08

E7.6-10374 III
CR-147967

DESIGN AND EVALUATION OF A COMPUTER BASED SYSTEM TO MONITOR AND GENERALISE, BY AREAS, DATA FROM ERTS PRECISION IMAGERY TAPES

Professor Keith M. Clayton
University of East Anglia
School of Environmental Sciences
Norwich NR4 7TJ, England

September 1975

Type III Report for period 23 January 1974 - 22 July, 1974

Department of Trade and Industry
Abell House
John Islip Street
London SW1P 4LN, England

(E76-10374) DESIGN AND EVALUATION OF A COMPUTER BASED SYSTEM TO MONITOR AND GENERALISE, BY AREAS, DATA FROM ERTS PRECISION IMAGERY TAPES Final Report, Jul. 1972 - Jul. 1974 (East Anglia Univ.) 156 p

N76-25619
HC $6.75

G3/43

UNCLAS
00374

1035A

Best Copy Available

1. SR No. MMC 035 2. Type of Report III	3. Recipient's Catalog No.
4. Title Design and evaluation of a computer based system to monitor and generalise, by areas, data from ERTS precision imagery tapes.	5. Report date 14 September 1975 6. Period covered July 1972-July 1974
7. Principal Investigator Professor Keith M. Clayton	8. No. of Pages 58 plus appendix
9. Name and Address of Principal Investigator's Organization School of Environmental Sciences University of East Anglia Norwich NR4 7TJ, England	10. Principal Invest.Rept. No. 11. GSFC Technical Monitor George J. Ensor
12. Sponsoring Agency Name and Address Department of Trade and Industry Abell House, John Islip Street London SW1P 4LN	13. Key Words (Selected by Principal Invesigator) LANSAT CCT, generalisation regionalisation, computer analysis
14. Supplementary Notes	

15. Abstract The author has identified the following significant results.

This Final Report describes an objective system for regionalisation using the ERTS-1 (or LANDSAT) computer compatible tapes. A range of computer programs for analysis of these tapes has been developed and these are described. Although some work has been carried out using individual pixels, the emphasis is on a level of generalization appropriate to a satellite system with repetitive global coverage. Work on establishing the main variables influencing regionalisation is incomplete, but the major variables are land/water ratios, and vegetation cover. The scale or texture of the pattern of change in these variables varies a good deal across the earth's surface, and it seems best if the unit of generalization adopted varies in sympathy with the surface being analysed. Computer generalization and analysis cannot keep pace with data collection without unreasonable investment in machine power, but it offers a degree of data reduction which is highly attractive.

CONTENTS.

CONTENTS (continued)

0. General Appraisal

0.1 Research Finance

The research carried out for two years from September 1972 was financed by the United Kingdom Department of Trade & Industry. Since October 1973 a further Research Associate has been working on the material: He is funded by a Research Grant from the UK Natural Environment Research Council.

0.2 History and concept

In common with other investigators, the original proposal was drawn up well in advance of the launch of ERTS-1, and without access to any simulated imagery or computer-compatible material. We were aware 1) that it would be pointless to duplicate existing United States work in the use of multispectral imagery and felt 2) that too much attention was being paid to interpretation (by whatever technique) at the limits of resolution of the system. Conversely it seemed that there might be room for a United Kingdom contribution that touched on two other aspects of satellite imagery, viz: 1) the considerable data-handling problem, suggesting that a computer-compatible approach would be of value, and 2) the unique opportunity offered to achieve global or near-global coverage on a consistent basis. We were also aware of the potential for monitoring change provided for cloud-free areas by the 18-day repetition of imagery. This did not form a major focus of our original proposal, but was borne in mind in considering approaches that might reduce the amount of data requiring detailed interpretation.

The original design proposed adoption of rectangular, preferably square, areas and the extraction of generalized information from these areas. They were designated UTA's (Unit Target Areas) and were seen as the basic building block of a global system for generalization and the monitoring of major change in successive images. With little more to guide us than a 'feel' for an appropriate scale, we proposed that initially the UTA should be 50 x 50 km. It was argued that, since botanical distributions based on a 10 km grid provided very satisfactory maps for the British Isles at about 1:5 000 000, continent-wide or global maps at 1:20 000 000 or smaller could well be based on a UTA of 50 x 50 km. We further argued 1) that the wish to escape recognition of such conventional elements of air photo recognition as towns, fields, ridges and valleys (as opposed to agricultural areas, mountain ranges or deserts) required a reasonably large UTA;

while 2) anticipated problems of precise location of UTA's suggested that smaller areas would suffer from more 'noise' due to inaccurate replication of the area on successive passes; 3) finally, a larger UTA reduced the size of the data by the largest factor.

An early decision involved our choice of computer on which to develop the system. The local university machine is an ICL and had the particular disadvantage that it required 7-track tapes. Although these were offered by NASA, it seemed that the 9-track tapes might be easier to handle. Additionally, it seemed sensible to establish a system on a machine readily available elsewhere, particularly in the USA. As we have access to a 'regional' computer - the IBM 370/165 at the University of Cambridge, and a terminal has been installed at the University of East Anglia during this contract, it seemed the obvious choice. We have suffered problems of slow turn round, particularly in the early stages of the program development when the terminal was not installed. Nevertheless, it was the right choice.

Like other investigators, whether at home or abroad, we made a slow start due to the late arrival of ERTS-1 products, particularly the computer tapes. Since not everything can be pursued at once, we concentrated on program development to extract gray-scale data for UTA's from the tapes. Once histograms were available, we spent part of our effort examining ways of handling them. All through this stage (which lasted over a year) we in general assumed that - 1) the tapes matched the images we had of the same scenes; 2) that the geographical coordinates on the tapes were accurate within 1-2 km and so were not likely to distort our UTA histograms, and 3) that the histograms produced by summation of the number of occurence of each value across the entire 128 points of the scale were a valid basis for analysis and comparison of successive images.

As our techniques were developed it became clear that these assumptions were not wholly justified. One disappointing feature of research at this stage was the discovery that the slow speed at which results from other investigators were becoming known to us would force us to examine some quite elementary aspects of the digital tapes. We had hoped to be able to build on results of other, larger, teams, but this was rarely possible, partly because so few seemed concerned with analysis, rather than interpretation, of the digital tapes.

Once it became clear that problems of location and of histogram characteristics would require investigation, much of the effort we were devoting to generalization had to go back to the individual pixel values. From this we were able to map by line-printer such features as fields, reservoirs and rivers in our Central Californian scene. From this we learnt that the photo images and the tapes did not coincide in coverage, while we also noted errors in position of up to 10 km using the latitude, longitude grid as shown round the edge of the scene and recorded on the computer tape. As these errors are large compared with the size of UTA we propose to use, it has become necessary to adjust the location information for each pass so that our UTA's may be accurately located. Our reports have also noted progress in smoothing the histogram valued to provide a satisfactory basis for the comparison of successive scenes of the same area, or different parts of the same image.

In the generalization of information from ERTS images, one approach is to map, by percentage area or other suitable measure, such simple variables as water, or snow and ice. An alternative approach we have been trying uses arbitrary square UTA's. The gray-scale histograms of these are compiled and they are then clustered using a cluster-analysis program. These are indications that the major regional groupings seen on ERTS imagery may fall-out from such an analysis without the need to use a contiguity constraint. If so (and we are still investigating the effect of varying the areas used as the building blocks of this process) then this will be a most satisfactory objective classification technique for mapping at a small scale.

Regionalization at a global or continental scale is at present likely to be based rather subjectively on climate, vegetation and relief. Since vegetation is in large part dependent on climate (including climate as modified by relief) the relationships are complex and the relative weight of the three variables varied from one person to another, and from one area to another. An objective classification based on statistical analysis of gray-scale values could provide a welcome alternative approach to geographical problems of global/ continental regionalization.

In developing this approach we have been aware that there are other visual characteristics of the earth's surface that would be relevant to our classification. Two of the most likely variables are the presence of rectangular or other checkerboard patterns which are most likely to result from agriculture; and the texture of ridge and valley patterns in mountainous or upland areas. The first are most readily derived from edge-detection techniques while the latter can be more elusive, although where they are well marked, we have been able to detect them by searching for 'edges'. Fourier transformation may also be of value, although rather slow. Edge detection will also be effective in picking up lake and snow edges in areas where these occur. However, summation of approximate water areas by count of very low gray-scale values is probably more appropriate for the high degree of generalization we seek.

0.3 Conclusion

In conclusion we are convinced that too little attention is being paid to the value of ERTS imagery as a consistent data-base for global/continental generalization (regionalization and mapping). If the information can be analysed in an objective manner, the combination of consistent data and replicable analysis will be a notable addition to the information we have on global distributions and patterns.

0.4 Form of this Report

The remainder of this Final Report to NASA consists of material drawn from earlier Progress Reports, updated where necessary. In this way the general points made in this General Appraisal and the Introduction are developed in more detail. The main publications of the research team are also included. The delay in submission is regretted. It is largely the result of repeated delays in commissioning a printer/plotter; it had been hoped to demonstrate that this represented a cheap solution for graphic output that would match the computing speed of the IBM installation. The delays here have been concerned with the completion of hardware and have been wholly outside the control of the team.

1. INTRODUCTION

1.1 Objectives

The aim of this project is to devise a system for generalising remotely sensed data for areas of the earth's surface at a scale appropriate to a world data collection system. Successful generalisation will allow the identification of changes in surface cover over successive orbits, and may be used for small scale mapping of these changes.

This system, if satisfactory, could be used as a first stage data filter, selecting for further analysis those images which show significant changes from earlier imagery. It is desirable that the system uses data which has undergone the minimum amount of pre-processing, and hence bulk (system corrected) ERTS imagery has been used throughout.

Since it is likely that any operational earth observation satellite will, like ERTS-1, return the imagery as a digitized picture, there are many advantages in devising a system capable of processing this raw digital data. Among these advantages are the reduction in slow and costly photographic processing, the potential for developing a fully automatic system, and the ability to make use of the full dynamic range of the sensors without the degradation in radiometric fidelity introduced by photographic processing. Also future systems will possibly have sensors which do not yield information suitable for photographic presentation.

1.2 Generalisation of the Imagery

Each ERTS frame contains 3×10^7 pixels (picture elements) for the four MSS bands, whereas a regional map at a scale of, say, $1 : 10^6$ could show $10^2 - 10^3$ independent data points in an equivalent area. It is thus obvious that some generalisation of the ERTS imagery is required for regional scale investigations.

This generalisation can be introduced in two ways. The conventional approach is to classify the imagery into surface types, and to use these surface types as a basis for generalisation. One disadvantage of this approach is the expense in computing power required to classify each pixel quite apart from the difficulty of recognizing 'standard types' for such classification. An alternative system is to generalise the raw data prior to classification and mapping on a regional scale. This second approach is adopted for this project, since it offers the capability of efficiently monitoring the imagery on a regional scale.

The types of feature which should be significant on this scale are, for example, large grass burns in savanna regions, or snow cover in the catchment area of a reservoir. It is hoped that the system described here could identify this type of change against a background noise of smaller scale changes.

1.3 Generalisation Units

One of the aims of the project is to determine the area over which the data can be effectively generalised and yet still give a useful representation of the image. It is probable that this area will be a function of the terrain type and the features of interest.

Initially, square Unit Test Areas (UTAs) covering 2.5×10^5 hectares have been used. These UTAs are defined by the geographical coordinates of their centre points, and have edges parallel to lines of latitude and longitude. In some cases it might be desirable to use a UTA coinciding with geographical boundaries, and to allow for this UTAs with polygonal outlines may also be used.

1.4 Method of Generalisation

The generalisation method which has been adopted characterizes each UTA by its gray scale frequency distribution. The parameters used for generalisation could be quantities derived from those distributions, such as the mean and the standard deviation, or they could be the histogram that represents that distribution. The latter will be used in this investigation, because the gray scale histograms are the simplest way of handling the information contained in the frequency distribution. For most purposes the four MSS bands will be treated independently, but greater sensitivity could be achieved by using a frequency distribution in four dimensional measurement space.

More complex methods of generalisation such as the power spectrum of the structure in the image could give parameters related to the photogrametric use of texture. The use of such alternative means of generalisation will be investigated at a later stage of the project.

1.5 Data Base For this Investigation

Two test sites have been chosen, although adequate coverage is only available for one of these, which is an area in the Central Valley of California, and includes part of the Sierra Nevada and the Coastal Range. The second test area is in Eastern England.

The 70mm bulk negatives for each of the scenes including part of the test area have been examined, and CCT's obtained for several of the scenes. The centre points, scene identifiers and local tape code names for these tapes are listed in the table below.

TABLE I.

Local code	NASA scene identifier	Centre point	Date	Area
Tape B	1038 - 18114	37 27 N, 120 22 W	30 Aug 73	C. valley, Calif.
Tape C	1031 - 10334	53 N, 0 W	23 Aug 72	E. England, The Wash
Tape D	1056 - 18114	37 23 N, 120 22 W	17 Sep 72	Calif. C. valley
Tape E	1308 1 18122	37 34 N, 120 37 W	27 May 73	Calif. C. valley
Tape F	1228 - 10293	51 40 N, 0 06 W	8 Mar 73	S.E.England London area
Tape G	1308 - 18120	38 58 N, 120 08 W	27 May 73	Calif. Lake Tahoe area
Tape H	1307 - 18071	36 08 N, 119 30 W	26 May 73	Calif. S. C. valley (Bakersfield)
Tape I	1307 - 18064	37 34 N, 119 06 W	26 May 73	Calif. Lake Mono & Sierra Nevada

2. UTA Location and Compilation

2.1 Introduction

Of primary importance for automatic generalisation of satellite imagery, is the ability (i) to align the digital data to geographical coordinates, and (ii) to select those pixels which lie within a specified target area.

There are several ways in which this alignment could be made, but for an automatic system operating on a global scale, the most hopeful in terms of computing time is to use the information available on the satellite position and attitude. This is the method used by NASA for meteorological satellites and for ERTS bulk (system corrected) imagery. From predicted or measured orbital parameters the ERTS processing facility calculates the latitude and longitude of both the sub-satellite point and the central point of the image. The latter corresponds to the intersection of the principal axis of the RBV cameras with the earth's surface and depends on the attitude of the satellite. These two points are then used to superimpose a latitude-longitude grid on the image. The pre-launch estimates of the accuracy of this grid were about 500 m, which implies that a 50km square UTA could be located with a 1% error. This is sufficiently accurate for the sort of changes that could be detected by this method.

2.2 Techniques for compiling the UTAs.

UTAs are defined by the geographical coordinates of their corner points, and are compiled from the CCTs in two stages:

i) The conversion from latitude - longitude coordinates to scan line, picture element coordinates (tape coordinates) is performed by linear interpolation between the tape coordinates corresponding to the intersection of the latitude-longitude grid with the edge of the frame using the grid information contained in the annotation record at the start of the tape.

ii) The digital data corresponding to a UTA is compiled by selecting from the CCTs those points which fall between lines connecting the corner points of the UTA. The UTA data is then stored as a separate data file consisting of segments of scan line of varying length.

2.3 Results

The accuracy of the coordinate conversions performed by this method have been tested using easily identified control points for the Californian images (Tapes B, D, and E). The geographical coordinates were taken off the USGS 1 : 250, 000 maps, and converted to tape coordinates . A section of the image surrounding the tape coordinate was displayed using line printer gray-scale plots, which gave an indication of the positional errors.

Figure 1 shows the results obtained for the dam of Honey Lake in California. The positions found for the dam on tapes B3 and D3 are marked by open circles, and show errors in the range of 5 to 10 km. Tests for other parts of these images show that the error has a large constant component, and a smaller component which varies across the image and has a magnitude of about 1 km.

2.4 Discussion

The source of the major error is almost entirely in the latitude-longitude grid data provided by NASA, and similar errors are found from the grid superimposed on the photographic products. It is possible that the minor error is due to the coordinate conversion algorithm. This is particularly likely near the corners of the images where grid marks are sparse, and linear interpolation is less reliable because of geometric distortion.

These errors in the positioning of the digital data relative to the UTAs are too large to allow reliable monitoring for changes in the imagery.

Two solutions are possible for the purposes of this investigation. Either the UTAs can be defined purely in terms of tape coordinates and subsequent images aligned by use of a correction which initially would have to be determined manually, or the geographical coordinates could be converted to tape coordinates with due allowance made for the error in the grid. The advantage of the first is that it removes the need for unpacking the annotation record, and so reduces the core store required for programs, which would bring useful savings in computer usage. The second alternative allows geographically significant (or at least interesting) arees to be used more easily, but at the expense of increased computer usage.

An interesting excercise in this context is the use of a cross-correlation technique to align one image with another. This is a standard technique in crystallographic and in biological image processing where the image consists of well defined objects seen against a uniform background, but it is not immediately obvious that it will be successful in aligning remotely sensed images where there is no clear distinction between object and background. Density slicing is one means of reducing images to a collection of objects on a uniform background, and this offers the most promising line of action. This approach is currently being pursued.

Alternatively, the tape coordinates of features recognisable on the photographs could be taken directly from a coordinate grid superimposed on the photograph, provided that the tape and photographic imagery are coincident. However, it appears that there is an along-track displacement of tape imagery relative to the photographic imagery by as much as 8 km, and so there is no advantage is using this method. The relative displacement of the two types of imagery has been found for all cases for which we have CCTs.

2.5 Display Techniques

In order to locate recognisable ground features in terms of tape coordinates, some form of display of the data on the tapes is needed. Two forms of picture display are available which are to some extent complementary, and both will be maintained.

i) Line printer gray scale pictures simulate the photographic imagery, and thereby provide a rapid display of the data. However, they suffer from the disadvantages of mapping at too small a scale, and of a limited range of gray tones. For speed, no allowance is made for the skew on the imagery, or for the distortions imposed by the use of a standard line printer, although in principle this could be done. Therefore, the resulting map is not geometrically accurate, and is not comparable, in the overlay sense, with conventional maps.

ii) Contour maps of the gray scales drawn by the graph plotter have corrections for the skew and sampling effects, and have an easily variable scale. In addition there is a wider range of gray scale information available. The disadvantages are the time taken to produce a single display, the difficulty in contouring areas of rapidly changing gray tone and the lack of an immediate interpretation of the display.

Figure 1. provides an example of a gray scale printout and Figure 2. shows a sample contour map.

3. Gray Scale Sampling

3.1 Introduction

Radiometric errors in the CCT imagery impose limitations on the use of the gray scale histograms. It is necessary to investigate the magnitude of these errors, so that the significance of particular differences between histograms can be assessed.

The sources of radiometric error to be investigated are:-

i) those related to the conversion from radiance to gray scale levels on the CCTs

ii) those caused by the uneven response of the sensors.

3.2 The MSS Transfer Function

A detailed description of the MSS transfer function, which determines the conversion from radiance to gray-scale values is given in the section "System Performance" in the ERTS Data Users' Handbook. A brief summary will be given here.

The radiance, R, to tape count (gray scale values on the CCT), T, transfer function is always linear, but as will be seen, the _a priori_ probability of any pixel having a tape count T is a function of T itself: that is, the probability distribution (as opposed to the frequency distribution), is a non-uniform function of T. The cause of this non-uniformity is the discrete nature of the gray scales, and becomes more apparent when intermediate transfer functions, f_1 between R and the sensor count, S, and f_2 between S and T, are considered.

For band 7 alone, the transfer function f_1 is linear and S has a range of 0 to 63. The second transfer function f_2, which is related to the processing of the raw sensor data by the Special

Processing Subsystem (SPS), is also linear and T has the same range as S. Among the operations performed by SPS is radiometric calibration of the sensor data and this could cause f_2 to differ from a one-to-one mapping of S onto T. If, for example, two different sensor counts, S_1 and S_2, are always mapped into a single value, T_1, then the a priori probability of obtaining T_1 will be twice the mean probability.

The non-uniformity introduced by this effect will be spread over adjacent gray levels only and could be removed by smoothing with a simple running mean technique. In practice the calibration function changes continually, and this smoothing does not appear to be necessary. This is shown by Fig.3 which is the unsmoothed band 7 histogram for a 50 km square area in the central valley in California.

The sensors for bands 4, 5 and 6 are usually operated in the compressed data acquisition mode, in which the sensor transfer function f_1 is non-linear, with the sensor count scale compressed for high radiance values. The range of S is again $0 \leqslant S \leqslant 63$. To correct for the non-linearity of f_1, the SPS transfer function f_2 is also non-linear and can be approximated by a quadratic of the form

$$T = 0.025\, S^2 + 0.45\, S \qquad (1)$$

This gives a tape count scale which is linear with a range $0 \leqslant T \leqslant 127$, or twice the range of S. Because of this difference in the ranges of S and T, the transfer function F_1 cannot be a one-to-one mapping. Without the smearing introduced by the changing calibration function which modifies equation (1), the probability of obtaining at least half of the T values would be zero. Because of the non-linearity low T values will correspond to two or more S values, whereas at the high end of the radiance scale only one in every two of three T values would have a non zero probability. Thsi is illustrated by Figure 4 which shows the mapping of the 64 S values into T on the assumption that f_2 is represented exactly by (1). This distribution shows that the a priori probability of obtaining T = 0 is approximately six times greater, on average, than that of obtaining any particular value of T greater than 120.

3.3 The Effects of the non-uniform probability on the histogram

The gray scale histograms derived from the imagery are the product of this non-uniform probability distribution and the actual gray scale frequency distribution. The effect on the histograms is illustrated by Figure 5. Here each line shows the histogram for 12 scan lines of the band 5 for Tape D3. The tape counts have been combined in pairs to give a range of 64 values. One of the most prominent features of the histograms are the columns with anomolously low frequencies which persist over several sets of 12 scan lines, and which are associated with gray levels for which the a priori probability is low or zero. Much of this irregularity is due to the stretching of the original 64-scale gray levels to the 128 gray levels on the tapes as described by Thomas (1973).

Because the exact form of the mapping f_2 from S to T changes from scan line, and from date to date, the lower limit of the detail in the histograms which can be accepted as real is approximately the range of T which corresponds to a single value of S. Because of the non-linearity of f_2, this is a function of T, i.e. smaller changes can be detected in the radiance of dark features than can be detected in brighter features.

The irregularity of the histograms can be reduced, and at the same time the redundancy in the grey scale values due to the expansion of the range by f_2, can be removed by smoothing the histograms with a suitable convolution function. The aim of this convolution is to derive a 64 level scale for which each level has an equal *a priori* probability. The probability of obtaining a value between T and T + ΔT is related to the probability of obtaining a value between S and S + ΔS by

$$p(T).\Delta T = p(S).\Delta S \frac{dT}{dS} \qquad (2)$$

For a uniform distribution in S, this reduces to

$$p(T).\Delta T = \frac{1}{64} \cdot \frac{dT}{dS} \qquad (3)$$

Differentiating (1) yields

$$\frac{dT}{dS} = 0.05\,S + 0.45$$

which, on substituting for S gives

$$\frac{dT}{dS} = 0.05\sqrt{81 + 40T} \qquad (4)$$

The width of the 64 intervals in T which have equal *a priori* probabilities is found from (3) and (4)

$$\Delta T = 0.05\sqrt{81 + 40T} \qquad (5)$$

The centre points of these intervals is given by

$$T_n = 0 \qquad n = 1$$

$$T_n = \frac{1}{2}\Delta T_n + \sum_{i=1}^{n-1} \Delta T_i \qquad (6)$$

Preliminary tests have been made using a trapezoidal convolution function which has a width at half height given by (5), outer edges with slopes of +1 and - 1, and an equal weighting for all points near the centre of the interval (Fig.6)

This convolution function has the advantages that it is smoothly varying over the range of T, and allows for the one-to-two or one-to-three mapping at high T values by weighting equally all values near the centre of the window. Figure 7 shows a histogram both unsmoothed and smoothed using this convolution function. The full 128 T values have been used in plotting the histogram, although at most 64 of them are independent, and a 64 point non-linear scale would be more approprate.

3.4 Identification of Scanner Irregularities

The multi-spectral scanner that produces the radiometric data has 6 independent sensors in each band, and thus any unevenness in the response of those sensors will produce a banding effect on the image which would repeat every 6 lines. In an effort to identify the magnitude of this effect, data from the CCTs were subjected to the technique of power spectrum analysis.

This technique of analysis takes a given sequence of data and reduces it to a series of wave forms. Each wave, which is of the form of a sine wave, is characterised by two parameters, a wavelength (the distance between two successive peaks), and an amplitude (the height of the wave form). The power associated with each wavelength is a measure of the contribution of that wavelength to the total signal.

Thus any wavelength corresponding to a pronounced banding effect will plot as a peak on the graph of power against wavelength, which will thus identify strong periodicities present in the data.

The analysis was implemented by use of the program published by Davis (1973). The data input was a string of gray-scale values at right angles to the scan lines, parallel to the flight path of the satellite, with a sampling interval of one scan line. Thus any periodicities with a wavelength of 6 times the sampling interval could be associated with unevenness in the sensors.

The resulting power spectra for all four spectral bands were calculated for several images over lengths of 300 scan lines. The results for an area of the North Sea (Tape C4) and for an area of the Sierra Nevada of California (Tape D3) are shown as Figures 8 and 9. Results, not shown here, were obtained for adjacent scan positions, which demonstrated that the spectra were identical or very similar if a small horizontal shift was applied. The resultant plots have been smoothed by application of the Hanning triangular window (Davis 1973), and power is plotted on a logarithmic scale.

The results for the Wash image (Figure 8), shows that there is a distinct peak near the wavelength of 6 scan lines, identifying the unevenness of the sensors. On this image all the values are very low, reflecting the fact that the image consists of open sea which is uniformly dark. Because of the low input values the resultant spectra are very noisy. However, the peak at wavelength 6 is markedly larger than the other peaks, which are associated with noise. The height of this peak decreases as the mean radiance increases, implying that the magnitude of the unevenness is similar on all bands, but that its relative importance varies inversely with radiance. The graph suggests that the variability attributable to this effect is of the order of ½ to 1 gray scale value.

The results for the Sierra Nevada (Figure 9) show the power spectra in the presence of a much stronger signal. The parallelism of the spectra for the four bands shows that, for this image, most of the power is related to surface features. Once again a peak is visible at a wavelength of 6, and suggests an irregularity of about 1 gray scale value, that can be attributed to the scanner.

3.5 Discussion

It has been shown that both the non-uniformity of the radiance to gray scale transfer function, and the use of 6 independent sensors for each spectral band, cause detectable effects in the imagery.

The irregularity of the histograms is accounted for by the non-uniform distribution of *a priori* probabilities of gray scale values. This is most significant at high radiance values, where it is equivalent to an uncertainty of about 3 in the gray scale.

The banding introduced by the unequal responses of the 6 sensors causes an uncertainty of about 1 in the gray scale. This is most significant at low radiance values where the errors introduced by the transfer function are small.

These two effects impose a lower limit of 2 gray scale values in the width of the classes which can reliably be used in preparing and comparing gray scale distribution histograms. An algorithm has been suggested which divides the full range of T into 64 classes with equal _a priori_ probability, but the importance of sensor unevenness at low radiance levels implies that 32 classes would be more suitable.

4. The Analysis of Gray-Scale Histograms

4.1 Introduction

Initial examination of histograms for several areas showed that there was a possibility of equating components of the histogram with surface types. Thus the possibility of splitting histograms, and thereby deriving estimates of the percentage of any image covered by each surface type was investigated.

The investigation concentrated initially on an area which showed a simple bimodal histogram, which could be readily interpreted. This area was part of image C4, which covered part of eastern England and the adjacent sea area. The land in this area consists mainly of small fields of the same order of magnitude, or smaller, than the 70 m resolution of the scanner, and thus the overall impression given by the land is of a uniform gray speckle. It was thus not expected that this technique would be able to distinguish between different areas on the land. In contrast, the sea on the image shows up as a dark area, visible on all four bands, although the greatest contrast is offered by band 7, in which the gray scale values for the sea are nearly all either 0, 1, or 2. However, on bands 4 and 5 the considerable areas of mud-flat, turbid water and sand bank are readily visible. It was hoped that for these bands it might be possible to split the histogram into three components, thereby identifying proportions not only of land and sea but also of turbid water. However, the histograms were in fact monomodal showing that the two distributions that could be identified on bands 6 and 7 were overlapping, and that it was not possible to accurately identify even two components of the histogram.

4.2 Technique adopted

A large number of techniques have been developed for splitting an observed frequency distribution into the sum of two or more component distributions. Clark (pers.comm.) has reviewed these methods, and recognised three main groups; graphical, analytical, and numerical.

Graphical techniques rely on the subjective interpretation of discontinuities in the cumulative frequency curve, and are thus subject to a high degree of operator error. Analytical techniques are exceptionally difficult to calculate, and do not on every occasion yield a solution. Numerical techniques are much more varied, but nearly all use iterative methods to improve upon initial estimates of parameters describing the component populations. They are thus suitable for inclusion in an automated data handling system.

Of the several alternatives available, it was decided to adopt that of Jones and James (1972), which uses a maximum likelihood method of estimation. The main reason for this decision were as follows:

a) Although initial estimates of the parameters to be estimated are required, the method will converge to a stable solution even if these are considerably in error, and it would thus always be possible to provide a complete null set of initial estimates.

b) The published work included a FORTRAN program, and so made it possible to examine the capabilities of this approach without the investment of a large programming effort at this exploratory stage.

c) It is possible to extend the existing program to include types of component distribution other than the Normal. The published work includes both the circular Normal distribution, and the mechanism for adding other distributions. One major disadvantage is that at present the existing program will handle only two components.

4.3. Brief Outline of the Maximum Likelihood Method

The maximum likelihood method relies on the fact that for any observation set, X, and a statistical model with parameter set Θ, it is possible to calculate the a posteriori probability of X occurring as the outcome of the stochastic process defined by the model. Thus for any model it is possible to find, in principle, a parameter set Θ that will maximise the a posteriori probability, or likelihood, L, of the observation set X.

In the case of a set of parameters refering to a mixture of distributions, the model becomes:

Let the j^{th} underlying probability distribution be $f_j\ (x_i)$, where $j = 1, \ldots, m$, and the subscript i, $i = 1, \ldots, n$ refers to the observations. Thus for each component distribution, the probability of the observation set is:

$$P_j\ (X) = \prod_{i=1}^{n} f_j\ (x_i)$$

If the proportion of each component distribution is Q_j, then the total a posteriori probability of the observation set is given by

$$L\ (x) = \sum_{j=1}^{m} Q_j \prod_{i=1}^{n} f_j\ (x_i)$$

Usually, however, it is easier to transform these probabilities into their logarithms, and replace the multiplication by an addition, so that the transformed expression to be maximized becomes

$$\lambda(x) = \sum_{j=1}^{m} Q_j \sum_{i=1}^{n} \log f_j(x_i)$$

The parameter set Θ for the model consists of (m-1) values of Q_j, and the parameters for each of the sub populations f_j. Note that because the Q_j's must sum to unity, (m - 1) values fully determine set.

In the case to be considered here, the underlying distributions are all considered to be Normal, and thus have the form

$$f(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-(x_i-\mu)^2/2\sigma^2\right\}$$

where μ and σ are the mean and standard deviation. Where there are two such distributions, only a single proportionality parameter is required, which we will call α.

Thus the total statistical model to be considered is

$$f(x) = \alpha\left[\frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\right] + (1-\alpha)\left[\frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right\}\right]$$

And the parameter set thus consists of the five parameters

$$\Theta = \{\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2\}$$

Because the derivatives are non-linear, it is not possible to reach the maximum likelihood solution by direct evaluation, but it is necessary to use a numerical technique. The Jones & James Program uses the steepest ascent method to provide a rapid convergence towards the maximum, and thereafter changes to the Newton Raphson method which is superior near that maximum, but also more sensitive to poor initial estimates.

4.4 Application to the Problem

In order to be able to use the maximum likelihood method it is necessary to have an observation set, rather than a grouped histogram. However, the histograms of any one area are the result of the examination of a large number of pixels, and it was thus considered convenient to convert the histogram to a reduced observation set, purely as a form of data reduction. Because of the integer nature of the gray-scale values, it was possible to reconstruct an observation set with approximately the same proportion of values at each gray scale level as in the original data set. The program was set up to accept a data set of up to 200 points, which was constructed from histograms typically derived from 80,000 points. Repeated analysis using a larger data set of 400 points yielded almost identical component distributions, confirming that this data reduction does not distort the results.

The efficiency of the program was improved by the use of integer, rather than real, arrays for storing the data set. Although this required the use of the FLOAT operation for calculations of the probability function, it permits a more efficient storing of the data. A histogram drawing routine was provided, which used the line printer to give a plot of the input histogram, the histogram derived from the initial estimates, and the histogram fitted by the program.

4.5 Results

Results were obtained from the program for histograms derived from all 4 spectral bands. The data from band 7 yielded the clearest results, since the bimodality of that histogram was extreme, with virtually no overlap between the two components. Initial estimates of the parameters were chosen visually, and a total of 30 iterations were required to produce stable estimates. A sample output is included as figure 10, and a summary of the results for band 6 and 7 is given below:

Results of Maximum Likelihood Estimation of Parameters of Mixed Normal Distributions

Parameter	Band 7 Initial	Band 7 Final	Band 6 Initial	Band 6 Final
Proportion of pop$^{\underline{n}}$ 1	.7	0.717	0.7	0.769
mean of " 1	1.0	1.255	3.0	2.856
s.d. of " 1	1.0	0.712	.4	1.478
mean of " 2	23.0	22.829	20.0	20.562
s.d. of " 2	5.0	5.576	3.0	3.36

As can be seen from the table, for band 7 population 1, which was identified as the sea component, covers 71.7% of the area. The corresponding estimate for band 6 however, was higher, at 76.9%.

The results for bands 4 and 5 were considerably less encouraging. The likelihood response surface proved to be very flat, and thus no stable solution seemed possible. In particular, the estimates of the standard deviations seemed most unstable.

Furthermore the most stable parameter, the proportion of the data in population 1 converged towards a value greater than one. In a different context such a result, and the corresponding negative proportion for population 2 can be interpreted as representing a single parent population from which a distinctive sub-population has been removed. However, in this case the concept of a negative (subtractive) population has no interpretable meaning, and the results are thus of little value.

4.6 Discussion

This pilot study has shown that the automated splitting of gray-scale histograms is possible, and can yield useful results. It has also shown that the results yield discrepancies between bands which were sufficient to cast doubt upon the applicability of the method in general. In the light of these considerations, and the considerable computational time required to perform the analysis, it was felt that this approach, while possibly useful in specific instances, was not suitable for routine application to ERTS data.

The alternative histogram splitting techniques that are available might produce results more rapidly, and be more readily applied. In particular methods that can handle more than two components need to be examined. However, at this exploratory stage it is necessary to make considerable checks on the input histograms, to correlate these with ground truth.

At the time of this part of the study, the problem of correlating tape coordinates with geographical areas, had not been solved. Thus the identification of the spectral responses of land types necessary for the interpretation of the histograms could not be pursued. This technique was not therefore tested further until more background information became available.

It is however pertinent to consider the question whether other analyses might not yield similar results as cheaply. In particular if it were possible to map ground types, it would be a simple matter to estimate their area. The possibility of using either cluster analysis to group points (or collections of points) into hopefully recognizable groups, or some form of discriminant function to map points onto _a priori_ groups might be equally useful. Some of these techniques are discussed below.

5. THE USE OF CLUSTERING TECHNIQUES

As has been emphasised in earlier reports, clustering techniques have been used in an attempt to identify regional subdivisions at a suitably generalized level within an ERTS scene. This approach has produced interesting results, and as they have been summarised in the papers included in the Appendix to this Report, they are not duplicated here. The three main papers concerned are:

Objective generalization using ERTS images. (to be published in: Environmental remote sensing practices and problems, Edward Arnold, 1975 or 1976.

Methodological questions in the digitized analysis of ERTS data, Journal British Interplanetary Society, 1975

The relative performance of some unsupervised clustering techniques for the per-field classification of LANDSAT data. Read to the Purdue Symposium, June 1975, and to be submitted to Pattern Recognition.

