

NAS5-29271

SATELLITE RAINFALL RETRIEVAL BY LOGISTIC REGRESSION

by
Long S. Chiu

1N-47
64418
F 101

prepared for
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland
under NASA-SBIR contract NAS5-29271
Final Report

(NASA-CR-180485) SATELLITE RAINFALL
RETRIEVAL BY LOGISTIC REGRESSION Final
Report (Applied Research Corp.) 101 p
CSCL 04B

N88-21607

Unclas
G3/47 0064418

Applied Research Corporation
8201 Corporate Drive, Suite 920
Landover, MD 20785

September 1986

Satellite Rainfall Retrieval by Logistic Regression

TABLE OF CONTENTS

	PAGE
1. Objective	3
2. Review of Satellite Estimation Techniques	4
3. The Logistics Model	6
4. GATE Rainfall Statistics.	7
4.1 The Data.	7
4.2 The Mixed Distribution Model.	8
4.3 Intermittency	8
4.4 Spatial and Temporal Rain Structure11
4.5 Cloud Height.14
5. Rain Field Models20
5.1 Simulation Model.20
5.2 Stochastic Regression23
6. Scale Dependence of Rainfield Parameters.23
7. Summary and Estimate of Technical Feasibility24
Appendix: Criterion for Independence of Conditional Probabilities26
Attachment A: "Satellite Rainfall Retrieval by Logistic Regression".A-1
Attachment B: "Estimating Time Mean Areal Average Rain: A Mixed Distribution Approach".B-1
Attachment C: "A Rain Field Simulation Model".C-1
Attachment D: "On Lognormality of Rainrates"D-1

PROJECT SUMMARY

The objective of this project is to investigate the potential use of logistic regression in rainfall estimation from satellite measurements. Satellite measurements provide covariate informations in terms of radiances from different remote sensors. The logistic regression technique can effectively accommodate many covariates and test their significance in the estimation. The outcome from the logistic model is the probability that the rainrate of a satellite pixel is above certain threshold. By varying the thresholds, a rainrate histogram can be obtained and from which the mean and variance estimated.

A logistic model is developed and applied to rainfall data collected during GATE, using as covariates the fractional rain area and a radiance measurement which is deduced from a microwave temperature-rainrate relation. It is demonstrated that the fractional rain area is an important covariate in the model, consistent with the use of the so-called 'Area Time Integral' in estimating total rain volume in other studies.

In order to calibrate the logistic model, simulated rain fields generated by rainfield models with prescribed parameters are needed. A stringent test of the logistic model is its ability in recovering the prescribed parameters of simulated rain fields. A rain field simulation model which preserves the fractional rain area and lognormality of rainrates as found in GATE is developed. The simulated rain fields are quite realistic. A stochastic regression model of branching and immigration whose solutions are lognormally distributed in some asymptotic limits has also been developed. This model makes no assumption about the law of proportionate effect which is often quoted to achieve lognormality.

This study has demonstrated the effectiveness of the logistic technique in examining a large number of covariates and in testing their significance. By identifying important covariates and the way in which they enter the estimation procedure, this technique will be useful in the design of a system of remote sensors for the measurement of rainfall from space and in the development of satellite rainfall retrieval algorithms.

I. Objective

The Earth distinguishes itself from other planets in the presence of water substances. The heat stored in various forms of water substances, the heat transported by atmospheric water vapor and by the oceans, the heat released during the transformations between the different phases have shaped Earth's climate to a large extent. Water vapor is the working substance of Earth's atmosphere: created to remove excess heat from the oceans and over land in the form of evaporation; participates in the radiative heating of the atmosphere by emission in the long wave regime of the atmospheric spectrum; transports excess heat in the tropics and deposits it in the high latitudes thus modulating the extreme heat and cold on Earth. In the final stage of this branch of the water cycle, it changes phase and is deposited in the form of precipitation over the Earth's surface.

Because of the scale of variability, precipitation is probably one of the least known but yet most sensitive parameter in the water budget over land and oceans (Miller 1977, Laevastu, et al., 1969). A knowledge of the amount and distribution of precipitation is crucial to our understanding of the large scale dynamics of the oceans and atmosphere. Strong empirical as well as theoretical evidence have suggested that condensational heating of the tropical atmosphere, as indicated by the amount of precipitation, is instrumental to circulation anomalies world wide (Horel and Wallace 1981, Gill 1982).

Precipitation and the antecedent latent heat release has been incorporated into General Circulation Models (GCM's) of the Earth's atmosphere for some time, but the intensity and distribution is still poorly modeled. A detailed global rainfall data set is therefore needed to calibrate the GCM's for mean and anomalous conditions. To accomplish this, a satellite rainfall monitoring mission to measure precipitation over the tropics, the Tropical Rainfall Measuring Mission (TRMM), has been proposed (Theon, et al., 1986). The objective is to obtain at least 3 years of monthly mean rainfall data over the tropical regions.

To achieve this, a retrieval algorithm by which satellite measurements can be converted to rainfall data is needed. The ultimate objective of our work is to develop such an algorithm. The immediate objective is to investigate the potential use of logistic regression in rainfall estimation from space. Since rainfall is not directly measured, the information available are covariate information in terms of radiances from satellite sensors. The logistic regression technique is especially suited for this purpose since it can effectively accommodate a large number of covariates and readily test their significance. A secondary objective is to study the statistics of rain fields which will be useful in interpreting problems such as the "beam filling" and estimate biases are due to sampling.

In section 2, the techniques of estimating rainfall from space are briefly reviewed. The need for multispectral estimation techniques is stressed. Section 3 discusses the logistic model and demonstrates its use in identifying important covariates. The scenario of concomittant observations of microwave and fractional rain area data, which may be obtained from visible or infrared measurements, is investigated. A major finding is the importance of rain area in estimating total rainfall. Since observation of the fractional area is dependent on the foot print size of the observation, the statistics of rainfall fields are examined in section 4 using the GATE data as an example. To calibrate the logistic technique, a simple rain field simulation model and a point process regression model which exhibit statistical properties of the GATE rainfall data are developed in section 4. The regression model is capable of producing rainfall rate with a lognormal distributions in some asymptotic limits. These limiting conditions are satisfied in the GATE data for large area averaged conditions. The dependence of statistical parameters in rain fields on scale is addressed in section 5. Section 6 summarizes our findings and makes recommendations for future work.

2. Review of Satellite Estimation Techniques

The need for satellite monitoring of global rainfall has been stressed by Atlas and Thiele (1982) and Austin and Geotis (1980). Barrett and Martin (1981) have reviewed the various estimation techniques. Another good source of reference is contained in the preprint volume of the second conference of satellite meteorology in which two sessions are devoted to the estimation of rainfall from space.

The source of satellite data is basically derived from three regions in the atmospheric spectrum: the visible (VIS), infrared (IR), and microwave windows. The techniques which use information in the visible part of the spectrum rely on identifying cloud types and assigning rainrates to them. This cloud type-rain rate relation is dependent on the local climatology, and hence, this method must be calibrated regionally.

The infrared techniques rely on information on cloud top temperatures which are indicators of cloud heights. The implicit assumption is that the rain-bearing clouds are tall cumulus clouds. Arkin (1979) developed an index of precipitation which is the number of pixels within an area in an IR satellite imagery with temperatures below 235 degrees Kelvin. This index represents the fractional area of high convective clouds within the area. When compared with rainfall data measured during GATE, a correlation coefficient of 0.87 is obtained. Arkin's index of precipitation has been adopted for local calibration of rainfall during the Tropical Ocean Global Atmosphere (TOGA) experiment. However, at middle to high latitudes, rainfall from large-scale low-level stratiform clouds becomes increasingly dominant, and this cloud area index becomes less effective in estimating rainfall in those regions.

A more direct approach relies on the radiative properties of rain drops in the microwave portion of the spectrum. By modeling the vertical structure of a rain cloud, a rainfall rate-microwave temperature relation can be established. Hence, a rainfall rate can be estimated from an observation of the microwave emission. There are several pitfalls in this approach.

1. Unfilled Field of View (FOV)--Microwave measurements usually have large foot print sizes, and, hence, the field of view of the foot print is usually not filled with rain. A bias is introduced if the measurements from the unfilled beam is used to retrieve rainfall through the microwave temperature-rainfall rate relation.
2. Saturation--The microwave measurements become saturated at high rainrates. At 19 GHz, the beam becomes saturated at rainrates above 15-20 mm/hr. Although only a small fraction of the measurements are contained in this portion of the the rain spectrum, the high rainrates account for a large fraction of the total rainfall.
3. Rainfall Rate-Microwave Temperature Relation--In deriving the rainfall rate-microwave temperature relation, a cloud model has to be assumed. Such a relationship is rather sensitive to the assumed parameters, such as profile of ice and liquid water content. Rather different relationships are found for different modeling assumptions. For example, the relationship presented by Wilheit, et al. (1977), showed a monotonic increase of microwave temperature as a function of rainfall rates in the range from 0 to about 15 mm/hr at 19 GHz whereas that of Wu and Weinman (1984) shows a decrease.

Estimation schemes which combine information from the different atmospheric channels, seems to yield good estimates. Lovejoy and Austin (1979) developed an algorithm which delineates rain areas from visible and infrared measurements. Radar detected rain patterns are used as ground truth and a statistical pattern recognition technique is used to establish rain area characteristics in the visible and infrared. Once the rain areas are calculated, the rainfall is obtained by multiplying the area by a climatological mean rainfall rate. This multi-spectral approach has had many successful applications and has been adopted for operational satellite rain estimation by the Atmospheric and Environmental Service of Canada.

It is argued that if information from different channels are combined, a better estimation scheme can be developed. Since the resolutions of the sensors are quite different, it is necessary to identify the important covariates as well as the way through which they enter the estimation scheme. In what follows, a logistic model is described and the scenario of concomitant microwave IR/VIS observations which delineate rain area is examined.

3. The Logistic Model

The logistic model is useful in determining the relationship between the distribution of a random variable and a set of covariates. It has been applied in various forms in reliability testing and the analysis of survival data (Cox and Oakes 1984). Detail treatment of the logistic model is given by Cox (1970). The model is briefly described below.

Let R be the random variable which stands for rainrate and let

$$\underline{z} = (z_1, \dots, z_p)$$

be the vector of covariables related to R . Suppose we are interested in estimating the probability

$$P(R \in I)$$

where I is a rainrate interval. Let X be defined by

$$X = \begin{cases} 1, & r \text{ in } I \\ 0, & \text{Otherwise} \end{cases}$$

Then

$$P(R \in I) = P(X = 1).$$

In many respects the simplest way to express the dependence of this probability on explanatory variables or covariates is to postulate the model [Cox (1970)].

$$P(X = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 z_1 + \dots + \beta_p z_p)}}$$

$$P(X = 0) = \frac{1}{1 + e^{(\beta_0 + \beta_1 z_1 + \dots + \beta_p z_p)}}$$

This is the logistic model. This model allows great flexibility in the choice of the covariates and in mathematical manipulations.

The parameters are estimated by maximizing a likelihood function and the significance of the covariates are readily tested by a likelihood ratio. The interested reader is referred to our paper (Chiu and Kedem 1986) for a more detailed discussion. This paper is attached (attachment A) with this report.

The scenario of a TRMM-like system of sensors which can provide microwave measurements and fractional rain area within a microwave foot print size pixel is examined. The data we use are the rainfall data collected during GATE. The GATE data are binned at 4 kms by 4 kms and are given at 15 minute intervals. A detail description of the data is given in the next section. The microwave temperature is mimicked through a microwave temperature-rainfall rate relation (see attachment A). The microwave measurements are assumed to have a resolution of about 32 kms on the side, somewhat similar to the resolution of the Electrically Scanning Microwave Radiometer (ESMR) which was flown on board the Nimbus V satellite. From the 4 kms by 4 kms rainfall rates, a temperature is computed. The temperatures of 64 (32/4 or 8 pixels on the side) neighboring pixels are averaged to obtain the microwave temperature (T). The fractional rain area with rainrates above 1 mm/hr (F) is obtained by counting the number of high resolution pixels (4 kms on the side) with rainrates above 1 mm/hr and dividing by the total number (64) in a large microwave pixel (32 kms on the side). Another index, F1, which is the fractional area with rainrate in excess of 20 mm/hr, is also used. This index mimicks Arkin's index of high clouds which produce heavy rainfall. To test the usefulness of the logistic technique, another parameter, TL, is also included in the estimation. TL is the microwave temperature T at a lag of 1 time units (15 minutes). The results are summarized in table 2 in Chiu and Kedem (1986)(attachment A). The results show that the inclusion of TL does not improve the model significantly. This is probably due to persistence in the time series so that there is not much new information in TL as most of it is contained in T. The results also show that T is the best regressor in the model. Since T is derived from the rain field, this result cannot be taken literally. An interesting finding is the importance of F in the model. This is a better regressor than F1, but, when the two parameters F and F1 are combined, a better model is obtained. This is consistent with our finding about the contribution of the rain area in determining the total rainfall, a point which we shall return to in section 4.

4. GATE Rainfall Statistics

The fractional area of rain within a pixel is dependent on the pixel size and the spatial variability of the rain field. Hence, the structure and statistical properties of the rain field need to be studied.

4.1 The Data

The study of the statistical properties of the rain field is based on data collected during GATE. This is one of the most comprehensive rain measurements made over the ocean.

1. GATE Surface Rainfall Data--The GATE is an observational program conducted in the summer of 1974. During three roughly tri-weekly periods, each termed a phase, detailed rainfall measurements from rain gauges and radars on an array of research vessels were made over an area called the B-scale. The center of the B-scale area is located at 8.5N, 23.5E and encompasses an area of about 200 km in diameter. Arkell and Hudlow (1977) composited the radar measurements from ships and presented an atlas of the radar echoes at 15 minute intervals. Patterson, et al. (1979), converted the radar measurements to rainrates and presented rainrate data in 4 by 4 km² bins.
2. CAPPI--For the height of the rain column, we used the Constant Altitude Plan-Position Indicator (CAPPI) radar data taken onboard the research vessel the "Oceanographer," which was positioned at the center of the B-scale area in GATE, but was moved to the Southeast quadrant. The original data was taken from the plane position indicator (PPI) for elevation angle of about 1.5 to 22 degrees. Pylowany, et al. (1979), converted the data from elevation-distance coordinate to constant altitude plane position co-ordinate, with a vertical resolution of about 1 km. The maximum echo height reported is 12 kms, i.e., at higher heights are truncated at 12 kms. This data covers 3 convectively active days in each phase of GATE.

4.2 The Mixed Distribution Model

An objective of TRMM is to obtain monthly averages of rainfall. If rainfall rates can be described by a class of statistical distribution, the estimation procedure can be simplified since only a few parameters of the distribution need to be estimated. We examined the GATE data and found that the rainrates can be described by a mixed distribution (attachment B). The mixed distribution consists of a finite probability of no rain and a continuous distribution for the rainy part. Conditional on rain, it was shown that the lognormal distribution provides an excellent fit to the data. A detailed description of the model and its application to sampling studies in GATE can be found in attachment B of this report.

4.3 Intermittency

Intermittency refers to sporadic changes in a field of turbulence. It expresses the fact that turbulence does not fill the whole space in a turbulent flow. This is an important aspect of turbulent flows, which despite much work, is far from being completely understood (Schertzer and Lovejoy 1985). A measure of intermittency is the fraction of time in which an event occurs over a period of distance (Tennekes and Lumley 1974). For extreme events, we expect this measure of intermittency to increase as turbulence sets in through flow instability, reaches some peak value and then decreases as the energy of the turbulent flow is cascaded to smaller scales through dissipative losses.

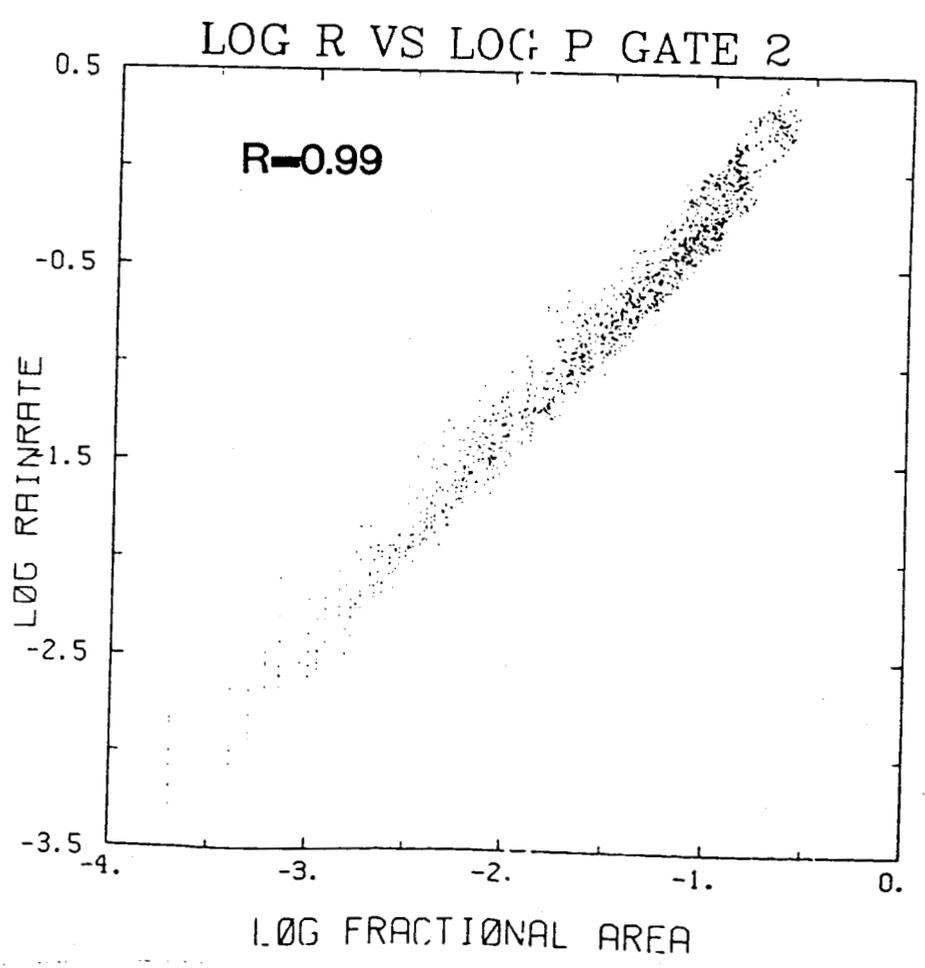
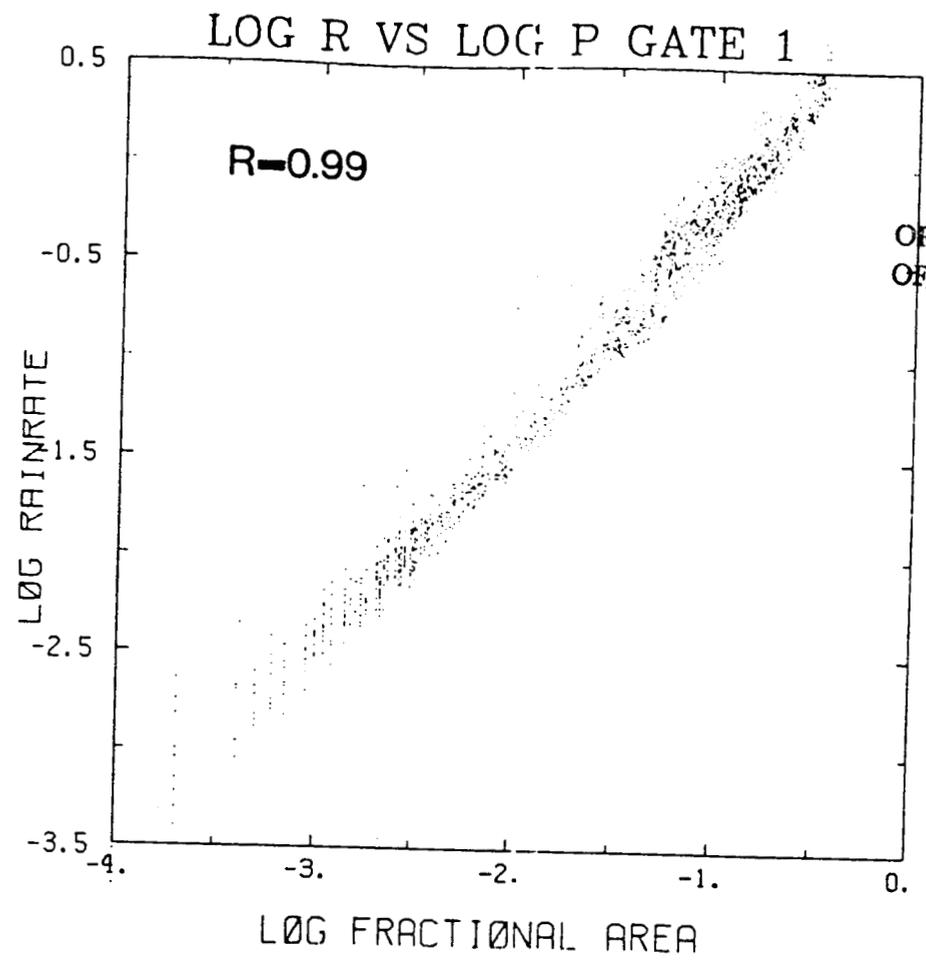
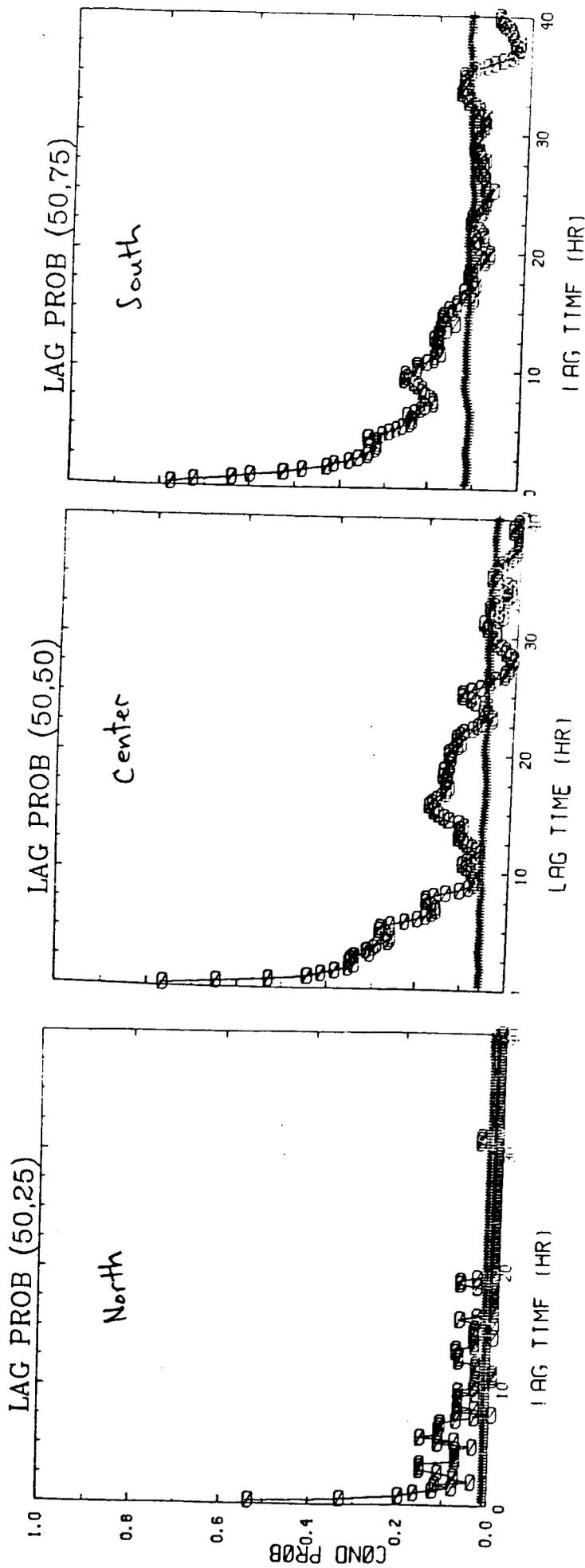


Fig. 1



ORIGINAL PAGE IS
OF POOR QUALITY

Figure 2. Conditional probability of observing $R > 1\text{mm/hr}$ conditional on observing $R > 1\text{mm/hr}$ at a lag time earlier for a pixel at the northern (left), central (middle) and southern (right) part of GATE 1.

In a rough sense, we can consider the rain fields as fields of turbulence. Precipitation can be considered an "extreme" event, an index of moist instability. The fraction of time/space that this event occurs is a measure of intermittency.

An important parameter in the estimation of total rainfall from a GATE scene is the fractional rain area (Chiu, et al., 1986). Figure 1 shows scatter diagrams of the average rainfall rate for a GATE scan and the fractional rain area with a rainfall rate of 1 mm/hr and above on logarithm scales. The correlations between the two variables are extremely high for both phases of GATE. It is interesting to note that this correlation of 0.99 is higher than the correlation of 0.87 between rainrate and Arkin's cloud index. Since the rainfall total (R) for a GATE scene is the product of the fractional area (p) multiplied by the average rainrate for the rainy pixels (α), or $R = p\alpha$, we can take the logarithm of both sides and compute the variance of $\log R$ as a sum of the variance of $\log p$ and $\log \alpha$. The contributions from the various terms are given below for GATE 1 and 2.

$$\text{var}(\log R) = \text{var}(\log p) + \text{var}(\log \alpha) + 2 \text{cov}(\log p \log \alpha)$$

(100%)	(77%)	(3%)	(20%)	GATE1
(100%)	(77%)	(3%)	(20%)	GATE2

We pointed out that this index of fractional rain area is equivalent to the so called "Area Time Integral (ATI)" used in radar meteorology to estimate rain volume. The ATI is the time integral of the area of radar echoes. It is shown that the total rain volume of a system can be obtained by multiplying the ATI by some climatological mean rainfall rate (Doneaud, et al. 1981). Jackson (1986) examined the contribution of the number of rain days in a month and the average intensity of rainfall during raindays in tropical stations to the monthly rainfall. It is found that the number of raindays is the dominant factor in determining the monthly total. These are consistent with our results on the analysis of GATE data and the logistic model.

4.4 Spatial and Temporal Rain Structure

Because of the phenomena of intermittency in rain fields, it is difficult to define the usual characteristic functions of a turbulent field such as correlation or autocorrelation functions. For example, the autocorrelation functions will have a long tail at long separations due to the abundance of no rain observations.

We have examined the structure of the rain field in terms of conditional probabilities.

Figure 2 shows the probability of observing a rainrate of 1 mm/hr at a fixed location (4 km by 4 km pixel) in GATE at different time lags conditional on observing such an event at time zero. It can be seen that the conditional probability drops off rather rapidly but reaches another secondary maximum in about 10-12 hours. The condition for independence is derived in the appendix and is plotted on the same graph. This assumed no sampling error or persistence in the data.

