NASA TECHNICAL MEMORANDUM

NASA TM 78184

(NASA-TM-78184) SOME OBSERVATIONS ABOUT LANDSAT DIGITAL ANALYSIS (NASA) 40 p HC A03/MF A01 CSCL 05B

N78-30634

Unclas

G3/43 28609

SOME OBSERVATIONS ABOUT LANDSAT DIGITAL ANALYSIS

By Robert R. Jayroe, Jr.
Data Systems Laboratory

August 1978

NASA

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama

1. Report No. NASA TM- 78184	2 Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Some Observations About Landsat Digital Analysis		5. Report Date August 1978 6. Performing Organization Code
7. Author(s) Robert R. Jayroe, Jr.		8. Performing Organization Report No.
9. Performing Organization Name and Address George C. Marshall Space Flight Center Marshall Space Flight Center, Alabama 35812		10. Work Unit No. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Technical Memorandum 14. Sponsoring Agency Code
15. Supplementary Notes Prepared by Data Systems Laboratory, Science and Engineering		

16. Abstract

The results reported in this study support several hypotheses concerning Landsat data that need to be verified further or disproved by examining additional data and by independent investigation performed by other Landsat data users. The hypotheses are as follows:

1) Landsat does not discriminate vegetation types, but mostly sees chlorophyl and canopy cover.

2) A majority of the features in the ground scene possess linearly proportional amounts of "color" from each spectral band.

3) The data are continuous and as a result there is no true separability of ground scene features in the data, but some features possess an excess of color in a particular band pair.

4) There are relatively few features present in the spectral data, and these do not correspond to the conventional definitions that are used.

5) Aside from seasonal effects, in a distributional sense all Landsat data are essentially the same. The only difference is the way the data are spatially arranged in the image.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Remote sensing Multiimage distributions	Unclassified-Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	40	NTIS

MSFC - Form 3292-1 (Rev December 1972)

For sale by National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Concluded)

TECHNICAL MEMORANDUM 78184

SOME OBSERVATIONS ABOUT LANDSAT DIGITAL ANALYSIS

I. INTRODUCTION

A considerable amount of time, money, and effort has been spent on computer analysis of Landsat image data for various applications such as agriculture, land use, water quality, and geology. A major tool of these efforts has been the classification program (of which there are many variations) that uses training areas as an aid for extracting various types of ground scene features such as wheat, corn, forest, urban areas, water, etc. Thus, not only have the data been used for a variety of applications, but within each application there is usually a broad range of information which is of interest to a particular user in terms of various types of inventories and maps.

Another type of inventory, which is equally important, is an inventory of observations concerning the data and results, as well as how the data are being analyzed. The first paragraph provides an important question concerning the balance between information content in the data versus the various types of information that users are attempting to extract for a myriad of applications. The number of applications tends to give the impression that the information content may be limitless. However, one indication of information content can be obtained by examining the success experienced by various investigators for selected applications.[1] From a statistical universe of 224 approved Landsat investigations, the average reported crop inventory accuracy in the agricultural discipline was 74 percent, while the average accuracy of the crop classification map was 63 percent. Generally, map accuracies tend to be 10 to 20 percent lower than inventory accuracies, because some of the error in the inventories tend to be random and cancel. For an inventory it is important for the proportion of crop type to be correct, whereas in a map, each individual picture

1. Data provided by Dr. Peter A. Castruccio, President of ECO Systems International Inc., P.O. Box 225, Gambrils, Maryland 21054.

element needs to be correct. The average mapping performances reported by investigators in the land use discipline for the categories of urban, crop land, forest, and water are presented in Table 1. The main observations are that the accuracies are not as high as desired and that two additional factors (season that the data were acquired and inventory area) need to be considered. The importance of seasonal effects and test site size are discussed in later sections.

TABLE 1. AVERAGE LAND USE MAPPING ACCURACIES

Category	Accuracy (%)
Urban	30
Crop Land	55
Forest	75
Water	86

Another indication of information content is provided by the way the imagery is analyzed and by considering how Landsat provides information. Basically, the two methods that Landsat uses to provide information are (1) through the spectral distribution of the multiband data and (2) through the spatial distribution (in image form) of the multispectral data. The most important step in the computer analysis is to make an image of the digital data so that data representing selected categories (wheat, corn, water, forest, etc.) can be located and used for training the decision process of a classifier to perform an inventory or produce a map of those selected categories. If an investigator is denied the opportunity of making an image of the data and forced to extract whatever information is possible from the spectral distributions, then the investigator is usually at a loss. From necessity, there usually has to be much overt, and sometimes covert, photointerpretation applied to the machine processing of digital imagery. Thus, there is not only a question of information content, but there is also a question of where the information is located. The majority of the information could be simply in the way the data are arranged and percieved in picture form rather than in the spectral distributions, which is the case with a black and white image.

Because photointerpretation is a well established information extraction process, investigators have tended to start with this process and assume that there is a one-to-one correspondence with the information contained in the spectral distributions. Since this does not appear to be the case, there is a need to determine and relate what information can and cannot be obtained from either type of distribution. In this report, the approach of starting with the spectral distribution information was used in an attempt to relate the spectral information to the spatial information as a check on the information correspondence.

II. TEST SITE DESCRIPTIONS

Two different test areas were examined using seasonal data. One test area was a Large Area Crop Inventory Experiment (LACIE) supersite in Finney County, Kansas that was 196 picture elements wide in the east-west direction and 117 picture elements long in the north-south direction. According to the ground truth information, winter wheat was planted in September 1975 and harvested sometime during or after June 1976, while other crops such as soybean, corn, and sorghum were planted in June 1976. The second test area was a region of southwest Alabama west of Mobile Bay. The size of this test site was 1000 by 1000 picture elements. The supersite in almost pure vegetation, while the Alabama test site contains a great variety of ground scene features.

Figure 1 is a color composite of the LACIE supersite. The apparent blurriness of the image is due to the small size of the site and the resulting magnification that was used to make the photograph. Each picture element was replicated 400 times in a 20 by 20 array. Figure 2 is a color composite of the Mobile Bay area, and in this case each picture element was replicated 4 times in a 2 by 2 array.

III. ANALYSIS OF SUPERSITE DATA

Figure 3 shows scatter diagrams of multispectral scanner (MSS) band 4 (0.5 to 0.6 μ) versus MSS band 6 (0.7 to 0.8 μ) in monthly order to illustrate the planting-harvesting cycle. The year associated with the data is not important provided that the farming practices do not change. The figure shows that

Figure 1. LACIE supersite color composite.

ORIGINAL PAGE IS
OF POOR QUALITY

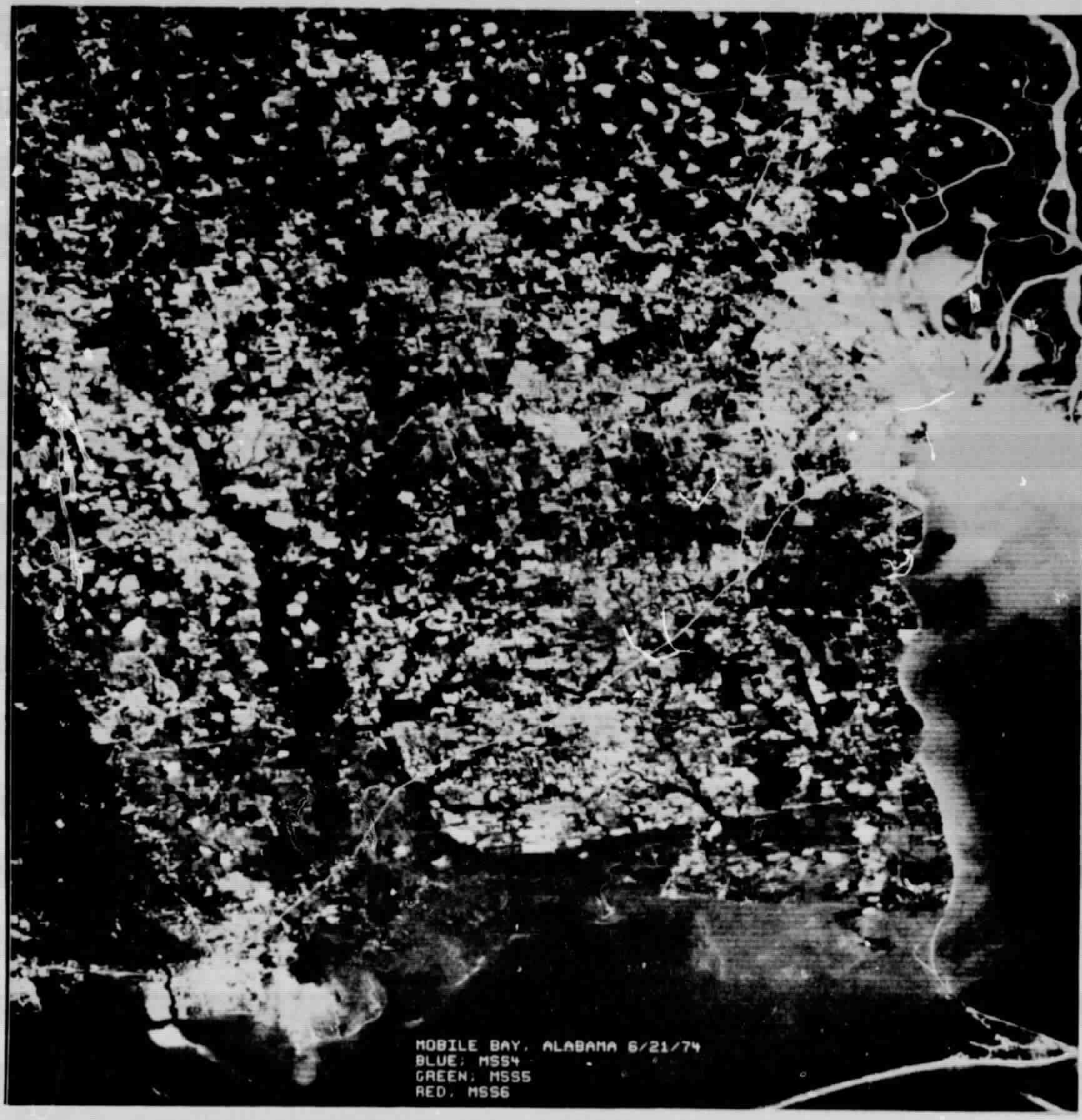


Figure 2. Mobile Bay color composite.

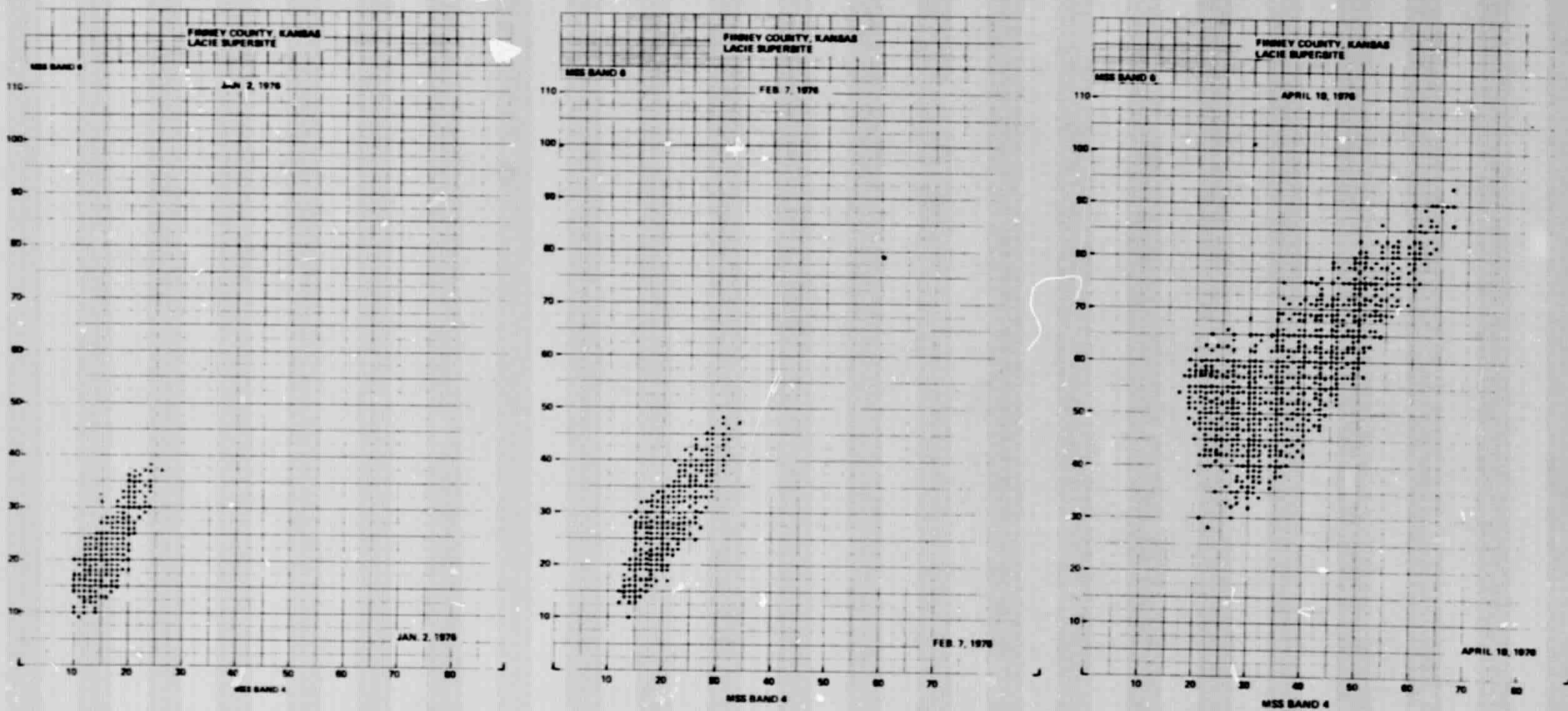


Figure 3. Supersite seasonal scatter diagrams of MSS band 4 versus MSS band 6.

