thus the null hypothesis H_0: $F^1 = F^2$ is not true and must be rejected.

The χ^2 Test applied to the Wilcoxon Experiment.

In the Wilcoxon experiment we have the following information:

$k = 3 \qquad j = 1,2,3.$

$n_1 = 430 \qquad n_2 = 4890 \qquad n_3 = 1065 \qquad n = \sum_{1}^{3} n_i = 6385$

$P_1 = .10 \qquad P_2 = .80 \qquad P_3 = .10$

$c_{.05,3} = 5.99$

$$\chi_3^2 = \frac{[430 - (6385)(.10)]^2}{6385\ (.10)} + \frac{[4890 - (6385)(.80)]^2}{6385\ (.80)} + \frac{[1065 - 6485\ (.10)]^2}{6385\ (.10)}$$

$$= \frac{[-208.5]^2}{639} + \frac{[-218]^2}{5108} + \frac{[426.5]^2}{639} = 68.03 + 9.30 + 284.67 = 362.00 > 5.99$$

Thus, the percentages of 10, 80; and 10% do not hold at the 5% level of significance and hence the null hypothesis H_0: $F^1 = F^2$ is not true and must be rejected.

4.6. Summary on Diurnal Analyses.

Two non-parametric tests were used to detect if there is a diurnal variation in oceanic rainfall using GATE data. Non-parametric methods are chosen because of their model free characteristic.

The test was undertaken at grid points where data are available. Of the 10,000 grid points in the study area, 1872 were excluded because data are not available; these points are around the boundary of the square. In addition, there are 1743 grid points where no conclusion was obtained due to insufficient frequency of rain in the midnight and at noon during the experiment period. The analysis was done at the rest of 6385 grid points.

The Kolmogorov-Smirnov test was used to test whether there is a difference in the rainfall distributions between noon and midnight. At one out of 10 chance of error, it was found that out of 6385 grid points, the rainfall distributions at noon and at midnight are different in 960 grid points and are same at 5,425 grid points, which are 15.0 and 85.0%, respectively. If there were no difference in the rainfall distributions, these percentages would have been 10 and 90% respectively. A chi-square test at one out of 20 chance of error shows that this split of percentage was not followed. There are more grid points where the assertion of difference is made than expected. Thus, overall the rainfall distributions at noon and at midnight are different. This concludion could have been made at practically any significant level.

The Wilcoxon signed-rank test was used to test if the noon rainfall is less than, equal to, or greater than the midnight rainfall. It was found that, at one out of 10 chance of error, the numbers of grid points fall in these three categories are, respectively, 430, 4,890, and 1,065. In terms of percentage, they are 6.73, 76.5, and 16.68%, as compared to 10, 80, and 10%, respectively, which are expected if the assumption of no difference had been true. At one out of 20 chance of error, the chi-square test concluded that the midnight rainfall is not equal to the noon rainfall; the noon rainfall is convincingly greater than the midnight rainfall. This conclusion could have been made at practically any significantly level.

From the analysis it was also found that the Wilcoxon signed-rank test is more sensitive than the Kolmogorov-Smirnov test in detecting the difference in the midnight and noon rainfalls. This is to be expected by the nature of these two tests.

BIBLIOGRAPHY

1. Allison, Lewis J., William R. Bandeen, Charles R. Laughlin, and Raymond Wexler, "Remote Sensing of the Atmosphere fron Environmental Satellites," X-901-77-132 Preprint, Goodard Space Flight Center, Greenbelt, Maryland, (June, 1977).

2. Arnold, A., E.B. Dobson, J. Goldhirsh, I. Katz, T.G. Konrad, and W.L. Vann, "Radar Derived Spatial Statistics of Summer Rain Experiment Description," NASA Contractor Report, NASA CR-Z59Z (September, 1975).

3. Bark, L. Dean, and A.M. Feyerhern, "Probability Models for Simulating Temperatures and Precipitation", Kansas State University, Manhattan, Kansas.

4. Bark, L. Dean, N.T. Ison, and A.M. Feyerherm, "Wet Period Precipitation and the Gamma Distribution," Journal of Applied Meteorology, X (1971), 658-665.

5. Bowen, E.G., and Paul T. Schickedanz, "The Computation of Climatological Power Spectra", The Embassy of Australia and Illinois State Water Survey, (March, 1977).

6. Bowman, K.O., and L.R. Shenton, "Comments on the Gamma Distribution and Uses in Rainfall Data", Computer Center, University of Georgia and Oak Ridge National Laboratory, Oak Ridge, Tennessee.

7. Bowman, K.O., and L.R. Shenton, "Further Remarks on Maximum Likelihood Estimation for the Gamma Distribution", Technometrics, XIV (1972), 725-733.

8. Bripan, Maurice, "Describing and Testing for Heavy Tailed Distributions", Colorado State University, Fort Collins, Colorado.

9. Carter, Michael C., Lee W. Falls, and William O. Williford, "Probability Distributions for Thunderstorm Activity at Cape Kennedy, Florida", NASA Technical Memorandum, NASA TM X-53867 (May, 1970).

