10259-1-X

Reprinted From

PROCEEDINGS of the SEVENTH INTERNATIONAL SYMPOSIUM on REMOTE SENSING OF ENVIRONMENT

17-21 May 1971

Sponsored by

Center for Remote Sensing Information and Analysis
Willow Run Laboratories, Institute of Science and Technology
The University of Michigan

Water Quality Office
Environmental Protection Agency

Office of Space Science and Applications
National Aeronautics and Space Administration

Federal Highway Administration
U. S. Department of Transportation

U. S. Geological Survey
U. S. Department of Interior

Air Force Cambridge Research Laboratories
Air Force Systems Command
U. S. Air Force

Agricultural Research Service
U. S. Department of Agriculture

Infrared and Optics Laboratory
Willow Run Laboratories
INSTITUTE OF SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF MICHIGAN
Ann Arbor, Michigan

CORRELATION STUDIES WITH GROUND TRUTH AND MULTISPECTRAL DATA: EFFECT OF SIZE OF TRAINING FIELD

C. B. Roth and M. F. Baumgardner

Agronomy Department and Laboratory for Applications of Remote Sensing
Purdue University
Lafayette, Indiana

1. INTRODUCTION

One of the problems in remote sensing is the quantification and precise location of ground observation data such that it can be correlated with multispectral data acquired from aerospace platforms. A technique which has been found to be very effective in soils studies involves the gridding and sampling of an area to be studied such that the exact location where soil samples are obtained can be related precisely to a known address on a magnetic tape containing multispectral scanner data of the area (3). Once this has been accomplished, analytical data (ground truth) of each soil sample from a specific grid point can be examined and related to quantitative multispectral data (from an airborne scanner) corresponding to the appropriate address or location on the magnetic tape.

In the implementation of this technique wide differences in correlation between soils analytical data and multispectral data can be obtained. An important factor which affects the correlation values is the size of the area from which multispectral measurements are made to obtain quantitative radiation values for correlating with ground observation data. In previous studies multispectral scanners have been used very effectively to map surface soil patterns (5). Other studies have demonstrated high correlation between soil organic matter content and spectral properties of soils (2, 3).

The purpose of this study was to investigate the effect of the size of training sets of multispectral scanner data upon the correlation with soil organic matter content as determined by laboratory analysis.

2. PROCEDURE

A soil test area was selected for remote sensing studies in Tippecanoe County, Indiana. Lying in a transitional zone between soils formed under deciduous hardwood forests (Alfisols) and those formed under prairie vegetation (Mollisols), the test field covered an area of approximately 45 hectares. The field was gridded at intervals of 46 meters. At each grid point a 1-kg surface soil sample was taken to a depth of 1 cm. The organic matter content was determined in the laboratory by a modified Walkley-Black method (4, 7) for each of the 193 samples. Samples occurring on the perimeter of the field were deleted, leaving only 139 samples for use in this study.

Multispectral scanner data were obtained over this area on May 6, 1970 at an altitude of 1000 meters. Two rates of digitization were used. At the low digitization rate (LDR) every eighth scan line was digitized at the rate of 220 samples per scan line. At the high digitization rate (HDR) every third scan line was digitized at the rate of 440 samples per scan line.

Statistical studies were made to determine the correlation between soil organic matter content and spectral response in the spectral channels shown in Table I.

Five spectral classes related to different levels of soil organic matter content were selected to coincide with the soil-color organic matter content guide devised by Alexander (1).

Correlation coefficients were obtained for soil organic matter content versus different size training sets for multispectral scanner data digitized at 220 samples per scan line for every eigth scan line (LDR) and at 440 samples per scan line for every third scan line (HDR).

3. RESULTS AND DISCUSSION

A panchromatic aerial photograph (Figure 1) was obtained of the test field, known as Soil Test Area 5 (STA 5), on May 6, 1970 at the time of the multispectral scanner flight. Surface soil patterns observed in this photograph are very similar to the mapping units outlined in the soil map (Figure 2) prepared by conventional soil survey techniques.

Five levels of soil organic matter were delineated by computer-implemented analysis of multispectral scanner data (Figure 3). Pattern recognition techniques developed at the Laboratory for Applications of Remote Sensing (LARS) were used in this analysis (6). The patterns in the three computer printouts (Figure 3) have a striking similarity to the soil patterns of the photograph (Figure 1). The only difference in the data used to generate these printouts was in the size of the training sets of multispectral scanner data.