6. Edge Detection

6.1 Introduction

Of the several techniques available for the computer processing of pictures, one of the most fundamental approaches is to split the picture into non-overlapping zones by the detection of edges between approximately uniform areas (Rosenfeld, 1969). There have thus been a considerable number of techniques developed for detecting edges in pictures.

It was decided to apply some of these techniques to the ERTS Imagery, in an attempt to identify the kinds of feature that could be distinguished by their sharp edges. In particular it was felt that this technique would be useful for identifying areas on the tape where readily identifiable features, such as lakes, are lacking. The technique would thus be potentially very useful in the irrigated zone of the Central Valley of California, where large spatial units occur in a regular rectangular pattern.

6.2 Techniques

Commonly, edge detection procedures consist of two steps, the first an initial detection of potential edges, and the second a process designed to 'clean up' the resultant image. Such a scheme will be adopted here, though at present only the first step has been implemented. The standard technique in finding edges is to locate the places where there is an abrupt change in the picture function (i.e. in the gray scale values). This is performed by examining the local slope of the picture function, and where this slope exceeds a given threshold value, an 'edge' is detected. Various more sophisticated filters have been proposed (eg. by Rosenfeld 1970), which examine the differences between average values either side of the potential edge, or which apply matched filters to each point in turn. However, for the purposes of this investigation it was felt initially that a fairly crude definition of an edge would suffice, and that any cleaning up of the picture that this would require could be carried out subsequently.

The operational procedure adopted was to examine each point in turn, and where the gradient at this point exceeded a threshold value, to demark that point as belonging to an edge. Thus an initial program was written which calculated the maximum slope at each point, using the function:

$$D_{ij} = \text{Max}\{\,|a_{i+k,\,j+l} - a_{ij}|\,\} \qquad k = -1, 0, 1 \quad l = -1, 0, 1$$

Where a_{ij} is the gray scale value at point ij. This value D_{ij} is thus the maximum difference between each point and its eight adjacent neighbours.

6.3 Results

A program was written to calculate these values of D_{ij}, and produce a binary image of edge points. This has been applied to a portion of the central valley of California (Tape D3). Fig 14 shows a gray scale print out of a portion of the analysed area, and Fig.15 the edges detected in that area.

As can be seen, this operation results in edges which closely correspond with the field pattern as revealed in the gray scale printout but has the unfortunate property that all edges are double, because the programme has identified each edge twice. This effect can be explained by reference to the one dimensional case. If we scan along a line, and find a large difference between points i and i + 1, point i will be labelled an edge. However, this same difference will also be located when examining point i + 1, so that both points i and i + 1 are labelled as belonging to an edge.

To circumvent this problem, differences were calculated only for one quadrant of directions, so that only three adjacent points were examined. Thus the edge identifying function becomes

$$D_{ij} = \text{Max}\left\{ (|a_{i-1,j} - a_{ij}|), (|a_{i,j-1} - a_{ij}|), (|a_{i-1,j-1} - a_{ij}|) \right\}$$

Two thresholds were used for the plotting of D_{ij}. Where D_{ij} was greater than 6, a dot was plotted, and where greater than 12 an asterisk was plotted. A result of this program, for the same area as for the other two pictures, is given in fig.16. Note that a much cleaner image results, and that very few edges are actually lost.

6.4 Discussion

The program to perform the edge detection has been used as a standard display technique. It has been found advantageous to use both a thresholded, compressed, map of edges, and a straight pseudo-gray scale map of edge intensity, as an aid to displaying the data.

Experience shows that the value of a threshold, which filters out much of the noise, cannot be set *a priori* but must be adjusted to the needs of a particular image. Work with the image of the Wash area, using band 5, suggests that a threshold of $D_{ij} \geqslant 3$ shows up many of the boundaries in the water area, but labels virtually the entire land area as "edge". An alternative, higher value of $D_{ij} \geqslant 6$ results in a totally "edge-less" area, but a more satisfactory picture on the land.

The routine has also been used to detect edges of lakes in the Sierra Nevada portion of the California image, where it also detects stream lines. However, it is not satisfactory for detecting the ridge lines that we know are present, largely because the features change less abruptly than the small 2 x 2 neighbourhood of cells considered here.

A further development has been to apply a further thresholding technique to the thresholded image, in an attempt to produce a "cleaner" picture. If we set the value at cell Mij equal to 1 if Dij $\geq$ T where T is same theshold, and equal to 0 if Dij D<T, we produce a simple thresholded map of the gradients. We then "clean" this map by examining the sum of all cells adjacent to each cell, i.e.

$$S = \sum_{k=-1}^{+1} \sum_{l=-1}^{+1} M_{i+k,j+l}$$

If the sum is equal to a given threshold, Ts, we can assume the point ij is part of a continuous edge, but if it is lower, we assume it is a pseudo-random effect, due to features in the image smaller than our field of interest.

Work by Yamada and Farango (1965) suggests a value of Ts = 4 would be sufficient to identify linear edges, but we found this to be too severe, and that a value of Ts = 3 is better for enhancing linear edges.

However, these cleaned maps often lost much useful information, particularly where the image does not contain strong, clear lines as was the case with both the British imagery, and the Californian imagery away from the Central valley. In contrast, in the central valley, where the main feature is that of sharp edges between fields, this "cleaning up" operation was found to be most valuable. Even here, however, it was not possible to generate closed boundaries round uniform areas, as was the original hope. Nevertheless, these edge detection routines have proved exceedingly useful techniques for displaying the imagery in a way that appears far less cluttered than the pseudo-gray scale technique.

7. COMPUTER PROGRAMS

This section does not attempt to give a full listing of the programs developed for this research: these will be listed and documented in the Final Report to NERC at the end of December, 1975. However, it is thought that this detail will enable the range and scope of the software to be judged.

The section is divided into three parts; first a record of the various sub-routines which are used extensively in the programs which are described in abstract in the second and third sections.

7.1 Sub-routines

1. INREAD:- reads and unpacks an annotation record in the ERTS CCT format.
2. ANNRD:- reads and unpacks an annotation record including the latitude-logitude grid data.
3. HEADS:- produces a line printer listing of the relevant information in identification and annotation records.
4. BULK1:- Constructs the mask for compiling square UTA's with sides parallel to latitude and longitude grid.
5. BULK2:- compiles a UTA using the masks generated by either BULK1 or BULK5.
6. BULK5:- constructs the mask for compiling polygonal UTA's with a maximum of eight sides.
7. BULK3:- a self-contained program which will compile histograms (smoothed and unsmoothed).
8. BULK4:- will contour on the graph plotter rectangular areas of the images defined by tape coordinates.
9. Other routines include programs to unpack and pack the digital data into ¼ words as used on the original CCT's, routines for converting the CCT's to ICL readable format, and for producing grey scale pictures on the line printer.

7.2 Computer programs available at Cambridge, UK, for the IBM system.

These are referred to by DSN's on the Archive tape (KMC 201).

KMC2.CLUS.VELDMAN LABEL = 1

This program uses the clustering method described by Veldman (1967) to perform an hierachical cluster analysis of data from the ERTS tapes. The prgram can handle only 176 data points with up to 128 variables per observation. These points are assumed to lie in a grid 16 across by 11 down, representing the 200 x 200 'fields' derived from the ERTS CCTs. The program produces a map of the resultant classification for the line printer for the last 10 stages of the clustering.

KMC2. CLUS726.ITER LABEL = 2

Program uses the Swain iterative clustering algorithm for 726 data points, i.e. for 100x100 fields, and produces a map of the results. The iterative algorithm prints out the transition matrix for each iterative step, and a final interdistance matrix for the stable cluster means.

KMC2.CLUS2880.ITER LABEL = 3

Uses the same algorithm as CLUS726.ITER, but takes 50 x 50 fields - up to 2880 of them, and maps the output onto the line printer. Since this program will almost certainly fail, due to its time requirements, it includes a facility for restarting from an intermediate solution.

KMC2.CLUS2880.NAGY LABEL = 4

Clusters up to 2880 data points - from the 50 x 50 fields - using the Nagy update of the Bonner chain algorithm. Options to the mapping routine permit the same program to handle also the 726 data points of the 100 x 100 fields, and the 176 points of the 200 x 200 fields.

KMC2.TWOTAPES LABEL = 5

Examines an area covered by two different CCT's. The tape coordinates of a common point are required, and the displacement between the two assumed to be parallel to satellite path and scan direction. The tape coordinates of the area to be examined on the first tape are also required. Pseudo-gray-scale print outs of the two examined areas are produced, and a map; also differences between the two maps. A histogram of the differneces is also produced. The program has facilities for performing multiple tasks for one entry of the program.

KMC2.LINER LABEL = 6

Uses a local edge detector to define sharp gradients in the image. The local slope can be mapped as a pseudo-grey scale to produce a visual plot of high gradient areas. The gradient map can also be 'thresholded' to produce maps of 'outlines', which can be variously thresholded, cleared or compressed.

KMC2.FOURIER LABEL = 7

Using the program written by Davis, this takes the Fourier transform of a line of data taken down the scan. The object is to identify the magnitude of the 6-line problem.

A graph of the power spectrum is produced, and facilities exist for performing this at multiple points in the tape, and for a choice of spectral bands.

KMC2.CLUS.ERTS.NAGY LABEL = 8

Performs the Nagy classification algorithm to individual pixels from an ERTS scene. These are mapped out on the line printer. Facilities exist for the suppression of the 6-line streak pattern.

KMC2.BAYES.COMPILER LABEL = 9

Compiles a data file of mean and covariance matrices from an ERTS tape from a set of coordinates which identify training areas for a Bayesian classifier. The training areas are assumed to be identified a priori. The output is a set of files to be used as 'training data' for a later Bayesian classification.

KMC2.OLDMAP.WASH LABEL = 10

Performs a point-by-point classification of an ERTS image using heuristic multi-dimensional density slicing. Present version set up to identify water features of the Wash area. The output is to the line printer in the first instance, and then to the graph plotter for final output. Elementary geometric corrections are included.

KMC2.BAYES.CLASS LABEL = 11

This program performs a maximum likelihood Bayesian classification on a portion of an ERTS image, using as training data the files produced by program KMC2.BAYES.COMPILER. Output is a classified map sent to the line printer.

KMC2.HISTO.GRAMS LABEL = 12

Compiles the 128-level histograms for regular 50 x 50 fields for a single ERTS tape. It thus needs to be run four times to compile the histograms from an entire ERTS image. The data are output to a file which is usually held on magnetic tape.

KMC2.TAPE.LOOKER LABEL = 13

Using the utility routines written by Ian Hill to unpack the header and annotation blocks from a tape, and to unpack and list the first scan line of that tape.

KMC2.GRADIENT

Produces a gradient image from a given ERTS tape, output onto a second tape.

KMC2.PROFILES

Plots single scane lines from an ERTS tape to the pine printer. Has the option for multiple plots from one program entry.

KMC2.COMPFILE

Compiles data files for 100 x 100 and 200 x 200 fields from the smaller fields produced by KMC2.HISTO?GRAMS.

7.3 <u>Support programs available at UEA</u>, for use on the ICL system.

MAPPING: Uses the graph plotter to produce a shaded map of a classified ERTS image. The routine allows for up to 13 shading types, and the boundaries round all non-identical types are drawn. The main program is set up to handle any of the three formats of data, i.e. 200 x 200, 100 x 100 or 50 x 50 fields, and any degree of skew. It has been adapted to plot small amounts pf pixel by pixel data.

LISTING: A catalogue program to list, with comments, the current holdings of ERTS imagery.

HADCALL: Routines to calaculate the Hadamard transform of a given data string using the Fast alogorithm. Intended for eventual use as part of a texture investigation.

MAXLIKE: Uses the James and James maximum likelihood program for splitting one-dimensional (marginal) histograms into two normal components. The printed program has been amended to take as its input a calculated histogram.

SLINK: The Sitson single link cluster analysis, adapted to accept the means of 200 x 200 fields. This program is run on the CDC 7600 at Manchester, and includes an option to draw a dendogram of the results.

QUICKPLOT: Draws on the ICL graph plotter, a single ERTS scan line, from data stored on punch cards.

NLMA: The Salmon non-linear mapping algorithm. Released to A.C.Armstrong on a personal basis from Imperial College, London. Used to project the means of the 200 x 200 fields from the original 4-dimensional space onto a 2-dimensional space. Modified to use the graph plotter as its main output device, rather than the line printer.

* * * * * * * * * * * * * * * * * *

Publications referred to in sections 0 - 7.

Davis, J.C. (1973) Statistics and data analysis in geology, (John Wiley & Sons, New York).

Harmuth, H.F. (1969) Application of Walsh functions in communications, IEEE Spectrum, 6, 82-91.

Jones, T.A. & W. R. James, (1972) Maxlike: FORTRAN IV program for maximum likelihood estimation. Geocom. Programs, 5.

L.A.R.S. (1968) Remote spectral sensing in agriculture. Research Bulletin, 844, Purdue University.

Openshaw, S. (1974) A regionalisation program for large data sets. Computer Applications, (University of Nottingham).

Owen-Jones, E.S. & N.D.E. Custance, (1974) Digitised analysis of Skylark rocket imagery. Journal British Interplanetary Soc., 27(1), 18-22.

Rosenfeld, A. (1969) Picture processing by computer. (Academic Press, New York).

Rosenfeld, A. (1970) A non-linear edge detection technique, Proc. IEEE, 58, 814-816.

Sibson, R. (1973) SLINK: an optimally efficient algorithm for the single-link cluster method. Computer Journal, 16(1), 30-34.

Thomas, V.L. (1973) Separation and physical characteristics of the ERTS MSSS system corrected computer compatible tapes. (Goddard Space Center, Document X-563-73-206, Greenbelt, Md.)

Veldman, D.J. (1967) Fortran programming for the behavioural sciences, (Holt, Reinhart, Winston, New York).

Ward, J.H. (1963) Hierachical grouping to optimize on objective function. Journal American Statistical Assoc., 58, 236-244.

Yamada, S. & J.P. Farango, (1965) Experimental results for local filtering of digitized pictures. Report 4, Dept. Computer Sciences, University of Illinois.

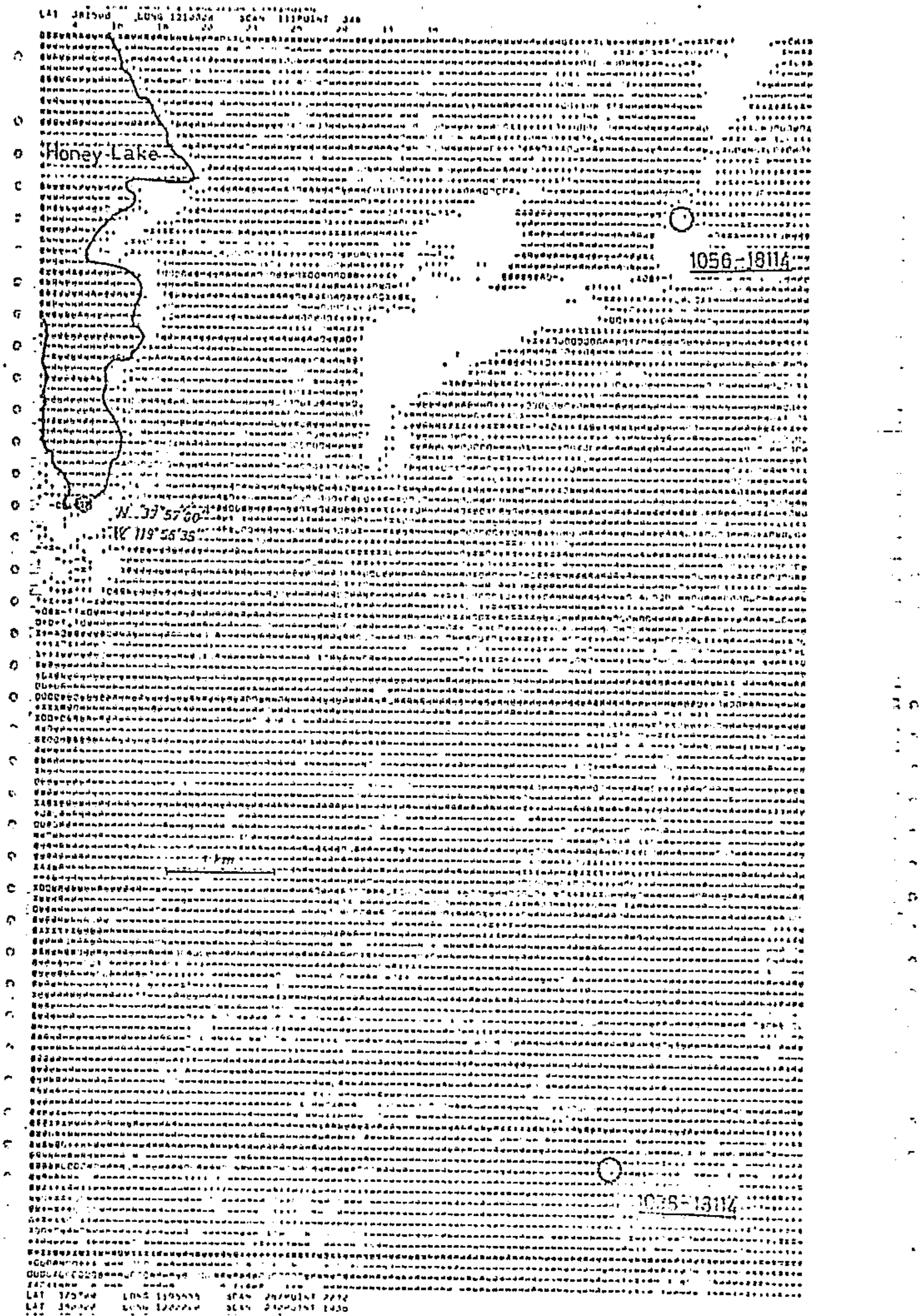


Figure 1. Position errors for the dam wall of Honey Lake.

The open circles show the position of the dam (marked by the hexagon) as found from the latitude - longitude grid included with the CCT annotation data.

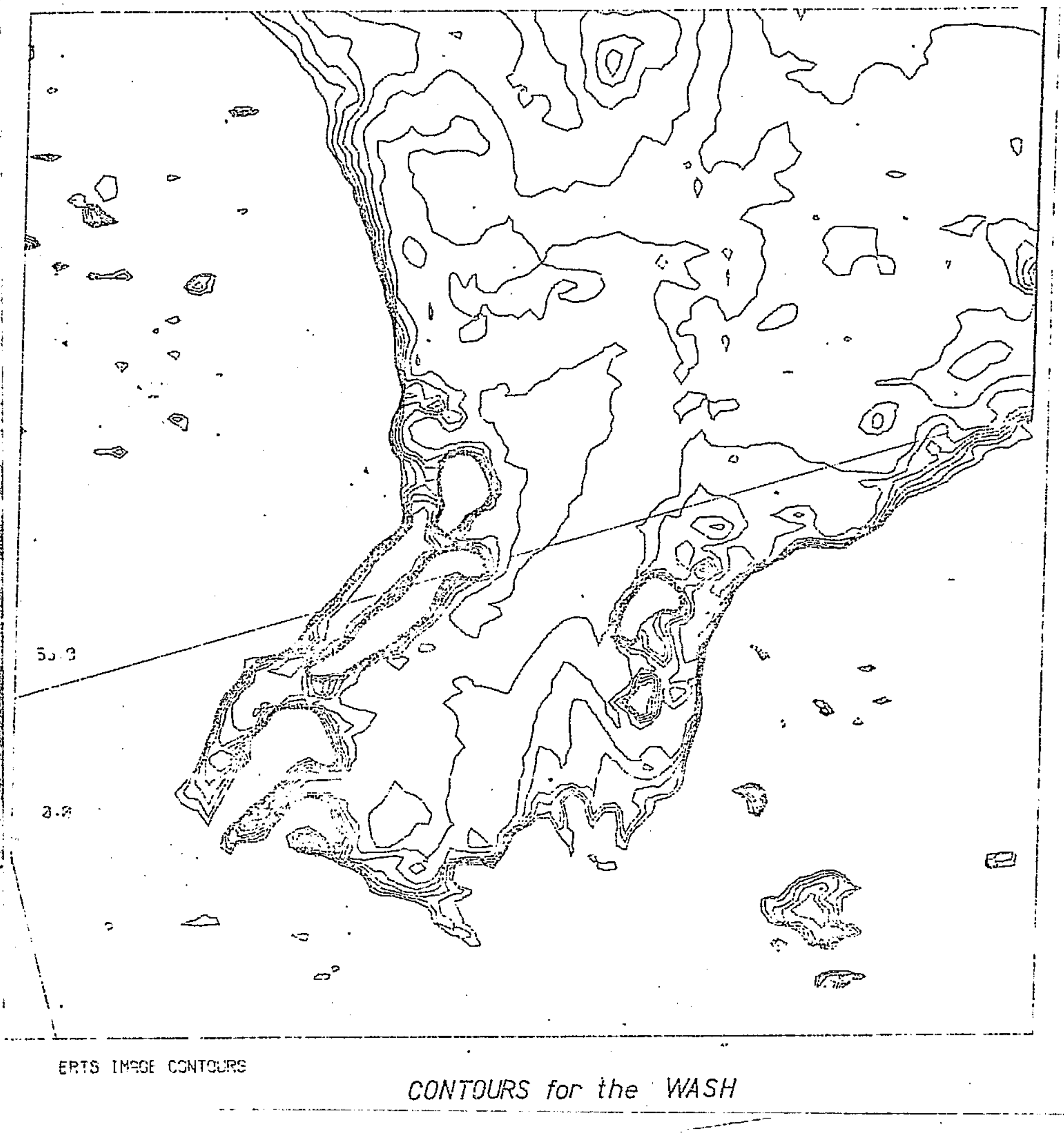


Figure 2. This contour map of gray scale values for the sea area near the Wash, U.K, gives a good indication of the distribution of sediment laden water.

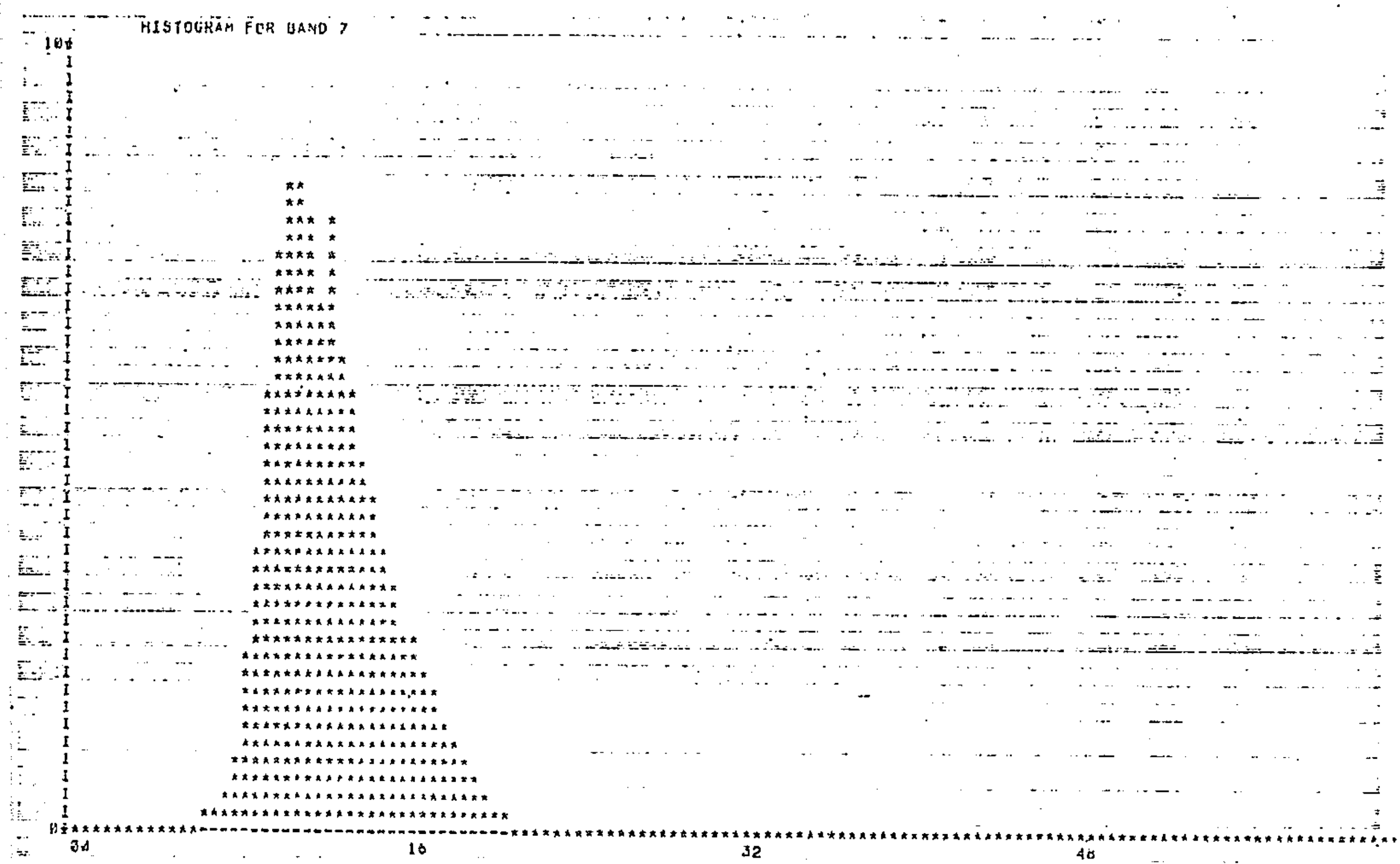


Figure 3. Unsmoothed gray scale histogram for MSS band 7 taken from a UTA in the Central Valley, California. Note how much smoother this distribution is than the corresponding distribution for band 5 shown in Fig.7.

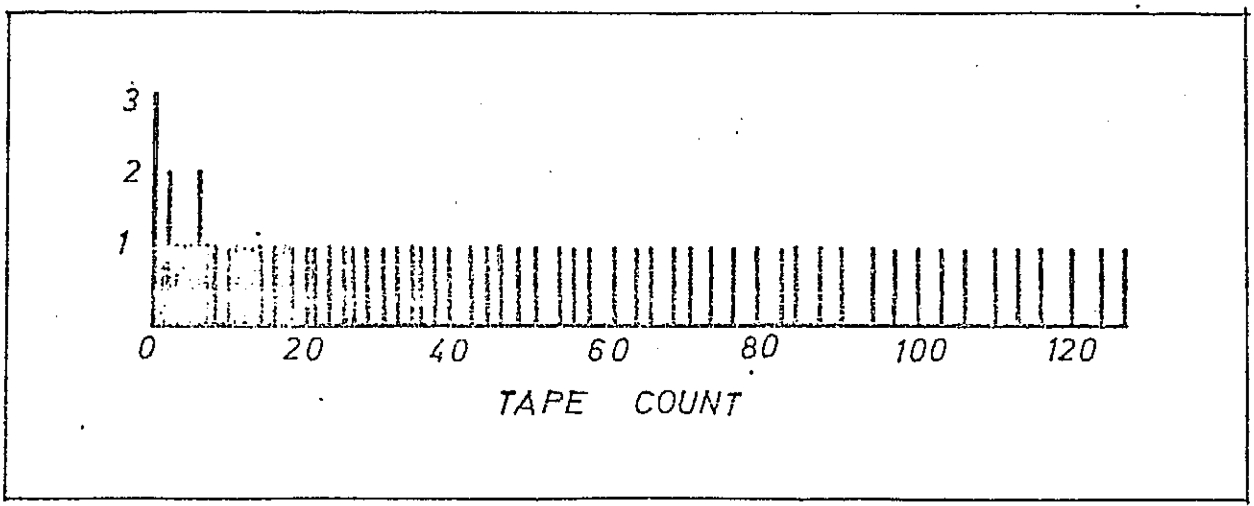


Figure 4. The mapping of the 64 sensor count values onto the 128 tape count values using the function described in section 3. The vertical axis gives the number of sensor counts mapped into each tape count.

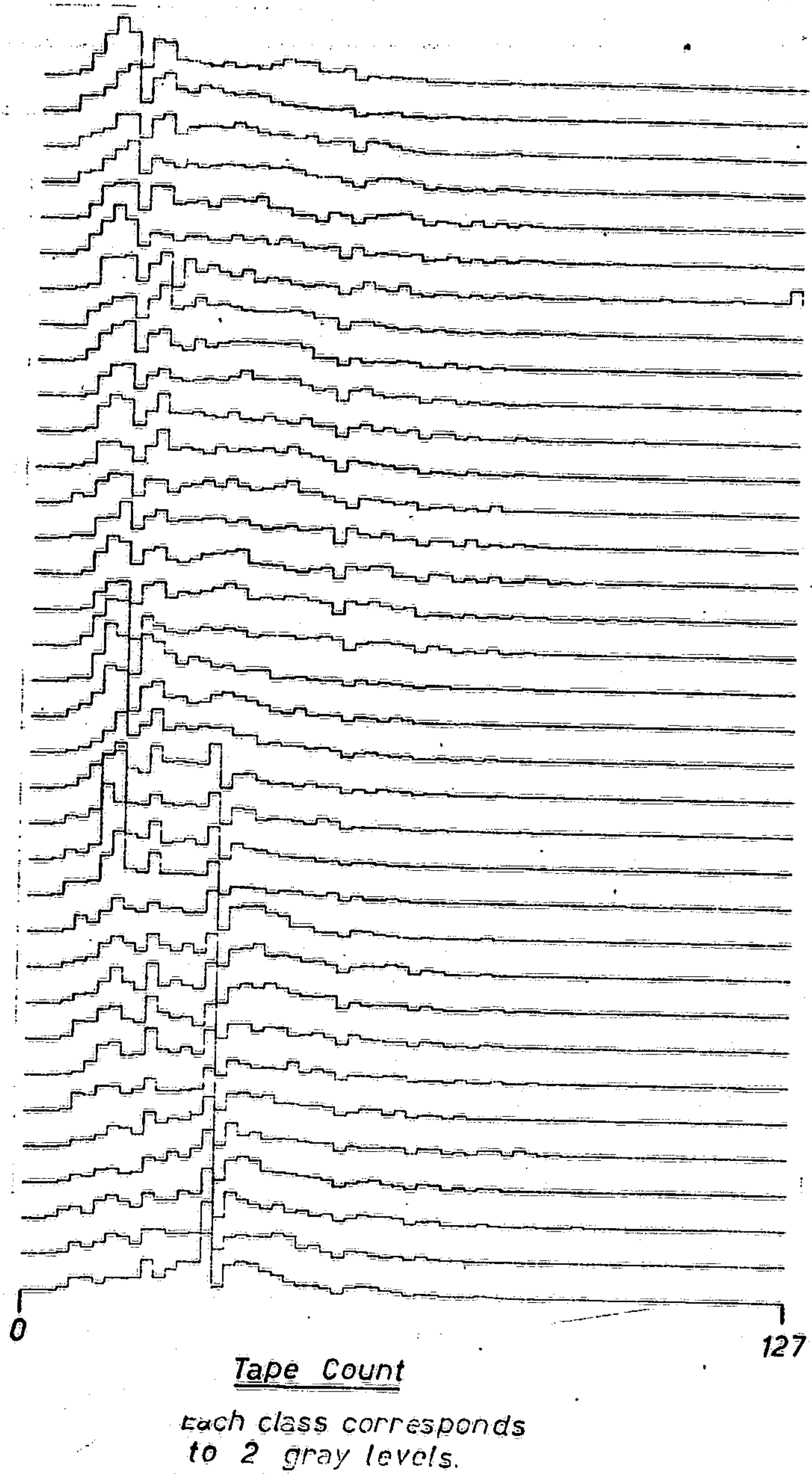


Figure 5. Gray scale histograms for successive groups of twelve scan lines for an area in the Sierra Nevada mountains.

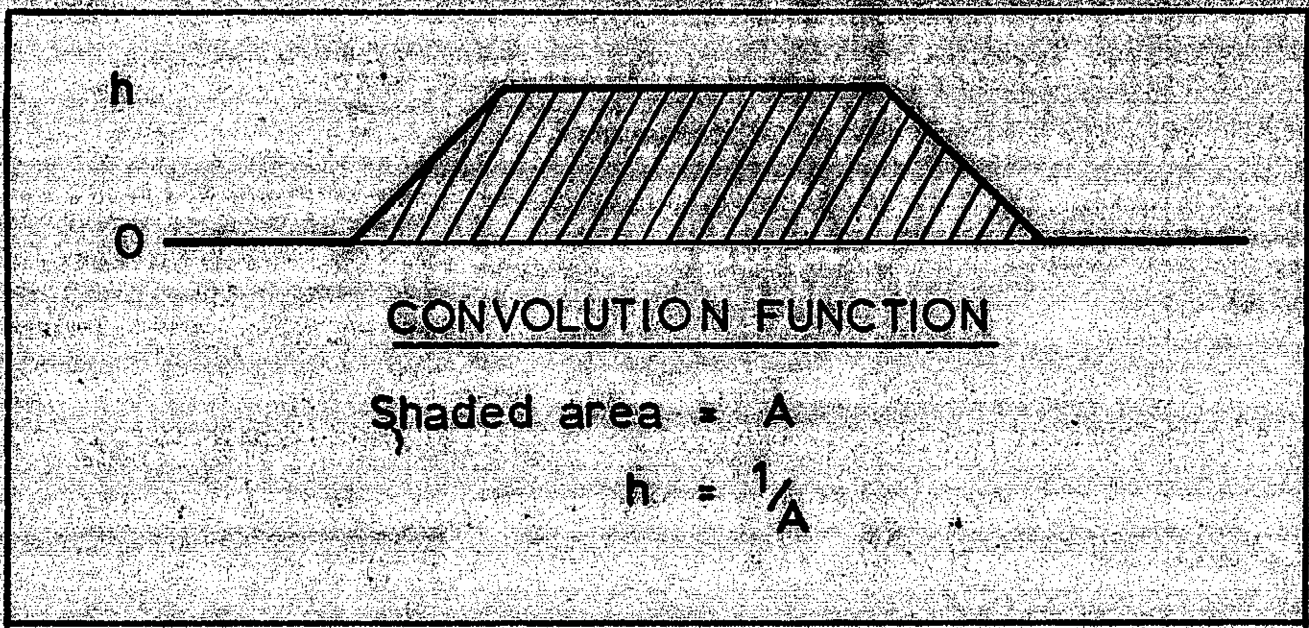


Figure 6. The convolution function used to remove the effects of uneven tape count probability distribution. The width of the function at half height is given by the equation for ΔT in section 3.