ORIGINAL PAGE IS
OF POOR QUALITY

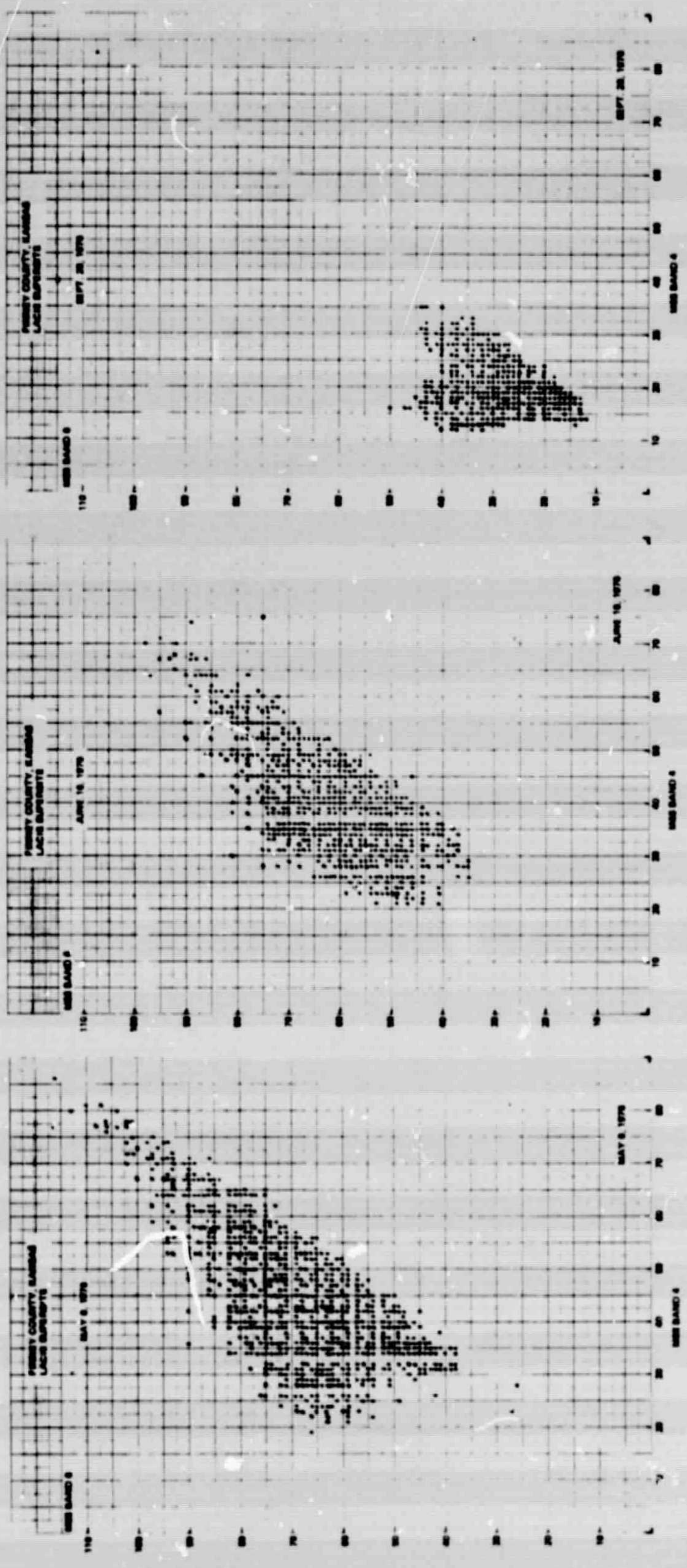


Figure 3. (Continued).

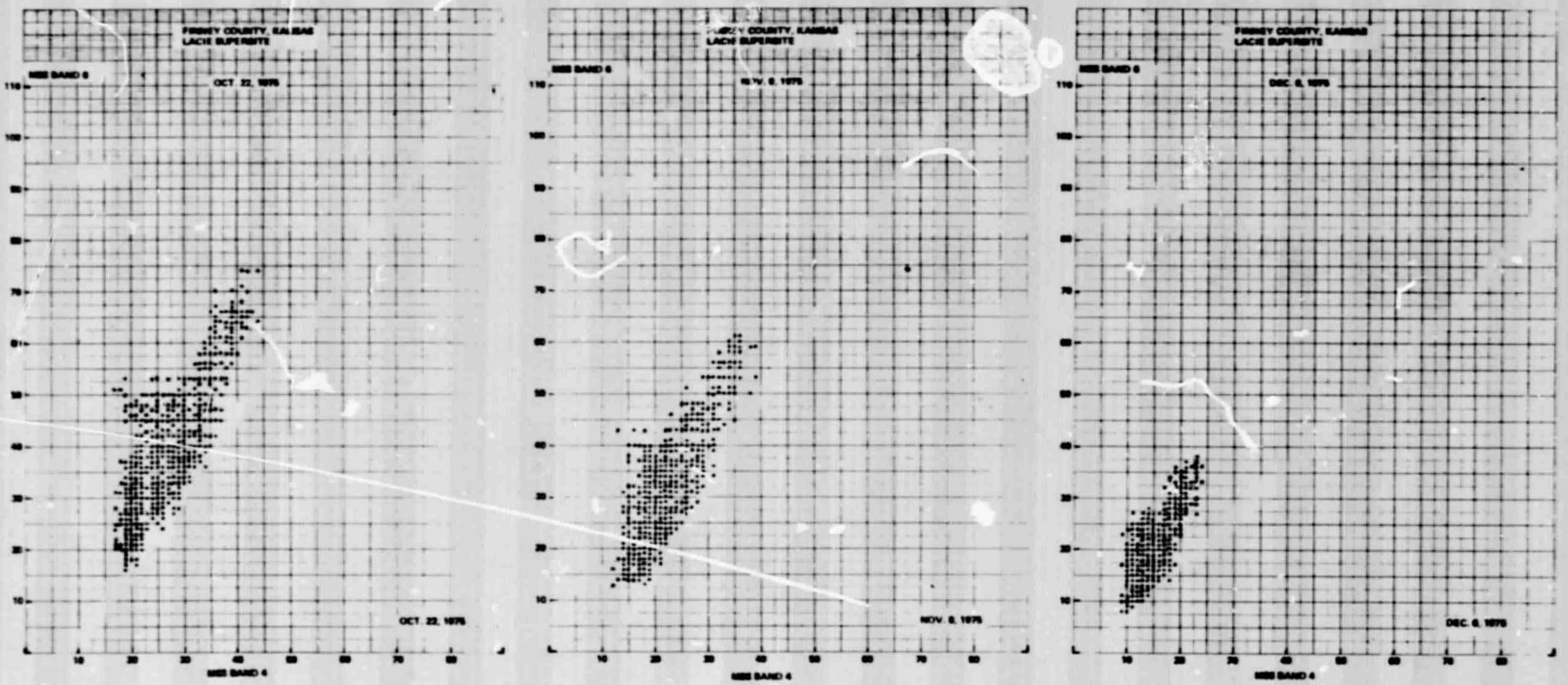


Figure 3. (Concluded).

during the winter, the distributions are compact and close to the origin, and that during the peak of the growing season the distributions move away from the origin and considerably increase in extent. This observation closely parallels the comment that is occasionally heard from the user community concerning the dimness of their photographic products and that the data have to be stretched before the product is useful. It is highly probable that these comments were generated from data acquired during the winter season. It should be recognized that if the sensor were adjusted to increase the dynamic range of the winter data, the result would be that the peak growing season data would saturate the sensor. Figure 3 also shows that sometime between June and September 1976, harvesting and planting took place because the distribution shrank in extent and moved toward the origin in September. but then moved away from the origin again and increased in extent in October. After that, the distributions start shrinking and moving toward the origin.

Figure 4 shows a monthly composite of the scatter diagrams. The most striking feature of the composite is the presence of a minimum or maximum reflection line. That is, for a given amount of reflection in band 4 there is a minimum of reflection found in band 6, or for a given amount of reflection in band 6 there is a maximum amount of reflection found in band 4. The equation of this line is given by MSS band 6 = $1.4 \times$ MSS band 4.

Two approaches can be taken which are helpful in interpreting the meaning of this line. The first approach is to collapse the data in band 6 on this line by replacing the band 6 data values with 1.4 times the band 4 data values. In this case, the modified band 6 will be lineraly dependent on band 4. A color composite can be made by photographing band 4 in blue, band 5 in green, and the modified band 6 in red. This color composite should approximate a black and white image. The result is shown in Figure 5. The reasonableness of the black and white image approximation can be established by examining the scatter diagrams of MSS band 4 versus MSS band 5 (0.6 to $0.7\ \mu$) shown in Figure 6. These distributions show the same monthly characteristics as those shown in Figure 3, plus they exhibit a high degree of linear dependence. The only thing that has been accomplished in the color composite of Figure 5 is that the excess red contained in the color composite of Figure 1 has been removed. It is interesting to notice that the fields which contain an excess of red also contain very little of any other color. This tends to indicate that, except for the excess red, a black and white image could essentially supply the same information.

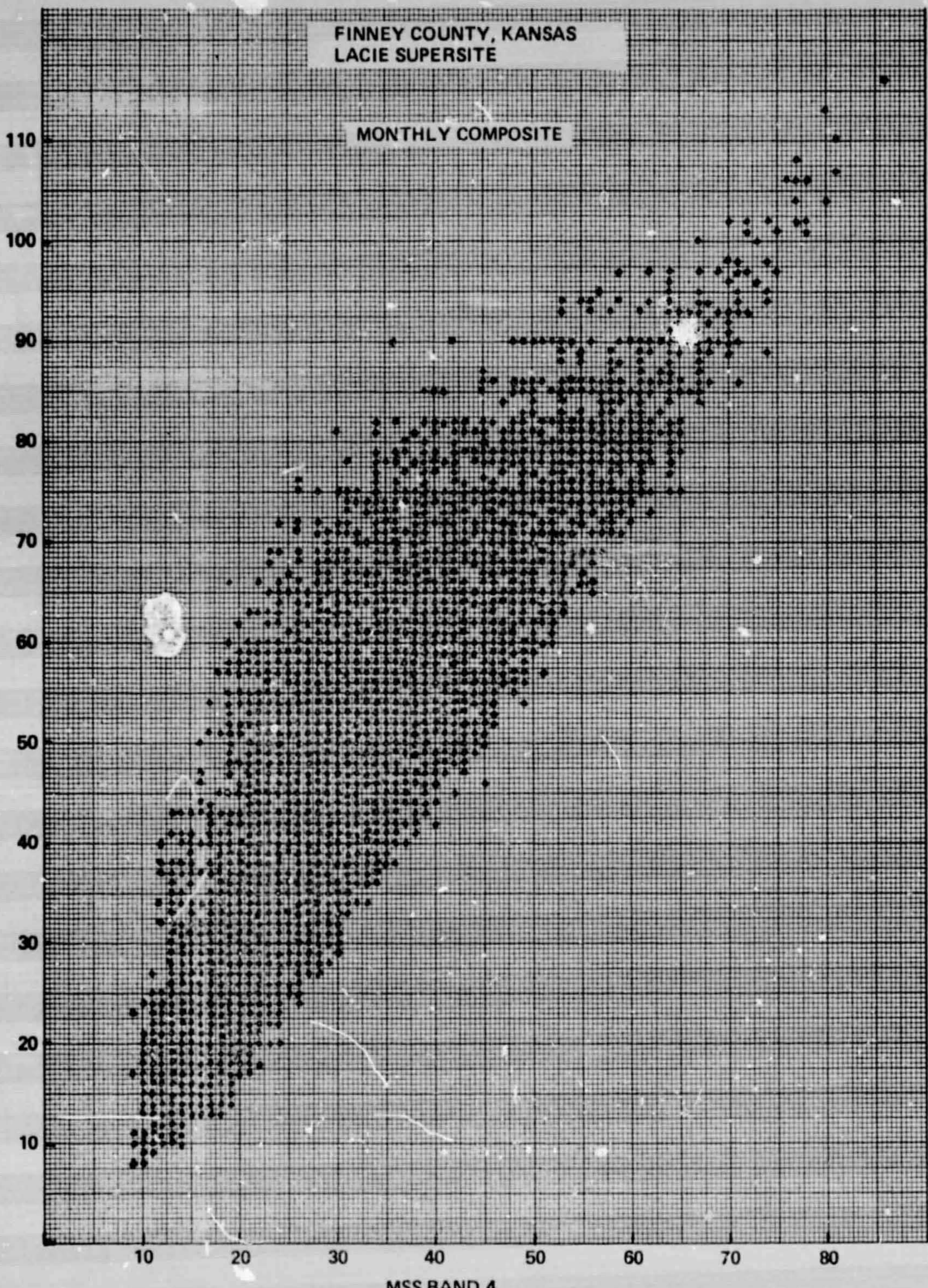


Figure 4. Supersite monthly composite scatter diagram of MSS band 4 versus MSS band 6.