A remote sensing unit (RSU) represents the instantaneous field of view or resolution element of the scanner system. For this data (Figure 3) each RSU represents an area of approximately nine square meters. The computer printouts were produced from data in only two wavelength channels, 0.50-0.52 micrometers and 0.66-0.72 micrometers.

Visual comparison of Figures 3 a, b, and c reveals few differences. Where differences are found there seems to be no consistant shift from a higher organic matter to a lower organic matter content or vice versa. However, a comparison of the correlation coefficients of spectral reflectance versus soil organic matter content (Table II) shows an increase in r^2 values with increasing number of RSU's as training sets. In stepwise regression analysis channel 5 (0.66-0.72 μm) was the single best channel for predicting organic matter content in all training set sizes except for the single RSU. Channels 1 (0.40-0.44 μm) and 2 (0.50-0.52 μm) were also generally high in the selection of the best two or three channels. A training set size of 4 RSU's where the two best channels (5 and 1) were used gave an r^2 of 0.62. Training set sizes of 25 and 144 RSU's with the use of channels 5 and 2 produced r^2 values of 0.67 and 0.72, respectively.

Correlation coefficients (spectral response versus soil organic matter content) were obtained for a wide array of training set sizes and channel selections for both the low (LDR) and high (HDR) digitization rates (Tables II and III). In general, it is observed that HDR data produced much higher r^2 values. With both LDR and HDR data the two best channels gave significant increases of r^2 values over those obtained with the single best channel. With LDR data the addition of third and fourth channels gave some increase in r^2 values. With HDR data the addition of channels beyond the best two gave very little improvement in correlation.

The 25-RSU training set size consistently produced the maximum r^2 values with LDR data (Table III). Larger training set sizes tended to reduce the value of r^2. This may have been caused by the overlap of multispectral scanner data from one training set with that obtained from another when 64 or more RSU's were used.

Although the maximum r^2 values were not obtained with the 25-RSU training set size for HDR data, this size closely approached maximum correlation coefficient values (Table II).

A comparison is presented between LDR and HDR data for two computer printouts each of which displays spectral patterns representing five levels of soil organic matter (Figure 4). The patterns in Figure 4a were produced from LDR data. The training set size was 4 RSU's, representing an area of approximately 200 square meters. Patterns in Figure 4b were produced from HDR data and training set sizes of 25 RSU's, representing an area of approximately 225 square meters. Spectral channels 2 (0.50-0.52 μm) and 5 (0.66-0.72 μm) were used for both illustrations. Visual comparison of the two indicates that Figure 4a gives sharper, more clearly-defined boundaries between classes. Perhaps more homogeneity is shown within a single pattern in Figure 4a than in 4b. However an examination of the r^2 values (spectral response versus organic matter content) reveals a significant difference, 4a having an r^2 of 0.31 and 4b an r^2 of 0.67.

In all phases of this study the results obtained from the high digitization rate proved to be significantly superior to those obtained from the low digitization rate.

A further comparison was made, using HDR data, of the patterns of five levels of soil organic matter produced from data of (1) the single best spectral channel, (2) the best two channels, and (3) all size channels utilized in this study (Figure 5). Channel 5 (0.66-0.72 μm), the single best channel selected by step-wise regression analysis, was used to produce the computer display in Figure 5a. Channels 5 and 2 (0.50-0.52 μm) were used for Figure 5b, and all six channels (Table I) were used for Figure 5c. Visual comparison of the three computer displays shows that the single best channel (Figure 5a) gave the sharpest, most homogeneous organic matter classes. Between the broad light colored and dark colored areas a narrow transitional band is delineated. As more spectral channels are used, the broad patterns remain similar, but much of the homogeneity within a single level or class of organic matter is lost and much more scattering of computer symbols representing different levels of organic matter occurs. This is particularly striking in Figure 5c where six spectral channels were used. However, when correlation coefficients (spectral response versus soil organic matter content) are compared, increasing the number of channels improves the correlation. The single best channels, channel 5, (Figure 5a) gave an r^2 of 0.44, channels 5 and 2 (Figure 5b) gave an r^2 of 0.67, and all six channels (Figure 5c) produced an r^2 of 0.69 (Table II).

An array of computer displays produced from LDR data is presented in (Figure 6) which five levels of organic matter are correlated with spectral response from (1) single best channel, (2) two best channels, and (3) all six channels. The patterns produced by channel 5 (Figure 6a) are homogeneous and distinct. Spectral patterns representing five levels of organic matter and produced by the best two channels (Figure 6b) have the same broad patterns, but the boundaries between classes is less distinct and there is less homogeneity within broad areas. When six spectral channels are used, the patterns become even less well defined and more heterogeneous (Figure 6c). The r^2 values for these three examples are 0.19, 0.39, and 0.48 for one channel, two channels, and six channels respectively.