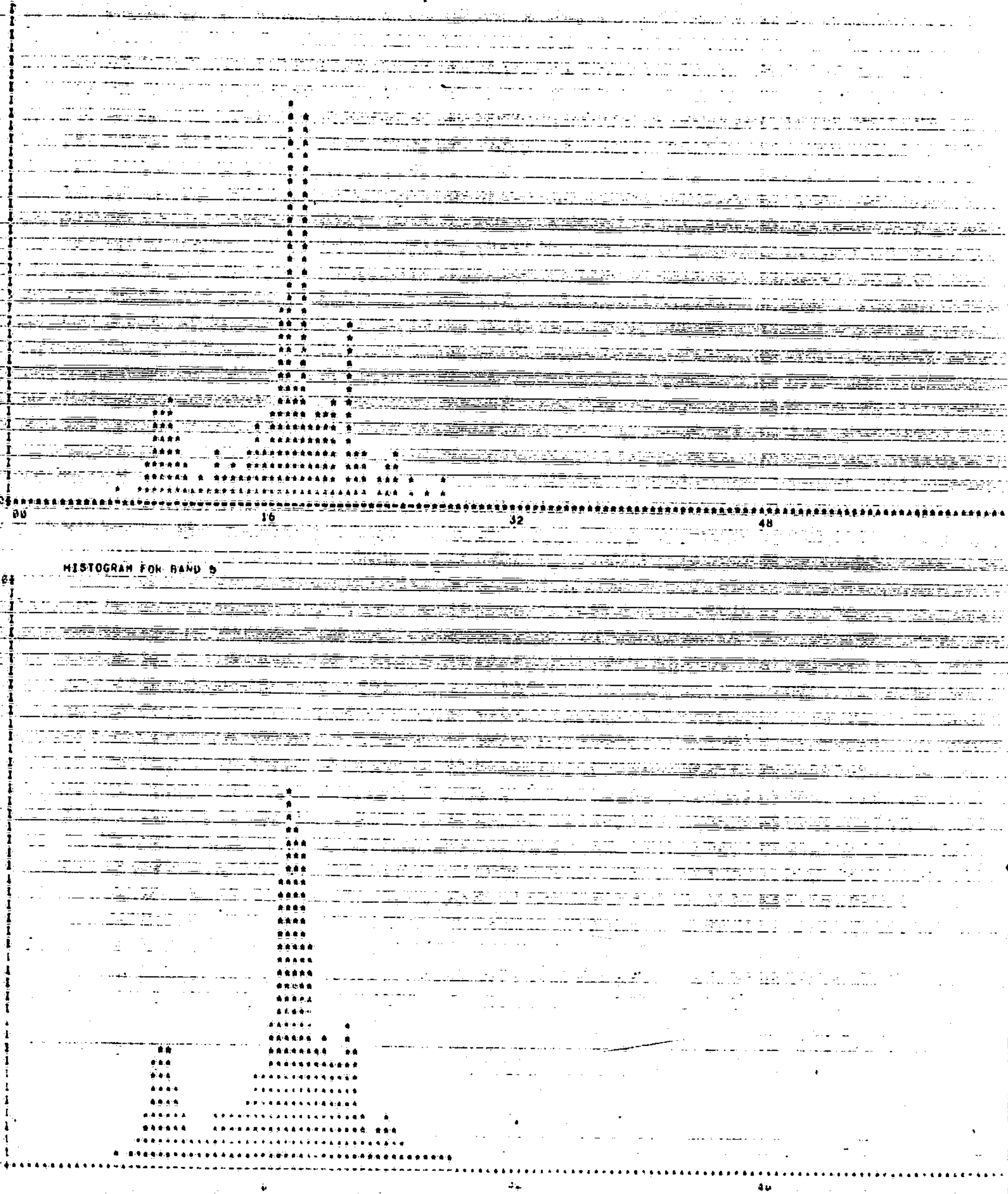


Figure 7. Smoothed and unsmoothed gray scale histograms for a UTA in the Central Valley.

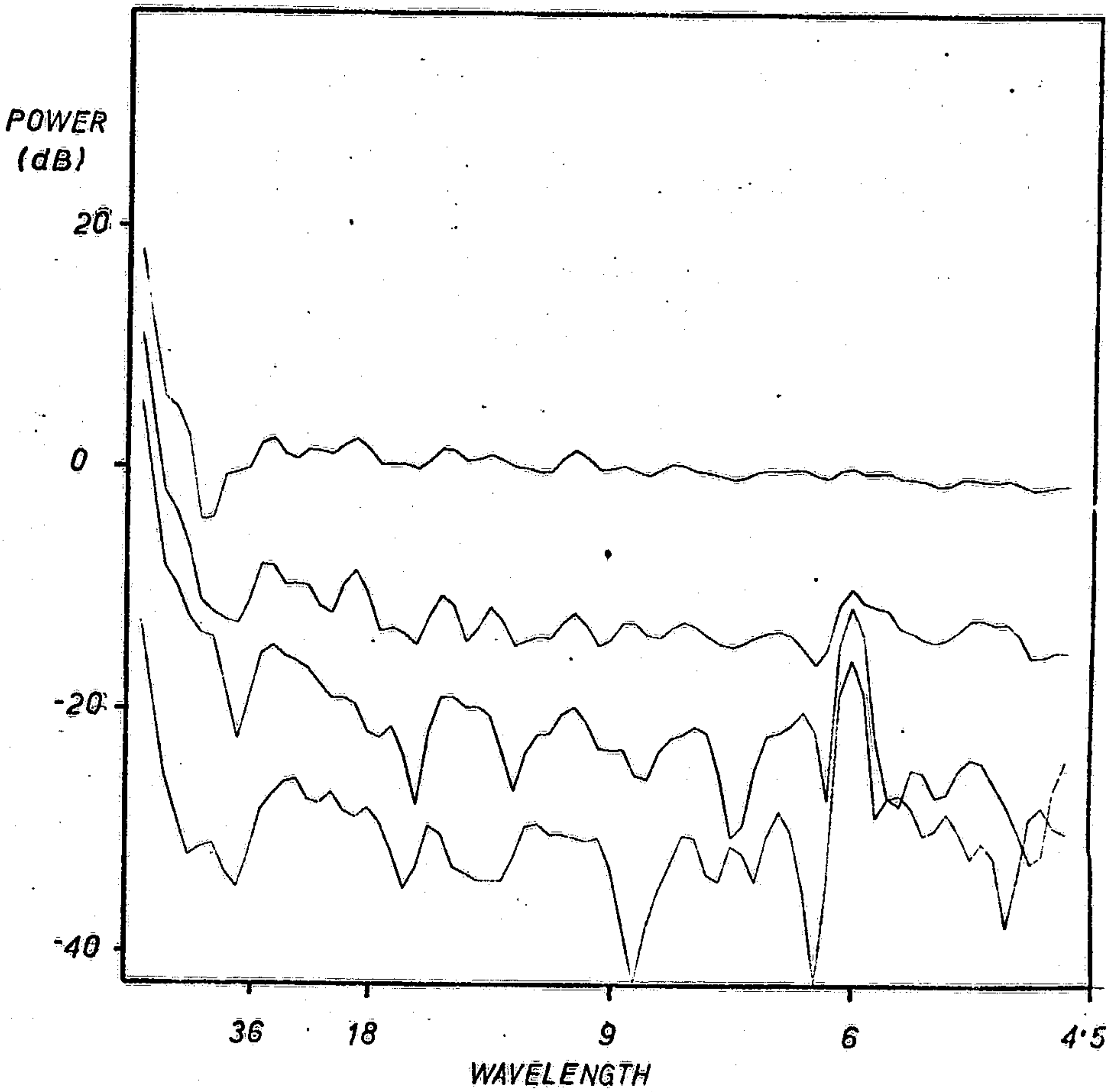


Figure 8. The power spectra for a strip one pixel wide and parallel to the edge of the frame for the sea area included in the U.K. CCT image. The spectra are plotted as a function of wavelength in scan lines, and the power is plotted as dB relative to the noise power in band 4.

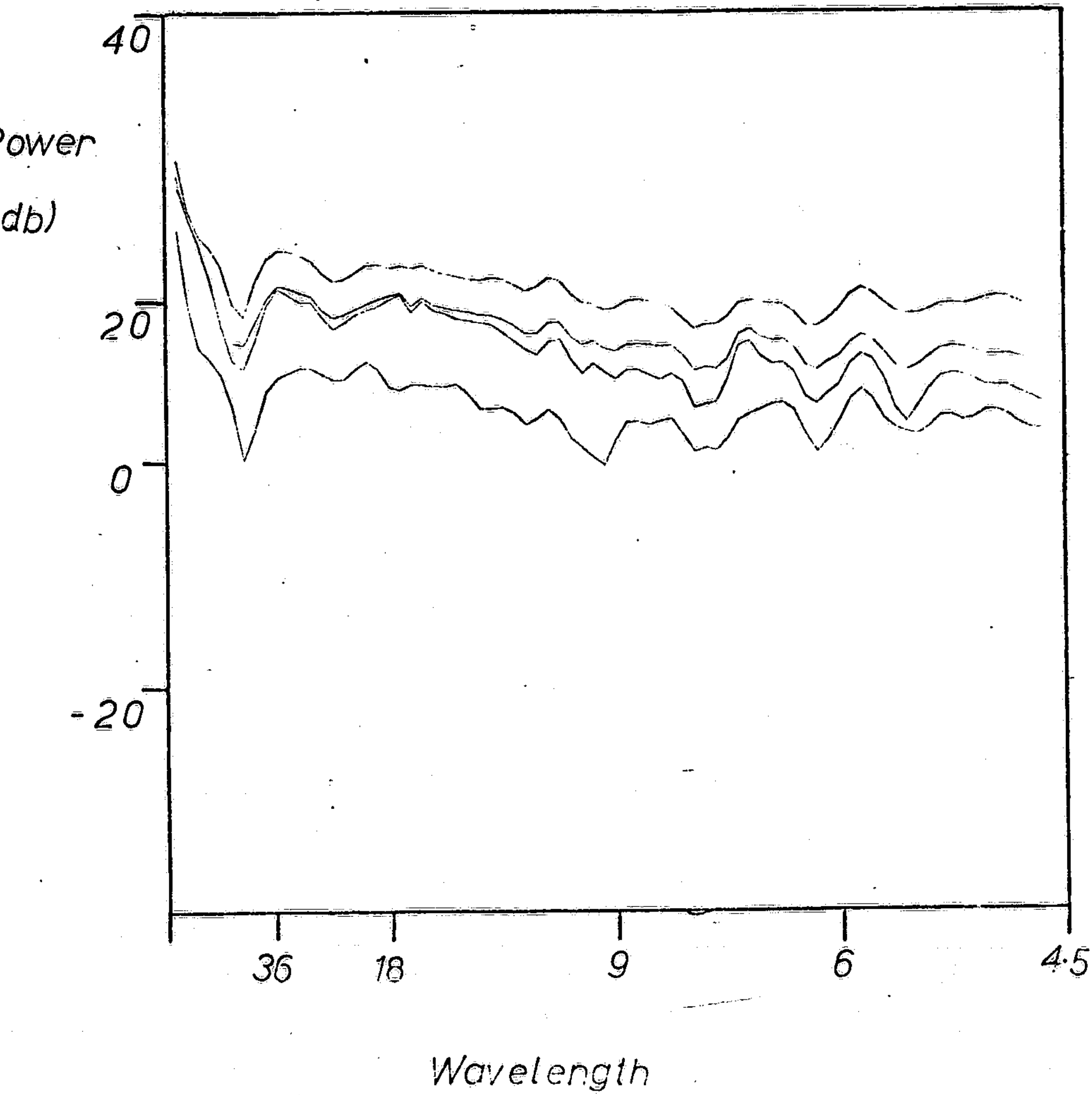


Figure 9. The power spectra for a strip through the Sierra Nevada range. Axes are as for Figure 8.

```
DOCUMENT RESULTS       , NORMCE390         : LP00 ON 20/03/74 AT 14.47
      SPLITTING OF ERTS DIGITAL HISTOGRAM - WASH BAND 7
            ESTIMATION OF PARAMETERS OF MIXED NORMAL DISTRIBUTIONS
          40     CLASSES IN INPUT HISTOGRAM
         200     DATA POINTS
       MAXIMUM NUMBER OF ITERATIONS
           8     STEEPEST ASCENT
          10     NEWTON-RAPHSON

                       INITIAL ESTIMATES
ALPHA          0.700
MU1            1.200
SIGMA1         0.500
MU2           23.000
SIGMA2         5.500
       MEAN OF INPUT HISTOGRAM          8.53
       STANDARD DEVIATION OF INPUT HISTOGRAM      10.6525
       HISTOGRAM CONTAINS 80787  DATA POINTS
   NO. OF DATA POINTS ORIGINALLY     200    ADJUSTED TO    210  BY CONVERSION OF HISTOGRAM
```

Figure 10. Sample output from the histogram splitting program.

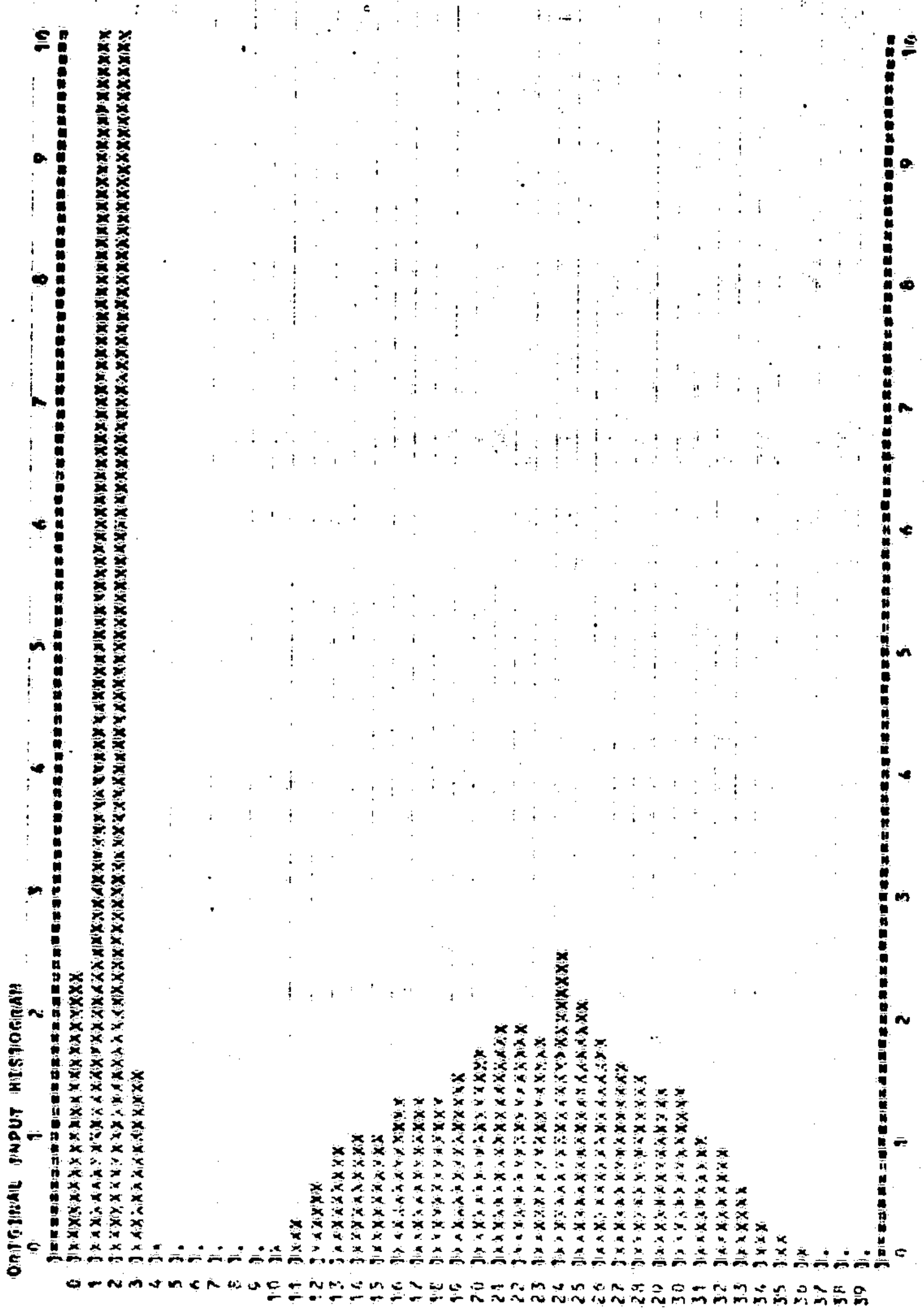


Figure 10 (part 2)

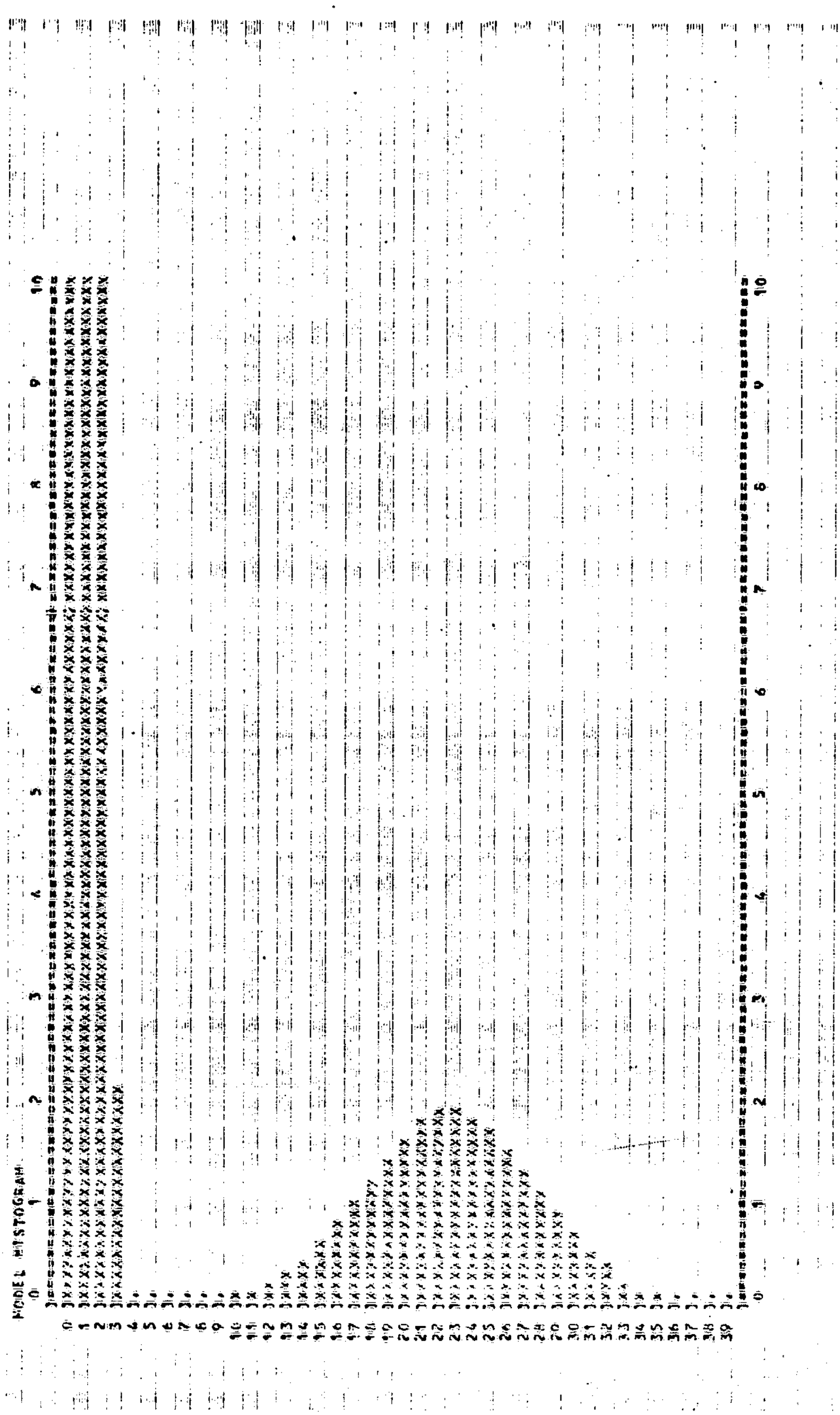


Figure 10 (part 3)

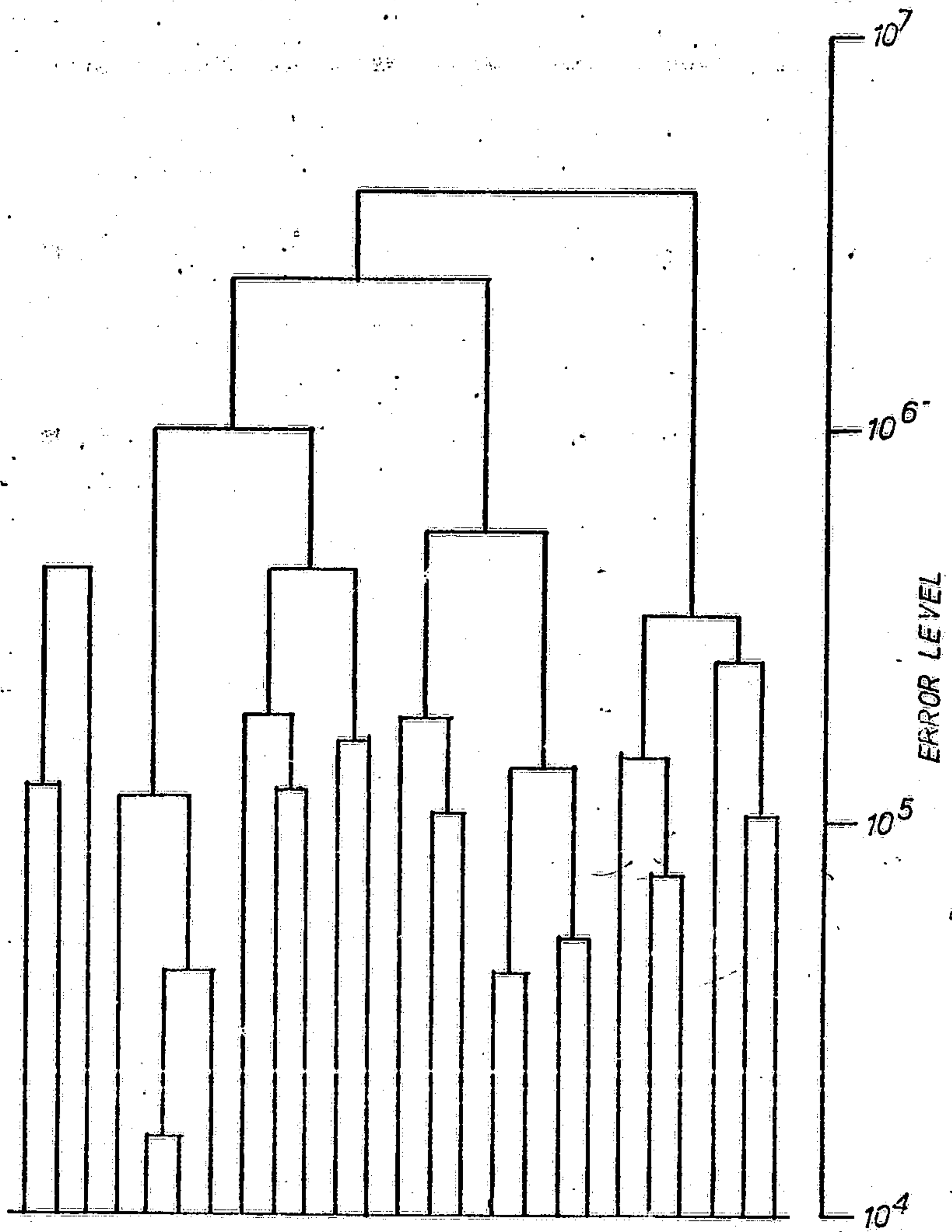


Figure 11. Dendrogram showing results of clustering the experimental set of OTU's. The error level is the square of the distance between the last two clusters amalgamated.

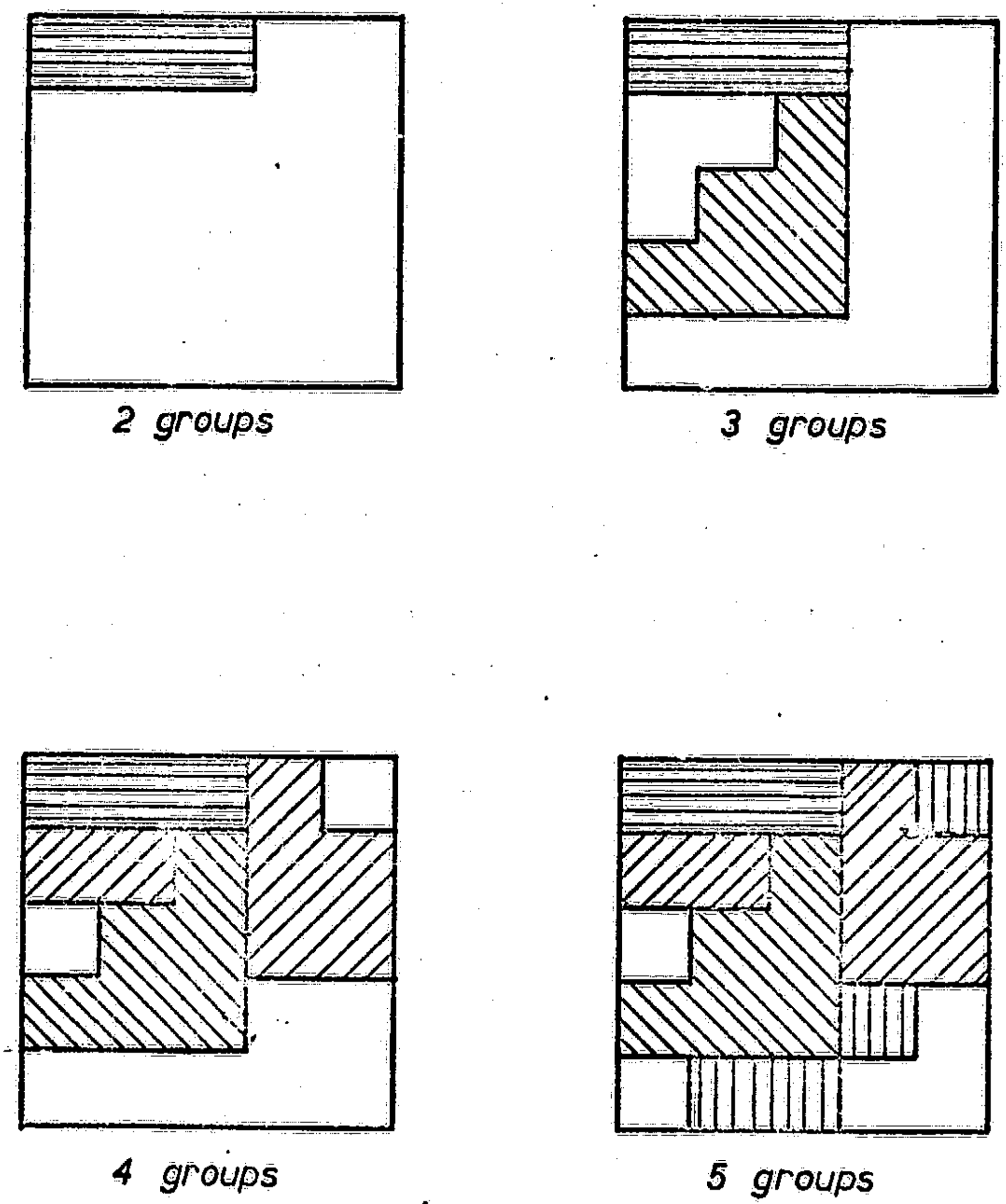


Figure 12. Maps of the groups produced in the last four steps of the clustering procedure.

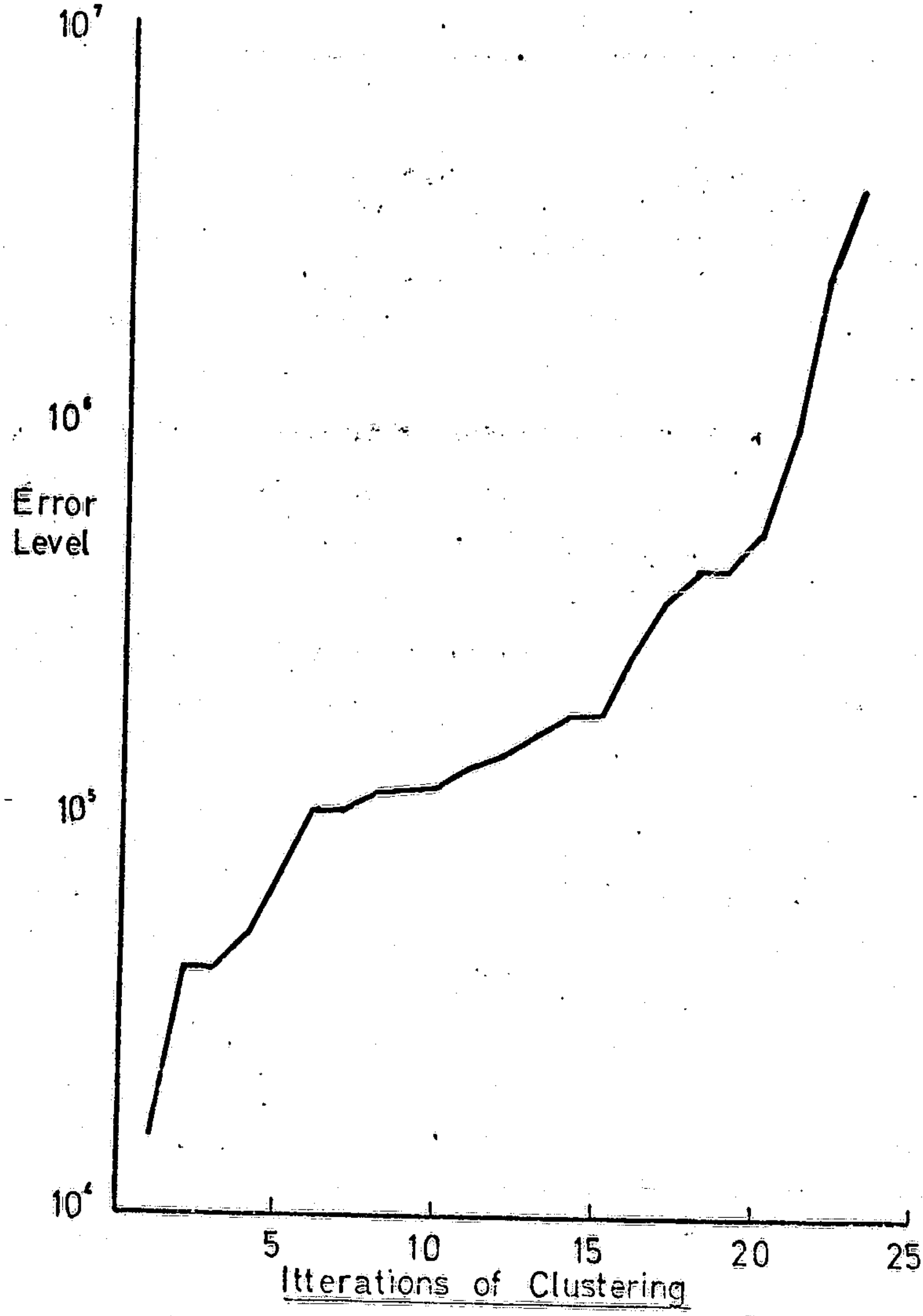
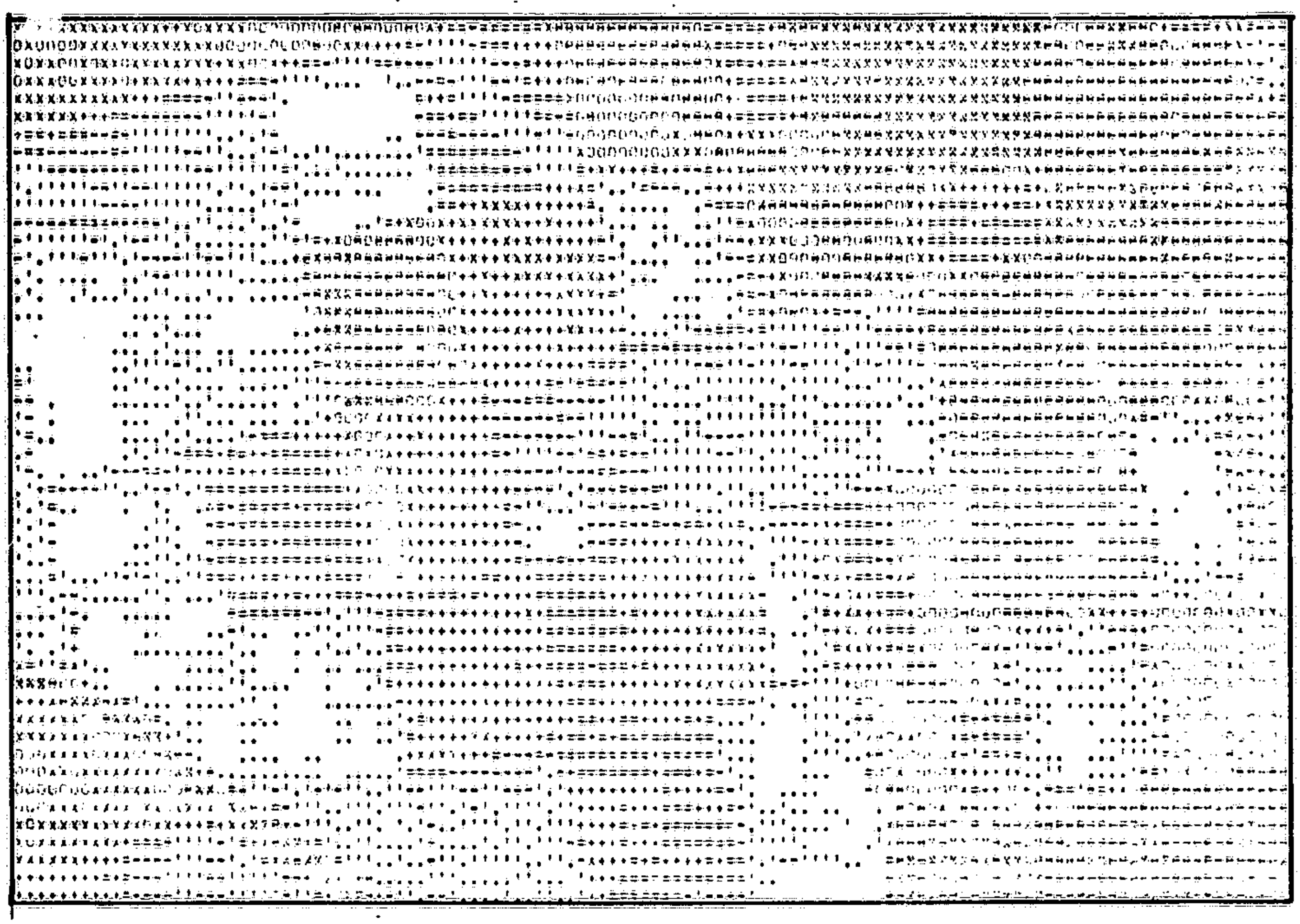
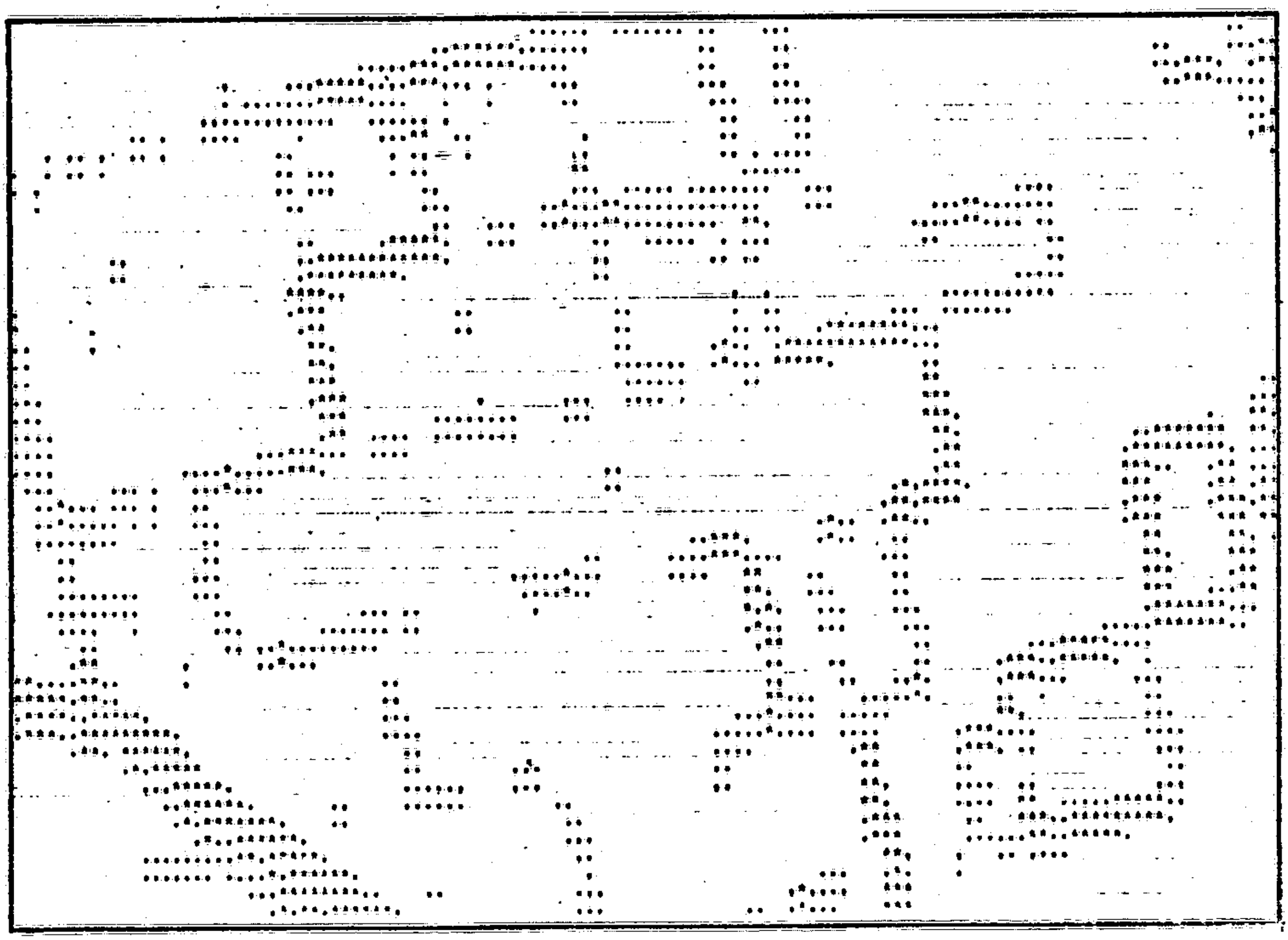


Figure 13. Graph showing the square of the distance between clusters as a function of the number of clusters.