Figure 5. Supersite black and white image approximation.

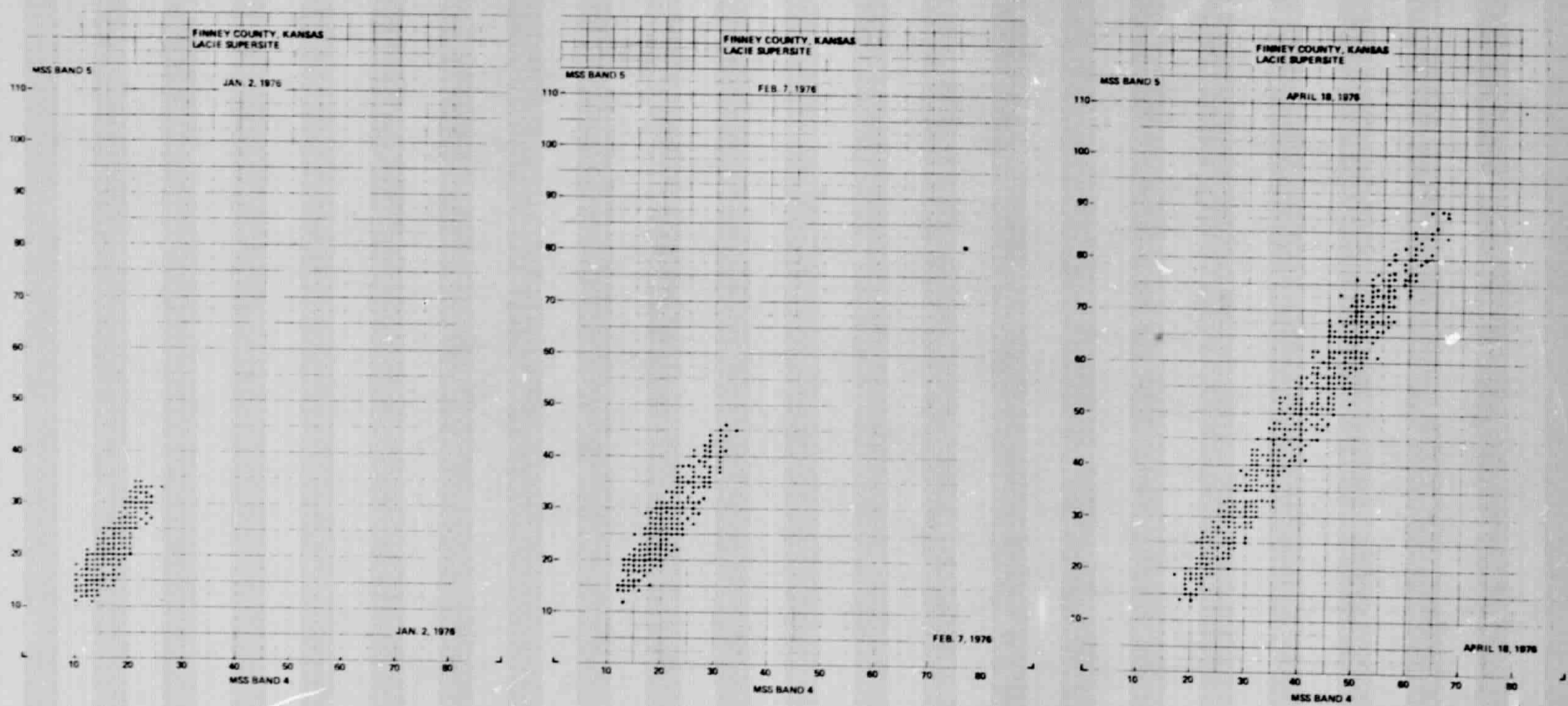


Figure 6. Supersite seasonal scatter diagrams of MSS band 4 versus MSS band 5.

ORIGINAL PAGE IS
OF POOR QUALITY

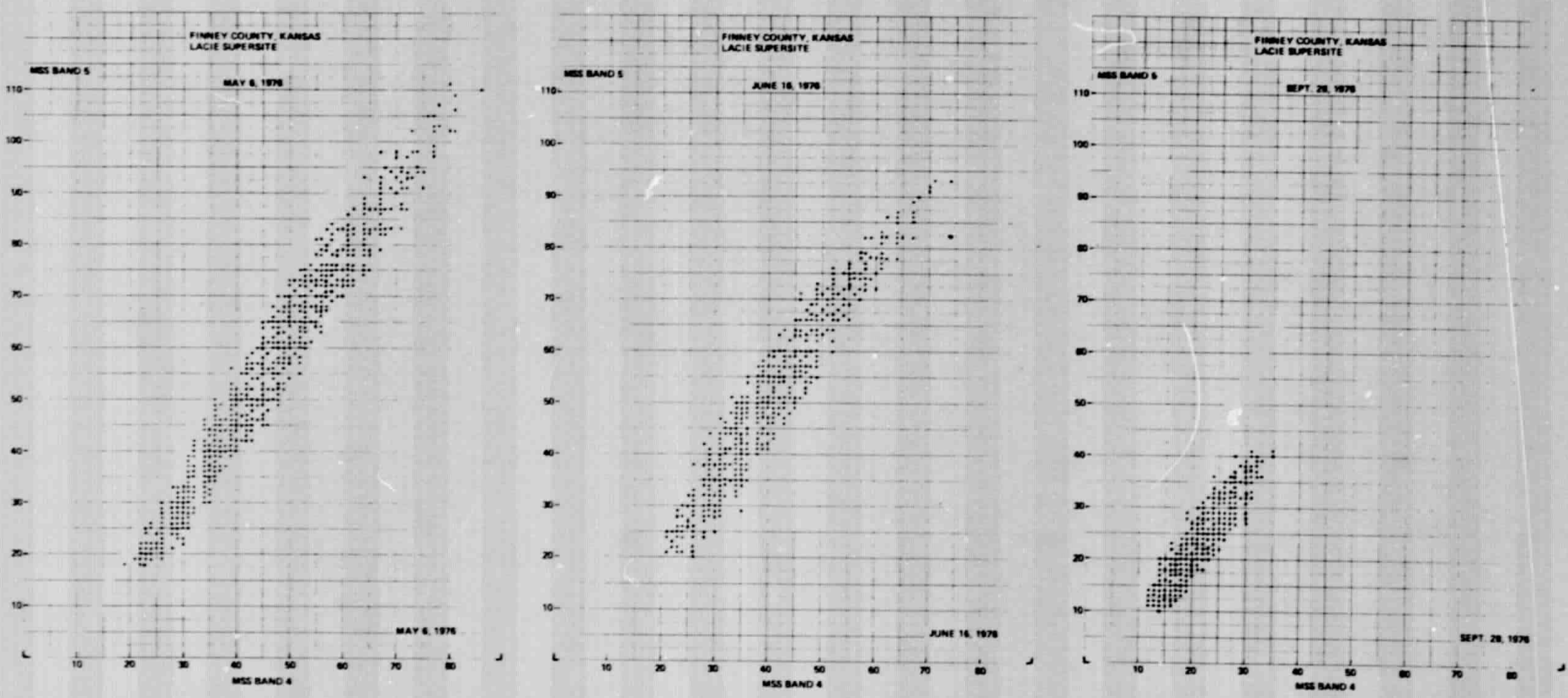


Figure 6. (Continued).

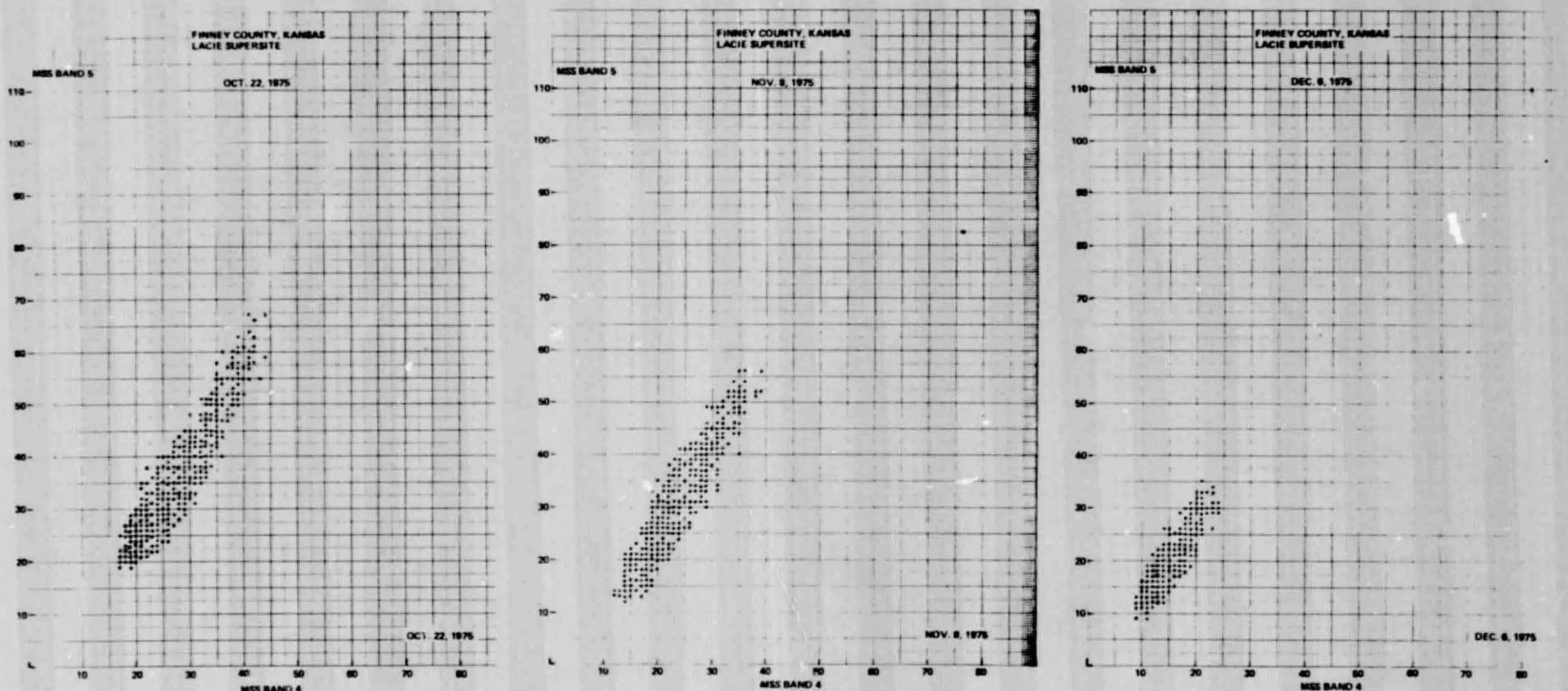


Figure 6. (Concluded).

The second approach is to examine the band 6 image after subtracting 1.4 times the band 4 data from the band 6 data. The resulting single band black and white image is shown in Figure 7, the white areas correspond to the excess red areas of Figure 1. Figure 7 should also be compared with the ground truth map in Figure 8 that shows the location of the wheat fields (also indicated in white) that were planted in September 1975. The ground truth map was converted to the coordinate system of the Landsat data and, as a result, does not cover the entire portion of the image. However, the portion of the Landsat image that is covered by the ground truth map indicates excellent agreement with the wheat fields. It would appear that to extract wheat information, it is not necessary to use training areas and a classification program, but instead a simple image-difference image and unsuperivsed density slicing could do just as well. However, this is not the case.

Figure 9 shows the modified band 6 for the supersite during April, May, June, and September, and in the annotation C3 is MSS band 6 and C1 is MSS band 4. These images show the wheat fields increasing in brightness through May, decreasing in brightness in June, and practically disappearing in September. However, fields identified as non-wheat (corn, sorghum, soybean) increase with a similar brightnesses in September. Thus, when different crop types are growing at the same time, it is extremely difficult to distinguish them because of the continuous nature of the data and the linear dependences of the spectral bands. The results suggest that what is being observed are chlorophyl and canopy cover, and that the reflection line discussed is in actuality a chlorophyl line. The results also suggest that high classification accuracies can be achieved, provided a small test site is chosen at the right time to eliminate confusing features.

IV. ANALYSIS OF MOBILE BAY DATA

If the results discussed in Section III are valid, then the same approach should work for another test site. The supersite area can be characterized as being relatively small and as containing almost pure vegetation. In contrast, the Mobile Bay area is much larger in size, and contains a large variety of vegetation and other ground scene features.

