General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

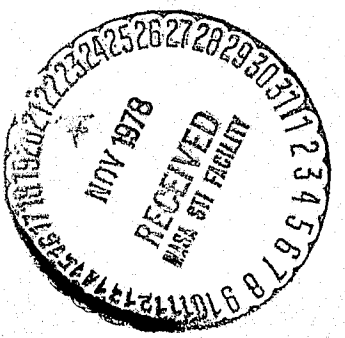
Produced by the NASA Center for Aerospace Information (CASI)

NASA

Technical Memorandum 79631

A Statistical Technique for Determining Rainfall Over Land Employing Nimbus-6 ESMR Measurements

E. Rodgers, H. Siddalingaiah,
A. T. C. Chang, and T. Wilheit



AUGUST 1978

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

(NASA-TM-79631) A STATISTICAL TECHNIQUE FOR DETERMINING RAINFALL OVER LAND EMPLOYING NIMBUS-6 ESMR MEASUREMENTS (NASA) 41 p HC A03/MF A01 CSCL 04B

N79-10671

Unclas
G3/47 37157

TM 79631

A STATISTICAL TECHNIQUE FOR DETERMINING RAINFALL OVER LAND EMPLOYING NIMBUS-6 ESMR MEASUREMENTS

by

Edward Rodgers
NASA/GSFC
Greenbelt, Maryland 20771

Honnappa Siddalingaiah
Computer Sciences Corporation
Silver Spring, Maryland 20910

A.T.C. Chang
NASA/GSFC
Greenbelt, Maryland 20771

Thomas Wilheit
NASA/GSFC
Greenbelt, Maryland 20771

August 1978

GODDARD SPACE FLIGHT CENTER

Greenbelt, Maryland 20771

PRECEDING PAGE BLANK NOT FILMED

A STATISTICAL TECHNIQUE FOR DETERMINING RAINFALL OVER LAND EMPLOYING NIMBUS-6 ESMR MEASUREMENTS

Edward Rodgers
NASA/GSFC
Greenbelt, Maryland 20771

Honnappa Siddalingaiah
Computer Sciences Corporation
Silver Spring, Maryland 20910

A.T.C. Chang
NASA/GSFC
Greenbelt, Maryland 20771

PRECEDING PAGE BLANK NOT FILMED

Thomas Wilheit
NASA/GSFC
Greenbelt, Maryland 20771

ABSTRACT

At 37 GHz, the frequency at which the Nimbus-6 Electrically Scanning Microwave Radiometer (ESMR-6) measures upwelling radiance, it has been shown theoretically that the atmospheric scattering and the relative independence on electromagnetic polarization of the radiances emerging from hydrometeors make it possible to monitor remotely active rainfall over land. In order to verify experimentally these theoretical findings and to develop an algorithm to monitor rainfall over land, the digitized ESMR-6 measurements were examined statistically.

Horizontally and vertically polarized brightness temperature pairs (T_H, T_V) from ESMR-6 were sampled for areas of rainfall over land as determined from the rain recording stations and the WSR-57 radar, and areas of wet and dry ground (whose thermodynamic temperatures were greater than 5°C) over the Southeastern United States. These three categories of brightness temperatures were found to be significantly different in the sense that the chances that the mean vectors of any two populations coincided were less than 1 in 100. Since these categories were significantly different, classification algorithms were then developed. Three decision rules were examined: the Fisher linear classifier, the Bayesian quadratic

classifier, and a non parametric linear classifier. The Bayesian algorithm was found to perform best, particularly at a higher confidence level. An independent test case analysis showed that a rainfall area delineated by the Bayesian classifier coincided well with the synoptic scale rainfall area mapped by ground recording rain data and radar.

CONTENTS

		Page
ABSTRACT		iii
1.	INTRODUCTION	1
2.	THE ESMR-6 SYSTEM	4
3.	DATA SAMPLING	4
4.	STATISTICAL ANALYSIS	7
5.	CLASSIFICATION ALGORITHMS	16
6.	ERROR ANALYSIS	20
7.	ALGORITHM EVALUATION	22
8.	CONCLUSION	29
REFERENCES		31

ILLUSTRATIONS

Figure		Page
1	Computed horizontally and vertically polarized brightness temperature at 37.0 GHz as a function of rain rate	3
2	ESMR-6 horizontally polarized T_B (1655 GMT January 6, 1976) superimposed on the PPI WSR-57 radar image at Waycross, Georgia (1700 GMT January 6, 1976)	6
3	Vertically polarized vs. horizontally polarized EMSR-6 T_B for each sampled category (rain over land, and wet and dry land surfaces)	8
4	Same as Figure 3 except for surfaces whose thermodynamic temperatures are greater than 15°C	10
5	Same as Figure 3 except for surfaces whose thermodynamic temperatures are between 5-15°C	11
6	Marginal densities (histograms) of the total sampled horizontally and vertically polarized T_B's from the three populations	14
7	Rainfall over the Southeast United States as delineated by the WSR-57 radar and hourly rainfall reporting stations. Time of the data is approximately 1630 GMT September 14, 1976	23
8	ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 70%. Time of Nimbus-6 pass— 1630 GMT September 14, 1976	25
9	ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 80%. Time of Nimbus-6 pass— 1630 GMT September 14, 1976	26
10	ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 80%. Time of Nimbus-6 pass— 1645 GMT August 27, 1976	28

TABLES

Table		Page
1	Dates of Synoptic Rain Cases Used to Develop ESMR-6 Classification Algorithms	5
2	Elementary Statistics of Sampled Data (Surface Temperature $\geq 5°C$)	9
3	Elementary Statistics of Sampled Data (Surface Temperature $>15°C$)	12
4	Elementary Statistics of Sampled Data (Surface Temperature 5-15°C)	12
5	Chi-Square Test for Normality	13
6	Covariance Matrices of Sampled Data	15
7	Significance Between Means (F Test)	17
8	Simultaneous Confidence Intervals for Differences Between Mean Brightness Temperatures Representing Rain (R), Dry (D), and Wet (W) Areas	17
9	Probabilities of Misclassification: Theoretical Computation	21
10	Bayesian Classification Error Matrix Determined from Sampled Data	21
11	Probabilities of Misclassification: Theoretical Computation ($<15°C$)	22
12	Probabilities of Misclassification: Theoretical Computation ($>15°C$)	22

A STATISTICAL TECHNIQUE FOR DETERMINING RAINFALL OVER LAND EMPLOYING NIMBUS-6 ESMR MEASUREMENTS

1. INTRODUCTION

Precipitation is a fundamental meteorological parameter and it functions as an indicator, determinant, or component of the distribution and amount of latent heat release which is critical to the understanding of storm and global atmospheric energetics and of the total hydrological cycle. The ability to monitor the coverage and movement of rain over land areas is important because of the direct impact of rain on crop production and also its influence on insect breeding areas and migration (Idso et al. 1975). Moreover, the destructive effects due to heavy rainfall could be reduced by advance warnings furnished by satellites that map regions of heavy rain.

Since the advent of the polar orbiting and geosynchronous satellites, quantitative techniques have been developed to estimate rainfall indirectly. Estimations of rainfall have been made by correlating rain rate and amount with either cloud cover and type, cloud brightness, or cloud temperature utilizing visible and infrared sensors on board these satellites (Barret, 1970, 1973; Martin and Scherer, 1973; Martin et al., 1975; Follansbee and Oliver, 1975; Oliver and Scoffield, 1976; and Griffith et al., 1976). However, all these techniques suffer from being only indirectly related to rainfall.

The microwave technique developed by Wilheit et al. (1977) has a direct physical relationship with rain rate but only over ocean areas. This technique establishes a relationship between rain rate in the dynamic range of 1-20 mm hr^{-1} and brightness temperatures (T_B) measured by the Electrically Scanning Microwave Radiometer on board Nimbus-5 (ESMR-5), which senses at 19.35 GHz upwelling radiation emitted by the earth and its atmosphere.

Meneely (1974) demonstrated that rainfall rate and coverage cannot be delineated using ESMR-5 measurements over land areas. This is because the rain has only a weak effect on the upwellint T_B from the land and the effect of soil moisture is comparable. Thus, although rain-like patterns can be discerned in the data, they correspond to both active rain areas and

areas with moist soil. McFarland and Blanchard (1977), however, did demonstrate that rain amounts over land could be estimated indirectly by monitoring temporal changes in the ESMR-5 T_B.