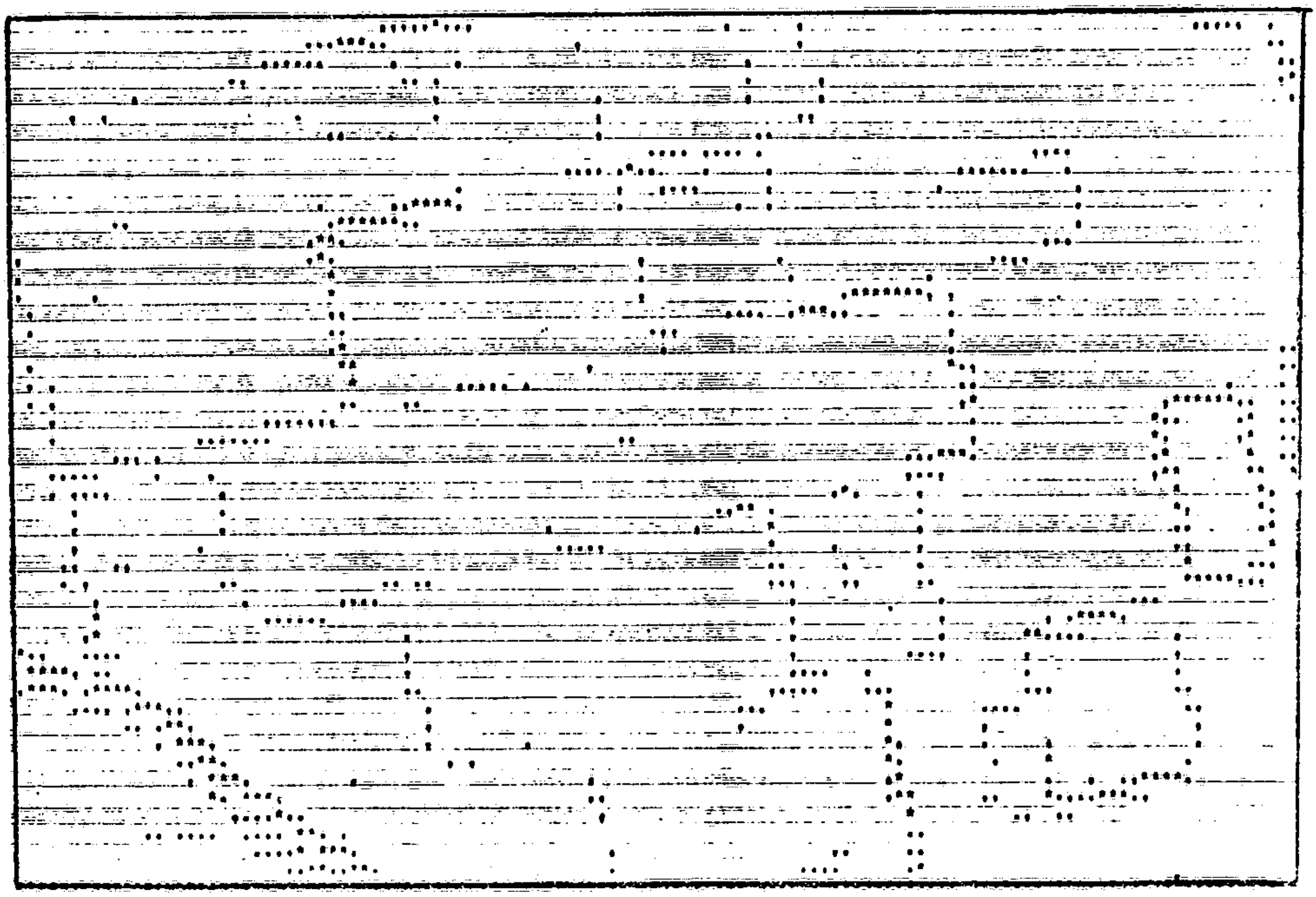
Grey scale map of area in Central Valley

Figure 14. A computer print out of an area in the Central Valley. The field pattern is clearly visible in this band 7 image and an irrigation canal crosses the south-west corner of the area.



Field boundaries using 8 nearest pixels

Figure 15. Sharp boundaries detected in the area shown in Fig.14 using the eight nearest neighbours method. A double threshold is used so that stars represent a larger difference in gray scale value.



Field boundaries using 3 nearest pixels in NW quadrant

Figure 16. Sharp boundaries detected using the three nearest neighbours in the north west quadrant. Note that the boundaries are more sharply defined than those in Figure 15, and that no significant features shown in Figure 15 are not also shown here.

Appendix: PUBLICATIONS & PAPERS OF THE PROJECT.

1. A computer based system for reading and generalising remote sensing data from ERTS-1. J. R. TARRANT, Proc. Eighth International Symposium on Remote Sensing of Environment, 1972, pp1425-1428.

2. Synoptic observations at a global scale. K. M. CLAYTON, Journal British Interplanetary Society, 27(1), 1974, pp 23-28.

3. A computerised system for identifying changes in the Earth's surface cover. I.E.HILL & J.R.TARRANT, Journal British Interplanetary Society, 27(1), 1974, pp 45-47.

4. Generalisation of ERTS data for global-scale investigations. I. HILL, A.C. ARMSTRONG & K.M.CLAYTON, in: European Earth-Resources Satellite Experiments, (Proceedings of the Frascati Symposium, 1974), 1974, pp 75-80.

5. Objective generalisation using ERTS images. A.C.ARMSTRONG & K.M.CLAYTON, Presented to the Second Bristol Symposium on Remote Sensing of Man's Environment. to be published in: Environmental remote sensing practices and problems, ed. E.C. Barrett & L.F. Curtis, (Edw. Arnold, London), 1975(?).

6. A common U.K. format for ERTS digital tapes? A.C. ARMSTRONG & I. E.HILL, Journal British Interplanetary Society, 28, 1975, pp 473-476.

7. Methodological questions in the digitised analysis of ERTS data. A. C. ARMSTRONG, Journal British Interplanetary Society, 28 (9&10), 1975, pp 608-612.

8. The relative performance of some unsupervised clustering techniques for the per-field classification of LANDSAT data. A. C. ARMSTRONG, presented to the Purdue Symposium on Machine Processing of Remotely Sensed Data, June 3-5, 1975. to be submitted to: Pattern Recognition.

9. Rapid digital mapping of coastal areas from LANDSAT. A.C. ARMSTRONG & P. BRIMBLECOMBE, forthcoming in Cartographic Journal, 1975.

195600-1-X

Reprinted From

PROCEEDINGS of the EIGHTH INTERNATIONAL SYMPOSIUM
ON
REMOTE SENSING OF ENVIRONMENT

2 - 6 October 1972

A COMPUTER BASED SYSTEM FOR READING AND GENERALISING REMOTE SENSING DATA FROM ERTS-1

J. R. Tarrant

School of Environmental Sciences, University of East Anglia
Norwich, Norfolk, United Kingdom

ABSTRACT

The paper presents an outline of a research project using data from ERTS-1. It involves reading the computer compatible tapes supplied by NASA for a limited number of scenes. Extracted from these are data applicable to a number of smaller areas (Unit Target Areas) of 50 km side and for each a frequency histogram is produced for the response within the four wave bands of the MSS system and the three wavebands of the RBV system. These histograms are then compared for successive orbits to detect and monitor change.

---

1. INTRODUCTION

As remote sensing platforms develop and become more remote from the surface of the earth that they are sensing, the resolution of the sensors becomes increasingly important. As the altitude of operation of the platforms has moved higher, from low flying aircraft to high altitude reconnaissance planes, rockets and balloons to finally the orbiting satellite, the ability to detect small objects has for the moment been sacrificed in order to scan larger sections of the earth's surface more rapidly. As many of the world's centers for research in remote sensing started in the field of interpretation of aerial photographs, normally flown at altitudes of less than 15,000 ft. and with camera and film systems which were fully recoverable, it is hardly surprising that the majority of research effort direct to the use of sensors mounted on orbiting platforms should continue to be concentrated at or near the absolute resolution of the system. Indeed some of the research proposals made to NASA for co-operation in the ERTS scheme would appear to be beyond this limit of resolution. Because an unmanned satellite must, at present, use non-recoverable methods of remote sensing the resolution of the imagery is determined largely by the fact that it is built up from scanning sensors, each scan line being broken up into a large number of units known as instantaneous fields of view. In the ERTS program the instantaneous field of view is of the order of a square 300 feet on the side. The response from such an area makes up an average response from all the surfaces within it. A typical research project involving the use of high altitude aircraft with multispectral scanners might involve sampling a number of small areas in the imagery and producing spectral response signatures from these samples and matching them to known signatures in order to allocate each sample plot to a known type of land use or land surface. Such work, with also visual interpretation of the photographic imagery built up from the scan line data, continues in many forms with the ERTS project. There are two important considerations in using this data from ERTS:

1.1 It is obviously important to adapt the analysis of the remotely sensed data to the scale of the data gathering system. A satellite system such as ERTS is perhaps a forerunner to a world-wide data gathering and monitoring system. On this scale looking at instantaneous fields of view would seem rather like looking at needles in haystacks.

1.2 The potential volume of data from a satellite based system is very large. The data acquisition rate from ERTS-1 alone, with its limited real time and recording capability, is estimated at 15×10^9 picture elements a week for a year (Erikson, 1972). Repetitive coverage is an important

58-1

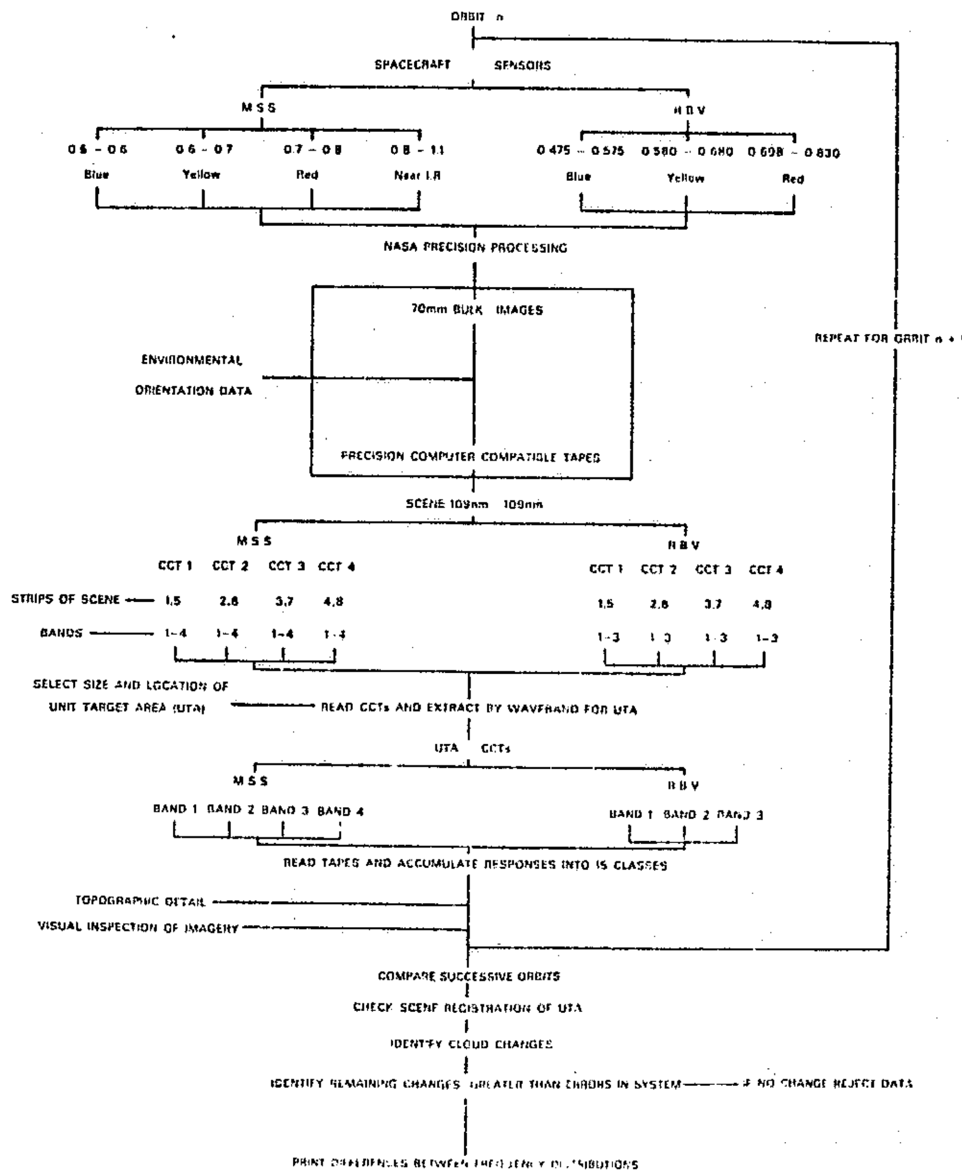


FIGURE 1. DATA FLOW

58-2

asset of a satellite system and therefore an effective sampling structure is difficult to establish and it is valuable if all the available data can be scanned to detect change over successive orbits.

2. THE OBJECTIVE OF THE RESEARCH PROJECT

2.1 With many research establishments having a long and distinguished history in the field of remote sensing they have accumulated necessary capital equipment and expertise to continue to develop research efforts along traditional lines. The first objective was to identify a new direction for research with satellite data which would not require large investments in capital equipment and manpower and yet at the same time would not be repeating experiments completed elsewhere.

2.2. Remote sensing data have little intrinsic value until they are used by someone else to make decisions or to facilitate action which benefits man (Schrumpf, 1972). The second objective was to devise a system making remote sensing data more readily available for such uses. The data needs to be available in areal units which have a scale appropriate to a data gathering system for the whole world. As speed is essential, not only to potential users but also because of the 18 day ERTS cycle, it may be desirable for the visual inspection of the photographic imagery to be circumvented, and produced only for those areas which are picked out as interesting by an automated data scanning system. The design of such a system became the second objective of the research.

3. THE DATA SCANNING SYSTEM

3.1 THE DATA USED

The precision computer compatible tapes, containing data from the four MMS bands and the three RBV bands, or such of these as are available, are used as the data input (fig.1). The precision data is required because of their greater radiometic accuracy and because the scene they represent is spatially registered on the basis of ground information leading to a reduction in marginal errors when identifying the same area of ground over successive orbits. On the other hand this goes against one of the objectives of the research, to avoid the photographic imagery, as the precision computer compatible tapes are formed by scanning the 9 inch square plates resulting from the precision processing. With ERTS-B we hope to use some of the bulk data but at this early stage of development accurate scene registration is essential. The eight computer compatible tapes are available for each scene covering approximately an area 100 nautical miles square and these contain all the MMS and the RBV data. These tapes have to be read and searched to find the data referring to the selected areas of the earth's surface.

3.2 THE UNIT TARGET AREAS

The scanners, by which the precision imagery is searched to create the computer compatible tapes, have an instantaneous field of view which is called a picture element. It is the intention of this project to combine these picture elements into much larger areas (Unit Target Areas) and to produce the average spectral response within these areas which can then be compared over successive orbits. The size of this unit target area will be a matter of experiment but it needs to be an appropriate scale for mapping large areas of the earth. Initially the UTA has been fixed at 50 km square and will be so selected that each falls within one scene examined by ERTS-1. Two such unit target areas are being used in the first instance, one in the United Kingdom and the second in Central California. This second area was chosen as an area with less cloud cover than in the United Kingdom and as an area where there would be little difficulty in obtaining and interpreting the ERTS-1 data.

3.3 UNIT TARGET AREA DATA ANALYSIS

The computer compatible tapes are scanned for each unit target area and the responses for each wave band for each sensor are accumulated into 13 class intervals. The resulting seven histograms will be printed using a CRT display unit. These histograms become the basis of all later work as they form a type of spectral response signature for the unit target area. Known radiometric errors in the sensors and in the precision processing can now be used to give confidence bands to each of the histograms. The computer compatible tapes for later orbits will then be examined and the unit target area data extracted after careful scene matching. This second, and all successive, response histograms can be compared to the first to isolate changes in the spectral response of the unit target area over time.

3.4 IDENTIFICATION OF CHANGES

If the histograms show change greater than the confidence limits the photographic products can be examined together with ground truth to discover the nature of the change. The most obvious change will be in the amount of cloud cover. Changes in the seven histograms which are a response to changes in cloud cover can be isolated partly by visual inspection and partly by a separate examination of cloud free sections of the unit target area. Once the nature of changes brought about by cloud cover are recognised future analysis will be able to include a calculation of the percentage of cloud cover automatically while looking for other significant changes in the responses. Should no such changes be recognisable then the data for that orbit can be rejected, the photographic evidence not requested, and the next orbit can be compared directly with the first.

4. APPLICATIONS

No data has been analysed so the results are those anticipated both in the long and the short term from such a system. This research program is designed to provide a method of analysis, not to produce directly usable results from the data. This is in keeping with the experimental nature of the ERTS project. In order to provide for as wide an application as possible the software is being produced for an IBM 370/165 computer system using the nine track computer compatible tapes. Future applications for this research are numerous and include the identification and monitoring of large scale environmental changes stemming from such things as seasonal crop changes and burning, the development of irrigation programs, pollution of coastal waters and many more. As the unit target areas are small and rectangular they lend themselves to computer mapping in various forms. For some purposes the size of the unit target area may be changed and the computer programs will allow for such alterations. Eventually an environmental watch could be preserved over large sections of the globe without having to process the photographic products continuously. Such long term developments assume that future satellite systems have better scene registration without special precision processing or that this processing will not involve the preparation of photographic products and it also assumes that large scale processing of computer tapes will be possible at far greater rates than at present. Such developments do not seem unlikely along with increasing accuracy of the sensors in use.

REFERENCES

Erikson, J. D., 'A summary of Michigan program for earth resources information systems', 4th Annual Earth Resources Program Review, NASA, 1972, 28-1.

Schrumpf, B. J., 'Resources analysis and land use planning with space and high altitude photography', 4th Annual Earth Resources Program Review, NASA, 1972, 116-1.

ACKNOWLEDGEMENTS

This project is funded by a research contract from the United Kingdom Department of Trade and Industry.

Journal of the British Interplanetary Society, Vol. 27, pp. 23-28, 1974.

SYNOPTIC OBSERVATIONS AT A GLOBAL SCALE

KEITH CLAYTON
School of Environmental Sciences, University of East Anglia, Norwich, NOR 88C, England.

• Earth observation satellites are known to provide a basis for mapping both permanent and more ephemeral features of the Earth's surface. While emphasis so far has been on their potential for detecting hazards such as floods, forest fires, etc., it is here suggested that their role in providing a uniform data base for mapping global distributions has been underestimated.

1. INTRODUCTION

OVER THE PAST few years talk of what might be done with satellite imagery of the Earth's surface has greatly outweighed practical applications. To some extent this situation has continued longer than we might have anticipated, with the delayed launch of ERTS 1 and the slow arrival of imagery in the UK. I apologise for adding to this class of literature, but I feel that, despite its volume, it has left considerable areas unexplored.

2. PHOTO-INTERPRETATION

So far two major areas of application have been examined; the scene being subjected to a suitable modification of conventional air photo interpretation. Details of land use, land form, and in suitable areas such elements as geology or soil, are interpreted and mapped from the scene, using either a single wave-band image or perhaps more effectively some form of combined image such as that produced by an Addcol viewer. In little-known areas new information is readily obtained by this approach, while even in well-mapped areas new developments that post-date the latest maps will be found, or ephemeral features such as floods or savanna burns may be mapped. Most of this information is produced by interpretation at or near the limits of resolution of the imagery, and the investigator is aware that improved resolution would add greatly to the information he could extract.

A more novel aspect of this approach is the occasional value of the small-scale and large area covered of the imagery for revealing features of such a size that they are not readily appreciated on conventional aerial photography. The most obvious and common examples are the recognition of major linear geological features, and a number of examples are already recorded in the literature.

3. MONITORING CHANGE

The second major application of satellite imagery of the Earth's surface is the value of its repetitive coverage in monitoring changes over time. There is no denying the fact that even the 18-day cycle of ERTS 1 allows the detection of relatively short-term changes such as floods or forest fires, as well as the surveillance of seasonal changes in vegetation (natural or cultivated) and such things as snow cover. Comparison of successive scenes will give rise to an impressive literature and some interesting (and potentially useful) examples are already on record. An example is the chance occurrence of a major savanna fire in the northern

58-5

part of South Africa on successive passes of ERTS 1. Since the fire occurred in the area covered by two adjacent scenes, change after one day was observed. We already have satellite scenes showing major floods, and will soon add valuable records of ephemeral vegetation growth in desert areas (potential sources of locust infestations) or a record of the extent of landslides in some mountainous area after a major earthquake.

Such successes must not blind us to the problems of regular surveillance of change by satellite imagery. Were ERTS 1 technically capable of monitoring every scene and returning the data to Earth, even this system would produce nearly half a million images a year. For floods or forest fires, coverage on a one-day, rather than an 18-day, basis would really be needed, and this would put up the annual total of scenes to ten million. In addition, there are currently four images for each scene, further increasing the number of images involved in global surveillance. The time taken to receive and look at these images must also be taken into account. If the returns are worthwhile a system can be of course developed, even if it consists of no more than a vast team of trained photointerpreters. I have not considered the additional complications of cloud cover at the scale of real-time global monitoring: we can suppose that new techniques will come along to allow penetration of cloud.

Whether we are considering a monitoring programme *per se* or are just facing up to using the voluminous output of an Earth satellite system, there is a strong case for automating our initial reconnaissance of the output. This is the first objective of work in progress at the University of East Anglia [1]. It rests on the supposition that, once one good quality scene is available for an area, we are only likely to want to spend time looking at other scenes if they are different, if they show seasonal or other more irregular changes. If we can monitor the satellite output by machine, or by machine-aided techniques, we have the chance of identifying duplicate (or cloud-covered) scenes, and so may greatly reduce the number of scenes which deserve more detailed work. This is data-reduction by the recognition and elimination of duplicate data.

4. GENERALIZATION OF SATELLITE IMAGERY

This review has covered what I see as the two main approaches we have so far seen in the utilization of Earth resources satellite imagery. In each case the data is reduced to a manageable amount, either by limiting interest to a restricted area or by concentrating on change and discarding those features that are fixed. It does not seem to have been appreciated that these two approaches do not provide information that may readily be handled at the continental or global scale. Yet surely we will be failing to exploit the most fundamental feature offered by orbiting satellites if we fail to work at a continental or global scale alongside the massive amount of more detailed interpretation already underway. I shall not be concerned with the usefulness of such generalization but rather with the techniques by which global synthesis might be brought about. I am quite convinced that generalized mapping and interpretation at continental and global scales is a most worthwhile aim, and have no doubt that the new information which can be extracted from satellite imagery will open up new areas of research and allow new approaches to environmental relationships at and near the global scale.

What we must first understand is the need to generalize and reclassify the images we obtain from orbiting satellites if we are to work at the global scale. The scenes are of course generalized to some extent by the scale of the imagery, and this is reflected in the limits of resolution which makes it difficult, for example, to discern city blocks and of course quite impossible to see individual buildings. Yet in suitable areas, where fields are large, individual fields can be seen, so that the imagery links to some extent with our ground-level (or air photo) view of our bit of the world as a mosaic of fields and built-up areas. Even this suggestion of a field scale suggests that at the margin of interpretability the ERTS 1 images

58-6

could yield data mappable at about 1:100,000. The small fields typical of England are identified on our 1:25,000 maps, but as the pre-war sheets of the Land Utilization show, land-use may be mapped at 1:63,360 on the basis of undistorted representation of field units as mapped at a larger (1:10,560) scale. I am not suggesting that there are many landscape elements that might be extracted from the ERTS 1 imagery at a 1:100,000 scale, but this seems a reasonable estimate of the largest scale at which current ERTS 1 imagery might be mapped.

If we turn from the consideration of large scales to the world scale, we move from 1:100,000 to 1:50,000,000, which is about the largest scale on which the entire world might be shown on a single sheet. This scale shift of 500 times carries with it a number of implications. First an individual ERTS 1 scene is going to occupy an area of no more than 3-4 millimetres square, so that in practical terms it will hardly be possible to divide a single scene into more than about 3 categories. Indeed, even the recording of a yes/no situation (e.g. presence or absence of a visible area of surface water) will produce a map with as much information as a botanist's distribution map (Fig. 1) based on presence or absence

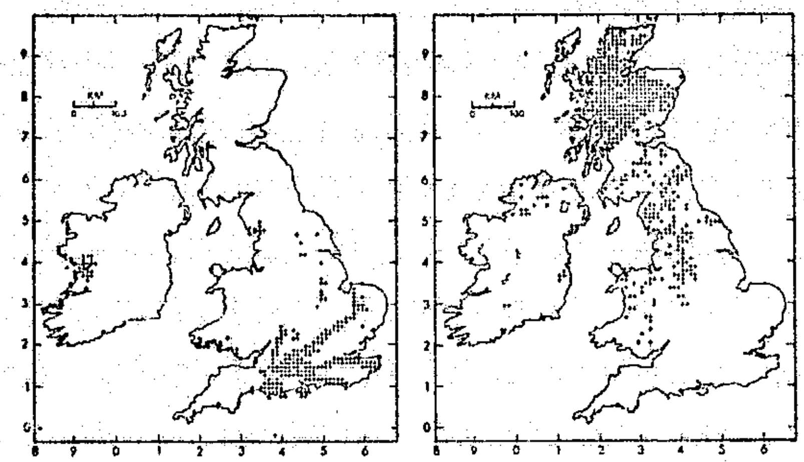


Fig. 1. Distribution for two plants plotted by 10km grid squares on a present/absent basis. On the left *Asperula cynanchica* L. is closely correlated with limestone outcrops, and in particular the chalk of south eastern England. On the right an upland distribution is represented by the distribution of *Vaccinium vitis-idaea* L (Cowberry). (After *Atlas of the British Flora* ed. F. H. Perring & S. M. Walters, 1962).

of a particular plant in grid squares with 10km sides. ERTS 1 imagery is far too incomplete to allow global mapping even of this simple type, but it will become feasible with some future satellite systems. We should look ahead to that system and start thinking about what sort of distributions would be worth mapping at the global scale, and what data-processing system will extract and map the information from about 20,000 scenes (6000 for land areas only) most efficiently.

5. GENERALIZATION AT THE GLOBAL SCALE

A second aspect of mapping at the continental scale is the classification of the information portrayed by Earth resources satellites. For most of the phenomena we are likely to map, the small scale of the ERTS imagery will involve the mapping of associations of elements, and these associations will tend to become increasingly diverse and hence increasingly generalized as descriptive units at smaller scales. We are familiar with this in terms of thematic maps where generaliza-

58-7

tion and a classification of the data are both involved when map scale is reduced. We have already referred to Land Use, portrayed as field units on the 1:63,360 scale. The '10-mile' (1:625,000) Land Use map of Great Britain was able to use the same classification as the 1:63,360 map, but had to agglomerate small parcels of each category into irregular dots whose area was balanced to yield the correct proportion of each land use category across areas about a centimetre square. The same approach cannot be used to produce a map at any smaller scale, and a land use map of Western Europe at (say) 1:5,000,000 must adopt a new classification based upon Land Use types or complexes. At a still smaller scale, such as a world map, it would be difficult to sustain more than two or three subdivisions within the cultivated land category, and the rest of the subdivisions would necessarily lean heavily on the major natural types of vegetation, which are attractive units at this scale because they cover wide areas and yet we can conceive what they look like. However, many of the distributions we are likely to discern on ERTS 1 imagery do not fall into such convenient categories and we shall have to build up complex units that can be expressed in quantitative terms.

For some purposes these synthetic maps are effective illustrations of the distribution of complex natural phenomena – soils, vegetation, landform, for example. But they are at the same time difficult to compile objectively and perhaps even more difficult to analyze. My own preference is for the compilation of single variable maps, either on a present/absent basis as already suggested for open water, or on a simple quantitative scale (Fig. 2). While some of the possibili-

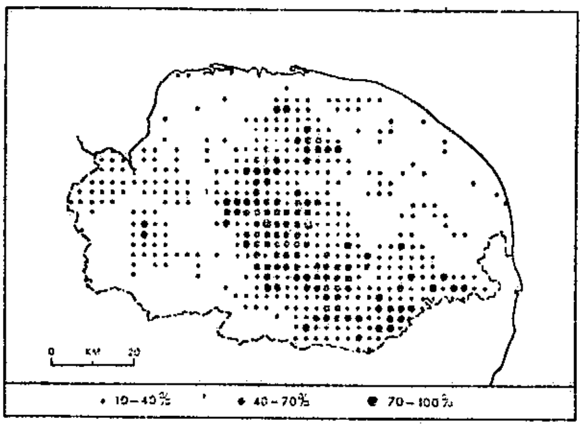


Fig. 2. The frequency of pits and ponds in Norfolk plotted in four groups on a semi-quantitative basis. The percentages refer to the proportion of sample areas containing pits or ponds (After A. R. Cartwright, unpublished MPhil thesis, University of East Anglia, 1972).

lities seem rather simple at first sight (such as my open-water example) we may still have a lot to learn from the assembly of this information in a uniform and consistent way for the whole world. The analagous map of biological distributions I have already referred to reveal far more than might at first be expected from data that records no more than the presence or absence of a certain species over units of 100km^2. Part of the value will be the consistent nature of the data base over the whole globe (Fig. 3). Existing world maps depend on a data base that is unnecessarily complex and detailed in a few places and almost non-existent over wide areas. With the best will in the world the compiler cannot eradicate these inconsistencies. It is true that ERTS 1 data will be most easily used in areas where 'ground truth' is best established. Nevertheless, it provides us with a far more consistent data base than we have ever had before.

58 - 8

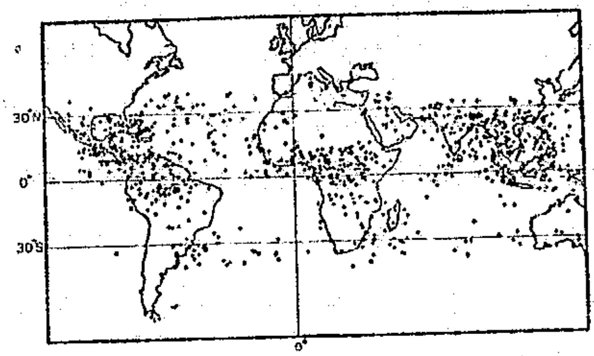


Fig. 3. Lightning: a distribution map for tropical and sub-tropical areas. The data was derived from computer analysis of night-time records from NIMBUS II, 1-15 July, 1966. There is some variation in the frequency of pass with latitude, but for any one latitudinal belt coverage is uniform. This is the first time that it has been possible to map lightning without distortion imposed by the varying density of observing stations. (Redrawn from *Nature.* 232 (5312), 20 August, 1971).

Finally an important feature of any global mapping or satellite-derived data is the opportunity of repeated mapping to pick up seasonal or annual changes. I discount the possibility of concerning ourselves with shorter term changes, partly because of the data-handling problems I have already discussed, but partly because it is important to match scales in time and space. Such generalized changes may usefully be referred to as synoptic, and represent a potential in world-wide mapping that is entirely new. The nearest approaches are in the pioneer generalized cloud cover maps produced by the repeated superimposition of meteorological satellite images or the mapping of parameters such as albedo from radiometers on meteorological satellites (Fig. 4). There is no doubt that they contributed a new view of the global circulation and a new analytical device. I feel it would be a pity if the inevitable fascination with the margin of interpretability of satellite imagery led us to neglect the potential of global synthesis.

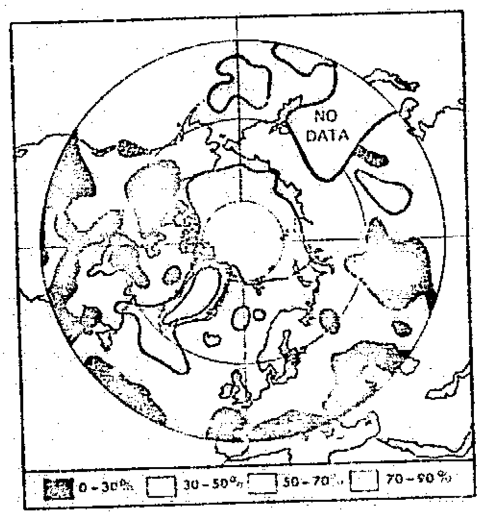


Fig. 4. Albedo mapped for a large part of the northern hemisphere on the basis of data from the satellite Nimbus II. The uniformity of the data base, regardless of land/sea differences or latitude, should be noted. (Redrawn from *Journal of Applied Meteorology*, 9, 1970).

58-9

Keith Clayton

REFERENCE

1. I. E. Hill, J. R. Tarrant, A computerised system for identifying changes in the Earth's surface cover.

(Presented at the Symposium of the British Interplanetary Society on 'Earth Observation Satellites' held at University College London, 10-12 April 1973)

58-10

Journal of the British Interplanetary Society, Vol. 27, pp. 45-47, 1974.

A COMPUTERISED SYSTEM FOR IDENTIFYING CHANGES IN THE EARTH'S SURFACE COVER

I. E. HILL and J. R. TARRANT
School of Environmental Sciences, University of East Anglia, Norwich, NOR 8CC, England.

The need for and design requirements of a rapid data filtering system for selecting satellite imagery containing significant new information on the Earth's surface cover are discussed. One possible method of performing this filtering is described.

1. INTRODUCTION

THIS PAPER DESCRIBES a computer based system we are developing for the analysis of imagery from the ERTS-1 project. Our aim is to test techniques for the automatic recognition of changes in the Earth's surface cover and to develop these techniques to a state in which they could be incorporated as a 'User Service' in an operational remote sensing facility. In this context 'Users' can be regarded as those who would utilize Earth imagery as an additional source of information to aid, for example, planning and administrative duties, in much the same way as meteorologists use NIMBUS satellite data. A water supply authority might combine satellite imagery with *in situ* measurements to estimate the snow cover in the catchment area of a dam, and use this estimate to decide whether water should be released from the dam.

Although the computer system had been designed around ERTS-1 and the data products made available by NASA, there is no reason why the techniques should not be extended to any other satellite based remote sensing platform, and for that reason the discussion in this paper will not be restricted to the ERTS system.