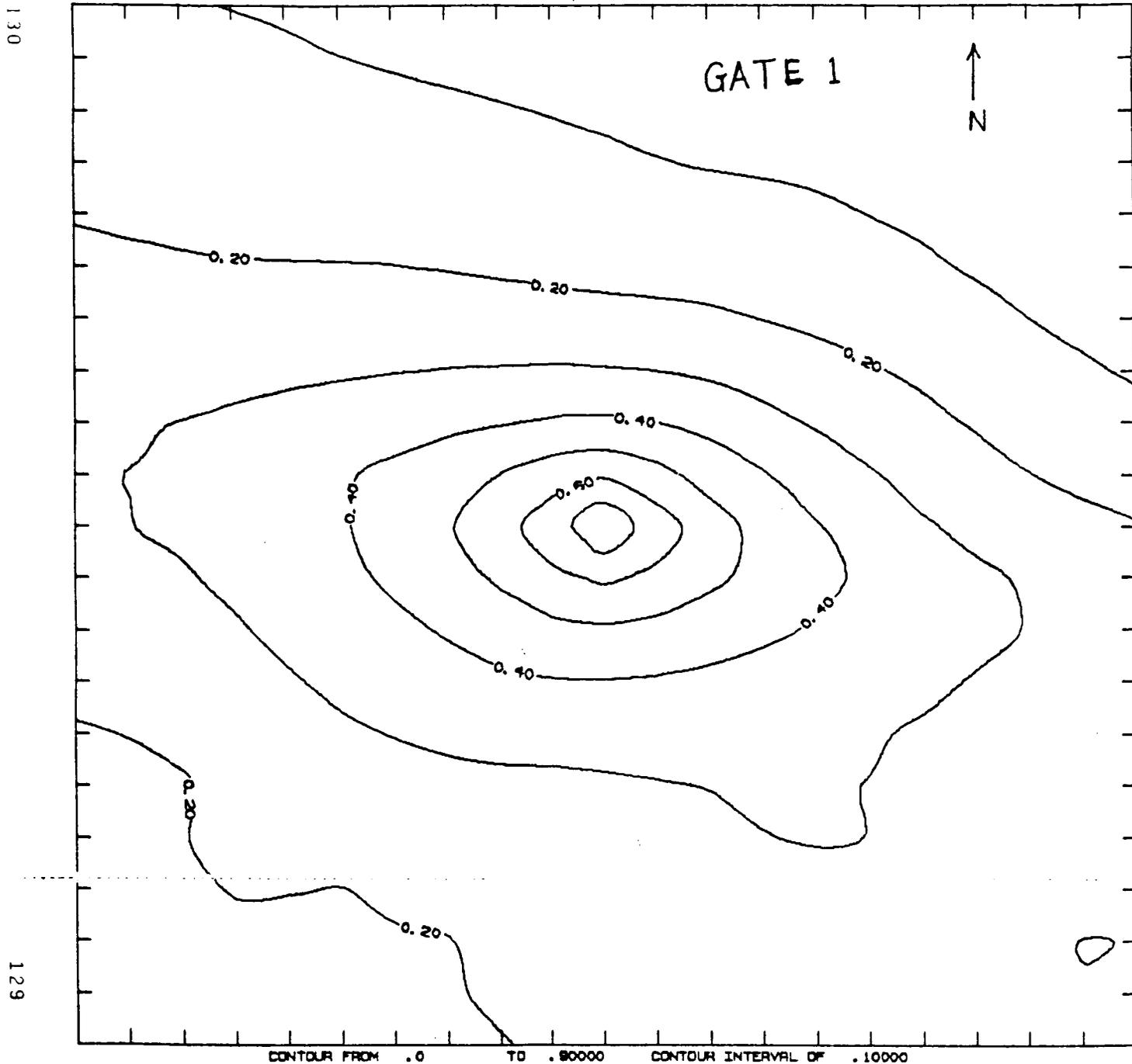


Figure 3A. Lines of constant probability of observing $R > 1\text{mm/hr}$ conditional on observing $R > 1\text{mm/hr}$ at a distance for GATE 1. The distance between two marks on the boundaries are 12 kms apart.

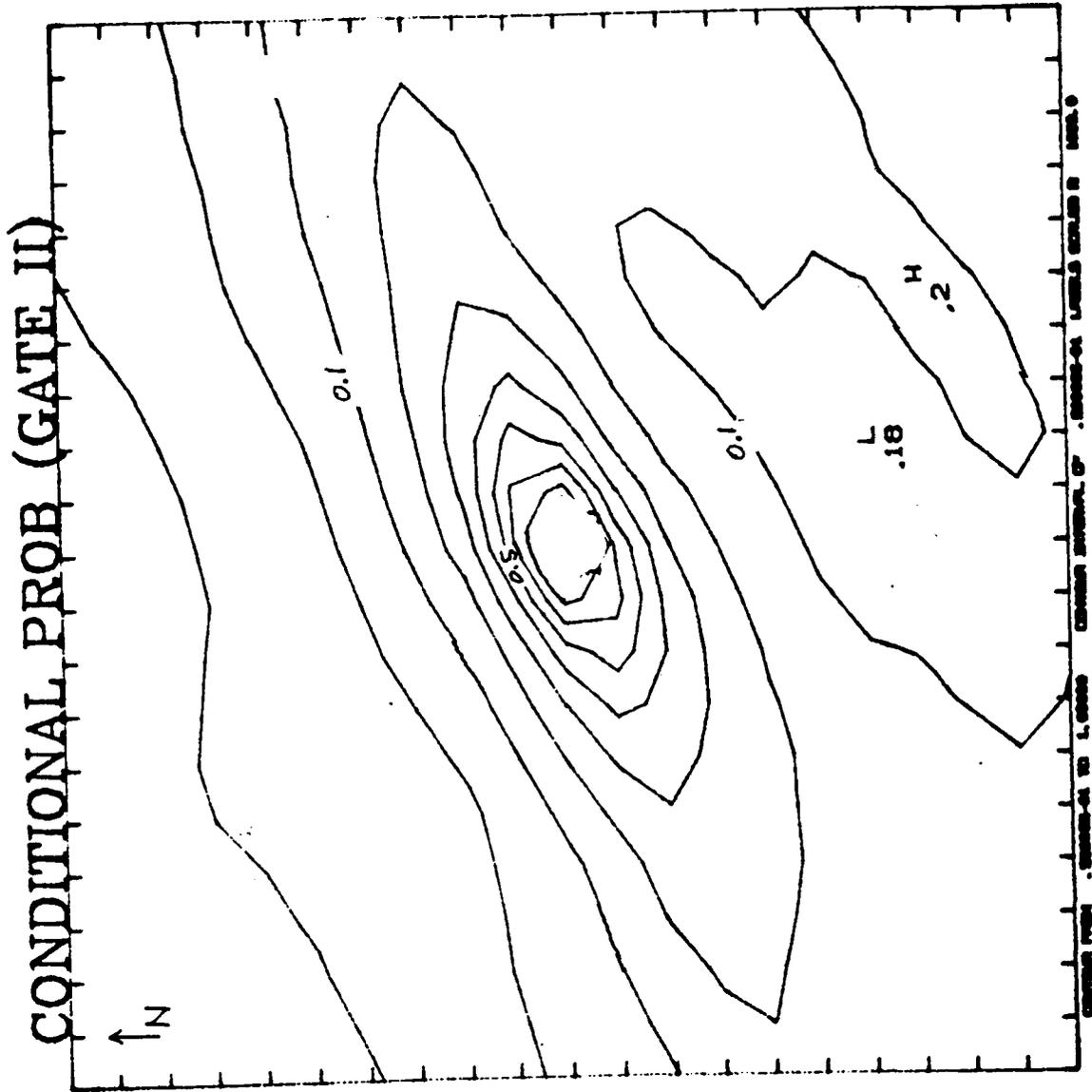


Fig 3B Same as A3 except for GATE 2.

ORIGINAL PAGE IS
OF POOR QUALITY

A similar calculation is performed on the conditional probability on space. Figure 3 shows lines of constant probability as a function of distance conditional on the event of having 1 mm/hr at a 4 km by 4 km pixel for GATE 1 and 2. The anisotropy in space is clearly discernible for GATE 2. The more less east-west orientations of lines of constant probability is consistent with the meteorological conditions in GATE of the passage of elongated rain bands oriented in the east-west direction.

4.5 Cloud Height

In the retrieval of rainfall rate from microwave temperature measurements, a number of parameters enters into the retrieval. Since these parameters are quite variable, errors are introduced into the estimation scheme if some constant value is used. An important parameter is the height of the rain column. The bias due to the rain cloud height can be estimated as follows. The attenuation of microwave radiation (or change in the optical thickness τ) in the presence of rain can be written as

$$\Delta\tau = ahR^b \quad (1)$$

where a and b are functions of frequency, drop size distribution and temperature of the drops. Olsen, et al. (1978), have examined the dependence of a and b over a broad range of frequencies and for different drop size distribution at various temperatures. At a frequency of about 20 GHz and 0 degrees Celsius,

$$.05 < a < 0.09$$

and

$$0.9 < b < 1.1$$

with R the rainrate, in mm/hr, h , the effective height of the rain column, in km. Oftentimes h is defined in terms of the attenuation as

$$\Delta\tau/aR^b$$

We would like to get some idea of the distribution of the height of rain columns and an estimate can then be made of the bias in using a climatological height in the estimation from microwave sensors.

From equation (1) above (for simplicity, assume $b = 1$), an estimate of the rainrate using a climatological cloud height, $\langle h \rangle$, where $\langle \rangle$ denote ensemble averaged quantities, is

$$R(\langle h \rangle) = (\Delta\tau/a)1/\langle h \rangle$$

The bias in percent can now be written as

$$B = (\langle R \rangle - R(\langle h \rangle)) / R(\langle h \rangle) - 1$$

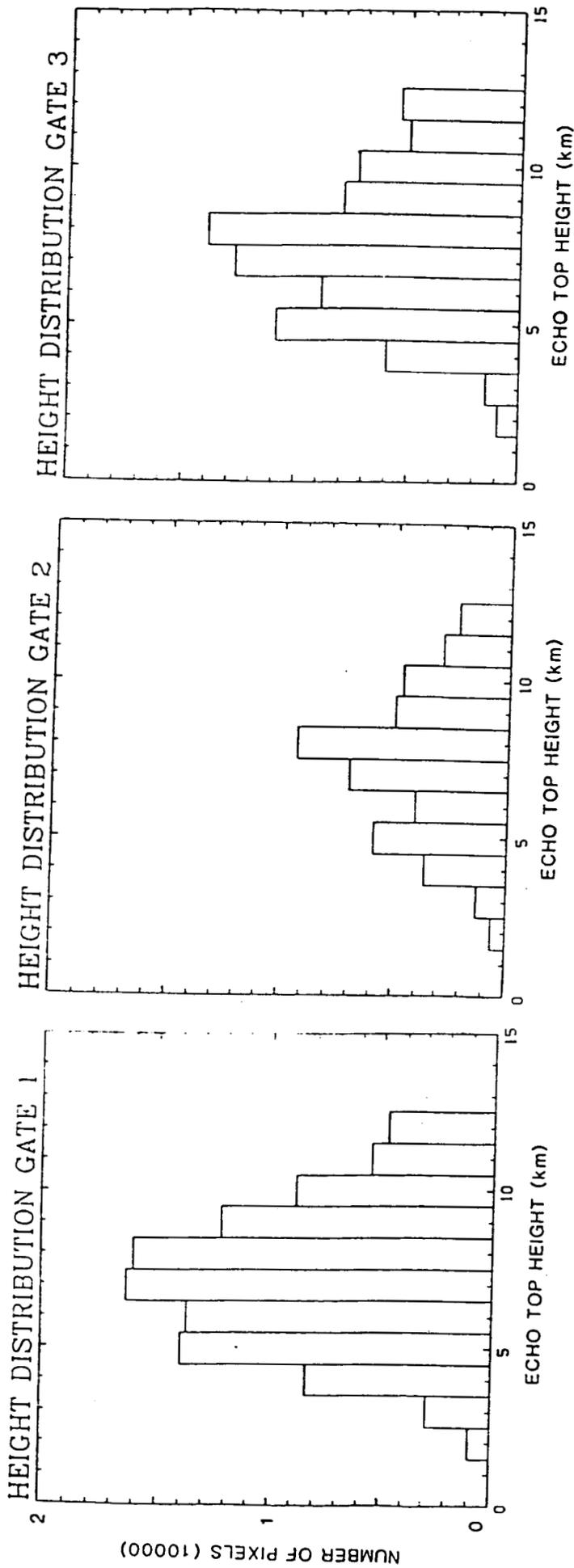


Figure 4. Histogram of cloud heights for 3 consecutively active days in each phase of GATE.

HEIGHT-RAINRATE RELATION

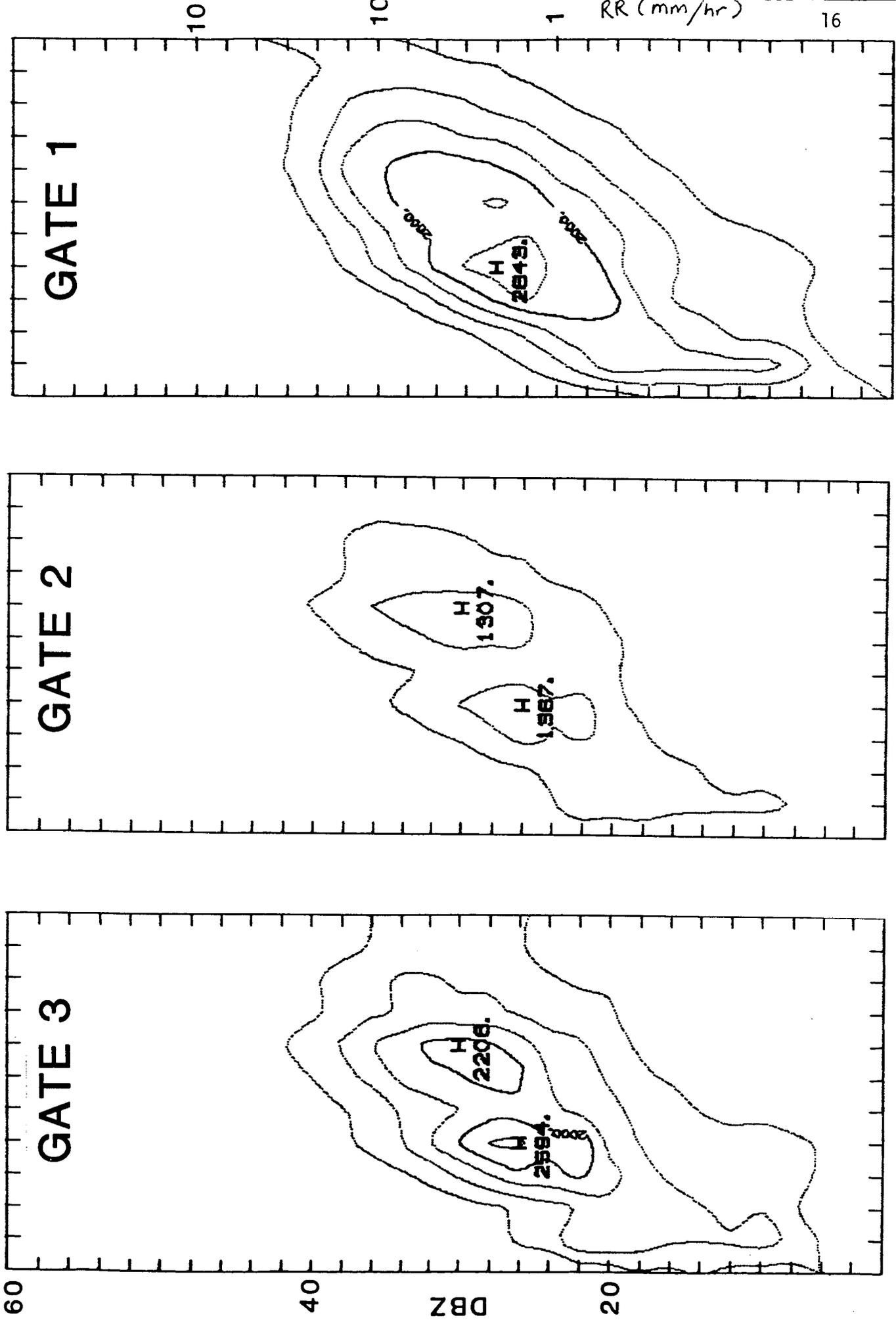


Fig 5 2 Dimensional histogram 1 of cloud height, and low level radar reflectivity for GATE 10 HEIGHT (km)

Table of 2-D histogram of Heights

FILE: GATE1H DATA A NASA GSFC/VPEVM 470/V6 - VM/SP 3.1 CMS

DBz	2D DISTRIBUTION GATE 1											
	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5
0.0	429	88	39	69	99	56	43	26	8	7	2	1
2.0	828	268	140	256	258	157	100	54	45	10	6	3
4.0	745	572	269	338	328	184	133	124	56	48	7	11
6.0	521	1154	607	515	628	355	232	141	73	43	27	8
8.0	500	1663	795	765	758	504	332	192	129	68	38	13
10.0	484	1661	839	731	716	586	327	222	129	80	41	12
12.0	583	1664	1219	679	855	666	448	334	171	98	48	28
14.0	476	1656	1382	921	1065	897	634	448	242	128	74	53
16.0	508	1741	1698	1297	1224	1222	863	587	355	207	127	75
18.0	221	1871	1893	2056	1548	1548	1204	735	562	331	160	85
20.0	144	1386	1795	2341	2437	1776	1582	1110	721	404	243	154
22.0	70	883	1609	2414	2312	2409	1839	1328	922	542	335	174
24.0	14	560	1134	2501	2820	2482	2419	1786	1188	756	384	211
26.0	5	248	791	2080	2843	2319	2547	2070	1441	816	549	285
28.0	3	57	490	1584	2565	2220	2442	2255	1506	1060	579	376
30.0	1	30	240	1057	2119	2103	2370	2355	1684	1184	649	388
32.0	0	12	122	614	1543	1742	2191	2199	1731	1244	739	466
34.0	0	8	49	310	1075	1361	1789	1929	1642	1221	679	572
36.0	0	1	17	134	579	766	1345	1664	1355	1032	692	533
38.0	0	0	5	52	263	450	712	961	822	691	486	546
40.0	0	0	0	12	96	196	347	534	484	433	387	537
42.0	0	0	0	7	50	67	158	207	257	263	241	498
44.0	0	0	0	1	17	29	40	70	115	120	102	324
46.0	0	0	0	0	3	2	7	16	24	34	23	95
48.0	0	0	0	0	0	1	0	2	0	3	0	9
50.0	0	0	0	0	0	0	0	0	0	0	0	1
52.0	0	0	0	0	0	0	0	0	0	0	0	0
54.0	0	0	0	0	0	0	0	0	0	0	0	0
56.0	0	0	0	0	0	0	0	0	0	0	0	0
58.0	0	0	0	0	0	0	0	0	0	0	0	0
250.0	0	0	0	0	0	0	0	0	0	0	0	0

ORIGINAL PAGE IS
OF POOR QUALITY

20 DISTRIBUTION	GATE 3											
0.0	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.
0.0	313	70	77	61	161	20	9	41	15	8	13	7
2.0	498	368	139	169	382	58	31	122	28	12	22	14
4.0	503	440	245	388	439	76	104	164	34	38	20	6
6.0	309	914	574	729	631	264	238	240	94	70	75	30
8.0	304	1181	982	945	676	401	372	290	133	130	77	45
10.0	350	1075	814	853	610	423	401	329	137	128	87	51
12.0	567	1015	1059	669	686	487	436	294	190	137	94	71
14.0	381	1070	1295	913	698	733	445	321	305	183	146	90
16.0	453	1310	1289	1224	987	842	593	443	359	346	209	129
18.0	200	1454	1359	1696	1349	1152	929	659	533	458	296	208
20.0	96	1468	1042	2091	2478	1351	1269	1102	610	461	270	330
22.0	33	731	817	1869	2143	1658	1472	1210	763	528	386	379
24.0	12	578	573	1394	2594	1730	2198	1741	942	743	525	551
26.0	3	215	415	1485	2536	1719	2132	2074	1110	917	552	540
28.0	0	89	221	1083	1978	1474	1976	2206	1082	947	683	586
30.0	1	21	142	662	1453	1208	1877	2093	1192	1050	729	579
32.0	0	7	57	423	987	1071	1619	1781	1048	1078	658	569
34.0	0	4	18	239	668	714	1119	1468	911	837	530	505
36.0	0	0	8	113	380	466	858	1052	678	653	385	408
38.0	0	0	7	52	173	273	510	728	405	435	315	418
40.0	0	0	0	14	78	125	278	454	246	256	245	420
42.0	0	1	1	2	23	55	109	224	175	185	192	331
44.0	0	0	0	1	8	14	37	111	87	123	120	292
46.0	0	0	0	0	2	3	11	35	30	37	58	148
48.0	0	0	0	0	0	1	1	9	12	9	11	50
50.0	0	0	0	0	0	0	0	1	0	1	0	11
52.0	0	0	0	0	0	0	0	0	0	0	0	1
54.0	0	0	0	0	0	0	0	0	0	0	0	1
56.0	0	0	0	0	0	0	0	0	0	0	0	0
58.0	0	0	0	0	0	0	0	0	0	0	0	0
250.0	0	0	0	0	0	0	0	0	0	0	0	0

ORIGINAL PAGE IS
OF POOR QUALITY

where

$$\langle R \rangle = \Delta\tau/a \int 1/h p(h)dh$$

the factor $\Delta\tau/a$ cancels out, and

$$\langle R \rangle / R(\langle h \rangle) = \langle h \rangle \langle 1/h \rangle$$

To calculate these quantities, the distribution of $h P(h)$ is needed. The data we used to calculate $P(h)$ are the so called "CAPPI" (constant Altitude Plane Position Indicator) data of GATE (pytlowany, et al., 1978). Figure 4 shows the histograms of height obtained from 3 days of data in each phase of GATE. We have taken individual pixels in calculating the statistics as opposed to earlier works which counts a rain cloud as an entity (e.g., Houze and Cheng, 1979). Our emphasis here is the estimation and correction of the bias associated with satellite retrieval. Because of the noise in the radar reflectivity, we have set a low threshold of 24 dbz corresponding to a rainrate of 1 mm/hr. The histograms show bimodal distributions in GATE 2 and 3, with peaks at 5 and 8 kms respectively whereas this feature is absent in GATE 1. The double peaks are also present if the statistics is calculated over cloud clusters (Houze and Cheng 1979).

We noted that the bias is extremely sensitive to the population at the low cloud heights. If the threshold value is changed to the lowest detectable level, the whole histogram rises over all ranges in height. The increase in population at the low height will increase the bias substantially (from about 25 to 50 percent).

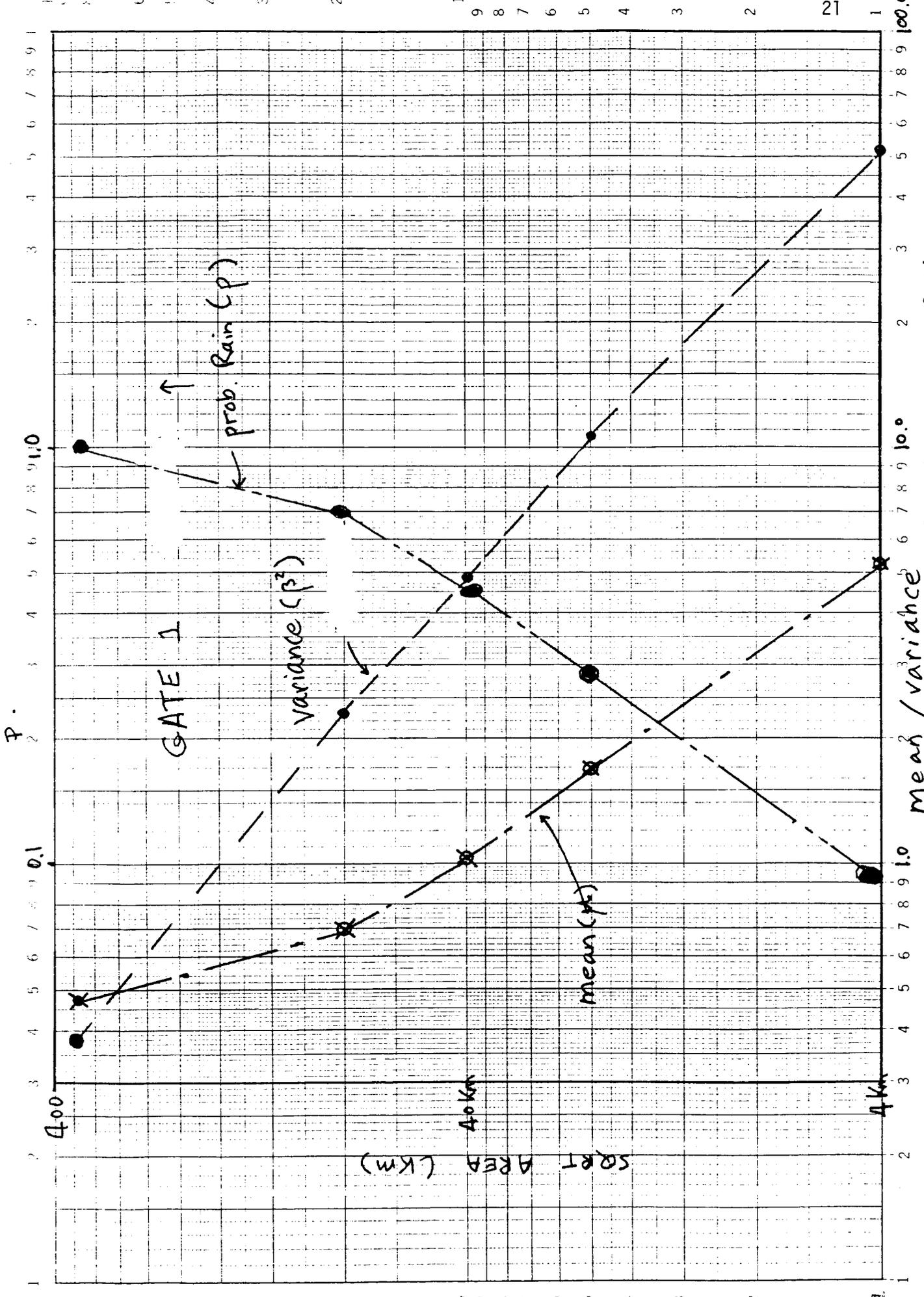
Another point is that R and h are related: one expects a higher rainrate associated with higher cloud top. Adler and Mack (1984) have examined the usefulness of this relation and other environmental information to estimate rainfall. Figure 5 shows a two dimensional distribution of distribution of h and radar reflectivity for the same GATE CAPPI data. The shape of the loci of the maxima of the distributions agree well with the rainrate--cloud height relation observed in tropical storms (Adler and Mack 1984, their Figure 1).