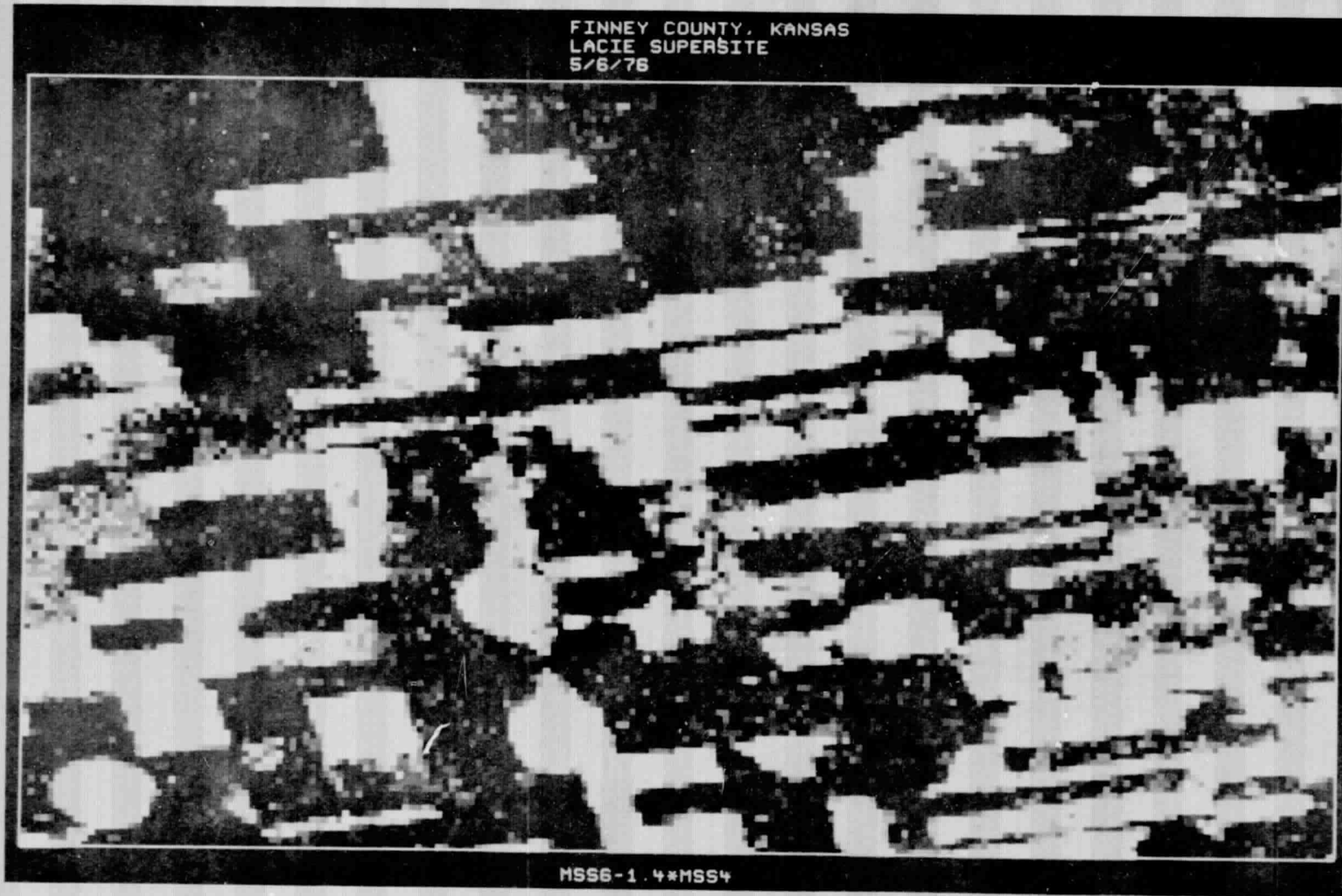


Figure 7. Wheat enhanced MSS band 6.

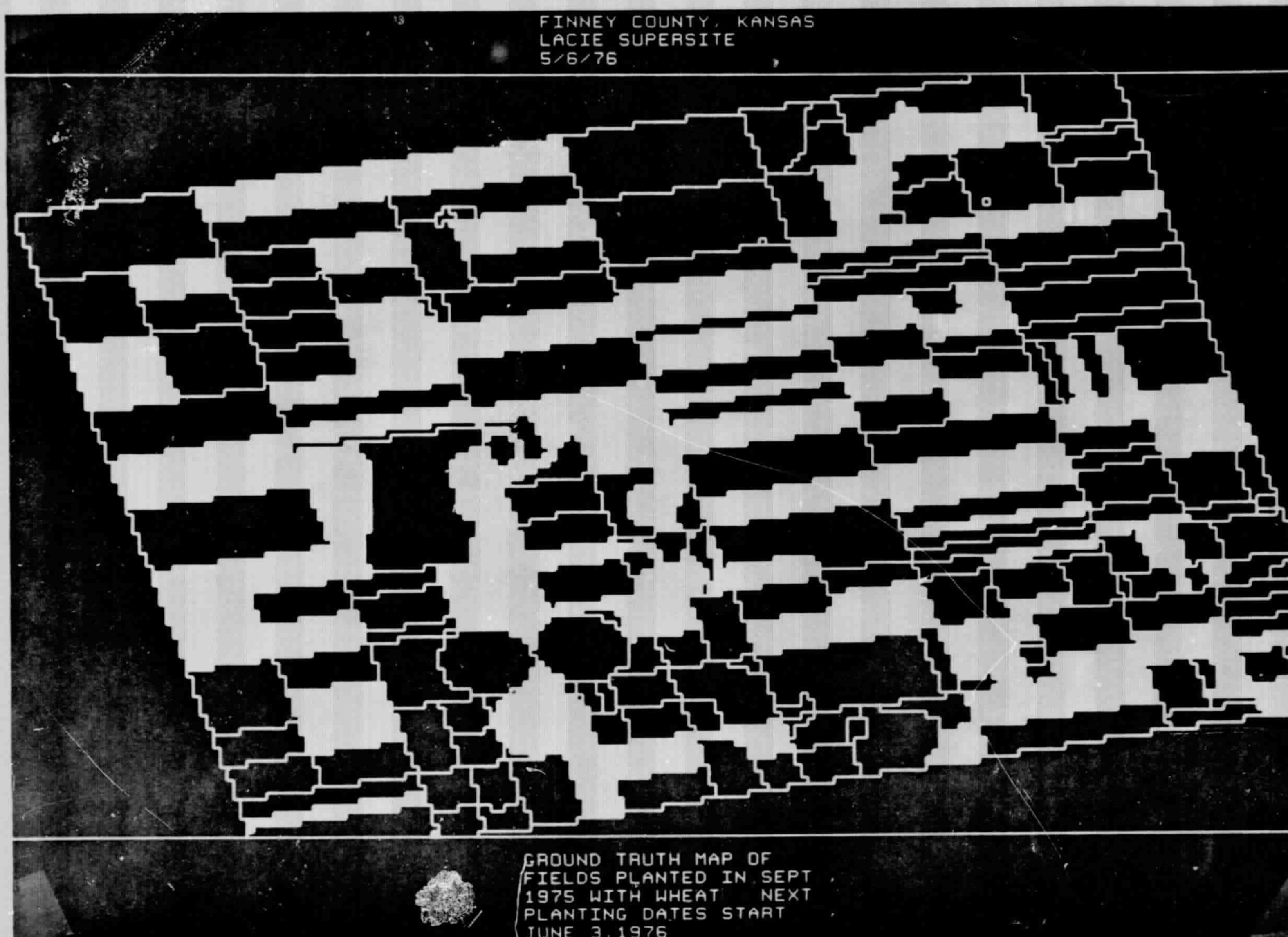


Figure 8. Supersite ground truth map.

ORIGINAL PAGE IS
OF POOR QUALITY

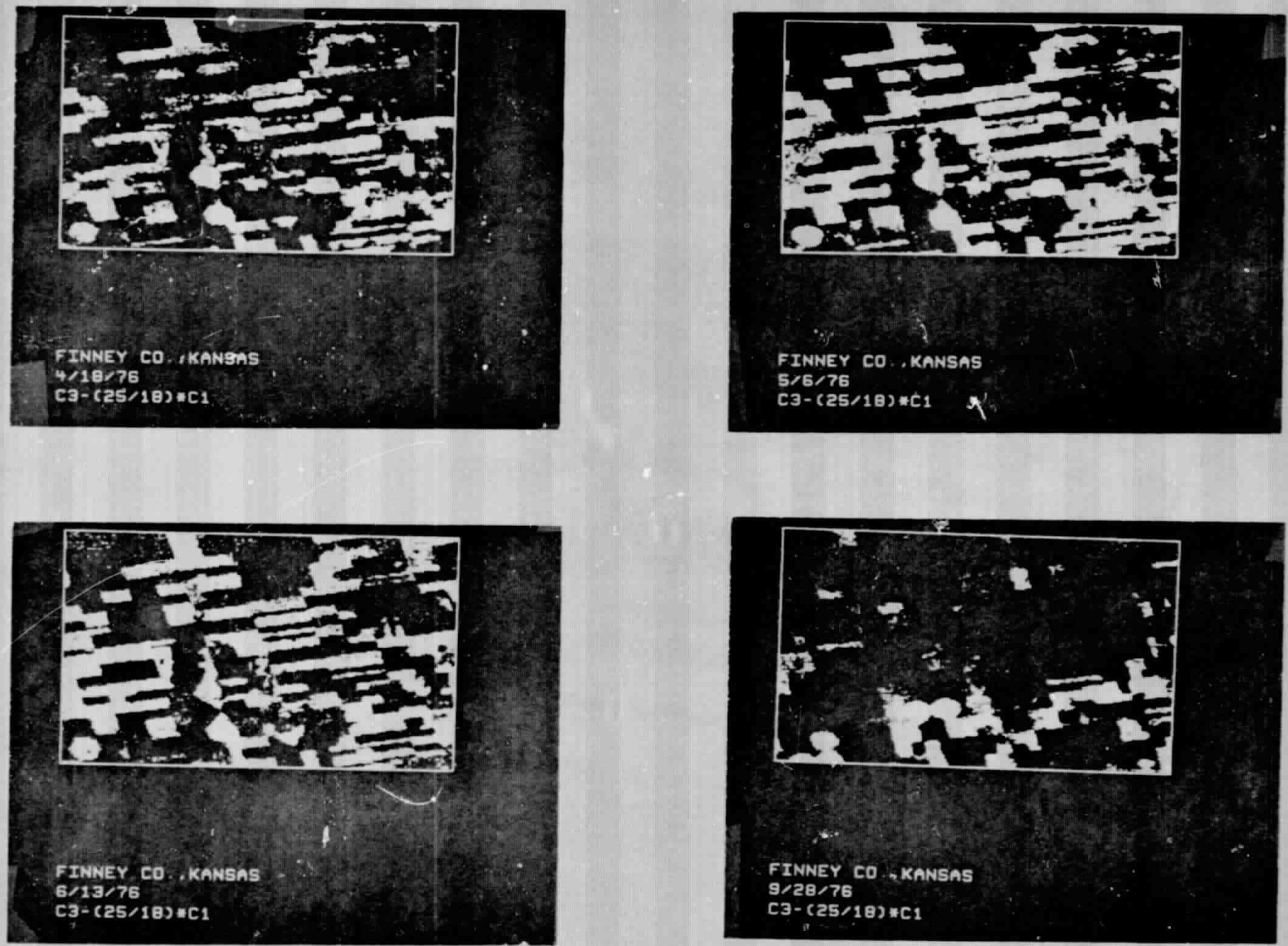


Figure 9. Monthly observations of vegetation enhanced MSS band 6.

Figure 10 shows the scatter diagrams of MSS band 4 versus MSS band 6 for four different dates. All of the data is cloud-free except for the June, 1974 data (as can be seen in the color composite image of Fig. 2). If transparencies are made of the scatter diagrams and they are overlaid, there is uncertainty in the location of the so-called chlorophyl line, even though apparent excesses of red (band 6) exist. To locate a chlorophyl line, photointerpretation was applied to the image in Figure 2 to locate a cloud-free and predominantly vegetated area. A 200 by 200 picture element area was selected which centered on Interstate 10 midway between Mobile, Alabama (upper right corner), and Pascagoula, Mississippi (lower left corner). This area is cloud-free for the June pass and predominantly vegetative, although it may contain a small portion of the highway and possibly a small amount of water. Figure 11 shows the scatter diagrams for this area, indicating that it has seasonal properties very similar to the supersite data area. To visualize the effect of clouds on the distributions, this same area was examined for an October, 1972, pass when it was approximately 50 percent cloud-covered (Fig. 12). It is interesting to notice that the effects of clouds appear linearly all along the distribution. If Figure 12 is compared with the scatter diagrams in Figure 10, there is the suggestion that many different things in the ground scene may appear linearly all along the distribution. Figure 13 is a monthly composite of Figure 11 and indicates the presence of a chlorophyl line. The line is not precise, however, because the 200 by 200 picture element area contains some data that are not vegetation. The chlorophyl line, however, does become more apparent when Figure 13 is overlaid with the November, January, April, and June data in Figure 10. If the chlorophyl line is used for removing all data above it in Figure 10, the data that are left lie in a linearly dependent band. Literally, everything in the world that does not contain chlorophyl appears forced to lie in this linearly dependent band. It is interesting to observe that if Figure 4 from the supersite data is overlaid with the Mobile data of Figure 13, there is very little difference. It is also interesting to observe that it does not appear to matter where the band 6 data are located in relation to band 4 data, but rather how far the band 6 data are located from the "so-called" chlorophyl line. Again, this strongly suggests that what an investigator sees in the data is chlorophyl and canopy cover rather than vegetative species.