Savage and Weinman (1975) and Savage et al. (1976) demonstrated theoretically that at 37.0 GHz (the frequency at which the Nimbus-6 ESMR [ESMR-6] sensor measures upwelling radiance) the scattering by hydrometeors is strong enough to provide a qualitative estimate of rain coverage over land. Furthermore, Weinman and Guetter (1977) demonstrated from a theoretical consideration that the upwelling radiation at 37.0 GHz emerging from hydrometeors was essentially unpolarized and therefore was in contrast to a wet surface background. According to the electromagnetic theory, if the emissivity of a surface is reduced by increasing its dielectric constant (as by adding moisture), then the emissivity will be highly polarized when viewed obliquely. These results are demonstrated in Figure 1 which displays theoretically calculated bipolarized 37.0 GHz T_B at 50° incidence angle with the earth surface for a given rain rate. These T_B's were derived from a radiative transfer model with Lambertian reflection (Born and Wolf, 1975) from land surfaces at a thermodynamic temperature of 299.1°K and with a fixed dielectric constant and an atmospheric freezing level at 4 km (Wilheit et. al., 1977). It is seen from this figure that as rain rate increases (beyond 4 mm hr^{-1}) the T_B decreases due to strong backscattering by the large raindrops. Also, the polarization difference becomes smaller. Moreover, Hall et al. (1978) inferred theoretically that information analogous to that provided by the National Weather Service radar summary charts can be produced when both ESMR-6 and the Temperature Humidity Infrared Radiometer (THIR) 11.5 μm data on board Nimbus-6 are used. Thus, it is reasonable to conclude from these theoretical considerations that rain coverage over dry land surfaces can be at least qualitatively monitored employing 37.0 GHz radiometer measurements from ESMR-6,

It is the purpose of this paper to substantiate the above conclusions and to arrive at an algorithm for the detection of rain over land by statistically analyzing ESMR-6 data. This statistical analysis will be performed by first sampling three categories of ESMR-6 T_B's

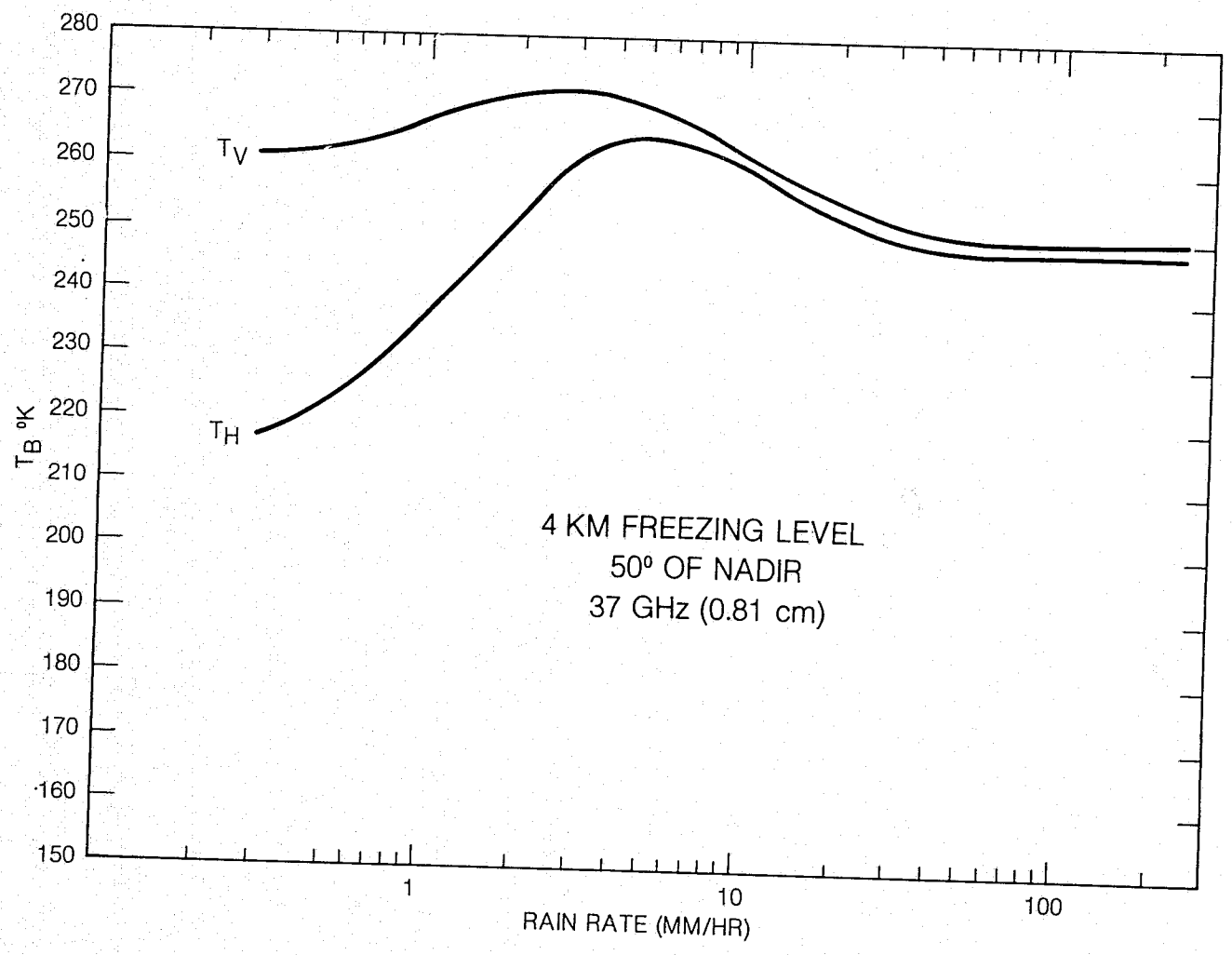


Figure 1. Computed horizontally and vertically polarized brightness temperature at 37.0 GHz as a function of rain rate. Freezing level assumed at 4 km and viewed at an incident angle of 50° with the earth surface.

(representing rain over land, wet land surfaces without rain, and dry land surfaces), then testing these populations for unqueness and separability, and finally developing a classification algorithm to delineate rain over land.

2. THE ESMR-6 SYSTEM

The ESMR-6 system flown aboard Nimbus-6 (Wilheit, 1975) receives the thermal radiation upwelling from the earth's surface and atmosphere in a 250 MHz band centered at 37 GHz. The antenna beam scans electrically an arc of 70° in 71 steps ahead of the spacecraft along a conical surface with a constant earth incidence angle of 50° every 5.3 seconds. The nominal resolution is 20 km crosstrack and 45 km downtrack. The instrument measures both horizontal and vertical polarization components by using two separate radiometric channels. The data are calibrated using warm (instrument ambient) and cold (cosmic background) inputs to the radiometer. Calibration errors have been observed in the data which appear to arise from a modulation of the loss in the antenna related to the sun angle and from an unexplained excess of noise in the data from the warm calibration source. An empirical correction was applied to these errors for all the data used in this study.

The T_B as observed from the satellite is dependent upon the emission from the earth's surface modified by the intervening atmosphere. The emissivity, being a function of the dielectric constant, is variable over land surfaces (depending on vegetation, soil type, soil moisture, etc.) and generally is large (Ca. 0.9). In rain situations three constituents contribute significantly to the absorption: molecular oxygen (Meeks and Lilly, 1963), water vapor (Staelin, 1966) and liquid water droplets (Mie, 1908; Gunn and East, 1954). Water droplets contribute more significantly to absorption and re-emittance than the other constituents and are the only source of scattering at this frequency. Ice crystals are essentially transparent.

3. DATA SAMPLING

Simultaneous ground stations and radar measurements of rain and ESMR-6 T_B were needed in order to develop an algorithm which classified a given ESMR-6 instantaneous field of view (IFOV) as rain over land, dry land surface, or wet land surface. Eight daytime synoptic scale rainfall cases over the southeastern United States were used where surface rainrate

data taken from stations reporting hourly rainfall amounts and from the WSR-57 radar coincided with Nimbus-6 overpass to within 5 minutes. The surface temperature in each of these cases was not less than 5°C. Rain areas were sampled within areas delineated as rain by either the WSR-57 radar (rain rates $\geqslant$ 2.5 mm hr^{-1}) and/or the stations reporting hourly rainfall amounts. The dates and time of the occurrance of these cases are given in Table 1. Wet land surfaces were sampled upwind and adjacent to the raincells observed on the WSR-57 radar and dry land surfaces were sampled over areas where rain had not fallen within a 24-hour period previous to the Nimbus-6 pass.

Table 1

Dates of Synoptic Rain Cases Used to Develop ESMR-6 Classification Algorithms

Case	Date	Time
1	31 July 1975	1620 GMT
2	4 August 1975	1635 GMT
3	1 October 1975	1700 GMT
4	7 November 1975	1700 GMT
5	12 November 1975	1700 GMT
6	29 December 1975	1717 GMT
7	3 January 1976	1715 GMT
8	6 January 1976	1655 GMT

Figure 2 illustrates the sampling technique. The figure shows the ESMR-6 horizontally polarized T_B's (°K) measured at approximately 1655 GMT January 6, 1976 together with rainfall data as delineated by the WSR-57 radar (located at Waycross, Georgia at 1700 GMT) and by stations reporting hourly rainfall amounts. The ESMR-6 T_B's are within the field of view of the radar where the circle shows the outer bounds of the PPI image at a 232 km radius. The shaded area represents rain (rainrate $\geqslant$ 2.5 mm hr^{-1}). The large dots are hourly rain recording stations where rain amounts (in inches) for hours ending at 1700, 1600, and 1500 GMT are displayed according to model in the figure. If no rain had fallen during that period, no measurements are shown. Station models reporting temperature, present weather, cloud type and amount, and wind direction and speed for 1800 GMT are also given. The small dots