5. Rain Field Models

5.1 Simulation Model

To extend the data base beyond the scope of GATE for the purposes of sampling studies and the calibration of the logistic model, a simulation model of rain field is developed which preserves the characteristics of GATE rainfall: namely, fractional rain area and lognormality of the rainy part of the distribution. A description of the model is given in attachment C. This model is capable of producing realistic rain fields.

Laughlin (1982) examined the errors associated with satellite sampling and computed the temporal autocorrelation function for different area averages for GATE. From the autocorrelation



Dependence of the probability of rain (p) mean and variance on resolution

Fig 6

ORIGINAL PAGE IS
OF POOR QUALITY

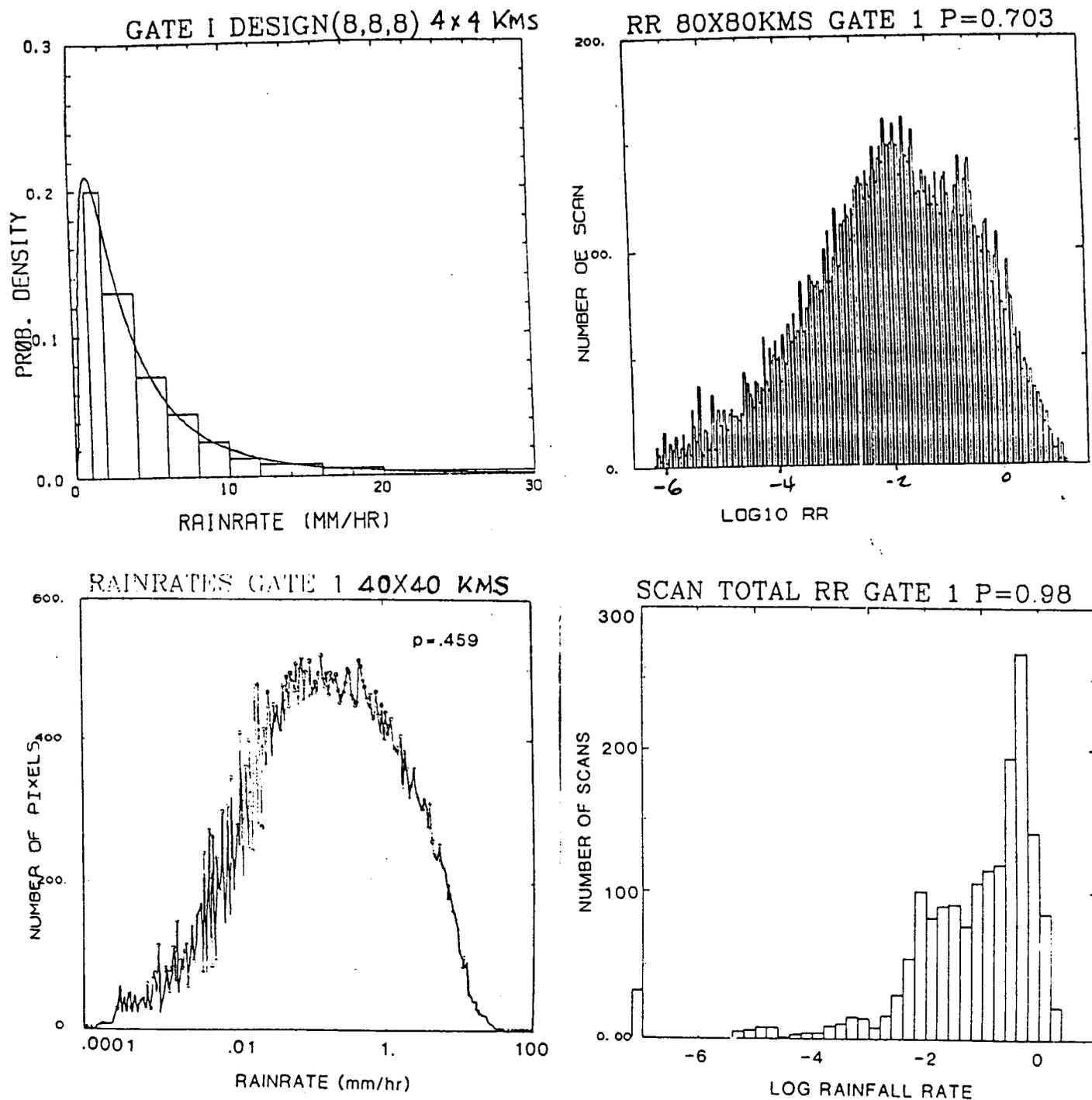


Fig. 7. Histogram of rainrate for different resolutions. The histogram with 4 km resolution is plotted on a linear scale of rainrate whereas the others are in logarithm scale to depict the resemblance to the lognormal distribution.

functions, sampling requirements for different area averages are calculated. The temporal autocorrelation function for different areal averages in our model is also calculated. Our results are similar to those computed by Laughlin (1982)(See attachment C.).

5.2 Stochastic Regression Model

A regression model of replacement and immigration is also developed (Kedem and Chiu 1986)(attachment D). In this model, the number of raindrops within a rain volume is considered a random variable which can be changed by replacement and/or immigration. The model takes the form

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n,i} + I_n, \quad n = 1, 2, \dots$$

where x_n is the number of drops at the n th step which can be replaced by y fresh drops, and I denotes the number of immigrants entering the rain volume. It can be shown that if

$E(Y_{n,i})$ is small but greater than zero; and

$E(I_n)$ is close to but less than unity

then x_n follows a lognormal distribution. This provides a justification for the use of the lognormal distribution in fitting rainfall data. It also bypasses the use of the law of proportionate effect often quoted to achieve lognormality. When the model parameters are estimated from the GATE data, it was found that these conditions are satisfied for large area averages. Since the sampling frequency is 15 minutes during GATE, this result suggests that there is a spatial and temporal range in which the lognormal distribution can provide a good description of the rainfall rates. The range over which the lognormal distribution provides a good fit to the data is investigated in the following section.

6. Scale Dependence Of Rain Field Parameters

The three parameters of a mixed lognormal distribution that describe a rainfall distribution are dependent on the scale of averaging. The threshold that define extreme events is (in this case precipitation), therefore, also dependent on the averaging time/area. An obvious question then is over what range in time and space does lognormality provide a good description of rainrate distributions. As a practical concern, it is of interest to examine the dependence of the intermittency factor on the pixel size which is determined by the resolution of satellite sensors and the altitude of the orbits.

We have examined the GATE data for different area averages. The three parameters, p , α , β^2 in the mixed distribution model of the GATE rain field have been computed for different averaging areas in the range from 4 kms to about 350 kms (whole of GATE B-scale area) on the side. Figure 6 shows the results on a log-log scale. The linear relation between the log of the parameters and the square root of the averaging area is clearly discernible at least over the range from areas of 4 kms to 80 kms on the side. The linear dependence suggests a power law dependence of the parameters on the averaging area for sampling frequency of 15 minutes.

Figure 7 shows the histograms of rainrates for square pixels of 4, 40, 80, and about 350 kms on the side. The histograms are calculated on a logarithm scale. The logarithm scale is used because a lognormal distribution on a linear scale is a normal distribution on log scale. Another advantage of using the logarithm scale is that the no rain category appears at minus infinity. Hence a threshold for the occurrence of events can be defined with no ambiguity.

The general shift from the high values towards the low values are noted as the resolution decreases. the skewness in the curve is also increased accordingly. the spatial averaging process smoothes out the high rainrates and inflates the population at the low rainrate portion. These shifts occur when nonrainy pixels are averaged with rainy pixels.

7. Summary And Estimate of Technical Feasibility

A logistic regression model has been developed to estimate the probability of rainfall given covariate observations such as radiometric measurements. The parameters of the model are estimated by maximizing a likelihood function. The significance of the estimators of the model can be readily tested by a ratio of the likelihoods. This method of testing allowed identification of important covariates as well as the way in which the covariates enter into the estimation. The logistic model has been tested on the rainfall data collected during phase 1 of GATE and successfully predict the observation for phase 2 of GATE. A major finding is the usefulness of the fractional rain area within a pixel. This parameter gives a better regression model than that which uses only the fractional area of heavy precipitation. The index of heavy precipitation area is interpreted as the cloud index of Arkin in estimating rainfall through the use of infrared measurements.

To investigate further this relation, a correlation analysis was performed on the logarithm of GATE rainfall data and the logarithm of the fractional rain area. Correlation coefficients of 0.99 are obtained for both phases of GATE. These coefficients are larger than the value of 0.87 between the cloud index proposed by Arkin and the total rain volume.

To estimate the mean and variance of areal average rainfall, a mixed distribution model was proposed and was found to model the distribution of rainfall data in GATE quite well. The parameters of the mixed distribution model consists of two parts: a discrete probability of no rain and a continuous distribuiton which describes

the rainy part of the mixed distribution. It was found that the rainy part of the distribution is fairly well described by a lognormal distribution. The discrete part of the mixed distribution is interpreted as a measure of intermittency which found familiarity in the study of turbulent flows.

Because of the nature of intermittency, we propose the use of the conditional probability in describing the rain field. The probability of rain conditional on rain at a different time/space for the GATE period is computed. The anisotropy in space is clearly discernible for GATE 2.

To broaden the data base for the testing of the logistic model, a data set of three dimensional rain cloud structure derived from radar echoes during GATE is used to compile a data set of cloud height and surface rainfall. The conditional probability distribution of cloud height and surface rainfall is calculated. The relationship between surface rainfall and cloud height is consistent with earlier results on tropical cloud systems.

A model is developed to simulate rain fields observed in GATE. The simulation model preserves the lognormality and intermittency characteristics of GATE and the temporal autocorrelation function computed from rainfields generated by the model is very similar to that of Laughlin (1982) in estimating the sampling errors associated with satellite observations.

A regression model of replacement and immigration is also developed which is capable of producing a lognormal distribution in some asymptotic limits. These asymptotic conditions are observed in GATE for large area averages (40 kms on the side) but not for small area averages (4 kms on the side).

Since the GATE data is taken every 15 minutes, this suggest that the lognormal distribution is a valid approximation within some range of averaging in time and space. This range of validity is investigated by computing the parameters of the mixed lognormal distribution model for different area averages in GATE. It was found that these parameters varies as a power of the averaging area at least over the range from areas of 4 to 80 kms on the side.

We have demonstrated the feasibility of using the logistic regression in identifying important covariates in the estimation of rainfall. A logical next step is to refine the logistic technique by the method of partial likelihood (Cox 1975). This method allows the disposition of the assumption of independence of the estimators. To examine the contribution of the radiometric data from the different atmospheric channels, we need to put together a data of concurrent visible, infrare and microwave data. Model generated rain fields are also needed to calibrate the logistic technique. Rainfall statistics derived from analyses of the rainfall data sets will proved to be useful in providing the required constraints for these simulation models.

**APPENDIX: Criterion for Independence
of Conditional Probabilities**

We want to compute the lag time between observations such that the observed events becomes statistically independent. Assuming stationarity, the condition for statistical independence can be obtained as follows. Let A(B) be the event that the rainrate (R) at time $t(t - \tau)$ in a fixed location be greater than some prescribed value, say R_0 , i.e.,

$$A: R(t) > R_0$$

$$B: R(t - \tau) > R_0$$

The probability of A conditional on B can be written as

$$P(A|B) = P(A \cap B)/P(B)$$

if A and B becomes statistically independent, then

$$P(A \cap B) = P(A) P(B)$$

so the condition for statistical independence is

$$P(A|B) = P(A)$$

where $P(A)$ is the probability of rainrate greater than R_0 .

REFERENCES

- Adler, R., and Mack, R., "Thunderstorm Cloud Height-Rainfall Rate Relations for Use with Satellite Rainfall Estimation Techniques," J. Clim. Appl. Meteor., 23(2), 280-296, 1984.
- Arkell, R., and Hudlow, M., "GATE International Meteorological Radar Atlas," U.S. Dept. of Commerce, NOAA, U.S. Government Printing Office, Washington, DC, 1977.
- Arkin, P.A., "The Relationship Between Fractional Coverage of High Cloud and Rainfall Accumulation During GATE over the B-Scale," Mon. Weath. Rev., 107, 1382-1387, 1979.
- Atlas, D., and Thiele, O., "Precipitation Measurements from Space," Workshop Report, NASA/Goddard Space Flight Center, Greenbelt, MD 20771, 1981.
- Austin, P. M., and Geotis, S., Air Sea Interaction, eds. F. Dobson, L. Hesse, and R. Davis, "Precipitation Measurements over the Oceans," Plenum Pub. Corp., 227 West 17th St., New York, NY 10011, 1980.
- Barrett, E. C., and Martin, D. M., "The Use of Satellite Data in Rainfall Monitoring," Academic Press, New York, NY, 340, 1981.
- Chiu, L. S., and Kedem, B., "Satellite Rainfall Retrieval by Logistic Regression," 2nd Conf. Satellite Meteorology, 224-227, Amer. Meteor. Soc., Boston, MA, 1986.
- Cox, D. R., "Partial Likelihood," Biometrika, 62, 269-276, 1975.
- Cox, D. R., and Oakes, D., Chapman and Hall, "Analysis of Survival Data," New York, NY 10017, 201, 1984.
- Cox, D. R., Methsen, "The Analysis of Binary Data," London, 1970.
- Doneaud, A. A., Smith, L., Dennis, A. S., and Sengupta, S., "A Simple Method for Estimating Convective Rain Volume over an Area," Water Resour. Res., 17, 1676-1682, 1981.
- Gill, A. E., "Atmosphere-Ocean Dynamics," Academic Press, NY, 662, 1982.
- Horel, J. D., and Wallace, J. M., "Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation," Mon. Weath. Rev., 109, 813-829, 1981.
- Hudlow, V. L., M. D., Pytlowany, P. J., Richards, F. P., and Hoff, J. D., "GATE Radar Rainfall Processing System," NOAA Technical Memorandum, EDIS 26, NOAA, Washington, DC, 1979.

- Jackson, I. J., "Relationship Between Raindays, Mean Daily Intensity, and Monthly Rainfall in the Tropics," J. Climatology, 6, 117-134, 1986.
- Laevastu, T., Clarke, L., and Wolff, P. M., "Oceanic Part of the Hydrological Cycle," WMO Reports on WMO/IHD Projects, 11, WMO, Geneva, Switzerland, 1969.
- Lovejoy, S., and Austin, G. L., "The Delineation of Rain Areas from Visible and IR Satellite Data for GATE and Mid-Latitudes," Atmosphere-Ocean, 17(1), 77-92, 1979.
- Miller, D. H., "Water at the Surface of the Earth," Academic Press, New York, NY 10003, 1977.
- Olsen, R. L., Rogers, D. V., and Hodge, D. B., "The αR Relation in Calculation of Rain Attenuation," IEEE Trans. Ant. and Prop., AP-26(2), 318-329, 1978.
- Pytlowany, P. J., Richards, R. P., and Hudlow, M. D., "GATE Radar CAPPI Derivation for Three-Dimensional Reflectivity Analyses," NOAA Tech. Memorandum, EDIS 27, NOAA, Washington, DC, 1979.
- "Radar Characteristics of Tropical Convection Observed During GATE: Mean Properties and Trends over the Summer Season," Mon. Weath. Rev., 105, 964-980, 1977.
- Schertzer, D., and Lovejoy, S., "The Dimension of Intermittency of Atmospheric Dynamics," Turbulent Shear Flows, 4, 7-33, Springer-Verlag, Berlin, Heidelberg, 1985.
- Theon, J., North, G., Wilheit, T., and Thiele, O., "Tropical Rainfall Measuring Mission," Workshop Report, Nov. 1985, NASA/Goddard Space Flight Center, Greenbelt, MD (to be published), 1986.
- Wilheit, T. T., Chang, A. T. C., Rao, M. S. V., Rodgers, E. B., and Theon, J. S., "A Satellite Technique for Quantitatively Mapping Rainfall over the Oceans," J. Appl. Meteor., 16, 551-560, 1977.
- Wu, R., and Weinman, J. A., "Microwave Radiance from Precipitating Clouds Containing Aspherical Ice Combined Phase, and Liquid Hydrometeors," J. Geophys. Res., 89(D5), 7170-7178, 1984.

ORIGINAL PAGE IS
OF POOR QUALITY

P4.31

Reprinted from Preprint Volume, Second Conference on
Satellite Meteorology/Remote Sensing and Applications,
May 13-16, 1986, Williamsburg, Va. Published by the
American Meteorological Society, Boston, Mass.

SATELLITE RAINFALL RETRIEVAL BY LOGISTIC REGRESSION

Long S. Chiu
Applied Research Corporation
8201 Corporate Drive, Suite 920
Landover, Maryland 20785

and

Benjamin Kadem
Department of Mathematics
University of Maryland
College Park, Maryland 20742

1. INTRODUCTION

The retrieval of meteorological quantities from satellite observations is based on covariate information such as radiometric measurements or physical quantities derived from them. The covariate information is influenced by factors other than the desired meteorological variable. The situation is further complicated by the different resolutions of the different sensors. It is useful to identify important covariates for the prioritization of transmission of data and to ascertain the possibility of on board processing.

Accurate measurement of tropical rainfall is crucial for the advancement of our understanding of the large-scale dynamics of the ocean/atmosphere system. An account of rainfall monitoring techniques from satellites is given by Barrett and Martin (1981). A satellite mission for the monitoring of tropical rainfall has been proposed to NASA (Theon et al 1986). Three instruments are proposed for the mission: a radar, an Advanced Very High Resolution Radiometer (AVHRR) and a microwave instrument, possibly an Electrically Scanning Microwave Radiometer (ESMR). The expected outcome from this mission is at least three years of rainfall data derived from concomitant covariate observations.

In the following a logistic model that can effectively accommodate covariate information, but which has not been used in the context of rainfall estimation, is described. A major difference between linear regression and logistic regression is that the former technique maximizes the variance explained while in logistic regression a likelihood function, or probability of an event, is maximized. The output from such a model is the distribution of rainrate categories from which standard errors can be estimated. The significance of the covariates can be tested rather readily. An example of the logistic model is given for the scenario of the proposed tropical rainfall monitoring mission from which microwave observations and fractional rain area measurements may be available.

2. THE LOGISTIC MODEL

The logistic model is useful in determining the relationship between the distribution

of a random variable and a set of covariates. It has been applied in various forms in reliability testing and the analysis of survival data (Cox and Oakes, 1984). A detailed treatment of the logistic model is given by Cox (1970). We are interested in the relationship between rainfall rate averaged over an area R and t , the vector of covariate variables related to R . For the event $R \geq R_0$, the logistic model is given by

$$P(R > R_0) = [1 + \exp(-b't)]^{-1}$$

where $P(R > R_0)$ is the probability that the rainfall rate R exceeds R_0 and

$$\underline{b} = (b_0, b_1, b_2, \dots, b_k)$$

is a vector of constants. From n observations of R , the b 's can be estimated by the method of maximum likelihood. Let R_0 be fixed so that the a binary variable Y can be defined as

$$Y = \begin{cases} 1 & R \geq R_0 \\ 0 & \text{otherwise} \end{cases}$$

The logistic model becomes

$$P(Y = 1) = [1 + \exp(-(b_0 + b_1 t_1 + \dots + b_k t_k))]^{-1}$$

where the t 's are covariate variables. We assume Y_1, Y_2, \dots, Y_n are conditionally independent given the covariate information. Then the likelihood function $L(\underline{b})$ is given by

$$L(\underline{b}) = \prod_{i=1}^n \frac{Y_i \exp(\underline{t}_i' \underline{b})}{[1 + \exp(\underline{t}_i' \underline{b})]}$$

and the asymptotic covariance matrix is given by

$$(-E(\partial^2 \log L(\underline{b}) / \partial b_i \partial b_j))^{-1}$$

where E is the expected value. To test the significance of the regression coefficients, we use the likelihood ratio test

$$\lambda = -2 \log L_0 / L_1$$

where L_1 is the maximized likelihood under the full model and L_0 is the maximized likelihood under the hypothesis that the some of the regression coefficients are zero. If q of the b 's are assumed to vanish, then λ follows asymptotically a chi-square distribution with q degrees of freedom.

3. A SCENARIO

We consider the scenario when concomitant observations of microwave temperature and fractional rain area are available. The rainfall data collected during the phase I of GATE are used. The basic data are radar-estimated rainrates on 4 by 4 km² pixels and measurements are made at 15 minute intervals. From the basic data, temperature and fractional rain area data for the scenario are generated as follows.

We assume that the microwave instrument measures the temperature over an area of 32 by 32 km² (i.e. 8 by 8 pixels) which is the unit area for our scenario. To calculate a mean temperature over the 32 by 32 km² box, a simple relation between the rainrate (r) and temperature (T_R)

$$T_R(r) = T_{av}(1-x) + T_s \epsilon x + (1-\epsilon) T_{av}(1-x)x \quad (1)$$

where T_{av} is the average temperature of the atmospheric column (=270K), T_s is the surface temperature (=290K), ε is surface emissivity (=0.5), x=exp(-τ) is optical thickness, is used. τ is approximated as τ = 0.2r, with r in mm/hr. The dependence of τ on the height of the rain column is ignored in this case. A functional relation between R and T_R is shown in fig. 1. From the rainrate at each pixel, a microwave temperature is computed. The microwave temperatures are then averaged over the 32 by 32 km² box to yield the average temperature (T). The fractional rain area (F) is obtained by dividing the number of pixels with rainrate in excess of 1 mm/hr by 64. The box averaged rainrate (R) is obtained by averaging the rainrates over the 64 pixels. Fig. 1 shows the scattered diagrams of R versus T. The fact that T_R(R) is greater than T follows from Jensen's Inequality (Feller, 1966). Fig. 2 shows the relationship between F and R. The strong correlation between R and F is also noted by Lovejoy (1980) for the phase III of GATE for the whole GATE area. Also included in our analysis are fractional rain area with rainrates in excess of 20 mm/hr (F₁). The data have been extracted from a 32 by 32 km² grid box in the center of the GATE area. Characteristics of the time series are summarized in table 1. Data from another box located approximately 100 km to the south of the first is used for validation.

Table 1. Characteristics of the time series

Variable	mean	s.d.	minimum	maximum
R (mm/hr)	.44	1.48	0.0	17.5
T (K)	151.5	16.67	145.0	263.1
F	.072	.17	0.0	1.0
F ₁	.006	.03	0.0	.47

ORIGINAL PAGE IS
OF POOR QUALITY

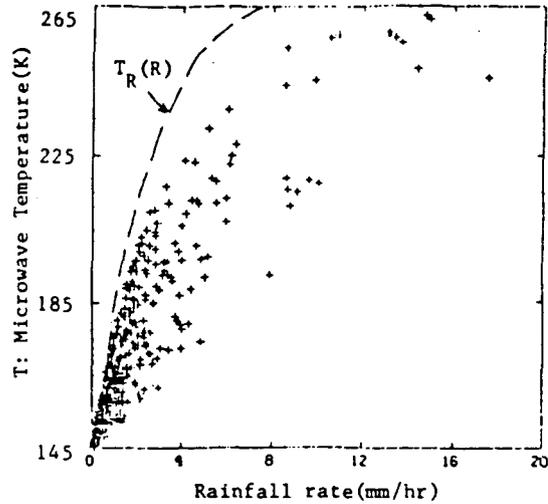


Fig. 1. Scatter diagram of rainfall rate over the box (R) and microwave temperature (T). The dashed curve shows that functional relationship between T_R and R.

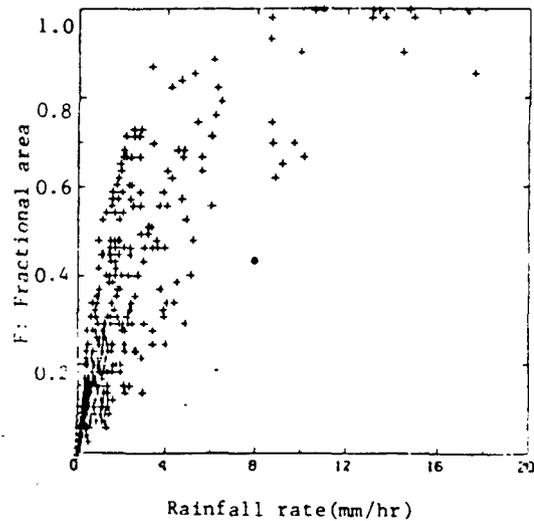


Fig. 2. Scatter diagram of rainfall rate (R) and fractional rain area with rainrates in excess of 1 mm/hr (F).

4. RESULTS

The full logistic model is of the form

$$P(R>1) = [1 + \exp(-(b_0 + b_1 F + b_2 F_1 + b_3 T + b_4 T_L))]^{-1}$$

where R, F, F₁ and T are rainrate, fractional area with rainrate in excess of 1 mm/hr, fractional area with rainrate in excess of 20 mm/hr and the temperature over the 32 by 32 km² box. T_L is T lagged at 1 time unit (i.e. 15 minutes). A total of 10 different models have been run and the regression coefficients are presented in table 2. The maximum log likelihood ranges from -309.6 for a model with F₁ as the only regressor (model 10) to -28.6 for the full model (model 1). From the table, important covariates can be identified. For illustration purposes, we consider models 1 and 8. The hypothesis we want to test is

$$H_0: b_2 = b_3 = b_4 = 0$$

ORIGINAL PAGE IS
OF POOR QUALITY

The likelihood ratio test yields

$$\lambda = -2 [-159.1 + 28.6] = 260$$

and the 5% significance level of $\chi^2(3)$ is 7.81. Hence H_0 is rejected. To see if covariate T_L contributes to the estimation, the hypothesis $H_1: b_4 = 0$ is tested. To do this, we compare model 1 and 2 and obtain the λ value of $-2[-30.1 + 28.6] = 3$. The 5% level for $\chi^2(1)$ is 3.8 and H_1 has to be accepted. In this way, it is readily seen that b_1, b_2, b_3 are highly significant.

The goodness of the model is tested by applying it to the validation time series which is taken from an area to the south of the center of GATE (see section 3). Model 2 is adopted for validation, i.e. we use

$$P(R > 1 \text{ mm/hr}) = [1 + \exp(-(-207.8 - 61.7F + 307F_1 + 1.34T))]^{-1}$$

The values of F, F_1 and T taken from the location designated for validation are substituted in (2) and the probability calculated. We defined as a goodness of fit criterion the mean square error

$$\text{MSE} = 1/n \sum_{i=1}^n (P(Y_i=1) - Y_i)^2$$

This is 0.005 for the prediction using model 2. It can be seen from fig. 3 that the prediction matches the observations very well.