4. CONCLUSIONS

Several conclusions important to the interpreter of earth resources data obtained by aerospace remote sensing techniques emerge from this study:

1. There is a high correlation between multispectral response from soils and the content of soil organic matter.

2. The size of the training set for computer implemented analysis of multi-spectral data had an important effect on the correlation between soil reflectance and organic matter content.

3. The high rate of digitization of the multispectral scanner data gives much greater correlation coefficient values (spectral reflectance versus soil organic matter content) than does the low rate of digitization.

4. The selection and number of channels has a profound influence on the correlation between soil reflectance and organic matter content.

Much more research is necessary before these relationships can be fully understood and explained.

LITERATURE CITED

1. Alexander, J. D. 1968. Color chart for estimating organic matter in mineral soils in Illinois. University of Illinois, Urbana.

2. Baumgardner, M. F., R. W. Leamer, and J. R. Shay. 1970. Remote sensing techniques used in agriculture today. From Aerospace Science and Agricultural Development, American Society of Agronomy, Madison, Wisconsin.

3. Baumgardner, M. F., S. J. Kristof, C. J. Johannsen, and A. Zachary. 1970. Effects of organic matter on the multispectral properties of soils. Proc. Indiana Acad. Sci. 79:413-422.

4. Graham, E. R. 1959. An explanation of theory and methods of soil testing. Missouri Agricultural Expt. Sta. Bul. 734.

5. Kristof, S. J. 1971. Preliminary multispectral studies of soils. Jour. Soil and Water Conservation 26:15-18.

6. Laboratory for Agricultural Remote Sensing. 1970. Remote multispectral sensing in Agriculture. Vol. 4 (Annual Report). Res. Bul. No. 873, Agricultural Experiment Station, Purdue University.

7. Walkley, A. and I. A. Black. 1934. An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Sci. 37:29-38.

TABLE I. LOCATION AND RANGE OF SPECTRAL (WAVELENGTH) CHANNELS.

Channel Number	Range (in micrometers)
1	.40 - .44
2	.50 - .52
3	.55 - .58
4	.58 - .62
5	.66 - .72
6	.80 - 1.00

TABLE II. EFFECT OF SIZE OF TRAINING SET ON THE CORRELATION OF REFLECTANCE FROM HDR VERSUS SOIL ORGANIC MATTER CONTENT.

CONDITIONS: High Digitization Rate (HDR): every third scanline at 440 samples/line
Date of scanner flight: May 6, 1970
Area represented by each RSU*: Approximately 9 sq. m.
Number of training samples: 139

No. of RSU's*	Sequence of Channel Selection	r^2 value after each selected channel was forced into the correlation					
		1st	2nd	3rd	4th	5th	6th
1	4-1-5-2-6-3	.42	.56	.58	.59	.61	.61
4	5-1-4-2-6-3	.43	.62	.63	.64	.64	.64
9	5-1-2-6-3-4	.43	.64	.65	.66	.66	.66
25	5-2-6-3-4-1	.44	.67	.68	.69	.69	.69
64	5-2-3-4-1-6	.45	.71	.71	.71	.72	.72
100	5-2-3-4-1-6	.45	.72	.72	.72	.72	.72
144	5-2-3-4-1-6	.45	.72	.72	.72	.72	.72

* Remote sensing unit (resolution element)

TABLE III. EFFECT OF SIZE OF TRAINING SET ON THE CORRELATION OF REFLECTANCE FROM LDR VERSUS SOIL ORGANIC MATTER CONTENT.

CONDITIONS: Low Digitization Rate (LDR): every eighth scan line at 220 samples/line
Area represented by each RSU*: Approximately 50 sq. m.
Date of scanner flight: May 6, 1970
Number of training samples: 139

No. of RSU's*	Sequence of Channel Selection	r^2 value after each selected channel was forced into the correlation					
		1st	2nd	3rd	4th	5th	6th
1	4-2-6-5-1-3	.18	.33	.34	.34	.34	.34
4	4-2-6-1-5-3	.15	.31	.35	.37	.38	.39
9	4-2-6-1-5-3	.19	.37	.42	.43	.44	.45
25	4-2-1-6-5-3	.19	.39	.45	.48	.48	.48
64	4-2-1-6-3-5	.17	.37	.45	.47	.47	.47
100	5-2-4-1-6-3	.19	.32	.38	.44	.46	.46
144	5-2-4-1-6-3	.21	.35	.39	.45	.46	.47