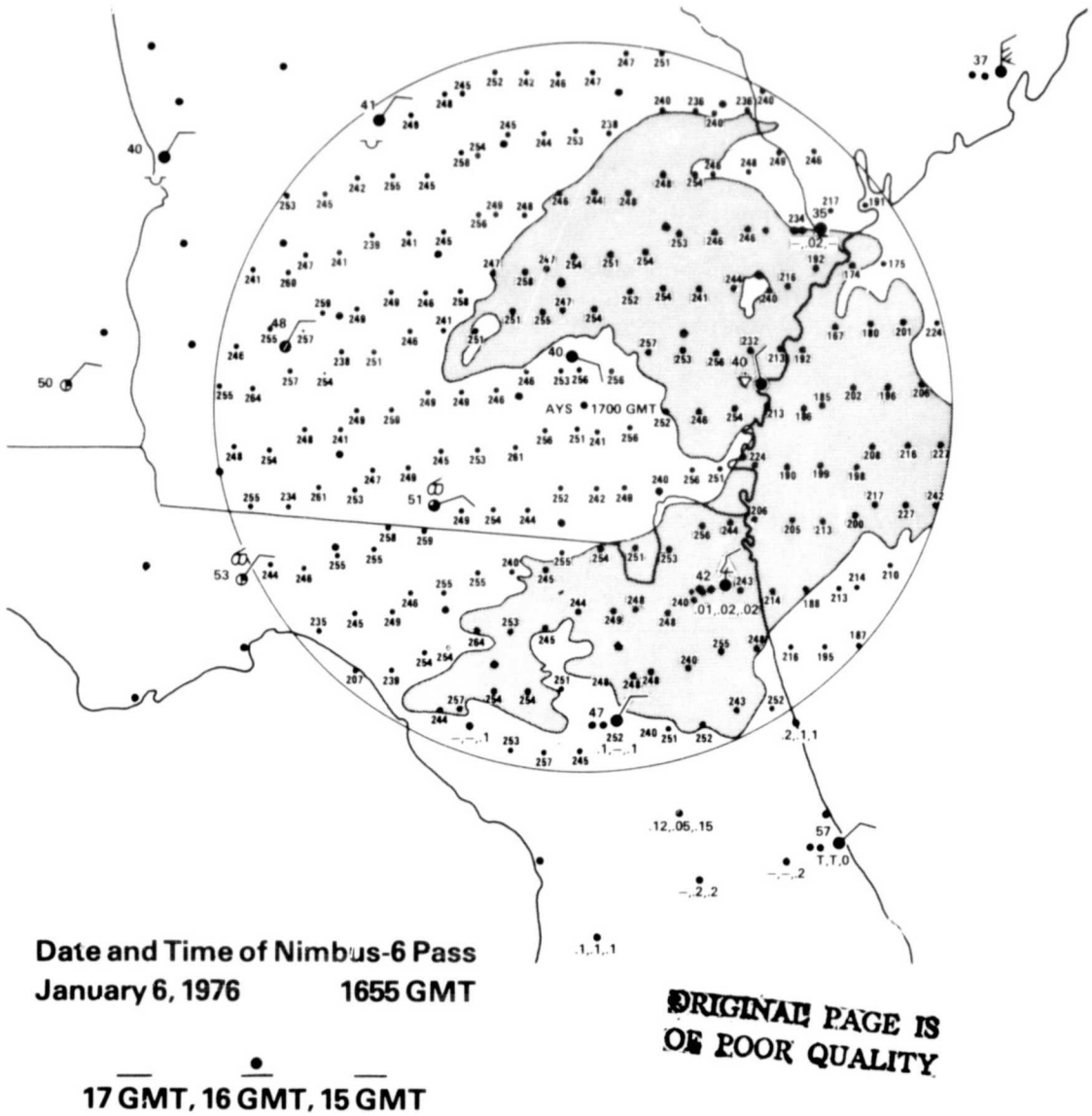


Figure 2. ESMR-6 horizontally polarized T_B (1655 GMT January 6, 1976) superimposed on the PPI WSR-57 radar image at Waycross, Georgia (1700 GMT January 6, 1976). Shaded area delineates rain rates $\geqslant$ 2.5 mm hr^{-1}. Circle shows outer bounds of PPI image (232 km radius). Large dots locate hourly rain recording stations. Station model (lower left hand corner) denotes the time hourly rainfall (inches) ended.

locate the center of the ESMR-6 footprints. For this case, ESMR-6 T_B's representing rain over land was sampled within the shaded area. The T_B's representing wet land surfaces were sampled southwest of the shaded area since the rain area was moving northeast, and T_B's representing dry land surfaces were sampled over western Georgia where rain had not fallen within 24 hours of the Nimbus-6 pass.

4. STATISTICAL ANALYSIS

Elementary statistics of the total sampled data (ESMR-6 measurements where surface thermodynamic temperatures were greater than 5°C) are presented in Table 2. The table gives for each category the sample size, the mean and standard deviation of the horizontally and vertically polarized T_B, the correlation between horizontally and vertically polarized T_B, and the mean difference between polarized T_B. These data are also shown as a scatter plot in Figure 3. In this figure the "C" represents the mean of the population and each frequency concentration ellipse encompasses 68 percent (one standard deviation) of the data within the population. The ellipses reveal the extent of scattering of data from each population, the correlation between the dual polarization T_B's, T_H and T_V, within each population (the higher the correlation the larger the eccentricity of the ellipse), and the extent of overlap among the populations. The three concurrent lines drawn in this figure are the Fisher (1938) linear discriminant lines which separate two-by-two the rain over land area (S_R), the dry land surface (S_D), and the wet land surface (S_W) populations represented by the T_B pairs (T_H, T_V).

It can be seen from Figure 3 and Table 2, that T_B's from rain areas over land are colder than those T_B's from dry land surface areas. Further, the difference between the mean horizontally and vertically polarized T_B's from rain areas over land (6.45°K) is much smaller than that for wet land surfaces (16.81°K). This is in accordance with theoretical findings that hydrometers are essentially unpolarized (Wienman and Guetter, 1977) whereas wet land surfaces are polarized. It is also seen from Figure 3 that the largest overlap occurs between the data obtained from rainfall areas and wet land surfaces. The reason for this is that sometimes

ORIGINAL PAGE IS
OF POOR QUALITY

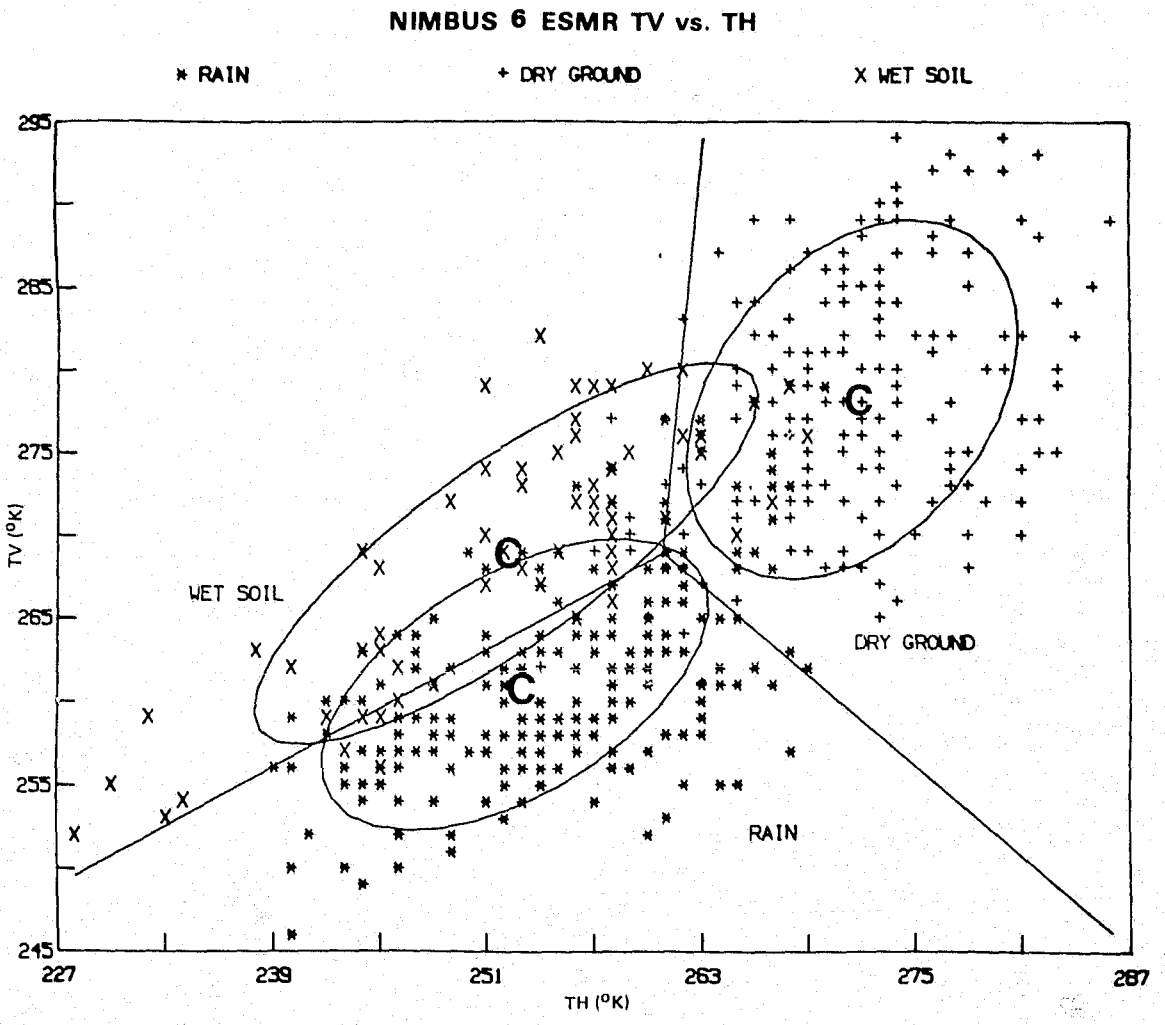


Figure 3. Horizontally polarized vs. vertically polarized ESMR-6 T_B for each sampled category (rain over land, and wet and dry land surfaces). The C's represent the mean points of the populations and the sllipses encompass 68 percent (one standard deviation) of the data from the respective categories. The three concurrent lines are the Fisher linear discriminant lines which separate two-by-two the three populations representing rain over land and wet and dry land surfaces.

Table 2

Elementary Statistics of Sampled Data (Surface Temperature $\geq 5°C$)

	Rain Area		Dry Ground		Wet Soil	
Sample Size: N	216		189		66	
	T_{HR}	T_{VR}	T_{HD}	T_{VD}	T_{HW}	T_{VW}
Mean: μ	254.53	260.98	271.46	278.18	252.05	268.86
Mean Brightness Temperature Difference	6.45		6.72		16.81	
Standard Deviation: d	7.21	5.81	6.18	7.20	9.41	7.64
Sample Correlation Coefficient Between T_H and T_V: ρ	0.55		0.37		0.82	

in sampling rain over land the total upwelling radiance received by the radiometer contains a direct surface contribution. This may occur when an IFOV of the ESMR-6 measurement is partially filled with moderate to heavy rain or when it is completely filled with light rain (background being wet land surface). Consequently, the T_B's for each category are somewhat similar, thus producing the overlap between rain over land and wet land surface classes.