V. COLOR EXPERIMENT

A simple experiment was devised to determine if linear combinations of bands 4, 5, and 6 could provide a means of automatic enhancement of the imagery for interpretation. Figure 14 shows a scatter diagram of band 4 versus band 5,

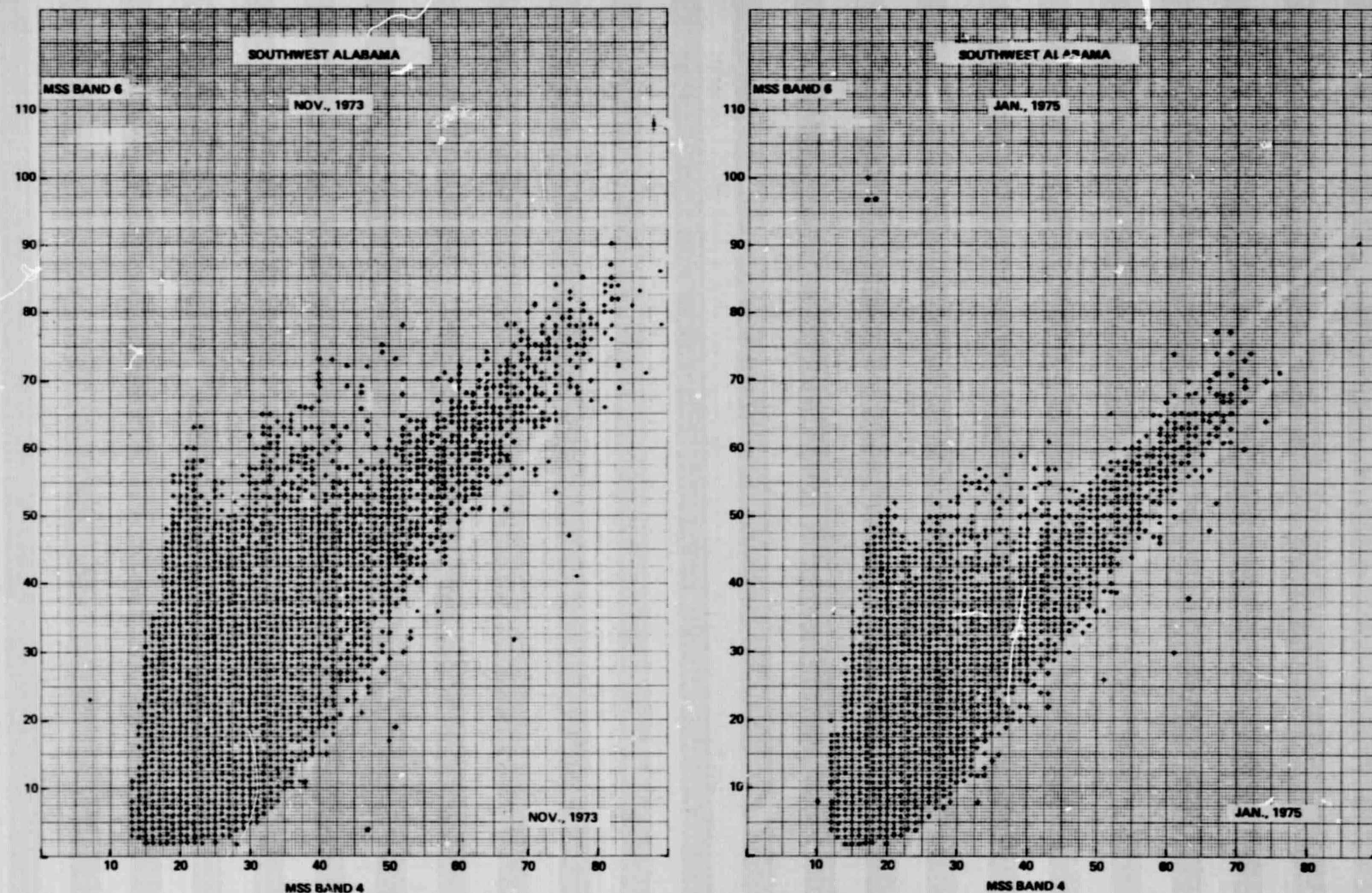


Figure 10. Mobile Bay seasonal scatter diagrams of MSS band 4 versus MSS band 6.

ORIGINAL PAGE IS
OF POOR QUALITY

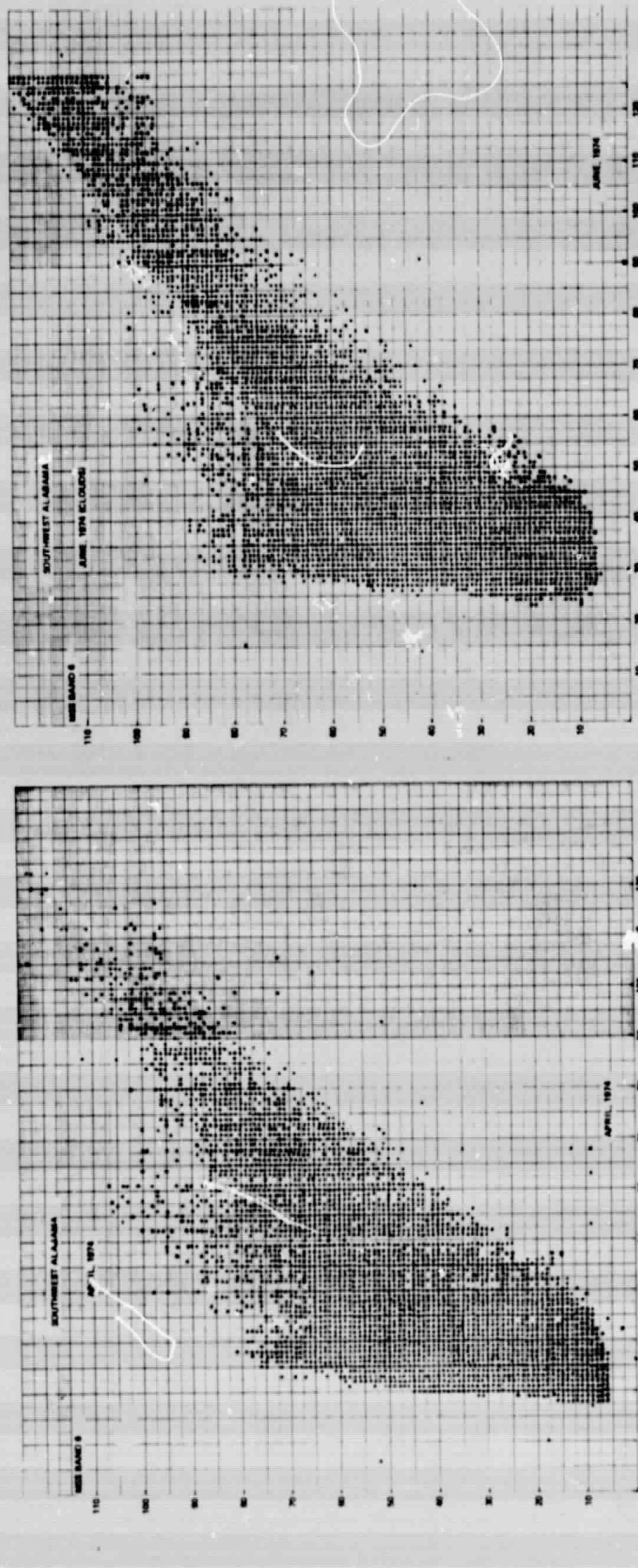


Figure 10. (Concluded).

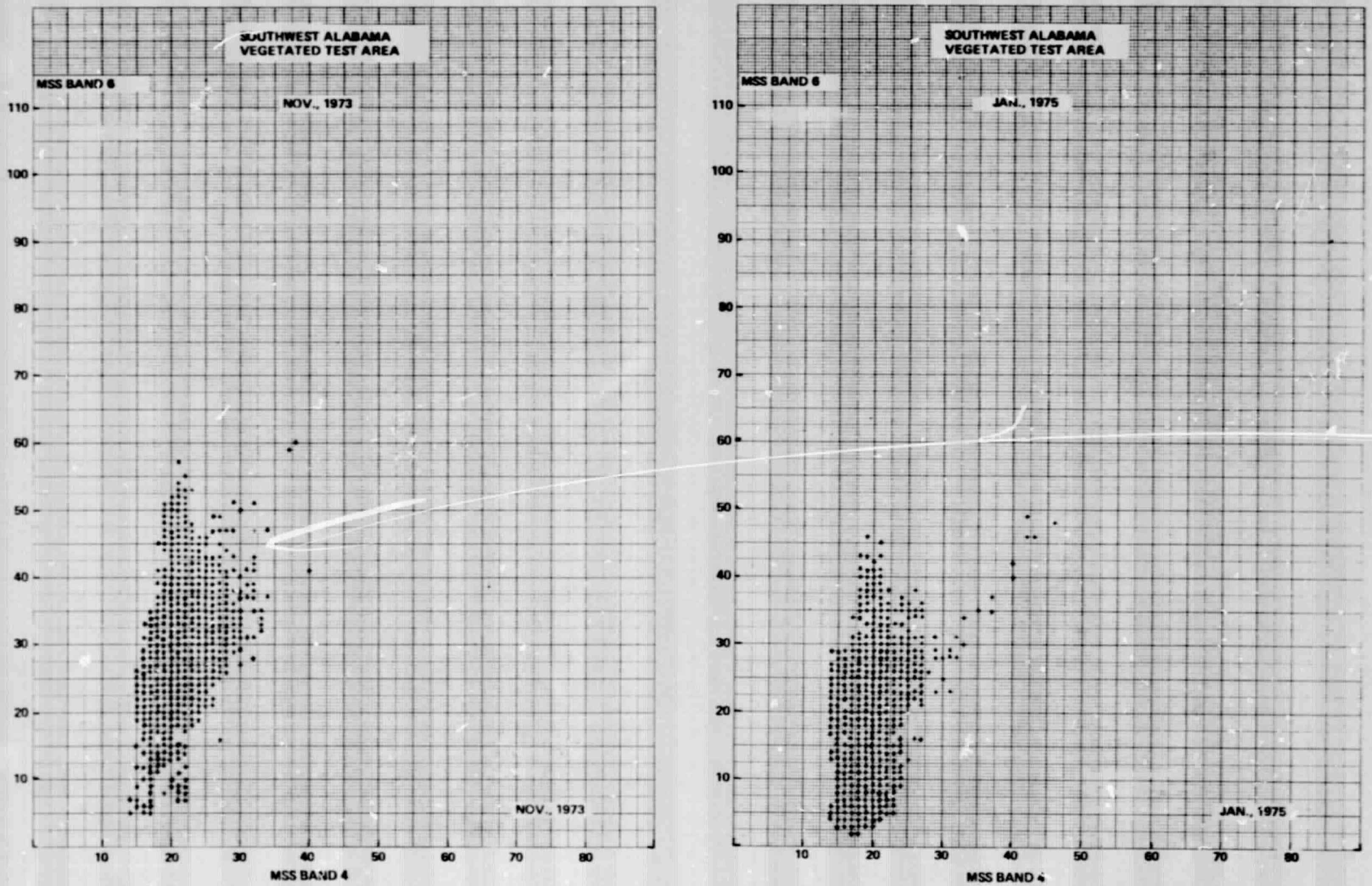
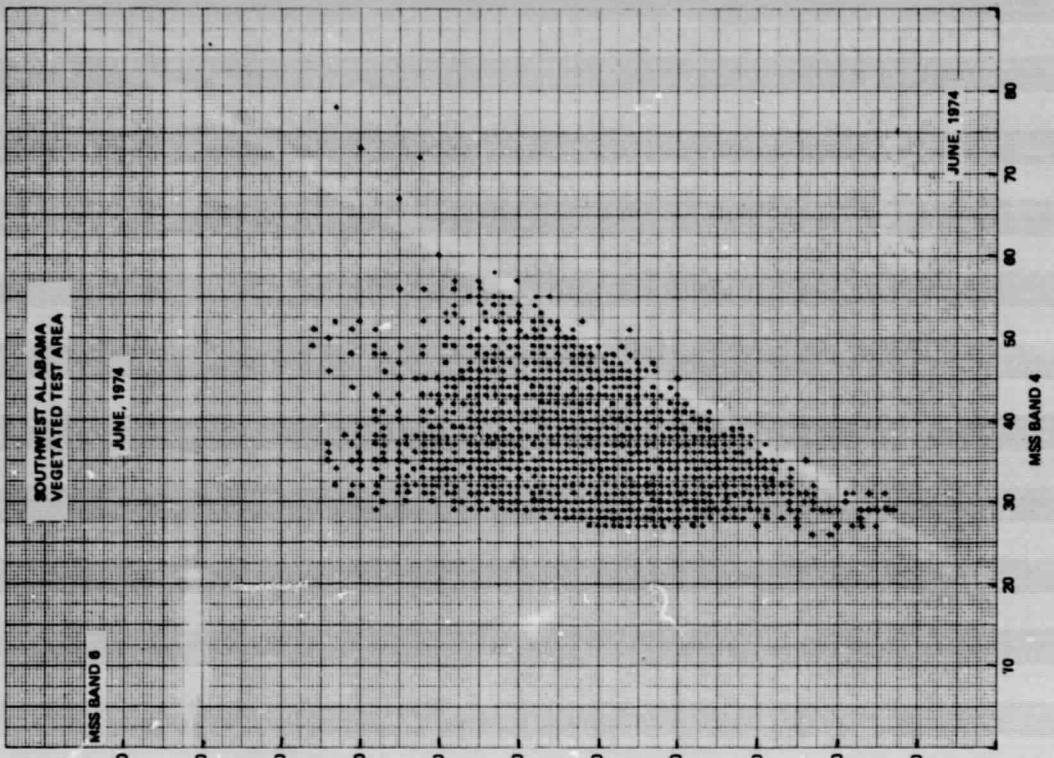


Figure 11. Vegetated area seasonal scatter diagrams of MSS band 4 versus MSS band 6.