* Remote sensing unit (resolution element)

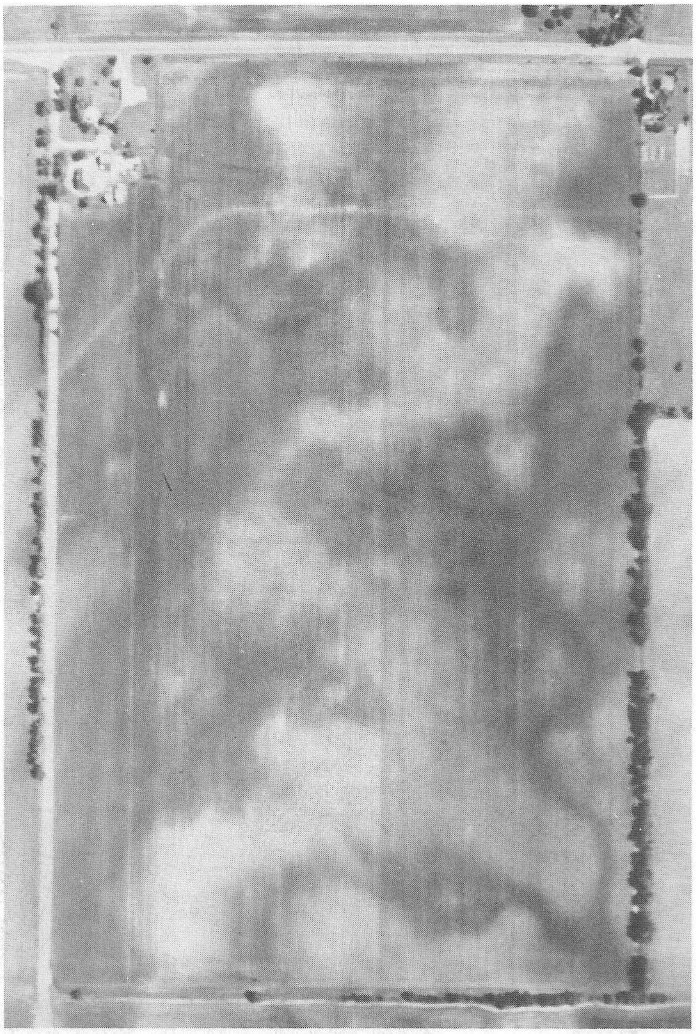


FIGURE 1. AERIAL PANCHROMATIC PHOTOGRAPH OF SOIL TEST AREA 5.