Since the surface emission is given by ϵT_S, where ϵ is the surface emissivity and T_S is the surface thermodynamic temperature, there is an influence of T_S on ESMR-6 measured dry land surface T_B. A decrease in T_S results in a decrease in T_B from dry ground and consequently, the T_B contrast between dry land surfaces and rain over land will also decrease. These effects can be observed in Figures 4 and 5 and Tables 3 and 4. The figures and tables are identical to Figure and Table 2 respectively except that Figure 4 and Table 3 correspond to cases where the surface thermodynamic temperatures were above 15°C while Figure 5 and Table 4 correspond to cases where the surface thermodynamic temperatures were between tween 5 and 15°C. It is clear from Figure 5 and Table 4 that rain over land

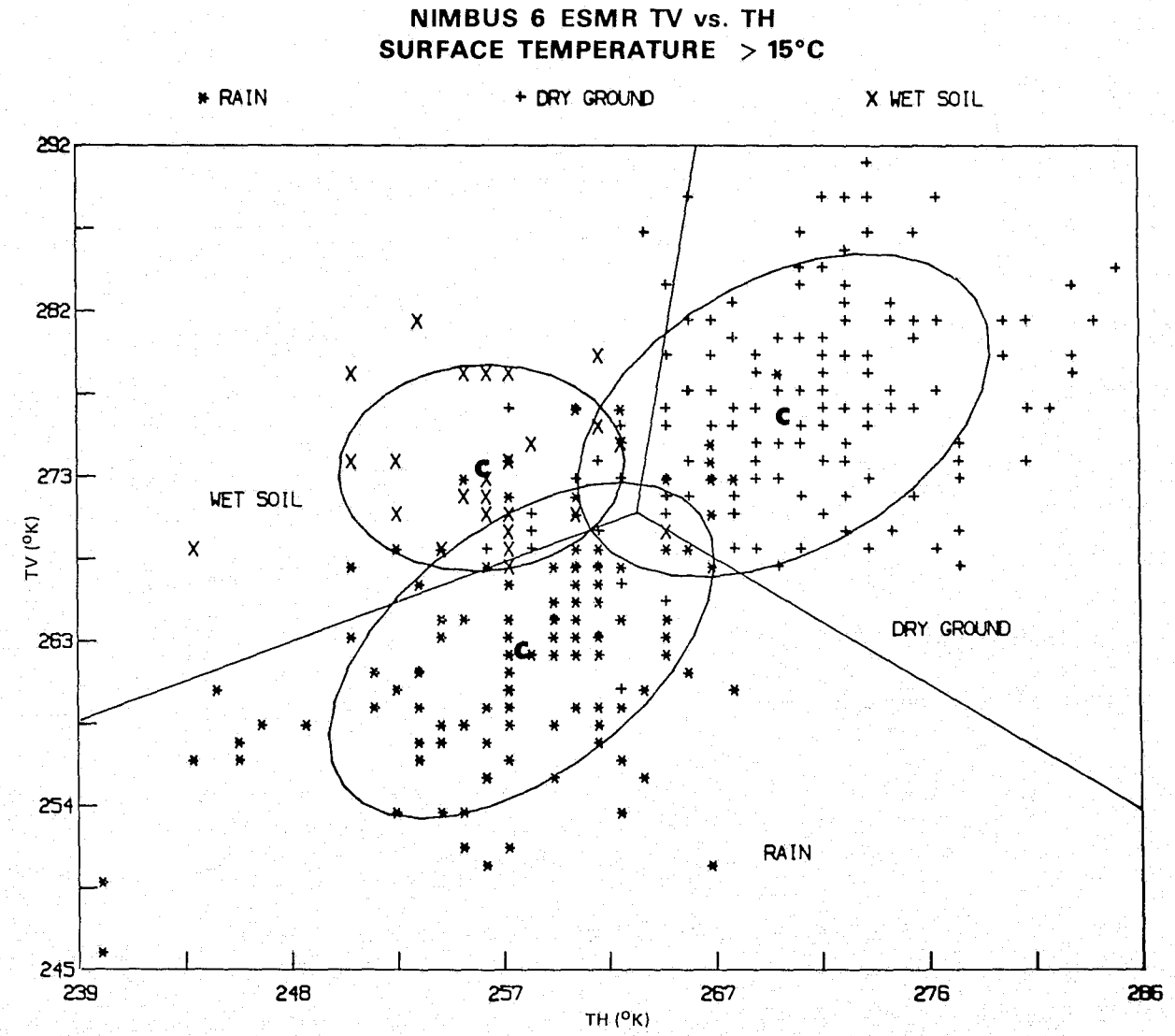


Figure 4. Same as Figure 3 except for surfaces whose thermodynamic temperatures are greater than 15°C

ORIGINAL PAGE IS
OF POOR QUALITY

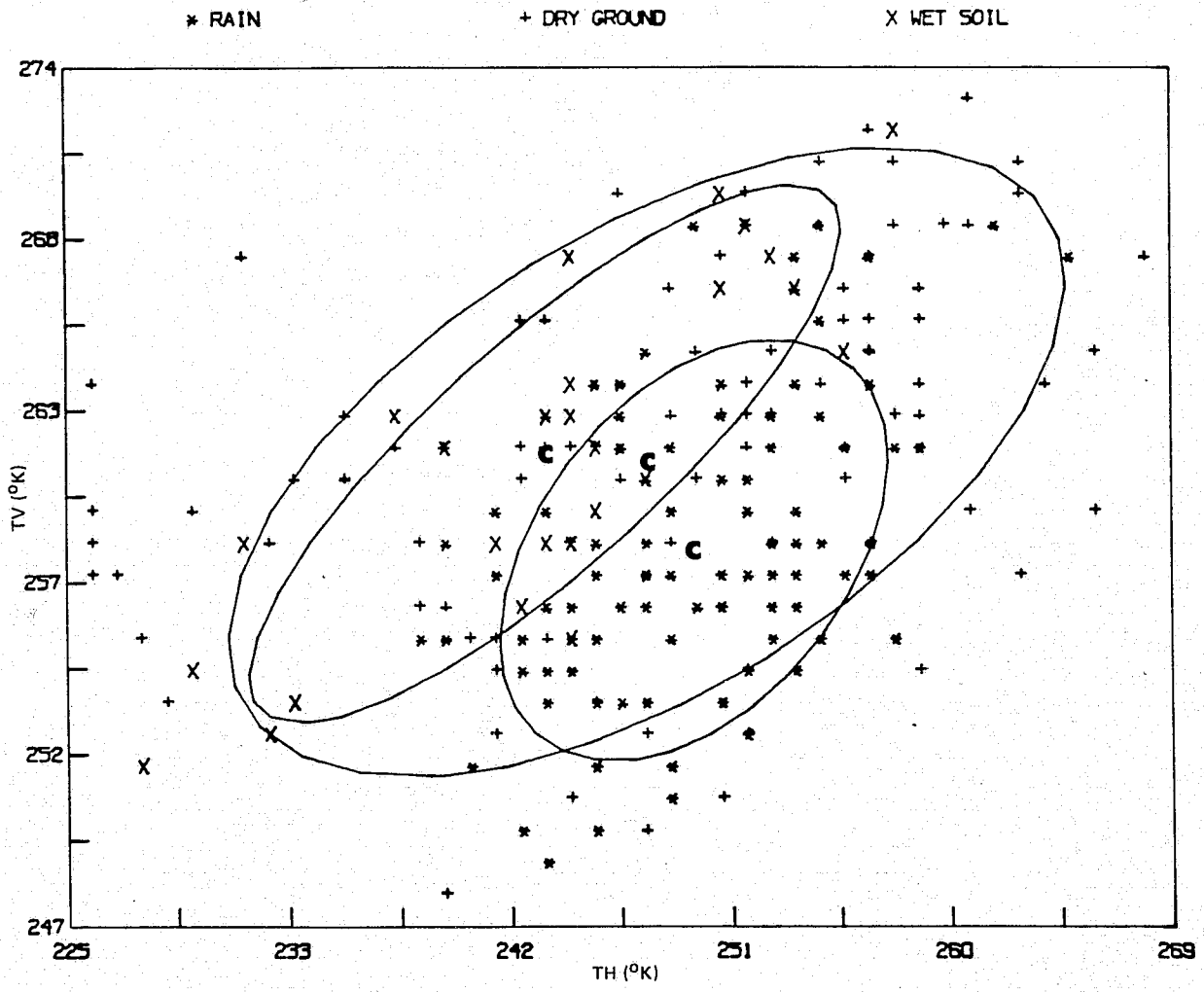


Figure 5. Same as Figure 3 except for surfaces whose thermodynamic temperatures are between 5-15°C

Table 3

Elementary Statistics of Sampled Data (Surface Temperature > 15°C)

	Rain Area		Dry Ground		Wet Soil	
Sample Size: N	112		145		26	
	T_{HR}	T_{VR}	T_{HD}	T_{VD}	T_{HW}	T_{VW}
Mean: μ	258.57	263.22	270.22	276.61	256.81	273.62
Mean Brightness Temperature Differences	4.65		6.39		16.81	
Standard Deviation: d	5.70	6.36	6.07	6.16	4.22	3.90
Sample Correlation Coefficient Between T_H and T_V: ρ	0.53		0.42		0.05	

Table 4

Elementary Statistics of Sampled Data (Surface Temperature 5-15°C)

	Rain Area		Dry Ground		Wet Soil	
Sample Size: N	104		98		28	
	T_{HR}	T_{VR}	T_{HD}	T_{VD}	T_{HW}	T_{VW}
Mean: μ	249.92	258.78	248.11	261.57	244.04	261.82
Mean Brightness Temperature Difference	8.86		13.46		17.78	
Standard Deviation: d	5.16	4.40	11.08	6.58	7.87	5.59
Sample Correlation Coefficient Between T_H and T_V: ρ	0.43		0.58		0.84	

is difficult to delineate from dry land surfaces when the surface thermodynamic temperature is below 15°C. Since the populations in Figure 5 cannot be separated, the Fisher linear discriminant lines are not drawn.