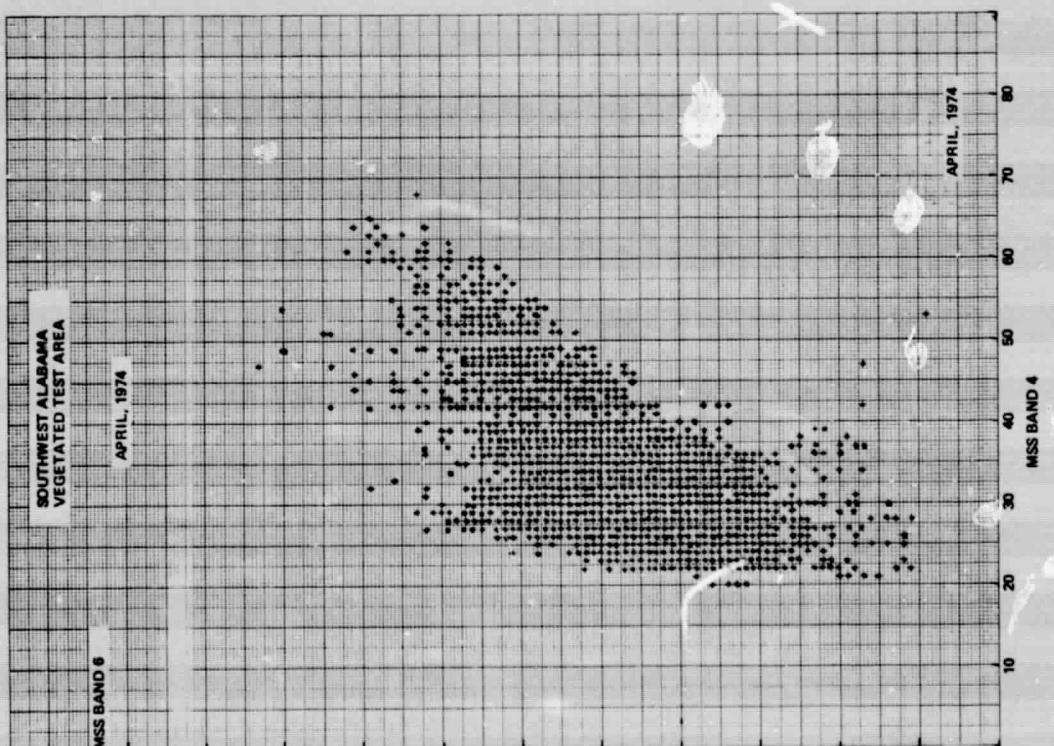


Figure 11. (Concluded).

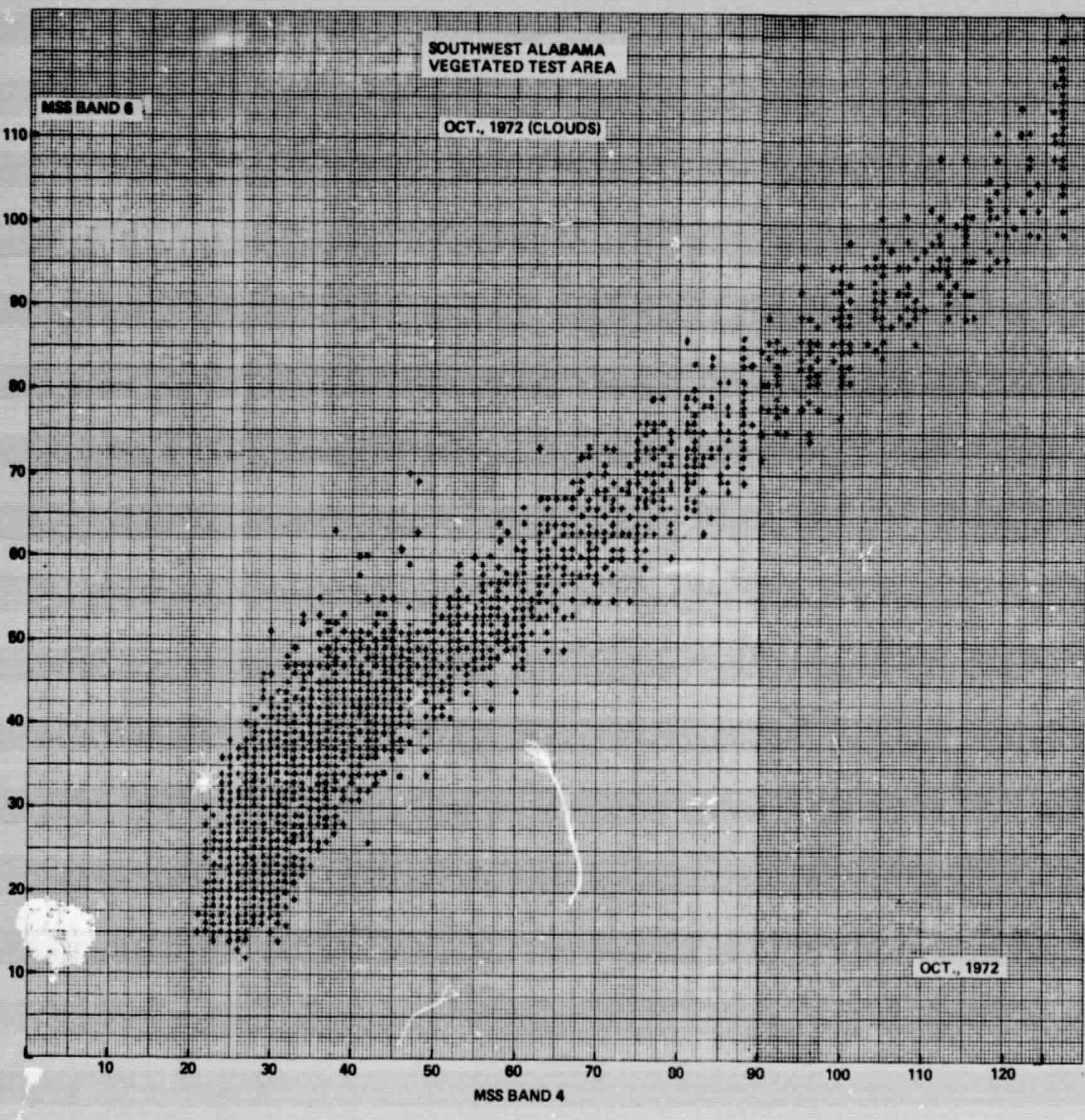


Figure 12. Scatter diagram of MSS band 4 versus MSS band 6 for vegetated area with 50 percent cloud cover.

ORIGINAL PAGE IS
OF POOR QUALITY

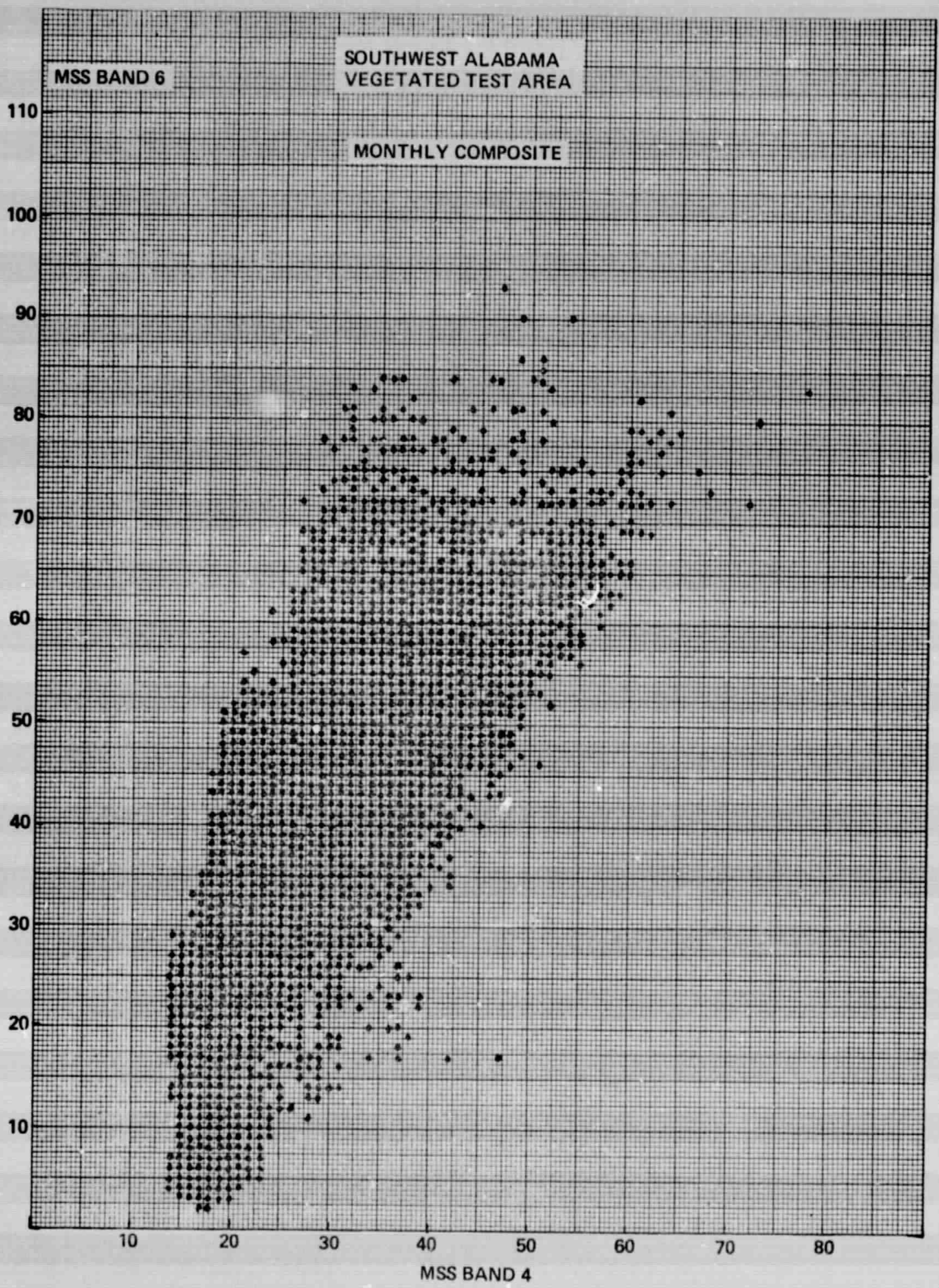


Figure 13. Mobile Bay seasonal composite scatter diagram of MSS band 4 versus MSS band 6.

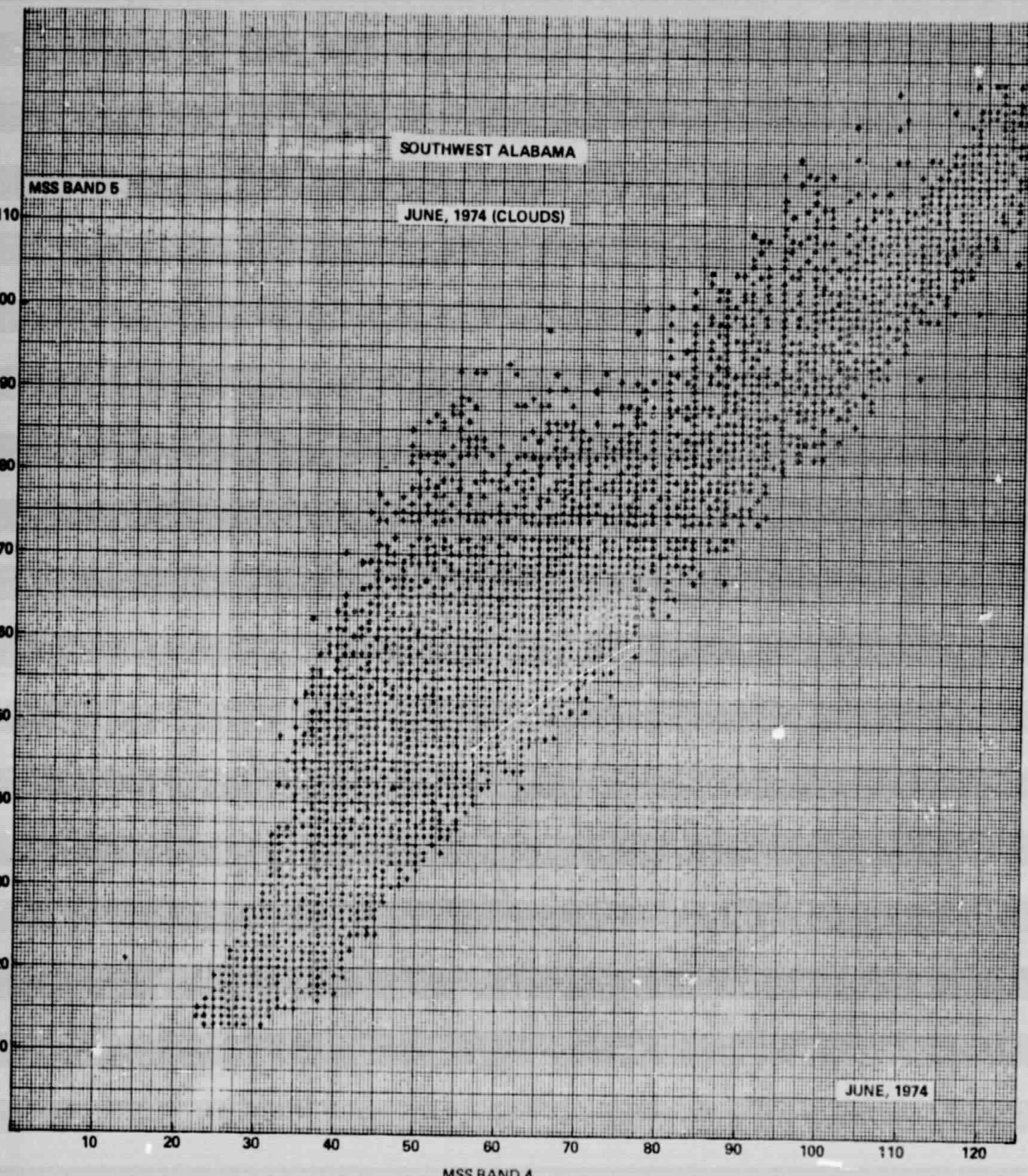


Figure 14. Mobile Bay scatter diagram of MSS band 4 versus MSS band 5.