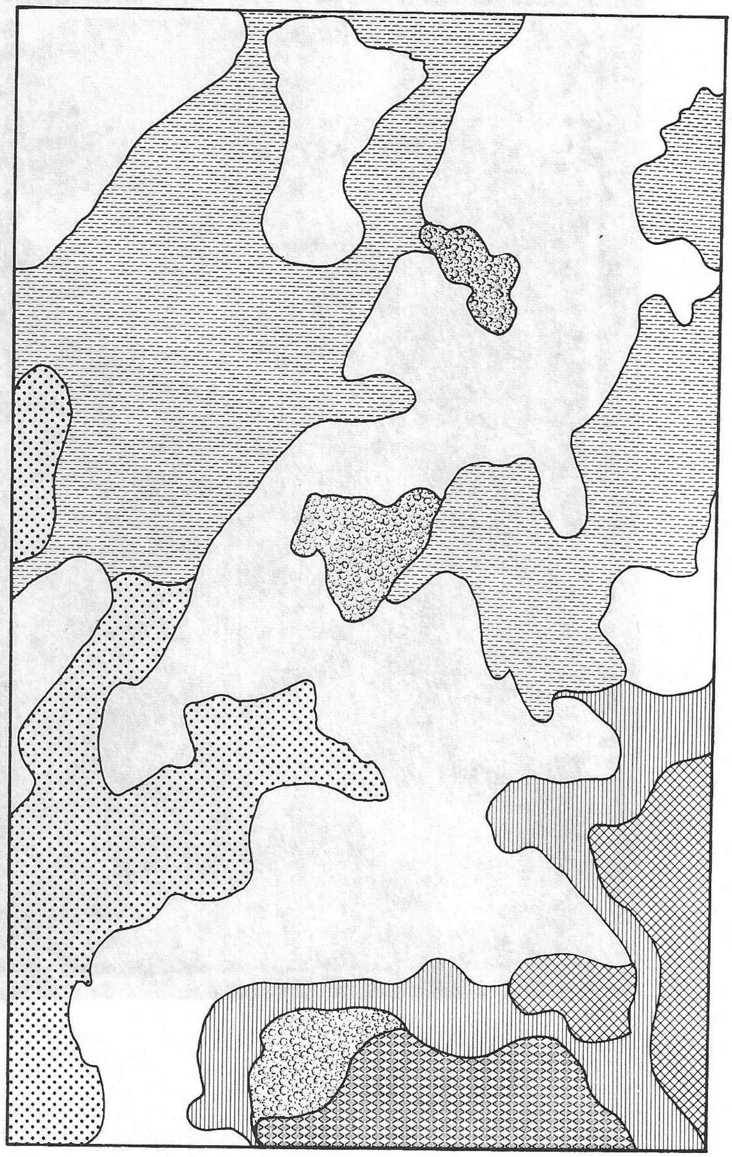
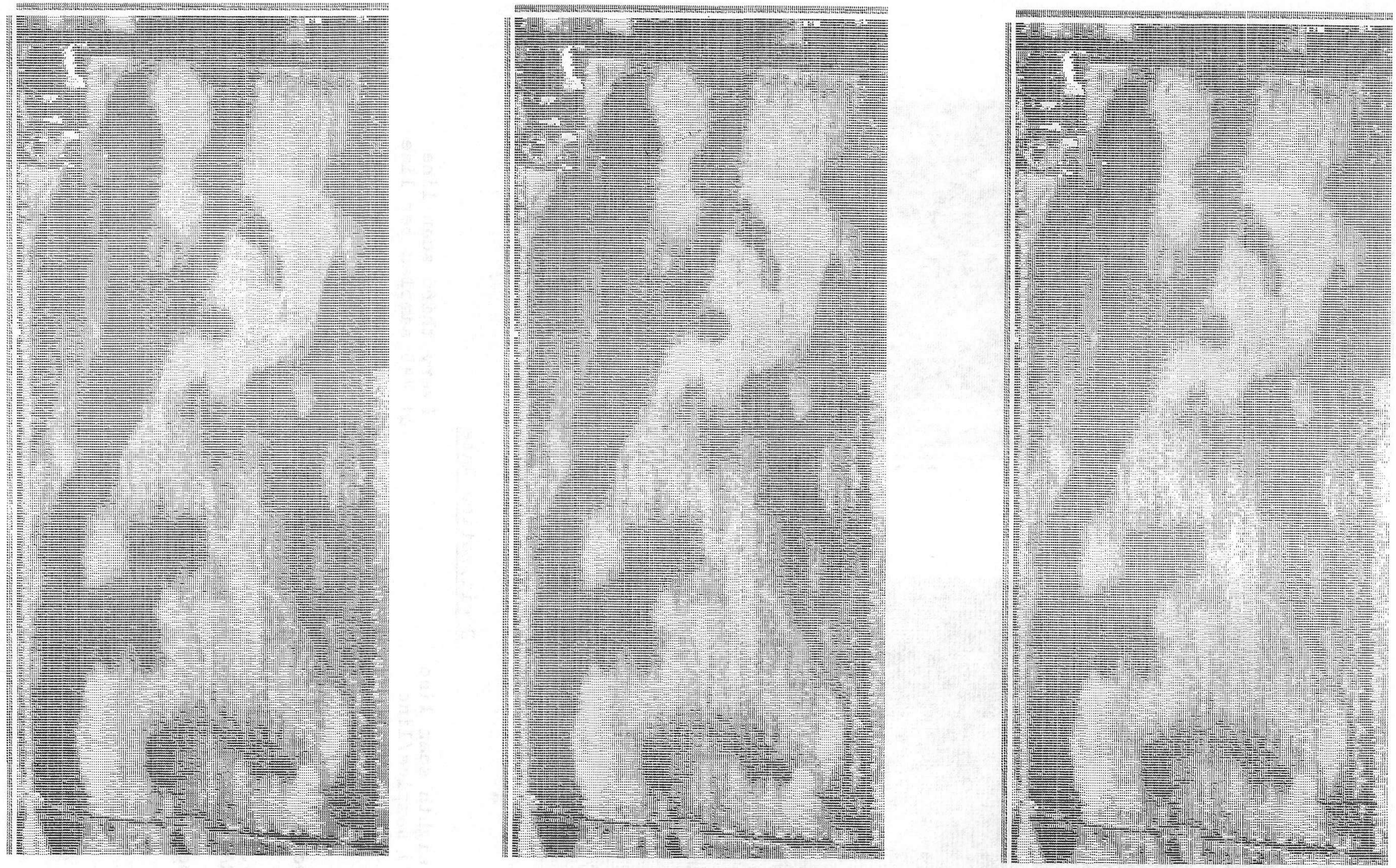


FIGURE 2. SOIL MAP OF SOIL TEST AREA 5.