Table 2. Regression Coefficients for Different Logistic Models

parameter	F	F_1	T	T_L		
regression	b_1	b_2	b_3	b_4	maximized	
coeff. b_0					log	
model					likelihood	
1	-217.7 (46.9)	-66.8 (16.7)	309.8 (61.0)	1.33 (.30)	0.077 (.046)	-28.6
2	-207.8 (44.8)	-61.7 (16.0)	307.0 (59.4)	1.34 (.29)	---	-30.1
3	-17.7 (4.34)	16.6 (3.2)	249.7 (31.4)	---	0.071 (.028)	-62.8
4	-7.27 (.68)	22.5 (2.4)	240.5 (29.3)	---	---	-65.9
5	-166.2 (18.5)	-62.8 (8.6)	---	1.10 (.12)	---	-76.2
6	-48.1 (3.46)	---	---	.33 (.04)	-.035 (.027)	-115.0
7	-47.9 (3.4)	---	---	.29 (.02)	---	-116.1
8	-4.9 (.27)	21.0 (1.4)	---	---	---	-159.1
9	-34.9 (2.1)	---	---	---	.21 (.01)	-188.1
10	-3.1 (.12)	---	249.3 (19.5)	---	---	-309.6

The s.d.s of the coefficients appear in parentheses

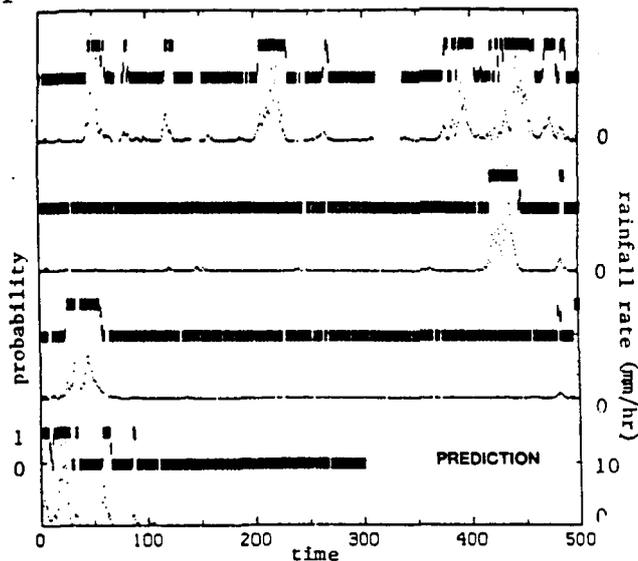


Fig. 3. Time series of predicted probability of exceeding 1 mm/hr and observed rainrate in another location.

5. DISCUSSION

In our example, the microwave temperature data, which integrates the effect of partially filled non-uniform field of view of the microwave sensor, are not independent since they are derived from a rainrate-temperature relationship but in the presence of T , the addition of T_L as a regressor does not improve the model significantly. This can be seen by comparing the maximized likelihood in models 1 and 2 and models 6 and 7. The interpretation is that most of the information is contained in T and the addition of T_L adds redundant information. In the absence of T , the inclusion of T_L improves the model as the maximized log likelihood is increased from -65.9 (model 4) to -62.8 (model 3). This increase is significant by the log likelihood ratio test.

The relationship between average rainrate and fractional rain area has strong implications for rainfall estimation. From infrared imagery, Arkin (1979) accounted for a large fraction of the rainfall variance by considering the number of pixels below a threshold temperature. Implicit is the assumption that convective rainfall, which is in the heavy rainrate portion of the rain spectrum, is produced in deep cumulus clouds. How important the heavy rainrates are in determining the total rainfall can be examined by considering F_1 . If we compare the maximized log likelihood in model 8 and 10, we see that model 8, with fractional low rainrates (>1 mm/hr) area as the only regressor is better than model 10 which uses only F_1 in the maximum log likelihood sense. If both are used, a significantly better model is obtained (model 4).

In this report, the potential of logistic models in rainfall estimation is demonstrated. We plan to extend our analysis using actual satellite observations such as the ESMR measurements taken on board the NIMBUS V satellite.

6. REFERENCES

- Arkin, P. A., 1979: The Relationship Between Fractional Coverage of High Cloud and Rainfall Accumulations during GATE over the B Scale, Mon. Weath. Rev., 107, 1382-1387.
- Barrett E. C. & D. W. Martin, 1981: The Use of Satellite Data in Rainfall Monitoring, Academic press, New York, NY, 10003, 340pp.
- Bras, R. L. & I. Rodriguez-Iturbe, 1976: Net Work Design for the Estimation of Areal Mean of Rainfall Events, Water Resources Res., 12, 1185-1195.
- Cox, D. R., 1970: The Analysis of Binary Data, Methsen, London.
- Cox, D. R. & D. Oakes, 1984: Analysis of Survival Data, Chapman and Hall, New York, NY, 10017, 201pp.
- Feller, W., 1966: An Introduction of Probability Theory and its Application, Vol 2, Wiley, New York, NY, 626pp.
- Lopez, R. E., J. Thomas, D. O. Blanchard & R. L. Holle, 1981: Estimation of Rainfall over an Extented Region Using only Measurements of the Area Covered by RADAR Echoes, 21st Conf. on Radar Meteorology, Amer. Meteor. Soc., Calgary, Canada.
- Lovejoy, S., 1981: The Remote Sensing of Rain, Ph.D. thesis, Dept. of Physics, McGill University, Montreal, Canada, 102pp.
- Theon, J., G. N. North, T. Wilheit & O. Thiele, 1986: Tropical Rainfall Monitoring Mission Workshop Report, NASA, Nov., 1985, Calverton, Maryland. (To be published)

Attachment B

Estimating Time Mean Areal Average Rainfall:

A Mixed Distribution Approach

Abstract

A technique to determine the time mean areal averaged rainfall is developed. The approach taken is to model the distribution of rainrate by a mixed distribution. The model is tested on rainfall data collected during GATE (GARP -Global Atmospheric Research Program- Atlantic Tropical Experiment). Sampling designs which select only a portion of the rain data are used. It was found that a lognormal distribution provides an excellent fit to the rainy portion of the distribution. The results are insensitive for sampling frequencies in the range of half to a few hours in time and 16 to 40 kms in space. Sampling errors are about 10% of the mean or less for sampling designs which mimic observations by satellites that are polar orbiting or have a low inclination. An important parameter in the model is the probability of rain which correlates significantly with the average rainfall. This is consistent with earlier results such as those which relate the number of rain days and rain intensity to monthly rainfall and the use of the Area Time Integral (ATI) in estimating rain volume. The need for microwave sensors in satellite rainfall monitoring systems is stressed and an algorithm for estimating monthly mean rainfall from microwave sensor measurements such as the Electrically Scanning Microwave Radiometer (ESMR) or a radar is proposed.

1. Introduction

The latent heat released during the process of precipitation constitutes a major component in the forcing of atmospheric circulations (Lorenz 1967). Theoretical as well as empirical studies have shown that variations in tropical forcing are instrumental to anomalous weather patterns world wide (Horel and Wallace 1981, Gill 1982). Accurate measurements of precipitation as an index of atmospheric variability are therefore useful both as a tool in diagnostic as well as prognostic studies of atmospheric circulations.

Over land the problem of estimating time mean areal average rainfall has occupied hydrologists for a long time (Eagleson 1967, Rodriguez-Iturbe & Mejia 1974, Bras & Rodriguez-Iturbe 1976, Bras & Colon 1978). The interest is in river/ground water flow, flood forecasting and catchment hydrology. A major emphasis is the modeling of the rain field as a two dimensional random field. Once the parameters of the random field are estimated, the mean and variance of rainfall total can be calculated. The applicability of various mapping techniques to fill in missing data has been assessed by Creutin and Obled (1982) and approaches to network designs have been summarized by Moses (1982).

Because of the huge extent of the tropical oceans and the errors associated with in situ measurements on board ships, satellite observation is probably the ultimate mode by which precipitation measurements can be made over the vast oceans (Austin and Geotis 1982; Atlas and Thiele 1981). A review of various satellite rainfall estimation techniques is given by Barrett and Martin (1981).

The method of sampling by satellites differs from that by networks of land based rain gauges. The former provides snap shots of precipitation information, in terms of radiances from different sensors, while the latter gives continuous rain gauge measurements at isolated stations.

An alternative approach to modeling the temporal and spatial structure of the rain field is to consider the distribution of rainfall categories in the estimation of time areal mean rainfall. If one considers continuous sampling at a fixed location, it is obvious that the rain volume can be estimated either through integrating the time series of rainfall rate or via computing the mean of the rainrate distribution. Once the distribution of rainfall rates is obtained, the mean and variance of the total rainfall can be estimated.

The climatology of heavy rainfall statistics at points or rainfall statistics along lines has been studied because of their importance in microwave communication (Rogers 1976, Drufuca & Rogers 1978, Lin 1976, Freeny & Gabbe 1969). The climatology of rainfall statistics for the whole rain spectrum has also been compiled for climatic studies. A common feature of these cumulative distributions of rainfall is that their functional forms are quite similar for a diversity of geographic regimes (Jones and Sims 1978). Oftentimes, a lognormal distribution is quoted.

The estimation of time mean areal average rainfall is determined by two factors: how often does it rain and how hard does it rain when it rains? An approach that address the first question is the use of the so called "Area Time Integral" (ATI) in estimating rain volume (Lopez 1976, Donuead et al 1982a). The ATI is the integral over time of the area of precipitation as seen by radar. The use of a convective index (Arkin

1979) and the delineation of rain area from visible and infrared satellite imageries (Lovejoy & Austin 1982) seem to fall into this category.

Jackson (1986) examines the two factors in toto by studying the relationship between the number of raindays in a month, the average rainfall intensity in raindays and the monthly total rainfall.

In this report, we propose a mixed distribution model for the estimation of time mean areal average rainfall. The model is structured so that both factors can be combined in a single formulation. A mixed distribution is described (section 2) and applied to rainrate data collected during GATE. The GATE rainfall data and estimation procedures are described in sections 3 and 4. Section 5 presents our results for different sampling designs. The relative importance of the different contributing factors in the estimation scheme are examined in section 6. Section 7 discusses and concludes our findings.

2. Mixed Distribution

Most statistical distributions encountered in practice are either discrete or continuous. In the discrete case, the random variable assumes a finite (or countable) number of values while in the continuous case, the variables assumes all values in the interval which can be finite or infinite. However, there are situations when the random variable assumes distinct values with positive probability and other values in the continuous interval. Such a random variable is said to have a mixed distribution. An example of a mixed distribution comes from the reliability and life time testing of light bulbs. When a light bulb is turned on at time zero, there is a positive probability that it will be burnt out immediately. If the light bulb is not burnt out it is left on for an hour. The probability that the light bulb may be burnt out during the hour is positive. Hence the distribution of X has a jump at $X=0$ while in the interval $(0,1]$, it is continuously differentiable (see Hogg and Tanis, 1977). The mixed distribution can be considered a special case of a mixture distribution.

In the case of rainfall rate sampling, the probability of measuring no rain at any instance is large. Many previous studies have focused on the estimation of the raining portion of the distribution. It turns out, as we shall demonstrate in this paper, that the no rain probability is an important parameter in the estimation.

The mixed distribution model of rainfall rates can be described as follows: Let R be the rainfall rate sampled in space and time. The cumulative probability distribution (CPD) can be written as

$$F (R) = P (R < r)$$

where $P (R < r)$ is the probability that the rainfall rate R is less than

some fixed r . Let $P(R=0) = 1 - p$. The conditional density of R given $R > 0$ is

$$f(r) = 1/p \partial F / \partial r$$

It follows that the generalized density $g(r)$ takes the form

$$g(r) = \begin{cases} 0 & r < 0 \\ 1 - p & r = 0 \\ p f(r) & r > 0 \end{cases} \quad (1)$$

where f is the density of R conditional on $R > 0$. Thus the CPF can be written as

$$F(r) = (1-p) + p \int_0^r f(x) dx, \quad r > 0$$

The expected mean of R is

$$E(R) = p \int_0^{\infty} x f(x) dx$$

and the variance

$$\text{Var}(R) = p \left\{ \int_0^{\infty} x^2 f(x) dx - p \left[\int_0^{\infty} x f(x) dx \right]^2 \right\}$$

The above mixed distribution can be described by several parameters, p and $\underline{\theta}$, ($\underline{\theta}$) = ($\theta_1, \theta_2, \dots$) such that

$$f(r) = f(r, p, \underline{\theta})$$

For a sample size of n which consists of m raining measurements and $n-m$ non-raining measurements, the likelihood function of p and $\underline{\theta}$ is given by

$$L(p, \underline{\theta} \dots) = (1-p)^{n-m} p^m f(r_1, \underline{\theta}), \dots, f(r_m, \underline{\theta})$$

The parameters can be estimated by various techniques such as the method of moments or maximum likelihood.

The maximum likelihood estimate of p is

$$\hat{p} = m/n$$

which is independent of any distribution model (i.e. f).

3. The data and sampling design

This technique has been tested by applying it to rainfall rate data collected during the GATE (GARP -Global Atmospheric Research Program- Atlantic Tropical Experiment). The GATE is an observational program conducted in the summer of 1974. During three roughly tri-weekly periods, each termed a phase, detailed rainfall measurements from rain gauges and radars on an array of research vessels were made over an area called the B-scale. The center of the B-scale area is located at 8.5N , 23.5E and encompasses an area of about 200 km in diameter. Arzell and Hudlow (1977) composited the radar measurements from ships and presented an atlas of the radar echoes at 15 minute intervals. Patterson et al. (1979) converted the radar measurements to rainrates and presented rainrate data in 4 by 4 km² bins.

To examine the spatial and temporal structure of the rain field various sampling designs have been used for the sampling. A design is described by 3 indices (n,k,l). The first index (n) denotes sampling frequency in time and the latter two (k,l) sampling frequencies in the east-west (x) and north-south (y) direction in space respectively. For example, the design (1,10,10) denotes sampling continuously in time (i.e. all 15 minute scans) but sampling spatially only every tenth pixel (40 km apart) in the x and y direction. This mimics the sampling by a raingauge network that continuously measures the rainrate with gauges placed 40 kms apart. The design (48,1,1) samples all pixels at an instance, but the time observations are taken only every 12 hrs (48 x 15 minutes). This mimics the sampling by a densely scanning sensor on board a polar orbiting satellite which passes the same location twice per day at the same local times (e.g., 12 a.m. and 12 p.m.).

4. Estimation procedure

Once the rainrate data are sampled, the parameters of the mixed distribution have to be estimated. The lognormal distribution has been adopted for the raining portion of the mixed distribution here. Much research effort has been devoted to modeling the rainrate distribution. Lognormality follows from the law of proportionate effects (Aitchison and Brown 1963) and physical cloud models have been proposed which can explain the lognormal distribution of cloud sizes (Lopez 1977). Studies using the GATE radar data have shown that rainrates, size of radar echoes and their durations follow lognormal distributions (Houze and Cheng 1977, Houze and Betts 1981).

4.1 Lognormal distribution

The lognormal distribution can be written as:

$$f(r) = 1/(r \sigma \sqrt{2\pi}) \exp [-(\log r - \mu)^2 / 2 \sigma^2] , r > 0 \quad (2)$$

The mean α and variance β^2 of the lognormal distribution are

$$\alpha = \exp (\mu + \sigma^2/2)$$

$$\beta^2 = \exp (2 \mu + \sigma^2) [\exp(\sigma^2) - 1]$$

(see Johnson and Kotz, p. 115, 1970). Consequently, The mean and variance of the complete mixed distribution is given by

$$E(R) = p \exp (\mu + \sigma^2/2)$$

$$\text{Var}(R) = p \exp (2 \mu + \sigma^2) [\exp (\sigma^2) - p]$$

$$= p \alpha^2 [\exp(\sigma^2) - p]$$

A thorough discussion of the lognormal distribution is given by Aitchison and Brown (1963).

4.2 Minimum χ^2 Estimation

We have grouped the GATE rainrate data into different rainfall rate categories. The categories are 0-1, 1-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-16, 16-20, and >20 mm/hr. The first category was chosen because it is difficult to distinguish non-raining pixels and pixels with only a trace of rain which may be due to noise in radar reflectivity. This low cutoff at 1 mm/hr has been used in earlier studies (Austin & Geotis 1979). Because of this truncation, the estimates are slightly lower (about 2% of the estimated mean) than the means calculated directly even after adjustments have been made. For a lognormal distribution with typical parameters found in our study, the interval (0,1) contains about 10 to 15% of the rainy pixels.

Minimum chi square estimation is used in our procedure. This procedure is asymptotically equivalent to the maximum likelihood method obtained from (1) (Berkson 1980). The χ^2 variate can be written as:

$$\chi^2 = \sum_{i=1}^9 (o_i - e_i)^2 / e_i \quad (3)$$

where o_i 's are the number of raining pixels observed in the i th category and e_i are the corresponding frequencies from a lognormal distribution with parameters μ and σ .

The truncated distribution R_T for $R > 1$ mm/hr can be written as:

$$R_T = f(r) / \int_1^{\infty} f(r) dr$$

and so

$$P(R_T < a) = \left[\int_1^a f(r) dr \right] / \left[\int_1^{\infty} f(r) dr \right]$$

If the number of rainy pixels greater than 1 mm/hr is N , then

$$e_1 = N [\phi ((\log 2 - \mu) / \sigma) - \phi (-\mu / \sigma)] / [1 - \phi (-\mu / \sigma)]$$

where ϕ is the distribution function of the standard normal distribution. Similar expressions can be obtained for the other e_i 's.

The χ^2 estimation procedure can also shed some light on the complex structure of rainfall. The minimum χ^2 value can be inflated or deflated for statistically dependent data even though the fit to the distribution is still good (see appendix). The dependence of the observations (o_i 's) is introduced in the sampling process. For the (48,1,1) design, too much spatial dependence may be introduced while for the (1,10,10) design, too much temporal dependence may be introduced.

4.3 Standard Error

The expected mean and variance of the mixed lognormal distribution are given in subsection 4.1. Rewriting the expected mean as

$$E(R) = p \alpha \quad (4)$$

where α is the mean over the lognormal distribution (conditional on rain) and $\alpha = \alpha(\underline{\theta})$. In this case $\underline{\theta} = (\mu, \sigma)$. Since p and $\underline{\theta}$ are asymptotically independent, i.e., p and θ become statistically independent if the number of observations and the number of rainfall categories goes to infinity, the variance of $E(R)$ can be expressed as a sum of the variance of p and α . Consider the functional

$$h(p, \alpha) = p \alpha$$

If p and α are independent an expansion in the form of a Taylor series gives

$$\begin{aligned} h(\hat{p}, \hat{\alpha}) &= p \alpha + (\hat{p}-p) \frac{\partial h}{\partial p} + (\hat{\alpha} - \alpha) \frac{\partial h}{\partial \alpha} + \dots \\ &= p \alpha + (\hat{p}-p) \alpha + (\hat{\alpha} - \alpha) p \end{aligned}$$

and so

$$\text{Var}(\hat{p} \hat{\alpha}) = \alpha_0^2 \text{var}(\hat{p}) + p_0^2 \text{var}(\hat{\alpha})$$

If we consider the rain/no rain sequence as the outcome of a Bernoulli trial with success rate p , the variance of p can be estimated as

$$\text{var}(\hat{p}) \doteq \hat{p}(1-\hat{p})/m$$

The variance of α is (Aitchison and Brown, p46, 1963)

$$\text{Var}(\hat{\alpha}) \doteq \hat{\alpha}^2/m (\hat{\sigma}^2 + \hat{\sigma}^4/2)$$

Hence an approximate expression for the variance of $E(R)$ is

$$\text{Var}(E(\hat{R})) = \hat{p}^2 \hat{\alpha}^2/m (\hat{\sigma}^2 + \hat{\sigma}^4/2) + \hat{\alpha}^2 \hat{p}(1-\hat{p})/m \quad (5)$$

Although the assumption about the independence of p and α is not strictly valid, this expression provides an estimate of the standard error which is a good approximation to sampling errors obtained from ensembles of different sampling designs.

5. Results

The technique outlined in section 4 is applied to GATE data. Table 1 summarizes the results for the sampling design (8,8,8) for GATE 1. The χ^2 value is 6.74. For 6 degrees of freedom, the $\chi^2_{6, 95\%}$ is 12. Hence the hypothesis that the observed histogram can be fitted by a lognormal distribution cannot be rejected at the 95% level. With $\mu = 1.14$ and $\sigma^2 = 1.05$, the mean and variance of the lognormal distribution is 5.28 and 51.5 respectively. Fig. 1 shows the observed histogram for the design (8,8,8) and a fit to a lognormal distribution.

The GATE data have been sampled by various designs, with sampling frequencies of 1 to a few hours in time and 8 to 40 km in space. The results for GATE 1 and 2 are summarized in table 2. Within this frequency range of sampling in space and time, the χ^2 values are small and the lognormal distribution provides a good fit to the data. It is noted that these are sample estimates since each histogram is but one realization of a sampling design.

5.1 Sensitivity to saturation at high rainrates

A major problem associated with passive microwave sensors is the saturation at high rainrates. A test was conducted using the sample obtained from the (8,8,8) design, but with only 8 categories instead of 9. The two heavy rainrate categories are combined and the χ^2 statistics computed. The results for the two run are very similar, the estimated mean rainrates are within 5% of each other. Since only 8 categories are used, the degrees of freedom are reduced and the 95% confidence level is accordingly higher. This sensitivity test serves to illustrate the

possibility of applying this method to estimate rainfall from existing microwave measurements such as by the Electrically Scanning Microwave Radiometer flown on board NIMBUS V since this technique is not sensitive to the problem of saturation at the high rainrates.

5.2 Comparison with Gamma Distributions

In this subsection, we compare the χ^2 statistics between a lognormal and Gamma distribution. The Gamma distribution can be written as

$$f(r) = \lambda^\alpha / \Gamma(\alpha) r^{\alpha-1} \exp(-r\lambda), \quad r > 0, \quad \alpha, \lambda > 0$$

where $\Gamma(\alpha)$ is the Gamma function. A procedure as outlined in section 4 was carried out and the results for lognormal and Gamma distribution for some selected designs are given in table 3. The lognormal distribution consistently gives a better fit to the observed histogram in the minimum χ^2 sense. However as we shall demonstrate later, the exact choice of the rainrate distribution is not crucial in the estimation scheme.

5.3 Satellite sampling

To mimic the satellite sampling of rainfall by a polar orbiting satellite, the design (48,1,1) is applied to GATE. This is equivalent to sampling at roughly 12 hour intervals. Within the GATE period, there are periods when observations are missing. This design samples every 48 snapshots in the sequence, paying no attention to missing periods. Hence not all samples are at intervals of 12 hours. As a comparison, sample estimates from the designs (24,1,1), (72,1,1) and (96,1,1), i.e., sampling at intervals of 6, 18 and 24 hours are made for GATE 1 and 2.

The results are summarized in table 4. Although the χ^2 values are large, probably due to over-sampling in space and inadequate sampling in time, the estimated rainrates are quite close to the actual mean values of 0.45 and 0.37 for GATE 1 and GATE 2.

Since the design (48,1,1) samples every 48th snap shot, 48 distinct estimates from this design can be realized; i.e., the first estimate is derived from sampling the 1st, 49th, 97th, ..., etc snap shots, the second from the 2nd, 50th, 98th, ..., and so on to the 48th estimate. The estimated means from these sample designs form a sample distribution. The histogram of these estimated means are shown in fig 2 (left column). The means and standard deviations of these distributions are computed and indicated in the figures. It should be noted that the members of the sampling ensemble are not independent.

If the local diurnal cycle can be described entirely in terms of the first harmonic, sampling twice a day at 12 hours intervals is sufficient to specify the diurnal cycle. However any higher harmonics would introduced a bias. It is therefore of interest to examine the sampling errors associated with sampling frequencies slightly less than 12 hours so that the diurnal cycle is sampled through the course of about a month.

A unique feature associated with the proposed Tropical Rainfall Measuring Mission (Theon et al, 1986) is a revisiting time of the satellite every roughly 10 hours, giving a total of about 80 partial visits (30 complete views) of a 600 by 600 km² grid box. We mimic this strategy by the sampling design of (40,1,1). This design will actually give more than 30 complete visits per month. The important point here is that it will sample through the diurnal cycle. The histograms for the 40

estimated means are given in fig. 2 (middle column). There is a reduction in the standard deviation of the estimated means of the (40,1,1) design compared to the (48,1,1) design even the number of estimated means is less in the former case.

5.4 Network Sampling

To mimic the sampling by a network of gauges, the rainrates in GATE 1 and 2 are sampled by the (1,10,10) design. Similar to the procedure described in section 5.2, 100 samples are obtained from the (1,10,10) design. The 100 different samples are obtained by sampling which starts at different locations in space. From the 100 sample estimates of the rainfall rates, the sample means for GATE 1 and 2 are 0.446 and 0.367 mm/hr and the s.d.s are 2.6% and 2.2% of the means respectively. The normality of the estimated rainfall rates are tested by using a minimum χ^2 test similar to that described in section 4.2. The estimated rainfall rates are divided into 10 equal interval categories and the χ^2 values are computed to be 4.9 and 7.2 respectively for GATE 1 and 2 compared to $\chi^2_{7,.95}=14$. The hypotheses of normality therefore must be accepted at the 95% level.

6. Correlations and analyses of variance

In estimating the standard error, the independence of p and α is assumed. This assumption can be examined by considering the correlation between the mean rainrate conditional on rain (α) and p for each of the 15 minute observations. p is calculated as the percent of pixels with rain rate in excess of 1 mm/hr to the total number of pixels. α is calculated as R/p and the condition of $p=0$ is not considered in the calculation.

The linear correlation between α and p is 0.58 (0.52) whereas that between p and R is 0.94 (0.94) for GATE 1 (2). Similar relations are also found in the GATE 3 data (Lovejoy 1982). The correlation coefficients between the logarithm of the quantities are higher. The results are summarized in fig. 3 which shows the scatter diagrams between the three quantities. The correlation coefficient between $\log p$ and $\log R$ is 0.99 for both GATE 1 and 2 whereas that between $\log \alpha$ and $\log p$ is 0.63 and 0.51 respectively for GATE 1 and 2.

The histograms of R , p and α are given in fig. 4. The distribution of p is skewed. Since the value of p lies between 0 and 1, a fit to a beta distribution may be appropriate. There is zero probability that the whole of GATE area is totally covered with rain. Obviously, the parameters of the distribution are dependent on the size of the area. Chiu and Kedem (1986) examined the fractional area for an area of about 40 by 40 km² using the same GATE data. In this instance, there are times when the smaller area (40 by 40 km²) are fully covered. There are times when the GATE area is totally rain free for a cutoff of 1 mm/hr. If a lower cutoff is used, e.g. 0 mm/hr, the

fractional rain free time is accordingly reduced.

We also examined the contribution of variance of p and α to that of R . If we take the logarithm of the equation

$$\bar{R} \text{ (mm/r)} = p \alpha \text{ (mm/hr)}$$

we get

$$\log \bar{R} = \log p + \log \alpha$$

the variance of which is

$$\text{var}(\log \bar{R}) = \text{var}(\log p) + \text{var}(\log \alpha) + 2 \text{cov}(\log p \log \alpha)$$

(77%)	(2%)	(21%)	GATE 1
(78%)	(2%)	(20%)	GATE 2

The contributions by each term are given in parenthesis for GATE 1 and 2.

It can be seen that the variance of $\log p$ dominates the variance in $\log \bar{R}$.