and again a minimum/maximum reflection line is observed whose equation is given by

$$\text{MSS band } 5 = 1.2 \times \text{MSS band } 4 - 23.6 \quad .$$

This line was subtracted from the band 5 data, and the resulting image is shown in Figure 15. The chlorophyl line was subtracted from band 6, and the resulting image is shown in Figure 16. Figures 15 and 16 appear to be negatives of one another; i.e., Figure 15 intensifies areas that show exposed soil and roadways while Figure 16 intensifies vegetated areas. Next, a color composite image was made by photographing band 4 in blue, the data of Figure 15 in green, and the data of Figure 16 in red. The resulting color image shown in Figure 17 is the same as the color image in Figure 2, except the red and green data that are linearly dependent on the blue data have been removed. An interpretation of Figure 17 can be aided by examining the scatter diagrams of band 4 versus the modified bands 5 and 6 (shown in Figures 18 and 19, respectively), and correlating the results with Figure 2.

If Figures 18 and 19 are overlaid, the region of data overlap tends to indicate the presence of vegetation. It is interesting to note that two colors (red and green) correspond to vegetation, and that as the red color decreases in brightness with an increasing brightness in blue, the green color increases in brightness. This suggests the interpretation that when the vegetation color changes from green to brown in the ground scene, the colors in the image will change from red to green.

An examination of Figure 19 indicates that it may be trimodal in blue and green. The peak in the green region was associated with vegetation, and therefore the cyan and blue peaks are associated with nonvegetation. The cyan peak appears to be associated with exposed soil (beaches, excavation areas, plowed fields), highways, silt laden water, and urban development. The blue color indicates the brightest reflectance and is represented by clouds, urban development, and silt laden water. It is interesting to observe that water occurs all along the blue axis, depending on the silt content ranging from black (no silt) through cyan to bright blue (large silt contents).

The definition of the features (with the possible exception of vegetation) shown in Figure 17 are difficult to precisely define such that there is a one-to-one correspondence between color and feature. If the spectral data do represent

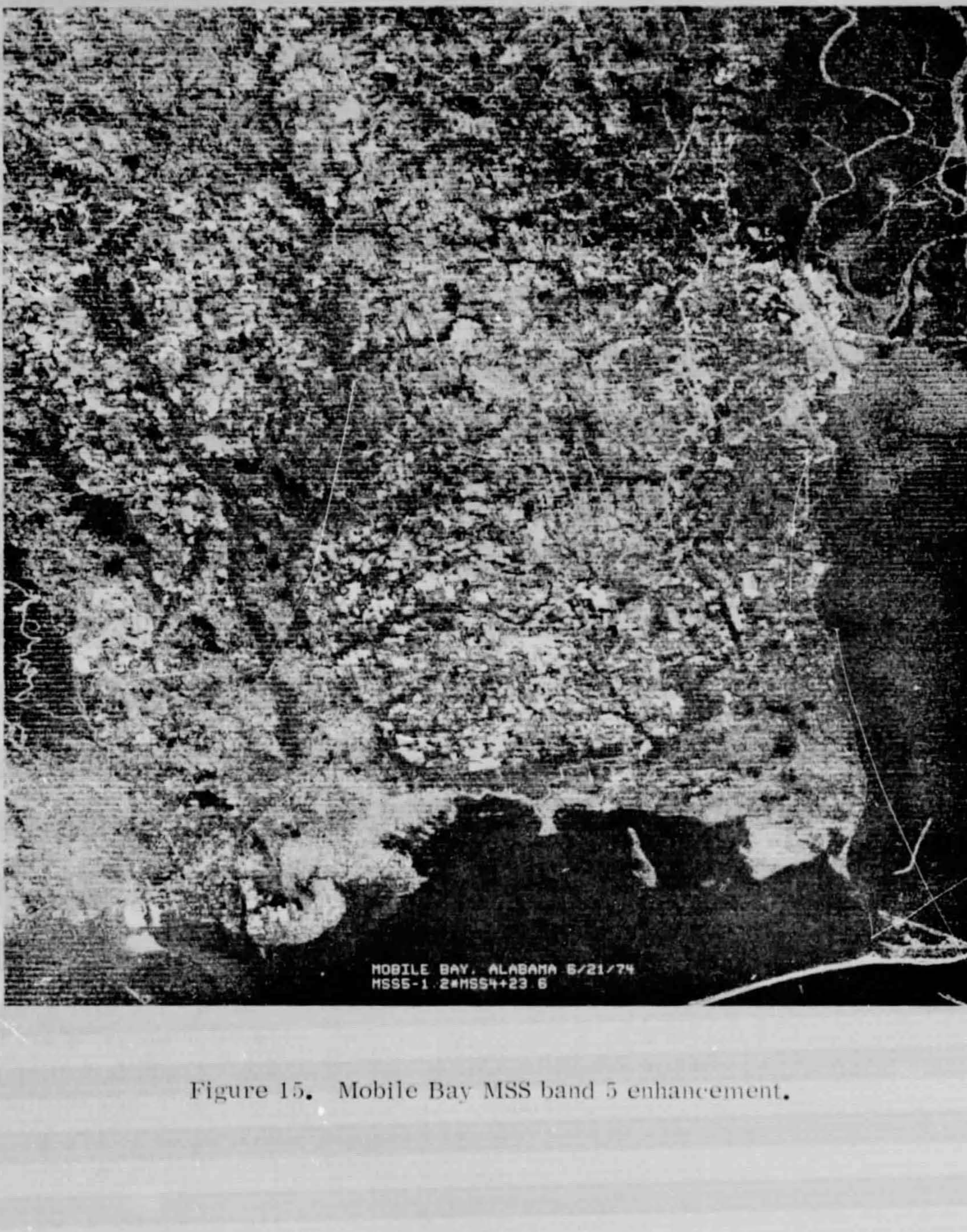


Figure 15. Mobile Bay MSS band 5 enhancement.

ORIGINAL PAGE IS
OF POOR QUALITY

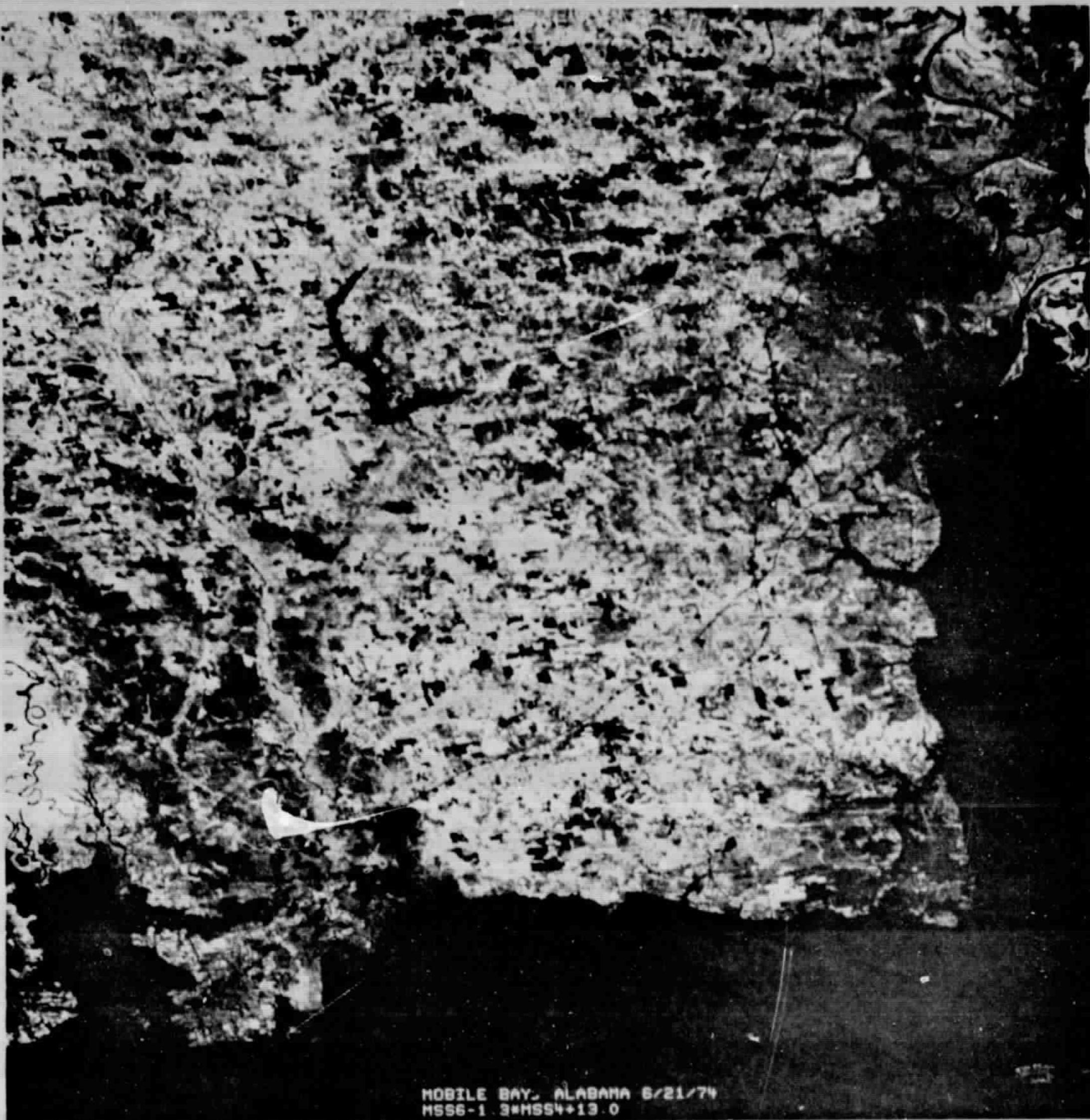


Figure 16. Mobile Bay MSS band 6 enhancement.