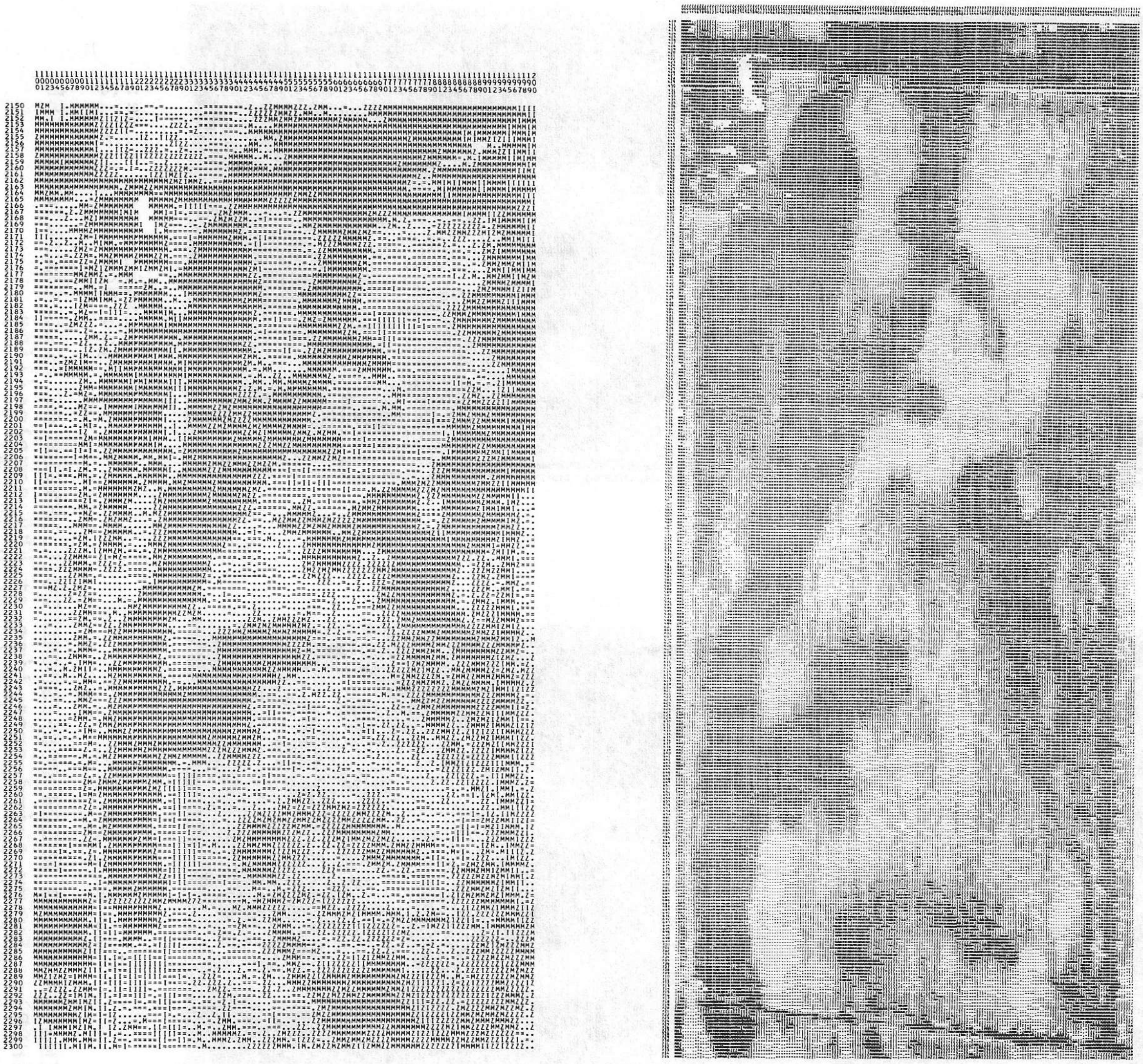


a. Training set size: 4 RSU's (2 x 2)

b. Training set size: 25 RSU's (5 x 5)

c. Training set size: 144 RSU's (12 x 12)