7. Discussions and Conclusions

It is demonstrated that the mixed distribution model provides a good estimate of time mean areal average rainfall at least for GATE type situations. The advantage of this model is its simplicity. Once the rainrates are sampled, the parameters can readily be estimated.

The mixed distribution approach suggests a retrieval algorithm for the estimation of monthly rainfall from satellites. If a functional relation exists between rainfall rates and radiance measurements, such as that proposed by Wilheit et al (1977), one would then accumulate the radiance measurements and compute histograms of radiance for the month. The histogram in radiance is then transformed into rainfall rates by the radiance-rainfall rate relation. The parameters of the lognormal distribution of the resultant rainfall rate histogram is then estimated to get the mean and variance. Consideration must be given to other factors such as beam filling and the variation of pixel size as a function of beam position.

To mimic satellite and rain gauge network sampling, various sampling designs have been devised. The sampling errors are about 10% for sampling by a polar orbiting satellite ((48,1,1) design). The sampling errors are reduced to about 5% for a satellite observation at low inclination ((40,1,1) design). McConnell & North (1987, this issue) examine sampling errors for four rainrate categories which contribute about equally to the total rainfall for sampling every 60 minutes of the same data. They found that the sampling errors in each of the rainrate categories are about 10%. If the categories are independent, the error in the total is reduced by $\sqrt{4}$, which is consistent with the 5% error found in this study.

We found a very strong correlation between the average rainfall rate and the fractional rain area in the GATE area. Chiu & Kedem (1986) had examined the usefulness of the fractional area with rainrates in excess of 20 mm/hr to estimate total rain volume. The high cutoff mimics the cloud index of Arkin (1979) to delineate fractional high cloud area. They found that the fractional light rain (rainrates greater than 1 mm/hr) area gives a better model than that which uses the fractional heavy rain area (rainrates greater than 20 mm/hr) alone. But when the two variables are used together, a much better model is obtained.

Jackson (1986) found that the monthly total rainfall in some tropical stations is strongly related to the number of raindays but bears little relation to the average daily intensity. A fair amount of skill has been achieved in the prediction of rain amount by the rain area as depicted in satellite visible and infrared imageries (Lovejoy & Austin 1979). Radar meteorologists have also found that the so called "Area Time Integral" (ATI) is a useful indicator of rain volume (Lopez 1982, Doneaud et al. 1982a). Doneaud et al (1982b) have applied the idea to rain gauge measurements. They also found that the percent of time when it rains is significantly related to the total rainfall. These are consistent with our findings of the importance of the parameter p . If we consider a design which samples all pixels in time and space, the estimated p is equivalent to the ATI. The improvement over the ATI technique by the mixed distribution would be derived from a knowledge of the distribution of the rainrates α conditional on rain. It provides

an estimate of the average rainrate intensity which replaces the climatological average often used in rain total estimates.

The importance of the fractional rain area in rainfall estimation has strong implications on satellite rainfall monitoring. Because of the absorption properties of raindrops, microwave sensors can clearly distinguish between rainy and non rainy areas. This special feature points to the need of microwave sensors, either active or passive, in the remote sensing of rain. These measurements, when used in conjunction with measurements from geostationary satellites such as GOES, can provide accurate monthly mean rainfall measurements.

Perhaps the most important conclusion that we can draw from this work is that, to the extent that the GATE data are representative of oceanic rainfall in the tropics, revisiting an area of roughly the GATE dimension (350 by 350 km²) at a repetition rate of about once every 10 to 12 hours provides an excellent estimate (of the order of 5 to 10% sampling error) for the area average three week mean rainrate for the region. This is within the capability of a single space platform with scanning sensors in a low inclination (tropical) orbit. This result is in good agreement with the work of Laughlin (1981) who used a rather different (Markov process) approach but based also upon the same GATE data.

Appendix: Remarks on the use of χ^2 for dependent data

There is ample evidence that the rainrate is lognormally distributed as illustrated by the small values in χ^2 and the excellent fit. When the χ^2 value is large (even though the estimated parameters obtained from the minimum chi square estimation are very similar) it is usually associated with sampling designs that sample the rainrate at points in time or space that are close to each other. This may not mean that the fit to the lognormal distribution is not good, but may suggest dependence in the sample. This can be understood as follows.

Let $p_i = o_i/N$, with $i=1, \dots, 9$, and let $p_i = E(p_i) = e_i/N$, where $E(x)$ is the expected value of x . Define the vectors $\underline{p} = (p_1, \dots, p_8)'$, $\hat{\underline{p}} = (\hat{p}_1, \dots, \hat{p}_8)'$ and $\underline{1} = (1, \dots, 1)'$ and put

$$\underline{A} = \text{diag}(1/p_1, \dots, 1/p_8) + \underline{1} \underline{1}' 1/p_9$$

then we have

$$\chi^2 = \sum_{i=1}^9 (o_i - e_i)^2 / e_i = N (\underline{p} - \hat{\underline{p}})' \underline{A} (\underline{p} - \hat{\underline{p}}) \quad (*)$$

Assuming that the rainrates satisfies some dependence condition (e.g. finite-dependence as discussed by Anderson 1971, p427) so that for large N , $\sqrt{N}(\underline{p} - \hat{\underline{p}})$ converges to a normal distribution,

$$\sqrt{N} (\underline{p} - \hat{\underline{p}}) \xrightarrow{L} N(\underline{0}, \underline{V})$$

then for sufficiently large N

$$\chi^2 \approx \sum_{i=1}^8 \lambda_i z_i^2$$

where the z_i^2 are independent $\chi^2(1)$ variables. If the sampled rainrates are independent, $\lambda_i = 1$ for all i and χ^2 is distributed as a chi-square variable with 8 degrees of freedom. But if the sampled rainrates are dependent, $\lambda^2 \neq 1$ and (*) can be inflated or deflated since its

asymptotic expected value is $\sum_{i=1}^8 \lambda_i$. The practical outcome emerging from this discussion is that large values of (*) may indicate dependence despite a possible perfect fit. As the rainrates are sampled further apart in time and space, they become reasonably independent and the distribution of (*) is close to a chi square distribution with 8 degrees of freedom adjusted for the number of unknown parameters. See also Kedem and Slud (1981) who discuss a similar quadratic form whose values are inflated due to dependence of the data.

References:

- Aitchison, J. & J. A. C. Brown, 1963: The Lognormal Distribution, Cambridge Univ. Press, London.
- Anderson, T. W., 1971: The Statistical Analysis of Time Series, Wiley, 704 pp.
- Arkell, R. & M. Hudlow, 1977: GATE International Meteorological Radar Atlas, US dept of Commerce, NOAA, US government printing office, Washington, D.C., 20402.
- Arkin, P. A., 1979: "The Relationship between Fractional coverage of High Cloud and Rainfall Accumulation during GATE over the B Scale", Month. Weath. Rev., 107, 1382-1387.
- Austin P. M. & S. G. Geotis, 1980: "Precipitation Measurements over the Oceans", Air Sea Interaction, eds. F. Dobson, L. Hasse & R. Davis, Plenum pub. corp., 227 West 17th Street, N.Y., New York, 10011.
- Austin P. M. & S. G. Geotis, 1978: "Evaluation of the Quality of Precipitation Data from a Satellite-borne Radiometer", Final report under NASA grant NSG 5024, Massachusetts Institute of Technology, Cambridge, MA 02139, 30pp.
- Atlas, D. & O. Thiele, 1981: Precipitation Measurements from Space, Workshop Report, Oct, 1981, NASA, Goddard Space Flight Center, Greenbelt, Maryland, 20771.
- Barrett E. C. & D. M. Martin, 1981: The Use of Satellite Data in Rainfall Monitoring, Academic Press, New York, NY, 340 pp.
- Berkson, J., 1980: "Minimum Chi-Square, not maximum likelihood", Annals of Statistics, 81, 457-469.
- Bras R. L. & R Colon, 1978: "Time-Averaged Areal Mean of Precipitation: Estimation and Network Design", Water Resour. Res., 14, no 5, 878-888.

- Bras R. L. & I. Rodriguez-Itrurbe, 1976: "Network Design for the Estimation of Areal Mean of Rainfall Events", Water Resour. Res., 12(6), 1185
- Chiu L. S. & B. Kedem, 1986: "Satellite Rainfall Retrieval by Logistic Regression", 2nd Conf. on Satellite Meteorology, Williamsburg, VA, American Meteorological Society, Boston, MA.
- Creutin J. D. & C. Obled, 1982: "Objective Analysis and Mapping Techniques for Rainfall Fields: An Objective Comparison", Water Resour. Res., 6, 117-134.
- Doneaud A.A., S. Ionescu-Niscov, D. L. Prignitz & P. L. Smith, 1982: "The Area-time Integral as an indicator for convective rain volume", 21st. Conf. on Radar Meteorology, Edmonton, Calgary, Canada, Amer. Meteoro. Soc., Boston , MA.
- Doneaud A. A., P. L. Smith, A. S. Dennis & S. Sengupta, 1981: "A Simple Method for Estimating Convective Rain Volume over an Area", Water Resour. Res., 17,1676-1682.
- Drufuca G. & R.R. Rogers, 1978: "Statistics of Rainfall over Paths form 1 to 50 km", Atmospheric Environment, 12, 2333-2342.
- Eagleson, P. S., 1967: "Optimal Density of Rainfall Networks", Water Resour. Res., 3, 1021-1033.
- Feuerverger, A., 1979: "On some methods of analysis for weather experiments", Biometrika, 66, 655-658.
- Freeny A. E. & J. D. Gabbe, 1969: "A Statistical Description of Intense Rainfall", Bell System Technical Journal, 48, 1789-1851.
- Gill, A. E., 1982: Atmosphere-Ocean Dynamics, Academic Press, NY, 622pp.
- Horel, J.D. & J. M. Wallace, 1981: "Planetary-Scale Atmospheric phenomena Associated with the Southern Oscillation", Month. Weath. Rev., 109, 813-829.

- Houze R. A. & A. K. Betts, 1981: "Convection in GATE", Rev. Geophys & Space Phy., 19, no 4, 541-576.
- Hogg, R. V. & E. A. Tanis, 1977: Probability and Statistical Inferences, MacMillan, New York, NY, 10022.
- Houze R. A. & C-P Cheng, 1977: "Radar Characteristics of Tropical convection observed during GATE: Mean Properties and Trends over the Summer Season", Monthly Weather Review, 105, 964-980.
- Jackson, I. J., 1986: "Relationship between Raindays, Mean Daily Intensity and Monthly Rainfall in the Tropics", J. Climatology, 6, 117-134.
- Johnson, N. C. & S. Kotz: Continuous Univariate Distributions, vol 1, Wiley, New York, New York.
- Jones M. A. & A. L. Sims, 1978: "Climatology of instantaneous Rainfall Rates", J. Applied Climatology, 17, 1135-1140.
- Kedem, B. & E. Slud, 1981: "On goodness of fit of time series models: An Application of Higher Order Crossings", Biometrika, 68, 551-556.
- Lin, S.H., 1976: "Rain-Rate Distribution and Extreme-Value Statistics" B. S. T. J., 55 (8), 1111-1124.
- Lopez R. E., 1977: "The Lognormal Distribution and Cummulus cloud populations", Mon. Weath. Rew., 105, 865-872.
- Laughlin, C., 1981: On the effect of temporal sampling on the observation of mean rainfall, Precipitation Measurements from Space, Atlas D. & O. Thiele, Editors, NASA/GSFC, Greenbelt, Maryland, 20771.
- Lorenz, E. N., 1967: The Nature and Theory of the General Circulation of the Atmsphere, WMO, Geneva, 161pp.
- Lovejoy S., 1981: "The Remote Sensing of Rain", Ph. D. Thesis, Dept. of Physics, McGill Univ., Montreal, Canada, 102pp.

- Lovejoy S. & G. L. Austin, 1980: "The Estimation of Rain from Satellite-borne Microwave Radiometers", Quart. J. R. Met. Soc., 106, 255-276.
- Lovejoy S. & G. L. Austin, 1979: "The Delineation of Rain Area from Visible and IR Satellite Data for GATE and Mid-latitudes", Atmosphere & Oceans, 17(1), 77-92.
- Moses, M. E., 1982: "Concepts and Techniques in Hydrological network Design", Operational hydrology report no. 19, WMO-no 580, World
- Okamoto, K., 1982: "Remote Sensing of Precipitation by Weather Radar system at Space Station", 21st. Conf. on Radar Meteorol., Edmonton, Calgary, Canada, 263-269, American Meteorology Society.
- Patterson, V. L., M. D. Hudlow, P. J. Pytlowany, F. P. Richards & J. D. Hoff, 1979: GATE Radar Rainfall Processing System, NOAA Technical Memorandum, EDIS 26, NOAA, Washington, D.C.
- Rodriguez-Iturbe, I. & J. M. Mejia, 1974: "The design of Rainfall Networks in Time and Space", Water Resour. Res., 10(4), 713.
- Rogers R. R., 1976: "Statistical Rainstorm Models: Their Theoretical and Physical Foundations", IEEE Trans. on Antennas and Propagation, July, 547-566.
- Sims A.L. & D. M. A. Jones, 1975: "Frequencies of Short-Period Rainfall Rates along Lines", J. Applied Meteorol., 14, 170-174.
- Theon, J., G. North, T. Wilheit & O. Thiele, 1986: "Tropical Rainfall Monitoring Mission" workshop report, Calverton, Maryland, Nov., 1985., NASA (in preparation).
- Wilheit, T. T., A. T. C. Chang, M. S. V. Rao, E. B. Rodgers, J. S. Theon, 1977: "A satellite technique for Quantitatively Mapping Rainfall over the Oceans", J. Appl. Meteorol., 16, 551-560.

Table 1. Results for (8,8,8) sampling

class	o_j	e_j	
1-2	453	450	
2-4	590	598	
4-6	325	324	
6-8	207	188	
8-10	116	116	
10-12	60	76	
12-16	82	88	
16-20	52	46	
>20	80	79	
total	1965	1965	$\chi^2=6.74$

$$\mu = 1.14, \quad \sigma^2=1.047$$

$$\alpha = 5.28, \quad \beta^2=51.5$$

Table 2: Estimated means, minimum χ^2 and fraction of rain for different designs.

GATE 1

n, (k,1)	(4,4)	(6,6)	(8,8)	(10,10)
2	.44 (11.1) 8.3	.44 (15.3) 8.2	.45 (4.2) 8.2	.46 (2.4) 8.7
4	.44 (3.7) 8.3	.44 (13.9) 8.1	.45 (3.8) 8.4	.46 (9.4) 8.4
6	.45 (7.9) 8.3	.43 (2.9) 8.1	.45 (3.8) 8.2	.46 (2.1) 8.9
8	.44 (2.7) 8.3	.44 (4.9) 8.0	.44 (6.7) 8.3	.44 (7.9) 8.3
10	.45 (4.3) 8.3	.45 (4.5) 8.2	.45 (5.2) 8.3	.43 (3.4) 8.1

GATE 2

2	.37 (59.9) 6.8	.36 (19.8) 6.8	.37 (36.9) 6.9	.36 (10.3) 6.9
4	.37 (50.4) 6.9	.36 (8.4) 6.9	.36 (23.8) 7.1	.35 (9.7) 7.0
6	.38 (27.6) 7.0	.37 (12.1) 6.9	.39 (16.9) 7.1	.36 (18.1) 7.0
8	.36 (23.8) 6.8	.34 (4.9) 6.9	.35 (17.4) 7.0	.37 (9.3) 7.2
10	.38 (19.3) 7.1	.37 (8.5) 7.0	.39 (7.7) 7.2	.37 (5.4) 7.2

Estimated rainrate in mm/hr on top line. Minimum chi square value in parentheses. The estimated rain probability, p, appeared in the lower right hand corner, in percent.

Table 3. Comparison between gamma and lognormal distribution fit to various designs

design	n*	lognormal			Gamma		
		μ	σ^2	χ^2	α	λ	χ^2
(30,10,10)	333	1.00	1.16	6.04	0.29	0.12	5.57
(20,10,10)	456	1.10	1.07	7.76	0.37	0.13	12.14
(10,10,10)	972	1.09	1.18	3.39	0.30	0.10	13.73
(5,10,10)	1976	1.06	1.21	6.80	0.34	0.10	32.46
(10, 5, 5)	3936	1.12	1.12	8.77	0.35	0.12	44.24
(5, 5, 5)	7889	1.11	1.13	16.83	0.35	0.12	85.87
(10,20,20)	219	1.32	1.00	4.98	0.49	0.12	12.39
(5,30,30)	263	1.09	1.41	6.53	0.26	0.09	4.09
(5,20,20)	461	1.19	1.07	0.80	0.41	0.12	7.21

n* is the number of raining pixels

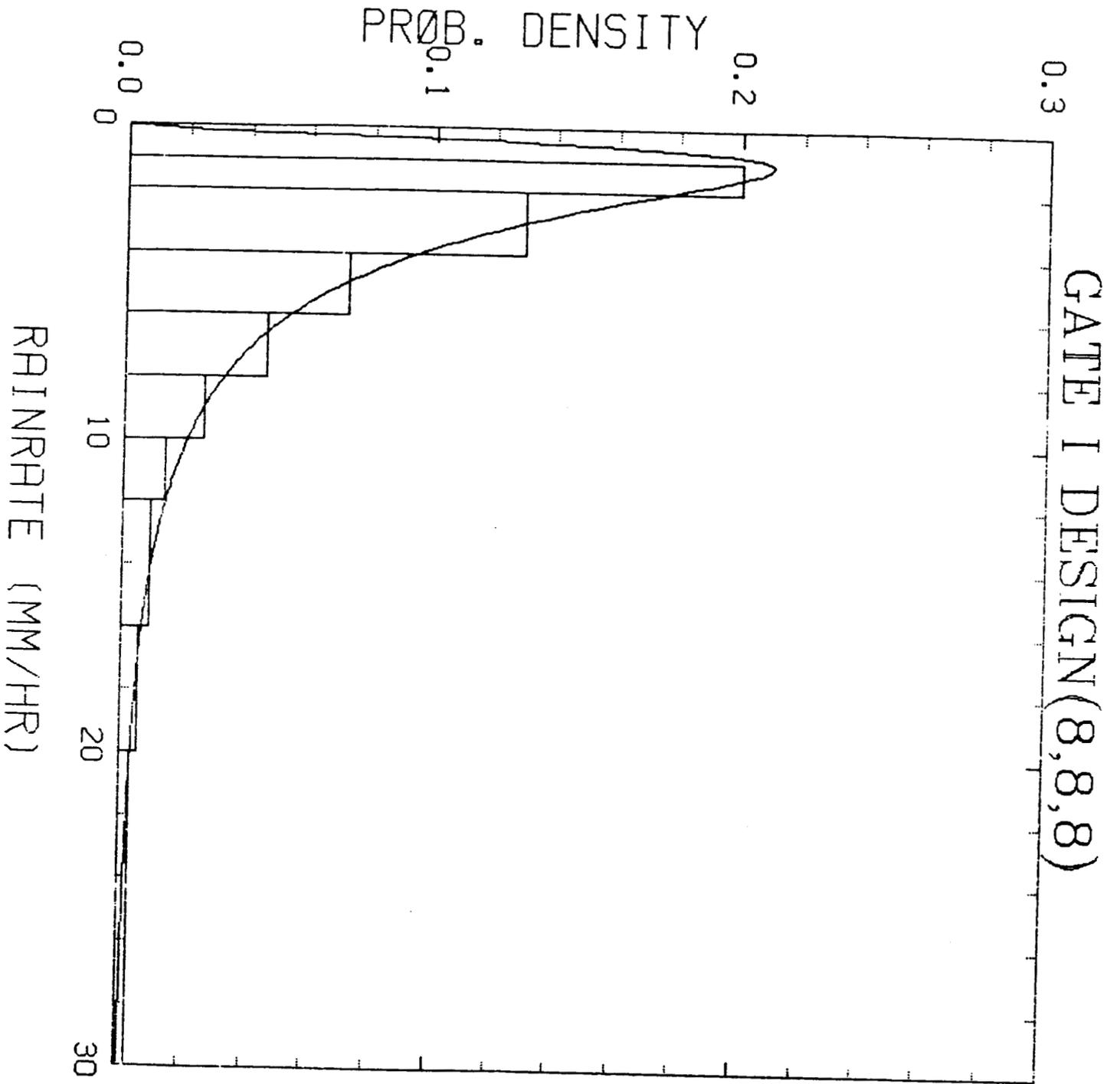
Table 4. Comparisons of Estimates from sampling designs for GATE 1 and 2.

design	GATE I					GATE II				
	n*	p	<R>	st.err	χ^2	n*	p	<R>	st.err	χ^2
(24,1,1)	42237	.083	.448	.0079	27.4	30111	.069	.364	.0083	148.1
(48,1,1)	22976	.088	.514	.0119	48.1	14156	.066	.317	.0107	77.4
(72,1,1)	14533	.086	.457	.0135	20.9	8826	.061	.316	.0141	77.2
(96,1,1)	11622	.089	.572	.0187	22.5	6409	.058	.282	.0151	46.7

n* number of raining pixels

Figures

1. Histogram of rainfall rate sampled from GATE by the design (8,8,8). The curve is a lognormal fit to the histogram with parameters $\mu = 1.14$ and $\sigma^2 = 1.05$ which are estimated by the method of minimum chi square.
2. Histograms of the estimated means from sampling designs of (48,1,1), (40,1,1) and (1,10,10) (left, middle and right column respectively) for GATE 1 (upper) and GATE 2 (lower). The total number of samples are 48,40 and 100 for the three designs. The means and standard deviations are included in the upper right hand corners.
3. Scatter diagram of the logarithm of the average rainfall rate (\bar{R}) and fractional rain area in the GATE area for GATE 1 (upper) and GATE 2 (lower). A cutoff value of 1 mm/hr is used to distinguish between rainy and dry pixels. The correlation coefficients are indicated on the upper left hand corners.
4. Scatter diagram of the logarithm of average intensity of the rainy pixels (α) and fractional rain area (p) for GATE 1 (upper) and GATE 2 (lower). The correlation coefficients are indicated on the upper left hand corners.
5. Histograms of the average rainfall rate (\bar{R}), fractional area (p) over the GATE area and average intensity of the rainy pixels (α) (left, middle and right columns) for GATE 1 (upper) and GATE 2 (lower). The means and standard deviations are indicated on the upper right hand corners. The numbers below the means and standard deviations on the histograms of \bar{R} indicate there are 1622 (1419) observations out of 1716 (1512) with $p \neq 0$ in GATE 1 (2).



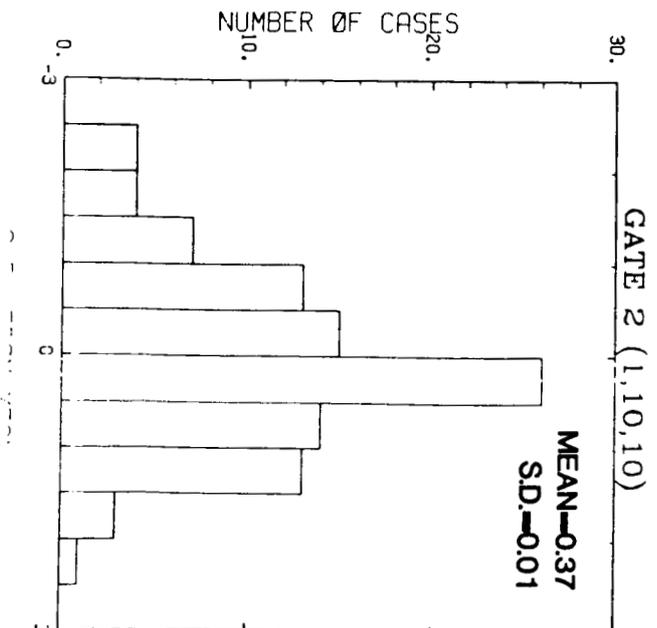
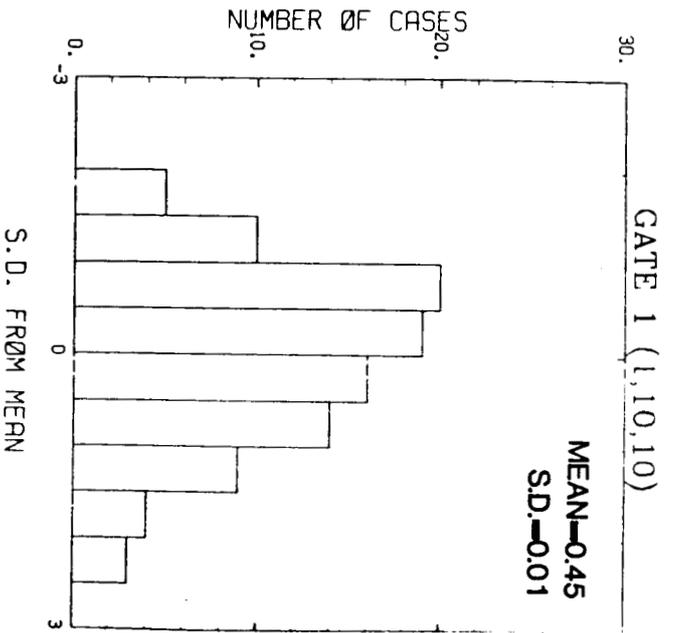
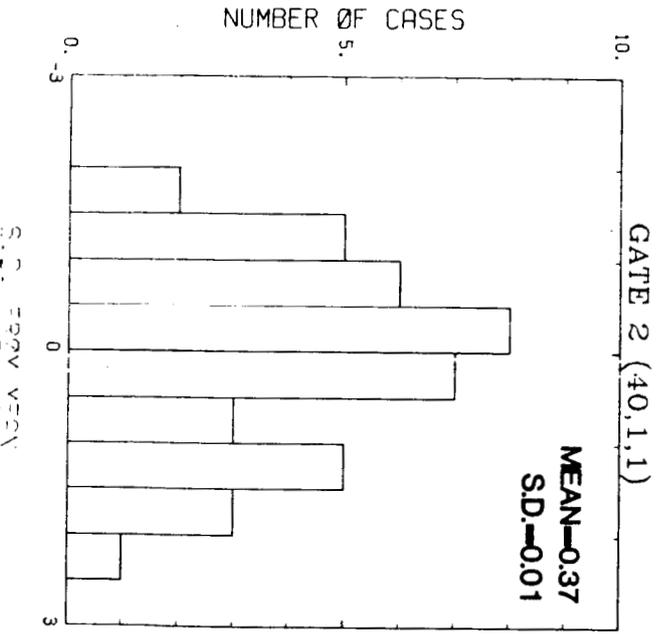
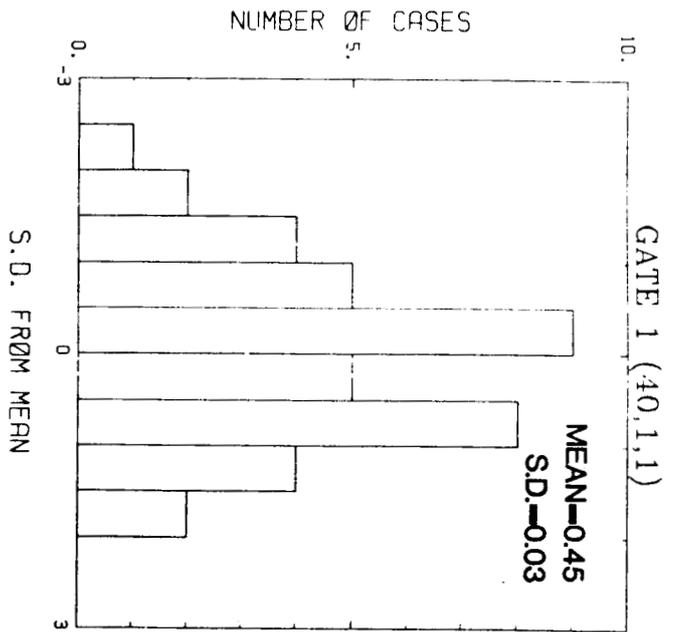
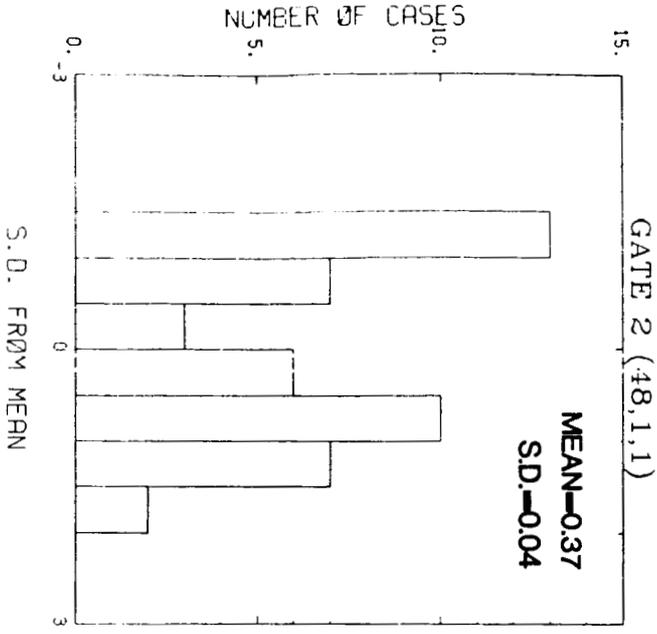
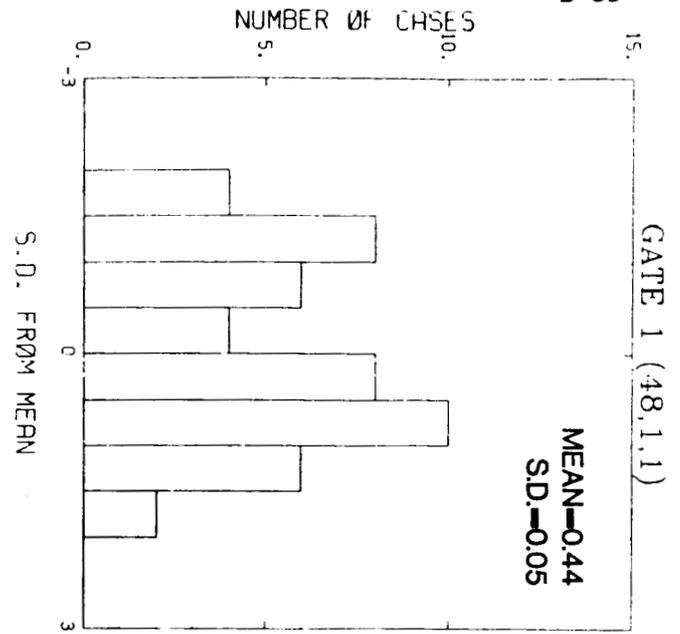