2. THE NEED FOR A DATA FILTER

The applications of the data from an operational remote sensing satellite will be determined by the advantages of this imagery over that derived from other platforms. This might seem an unnecessary statement but surprisingly little work is being done to realise the potential of one of the major advantages of satellite imagery, namely regular repetitive cover of large areas of the Earth's surface. Satellite platforms are well suited to monitoring changes on a regional scale, such as, for instance, crop ripening or the spread of crop disease in a large agricultural area. If information of this kind is to be of practical use to administrators, it must be available in a compact form and, most importantly, in near real time, both of which imply the need for an efficient data reduction system. We are concerned here with only the first stage of this process, an initial filtering stage for selecting for further analysis only those images which show significant changes in the Earth's surface cover.

The choice of a computer based system rather than a manual one is dictated by the quantity of data an operational satellite platform could produce and by the format of this data. A satellite covering the whole of the U.K. once a week with imagery in ten different spectral bands would require an analysis rate of 30

58-11

or 40 images daily. Much of this information will be irrelevant to any particular user not only by virtue of its geographic locality or spectral band, but also because it is effectively identical to earlier imagery. Rejection of images with no new information requires a well defined criterion for deciding what constitutes a significant change and this criterion must be easily updated to allow for different fields of interest of different users and variations in the type of change expected with, say, season or locality. Provided a fast and simple algorithm can be made to implement this criterion, a conventional computer can easily perform this type of repetitive analysis on a large data base. A team of trained photo-interpreters would do the selection as well as a computer, if not better, but only at the expense of many man-hours which could be better employed in detailed analysis of the imagery.

A more important reason for using a computer for the first stage analysis is the digital nature of the raw data. Most modern sensors are capable of providing digital data either in place of, or alongside, photographic images and transmission of data from the satellite to the ground receiving station will almost certainly be in a digital form. There is little advantage in making hard copy of the image before deciding whether it will be needed, especially as many of the changes of interest will not be easily detected in a single band but will require colour composites of several bands. In addition digital data permits a greater sensitivity in the identification of change as it allows use of the full dynamic range of the sensors with none of the compression of the grey scale produced by photographic techniques.

One final advantage of using a computer-based first stage lies in the generalisation of data to a regional scale. The algorithm for identifying changes can be made more sensitive to small changes in brightness over a large area than to large changes in a localised area, but the eye is particularly sensitive to contrast between adjacent areas and cannot easily compare large areas.

3. THE DATA FILTERING ALGORITHM

There is obviously more than one way to perform this initial filtering for change but only one possible method will be considered here. Initially we will be concerned with recognising changes of any kind but the system could be modified to deal with specific types of change of interest to the users.

The region of interest could be a county, a mountain range, an agricultural area or indeed any other area contained within reasonably smooth boundaries. For simplicity the Unit Test Areas (UTA) used for developing the programme are 50 km x 50 km squares. The boundary of the UTA is used to generate a mask for the image which classifies each picture element as either in or out of the UTA. This step requires registration of the image with geographic co-ordinates but the accuracy required for sufficiently large UTA's can be obtained, for ERTS at least, if the geographical position of the image is derived from the estimated position and attitude of the satellite. In practice with ERTS-1 the UTA is located relative to the latitude and longitude grid provided with the image, but no ground truth is used by NASA in establishing this grid so the process is fully automatic.

The picture elements in the UTA are used to derive a grey scale frequency distribution for each spectral band sensor on the platform. The histograms of these frequency distributions can be regarded as brightness 'signatures' associated with the UTA in the same way as spectral signatures are commonly used in identifying crops. Changes in the surface cover alter the shape of the histograms and thus can be monitored by comparison of each histogram with those of earlier passes over the UTA or even with a generalized signature for the UTA obtained from several earlier passes. It is expected that with experience specific types of change will be recognisable by considering the associated changes in several spectral bands, while multi-dimensional grey scale distributions for combinations of spectral bands, will provide even greater sensitivity although at greater expense.

The advantages of this method of comparison over the more obvious point for

58-12

point comparison lie in the huge reduction in the number of decisions to be made and in the generalization of the data over a large area. The major disadvantage is the possibility of significant changes being obscured by larger but less important changes. For example, cloud cover over part of the UTA could well make changes in the unobscured part undetectable by this process although still readily detectable by other means.

Comparison of the histograms can be done by eye or automatically but in either case the process is one of looking for changes in shape or position of the histogram greater than the limits imposed by the system. These are:

(i) the accuracy with which the UTA is positioned relative to the imagery.
(ii) the radiometric accuracy of the imagery.

The first of these depends on how well the position and attitude of the satellite is known, the geometric accuracy of the sensors and projection effects and for the ERTS-1 multi-spectral scanner these amount to an RMS error of approximately 4% of the area of a 50 km square UTA. This is probably an over-estimate as it assumes none of the error is systematically repeated with successive passes over the UTA and that all elements in the UTA have the same position error. The radiometric accuracy of the imagery depends on the accuracy of sensor sampling and calibration but also on atmospheric effects. For ERTS-1 digital data the radiometric accuracy of the imagery is 2% of full scale brightness but this does not include atmospheric effects. It is not yet known to what extent atmospheric effects will affect the comparison of the grey scale histograms but it will probably be necessary to include a correction for the atmosphere in an operational system. These estimates of the limitations on the system suggest that for a single band it should be possible to detect either a large change in brightness of at least 5% of the area (e.g. by cloud or inundation) or a change in brightness of a larger fraction of the UTA by 4% of full scale brightness.

4. PROPOSED TESTS OF THE SYSTEM

The computing system is still in the early stages of development and will require extensive testing before any conclusions can be reached about the usefulness of this technique. A test area in central California has been chosen which includes terrain types ranging from agricultural land in the San Joaquin Valley to the peaks of the Sierra Nevada range. Digital data for pairs of successive passes over the area in each season will be used and should provide both short term and seasonal changes in agricultural and natural vegetation for testing purposes. In addition it is hoped to use imagery taken over East African savanna regions where grass burns will be used as a source of change in the surface cover.

(Presented at the Symposium of the British Interplanetary Society on 'Earth Observation Satellites' held at University College London, 10-12 April 1973)

58-13

GENERALISATION OF ERTS DATA FOR GLOBAL-SCALE INVESTIGATIONS

I.E. Hill, A.C. Armstrong and K.M. Clayton
University of East Anglia, Norwich, UK

ABSTRACT

Information contained in ERTS imagery is of use on many spatial scales. To achieve a global scale, generalisation to units of the order of 50 km x 50 km is necessary. This involves problems of both data retrieval and efficient and meaningful generalisation. A system for the computer-based processing of ERTS digital tapes at this scale is discussed. Some results based on analysis of the grey-scale histograms are presented and possibilities for future developments are discussed.

1. INTRODUCTION

This paper describes a computer-based system that we, at the University of East Anglia, are developing for the analysis of imagery from the ERTS project. The starting point for our work is a belief that, although conventional air-photo interpretation techniques can be, and have been, used to interpret remotely sensed data from orbital platforms, the differences between the data obtained from aircraft and from satellites warrant the development, or at least the testing, of new methods of analysis. These new methods should be designed to work in parallel with, rather than in place of, conventional analyses, with the aim of exploiting the additional information contained in the satellite imagery.

The advantages of satellite platforms over aircraft platforms are their ability to provide (i) repetitive coverage at regularly spaced points in time and (ii) global or regional coverage with a uniform data base, whereas a major disadvantage of satellite platforms is the limited spatial resolution obtainable from such altitudes. Repetitive coverage provides the opportunity for monitoring changes in surface reflectance, due to either transient phenomena such as flooding and burns, periodic effects due to seasonal changes in, for example, vegetation, or long-term changes, such as urban development, forest clearance, etc. The poor resolution and the global coverage together imply that satellite imagery would be more usefully used looking for changes on a large spatial scale than changes involving areas near the limits of resolution of the sensor.

The work described here has accordingly been aimed at developing a system for monitoring satellite imagery for changes that are significant on a regional scale. Probably one of the most important uses of such a system would be a first-stage data filter for an operational Earth-observation satellite. There it would be used to select, for further processing, only those images that differ significantly from earlier imagery of that area.

2. TECHNIQUES

Because it is proposed that the system be used as a first-stage data filter, it is important that it can make use of data that has undergone the minimum amount of pre-processing. It is probable that any satellite platform for monitoring the Earth's surface on an operational and repetitive basis will return the data to Earth as a digitised picture and, for this reason, a computer-based system has been chosen. Other reasons for the use of digitised data are the ability to make use both of the spectral information provided by multispectral scanners without producing colour-composite photographs, a slow process when compared with the data-acquisition rate, and of the full dynamic range of the sensors, without the degradation in radiometric fidelity introduced by photographic processing.

For this project large general-purpose computers (an IBM 370/165 and an ICL 1903E) have been used and the programs have been written in Fortran. While a large computer and a high-level programming language are convenient in the experimental stages of a project such as this, it seems likely that a smaller special-purpose machine, perhaps with some special hardware functions, or with some analogue processing, could be more efficient for an operational system.

The first stage in the analysis of the imagery is generalisation of the data to a scale appropriate for a regional survey. Each ERTS frame contains 3×10^7 picture points for the four MSS bands, whereas a regional map at a scale of, say, $1:10^6$ could show $10^2 - 10^3$ independent data points in a equivalent area. The second stage is the recognition step where the data is compared with that for previous passes over the area, and significant changes are identified. Note that in most other computer-based systems for analysing remotely sensed data, these two steps are performed in the reverse order. For example, classification into surface types is usually done using the spectral signature for each picture point, and generalisation to large areas (individual fields) is only made when the classification results are displayed. It is hoped that by suitable choice of the area over which this initial generalisation is made, it will be possible to mask the effects of large changes taking place in a small area, and so obtain a system that is sensitive only to changes significant on a larger scale. For experimental purposes, Unit Test Areas (UTAs) 50 km square have been chosen.

The simplest possible form of generalisation is to reduce the data for each UTA to a single number for each spectral band; this number could be the mean radiance, in which case the generalisation is equivalent to reducing the resolution of the sensor to match the UTA, or it could be some more complex parameter. For most purposes, this would be too drastic a reduction in the information content of the data, although this level of generalisation could possibly be used to recognise areas with a high degree of cloud cover.

The form of generalisation used in this experiment is to characterise each UTA by the grey-scale frequency distribution for all the picture elements within the UTA. The parameters obtained from this generalisation could be either the grey-scale histogram itself or quantities that can be derived from it such as the mean, halfwidth, skewness or number of modes. Initially the histograms have been used, but one of the aims of the experiment is to determine which parameters can most usefully be used to monitor changes in surface cover.

The histograms can be compared visually, or by means of statistical tests. In the latter case the Chi-squared and Kolmogarov-Smirnov tests have been used.

3. PRELIMINARY RESULTS

In the previous sections, the aims and lines of investigation of the project have been outlined. Regrettably, it is still too early in the experiment to say how effective this technique will be, but in the remainder of this paper some early results and problems will be discussed.

The area that has been chosen as a test site is covered by one ERTS frame, and is shown in Figure 1. The site covers part of the Central Valley in California and extends as far as the coastal ranges in the southwest. A large area of the Sierra Nevada range is included on the eastern side of the frame. Digital data on computer-compatible tapes have been obtained for two successive passes over the area in later summer 1972, and a third set of data has been obtained for spring 1973. In addition, a second test area in central England was selected, but repetitive coverage is not yet available for this area.

One essential requirement of the technique is that the grey-scale distribution in unchanged areas is sufficiently repeatable for changes in other areas to be recognised. Figure 2 shows the histograms for all four bands for a UTA containing predominantly agricultural land in Central Valley. The only changes visible on photographic prints of the imagery relate to a few fields which together make up 1% or less of the UTA. The histograms have a similar shape for both scenes 1038-18114 and 1056-18114, but there is a small

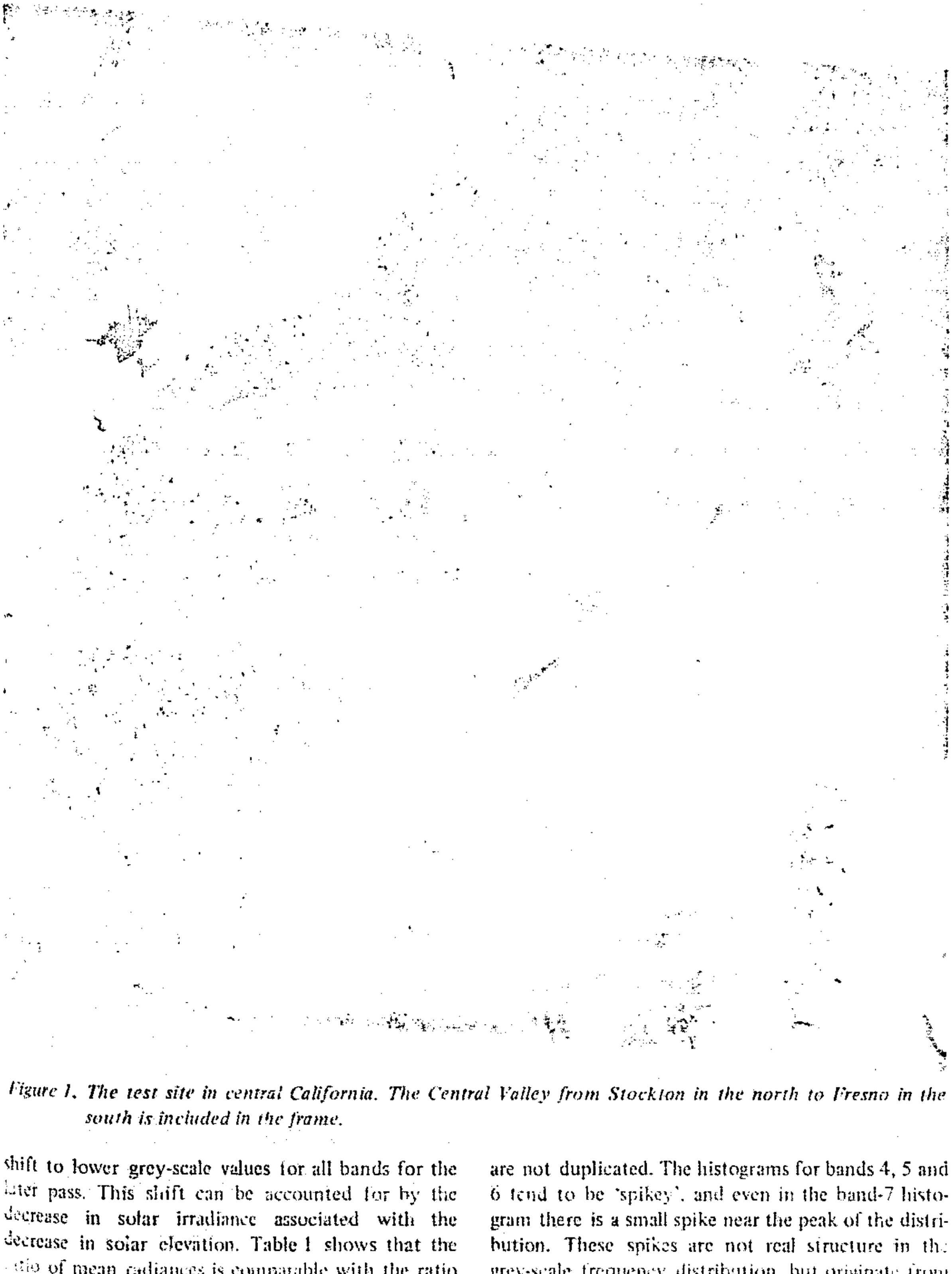


Figure 1. The test site in central California. The Central Valley from Stockton in the north to Fresno in the south is included in the frame.

shift to lower grey-scale values for all bands for the later pass. This shift can be accounted for by the decrease in solar irradiance associated with the decrease in solar elevation. Table 1 shows that the ratio of mean radiances is comparable with the ratio of the sine of the elevation angle.

Although there is very good agreement between the large-scale structure of the histograms, particularly for band 7, there are many smaller features which are not duplicated. The histograms for bands 4, 5 and 6 tend to be 'spikey', and even in the band-7 histogram there is a small spike near the peak of the distribution. These spikes are not real structure in the grey-scale frequency distribution, but originate from the mapping of the sensor grey scale with its 64 values to the CCT grey scale which has 128 levels for bands 4, 5 and 6, and 64 levels for band 7. Because of the nonlinear transfer function and expanded CCT

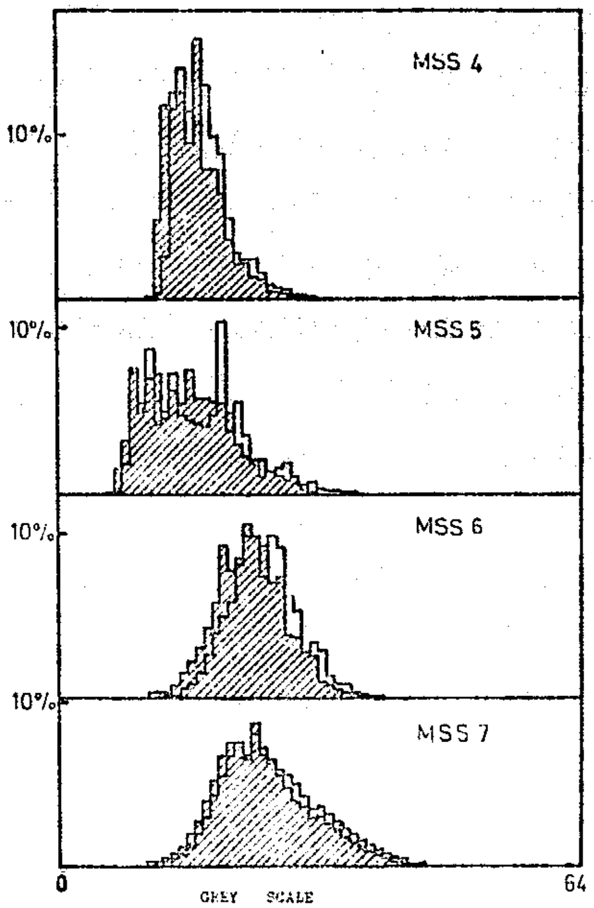


Table 1.

Band	Mean Radiance for UTA from Scene 1038-18114	1056-18114	Ratio
4	18.7	17.8	1.06
5	18.0	17.0	1.06
6	24.8	23.3	1.06
7	26.0	24.8	1.05
Sun Elevation	52°.1	47°.4	
sine (El)	0.792	0.736	1.07

Figure 2. Grey-scale histograms for a UTA in the Central Valley. The open histograms are derived from ERTS scene 1038-18114, and the shaded histograms are from 1056-18114. The horizontal scale for all bands except 7 is divided into 64 grey levels, each of which corresponds to two CCT grey levels.

grey scale for bands 4, 5 and 6, and because of the radiometric corrections applied to all bands during the transfer, there is not a one-to-one mapping of sensor grey levels to CCT grey levels. This is illustrated by Figure 3, which shows the grey-scale frequency distributions for successive groups of six scan lines down a strip one quarter of the image in width. The features to be noted are the blank columns corresponding to two grey-scale levels for this band-5 data which continue through the distribution for several successive lines. The calibration process changes the mapping from sensor levels to tape levels as the sensitivity of the sensor changes, and over a sufficiently large number of lines this tends to smooth the frequency distribution. Alternatively, the histograms may be smoothed, by use of a suitable weighting function, to give about 32 independent grey levels. However, the smoothing function for bands 4, 5 and 6 is dependent on the actual grey level and results in a nonlinear relationship between tape grey scale and radiance. The effect of the nonlinearity is to weight changes in low-reflectance surface cover more heavily than changes in the upper half of the radiance scale.

One part of the computer system which has been extensively tested is the segment that aligns the digital data relative to the UTA. The confidence that can

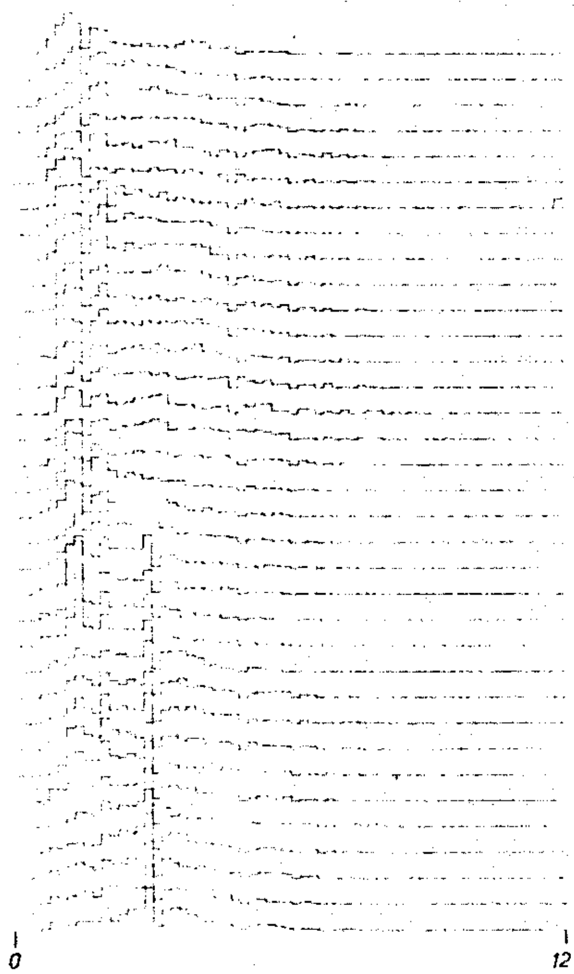


Figure 3. Grey-scale histograms for successive groups of six scan lines.

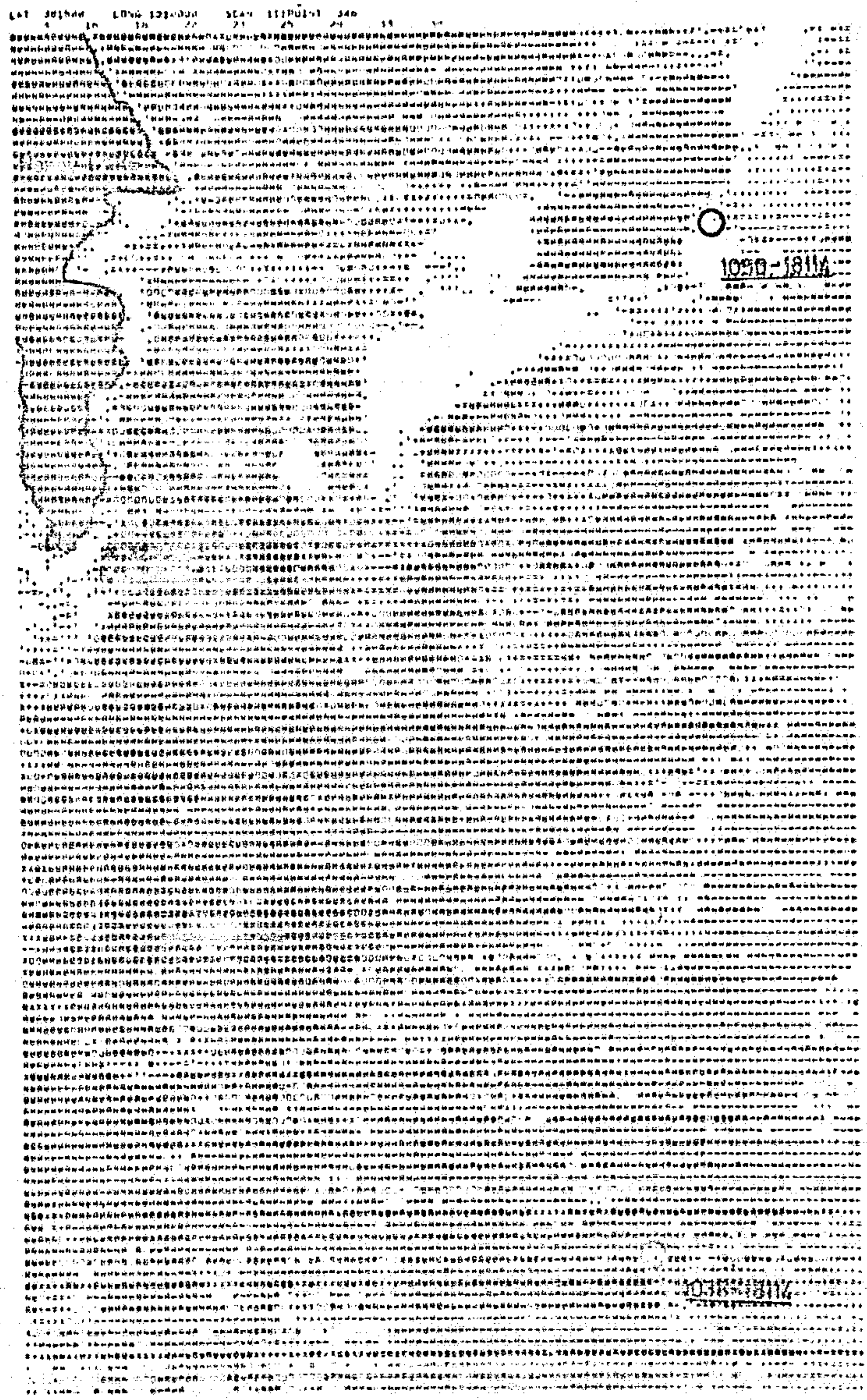


Figure 4. Position errors for the dam wall at Honey Lake. The open circles show the position of the dam (marked by the hexagon) as found from the latitude-longitude grid included with the CCT annotation data.

58-18

be placed in changes in surface reflectance detected by this system depends on the accuracy of this alignment. For example, with 50 km square UTAs an r.m.s positional error of 1 km places a lower limit on the fraction of the area for which changes can be reliably detected, of above 5% of the total area of the UTA. The proposed computer system uses the latitude and longitude grid information in the annotation record of the CCT to convert the co-ordinates of the corners of the UTA to tape co-ordinates (scan line, picture element). The pre-launch estimate of the accuracy of this conversion was about 500 m r.m.s.; for the three scenes for which tests have been made, the errors are in the range 5–10 km. Figure 4 shows the positions found for the dam of Honey Lake, California, from the annotation data for scenes 1038-18114 and 1056-18114. This error is sufficiently large to obscure any changes involving less than about one quarter of the area of the UTA, and so, for most purposes, is large enough to prevent this monitoring technique from working.

The UTA could of course be positioned more accurately relative to the digital data by use of several control points on the ground, but this has the disadvantage that, for an operational system working on a world-wide scale, the number of control points would soon become unmanagable. The ideal solution for an operational satellite platform would be improved attitude and position information and control, but this might not be technically feasible. One alternative method, which is being tested, is to align one image with the ground by manual methodsand then to use this primary image as a correlation mask for successive passes.

As a final example of the sort of difficulties that need to be overcome before the computer system can be used operationally, Figure 5 shows the effects of clouds and cloud shadow on the grey-scale distribution. The histograms are for the same part of the Sierra Nevada range, but in one case the area is partially cloud covered. This produces a sharp peak at high radiance, as expected, but in addition the cloud shadow, and structure within the cloud, are sufficient to make the remainder of the histograms totally different. The implication of this is that changes in the cloud-free area can only be detected if some spatial information is used to remove the whole cloud-affected area from both images before preparing the histograms.

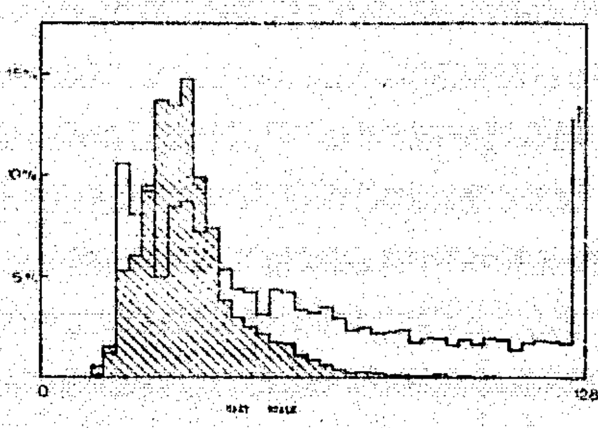


Figure 5. Grey-scale histograms for an area in the Sierra Nevada range. The histogram for scene 1038-18114 (unshaded) is plotted with only the unsaturated part of the image used in calculating the frequencies for each grey-scale class. The grey-scale classes correspond to three CCT grey levels.

4. CONCLUSIONS

The results presented above tend to emphasise the difficulties that have been met in developing the computerised system, and give a rather negative outlook on the usefulness of the technique. At this stage, however, much development is still required, and no tests on the effectiveness of the system will be possible before difficulties of the type described have been solved. Other work proceeding in parallel with that described here does, however, give some indication that use of grey-scale histograms as a method of generalising the data is a satisfactory solution. In particular, it seems possible to use the histograms alone as a means of dividing the image into surface types by using standard procedures.

58-19

OBJECTIVE GENERALISATION USING ERTS IMAGES.

by A. C. Armstrong & K. M. Clayton. (forthcoming in: Environmental remote sensing practices and problems, 1975).

1) OBJECTIVES

An orbiting satellite observatory, such as ERTS, produces data whose qualities differ from any previously available. The resolution of the scanners is lower than that of conventional air-borne systems, and thus it is difficult to apply conventional techniques in their interpretation. Compensating for this lower resolution is the repetitive global coverage that the satellite potentially offers. Working normally, such a system produces an enormous quantity of data; a quick calculation shows that ERTS-1 has the capability of filling 10^5 reels of magnetic tape a year. This figure suggests the value of any approach which allows a reduction in the volume of the data. Once a reduction in the volume of data is achieved it becomes possible to take advantage of the consistent data base offered for global investigations.

The approach we discuss makes use of computer classification techniques to achieve this data reduction. Using the ERTS Multispectral Scanner (MSS) as its data source, it attempts a mapping of the earth surface on an objective basis. Such a classification would be distinct, both in objectivity and in scale from anything possible before the advent of remote sensing satellites.

2. Characteristics of ERTS data

In a sun-synchronous orbit at a height of 910 Km ERTS performs 14 orbits a day, retracing its path every 18 days to provide precisely-controlled repetitive imagery. The sensing systems on board are a Return Beam Vidicon Camera (RBV) and a Multispectral Scanner (MSS). We are concerned with the MSS data although the techniques developed could be applied equally to the RBV or any other multispectral digital data.

The MSS scans an angle of 2.89° in a direction perpendicular to the satellite path, the forward motion of the satellite ensuring a North to South scan. The system produces no images during the South to North half of each orbit (which is in darkness). Each scan line is 185 Km in length and the continuous stream of data is split into sections 185 Km long, called 'Scenes'.

The rotation of the Earth beneath the satellite imparts a skew on the image which varies inversely with latitude. For Californian imagery we use the skew is 5° so that successive scan lines are displaced to the left by about 20m. For most practical applications it is convenient to neglect this skew, assuming a rectangular array of points for analysis, and correcting for the skew in the final display stages.

The MSS senses the reflected radiation infour spectral regions, conventionally called bands 4 to 7 (table.1). The nominal instantaneous field of view is 79m, although the system is capable of detecting smaller objects. Each scan line is moved on 79 m. The resolution of the system is thus much coarser than that of conventional airborne sensing systems.

The data are made available as photographic images or digitised as Computer Compatible Tapes (CCTs). Because of the capability of numerical processing to provide objective generalisation, the CCTs were used as the primary data source.

Each scene in CCT form contains 2340 scan lines each with 3240 picture elements (pixels). The four spectral bands are interleaved, so that the multispectral information for any one point is available at one point in the tape. Scan lines are split into four equal portions, and placed on four magnetic tapes, thus dividing each scene into four North-South strips, 185 x 48 Km. (Fig. 1).

A fuller description of the ERTS system can be found in NASA (1971) and of the CCT format in Thomas (1973).

2) CLASSIFICATION PROCEDURES IN REMOTE SENSING

Because of the size of the data sets involved, remote sensing applications have used only a limited number of the many classification strategies available. Those procedures that have been used can be considered initially in terms of two fundamental dichotomies, between the supervised and unsupervised, and between per-point and per-field approaches. (Fig.2)

The greatest effort so far has gone into per-point classification procedures, which take each data point in turn, and either assign it to the nearest cluster, or classify the point by application of an *a priori* decision function, thus ignoring the spatial arrangement implicit in such remotely sensed data. Per-field classification procedures, on the other hand, take note of the spatial characteristics of the data, and attempt to classify not individual points but groups of spatially contiguous points, or 'fields'. These fields can be assembled either by examination of the data, or by some arbitrary system.

The use of supervised procedures stems from earlier work which sought to establish *empirical* catalogues of 'spectral signatures'. However, such work has shown spectral signatures to vary considerably with scene, and to be insufficiently distinct for the establishment of such catalogues. [illegible] inclusion still be found in the supervised technique, which [illegible] of the data for which 'ground truth' is available, and then this can be used [illegible] the statistics describing the spectral responses of the classes required. This procedure [illegible] [illegible] of [illegible] reduced [illegible]

Typical among these schemes is the Bayesian classifier adopted by LARSYS (Lindenlaub, 1973), which requires the estimation of the mean vector and variance-covariance matrix for each land use type. These statistics are calculated from a training set. A second sample for which ground truth is also available is designated the 'testing set' and the performance of the classifier on this test data is then assessed, and this is taken as an estimate of the performance of the classifier over the whole data set.

Statistically, this procedure is fraught with difficulties. Training sets are often compiled arbitrarily, in violation of normal sampling assumptions, thereby yielding incorrect estimates of the population parameters, and subsequently poor performance of the classifier (Basu and Odell 1974). Further, both training and testing sets are usually compiled from areas where land use types are clearly defined, whereas large sections of the image may well be considerably less clear.

Because of these statistical problems, Nagy et al (1971) argued it is more efficient to classify the data initially, and then to collect ground truth for each of the classes produced. This alternative requires fewer *a priori* decisions and less additional ground-based information, thereby being far more efficient in exploratory investigations.

This investigation uses a per-field, unsupervised procedure. The use of a per-field approach seemed to offer the greatest possibility of generalisation, and permits a varying level of generalisation by altering the size of the fields, and the unsupervised analysis is preferable in this exploratory situation. In particular, since it was one of the aims of the investigation to determine the kind of feature that can be extracted at this scale, the use of a training set was not appropriate.

The classification procedure used in this investigation involves three data files, and three decision blocks (Fig 3). The original data on the CCT supplied by NASA is recompiled into a series of subsets, the fields. From these fields, summary measures are computed, producing a feature file for each field which is input into the classification algorithm, and the end result is a map of the area, with each field assigned to a category.

3) FIELD SELECTION

It is obviously unrealistic, and in terms of this investigation unnecessary, to classify individual pixels. Since our aim is generalisation, it is convenient to reduce the data set at this stage, by considering groups of pixels, the fields which can then be classified. Fields were arbitrarily defined as a square block of pixels using tape coordinates, referring to parallelogram-shaped area on the ground. This arrangement is computationally convenient, and also in line with schemes used by other workers for per-field studies. It is further easy to increase the level of generalisation by amalgamating these fields, without returning to the original data source.

Several factors influence the size of field that is appropriate. The resolution of the required classification dictates a maximum size of field, and the resolution of the sensing system imposes a lower limit. The scale of the analysis will affect the size that is appropriate, for in general the larger the area to be analysed, the larger the fields.

ORIGINAL PAGE IS
OF POOR QUALITY

This leads to the practical consideration of computer storage space. While this factor is not independent of the classification algorithm used, there is in nearly every case an upper limit on the number of data points that can be handled by any one program, and even if this space is not a limiting factor it is likely that machine time will become a limiting factor instead.

With these considerations in mind, and noting the original purpose of the investigation, three sizes of field were used (table 2). The basic field was taken as a size of 100 x 100 pixels, yielding 736 fields to one ERTS scene, in an array 23 down and 32 across. These basic fields were then amalgamated to produce larger fields containing 200 x 200 pixels (16 across by 11 down). For pragmatic reasons these fields were not exhaustive of the original image. Thus the 8 fields across each CCT (1/4 image) left a narrow strip 10 pixels wide that was not considered. The increased speed of processing gained by restricting the field compilation program to a single source magnetic tape was the important factor. Similarly only the first 2300 of the 2340 scan lines were used. The clustering algorithm thus dealt with complete fields only, so that scale problems that might arise with incomplete fields were avoided. In future, for operational implementation, it would probably be preferable to adopt an arrangement that more completely fitted the tapes, using, for example, fields 101 x 117 pixels in size.