Figure 6 displays the marginal densities (histograms) of the total sampled horizontally and vertically polarized T_B's from the three populations. Table 5 presents the results of the chisquare test (Cochran, 1952) performed to validate the normal distribution of the data. Since each observed chi-square value in Table 5 is comparable to the corresponding critical (table) value at one percent, it is assumed that each marginal distribution of the data is Gaussian. Therefore, it is reasonable to assume that the data from each of the populations S_R, S_D or S_W satisfy the bivariate Gaussian density distribution:

$$f(\vec{x}) = \frac{1}{2\Pi\sqrt{|c|}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{u})^T\, c^{-1}(\vec{x}-\vec{u})\right] \qquad (1)$$

Where $\vec{x}$ is the two dimensional column vector (T_H, T_V), $\vec{u}$ is the mean of $\vec{x}$, c is the covariance matrix of the population, c^{-1} is the inverse of c, $|c|$ is the determinant of c, and $(\vec{x}-\vec{u})^T$ is the transpose of $(\vec{x}-\vec{u})$. The $\vec{u}$ and $\vec{c}$ are estimated using the sampled data from each class. The $\vec{u}$'s are provided by Table 2 and c, c^{-1}, and $|c|$ by Table 6.

Table 5

Chi-Square Test for Normality

	Rain		Dry		Wet	
	T_H	T_V	T_H	T_V	T_H	T_V
Number of Cells	8	8	8	9	6	5
Degrees of Freedom	5	5	5	6	3	2
Table Value of χ^2 at 0.01	15.09	15.09	15.09	16.81	11.34	9.21
Observed Value of χ^2	14.57	28.17	18.99	11.59	8.93	10.33

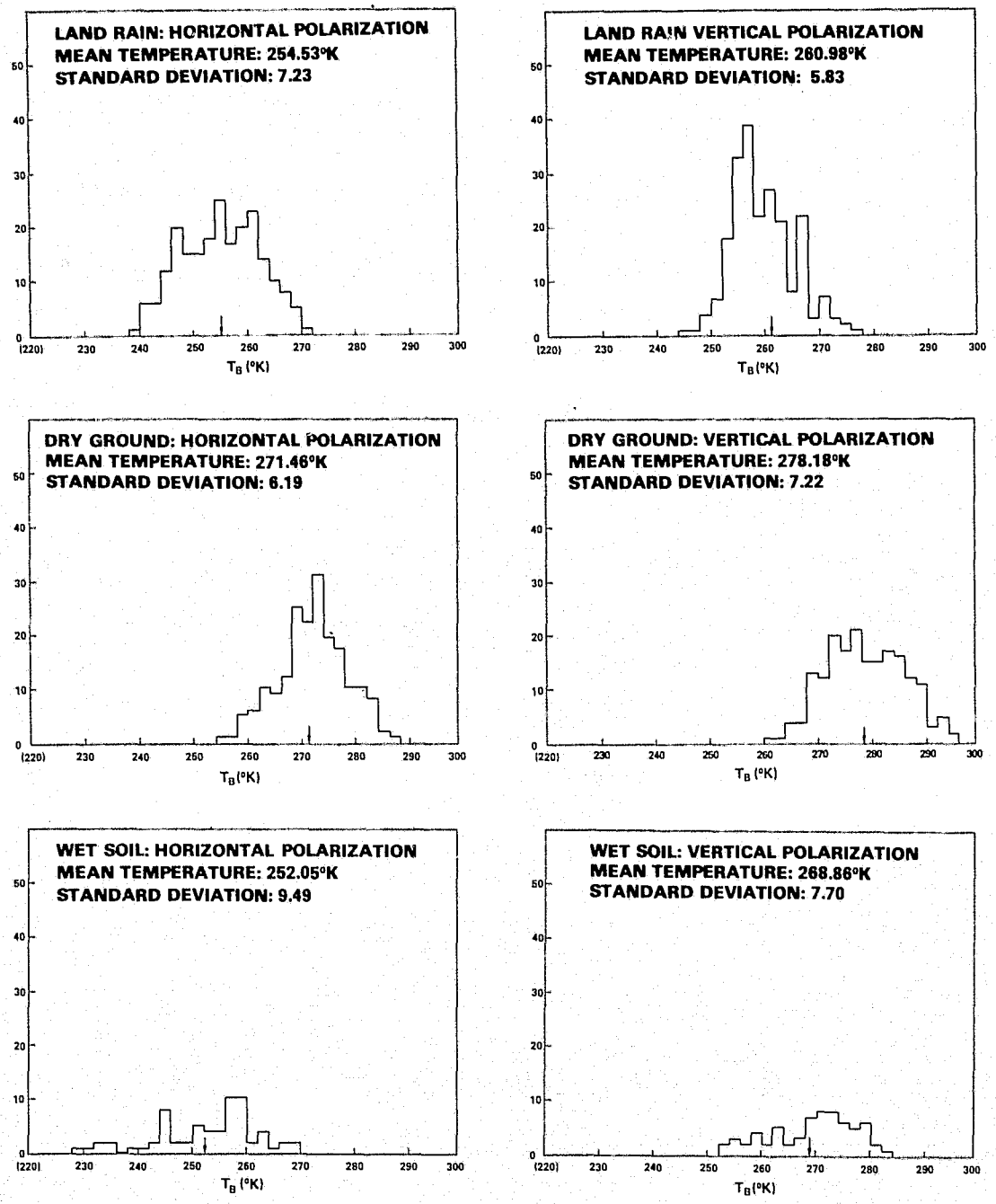


Figure 6. Marginal densities (histograms) of the total sampled horizontally and vertically polarized T_B's from the three populations

Table 6

Covariance Matricies of Sampled Data

	Matrix		Inverse	
	T_{HR}	T_{VR}	T_{HR}	T_{VR}
T_{HR}	52.23	23.02	0.0273	−0.0185
T_{VR}	23.02	33.93	−0.0185	0.0420

Covariance Matrix Determinant: 1242.424

Rain Area

	Matrix		Inverse	
	T_{HD}	T_{VD}	T_{HD}	T_{VD}
T_{HD}	38.36	16.51	0.030	−0.010
T_{VD}	16.51	52.14	−0.010	0.022

Covariance Matrix Determinant: 1727.7496

Dry Ground

	Matrix		Inverse	
	T_{HW}	T_{VW}	T_{HW}	T_{VW}
T_{HW}	90.39	59.73	0.034	−0.035
T_{VW}	59.73	58.28	−0.035	.0.053

Covariance Matrix Determinant: 1700.35248

Wet Soil

Prior to employing the data in Table 2 for the purpose of developing classification algorithms, the data were examined to verify whether the three populations were statistically distinguishable from one another. To accomplish this, an F (variance ratio) test, in terms of Hotelling's T^2 and Mahalanobis's D^2 (Kshirsagar, 1972) was performed to determine the significance of the differences between the means of any two classes. Then the simultaneous confidence intervals were estimated for these differences by Scheffé's procedure (Scheffé, 1943; Bennett, 1951).

Table 7 displays D^2 and T^2 as well as the computed and table (critical) values of F. The difference between the means of any two classes is highly significant since the observed value of F is much higher for each pair of classes than the corresponding critical (table) value of F at the 1 percent confidence level. That is, the probability that the mean vectors of any two populations are identical is less than 1 in 100.

Table 8 shows the estimated confidence intervals. It can be seen that only the interval for the differences between the wet land surface and rainfall over land mean horizontal polarization T_B's contains zero. Therefore, the three populations are distinguishable from one another when the dual polarization information is taken into consideration. However, the lower bounds of the mean differences between rainfall over land and wet land surface T_B's is smaller than those of the other two pairs. This indicates that it will be more difficult to distinguish an area of rain over land from wet land surfaces.

5. CLASSIFICATION ALGORITHMS

Since the populations were found to be statistically distinguishable and satisfied the Gaussian frequency distribution, three classification techniques were considered with the purpose of developing an efficient and effective classification algorithm to detect and delineate active rainfall over land from dry and wet land surfaces. The three techniques are: the Bayesian classifier, the Fisher linear discriminant classifier, and a non-parametric linear discriminant classifier.

Table 7

Significance Between Means (F Test)

Between the Means of ↓	Mahalanobis's Distance Squared D^2	Hotelling's T^2	The Observed Variance Ratio F	Table Value of F at 1%
Rain and Dry	9.13	920.35	459.04	3.83
Dry and Wet	6.03	295.13	146.28	3.87
Rain and Wet	4.00	202.06	100.67	3.86

Table 8

Simultaneous Confidence Intervals for Differences Between Mean Brightness Temperatures Representing Rain (R), Dry (D), and Wet (W) Areas

Polarization	Rain–Dry	Dry–Wet	Wet–Rain
Horizontal	$14.88 \leqslant \mu_{HD} - \mu_{HR} \leqslant 18.98$	$15.89 \leqslant \mu_{HD} - \mu_{HW} \leqslant 22.93$	$-1.15 \leqslant \mu_{HR} - \mu_{HW} \leqslant 6.11$
Vertical	$15.00 \leqslant \mu_{VD} - \mu_{VR} \leqslant 19.20$	$6.04 \leqslant \mu_{VD} - \mu_{VW} \leqslant 12.60$	$4.93 \leqslant \mu_{VW} - \mu_{VR} \leqslant 10.83$

The Bayesian classifier is a parametric classifier (i.e., it assumes the functional form of the relevant density function). The non-parametric linear discriminant classifier does not assume a density function (Bond and Atkinson, 1972). The Fisher linear discriminant classifier may be either parametric or non-parametric (Fisher, 1938). All three methods are termed supervised in the sense that it is necessary to use known sample data for the various classes to train the algorithms. Algorithms were developed using all three classifiers and tested using independent data. It was found that the results from the Bayesian classifier were superior to the other two methods. Hence, only the Bayesian classification technique will be described in the following.