fig 2

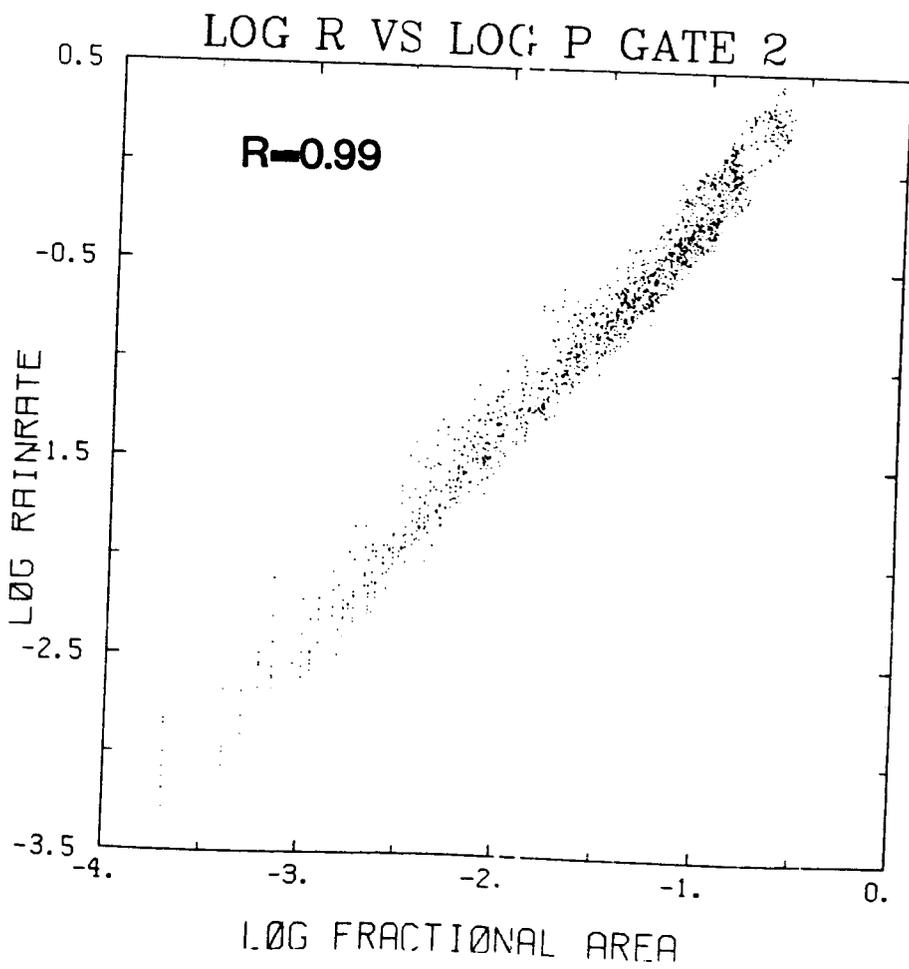
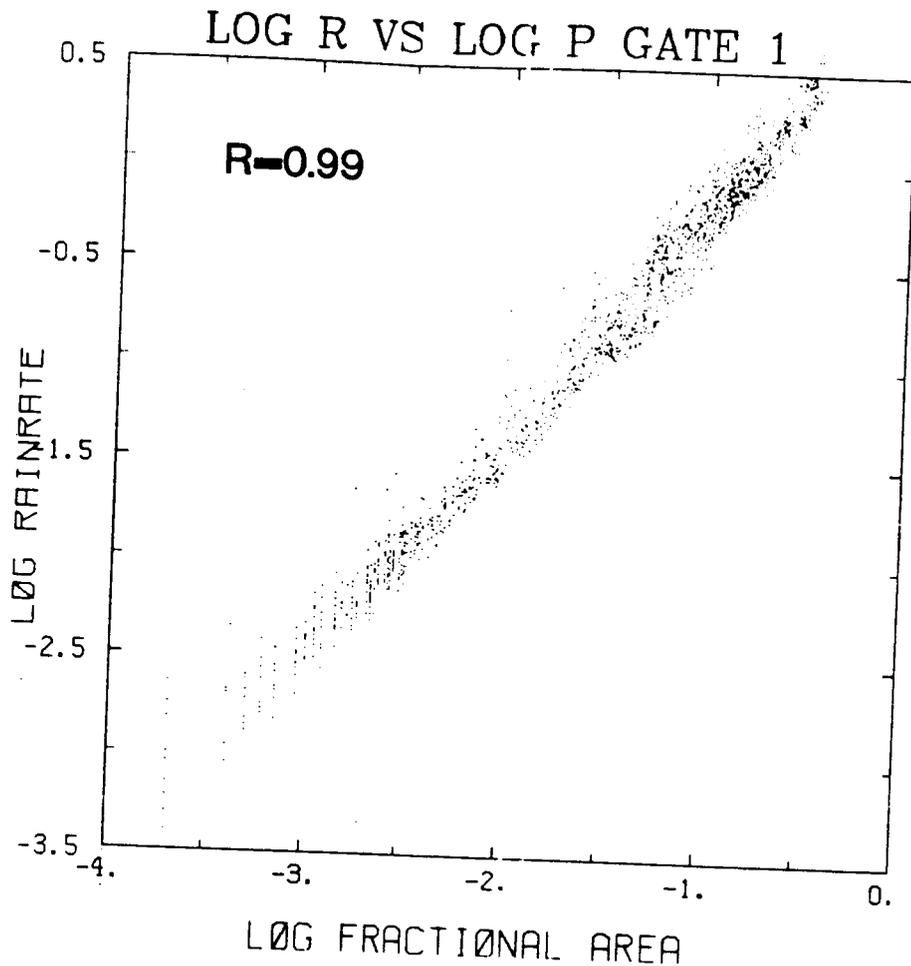
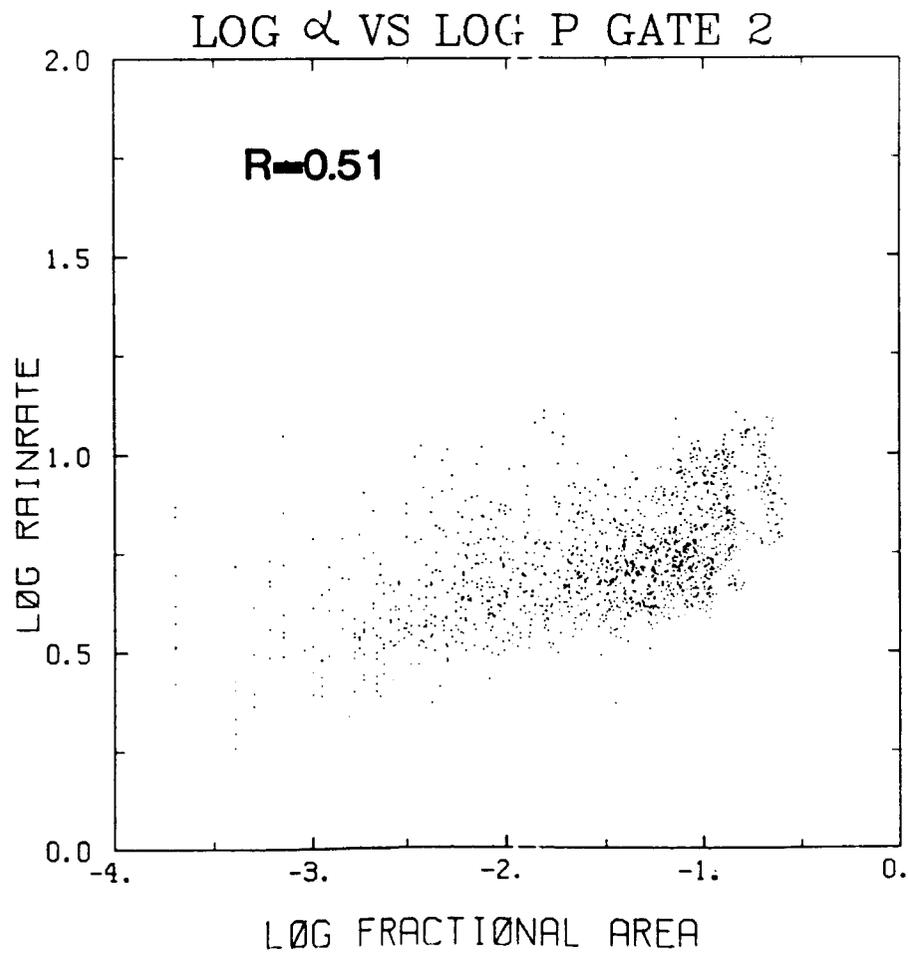
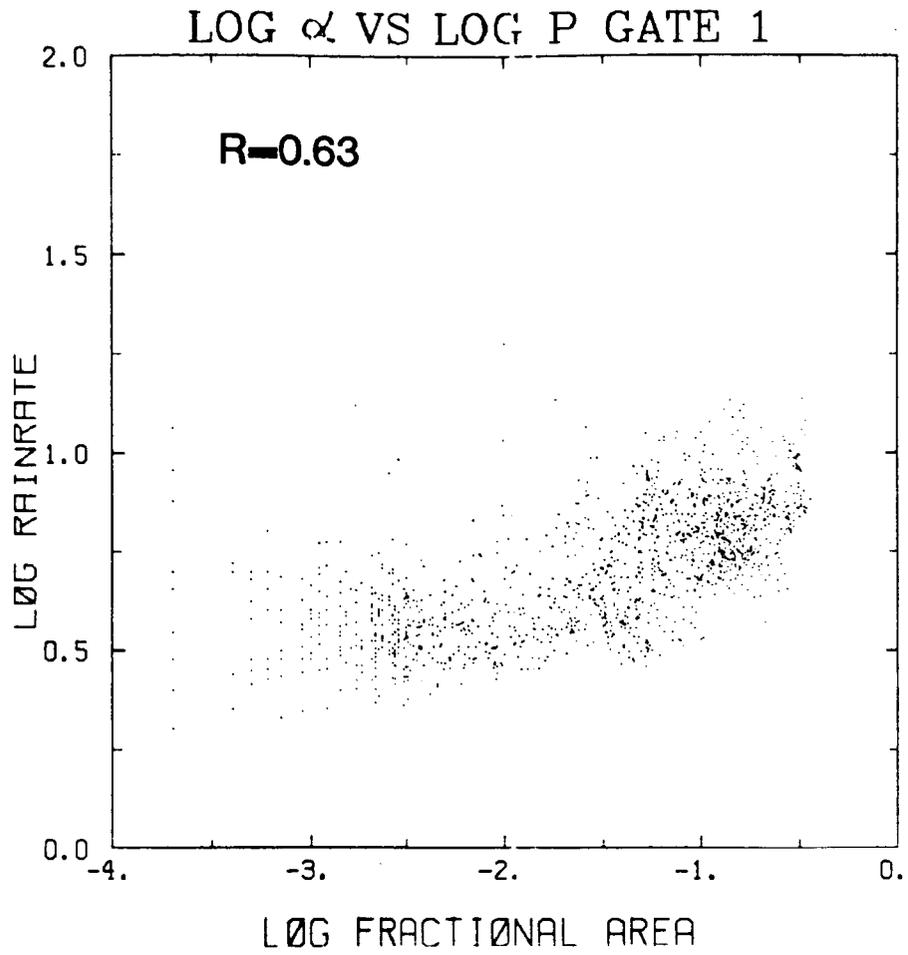
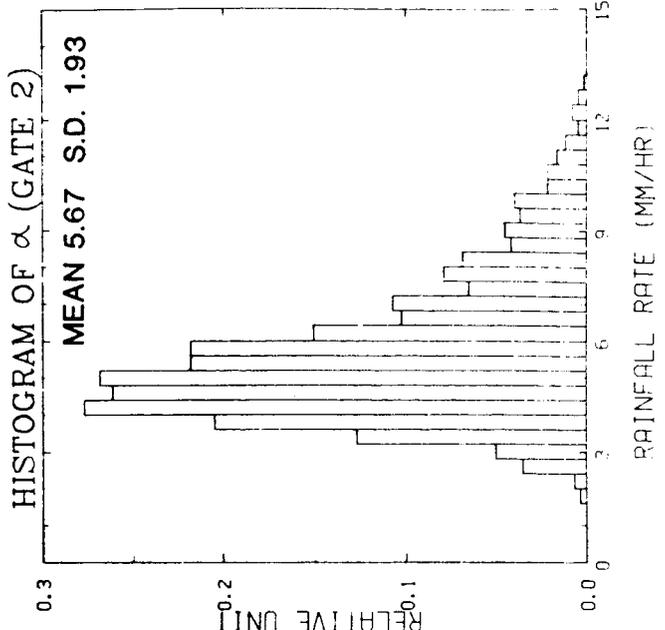
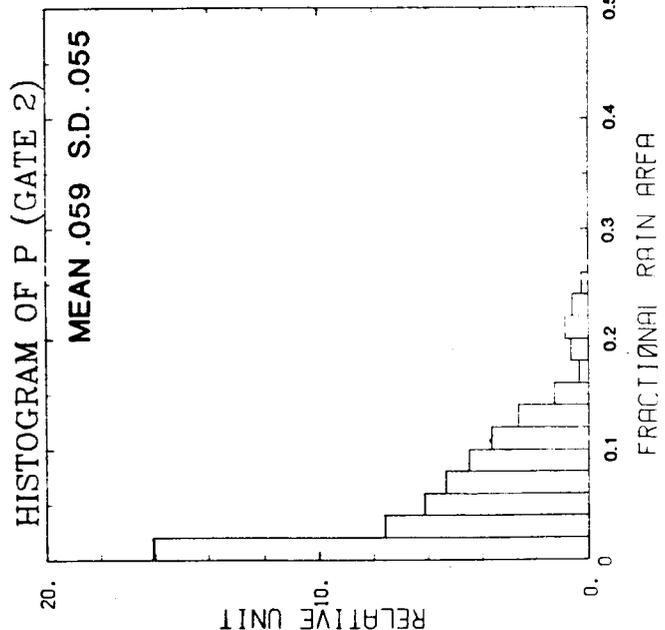
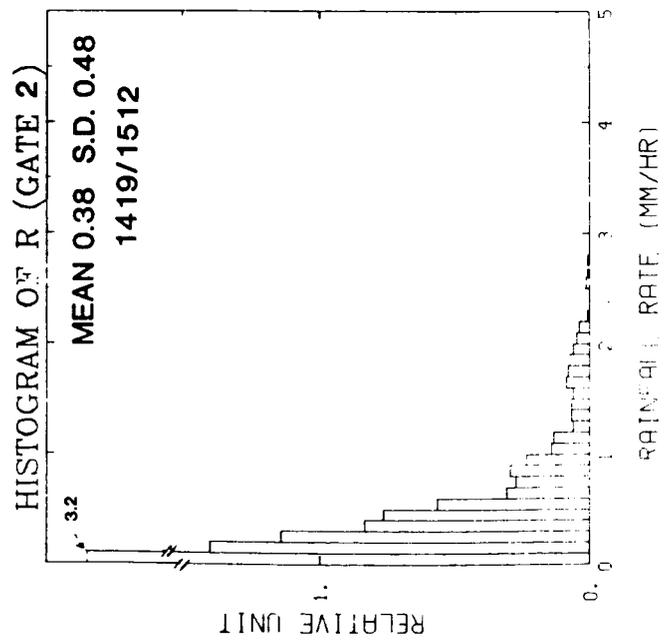
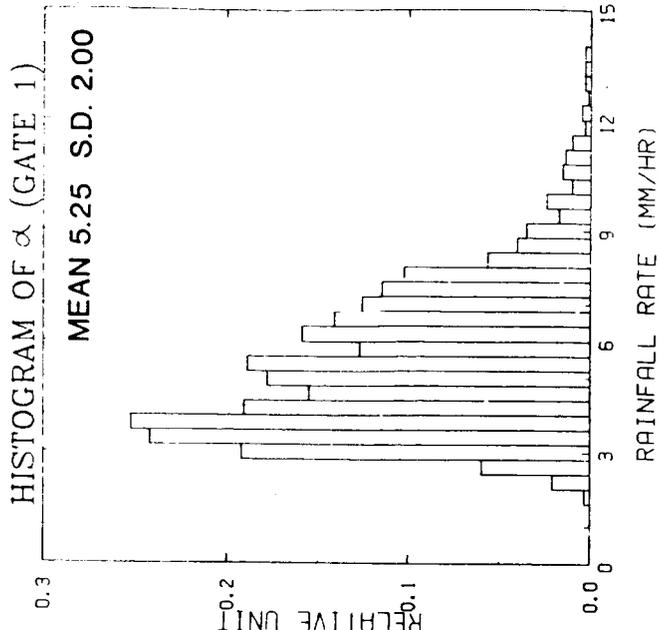
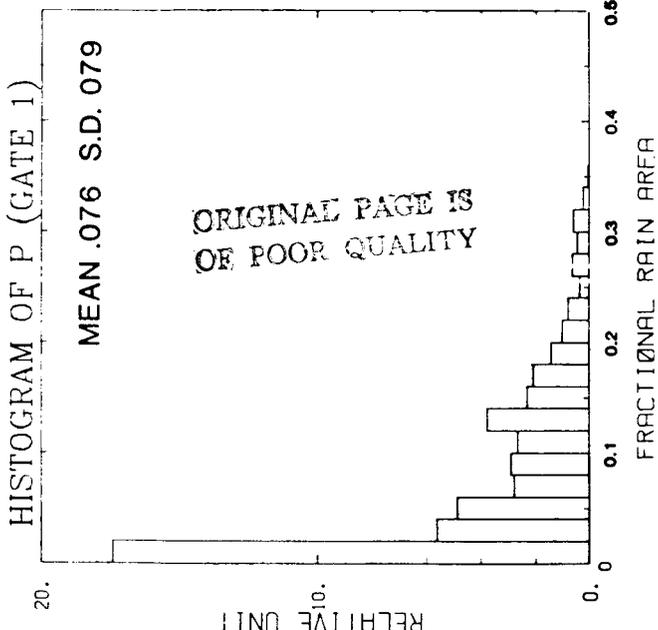
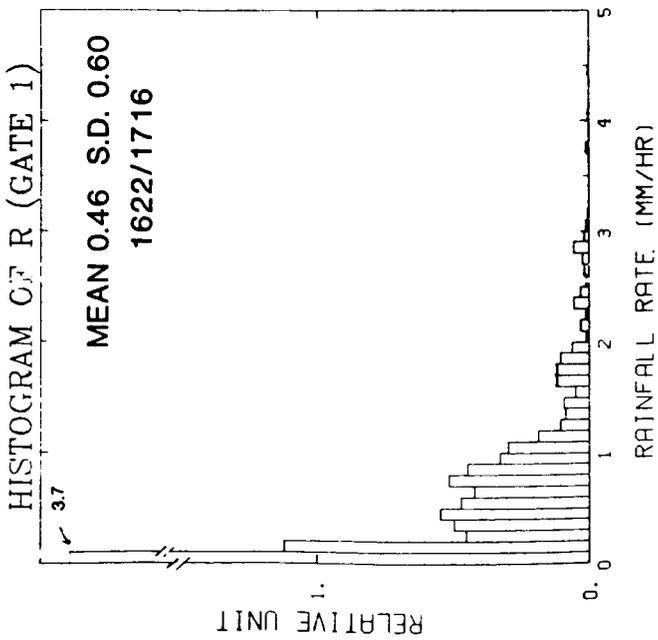


fig 3





Attachment C: A Rainfield Simulation Model

A. Simulation of Rain Field Snapshots

In the absence of real rainrate data, it is useful to generate artificial data by stochastic models which preserve certain specified statistics. Also, such a model is very helpful in assessing the outcomes of controlled experiments. We have developed such a model and intend to use it in the next phase. A source FORTRAN program is attached.

A.1 A Stochastic Rain Field Model

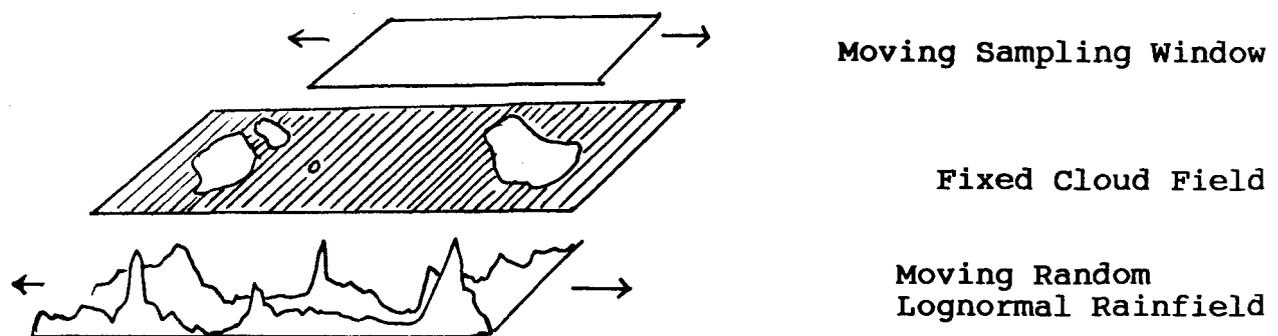
In what follows, we describe a simulation model which generates artificial "radar" snapshots of a rain field.

Our model is made of three parts, one of which is fixed while the others move in relation to the fixed part. The three parts are (See figure A1.):

- (a) Spatial random rainfield (moving);
- (b) Cloud field (fixed); and
- (c) Moving window (moving).

This is a very flexible model which can accommodate any kind of cloud and rain fields.

Figure A1



A.2 Random Rain Field

This field consists of a spatial moving average with specified distribution for its rainrate (in this case lognormal) and specified spatial correlation. This is the bottom part and should be thought of as an infinite random field which is being constantly shifted. For

example, we can use a field of the form

$$R(i, j) = \exp[y(i, j) + 1.140]$$

where

$$Y(i, j) = E(i, j) + 0.1084[E(i - 1, j) + E(i + 1, j) + E(i, j - 1) + E(i, j + 1)], \quad i, j = 0, \pm 1, \pm 2, \dots,$$

where $E(i, j)$ is white Gaussian noise. In this case, $R(i, j)$ has a lognormal distribution

$$\Lambda(\mu_R, \sigma^2_R)$$

with parameters

$$\mu_R = 1.14$$

and

$$\sigma^2_R = 1.047$$

The coefficient 0.1084 is needed for stationarity requirements. We can easily change this model to suit any correlation requirement.

A.3 Cloud Field

The cloud field covers a certain large area (e.g., GATE area) and consists of clouds whose areas are very close to being lognormally distributed. It is a fixed field located above the rainfield. The "clouds" are to be thought of as "holes" or "windows" through which we see rain. At a given time constant, what we see through a given cloud is precisely its content. This content keeps changing since the rainfield is moving.

Here is how a cloud is generated in a field of area 10^4 pixels. Consider, for example, an interval at length 100 from which a point is selected at random. From that point, we measure a random length whose distribution is lognormal with parameters μ, σ^2 . Let X be the part of this length which overlays with the interval $(0, 100)$. Then, by properly conditioning X , we have

$$E(X) = [100 \exp(\mu + \frac{1}{2} \sigma^2) - \exp(2\mu + 2\sigma^2)]/200$$

The square of this quantity (by independence) can be thought of as the average size of a "random cloud." Let

$M \equiv$ number of clouds

Then the fractional rainy area over a field of area 10^4 is given by

$$\frac{M \times E^2(X)}{10^4}$$

Thus, we can control the probability of rain by μ , σ^2 , and M . The table below illustrates this fact.