FIGURE 3. COMPUTER DISPLAYS ILLUSTRATING PATTERNS WHICH CORRESPOND TO 5 LEVELS OF SOIL ORGANIC MATTER.
High Digitization Rate (HDR): Every third scan line at 440 samples/line
Spectral Channels: 2 (0.50-0.52 μm) and 5 (0.66-0.72 μm)



a b

<u>Digitization Rate</u>

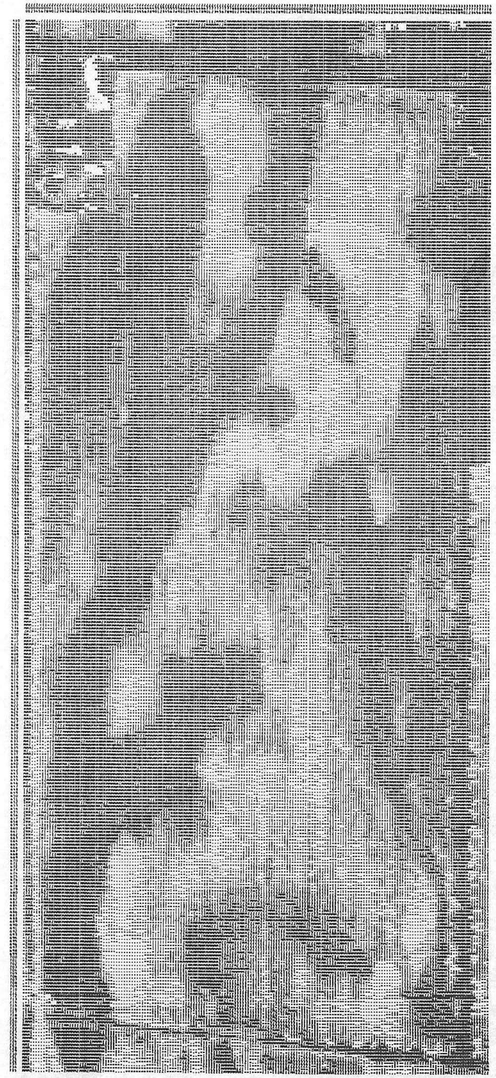
Every eighth scan line at 220 samples/line — Every third scan line at 440 samples per line

<u>Training Set Size</u>

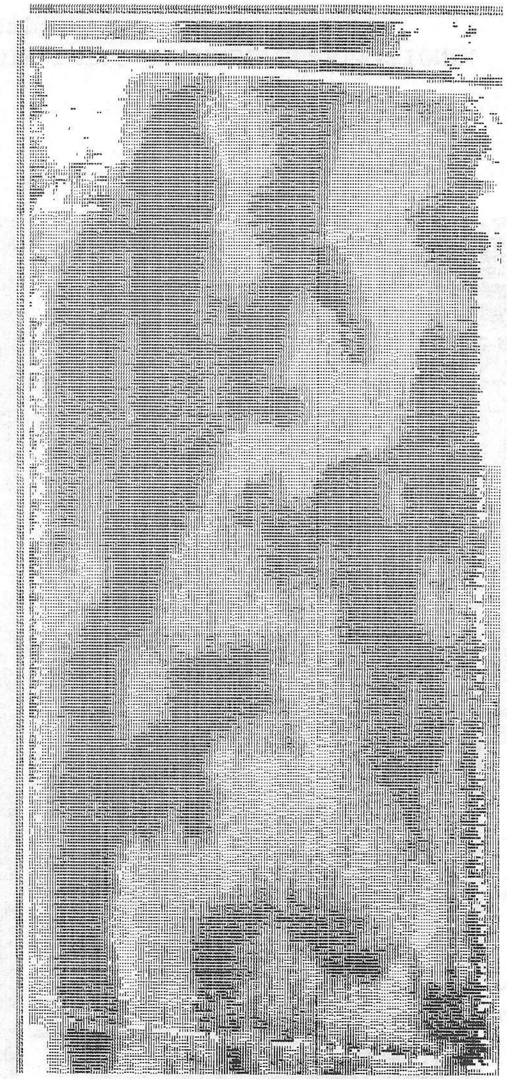
4 RSU's (approx. 200 sq. m.) — 25 RSU's (approx. 225 sq. m.)