10. Chernoff, Herman, "The Efficient Estimation of a Parameter Measurable By Two Instruments of Unknown Precisions", Optimizing Methods in Statistics, (1971), 1-27.

11. Clements, Arthur E., and Benjamin M. Herman, "Removal of Seeing and Instrumental Blur Effects from Astronomical Scanner Observations", Applied Optics, XIII (December, 1974), 2947-2956.

12. Cohen, A. Clifford, F. Russell Helm, and Merritt Sugg, "Tables of Areas of the Standardized Pearson Type III Density Function", NASA Contractor Report, NASA CR-61266 (March, 1969).

13. ______________, The Space Applications Board Assembly of Engineering, National Research Council, Practical Applications of Space Systems: Supporting Paper 1, Weather and Climate, Washington, D.C.

14. ______________, A Note on a Gamma Distribution Computer Program and Computer Produced Graphs", NOAA Technical Report, EDS 24 (May, 1977).

15. Crosby, D.S., "Estimation of the Mean and Covanance Matrix of Cloud Effect for Remote Soundings of Atmospheric Temperature Profiles", National Environmental Satellite Service, NOAA and The American University, Washington, D.C.

16. Durand, David, and J. Arthur Greenwood, "Aids for Fitting the Gamma Distribution by Maximum Likelihoods", Technometrics, II (1960), 55-65.

17. Falcone, V.J., and K. Mano, "Statistical Methods of Indirect Probing of the Atmosphere" Air Force, Cambridge Research Laboratories, L.G. Hanscom Field, Bedford, Mass.

18. Falls, Lee W., "The Beta Distribution: A Statistical Model for World Cloud Cover, "Journal of Geophysical Research, 79 (March, 1974), 1261-1264.

19. Falls, Lee W., "A Computer Program for Standard Statistical Distributions", NASA Technical Memorandum, NASA TM X-64588 (April, 1971).

20. Gihman, I.I., "On the empirical distribution function in the case of grouping of data", Doklady Akad, Nauk SSSR (NS), Vol. 82 (1952, pp. 837-840.

21. Gihman, I.I., "Some results on empirical distribution functions", Mat. Sborn. Univ. Kiev., Vol. 8 (1954).

22. Gihman, I.I., "On the Empirical Distribution Function in the Case of Grouping of the Data", Trans. by the National Bureau of Standards.

23. Glivenko, V., "Sulla determinazione empirica delle leggi di probabilita", Giorn. del l'Instit. degli att., Vol. 4 (1933), pp. 92-99.

24. Gnedenko, B.V., and V.S. Mihalevic, "Two Theorems on the Behavior of Empirical Distribution Functions", Trans. by the National Bureau of Standards.

25. Gnedenko, B.V., and V.S. Mihalevic, "On the Distribution of the Number of Excesses of One Empirical Distribution Function Over Another", Trans. by The National Bureau of Standards.

26. Gnedenko, B.V. and V.C. Korolyuk, "On the Maximal deviation between two empirical distributions", Doklady Akad, Nauk SSSR, Vol. 80 (1951), pp. 525-528.

27. Gnedenko, B.V. and V.S. Mihalevic, "Two theorems on the behavior of empirical distribution functions", Doklady Akad. Nauk SSSR, Vol. 85 (1952), pp. 25-27.

28. Gnedonko, B.V., "Unber die Nachprufung Statisticher Hypothesen mit Hilfe der Variationsreihe", Bericht uber die Tagung Wahr, and Math. Stat. Berlin, 1954, pp. 97-107.

29. Grantbam, Donald D., and Iver A. Lund, "Persistense Runs and Recurrence of Precipitation", Journal of Applied Meterology, XVI (March, 1977), 346-358.

30. Green, Richar N., "Simulation Studies of Wide and Medium Field of View Earth Radiation Data Analysis", NASA Technical Paper 1182 (1978).

31. Gringorten, Irving I., "Modelling Conditional Probability", Journal of Applied Meterology, (December, 1970), 646-798.

32. Gumbel, E.J., and C.K. Mustafi, Tables of the Probability and Density Functions for the Bivariate Extremal Distributions, Department of Industrial Engineering, Columbia University, New York, 1968.

33. Hoff, John D., Michael D. Hudlow, Vernon L. Patterson, Peter J. Pyhowany, Frank P. Richards, "Gate Radar Rainfall Processing System", NOAA Technical Memorandum, (August 1978).

34. Hogg, Robert V., and Elliot A. Tanis, Probability and Statistical inference, Macmillan New York 1977.

35. Hudlow, Michael D., and Frank Richards, "Use and Abuse of the Gate Digital Radar Data", Center for Experiment Design and Data Analysis, National Oceanic and Atmospheric Administration, Washington, D.C.

36. Hudlow, Michael D., "Precipitation Climatology for the Three Phases of GATE", Center for Experimental Design and Data Analysis, National Oceanic and Atmospheric Administration, Washington, D.C.