The set of 176 fields could easily be handled by a hierarchical clustering program, but such a algorithm could not easily handle the larger set of 736 fields, for which a different, (iterative) clustering technique was implemented.

Some results were also obtained using a smaller field, 50 x 50 pixels in size, producing a total of 2880 fields to an ERTS scene in an array 64 across and 45 down. The size of this data set meant that far fewer analyses could be performed using this set. However, using all three sets and a common algorithm it was possible to investigate the effects of different field size on the achieved generalisation.

4) FEATURE SELECTION

Any single field contains a vast amount of information, and it is both impossible, and unnecessary to classify fields using all this information. Feature selection is the process whereby we summarise this vast amount of information and reduce it to a smaller more manageable set, thus eliminating as much of the redundant information as possible, while retaining as much of the essential as is required.

Obviously from any large field any number of features can be computed, and the choice must be to some extent arbitrary. Several techniques exist for dealing with the selection of an optimum sub-set of features, developed largely for selecting the best combination of spectral bands for a per-point classification of multispectral data (eg. Swain & King 1973). In this study the use of such techniques is not directly applicable, since there is no clearly defined set from which to form an optimum sub-set, but rather an infinity of possible features from which we must arbitrarily choose. This choice is made by considering the kind of features in the image we wish to extract and on our a priori notions of how these might be detected.

Several possibilities suggest themselves as suitable variables. It is simple to compute the mean reflectance in each spectral band, and then to use a 'mean spectral signature' of each field, thus reducing the data, to a total of four variables per field. Although this set is potentially very efficient, there is no _a priori_ reason for thinking that the mean spectral signature for a field should be any more successful than the point spectral signature, particularly over the large spatial scales involved. Surprisingly this feature set has been used relatively successfully, but the generalisation is unnecessarily extreme, and it is unlikely that this would provide an acceptable basis for routine classification.

A more complete feature set includes not only the mean reflectances but also the frequency distribution of reflectance values. Rather than attempt to model these frequency distributions, with the attendant problems of finding a model of sufficient generality and of estimating its parameters, it was decided to examine the distributions directly. For each field the histogram of reflectance values in each of the spectral bands was compiled, and the data contained in these histograms used as the feature set. This was achieved by splitting the range of reflectance values into 32 equal classes, and recording the number of pixels with a value in each class. The values for all four bands are put into one vector, 128 numbers long, which became the feature vector. This procedure has certain advantages. Firstly it deals very simply with polymodal distributions. Secondly, since it deals with uniform fields, the values can be the total frequency of occurrence and require no further normalisation. This has the result that the length of each axis in the feature space is directly related to the areal coverage on the ground within that spectral range, and thus the variables are automatically scaled in their order of importance in terms of areal coverage. Thirdly this set still retains a large amount of information, such that it is easy to submit it to further compression if required. For example, it would be easy to estimate the mean and variance of each spectral band from this feature set. Thus the feature file derived at this stage acts as a secondary data source, so that further features can be extracted without reference to the original CCT data.

This feature set results in a great compression of the data. For a 100 x 100 field some 40,000 numbers are summarised by 128. This considerable saving can be realised with only a modest expenditure of computer time. Further, this set is small enough to be input into most clustering programs without creating excessive demands on computer time or space.

Several other feature sets have been used by other investigators. Various transforms have been used as measures of texture, for example by Kirvida (1973) who compared the use of the Fourier, Karhunen-Loève, Walsh and Slant transforms for this purpose. However, Kirvida calculated these transforms over small fields, 8 x 8 pixels in size, and their extension to larger fields would require greatly increased computation times, and the storage of a very large number of transform coefficients. In a similar way Haralick and his co-workers (Haralick and Shanmugam 1974, Haralick, Shanmugam and Dinstein 1973) have measured texture using the spatial grey-tone dependence matrix, defined over fields 64 x 64 pixels in size.

There is thus no single satisfactory set of features available for describing large fields that can be implemented in reasonable computing times. The hybrid approach discussed by Gramenopoulos (1974) presents an interesting alternative, using an optical system to produce a Fourier transform of fields, which is then converted electro-optically into a digital signal suitable for input into a classifier.

5) CHOICE OF CLUSTERING ALGORITHM

A great variety of clustering strategies could be implemented for this study, and some choice must be made between them to decide which are best for the purposes in hand. Published techniques were used and modified only where necessary, so that the performance of the classifiers could be related to previous studies, and no attempt was made to add yet another strategy to the already long list of strategies available.

Two main groups of classification procedures can be identified. Firstly, hierarchical cluster analyses, such as have frequently been implemented in the social and biological sciences, constitute a body of standard techniques. To these certain modifications have been made to cater for the special class of spatially organised data. Second, there is a diffuse body of other strategies, most of which involve rather more *a priori* assumptions about the data, but can also handle larger data sets.

Common to all clustering techniques is that they require some measure of the distance between points, and between clusters, and several distance measures are available. For distances between points, and between points and cluster centres the traditional Euclidean distance is most frequently used, although the alternative 'city-block' distance has a certain popularity because of greater ease of computation. The distances between clusters of points is a considerably more complex problem, since it involves not only the absolute distance between cluster centres, but also the dispersion of the observations around those centres. A review of these measures is given by Wacker & Landgrebe (1972), who conclude that in general there is no strong reason for choosing any one in preference to any other.

5a) CLASSICAL CLUSTER ANALYSIS

This body of techniques uses variations of the algorithm described by Ward (1963), for producing hierarchical classifications. This method examines the observations in the variable space, and finds the two observations closest together. These two observations are replaced by their mean position, and the whole set of points, reduced by one in number, is examined again to find the next smallest distance. Thus at each iteration the number of points is diminished by one, and a hierarchical classification results. (Fig.4)

This group of techniques is available in many versions, and can be considered as the standard classification method. However, they require the calculation and continued re-accessing of a matrix of distances between all points in the variable space. Thus to classify n points, they typically require the use of an n by n distance matrix which is updated and scanned (n-1) times. This imposes an upper limit on the number of observations that can be handled. Most computer systems can cope with only of the order of 200 points, and even the use of back-up storages does not permit the extension of the algorithm far beyond this limit (Openshaw, 1974). This technique was implemented for the classification of the 176 fields 200 by 200 pixels in size, using the program published by Veldman (1967). This program has been used by Owen Jones (pers.comm.) in the remote sensing field, and yields what can be considered as standard results. It could not, however, be extended to cope with the larger number of fields 100 by 100 pixels in size.

It is possible to use the spatial characteristics of the data to achieve a reduction in storage requirements, and thus handle a larger data set. Thus by joining only spatially contiguous points, it is possible to produce areally tight clusters. This permits the storing of only the distances between contiguous points. Openshaw (1974) has used this technique to produce a program that can cluster 1200 points in the same space the Veldman program clusters 200. This saving is considerable, but the approach presents conceptual difficulties. It is by no means obvious that we should expect adjacent fields necessarily to belong to the same class. Further there are grounds for wishing to place spatially non-contiguous areas in the same class. Thus for example if the image consisted of a large agricultural area crossed by one distinctive flood plain, we would want to produce only two classes; flood plains and agricultural areas, and not divide the agricultural area into two classes separated by the flood plain. Such a classification could perhaps be achieved by relaxing the contiguity constraint at some (subjectively - determined) late stage in the analysis, such as was done by Owen-Jones & Cust-nce (1974). However, in view of this difficulty, this scheme was not pursued.

5b) ALTERNATIVE CLASSIFICATION STRATEGIES

Of the several alternatives to the hierarchical technique, two procedures developed for remote sensing purposes were investigated. These were the iterative clustering method described by Swain (1972), and the chain method described by Nagy & Tolaba (1972).

<u>The Iterative Clustering Method</u>

This technique requires the setting up of an initial set of cluster centres, which may be determined *a priori* or derived by taking a sample of the data. The observations are then assigned to the nearest cluster centre, and the procedure repeated, again scanning through the data and assigning each point to the nearest cluster. (Fig 5). It can be shown that this procedure converges to a stable solution which is influenced only slightly by the initial estimates of the cluster centres. This technique has been implemented in the LARSYS package of remote sensing programs (Swain, 1972).

To ensure that all the initial cluster centres were within the 'cloud' of points, every n/m^{th} data point (where n is the number of observations and m the number of clusters) was arbitrarily assigned to a cluster centre. Because the program needs store only the cluster centres, and can place the observation points on a rewindable back up store (magnetic tape or disc unit), this procedure was capable of handling large numbers of points with ease.

Although the algorithm produces as many clusters as there are initial centres, these clusters may overlap, and it is thus necessary to investigate the distance between pairs of clusters to determine how many 'separable' groups are available. The problems of defining the inter-cluster distance has already been mentioned. Most of the traditional measures of the distance between two multivariate groups require the inversion of either one common or two different variance-covariance matrices. The calculation time required to invert a least n x (n-1)/2 matrices with dimensions 128 x 128 becomes prohibitive and also increases the space requirements considerably. Thus an alternative measure was devised, which was the city block distance between the two centres, weighted by the standard deviations along each variable axis. It is given by

$$D_{ij} = \frac{1}{N} \sum_{k=1}^{N} \frac{|\bar{X}_{ik} - \bar{X}_{jk}|}{(\sigma_{ik} + \sigma_{jk})}$$

where $\bar{X}_{ik}$ is the mean of the i^{th} group on variable k

σ_{ik} is the standard deviation of the i^{th} group on variable k

N is the number of variables

This measure takes on a value of less than 1.0 if, on the average, the clusters overlap. It has a lower bound of zero, but no upper limit. In practice, values lower than 0.5 seem to represent clusters that overlap considerably. However, no inferential statements can be made about this measure, and it remains essentially a descriptive statistic.

Using this measure it is possible to construct a 'separability matrix' showing the distances between all pairs of cluster centres. This matrix is examined manually, and the subjective interpretation involved was aided by taking into account the spatial distribution of the classified points. There was thus a considerable degree of freedom left in the preparation of the final classified map.

The Chain Clustering Method

This method was used by Nagy & Tolaba (1972), based on an algorithm described by Bonner (1964). This method examines each point in turn, and calculates the distance between it and all clusters. If the point is further away from all clusters than a given threshold value, then it forms a new cluster. If it is nearer than this threshold, it is added to the cluster, and the cluster centre recalculated. Nagy and Tolaba discuss the use of a 'priority list' which permits the searching through the list of clusters in an efficient manner, and also the deletion of clusters which do not grow and thus represent anomalous points. They further discuss the technique of amalgamating similar points into 'fields' before they are compared to clusters. This technique is not applicable in this case, since the data is already in an amalgamated form, and this modification was not included. A further modification that was implemented for this analysis was the recording of the deleted clusters, and the subsequent allocation of the deleted clusters to the remaining clusters at the end of the analysis. (Fig 6).

This technique thus requires only the same order of storage space as the iterative algorithm, but because it examines each data point only once, it is several times faster in execution.

5c) VISUAL ANALYSIS

One strategy that was adopted for this investigation was a technique that simply involved the visual inspection of scatter diagrams. For each field the mean value for each band was calculated and these values for each pair of bands plotted as a two-dimensional scatter diagram. On these diagrams, clusters of points can be visually determined, and then mapped. This technique allows the rapid production of a classified map based on a minimum amount of information, which is potentially useful.

5d) SUMMARY

To summarise this investigation has available four clustering techniques, the hierarchical, classical method, the iterative and chain methods, and the visual inspection of scatter-plots. With this body of techniques, we are in a position to examine some results. First, however, it is necessary to discuss the data available.

6) THE DATA

Three sets of data were available for the analysis, two of central California and one of eastern England. (Table 3. figs 7,8,9)

The two of central California include a complete transect from the High Sierra Nevada almost to the coast, thus including a large section of the central valley, characterized by irrigated agriculture. The two images are of essentially the same area, one taken in September '72, and the other in the following May, and this pair provides a test of the reproducibility of the results, and be used to identify the kind of effect [illegible] seasonality introduces into the data.

The image of eastern England provides a contrasting scene, with much lower quality. The low resolution of the sensors and the hazy atmosphere contribute to an image which is visually very poor. However, the ability to use all four spectral bands simultaneously permits the analysis to differentiate features which are not obvious on any of the single band images.

7. Results.

Some results of applying these techniques to the images are shown as figs 10-14, and table 4 acts as a guide to these analyses. They should be compared to the images from which they were derived, figs 7-9. Although the plotted classifications have been geometrically corrected they do not overlay the images directly due to a lack of correspondence between the CCTs and the imagery. The imagery has an along-track displacement with the respect to the CCTs of an average of 120 scan lines at the north and about 50 scan lines on the south. There is however no cross-scan displacement.

The first major conclusion that can be drawn from all the analyses is that in general between 4 and 6 groups could be produced by these techniques. Wherever possible the analyses have been presented with a standard number of 5 groups. This standardisation was less easy to achieve with the iterative technique which if given 5 initial cluster cents would produce less than 5 distinctive groups. For this technique it was considered best to produce 10 groups and then to reduce the number of clusters by examination of the separability matrix.

7a) Image D

The 200 x 200 fields were classified using the hierarchical technique, using the hierarchical technique, taking the data for all spectral bands, for each band separately, and each combination of bands. The results using all the data, band 4 only, and bands 5 and 7 together are shown in fig. 10. The classification using all the data (i.e. bands 4,5, 6 & 7) can be considered the standard, against which the performance of the reduced data sets can be compared.

The classification using all the data shows a remarkable spatial cohesion to the results, and all the features observed can be interpreted satisfactorily. Working from NE to SW, the groups were identified as : 1) the high Sierra Nevada, with its cover of snow, bare rock and high mountain vegetation; 2) a zone of heavily forested mountains; 3) a zone of mixed grassland and woodland (Chaparral), which also recurs in the Coast Ranges; 4) as the woody component of the previous class becomes more scarce it is replaced by a uniform Dry Grassland which blankets the landscape in a uniform layer which occurs as a strip down both sides of the Central Valley, and also in the upper Monterey valley in the extreme southwest of the image; 5) the remainder of the image is taken up by the irrigated agricultural area of the Central Valley. It is worth noting that the area of uncultivated land in the left centre of the agricultural zone is correctly classified as belonging to the Dry Grasslands.

Several features that are obvious on the imagery are not picked out by the analyses and require further comment.

The urban areas of Fresno and Modesto are fairly distinctive features, and several smaller towns can also be distinguished, which are not detected by the analyses, because although they are texturally very different from the adjacent agricultural zones, they are not spectrally distinct.

Similarly the reservoirs that are a very prominent feature of the foothills and of the coast range are not detected, for although visually prominent, these reservoirs are areally small, and cover only small proportions of fields 200 x 200 pixels in size. Further, being in general long and thin, their influence on the grey scale histograms is often spread over several fields.

Classifications using single bands (of which fig 10b is an example) were considerably less successful, for only the Chaparral, Forest, and High Sierra subdivisions were even approximately delimitated. The combinations of bands 5 & 7 (fig 10c) produces reasonable results, but are confused when compared to the classification using all the data. These analyses show that the full combination of all four spectral bands is required for a coherent classification and that attempts to reduce the size of the feature set by considering only a subset of bands are not likely to be successful.

Mean reflectance values of the fields for each pair of spectral bands were plotted on scatter diagrams (fig.11). These showed high [illegible] and the two [illegible] of detected in plots [illegible] bands 5 & 7 [illegible] (fig [illegible]). The [illegible] agricultural zone, although the Forested area of the Sierra Nevada is confused with the Grassland zone.

The results of applying the iterative technique to fields of differing sizes is shown by fig 12. Although the regionalisation using 200 x 200 fields was not identical to that produced by the hierarchical technique, the agreement was good. The delimitation of the Grassland and Chaparral zones on the north eastern side of the central valley is almost identical, although there is a confusion between the High Sierra and the Dry Grasslands. The agricultural area is divided into two zones, an inner one consisting of smaller fields (showing red on the false colour image fig 7), and the other showing the influence of irrigation water, showing blue on the image.

Using the smaller 100 x 100 fields (fig 12b) the overall pattern is the same, preserving the three-fold division into agriculture, grassland, and forest, but both the agricultural and mountain zones are subdivided. It can be seen that this use of smaller field does not produce a cluster map of the same classes with better resolution, but rather a different kind of classification. As the size of the field is diminished so the analysis begins to pick out individual components of the landscape. This effect is even more marked when considering the 50 x 50 fields (fig.12c). In this analysis the mountains are clearly defined, and the transitional nature of the boundaries between Grassland, Chaparral and Forest is shown by their interdigitation.

7b) Image E.

The results of both the hierarchical and iterative techniques applied to image E are similar to those obtained from image D. The one major exception is that the agricultural zone and the adjacent grasslands are not differentiated, reflecting the fact that in springtime the grassland is growing vigorously and is thus spectrally similar to the irrigated crops of the valley floor. This effect is particularly marked in the infrared image (fig.8B). Thus in this analysis the seasonal variation confuses the classification procedure, demonstrating the necessity of both multispectral and multitemporal data for a complete analysis. This effect is visible in both the hierarchical and iterative classifications, and this consistent behaviour in the face of a data change demonstrates their validity for the application.

The result of applying the chain algorithm to the 100 x 100 fields for image E is shown in fig 13e. The results are more confused than for either of the other analyses, although the general lines of the threefold division into agriculture, grassland and mountain is visible. This result gives the general impression of a more 'noisy' classification. Further this result was obtained only after a certain amount of 'juggling' with the threshold and cluster deletion criteria, which indicates that this technique would not be suitable for unsupervised application on a routine basis.

7c) Image C.

The classifications of image C of eastern England (fig 14) are in general less clear, and the most persistent feature of the results if the obvious division between land and sea. The simple hierarchical technique applied to the 200 x 200 fields (figure 14a) shows three distinct groups in the water; the shallow water of the Wash, the sediment plume emanating from the Humber, and the clearer water of the North Sea. Figure 14b was prepared using the smaller 100 x 100 fields. It is confused on the west by the cloud cover and the associated cloud shadows, but over the rest of the image several parts of the land area are distinguished. The sea again divides into the darker more open sea and the narrower coastal water with considerable turbidity. The fields which comprise roughly equal portions of land and sea are so distinctive that they form a separate category which is almost confined to the coast. Two other groups of areas fall into the same category; it includes a combination of clouds and cloud shadows but the other is not a statistical artifact since it is located precisely over the coniferous woodlands of Cannock Chase and Sherwood Forest. The rest of the land area divides into two unequal portions; in general the smaller of these is the most arable part of the landscape, notably West Norfolk and the fens edge south of Lincoln, but the relationship between the proportion of arable land and this division is not consistent across the image and some further unknown factor must be involved.

ORIGINAL PAGE IS
OF POOR QUALITY

8) DISCUSSION

Of the classification techniques used so far, by the far best in terms of the interpretability of the results is the classical hierarchical method. This method is, however, severely limited by its inability to handle the 100 x 100 pixel fields, which reduce its utility in dealing with small scale landscapes, such as that in the East of England. The Chain algorithm on the other hand is by far the most efficient in terms of the usage of computer time and space, but produces considerably lower quality results. The iterative technique seems a good compromise, but it is hard to see how this could be implemented on a special purpose machine for routine applications to the ERTS imagery. However, all these clustering techniques take considerably less time to execute than the field and feature compilation stages.

The inability of single bands to produce effective classifications points out the value of using the multispectral nature of the data to the full, and the seasonal effects noted demonstrate the need for multi-temporal data. However, the full 128-variable feature set has some considerable redundancy, and experiments involving its reduction to a 64 variable set produced almost identical classifications. We conclude that reductions in the size of the feature set are best made by reducing the number of variables per spectral band, and not by reducing the number of bands.

The effect of varying the size of field used is not just to vary the resolution of the achieved classification, but determines the type of class produced. The choice of an appropriate size depends on the scale of the features in the landscape, and thus the 200 x 200 pixel fields which were suitable for California yielded poor results in eastern England, where the smaller fields were more effective.

These techniques have all been developed in the context of a large general purpose computer. The average total computer time to compile the field feature files and to classify them on the IBM 370/165 at Cambridge is of the order of 20 minutes central processing time, of which a large amount is taken up with the tape handling routines. Obviously for routine applications some closer approach to real time analysis is required, and here the way forward seems to be with special-purpose small computers. In this context the most promising technique seems to be the chain algorithm, which requires little store and examines the data only once, as it is read in. We can thus conclude that, at least with present generation of computers, the volume of data to be processed will force us to use techniques that are not necessarily the best, but which offer some hope of coping with that vast quantity of data.

9) References.

Basu, J.P. & P. L. Odell 1974 Effect of intraclass correlation among training samples on the misclassification probabilities of Bay's procedure.
Pattern Recognition v.6 pp.13-16

Bonner, R.E. 1964 On some clustering techniques
IBM Journal of Research and Development v.8 pp.22-32

Gramenopoulos, N. 1974 *Automated thematic mapping and change detection of ERTS-A imagery.*
Optical Systems Division, Itek Corporation
Final Report prepared for Goddard Space Flight Center
Greenbelt, Maryland.

Haralick, R.M. & K.S. Shanmugam 1974 Combined spectral and spatial processing of ERTS imagery data
Remote Sensing of Environment v.3 pp.3-13

Haralick, R.M., K.S. Shanmugam & I.Dinstein 1973 Textural features for image classification.
IEEE Transactions on Systems, Man and Cybernetics.
v.SMC-3, pp.610-621

Kirvida, L. 1973 *Final report on automatic [illegible] for [illegible] identification ([illegible]).*
Honeywell, Inc. Systems and Research Division,
Minneapolis, Minnesota.

Lindenlaub, J.C. 1973 Guide to multispectral data analysis using LARSYS
LARS Information Note 062773
[illegible] Applications of Remote Sensing
Purdue University, Lafayette, Indiana.

Nagy, G. & J. Tolaba 1972 Nonsupervised crop classification through

airborne multispectral observations.

IBM Journal of Research and Development. v.16 pp.138-53

Nagy, G., G. Shelton & J. Tolaba 1971 Procedural questions in

signature analysis.

IBM Research RC 3517 Yorktown Heights, New York. also

Proceedings of the Seventh Symposium on Remote Sensing of the

Environment. The University of Michigan, Ann Arbor, Michigan,

1971. p.1387

NASA 1972 ERTS Data Users Handbook.

NASA Document 71 SD4249

Openshaw, S. 1974 A Regionalisation Program for large data sets.

Computer Applications vol.1 no.3/4 p.136-160

Owen Jones, E. & H. Constance. Digitised Analysis of Skylark Rocket Imagery

Journal British Interplanetary Society, v.27 pp.19-32

Swain, P.H. 1972 Pattern recognition: a basis for remote sensing data analysis

LARS Information Note 111572

Laboratory for Applications of Remote Sensing

Purdue University, Lafayette, Indiana.

Swain, P.H. & R.C.King 1973 Two effective feature selection

criteria for multispectral remote sensing

LARS Information Note 042673

Laboratory for Application of Remote Sensing

Purdue University, Lafayette, Indiana

Thomas, V.L. 1973 Generation and physical characteristics of the

ERTS MSS system corrected computer compatible tapes

Goddard Space Flight Center, Document X-563-73-206

Greenbelt. Maryland.

Wacker, A.G. & Landgrebe, D.A. 1972 Minimum distance classification in remote sensing.
Proc. First Canadian Symposium for Remote Sensing.
also *LARS Print* 030772
Laboratory for applications of remote sensing,
Purdue University, Lafayette Indiana

Ward, J.H. 1963 Hierarchical grouping to optimize an objective function. *Journal American Statistical Association*
v.58 pp.236-44

Veldman, D.J. 1967 *Fortran Programming for the Behavioural Sciences*
Holt, Rinehart and Winston, New York.

10) Acknowledgements

Much of the work involved in developing the routines for handling the data tapes was done by Dr.I.[illegible].Hill, now of the Experimental Cartography Unit, London, without whose contribution this work could not have been done.

The computing services were provided by both the University of Cambridge and the University of East Anglia, and thanks are due to the advisory services of both of these establishments for their help and advice.

Dr.Tarrant & Dr.Haddon of the University of East Anglia have provided many valuable comments and criticisms of this work as it has progressed. This research has been funded by both the Department of Industry (formerly DTI) and the Natural Environmental Research Council.

Table 1

The Spectral Bands on the ERTS MSS

Band	Wavelength Range (Micrometres)		Colour
4	.5	.6	Green
5	.6	.7	Red

Table 2.

Fields used in the study

Size of Field in pixels	Horizontal Size (km)	Vertical Size (km)	Number in ERTS scene	Number across	down
50 x 50	2.8	3.7	[illegible]	65	45
100 x 100	5.7	7.4	736	32	23
200 x 200	11.4	14.8	176	16	11

TABLE 3

The Data Available

LOCAL TAPE NUMBER	NASA FRAME NUMBER	CENTRE COORDINATES N	CENTRE COORDINATES W	DATE	AREA
C	1031 - 10334	53 - 13	00 - 00	23 Aug.72	Eastern England
D	1056 - 18114	37 - 23	120 - 22	17 Sept.72	Central Valley California
E	1308 - 18122	37 - 34	120 - 37	27 May 73	Central Valley California

Table 4 Guide to Analyses Presented

Fig.	Image	Technique	No.of Groups	Field Size	Comments
10a	D	Hierarchical	5	200 x 20	all bands
10b	D	Hierarchical	5	200 x 20	band 4 only
10c	D	Hierarchical	5	200 x 20	bands 5 & 7
10d	D	Visual Inspection	4	200 x 20	see fig 11 for scatter diagram
12a	D	Iterative	5	200 x 200	
12b	D	Iterative	7	100 x 100	
12c	D	Iterative	7	50 x 50	
13a	E	Hierarchical	5	200 x 200	
13b	E	Iterative	5	100 x 100	
13c	E	Chain	6	100 x 100	
14a	C	Hierarchical	5	200 x 200	
b	C	Iterative	6	100 x 100	

6/1

Journal of the British Interplanetary Society, Vol. 28, pp. 473-476, 1975.

A COMMON U.K. FORMAT FOR ERTS DIGITAL TAPES?

A. C. ARMSTRONG and I. E. HILL*
School of Environmental Sciences, University of East Anglia, University Plain, Norwich, England.

The need for a common U.K. reformatting scheme for ERTS digital data is presented, and the merits of various schemes discussed. Practical details of the reformatting are discussed in detail, and recommendations for a possible scheme are made.

1. INTRODUCTION

RECENT DISCUSSIONS HAVE suggested that a common U.K. scheme for the reformatting of ERTS digital tapes would be appropriate. The following is intended as a basis for discussion of this point.

There are several disadvantages associated with the format used by NASA for the system corrected MSS data, as is evident from the number of users who have found it convenient to reformat the data before further use. Some of these users have taken advantage of such a reformatting step to modify the data, by introducing, for example, scale and geometric corrections, which may be useful for specific purposes, but have little value in the general case.

At present there are only a small number of groups in the U.K. who are using the ERTS digital data, and thus if a standard reformatting scheme were to be established now, it is more likely to be adopted by subsequent investigators. This would permit a considerable saving of computing effort, by allowing the easy transfer of software between users. If however the wrong decision is made at this stage, the penalties are severe, so care must be taken to ensure that the adopted scheme does not preclude new uses of the data.

The most awkward features of the NASA format are:

(i) The segmentation of scan lines into quarters, each of which is stored on a physically separate tape.

(ii) The interleaving of the spectral data using adjacent pairs of pixels.

Of the several possible changes which could be made to the data format the following seem the most useful:

(i) The reconstitution of the scan lines.

(ii) The reformatting of the interleaved spectral data.

(iii) The repacking and reformatting of the annotation block.

Each of the possibilities will be discussed in turn.

The choice of format is however also dependent to a degree on the characteristics of the magnetic tape systems available, and this is a factor over which the potential user usually has little or no control, i.e. he must use such facilities

* Now at Experimental Cartography Unit, London.

58 - 51

as are available to him.

Using 9-track tapes at 1600 b.p.i. it is possible to store a complete image as 4 sequential files, identical to the original NASA tapes, on a single 2400 ft. tape, and this saving in space is obviously an advantage. However storing all the data as separate files on one tape means that if an area covered by two or more quarter strips needs to be examined, accessing that data becomes difficult. Further, the availability of 1600 b.p.i. tape drives is limited, and it seems advisable to define a scheme that will fit any system, and which permits the storing of an image on more than one tape.

2. THE RECONSTITUTION OF SCAN LINES

The reconstitution of individual scan lines has been the most commonly implemented reformatting scheme, as it permits the examination of any small area of an image without accessing several tapes. It further makes the processing of an entire image as a unit very much easier, since most image processing systems require an unsegmented raster scan. Lastly such a reformatting restores the data to the form in which it is collected, which seems a desirable feature. However, such a scheme raises certain difficulties.

Such a scheme requires a computing power that may well be beyond that available to many U.K. users. To compile scan lines from four separate source tapes would require either a multi-pass system with three tape drives, or else a single-pass system with five simultaneous tape drives. The reformatting can also be done using a single tape drive, and holding the four tape files on a temporary disc store. This requires more disc space than is usually available to most users, although the Image Analysis Group at Harwell have indicated that they can perform such an operation on their installation (Carter, 1974, personal communication).

An alternative way of implementing the reformatting scheme would be to create a small central special purpose computing facility, dedicated to performing this task. If no further reformatting were required, this system could be implemented with very little direct access core, but with either multiple tape drives, or a combination of a single tape drive and a disc pack, necessary to do the task. It would be possible to build such a system at an estimated cost of the order of £30,000. Such a facility would, however, only be justified if there were a greater use made of ERTS digital data in U.K. than is now the case.

If the scan lines are reconstituted, a second problem arises as to the format of those reconstituted lines. At present, the quarter scan lines require 3296 8-bit characters to be stored. If the reconstituted scan lines were to occupy one record, they would then be 13584 characters long, and would require large buffers which might be a problem where only small computers are available.

It is suggested that the reconstitution of the scan lines be achieved by writing the existing four data records sequentially onto the tape, instead of writing one long record. The reconstituted image could then be stored on as many tapes as the user requires, by splitting the image at given scan line intervals.

3. REFORMATTING THE SPECTRAL INFORMATION

Within each scan line, the data at present are stored in 8 character groups, with the spectral information for two pixels interleaved in the order:

$$(4, 4, 5, 5), \quad (6, 6, 7, 7)$$

thus making it necessary to read 8 characters in order to gain the spectral information relevant to one pixel.

Two reformatting schemes seem possible, either the data can be reformatted pixel by pixel, or band by band. In the first of these, all the information relevant to one pixel is stored as successive characters in the order:

(4, 5, 6, 7)

and in the second case the four records for each scan line would consist of the complete scan line for one spectral band.

For most digital applications, multispectral data are required, and a pixel by pixel repacking seems appropriate. A band by band repacking scheme would however be more suitable for pictorial processing, but in view of the low cost of the NASA photographic products, this advantage does not outweigh the increased difficulty of digital analysis.

A third alternative that has been adopted by some workers is to produce tapes which cover the image but contain only the data for a single spectral band. Such a procedure requires the compilation of a reformatted tape from four separate source tapes, and the problems already noted with reconstituting the scan lines equally apply. In addition such a scheme effectively precludes multispectral analysis of the data.

4. ANNOTATION AND IDENTIFICATION RECORDS

The repacking and reformatting of the heading blocks on the tapes would result in some increased efficiency in the reading and processing of the tapes. However, these records form such a small part of the total data set that it is felt they are best left alone.

Once scan lines are reconstituted it is necessary to reproduce the annotation block only at the head of each tape used, and to include such information as the reformatting makes necessary, for example the number of the first scan line on the tape.

Further if the heading blocks were to be reformatted, it would probably be advantageous to add to it data not currently present in the header. Data concerning satellite ephemeris is available from NASA, and it would be useful in some circumstances to have this data available. Thus if this data could be made available it would be as well to add it to the header block at this stage. Probably the easiest way to achieve this is to add a further header record at the beginning of each tape, into which the user can place such information as he cares to add. This block can, of course, be left empty if the user so requires.

5. DISCUSSION

The major advantage of reformatting the data seems to lie in the reconstitution of the scan lines. It is felt that unless this can be achieved the best course is to leave the data as it is supplied by NASA. However, this reformatting can be achieved, and it seems reasonable to put forward the following scheme for the consideration of the remote sensing community in the U.K.

1. The annotation and header blocks be left as they are.
2. An additional record block be added after the annotation block to contain such additional information as the user requires.
3. The scan lines be reconstituted, but left as quarter length blocks.
4. The data within each record be reordered so the information for each pixel appear as successive characters on the tape.
5. If demand warrants, a small, central, dedicated computer be set up to perform this reformatting.

Journal of the British Interplanetary Society, Vol. 28, pp. 608-612, 1975.

METHODOLOGICAL QUESTIONS IN THE DIGITISED ANALYSIS OF ERTS DATA

A. C. ARMSTRONG
School of Environmental Sciences, University of East Anglia, Norwich, England.

Methodological questions that arise when using ERTS digital data to derive 'land types' over large areas are discussed. Some of the disadvantages of per-point classifications as usually implemented are outlined, and the advantages of per-field classifications discussed. The superior performance of an unsupervised classification scheme in an exploratory situation is outlined. The advantages of an unsupervised per-field analysis are, however, gained only at the expense of greatly increased computing effort.

1. INTRODUCTION

IT IS THE purpose of this paper to examine some of the methodological issues that arise in the digital analysis of ERTS data for regional studies. This paper thus acts as a methodological companion to the results presented by Armstrong and Clayton [1] and the technical discussion of Armstrong (mss, 1975). Although developed in the context of ERTS-1* data, the discussion is applicable to the analysis of any remotely sensed multispectral data.

Two major points will be examined, the advantages of analysing groups of pixels rather than individual pixels, and the choice between supervised and unsupervised analyses. However, in order to give some background to the paper, the ideas behind the investigation are first outlined.

The potential of ERTS to cover the Earth every 18 days has two main implications. Firstly, the sheer volume of data introduces a need for automated data processing [2], and secondly, for the first time, a consistent data base is available for global scale investigations, offering the possibility of mapping features over wide areas in a way that was not previously possible [3].

With these considerations in mind we have been concerned with developing a scheme that has the potential for classifying areas covered by several ERTS frames. The classifications would be of the kind to permit comparisons with those shown, for example, in regional geographies and atlas maps. It is not known *a priori* exactly what types of feature can be distinguished in this way, and we are thus in a learning situation where we hope to produce classifications and then to *a posteriori* identify the classes.

A problem that arises in identifying the classes is that the landscape is an integrated system, and that when one environmental variable changes, then others change in sympathy. It thus becomes difficult to know whether we are mapping soils, geology, topography, or vegetation, since these variables are so interrelated. We thus cannot state exactly which variables we are mapping, and call the resultant classes 'land types' without attempting at this stage, to further qualify them.

* Now called LANDSAT-1.

58-54

2. PER-POINT CLASSIFICATIONS

Although per-point classification has several disadvantages, it is the technique that has been investigated the most thoroughly, so that it is now a relatively routine procedure, supported by several well-known software packages, and even some special purpose hardware.

The main feature of per-point classification is the tremendous detail it offers. When remote sensing data were collected only by special flights of special purpose aircraft, and was thus available only on a limited basis, it seemed better to analyse the data in the most detailed manner possible. However, with the advent of satellite sensing systems, an almost overwhelming flow of data has changed the situation, so that the choice is offered of low resolution broad scale studies, or high resolution detailed study. Each ERTS frame contains approximately 7 million pixels, each referring to a ground resolution element of 56 x 79 m. This level of detail means that it is possible to classify only a small area if we wish to present the results as a map with single pixel resolution. We note, however, that it is possible to classify any arbitrary large area to produce summary statistics.