μ	σ^2	EX	M	Probability of Rain
1	1	4.208	50	0.088
1	1	4.208	40	0.071
1	0.5	3.3899	70	0.08
1	2	5.3719	28	0.08

The probability of rain is fixed over the cloud field but can obviously change for subfields.

The truncation at, say, 100 is needed for the rainy area under study in real life is usually an arbitrary area taken from a much larger area by truncation.

A.4 The Sampling Window

The third part is a moving window which moves at random over the cloud field. Each time the context of the window is observed, we call it a snapshot.

Figure A2 shows a typical snapshot with a sampling window of 20 x 20 pixels. The zeroes denote the no rain areas, and the rainrate are given in mm/hr.

A.5 Comparison with Laughlin's Results

To estimate the error in satellite sampling, Laughlin (1982) computed the temporal autocorrelation function as a function of average areas using the GATE data. Since our model parameters are constrained by the GATE observations, a calculation similar to Laughlin was carried out and the results presented on figure A3. The results are very similar to those of Laughlin, as anticipated.

.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	1.1	2.8	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	2.5	3.6	4.5	.0	.0	.0	.0
.0	.0	3.0	.8	6.4	2.8	1.9	2.5	.0	.0	.0	.0	.0	.0	2.8	5.3	9.7	.0	.0	.0	.0
.0	.0	2.0	.2	5.6	2.9	1.5	2.4	2.1	6.8	2.2	8.9	7.4	18.1	9.1	17.8	.0	.0	.0	.0	.0
.0	.0	1.4	1.5	6.8	18.0	12.1	1.5	6.0	1.6	.0	.0	.0	.0	1.0	.0	.0	.0	.0	.0	.0
.0	.0	.037	.119	.4	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.0	.0	5.3	16.2	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
3.0	1.4	1.0	5.7	.7	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
.9	.7	.92	1.3	4.5	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
5.1	2.1	.6	11.1	.9	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
4.9	8.7	1.1	.6	12.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
5.0	5.1	2.2	3.4	3.5	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
10.2	2.0	.4	3.3	.8	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0

Fig. A2. A snapshot generated from the simulation model with $\mu = 1.14$ and $\sigma^2 = 1.047$. These values are derived from GATE rainfall measurements.

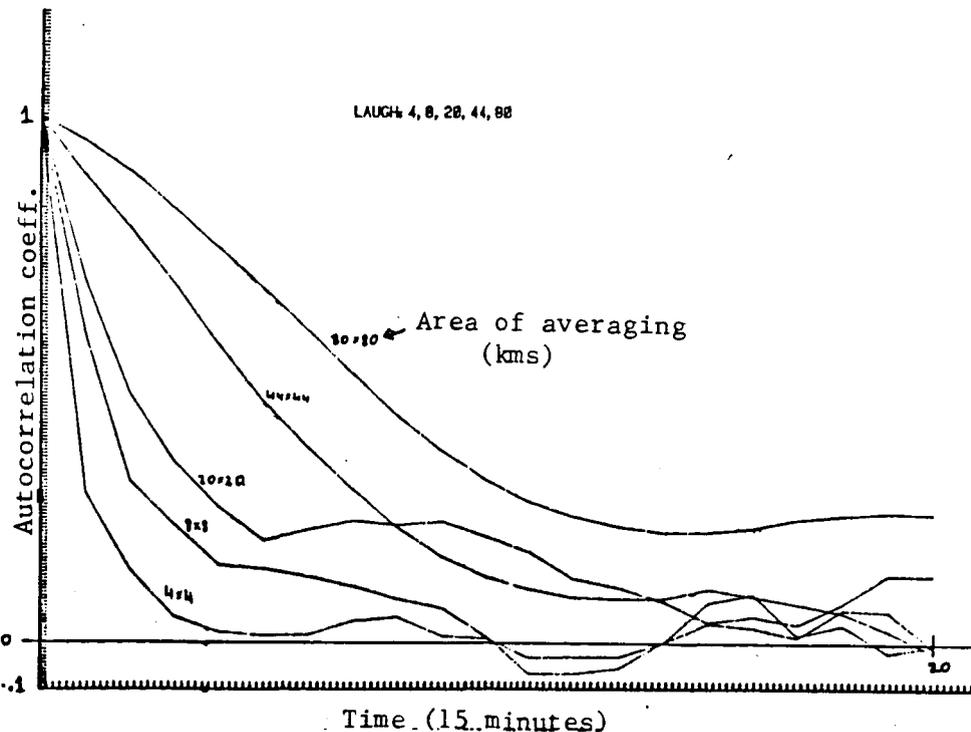


Fig. A3. Temporal autocorrelation functions for different averaging areas. The size of the averaging boxes are indicated on the graph (in kms).

Program listing to generate Rain fields from the Simulation Model

```

C
C THIS PROGRAM CALCULATES RAINFALL FIELDS FROM BEN KEDEM'S TIME-
C DEPENDENT RAINFALL RATE MODEL AND WRITES THEM TO TAPE FOR FURTHER
C ANALYSIS.
C
C INPUT : DSEED0 - ARRAY OF SEEDS FOR THE RANDOM NUMBER GENERATOR FOR
C           CALCULATING RAIN FIELDS
C           DELTAT - TIME IN HOURS OF EACH TIME STEP (USEFUL RANGE .25
C                 TO 24.)
C           NCDAYS - NUMBER OF DAYS FOR WHICH RAIN FIELDS ARE TO BE
C                 CALCULATED
C           NSTEPS - NUMBER OF TIME STEPS FOR WHICH RAIN FIELDS ARE
C                 TO BE CALCULATED
C           NCALLS - NUMBER OF CALL TO BE MADE TO THE RAINFALL CALCULATING
C                 PROGRAM
C           N - NUMBER OF POINTS IN EACH DIMENSION OF THE RAINFALL ARRAY
C           RESIZE - SIZE OF THE SIDES OF PIXELS IN KM
C           GPHASE - GATE PHASE OF THE SIMULATION INPUT PARAMETERS
C           AVAREA - AVERAGE AREA OF CONTIGUOUS RAIN PATCHES
C           AVFTR - AVERAGE FRACTION OF A REALIZATION THAT HAS RAIN
C
C INTERNAL : I - LOOP AND ARRAY INDEX
C           SEED - ARRAY OF SEEDS FOR RANDOM NUMBER GENERATORS TO BE
C                 PASSED TO SUBROUTINES IN ORDER TO LEAVE THE DSEED0
C                 ARRAYS UNCHANGED
C           TIME - CUMULATIVE TIME OF THE REALIZATION
C
C OUTPUT : RR - TWO-DIMENSIONAL ARRAY OF RAINFALL RATES
C
C SUBROUTINES : EKRAIN - CALCULATES RAINFALL RATE ARRAYS AND WRITES
C                 THEM TO TAPE
C
C
C REAL*4 RR(128,128),RESIZE,RINC
C REAL*4 TIME,DELTAT,AVAREA,AVFTR,GPHASE
C REAL*8 DSEED0(8),SEED(4)
C INTEGER*4 N,NSTEPS,NDAYS,NCALLS,IDAY
C INTEGER*4 I,I1,J,L
C DATA NDAYS/2/,N/128/,NSTEPS/24/
C DATA AVFTR/.09/,DELTAT/1.00/
C DATA AVAREA/450./,RESIZE/4.0/,RINC/5.0/
C DATA DSEED0/3141592.D0,2314159.D0,9231415.D0,5923141.D0,
C 1592314.D0,4159231.D0,1415923.D0,1341592.D0/
C INITIALIZE VARIABLES
C   IDAY=0
C   TIME=0.0
C   NCALLS=NINT((FLOAT(NDAYS)/2.0)+0.4)
C   WRITE (6,9) NCALLS
C 9   FORMAT(' ', 'NCALLS= ',I3)
C   DO 11 J=1,4
C     SEED(J)=DSEED0(J)
C 11 CONTINUE
C LOOP OVER CALLS TO RAINFALL MODEL
C DO 20 I1=1,NCALLS
C   CALL EKRAIN(N,RESIZE,GPHASE,DAY,TIME,AVAREA,AVFTR,SEED,DELTAT,
C   CR,ICAY,ICS)
C   IF (ICS.NE.0) THEN
C     WRITE (6,99)
C 99   FORMAT(' ', 'I/C ERROR')
C     GO TO 100
C   END IF
C 20 CONTINUE
C CALL LEVPTC(N,RINC,RR)
C 100 STOP
C END
C
C SUBROUTINE EKRAIN(NUM,RESIZE,GPHASE,DAY,TIME,AVAREA,AVFTR,SEED,

```

ORIGINAL PAGE IS
OF POOR QUALITY

CDELTA,RR,ICAY,ICS)

C
C
C
C
C
C

BEN KEDEM, UNIVERSITY OF MD
AUG. 20, 1986

THIS PROGRAM GENERATES RAINFALL FIELDS WITH GATE-LIKE STATISTICAL
PROPERTIES.

REAL*4 LGN1(500),LGN2(500),U1(500),U2(500),Y(300,300),E(300,300)
REAL*4 RESIZE,FP(128,1:*),GPHASE,DAY,TIME,AVAREA,AVFTP,DELTA
REAL*4 XM,SIGMA,SIGSC,U(90000)
REAL*8 DS1,DS2,DS3,DS4,SEED(1:*)
INTEGER*4 IV(200,200),N,II,JJ,M,MMM,MINI,MINJ,MAXI,MAXJ,I1,J1
INTEGER*4 I11,J11,NSNP,NDAYS,ICS,NUM,IDAY
CHARACTER*7 FILELE(30)

C

CHARACTER CONSTANTS FOR FILE PARAMETER IN OPEN AND DD STATEMENTS
DATA FILELR/'FT11F01','FT11F02','FT11F03','FT11F04','FT11F05',
C'FT11F06','FT11F07','FT11F08','FT11F09','FT11F10','FT11F11',
C'FT11F12','FT11F13','FT11F14','FT11F15','FT11F16','FT11F17',
C'FT11F18','FT11F19','FT11F20','FT11F21','FT11F22','FT11F23',
C'FT11F24','FT11F25','FT11F26','FT11F27','FT11F28','FT11F29',
C'FT11F30'/

C

SPECIFY PARAMETERS AND DS'S. (DSEEDS)
DATA DS1/4257/,DS2/501/,DS3/9074/,DS4/419/
DS1=SEED(1)
DS2=SEED(2)
DS3=SEED(3)
DS4=SEED(4)
XM=1.0
SIGSC=0.5
MMM=NINT((NUM*NUM*RESIZE*RESIZE*AVFTP)/AVAREA)

9 FORMAT(' ',# OF RAIN PATCHES ',I5)

SIGMA=SQRT(SIGSC)

N=500

DC 5 I=1,200

DC 6 J=1,200

IV(I,J)=0

6 CCNTINUE

5 CCNTINUE

CALL GGUBS(DS1,N,U1)

CALL GGUBS(DS2,N,U2)

CALL GGNLG(DS3,N,XM,SIGMA,LGN1)

CALL GGNLG(DS4,N,XM,SIGMA,LGN2)

C GENERATE CLOUD FIELD OVER 50 BY 50 AREA

DC 777 M=1,MMM

MINI=INT(200*U1(M))

MINJ=INT(200*U2(M))

MAXI=MINI+INT(LGN1(M))

MAXJ=MINJ+INT(LGN2(M))

IF(MINI .LE. 1) MINI=1

IF(MINJ .LE. 1) MINJ=1

IF(MAXI .GE. 200) MAXI=200

IF(MAXJ .GE. 200) MAXJ=200

DC 11 I=MINI,MAXI

DC 12 J=MINJ,MAXJ

IV(I,J)=1

12 CCNTINUE

11 CCNTINUE

777 CCNTINUE

C DISTORT RECTANGULAR RAIN CLOUDS

DC 51 J=1,199

DC 52 I=1,199

IF(IV(I,J) .NE. IV(I+1,J+1)) IV(I,J)=INT(U1(I)+0.5)

IF(IV(I,J) .NE. IV(I,J+1)) IV(I,J)=INT(U2(J)+0.5)

52 CCNTINUE

51 CCNTINUE

C SPECIFY BACKGROUND RAIN

NNOR=90000

CALL GGNPM(DS3,NNOR,U)

K=0

DC 21 I=1,300

DC 22 J=1,300

K=K+1

E(I,J)=U(K)

22 CCNTINUE

```

21 CONTINUE
   DD 23 I=2,299
   DC 24 J=2,299
   XP=E(I,J)+.1084*(E(I-1,J)+E(I+1,J)+E(I,J-1)+E(I,J+1))
   Y(I,J)=EXP(XP+1.140)
24 CONTINUE
23 CONTINUE
C
C GET SNAPSHOTS BY MOVING THE 128 BY 128 WINDOW OF CLOUDS OVER THE
C MOVING RAINFIELD. THE ADVECTION IS ACCOMPLISHED BY INCREMENTING II
C AND JJ. THE RAIN RATE AT THE BOUNDARIES OF THE CLOUDS CHANGES VIA
C THIS JOINT DYNAMIC. WRITE EACH SNAPSHOT TO TAPE AFTER IT IS
C GENERATED.
   II=0
   JJ=0
   I1=1
   J1=1
   DC 65 NDAYS=1,2
   IDAY=IDAY+1
   OPEN(UNIT=11,ERR=200,STATUS='NEW',FILE=FILELE(IDAY),
   ACCESS='SEQUENTIAL',FORM='UNFORMATTED',IOSTAT=IOS)
   DC 66 NSNP=1,24
   II=II+1
   III=II+127
   JJ=JJ+1
   JJJ=JJ+127
   DC 70 I=1,128
   DC 71 J=1,128
   PR(I,J)=Y(II+I+I1, JJ+J+J1)*IV(II+I, JJ+J)
71 CONTINUE
70 CONTINUE
   I1=I1+1
   J1=J1+1
   TIME=TIME+DELTAT
   WRITE (UNIT=11) GFHASE, DAY, TIME, RESIZE, AVAREA, AVFTR, XM, SIGMA,
   C((RR(I,J), I=1,128), J=1,129)
66 CONTINUE
   CLOSE(UNIT=11,ERR=300,STATUS='KEEP',IOSTAT=ICS)
65 CONTINUE
100 SEED(1)=DS1
   SEED(2)=DS2
   SEED(3)=DS3
   SEED(4)=DS4
200 IF (IOS .NE. 0) THEN
   WRITE (6,299)
299   FCRMAT(' ','I/C ERROR ON OPEN')
   GO TO 400
   END IF
300 IF (ICS .NE. 0) THEN
   WRITE (6,399)
399   FCRMAT(' ','I/C ERROR ON CLOSE')
   GO TO 400
   END IF
400 RETURN
   END
C
   SUBROUTINE LEVPTC(N,RINC,R)
C
C R. L. MARTIN ARC & GSFC
C OCT. 31,1985
C
C THIS SUBROUTINE TAKES A 2 DIMENSIONAL ARRAY OF VALUES, UP TO 128 BY
C 128 AND BINS THEM IN 10 EQUALLY SPACED BINS RINC IN SIZE, PLUS A ZERO
C LEVEL. VALUES OUTSIDE THE RANGE ARE SET EQUAL TO 0 OR 10. THE LEVELS
C ARE THEN PRINTED OUT AS CHARACTERS TO REPRESENT THE CONTOUR LEVELS OF
C THE ARRAY.
C
C INPUT : N - SIZE OF THE SQUARE ARRAY, R.
C         RINC - SIZE OF BINS TO PLACE THE DATA IN
C         R - 2 DIMENSIONAL ARRAY OF DATA VALUES
C
C INTERNAL : I - LOOP INDEX
C           J - LOOP INDEX
C           LP - ARRAY OF INTEGER VALUES TO REPRESENT THE BINNED DATA
C           R - ARRAY OF CHARACTERS TO REPRESENT THE BINNED DATA
C
C RETURNS : NOTHING

```

ORIGINAL PAGE IS
OF POOR QUALITY

```

REAL*4 R(128,1:*),RINC
INTEGER*4 N,LP(128),I,J
CHARACTER*1 P(128)
WRITE(6,1)
1 FORMAT('1', 'RELATIVE RAINFALL RATE')
DC 5 I=1,N
P(I)='.'
5 CCNTINUE
WRITE(6,41)(P(J), J=1,N)
DC 10 I=1,N
DC 20 J=1,N
LP(J)=0
20 CCNTINUE
DC 30 J=1,N
IF (R(I,J) .GT. 0.0) LP(J)=INT((R(I,J)/RINC)+1.0)
IF (LP(J) .GT. 10) LP(J)=10
30 CCNTINUE
DC 40 J=1,N
GC TC (101,102,103,104,105,106,107,108,109,110,111) LP(J)+1
101 P(J)=' '
GC TC 40
102 P(J)=':'
GC TC 40
103 P(J)='-'
GC TC 40
104 P(J)='+'
GC TC 40
105 P(J)='/'
GC TC 40
106 P(J)='H'
GC TC 40
107 P(J)='X'
GC TC 40
108 P(J)='#'
GC TC 40
109 P(J)='W'
GC TC 40
110 P(J)='2'
GC TC 40
111 P(J)='*'
40 CCNTINUE
WRITE(6,41)(P(J), J=1,N)
41 FORMAT(' ', ' ', ' ', ' ', 12E A, ' ')
10 CCNTINUE
DC 50 I=1,N
P(I)='.'
50 CCNTINUE
WRITE(6,41)(P(J), J=1,N)
RETURN
END

```

THIS PAGE IS
 OF POOR QUALITY

ATTACHMENT D

ON THE LOGNORMALITY OF RAINRATE

Benjamin KedeM
Department of Mathematics
University of Maryland
College Park, Maryland

and

Long S. Chiu
Applied Research Corporation
Landover, Maryland

Submitted to Proceedings of the National Academy of Sciences

(To appear)

ABSTRACT

A stochastic regression model is used in modeling rainrate. Under some conditions on the model parameters, it is shown that rainrate is asymptotically lognormal. An application of the model to the GATE data shows a remarkable agreement between the assumed and estimated model parameters for rainrate averaged over sufficiently large area and a sampling interval of 15 minutes.

1. INTRODUCTION

There is ample evidence based on observations that rain characteristics tend to be approximately lognormally distributed. This observation is shared by quite a few research workers who considered different data sets. These pertain to the duration of rainfall and amount, and to horizontal and vertical cloud extent in tropical and extratropical regions under a wide variety of convective conditions (Biondini 1976, Lopez 1977, Houze and Cheng 1977, Chiu et al. 1986). The question is then what makes the lognormal distribution so prevalent when it comes to rain systems and whether there is any theoretical basis for these observational findings. On practical grounds, we may ask whether at all it even makes sense to fit a lognormal distribution to rain characteristics and under what conditions. This is the subject of the present note. We will focus on the lognormality of rainrate.

Many authors believe that the lognormal distribution is a natural outcome of the so called law of proportionate effect (Aitchison and Brown 1963, p. 22). Accordingly, $\{X_j\}$ satisfies the law of proportionate effect if

$$X_j - X_{j-1} = \epsilon_j X_{j-1}$$

where the ϵ_j 's are mutually independent and are also independent of the X_j 's. While the law of proportionate effect is of funda-

mental importance in motivating the lognormal distribution, the independence assumption on the ε_j is quite restrictive and can in fact be relaxed. It is sufficient that the ε_j 's obey conditions which guarantee the asymptotic normality of sums in terms of these variates. For this to hold, they need not be independent and may even be dependent on the X_j 's.

In the present note, we discuss a certain type of dynamic regression model which together with less restrictive conditions, yields the lognormality of rainrate asymptotically. The model has a strong intuitive appeal and is quite flexible in that it requires only a few parameters which can be easily estimated from data. Using a novel estimation procedure, the model is fitted to the GATE (GARP-Global Atmospheric Research Program-Atlantic Tropical Experiment) data. It is shown that some requirements for asymptotic lognormality are satisfied by the data. Furthermore realizations produced by the model appear to be very similar to those produced by real rainrate data.

It should be emphasized that our result is model based and that by itself does not constitute a "proof" that rainrate is precisely lognormally distributed. We merely provide reasonable conditions which lead to lognormality, and indeed some of our conditions are well supported by the GATE data. It seems to us that the present approach is an improvement over the approach which solely relies on the law of proportionate effect.

2. A STOCHASTIC MODEL FOR RAINRATE

To unravel the lognormal mystery, we begin with a rather naive notion of a rain element. Conditional on rain, we conceive of a rain element as a small volume in space containing small droplets of water which have the following dynamics. Let time be discrete. At the $n-1$ time step, some droplets give rise to a

new generation of droplets through a complicated physical process, some droplets leave the volume while new ones, called immigrants, arrive to join the folks of the new generation. It is really a process of replacement and immigration where the replacement refers to droplets already in the volume. The droplets are being replaced by a non-negative number of droplets where zero could mean complete departure or emigration. Thus at time n , the number of droplets in the volume in space is the sum of the replacement droplets and the immigrants. Let X_{n-1} stand for the (random) number of droplets in the volume at time $n-1$ and suppose the i th droplet there is replaced by $Y_{n,i}$ fresh droplets while I_n denotes the number of immigrants. Then at time n , the rain element contains

$$X_n = \sum_{i=1}^{X_{n-1}} Y_{n,i} + I_n, \quad n = 1, 2, \dots \quad (1)$$

droplets with the convention that $\sum_1^0 = 0$. For (1) to cover dry periods and shifts from dry (wet) to wet (dry) periods the following interpretation is adopted. Most of the time when it is not raining, the rain element is dry and both X_n and I_n vanish. The rain element becomes active as soon as I_n admits a positive value. This sets the X_n , and hence the $Y_{n,i}$, in motion until the X_n vanish. The process restarts when I_n admits again a positive value. I_n can be thought of as the part of the process responsible for the occurrence of rain storms while $\sum Y_{n,j}$ pertains to the duration and amount of rain.

The most important parameters associated with the dynamic model (1) are

$$EY_{n,i} = m, \quad EI_n = \lambda, \quad n, i = 1, 2, \dots$$

No further assumption is needed for the present use of the model except for A1 and A2 below.

When the occurrences of rain are not too frequent, we expect λ to be small and close to zero. When it does rain, it usually persists for a while before it stops. This means that m should be close to 1 but still strictly less than 1. If m is greater than or equal to 1, the duration and amount can be explosive. Thus an indication of goodness of fit of (1) to rainrate data is small λ , and m close to but smaller than unity. It is interesting to apply the model to real data to see if these conditions are met.

When $\{Y_{n,i}\}$, $\{I_n\}$ are families of mutually independent non-negative integer valued random variables, the process $\{X_n\}$ is called a Galton-Watson Process with Immigration (Athreya and Ney 1972, p.263). This type of process was introduced as early as 1915 by Smoluchowski whose work is reported by Chandrasekhar (1943, chapter 3). Smoluchowski used the model to study the fluctuations in the number of particles contained in a small volume which exhibit random motion. However, we do not necessarily require the Y 's and I 's to be independent.

There is a well known device which transforms (1) into a more convenient regression equation which takes into account past values of X_n (Heyde and Seneta 1972, Winnicki 1986). Let \mathcal{F}_n be the σ -field generated by the random variables (X_0, X_1, \dots, X_n) , and note that

$$E(X_n | \mathcal{F}_{n-1}) = m X_{n-1} + \lambda$$

Define ε_n by the difference

$$\varepsilon_n = X_n - E(X_n | \mathcal{F}_{n-1})$$

and write (1) as

$$X_n = m X_{n-1} + \lambda + \varepsilon_n \quad (2)$$

Then $\{X_n\}$ is seen to be a stochastic difference equation where ε_n is a martingale difference (Lai and Wei 1982); i.e., ε_n is \mathcal{F}_n -measureable and $E(\varepsilon_n | \mathcal{F}_{n-1}) = 0$ for every n . An important example is the case of independent ε_n with mean 0 which is not required here. Other than its formal importance as expressed in (2), martingale differences follows the Central Limit Theorem under quite general conditions.

Since X_n refers to the density of droplets in the rain element, it is related to the rainrate. But multiplication of (2) by a constant leaves the model intact and we can actually think of X_n as representing rainrate. We therefore model rainrate dynamics by (2) where X_n admits only non-negative values.

3. CONTINUITY ASSUMPTION

In its present form, equation (2) is a fairly general model which could represent a wide range of physical and statistical processes. In order to ensure the lognormality of X_n , some more assumptions are needed.

Let $\{X_n\}$ $n = 0, 1, \dots$, be the stochastic process (2) which stands for the rainrate process at a given rain element. Assume that the X_0, X_1, X_2, \dots , are readings at time $0, T, 2T, \dots$, where the sampling interval T is small. The main assumption we shall adhere to is that of continuity: when the sampling interval T is sufficiently small we require that, conditional on rain, X_n and X_{n-1} be close to each other as is the case with many continuous phenomena in nature. This implies that the ε_n , the "errors", are themselves small. For normality we also require the sum of squares of the ε_n to explode. More precisely, conditional on rain (i.e., positive X_n 's) we assume

$$A1: \quad |X_i - X_{i-1}| \ll X_{i-1}$$

$$A2: \quad (1/n) \sum_{i=1}^n E[(\varepsilon_i/X_{i-1})^2 | \mathcal{F}_{i-1}] \rightarrow c^2 > 0, \quad n \rightarrow \infty$$

Since $E(\varepsilon_i/X_{i-1} | \mathcal{F}_{i-1}) = 0$, and since by A1 ε_i/X_{i-1} is essentially bounded as $m \rightarrow 1$ and $\lambda \rightarrow 0$, it follows that (McLeish 1974, Basawa and Prakasa Rao 1980, p.388)

$$(1/\sqrt{n}) \sum_{i=1}^n \varepsilon_i/X_{i-1} \xrightarrow{d} N(0, c^2), \quad n \rightarrow \infty.$$

4. ASYMPTOTIC LOGNORMALITY OF RAINRATE

Let $\chi[A]$ be the indicator of the event A , and define δ_n by

$$\delta_n = \frac{X_n - X_{n-1}}{X_{n-1} + \chi[X_{n-1} = 0]}$$

Then (2) can be written as

$$\begin{aligned} X_n &= (1+\delta_n)X_{n-1} + I_n \chi[X_{n-1} = 0] \\ &= (1+\delta_n)(1+\delta_{n-1})X_{n-2} + (1+\delta_n)I_{n-1} \chi[X_{n-2} = 0] + I_n \chi[X_{n-1} = 0] \\ &\vdots \\ &= \prod_{j=1}^n (1+\delta_j) X_0 + \sum_{k=2}^n \prod_{j=k}^n (1+\delta_j) I_{k-1} \chi[X_{k-2} = 0] + I_n \chi[X_{n-1} = 0] \end{aligned} \quad (3)$$

Thus, conditional on rain, it follows that

$$X_n = (1+\delta_n)(1+\delta_{n-1}) \dots (1+\delta_1) X_0 \quad (4)$$

from which we obtain by A1 that

$$\log(X_n/X_0) \approx \sum_{i=1}^n \delta_i \quad (5)$$

or

$$\log(X_n/X_0) + \sum_{i=1}^n [(1-m)-\lambda/X_{i-1}] \approx \sum_{i=1}^n \varepsilon_i/X_{i-1} \quad (6)$$

Therefore for m sufficiently close to 1 and λ close to 0, A1 and A2 imply that for large n

$$\left(\frac{X_n}{X_0}\right)^{1/\sqrt{n}} \left[1+(1-m) - \frac{\lambda}{n} \sum_{i=1}^n \frac{1}{X_{i-1}}\right]^{\sqrt{n}} \sim \wedge(0, c^2) \quad (7)$$

where $\wedge(0, c^2)$ denotes the lognormal distribution with parameters 0 and c^2 (Aitchinson and Brown 1963). When $m \rightarrow 1$ and $\lambda \rightarrow 0$, we obtain the useful approximation

$$\left(\frac{X_n}{X_0}\right)^{1/\sqrt{n}} \sim \wedge(0, c^2) \quad (8)$$

The 0 parameter is expected if we assume that X_n for large n is independent of X_0 and that the two are identically distributed. Under these conditions both $X_n^{1/n}$ and $X_0^{1/n}$ are asymptotically $(\mu, 1/2c^2)$ for some μ .