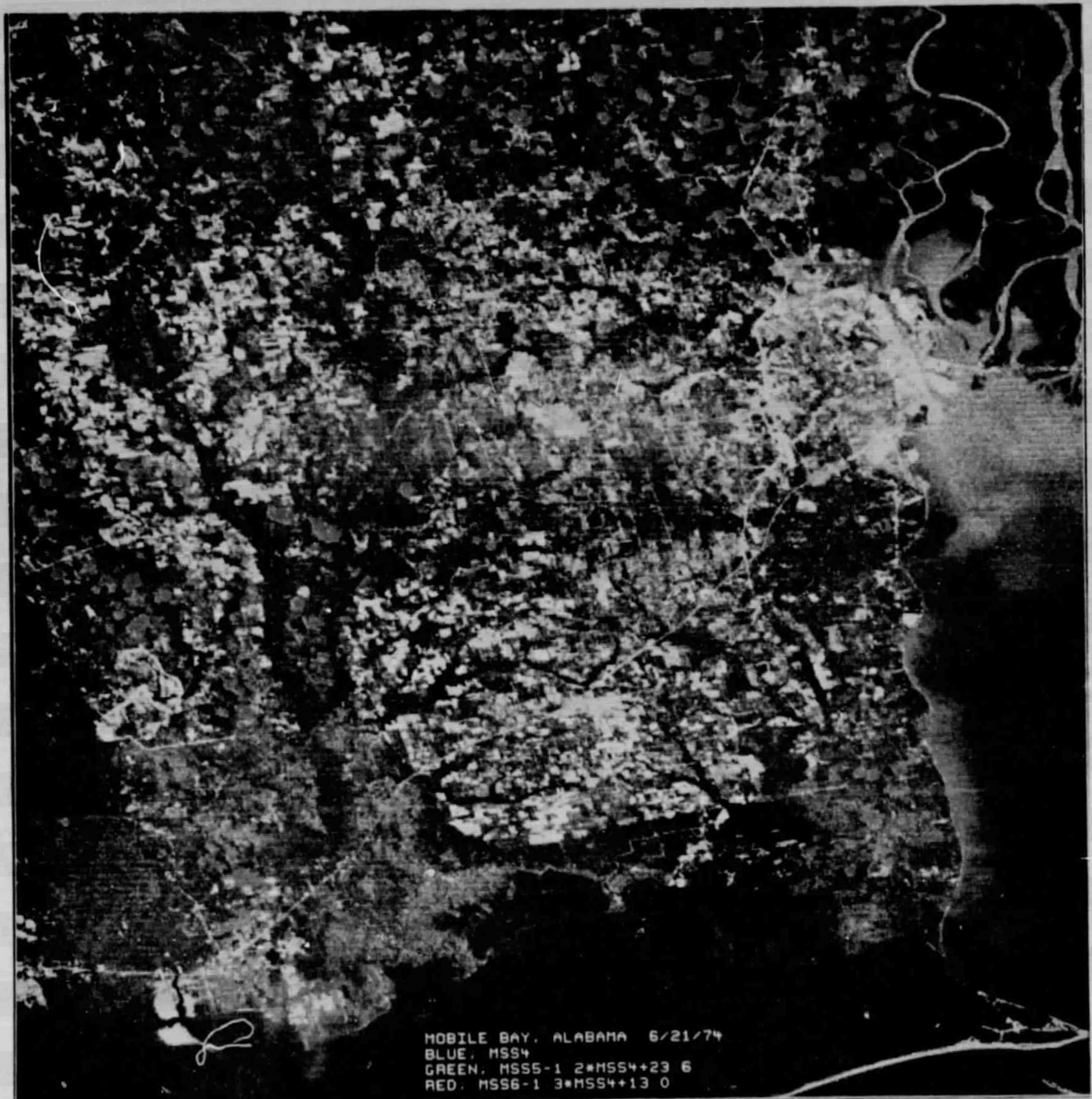


Figure 17. Mobile Bay color composite of MSS band 4 and enhanced MSS bands 5 and 6.

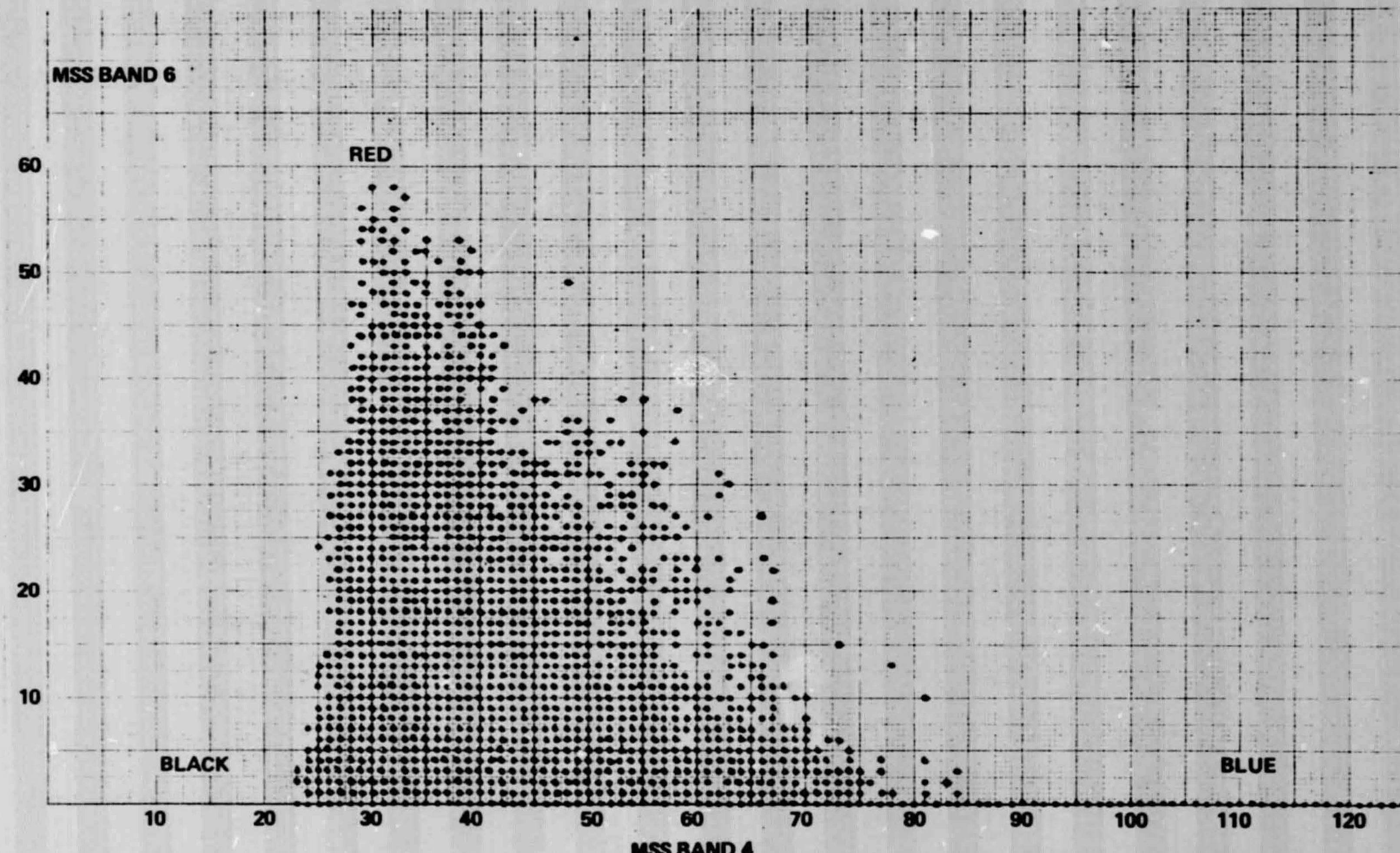


Figure 18. Mobile Bay scatter diagram of MSS band 4 and enhanced MSS band 6.

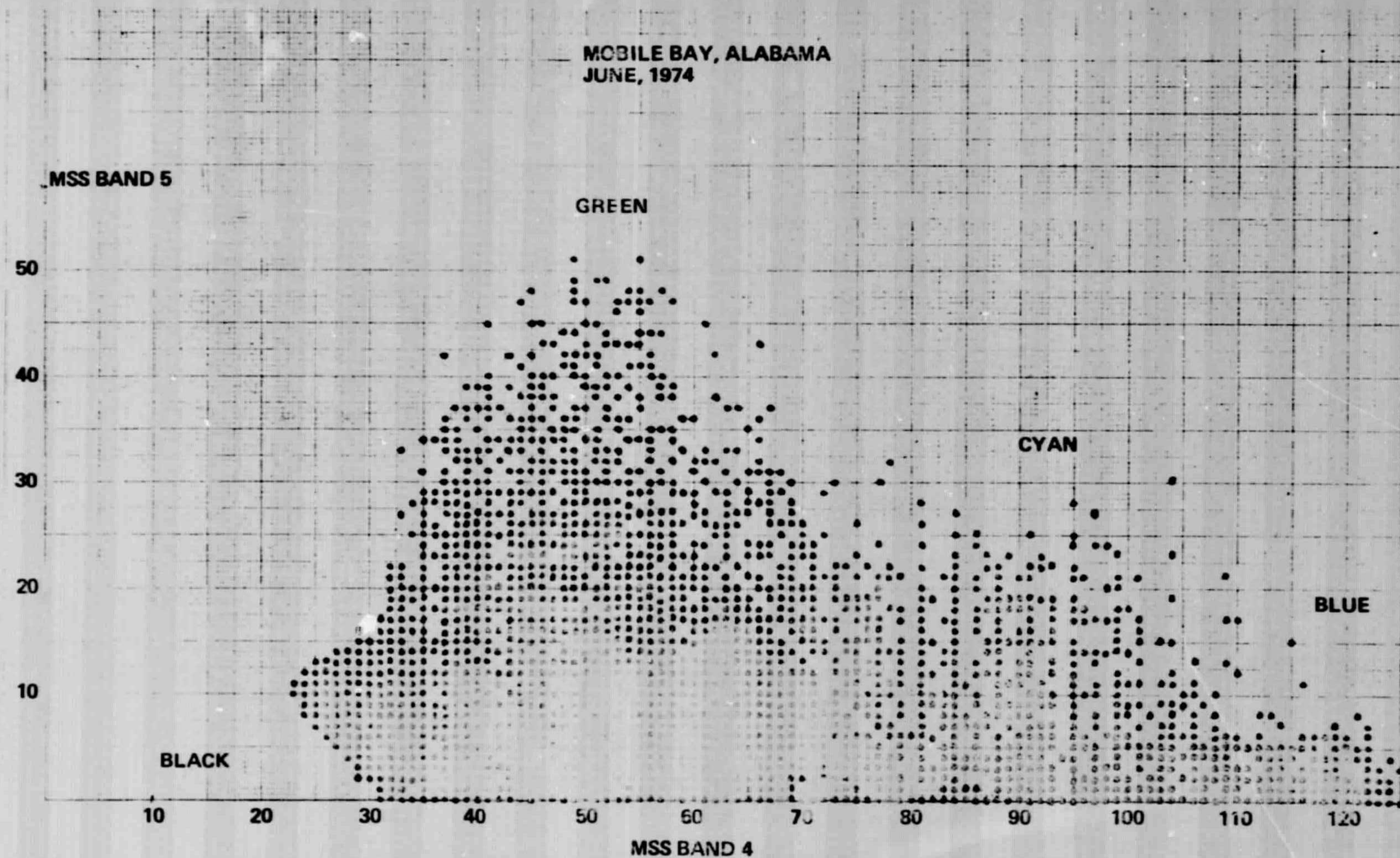


Figure 19. Mobile Bay scatter diagram of MSS band 4 and enhanced MSS band 5.

particular features in the ground scene, then this recognition should be apparent in the distributions of the data, and it should not be necessary to resort to photointerpretation. The results tend to indicate that there are relatively few features present in the data, and that if their definitions can be developed, they will be very broad definitions that do not correspond to those commonly used. It was also noted that large portions of the different band pair distributions are linearly dependent, although there may be excesses of brightness in some regions of the data. The portions that are linear essentially convey the same information as a black and white image, and there are a multitude of things in the ground scene that can have the same grey scale of brightness. It appears that the brightness in excess of linear dependence may provide a key to feature extraction.

VI. SUMMARY AND RECOMMENDATIONS

As a result of this study, there are several hypotheses that can be set forth concerning Landsat data that need to be further verified or disproved by examining additional data and by independent investigation by other Landsat data users. The hypotheses are as follows:

1) Landsat does not discriminate vegetation types, but mostly sees chlorophyl and canopy cover.

2) A majority of the features in the ground scene possess linearly proportional amounts of "color" from each spectral band.

3) The data are continuous, and, as a result, there is no true separability of ground scene features in the data; however, some features possess an excess of color in a particular band-pair.

4) There are relatively few features present in the spectral data, and these do not correspond to the conventional definitions that are used.

5) Aside from seasonal effects, in a distributional sense all Landsat data are essentially the same. The only difference is the way the data are spatially arranged in the image.

If these hypotheses are true, the best existing analysis tool is still probability photointerpretation.

This report has presented arguments which support the previously mentioned hypotheses. However, a strong counter argument can be developed by compiling a list of reports that demonstrate the achievement of very high computer classification accuracies. It is suggested that the counter argument can be diluted if the following observations are considered. First, investigators tend to report successes instead of failures, and there is no way of knowing the proportion of successes and failures. Second, the choices of test area size and season can play an important role in the achievement of high classification accuracies by the elimination of confuser features. Thus, a great majority of the error can be eliminated simply by choosing the right data set at the right time before classification is attempted. Third, very few results seem to be reported on the machine processing of very large areas, presumably because of the signature extension problem. Again, this observation can be explained on the basis that the larger the test area, the higher the probability that it will contain confuser features.

It appears that the majority of the machine feature extraction problem is contained in the data, and is principally due to the continuous nature of the data and linear dependences of its spectral components. All of the commonly used classification techniques have been developed on the assumption that naturally occurring ground scene features will exhibit some degree of feature separability or clustering in the spectral distributions, and, in this sense, the techniques are more advanced than the data to which they are being applied. This assumption of separability appears to be justified when data from training areas are examined. However, as more and more areas are included in the training, the distributional separation between different ground scene features tends to become less distinct and usually disappears when the entire test site is considered.

The main conclusions suggested by this report indicate that the following approach may be worthwhile to explore. First, a model needs to be developed for explaining the spectral distributional behavior of Landsat data so that limitations on its use will be understood, as well as how to best process the data. This model should also have the capability of providing information without having to resort to photointerpretation. Second, the Landsat data should be combined with other types of spectral imagery in an attempt to reduce the linear dependences of the spectral bands.

APPROVAL

SOME OBSERVATIONS ABOUT LANDSAT DIGITAL ANALYSIS

By Robert R. Jayroe, Jr.

The information in this report has been reviewed for technical content. Review of any information concerning Department of Defense or nuclear energy activities or programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

J. T. Powell

J. T. POWELL
Director, Data Systems Laboratory