A second major feature of the pixel-by-pixel classification is that because it considers only individual points, it is very sensitive to noise in the system. Under normal circumstances the noise level in the ERTS data is low, but in some cases it can become quite large. For example, we have been examining the imagery of the Wash area, in an attempt to define areas of sediment-laden water. For this purpose, the data in band five contains the most useful information, and we have found that the spectral levels associated with turbid and non-turbid water differ only by about two gray levels, which is only just greater than the noise introduced by the variability of the six sensors used. The resultant classified output thus shows a pronounced six-line streaking. Thus in this context at least a per-point analysis can be very sensitive to noise in the system. The situation is likely to be worse where examining data that is inherently more noisy, as for example the S-192 on board SKYLAB.

A further disadvantage of the pixel by pixel approach is the very limited amount of information available to the classifier, which in the case of ERTS is the values of the gray scale in the four spectral bands. Any classification scheme can thus only use the 'spectral signature' of the pixel, which can only be related to other 'spectral signatures', as a basis for classification. Because of the difficulties associated with building up an acceptable library of spectral signatures that has any general validity, these pixel-by-pixel approaches usually require the input of further ground truth for the image in question.

By examining each pixel in turn, without reference to other data points, the classifier ignores one of the most important features of the data set, that it is spatially organised. This spatial organisation of the data is important in the way the eye recognises patterns. For instance, when we recognise an agricultural area in an ERTS scene of the USA, it is by the arrangement of the fields, and not by the spectral characteristics of the crops (which we probably cannot identify in any case).

3. THE PER-FIELD APPROACH

Because the per-point analysis of remotely sensed data ignores the spatial aspects of the data, a per-field approach offers many advantages. Not only does it correspond more nearly to the way the eye observes an image, but it permits a far wider range of analyses to be implemented. However, the approach is expensive in terms of computer time and core requirements, and is less easy (in general) to implement on special purpose hardware.

The essential feature of the per-field approach is that it considers not just individual pixels, but neighbourhoods of pixels. Usually the neighbourhood is defined as a matrix of pixels n x m in size, centred over the point in question.

58-55

It is worth noting that it is not necessary to loose any of the resolution of the system by using such a procedure, for we can locate one of these neighbourhoods over each pixel in turn. This is essentially the operation performed by the local gradient operator, used to locate boundaries in the data. However, in many cases it is useful to use non-overlapping neighbourhoods, thereby achieving a generalisation of the data.

Once the per-field approach has been adopted, several further advantages accrue. Firstly, it is possible to derive many more features for classification than was available to the per-point classifier. Indeed, one of the major problems of per field classifications is the embarrassingly large set of potential features.

One of the most important features that can be derived from fields, is information concerning image texture. Texture is a property of an image that presents a great many problems, for several reasons. Perhaps the most important is that whereas the eye integrates data over an area to see 'texture', there is no adequate model of what the eye/brain system does in this operation, and thus no adequate statistical model of texture. Various attempts to measure texture in the literature include the use of various transforms, such as the Karhunen-Loève (principal components), Fourier, Walsh and Slant transforms [4]; measures based on transition matrices for gray scale values at specified spatial relations [5]; optical-analog techniques [6, 7]; and the application of two-dimensional time series techniques [8]; as well as a large number of simple measures based on gray scale distributions, gradient distributions etc. [9]. All these methods, however, use only statistical models, and do not attempt to model texture *a priori* and then calibrate the model. Some of the difference between workers is probably due to the fact that the eye sees many different kinds of things that are put under the blanket term 'texture', and the different investigators are analysing different aspects of this problem. However, it is the experience of all the investigators, that the use of texture, however it is measured, yields satisfactory classifications.

A second possibility that arises from the use of per-field classification schemes is that it is possible to alter the scale of the investigation simply by changing the field size. We have found that depending on the kind of feature that is of interest a particular scale of analysis is appropriate, and the field size can be adjusted accordingly. It is possible to envisage a hierarchy of classification scales, the lowest case being that of the degenerate neighbourhood containing only one pixel, and the results of one level of the hierarchy being input into the next. For example, some workers have discussed the possibility of using per-point classifiers as inputs to a per-field classifier for the identification of urban areas.

The list of further possibilities is potentially endless. Certainly the evidence points to the fact that moving from a per-point to a per-field scheme has the two major advantages that it is possible to use many more aspects of the data, and to examine the data at any scale required (subject, of course, to the resolution of the system). However, to pay for this greater flexibility per-field classifiers are usually expensive in terms of computer time, especially in the time required to find the information relevant to any one field. Frequently it is necessary to arrive at a per-field classification *via* several stages of computation. It is difficult to see how, at the present state of the art, a per-field classifier could be built to handle the flow of data from a system such as ERTS in anything approaching real time.

4. SUPERVISED VERSUS UNSUPERVISED CLASSIFICATIONS

The second major area of methodological debate in the analysis of ERTS data is between the supervised and unsupervised classification techniques. The popularity of the supervised classifier arises largely for historical reasons, in that these techniques were developed largely in the USA for the classification of crops, in carefully controlled conditions, for which large amounts of 'ground truth' was available. At an early stage in the development of Remote Sensing techniques, it was useful and easy to examine conditions on the ground and relate these conditions

58-56

to the outputs of the sensors, and from this line of approach the idea of spectral signatures arose. However, the difficulties associated with building a library of spectral signatures has lead to a situation where an input of ground truth is required for each image analysed. Thus the traditional supervised classification scheme requires the delimitation of 'training sets' which are used to teach the classifier. Despite the objections that will be raised, this scheme is known to work efficiently where sufficient ground truth is available.

Several workers, among them Nagy *et al* (1971) [10], have pointed out several quite fundamental objections to supervised analysis technique as it is usually applied.

Firstly, the training set is usually gathered without any reference to the assumptions of sampling theory. Often the training set is derived from areas which are thought to be particularly well defined, and all boundary effects are ignored. Basu & Odell (1974) [11] have shown that the lack of independence between sample members used to 'train' the classifier, can seriously affect the performance of the classifier. Similar considerations also apply to the 'testing set' used to evaluate the classifiers performance.

Secondly, it is very much more efficient to classify the data, and then *a posteriori* to identify the classes produced. Nagy & Tolaba (1972) [12] have shown that in this way it is possible to classify large areas with considerably less input of ground truth than the traditional supervised analysis.

Thirdly, and most importantly in the context of ERTS, the supervised approach requires a high degree of prior knowledge of the situation being sensed. The Bayesian classifier normally used requires an estimate of the mean and covariance matrices of every type to be mapped, and these can only be estimated from the training sets. There is thus no provision for starting a new class of object, no matter how strong its signal; and the only way the classifier can cope with it is either to force it into one of the existing classes, or put it into a 'don't know' class.

This last characteristic of the supervised approach renders it highly unsatisfactory for many applications. One of the values of data such as ERTS is that it is available for areas where ground truth is notably lacking, and expensive to collect. It is thus desirable to develop techniques that are applicable in such situations, and unsupervised techniques seem to offer the best possibilities.

The use of unsupervised techniques does not, moreover automatically lead to the rejection of all prior knowledge of the situation, rather to its use in a rather less direct fashion. When embarking on a classification exercise, a great many decisions have to be made, often quite arbitrary, such as what field size to use, what features to extract, the thresholds to be used by whatever classifier to be used, etc. and these decisions are influenced by such prior knowledge as is available of the scene being analysed, and the purpose of the investigation. Further, where we have no 'testing sets' to evaluate the classifiers performance, this prior information will be used to assess the progress of the classifier.

5. DISCUSSION

This paper has presented the two fundamental dichotomies in remote sensing data analysis, between per-field and per-point analyses, and between supervised and unsupervised classifiers. Whereas the per-point, supervised analysis has been developed to a point where its application to well defined problems is virtually routine, the per-field analysis offers immensely more possibilities in terms of scale of result and features available, and the unsupervised techniques offer advantages in application to new situations.

However, the greater flexibility and usefulness of the unsupervised per-field classifier is gained at the expense of increased computing time, so much so that it is difficult to see how such a set of techniques could ever be applied on a routine basis to a continuous data stream such as that potentially available from ERTS. Nevertheless, the great successes of the per-point supervised analysis must

58-51

not be allowed to dominate remote sensing data analysis to the extent that the alternatives are overlooked.

ACKNOWLEDGEMENTS

This research was funded by a grant from the Natural Environment Research Council. Thanks are due to Professor K. M. Clayton, Dr. J. R. Tarrant, and Dr. I. F. Hill for their comments, help, and advice.

REFERENCES

1. A. C. Armstrong and K. M. Clayton (in press). 'Objective generalisation using ERTS images', in E. C. Barrett and L. F. Curtis (eds). *Environmental Remote Sensing Practices and Problems.* Arnold, London.
2. I. F. Hill and J. R. Tarrant (1974). 'A Computerised System for Identifying Changes in the Earth's Surface Cover'. *JBIS* v. 27, p. 45-47.
3. K. M. Clayton (1974). 'Synoptic observations at a global scale'. *JBIS* v. 27, p. 23-28.
4. L. Kirvida (1973). *'Automatic classification of remotely sensed data'.* Final Report on contract NAS 5-21866, prepared for Goddard Space Flight Center. Honeywell Inc., Minneapolis, Minnesota.
5. R. M. Haralick and K. S. Shanmugam (1974). 'Combined spectral and spatial processing of ERTS imagery data'. *Remote Sensing of Environment*, v. 3, p. 3-13.
6. H. Kaizer (1955). *A Quantification of texture in aerial photographs.* Technical note 121, Boston University Optical Research Lab.
7. N. Gramenopoulos (1974). *Automated thematic mapping and change detection of ERTS-A images.* Final Report on contract NAS 5-21766, prepared for Goddard Space Flight Center, Itek Corporation, Lexington, Mass.
8. S. N. Jayaramamurthy (1973). *Computer methods for analysis and synthesis of visual texture.* Report 601, Department Computer Science, University of Illinois, Urbana, Illinois.
9. J. K. Hawkins (1970). 'Textural properties for pattern recognition'. B. Lipkin and A. Rosenfeld (eds) *Picture Processing and Psychopictorics.* Academic Press, New York.
10. G. Nagy, G. Shelton and J. Tolaba (1971). Procedural questions in signature analysis. *Proceedings 7th International Symposium on Remote Sensing of Environment*, v. II, p. 1387-1401, Ann Arbor, Michigan.
11. J. P. Basu and P. L. Odell (1974). Effects of intraclass correlation among training samples on the misclassification probabilities of Bayes' Procedure. *Pattern Recognition* v. 6, p. 13 16.
12. G. Nagy and J. Tolaba (1972). 'Non supervised crop classification through airborne multispectral observations'. *IBM J. of Research and Development*, v. 16, p. 138-53.

(Presented at the Symposium of the British Interplanetary Society on 'European Participation in Earth Resources (Space) Projects' held at University College London on 9 April 1975.)

58 58

8/1

The relative performance of some unsupervised clustering techniques for the per-Field classification of LANDSAT data

by A. C. ARMSTRONG

ABSTRACT:

Three algorithms for the per-field analysis of LANDSAT data are described. Methods of generating per-field data sets are briefly outlined, and the performance of the three algorithms discussed. At any given spatial scale, the traditional hierarchical algorithm produces better classifications than a multi-pass iterative technique, which is in turn better than a single pass chain algorithm. However, the computation requirements of these algorithms is such that for large data sets, the hierarchical technique is too expensive, and the iterative technique always more expensive than the chain.

The value of an unsupervised approach to the large scale generalisation of ERTS data has already been demonstrated by Armstrong & Clayton (1). In this paper the comparative performance of three algorithms for these operations are discussed. As in most computing applications, there is a trade-off between the computing time required to achieve the results, and the quality of those results. However in this case the problem is confounded with that of the scale of the investigation which affects the suitability of the algorithms.

The Algorithms.

The three algorithms discussed here are by no means an exhaustive sample of the many clustering techniques available, but represent three basic types of approach. These algorithms are:

(i) the classical hierarchical technique for maximising an objective function, derived initially by Ward (2).

(ii) An iterative technique described by Swain (3) which makes an initial guess at the results, and then improves upon that initial estimate.

(iii) the chain technique described by Bonner (4) and implemented in the remote sensing field by Nagy et al (5).

These three techniques represent three levels of data manipulation: the first technique takes all the data and from it produces a derived data set, the interdistance matrix, on which it performs all subsequent calculations; the second examines each data point in turn, and does so several times; the third performs a single assignment operation to each point in turn, thus scanning the data file only once.

Although it is not necessary to describe these algorithms in detail, since they are well described elsewhere, some of the problems that arise when implementing these techniques for remote sensing applications will be discussed.

(i) The Hierarchical Technique.

This algorithm takes the data, and from it calculates a matrix of distances between each point. All subsequent steps in the algorithm are based on this interdistance matrix, and thus this initial stage is critical to the results obtained. In some applications it is conventional to normalise the measurement vectors before this step, so that all variables are given equal importance in the analysis. This procedure has the effect of reducing variables collected on differing measuring schemes to a common standard, but might equally swamp out any signal present with noise, or equally enhance subtle patterns that might be otherwise overlooked. It is thus necessary to have some *a priori* knowledge of the characteristics of the data before this step can be taken. One way of ensuring that the input data has well defined characteristics is to perform a principal components transformation of the data before this step in the analysis. Although such a transformed data set would enable the use of an optimally efficient subset of the data, it requires large amounts of computing time to be implemented, and it is not suited to remote sensing applications which normally involve large data sets.

A second consideration at the stage of calculating the interdistance matrix is the choice of an appropriate distance measure. Usually Euclidean distance is taken without much further consideration. However other schemes can seem equally attractive for specific applications. City-block distance has often been used as a cheaper alternative to Euclidean distance, but in view of the critical nature of this step, and its relatively small contribution to the total computing effort, the saving is not sufficiently great to argue for its use unless a priori reasons indicate that it is appropriate.

Once the critical step of defining the interdistance measure has been calculated, the algorithm follows a standard procedure, of scanning the interdistance matrix for the two points closest together, then replacing those two points by their weighted mean. Thus to cluster n points this technique requires that a matrix of interdistances, which is (n-1)×(n-2) in size, be stored, and scanned (n-1) times. There are thus very real limits on the size of data set that can be handled by this algorithm. Most computer systems can hold up to 200 points, but beyond this limit resources tend to become strained. Openshaw (6) has shown that if, for a spatially organised data set, we impose the constraint that only contiguous points can be joined, it becomes unnecessary to store most of the interdistance matrix, and that a considerable saving can then be achieved, such that up to 1200 points can be easily handled. However, imposing such a constraint upon the data requires that we a priori know adjacent points to be similar, and also raises problems when non contiguous areas are in fact similar.

The hierarchical technique was implemented for this study using the FORTRAN program published by Veldman (7) and the data were all treated in an unnormalised fashion, using the Euclidean definition of distance throughout.

(ii) <u>The iterative technique</u>

The iterative technique discussed by Swain (3) has the advantage that it does not require large amounts of computer store, since the data are always accessed sequentially, and can thus be stored on a temporary device such as magnetic tape or disc. This algorithm takes some initial estimates of the cluster centers to be derived, the data are then assigned in turn to the nearest of these cluster centers. Each cluster is then replaced by the mean of the data points assigned to it, and the process repeated until a stable assignment of points achieved. This technique however does have several disadvantages:

(i) an a priori estimate of the number of clusters to be extracted is required. Often the investigator has no idea of this number, and so usually a value in excess of that required is given. Usually it requires several attempts to find a useful classification.

(ii) initial estimates of the cluster centers must be made. Unless some prior information is available, these must be determined arbitrarily. We have found that these are best determined by assigning some of the data points, evenly spaced through the data set, to these cluster centers. The use of any other arbitrary scheme usually results in some of the initial centers being outside the cloud of data points, and remaining empty throughout, so that they involve additional computation with no effect on the results.

(iii) the technique tends to use large amounts of computer time. To classify n points into m clusters requires the (n x m) distances be classified every iteration. It was found that in most circumstances the time taken to perform the analysis did not greatly depend on the number of cluster centers, but on the number of data points considered, and that a limit of 1000 points was all that could be routinely used. However, some results are presented here for a larger data set which involved much larger amounts of time than were routinely available. It is however possible to reduce the time requirements of this technique by stopping the procedure before convergence. Examining the reclassifications that occur at each iteration we observed that the matrix of transitions soon becomes very nearly empty, and that all the reclassifications are between two or three clusters. We also observe that these clusters are very close together in the feature space. Thus we find that the algorithm converges to a stable solution for a large proportion of the data set (over 95%) within 5 or 6 iterations, and that it takes another 20 or so iterations to resolve the residual uncertainties. If we accept this small level of uncertainty in the results, we can achieve a solution that will probably be as good for our purposes within a much shorter time.

(iv) there is no guarantee that the resultant clusters do not overlap in the measurement space. It thus becomes necessary to examine the 'separability' of the derived clusters, so that all clusters that are not truly separable can be amalgamated. In the original scheme discussed by Swain this separability was determined by the Swain-Fu distance. However this distance measure, and indeed all other measures commonly used, require that at least one variance-covariance matrix be inverted for each distance measure calculated. Since this technique was intended for use with data sets measured on many variables (up to 128), it becomes impractical to store many of these large matrices, and exceedingly expensive in computing terms to invert them. For this investigation an alternative distance measure was devised, which had the form:

$$D_{ij} = \frac{1}{n} \sum_{k=1}^{n} \frac{|x_{ik} - x_{jk}|}{\sigma_{ik} - \sigma_{jk}}$$

where D_{ij} is the distance between clusters i and j

x_{ik} is the mean of cluster i on the k-th variable

$_{ik}$ is the standard deviation of the k-th variable in cluster i.

This distance measure has no distribution theory attached to it, and so it remains essentially a descriptive, and not an inferential statistic. It has a lower bound of zero, but no upper bound. Values of less than 1.0 indicate that on average two clusters overlap to some extent, and values of 0.5 or less indicate considerable overlaps. In practice, the relative values of this statistic was found to be the most valuable, and the mean of its value gave an overall indication of the clarity of the classification that could be derived from the data set in question.

(iii) <u>The Chain Algorithm</u>

This algorithm was first described by Bonner, (4) and includes some of the modifications discussed by Nagy and Tolaba (8) for use in the remote sensing context. This technique involves scanning the data in sequence, and for each point calculating the distance between that point and a list of cluster means. If the point is nearer to the nearest cluster than a given threshold value then it is joined to that cluster, but if not, it is used to initiate a new cluster. The only *a priori* information that the algorithm requires is the value of this threshold, which has the dimensions of distance in the measurement space. It was found that in practice the value of this threshold could only be determined by trial and error, and further that the performance of the classifier was sensitive to this value over quite a narrow range which could only be determined by examining the effects on the analysis of varying the threshold. The sensitivity of the results to this parameter which can not be determined *a priori* was the major disadvantage of this technique, although this was the sole problem of implementation.

Thus, this technique starts with an empty list of clusters, which is gradually filled as the points are examined in turn. As the programme proceeds, the number of distances calculated for each point increases as the number of cluster centers increases. However, some of the new clusters that are formed are only aberrant single points, and to prevent the programme maintaining these aberrant points as cluster means, after specific numbers of points have been examined all cluster means are examined, and those including only a specified small portion of the data, or which have not been used within a certain period of time, are deleted from the cluster list. Although this procedure involves the estimation of two further parameters, for which we have no *a priori* guides, the performance of the classifier is not sensitive to their values and the overall efficiency of the programme is much improved by the inclusion of this deletion routine.

The Data

Although many investigators have classified ERTS data point by point, considerably less effort has been made towards a generally acceptable scheme that operates on groups of pixels, or 'fields'. In order to achieve generalisation over large areas, it is efficient to amalgamate the data into fields before classification is attempted. Once such fields have been defined it is also necessary to extract features from those fields so that classification can be achieved.

Field definition involves defining the arrays of pixels that are to be the basic units to be classified. Convenience dictates that regular arrays of pixels constitute the fields to be used. In this study three basic sizes of field were used, all being square arrays of pixels which refer to parallelogram shaped areas on the ground, covering 50 x 50, 100 x 100, and 200 x 200 pixels respectively (Table 1). It was possible to use this hierarchy of fields to examine the effect of varying field sizes on the performance of the classifiers.

There are few *a priori* guides as to what might be an effective set of features that might be put into the classifier. Experience to date with LANDSAT imagery has shown that for most applications the use of data from all four spectral bands is better than analyses of single bands. For this study the information contained in the gray scale histograms for each field was used as the feature file for input into the classifier. For each field the four spectral bands were each split into 32 equal classes and the number of pixels with values in each of the total of 128 classes recorded. This vector of 128 numbers was then used as the feature vector for each field.

One advantage of this feature scheme is that it is possible to amalgamate the data for groups of fields simply by adding the feature vectors. It was thus possible to perform analyses at the three scales of field from the one data set, that using the smallest fields, thereby enabling a considerable efficiency to be implemented. Such an efficiency would not be achieved by some alternative schemes, such as the use of various transforms to measure texture, which would require re-computing for every field size used.

This feature set has the further advantage that, because it represents the frequency distribution directly, it is not in any way influenced by problems that arise in the processing of polymodal distributions, as are measures based on moments of the distributions. Further, for most purposes the moments of the distributions could, if required, be estimated from this data, and thus the feature file compiled for input for the classifier can act as a valuable secondary data source.

Lastly this feature file results in a measurement space that does not require normalisation. The values on each of the 128 variables are measured in the same units, and the variables are directly scaled by their areal importance.

Results

Three sets of images were used to test the performance of the algorithms, two of California, and one of Eastern England. (Table 2). The images of Central California show a very clear regionalisation into a series of parallel NW-SE trending zones, which also include a repetition of some zones either side of the irrigated area in the central Valley. The two images are of the same area, separated by a period of

9 months, the first being taken in August '72, and the second in May '73. The contrast between the images highlights the effects that seasonality may have on the imagery.

The image of Eastern England is considerably less clear, due to the presence of cloud in the West of the image, and a very large amount of atmospheric haze. This image represents a much more severe test of the classifiers.

The performance of the three algorithms in terms of their consumption of computing resources is shown in Table 3. As can be seen the three algorithms differ greatly in their requirements of computing time, although it was found that these time requirements did not vary significantly with the data sets used. Several points can be made in the light of this table.

Firstly it is not possible to use the hierarchical technique for anything other than the largest fields, since the requirements of core space become prohibitively large. Secondly the iterative technique becomes very expensive to operate in terms of the consumption of time, with large numbers of points, and that it is not possible to use this algorithm for routine application to the small 50 x 50 fields.

Although it was not possible to devise an objective criterion of the accuracy of the classifications, it was possible to give a subjective evaluation in terms of the interpretability and spatial coherence of the classes, which were in turn related to our <u>a priori</u> knowledge of the regional geographies of the two areas. In all cases it was seen that

the 'best' classifier was always the most expensive, that is the hierarchical technique always produced better classes than the hierarchical, which was in turn always better than the iterative. However the differences between the iterative and chain techniques when applied to the large data set of 2800 fields are small, and indicate that in view of the computational economy it offers, the chain technique should be used. Thus in defining the size of fields to be used, the investigator in effect determines the algorithm that is appropriate.

The problem of the choice of scale is however difficult, for it is possible to observe that altering the field size is equivalent to reducing the resolution of the classification scheme, and we find that in altering the resolution we are not only able to locate the same feature more accurately but also to see smaller features. This change is most marked if we compare the classifications using the large 200 x 200 fields and the small 50 x 50 fields for the images of California. The classification using the large fields for this reproduces almost exactly the broad zonations that we see in the image. However, the low resolution of the system means that boundaries of these zones are located only crudely. Moving to smaller fields however does not just reproduce the same classification with better definition, but produces a different kind of classification. The features that form part of the broad classes defined using the large fields become individual classes in their own right, so that a far more complex pattern arises. Thus by reducing the scale of the input, the classifiers respond by changing the scale of the classes produced.

This problem of field size is not independent of the image used. We have noted that the 200 x 200 fields preferably classified using the hierarchical technique, produce very satisfactory classification of California. The same scheme however when applied to the image of Eastern England is considerably less successful, and it is possible to interpret only three groups from the classifier; clear sea, turbid water, and land. Moving to a smaller field size greatly improves the performance of the classifier, which then starts to produce coherent and interpretable classes in the land. However even at this smaller scale the algorithms do not function very well with this poor quality image.

It is possible to conclude that the choice of an appropriate scale depends not only upon the required resolution of the classified map, but also upon the nature of the underlying terrain. Where the scale of the underlying spatial elements is small, as with the English imagery, small fields are indicated, but where the landscape is made up of larger units, so larger fields can be used.

References

1. Armstrong, A.C. and Clayton, K.M. (in press) Objective generalisation using ERTS images. In; Environmental Remote Sensing Practices and Problems. ed. E.C. Barrett & L.F. Curtis.

2. Ward, J.H., 1963. Hierarchical grouping to optimise an objective function. J.Am.Stat. Ass. v.58, p. 236-44.

3. Swain, P.H., 1972, Pattern recognition: a basis for remote sensing data analysis. LARS Inf.Note 111572

4. Bonner, R.E. (1964) On some clustering techniques. IBM Jour. Res. and Dev. v.8, p. 22 - 32.

5. Nagy, G., Shelton, G. & Tolaba, J. (1971), Procedural questions in Signature Analysis. IBM Research RC 3517.

6. Openshaw, S., 1974, A Regionalisation Program for large data sets. Computer Applications, v.1, no. 3/4, pp. 136 - 60.

7. Veldman, D.J., 1967, Fortran Programming for the Behavioural Sciences. Holt, Rinehart & Winston, New York.

8. Nagy, G. & Tolaba, J., (1972) Nonsupervised crop classification through airborne multispectral observations. IBM Jour.Res. & Dev., v. 16, pp. 138 - 53.

Table 1

Fields Used for this study

Size of Fields in Pixels	Horizontal Size (Km)	Vertical Size (Km)	Number in an ERTS scene
50 x 50	2.8	3.7	2806
100 x 100	5.7	7.4	704
200 x 200	11.4	14.8	176

Table 2

Images discussed

Local code	ERTS frame	Date	Area
Image D	1056-18114	17 Sept 1972	Central California
Image E	1308-18122	27 May 1973	" "
Image C	1031-10334	23 August 1973	Eastern England

Table 3

Performance of the Algorithms

Algorithm	Number Points	Core	CP Time	Tape I/O time
Hierarchical	176	190	31 s	15 s
Iterative	176	60	60 s	15 s
	704	76	4 m	24 s
	2806	114	19 m	13 m
Chain	176	60	15 s	10 s
	704	80	20 s	24 s
	2806	98	60 s	30 s

Note:

Core measured in Kbytes (i.e. words x 4 x 1000)

Machine used was IBM 370/165

The times are total amounts of time used, which are considerably less than the total real time that the jobs are in the machine.

9/1

RAPID DIGITAL MAPPING OF COASTAL AREAS FROM LANDSAT

Adrian C. Armstrong and Peter Brimblecombe

School of Environmental Sciences
University of East Anglia
Norwich NR4 7TJ

Abstract

The value of repeated imagery from satellite platforms can only be realised if coupled to rapid mapping systems. Techniques for producing rapid, user oriented maps from LANDSAT data are discussed, and examples derived from the mapping of the coastal and off-shore areas of eastern England.

58-15

1. Introduction:

Conventional surveying and cartographic techniques are usually considered excessively expensive and far too slow for the monitoring of rapidly changing areas. It is the aim of this paper to show how remotely sensed data can provide the basis for cheap rapid cartography, which can be tailored to suit a potential user's needs. Remotely sensed data, particularly from the LANDSAT series of satellites (previously known as ERTS) has proved its utility over large areas of the world, and particularly in the United States (for example papers see Freden, Mercanti & Becker, 3) However its utility in the British context is less clear, since excellent cartographic coverage of most environmental features already exists, and the need for repeated coverage with a spatial resolution of only 80 m is questionable. We will attempt to show here that some very simple digital techniques are capable of providing rapid, useful cartography of changing areas. The example we consider is that of mapping the changing sand banks and suspended sediment distributions in coastal areas, providing maps that are relevant to navigation, sedimentology, recreational planning and coastal protection. The major advantage of this method of mapping is the speed with which it can produce a specialist map of an area, although in doing so it may sacrifice some of the absolute cartographic standards that are normally expected. The maps can be produced at an extremely low cost and the system will allow regular re-mapping at frequent intervals.

2. Data Characteristics:

The specifications of the LANDSAT data collection system has been described in detail in several sources[7] so only a brief summary will be given here.

The two satellites are in near polar, sun-synchronous orbits at a height of 950 km, and provides the potential for covering the earth every 19 days.

The data is collected by two sensing systems, the Return Beam Vidicon camera (RBV) and the Multi-Spectral Scanner (MSS). For technical reasons, the data has been restricted to that collected from the MSS in most cases, and that is the sole data source that will be considered here. The multispectral scanner is a device for recording the amount of radiation reflected from a small area of the earth's surface. The optics of the sensor focus on an area about 80 m square and the amount of light returning from this area is recorded as a voltage in sensors which operate in four spectral bands (see Table 1). A mirror which oscillates over a small angle ($\pm 2.8^{o}$) alters the position of the Ground Resolution (GRE) being sensed by the element system, scanning a strip across the earth beneath the satellite. The responses of the sensors are sampled at regular intervals, corresponding to a value for every 56 m of the ground. The forward motion of the sensor platform moves the scanner forward so that successive scan lines are located parallel to one another. In order to ensure complete coverage of the surface the MSS senses six such strips simultaneously. These signals are digitized to a 64 value scale, and telemetered to a receiving station on the earth, where this raw data string is computer processed by NASA. The continuous data stream is split into separate scenes for convenience. The data is available in two forms, images and bulk processed digital data on computer compatible tapes (CCTs).

Images can be created from the data stream by a procedure which is essentially the reverse of the scanning. A fine light beam is moved in a regular scanning raster corresponding to the original scan of the sensor and its intensity is varied in the same way as the intensity recorded by the satellite based sensor. This produces an image in much the same way as on TV screen, and a permanent copy can be produced photographically. Such an image has most of the qualities of an air photograph.

The digital data source is subject, at this stage, to certain corrections. There is a slight variation in the speed of the scan as generated by the mirror, so it becomes necessary to ajust the number of data points in each scan line to keep these constant for a given scene (usually 3240 points per line). Secondly the variations in sensor response are removed by reference to a constant calibration source. Thirdly the digital values in bands 4, 5 and 6 are "stretched" from their original 64 unit scale to a 128 unit scale by a non-linear transform which gives a linear relationship between the recorded value (grey scale) and the absolute amount of radiation sensed.

The major geometric distortion in the imagery is caused by the earth's rotation beneath the satellite. Although the satellite maintains a constant path, the resultant scan lines are displaced to the west as the satellite continues its forward motion. Although a more complete analysis of the geometry of the data includes several further corrections,[5] to a first approximation, we can represent the data for a single scene as a parallelogram, with 3240 pixels (picture elements) across each scan and with 2340 lines to each scan, covering an area 185 x 185 km, (100 x 100 nautical miles). There are more than 7 million pixels in a single image, each one covering an area 79 x 56 m, with the values of the reflectance in four spectral bands recorded at each of these points. In view of the size of the data set (each scene of 7 million data points is collected in 25 seconds) the need for computer handling of the data is obvious, if even a small fraction of the information is to be used.

3. Basis of Computer Mapping:

The basis of computer mapping is the ability to characterise the response of a category of surface features in spectral terms. There are several ways of approaching this problem, the three most important being:-

(i) an ad hoc approach which arbitarily splits the four dimensional response-space into groups.

(ii) Bayesian pattern recognition techniques which identify areas of each ground type, and then use these as training areas to develop statistics which are used to classify all additional points.

(iii) Unsupervised classifications, which groups all similar data points together so that they can be a-posteriori related to ground types.

Of these methods the second has been used by the LARSYS package[4] and the third by Nagy and Tolaba[6]. The first technique is by far the simplest, and has been used by many workers including for example David, Deries & Verger[2].

It is our experience that the ad hoc multi-band density slicing procedure yields the fastest and most accurate mapping technique for coastal regions, and furthermore has the advantage of great flexibility, and minimal requirements on computer time and core space. The alternatives may be more elegant, and offer the possibility of transfer to other data sets, but this advantage is bought at the expense of greater complexity. Where the object is to produce rapid, cheap cartography, the first of these seems to be the obvious choice.

A convenient starting point is to look at the intensity levels in the four spectral bands over the region of interest. In the present study a number of scan lines across the Wash were examined and it was possible to identify areas of land, wet dark sand, dry bright sand, shallow water, and clear water, from some knowledge of the spectral response of these surface-types and their position along the scan line. For example very low values in band 7 are associated with water, which has a low reflectance in this part of the spectrum. High values of reflectance in band 4 over the water scanning scales, and to displace every twentieth line one position to the [illegible] corrects for skew. This removes all but 3% of the vertical scale distortion and all but 5% of the skew, giving a map with a scale of very nearly 1:22,000 (to be exact 1:22,047 or about 3 inches to 1 mile). Although the resolution of the map is low, objects and boundaries can be positioned relative to each other within about 100 m and every data point (excluding a very small number of spurious points arising from electronic noise) represents a "real" piece of information about the location and is not based on interpolation as is often the case in map production. Reducing the output dimensions by a factor of ten gives a scale of nearly 1:230,000 and a map of potentially great utility.

4 Results

The results present here ate two maps showing the Wash, and the Humber Estuary. The data was derived from LANDSAT image ERTS-1031-10334, taken on 23rd August, 1974 at 9.40 am. These maps were originally output into a line printer and hand coloured. Since these original outputs were some 4 m^2 in area, photo-reduction was essential.

For the purposes of this paper, the output has also been agreed by a graph plotter, to produce a black and white map. The graph plotter has the potential of bing able to accept considerably more sophisticated geometric corrections, a continuous range of scaling, and a larger range of symbols, so for the production of high quality maps production of high quality maps such a system would be preferable. David et al.[2] have in fact implemented such a system using a plotter with six different coloured pens. However it is our experience that plotter systems are slow and expensive compared to the line printer, and that they are best used only for production of final maps, and then only if economically justified.

The accuracy of the resulting maps is difficult to assess quantitatively, since they were collected in the absence of ground truth, and it is possible only to make some qualitative statements. The presence of sea fog off the North Norfolk coast prevents the estension of the map into that area. This demonstrates one of the main restrictions of remotely sensed data, that they are affected storngly by atmospheric conditions (except for such systems as SLAR). Another error has been a relatively high rate of mis-classification of the land. In particular the confusion between urban areas and mud flats such that most of the urban area of Hull has been classified as mudflat. This is due to the low coverage of green vegetation in both which result in low IR reflectances in both regimes. However for the present task of mapping coastal areas this has not proved troublesome, for once the coastling has been established the misclassification of features behind the shore line is of no interest. The accuracy of the rest of the seems to be associated with areas of turbid water. Figure 1 shows a typical scan line and the surface categories into which various spectral classes were assigned. This information was used to construct a decision tree (Fig. 2) and every point in the area could be classified using the sequence of decisions.

Once the decision tree has been established it is a simple matter to read a scan line at a time, and then to classify each of the points along the scan line. If the output of the single line is sent straight to a plotting device at this stage there is no need to store the classified image and it is thus possible to use a classifier of this type on a computer with a very small amount of core (less than 20K core words). However, the nature of the output presents a problem. The fastest and easiest technique is to use the computer line printer and this is the device that has been used most frequently in the present work. Alternative plotting devices such as multicolour plotters as used by David[1], or the colour TV screens used by various American workers appear to offer much better quality but at an increased cost.