The Bayesian classifier is a Gaussian parametric maximum likelihood quadratic classifier which requires the knowledge of the *a priori* probabilities for the occurrence of each class (Duda and Hart, 1973; Fu et al., 1969). It minimizes the average loss due to misclassification by assuming that each misclassification is equally costly.

It minimizes the conditional average loss:

$$L(\vec{x}, S_k) = \sum_{i=1}^{3} \lambda(S_k \mid S_i)\, P(S_i \mid \vec{x}) \tag{2}$$

where $\lambda(S_k \mid S_i)$ is the loss incurred when a measurement $\vec{x} = (T_H, T_V)$ actually belonging to class S_k is placed in class S_i and $P(S_i \mid \vec{x})$ is the *a priori* probability of the class S_i occurring having observed $\vec{x}$. The symmetrical loss function $\lambda(S_k \mid S_i)$ is given by:

$$\lambda(S_k \mid S_i) = \begin{cases} 0 & i = k \\ 1 & i \neq k \end{cases} \quad \text{if} \quad i, k = 1, 2, 3 \tag{3}$$

Hence, all misclassifications are equally costly and equation 2 now reduces to:

$$L(\vec{x}, S_k) = 1 - P(S_k \mid \vec{x}) \tag{4}$$

where $P(S_k \mid \vec{x})$ is the conditional probability that the class S_k to which $\vec{x}$ is assigned is correct.

The likelihood function $P(S_k \mid \vec{x})$ is given by the Bayes' rule:

$$P(S_k \mid \vec{x}) = \frac{P(\vec{x} \mid S_k)\, P(S_k)}{\sum_{j=1}^{3} P(\vec{x} \mid S_j)\, P(S_j)} \tag{5}$$

where $P(\vec{x} \mid S_j)$ is the bivariate Gaussian probability density function of $\vec{x}$ given that $\vec{x}$ is in S_j, and $P(S_j)$ is the *a priori* probability of the class S_j occurring. Sample data sizes given in Table 2 provide the values of $P(S_j)$. They are 0.459, 0.401, and 0.140 for the classes S_R, S_D, and S_W respectively.

Since the loss given by equation 4 is to be minimized, the quadratic discriminant functions are:

$$g_i(\vec{x}) = P(S_i)\, P(\vec{x} \mid S_i), \quad i = 1, 2, 3 \tag{6}$$

These functions, considering the relationship in equation 1, lead to the following decision rule. The measurement $\vec{x}$ belongs to the class S_k if:

$$\begin{aligned} &2 \ln P(S_k) - \ln |c_k| - (\vec{x} - \vec{\mu}_k)^T c_k^{-1} (\vec{x} - \vec{\mu}_k) > \\ &2 \ln P(S_i) - \ln |c_i| - (\vec{x} - \mu_i)^T c_i^{-1} (\vec{x} - \vec{\mu}_i) \end{aligned} \tag{7}$$

for all $i \neq k$, where c_j and $\vec{\mu}_j$ are the covariance matrix and the mean vector of the class S_j. Then, by substituting the relevant values into equation 7, one arrives at the following Bayesian algorithm. The pixel corresponding to the given vector (T_H, T_V) is rainfall over land, dry land surface, or wet land surface respectively, depending on which of the following values is the largest:

$$\begin{aligned} P_R(T_H, T_V) = &-0.027\, T_H^2 + 0.038\, T_H\, T_V - 0.042\, T_V^2 \\ &+3.826\, T_H + 12.250\, T_V - 2094.097 \end{aligned} \tag{8}$$

$$\begin{aligned} P_D(T_H, T_V) = &-0.030\, T_H^2 + 0.020\, T_H\, T_V - 0.022\, T_V^2 \\ &+10.720\, T_H + 6.811\, T_V - 2412.165 \end{aligned} \tag{9}$$

$$\begin{aligned} P_W(T_H, T_V) = &-0.034\, T_H^2 + 0.070\, T_H\, T_V - 0.053\, T_V^2 \\ &-1.678\, T_H + 10.846\, T_V - 1261.721 \end{aligned} \tag{10}$$

The quadratic function

$$Q_k(\vec{x}) = (\vec{x} - \vec{\mu}_k)^T c_k^{-1} (\vec{x} - \vec{\mu}_k) \tag{11}$$

has a chi-square distribution with two degrees of freedom (Scheffé, 1959). Therefore, a confidence value F can be associated with each classified pixel. F is given by:

$$F(\vec{x}) = 255 \left(1 - \frac{\sqrt{Q_K(\vec{x})}}{n_\sigma}\right) \tag{12}$$

where n_σ is the distance, in terms of standard deviation from the mean, to which zero confidence value is assigned.

6. ERROR ANALYSIS

An error estimate was made in order to evaluate quantitatively the performance of the Bayesian classification algorithm. The error rates were computed according to the asymptotic formulas given by Okamoto (1963), assuming that the population satisfies the Gaussian distribution, have different means, and have the same covariance matrices. The results are shown in Table 9. Virtually all of the misclassification probability in each case was accounted for by the first term of the asymptotic expansion:

$$\Phi(-\frac{1}{2}\Delta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{1}{2}\Delta} \exp\left[\frac{-z^2}{2}\right] dz \tag{13}$$

where

$$\Delta^2 = \frac{N_1 + N_2 - 5}{N_1 + N_2 - 2} D^2 - \frac{2(N_1 + N_2)}{N_1 N_2} \tag{14}$$

D is Mahalanobis' distance

and N_1, N_2 are sample sizes of the populations under consideration. Only a small fraction is contributed by the rest of the third order terms. From Table 9 it is clear that the chance of incorrectly classifying wet land surfaces or dry land surfaces as rain over land is nearly 23 percent. But when a given pixel is classified as a raining area and each of the eight contiguous pixels that cluster around it is also classified as rain over land, then the chance of misclassification of that central pixel is reduced to 7.7×10^{-6} percent assuming each pixel is independently classified.

Table 10 displays the actual probabilities that the Bayesian algorithm classify the sampled training data into the various population is as indicated. The average accuracy is the mean of the diagonal elements of the corresponding error matrix, and these averages compare well with the estimated average.

Table 9

Probabilities of Misclassification:
Theoretical Computation

Classified ↓ →	Rain	Dry	Wet
Rain	77.15	6.66	16.19
Dry	6.67	82.08	11.25
Wet	16.28	11.29	72.43

Average accuracy: 77.22 percent

Table 10

Bayesian Classification Error Matrix
Determined from Sampled Data

Classified ↓ →	Rain	Dry	Wet
Rain	89.35	6.02	4.63
Dry	7.41	91.53	1.06
Wet	27.27	15.15	57.58

Average accuracy: 79.49 percent

Tables 11 and 12 show the estimated error matrices corresponding to data which came from land areas where the surface thermodynamic temperature was less than or greater than 15°C respectively. It is apparent from the tables that the classifications are not definitive when the surface thermodynamic temperature is between 5°C and 15°C.

Table 11

Probabilities of Misclassification:
Theoretical Computation (<15°C)

Classified → / Rain	Rain	Dry	Wet
Rain	45.99	36.23	17.78
Dry	36.26	23.99	39.75
Wet	18.03	41.86	40.11

Average Accuracy: 36.70 percent

Table 12

Probabilities of Misclassification:
Theoretical Computation (>15°C)

Classified → / Rain	Rain	Dry	Wet
Rain	74.83	11.76	13.41
Dry	11.75	77.66	10.59
Wet	13.59	10.72	75.69

Average accuracy: 76.06 percent

7. ALGORITHM EVALUATION

A case not previously used in sampling was tested to verify qualitatively the performance of the Bayesian classification algorithm. This case consisted of a synoptic scale rain pattern over the Southeastern United States (14 September, 1976) which was observed by the ESMR-6 sensor (surface thermodynamic temperature $\geq 15°C$). Figure 7 shows the rainfall area