37. Jacobson, R.W., and W.M. Gray, 1976: Diurnal Variation of Oceanic Deep Comulus Convection, Colorado State University, Atmospheric Science Paper No. 243, Fort Collins, Colorado.

38. Jenkins, Gwilym M., and Donald G. Watts, Spectral Analysis and its applications Holden-Day Inc. C 1968.

39. Kaplan, L.D., 1959: Inference of Atmospheric Structure from Remote Radiation Measurements, Journal the Optican Society of America, P. 49, 10, 1004-1007.

40. Kidder, S.Q., 1976: Tropical Oceanic Precipitation Frequency from Nimbus-5 Microwave Data, Colorado State University, Atmospheric Science Paper No. 248, Fort Collins, Colorado.

41. Kolmogorov, A.N., "Confidence limits for an unknown distribution function", Ann. Math. Stat., Vol. 12 (1971), pp. 461-463.

42. Kolmogorov, A.N., "Sulla determinazione empirica di une legge di distribuzione", Giorn. dell 'Instit. degli att., Vol. 4 (1933), pp. 83-91.

43. Koopmans, Lambert Herman, The Spectral analysis of Time Series, Academic Press, New York 1974.

44. Kuelbs, J., "A Counterexample for Banach Space Valued Random variables", The Annals of Probability, IV (1976), 684-689.

45. Mielke, Paul W., "Another Family of Distributions for Describing and Analyzing Precipitation Data", Journal of Applied Meteorology, XII (1973), 275-280.

46. Peers, H.W., and B.L. Welch, "On Formulae of Confidence Points Based on Integrals of Weighted Likelihoods", University of Leeds.

47. Samuel, Ester, "An Empirical Bayes Approach to the Testing of Certain Parametric Hypotheses", The Hebrew University of Jerusaleum, (June, 1963).

48. Sasyo, Yoshio, "On the Probabilistic Analysis of Precipitation Particles", M teorological Research Institute of Japan.

49. Simpson, Joanne, "Use of the Gamma Distribution in Single Cloud Rainfall Analysis", Monthly Weather Review, 100 (1972), 309-312.

50. Smirnov, N.V., "Probabilities of Large Values of Non parametric one-Sided Goodness of Fit Statistics", Trans. by Wassily Hoeffding.

51. Smirnov, N.V., "Sur la distribution de w^2", C.R. Acad. Sci. Paris, Vol. 202 (1936), pp. 449-452.

52. Smirnov, N.V., "On the distribution of the w^2 criterion of von Mises", Rec. Math. (NS), Vol. 2 (1937), pp. 973-993.

53. Smirnov, N.V., "Sur les ecarts de la courbe de distribution empirique" (Russian, French summary), Rec. Math, (NS) (Mat. Sborn.), Vol. 6 (98) (1939), pp. 3-26.

54. Smirnov, N.V., "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", Bull, Math. Univ. Moscow, Vol. 2, No. 2 (1939), pp. 3-14.

55. Smirnov, N.V., "Approximate laws of distribution of random variables from empirical data", Uspehi Matem. Nauk, Vol. 10 (1944), pp. 179-206.

56. Smirnov, N.V., "Table for estimating the goodness of fit of empirical distributions", Ann. Math. Stat., Vol. 19 (1948), pp. 279-281.

57. Smirnov, N.V., "On the Cramer-von Mises criterion", Uspehi Matem. Nauk (NS), Vol. 4, No. 4 (32) (1949), pp. 196-197.

58. Smirnov, N.V., "On the construction of confidence regions for the density of distributions of random variables", Doklady Akad, Nauk SSSR (NS), Vol. 47 (1950), pp. 189-191.

59. Staelin, D.A., 1966: "Measurements and Interpretation of the Microwave Spectrum of the Terrestrial Atmosphere Near 1 Centimeter Wave length", J. Geophys. Res., 71, 2875-2881.

60. Staelin, D.A., 1969: Passive remote sensing at microwave wave lengths, Proc. IEEE, 57, 427-439.

61. Thom, H.C.S., "Approximate Convolution of the Gamma and Mixed Gamma distributions", Monthly Weather Review, 96 (1968), 883-886.

62. Van Vleck, J.H., 1947: Absorption of Microwaves by water vapor, Phys. Rev., 71, 425-433.

63. Welch, B.L., "The Significance of the Difference Between Two Means When The Population variances are Unequal", University of Leeds.

64. Wilheit, T.T., M.S.V. Rao, T.C. Chang, E.B. Rogers and J.S. Theon, 1975: The Remote Sensing of Rainfall Rates over the Oceans, Presented at the Second Conference on Atmospheric Radiation of the American Meterological Society, Arlington, VA., October 29, 1975.

65. Wilheit, T.T., 1972: The electrically Scanning Microwave Radiometer (ESMR) Experiment in the Nimbus-5 User's Guide, NASA/Goddard Space Flight Center, Greenbelt, MD., pp. 59-105.