5. STATISTICAL ESTIMATION OF m and λ

A great deal of the foregoing discussion depends on m being close to but strictly smaller than 1, and λ positive but close to 0. To verify these conditions, the parameters should be estimated as precisely as possible. Fortunately, this estimation problem is a special case of a general problem investigated in detail by Lai and Wei (1982) who give conditions under which the

least squares estimates converge almost surely (that is, with probability one. This is abbreviated a.s.) to the respective parameters. Winnicki (1986) has suggested that m and λ should be estimated from the weighted model

$$\frac{X_n}{\sqrt{X_{n-1}+1}} = m \frac{X_{n-1}}{\sqrt{X_{n-1}+1}} + \lambda \frac{1}{\sqrt{X_{n-1}+1}} + \varepsilon_n^* \quad (8)$$

where $\varepsilon_n^* \equiv \varepsilon_n / \sqrt{X_{n-1}+1}$, by minimizing the sum of squares of the ε_n^* . The estimates obtained in this way are called weighted least squares and are shown, under some conditions, to be superior to the ordinary least squares when m is close to 1. Now, the Lai and Wei (1982) theory can be applied to the stochastic regression model (8) since ε_n^* in (8) is still a martingale difference. This is done next.

Denote the weighted least squares estimators by \hat{m} , $\hat{\lambda}$ and the design matrix by X_n . Then

$$X_n = \begin{bmatrix} \frac{X_1}{\sqrt{X_1+1}} & \frac{1}{\sqrt{X_1+1}} \\ \frac{X_2}{\sqrt{X_2+1}} & \frac{1}{\sqrt{X_2+1}} \\ \vdots & \vdots \\ \frac{X_n}{\sqrt{X_n+1}} & \frac{1}{\sqrt{X_n+1}} \end{bmatrix}$$

Define a 2x2 matrix A by, $A \equiv X_n' X_n$ and let $\lambda_{\min}(n)$ and $\lambda_{\max}(n)$ be, respectively, the smaller and larger eigenvalues of A . Then the relevant result of Lai and Wei (1982) can be stated as follows, assuming model (8). Assume

$$(a) \quad \sup_n E \left[|\varepsilon_n^*|^a \mid \mathcal{F}_{n-1} \right] < \infty \quad \text{a.s. for some } a > 2,$$

and that

$$(b) \quad \lambda_{\min}(n) \rightarrow \infty \quad \text{such that as } n \rightarrow \infty \\ \log \lambda_{\max}(n) = o\left[\lambda_{\min}(n)\right] \quad \text{a.s.}$$

Then

$$(\hat{m}, \hat{\lambda}) \rightarrow (m, \lambda) \quad \text{a.s.}$$

Thus, when (a) and (b) are satisfied, the result guarantees a strong sense of convergence of the weighted least squares estimates. The estimates themselves are given in Winnicki (1986) as

$$\hat{m} = \frac{\sum_{i=1}^n X_i \sum_{i=1}^n \frac{1}{X_{i-1}+1} - n \sum_{i=1}^n \frac{X_i}{X_{i-1}+1}}{\sum_{i=1}^n (X_{i-1}+1) \sum_{i=1}^n \frac{1}{X_{i-1}+1} - n^2} \quad (9)$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n X_{i-1} \sum_{i=1}^n \frac{X_i}{X_{i-1}+1} - \sum_{i=1}^n X_i \sum_{i=1}^n \frac{X_{i-1}}{X_{i-1}+1}}{\sum_{i=1}^n (X_{i-1}+1) \sum_{i=1}^n \frac{1}{X_{i-1}+1} - n^2} \quad (10)$$

where n is the series size.

Since observed rainrate is finite, condition (a) is automatically satisfied. To verify condition (b) analytically is difficult in general but it can be verified from data. The rainrate data we have in mind are described in the next section. For rainrate averages obtained from squares of 32 by 32 km² at 15 minute intervals, the results from two different time series are given in Table 1. The series size ranges from $n=100$ to $n=1700$, and it is seen that condition (b) is satisfied since $\lambda_{\min}(n)$ tends to infinity faster than $\log(\lambda_{\max}(n))$. Similar results were obtained for other time series and so, for all practical purposes, the door is now open to the actual estimation of m, λ using these data.

Table 1. Two cases for which condition (b) is satisfied. The rainrate series are sampled every 15 minutes over a square of 32x32 km²

n	$\lambda_{\min}(n)$	$\lambda_{\max}(n)$	$[\log \lambda_{\max}(n)]/\lambda_{\min}(n)$
FIRST TIME SERIES			
100	6.368	97.212	0.719
200	53.972	185.318	0.097
400	108.249	351.806	0.054
600	142.773	533.428	0.044
800	151.526	722.730	0.043
1000	421.242	901.590	0.016
1200	438.366	1084.106	0.016
1500	475.987	1359.853	0.015
1700	514.737	1540.269	0.014
SECOND TIME SERIES			
100	66.853	132.213	0.073
200	117.430	212.229	0.046
400	224.335	474.801	0.027
600	322.439	600.600	0.020
800	389.040	732.568	0.017
1000	526.569	945.093	0.013
1200	576.146	1094.382	0.012
1500	706.008	1363.871	0.010
1700	729.100	1537.641	0.010

6. APPLICATION TO GATE DATA

We applied the model to rainfall data collected during GATE. GATE was conducted in the Summer of 1974. During roughly three tri-weekly periods, detailed rainfall measurements from rain gauges and radars on an array of research vessels were made over an area called the B-Scale. The B-Scale encompasses an area of about 400 kms in diameter. Arkell and Hudlow (1977) composited the radar ship data and presented 15 minutes radar reflectivity scan data. Patterson et al. (1979) converted the radar reflectivity data to rainrates which are binned into 4 by 4 km² pixels. This data set is probably as yet one of the most extensive rainfall measurements made over the oceans.

Time series of rainrate for individual pixels (4 by 4 km² resolution) and for area averages (10 by 10 pixels or 40 by 40 km²) have been extracted from the first tri-weekly period in GATE (called Phase 1). The parameters of the model are estimated by the method of weighted least squares described above. Tables 2-5 give the estimated m and λ for 10 by 10 pixel arrays and for individual pixels situated at the center of the GATE area.

The results for large area averages of 10 by 10 pixels are shown in table 2 and 3. For each 10 by 10 pixel array throughout the GATE area a time series was obtained from which m and λ are estimated using (9) and (10). The estimated m are very close to but less than 1 except for some boundary points where there are missing data. At the four corners, there are no data at all in the 10 by 10 pixel array. The λ field in table 3 shows small values except at the boundaries where again the problem of missing data is encountered. We see that for large area averages sampled (really visited!) at $T = 15$ minute intervals the results are very satisfactory and so a lognormal fit makes good sense.

For individual pixels (4 by 4 km²) m is still fairly large although not as close to 1 as in the 10 by 10 pixel array case, but λ is large as seen in tables 4 and 5 respectively. The reason for this can be attributed to the sampling interval of 15 minutes: for smaller pixels we need to sample more often than 15 minutes to achieve similar results. This suggests that the model approaches the lognormal limit for large aggregates at the 15 minute sampling rate, and more generally, that there exists a time scale which corresponds to a spatial scale. This dependence of the model parameters on the averaging area can be seen very clearly from Figures 1 and 2 where \hat{m} and $\hat{\lambda}$ are given as a function of the pixel size (i.e. the averaging area) while the sampling interval is fixed at $T = 15$ minutes. The pixel sizes examined are 4x4, 8x8, 16x16, 24x24, 32x32, 40x40 and 352x352 km². We therefore conclude that lognormality of rainrate can already be observed fairly closely by averaging over pixels whose area is roughly as small as 40x40 km² where the sampling frequency is 15 minutes. This finding is enhanced by a histogram plot in Figure 3 derived from about 60000 40x40 km² GATE pixels. The figure displays the distribution of the rainrate areal average on a logarithmic scale. The distribution appears to be fairly symmetric in support of the above discussion.

Table 2. ESTIMATED m FOR 10 BY 10 PIXEL AVERAGES

Each number represents estimates for a 10 by 10 pixel average. The total area covers the whole of the GATE area. The four corners contains no data.

---	.40	.69	.89	.88	.92	.86	.85	.62	---
.55	.96	.98	.97	.97	.94	.93	.96	.96	.70
.94	.95	.96	.98	.97	.92	.97	.96	.97	.95
.94	.95	.96	.98	.99	.98	.98	.98	.98	.97
.94	.97	.98	.98	.99	.99	.99	.99	.99	.98
.96	.97	.97	.98	.98	.99	.99	.99	.99	.98
.95	.97	.98	.98	.98	.99	.99	.99	.99	.97
.97	.97	.98	.98	.99	.99	.99	.98	.98	.96
.79	.97	.98	.99	.98	.98	.99	.98	.98	.81
---	.97	.98	.98	.99	.99	.99	.99	.99	---

Table 3. ESTIMATED LAMBDA FOR 10 BY 10 PIXEL AVERAGES

Each number represents estimates for a 10 by 10 pixel area average. The area covers the whole of the GATE area. Data are missing in the four corners of GATE.

---	.077	.029	.019	.019	.019	.028	.049	.233	---
.050	.024	.015	.028	.029	.033	.049	.038	.074	.506
.031	.025	.045	.047	.047	.089	.053	.068	.066	.086
.041	.073	.081	.071	.048	.055	.075	.084	.080	.089
.072	.094	.098	.080	.057	.059	.065	.074	.076	.094
.086	.099	.096	.082	.079	.056	.076	.073	.076	.115
.134	.092	.101	.079	.071	.061	.069	.086	.083	.157
.109	.106	.086	.098	.078	.083	.100	.124	.111	.170
.817	.125	.113	.109	.114	.106	.108	.140	.1451	.350
---	1.124	.212	.143	.123	.114	.164	.1671	.502	---

Table 4. ESTIMATED m FOR INDIVIDUAL PIXELS.

Each number represents the estimates for a 4 km by 4 km area average. The total area is 40 kms by 40 kms situated at the center of the GATE area.

.91	.93	.92	.93	.90	.92	.87	.88	.88	.92
.87	.91	.90	.91	.93	.91	.92	.87	.89	.88
.89	.91	.92	.93	.94	.94	.93	.90	.85	.89
.91	.90	.86	.85	.91	.93	.94	.94	.92	.91
.90	.86	.83	.88	.92	.95	.94	.90	.94	.90
.93	.90	.83	.88	.89	.94	.91	.92	.92	.94
.82	.80	.82	.87	.88	.90	.87	.92	.92	.93
.88	.83	.82	.89	.90	.85	.87	.92	.91	.90
.89	.83	.82	.89	.91	.92	.87	.89	.88	.92
.90	.86	.91	.91	.93	.89	.86	.86	.91	.91

Table 5. ESTIMATED LAMBDA FOR INDIVIDUAL PIXELS.

Each number represents the estimate for an individual pixel. The total area covers an area of 40 kms by 40 kms situated at the center of the GATE area.

.86	.40	.57	.49	.66	.53	.81	.68	.72	.40
.50	.38	.57	.52	.43	.57	.48	.88	.75	.60
.44	.41	.37	.40	.38	.41	.50	.68	1.05	.60
.84	.39	.53	.51	.40	.37	.38	.42	.56	.54
.86	.52	.62	.34	.32	.22	.34	.61	.50	.75
.28	.39	.57	.38	.39	.21	.37	.50	.53	.45
.65	.67	.61	.41	.38	.36	.57	.40	.54	.45
.81	.50	.47	.37	.33	.57	.53	.36	.50	.61
.26	.45	.65	.45	.42	.33	.49	.40	.52	.40
.24	.52	.39	.47	.37	.46	.60	.56	.41	.45

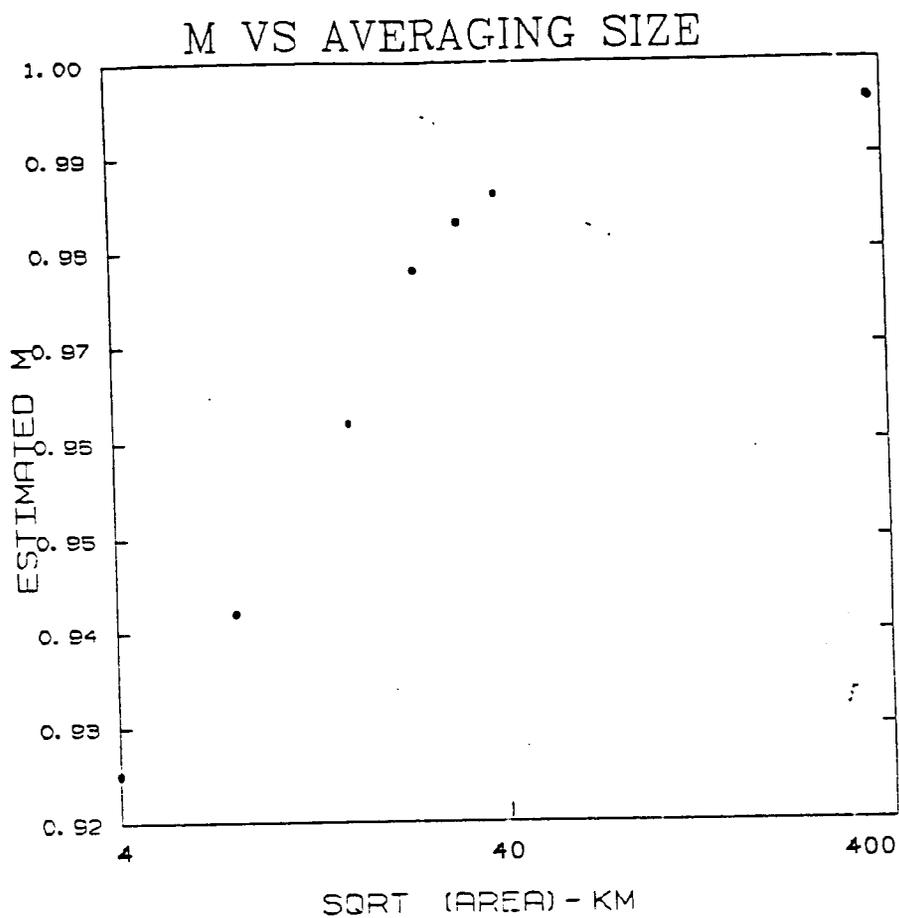


Figure 1. The monotone increase in \hat{m} as a function of the pixel size.

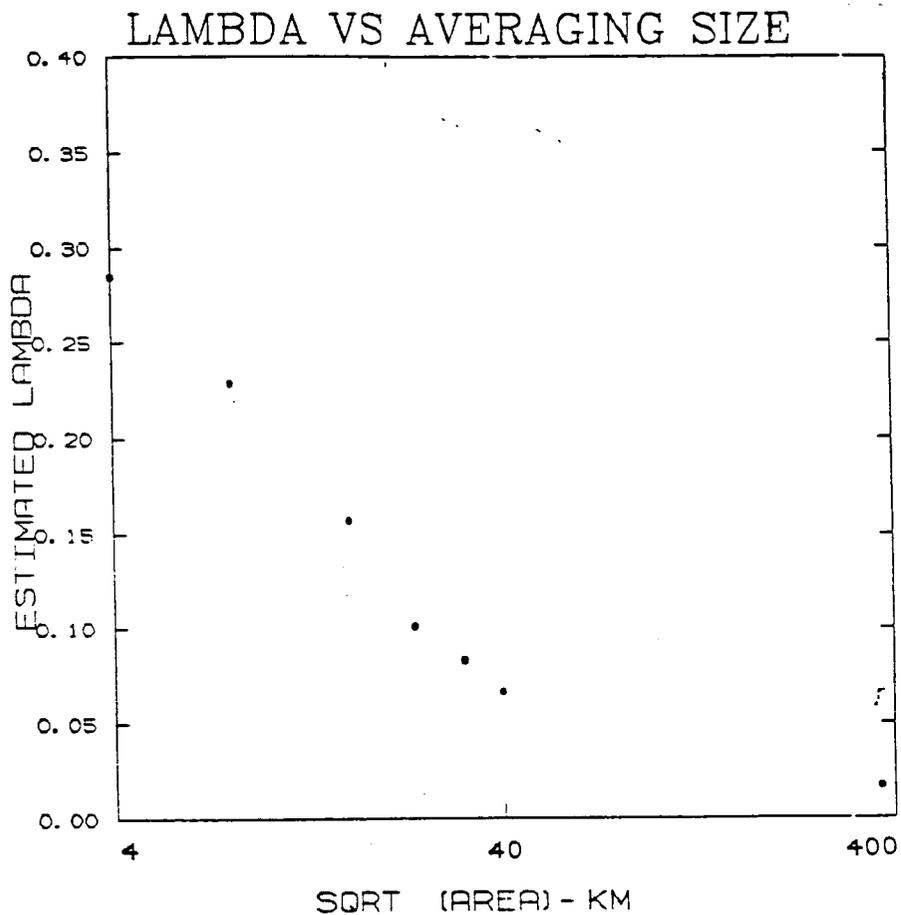


Figure 2. The monotone decrease in $\hat{\lambda}$ as a function of the pixel size.

7. SIMULATION

We end this note with a short graphical comparison between time series from (1) and a typical time series from the GATE data. It should be noted that in the foregoing discussion we made no restrictions on the Y's and I's in (1) except for the requirements that they be non-negative integers. In fact (2) is a more general model since even this last restriction is removed. Thus, if (1) is capable of producing realizations which resemble real rainrate data, this shows all the more the adequacy of (2) which is the model we used all along in the foregoing discussion.

Now, there are many ways to simulate (1). One simple and fast way is to take the Y's and I's as independent Poisson random variables with parameters m and λ respectively. By this process we generated the time series in Figure 5. Figure 4 shows a typical time series from GATE which constitutes 100 hours. The sudden bursts of rain storms, duration, intensity, decay and inter arrival times between storms in the real and simulated realizations are quite intriguingly similar.

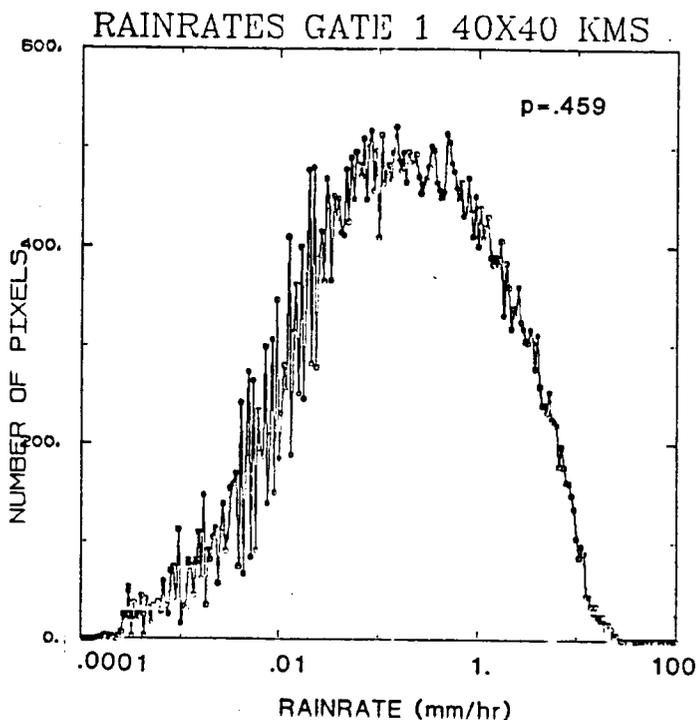


Figure 3. A histogram obtained from a large number of $40 \times 40 \text{ km}^2$ GATE pixels. The observations are log-areal averages.

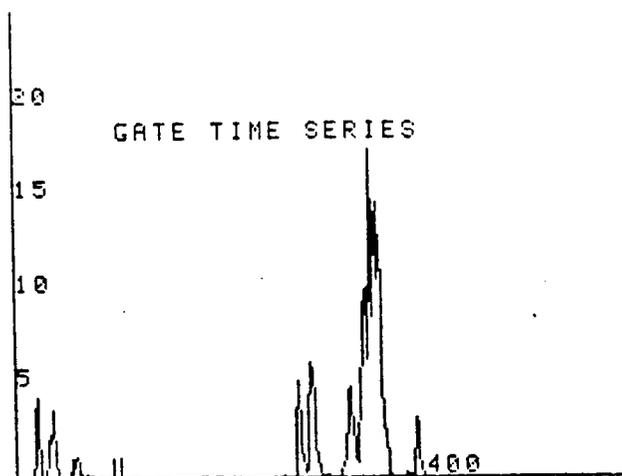


Figure 4. 400 observations from a typical GATE time series taken every 15 minutes. The pixel size is 32 km^2 .

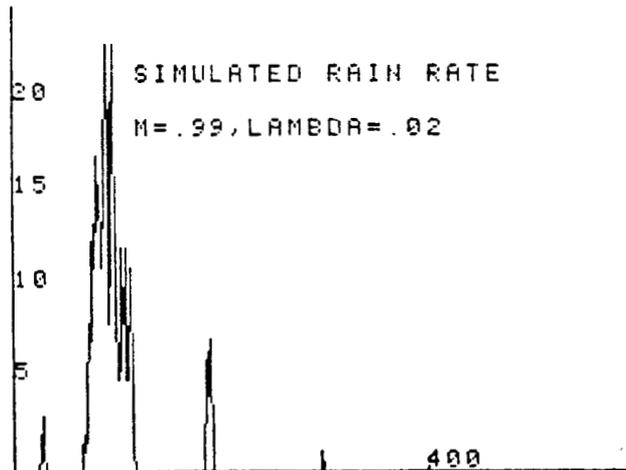
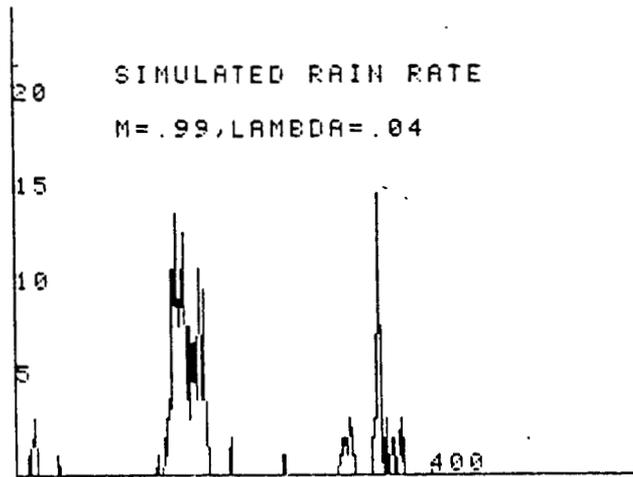


Figure 5. Realizations from (1). $\{Y_{n,i}\}$, $\{I_n\}$ are independent Poisson random variables with parameters m, λ respectively.

SUMMARY

The puzzling experimental fact that rainrate tends to follow a lognormal distribution was explained with the aid of a model. Accordingly, under some conditions, as a rain storm develops, rainrate tends to follow a lognormal distribution. The conditions on the model parameters are shown to be satisfied fairly closely by the GATE data for time series which consist of rainrate averages over sufficiently large pixels observed every 15 minutes. A variant special case of the model is capable of producing realizations which appear to be very similar to real rainrate time series. Another fact is that the eigenvalue conditions necessary for the almost sure convergence of the weighted least squares estimates are well satisfied by the GATE data. In light of all these consistencies it is hoped that the model (2) can serve in settling other intriguing facts about rain.

Acknowledgement: The Authors wish to express their gratitude to J. Winnicki for many useful discussions. This research is partially supported by a NASA-SBIR grant under contract number NAS5-29271

REFERENCES

Atkinson, J. and Brown, J. A. C., The Lognormal Distribution, Cambridge University Press, London (1963).

Arkell, R. and Hudlow, M., GATE International Meteorological Radar Atlas, U.S. dept. of Commerce, NOAA, U.S. government printing office, Washington, D.C. (1977).

Athreya, K. B. and Ney, P. E., Branching Processes, Springer-Verlag, New York (1972).

Basawa, I. V. and Prakasa Rao, L. S., Statistical Inference for Stochastic Processes, Academic Press, New York (1980).

Biondini, R., "Cloud motion and rainfall statistics," J. Appl. Meteor., 15, (1976) 205-224.

Chandrasekhar, S., "Stochastical problems in physics and astronomy," Rev. Mod. Phys., 15, (1943) 1-89.

Chiu, L. S., Kedem, B. and North, G. R., "Estimating the mean areal average rainfall: a mixed distribution approach." To appear.

Heyde, C. C. and Seneta, E., "Estimation theory for growth and immigration rates in a multiplicative process," J. Appl. Prob., 9, (1972) 235-256.

Houze, R. A. and Cheng, C. P., "Random characteristics of tropical convections observed during GATE: mean properties and trends over the summer season," Month. Weath. Review, 105 (1977) 964-980.

Lai, T. L. and Wei, C. Z., "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," Annal. of Stat., Vol. 10, (1982) 154-166.

Lopez, R. E., "The lognormal distribution and cummulus cloud populations," Month. Weath. Review, 105 (1977) 865-872.

McLeish, D. L., "Dependent central limit theorems and invariance principles," *Annal. of Prob.*, Vol. 2, (1974) 620-628.

Patterson, V. L., Hudlow, M. D., Pytlowany, P. J., Richards, F. P. and Hoff, J. D., GATE Radar Rainfall Processing System, NOAA Technical Memorandum, EDIS 26, NOAA, Washington, D.C. (1979).

Smoluchowski, M., "Drei vortrage uber diffusion, Brownsche Bewegung und Wagulation von Kolloidteilchen," *Physik. Zeits.*, 17 (1916) 557-585.

Winnicki, J., "A unified estimation theory for the branching process with immigration," Ph.D. thesis, Dept. of Math., Univ. of MD., College Park (1986).