ORIGINAL PAGE IS
OF POOR QUALITY

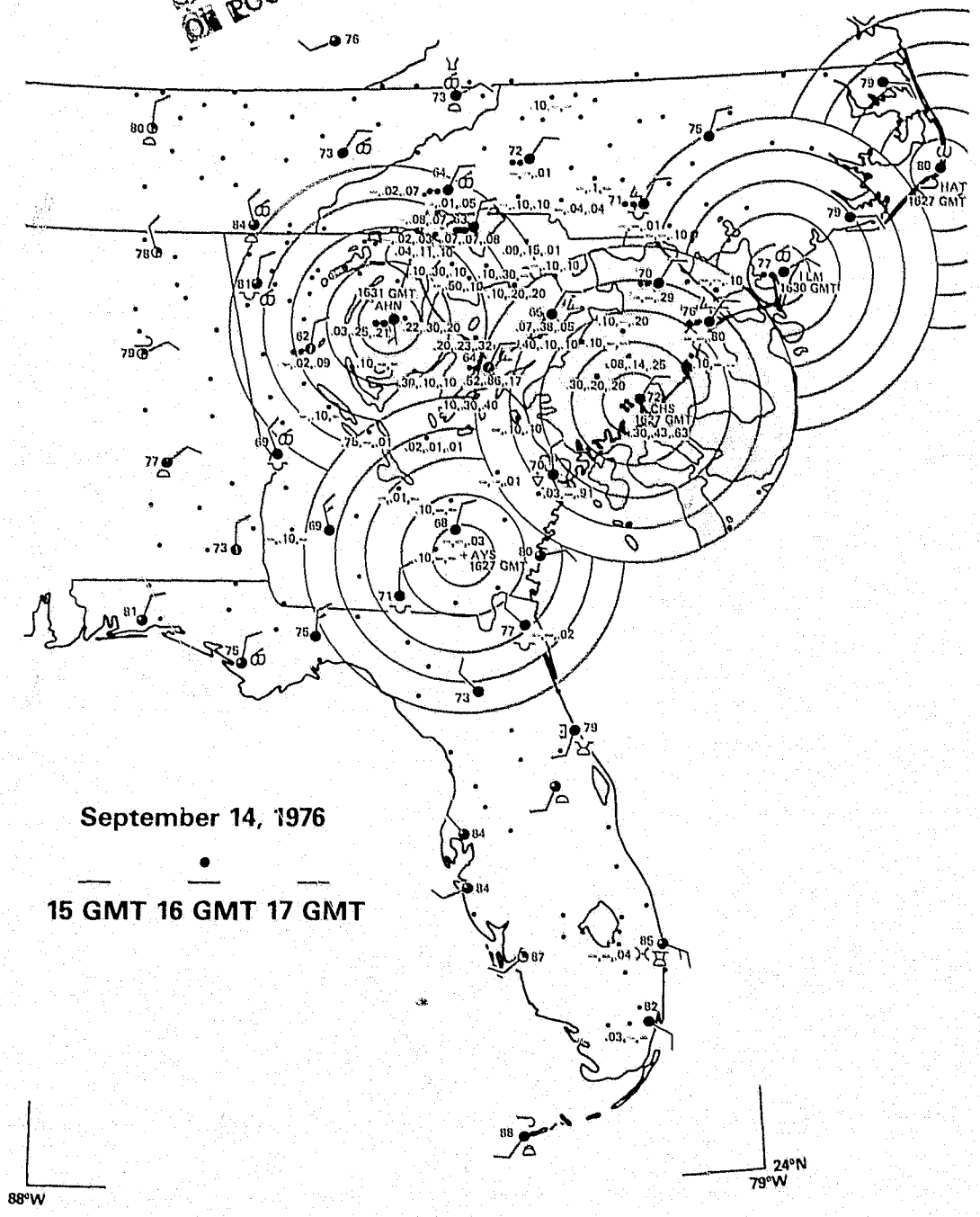


Figure 7. Rainfall over the Southeast United States as delineated by the WSR-57 radar and hourly rainfall reporting stations. Time of the data is approximately 1630 GMT, September 14, 1976. Shaded areas represent WSR-57 observed rain (rain rates $\geq$ 2.5 mm hr.). Dots represent hourly rainfall reporting stations.

delineated by the WSR-57 radars and hourly rainfall reporting stations. The approximate time of the radar PPI images was 1630 GMT (within 5 minutes of the Nimbus 6 pass). The reporting times of the hourly precipitation amounts were 1500, 1600, and 1700 GMT. The shaded area within the WSR-57 radar PPI range (232 km) is rainfall area with rain rates greater than 2.5 mm hr^{-1}. The radars were located at Waycross and Macon, Georgia; Charleston, South Carolina; and Wilmington and Cape Hatteras, North Carolina. Surface station data (present weather, temperatures, cloud type and amount, precipitation amount in three hours, and wind velocity and direction) were taken at 1800 GMT. Hourly rainfall is also shown. (See model in Figure 7.)

The Bayesian (70 percent and 80 percent confidence) classification maps are seen in Figures 8 and 9 respectively. Areas of clouds most likely producing rain are delineated by the Nimbus-6 THIR 11.5 μm channel where equivalent black body temperatures (T_{BB}) $\leq$ 270°K (Shenk et al., 1976). Rain areas in the absence of rain producing clouds are considered misclassifications. Regions only covered by clusters of contiguous pixels classified into a single individual class are shown, since the probability of misclassifying clusters is much less than that of a single pixel.

It is seen by comparing the two Bayesian classification maps at 70 percent and 80 percent confidence level (Figures 8 and 9 respectively) with the map delineating observed rain (Figure 7) that they agree well, particularly at the 80 percent confidence level. No attempt was made to verify wet land surfaces. The 80 percent confidence Bayesian classification, however, did not delineate rain over eastern South Carolina as well as the 70 percent confidence classification.

The main discrepancies found between the ESMR-6 observed rainfall and ground observed rainfall is seen over North Carolina and Southwestern Georgia. The rainfall indicated by ESMR-6 over North Carolina may be suspended liquid water in the clouds and/or

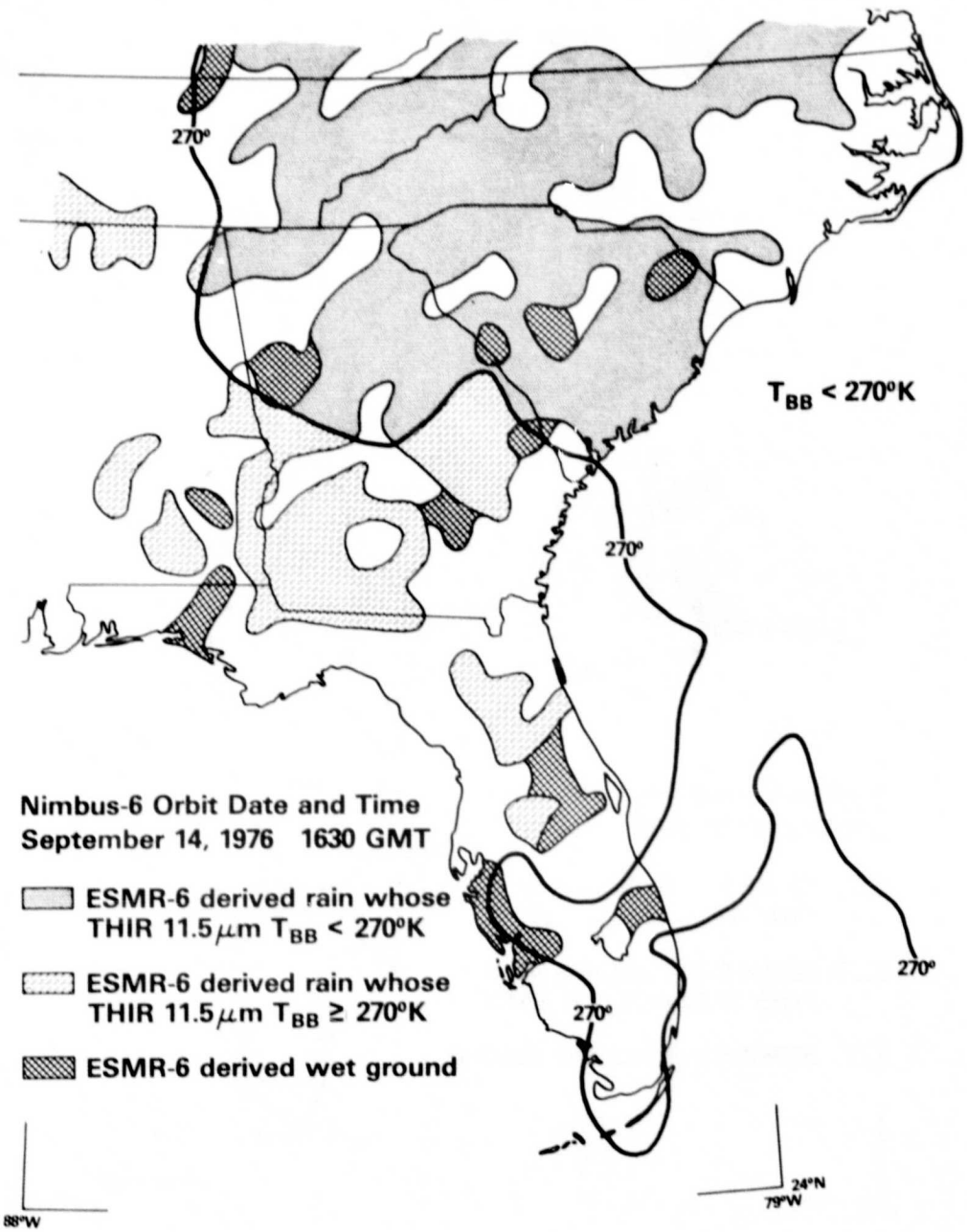


Figure 8. ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 70%. Line depicts equivalent blackbody temperatures (T_{BB}) of 270°K as measured by the Nimbus-6 THIR 11.5 channel. Areas that have $T_{BB} \leq 270°K$ represent cloud cover. Time of Nimbus-6 pass—1630 GMT, September 14, 1976.