FIGURE 4. COMPARISON OF DATA FROM DIFFERENT DIGITIZATION RATES BUT OF COMPARABLE GEOGRAPHICAL AREA FOR PRODUCING PATTERNS WHICH CORRESPOND TO 5 LEVELS OF SOIL ORGANIC MATTER.
Spectral Channels: 2 (0.50-0.52 μm) and 5 (0.66-0.72 μm).

a. Channel 5

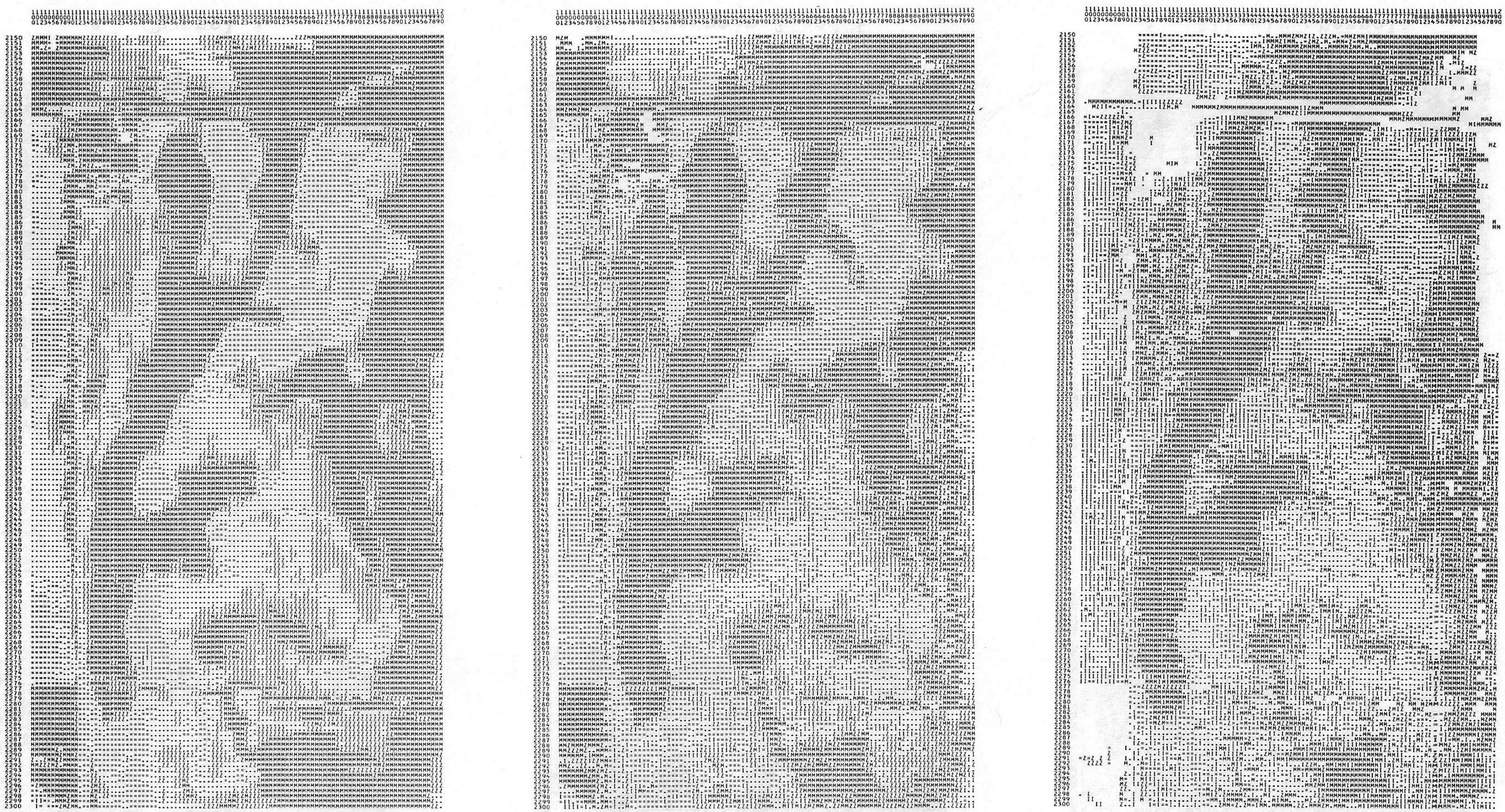


b. Channels 2 and 5



c. All six channels

FIGURE 5. COMPARISON OF SOIL ORGANIC MATTER PATTERNS PRODUCED FROM DATA OF DIFFERENT CHANNEL COMBINATIONS. High Digitization Rate (HDR): Every third scan line at 440 samples per line.



a. Channel 5

b. Channels 2 and 5

c. All six channels

FIGURE 6. COMPARISON OF SOIL ORGANIC MATTER PATTERNS PRODUCED FROM DATA OF DIFFERENT CHANNEL COMBINATIONS. Low digitization rate (LDR), every eighth scan line at 220 samples per line.