Using the line printer as the main graphic output device, we wrote out a symbol that corresponded to a surface type at each point along the scan line. The restriction on the size of a line that can be output on the line printer means that it is necessary to reconstruct the image from several sheets of line printer output. Furthermore, the geometries of such output and that of the scanner area are not identical. The spacing of the text symbols is ten to the inch in a horizontal direction and six to the inch in a vertical direction, which does not correspond to the 79 x 56 m size of the original pixels and no allowance at all is made for skew. To correct for these two effects, it has been found simplest to omit every sixth line of data, which allows for the differences in line printer and classification is less easy to establish. The classification that was originally established for the Wash was used without modification on the Humber which comes from a quite different part of the same scene. In particular it is observed that nearly all the shallow water in the Humber is classified as mud, and the clear, but still turbid water in the deep channel is classified as shallow water. This implies that in this area the water is reflecting more light in all channels, most significantly in the infra-red. This is probably due to the higher turbidity of the Humber compared with the waters of the Wash. Although the labels on a particular class might be incorrect, their spatial arrangement makes good sense in terms of the known mechanics of the river. It would be a simple matter to device a modified decision tree for the Humber and produce a map which classified the areas in a more satisfactory manner.

An inaccuracy induced by the nature of the scanning system itself, arises from the fact that the six scanners which scan across the image in one spectral band do not have exactly the same response. This effect has been partially removed during the production of the computer compatible tapes by NASA, but some still remains. (Further corrections have been introdu ced by NASA subsequent to the production of the data set used in this example.) The effect of this is to introduce a streaky pattern of noise on the map, which has a tendency to repeat itself every six lines, and is particularly evident in divisions between shallow, turbid and clear water, where the classification involves the arbitary division of a continuous scale into three classes. The noise level is often two grey scale units and where classification boundaries are only three units apart the noise can prove a very noticeable feature. This noise can be removed by a simple filtering process which multiplies the values on each line by a constant correction factor.

$$\frac{1}{(\bar{g}_{ii} / \bar{g}_{ii})}$$

where $\bar{g}_{ij}$ is the mean value of the reflectance in band i and the j is the line number modulo 6. However the correction has not been applied to the image used here.

The detail that can be seen in the map is demonstrated by fig. 4 which shows a detail from a map of the Hunstanton area. Here we see a feature classified as a mud bank running at about 45^o to the coastline, which could easily be considered a random classification. However detailed examination of air photographs and a field check revealed that the feature is a mussel bed some 80 m across. In this instance the feature is very near the resolution limit of the scanner optics, yet it has been mapped and classified. Two types of beach material were also evident, the darker wet sand near the sea edge being easily distinguishable from the light coloured aeolian sand further inland. However these features are so small that it is virtually impossible to identify them on the photographic images from the satellite data. Thus to extend LANDSAT data towards its resolution limit it is necessary to use computer compatible tapes rather than photography. The tapes have a two fold advantage; not only do they allow examination of each individual pixel, but also to quantify information about the spectral properties of the pixel.

5. Discussion:

It is possible to produce rapid and relatively accurate maps of coastal areas. It is our belief that now the data handling routines have been established it would be possible to produce a map from a given tape of LANDSAT data of specified classes, within about a week (i.e. a total of six jobs on the computer). Whereas this technique is not capable of initial mapping of previously unmapped areas to within acceptable standards, it is quite capable of producing thematic maps of specific surface types which can be controlled in absolute location by reference to other more conventional cartography. The advantage of this technique is not its locational accuracy or fine resolution, but its ability to produce a specialist map to specifications desired by the user. Such maps would be ideally suited for resource inventories and the monitoring of rapidly changing areas. Possible the main limitation of the method is that the maps are essentially maps of spectral classes and the greatest difficulty arises in assigning a spectral class unambiguously a surface type. Thus it is necessary for the user to be aware of the limitations of the classification system, or where the maps are to be used by non-specialists, that spurious data points and areas of mis-classification be cleaned up by hand or further computer routines. With the construction of a receiving station at Fucino, Italy, it is potentially possible to obtain coverage of a given area of the British Isles every 18 days and thus, within the limitations of cloud conditions, regular monitoring by the methods described is eminently feasible.

Acknowledgements:

Computer facilities were provided by University of Cambridge Computer Centre and the University of East Anglia Computer Centre, A.C.A. was funded by a NERC during this research. Thanks are due to Prof. Clayton for encouragement and advice, Dr. Ian Hill for his initial work on handling the data tapes, and to the staff of the UEA Computer Centre for their technical assistance.

References

1. D-J. David 1975 Automatic cartography of ERTS teledetection data. Paper presented to British Interplanetary Society Symposium on "European participation in Earth Resources (Space) Projects", London 9th April 1975. to appear in the Journal, British Interplanetary Society.
2. D-J. David, J. Deries & Verger 1974. Cartographie Automatique des marais et wadden a partir des donnees multispectrales d'ERTS 1. Memoires du Laboratoire de Gemorphologie de L'Ecole pratique des Hautes Etudes, no. 26. Dinard, France
3. S.C. Freden, E.P. Mercanti & M.A. Becker (compilers) 1974 Third Earth Resources Technology Satellite - 1 Symposium. NASA, Goddard Space Flight Center, Greenbelt, Maryland. 3 vols.
4. J.C. Lindenlaub 1973. Guide to multispectral data analysis using LARSYS LARS Information Note 062873. Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
5. V. Kratky 1974. Cartographic accuracy of ERTS Photogrammetric Engineering v. 60 no. 2 p. 203-12
6. G. Nagy & J. Tolaba 1972. Nonsupervised coop classification through airborne multispectral observations IBM Journal of Research & Development v. 16 p. 138-53.
7. NASA ERTS Users Handbook NASA Document 71 SD4249 first issued 1971, with various updates 1971 - 1972 NASA, Goddard Space Flight Center, Greenbelt, Maryland.

Table 1.

The bands of the LANDSAT Multispectral Scanner (MSS)

Band	Wavelength (micrometers)	Colour
4	.5 - .6	Green
5	.6 - .7	Red
6	.7 - .8	Infra-red
7	.8 - 1.1	Infra-red

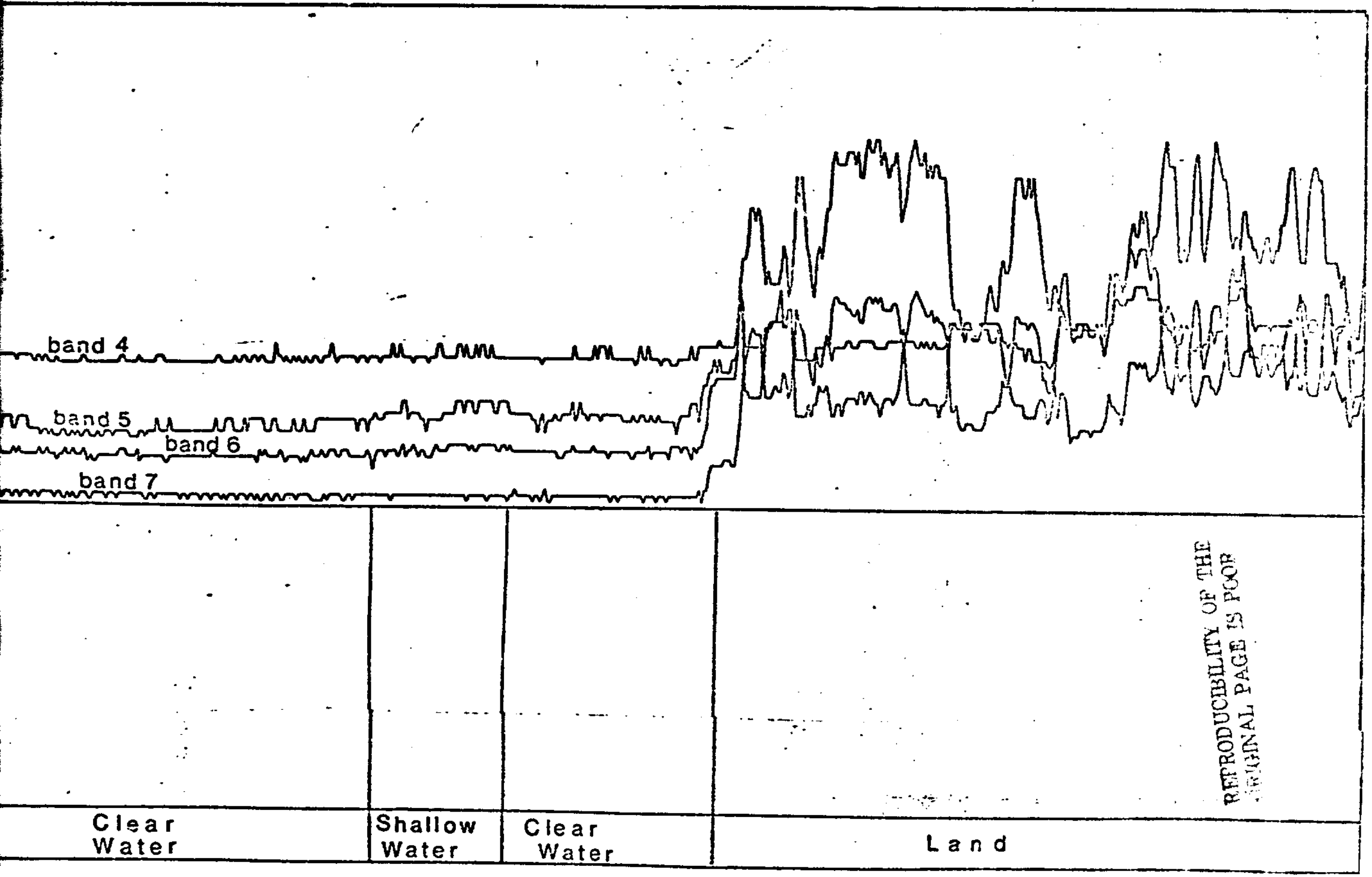
Figure Captions

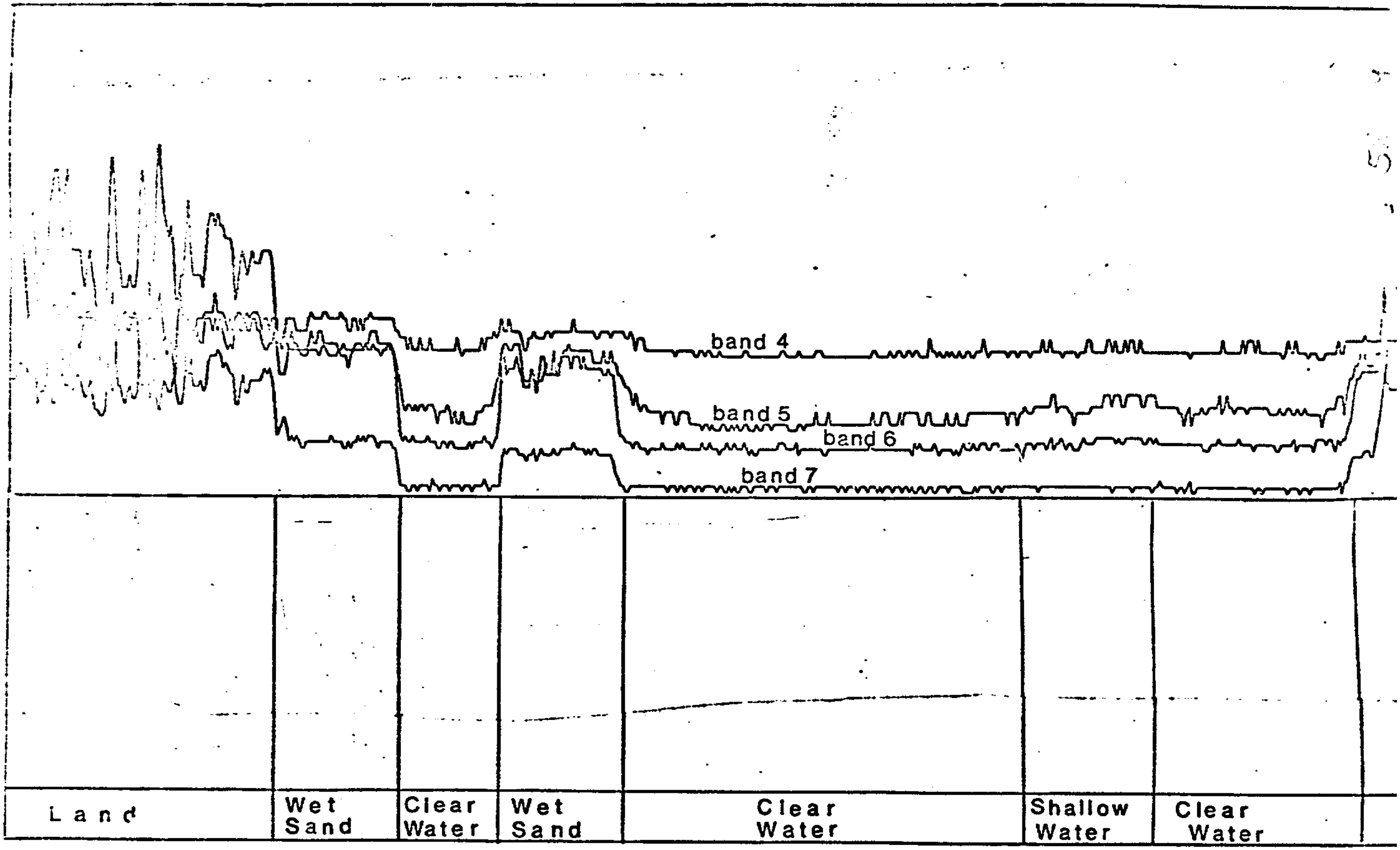
Fig.1. An example scan line across the Wash.

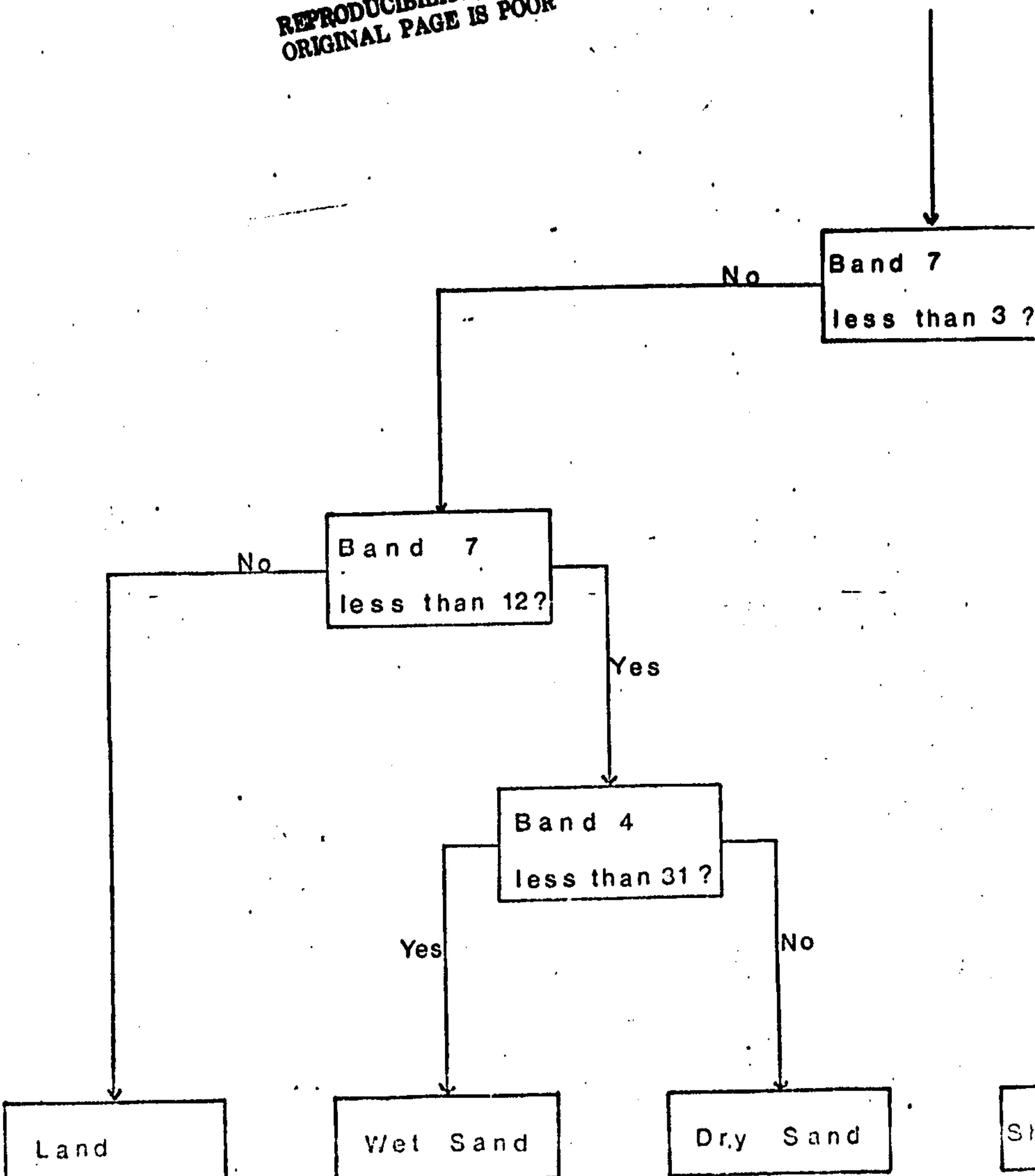
Fig.2. Decision tree used to categorise each pixel.

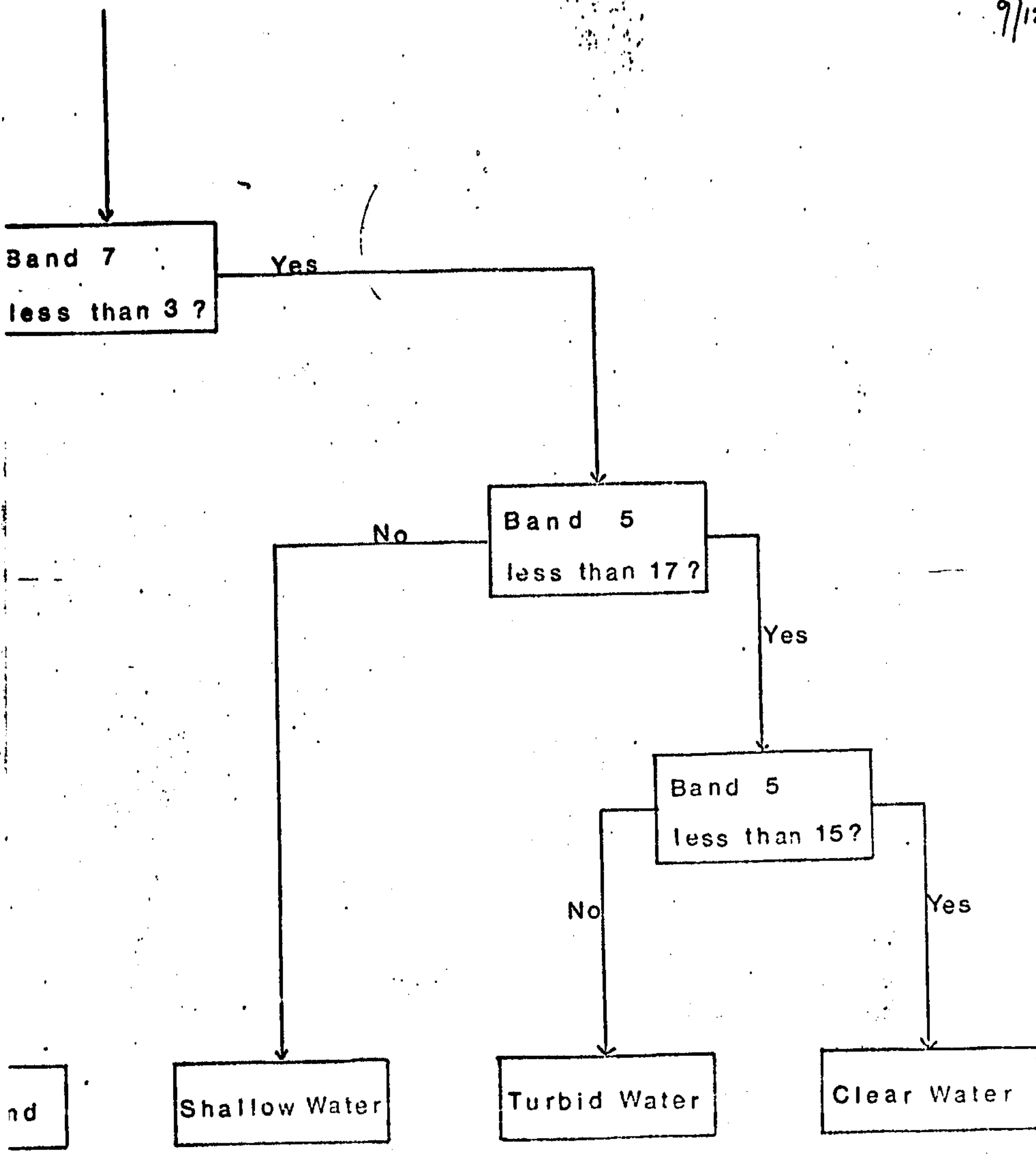
Fig.3. Example outputs from the classification program.

Fig.4. Example of the accuracy of the classifier. The Hunstanton area as classified by the program, and by visual inspection of air photographs.


band 4
band 5
band 6
band 7
Clear Water
Shallow Water
Clear Water
Land


band 4
band 5
band 6
band 7
Land
Wet Sand
Clear Water
Wet Sand
Clear Water
Shallow Water
Clear Water

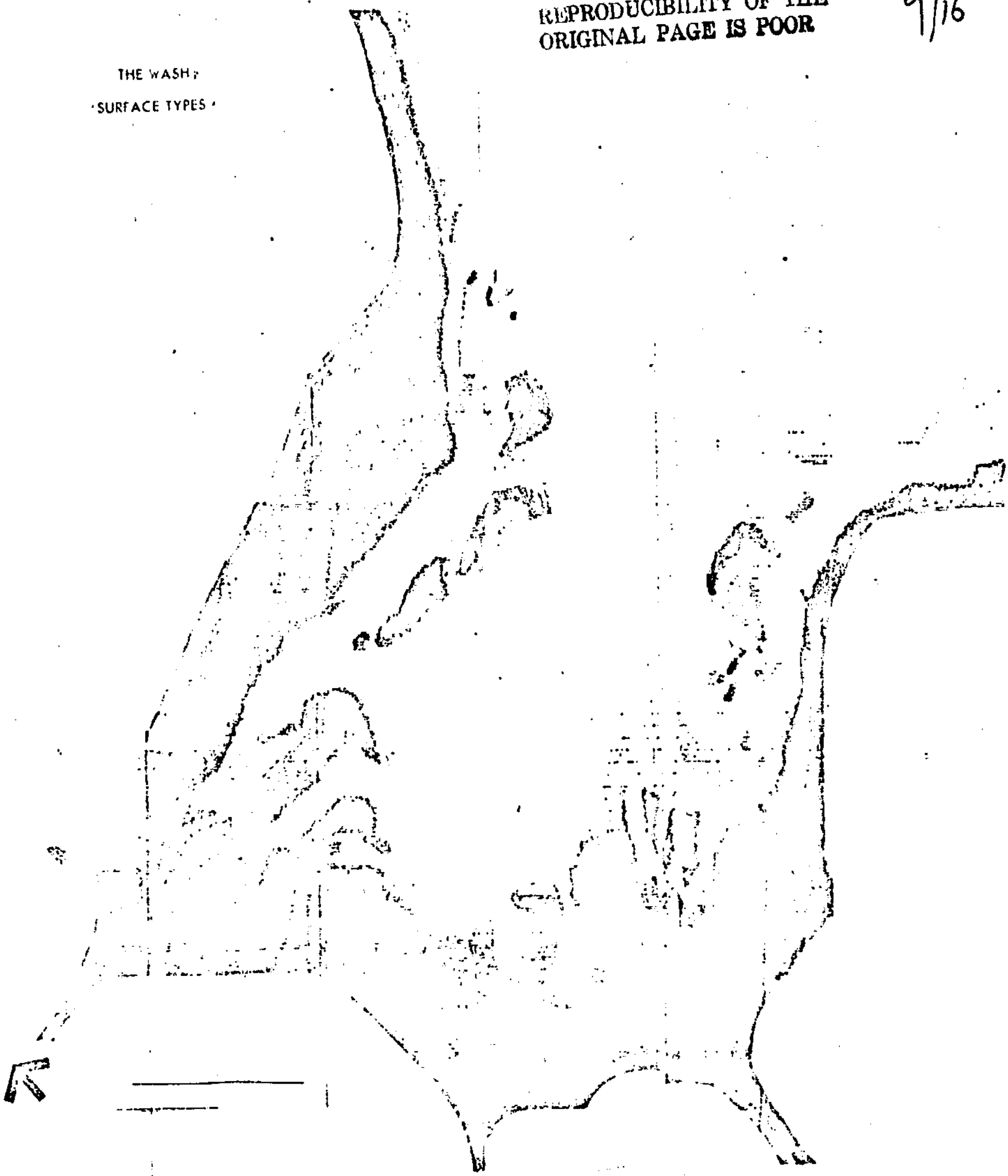

Band 7
less than 3 ?
No
Band 7
less than 12?
No
Yes
Band 4
less than 31 ?
Yes
No
Land
Wet Sand
Dry Sand
Sh


Band 7
less than 3 ?
Yes
Band 5
less than 17 ?
No
Yes
Band 5
less than 15?
No
Yes
nd
Shallow Water
Turbid Water
Clear Water

9/15

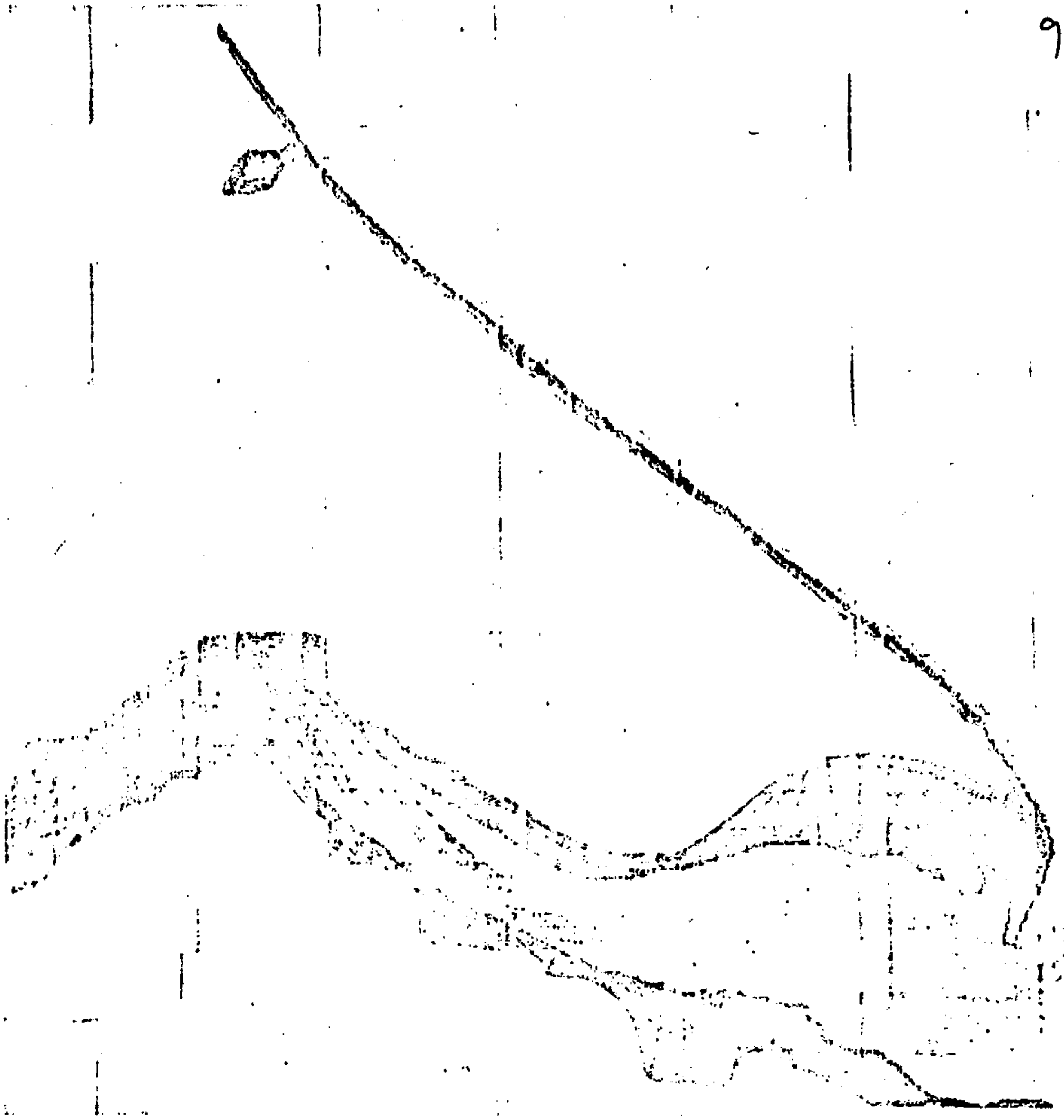
58

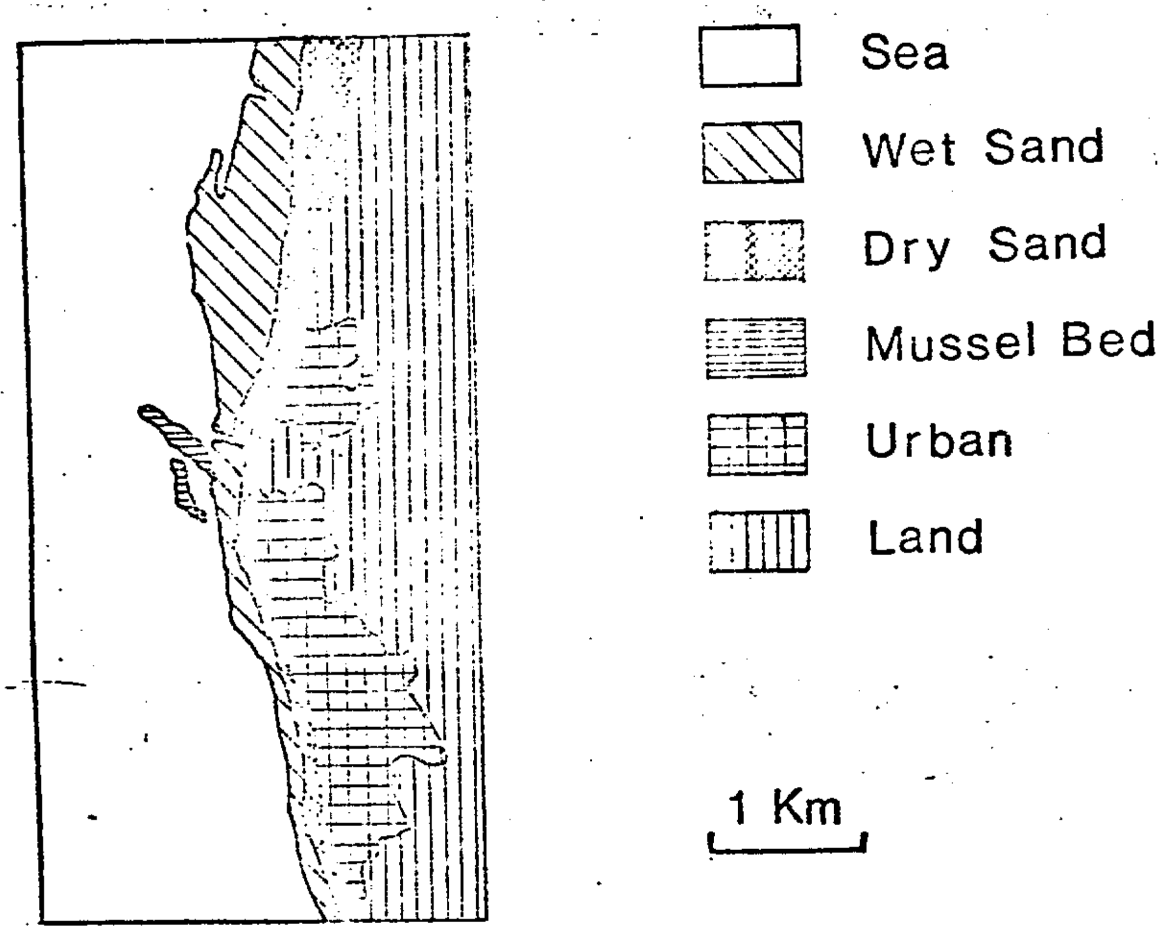
9/16



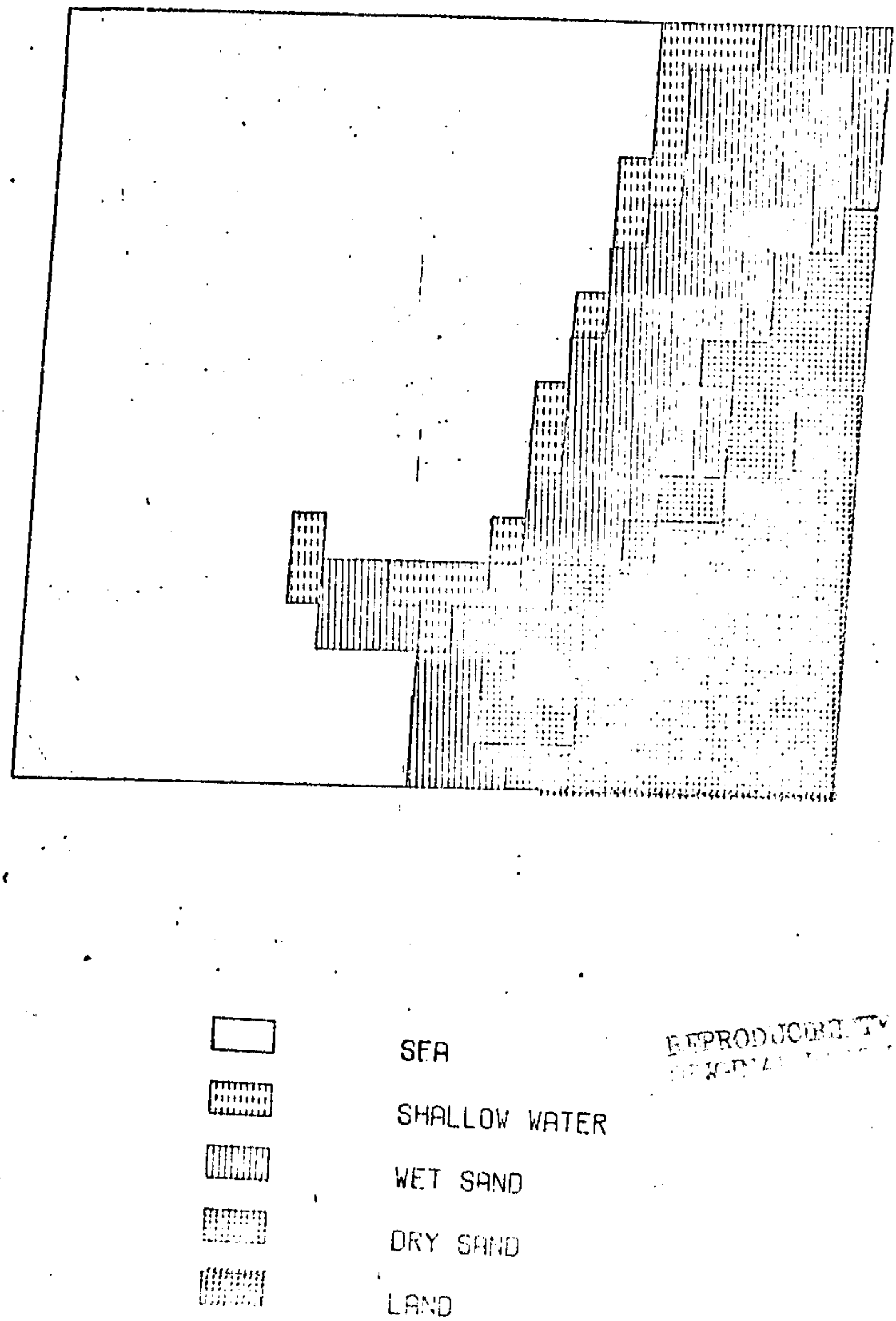
58-97

9/17




Sea
Wet Sand
Dry Sand
Mussel Bed
Urban
Land
1 Km

9/19



58-95