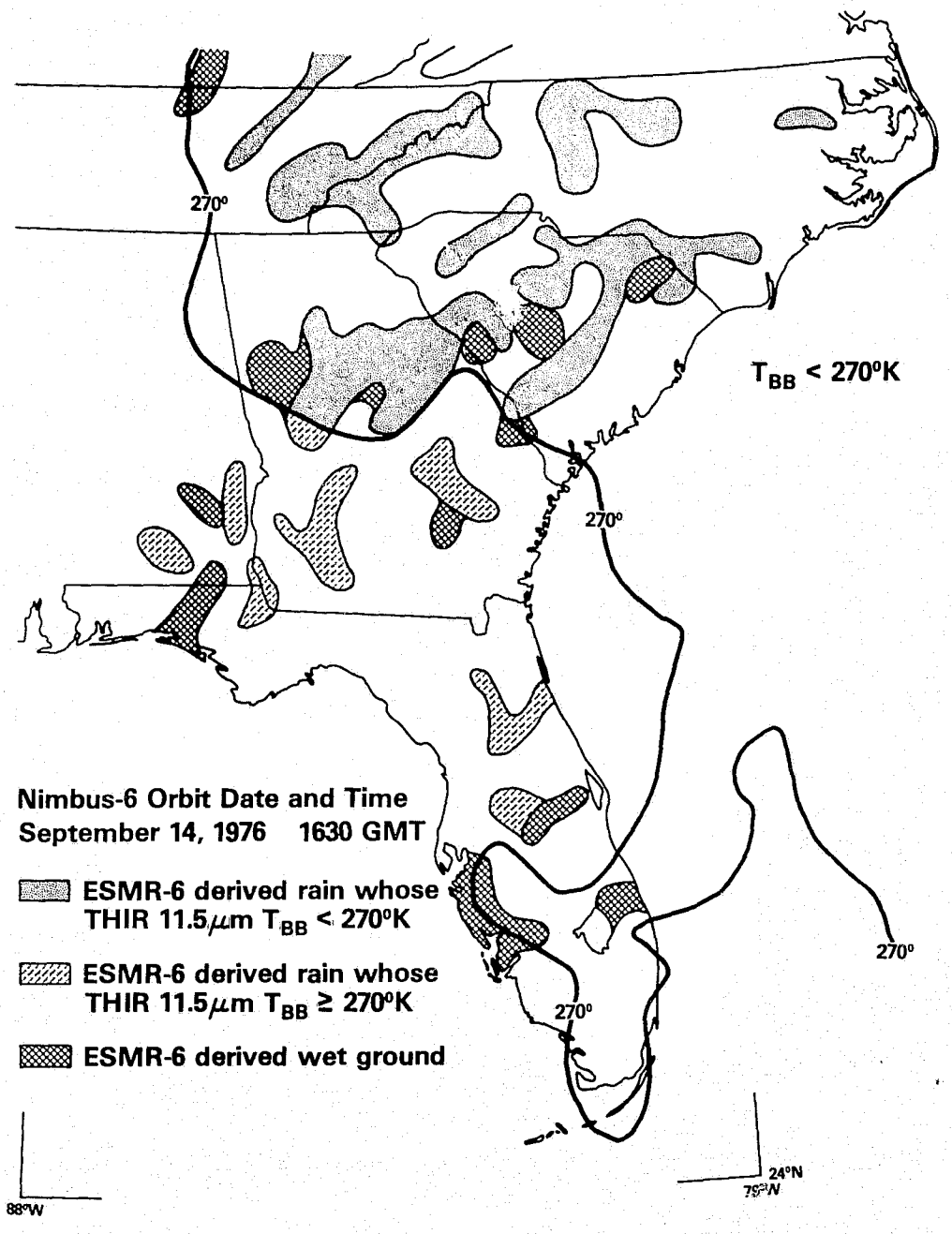


Figure 9. ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 80%. Line depicts equivalent blackbody temperature (T_{BB}) of 270°K as measured by the Nimbus-6 THIR 11.5 channel. Areas that have $T_{BB} < 270°K$ represent cloud cover. Time of Nimbus-6 pass—1630 GMT, September 14, 1976.

virga ahead of the rain area (the area of rain was moving northeastward towards North Carolina). The ESMR-6 delineated rain over Southwestern Georgia, which was upstream from the rain area, may be due to wet land surfaces produced by the rain that fell a few hours prior to the Nimbus-6 pass.

The Bayesian classification algorithm was applied to another test case (1645 GMT August 27, 1976, surface thermodynamic temperatures were $\geq 15°C$) over the same geographical area as the previous case in order to determine whether the surface characteristics (vegetation, soil moisture, and surface roughness) had influenced the classification performed in the previous case. During this period, the area was under the influence of a Bermuda high and there was only convective rainfall in the area, particularly along the Gulf States. Figure 10 shows the 80 percent confidence level Bayesian classification map superimposed over the surface station models. The reporting time for these stations was 1800 GMT. The figure shows that the only areas classified as rain over land were along the Gulf coast. The regions in the previous case where the algorithm showed rainfall were classified as dry land surfaces. Hence, there were no influences of extraneous surface characteristics on the outcome of the previous case study.

However, contradicting results occurred when the Bayesian classification algorithm was applied to a night time Nimbus-6 pass over the same geographical area (0525 GMT, September 13, 1976) where surface thermodynamic temperatures were $>15°C$ and there was no synoptic scale rainfall. Almost all pixels were classified by the algorithm as rain over land. An examination of the ESMR-6 vertically polarized T_B's showed that the temperatures were below 0°C. This anomally may be attributed to the change in the surface emissivity caused by the presence of dew on the vegetation. The 0600 GMT National Weather Service map indicated that the conditions were ideal for the formation of dew. A large anticyclone centered over Virginia produced clear skies, winds less than 5 kts and dew point temperature

ORIGINAL PAGE IS
OF POOR QUALITY

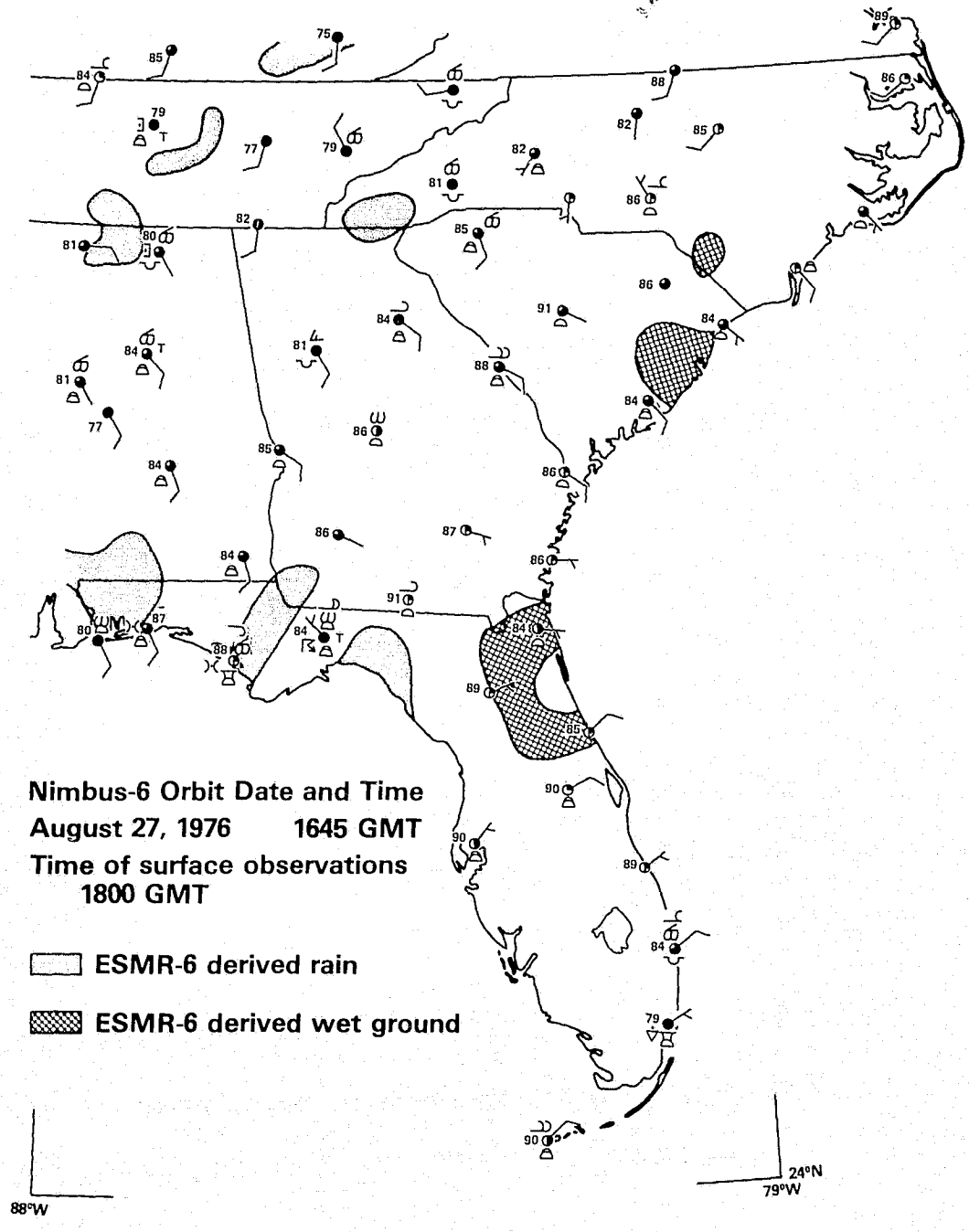


Figure 10. ESMR-6 derived rainfall distribution using the Bayesian classifier with a confidence level of 80%. Time of Nimbus-6 pass—1645 GMT, August 27, 1976.

differences of less than 3°C over the majority of the reporting stations in the Southeast United States. Therefore, the classification algorithm trained by data sampled from Nimbus-6 day time passes can be employed only when dew is absent.

8. CONCLUSION:

Statistical analyses were performed on the sampled ESMR-6 data for the purpose of detecting rainfall areas over land from dry and wet land surfaces. It was found that synoptic scale rainfall over land, where surface thermodynamic temperatures were greater than 15°C and where the vegetation was not covered with dew, could be delineated despite the large ESMR-6 IFOV. However, there was some ambiguity in distinguishing between rainfall over land and wet land surfaces.